

MODELING FINANCIAL FRAUD DETECTION USING RADIAL BASIS FUNCTION NETWORK

Ph.D. THESIS

BY

**ALABI, Isiaq Oludare
(13/52HG003)**

SUPERVISOR:

DR. R. G. JIMOH

**DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF COMMUNICATION & INFORMATION SCIENCES
UNIVERSITY OF ILORIN, ILORIN, NIGERIA**

May, 2018

CERTIFICATION

This is to certify that this thesis has been read and approved as meeting part of the requirements of the Department of Computer Science, University of Ilorin, Ilorin, Nigeria for the award of Doctor of Philosophy degree.

Dr, R. G. Jimoh
(Supervisor)

Date

Dr. Tinuke O. Oladele
(Departmental Postgraduate Coordinator)

Date

Dr. D. R. Aremu
(Head of Department)

Date

Dr. XX
(Internal Examiner)

Date

Prof. XX
(External Examiner)

Date

DEDICATION

To Almighty Allah, the Creator of all things, who enabled me to accomplish this feat, my late parents, siblings and especially those God used to facilitate my sail through the programme.

DECLARATION

I, ALABI, Isiaq Oludare (13/52HG003), hereby declare that this thesis entitled MODELING FINANCIAL FRAUD DETECTION USING RADIAL BASIS FUNCTION NETWORK is a record of my research work. It has neither been presented nor accepted in any previous application for a higher degree. All sources of information have been specifically acknowledged.

In addition, the research work has been ethically approved by the University Ethical Review Committee.

Alabi, I. O.

Date

ACKNOWLEDGEMENTS

Special thanks to Almighty God, and those who have contributed in one form or the other to the successful completion of this study, notably my Supervisor, Dr. R. G. Jimoh, for his unrelenting pieces of advice; all the academic staff of the Computer Science Department of the University, notably the Head of Department; members of staff of the Unilorin Postgraduate school and the Computer Science Department Postgraduate Coordinator. Profound gratitude also goes to my family members for their love, sacrifice, unflinching support and endurance. The contributions of my brothers: Engr. Ajao Ibiyeye and Mr. AbdulGaniy Alabi; my long time bosom friends: Alhaji AbdulGaniy Salman, Dr. Lukuman Aminu and Alh. Dauda Adeniyi, others are Dr. K. Rauf, and Dr. W. B. Yahya for their moral support and kind encouragements. Also, I gratefully acknowledge the insightful suggestions and helpful comments of those who reviewed the thesis and made useful contributions as well as many more people too numerous to mention.

ABSTRACT

Fraudulent activities with intent to defraud are pervasive in every system. The traditional statistical and long investigative auditors' methods of fraud detection are becoming obsolete in dealing with the present trend of financial malpractices. The existing fraud detection methods are inadequate in terms of cost and time to detect emerging frauds. This study aimed at developing an enhanced Radial Basis Function (RBF) network to efficiently detect financial frauds. The specific objectives were to: (i) determine the input features relevant to fraud occurrences; (ii) create a series of RBF models; (iii) select an optimum model amongst the candidate models; (iv) evaluate the chosen model; and (v) compare the model with some existing ones.

The methodology employed RBF algorithm with radially symmetric Gaussian activation functions to classify previously unseen data features into their respective true categorical classes. A set of 20 input attributes, minimum hidden nodes and weight adjustments were observed for different runs until the network became divergent. The model was implemented in *R* vectorized software packages and was trained 200 times with 1000 online German bank credit transactions. The developed RBF model for fraud detection was compared with three other models: Multi-layer perceptron back-propagation (MLP), Dynamic decay adjustment (DDA) and General nonlinear regression (GNLR).

The findings of the study were that:

- i. twenty input features were ranked in order of their importance relative to the network's output; 4 of the variables, namely, credit history, nationality, guarantor and job status were the most important in that order, while 5 variables: accounts type, telephone number, co-applicant's guarantor and savings bond are the variables with negative contribution;
- ii. a total of 200 RBF base models were generated with randomly selected 850 credit transactions out of 1000 at ten different epochs and hidden nodes;
- iii. the RBF model for fraud detection was developed and of the 200 candidate base models, an optimum model was obtained at 600 iterations alongside 840 hidden nodes with a misclassification error rate of 6.9% and about 93% degree of accuracy. RBF models are of the form:

$$y(x) = \sum_{i=1}^h \omega_{ij} \phi_i(x),$$
 where ω_{ij} is output weight, and $\phi_i(x)$ is the Gaussian activation function, $\frac{1}{1+e^{-\phi_j(x)}}$, with output $y \in \{0,1\}$;

- iv. the chosen RBF model was evaluated and its performance yielded 7.18% misclassification error rate, 92.82% prediction accuracy, 89.71% sensitivity and receiver operating characteristic (ROC) of 98.6 %; and
- v. RBF had the highest average accuracy of 75.90% compared to DDA (75.04%), MLP (74.60%) and GNLR (74.53%). Further, RBF and DDA attained the highest accuracy

score of (85.71%) and (79.22%) respectively, at iteration 4, while MLP and GNLR attained their highest accuracy at iterations 7 and 5 respectively.

The study concluded that RBF network compared to other similar models trained faster, minimized the cost and time of fraud detection. The study recommended that RBF network is an efficient fraud detection model and its parameters must be adjusted closely to zero tolerance as a little threshold tolerance could imply a significant cost.

TABLE OF CONTENTS

CERTIFICATION	ii
DEDICATION	iii
DECLARATION	iv
ACKNOWLEDGEMENTS	v
ABSTRACT	.vi
LIST OF TABLES	.11
LIST OF FIGURES	.12
ABBREVIATIONS AND NOTATIONS	13
CHAPTER ONE: INTRODUCTION	.15
1.1 Background to the Study	15
1.2 Statement of the Problem	18
1.3 Aim and Objectives	18
1.4 Significance of the Study	19
1.5 Scope and Limitation	19
1.6 Operational Definition of Terms	20
1.7 Chapters' Layout	21
CHAPTER TWO: LITERATURE REVIEW	.22
2.1 Fraud	22
2.2 Types of Fraud	23
2.3 Reasons for Committing Frauds	25
2.4 Fraud Indicators	28
2.5 Fraud Perpetrators	30
2.6 Challenges in Fraud Detection	30
2.7 Fraud Detection Responsibility	31
2.8 Methods of Detecting/Exposing Frauds	31
2.9 The Essence of Data Mining in Fraud Detection	32
2.10 Modelling Fraud Detection	35
2.10.1 Network Topology Size	41
2.11 Model Selection for Fraud Detection	42
2.12 Fraud Classification with Ensemble Methods	45
2.13 RBF Model Robustness	46
2.14 Classifier Issues	46
2.15 RBF Network Benefits	47

2.16	RBF as a Binary Classifier	49
2.17	Summary	51
CHAPTER THREE: METHODOLOGY		52
3.1	The Research Framework	52
3.2	The Framework Design	54
3.2.1	The System Monitors	57
3.2.2	Data Attributes Selection and Transformation	57
3.2.3	Data Access, Analysis and Inductive Training	58
3.2.4	Enterprise Data Warehouse Configuration	59
3.2.5	Transactions Query Portal	59
3.3	Determining the Size of the RBF-ANN Topology	59
3.4	The RBF Network Overview	61
3.4.1	Input Selection	63
3.4.2	Parameters Estimation	64
3.5	Obtaining the Cluster Centres	65
3.6	Determining the Number of Hidden Nodes	66
3.7	Base Models Creation / Training	69
3.7.1	Model's Training and Termination	72
3.8	Models Aggregation and the Final Model Selection	72
3.9	Misclassification Error Rate	73
3.9.1	The Model's Performance Criteria	75
3.9.2	Misclassification Cost	76
3.9.3	Cost Matrix	77
CHAPTER FOUR: RESULTS AND DISCUSSION		80
4.1	Test Environment Requirement	80
4.2	Dataset Description	81
4.3	Measure of Input Variables on the Network Response	82
4.4	Input Normalization	87
4.5	RBF Network Requirement	87
4.6	Model Training	88
4.7	Parameters Initialization	89
4.7.1	Model Training Optimization	89
4.7.2	Training Model Outcome	94
4.8	The RBF Network Architecture	99

4.9	Results Overview	101
4.10	Comparison with Other Classifiers	102
CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS		107
5.1	Conclusion	107
5.2	Contributions to Knowledge	109
5.3	Recommendations	109
5.3.1	Action Plan: What to Do	110
REFERENCES		112
APPENDIX A		124
APPENDIX B		153
APPENDIX C		155

LIST OF TABLES

	PAGE
Table I: Abbreviations and Notations	13
Table 1.1: Operational Definition of Terms	20
Table 2.1: Summary of Fraud Profiles in Different Sectors	24
Table 2.2: Percentage of Firms Affected by Fraud in 2016	24
Table 2.3: Percentage of Fraud Perpetrators	29
Table 4.1: Input Variables Description	81
Table 4.2: Input Variables Importance using Garson Algorithm	84
Table 4.3: Key Input Variables Based on Ranked Variable importance Values	86
Table 4.4: Table of Average Misclassification (Trained) Error Rate (MER)	90
Table 4.5: Table of Average Misclassification (Test) Error Rate	92
Table 4.6: Estimates of Various Performance Indices	96
Table 4.7: Comparison of Neural Network Techniques Accuracies (Training)	102
Table 4.8: Comparison of Neural Network Techniques Accuracies (Test)	104

LIST OF FIGURES	PAGE
Figure 2.1: A Fraud Triangle	25
Figure 2.2: Fraud Risk Factors	27
Figure. 2.3: A Basic RBF Neural Network Schematic Architecture	47
Figure 3.1: Data Mining Process Flow	52
Figure 3.2: The Flow of RBF-ANN Fraud Detection Framework	55
Figure 3.3 A RBF Network with n Inputs, h Hidden Units	59
Figure 3.4: A 2-class Misclassification Cost Matrix	77
Figure 4.1: A 1-layer RBF Network	83
Figure 4.2: Overall Attributes Performance Chart	85
Figure 4.3: Confusion Matrix	95
Figure 4.4: Plots of Radial Basis Function Trained	97
Figure 4.5: RBF Network Parameters and Architecture Summary	99
Figure 4.6: Comparison of Neural Network Techniques (Training)	103
Figure 4.7: Comparison of Neural Network Techniques (Test)	105

ABBREVIATIONS AND NOTATIONS

Table I: Abbreviations and Notations

Abbr. / Notations	Full Meaning
ACFE	Association of Certified Fraud Examiners
AdaBoost	Adaptive Boost
AMD	Advanced Micro Devices Processor
AML	Anti-Money Laundering
ANN	Artificial Neural Network
AUC	Area Under Curve
C4.5	A Variant of Decision Tree for Continuous Variables
CART	Classification and Regression Tree
CCTV	Closed-Circuit Television
COSO	Committee on Sponsoring Organizations of the Treadway Commission
CPU	Central Processing Unit
CRAN	Comprehensive R Archive Network
DDA	Dynamic Decay Adjustment
DMR	Data Mining in R
EBS	Enron Broadband Services
eCommerce	Electronic Commerce
ETL	Extraction, Transformation Tool
FBI	Federal Bureau of Investigation
FDD	Financial Fraud Detection
FFR	Fraudulent Financial Reporting
FNN	Feed-forward Neural Network
FPR	False Positive Rate
FRCN	Financial Reporting Council of Nigeria
GHz	Giga-Hertz
GNLR	Generalized Non-linear Regression
GNM	Generalised Nonlinear Model
IB1	Instance-based (Learning) version 1
ID3	Iterative Dichotomiser-3
IFAC	International Federation of Accountants
INTERPOL	International Police
ISA	International Standards on Auditing
IT	Information Technology
KDD	Knowledge Discovery in Databases
KPMG	Klynveld Peat Marwick Goerdeler, A Professional Service Company
LDA	Linear Discriminant Analysis
LR	Linear Regression
MER	Misclassification Error Rate
ML	Machine Learning
MLP	Multi-layer Perceptron
MMM	Mavrodi Mondial Movement
MSE	Mean Square Error

NASB	Nigerian Accounting Standards Board
NASDAQ	National Association of Securities Dealers Automated Quotations Exchange
OLAP	Online Analytical Processing
PDE	Partial Differential Equation
POS	Point-of-Sale
PPV	Positive Predictive Value
PR	Public Relations
RAID	Redundant Array of Inexpensive Disks
RAM	Random Access Memory
RBF	Radial Basis Function
ROC	Receiver Operating Characteristics
RSNNS	Neural Networks in <i>R</i> using the Stuttgart Neural Network Simulator
SEC	Securities and Exchange Commission
SMP	Symmetric Multi-processing
SOP	Standard Operating Procedure
SPSS	Statistical Package for the Social Sciences
SSE	Summed Squared Error
SVM	Support Vector Machine
TCP/IP	Transmission Control Protocol/Internet Protocol
TPR	True Positive Rate
UCI	University of California Irvine
VIP	Very Important Person

CHAPTER ONE

INTRODUCTION

This chapter provides a background to the study, giving the rationale for the use of Radial Basis Functions (RBF) model. An outline of the problem statement was made, stating the aim, objectives as well as the significance, scope and limitations of the study.

1.1 Background to the Study

Financial fraud is an intentional act to unduly obtain something of value under false pretence, for the fraudster's benefit. Usually, fraud involves one or more persons who intentionally act under false pretence to deprive another of something of value, for their own benefit. Fraud is as old as humanity itself and could take an unlimited variety of forms (Gupta & Gill, 2013).

Legends of financial scam are pervasive and are adversely affecting economies worldwide of which many people have been deprived of substantial amount of valuables. In recent years, the development of new technologies has also provided further ways in which criminals may commit fraud (Bolton & Hand, 2002).

Fraud being criminal in nature, it is an undesirable act, even in statutory laws. Nonetheless, fraudulent activities are assuming sophisticated dimensions, whereby individuals, governments, agencies and businesses are deprived of substantial material benefits. Meanwhile, detecting and preventing fraud is not a simple task (Coderre, 2009). Much research efforts and other resources have been deployed to prevent fraud, yet the syndrome persists (Nigrini, 2011). Then, a recurring question is how can this behaviour be eliminated or nipped before it materializes?

Fraud detection and prevention are concurrent processes in combating fraud malaise; while fraud detection is the spotting of false claim, act or data; fraud prevention is the bursting of the crime before it materializes by raising alarms, thus preventing it from occurring (KPMG,

2016). Fraud is an adaptive crime, so it needs special methods of intelligence gathering and data analysis in order to detect and prevent frauds (Nigrini, 2011). Indeed, fraud detection comes to fore once fraud prevention has failed; hence, fraud detection must be routinely active in a financial transactions system.

Data mining methods are assisting auditors or crime investigators in prevention and detection of frauds because data mining uses supervised or unsupervised training data (past cases) of fraud and/or normal transactions to build models (test algorithms) to identify and detect the risk of fraud. Also, new techniques are being designed for preventing fraudulent financial reporting (Gupta & Gill, 2013). Likewise, data mining is a methodology for the extraction of knowledge from data (Yue, Wu, Wang, Li, & Chu, 2007).

Researchers and business professionals have acknowledged that data mining is playing a key role in fraud detection due to its capability to extract knowledge from huge data heaps, a property which makes its use indispensable for auditors and crime investigators to track fraudulent practices (Yue et al., 2007). Similarly, past cases of fraud can be used to build models to identify and detect the risk of fraud as new techniques are being designed for preventing fraudulent financial reporting (Gupta & Gill, 2013).

Previous researchers have proposed a number of fraud detection techniques, yet some problems persist, for instance, most of the existing fraud detection systems do not timely alert when the fraud is committed, until some later time when it is almost too late to track culprits, perhaps due to their computational complexities or other deficiencies (Khac & Kehadi, 2010). In some situations where a fraud detection system alerts, it might be too rigid to keep pace with the current fraud trends, whereas fraud detection models must be dynamic to encompass emerging and future fraud trends (Brause, Langsdorf, & Hepp, 1999).

Therefore, the researcher developed a model with possibility of handling larger datasets in detecting fraudulent activities (such as money laundering, electronic commerce scam, dubious insurance, mortgage and health claims, etc.) using RBF network classification approach. RBFs are classical functions that can be employed to approximate models (linear or non-linear) of a neural (single) network, RBF methods approximate multivariate functions with its remarkable properties that render it highly efficient, adjustable smoothness as well as its powerful convergence properties (Buhmann, 2004).

Though, the rationale of ensemble modelling was to combine multiple inductive models for the same domain, it was intended to obtain better prediction quality, thereby reinforcing strengths and compensating weaknesses (Cichosz, 2015) RBF is a variant of ANN, where each hidden unit executes a radial activated function and the output unit executes a weighted sum of hidden unit outputs.

Early application of RBF was in fire detection, where several measured parameters (the flame colour, spectrum, intensity, direction, etc.) were used to model a fire detector device (Buhmann, 2004). In nuclear physics, RBF was used to interpolate the data that come from the raster of a screen of a robot's eye (Eckhorn, 1999; Kremper, Schanze & Eckhorn, 2002). Other applications include RBF as interpolation (Broomhead & Lowe, 1988; Matej & Lewitt, 1996), time-series modelling, classification and control engineering (Sanner & Slotine, 1996).

Advantages of RBF over other approximating algorithms (e.g. partial differential equations, [PDEs], regression, etc.) are its convergence power and its easier formulation of interpolants; more remarkably resilience against irregular data distributions. Above all, RBFs have a variation-diminishing property to guarantee approximations smoothness (Buhmann, 2003). All these strengths of RBF inform its choice as the base model in this study.

1.2 Statement of the Problem

The need for internet-based transactions and the associated security issues necessitated rigorous transactions verification to determine their veracity (Cai, LeKhac, & Kechadi, 2012). It is worthy of note that, fraud is endemic and pervasive in virtually every system with increasing sophistication as new technologies emerge (Bolton & Hand, 2002). Since operating on internet inevitable, it then becomes necessary to devise a means of ensuring secured financial transactions.

Most of the existing fraud detection systems use traditional statistical and/or long investigative auditors' methods to detect fraud which takes a longer time to track offenders, perhaps due to their computational complexities or other deficiencies (Ngai, Hu, Wong, Cheng & Sun, 2011; Khac & Kechadi, 2010; Phua, Lee, Smith, & Gayler, 2005; and Krambia-Kapardis, Christodoulou, & Agathocleous, 2010). Some others require the specification of confidence thresholds, thereby tolerating some level of frauds, which is undesirable where zero-tolerance losses is desirable (Hand, Mammila, & Smyth, 2001). Others have high dimensionality, i.e. a larger number of measured attributes that yields more computations that could degrade accuracy especially with irrelevant attributes (Kohavi & John, 1997).

Finally, some fraud detection methods are not dynamic to encompass emerging and future fraud trends (Brause, Langsdorf, & Hepp, 1999). The proposed RBF network model is an improvement over the existing models by providing faster response with large datasets.

1.3 Aim and Objectives

The aim of this research was to develop a radial basis function neural network model to detect fraud in financial transactions.

The specific objectives were to:

- i. determine the basic factors of input features relevance to fraud occurrences;
- ii. create a series of RBF classification models for fraud detection using a published bank credit dataset;

- iii. select an optimum RBF model using a set of metrics;
- iv. evaluate the appropriateness of the chosen RBF model; and
- v. compare the RBF model with some existing methods.

1.4 Significance of the Study

The significance of this study lies in the fact that propensity of fraudulent activities are pervasive in virtually every system. The dynamism of fraudulent acts called for learning and intelligent systems (or models) as proposed in this study. Data mining techniques are becoming more proactive rather than reactive measures of fraud detection. These techniques have found relevance among business professionals, policy makers, as well as in substantive applications such as decision support systems and operational business applications among others, thereby minimizing the time and cost of fraud detection.

Adopting computationally efficient models and frameworks rather than mere reliance on primitive intuition, manual long investigations and/or windy auditors' findings. Above all, the study would contribute to early fraud detection and prevention which is of immense significance to national financial stability and international security.

1.5 Scope and Limitation

This study pays attention to outlier detection, since financial transactions can be classified into normal or abnormal transactions. Hence the study becomes generic with similar situations that have categorical characteristics but with distinct parameters, e.g. detection of a certain diseases in patients. For instance, the same approach and underlying models used to develop a fraud-detection capability for a bank can be used to develop medical insurance fraud detection applications. The difference is how the models are parameterized.

1.6 Operational Definition of Terms

The relevant operational terms and notations used in this study are defined in Table 1.1.

Table 1.1: Operational Definition of Terms

Terms	Definition
Artificial Neural Network (ANN)	A set of statistical learning algorithms based on human central nervous system, which are used to estimate large number of unknown inputs, or simply parallel distributed processors.
Classifier	A model, scheme or algorithm for outlier detection or data classification.
Data Mining	Methodology for the extraction of knowledge from large dataset.
Dashboard	A query portal.
Data Warehouse	A repository of heterogeneous data from a variety of sources, organized under a unified schema at a single location to facilitate data mining.
Epoch	Simulation iteration or run.
Evaluation Matrix	This is intended at examining classification performance of the model based on vector of predictor variables.
Financial Fraud Detection (FFD)	Detecting frauds involving quantifiable monetary values. Financial fraud and fraud are used interchangeably in this study.
Fraud	To intentionally act under false pretence to deprive another of something of value, for his or her own benefit.
Fraudulent Financial Reporting (FFR)	Involves intentional misstatements, including omissions of amounts or non-disclosures in financial statements, to deceive financial statement users.
IB1	A Nearest-neighbour classifier that uses normalized Euclidean distance to find the training instance closest to the given test instance, and predicts the same class as this training instance.
Inductive bias	The selection criteria used by the learning RBF-ANN network to select the appropriate model after models creation.
Logit Model	Otherwise called logistic (or logit) regression, a regression model where the dependent variable is categorical.
Online Analytical Processing (OLAP)	Analysis techniques with functionalities such as summarization, consolidation, and aggregation, as well as the ability to view information from different perspectives.

<i>R</i>	<i>R</i> vectorized statistical package; <i>R</i> code snippets.
Radial Basis Functions (RBF)	Measures to approximate multivariate functions.
The Model	Radial basis function-artificial neural network (RBF-ANN).
X ; <i>x</i>	Bold-capital letters were used for vector quantities; italicized-small letters were used for scalar values.

1.7 Chapters' Layout

This thesis is organised as follows: Chapter One outlines the study background, the aim and objectives of the study, scope as well as the significance of study. Chapter Two is the review of relevant and related literatures to the study, including highlights of the merits and gaps identified and the summary of the entire literatures reviewed. Chapter Three presents and discusses different classification methods and the methodological framework adopted for this research as well as the expected results. Chapter Four contains the dataset used for experimentation, algorithms and computational requirements for developing the RBF network as well as data analysis and discussion of the results. Finally, conclusion and recommendations are contained in Chapter Five.

CHAPTER TWO

LITERATURE REVIEW

This chapter presents the review of relevant literatures for financial fraud detection, forming the premise for this research. It highlights a variety of previous researchers' efforts, and perspectives in fraud prevention and detection; their views were underscored while noting their strengths and weaknesses.

It is noteworthy that if objects can be classified exactly by a Supervisor, then is it necessary to replace exact classification by some approximations —the thrust of this study. Perhaps, part of the reasons is the fact that the art of classification is wide and diverse across multi-discipline, aside from the fact that criminal activities are shrouded in secrecy thereby making them difficult to detect. However, the study seeks to proffer more probes into this assertion in later sections.

2.1 Fraud

Although, fraud does not have a specific conventional definition, Wang, Liao, Tsai and Hung (2006) said fraud is “a deliberate act contrary to law or policy with intent to obtain deserved benefit.” (p. 1122). Fraud was (and still is) a very serious, underrated, and somewhat ignored threat to private and public entities for most of the nineteenth century (Silverstone & Davia, 2005). As a consequence, fraud thrived relatively undeterred and has cost many entities enormous sums of money and valuables. As much as there is a growing concern to combat fraud more aggressively, given that, it has always been human nature to wait until an incident happens before doing something about the issue, even in businesses and government milieu. Many business owners will not contemplate paying for proactive fraud services, but might end up to pay a lot more when fraud is perpetrated through the cost of hiring lawyers, accountants, support staff, etc. (Coderre, 2009).

Hitherto, traditional methods of data analysis had been used to detect fraud, requiring complex and time-consuming investigations involving different domains of knowledge like finance, economics, business practices and law (Mena, 2003). However, data mining provides functionality not provided by general database query techniques that cannot deal with the large number of records with high dimensional structures (Olson & Delen, 2008).

Khac and Kechadi (2010) also asserted that most financial fraud solutions are premised on sets of predefined rules and thresholds, perhaps based on statistical means and standard deviations, these are hardly enough to trap recent sophisticated fraud means.

Fraud instances can be similar in content and appearance but usually are not identical (Palshikar, 2002). Researchers have observed that a typical organization loses up to 5% of its revenue to fraud each year (Gupta & Gill, 2012), even though, about 60 per cent of fraudulent activities were with insiders' accomplice (COSO, 1999). Cases of frauds are quite alarming. Amazingly most criminals walk the streets freely in most cities unperturbed, they depraved more victims because their identities are either shielded or difficult to verify (Silverstone & Davia, 2005).

2.2 Types of Fraud

Professional bodies such as International Standards on Auditing (ISA) broadly classified frauds into two: *Management fraud* (i.e. misstatements resulting from fraudulent financial reporting), and *Employee fraud* (i.e. misstatements resulting from misappropriation of assets) (DeZoort & Lee, 1998). Frauds appear in a variety of forms. Therefore the approach to fraud detection must be dynamic (Coderre, 2009). Researchers often prescribe a specific model for peculiar fraud type, however, the ability to detect different types of fraud, rates a fraud detection model highly.

Specifically, fraud examples include: Money laundering, Credit card scam, Pension fund diversion, False health insurance claims, etc. The list is endless. Table 2.1 shows a variety of fraud types in some sectors of a nation's economy and the level of anti-fraud investment. The table shows that the level of commitment to fraud detection in some sectors was diminishing. Also, see Table 2.2 indicating the level of organisations that have fallen prey to different fraud types recently according to Kroll's (2016) report.

Surprisingly, there were instances when fraud had been indicated, but the will to eradicate or prevent further frauds was not there!

Table 2.1: Summary of Fraud Profiles in Different Sectors (Source: Ahmed, Mahmood, & Islam, 2015)

Sector	Level of Financial Loss	Investment on Anti-fraud
Natural Resource	Medium	High
Medical, Healthcare	High	High
Retail, Wholesale	Medium	Low
Financial Services	High	Medium
Manufacturing	High	Low
Information Technology	Medium	Low

Table 2.2: Percentage of Firms Affected by Fraud in 2016 (Source: Kroll, 2016)

Fraud Types	Percentage of Firms Affected in 2016
Theft of Physical Assets	22%
Vendor, Supplier or Procurement Fraud	17%
Information Theft	15%
Management Conflict of Interest	12%
Regulatory or Compliance Breach	12%
Corruption and Bribery	11%
Internal Financial Fraud	9%
Funds Misappropriation	7%
Money Laundering	4%
IP Theft	4%
Market Confusion	2%

2.3 Reasons for Committing Frauds

Fraud includes a wide variety of acts characterized by the intent to deceive or to obtain an unearned benefit. Fraud is not something that will go away on its own — it must be discovered and stopped or it will continue to grow (Coderre, 2009). Legally, fraud can lead to a variety of criminal charges, including theft, embezzlement, and larceny. Each charge has its own specific legal definition and required criteria, and all of the charges can result in severe penalties and a criminal record (Coderre, 2009). Coderre (2009) proposed a Fraud Triangle to explain major fraud motivations as captioned in Figure 2.1. An effective fraud risk identification process should include an assessment of the incentives, opportunities and rationales to commit fraud

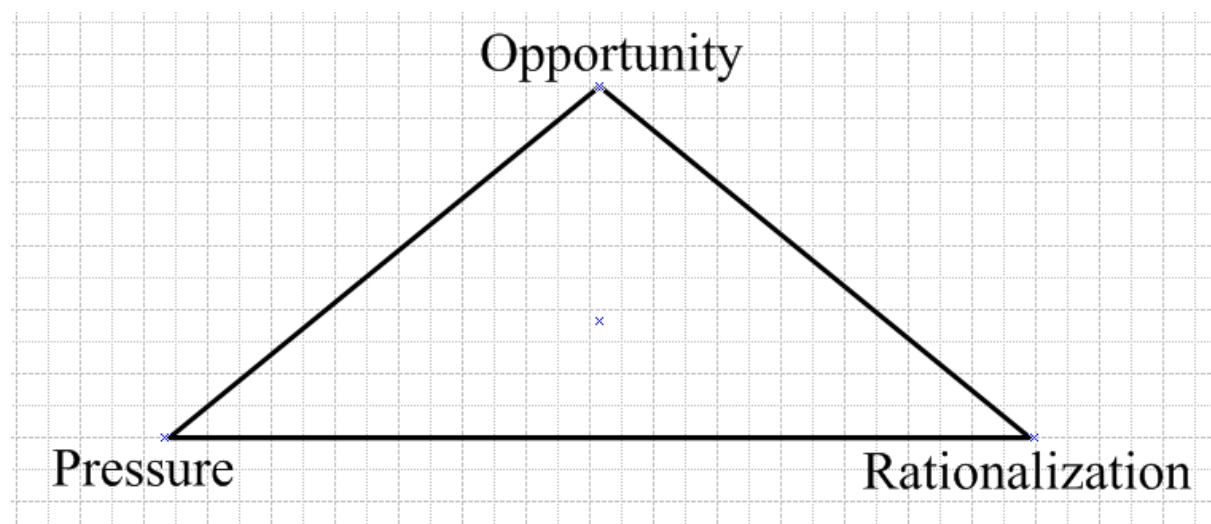


Figure 2.1: A Fraud Triangle (Source: Coderre, 2009)

The author also explained that the *opportunity* exists when there are weak controls and/or when an individual is in a position of trust. While the *pressures* on those who commit fraud are often financial in nature, unrealistic corporate targets may also influence a person to commit fraud to meet the targets. The *rationalization* for fraud often includes some mind-sets such as: thinking that the act is not criminal, justifying the act with words like “some others are doing it”, so it is a norm. Phua *et al*, (2005) stated that external fraudster has three basic profiles: average offender, criminal offender, and organised crime offender. Average offenders

display random and/or occasional dishonest behaviour when there is opportunity, sudden temptation, or when suffering from financial hardship. In contrast, the more risky external fraudsters are individual criminal offenders and organised/group crime offenders (professional/career fraudsters) because they repeatedly disguise their true identities and/or evolve their modus operandi over time to approximate legal forms and to counter detection systems. An effective fraud risk identification process should include an assessment of the incentives, opportunities and rationales to commit fraud (Interpol, 2017).

However, development of new technologies is opening up new vistas for innovative ways in which criminals may commit fraud (Bolton & Hand, 2002). Notwithstanding, fraud detection is more of monitoring the behaviour of populations of users of a system in order to estimate, detect, or deter undesirable behaviour (Phua, Lee, Smith, & Gayler, 2005). Silverstone and Davia (2005) opine that it can be argued that fraudulent practices can never be eliminated, but the onus is on the management to create the most effective system possible to discourage it.

Economically, there have been significant financial losses due to fraud in online businesses recently, which increased from US\$5.2 billion in 2008 to US\$8.6 billion in 2009 (Louzada & Ara, 2012). An apposite illustration is the Ponzi scheme perpetuated by former NASDAQ chairman Bernard Madoff, which led to the loss of approximately US\$50 billion worldwide (FBI, 2008), and in Nigeria, Mavrodi Mondial Movement (MMM), a global financial network founded by Andrew Marc despite several official warnings through the Central Bank of Nigeria and the Securities and Exchange Commission scammed a sumptuous amount of money in the illicit scheme. Another example is that of Joseph Hirko, former co-chief executive officer of Enron Broadband Services (EBS), who has avowed to forfeit approximately US \$8.7 million in restitution to Enron victims through the United States of

America's Securities and Exchange Commission's Enron Fair Fund after pleading guilty to wire fraud (FBI, 2008), The overall annual losses through financial fraud are enormous.

Therefore, it is essential that prevention technologies and fraud detection methods are developed and updated continuously, in order to identify fraud cases as soon as it is perpetrated (Bolton & Hand, 2002).

Figure 2.2 depicts classical financial fraud types across different sectors of global economies such as banks, insurance mortgages, securities and commodities, etc., as well as some of the factors or motivations for fraud ranging from eCommerce activities, frequent staff turnover, collaboration among businesses, outsourcing or offshoring of employees, entrance into a new riskier market or new businesses outrightly. These are some of the elements, among others on which fraud threats could emanate.

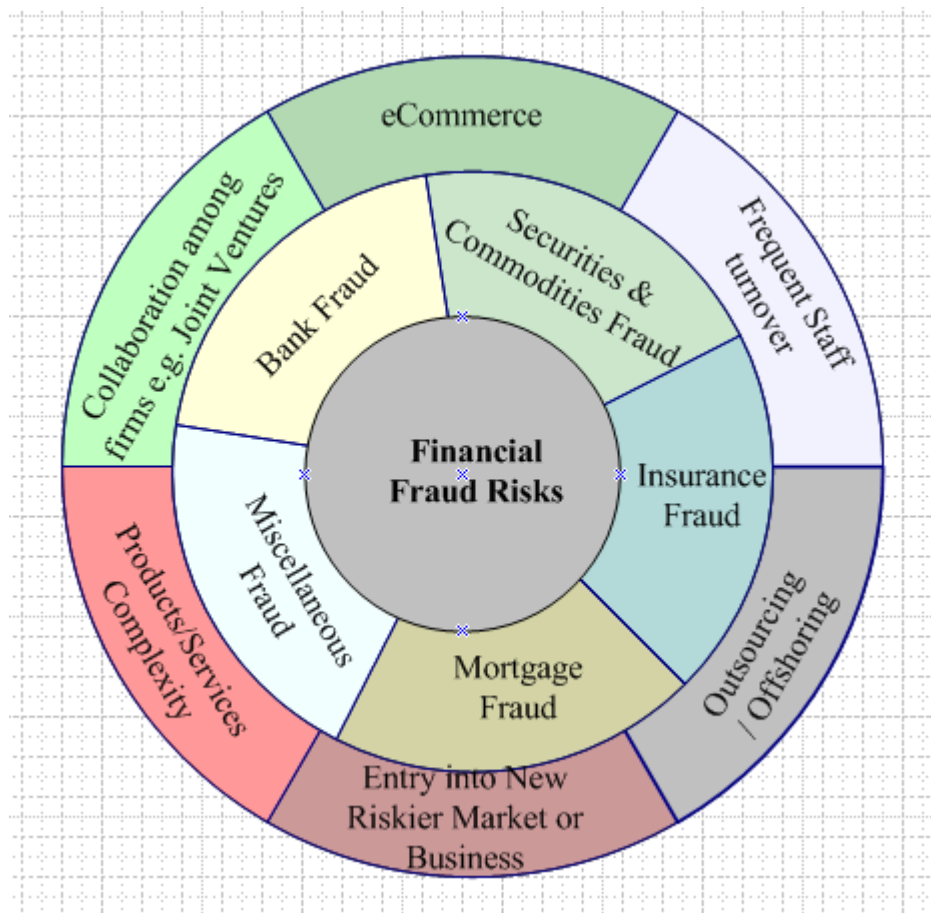


Figure 2.2: Fraud Risk Factors

2.4 Fraud Indicators

A red flag is an indicator (or symptom) of potential problems with undesirable characteristic that stands out to an analyst. Red flags differ depending on meanings ascribed to it. In financial services sector, Red flags are indicators to signal reasonable suspicious activities, such as money laundering, terrorist financing, *et cetera*, for further scrutiny.

Gullkvist and Jokipii (2012) posit that the perceptions and assessment of materiality when judging fraud risk factors (red flags) differ among financial professionals. In the business world for instance, if inventory of a business entity is on the rise, without corresponding increase in its fortunes. It may mean that its products/services are not selling. Also, if a Red flag is flared when a large accounts receivable figure is reported, it could mean that huge

profits has been made by the firm, if and only if it can be collected. Rather, a Red flag here should indicate that collections process needs to be strictly enforced in line with the credit policies. Largely, professionals exercise materiality judgement in evaluating the importance of the Red flags.

In the past, researchers and financial professional bodies, such as Association of Certified Fraud Examiners (ACFE), International Standards on Auditing (ISA), International Federation of Accountants (IFAC), Financial Reporting Council of Nigeria (FRCN)— formerly Nigerian Accounting Standards Board (NASB), often identify (new) and revise (old) fraud risk indicators and taxonomies. These Red flags have been raising auditors' sensitivity to fraud occurrence (Krambia-Kapardis et al., 2010).

To eradicate fraud, financial professionals and regulators often prescribe the use of fraud risk indicators or *Red flags* concept as they can be monitored (Pincus, 1989). Various researchers have worked on the *Red flags* concept (Apostolou, Hassell, Webber & Sumners, 2001; Hackenbrack, 1993, Loebbecke, Eining & Willingham, 1989; Majid, Gul, & Tsui, 2001; Mock & Turner, 2005; Moyes, 2007; Smith , Omar, Idris & Baharuddin, 2005). Their results indicated that fraud risk indicators are crucial in fraud detection.

One could isolate the factors that indicate a fraud risk or a high probability of fraud, then develop rules (or controls) and use them to flag only those claims or requests susceptible to be fraudulent. Data mining techniques are replete with various techniques of doing such. By this, fraud investigators can identify the symptoms of fraud before large losses occur. Continual routines that monitor key symptoms and track fraud risk trends can also be a major deterrent, thereby preventing or tracking fraud almost as soon as it occurs.

2.5 Fraud Perpetrators

Kroll (2016) findings revealed that the biggest fraud threat to organisations comes from within, where four in five, representing 81% of fraud perpetrators, had the hands of insiders in the crime. Others came from senior or middle management employees, while a negligible percentage was committed by junior employees or intermediaries. See Table 2.3.

Table 2.3: Percentage of Fraud Perpetrators (Source: Kroll, 2016)

GROUP	PERCENTAGE OF FIRMS WHERE SOMEONE IN THIS GROUP WAS INVOLVED
Junior Employees	45%
Vendors/Suppliers	18%
Agents or Intermediaries	23%
Senior or Middle Management	36%
Joint Venture Partners	8%
Regulators	7%
Customers	5%
Government Officials	3%
Others	3%

2.6 Challenges in Fraud Detection

According to Kroll (2016), the following encumbrances were identified in physical nabbing of fraudsters:

1. **Less visibility (Opacity):** Fraudsters often take cover under corporate veil by not divulging their true identity.
2. **Audit limitations:** Quite often, many organisations rely on internal or third-party audits to root out fraud and/or financial misappropriation or anomalies.
3. **Globalization:** Globalization of businesses increases fraud risk, for instance, in a global marketplace where many international businesses have thousands of companies in their supply chain, risks become more difficult to identify and keep under control.
4. **Vague details:** Transactions are described in vague terms such as: “strategic sales advice”, or “VIP or confidential” or “Public Relations (PR)”.

2.7 Fraud Detection Responsibility

As stated earlier, the development of new technologies is opening up new vistas for innovative ways in which criminals may commit fraud (Bolton & Hand, 2002). Fraud detection is more of monitoring the behaviour of populations of systems users in order to estimate, detect, or deter undesirable behaviour (Phua et al, 2005).

Management “owns” the systems, people, and records that constitute the controls (Mena, 2003). Therefore, managers should have a complete picture, i.e. knowledge of the business risks and controls, including the authority to adjust business operations. Hence, management rather than the auditors, should be responsible for the detection of fraud.

2.8 Methods of Detecting/Exposing Frauds

Kroll (2016) study shows that whistleblowers were responsible for exposing at least 41% of fraud cases that were detected aside from internal and external sources of detection such as:

- i. Internal controls
- ii. Auditing
- iii. Cyber and security information
- iv. Compliance procedure and ethics
- v. External sources e.g. vetting suppliers, vendors and other third party clients

Although, it can be argued that fraud can never be eliminated, the onus is on management to create the most effective system possible to prevent it (Silverstone & Davia, 2005). As part of the new emphasis on internal controls and corporate security in helping management prevent and detect fraud, financial fraud detection techniques come in. A corporate fraud policy

formally sets out what an employee is expected to do when he or she suspects that fraud is occurring or something unusual has occurred. Fraud detection procedure also increases employees' awareness, the gravity of fraud and management's refusal to tolerate such acts.

2.9 The Essence of Data Mining in Fraud Detection

Data mining is a methodology for the extraction of knowledge from large data (Spiliopoulou, 2011). Data mining is a relatively new discipline lying at the intersection of various disciplines such as statistics, machine learning, databases, pattern recognition, artificial intelligence, etc. It is capable of transforming research and industrial fields (Mattison, 1999). KPMG (2016) surveys show that technology as proactive analytics was found to be a significant enabler for the fraudsters investigated in its report. Data mining attempts to formulate, analyze and implement basic induction processes that facilitate the extraction of meaningful information and knowledge from unstructured data by filtering irrelevant information, and extracting meaningful knowledge. It is a tool that can sift through tonnes of transactions in order to find suspicious transactions.

Rather than relying merely on fraud examiners' intuition, data mining techniques are used to combine powerful analytical techniques with known business processes to turn warehoused data into the insight to mitigate the risk of fraud and abuse. Typically, organizations with local or branch offices allow the central office to store and mine data for the entire organization and grant access to local offices.

The fast-growing, tremendous amount of data, collected and stored in large and numerous data repositories, has far exceeded our human ability for comprehension without powerful tools. As a result, data collected in large data repositories (Data Warehouses) soon become "data tombs"—data archives that are seldom visited. Hence, data mining tools can turn such data

into “golden nuggets” of knowledge. The first industries to apply data analysis techniques to prevent fraud were the telephony companies, the insurance companies and the banks (Field & Hobson, 1997).

Hitherto, the traditional data analysis techniques such as regression analysis, cluster analysis, numerical taxonomy, multidimensional analysis, time series analysis, nonlinear estimation techniques, and other multivariate stochastic models have been applied to solving many practical problems, yet they are oriented toward the extraction of quantitative and statistical data elements (Sumathi & Sivanandam, 2006). Researchers have recently turned to ideas and methods developed in statistics, database technology, artificial intelligence, pattern recognition, machine learning, information theory, knowledge acquisition, information retrieval, high-performance computing, and data visualization. Lately, these techniques have also been applied to other emerging data types such as spatial data domains, text-based domains, and multimedia (image) domains.

Meanwhile, data mining differs from traditional statistics in several ways. For example, statistical inference is assumption-driven while data mining is discovery-driven in the sense that patterns and hypothesis are automatically extracted from data (Sumathi & Sivanandam, 2006). Machine learning (ML) is the backbone of the data mining concept, due to its capabilities to gain physical insight into a problem, and participate directly in data selection and model search steps.

Data mining is a process, involving data pre-processing (dimensionality reduction, cleaning, noise/outlier removal), post-processing (comprehensibility, summary and presentation), good understanding of problem domains and domain expertise. As such, data mining seeks to extract qualitative models, which can easily be translated into logical rules or visual

representations. Data mining is a step in Knowledge Discovery in Databases (KDD) process, a process that is interactive and semi-automated, starting with raw data and ending with insights, rules or a predictive model. Distinctively, the various techniques of data mining include classification (decision trees, naive Bayes classifier, k -nearest neighbour, neural networks), clustering (k -means, hierarchical clustering, density-based clustering), association (one-dimensional, multidimensional, multilevel association, constraint-based association).

Due to the dramatic increase of fraud cases which results in loss of substantial worth of valuables globally annually, several modern techniques in detecting fraud are continually developed and deployed in many business enterprises, where many service agencies have incorporated data mining into their investigating and auditing processes (SPSS, 2000). Data mining has become a strategic tool to prevent, detect and react to anomalies. As fraud threats are constantly changing, organisations need to conduct regular fraud risk assessments, modifying how to prevent and detect frauds (KPMG, 2017). Financial fraud detection (FFD) is vital to prevent fraud by distinguishing fraudulent financial data from authentic data, thereby alerting fraudulent behaviour or activities and enabling decision makers to develop appropriate strategies to decrease the impact of fraud (Ngai, et al, 2011).

According to Sherly and Nedunchezian (2010), fraud detection is important in today's computing environment. The increase in internet usage, electronic commerce and security vulnerability inherent in most systems make fraud detection a topic of interest for researchers, especially for governments, academia and corporate business enterprises.

As new data types are emerging, data mining techniques can be, and, have been, applied to other data types including spatial data domains, text-based domains, and multimedia (image)

domains. Data mining algorithms can look at numerous multidimensional data relationships concurrently, highlighting those that are dominant or exceptional. As a result, corporate enterprises can identify fraudulent activity by mining their existing data. This explains why some retail businesses use digitized closed-circuit television (CCTV) and point-of-sale (POS) terminals data to discourage fraudulent transactions (Palshikar, 2002). Similarly, many service agencies have incorporated data mining into their investigating and auditing processes (SPSS, 2000) since data mining combines powerful analytical techniques with known business knowledge to turn data already acquired into the insight to identify likely instances of fraud and abuse.

2.10 Modelling Fraud Detection

A model is a high-level global description of a dataset. The expressive power of a model increases as it represents more complex functions, i.e. as we increase the expressive power of a model, we can generally continue to get a better fit to the available data. If a model is over simplified, it might not be expressive of the data (Hand et al., 2001). Meanwhile, a model is built using machine learning by devising a set of attributes that are trained to produce a model that generalizes well on new data objects, then the model evaluation is carried out to determine how well a model is expected to perform on data objects beyond those in the training set, usually to assess its performance on a separate test set of unseen data.

As such, fraud detection is a typical classification issue; hence classification modelling tries to find models (or rules) that predict the values of one or more variables in a data set (target) from the values of other variables in the data set (inputs). After finding a good model, a data mining tool applies the model to new data sets that may or may not contain the variable(s) being predicted. When applying a model to new data, each record in the new data set receives a score based on the likelihood the record represents some target values. Supervised neural

networks such as fuzzy neural nets, and combinations of neural nets and rules, have been extensively explored and used for detecting fraud in mobile phone networks and financial statement fraud (Green & Choi, 1997; Estevez & Perez, 2006).

Literatures have revealed that statistical modelling techniques such as logistic (logit) regression, linear and quadratic discriminant analysis are widely used for modelling and prediction purposes, but their predetermined functional form and restrictive (often unfounded) model assumptions limit their usefulness (Wang et al., 2010).

The concept of logistic regression for analyzing a dataset in which there are one or more independent variables that determine a binary outcome (a dichotomous variable) is correlated with the study being investigated. This idea was used to estimate parameters for the RBF network as discussed in chapter 4.

To detect a novel type of fraud may require the use of supervised or an unsupervised machine learning algorithm. Notable studies among unsupervised learning in fraud detection include Peer Group Analysis and Break Point Analysis applied to spending behaviour in credit card accounts (Bolton & Hand, 2001). Peer Group Analysis monitors individual objects behaviour over time and detects those that tend to behave in a way different from objects to which they had previously been similar, while Break Point analysis operates on the account level by comparing changes in the recent and previous transactions information on a single account to detect anomalous features of the account. The down side of this method is that it focuses on local patterns rather than global models and the comparisons are made over time.

Khac & Kechadi (2010) presented a case study on anti-money laundering (AML) detection, where data mining and natural computing techniques were combined. Specifically, *k*-means clustering was used and the value of *k* was determined by AML experts in creating suspicious/unsuspicious groups, while resource-bounded outlier detection using clustering methods was suggested by Torgo and Soares (2010). Likewise, Glancy and Yadav (2011) proposed a computational model for fraud detection to circumvent the manipulation of information in annual filings with the Securities and Exchange Commissions (SECs). Further, Cai, LeKhac, and Kechadi (2012) proposed a combined classification model tagged G-KM-NC to forecast and classify financial datasets using C4.5, Gaussian non-linear classification and Multilayer perceptron techniques.

Dorrnsoro, et al. (1997) used Fisher's discriminant analysis for fraud detection in credit card operations. The linear nature of Fisher's pattern transformation makes it unsuitable for a problem such as fraud detection which one hardly expects to be linear rather a Nonlinear Discriminant Analysis (NLDA). Neural models were favoured because of its complexity and behaviour with respect to local minima and model training. Complexity estimates depend on the particular criterion function used.

Also Murad and Pinkas (1999) presented three-level-profiling for fraud detection. The three-level-profiling method operates at the account level and points to any significant deviation from an account's normal behaviour as a potential fraud. In order to do this, 'normal' profiles were created based on data without fraudulent records (semi supervised). In the same field, Burge and Shawe-Taylor (2001) used behaviour profiling for the purpose of fraud detection.

The gap here is its low dimensionality as stated earlier and it is not suitable in a multi-variate cases being considered.

Cox, Eick and Wills (1997) combined human pattern recognition skills with automated data algorithms. In their work, information was presented visually by domain-specific interfaces, combining human pattern recognition skills with automated data algorithms. Maranzato, Pereira, Naubert, and Lago (2010), and Wilson (2009) used the logistic regression method as a tool to discriminate fraudulent actions from legitimate actions for insurance companies and e-commerce. Field and Hobson (1997) presented a neural network based fraud management technique hinged on profiling techniques. Fawcett and Provost (1997) presented a rule-based tool for fraud detection using a series of machine learning methods.

Fraud detection with Bayesian networks was presented in Ezawa (1996), where the author used a Bayesian network as a normative expert system. The author focused on the unbalanced ratio of fraud cases to legitimate cases with different misclassification costs. A critical factor is whether a fraud detection model is accurate enough in order to provide correct classification of a case being fraudulent or legitimate, since fraud detection tools with largest predictive capability are always required in practice.

In this context, some other researchers have drawn comparisons among these techniques; for instance, Shen et al., (2007) compared logistic regression, neural networks and regression trees. Their results show that the proposed model of neural networks and logistic regression approaches outperform decision tree in solving the problem under investigation. Taniguchi, Haft, Hollm, and Tresp (1998) applied neural networks techniques, Gaussian mixture and Bayesian networks, reached similar conclusion and suggests a combination of three methods.

Though Bayesian networks are more accurate and require a short training time, they are slower in the application for new instances.

Viaene, Dedene and Derrig (2005) stated that Machine learning and artificial intelligence solutions are increasingly explored for the purpose of fraud prediction and diagnosis, especially in insurance domain. Among the considered techniques used were neural network classifiers trained according to MacKay's (1992) evidence framework approach to Bayesian learning.

Chang and Chang (2012) used decision trees (C4.5) and the instance-based learning algorithm to construct early fraud detection methods for classifying fraudsters and legitimate users. This is unsuitable where multiple attributes are being considered. Though, there are far too many attributes for the learning scheme to handle, and some might be irrelevant or redundant. Aha and Kibler (1991) proposed an extended model — IB1 (Instance-based learning1), which is one of the instance-based learners that implemented a nearest neighbour algorithm, using Euclidean distance function with equal weighting.

Chang & Chang (2012) used decision trees (C4.5) and the instance-based learning algorithm to construct early fraud detection methods for classifying fraudsters and legitimate users. Decision trees could be adopted for both nominal and categorical data attributes and are simple to construct, understand and interpret. However, small changes in the data will cause a significant change in the decision tree, it soon becomes over-fit and lastly it is unsuitable where multiple attributes are being considered as there are far too many attributes for learning schemes to handle, and some might be irrelevant or redundant.

Methods like rule mining (Fawcett & Provost, 1997; Adomavicious & Tuzhilin, 1999), clustering (Oh & Lee, 2003), Bayesian network (Buschkes et al., 1998), neural network (Manikopoulos & Papavassiliou, 2002; Daskalaki et al., 2003), latent Dirichlet allocation (Xing & Girolami, 2007), and decision tree (Daskalaki et al., 2003) are some examples in this context. A comprehensive survey of data mining techniques applied to various fraud detection problems was presented in Phua et al. (2005).

Statistical methods (parametric and non-parametric) along with artificial intelligence and machine learning have been widely used to make classifiers (Desai et al., 1996; Thomas, 2000; West, 2000). Linear Discriminant Analysis (LDA) is the oldest and most common statistical tool of handling classification problems (Lee et al., 1999). It has been employed in fraud detection and credit scoring (Desai et al., 1996; Daskalaki et al., 2003; Lee et al., 2006). The advantage of Linear Regression (LR) in comparison with LDA is that LR does not make any assumption about the distribution of the independent variables. Furthermore, the result of classification is not in the **YES/NO** format, but is an estimated probability of each observation belonging to a given class, and it is easy to interpret (Timm, 2002).

In recent times, neural network has been widely used due to its ability to model complex and non-linear models, as it does not have any strict limitations and rigorous assumption for the type of input data (Stern, 1996; Anderson & Rosenfeld, 1998). Its downsides include long learning time, over-fitting error, and black box characteristics (Bishop, 1995; Hippert et al., 2005).

Thus far, there have been lots of comparisons among algorithms in order to find the most appropriate classification algorithm in the literature. Caruana and Niculescu-Mizil (2006)

presented an extensive empirical comparison among ten supervised methods including: neural nets, decision trees, support vector machine (SVM), logistic regression, naïve bayes, memory-based learning, random forests, bagged trees, boosted trees, and boosted sums using a variety of performance criteria to evaluate the learning and generalization capabilities of the mentioned methods. Their results show a relative superiority of some methods in terms of some criteria. The salient problems are model input data and parameter adjustments.

Iroko (2015) proposed a method for the selection of efficient hidden layers and activation function that would yield optimum classification results especially for binary class problems as the case under consideration. The study concluded that the linear activation function performs better than all the other activation functions considered under various parameters at the 10th hidden layers.

Wiese and Omlin (2009) proposed some dynamic machine learning methods using SVM and memory recurrent neural network features to prevent high rate of misclassifications inherent in most feed-forward neural networks whereby in addition to each transaction attributes, time is factored in.

2.10.1 Network Topology Size

The size of a satisfactory neural network topology depends on the number of training cases, noise level and the complexity of the classification function to be learned. Also the size of hidden nodes depends on the stopping criteria chosen, (i.e. early stopping, weight decay, bootstrap or estimate the generalization error) in order to avoid over-fitting. The tendency is to select the one with the minimal generalization error.

Hence, RBF standardized epochs was adopted for early stopping strategy, while the data was split into 85:15% proportion. The network was trained 200 times each time. The obvious issue here is having to retrain the network many times, but it yielded a good generalization.

2.11 Model Selection for Fraud Detection

As stated earlier, many models have been explored for fraud detection, the widely used ones would be highlighted along with their defects. First, the Neural networks — the role of neural networks was to provide general and efficiently scalable parameterized nonlinear mappings between a set of input variables and a set of output variables (Bishop, 1995). Neural networks have shown to be very promising alternatives for modelling complex nonlinear relationships (Desai et al., 1996; Lacher et al., 1995; Lee et al., 1996; Mobley et al., 2000; Piramuthu, 1999; Salchenberger et al., 1997; Sharda & Wilson, 1996).

As much as the modelling flexibility of neural networks is very attractive for modelling complex and non-linear models (Stern, 1996; Anderson & Rosenfeld, 1998), yet some practical issues persist when implementing neural networks, such as the impact of the initial weight choice, how to set the weight decay parameter, and how to fit the noise in the training data. Other defects include long learning time, over-fitting error and black box characteristics (i.e. lack of explanatory power) (Bishop, 1995; Hippert et al., 2005). Nonetheless, they are crucial to the success of any neural network implementation.

Viaene et al. (2005) argue in support of Bayesian learning algorithm, but they emphasise the importance of input relevance, arguing that it is not uncommon for domain experts to ask which inputs are relatively more important or contribute most to fraud detection. As such, methods for input selection are not only capable of improving the human understanding of the problem, but also allow for more efficient and lower-cost solutions. In practice, adding inputs

(even relevant ones) beyond a certain point can actually lead to a reduction in the performance of a predictive model (Viaene et al, 2005).

Statistical method such as Linear Discriminant Analysis (LDA) is the oldest and most common statistical tool in handling classification problems (Lee et al., 1999) that has been employed in fraud detection and credit scoring (Desai et al., 1996; Daskalaki et al., 2003; Lee et al., 2006). However, they are deficient in handling fraud cases *per se*. According to Reichert et al. (1983) two disadvantages exist for LDA: assuming multivariate normality of metric independent variables and homogeneous covariance matrix between categorical dependent variable (classes). Furthermore, the result of its classification is not in a binary format, as required for fraud detection, but it was an estimated probability of each observation belonging to a given class, and it is easy to interpret (Farvaresh, & Sepehri, 2011).

Decision tree methods such as CART (Classification and Regression Tree), ID3 (Iterative Dichotomiser-3), for categorical outcomes; and C4.5 (for continuous variables) have also been used in fraud classification problems (Rosset et al., 1999; Shao et al., 2002; Daskalaki et al., 2003; Lee et al., 2006) and are suitable for discrete-valued problems.

The neural network and Bayesian network comparison study (Maes et al., 2002) uses the STAGE algorithm for Bayesian networks and back-propagation algorithm for neural networks in credit transactional fraud detection. Comparative results show that Bayesian networks were more accurate and much faster to train, but Bayesian networks are slower when applied to new instances.

Chen et al. (2006) used Support vector machine (SVM) and Neural networks to show that when the data records are small, Support vector machines (SVM) can offer a better performance than neural network does. But, SVM tries to find a linear optimal hyper-plane to separate positive and negative cases by solving a quadratic optimization problem, but the data are not often linearly separable. In order to enhance the feasibility of the linear separation, the input space is transformed via a non-linear mapping into a higher dimensional feature space by using a kernel function (Steinwart & Christmann, 2008).

So far, there have been lots of comparisons among algorithms in order to find the most appropriate classification algorithm in the literature. Caruana and Niculescu-Mizil (2006) presented an extensive empirical comparison among ten supervised methods including: neural nets, decision trees, SVM, logistic regression, naive Bayes, memory-based learning, random forests, bagged trees, boosted trees, and boosted sums, using a variety of performance criteria to evaluate the learning and generalization capabilities of the methods. Their results showed a relative superiority of some methods in terms of some criteria. Model input data and parameter adjustments were outstanding problems.

Though the modelling flexibility of neural networks makes them a very attractive and interesting alternative for pattern learning purposes, unfortunately, many practical problems still remain when implementing neural networks, such as:

- i) Determining the impact of the initial weight choice,
- ii) How to set the weight decay parameter,
- iii) How to avoid the neural network from fitting the noise in the training data.

Nonetheless, they are crucial to the success of any neural network implementation. Recently, NN has been widely used due to its ability for modelling complex and non-linear models, and also because it does not have any strict limitations and rigorous assumption for the type of input data (Stern, 1996; Anderson & Rosenfeld, 1998). The downsides, on the other hand, are its long learning time, over-fitting error, and black box characteristics (Bishop, 1995; Hippert et al., 2005).

Some Artificial Intelligence techniques have been used to detect fraud. These include neural networks, case-based reasoning, probabilistic algorithms, adaptive diagnostic algorithms based on nearest neighbour classifiers, decision trees, hidden Markov models among others. However, in general, the results of these techniques are not very user friendly, i.e. it does not explain the results about why some transactions are considered abnormal or fraudulent. Additional approaches used in conjunction with these and other analytical techniques include case-based reasoning, fuzzy logic, genetic algorithms, and fractal-based transforms. Fractal-based transforms (relatively new as data analysis tools) are interesting as they are also used as aggressive, lossless data compression algorithms.

2.12 Fraud Classification with Ensemble Methods

There are many literature that recommend single classifiers like Neural net, Decision Tree, SVM, and other learning methods (Daskalaki et al., 2003; Hung et al., 2006; Coussement & Poel, 2008; Hilar & Mastorocostas, 2008). There are also some good empirical evaluations that strongly support ensemble methods (Kim et al., 2003; Caruana & Niculescu-Mizil, 2006). Due to the diversity among base classifiers, Kim (2009) and Bauer and Kohavi (1999) have noted that ensemble methods generalizes better by combining base classifiers. Based on these premises, the Radial Basis Function is proposed in this work with higher predictive capability and accuracy.

It is agreed that the combination of data based model, and ensemble modelling techniques yield good prediction results. It is conceptually interesting and practically useful to enhance the predictive capabilities of inductive models with corresponding little or no risk of over-fitting or parameter tuning. Single base models are refined with ensembles techniques. The downside of ensemble technique is, of course, increased computation of creating base models and aggregating them for quality prediction.

2.13 RBF Model Robustness

An evolutionary algorithm as RBF network can be used to improve the robustness and accuracy of the previously mentioned techniques. A RBF model is an approximation of a true model; the study aims to find a model that is as close enough and represents better the issue at hand. In view of this, it would be re-assuring if our model is insensitive to small changes in the training data as mitigated by higher processing power of modern processors. The score functions and underlying assumptions (if any) must be slightly relaxed so that the model's parameters and predictions do not vary drastically.

2.14 Classifier Issues

According to Michie, Spiegelhalter and Taylor (1994), the following issues are peculiar to most classifiers, *viz*: Accuracy, speed, learning rate and comprehension.

Accuracy: There is the reliability of the rule, usually represented by the proportion of correct classifications.

Speed: Quite often, the speed of the classifier is a major consideration, hence a classifier that is 89% accurate may be preferred over one that is 93% accurate if it is many times faster in testing.

Comprehensibility: How clearly understandable the classification procedure is.

Learning rate: A preferred classifier must learn a classification rule quickly with minimal number of observation attributes, or make adjustments to an existing rule in real time, in order to formulate a model.

2.15 RBF Network Benefits

The proposed RBF model also called radial basis function artificial neural network (RBF-ANN), since it is a variant of artificial neural network, could interact with online or operational transactions rather than mere historical warehoused data to analyze transactions in order to detect fraud and trigger timely alerts as necessary. A feed-forward radial basis function neural network with three-layers was introduced by Ghosh and Reilly (1994), the results showed more accuracy with shorter training time, but are slower on the application of new instances.

In this research work, RBF-ANN was optimized and applied to detect fraud in a less complex, reliable and faster computations. An intelligent fraud monitoring and detection system based on this model can be integrated in interfaces over Transmission Control Protocol/Internet Protocol (TCP/IP) communication in practice. The results indicated that a fraud detection system based on mining of the operational data in this manner is realizable, resilient and robust.

Conceptually, a radial kernel (also known as Gaussian or RBF) depends on the Euclidean distance between the original attribute value vectors (i.e., the Euclidean norm of their difference) i.e.,

$$K(x_1, x_2) = e^{-\gamma \| \mathbf{a}_{1:n}(x_1) - \mathbf{a}_{1:n}(x_2) \|^2} \quad (2.1),$$

where $\gamma > 0$ is a parameter,

$\mathbf{a}_{1:n}(x)$ = the attributes vector values, i.e., $a_1(x), a_2(x), \dots, a_n(x)$

This objective function makes the contribution of instance x to the prediction calculated for instance x_* , to be proportional to $K(x, x_*)$, which exponentially decay with increasing distance between them. The resulting prediction is therefore highly localized and, as such, can be well fitted to arbitrarily complex target functions or concepts. The RBF prediction function is:

$$g(x_*) = \sum_{x \in T} \eta_x k(x, x_*) + w_{n+1} \quad (2.2)$$

RBF is an approximating and interpolating function of large, say n , of radial basis functions, each with different centres x_i and weights w_i . The weights can be approximated with linear least squares using linear algebra, which makes analysis easier and computations faster (Orr, 1999), See Figure. 2.3.

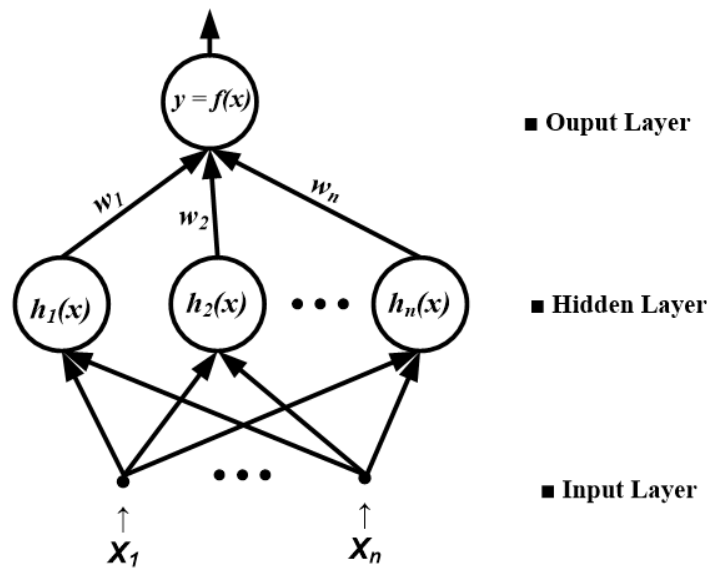


Figure 2.3: A Basic RBF Neural Network Schematic Architecture

Typically, Figure 2.3 indicates how a conventional RBF network feeds-forward the input vector $\mathbf{x}$ to m basis functions whose outputs are linearly combined with weights into the network output. Each hidden unit receives input from neighbours or external units and use this to compute an output signal which is propagated to other units along with adjustment of the weights. The system is inherently parallel as many units can carry out their computations at the same time.

Since RBF is an abridged ANN, it is also called RBF network, where the basis functions perform the activation function of the network for approximating and interpolating. RBF was first introduced by Broomhead and Lowe (1988) to act as a kernel in support vector machines (SVM) and as a simple neural network (Hardy, 1990). RBF performs better where other methods falter aside, it easily converges (Buhmann, 2004).

Another feature of RBF model is that it is non-parametric in nature, i.e. only very little, if at all, is known about the true function which is being estimated; thus enabling the resulting model to be broadened to allow all class of functions. Also, RBF is a Feed-forward neural network (FNN) model that is frequently used for its global approximation properties and freedom of local minima.

2.16 RBF as a Binary Classifier

Radial basis function is especially a classification model that could explore online and off-line big data in detecting fraud activities (such as money laundering, electronic commerce scam, dubious insurance, mortgage and health claims, etc.), since fraud activities are generic with similar characteristics but with different parameters. For instance, the same approaches and underlying models used to develop a fraud-detection capability for a bank can be used to develop medical insurance fraud detection applications. Obviously, the difference is how the models are parameterized.

RBF networks are typically one hidden layer, feed-forward networks with radially symmetric Gaussian activation functions as opposed to MLP networks (Multi-Layer Perceptron) using Sigmoid activation. Besides, RBF information is represented locally in the network (in contrast to MLP, where it is globally represented); the networks trains faster and are readily interpretable. Depending on application, RBF model is adjustable to non-parametric, faster, dynamic options and provides an adaptive learning approach to fraud detection. It equally

provides a generic mechanism for adding new information to the model without retraining it, thereby saving computational time.

A salient feature of RBF model is that it is non-parametric in nature, i.e. only very little, if at all is known about the true function which is being estimated, thus enabling the resulting model to be broadened to allow all classes of functions. This is so because, neural networks and of course radial basis function networks are nonparametric. Estimating the accuracy of a classifier for supervised learning algorithms can be used to estimate its future prediction accuracy and to determine the level of confidence one should associate to optimize the classifier's output in the prediction system. A model with low support may have occurred by chance or be uninteresting. However, Confidence measures the reliability of the inference made by a model. The higher the confidence, the more likely it is for Y to be present in transactions that contain X , for instance.

For estimating the final accuracy of a classifier one would like an estimation method with low bias and low variance. In some application domains, to choose a classifier or to combine classifiers the absolute accuracies may be less important and one might be willing to trade off bias for low variance.

It is worthy to note that in some application domains, the absolute accuracies of a classifier may be less important and one might be willing to trade off bias for low variance, this is not the case with fraud detection where absolute accuracy is imperative in so far as the system has zero tolerance for frauds or losses. This would be discussed further under misclassification cost in section 3.9.2.

2.17 Summary

A total of one hundred and forty-one (141) literatures, including published books, journals, conference proceedings, corporate reports and other periodicals were reviewed in the chapter.

The outcome of these results are in the following conclusions:

- i. Fraud is a global phenomenon;
- ii. New technologies is opening new frontier for new crimes;
- iii. Fraud detection is a classification issue;
- iv. A wide-range model can be used to express anomaly in a dataset;
- v. Researchers have attempted multi-faceted approach to curbing fraud crimes;
- vi. Many techniques involving statistics, database technology, artificial intelligence, pattern recognition, machine learning, information theory, information retrieval, high-performance computing, and data visualization have been deployed;
- vii. Some of these fraud detection techniques can be broadly divided into statistical analysis, machine learning cum artificial intelligence and visualization.
- viii. Some of the techniques were not efficient in terms of feedback or were computationally slow; and
- ix. Of machine learning techniques, neural network was better than other networks in so far as accuracy and shorter training time are concerned.

These outcomes formed the basis of this study. Though, many techniques have been proposed to detect frauds (innately outliers) in transactions or systems generally, the proposed RBF method is unique due to its global approximation, accuracy and faster training properties.

CHAPTER THREE

METHODOLOGY

This chapter explains the methodological framework for this study. The framework consists of three inter-dependent phases and each phase is directly related to the research objectives outlined in chapter one. Each phase is further divided into sub-tasks. The modelling technique presented in this study is illustrated with *R* (version 4.3.2) codes listed in Appendix B. The reasons for the choice of *R*, is that *R* is a vectorised software, ideal for handling large dataset. The model was trained using a dataset from an online German bank credit data.

The RBF-ANN algorithm was used for base-models creation and classification, with their *R* implementations. Several DMR (Data Mining in *R*) packages, containing defined functions and other simple utilities were used. The datasets were randomly partitioned into training and test subsets, to build and evaluate the performance of the model. The training and test ratio data split was 85%:15%. These sufficed for the comparison of misclassification error (*me*) obtained from the corresponding class labels.

3.1 The Research Framework

The following is the extended data mining process flow adopted in this study (depicted in Figure 3.1), culminating in the overview in the next section:

1. Identify inductive bias (i.e. identify the attributes most useful for the model creation)
2. Obtain Datasets
3. Normalize data (dimensionality reduction, cleaning, noise/outlier removal) and regularize the parameters
4. Split the data into Training and Test dataset (85:15) respectively
5. Set a best performance threshold
6. Generate base models
7. Train and use the RBF algorithm to classify dataset
8. Score the classification model (the classifier)
9. Does the identified model measure up i.e. generalize the datasets? If No, go to step 6
10. End.

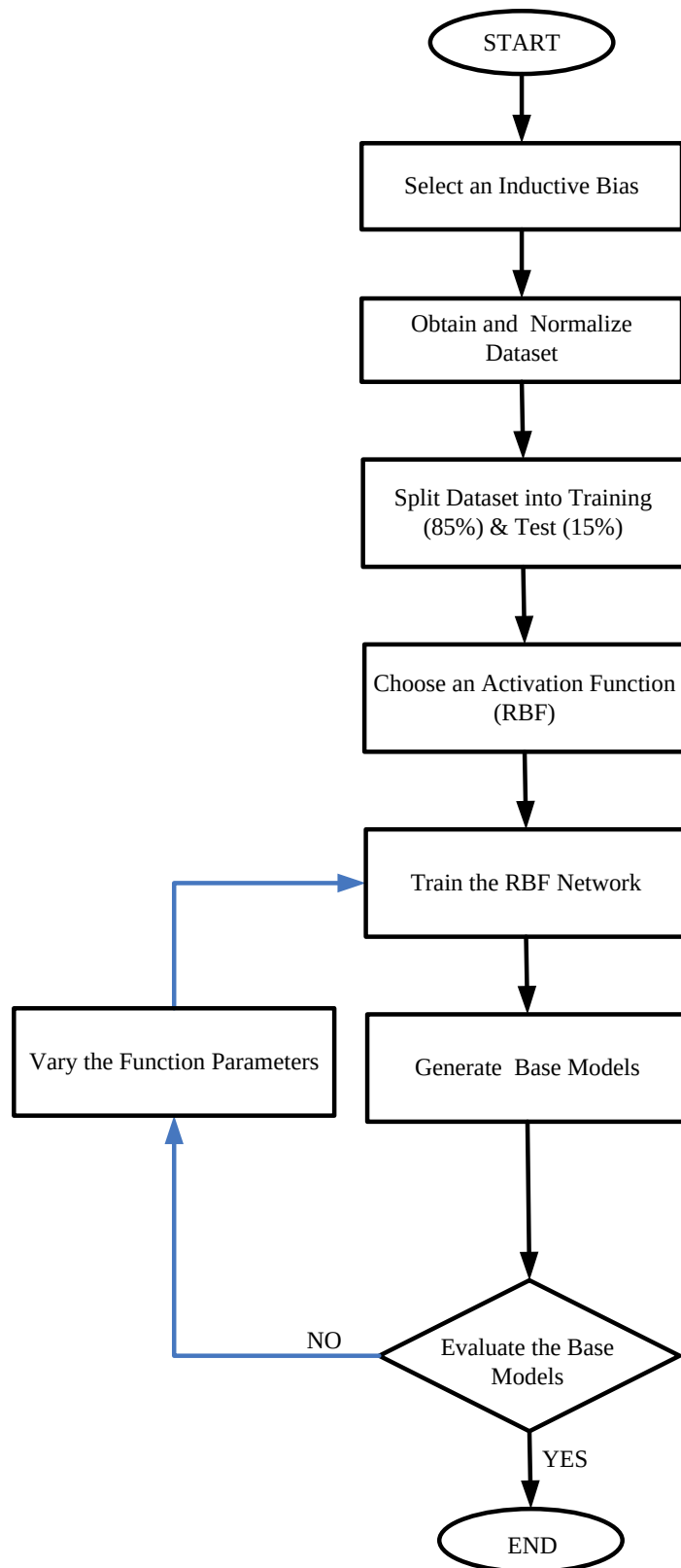


Figure 3.1: Data Mining Process Flow

To realize this process flow, the following procedure was adopted:

1. Train an RBF network with training examples, which consist of a pattern of activities for the input units together with the desired pattern of activities for the output units.
2. Determine how closely the actual output of the network matches the desired output, by varying the weight of each connection so that the network can produce a better approximation of the desired output.

In practice, section 3.2 framework could be applied for actual implementation of the strategy to gather data, normalize it and applying the RBF algorithm to collected dataset for fraud prediction. However, one can skip the section to continue with the RBF model development and evaluation framework as presented in section 3.3.

3.2 The Framework Design

In the architecture shown in Figure 3.2, data sources are spread in multiple locations (Location A, B, C, ..., N) usually local and remote data sources, which could be operational, internal or external data sources, which are subsequently aggregated in a single location (a Data Warehouse) via transactions Monitors, using a client-server configuration through TCP/IP protocol (Fig. 3.2, step1). The data is then pre-processed. The selected data is, cleaned and transformed as necessary under the guidance and knowledge of a domain expert. This process could be time-consuming, but could be mitigated in part if the data is already in a structured database, or a data warehouse, (Fig. 3.2, step 2). See the *R* code for the pre-processing in Appendix B.

This is a three-tier architecture framework whereby the data sources are aggregated in a single location on a data server.

The substantive data mining software, query applications and fraud detection algorithm are housed on an application server while the database and its associated drivers are on an independent server, to be discussed in later sections.

The model is simulated with the German Bank credit data (in other climes, synthetic data could be applied as most fraud transactions are classified, i.e. not available in public domain) and the performance evaluation of the model was derived from several runs and performance metrics generated, which would be compared with some other methods in order to show how prompt the model is in fraud detection.

In order to minimize the network topology, a certain number of training cases could be applied, the noise level noted and a minimum hidden nodes with weight decay observed for run, until the results become divergent (a stopping strategy). See Table 4.4.

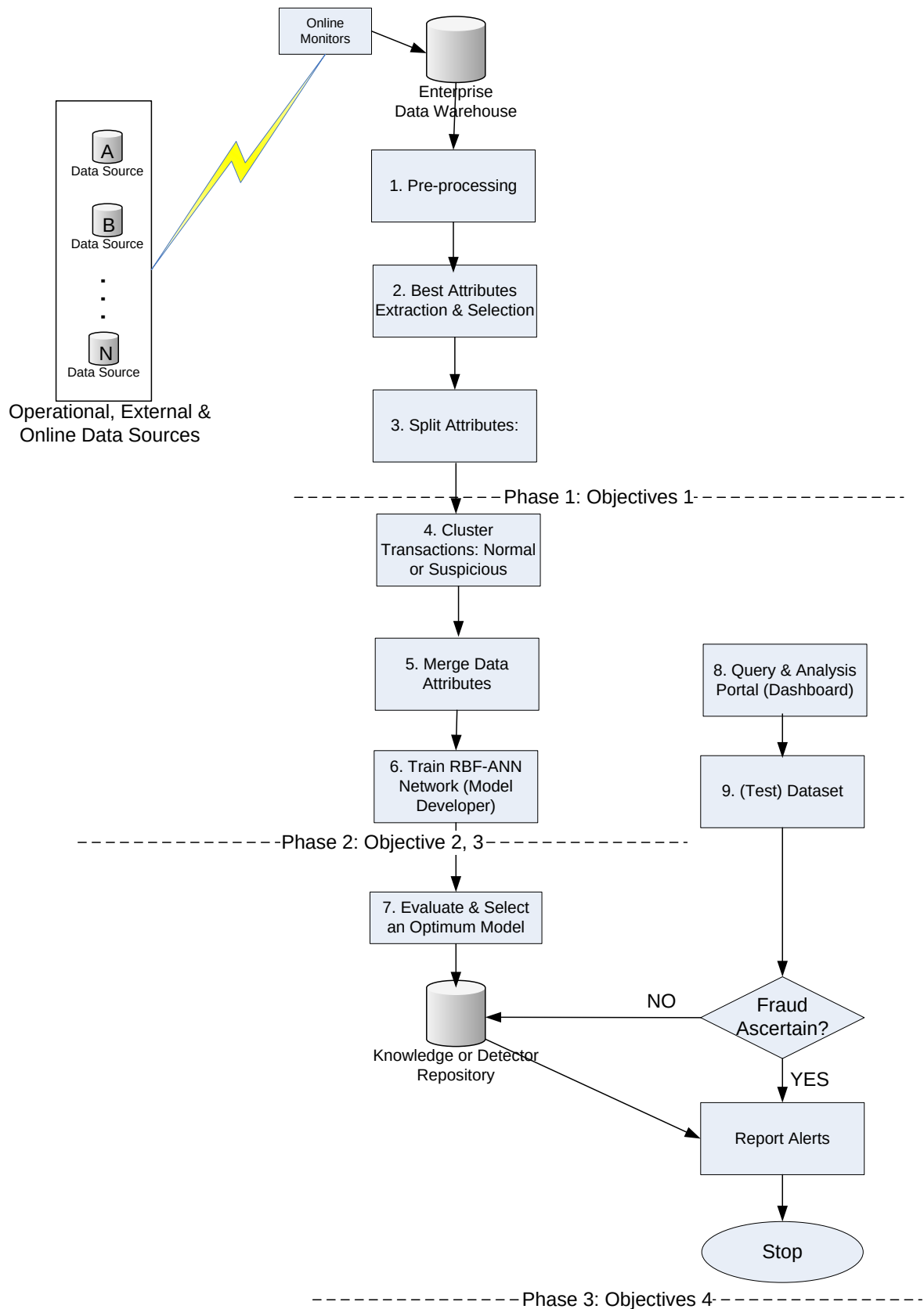


Figure 3.2: The Flow of RBF-ANN Fraud Detection Framework

3.2.1 The System Monitors

The monitoring system based on RBF neural network consists of a set of PCs (at least ten (10) Personal Computers) for mining operational transactions. Eighty-five per cent (85%) of the transactions are to be assigned as training samples set for the neural network. Each of these transactions are inputs of the neural network, of which the output status are binary values of 0 (legal transaction) and 1 (illegal transaction). If an input for a neural network is close, the corresponding transaction is identified as illegal; otherwise it is identified as a legal (or normal) transaction.

To test the predictive ability and the generalization of the derived model, we shall apply the remaining 15 per-cent of the test transactions.

A generic Extraction, Transformation and Load (ETL) tool or a specialized Warehouse Builder is used to size the platform. Quite often, data is loaded from source systems into data warehouse during off-peak hours using underutilized or standby nodes for this activity.

3.2.2 Data Attributes Selection and Transformation

Attributes selection is a pre-processing process, it requires some refined and computationally efficient algorithms to impact better quality on the final model. The objective is to determine data attributes relative to the strength of their relationship to the target attribute of the task under consideration (See Table 4.1). At end, smaller attribute subsets are identified to reduce the risk of over-fitting, that are easier to inspect, understand and apply to the training and test algorithm.

Heuristically, attributes selection is performed based on the area of interest prior to creating a fraud detection model, thereby making good models easier to find. This is where a domain expertise intuition comes in so as to determine suitably minimal attributes worthy of selection that are sufficient to detect suspicious transactions which could yield utmost performance (i.e.

to reduce the runtime). Specifically, we suggest that the dataset attributes should be split into two: namely the *main* parameters, and the unselected parameters of a transaction otherwise called *subordinate* parameters. Quite often, normal transactions are more frequent than the fewer abnormal transactions; so the class label or the network output is tagged “*normal*” or “*abnormal*”.

3.2.3 Data Access, Analysis and Inductive Training

The next is Figure 3.2 step 3, where a generic cluster technique is applied on the *main* parameters. The data mining engine accesses, splits the data to learn and subsequently predicts anomalous (or normal) transactions accordingly.

Normally, the *German Bank* datasets preliminary results indicate that most of the transactions are normal, while only a few are suspiciously fraudulent. We then merge the segregated subordinate attributes to the main attributes. These combined data were fed into the RBF-ANN network for training and the output is stored in the fraud knowledge repository or detector.

The function of the model developer is to dynamically generate and share new fraud detection updates as shown in the Figure 3.2. In this framework, the first instance of a detected fraud may have its exemplary data processed by the model developer, which is subsequently used to detect new frauds and shares it with the detector(s).

In this design, a data mining engine combined with feature extraction algorithms and dynamic machine learning algorithms serves as the model developer for the detector(s). It receives data for anomalous events from a detector, computes patterns from the data, compares them with stored normal transactions to verify the transaction and constructs the features accordingly.

3.2.4 Enterprise Data Warehouse Configuration

Typically, an enterprise data warehouse is deployed on a single Symmetric Multiprocessing (SMP) system, where programs/data are processed by multiple processors sharing the same operating system and memory or a cluster of similar systems that are cost effective with capability for a few CPUs (say 4 CPUs) featuring Intel or AMD processors hosting databases of about 50 terabytes with support for at least 16 or more nodes (with higher disk throughput in the range 2.5 to 10 GB per second). Emphasis should be on manageability, scalability and cost. The utmost advantage introducing SMP is to dynamically balance the workload among the processors in order to process the data faster.

Most of these configurations are suitable on Linux, Unix or Windows operating systems. To reduce the system downtime and for fault tolerance, different RAID (Redundant Array of Inexpensive Disks) schemes could be used, but RAID-5 is common in order to restore on inoperative or corrupt disks.

3.2.5 Transactions Query Portal

Transactions query portal or a dashboard is where casual users or analysts can verify the status of doubtful transactions (Fig. 3.2, step 8), the dataset is passed on to the trained RBF-ANN fraud-knowledge database to determine its suspicious level (Fig. 3.2, step 9). The level of the transaction suspiciousness would inform the fraud investigator(s) whether it is enough proof of fraud or to probe further.

3.3 Determining the Size of the RBF-ANN Topology

RBF is a feed-forward neural network, see Fig. 3.3, having a set of inputs and a set of outputs, in between the inputs and outputs are a layer of processing units, called *hidden nodes* (or neurons), each of which processes a radial basis function. The outputs are weighted sum of the hidden units' output. In this scenario of classifying datasets, the inputs represent the data

attributes while the output is a binary 2-class output (denoting Normal or Suspicious transactions). The purpose of an acceptable learning process is to produce a model that fits well into a variety of previously unknown data objects and makes the model generalizes well.

Figure 3.3, graphically illustrates a conventional RBF network feeds-forward; the input vector $\mathbf{x}$ to m basis functions whose outputs are linearly combined with weights into the network output. The information flows only in the forward direction, i.e. from inputs to outputs. The neurons that perform the final computation, i.e. whose outputs are the outputs of the network, are called output neurons which is the outcome of the hidden nodes numerical optimization.

The other neurons, which perform intermediate computations (hidden nodes), the network connections are the inter-activity or computations performed by each neuron such that the output of a node forms the input for other nodes; these are implemented as software programs, written in any convenient language (R in this study) and running on any computer.

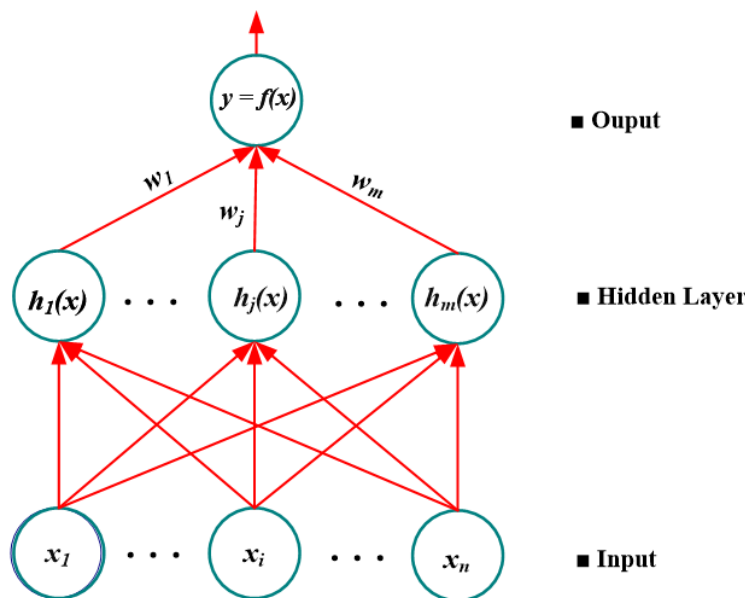


Figure 3.3: A RBF Network with n Inputs, Hidden Layers with m -basis Functions and the Output

As depicted in Figure 3.3, RBF neural network has n -inputs, h hidden nodes and m outputs, otherwise called **n - h - m neural network** such that each of the vector $\mathbf{x}$ feeds forward to m -

basis functions whose output are linearly combined with weights, ω_i into the network output $f(x)$. The input vector, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, in which $\boldsymbol{\omega} \in \mathbb{R}^{n \times m}$ is the output weight matrix, the network output is:

$$\mathbf{y} = f(x) = \mathbf{W}(\mathbf{x}) = \sum_{i=1}^h \omega_i \varphi_i(\mathbf{x}) \quad (3.1)$$

Where, $\varphi_i(\mathbf{x})$ is the activation function of hidden node i ; ω_i is the weight of node i and h is the number of hidden nodes.

3.4 The RBF Network Overview

The procedure adopted in this work to construct a RBF network for financial fraud classification models is a supervised learning scheme, whereby the network is trained with feature inputs $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ and the corresponding outputs $y_i \in \{0, 1\}$. The sole objective of the training algorithm was to ensure that a set of input features would yield the anticipated set of outputs using the RBF network framework; such that the developed final model could subsequently classify previously unseen data features into their respective true classes.

As part of the design requirements and the consequent model training, prediction and evaluation, there is the need to follow the following processes:

- i. obtain the relevant inputs, and identify the input data features (attributes) that have a significant influence on the model's response (Output),
- ii. train and test the RBF network,
- iii. find the appropriate number of hidden neurons, estimate other parameters to minimize the cost function during the network training, and
- iv. evaluate the model's predictive and generalization ability after training.

Depending on the network evaluation outcome, it might be necessary to iterate partly or wholly the entire processes (i – iv), just outlined.

The purpose of an acceptable learning process is to produce a model that fits well on a variety of previously unknown data objects and makes the model generalizes better. Machine learning (ML) is the kernel of any data mining technique, due to its capability to gain insight into a problem, data selection and model search steps.

The input vector, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, in which $\omega \in \mathbb{R}^{n \times m}$ is the output weight matrix, the network output is:

$$Y(\mathbf{x}) = \sum_{i=1}^h \omega_i \varphi_i(\mathbf{x}) \quad (3.1')$$

Where, $\varphi_i(\mathbf{x})$ is the activation function of hidden node i , ω_i is the weight of node i and h is the number of hidden nodes. The RBF network could assume a variety of activation functions for hidden nodes e.g. Thin-plate-spline ($x^2 \log(x)$); Gaussian (e^{-x}); multi-quadric ($(1+x)^{1/2}$); inverse multi-quadric ($(1+x)^{-1/2}$); and Cauchy $(1+x)^{-1}$, of which Gaussian function is most frequently used and this research also adopts, i.e.

$$\varphi_i(\mathbf{x}) = e^{-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{\delta_i^2}}, \quad (\text{for } i = 1, 2, \dots, h) \quad (3.2)$$

Where, $\mathbf{c}_i = (x_1, x_2, \dots, x_n)^T$ is the centre of the hidden node i ; δ_i is the constant extension of hidden node i . This could be expanded to become:

$$Y(\mathbf{x}) = e^{-\sum_{i=1}^h (x_i - \omega_i)^2 \div 2\omega_{h+1}^2} \quad (3.3)$$

Where the parameters $\omega_i, i = 1 \text{ to } h$, are the Gaussian centre positions and ω_{h+1} is its standard deviation as compressed in equation (3.2).

RBF special feature is that its response decreases (or increases) monotonically with distance from the centre, however, the beauty and the flexibility of the selected (Gaussian) activation

function here is its ability to fit different functions derived from the freedom to choose different values for the weights, as we are going to see in the subsequent section, the function varies during the learning process, thereby making the model non-linear.

3.4.1 Input Selection

The input variable selection process, i.e. the selection of the components of vector $\mathbf{x}$ in $Y(\mathbf{x}, \mathbf{w})$; can be accomplished in two steps, viz:

- i) input vector, $\mathbf{x}$ dimension reduction, and
- ii) the selection of relevant variables whose influence on the response variable is vital rather than the noise variables' influence.

This shows that the input selection process is of two prongs, i.e., the model's independent attributes dimension reduction and the rejection of inputs that are irrelevant to the model's prediction.

For practical purposes, comparisons between models that have different numbers of inputs can be carried out using several techniques that can be used as discussed in section 3.7. However, a simple but efficient principle, Garson algorithm (1991), was used to rank the variables in order of relevance to the output. See Table 4.1 for a list of the variables.

The summary of the procedure for discarding irrelevant variables is:

1. Collect the set of potential or candidate inputs,
2. Select the input that is most correlated to the output (perhaps in observation), and
3. Iterate for more relevant features selection.

Prior to training, the selected input variables are normalized and used as inputs of an RBF network and train as indicated in the next sections. The essence of this procedure is to reduce the dimension of the input dataset. The sample data was split into two: n_T (training) and n_Q

(test) in the ratio 85:15 respectively to construct a classification model based on n_T at different hidden layers and the resulting fitted model was used to classify the n_Q test data, as stated in section 3.6.

3.4.2 Parameters Estimation

Next to data features selection is the estimation of parameters, i.e., estimating the parameter of a model $g(\mathbf{x}, \mathbf{w})$ by finding the numerical values of the components of the parameter vector $\mathbf{w}$ that make the model optimum, with respect to a defined criterion. A plausible approach is to find an approximation of a function in the family of non-linear functions; Gaussians with adjustable centres was selected for reasons already enumerated in the previous sections. The method was to gradually vary the complexity of the models, by increasing the number of the Gaussians functions, the number of hidden neurons, etc.

The values of the $\mathbf{w}$ component of the parameters was computed by minimizing a cost function that represents the “distance” between the predictions of the model and the measured values.

The estimation of the parameters $\mathbf{w}$ of the model $g(\mathbf{x}, \mathbf{w})$ during the model’s training was done thus:

1. selection among a set of functions within which the model is sought;
2. training of one or several models within the chosen set; and
3. selection of the best model and the estimation of its performances;

The above steps are repeated until a satisfactory model is obtained by varying, the number of hidden neurons, the degree of the polynomial etc. The details are discussed in (Section 3.7), and the process is iterated to the second step of the previous task.

Since the desired model $g(\mathbf{x}, \mathbf{w})$ is non-linear with respect to its parameters, the gradient of the cost function is also non-linear; hence a more elaborate minimization techniques must be used to update the parameters iteratively. This is where RBF routine of RSNNS library comes in. The RSNNS (Neural Networks in R using the Stuttgart Neural Network Simulator) is a library containing many standard implementations of neural networks from within R. The model begins to learn starting with an empty network, i.e. a network consisting of input and output units, and which adds new units successively. The bias parameters are initialized around zero, so as to make the functions of the hidden nodes train slowly.

3.5 Obtaining the Cluster Centres

The k -mean algorithm, an indirect clustering approach based on inter-sample similarity measurement, is used to select k samples out of n samples as the initial cluster centre and assigns the other objects to clusters represented by cluster centres that are most similar to them according to the distances to the initial cluster centres. Then, it calculates the cluster centres of new clusters, one after the other, until the metric function, usually, *mean square error (mse)* begins to converge.

Suppose h initial cluster centres are created from the samples, and the first h of them are selected by default. c_i (a scalar value) is the centre of cluster i ; its corresponding mean square error is σ_i . The distance norms from all the sample inputs to the initial cluster centres are defined as:

$$D_i(\mathbf{x}) = \sigma_i \|\mathbf{x} - \mathbf{c}_i\|^2, i = 1, 2, \dots, h \quad (3.3)$$

Then recalculate the new cluster centre of each cluster until the first $D_i(\mathbf{x}) = \min D_i(\mathbf{x})$ appears,

$$\mathbf{c}_i(k+1) = \mathbf{c}_i(k) + v[\mathbf{x}(k) - \mathbf{c}_i(k)] \quad (3.4)$$

When the next $D_i(\mathbf{x}) = \min D_i(\mathbf{x})$ appears,

$$\mathbf{c}_i(\mathbf{k} + 1) = \mathbf{c}_i(\mathbf{k}) - \xi v [\mathbf{x}(\mathbf{k}) - \mathbf{c}_i(\mathbf{k})] \quad (3.5)$$

Subsequently,

$$\mathbf{c}_i(\mathbf{k} + 1) = \mathbf{c}_i(\mathbf{k}) \quad (3.6)$$

If v is the learning rate of the emerged cluster centre, ξ is the ratio of the penalty rate of the cluster centre ρ to v . then the *mse* of the emerged or cluster is:

$$\sigma_i(\mathbf{k} + 1) = \mu \sigma_i(\mathbf{k}) + (1 - \mu) \|\mathbf{x} - \mathbf{c}_i\|^2 \quad (3.7)$$

where, μ is a constant usually taken to be 0.999. The learning rate of a cluster centre is:

$$v = 1 - \frac{\sum_{i=1}^h \bar{v}_i}{\ln h}, \text{ where } \bar{v}_i = \frac{v_j}{\sum_{j=1}^h v_j} \quad (3.8)$$

If this equation converges, the iteration ends. If it does not converge, the distance between the samples and the cluster centres has to be circulated.

The inter-cluster distance, d_i is the distance from cluster centre i to other cluster centres, i.e.

$d_i = \min_j \|\mathbf{c}_j - \mathbf{c}_i(k)\|$ in which k is the overlap coefficient.

3.6 Determining the Number of Hidden Nodes

Quite often, an RBF network with a robust objective function is capable of approximating accurately a given function. However, this can be accomplished with a proper network, but the network size has to be determined, as it is unknown *ab initio* for a particular problem. Overcoming this challenge requires some adaptive growing technique as proposed by Lee et al. (1999) that dynamically adjusts the number of hidden nodes. The rules are:

- i) A node is added when the network is not converging, or the function value exceeds the termination criterion value and the centre of the new node is placed in the position of the training pattern with the largest residual and other parameters can be initialized randomly.

- ii) Delete the node whose outgoing weight is smaller than a prescribed threshold after some learning iterations as it does not give significant contribution to the network.

The success of this rule is dependent largely on the training pattern quality, especially when there are no outliers. However, an RBF network should not have too many initial nodes, as the outliers will be interpolated by some of these initial nodes during training. Hence a better approach is to start with fewer initial nodes. Then as the number of nodes increases, the approximating function will be more accurate in representing the normal training patterns. A larger residual in the training pattern is also regarded as an outlier.

If the cluster centre as well as the weight vector, $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^T$ of each hidden node is known. Hence, the output of hidden node i is:

$$O_{ji} = \phi_j(\|x_j - \mathbf{c}_j\|) \quad (3.9)$$

yielding the following output matrix for the hidden layer:

$$\mathbf{O} = \begin{bmatrix} O_{11} & O_{12} & \cdots & O_{1h} \\ O_{21} & O_{22} & \cdots & O_{2h} \\ \cdots & \cdots & \cdots & \cdots \\ O_{nl} & O_{nl} & \cdots & O_{nh} \end{bmatrix}, \quad \mathbf{y}_0 = \mathbf{O}\boldsymbol{\omega} \quad (3.10)$$

Since it would be erroneous to divide the sum-squared-training error by the number of base models in order to estimate the noise variance, besides some degrees of freedom allowed in fitting the model, the trained network output error estimates is:

$$\mathbf{e}_0 = \|\mathbf{y} - \mathbf{y}_0\| = \|\mathbf{y} - \mathbf{O}\boldsymbol{\omega}\| \quad (3.11)$$

In this algorithm, the winner learning rate self-adaptive to the original dataset information volume is contained in each cluster centre. The penalty rate of the competition centre is limited and in proportion to the learning rate. That is, when the centre cluster begins, the penalty rate for the competition centre is fairly high. This way, redundant competition centres

will move away from the dataset and become empty nodes that will not be learned anymore. As the computation progresses, the competition rate among cluster centres diminish too, so should the penalty rate for the competitions.

The output function of the network is a weighted sum of hidden units as stated earlier, i.e.:

$$\mathbf{W}_j(\mathbf{X}) = \sum_{i=1}^h \omega_{ij} \varphi_i(\mathbf{x}) \quad \text{for } j = 1, 2, \dots, k \quad (3.12)$$

Where, ω_{ij} are the output weights,

$\varphi_i(\mathbf{x})$ is the activation function of hidden node i .

$\mathbf{X}$ represents the input attributes vector

h is the number of hidden layers

k is the number of output units.

Typically, in a binary classification problem as this, the hidden layers of the network actively interact with one another such that the output of the hidden layer $j-1$ is the input of the hidden layer j . and that the radial basis function output is restricted to the interval (0,1) by the function:

$$Y_j(\mathbf{X}) = \frac{1}{1+e^{-\varphi_j(\mathbf{X})}}, \quad \text{for } j=1, \dots, k \quad (3.13)$$

Hence, the j th layer is the predicted response class, i.e.

$$\hat{y} = f(\mathbf{x}, \mathbf{w}) = \begin{cases} 1, & \text{if transaction is fraudulent} \\ 0, & \text{if transaction is normal} \end{cases} \quad (3.14)$$

Input data gathering, sampling and filtering is core to the model training, as well as the validation and test sets. The data set ought to be large and sufficient enough, with fewer

dimensions, as too many data attributes could lead to the model over-fitting, also too few parameters will not generalize better as it would not be able to learn enough training data.

Actually, attribute transformation is a kind of preprocessing task for inductive learning especially designed to enhance model diversity. Attribute transformation, though functionally dependent on the original data for easy analysis, in this study, we used multi-class encoding to achieve this (see Table 4.1). These transformations are required to adjust to the needs of our classification algorithm.

3.7 Base Models Creation / Training

In generating good quality base models, we need to create some sufficient, diverse and good base models by using any of the following methods:

- i) Different training sets
- ii) Different deterministic algorithms
- iii) Different parameter setups
- iv) Randomized or non-deterministic algorithm

One could use a different training set from the same domain to create each base model (or use different algorithm parameters). However, using the same training set and the same parameter setup will lead to a certain deterministic base model.

Other variant methods of creating base models by different training sets include: *Sampling instances* (drawing multiple training set samples); *Replicating instances* (some instances are selected randomly, to yield different models with unstable models especially); **Varying instance weights** (attached to data attributes), *sampling attributes* (using different attribute subsets) and by applying *Attribute transformation* (vary dataset attributes to obtain model

diversity). In this research, instance weights were varied, since RBF-ANN model is considered a stable algorithm as it does not react adversely when its parameters are perturbed (changes due to minor data variation), unlike some algorithms such as Decision trees and Regression. Besides, RBF network is weight-sensitive. Therefore, varying its weight vector is enough and sufficient to obtain different base models.

Hence, base models $m_1, m_2, \dots, m_h$ could be generated using the same training set, but different vectors of per-instance weights $\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(h)}$ with each vector $\omega^{(i)}$ containing weight $\omega^{(i)}x$ for each instance $x \in T$ (Cichosz, 2015). However, instead of generating the weights randomly, we used a refined weight adjustment scheme of AdaBoost (i.e. Adaptive boosting) algorithm, the choice of this technique stems from the fact that it is particularly suited for a 2-class classification tasks and the error level of the base models does not exceed 5% , above which it gives a negative weight. Aside from enhancing base models' creation and diversity, prediction quality is boosted, since base models are likely to be “under-fitted” rather than over-fitted. The boosting process is ideal for its actual “fitting” to the training set. Alternatively, the weights of the hidden layers could be assigned randomly; however, it is better to select weights using some refined weight adjuster techniques as discussed earlier.

There are other schemes to instantiate boosting for ensemble modelling algorithms. The most prominent of these especially for classification tasks are: AdaBoost (Adaptive boosting) algorithm, Multi-class AdaBoost and Random Forest. The *R* implementation of AdaBoost and its variants found in *R*-Project: <http://cran.r-project.org/web/packages/> was adopted in this research.

Typically, financial transactions are characterized by a set of features, say m (often numerous in dimensionality), some of which might not have correlation with fraud detection, e.g. telephone number as indicated in Table 4.1, while some of these transactions features are highly correlated to fraud detection, e.g. credit history. It is these significant highly correlated features that were selected to construct the desired model.

Meanwhile, to prevent noise and over-fitting of the algorithm, the selected features must be less than the sample size (the number of transactions) to be used for the algorithm training. t -statistics with unequal variances could be used to select sample features as enjoined by Yahya et al. (2011).

An arbitrary selection of hidden neurons for an ANN network could yield a suboptimal model with poor predictions (Lawrence *et al*, 1997). The researcher adopted a data-driven algorithm as proposed by Yahya *et al* (2012), to uniquely determine the number of hidden layers for accurate classification models' prediction. RBF networks as in other neural networks are embedded in a one-layer network topology and each layer performs a radial activated function. The output units are weighted sum of hidden unit output (Fig. 3.3). In this study, the input transactions variables was denoted by the vector $\mathbf{x} = (x_1, x_2, \dots, x_i)$, where by, $i = 20$, with two response class labels $y_i \in \{0,1\}$, i ranges from 1 to n (the size of the dataset). Hence the class label $y_i = 1$, if the i -th transaction is ***fraudulent*** and $y_i = 0$ if the i -th transaction is ***normal***. The interactivity between the input and the hidden nodes is described in section 3.4.

Due to scarcity of fraud data, a German bank credit data, a true representative of the problem, available online was used to construct a classification model based on n_T at different hidden

layers and the resulting fitted model was used to classify the n_Q test data. The outcome is discussed in section 4.1.

Actually, more than one dataset could be used over the domain of interest to ensure adequate learning by the neural network, but for reasons explained earlier in this section (See Appendix A for the dataset extract). Once trained, the neural network could be used in production for estimating the states of the previously unknown datasets.

3.7.1 Model's Training and Termination

The most popular method consists of monitoring the variation of the standard prediction error on a validation set, and in terminating training when the prediction error starts increasing. After input variable selection and training, model selection comes next in the model design methodology. Having obtained several trained (or candidate) models. The selected model should net a trade-off between learning and generalization capability by minimizing its cost function. This is illustrated in Chapter 4.

3.8 Models Aggregation and the Final Model Selection

Essentially, the prediction error is estimated as the criterion for model selection, i.e. the estimation of how well the trained model will perform on future unseen datasets, by selecting the model whose estimated prediction error is least. The estimation of the prediction error in the final selection of base models is where RBF-ANN comes in handy.

However, as a follow up to base models creation, the model aggregation and selection process follows. This is the combination of base models $m_1, m_2, \dots, m_n$ into a better model m_* using the RBF-ANN algorithm which is necessary to calculate $m_*(x)$ based on $m_1(x), m_2(x), \dots, m_n(x)$ for arbitrary $x_i \in X$. In the combined (aggregated) model, m_* is the representative of its base models.

Many model aggregation schemes are possible, viz: *Voting averaging*, *Probability averaging*, and *Weighted voting averaging*. The most widely used aggregation technique for classification tasks is *class label voting*, hence, the combined prediction is obtained by:

$$m_*(x) = \arg \max_{d \in C} \sum_{i=1}^n \prod_{h_{i(x)=d}} \quad (3.15)$$

where the $\prod_{\text{condition}}$ notation is used to denote an indicator function that yields 1 when the *condition* is satisfied and 0 otherwise, i.e.,

$$\lambda(r) = \begin{cases} 1, & \text{if } r \geq \theta \\ 0, & \text{otherwise} \end{cases} \quad (3.16)$$

Finally, the aggregation function is:

$$m_*(x) = \frac{1}{h} \sum_{i=1}^h m_i(x) , \text{ where } h \text{ is the number of hidden nodes.} \quad (3.17)$$

This is acceptable, more so, as researchers often times tend to aggregate model predictions rather than creating the models' representative. See Appendix B.2 for the *R* code implementation of the voting/average model aggregation.

3.9 Misclassification Error Rate

The prediction performance of each base model at different hidden layers was evaluated with the average *misclassification error rate* (MER) or the prediction error rate thus:

The prediction performance of each classification model at different hidden layers, H was determined using the average *misclassification error rate* (MER), $\tilde{\vartheta}_H$ given as:

$$\tilde{\vartheta}_H = \frac{1}{S \times n_T} \sum_{r=1}^S \sum_{i=1}^{n_T} I(y_T \neq \tilde{y}_T) \quad (3.18)$$

where, S is the cross-validation runs,

$I(\bullet) \in \{0,1\}$ is an indicator function.

H is the number of hidden nodes.

Averaged over the number of cross-validation runs s , where $I(\bullet)$, I an indicator function whose value is 1 if the predicted class label $\hat{y}_{iT}$ of the i th sample at the r th cross-validation run does not equal the true class label y_{iT} of the sample transactions and 0 if otherwise. Hence, $\hat{\vartheta}_H$ is the prediction error rate of the model with H number of hidden nodes.

Classification performance measures are usually computed by comparing the model's output (or predictions) on a dataset S with the true class labels of the instances from this dataset. There are various model's performance evaluation techniques, viz.: Confusion matrix, ROC (Receiver Operating Characteristic), Misclassification error, Mean classification error, weighted misclassification error, etc.

Usually as in our 2-class classification case, a model's performance are based on a scoring function, $\pi: X \rightarrow \mathcal{R}$ (maps all instances of x instances to a real number score, θ); and a class labelling function $\lambda: \mathcal{R} \rightarrow \{0,1\}$ converts the scores to a class label using a rule such as Eq. 3.15, i.e.,

$$\lambda(r) = \begin{cases} 1, & \text{if } r \geq \theta \\ 0, & \text{otherwise} \end{cases} \quad (3.19)$$

Where θ is a cut-off score (or threshold value), hence the model becomes $m(x) = \lambda(\pi(x))$.

Aside from the training dataset T , our test dataset Q from the same domain, we sought to select a model out of multitude of candidate models (generated base models), as in section 3.7. Since a model's performance is best decided by its minimal predictive error or accurate scoring, one approach is to embark on intermediate and final evaluation processes whereby the candidate models have to undergo intermediate evaluation process in order to rank them before choosing the appropriate highest performance model as regards generalization and other desirable metrics. As it is, this process often yield further refinement that could lead to

additional models to be evaluated as well until no more enhancement is possible. Thus, the test data to evaluate the candidate models' performance is justified and effectively utilized to arrive at a best quality model possible. Finally, a new dataset is used to test the chosen model.

However, this approach contrasts with k -fold cluster validation method, though this method reduces bias and variance estimators to its barest minimum. Nonetheless, we desire that the predictive performance of the formulated model can be obtained with different performance measures, in order to enhance the general approximation quality of the derived model.

Meanwhile, we shall consider some criteria, in so far as the model's performance is concerned, the following performance criteria come to fore, viz: accuracy (i.e. the proportion of correct classification); the classifier speed, comprehensibility and learning rate, dependent on the amount of observations sufficient to learn or establish a rule. As usual, some utility functions of RSNNS package with the same dataset could be used to illustrate these evaluation procedures.

3.9.1 The Model's Performance Criteria

In so far as the derived model evaluation is concerned, one measures the comparison between the model's prediction on the applied dataset S , and the true labels of the same dataset. By convention, researchers distinguish *intermediate evaluation*, whereby evaluation affects the adjustment to the derived model (e.g., attributes selection, changes in parameters and selecting a different classification model), in this case *validation* dataset is said to have been applied, whereas, when evaluating the finally created model, one applies the test set, and we say that the *final evaluation* is performed. Also, any arbitrary dataset is denoted by S , training set by T , and the test/validation dataset by Q .

Generally, misclassification error and accuracy measures indicate how often the derived model predicts accurately or otherwise when applied to previously unknown dataset. Misclassification error is an easy metric to interpret. Hence, this study adopted a weighted-miscalculation error since RBF-ANN is a weight sensitive algorithm. Now, since, misclassification error e , is a relative frequency of the model's wrong predictions from any arbitrary instance S . Suppose a weight w_x is assigned to each $x \in S$, the weighted misclassification error of model m with respect to the class c (normal or illegal transaction), on dataset S can be obtained as:

$$e_{c,S,w}(m) = \frac{\sum_{x \in S_{m \neq c}} w_x}{\sum_{x \in S} w_x} \quad (3.20)$$

Equation 3.20 is analogous to the total number of incorrect predictions made by the model ($\sum_{x \in S_{m \neq c}} w_x$) and the total number of predictions, ($\sum_{x \in S} w_x$), See Appendix B.3 for the *R* code of this algorithm. The algorithm accepts the supplied weight vector to determine the weighted miscalculation error. The evaluation is performed on the test dataset for which model prediction (“normal” or “abnormal”) is derived from the class label type of *predict* function of the *R* code.

3.9.2 Misclassification Cost

While considering misclassification error for financial fraud detection, zero-tolerance is often the desired optimum. However, misclassification cost is an additional performance criterion imposed for model evaluation, and ought to be incorporated into model development, to make it cost-sensitive.

In factoring the misclassification cost, real-valued cost is assigned for possible misclassification or we roughly estimate the actual **cost of wrong predictions** in terms of resources involved (human effort, money, power consumption, etc.) based on domain-

knowledge, though. Here, one is more concerned about those transactions that are fraudulent (1) but were predicted to be normal (0) and there should be none of such transactions, possibly.

Numerically, the cost function is a function of all adjustable parameters, $\mathbf{w}$ of the model; more so as training involves the numerical optimization of finding the parameter vector, $\mathbf{w}$ for which MSE (Mean Square Error), i.e., $\frac{1}{2} \sum (Y(x_i) - \hat{Y}(x_i, \mathbf{w}))^2$, over $i = 1, 2, \dots, h$ is minimum; where Y is the corresponding sample output (classes) of the model and $\hat{Y}$ is the predicted value of the model with parameters $\mathbf{w}$, (the model's weight).

If a model is linear with respect to its parameters such as radial basis functions with fixed centres and widths, the cost function $\hat{Y}(\mathbf{w})$ becomes quadratic with respect to the parameters, then the ordinary least squares method is suitable as it is simple and efficient to use. However, the resulting models are not parsimonious. But in situations as in this research where the model is non-linear with respect to its parameters (e.g., a feed-forward RBF network with adjustable centres and widths), the optimization case is multivariable non-linear, thus rendering the ordinary least squares method inapplicable.

3.9.3 Cost Matrix

Generally, misclassification costs can be specified as a $|C| \times |C|$ matrix (a square matrix), as in the case under consideration, a 2×2 matrix (see Figure 3.4), i.e. $C \in \{0,1\}$, where $C = 0$ is normal transaction and $C = 1$ is a fraudulent one; then, $c[x_1, x_2]$ is the misclassification cost of predicting class x_1 for an instance instead of a true class x_2 . The matrix is usually assumed to contain 1s on the main diagonal (i.e. $c[x_1, x_2] = 1, \forall x \in C$). Likewise, one can say that an instance x is a positive instance, when $c(x) = 0$ and one can say that an instance x is a negative instance, when $c(x) = 1$. As in Figure 3.4, positive integer numbers are used for the remaining entries, with 1 corresponding to the least expensive misclassification,

		Predicted Class	
		Legitimate (Normal = 0)	Fraudulent (Abnormal = 1)
Target Class Target Class	Legitimate (Normal = 0)	532 (TP)	61 (FP)
	Fraudulent (Abnormal = 1)	0 (FN)	257 (TN)

Figure 3.4: A 2-class Misclassification Cost Matrix for Fraud Transactions Detection (TP= True Positive; FP=False Positive; TN=True negative; and FN= True Negative)

The cost of classifying “normal” (or abnormal) transaction as “normal” (or abnormal) respectively is assigned 1, and the cost of misclassifying “abnormal” as “normal” or “normal” as “abnormal” is set to 7, because it is more expensive to predict an abnormal transaction as normal.

Since we are more concerned with a generic fraud classification model, a better approach to incorporating misclassification costs that is not algorithm dependent, is preferred based on *instance weighting*, as discussed in section 3.7, in which case RBF-ANN will convert its weight-sensitivity to cost-sensitivity.

Generally, the cost matrix c that assigns misclassification costs is defined as:

$$c[x, x_2] = \begin{cases} 1 & \text{if } x_1 = x_2 \\ \omega_{x_2} & \text{otherwise} \end{cases} \quad (3.21)$$

In fact, this represents the mean misclassification cost of model m as regards class c , provided weights are assigned to training instances according to their classes, where ω_{x_2} is a weight value corresponding to class x_2 . Hence, minimizing the weighted training set error is equivalent to minimizing the mean training set misclassification cost, thereby turning a weight sensitive algorithm into a cost-sensitive algorithm.

The RBF-ANN minimum cost selects the class with the lowest expected cost. Further, the relation between the minimum error and minimum cost rules must be the same for all errors and one (1) when a class is correctly assigned. Now, suppose that the expected cost, $c(i,j) = c$ for $i \neq j$ and $c(i,j) = 0$ for $i = j$, with π_i , the probability of incurring the cost, the expected cost is:

$$C_d = \sum_i \pi_i c(i, d) = \sum_{i \neq d} \pi_i c = c \sum_{i \neq d} \pi_i = c(1 - \pi_d) \quad (3.22)$$

However, we noted that misclassification costs are elusive in practice, even where there are substantial disparities in the assigned penalties or reward for committing right or wrong decision. They are difficult to quantify, they are often left at users' discretion.

Regardless of the assigned misclassification cost, the classification model with $\tilde{\vartheta}_n = \min(\tilde{\vartheta}_1, \tilde{\vartheta}_2, \dots, \tilde{\vartheta}_j)$ is chosen as the best model that fits the dataset and the number of hidden nodes that yielded this best model $n \in H$ is the optimal hidden nodes number for this classification model.

CHAPTER FOUR

RESULTS AND DISCUSSION

This section describes the process of obtaining the data used for the experimentation, its pre-processing, and the procedure for the data processing to achieve desired results. The outcome of the experiment was reported, while the salient points were highlighted.

The case study presented in this study uses CRAN (*Comprehensive R Archive Network*) packages to implement the RBF model approach to fraud detection. On occasions, other data mining packages such as Weka library could be invoked using *RWeka* interface of *R* package.

Other selected functions from the *DMR* (Data Mining in *R*) embedded in *R* Using the Stuttgart Neural Network Simulator (RSNNS) family of packages were applied as necessary, especially for model evaluation. All the packages used for this case study are listed in Appendix B.

As stated in Chapter one, the primary objective of this study was to detect and predict fraud generically using RBF network model with specific case study of a German Bank credit (available online). With this dataset, first, the study determined the contribution of each independent input variable to the network response i.e., to determine which of the inputs are most critical to fraud prediction or otherwise.

4.1 Test Environment Requirement

The testing environment was Windows 8 platform with 4 Gigabytes RAM, 3.4 GHz Intel Dual Core and *R* platform. As explained earlier in Chapter three, a three-tier architecture platform could be configured (in practice) whereby the data sources are aggregated in a single location on a data server. The substantive data mining software, query applications and fraud detection algorithm should be housed on an application server while the database and its associated drivers are on an independent server.

For the model's implementation, a number of *R* (version 4.3.2) statistical packages embedded in RSNNS library was employed such as *NeuralNetTools*, *Neuralnet*, *Utils*, *ggplot2*, *devtools* etc., which are freely available online at CRAN repository: <https://github.com/cbergmeir/RSNNS>.

4.2 Dataset Description

In this study, the dataset used consists of real **German Bank Credit** data from the UCI Repository of Machine Learning Databases (Hoffman, 2001). However, this dataset is of interest because it consists of a good mix of continuous and nominal attributes, a few missing values and no special knowledge is required to understand it. All attributes and values of real identities have been concealed due to the confidentiality of the data.

This dataset was edited to include several indicator variables to make it suitable for the algorithm under consideration in order to conform to the response categorical variables (See Table 4.1). Several attributes that are ordered categorical (such as attributes 13, 17, etc.) have been coded as integer. The predicted response class label $\hat{y}$ is dichotomously defined as follows:

$$\hat{y} = \mathbf{W}(\mathbf{x}) = \begin{cases} 1, & \text{if a credit transaction is fraudulent} \\ 0, & \text{if a credit transaction is normal} \end{cases} \quad (4.1)$$

The dataset consists of 1000 records of bank customers on whose 20 features attributes (factors), having a good mix of continuous and nominal attributes were taken. These attributes consist of several indicator variables that are suitable for the model development.

The training set consists of N samples, 850 samples of which 150 samples were used for the test/validation set.

Finally, the response variable is a binary valued variable, coded as 0 (for *normal transaction*) and 1 (for *abnormal or fraudulent transaction*), see Appendix A.2. The data attributes used are:

Table 4.1: Table of Input Variables Description

S. No	Variables	Description
1.	A1	Status of current account
2.	A2	Duration
3.	A3	Credit history
4.	A4	Credit purpose
5.	A5	Credit amount
6.	A6	Status of savings account
7.	A7	Present employment status
8.	A8	Installment rate in percentage of disposable incomes
9.	A9	Gender and marital
10.	A10	Guarantor(s)
11.	A11	Present address
12.	A12	Property ownership
13.	A13	Age (in years)
14.	A14	Other installment plans
15.	A15	Housing
16.	A16	Number of existing credits in this bank
17.	A17	Employment type
18.	A18	Number of dependents
19.	A19	Telephone
20.	A20	Foreign worker

4.3 Measure of Input Variables on the Network Response

As stated earlier, a RBF-ANN network is defined by:

$$W(x) = \sum_{i=1}^h \omega_i \varphi_i(x) \quad (4.2)$$

Where w_0 denote the intercept, $w_i = (w_1, w_2, \dots, w_h)$ the vectors consisting of all synaptic weight without the intercept, and $x = (x_1, x_2, \dots, x_n)$ the vector of all independents variable (Frank et al., 2010). After experimenting with the equation (4.1) on the data set, it is necessary to measure the contribution of each independent (input) variable and the best **performing variables** on the response variable. The Garson Algorithm (1991), (relative importance of these independents variables) was a better technique to adopt. The following 2-layer RBF-

ANN network results, i.e., a typical 20 : 1 : 1 network layer (Figure 4.1) shows the ranks of the input attributes that significantly contribute to the network response. See Appendix B1 for the *R* codes.

Quite often, large number of data objects also affects the accuracy of the classification because of outliers and noises as the outliers will be interpolated by some of these initial nodes during training; similarly, the number of non-relevant attributes could degrade the performance.

Here is the summary of the procedure for obtaining the relevant variables using a probe feature:

1. Choose the set of candidate inputs (primary and secondary variables).
2. Select the input that is most correlated to the output using observation space, and a designed algorithm (Garson, 1991).
3. Iterate for more relevant features for selection. Until the probability of the probe feature variable is higher.

Use the selected variables as inputs of a neural network, then train and test the model.

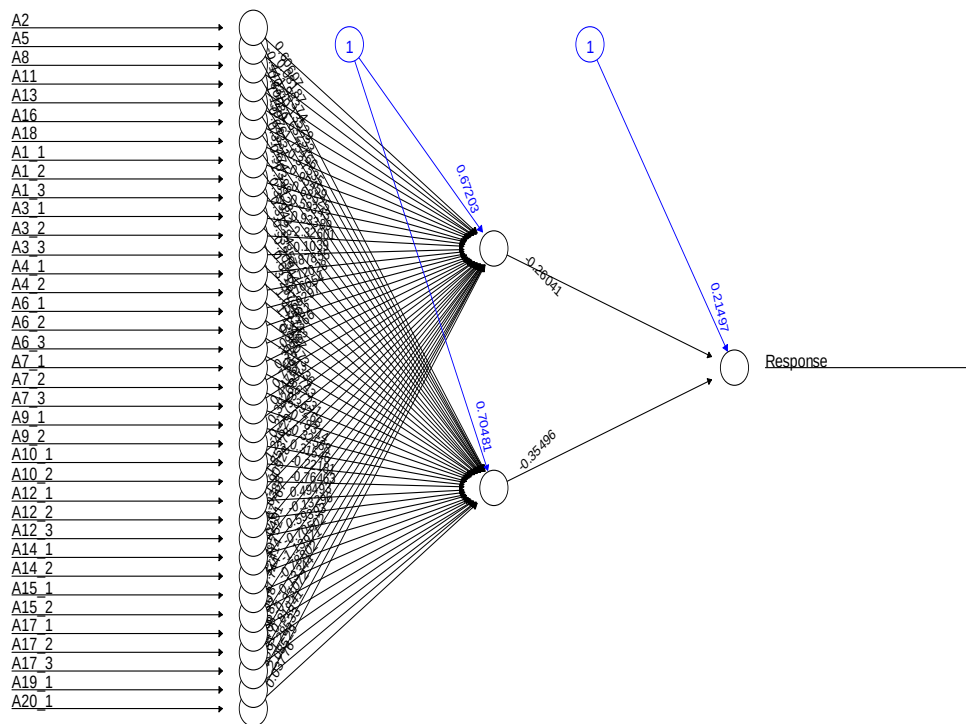


Figure 4.1: A 1-Layer RBF Network

Table 4.2 shows the input variables performance chart, i.e. the input variables that positively and negatively affect the network's response. For instance, Credit history (Variable 3), Employment type (Variable 17), and provision of loan Guarantor are strong indicators of positive response; that is, if any of these variables are not provided in any instance, the network issues a red flag (an indicator for fraud transaction).

On the other hand, Account status (variable A1), Employment status (variable 7) and Instalment rate (variable 8) were not strong enough to pre-empt fraud (see Figure 4.2).

Table 4.2: Input Variables Importance using Garson Algorithm (1991)

S/N	Independent Variable		Relative importance	Remarks on the Transaction Status
1	Duration in Month (A2)		0.014	Very weak positive relationship
2	Amount of Credit(A5)		0.0000	No relationship
3	Installment Rate (%) (A8)		0.0917	Very weak positive relationship
4	Present Residence in Years (A11)		0.0294	Very weak positive relationship
5	Age in Years (A13)		-0.0003	Very weak negative relationship
6	Number of Existing Credit (A16)		-0.0488	Very weak negative relationship
7	People Liable for Maintenance (A18)		0.0894	Very weak positive relationship
8	Status of existing Account	< 0DM (Ref., A1_1)	-0.1900	Weak negative relationship
		0 ≤ ... ≤ 200DM (A1_2)	-0.6001*	Intermediate negative relationship
		No checking Acc. (A1_3)	-0.7571*	Strong negative relationship
9	Credit History	No credit taken (A3_1)	0.3966*	Substantial positive relationship
		All credit payback duly (A3_2)	0.8276*	Strong positive relationship
		Delay in paying back (Ref.,A3_3)	0.3321	Substantial positive relationship
10	Purse of the credit	Tangible asset (A4_1)	0.0330	Very weak positive relationship
		Non-tangible (Ref., A4_2)	-0.0349	Very weak negative relationship
11	Saving Acc./Bonds	< 100DM (Ref., A6_1)	-0.2314	Very weak negative relationship
		100 ≤ ... ≤ 1000DM (A6_2)	-1.000	Perfect negative relationship
		No saving Acc. (A6_3)	-0.4297	Substantial negative relationship
12	Present Employment	Unemployed (Ref., A7_1)	-0.1354	Weak negative relationship
		1 ≤ ... < 4 years (A7_2)	-0.048	Very weak negative relationship
		4 ≤ ... ≤ 7 years (A7_3)	-0.2446	Weak negative relationship
13	Personal Status and sex	Sex (M=1, F=0) (A9_1)	0.0767	Very weak positive relationship
		Marital status (M/D/S/W=1, S=0) (A9_2)	0.3024	Substantial positive relationship
14	Guarantors	None (Ref., A10_1)	0.3842	Substantial positive relationship
		Co-applicant/guarantors (A10_2)	-0.8577	Strong negative relationship
15	Property	Real Estate (A12_1)	-0.3387	Substantial Negative relationship
		Car & others (A12_2)	-0.2895	Substantial Negative relationship
		No property (Ref., A12_3)	-0.2708	Substantial Negative relationship
16	Installment Plans	Bank/Store (A14_1)	0.1083	Weak positive relationship
		None (Ref., A14_2)	0.2322	Substantial positive relationship
17	Housing	Rent (Ref., A15_1)	-0.1995	Weak negative relationship
		Own/free (A15_2)	-0.3006	Substantial negative relationship
18	Job	Unskilled (Ref., A17_1)	-0.1162	Weak negative relationship
		Skilled (A17_2)	-0.0627	Very weak negative relationship
		Self-employed (A17_3)	-0.0509	Very weak negative relationship
19	Telephone	(None=0, Yes=1, A19_1)	-0.1607	Weak negative relationship
20	Foreign Worker	(Yes=1, No=0, A20_1)	0.4900*	Substantial positive relationship

The graphical representation of the variables importance rank is depicted in Figure 4.2.

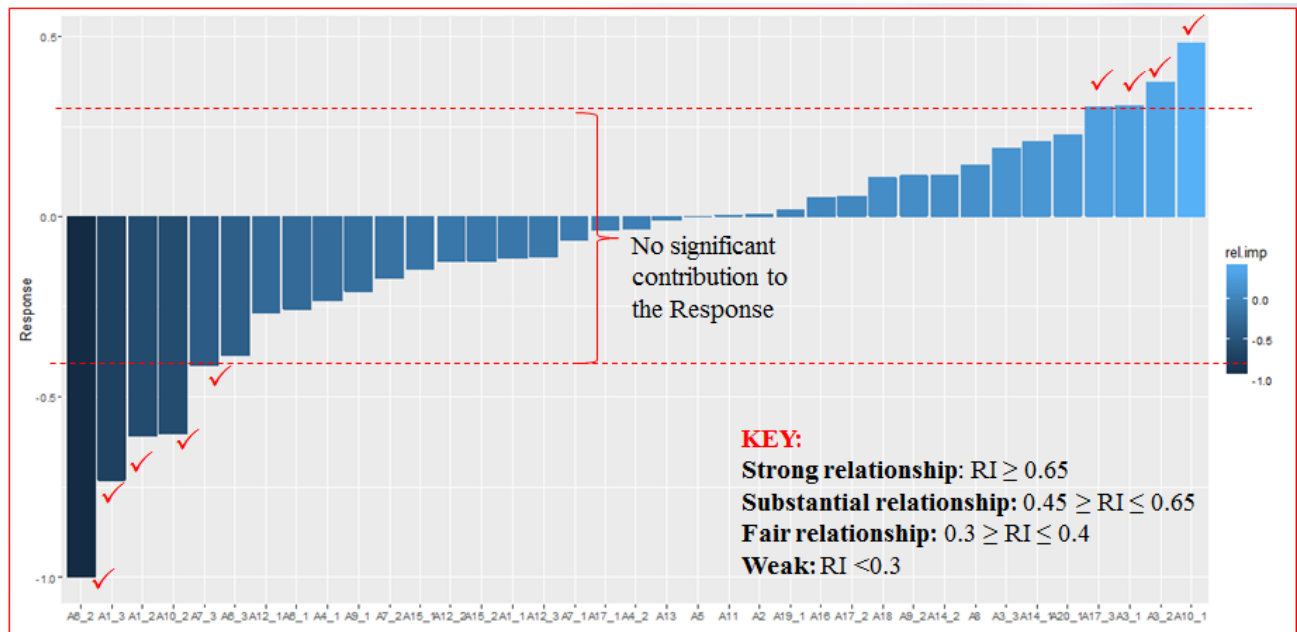


Figure 4.2: Overall Attributes Performance Chart

The distribution of the relationship measures is shown in Figure 4.2. The figure shows the ranking of the variables in order of importance relative to the network's output. The most important variables were ticked as variables that have positive relationship with the response variable; it also shows the variants that have negative relationship with the response variable. It also shows the inputs that do not significantly contribute to the response variable. Hence, judicious attributes selection is imperative in order to determine minimal attributes worthy of selection that are necessary and sufficient to detect suspicious transactions, while the candidate variables that are less relevant than the probe feature are discarded.

Table 4.3: Key Input Variables Based on Ranked Variable Importance Values

S. No.	Variables	Description	Variable Importance Value
1.	A10_1	Guarantors: None	0.3842
2.	A3_2	Credit History: All Credit Payback Duly	0.8276
3.	A3_1	Credit History: No credit Taken	0.3966
4.	A17_3	Job: Self-employed	-0.0509
5.	A20_1	Foreign Worker	0.4900
6.	A6_2	Saving Account/Bonds (btw 100 and 1000 DM)	-1.0000
7.	A1_3	No checking Account	-0.7571
8.	A1_2	Status of Existing Account : $0 \leq \dots \leq 200DM$	-0.6001
9.	A10_2	Guarantors : Co-applicant/guarantors	-0.8577
10.	A7_3	Present Employment: $4 \leq \dots \leq 7 \text{ years}$	-0.2446
11.	A6_3	No saving Account	-0.4297

4.4 Input Normalization

Prior to training, the input variables must be normalized and centred: if the inputs have very different orders of magnitude, the smallest ones will not be taken into account during training.

See the normalized dataset extracts in Appendix A2.

4.5 RBF Network Requirement

As mentioned in earlier, building a functional RBF network involves the following processes:

1. Find and select the relevant candidate inputs i.e., the input attribute features that are significant predictive factors of the model output,
2. Collect the data that is necessary for training and testing the network,
3. Determine the appropriate complexity of the model, i.e., the appropriate model parameters,
4. Estimate the parameters for which the cost function is minimum, i.e., training the network,
5. Evaluate or assess the generalization ability of the neural network after training.
6. Iterate processes (1) – (5) until an optimum model is obtained.

Noteworthy is the fact that data collection is core in training the desired model as well as the test and validation processes. Towards this, the data set must be numerous and sufficient enough to enable the model learn adequately. Also, too many parameters should be avoided to prevent over-fitting of the model; too few parameters will hinder the generalization ability of the model, as it might not have learn enough.

4.6 Model Training

Succinctly, model training entails a numerical optimization process, and for better

understanding, let us consider a simple linear model with an output $g(x, w)$ with say, N sample training set as:

$$J(w) = \frac{1}{2} \sum_{k=1}^N [(y_p(x^k) - g(x^k, w))]^2 \quad (4.3)$$

Where x^k is the input variables vector for instance k ,

$y_p(x^k)$ is the corresponding measured (target or actual) value of the model,

w is the weight vector, and

$g(x, w)$ is the model output.

The simple and efficient ordinary least squares methods can be used in such scenario, but for a feed-forward model as radial basis function with adjustable centres and non-linear with respect to its parameters, the optimization problem is multivariable non-linear, and ordinary least squares is not applicable to estimate its parameters.

Therefore the technique to be adopted in this case is to find an approximation of a function in the family of non-linear functions with respect to the parameters. Therefore, a set of Gaussian functions with adjustable centres is chosen, such that the complexity of the family of models is increased step-wise, by increasing the number of Gaussians, the number of hidden neurons, etc. as shown in Figure 4.5. Ultimately, one obtains the parameter vector w , for which $J(w)$ is the minimum by computing the values of the parameters w must be computed; this is done by

minimizing a cost function that mirrors the “distance” between the predictions of the model and the measured values.

Essentially, the estimation of the parameters $\mathbf{w}$ of the model $\mathcal{G}(\mathbf{x}, \mathbf{w})$, or the model’s training is accomplished using the following processes:

1. Select set of activation functions with which the model is sought (RBF in this case),
2. Train one or several models within the chosen activation function,
3. Select best model by the estimation of its performances, and
4. if that best model is not satisfactory, vary the functions parameters (e.g. the number of hidden neurons, the degree of the polynomial, etc.), and the process is iterated to step (2).

4.7 Parameters Initialization

For RBF networks, the parameters initialization is crucial, as they are localized functions; if they are initially located far from the domain of interest, or if their extension (standard deviation or dilation) is not appropriate, training will generally fail (Oussar & Dreyfus, 2002). The parameters related to the bias inputs must be initialized to zero, in order to ascertain that the function of the hidden neurons are initialized around zero, so that training starts very slowly.

4.7.1 Model Training Optimization

The adopted optimization techniques are based on the iterative computations at each iteration run, the gradient of the cost function with respect to the parameters of the model. The gradient thus computed is subsequently used for updating the values of the parameters found at the previous iteration.

As stated in section 4.2, the German credit dataset used for experimentation in this study consists of 1000 records of bank customers on whose 20 features attributes (factors) were normalized (See Appendix A2). The dataset was randomly split as the training set (85%) consists of 850 samples of which 150 samples (15%) were used for the test/validation set using ten (10) different RBF iterations, $r \in \{100, 200, 300, 400, 500, 600, 1000, 1500, 2000, 3000\}$. At each RBF iteration, r , twenty (20) RBF models were trained using the training data over a set of twenty (20) different hidden layers, H as shown in Table 4.4, i.e., $H = \{10, 25, 50, \dots, 860\}$.

Table 4.4: Table of Average Misclassification (Trained) Error Rate (MER) obtained from several RBF neural network runs at different hidden layers and at different RBF iterations. The asterisk (*) indicated the average MER of the best RBF model (i.e., with the least MER) at each RBF iteration run. The best model here is the RBF model with 840 Hidden nodes and 300 iterations and higher.

		Average Misclassification (Train) Error Rates (MERs) in percentage										
	Hidden				Number of rbf Iterations to learn, (r)							Average
S.No.	Nodes (H)	100	200	300	400	500	600	1000	1500	2000	3000	MER (%)
1	10	68.94	68.94	68.94	69.18	69.18	69.18	69.18	69.18	69.18	69.18	69.11
2	25	68.00	68.00	68.12	68.12	68.12	68.12	68.12	68.12	68.47	68.00	68.12
3	50	64.83	65.06	65.29	65.53	65.53	65.53	65.65	65.65	65.88	65.29	65.42
4	80	62.12	62.82	62.94	63.18	63.18	63.41	63.41	64.00	64.00	63.88	63.29
5	100	60.82	61.88	62.35	62.59	62.47	62.59	62.24	62.24	62.24	62.00	62.14
6	150	55.77	56.94	57.53	57.53	57.65	57.77	58.24	57.88	57.65	57.41	57.44
7	250	47.88	49.53	50.53	50.82	50.94	50.59	50.12	50.12	50.35	50.35	50.12
8	450	32.59	38.00	34.47	34.94	34.94	34.71	34.82	36.00	36.00	35.77	35.22
9	650	19.29	21.18	21.06	20.94	21.06	20.35	19.53	19.18	19.65	19.88	20.21
10	750	12.12	13.88	14.59	14.59	14.59	14.00	12.00	12.24	11.41	11.06	13.15
11	800	8.82	10.71	11.41	11.29	11.29	11.06	9.76	9.53	9.53	9.41	10.28
12	820	8.00	9.06	9.18	9.77	9.77	9.53	8.94	8.59	8.71	7.88	8.94
13	830	7.06	8.24	8.94	9.06	9.41	9.18	9.06	7.65	7.88	7.65	8.41
14	840	6.24	*7.29	*8	*8.24	*8.35	*8	*7.18	7.18	7.18	*6.47	7.41
15	842	*6.12	6.82	8.24	8.47	8.94	8.35	8.00	7.65	6.82	6.71	7.78
16	843	*6.12	*7.29	8.12	8.59	8.47	8.59	8.12	*6.94	*6.71	7.53	7.65
17	845	6.71	7.65	8.47	9.06	8.82	9.41	8.12	7.76	8.12	8.24	8.24
18	850	69.41	69.65	30.24	69.29	30.24	30.24	69.77	51.18	51.18	69.77	54.10
19	855	66.82	66.71	65.06	55.41	69.06	30.24	69.77	69.76	63.77	69.77	62.64
20	860	67.41	45.06	69.77	69.77	59.65	57.18	69.65	58.35	69.65	69.77	63.63

For each model, the class label of the unseen 15% test data was predicted using the fitted RBF network and the performance of the fitted model was assessed using the Average Misclassification Error Rate (MER) stated in equation (3.18) (see Table 4.5).

Table 4.5: Table of Average Misclassification (Test) Error Rate (MER) obtained from several RBF neural network runs at different hidden layers and at different RBF iterations. The asterisk (*) indicated the average MER of the best RBF model (i.e., with the least MER) at each RBF iteration run. The best model here is the RBF model with 860 Hidden nodes and 300 iterations and higher

S.No.	Hidden	Average Misclassification Error (Test) Rates in percentage										Average MER (%)
	Nodes (H)	100	200	300	400	500	600	1000	1500	2000	3000	
1	10	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33
2	25	70.67	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.26
3	50	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	70.67	71.26
4	80	66.67	68.67	70.67	71.33	71.33	71.33	71.33	71.33	70.00	69.33	70.20
5	100	68.67	69.33	70.67	71.33	71.33	71.33	70.67	70.67	70.67	70.00	70.47
6	150	69.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	71.33	70.67	71.06
7	250	68.00	70.67	70.67	70.67	70.67	70.67	70.67	70.67	70.00	70.67	70.34
8	450	68.00	70.00	70.00	70.67	70.00	70.67	70.00	70.67	70.00	*68	69.80
9	650	69.33	69.33	68.67	69.33	69.33	69.33	69.33	*69.33	69.33	68.67	69.20
10	750	70.00	68.67	68.67	*68.67	69.33	68.67	68.67	*69.33	69.33	69.33	69.08
11	800	67.33	68.00	68.67	69.33	*68	68.67	68.67	68.67	68.67	68.67	68.47
12	820	68.37	68.00	68.00	*68.67	69.33	68.67	69.33	68.67	69.33	69.33	68.78
13	830	68.67	69.33	68.67	68.67	68.67	68.67	69.33	68.67	*68.67	68.67	68.82
14	840	70.00	69.33	70.00	70.00	70.00	70.00	70.00	70.00	70.00	69.33	69.87
15	842	70.00	69.33	69.33	69.33	69.33	69.33	69.33	69.33	69.33	69.33	69.40
16	843	70.00	69.33	69.33	69.33	69.33	70.00	70.00	70.00	70.00	70.00	69.73
17	845	*66.67	68.00	68.67	*68.67	68.67	69.33	69.33	69.33	69.33	69.33	69.00
18	850	71.33	69.33	71.33	70.00	71.33	71.33	*54	71.33	71.33	71.33	70.96
19	855	71.33	66.00	66.67	70.67	71.33	68.67	68.00	71.33	71.33	71.33	69.67
20	860	59.33	72.00	70.67	71.33	71.33	*32.67	71.33	71.33	71.33	71.33	66.27

Consequently, the model with the best average predictive accuracy (i.e., with the least average MER) among the several fitted models are visible at each iteration as indicated by the asterisk (*), such a model becomes the best model at that iteration and the corresponding hidden layer becomes the optimum hidden layer at that run.

Aside from the misclassification error rate, other performance metrics used to assess the model's performance are: Sensitivity, Specificity, Receiver Operating Characteristics (ROC) and the estimated area under the ROC curve. Due to space problems, few of these were shown in Figure 4.4, especially of the final model.

4.7.2 Training Model Outcome

This section presents the result of the predictive performance of the various RBF classification models fitted in the previous section. Table 4.4 shows the results of all the fitted RBF networks at twenty different hidden layers, H and at different RBF iterations, r . The figures in the table indicate the average test sample predictions error rates of a total of 850 bank customers credit transactions constructed on the German bank credit dataset.

At each RBF iteration, r , the average MER of the network at different hidden layers H was reported and the average MER of the best model was asterisked. Hence, from the results presented in Table 4.1, the hidden layer that produced the best RBF model can be observed at each iteration. For instance, at 100, 400, 600 and 1500 iterations, the best models were those that have 842, 840, 840 and 843 hidden layers respectively. The corresponding misclassification error rates are: 6.12%, 8.24%, 8% and 6.9% respectively, which translates to an average of about 93% degree of accuracy for the four models.

Worthy of note is the outcome of the Confusion matrix, whereby the False Positive (FP) entry returned a value of zero (0), whereas the False Negative (FN) entry returned a value of sixty-one (61) (see Figure 4.3), signifying that the model predicted that there was no fraud and actually there was no fraud incident; the model would rather predict there were frauds where there was none than to allow a fraud incident to pass undetected. This is a very salient attraction of a desirable fraud detection system, whereby the false alarm (of 61 cases) is allowed but having zero-tolerance for fraud incidents is desirable. The sixty-one (61) suspected fraud cases in this situation, some of which are misclassified cases, but could be subjected to further scrutiny.

It was also observed in Table 4.4 that the computations converge easily from 200 iterations and the network become stable, specifically the best hidden layer, 840 yielded about 93% accuracy at 200 RBF iterations. Also, 840 hidden layers yielded the least average MERs of 7.29%, 8%, 8.24%, 8.35%, 8%, 7.18% and 6.47% at 200, 300, 400, 500, 600, 1000 and 3000 RBF iterations respectively. This is affirming the supremacy of the 840 hidden layer at 200 iterations or higher among other hidden layers band. This is the appropriate RBF model for the trained data using 200 iterations or higher. However, to conserve computation time, 200 RBF iterations is recommended for the efficient construction of this model since similar prediction outcome would be obtained at higher RBF iterations.

The execution of this implementation on an AMD Turion X2 Dual-Core CPU, 2.3 GHz, 32-bit on Windows 10 Pro platform takes about 180 seconds, this could be mitigated by higher processing power systems.

Next is the Confusion Matrix of the selected model, which shows the number of times the RBF network accurately/erroneously classified a transaction of class 1 (fraud) to be a member of class 0 (no fraud). See Figure 4.3 that follows:

		Predicted	
		0 (No Fraud)	1 (Fraud)
Target	0 (No Fraud)	532 (TP = a)	61 (FN = (b))
	1 (Fraud)	0 (FP = c)	257 (TN = d)

Figure 4.3: A Confusion Matrix

With the Confusion matrix, a number of other performance indices (See Table 4.6) can be computed. For brevity to suit the purpose of this study, the following performance metrics were calculated as follows:

$$\text{Prediction Accuracy} = \frac{a+d}{a+b+c+d}$$

$$\text{Misclassification Error} = \frac{b+c}{a+b+c+d}$$

$$\text{Sensitivity} = \frac{a}{a+b}$$

$$\text{Specificity} = \frac{d}{d+c}$$

$$\text{Negative Predictive Value} = \frac{d}{d+b}$$

$$\text{Precision} = \frac{a}{a+c}$$

There are other classification performance metrics; it was believed that these few metrics should suffice to justify the formulated classification model to detect fraud practices in financial transactions as applied to the German bank credit data.

Table 4.6: Estimates of Various Performance Indices of the Optimum RBF Model Selected for the Bank Credit Transactions

Performance Indices of the Best Trained RBF Model	Estimated Values (in %)
Misclassification Error Rate	7.18
Prediction Accuracy	92.82
Sensitivity	89.71
Specificity	100
Area under the ROC Curve (AUC)	98.6

In addition to the performance indices given above, many other graphical (visual) and analytic methods are possible. For example, the figure 4.4a is an Iterative error plot that indicates the summed squared error (SSE), i.e. the sum of the squared errors of all patterns for every epoch (iteration). This is indicative of the quality of the regression. It has target values on the x-axis and fitted/predicted values on the y-axis. A test set is provided, its SSE is also shown in the plot, normalized by dividing the SSE through the test set ratio (which is the amount of patterns in the test set divided by the amount of patterns in the training set).

The Figure 4.4b is a regression plot which illustrates the quality of the regression. It has target values on the x-axis and fitted/predicted values on the y-axis. The optimal fit would yield a line through zero with gradient one. This optimal line is shown, as well as a linear fit to the actual data.

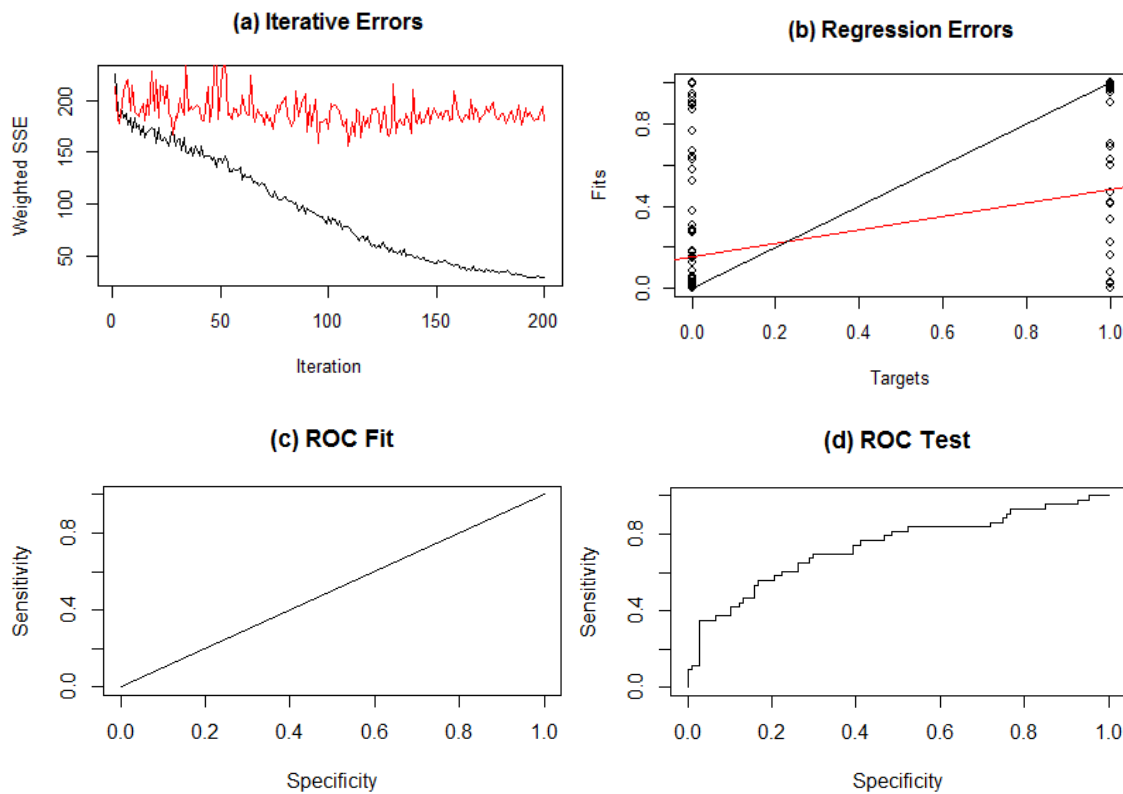


Figure 4.4: Plots of Radial Basis Function Trained with the German Credit Dataset. (a) The Iterative Error Plot of both Training (Black) and Test (Red) Error. (b) The Regression Plot for the Test Data. As a Classification is Performed, Ideally Only the Points (0,0) and (1,1) was Populated. (c) ROC Plot for Specificity against Sensitivity, on the Training Dataset. (d) Same as (c), but for the Test Data

As much as iterative and regression error plots are frequently used for classification problem analysis, nonetheless, the regression error plot is parsimonious with information than for a regression problem. Therefore, a function for displaying receiver operating characteristics (ROC) is shown (Figure 4.4c. ROC plots are strikingly appropriate for the analysis of binary classification problems with two classes as the issue at hand (fraud detection). The estimated area under the ROC curve is 0.9860, which is almost 100% coverage of the area under the ROC curve. This is giving more credence to the chosen model as a veritable classification for the dataset under consideration.

4.8 The RBF Network Architecture

The following snippet shows the extract of the RBF model characteristics. This summarizes the entire features (parameters) of the constructed model, such as the number of hidden nodes generated (878); the number of connections (31920); the weights associated with each hidden node; the activation function (radial basis function), etc.

```
SNNS network definition file V1.4-3D
generated at Wed May 10 20:28:04 2017
```

```
network name : RSNNSUntitled
source files :
no. of units : 878
no. of connections : 31920
no. of unit types : 0
no. of site types : 0
```

```
learning function : RadialBasisLearning
update function   : Topological_Order
```

```
unit default section :
```

act	bias	st	subnet	layer	act func	out func
0.00000	0.00000	i	0	1	Act_Logistic	Out_Identity

no.	typeName	unitName	act	bias	st	position	act func	out func	sites
1		Input_1	1.99928	0.00000	i	1, 0, 0	Act_Identity		
2		Input_2	0.49146	0.00000	i	2, 0, 0	Act_Identity		
3		Input_3	0.03237	0.00000	i	3, 0, 0	Act_Identity		
4		Input_4	1.04039	0.00000	i	4, 0, 0	Act_Identity		
5		Input_5	-0.73814	0.00000	i	5, 0, 0	Act_Identity		
33		Input_33	-0.49787	0.00000	i	33, 0, 0	Act_Identity		
34		Input_34	0.75723	0.00000	i	34, 0, 0	Act_Identity		
35		Input_35	-0.40913	0.00000	i	35, 0, 0	Act_Identity		
36		Input_36	-0.81003	0.00000	i	36, 0, 0	Act_Identity		
37		Input_37	0.18445	0.00000	i	37, 0, 0	Act_Identity		
38		Hidden_2_1	0.00000	0.20000	h	1, 2, 0	Act_RBF_Gaussian		
39		Hidden_2_2	0.00001	0.20000	h	2, 2, 0	Act_RBF_Gaussian		
40		Hidden_2_3	0.00000	0.20000	h	3, 2, 0	Act_RBF_Gaussian		
41		Hidden_2_4	0.00046	0.20000	h	4, 2, 0	Act_RBF_Gaussian		
42		Hidden_2_5	0.00000	0.20000	h	5, 2, 0	Act_RBF_Gaussian		
43		Hidden_2_6	0.00000	0.20000	h	6, 2, 0	Act_RBF_Gaussian		
44		Hidden_2_7	0.00004	0.20000	h	7, 2, 0	Act_RBF_Gaussian		
45		Hidden_2_8	0.00000	0.20000	h	8, 2, 0	Act_RBF_Gaussian		
46		Hidden_2_9	0.00026	0.20000	h	9, 2, 0	Act_RBF_Gaussian		
47		Hidden_2_10	0.00000	0.20000	h	10, 2, 0	Act_RBF_Gaussian		
48		Hidden_2_11	0.00003	0.20000	h	11, 2, 0	Act_RBF_Gaussian		
49		Hidden_2_12	0.00000	0.20000	h	12, 2, 0	Act_RBF_Gaussian		
50		Hidden_2_13	0.00002	0.20000	h	13, 2, 0	Act_RBF_Gaussian		

connection definition section :

target	site	source:weight									
38		37: 0.19663,	36:-0.69037,	35:-0.40654,	34:-1.24607,	33: 1.93498,	32:-0.35442,	31: 0.66158,	30: 4.85060,		
		29:-0.41699,	28:-0.66426,	27:-0.55182,	26: 1.51166,	25:-0.21278,	24:-0.18107,	23:-0.98154,	22: 0.41106,		
		21: 1.16611,	20:-0.76423,	19:-0.47816,	18:-0.47532,	17:-0.39188,	16:-0.34343,	15:-0.25885,	14: 0.44986,		
		13: 0.94864,	12:-0.21951,	11:-0.21401,	10:-0.77189,	9:-0.27166,	8:-0.65266,	7:-0.42202,	6:-0.68042,		
		5: 1.88517,	4: 0.94641,	3: 0.90351,	2:-0.93034,	1:-0.70803					
39		37: 0.18892,	36:-1.06493,	35:-0.42195,	34: 0.74168,	33:-0.46897,	32:-0.35865,	31: 0.64005,	30:-0.22393,		
		29:-0.39558,	28:-0.68619,	27:-0.37818,	26: 1.43634,	25:-0.23431,	24:-0.19715,	23: 1.16932,	22:-2.21582,		
		21:-0.87001,	20:-0.82439,	19: 2.34379,	18:-0.46798,	17:-0.39639,	16:-0.34049,	15:-0.25537,	14: 0.47958,		
		13: 1.03039,	12:-0.20949,	11:-0.22590,	10: 1.20146,	9:-0.27532,	8:-0.58245,	7:-0.43379,	6:-0.75695,		
		5:-0.81541,	4:-0.63529,	3: 1.00342,	2:-0.64776,	1: 0.39686					
40		37: 0.18199,	36:-0.64613,	35:-0.39728,	34: 0.74997,	33:-0.49707,	32:-0.44544,	31:-1.67980,	30:-0.21243,		
		29:-0.41268,	28: 1.48779,	27:-0.49496,	26:-0.66113,	25:-0.20837,	24:-0.19919,	23:-0.81342,	22: 0.40722,		
		21:-0.77041,	20: 1.43326,	19:-0.42111,	18: 2.20853,	17:-0.34500,	16:-0.34236,	15:-0.28815,	14: 0.45860,		
		13: 0.96486,	12:-0.22627,	11:-0.20250,	10: 1.28238,	9:-0.27640,	8:-0.61716,	7:-0.41608,	6: 1.03496,		
		5: 2.70619,	4: 1.06729,	3:-0.85294,	2: 1.35598,	1:-0.21681					
41		37: 0.19363,	36: 1.34511,	35:-0.39860,	34: 0.75012,	33:-0.50096,	32:-0.35267,	31: 0.66098,	30:-0.21966,		
		29:-0.37518,	28: 1.43435,	27:-0.53489,	26:-0.66141,	25:-0.24365,	24:-0.19410,	23:-1.09445,	22: 0.39891,		
		21: 1.16724,	20:-0.73646,	19:-0.44264,	18: 2.30278,	17:-0.37491,	16:-0.34763,	15:-0.25861,	14: 0.46819,		
		13: 1.01894,	12:-0.22530,	11:-0.22284,	10:-0.85762,	9:-0.26906,	8: 1.60835,	7:-0.44847,	6:-0.74510,		
		5:-0.19200,	4:-0.11237,	3:-0.96468,	2: 0.24485,	1:-0.04975					
42		37: 0.18651,	36: 1.19247,	35:-0.41280,	34: 0.81287,	33:-0.50573,	32:-0.34923,	31: 0.63594,	30:-0.21494,		
		29: 2.62474,	28:-0.71749,	27: 1.90502,	26:-0.62612,	25:-0.25442,	24:-0.18600,	23:-0.91818,	22: 0.41611,		
		21: 1.13550,	20:-0.69491,	19:-0.44871,	18:-0.54158,	17:-0.35212,	16:-0.34829,	15:-0.29909,	14: 0.49951,		
		13: 0.87150,	12:-0.21829,	11:-0.21125,	10: 1.17649,	9:-0.26239,	8:-0.67086,	7: 2.39688,	6: 2.67085,		

The network summary:

```

Class: rbf->rsnns
Number of inputs: 37
Number of outputs: 1
Maximal iterations: 200
Initialization function: RBF_Weights
Initialization function parameters: -4 4 0 0.2 0.04
Learning function: RadialBasisLearning
Learning function parameters: 0.001 0 0.001 0.1 0.8
Update function:Topological_Order
Update function parameters: 0
Patterns are shuffled internally: TRUE
Compute error in every iteration: TRUE
Architecture Parameters:
$size
[1] 840

```

All members of model:

```

[1] "nInputs"           "maxit"              "initFunc"
[4] "initFuncParams"    "learnFunc"          "learnFuncParams"
[7] "updateFunc"        "updateFuncParams"   "shufflePatterns"
[10] "computeIterativeError" "snnsObject"         "archParams"
[13] "IterativeFitError"  "IterativeTestError" "fitted.values"
[16] "fittedTestValues"   "nOutputs"

```

Figure 4.5: RBF Network Parameters and Architecture Summary

4.9 Results Overview

The result indicates that the number of normal (no fraud) 532 transactions are more than (abnormal) fraudulent 257 transactions. Sixty-one transactions were predicted to be fraudulent, but were actually not all fraudulent. Strikingly, zero (0) transactions were predicted to be normal in situations where they were indeed fraudulent. This is a desirable feature of this model, whereas some of those 61 transactions that were flagged to be fraudulent could be subjected to further analysis and investigations to really determine their true class. This way, it is better to pend a transaction as fraudulent when not actually fraudulent, than to certify a fraudulent transaction as non-fraudulent.

The most popular model selection strategy (or stopping criterion) in this study was to classify objects and observe the variation of the standard prediction error on a validation set, and in terminating training when the prediction error starts to increase as was the case in table 4.5, where the misclassification error begins to increase after 3000 iterations.

Recall that the Training Error obtained by applying the model to the dataset by which it was trained, and the Test Error (the error obtained from new unseen transaction to the model) test data. The test error is actually how well the model would behave in practice (production) on future data the model had not known. More often than not, Training error always underestimates Test error drastically, especially in a complex model as this. This explain the divergence in the results of the Training misclassification error rate and Test misclassification error rate as shown in Tables 4.5 and 4.6 respectively. However, this disparity can be mitigated (minimized) using a Surrogate convex loss algorithm (Ben-David, Loker, Srebro & Sridharam 2012). However, this is not discussed in this study.

Also, as most of the regression accuracy metrics are based on residual, logistic or classification metrics are based on error rate as illustrated in section 4.7. A list of classification accuracy metrics with varying nomenclature, e.g. Accuracy, Error Rate, Specificity, Sensitivity, TPR, FPR, PPV (Positive Predictive Value, False Discovery Rate, Likelihood Ratios, Negative Predictive Value, False Omission Rate, etc.). In some literatures, all these metrics could be grouped into Column ratios or Row ratios. For this study, MER, ROC, Sensitivity and Specificity metrics were used and deemed sufficient.

4.10 Comparison with Other Classifiers

There are several neural network methods, amongst which are: multi-layer perceptron, radial basis function, fuzzy logic, genetic algorithm, hidden Markov, random forest, self-organizing map, etc. However, for brevity only three of these techniques were selected for comparison with RBF.

Three Neural network (NN) variants, viz: Multi-layer perceptron back-propagation (MLP-BP or MLP for short), Dynamic decay adjustment (DDA) and General nonlinear regression (GNLR) were adopted to benchmark the results of RBF discussed in section 4.8. These neural networks techniques were selected for the comparisons because these classifiers are all suitable for non-parametric prediction among other common features. For instance, back-propagation networks are widely used by researchers and are fully connected. GNLR (General nonlinear regression) non-linear regression functions can reduce computation complexity and time as it uses a single smoothing factor to optimize the transfer function for all nodes in a one pass training through the network (Fung, Iyer, Brown & Wong, 2005); DDA is similar to RBF network with dynamic decay adjustment (DDA) algorithm, the network is essentially used for classification.

The same dataset, *German bank credit data*, was applied to all the algorithms, in order to compare the performance of the neural network techniques, training set 85% and 15% testing set was also retained. For the purpose of experimentation, MLP, RBF and DDA are available in the RSNNS package, while *gnm* (Generalised Non-linear Model) package was used for the generalised non-linear regression.

**Table 4.7: Comparison of Classification Accuracies (%) (Training)
of Four Neural Network Techniques. The Asterisk (*)
Indicates the accuracy of the Best Model for each Technique**

No. of Runs	NEURAL NETWORK TECHNIQUES			
	GNLR	MLP	DDA	RBF
1	74.68	77.27	74.19	77.92
2	79.87	76.62	75.22	76.62
3	70.13	70.13	72.03	74.03
4	75.97	70.78	79.22*	85.71*
5	81.82*	75.97	77.27	75.32
6	70.13	70.78	72.08	72.08
7	72.73	83.97*	76.62	75.97
8	78.57	79.22	77.27	79.22
9	74.68	74.68	76.62	75.32
10	74.03	75.97	74.03	76.62
Av. Accuracy	74.53	74.60	75.04	75.90

Table 4.7 shows the performance of the different neural network techniques trained over ten iterations; twenty (20) neural network models were trained using the training data, the average of the ten results of the classification accuracy was used to benchmark the performance of these neural networks. The classification accuracies obtained by these four neural network techniques were compared, with the best accuracy for each technique indicated with an asterisk (*).

The comparative results show that the average accuracies obtained by each neural network technique are somehow similar. The highest average accuracy score obtained was from RBF (75.90%), followed by DDA (75.04%) as expected, while MLP (74.60%) and GNLR

(74.53%) were the least. Also, it was noted that RBF and DDA attained the highest accuracy score of (85.71%) and (79.22%) respectively, at iteration 4, while MLP and GNLR attained their highest accuracy at iterations 7 and 5 respectively.

This outcome is expected as conformed by literatures; for instance, RBF has the simplest architecture and thus trained faster, whereas GNLR and MLP require higher attributes, susceptible to noise in order to perform better with a resultant lower computational speed.



Fig. 4.6: Comparison of Neural Network Classification Techniques (Training)

The graphical representation of the four neural network techniques comparison is shown Fig. 4.6. It is observed that out of the four classification techniques, RBF has the highest accuracy obtained at a least iteration (4 out of 10) because of its simple structure and fastest computation. Expectedly, next in performance is the DDA technique because of its structural affinity to RBF. Generally, the results of RBF outperforms the other three techniques in so far as binary classification is concerned, especially with fewer attributes, 20 in this study and adequate enormous training data is provided. Obviously, RBF can provide better results in most binary classification cases as in fraud detection.

Thereafter, the test data was applied to the four algorithms, the following results were obtained:

Table 4.8: Comparison of Classification Accuracies (%) (Test) by Four Neural Network Techniques. The asterisk (*) Indicates the Accuracy of the Test Model for each Technique.

No. of Runs	Neural Network Techniques			
	GNLR	MLP	DDA	RBF
1	46.67	47.27	56.69	64.68
2	59.76	46.52	75.22	79.87
3	57.32	59.04	73.32	70.13
4	63.04	67.53	77.22	81.82
5	41.07	55.19	86.27*	85.97
6	57.36	72.63	72.08	87.13
7	60.13	79.67*	76.62	94.73*
8	71.42*	78.03	72.77	88.57
9	70.16	62.36	78.12	74.68
10	70.45	60.74	44.24	54.03
Av. Accuracy	60.61	61.03	69.59	76.32

The predictive performance of the different neural network techniques (GNLR, MLP, DDA and RBF) was tested with the 15% dataset on ten iterations. The average of the ten results of the classification accuracy is shown in Table 4.8. The classification accuracies obtained by these four neural network techniques were compared, with the best accuracy for each technique indicated with an asterisk (*).

The highest average accuracy score obtained was from RBF (94.73%), followed by DDA (86.27%) as expected, while MPL (79.67%) and GNLR (71.42%) were the least. Also, note that RBF and DDA attained their highest accuracy score at iterations 7 and 5 respectively, while MLP and GNLR attained their highest accuracy at iterations 7 and 8 respectively.

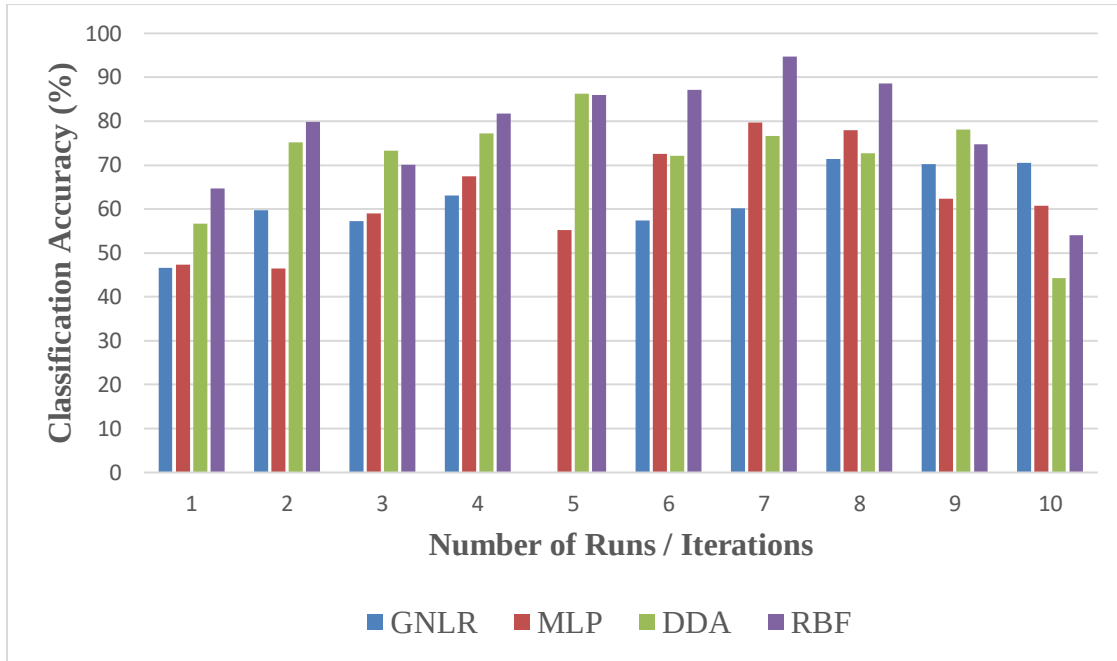


Fig. 4.7: Comparison of Neural Network Classification Techniques (Test)

The graphical representation of the four neural network techniques comparison is shown Fig. 4.7. It can be observed that out of the four classification techniques, RBF has the highest accuracy obtained at iteration 7, which tallies with Table 4.5, where hidden nodes = 850 at 1000 iterations, equivalent of the seventh iteration. Similarly, the next in performance to RBF technique is the DDA technique for reasons already explained. Likewise, the results of RBF outperforms the other three techniques, indicating that RBF suggests better prediction in most binary classification for previously unseen cases.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

This chapter presents the study summary and the drawn conclusions from the experimentation carried out as explained in Chapter four. The contribution to knowledge is outlined, while proposing recommendations on probable future work.

5.1 Conclusion

Fraud detection is applied once fraud prevention has failed. This study was solely concerned with fraud detection, given that fraud prevention could be vulnerable. Hence, fraud detection mechanism must evolve and be active continuously in a fraud detection system. Once fraudsters recall that a certain mode of fraudulent characteristics can be detected, they will re-shuffle their strategies and try some others.

RBF networks are feed-forward networks and their activation is not sigmoid as in Multi-layer Perceptron (MLP), but radially symmetric (often Gaussian). So, information is represented locally in the network (in contrast to MLP, where it is globally represented). Nonetheless, advantages of RBF networks in comparison to MLPs are mainly, that the networks are more interpretable, training is easier and faster, as better prediction becomes apparent from 200 epochs (see figure 4.4). Also, Initialization of an RBF network can be rather difficult and require prior knowledge. The initialization is performed in the current implementation and which heuristically vary the RBF weights by a call to successive call to InitFunc (RBF Weights). With different datasets, this parameter might be adjusted. Since we are dealing with a binary classification in this study, logistic output was chosen, as the dependent output, y -variable under consideration in this study is categorical, $y \in \{0,1\}$.

After designing a model framework outlined in section 3.2, several candidate models were created (trained) in section 4.7 and one was ultimately chosen (See section 4.7.2). As demonstrated, the model was fairly complex (Gaussian functions) to find the deterministic relations between the quantity to be modelled and the input attribute features that have a significant influence on its outcome (predictions). At the same time, a balance was struck to avoid over-fitting the model, such that the selected model enjoys a trade-off in its learning generalizing capability, i.e. if the model learns too much, it invariably incorporates noises and hence generalizes poorly. A bias was introduced to avoid this pitfall, see the model summary (figure 4.5) and ultimately the cost function was minimized.

As stated in section 1.4, financial frauds are abnormal activities which are generic with similar characteristics but distinct parameters. Notwithstanding the fact that a German bank credit dataset was used for this experiment, it is believed that the model would exhibit similar behaviour with Nigerian or other localized datasets.

The future work of this study should be able to make some comparisons of the level of frauds in other sub-sectors of the financial industry, such as Microfinance banks, Commercial banks and some specialized banks like Export/Import banks, Industrial and Agro-allied development banks where fraud propensity is assumed low.

Meanwhile, this report adhered to some data mining processes to harness data, pre-process it, then trained a RBF network model using a dataset from an online German bank credit data. Base models were created, with their *R* implementations, and in turn used in aggregation to build the required radial basis network model.

The input variables' importance was ascertained and the data summary shown. Also, it was shown how some of the input variables were distinguished as strong indicators or weak indicators of fraud. The dataset was split into 85:15 per-cent Training and Test data ratio. The

derived model's prediction shall be evaluated for prediction accuracy or misclassification error in the subsequent reports. However, further experiments were carried out to test the model's performance where the model was benchmarked against some three other neural network models. The researcher still believes that some more variables when explored further, could still be identified to positively/negatively affect the response of the network.

5.2 Contributions to Knowledge

Fraudulent propensity is innately inherent in every system globally. This research has adopted a computational approach to quickly detect fraudulent or abnormal transactions from its attribute features presentation. The model is expected to be incorporated as part of the front-end security checks to filter erroneous or outrightly fraudulent transactions and as an aid during forensic-based investigations or auditing.

Though the concept of detecting fraud is not new, many models and methods have been proposed. However, this study applied the use of Radial Basis Function, an interpolation approximation method to classify transactions into normal or otherwise, due to its fascinating features as stated in section 5.1. Also, see Appendix C for the list of publications.

This study would assist governments and private corporate organisations to proactively detect fraudulent business practices, especially now that the global economy is in recession, thereby minimizing the time and cost of fraud detection by adopting computationally efficient models rather than mere reliance on intuition, long manual investigations and/or windy auditors' findings.

5.3 Recommendations

It is recommended that RBF network is an efficient fraud detection model and its parameters must be adjusted with zero tolerance for fraudulent practices as a little threshold tolerance could imply a significant cost.

Similarly, a fraud detection model must be proactive rather than reactive with zero tolerance for fraudulent practises. A little threshold could imply a significant cost, i.e. if a fraud detection system permits, say 2-in-10 fraud chances, the permissive 2 chances could cost a fortune when a bad ‘guy’ aims and strikes appropriately.

The scan of all transactions, without exception for possible fraud detection and subsequent clearance is a continuous activity, and the process should be well positioned during financial operational processes and the processing/clearance must be processed swiftly, perhaps without notice by the customers by employing state-of-the art information technology devices as it is not desirable to describe fraud detection techniques in great details in the public domain, as this gives criminals the information that they require to evade detection.

Finally, since fraudulent practices are growing in sophistication due to economic contingencies, the researcher recommended that fraud detection processes should be regularly updated at fixed time intervals and fraud detection must be routinely active in a financial transactions system to checkmate dubious criminal tendencies.

5.3.1 Action Plan: What to Do

Fraud prone enterprises must implement procedures to identify, mitigate and manage fraud risks. Such procedures include:

- a) Protect key systems and establish protocols for urgent or confidential transactions;
- b) Use information technology (IT) filters to identify and block suspicious or illegitimate transactions;
- c) Institute anti-fraud standard operating procedure (SOP) to define and formalise the following;
 - i. Clearly outline the roles and responsibilities of staff;

- ii. Embark on fraud awareness training for staff;
- iii. Regularly conduct data and IT infrastructure assessments;
- iv. Routinely carry out fraud risk assessment;
- v. Clearly define reporting procedures and whistle-blower protection with independent hotlines;
- vi. Ensure specific investigation procedures are outlined by escalation protocols;
- vii. Promulgate disciplinary action for violators of procedures;
- viii. Ensure continuous or regular assessment, auditing and senior management oversight (monitoring); and
- ix. Have an up-to-date information security incident response plan.

Fraud which is on the rise is virulent. Perpetrators constantly adopt new methods on an ongoing basis. As one barrier is put up, fraudsters will seek and find an alternative weakness to exploit. This type of persistence and stealth is especially evident in the creative ways digital networks are constantly being attacked and often penetrated. Hence, businesses must implement procedures that can help them identify, mitigate and manage fraud risks. There is no absolute or perfect solution, however, energy and efforts have to be focused not only on prevention, but also on proactive responses in the event that such fraudulent efforts are able to circumvent processes and other preventive measures. Obviously, the benefit of a fraud risk management scheme outweighs its cost.

REFERENCES

- Abbass, H. A., Sarker, R. A. & Newton, C. S. (2002). *Data mining: A Heuristic approach*. Idea Group Publishing, London WC2E 8LU.
- Adomavicious, G., Tuzhilin, A. (1999). *User profiling in personalization applications through rule discovery and validation*. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA.
- Agyemang, M., Barker K., & Alhajj, R. (2006). A comprehensive survey of numeric and symbolic outlier mining techniques, *Intelligent Data Analysis*, 10(6), 521–538.
- Aha, D., & Kibler, D. (1991). Instance-based learning algorithms. *Machine Learning*, (6), 37–66.
- Ahmed, M., Mahmood, A. N., & Islam R. (2015). *A survey of anomaly detection techniques in financial domain*. Future Generation Computer Systems (2015), <http://dx.doi.org/10.1016/j.future.2015.01.001>.
- Anderson, J.A. & Rosenfeld, E. (1998). *Neurocomputing: Foundations of Research*. MIT Press, Cambridge.
- Apostolou, B., Hassell, J., Webber, S., & Sumners, G. (2001). The relative importance of management fraud risk factors. *Behavioral Research in Accounting* 13, 1–24.
- Aral, K.D., Güvenir, H. A., Sabuncuog, I. & Akar, A. R. (2012). A prescription fraud detection model. *Elsevier Computer Methods and Programs in Biomedicine*, 10(6), 37 – 46.
- Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning*, 36 (1–2), 105–139.
- Ben-David, S., Loker D., Srebro, N., & Sridharam, K. (2012). *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, U.K.
- Bergmeir, C. & Benítez, J.M. (2012), Neural networks in R using the Stuttgart Neural Network Simulator: RSNNS. *Journal of Statistical Software*, 46(7), 1-26. <http://www.jstatsoft.org/v46/i07/>
- Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

- Bolton, R., & Hand, D. (2002). Statistical fraud detection: A review (With Discussion). *Statistical Science*, 17(3), 235–255.
- Bose, R.K. Mahapatra (2001), Business data mining — A machine learning perspective. *Information Management*, 39(3), 211–225.
- Bramer, M. (2007). *Principles of data mining*. London: Springer-Verlag Limited.
- Brause, R., Langsdorf T., & Hepp M. (1999), *Neural data mining for credit card fraud detection*, 11th IEEE International Conference proceeding, (pp103 -106).
- Broomhead, D. S. & Lowe, D. (1988). Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2, 321-355.
- Buhmann, M. D. (2004). *Radial basis functions: Theory and implementations*. Cambridge: Cambridge University Press. ISBN 0-521-63338-9.
- Burge, P. & Shawe-Taylor, J. (2001). An unsupervised neural, network approach to profiling the behaviour of mobile phone, users for use in fraud detection. *Journal of Parallel and Distributed Computing*, 61, 915–925.
- Burl, M., Asker, L., Smyth, P., Fayyad, U., Perona, P., Crumpler, L., & Aubele, J. (1998). Learning to recognize volcanoes on Venus. *Machine Learning*, 30, 165-195.
- Buschkes, R., Kesdogan D., & Reichl, P. (1998). *How to increase security in mobile networks by anomaly detection*. Proceedings of the 14th Annual Computer Security Applications Conference (ACSAC '98).
- Cai, F., LeKhac, N., & Kechadi, M. (2012). Toward a new classification model for analysing financial datasets. In *Digital Information Management (ICDIM), 2012 Seventh International Conference on Data Mining*, (pp 1-6). IEEE.
- Caruana, R., & Niculescu-Mizil, A. (2006). *An empirical comparison of supervised learning algorithms*. Paper presented at the 23rd International Conference on Machine Learning, Pittsburgh, USA.
- Chang, W. & Chang, J. (2012), An effective early fraud detection method for online auctions. *Electronic Commerce Research and Applications*, 11, 346–360.
- Cheng C., Wang, D., & Shi, W. (2014). *The second monitor centre of China earthquake administration, Xi'an, Shaanxi, China*. Engineering Technology and Applications. London: Taylor & Francis Group.

- Cichosz, P., (2015). *Data mining algorithms: Explained using R*. Poland: John Wiley & sons Ltd.
- Coderre, D. (2009), *Computer-aided fraud prevention and detection*. New Jersey: John Wiley and Sons, Inc.
- COSO, Committee of Sponsoring Organizations of the Treadway Commission (1999). *Fraudulent financial reporting: 1987–1997*. An analysis of U.S. public companies. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Coussement, K., & Poel, D.V. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter- selection techniques. *Expert Systems with Applications*, 34(1), 313–327.
- Cox, K., Eick, S. & Wills, G. (1997). Visual data mining: Recognising telephone calling fraud. *Data Mining and Knowledge Discovery*, 1, 225–231.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press.
- Daskalaki, S., Kopanas, I., Goudara, M., & Avouris, N. (2003). Data mining for decision support on customer insolvency in telecommunications business. *European Journal of Operational Research*, 145, 239–255.
- Desai, V.S., Crook, J.N., & Overstreet, J. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95, 24–37.
- DeZoort, F.T., & Lee T.A. (1998). The impact of SAS No. 82 on perceptions of external auditor responsibility for fraud detection. *International Journal of Auditing*, 2, 167–82.
- Dorransoro, J. R., Ginel, F., Sánchez C., & Cruz C. S. (1997). Neural Fraud Detection in Credit Card Operations. *IEEE Transactions on Neural Networks*, 4(8), 827-834.
- Duda, R. O., Hart, P.E., & Green, D. (2001), *Stork. pattern classification*. New York: John Willey & Sons, Inc.
- Eckhorn, R. (1999). Neural mechanisms of scene segmentation: Recordings from the visual cortex suggest basic circuits for linking field models, *IEEE Transactions on Neural Networks*, 10, 1–16.

- Estevez, P., Held, C., & Perez, C. (2006). Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. *Expert Systems with Applications*, 31, 337–344.
- Ezawa, K. J., & Norton, S. W. (1996). Constructing bayesian networks to predict uncollectible telecommunications accounts. *IEEE Expert*, 11, 45–51.
- Farvaresh, H. & Sepehri, M. M., (2011). A data mining framework for detecting subscription fraud in telecommunication. *Engineering Applications of Artificial Intelligence*, 24, 182–194.
- Fawcett, T. (1997). AI approaches to fraud detection and risk management. Papers from the 1997 AAAI Workshop. Technical Report WS-97-07. AAAI Press.
- Fawcett, T., & Provost, F. (1997). Combining data mining and machine learning for effective fraud detection. In AAAI workshop on AI approaches to fraud detection and risk management, (pp.14–19).
- FBI, Federal Bureau of Investigation New York Division, Department of Justice, United States. (2008). <http://newyork.fbi.gov/dojpressrel/pressrel08/nyfo121108.htm>.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179 – 188.
- Field, S., & Hobson, P. (1997). Techniques for telecommunications fraud management. *Proceedings of international workshop on applications of neural networks to telecommunications*, 3, 107–115.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheu, C. J., (1992). Knowledge discovery in databases: An overview. *AI Magazine*, 13(3), 57–70.
- Frank, D. C., Esper, J., Raible, C. C., Büntgen, U., Trouet, V., Stocker, B., & Joos, F. (2010). Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. *Nature*, 463(7280), 527-530.
- Fung, C. C., Iyer V., Brown W., & Wong, K. W., (2005). Comparing the performance of different neural networks architectures for the prediction of mineral prospectivity. In *Machine Learning and Cybernetics, 2005*. Proceedings of 2005 International Conference, pp. 394-398.
- Garson, G. D. 1991. Interpreting neural network connection weights. *Artificial Intelligence Expert*. 6(4), 46-51.
- Ghosh, R., & Reilly, D. (1994). Credit card fraud detection with a neural-network. *Proceedings of the twenty-seventh annual Hawaii international conference on system sciences*, pp. 347-454.

- Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection, *Decision. Support System*, 50(3), 595–601.
- Green, B., & Choi, J. (1997). Assessing the risk of management fraud through neural network technology. *Auditing: A Journal of Practice and Theory*, 16(1), 14–28.
- Gullkvist, B. & Jokipii, A. (2012). Perceived importance of red flags across fraud types. *Critical Perspectives on Accounting*, (24), 44-61.
- Gupta, R., & Gill, N. S. (2013). Prevention and detention of financial statement fraud – An implementation of data mining framework. *International Journal of Advanced Computer Science and Applications*, 3(8), 65-76.
- Hackenbrack, K. (1993). The effect of experience with different sized clients on auditor evaluations of fraudulent financial reporting indicators. *Auditing: A Journal of Practice and Theory*, 16(1), 99 –110.
- Hamel, L.H. (2009). *Knowledge discovery with support vector machines*. New Jersey: John Wiley & Sons Inc.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and techniques*, 3 ed. Waltham, USA: Morgan Kaufmann Publishers.
- Hand D., Mannila H., & Smyth P. (2001). *Principles of data mining*. London, England: Massachusetts Institute of Technology Press,
- Hardy, R.L. (1990). Theory and applications of the multiquadric-biharmonic method, 20 years of discovery, 1968-1988, *Computational Mathematics Application*, 1(8/9), 163-208.
- Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The Elements of statistical learning: data mining, inference, and prediction*, 2nd ed. New York: Springer-Verlag.
- Herbrich, R., (2001). *Learning kernel classifiers: Theory and algorithms*. Cambridge: MIT Press.
- Hilas, C.S., & Mastorocostas, P.A. (2008). An application of supervised and unsupervised learning approaches to telecommunications fraud detection. *Knowledge-Based Systems*, 21(7), 721–726.
- Hippert, H.S., Bunn, D.W., & Souza, R.C. (2005). Large neural networks for electricity load forecasting: Are they over-fitted? *International Journal of Forecasting*, 21, 425–434.
- Hoffman, H. (2000). *UCI machine learning repository*. Institut f"ur Statistik und Okonometrie Universit at Hamburg. Retrieved from: <http://archive.ics.uci.edu/ml>.

- Hung, S. Y., Yen, D. C., & Wang, H.Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), 515–524.
- Interpol, International Police (2017). Retrieved from: <https://www.interpol.int/Crime-areas/Financial-crime/Social-engineering-fraud/Types-of-social-engineering-fraud> on Jan-13-2017.
- Iroko, A. P. (2015). *The choice of activation function in back-propagation artificial neural network for classification*. (Unpublished Master's thesis). University of Ilorin, Ilorin.
- Khac, N. L. & Kechadi, M. (2010). Application of data mining for anti-money laundering detection: A case study. *2010 IEEE international conference on data mining workshops proceeding*, 577-584.
- Kim, J., Ong, A., & Overill, R. (2003). Design of an Artificial Immune System as a Novel Anomaly Detector for Combating Financial Fraud in Retail Sector. *Congress on Evolutionary Computation*, Canberra, Australia.
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4), 995–1003.
- Kohavi, R., & John, H. G. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(2), 273–324.
- KPMG, Klynveld Peat Marwick Goerdeler, (2016). Global profiles of the fraudster: Technology enables and weak controls fuel the fraud. Retrieved from: <https://home.kpmg.com/xx/en/home/insights/2016/05/global-profiles-of-the-fraudster.html>
- Krambia-Kapardis, M., Christodoulou, C., & Agathocleous, M. (2010). Neural networks: The panacea in fraud detection. *Managerial Auditing Journal*, 7, 659–78.
- Kremper, A., Schanze, T. & Eckhorn, R. (2002). Classification of cortical signals with a generalized correlation classifier based on radial basis functions, *J. Neuroscience Method*, 116, 179–187.
- Kristin, R. N., & Philip, M. (1999). *Using data mining techniques for fraud detection*. California, USA: Federal Data Corporation, Bethesda, MD and SAS Institute Inc.
- Kroll, J. B. (2016). *Global fraud report 2015/16.*, New York: 600 Third Avenue

- Lacher, R. C., Coats, P. K., Shanker, S. C., & Fant, L. F. (1995). A neural network for classifying the financial health of a firm. *European Journal of Operational Research*, 85(1), 53–65.
- Langley, P. & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, 38(11), 55-64.
- Lavrac, N., Motoda, H., Fawcett, T., Holte, R., Langley, P. & Adriaans, P. (2004). Introduction: Lessons learned from data mining applications and collaborative problem solving. *Machine Learning*, 57(1-2), 13-34.
- Lawrence, C., Giles, C. L., & Tsoi, A. C. (1997). *Lessons in neural network training: overfitting may be harder than expected*. Proceedings of the Fourteenth national Conference on Artificial Intelligence, AAAI-97, July27-31, 1997, AAAI Press, Menlo Park, California, (pp. 540-545).
- Lee C., Chung P., Tsai J., & Chang C. (1999). Robust radial basis function neural networks. *IEEE transactions on systems: Man and Cybernetics*, 29(6), 674-685.
- Lee, K. C., Han, I., & Kwon, Y. (1996). Hybrid neural network models for bankruptcy predictions. *Decision Support Systems*, 18(1), 63–72.
- Lee, T. S., Chiu, C. C., Chou, Y. C., & Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50, 1113–1130.
- Lei, Z. W., Xu, Z. S. & Liu, F. (2010). The improvement of information neural network and its application. *Instrumentation Technology*, 5, 57-59.
- Loebbecke, J.K., Eining, M. M. & Willingham, J. J. (1989). Auditors' experience with material irregularities: frequency, nature and detectability. *Auditing: A Journal of Practice & Theory*, 1, 1–28.
- López, M. I., Colomer N., Ruisánchez, I., & Callao M. P. (2014). Validation of multivariate screening methodology. Case study: Detection of food fraud. *Elsevier Analytica Chimica Acta*, 8(27), 28 -33.
- Louzada, F. & Ara, A. (2012). Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool. *Elsevier*, 39(44), 11583 – 11592.
- MacKay, D. J. C. (1992). The evidence framework applied to classification networks. *Neural Computation*, 4(5), 720–736.

- Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002). *Credit card fraud detection using bayesian and neural networks*. Proceedings of the 1st International NAISO Congress on Neuro Fuzzy Technologies.
- Majid, A, Gul, F.A. & Tsui, J. (2001). An analysis of Hong Kong auditors' perceptions of the importance of selected red flag factors in risk assessment. *Journal of Business Ethics*, 3, 263–74.
- Manikopoulos, C., & Papavassiliou, S. (2002). Network intrusion and fault detection: A statistical anomaly approach. *IEEE Communications Magazine*, 76–82.
- Matej, S. & Lewitt, R.M. (1996). Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Transactions on Medical Imaging*, 15(1), 68-78.
- Maranzato, R., Pereira, A., Naubert, M., & Lago, A. P. (2010). Fraud detection in reputation systems in e-markets using logistic regression and stepwise optimization. *In ACM SIGAPP Applied Computing Review*.
- Mattison, R. (1999). *Web warehousing and knowledge management*. New York: McGraw-Hill.
- Mena, J. (2003). *Investigative data mining for security and criminal detection*. Elsevier Science (USA): Butterworth Heinemann.
- Meyer, D., Leisch, F. & Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, 55,169–186.
- Michie, D., Spiegelhalter, D. J., & Taylor, C.C. (1994). *Machine learning: neural and statistical classification*. London: Ellis Horwood.
- Mobley, B. A., Schechter, E., Moore, W. E., McKee, P. A., & Eichner, J. E. (2000). Predictions of coronary artery stenosis by artificial neural network. *Artificial Intelligence in Medicine*, 18(3), 187–203.
- Mock, T.J. & Turner, J.L. (2005). Auditor identification of fraud risk factors and their impact on audit programs. *International Journal of Auditing*, 5(9), 59–77.
- Moyes, G.L. (2007). The differences in perceived level of fraud-detecting effectiveness of SAS No. 99 red flags between external and internal auditors. *Journal of Business & Economics Research*, 6, 9–25.
- Murad, U. & Pinkas, G. (1999). *Unsupervised profiling for identifying superimposed fraud*. Proceedings of Principle and Learning of Knowledge Discovery in Databases, Dublin, U.K.

- Nauta, K. R., & Matkovsky, P. (1999). *Using data mining techniques for fraud detection*. North Carolina, USA: Federal Data Corporation, Bethesda, MD and SAS Institute Inc.
- Ngai, E.W.T., Hu, Y., Wong Y.H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50, 559-569.
- Nigrini, M. (2011). *Forensic analytics: Methods and techniques for forensic accounting investigations*. Hoboken, NJ: John Wiley & Sons Inc.
- Oh, H., Lee, W.S., (2003). An anomaly intrusion detection method by clustering normal user behavior. *Computers & Security*, 22 (7), 596–612.
- Olson, D. L. & Delen, D. (2008). *Advanced Data Mining Techniques*, Berlin Heidelberg: Springer-Verlag.
- Orr, M. J. (1999). *Introduction to radial basis function networks*. Recent advances in radial basis function networks, Edinburgh EH8 9LW, Scotland.
- Oussar, Y., & Dreyfus, G. (2002). Initialization by selection for wavelet network training, *Neurocomputing*, 34, 131–143.
- Palshikar, G.K. (2002). The hidden truth – Frauds and their control: A critical application for business intelligence. *Intelligent Enterprise*, 5(9), 46–51.
- Pérez, D. G. & Lavalley, M. M. (2011). *Outlier detection applying an innovative user transaction modeling with automatic explanation*. IEEE 2011 Electronics, Robotics and Automative Conference proceeding, (pp. 41 – 46).
- Pincus, K. V. (1989). The efficacy of a red flags questionnaire for assessing the possibility of fraud. *Accounting, Organizations and Society*, 14, 153–64.
- Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research*, 112(2), 310–321.
- Phua, C., Lee V., Smith, K., & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review*, 1–14.
- Rosset, S ., Murad U., Neumann, E., Idan Y., & Pinkas G. (1999). *Discovery of fraud rules for telecommunications—challenges and solutions*. Proceeding of the Special Interest Group on Knowledge Discovery in Databases..

- Salchenberger, L. M., Venta, E. R., & Venta, L. A. (1997). Using neural networks to aid the diagnosis of breast implant rupture. *Computers and Operations Research*, 24(5), 435–444.
- Sanner, R. M. & Slotine, J. E. (1994). Gaussian networks for direct adaptive control. *IEEE Transactions Network*, 3(6), 837-863.
- Sánchez, D., Vila, M. A., Cerda, L. & Serrano, J. M. (2009). Association rules applied to credit card fraud detection. *Elsevier Expert Systems with Applications*, 36, 3630–3640.
- Schölkopf, B., & Smola, A. J. (2001). *Learning with kernels*. Cambridge: MIT Press.
- Shao, H., Zhao H., & Chang, G. (2002). *Applying data mining to detect fraud behaviour in customs declaration*. Proceeding of First International Conference on Machine Learning and Cybernetics.
- Sharda, R., & Wilson, R. (1996). Neural network experiments in business failures prediction: A review of predictive performance issues. *International Journal of Computational Intelligence and Organizations*, 1(2), 107–117.
- Shen, A., Tong, R., & Deng, Y. (2007). *Application of classification models on credit card fraud detection*. Proceedings of the 10th International Conference on Service Systems and Service Management, (pp. 1–4).
- Sherly, K.K., & Nedunchezian, R. (2010). Boat adaptive credit card fraud detection system. *IEEE Trans. Dependable and Secure Computing*, 4(10), 738-763.
- Silverstone, H., & Davia, H. R. (2005). *Fraud 101: Techniques and strategies for detection*, 2ed. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Smith, M., Omar, N.H., Idris, S., & Baharuddin, I. (2005). Auditors' perception of fraud risk indicators, Malaysian evidence. *Managerial Auditing Journal*, 1, 73–85.
- Sokol, L., Garcia, B., Rodriguez, J., West, M., & Johnson, K. (2001). Using data mining to find fraud in HCFA health care claims. *Topics in Health Information Management*, 22(1), 1–13.
- SPSS, Statistical Package for the Social Sciences (2000). *Using data mining to detect fraud*. White paper -technical report. USA: SPSS Inc.
- Steinwart, I., & Christmann, A., (2008). *Support vector machines*. New York, USA: Springer.

- Spiliopoulou, M., Wang, D., Cook H., Pei J., Wang W., Zaiane O., & Wu X. (2011). *Proceeding of the 11th IEEE international conference on data mining workshops*. Vancouver, Canada.
- Stern, H.S. (1996). Neural networks in applied statistics. *Technometrics*, 38(3), 205–216.
- Sumathi, S., & Sivanandam, S. N. (2006). Introduction to data mining and its applications. *Studies in Computational Intelligence*, 29(3), 540- 560.
- Taniguchi, M., Haft, M., HollmTn, J., & Tresp, V., (1998). *Fraud detection in communications networks using neural and probabilistic methods*. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing 2, (pp. 1241–1244).
- Thomas, L.C. (2000). A survey of credit and behavioral scoring: forecasting financial risks of lending to customers. *International Journal of Forecasting*, 16, 149–172.
- Timm, N.H. (2002). *Applied multivariate analysis*. New York: Springer-Verlag.
- Torgo, L., & Soares C. (2010). *Resource-bounded outlier detection using clustering methods*. Proceedings of the 2010 Conference on Data Mining for Business Applications, IOS Press, Amsterdam, The Netherlands, (pp. 84–98).
- Turban, E., Aronson, J.E., Liang, T.P., & Sharda, R. (2007). *Decision support and business intelligence systems*, 8 ed. London, United Kingdom: Pearson Education.
- Viaene, S., Derrig, R.A., & Dedene, G. (2004). A case study of applying boosting naive Bayes to claim fraud diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 16(5), 612–620.
- Viaene, S., Dedene, G., & Derrig R. A., (2005). Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications*, 29 (2005) 653-666.
- Wang, J., Liao, Y., Tsai, T., & Hung, G. (2006). Technology-based financial frauds in Taiwan: Issue and approaches. *IEEE Conference on systems, man and cyberspace*, (pp. 1120–1124).
- Weiss, G.M. (2004). Mining with rarity: A unifying framework. *SIGKDD Explorations*, 6(1), 7–19.
- Wiese, B., & Omlin, C. (2009). Credit card transactions, fraud detection, and machine learning: Modelling time with LSTM recurrent neural networks. *Innovations in Neural Information Paradigms and Applications*, 231-268.

- Wilson, J. H. (2009). An analytical approach to detecting insurance fraud using logistic regression. *Journal of Finance and Accountancy*, 1, 17 -24.
- Xing, D. & Girolami, M. (2007). Employing latent Dirichlet allocation for fraud detection in telecommunications. *Pattern Recognition Letters*, 28, 1727–1734.
- Yahya, W. B., Oladiipo, M. O. & Jolayemi, E. T. (2012). A fast algorithm to construct neural networks classification models with high-dimensional genomic data. *Anaa Seria Informatica*, 10(1), 39-56.
- Yahya, W.B., Ulm, K., Fahrmeir, L. & Hapfelmeier, A. (2011). K-SS: a sequential feature selection and prediction method in microarray study. *International Journal of Artificial Intelligence*, 6(2011), 19-47.
- Yue, X., Wu, Y., Wang, L., & Chu C. (2007). A review of data mining-based financial fraud detection research. International conference on wireless communications. *Networking and Mobile Computing*, (pp. 5519–5522).
- Zhang, D. & Zhou, L. (2004). Discovering golden nuggets: Data mining in financial application. *IEEE Transactions on Systems, Man and Cybernetics*, 34(4), 156–166.

APPENDIX A

A.1 Raw German credit data extracts

A11 6 A34 A43 1169 A65 A75 4 A93 A101 4 A121 67 A143 A152 2 A173 1 A192 A201 1
A12 48 A32 A43 5951 A61 A73 2 A92 A101 2 A121 22 A143 A152 1 A173 1 A191 A201 2
A14 12 A34 A46 2096 A61 A74 2 A93 A101 3 A121 49 A143 A152 1 A172 2 A191 A201 1
A11 42 A32 A42 7882 A61 A74 2 A93 A103 4 A122 45 A143 A153 1 A173 2 A191 A201 1
A11 24 A33 A40 4870 A61 A73 3 A93 A101 4 A124 53 A143 A153 2 A173 2 A191 A201 2
A14 36 A32 A46 9055 A65 A73 2 A93 A101 4 A124 35 A143 A153 1 A172 2 A192 A201 1
A14 24 A32 A42 2835 A63 A75 3 A93 A101 4 A122 53 A143 A152 1 A173 1 A191 A201 1
A12 36 A32 A41 6948 A61 A73 2 A93 A101 2 A123 35 A143 A151 1 A174 1 A192 A201 1
A14 12 A32 A43 3059 A64 A74 2 A91 A101 4 A121 61 A143 A152 1 A172 1 A191 A201 1
A12 30 A34 A40 5234 A61 A71 4 A94 A101 2 A123 28 A143 A152 2 A174 1 A191 A201 2
A12 12 A32 A40 1295 A61 A72 3 A92 A101 1 A123 25 A143 A151 1 A173 1 A191 A201 2
A11 48 A32 A49 4308 A61 A72 3 A92 A101 4 A122 24 A143 A151 1 A173 1 A191 A201 2
A12 12 A32 A43 1567 A61 A73 1 A92 A101 1 A123 22 A143 A152 1 A173 1 A192 A201 1
A11 24 A34 A40 1199 A61 A75 4 A93 A101 4 A123 60 A143 A152 2 A172 1 A191 A201 2
A11 15 A32 A40 1403 A61 A73 2 A92 A101 4 A123 28 A143 A151 1 A173 1 A191 A201 1
A11 24 A32 A43 1282 A62 A73 4 A92 A101 2 A123 32 A143 A152 1 A172 1 A191 A201 2
A14 24 A34 A43 2424 A65 A75 4 A93 A101 4 A122 53 A143 A152 2 A173 1 A191 A201 1
A11 30 A30 A49 8072 A65 A72 2 A93 A101 3 A123 25 A141 A152 3 A173 1 A191 A201 1
A12 24 A32 A41 12579 A61 A75 4 A92 A101 2 A124 44 A143 A153 1 A174 1 A192 A201 2
A14 24 A32 A43 3430 A63 A75 3 A93 A101 2 A123 31 A143 A152 1 A173 2 A192 A201 1
A14 9 A34 A40 2134 A61 A73 4 A93 A101 4 A123 48 A143 A152 3 A173 1 A192 A201 1
A11 6 A32 A43 2647 A63 A73 2 A93 A101 3 A121 44 A143 A151 1 A173 2 A191 A201 1
A11 10 A34 A40 2241 A61 A72 1 A93 A101 3 A121 48 A143 A151 2 A172 2 A191 A202 1
A12 12 A34 A41 1804 A62 A72 3 A93 A101 4 A122 44 A143 A152 1 A173 1 A191 A201 1
A14 10 A34 A42 2069 A65 A73 2 A94 A101 1 A123 26 A143 A152 2 A173 1 A191 A202 1
A11 6 A32 A42 1374 A61 A73 1 A93 A101 2 A121 36 A141 A152 1 A172 1 A192 A201 1
A14 6 A30 A43 426 A61 A75 4 A94 A101 4 A123 39 A143 A152 1 A172 1 A191 A201 1
A13 12 A31 A43 409 A64 A73 3 A92 A101 3 A121 42 A143 A151 2 A173 1 A191 A201 1
A12 7 A32 A43 2415 A61 A73 3 A93 A103 2 A121 34 A143 A152 1 A173 1 A191 A201 1
A11 60 A33 A49 6836 A61 A75 3 A93 A101 4 A124 63 A143 A152 2 A173 1 A192 A201 2
A12 18 A32 A49 1913 A64 A72 3 A94 A101 3 A121 36 A141 A152 1 A173 1 A192 A201 1
A11 24 A32 A42 4020 A61 A73 2 A93 A101 2 A123 27 A142 A152 1 A173 1 A191 A201 1
A12 18 A32 A40 5866 A62 A73 2 A93 A101 2 A123 30 A143 A152 2 A173 1 A192 A201 1
A14 12 A34 A49 1264 A65 A75 4 A93 A101 4 A124 57 A143 A151 1 A172 1 A191 A201 1
A13 12 A32 A42 1474 A61 A72 4 A92 A101 1 A122 33 A141 A152 1 A174 1 A192 A201 1
A12 45 A34 A43 4746 A61 A72 4 A93 A101 2 A122 25 A143 A152 2 A172 1 A191 A201 2
A14 48 A34 A46 6110 A61 A73 1 A93 A101 3 A124 31 A141 A153 1 A173 1 A192 A201 1
A13 18 A32 A43 2100 A61 A73 4 A93 A102 2 A121 37 A142 A152 1 A173 1 A191 A201 2
A13 10 A32 A44 1225 A61 A73 2 A93 A101 2 A123 37 A143 A152 1 A173 1 A192 A201 1
A12 9 A32 A43 458 A61 A73 4 A93 A101 3 A121 24 A143 A152 1 A173 1 A191 A201 1
A14 30 A32 A43 2333 A63 A75 4 A93 A101 2 A123 30 A141 A152 1 A174 1 A191 A201 1
A12 12 A32 A43 1158 A63 A73 3 A91 A101 1 A123 26 A143 A152 1 A173 1 A192 A201 1
A12 18 A33 A45 6204 A61 A73 2 A93 A101 4 A121 44 A143 A152 1 A172 2 A192 A201 1
A11 30 A34 A41 6187 A62 A74 1 A94 A101 4 A123 24 A143 A151 2 A173 1 A191 A201 1

A11 48 A34 A41 6143 A61 A75 4 A92 A101 4 A124 58 A142 A153 2 A172 1 A191 A201 2
A14 11 A34 A40 1393 A61 A72 4 A92 A101 4 A123 35 A143 A152 2 A174 1 A191 A201 1
A14 36 A32 A43 2299 A63 A75 4 A93 A101 4 A123 39 A143 A152 1 A173 1 A191 A201 1
A11 6 A32 A41 1352 A63 A71 1 A92 A101 2 A122 23 A143 A151 1 A171 1 A192 A201 1
A14 11 A34 A40 7228 A61 A73 1 A93 A101 4 A122 39 A143 A152 2 A172 1 A191 A201 1
A14 12 A32 A43 2073 A62 A73 4 A92 A102 2 A121 28 A143 A152 1 A173 1 A191 A201 1
A12 24 A33 A42 2333 A65 A72 4 A93 A101 2 A122 29 A141 A152 1 A172 1 A191 A201 1
A12 27 A33 A41 5965 A61 A75 1 A93 A101 2 A123 30 A143 A152 2 A174 1 A192 A201 1
A14 12 A32 A43 1262 A61 A73 3 A93 A101 2 A123 25 A143 A152 1 A173 1 A191 A201 1
A14 18 A32 A41 3378 A65 A73 2 A93 A101 1 A122 31 A143 A152 1 A173 1 A192 A201 1
A12 36 A33 A40 2225 A61 A75 4 A93 A101 4 A124 57 A141 A153 2 A173 1 A192 A201 2
A14 6 A31 A40 783 A65 A73 1 A93 A103 2 A121 26 A142 A152 1 A172 2 A191 A201 1
A12 12 A32 A43 6468 A65 A71 2 A93 A101 1 A124 52 A143 A152 1 A174 1 A192 A201 2
A14 36 A34 A43 9566 A61 A73 2 A92 A101 2 A123 31 A142 A152 2 A173 1 A191 A201 1
A13 18 A32 A40 1961 A61 A75 3 A92 A101 2 A123 23 A143 A152 1 A174 1 A191 A201 1
A11 36 A34 A42 6229 A61 A72 4 A92 A102 4 A124 23 A143 A151 2 A172 1 A192 A201 2
A12 9 A32 A49 1391 A61 A73 2 A94 A101 1 A121 27 A141 A152 1 A173 1 A192 A201 1
A12 15 A34 A43 1537 A65 A75 4 A93 A103 4 A121 50 A143 A152 2 A173 1 A192 A201 1
A12 36 A30 A49 1953 A61 A75 4 A93 A101 4 A124 61 A143 A153 1 A174 1 A192 A201 2
A12 48 A30 A49 14421 A61 A73 2 A93 A101 2 A123 25 A143 A152 1 A173 1 A192 A201 2
A14 24 A32 A43 3181 A61 A72 4 A92 A101 4 A122 26 A143 A152 1 A173 1 A192 A201 1
A14 27 A32 A45 5190 A65 A75 4 A93 A101 4 A122 48 A143 A152 4 A173 2 A192 A201 1
A14 12 A32 A43 2171 A61 A72 2 A92 A101 2 A123 29 A141 A152 1 A173 1 A191 A201 1
A12 12 A32 A40 1007 A64 A73 4 A94 A101 1 A121 22 A143 A152 1 A173 1 A191 A201 1
A14 36 A32 A46 1819 A61 A73 4 A93 A101 4 A124 37 A142 A153 1 A173 1 A192 A201 2
A14 36 A32 A43 2394 A65 A73 4 A92 A101 4 A123 25 A143 A152 1 A173 1 A191 A201 1
A14 36 A32 A41 8133 A61 A73 1 A92 A101 2 A122 30 A141 A152 1 A173 1 A191 A201 1
A14 7 A34 A43 730 A65 A75 4 A93 A101 2 A122 46 A143 A151 2 A172 1 A192 A201 1
A11 8 A34 A410 1164 A61 A75 3 A93 A101 4 A124 51 A141 A153 2 A174 2 A192 A201 1
A12 42 A34 A49 5954 A61 A74 2 A92 A101 1 A121 41 A141 A152 2 A172 1 A191 A201 1
A11 36 A32 A46 1977 A65 A75 4 A93 A101 4 A124 40 A143 A152 1 A174 1 A192 A201 2
A11 12 A34 A41 1526 A61 A75 4 A93 A101 4 A124 66 A143 A153 2 A174 1 A191 A201 1
A11 42 A32 A43 3965 A61 A72 4 A93 A101 3 A123 34 A143 A152 1 A173 1 A191 A201 2
A12 11 A33 A43 4771 A61 A74 2 A93 A101 4 A122 51 A143 A152 1 A173 1 A191 A201 1
A14 54 A30 A41 9436 A65 A73 2 A93 A101 2 A122 39 A143 A152 1 A172 2 A191 A201 1
A12 30 A32 A42 3832 A61 A72 2 A94 A101 1 A122 22 A143 A152 1 A173 1 A191 A201 1
A14 24 A32 A43 5943 A65 A72 1 A92 A101 1 A123 44 A143 A152 2 A173 1 A192 A201 2
A14 15 A32 A43 1213 A63 A75 4 A93 A101 3 A122 47 A142 A152 1 A173 1 A192 A201 1
A14 18 A32 A49 1568 A62 A73 3 A92 A101 4 A122 24 A143 A151 1 A172 1 A191 A201 1
A11 24 A32 A410 1755 A61 A75 4 A92 A103 4 A121 58 A143 A152 1 A172 1 A192 A201 1
A11 10 A32 A43 2315 A61 A75 3 A93 A101 4 A121 52 A143 A152 1 A172 1 A191 A201 1
A14 12 A34 A49 1412 A61 A73 4 A92 A103 2 A121 29 A143 A152 2 A174 1 A192 A201 1
A12 18 A34 A42 1295 A61 A72 4 A92 A101 1 A122 27 A143 A152 2 A173 1 A191 A201 1
A12 36 A32 A46 12612 A62 A73 1 A93 A101 4 A124 47 A143 A153 1 A173 2 A192 A201 2
A11 18 A32 A40 2249 A62 A74 4 A93 A101 3 A123 30 A143 A152 1 A174 2 A192 A201 1
A11 12 A30 A45 1108 A61 A74 4 A93 A101 3 A121 28 A143 A152 2 A173 1 A191 A201 2

A14 12 A34 A43 618 A61 A75 4 A93 A101 4 A121 56 A143 A152 1 A173 1 A191 A201 1
A11 12 A34 A41 1409 A61 A75 4 A93 A101 3 A121 54 A143 A152 1 A173 1 A191 A201 1
A14 12 A34 A43 797 A65 A75 4 A92 A101 3 A122 33 A141 A152 1 A172 2 A191 A201 2
A13 24 A34 A42 3617 A65 A75 4 A93 A102 4 A124 20 A143 A151 2 A173 1 A191 A201 1
A12 12 A32 A40 1318 A64 A75 4 A93 A101 4 A121 54 A143 A152 1 A173 1 A192 A201 1
A12 54 A30 A49 15945 A61 A72 3 A93 A101 4 A124 58 A143 A151 1 A173 1 A192 A201 2
A14 12 A34 A46 2012 A65 A74 4 A92 A101 2 A123 61 A143 A152 1 A173 1 A191 A201 1
A12 18 A32 A49 2622 A62 A73 4 A93 A101 4 A123 34 A143 A152 1 A173 1 A191 A201 1
A12 36 A34 A43 2337 A61 A75 4 A93 A101 4 A121 36 A143 A152 1 A173 1 A191 A201 1
A12 20 A33 A41 7057 A65 A74 3 A93 A101 4 A122 36 A141 A151 2 A174 2 A192 A201 1
A14 24 A32 A40 1469 A62 A75 4 A94 A101 4 A121 41 A143 A151 1 A172 1 A191 A201 1
A12 36 A32 A43 2323 A61 A74 4 A93 A101 4 A123 24 A143 A151 1 A173 1 A191 A201 1
A14 6 A33 A43 932 A61 A73 3 A92 A101 2 A121 24 A143 A152 1 A173 1 A191 A201 1
A12 9 A34 A42 1919 A61 A74 4 A93 A101 3 A123 35 A143 A151 1 A173 1 A192 A201 1
A14 12 A32 A41 2445 A65 A72 2 A94 A101 4 A123 26 A143 A151 1 A173 1 A192 A201 1
A12 24 A34 A410 11938 A61 A73 2 A93 A102 3 A123 39 A143 A152 2 A174 2 A192 A201 2
A14 18 A31 A40 6458 A61 A75 2 A93 A101 4 A124 39 A141 A152 2 A174 2 A192 A201 2
A12 12 A32 A40 6078 A61 A74 2 A93 A101 2 A123 32 A143 A152 1 A173 1 A191 A201 1
A11 24 A32 A42 7721 A65 A72 1 A92 A101 2 A122 30 A143 A152 1 A173 1 A192 A202 1
A12 14 A32 A49 1410 A63 A75 1 A94 A101 2 A121 35 A143 A152 1 A173 1 A192 A201 1
A12 6 A33 A49 1449 A62 A75 1 A91 A101 2 A123 31 A141 A152 2 A173 2 A191 A201 1
A13 15 A32 A46 392 A61 A72 4 A92 A101 4 A122 23 A143 A151 1 A173 1 A192 A201 1
A12 18 A32 A40 6260 A61 A74 3 A93 A101 3 A121 28 A143 A151 1 A172 1 A191 A201 1
A14 36 A34 A40 7855 A61 A73 4 A92 A101 2 A121 25 A142 A152 2 A173 1 A192 A201 2
A11 12 A32 A43 1680 A63 A75 3 A94 A101 1 A121 35 A143 A152 1 A173 1 A191 A201 1
A14 48 A34 A43 3578 A65 A75 4 A93 A101 1 A121 47 A143 A152 1 A173 1 A192 A201 1
A11 42 A32 A43 7174 A65 A74 4 A92 A101 3 A123 30 A143 A152 1 A174 1 A192 A201 2
A11 10 A34 A42 2132 A65 A72 2 A92 A102 3 A121 27 A143 A151 2 A173 1 A191 A202 1
A11 33 A34 A42 4281 A63 A73 1 A92 A101 4 A123 23 A143 A152 2 A173 1 A191 A201 2
A12 12 A34 A40 2366 A63 A74 3 A91 A101 3 A123 36 A143 A152 1 A174 1 A192 A201 1
A11 21 A32 A43 1835 A61 A73 3 A92 A101 2 A121 25 A143 A152 2 A173 1 A192 A201 2
A14 24 A34 A41 3868 A61 A75 4 A92 A101 2 A123 41 A143 A151 2 A174 1 A192 A201 1
A14 12 A32 A42 1768 A61 A73 3 A93 A101 2 A121 24 A143 A151 1 A172 1 A191 A201 1
A13 10 A34 A40 781 A61 A75 4 A93 A101 4 A124 63 A143 A153 2 A173 1 A192 A201 1
A12 18 A32 A42 1924 A65 A72 4 A92 A101 3 A121 27 A143 A151 1 A173 1 A191 A201 2
A11 12 A34 A40 2121 A61 A73 4 A93 A101 2 A122 30 A143 A152 2 A173 1 A191 A201 1
A11 12 A32 A43 701 A61 A73 4 A94 A101 2 A121 40 A143 A152 1 A172 1 A191 A201 1
A12 12 A32 A45 639 A61 A73 4 A93 A101 2 A123 30 A143 A152 1 A173 1 A191 A201 2
A12 12 A34 A41 1860 A61 A71 4 A93 A101 2 A123 34 A143 A152 2 A174 1 A192 A201 1
A11 12 A34 A40 3499 A61 A73 3 A92 A102 2 A121 29 A143 A152 2 A173 1 A191 A201 2
A12 48 A32 A40 8487 A65 A74 1 A92 A101 2 A123 24 A143 A152 1 A173 1 A191 A201 1
A11 36 A33 A46 6887 A61 A73 4 A93 A101 3 A122 29 A142 A152 1 A173 1 A192 A201 2
A14 15 A32 A42 2708 A61 A72 2 A93 A101 3 A122 27 A141 A152 2 A172 1 A191 A201 1
A14 18 A32 A42 1984 A61 A73 4 A93 A101 4 A124 47 A141 A153 2 A173 1 A191 A201 1
A14 60 A32 A43 10144 A62 A74 2 A92 A101 4 A121 21 A143 A152 1 A173 1 A192 A201 1
A14 12 A34 A43 1240 A65 A75 4 A92 A101 2 A121 38 A143 A152 2 A173 1 A192 A201 1

A14 27 A33 A41 8613 A64 A73 2 A93 A101 2 A123 27 A143 A152 2 A173 1 A191 A201 1
A12 12 A32 A43 766 A63 A73 4 A93 A101 3 A121 66 A143 A152 1 A172 1 A191 A201 2
A12 15 A34 A43 2728 A65 A74 4 A93 A103 2 A121 35 A141 A152 3 A173 1 A192 A201 1
A13 12 A32 A43 1881 A61 A73 2 A92 A101 2 A123 44 A143 A151 1 A172 1 A192 A201 1
A13 6 A32 A40 709 A64 A72 2 A94 A101 2 A121 27 A143 A152 1 A171 1 A191 A202 1
A12 36 A32 A43 4795 A61 A72 4 A92 A101 1 A124 30 A143 A152 1 A174 1 A192 A201 1
A11 27 A32 A43 3416 A61 A73 3 A93 A101 2 A123 27 A143 A152 1 A174 1 A191 A201 1
A11 18 A32 A42 2462 A61 A73 2 A93 A101 2 A123 22 A143 A152 1 A173 1 A191 A201 2
A14 21 A34 A42 2288 A61 A72 4 A92 A101 4 A122 23 A143 A152 1 A173 1 A192 A201 1
A12 48 A31 A49 3566 A62 A74 4 A93 A101 2 A123 30 A143 A152 1 A173 1 A191 A201 1
A11 6 A34 A40 860 A61 A75 1 A92 A101 4 A124 39 A143 A152 2 A173 1 A192 A201 1
A14 12 A34 A40 682 A62 A74 4 A92 A101 3 A123 51 A143 A152 2 A173 1 A192 A201 1
A11 36 A34 A42 5371 A61 A73 3 A93 A103 2 A122 28 A143 A152 2 A173 1 A191 A201 1
A14 18 A34 A43 1582 A64 A75 4 A93 A101 4 A123 46 A143 A152 2 A173 1 A191 A201 1
A14 6 A32 A43 1346 A62 A75 2 A93 A101 4 A124 42 A141 A153 1 A173 2 A192 A201 1
A14 10 A32 A43 1924 A61 A73 1 A93 A101 4 A122 38 A143 A152 1 A173 1 A192 A202 1
A13 36 A32 A43 5848 A61 A73 4 A93 A101 1 A123 24 A143 A152 1 A173 1 A191 A201 1
A12 24 A34 A41 7758 A64 A75 2 A92 A101 4 A124 29 A143 A151 1 A173 1 A191 A201 1
A12 24 A33 A49 6967 A62 A74 4 A93 A101 4 A123 36 A143 A151 1 A174 1 A192 A201 1
A11 12 A32 A42 1282 A61 A73 2 A92 A101 4 A123 20 A143 A151 1 A173 1 A191 A201 2
A11 9 A34 A45 1288 A62 A75 3 A93 A103 4 A121 48 A143 A152 2 A173 2 A191 A202 1
A11 12 A31 A48 339 A61 A75 4 A94 A101 1 A123 45 A141 A152 1 A172 1 A191 A201 1
A12 24 A32 A40 3512 A62 A74 2 A93 A101 3 A123 38 A141 A152 2 A173 1 A192 A201 1
A14 6 A34 A43 1898 A65 A73 1 A93 A101 2 A121 34 A143 A152 2 A172 2 A191 A201 1
A14 24 A34 A43 2872 A62 A75 3 A93 A101 4 A121 36 A143 A152 1 A173 2 A192 A201 1
A14 18 A34 A40 1055 A61 A72 4 A92 A101 1 A122 30 A143 A152 2 A173 1 A191 A201 1
A14 15 A32 A44 1262 A63 A74 4 A93 A101 3 A122 36 A143 A152 2 A173 1 A192 A201 1
A12 10 A32 A40 7308 A61 A71 2 A93 A101 4 A124 70 A141 A153 1 A174 1 A192 A201 1
A14 36 A32 A40 909 A63 A75 4 A93 A101 4 A122 36 A143 A152 1 A173 1 A191 A201 1
A14 6 A32 A42 2978 A63 A73 1 A93 A101 2 A123 32 A143 A152 1 A173 1 A192 A201 1
A11 18 A32 A42 1131 A61 A71 4 A92 A101 2 A123 33 A143 A152 1 A173 1 A191 A201 2
A12 11 A32 A42 1577 A64 A72 4 A92 A101 1 A121 20 A143 A152 1 A173 1 A191 A201 1
A14 24 A32 A42 3972 A61 A74 2 A92 A101 4 A122 25 A143 A151 1 A173 1 A192 A201 1
A12 24 A34 A49 1935 A61 A75 4 A91 A101 4 A121 31 A143 A152 2 A173 1 A192 A201 2
A11 15 A30 A40 950 A61 A75 4 A93 A101 3 A123 33 A143 A151 2 A173 2 A191 A201 2
A14 12 A32 A42 763 A61 A73 4 A92 A101 1 A121 26 A143 A152 1 A173 1 A192 A201 1
A12 24 A33 A42 2064 A61 A71 3 A92 A101 2 A122 34 A143 A152 1 A174 1 A192 A201 2
A12 8 A32 A43 1414 A61 A73 4 A93 A103 2 A121 33 A143 A152 1 A173 1 A191 A202 1
A11 21 A33 A46 3414 A61 A72 2 A93 A101 1 A122 26 A143 A152 2 A173 1 A191 A201 2
A14 30 A31 A41 7485 A65 A71 4 A92 A101 1 A121 53 A141 A152 1 A174 1 A192 A201 2
A11 12 A32 A42 2577 A61 A73 2 A91 A101 1 A123 42 A143 A152 1 A173 1 A191 A201 1
A11 6 A34 A43 338 A63 A75 4 A93 A101 4 A123 52 A143 A152 2 A173 1 A191 A201 1
A14 12 A32 A43 1963 A61 A74 4 A93 A101 2 A123 31 A143 A151 2 A174 2 A192 A201 1
A11 21 A34 A40 571 A61 A75 4 A93 A101 4 A121 65 A143 A152 2 A173 1 A191 A201 1
A14 36 A33 A49 9572 A61 A72 1 A91 A101 1 A123 28 A143 A152 2 A173 1 A191 A201 2
A12 36 A33 A49 4455 A61 A73 2 A91 A101 2 A121 30 A142 A152 2 A174 1 A192 A201 2

A11 21 A31 A40 1647 A65 A73 4 A93 A101 2 A122 40 A143 A152 2 A172 2 A191 A201 2
A14 24 A34 A42 3777 A64 A73 4 A93 A101 4 A121 50 A143 A152 1 A173 1 A192 A201 1
A12 18 A34 A40 884 A61 A75 4 A93 A101 4 A123 36 A141 A152 1 A173 2 A192 A201 2
A14 15 A34 A43 1360 A61 A73 4 A93 A101 2 A122 31 A143 A152 2 A173 1 A191 A201 1
A12 9 A31 A41 5129 A61 A75 2 A92 A101 4 A124 74 A141 A153 1 A174 2 A192 A201 2
A12 16 A34 A40 1175 A61 A71 2 A93 A101 3 A123 68 A143 A153 3 A171 1 A192 A201 1
A11 12 A32 A43 674 A62 A74 4 A94 A101 1 A122 20 A143 A152 1 A173 1 A191 A201 2
A12 18 A30 A42 3244 A61 A73 1 A92 A101 4 A123 33 A141 A152 2 A173 1 A192 A201 1
A14 24 A32 A49 4591 A64 A73 2 A93 A101 3 A122 54 A143 A152 3 A174 1 A192 A201 2
A12 48 A30 A49 3844 A62 A74 4 A93 A101 4 A124 34 A143 A153 1 A172 2 A191 A201 2
A12 27 A32 A49 3915 A61 A73 4 A93 A101 2 A123 36 A143 A152 1 A173 2 A192 A201 2
A14 6 A32 A43 2108 A61 A74 2 A94 A101 2 A121 29 A143 A151 1 A173 1 A191 A201 1
A12 45 A32 A43 3031 A62 A73 4 A93 A103 4 A122 21 A143 A151 1 A173 1 A191 A201 2
A12 9 A34 A46 1501 A61 A75 2 A92 A101 3 A123 34 A143 A152 2 A174 1 A192 A201 2
A14 6 A34 A43 1382 A61 A73 1 A92 A101 1 A123 28 A143 A152 2 A173 1 A192 A201 1
A12 12 A32 A42 951 A62 A72 4 A92 A101 4 A123 27 A141 A151 4 A173 1 A191 A201 2
A12 24 A32 A41 2760 A65 A75 4 A93 A101 4 A124 36 A141 A153 1 A173 1 A192 A201 1
A12 18 A33 A42 4297 A61 A75 4 A91 A101 3 A124 40 A143 A152 1 A174 1 A192 A201 2
A14 9 A34 A46 936 A63 A75 4 A93 A101 2 A123 52 A143 A152 2 A173 1 A192 A201 1
A11 12 A32 A40 1168 A61 A73 4 A94 A101 3 A121 27 A143 A152 1 A172 1 A191 A201 1
A14 27 A33 A49 5117 A61 A74 3 A93 A101 4 A123 26 A143 A152 2 A173 1 A191 A201 1
A11 12 A32 A48 902 A61 A74 4 A94 A101 4 A122 21 A143 A151 1 A173 1 A191 A201 2
A14 12 A34 A40 1495 A61 A75 4 A93 A101 1 A121 38 A143 A152 2 A172 2 A191 A201 1
A11 30 A34 A41 10623 A61 A75 3 A93 A101 4 A124 38 A143 A153 3 A174 2 A192 A201 1
A14 12 A34 A42 1935 A61 A75 4 A93 A101 4 A121 43 A143 A152 3 A173 1 A192 A201 1
A12 12 A34 A44 1424 A61 A74 4 A93 A101 3 A122 26 A143 A152 1 A173 1 A191 A201 1
A11 24 A32 A49 6568 A61 A73 2 A94 A101 2 A123 21 A142 A152 1 A172 1 A191 A201 1
A14 12 A32 A41 1413 A64 A74 3 A93 A101 2 A122 55 A143 A152 1 A173 1 A191 A202 1
A14 9 A34 A43 3074 A65 A73 1 A93 A101 2 A121 33 A143 A152 2 A173 2 A191 A201 1
A14 36 A32 A43 3835 A65 A75 2 A92 A101 4 A121 45 A143 A152 1 A172 1 A192 A201 1
A11 27 A30 A49 5293 A61 A71 2 A93 A101 4 A122 50 A142 A152 2 A173 1 A192 A201 2
A13 30 A33 A49 1908 A61 A75 4 A93 A101 4 A121 66 A143 A152 1 A174 1 A192 A201 2
A14 36 A34 A43 3342 A65 A75 4 A93 A101 2 A123 51 A143 A152 1 A173 1 A192 A201 1
A12 6 A34 A48 932 A65 A74 1 A92 A101 3 A122 39 A143 A152 2 A172 1 A191 A201 1
A11 18 A30 A49 3104 A61 A74 3 A93 A101 1 A122 31 A141 A152 1 A173 1 A192 A201 1
A13 36 A32 A43 3913 A61 A73 2 A93 A101 2 A121 23 A143 A152 1 A173 1 A192 A201 1
A11 24 A32 A42 3021 A61 A73 2 A91 A101 2 A121 24 A143 A151 1 A172 1 A191 A201 1
A14 10 A32 A40 1364 A61 A73 2 A92 A101 4 A123 64 A143 A152 1 A173 1 A192 A201 1
A12 12 A32 A43 625 A61 A72 4 A94 A103 1 A121 26 A141 A152 1 A172 1 A191 A201 1
A11 12 A32 A46 1200 A65 A73 4 A92 A101 4 A122 23 A141 A151 1 A173 1 A192 A201 1
A14 12 A32 A43 707 A61 A73 4 A93 A101 2 A121 30 A141 A152 2 A173 1 A191 A201 1
A14 24 A33 A49 2978 A65 A73 4 A93 A101 4 A121 32 A143 A152 2 A173 2 A192 A201 1
A14 15 A32 A41 4657 A61 A73 3 A93 A101 2 A123 30 A143 A152 1 A173 1 A192 A201 1
A14 36 A30 A45 2613 A61 A73 4 A93 A101 2 A123 27 A143 A152 2 A173 1 A191 A201 1
A12 48 A32 A43 10961 A64 A74 1 A93 A102 2 A124 27 A141 A152 2 A173 1 A192 A201 2
A11 12 A32 A42 7865 A61 A75 4 A93 A101 4 A124 53 A143 A153 1 A174 1 A192 A201 2

A14 9 A32 A43 1478 A61 A74 4 A93 A101 2 A123 22 A143 A152 1 A173 1 A191 A201 2
 A11 24 A32 A42 3149 A61 A72 4 A93 A101 1 A124 22 A141 A153 1 A173 1 A191 A201 1
 A13 36 A32 A43 4210 A61 A73 4 A93 A101 2 A123 26 A143 A152 1 A173 1 A191 A201 2
 A14 9 A32 A40 2507 A63 A75 2 A93 A101 4 A124 51 A143 A153 1 A172 1 A191 A201 1
 A14 12 A32 A43 2141 A62 A74 3 A93 A101 1 A124 35 A143 A152 1 A173 1 A191 A201 1
 A12 18 A32 A43 866 A61 A73 4 A94 A103 2 A121 25 A143 A152 1 A172 1 A191 A201 1
 A14 4 A34 A43 1544 A61 A74 2 A93 A101 1 A121 42 A143 A152 3 A172 2 A191 A201 1
 A11 24 A32 A43 1823 A61 A71 4 A93 A101 2 A123 30 A142 A152 1 A174 2 A191 A201 2
 A12 6 A32 A40 14555 A65 A71 1 A93 A101 2 A122 23 A143 A152 1 A171 1 A192 A201 2
 A12 21 A32 A49 2767 A62 A75 4 A91 A101 2 A123 61 A141 A151 2 A172 1 A191 A201 2
 A14 12 A34 A43 1291 A61 A73 4 A92 A101 2 A122 35 A143 A152 2 A173 1 A191 A201 1
 A11 30 A32 A43 2522 A61 A75 1 A93 A103 3 A122 39 A143 A152 1 A173 2 A191 A201 1
 A11 24 A32 A40 915 A65 A75 4 A92 A101 2 A123 29 A141 A152 1 A173 1 A191 A201 2
 A14 6 A32 A43 1595 A61 A74 3 A93 A101 2 A122 51 A143 A152 1 A173 2 A191 A201 1
 A11 48 A30 A41 4605 A61 A75 3 A93 A101 4 A124 24 A143 A153 2 A173 2 A191 A201 2
 A14 12 A34 A49 1185 A61 A73 3 A92 A101 2 A121 27 A143 A152 2 A173 1 A191 A201 1
 A14 12 A31 A48 3447 A63 A73 4 A92 A101 3 A121 35 A143 A152 1 A172 2 A191 A201 1
 A14 24 A32 A49 1258 A61 A74 4 A93 A101 1 A121 25 A143 A152 1 A173 1 A192 A201 1
 A14 12 A34 A43 717 A61 A75 4 A93 A101 4 A121 52 A143 A152 3 A173 1 A191 A201 1
 A14 6 A30 A40 1204 A62 A73 4 A93 A101 1 A124 35 A141 A151 1 A173 1 A191 A202 1
 A13 24 A32 A42 1925 A61 A73 2 A93 A101 2 A121 26 A143 A152 1 A173 1 A191 A201 1
 A14 18 A32 A43 433 A61 A71 3 A92 A102 4 A121 22 A143 A151 1 A173 1 A191 A201 2
 A11 6 A34 A40 666 A64 A74 3 A92 A101 4 A121 39 A143 A152 2 A172 1 A192 A201 1
 A13 12 A32 A42 2251 A61 A73 1 A92 A101 2 A123 46 A143 A152 1 A172 1 A191 A201 1
 A12 30 A32 A40 2150 A61 A73 4 A92 A103 2 A124 24 A141 A152 1 A173 1 A191 A201 2
 A14 24 A33 A42 4151 A62 A73 2 A93 A101 3 A122 35 A143 A152 2 A173 1 A191 A201 1
 A12 9 A32 A42 2030 A65 A74 2 A93 A101 1 A123 24 A143 A152 1 A173 1 A192 A201 1
 A12 60 A33 A43 7418 A65 A73 1 A93 A101 1 A121 27 A143 A152 1 A172 1 A191 A201 1
 A14 24 A34 A43 2684 A61 A73 4 A93 A101 2 A121 35 A143 A152 2 A172 1 A191 A201 1
 A11 12 A31 A43 2149 A61 A73 4 A91 A101 1 A124 29 A143 A153 1 A173 1 A191 A201 2
 A14 15 A32 A41 3812 A62 A72 1 A92 A101 4 A123 23 A143 A152 1 A173 1 A192 A201 1
 A14 11 A34 A43 1154 A62 A71 4 A92 A101 4 A121 57 A143 A152 3 A172 1 A191 A201 1
 A11 12 A32 A42 1657 A61 A73 2 A93 A101 2 A121 27 A143 A152 1 A173 1 A191 A201 1
 A11 24 A32 A43 1603 A61 A75 4 A92 A101 4 A123 55 A143 A152 1 A173 1 A191 A201 1
 A11 18 A34 A40 5302 A61 A75 2 A93 A101 4 A124 36 A143 A153 3 A174 1 A192 A201 1
 A14 12 A34 A46 2748 A61 A75 2 A92 A101 4 A124 57 A141 A153 3 A172 1 A191 A201 1
 A14 10 A34 A40 1231 A61 A75 3 A93 A101 4 A121 32 A143 A152 2 A172 2 A191 A202 1
 A12 15 A32 A43 802 A61 A75 4 A93 A101 3 A123 37 A143 A152 1 A173 2 A191 A201 2
 A14 36 A34 A49 6304 A65 A75 4 A93 A101 4 A121 36 A143 A152 2 A173 1 A191 A201 1
 A14 24 A32 A43 1533 A61 A72 4 A92 A101 3 A123 38 A142 A152 1 A173 1 A192 A201 1
 A11 14 A32 A40 8978 A61 A75 1 A91 A101 4 A122 45 A143 A152 1 A174 1 A192 A202 2
 A14 24 A32 A43 999 A65 A75 4 A93 A101 2 A123 25 A143 A152 2 A173 1 A191 A201 1
 A14 18 A32 A40 2662 A65 A74 4 A93 A101 3 A122 32 A143 A152 1 A173 1 A191 A202 1
 A14 12 A34 A42 1402 A63 A74 3 A92 A101 4 A123 37 A143 A151 1 A173 1 A192 A201 1
 A12 48 A31 A40 12169 A65 A71 4 A93 A102 4 A124 36 A143 A153 1 A174 1 A192 A201 1
 A12 48 A32 A43 3060 A61 A74 4 A93 A101 4 A121 28 A143 A152 2 A173 1 A191 A201 2

A11 30 A32 A45 11998 A61 A72 1 A91 A101 1 A124 34 A143 A152 1 A172 1 A192 A201 2
A14 9 A32 A43 2697 A61 A73 1 A93 A101 2 A121 32 A143 A152 1 A173 2 A191 A201 1
A14 18 A34 A43 2404 A61 A73 2 A92 A101 2 A123 26 A143 A152 2 A173 1 A191 A201 1
A11 12 A32 A42 1262 A65 A75 2 A91 A101 4 A122 49 A143 A152 1 A172 1 A192 A201 1
A14 6 A32 A42 4611 A61 A72 1 A92 A101 4 A122 32 A143 A152 1 A173 1 A191 A201 2
A14 24 A32 A43 1901 A62 A73 4 A93 A101 4 A123 29 A143 A151 1 A174 1 A192 A201 1
A14 15 A34 A41 3368 A64 A75 3 A93 A101 4 A124 23 A143 A151 2 A173 1 A192 A201 1
A14 12 A32 A42 1574 A61 A73 4 A93 A101 2 A121 50 A143 A152 1 A173 1 A191 A201 1
A13 18 A31 A43 1445 A65 A74 4 A93 A101 4 A123 49 A141 A152 1 A172 1 A191 A201 1
A14 15 A34 A42 1520 A65 A75 4 A93 A101 4 A122 63 A143 A152 1 A173 1 A191 A201 1
A12 24 A34 A40 3878 A62 A72 4 A91 A101 2 A123 37 A143 A152 1 A173 1 A192 A201 1
A11 47 A32 A40 10722 A61 A72 1 A92 A101 1 A121 35 A143 A152 1 A172 1 A192 A201 1
A11 48 A32 A41 4788 A61 A74 4 A93 A101 3 A122 26 A143 A152 1 A173 2 A191 A201 1
A12 48 A33 A410 7582 A62 A71 2 A93 A101 4 A124 31 A143 A153 1 A174 1 A192 A201 1
A12 12 A32 A43 1092 A61 A73 4 A92 A103 4 A121 49 A143 A152 2 A173 1 A192 A201 1
A11 24 A33 A43 1024 A61 A72 4 A94 A101 4 A121 48 A142 A152 1 A173 1 A191 A201 2
A14 12 A32 A49 1076 A61 A73 2 A94 A101 2 A121 26 A143 A152 1 A173 1 A192 A202 1
A12 36 A32 A41 9398 A61 A72 1 A94 A101 4 A123 28 A143 A151 1 A174 1 A192 A201 2
A11 24 A34 A41 6419 A61 A75 2 A92 A101 4 A124 44 A143 A153 2 A174 2 A192 A201 1
A13 42 A34 A41 4796 A61 A75 4 A93 A101 4 A124 56 A143 A153 1 A173 1 A191 A201 1
A14 48 A34 A49 7629 A65 A75 4 A91 A101 2 A123 46 A141 A152 2 A174 2 A191 A201 1
A12 48 A32 A42 9960 A61 A72 1 A92 A101 2 A123 26 A143 A152 1 A173 1 A192 A201 2
A14 12 A32 A41 4675 A65 A72 1 A92 A101 4 A123 20 A143 A151 1 A173 1 A191 A201 1
A14 10 A32 A40 1287 A65 A75 4 A93 A102 2 A122 45 A143 A152 1 A172 1 A191 A202 1
A14 18 A32 A42 2515 A61 A73 3 A93 A101 4 A121 43 A143 A152 1 A173 1 A192 A201 1
A12 21 A34 A42 2745 A64 A74 3 A93 A101 2 A123 32 A143 A152 2 A173 1 A192 A201 1
A14 6 A32 A40 672 A61 A71 1 A92 A101 4 A121 54 A143 A152 1 A171 1 A192 A201 1
A12 36 A30 A43 3804 A61 A73 4 A92 A101 1 A123 42 A143 A152 1 A173 1 A192 A201 2
A13 24 A34 A40 1344 A65 A74 4 A93 A101 2 A121 37 A141 A152 2 A172 2 A191 A201 2
A11 10 A34 A40 1038 A61 A74 4 A93 A102 3 A122 49 A143 A152 2 A173 1 A192 A201 1
A14 48 A34 A40 10127 A63 A73 2 A93 A101 2 A124 44 A141 A153 1 A173 1 A191 A201 2
A14 6 A32 A42 1543 A64 A73 4 A91 A101 2 A121 33 A143 A152 1 A173 1 A191 A201 1
A14 30 A32 A41 4811 A65 A74 2 A92 A101 4 A122 24 A142 A151 1 A172 1 A191 A201 1
A11 12 A32 A43 727 A62 A72 4 A94 A101 3 A124 33 A143 A152 1 A172 1 A192 A201 2
A12 8 A32 A42 1237 A61 A73 3 A92 A101 4 A121 24 A143 A152 1 A173 1 A191 A201 2
A12 9 A32 A40 276 A61 A73 4 A94 A101 4 A121 22 A143 A151 1 A172 1 A191 A201 1
A12 48 A32 A410 5381 A65 A71 3 A93 A101 4 A124 40 A141 A153 1 A171 1 A192 A201 1
A14 24 A32 A42 5511 A62 A73 4 A93 A101 1 A123 25 A142 A152 1 A173 1 A191 A201 1
A13 24 A32 A42 3749 A61 A72 2 A92 A101 4 A123 26 A143 A152 1 A173 1 A191 A201 1
A12 12 A32 A40 685 A61 A74 2 A94 A101 3 A123 25 A141 A152 1 A172 1 A191 A201 2
A13 4 A32 A40 1494 A65 A72 1 A93 A101 2 A121 29 A143 A152 1 A172 2 A191 A202 1
A11 36 A31 A42 2746 A61 A75 4 A93 A101 4 A123 31 A141 A152 1 A173 1 A191 A201 2
A11 12 A32 A42 708 A61 A73 2 A93 A103 3 A122 38 A143 A152 1 A172 2 A191 A201 1
A12 24 A32 A42 4351 A65 A73 1 A92 A101 4 A122 48 A143 A152 1 A172 1 A192 A201 1
A14 12 A34 A46 701 A61 A73 4 A93 A101 2 A123 32 A143 A152 2 A173 1 A191 A201 1
A11 15 A33 A42 3643 A61 A75 1 A92 A101 4 A122 27 A143 A152 2 A172 1 A191 A201 1

A12 30 A34 A40 4249 A61 A71 4 A94 A101 2 A123 28 A143 A152 2 A174 1 A191 A201 2
A11 24 A32 A43 1938 A61 A72 4 A91 A101 3 A122 32 A143 A152 1 A173 1 A191 A201 2
A11 24 A32 A41 2910 A61 A74 2 A93 A101 1 A124 34 A143 A153 1 A174 1 A192 A201 1
A11 18 A32 A42 2659 A64 A73 4 A93 A101 2 A123 28 A143 A152 1 A173 1 A191 A201 1
A14 18 A34 A40 1028 A61 A73 4 A92 A101 3 A121 36 A143 A152 2 A173 1 A191 A201 1
A11 8 A34 A40 3398 A61 A74 1 A93 A101 4 A121 39 A143 A152 2 A172 1 A191 A202 1
A14 12 A34 A42 5801 A65 A75 2 A93 A101 4 A122 49 A143 A151 1 A173 1 A192 A201 1
A14 24 A32 A40 1525 A64 A74 4 A92 A101 3 A123 34 A143 A152 1 A173 2 A192 A201 1
A13 36 A32 A43 4473 A61 A75 4 A93 A101 2 A123 31 A143 A152 1 A173 1 A191 A201 1
A12 6 A32 A43 1068 A61 A75 4 A93 A101 4 A123 28 A143 A152 1 A173 2 A191 A201 1
A11 24 A34 A41 6615 A61 A71 2 A93 A101 4 A124 75 A143 A153 2 A174 1 A192 A201 1
A14 18 A34 A46 1864 A62 A73 4 A92 A101 2 A121 30 A143 A152 2 A173 1 A191 A201 2
A12 60 A32 A40 7408 A62 A72 4 A92 A101 2 A122 24 A143 A152 1 A174 1 A191 A201 2
A14 48 A34 A41 11590 A62 A73 2 A92 A101 4 A123 24 A141 A151 2 A172 1 A191 A201 2
A11 24 A30 A42 4110 A61 A75 3 A93 A101 4 A124 23 A141 A151 2 A173 2 A191 A201 2
A11 6 A34 A42 3384 A61 A73 1 A91 A101 4 A121 44 A143 A151 1 A174 1 A192 A201 2
A12 13 A32 A43 2101 A61 A72 2 A92 A103 4 A122 23 A143 A152 1 A172 1 A191 A201 1
A11 15 A32 A44 1275 A65 A73 4 A92 A101 2 A123 24 A143 A151 1 A173 1 A191 A201 2
A11 24 A32 A42 4169 A61 A73 4 A93 A101 4 A122 28 A143 A152 1 A173 1 A191 A201 1
A12 10 A32 A42 1521 A61 A73 4 A91 A101 2 A123 31 A143 A152 1 A172 1 A191 A201 1
A12 24 A34 A46 5743 A61 A72 2 A92 A101 4 A124 24 A143 A153 2 A173 1 A192 A201 1
A11 21 A32 A42 3599 A61 A74 1 A92 A101 4 A123 26 A143 A151 1 A172 1 A191 A201 1
A12 18 A32 A43 3213 A63 A72 1 A94 A101 3 A121 25 A143 A151 1 A173 1 A191 A201 1
A12 18 A32 A49 4439 A61 A75 1 A93 A102 1 A121 33 A141 A152 1 A174 1 A192 A201 1
A13 10 A32 A40 3949 A61 A72 1 A93 A103 1 A122 37 A143 A152 1 A172 2 A191 A201 1
A14 15 A34 A43 1459 A61 A73 4 A92 A101 2 A123 43 A143 A152 1 A172 1 A191 A201 1
A12 13 A34 A43 882 A61 A72 4 A93 A103 4 A121 23 A143 A152 2 A173 1 A191 A201 1
A12 24 A32 A43 3758 A63 A71 1 A92 A101 4 A124 23 A143 A151 1 A171 1 A191 A201 1
A14 6 A33 A49 1743 A62 A73 1 A93 A101 2 A121 34 A143 A152 2 A172 1 A191 A201 1
A12 9 A34 A46 1136 A64 A75 4 A93 A101 3 A124 32 A143 A153 2 A173 2 A191 A201 2
A14 9 A32 A44 1236 A61 A72 1 A92 A101 4 A121 23 A143 A151 1 A173 1 A192 A201 1
A12 9 A32 A42 959 A61 A73 1 A92 A101 2 A123 29 A143 A152 1 A173 1 A191 A202 2
A14 18 A34 A41 3229 A65 A71 2 A93 A101 4 A124 38 A143 A152 1 A174 1 A192 A201 1
A11 12 A30 A43 6199 A61 A73 4 A93 A101 2 A122 28 A143 A151 2 A173 1 A192 A201 2
A14 10 A32 A46 727 A63 A75 4 A93 A101 4 A124 46 A143 A153 1 A173 1 A192 A201 1
A12 24 A32 A40 1246 A61 A72 4 A93 A101 2 A121 23 A142 A152 1 A172 1 A191 A201 2
A14 12 A34 A43 2331 A65 A75 1 A93 A102 4 A121 49 A143 A152 1 A173 1 A192 A201 1
A14 36 A33 A43 4463 A61 A73 4 A93 A101 2 A123 26 A143 A152 2 A174 1 A192 A201 2
A14 12 A32 A43 776 A61 A73 4 A94 A101 2 A121 28 A143 A152 1 A173 1 A191 A201 1
A11 30 A32 A42 2406 A61 A74 4 A92 A101 4 A121 23 A143 A151 1 A173 1 A191 A201 2
A12 18 A32 A46 1239 A65 A73 4 A93 A101 4 A124 61 A143 A153 1 A173 1 A191 A201 1
A13 12 A32 A43 3399 A65 A75 2 A93 A101 3 A123 37 A143 A152 1 A174 1 A191 A201 1
A13 12 A33 A40 2247 A61 A73 2 A92 A101 2 A123 36 A142 A152 2 A173 1 A192 A201 1
A14 6 A32 A42 1766 A61 A73 1 A94 A101 2 A122 21 A143 A151 1 A173 1 A191 A201 1
A11 18 A32 A42 2473 A61 A71 4 A93 A101 1 A123 25 A143 A152 1 A171 1 A191 A201 2
A14 12 A32 A49 1542 A61 A74 2 A93 A101 4 A123 36 A143 A152 1 A173 1 A192 A201 1

A14 18 A34 A41 3850 A61 A74 3 A93 A101 1 A123 27 A143 A152 2 A173 1 A191 A201 1
A11 18 A32 A42 3650 A61 A72 1 A92 A101 4 A123 22 A143 A151 1 A173 1 A191 A201 1
A11 36 A32 A42 3446 A61 A75 4 A93 A101 2 A123 42 A143 A152 1 A173 2 A191 A201 2
A12 18 A32 A42 3001 A61 A74 2 A92 A101 4 A121 40 A143 A151 1 A173 1 A191 A201 1
A14 36 A32 A40 3079 A65 A73 4 A93 A101 4 A121 36 A143 A152 1 A173 1 A191 A201 1
A14 18 A34 A43 6070 A61 A75 3 A93 A101 4 A123 33 A143 A152 2 A173 1 A192 A201 1
A14 10 A34 A42 2146 A61 A72 1 A92 A101 3 A121 23 A143 A151 2 A173 1 A191 A201 1
A14 60 A34 A40 13756 A65 A75 2 A93 A101 4 A124 63 A141 A153 1 A174 1 A192 A201 1
A12 60 A31 A410 14782 A62 A75 3 A92 A101 4 A124 60 A141 A153 2 A174 1 A192 A201 2
A11 48 A31 A49 7685 A61 A74 2 A92 A103 4 A123 37 A143 A151 1 A173 1 A191 A201 2
A14 18 A33 A43 2320 A61 A71 2 A94 A101 3 A121 34 A143 A152 2 A173 1 A191 A201 1
A14 7 A33 A43 846 A65 A75 3 A93 A101 4 A124 36 A143 A153 1 A173 1 A191 A201 1
A12 36 A32 A40 14318 A61 A75 4 A93 A101 2 A124 57 A143 A153 1 A174 1 A192 A201 2
A14 6 A34 A40 362 A62 A73 4 A92 A101 4 A123 52 A143 A152 2 A172 1 A191 A201 1
A11 20 A32 A42 2212 A65 A74 4 A93 A101 4 A123 39 A143 A152 1 A173 1 A192 A201 1
A12 18 A32 A41 12976 A61 A71 3 A92 A101 4 A124 38 A143 A153 1 A174 1 A192 A201 2
A14 22 A32 A40 1283 A65 A74 4 A92 A101 4 A122 25 A143 A151 1 A173 1 A191 A201 1
A13 12 A32 A40 1330 A61 A72 4 A93 A101 1 A121 26 A143 A152 1 A173 1 A191 A201 1
A14 30 A33 A49 4272 A62 A73 2 A93 A101 2 A122 26 A143 A152 2 A172 1 A191 A201 1
A14 18 A34 A43 2238 A61 A73 2 A92 A101 1 A123 25 A143 A152 2 A173 1 A191 A201 1
A14 18 A32 A43 1126 A65 A72 4 A92 A101 2 A121 21 A143 A151 1 A173 1 A192 A201 1
A12 18 A34 A42 7374 A61 A71 4 A93 A101 4 A122 40 A142 A152 2 A174 1 A192 A201 1
A12 15 A34 A49 2326 A63 A73 2 A93 A101 4 A123 27 A141 A152 1 A173 1 A191 A201 1
A14 9 A32 A49 1449 A61 A74 3 A92 A101 2 A123 27 A143 A152 2 A173 1 A191 A201 1
A14 18 A32 A40 1820 A61 A73 2 A94 A101 2 A122 30 A143 A152 1 A174 1 A192 A201 1
A12 12 A32 A42 983 A64 A72 1 A92 A101 4 A121 19 A143 A151 1 A172 1 A191 A201 1
A11 36 A32 A40 3249 A61 A74 2 A93 A101 4 A124 39 A141 A153 1 A174 2 A192 A201 1
A11 6 A34 A43 1957 A61 A74 1 A92 A101 4 A123 31 A143 A152 1 A173 1 A191 A201 1
A14 9 A34 A42 2406 A61 A71 2 A93 A101 3 A123 31 A143 A152 1 A174 1 A191 A201 1
A12 39 A33 A46 11760 A62 A74 2 A93 A101 3 A124 32 A143 A151 1 A173 1 A192 A201 1
A11 12 A32 A42 2578 A61 A71 3 A92 A101 4 A124 55 A143 A153 1 A174 1 A191 A201 1
A11 36 A34 A42 2348 A61 A73 3 A94 A101 2 A122 46 A143 A152 2 A173 1 A192 A201 1
A12 12 A32 A40 1223 A61 A75 1 A91 A101 1 A121 46 A143 A151 2 A173 1 A191 A201 2
A14 24 A34 A43 1516 A64 A73 4 A92 A101 1 A121 43 A143 A152 2 A172 1 A191 A201 1
A14 18 A32 A43 1473 A61 A72 3 A94 A101 4 A121 39 A143 A152 1 A173 1 A192 A201 1
A12 18 A34 A49 1887 A65 A73 4 A94 A101 4 A121 28 A141 A152 2 A173 1 A191 A201 1
A14 24 A33 A49 8648 A61 A72 2 A93 A101 2 A123 27 A141 A152 2 A173 1 A192 A201 2
A14 14 A33 A40 802 A61 A73 4 A93 A101 2 A123 27 A143 A152 2 A172 1 A191 A201 1
A12 18 A33 A40 2899 A65 A75 4 A93 A101 4 A123 43 A143 A152 1 A173 2 A191 A201 1
A12 24 A32 A43 2039 A61 A72 1 A94 A101 1 A122 22 A143 A152 1 A173 1 A192 A201 2
A14 24 A34 A41 2197 A65 A74 4 A93 A101 4 A123 43 A143 A152 2 A173 2 A192 A201 1
A11 15 A32 A43 1053 A61 A72 4 A94 A101 2 A121 27 A143 A152 1 A173 1 A191 A202 1
A14 24 A32 A43 3235 A63 A75 3 A91 A101 2 A123 26 A143 A152 1 A174 1 A192 A201 1
A13 12 A34 A40 939 A63 A74 4 A94 A101 2 A121 28 A143 A152 3 A173 1 A192 A201 2
A12 24 A32 A43 1967 A61 A75 4 A92 A101 4 A123 20 A143 A152 1 A173 1 A192 A201 1
A14 33 A34 A41 7253 A61 A74 3 A93 A101 2 A123 35 A143 A152 2 A174 1 A192 A201 1

A14 12 A34 A49 2292 A61 A71 4 A93 A101 2 A123 42 A142 A152 2 A174 1 A192 A201 2
A14 10 A32 A40 1597 A63 A73 3 A93 A101 2 A124 40 A143 A151 1 A172 2 A191 A202 1
A11 24 A32 A40 1381 A65 A73 4 A92 A101 2 A122 35 A143 A152 1 A173 1 A191 A201 2
A14 36 A34 A41 5842 A61 A75 2 A93 A101 2 A122 35 A143 A152 2 A173 2 A192 A201 1
A11 12 A32 A40 2579 A61 A72 4 A93 A101 1 A121 33 A143 A152 1 A172 2 A191 A201 2
A11 18 A33 A46 8471 A65 A73 1 A92 A101 2 A123 23 A143 A151 2 A173 1 A192 A201 1
A14 21 A32 A40 2782 A63 A74 1 A92 A101 2 A123 31 A141 A152 1 A174 1 A191 A201 1
A12 18 A32 A40 1042 A65 A73 4 A92 A101 2 A122 33 A143 A152 1 A173 1 A191 A201 2
A14 15 A32 A40 3186 A64 A74 2 A92 A101 3 A123 20 A143 A151 1 A173 1 A191 A201 1
A12 12 A32 A41 2028 A65 A73 4 A93 A101 2 A123 30 A143 A152 1 A173 1 A191 A201 1
A12 12 A34 A40 958 A61 A74 2 A93 A101 3 A121 47 A143 A152 2 A172 2 A191 A201 1
A14 21 A33 A42 1591 A62 A74 4 A93 A101 3 A121 34 A143 A152 2 A174 1 A191 A201 1
A12 12 A32 A42 2762 A65 A75 1 A92 A101 2 A122 25 A141 A152 1 A173 1 A192 A201 2
A12 18 A32 A41 2779 A61 A73 1 A94 A101 3 A123 21 A143 A151 1 A173 1 A192 A201 1
A14 28 A34 A43 2743 A61 A75 4 A93 A101 2 A123 29 A143 A152 2 A173 1 A191 A201 1
A14 18 A34 A43 1149 A64 A73 4 A93 A101 3 A121 46 A143 A152 2 A173 1 A191 A201 1
A14 9 A32 A42 1313 A61 A75 1 A93 A101 4 A123 20 A143 A152 1 A173 1 A191 A201 1
A11 18 A34 A45 1190 A61 A71 2 A92 A101 4 A124 55 A143 A153 3 A171 2 A191 A201 2
A14 5 A32 A49 3448 A61 A74 1 A93 A101 4 A121 74 A143 A152 1 A172 1 A191 A201 1
A12 24 A32 A410 11328 A61 A73 2 A93 A102 3 A123 29 A141 A152 2 A174 1 A192 A201 2
A11 6 A34 A42 1872 A61 A71 4 A93 A101 4 A124 36 A143 A153 3 A174 1 A192 A201 1
A14 24 A34 A45 2058 A61 A73 4 A91 A101 2 A121 33 A143 A152 2 A173 1 A192 A201 1
A11 9 A32 A42 2136 A61 A73 3 A93 A101 2 A121 25 A143 A152 1 A173 1 A191 A201 1
A12 12 A32 A43 1484 A65 A73 2 A94 A101 1 A121 25 A143 A152 1 A173 1 A192 A201 2
A14 6 A32 A45 660 A63 A74 2 A94 A101 4 A121 23 A143 A151 1 A172 1 A191 A201 1
A14 24 A34 A40 1287 A64 A75 4 A92 A101 4 A121 37 A143 A152 2 A173 1 A192 A201 1
A11 42 A34 A45 3394 A61 A71 4 A93 A102 4 A123 65 A143 A152 2 A171 1 A191 A201 1
A13 12 A31 A49 609 A61 A72 4 A92 A101 1 A121 26 A143 A152 1 A171 1 A191 A201 2
A14 12 A32 A40 1884 A61 A75 4 A93 A101 4 A123 39 A143 A152 1 A174 1 A192 A201 1
A11 12 A32 A42 1620 A61 A73 2 A92 A102 3 A122 30 A143 A152 1 A173 1 A191 A201 1
A12 20 A33 A410 2629 A61 A73 2 A93 A101 3 A123 29 A141 A152 2 A173 1 A192 A201 1
A14 12 A32 A46 719 A61 A75 4 A93 A101 4 A123 41 A141 A152 1 A172 2 A191 A201 2
A12 48 A34 A42 5096 A61 A73 2 A92 A101 3 A123 30 A143 A152 1 A174 1 A192 A201 2
A14 9 A34 A46 1244 A65 A75 4 A92 A101 4 A122 41 A143 A151 2 A172 1 A191 A201 1
A11 36 A32 A40 1842 A61 A72 4 A92 A101 4 A123 34 A143 A152 1 A173 1 A192 A201 2
A12 7 A32 A43 2576 A61 A73 2 A93 A103 2 A121 35 A143 A152 1 A173 1 A191 A201 1
A13 12 A32 A42 1424 A65 A75 3 A92 A101 4 A121 55 A143 A152 1 A174 1 A192 A201 1
A12 15 A33 A45 1512 A64 A73 3 A94 A101 3 A122 61 A142 A152 2 A173 1 A191 A201 2
A14 36 A34 A41 11054 A65 A73 4 A93 A101 2 A123 30 A143 A152 1 A174 1 A192 A201 1
A14 6 A32 A43 518 A61 A73 3 A92 A101 1 A121 29 A143 A152 1 A173 1 A191 A201 1
A14 12 A30 A42 2759 A61 A75 2 A93 A101 4 A122 34 A143 A152 2 A173 1 A191 A201 1
A14 24 A32 A41 2670 A61 A75 4 A93 A101 4 A123 35 A143 A152 1 A174 1 A192 A201 1
A11 24 A32 A40 4817 A61 A74 2 A93 A102 3 A122 31 A143 A152 1 A173 1 A192 A201 2
A14 24 A32 A41 2679 A61 A72 4 A92 A101 1 A124 29 A143 A152 1 A174 1 A192 A201 1
A11 11 A34 A40 3905 A61 A73 2 A93 A101 2 A121 36 A143 A151 2 A173 2 A191 A201 1
A11 12 A32 A41 3386 A61 A75 3 A93 A101 4 A124 35 A143 A153 1 A173 1 A192 A201 2

A11 6 A32 A44 343 A61 A72 4 A92 A101 1 A121 27 A143 A152 1 A173 1 A191 A201 1
A14 18 A32 A43 4594 A61 A72 3 A93 A101 2 A123 32 A143 A152 1 A173 1 A192 A201 1
A11 36 A32 A42 3620 A61 A73 1 A93 A103 2 A122 37 A143 A152 1 A173 2 A191 A201 1
A11 15 A32 A40 1721 A61 A72 2 A93 A101 3 A121 36 A143 A152 1 A173 1 A191 A201 1
A12 12 A32 A42 3017 A61 A72 3 A92 A101 1 A121 34 A143 A151 1 A174 1 A191 A201 1
A12 12 A32 A48 754 A65 A75 4 A93 A101 4 A122 38 A143 A152 2 A173 1 A191 A201 1
A14 18 A32 A49 1950 A61 A74 4 A93 A101 1 A123 34 A142 A152 2 A173 1 A192 A201 1
A11 24 A32 A41 2924 A61 A73 3 A93 A103 4 A124 63 A141 A152 1 A173 2 A192 A201 1
A11 24 A33 A43 1659 A61 A72 4 A92 A101 2 A123 29 A143 A151 1 A172 1 A192 A201 2
A14 48 A33 A43 7238 A65 A75 3 A93 A101 3 A123 32 A141 A152 2 A173 2 A191 A201 1
A14 33 A33 A49 2764 A61 A73 2 A92 A101 2 A123 26 A143 A152 2 A173 1 A192 A201 1
A14 24 A33 A41 4679 A61 A74 3 A93 A101 3 A123 35 A143 A152 2 A172 1 A192 A201 1
A12 24 A32 A43 3092 A62 A72 3 A94 A101 2 A123 22 A143 A151 1 A173 1 A192 A201 2
A11 6 A32 A46 448 A61 A72 4 A92 A101 4 A122 23 A143 A152 1 A173 1 A191 A201 2
A11 9 A32 A40 654 A61 A73 4 A93 A101 3 A123 28 A143 A152 1 A172 1 A191 A201 2
A14 6 A32 A48 1238 A65 A71 4 A93 A101 4 A122 36 A143 A152 1 A174 2 A192 A201 1
A12 18 A34 A43 1245 A61 A73 4 A94 A101 2 A123 33 A143 A152 1 A173 1 A191 A201 2
A11 18 A30 A42 3114 A61 A72 1 A92 A101 4 A122 26 A143 A151 1 A173 1 A191 A201 2
A14 39 A32 A41 2569 A63 A73 4 A93 A101 4 A123 24 A143 A152 1 A173 1 A191 A201 1
A13 24 A32 A43 5152 A61 A74 4 A93 A101 2 A123 25 A141 A152 1 A173 1 A191 A201 1
A12 12 A32 A49 1037 A62 A74 3 A93 A101 4 A121 39 A143 A152 1 A172 1 A191 A201 1
A11 15 A34 A42 1478 A61 A75 4 A93 A101 4 A123 44 A143 A152 2 A173 2 A192 A201 1
A12 12 A34 A43 3573 A61 A73 1 A92 A101 1 A121 23 A143 A152 1 A172 1 A191 A201 1
A12 24 A32 A40 1201 A61 A72 4 A93 A101 1 A122 26 A143 A152 1 A173 1 A191 A201 1
A11 30 A32 A42 3622 A64 A75 4 A92 A101 4 A122 57 A143 A151 2 A173 1 A192 A201 1
A14 15 A33 A42 960 A64 A74 3 A92 A101 2 A122 30 A143 A152 2 A173 1 A191 A201 1
A14 12 A34 A40 1163 A63 A73 4 A93 A101 4 A121 44 A143 A152 1 A173 1 A192 A201 1
A12 6 A33 A40 1209 A61 A71 4 A93 A101 4 A122 47 A143 A152 1 A174 1 A192 A201 2
A14 12 A32 A43 3077 A61 A73 2 A93 A101 4 A123 52 A143 A152 1 A173 1 A192 A201 1
A14 24 A32 A40 3757 A61 A75 4 A92 A102 4 A124 62 A143 A153 1 A173 1 A192 A201 1
A14 10 A32 A40 1418 A62 A73 3 A93 A101 2 A121 35 A143 A151 1 A172 1 A191 A202 1
A14 6 A32 A40 3518 A61 A73 2 A93 A103 3 A122 26 A143 A151 1 A173 1 A191 A201 1
A14 12 A34 A43 1934 A61 A75 2 A93 A101 2 A124 26 A143 A152 2 A173 1 A191 A201 1
A12 27 A30 A49 8318 A61 A75 2 A92 A101 4 A124 42 A143 A153 2 A174 1 A192 A201 2
A14 6 A34 A43 1237 A62 A73 1 A92 A101 1 A122 27 A143 A152 2 A173 1 A191 A201 1
A12 6 A32 A43 368 A65 A75 4 A93 A101 4 A122 38 A143 A152 1 A173 1 A191 A201 1
A11 12 A34 A40 2122 A61 A73 3 A93 A101 2 A121 39 A143 A151 2 A172 2 A191 A202 1
A11 24 A32 A42 2996 A65 A73 2 A94 A101 4 A123 20 A143 A152 1 A173 1 A191 A201 2
A12 36 A32 A42 9034 A62 A72 4 A93 A102 1 A124 29 A143 A151 1 A174 1 A192 A201 2
A14 24 A34 A42 1585 A61 A74 4 A93 A101 3 A122 40 A143 A152 2 A173 1 A191 A201 1
A12 18 A32 A43 1301 A61 A75 4 A94 A103 2 A121 32 A143 A152 1 A172 1 A191 A201 1
A13 6 A34 A40 1323 A62 A75 2 A91 A101 4 A123 28 A143 A152 2 A173 2 A192 A201 1
A11 24 A32 A40 3123 A61 A72 4 A92 A101 1 A122 27 A143 A152 1 A173 1 A191 A201 2
A11 36 A32 A41 5493 A61 A75 2 A93 A101 4 A124 42 A143 A153 1 A173 2 A191 A201 1
A13 9 A32 A43 1126 A62 A75 2 A91 A101 4 A121 49 A143 A152 1 A173 1 A191 A201 1
A12 24 A34 A43 1216 A62 A72 4 A93 A101 4 A124 38 A141 A152 2 A173 2 A191 A201 2

A11 24 A32 A40 1207 A61 A72 4 A92 A101 4 A122 24 A143 A151 1 A173 1 A191 A201 2
A14 10 A32 A40 1309 A65 A73 4 A93 A103 4 A122 27 A143 A152 1 A172 1 A191 A201 2
A13 15 A34 A41 2360 A63 A73 2 A93 A101 2 A123 36 A143 A152 1 A173 1 A192 A201 1
A12 15 A31 A40 6850 A62 A71 1 A93 A101 2 A122 34 A143 A152 1 A174 2 A192 A201 2
A14 24 A32 A43 1413 A61 A73 4 A94 A101 2 A122 28 A143 A152 1 A173 1 A191 A201 1
A14 39 A32 A41 8588 A62 A75 4 A93 A101 2 A123 45 A143 A152 1 A174 1 A192 A201 1
A11 12 A32 A40 759 A61 A74 4 A93 A101 2 A121 26 A143 A152 1 A173 1 A191 A201 2
A14 36 A32 A41 4686 A61 A73 2 A93 A101 2 A124 32 A143 A153 1 A174 1 A192 A201 1
A13 15 A32 A49 2687 A61 A74 2 A93 A101 4 A122 26 A143 A151 1 A173 1 A192 A201 1
A12 12 A33 A43 585 A61 A73 4 A94 A102 4 A121 20 A143 A151 2 A173 1 A191 A201 1
A14 24 A32 A40 2255 A65 A72 4 A93 A101 1 A122 54 A143 A152 1 A173 1 A191 A201 1
A11 6 A34 A40 609 A61 A74 4 A92 A101 3 A122 37 A143 A152 2 A173 1 A191 A202 1
A11 6 A34 A40 1361 A61 A72 2 A93 A101 4 A121 40 A143 A152 1 A172 2 A191 A202 1
A14 36 A34 A42 7127 A61 A72 2 A92 A101 4 A122 23 A143 A151 2 A173 1 A192 A201 2
A11 6 A32 A40 1203 A62 A75 3 A93 A101 2 A122 43 A143 A152 1 A173 1 A192 A201 1
A14 6 A34 A43 700 A65 A75 4 A93 A101 4 A124 36 A143 A153 2 A173 1 A191 A201 1
A14 24 A34 A45 5507 A61 A75 3 A93 A101 4 A124 44 A143 A153 2 A173 1 A191 A201 1
A11 18 A32 A43 3190 A61 A73 2 A92 A101 2 A121 24 A143 A152 1 A173 1 A191 A201 2
A11 48 A30 A42 7119 A61 A73 3 A93 A101 4 A124 53 A143 A153 2 A173 2 A191 A201 2
A14 24 A32 A41 3488 A62 A74 3 A92 A101 4 A123 23 A143 A152 1 A173 1 A191 A201 1
A12 18 A32 A43 1113 A61 A73 4 A92 A103 4 A121 26 A143 A152 1 A172 2 A191 A201 1
A12 26 A32 A41 7966 A61 A72 2 A93 A101 3 A123 30 A143 A152 2 A173 1 A191 A201 1
A14 15 A34 A46 1532 A62 A73 4 A92 A101 3 A123 31 A143 A152 1 A173 1 A191 A201 1
A14 4 A34 A43 1503 A61 A74 2 A93 A101 1 A121 42 A143 A152 2 A172 2 A191 A201 1
A11 36 A32 A43 2302 A61 A73 4 A91 A101 4 A123 31 A143 A151 1 A173 1 A191 A201 2
A11 6 A32 A40 662 A61 A72 3 A93 A101 4 A121 41 A143 A152 1 A172 2 A192 A201 1
A12 36 A32 A46 2273 A61 A74 3 A93 A101 1 A123 32 A143 A152 2 A173 2 A191 A201 1
A12 15 A32 A40 2631 A62 A73 2 A92 A101 4 A123 28 A143 A151 2 A173 1 A192 A201 2
A14 12 A33 A41 1503 A61 A73 4 A94 A101 4 A121 41 A143 A151 1 A173 1 A191 A201 1
A14 24 A32 A43 1311 A62 A74 4 A94 A101 3 A122 26 A143 A152 1 A173 1 A192 A201 1
A14 24 A32 A43 3105 A65 A72 4 A93 A101 2 A123 25 A143 A152 2 A173 1 A191 A201 1
A13 21 A34 A46 2319 A61 A72 2 A91 A101 1 A123 33 A143 A151 1 A173 1 A191 A201 2
A11 6 A32 A40 1374 A65 A71 4 A92 A101 3 A122 75 A143 A152 1 A174 1 A192 A201 1
A12 18 A34 A42 3612 A61 A75 3 A92 A101 4 A122 37 A143 A152 1 A173 1 A192 A201 1
A11 48 A32 A40 7763 A61 A75 4 A93 A101 4 A124 42 A141 A153 1 A174 1 A191 A201 2
A13 18 A32 A42 3049 A61 A72 1 A92 A101 1 A122 45 A142 A152 1 A172 1 A191 A201 1
A12 12 A32 A43 1534 A61 A72 1 A94 A101 1 A121 23 A143 A151 1 A173 1 A191 A201 2
A14 24 A33 A40 2032 A61 A75 4 A93 A101 4 A124 60 A143 A153 2 A173 1 A192 A201 1
A11 30 A32 A42 6350 A65 A75 4 A93 A101 4 A122 31 A143 A152 1 A173 1 A191 A201 2
A13 18 A32 A42 2864 A61 A73 2 A93 A101 1 A121 34 A143 A152 1 A172 2 A191 A201 2
A14 12 A34 A40 1255 A61 A75 4 A93 A101 4 A121 61 A143 A152 2 A172 1 A191 A201 1
A11 24 A33 A40 1333 A61 A71 4 A93 A101 2 A121 43 A143 A153 2 A173 2 A191 A201 2
A14 24 A34 A40 2022 A61 A73 4 A92 A101 4 A123 37 A143 A152 1 A173 1 A192 A201 1
A14 24 A32 A43 1552 A61 A74 3 A93 A101 1 A123 32 A141 A152 1 A173 2 A191 A201 1
A11 12 A31 A43 626 A61 A73 4 A92 A101 4 A121 24 A141 A152 1 A172 1 A191 A201 2
A14 48 A34 A41 8858 A65 A74 2 A93 A101 1 A124 35 A143 A153 2 A173 1 A192 A201 1

A14 12 A34 A45 996 A65 A74 4 A92 A101 4 A121 23 A143 A152 2 A173 1 A191 A201 1
A14 6 A31 A43 1750 A63 A75 2 A93 A101 4 A122 45 A141 A152 1 A172 2 A191 A201 1
A11 48 A32 A43 6999 A61 A74 1 A94 A103 1 A121 34 A143 A152 2 A173 1 A192 A201 2
A12 12 A34 A40 1995 A62 A72 4 A93 A101 1 A123 27 A143 A152 1 A173 1 A191 A201 1
A12 9 A32 A46 1199 A61 A74 4 A92 A101 4 A122 67 A143 A152 2 A174 1 A192 A201 1
A12 12 A32 A43 1331 A61 A72 2 A93 A101 1 A123 22 A142 A152 1 A173 1 A191 A201 2
A12 18 A30 A40 2278 A62 A72 3 A92 A101 3 A123 28 A143 A152 2 A173 1 A191 A201 2
A14 21 A30 A40 5003 A65 A73 1 A92 A101 4 A122 29 A141 A152 2 A173 1 A192 A201 2
A11 24 A31 A42 3552 A61 A74 3 A93 A101 4 A123 27 A141 A152 1 A173 1 A191 A201 2
A12 18 A34 A42 1928 A61 A72 2 A93 A101 2 A121 31 A143 A152 2 A172 1 A191 A201 2
A11 24 A32 A41 2964 A65 A75 4 A93 A101 4 A124 49 A141 A153 1 A173 2 A192 A201 1
A11 24 A31 A43 1546 A61 A74 4 A93 A103 4 A123 24 A141 A151 1 A172 1 A191 A201 2
A13 6 A33 A43 683 A61 A72 2 A92 A101 1 A122 29 A141 A152 1 A173 1 A191 A201 1
A12 36 A32 A40 12389 A65 A73 1 A93 A101 4 A124 37 A143 A153 1 A173 1 A192 A201 2
A12 24 A33 A49 4712 A65 A73 4 A93 A101 2 A122 37 A141 A152 2 A174 1 A192 A201 1
A12 24 A33 A43 1553 A62 A74 3 A92 A101 2 A122 23 A143 A151 2 A173 1 A192 A201 1
A11 12 A32 A40 1372 A61 A74 2 A91 A101 3 A123 36 A143 A152 1 A173 1 A191 A201 2
A14 24 A34 A43 2578 A64 A75 2 A93 A101 2 A123 34 A143 A152 1 A173 1 A191 A201 1
A12 48 A32 A43 3979 A65 A74 4 A93 A101 1 A123 41 A143 A152 2 A173 2 A192 A201 1
A11 48 A32 A43 6758 A61 A73 3 A92 A101 2 A123 31 A143 A152 1 A173 1 A192 A201 2
A11 24 A32 A42 3234 A61 A72 4 A92 A101 4 A121 23 A143 A151 1 A172 1 A192 A201 2
A14 30 A34 A43 5954 A61 A74 3 A93 A102 2 A123 38 A143 A152 1 A173 1 A191 A201 1
A14 24 A32 A41 5433 A65 A71 2 A92 A101 4 A122 26 A143 A151 1 A174 1 A192 A201 1
A11 15 A32 A49 806 A61 A73 4 A92 A101 4 A122 22 A143 A152 1 A172 1 A191 A201 1
A12 9 A32 A43 1082 A61 A75 4 A93 A101 4 A123 27 A143 A152 2 A172 1 A191 A201 1
A14 15 A34 A42 2788 A61 A74 2 A92 A102 3 A123 24 A141 A152 2 A173 1 A191 A201 1
A12 12 A32 A43 2930 A61 A74 2 A92 A101 1 A121 27 A143 A152 1 A173 1 A191 A201 1
A14 24 A34 A46 1927 A65 A73 3 A92 A101 2 A123 33 A143 A152 2 A173 1 A192 A201 1
A12 36 A34 A40 2820 A61 A72 4 A91 A101 4 A123 27 A143 A152 2 A173 1 A191 A201 2
A14 24 A32 A48 937 A61 A72 4 A94 A101 3 A123 27 A143 A152 2 A172 1 A191 A201 1
A12 18 A34 A40 1056 A61 A75 3 A93 A103 3 A121 30 A141 A152 2 A173 1 A191 A201 2
A12 12 A34 A40 3124 A61 A72 1 A93 A101 3 A121 49 A141 A152 2 A172 2 A191 A201 1
A14 9 A32 A42 1388 A61 A73 4 A92 A101 2 A121 26 A143 A151 1 A173 1 A191 A201 1
A12 36 A32 A45 2384 A61 A72 4 A93 A101 1 A124 33 A143 A151 1 A172 1 A191 A201 2
A14 12 A32 A40 2133 A65 A75 4 A92 A101 4 A124 52 A143 A153 1 A174 1 A192 A201 1
A11 18 A32 A42 2039 A61 A73 1 A92 A101 4 A121 20 A141 A151 1 A173 1 A191 A201 2
A11 9 A34 A40 2799 A61 A73 2 A93 A101 2 A121 36 A143 A151 2 A173 2 A191 A201 1
A11 12 A32 A42 1289 A61 A73 4 A93 A103 1 A122 21 A143 A152 1 A172 1 A191 A201 1
A11 18 A32 A44 1217 A61 A73 4 A94 A101 3 A121 47 A143 A152 1 A172 1 A192 A201 2
A11 12 A34 A42 2246 A61 A75 3 A93 A101 3 A122 60 A143 A152 2 A173 1 A191 A201 2
A11 12 A34 A43 385 A61 A74 4 A92 A101 3 A121 58 A143 A152 4 A172 1 A192 A201 1
A12 24 A33 A40 1965 A65 A73 4 A92 A101 4 A123 42 A143 A151 2 A173 1 A192 A201 1
A14 21 A32 A49 1572 A64 A75 4 A92 A101 4 A121 36 A141 A152 1 A172 1 A191 A201 1
A12 24 A32 A40 2718 A61 A73 3 A92 A101 4 A122 20 A143 A151 1 A172 1 A192 A201 2
A11 24 A31 A410 1358 A65 A75 4 A93 A101 3 A123 40 A142 A152 1 A174 1 A192 A201 2
A12 6 A31 A40 931 A62 A72 1 A92 A101 1 A122 32 A142 A152 1 A172 1 A191 A201 2

A11 24 A32 A40 1442 A61 A74 4 A92 A101 4 A123 23 A143 A151 2 A173 1 A191 A201 2
A12 24 A30 A49 4241 A61 A73 1 A93 A101 4 A121 36 A143 A152 3 A172 1 A192 A201 2
A14 18 A34 A40 2775 A61 A74 2 A93 A101 2 A122 31 A141 A152 2 A173 1 A191 A201 2
A14 24 A33 A49 3863 A61 A73 1 A93 A101 2 A124 32 A143 A153 1 A173 1 A191 A201 1
A12 7 A32 A43 2329 A61 A72 1 A92 A103 1 A121 45 A143 A152 1 A173 1 A191 A201 1
A12 9 A32 A42 918 A61 A73 4 A92 A101 1 A122 30 A143 A152 1 A173 1 A191 A201 2
A12 24 A31 A46 1837 A61 A74 4 A92 A101 4 A124 34 A141 A153 1 A172 1 A191 A201 2
A14 36 A32 A42 3349 A61 A73 4 A92 A101 2 A123 28 A143 A152 1 A174 1 A192 A201 2
A13 10 A32 A42 1275 A61 A72 4 A92 A101 2 A122 23 A143 A152 1 A173 1 A191 A201 1
A11 24 A31 A42 2828 A63 A73 4 A93 A101 4 A121 22 A142 A152 1 A173 1 A192 A201 1
A14 24 A34 A49 4526 A61 A73 3 A93 A101 2 A121 74 A143 A152 1 A174 1 A192 A201 1
A12 36 A32 A43 2671 A62 A73 4 A92 A102 4 A124 50 A143 A153 1 A173 1 A191 A201 2
A14 18 A32 A43 2051 A61 A72 4 A93 A101 1 A121 33 A143 A152 1 A173 1 A191 A201 1
A14 15 A32 A41 1300 A65 A75 4 A93 A101 4 A124 45 A141 A153 1 A173 2 A191 A201 1
A11 12 A32 A44 741 A62 A71 4 A92 A101 3 A122 22 A143 A152 1 A173 1 A191 A201 2
A13 10 A32 A40 1240 A62 A75 1 A92 A101 4 A124 48 A143 A153 1 A172 2 A191 A201 2
A11 21 A32 A43 3357 A64 A72 4 A92 A101 2 A123 29 A141 A152 1 A173 1 A191 A201 1
A11 24 A31 A41 3632 A61 A73 1 A92 A103 4 A123 22 A141 A151 1 A173 1 A191 A202 1
A14 18 A33 A42 1808 A61 A74 4 A92 A101 1 A121 22 A143 A152 1 A173 1 A191 A201 2
A12 48 A30 A49 12204 A65 A73 2 A93 A101 2 A123 48 A141 A152 1 A174 1 A192 A201 1
A12 60 A33 A43 9157 A65 A73 2 A93 A101 2 A124 27 A143 A153 1 A174 1 A191 A201 1
A11 6 A34 A40 3676 A61 A73 1 A93 A101 3 A121 37 A143 A151 3 A173 2 A191 A201 1
A12 30 A32 A42 3441 A62 A73 2 A92 A102 4 A123 21 A143 A151 1 A173 1 A191 A201 2
A14 12 A32 A40 640 A61 A73 4 A91 A101 2 A121 49 A143 A152 1 A172 1 A191 A201 1
A12 21 A34 A49 3652 A61 A74 2 A93 A101 3 A122 27 A143 A152 2 A173 1 A191 A201 1
A14 18 A34 A40 1530 A61 A73 3 A93 A101 2 A122 32 A141 A152 2 A173 1 A191 A201 2
A14 48 A32 A49 3914 A65 A73 4 A91 A101 2 A121 38 A141 A152 1 A173 1 A191 A201 2
A11 12 A32 A42 1858 A61 A72 4 A92 A101 1 A123 22 A143 A151 1 A173 1 A191 A201 1
A11 18 A32 A43 2600 A61 A73 4 A93 A101 4 A124 65 A143 A153 2 A173 1 A191 A201 2
A14 15 A32 A43 1979 A65 A75 4 A93 A101 2 A123 35 A143 A152 1 A173 1 A191 A201 1
A13 6 A32 A42 2116 A61 A73 2 A93 A101 2 A121 41 A143 A152 1 A173 1 A192 A201 1
A12 9 A31 A40 1437 A62 A74 2 A93 A101 3 A124 29 A143 A152 1 A173 1 A191 A201 2
A14 42 A34 A42 4042 A63 A73 4 A93 A101 4 A121 36 A143 A152 2 A173 1 A192 A201 1
A14 9 A32 A46 3832 A65 A75 1 A93 A101 4 A121 64 A143 A152 1 A172 1 A191 A201 1
A11 24 A32 A43 3660 A61 A73 2 A92 A101 4 A123 28 A143 A152 1 A173 1 A191 A201 1
A11 18 A31 A42 1553 A61 A73 4 A93 A101 3 A123 44 A141 A152 1 A173 1 A191 A201 2
A12 15 A32 A43 1444 A65 A72 4 A93 A101 1 A122 23 A143 A152 1 A173 1 A191 A201 1
A14 9 A32 A42 1980 A61 A72 2 A92 A102 2 A123 19 A143 A151 2 A173 1 A191 A201 2
A12 24 A32 A40 1355 A61 A72 3 A92 A101 4 A123 25 A143 A152 1 A172 1 A192 A201 2
A14 12 A32 A46 1393 A61 A75 4 A93 A101 4 A122 47 A141 A152 3 A173 2 A192 A201 1
A14 24 A32 A43 1376 A63 A74 4 A92 A101 1 A123 28 A143 A152 1 A173 1 A191 A201 1
A14 60 A33 A43 15653 A61 A74 2 A93 A101 4 A123 21 A143 A152 2 A173 1 A192 A201 1
A14 12 A32 A43 1493 A61 A72 4 A92 A101 3 A123 34 A143 A152 1 A173 2 A191 A201 1
A11 42 A33 A43 4370 A61 A74 3 A93 A101 2 A122 26 A141 A152 2 A173 2 A192 A201 2
A11 18 A32 A46 750 A61 A71 4 A92 A101 1 A121 27 A143 A152 1 A171 1 A191 A201 2
A12 15 A32 A45 1308 A61 A75 4 A93 A101 4 A123 38 A143 A152 2 A172 1 A191 A201 1

A14 15 A32 A46 4623 A62 A73 3 A93 A101 2 A122 40 A143 A152 1 A174 1 A192 A201 2
A14 24 A34 A43 1851 A61 A74 4 A94 A103 2 A123 33 A143 A152 2 A173 1 A192 A201 1
A11 18 A34 A43 1880 A61 A74 4 A94 A101 1 A122 32 A143 A152 2 A174 1 A192 A201 1
A14 36 A33 A49 7980 A65 A72 4 A93 A101 4 A123 27 A143 A151 2 A173 1 A192 A201 2
A11 30 A30 A42 4583 A61 A73 2 A91 A103 2 A121 32 A143 A152 2 A173 1 A191 A201 1
A14 12 A32 A40 1386 A63 A73 2 A92 A101 2 A122 26 A143 A152 1 A173 1 A191 A201 2
A13 24 A32 A40 947 A61 A74 4 A93 A101 3 A124 38 A141 A153 1 A173 2 A191 A201 2
A11 12 A32 A46 684 A61 A73 4 A93 A101 4 A123 40 A143 A151 1 A172 2 A191 A201 2
A11 48 A32 A46 7476 A61 A74 4 A93 A101 1 A124 50 A143 A153 1 A174 1 A192 A201 1
A12 12 A32 A42 1922 A61 A73 4 A93 A101 2 A122 37 A143 A152 1 A172 1 A191 A201 2
A11 24 A32 A40 2303 A61 A75 4 A93 A102 1 A121 45 A143 A152 1 A173 1 A191 A201 2
A12 36 A33 A40 8086 A62 A75 2 A93 A101 4 A123 42 A143 A152 4 A174 1 A192 A201 2
A14 24 A34 A41 2346 A61 A74 4 A93 A101 3 A123 35 A143 A152 2 A173 1 A192 A201 1
A11 14 A32 A40 3973 A61 A71 1 A93 A101 4 A124 22 A143 A153 1 A173 1 A191 A201 1
A12 12 A32 A40 888 A61 A75 4 A93 A101 4 A123 41 A141 A152 1 A172 2 A191 A201 2
A14 48 A32 A43 10222 A65 A74 4 A93 A101 3 A123 37 A142 A152 1 A173 1 A192 A201 1
A12 30 A30 A49 4221 A61 A73 2 A92 A101 1 A123 28 A143 A152 2 A173 1 A191 A201 1
A12 18 A34 A42 6361 A61 A75 2 A93 A101 1 A124 41 A143 A152 1 A173 1 A192 A201 1
A13 12 A32 A43 1297 A61 A73 3 A94 A101 4 A121 23 A143 A151 1 A173 1 A191 A201 1
A11 12 A32 A40 900 A65 A73 4 A94 A101 2 A123 23 A143 A152 1 A173 1 A191 A201 2
A14 21 A32 A42 2241 A61 A75 4 A93 A101 2 A121 50 A143 A152 2 A173 1 A191 A201 1
A12 6 A33 A42 1050 A61 A71 4 A93 A101 1 A122 35 A142 A152 2 A174 1 A192 A201 1
A13 6 A34 A46 1047 A61 A73 2 A92 A101 4 A122 50 A143 A152 1 A172 1 A191 A201 1
A14 24 A34 A410 6314 A61 A71 4 A93 A102 2 A124 27 A141 A152 2 A174 1 A192 A201 1
A12 30 A31 A42 3496 A64 A73 4 A93 A101 2 A123 34 A142 A152 1 A173 2 A192 A201 1
A14 48 A31 A49 3609 A61 A73 1 A92 A101 1 A121 27 A142 A152 1 A173 1 A191 A201 1
A11 12 A34 A40 4843 A61 A75 3 A93 A102 4 A122 43 A143 A151 2 A173 1 A192 A201 2
A13 30 A34 A43 3017 A61 A75 4 A93 A101 4 A122 47 A143 A152 1 A173 1 A191 A201 1
A14 24 A34 A49 4139 A62 A73 3 A93 A101 3 A122 27 A143 A152 2 A172 1 A192 A201 1
A14 36 A32 A49 5742 A62 A74 2 A93 A101 2 A123 31 A143 A152 2 A173 1 A192 A201 1
A14 60 A32 A40 10366 A61 A75 2 A93 A101 4 A122 42 A143 A152 1 A174 1 A192 A201 1
A14 6 A34 A40 2080 A63 A73 1 A94 A101 2 A123 24 A143 A152 1 A173 1 A191 A201 1
A14 21 A33 A49 2580 A63 A72 4 A93 A101 2 A121 41 A141 A152 1 A172 2 A191 A201 2
A14 30 A34 A43 4530 A61 A74 4 A92 A101 4 A123 26 A143 A151 1 A174 1 A192 A201 1
A14 24 A34 A42 5150 A61 A75 4 A93 A101 4 A123 33 A143 A152 1 A173 1 A192 A201 1
A12 72 A32 A43 5595 A62 A73 2 A94 A101 2 A123 24 A143 A152 1 A173 1 A191 A201 2
A11 24 A32 A43 2384 A61 A75 4 A93 A101 4 A121 64 A141 A151 1 A172 1 A191 A201 1
A14 18 A32 A43 1453 A61 A72 3 A92 A101 1 A121 26 A143 A152 1 A173 1 A191 A201 1
A14 6 A32 A46 1538 A61 A72 1 A92 A101 2 A124 56 A143 A152 1 A173 1 A191 A201 1
A14 12 A32 A43 2279 A65 A73 4 A93 A101 4 A124 37 A143 A153 1 A173 1 A192 A201 1
A14 15 A33 A43 1478 A61 A73 4 A94 A101 3 A121 33 A141 A152 2 A173 1 A191 A201 1
A14 24 A34 A43 5103 A61 A72 3 A94 A101 3 A124 47 A143 A153 3 A173 1 A192 A201 1
A12 36 A33 A49 9857 A62 A74 1 A93 A101 3 A122 31 A143 A152 2 A172 2 A192 A201 1
A14 60 A32 A40 6527 A65 A73 4 A93 A101 4 A124 34 A143 A153 1 A173 2 A192 A201 1
A13 10 A34 A43 1347 A65 A74 4 A93 A101 2 A122 27 A143 A152 2 A173 1 A192 A201 1
A12 36 A33 A40 2862 A62 A75 4 A93 A101 3 A124 30 A143 A153 1 A173 1 A191 A201 1

A14 9 A32 A43 2753 A62 A75 3 A93 A102 4 A123 35 A143 A152 1 A173 1 A192 A201 1
A11 12 A32 A40 3651 A64 A73 1 A93 A101 3 A122 31 A143 A152 1 A173 2 A191 A201 1
A11 15 A34 A42 975 A61 A73 2 A91 A101 3 A122 25 A143 A152 2 A173 1 A191 A201 1
A12 15 A32 A45 2631 A62 A73 3 A92 A101 2 A121 25 A143 A152 1 A172 1 A191 A201 1
A12 24 A32 A43 2896 A62 A72 2 A93 A101 1 A123 29 A143 A152 1 A173 1 A191 A201 1
A11 6 A34 A40 4716 A65 A72 1 A93 A101 3 A121 44 A143 A152 2 A172 2 A191 A201 1
A14 24 A32 A43 2284 A61 A74 4 A93 A101 2 A123 28 A143 A152 1 A173 1 A192 A201 1
A14 6 A32 A41 1236 A63 A73 2 A93 A101 4 A122 50 A143 A151 1 A173 1 A191 A201 1
A12 12 A32 A43 1103 A61 A74 4 A93 A103 3 A121 29 A143 A152 2 A173 1 A191 A202 1
A14 12 A34 A40 926 A61 A71 1 A92 A101 2 A122 38 A143 A152 1 A171 1 A191 A201 1
A14 18 A34 A43 1800 A61 A73 4 A93 A101 2 A123 24 A143 A152 2 A173 1 A191 A201 1
A13 15 A32 A46 1905 A61 A75 4 A93 A101 4 A123 40 A143 A151 1 A174 1 A192 A201 1
A14 12 A32 A42 1123 A63 A73 4 A92 A101 4 A123 29 A143 A151 1 A172 1 A191 A201 2
A11 48 A34 A41 6331 A61 A75 4 A93 A101 4 A124 46 A143 A153 2 A173 1 A192 A201 2
A13 24 A32 A43 1377 A62 A75 4 A92 A101 2 A124 47 A143 A153 1 A173 1 A192 A201 1
A12 30 A33 A49 2503 A62 A75 4 A93 A101 2 A122 41 A142 A152 2 A173 1 A191 A201 1
A12 27 A32 A49 2528 A61 A72 4 A92 A101 1 A122 32 A143 A152 1 A173 2 A192 A201 1
A14 15 A32 A40 5324 A63 A75 1 A92 A101 4 A124 35 A143 A153 1 A173 1 A191 A201 1
A12 48 A32 A40 6560 A62 A74 3 A93 A101 2 A122 24 A143 A152 1 A173 1 A191 A201 2
A12 12 A30 A42 2969 A61 A72 4 A92 A101 3 A122 25 A143 A151 2 A173 1 A191 A201 2
A12 9 A32 A43 1206 A61 A75 4 A92 A101 4 A121 25 A143 A152 1 A173 1 A191 A201 1
A12 9 A32 A43 2118 A61 A73 2 A93 A101 2 A121 37 A143 A152 1 A172 2 A191 A201 1
A14 18 A34 A43 629 A63 A75 4 A93 A101 3 A122 32 A141 A152 2 A174 1 A192 A201 1
A11 6 A31 A46 1198 A61 A75 4 A92 A101 4 A124 35 A143 A153 1 A173 1 A191 A201 2
A14 21 A32 A41 2476 A65 A75 4 A93 A101 4 A121 46 A143 A152 1 A174 1 A192 A201 1
A11 9 A34 A43 1138 A61 A73 4 A93 A101 4 A121 25 A143 A152 2 A172 1 A191 A201 1
A12 60 A32 A40 14027 A61 A74 4 A93 A101 2 A124 27 A143 A152 1 A174 1 A192 A201 2
A14 30 A34 A41 7596 A65 A75 1 A93 A101 4 A123 63 A143 A152 2 A173 1 A191 A201 1
A14 30 A34 A43 3077 A65 A75 3 A93 A101 2 A123 40 A143 A152 2 A173 2 A192 A201 1
A14 18 A32 A43 1505 A61 A73 4 A93 A101 2 A124 32 A143 A153 1 A174 1 A192 A201 1
A13 24 A34 A43 3148 A65 A73 3 A93 A101 2 A123 31 A143 A152 2 A173 1 A192 A201 1
A12 20 A30 A41 6148 A62 A75 3 A94 A101 4 A123 31 A141 A152 2 A173 1 A192 A201 1
A13 9 A30 A43 1337 A61 A72 4 A93 A101 2 A123 34 A143 A152 2 A174 1 A192 A201 2
A12 6 A31 A46 433 A64 A72 4 A92 A101 2 A122 24 A141 A151 1 A173 2 A191 A201 2
A11 12 A32 A40 1228 A61 A73 4 A92 A101 2 A121 24 A143 A152 1 A172 1 A191 A201 2
A12 9 A32 A43 790 A63 A73 4 A92 A101 3 A121 66 A143 A152 1 A172 1 A191 A201 1
A14 27 A32 A40 2570 A61 A73 3 A92 A101 3 A121 21 A143 A151 1 A173 1 A191 A201 2
A14 6 A34 A40 250 A64 A73 2 A92 A101 2 A121 41 A141 A152 2 A172 1 A191 A201 1
A14 15 A34 A43 1316 A63 A73 2 A94 A101 2 A122 47 A143 A152 2 A172 1 A191 A201 1
A11 18 A32 A43 1882 A61 A73 4 A92 A101 4 A123 25 A141 A151 2 A173 1 A191 A201 2
A12 48 A31 A49 6416 A61 A75 4 A92 A101 3 A124 59 A143 A151 1 A173 1 A191 A201 2
A13 24 A34 A49 1275 A64 A73 2 A91 A101 4 A121 36 A143 A152 2 A173 1 A192 A201 1
A12 24 A33 A43 6403 A61 A72 1 A93 A101 2 A123 33 A143 A152 1 A173 1 A191 A201 1
A11 24 A32 A43 1987 A61 A73 2 A93 A101 4 A121 21 A143 A151 1 A172 2 A191 A201 2
A12 8 A32 A43 760 A61 A74 4 A92 A103 2 A121 44 A143 A152 1 A172 1 A191 A201 1
A14 24 A32 A41 2603 A64 A73 2 A92 A101 4 A123 28 A143 A151 1 A173 1 A192 A201 1

A14 4 A34 A40 3380 A61 A74 1 A92 A101 1 A121 37 A143 A152 1 A173 2 A191 A201 1
A12 36 A31 A44 3990 A65 A72 3 A92 A101 2 A124 29 A141 A152 1 A171 1 A191 A201 1
A12 24 A32 A41 11560 A61 A73 1 A92 A101 4 A123 23 A143 A151 2 A174 1 A191 A201 2
A11 18 A32 A40 4380 A62 A73 3 A93 A101 4 A123 35 A143 A152 1 A172 2 A192 A201 1
A14 6 A34 A40 6761 A61 A74 1 A93 A101 3 A124 45 A143 A152 2 A174 2 A192 A201 1
A12 30 A30 A49 4280 A62 A73 4 A92 A101 4 A123 26 A143 A151 2 A172 1 A191 A201 2
A11 24 A31 A40 2325 A62 A74 2 A93 A101 3 A123 32 A141 A152 1 A173 1 A191 A201 1
A12 10 A31 A43 1048 A61 A73 4 A93 A101 4 A121 23 A142 A152 1 A172 1 A191 A201 1
A14 21 A32 A43 3160 A65 A75 4 A93 A101 3 A122 41 A143 A152 1 A173 1 A192 A201 1
A11 24 A31 A42 2483 A63 A73 4 A93 A101 4 A121 22 A142 A152 1 A173 1 A192 A201 1
A11 39 A34 A42 14179 A65 A74 4 A93 A101 4 A122 30 A143 A152 2 A174 1 A192 A201 1
A11 13 A34 A49 1797 A61 A72 3 A93 A101 1 A122 28 A141 A152 2 A172 1 A191 A201 1
A11 15 A32 A40 2511 A61 A71 1 A92 A101 4 A123 23 A143 A151 1 A173 1 A191 A201 1
A11 12 A32 A40 1274 A61 A72 3 A92 A101 1 A121 37 A143 A152 1 A172 1 A191 A201 2
A14 21 A32 A41 5248 A65 A73 1 A93 A101 3 A123 26 A143 A152 1 A173 1 A191 A201 1
A14 15 A32 A41 3029 A61 A74 2 A93 A101 2 A123 33 A143 A152 1 A173 1 A191 A201 1
A11 6 A32 A42 428 A61 A75 2 A92 A101 1 A122 49 A141 A152 1 A173 1 A192 A201 1
A11 18 A32 A40 976 A61 A72 1 A92 A101 2 A123 23 A143 A152 1 A172 1 A191 A201 2
A12 12 A32 A49 841 A62 A74 2 A92 A101 4 A121 23 A143 A151 1 A172 1 A191 A201 1
A14 30 A34 A43 5771 A61 A74 4 A92 A101 2 A123 25 A143 A152 2 A173 1 A191 A201 1
A14 12 A33 A45 1555 A64 A75 4 A93 A101 4 A124 55 A143 A153 2 A173 2 A191 A201 2
A11 24 A32 A40 1285 A65 A74 4 A92 A101 4 A124 32 A143 A151 1 A173 1 A191 A201 2
A13 6 A34 A40 1299 A61 A73 1 A93 A101 1 A121 74 A143 A152 3 A171 2 A191 A202 1
A13 15 A34 A43 1271 A65 A73 3 A93 A101 4 A124 39 A143 A153 2 A173 1 A192 A201 2
A14 24 A32 A40 1393 A61 A73 2 A93 A103 2 A121 31 A143 A152 1 A173 1 A192 A201 1
A11 12 A34 A40 691 A61 A75 4 A93 A101 3 A122 35 A143 A152 2 A173 1 A191 A201 2
A14 15 A34 A40 5045 A65 A75 1 A92 A101 4 A123 59 A143 A152 1 A173 1 A192 A201 1
A11 18 A34 A42 2124 A61 A73 4 A92 A101 4 A121 24 A143 A151 2 A173 1 A191 A201 2
A11 12 A32 A43 2214 A61 A73 4 A93 A101 3 A122 24 A143 A152 1 A172 1 A191 A201 1
A14 21 A34 A40 12680 A65 A75 4 A93 A101 4 A124 30 A143 A153 1 A174 1 A192 A201 2
A14 24 A34 A40 2463 A62 A74 4 A94 A101 3 A122 27 A143 A152 2 A173 1 A192 A201 1
A12 12 A32 A43 1155 A61 A75 3 A94 A103 3 A121 40 A141 A152 2 A172 1 A191 A201 1
A11 30 A32 A42 3108 A61 A72 2 A91 A101 4 A122 31 A143 A152 1 A172 1 A191 A201 2
A14 10 A32 A41 2901 A65 A72 1 A92 A101 4 A121 31 A143 A151 1 A173 1 A191 A201 1
A12 12 A34 A42 3617 A61 A75 1 A93 A101 4 A123 28 A143 A151 3 A173 1 A192 A201 1
A14 12 A34 A43 1655 A61 A75 2 A93 A101 4 A121 63 A143 A152 2 A172 1 A192 A201 1
A11 24 A32 A41 2812 A65 A75 2 A92 A101 4 A121 26 A143 A151 1 A173 1 A191 A201 1
A11 36 A34 A46 8065 A61 A73 3 A92 A101 2 A124 25 A143 A152 2 A174 1 A192 A201 2
A14 21 A34 A41 3275 A61 A75 1 A93 A101 4 A123 36 A143 A152 1 A174 1 A192 A201 1
A14 24 A34 A43 2223 A62 A75 4 A93 A101 4 A122 52 A141 A152 2 A173 1 A191 A201 1
A13 12 A34 A40 1480 A63 A71 2 A93 A101 4 A124 66 A141 A153 3 A171 1 A191 A201 1
A11 24 A32 A40 1371 A65 A73 4 A92 A101 4 A121 25 A143 A151 1 A173 1 A191 A201 2
A14 36 A34 A40 3535 A61 A74 4 A93 A101 4 A123 37 A143 A152 2 A173 1 A192 A201 1
A11 18 A32 A43 3509 A61 A74 4 A92 A103 1 A121 25 A143 A152 1 A173 1 A191 A201 1
A14 36 A34 A41 5711 A64 A75 4 A93 A101 2 A123 38 A143 A152 2 A174 1 A192 A201 1
A12 18 A32 A45 3872 A61 A71 2 A92 A101 4 A123 67 A143 A152 1 A173 1 A192 A201 1

A12 39 A34 A43 4933 A61 A74 2 A93 A103 2 A121 25 A143 A152 2 A173 1 A191 A201 2
A14 24 A34 A40 1940 A64 A75 4 A93 A101 4 A121 60 A143 A152 1 A173 1 A192 A201 1
A12 12 A30 A48 1410 A61 A73 2 A93 A101 2 A121 31 A143 A152 1 A172 1 A192 A201 1
A12 12 A32 A40 836 A62 A72 4 A92 A101 2 A122 23 A141 A152 1 A172 1 A191 A201 2
A12 20 A32 A41 6468 A65 A71 1 A91 A101 4 A121 60 A143 A152 1 A174 1 A192 A201 1
A12 18 A32 A49 1941 A64 A73 4 A93 A101 2 A122 35 A143 A152 1 A172 1 A192 A201 1
A14 22 A32 A43 2675 A63 A75 3 A93 A101 4 A123 40 A143 A152 1 A173 1 A191 A201 1
A14 48 A34 A41 2751 A65 A75 4 A93 A101 3 A123 38 A143 A152 2 A173 2 A192 A201 1
A12 48 A33 A46 6224 A61 A75 4 A93 A101 4 A124 50 A143 A153 1 A173 1 A191 A201 2
A11 40 A34 A46 5998 A61 A73 4 A93 A101 3 A124 27 A141 A152 1 A173 1 A192 A201 2
A12 21 A32 A49 1188 A61 A75 2 A92 A101 4 A122 39 A143 A152 1 A173 2 A191 A201 2
A14 24 A32 A41 6313 A65 A75 3 A93 A101 4 A123 41 A143 A152 1 A174 2 A192 A201 1
A14 6 A34 A42 1221 A65 A73 1 A94 A101 2 A122 27 A143 A152 2 A173 1 A191 A201 1
A13 24 A32 A42 2892 A61 A75 3 A91 A101 4 A124 51 A143 A153 1 A173 1 A191 A201 1
A14 24 A32 A42 3062 A63 A75 4 A93 A101 3 A124 32 A143 A151 1 A173 1 A192 A201 1
A14 9 A32 A42 2301 A62 A72 2 A92 A101 4 A122 22 A143 A151 1 A173 1 A191 A201 1
A11 18 A32 A41 7511 A65 A75 1 A93 A101 4 A122 51 A143 A153 1 A173 2 A192 A201 2
A14 12 A34 A42 1258 A61 A72 2 A92 A101 4 A122 22 A143 A151 2 A172 1 A191 A201 1
A14 24 A33 A40 717 A65 A75 4 A94 A101 4 A123 54 A143 A152 2 A173 1 A192 A201 1
A12 9 A32 A40 1549 A65 A72 4 A93 A101 2 A121 35 A143 A152 1 A171 1 A191 A201 1
A14 24 A34 A46 1597 A61 A75 4 A93 A101 4 A124 54 A143 A153 2 A173 2 A191 A201 1
A12 18 A34 A43 1795 A61 A75 3 A92 A103 4 A121 48 A141 A151 2 A172 1 A192 A201 1
A11 20 A34 A42 4272 A61 A75 1 A92 A101 4 A122 24 A143 A152 2 A173 1 A191 A201 1
A14 12 A34 A43 976 A65 A75 4 A93 A101 4 A123 35 A143 A152 2 A173 1 A191 A201 1
A12 12 A32 A40 7472 A65 A71 1 A92 A101 2 A121 24 A143 A151 1 A171 1 A191 A201 1
A11 36 A32 A40 9271 A61 A74 2 A93 A101 1 A123 24 A143 A152 1 A173 1 A192 A201 2
A12 6 A32 A43 590 A61 A72 3 A94 A101 3 A121 26 A143 A152 1 A172 1 A191 A202 1
A14 12 A34 A43 930 A65 A75 4 A93 A101 4 A121 65 A143 A152 4 A173 1 A191 A201 1
A12 42 A31 A41 9283 A61 A71 1 A93 A101 2 A124 55 A141 A153 1 A174 1 A192 A201 1
A12 15 A30 A40 1778 A61 A72 2 A92 A101 1 A121 26 A143 A151 2 A171 1 A191 A201 2
A12 8 A32 A49 907 A61 A72 3 A94 A101 2 A121 26 A143 A152 1 A173 1 A192 A201 1
A12 6 A32 A43 484 A61 A74 3 A94 A103 3 A121 28 A141 A152 1 A172 1 A191 A201 1
A11 36 A34 A41 9629 A61 A74 4 A93 A101 4 A123 24 A143 A152 2 A173 1 A192 A201 2
A11 48 A32 A44 3051 A61 A73 3 A93 A101 4 A123 54 A143 A152 1 A173 1 A191 A201 2
A11 48 A32 A40 3931 A61 A74 4 A93 A101 4 A124 46 A143 A153 1 A173 2 A191 A201 2
A12 36 A33 A40 7432 A61 A73 2 A92 A101 2 A122 54 A143 A151 1 A173 1 A191 A201 1
A14 6 A32 A44 1338 A63 A73 1 A91 A101 4 A121 62 A143 A152 1 A173 1 A191 A201 1
A14 6 A34 A43 1554 A61 A74 1 A92 A101 2 A123 24 A143 A151 2 A173 1 A192 A201 1
A11 36 A32 A410 15857 A61 A71 2 A91 A102 3 A123 43 A143 A152 1 A174 1 A191 A201 1
A11 18 A32 A43 1345 A61 A73 4 A94 A101 3 A121 26 A141 A152 1 A173 1 A191 A201 2
A14 12 A32 A40 1101 A61 A73 3 A94 A101 2 A121 27 A143 A152 2 A173 1 A192 A201 1
A13 12 A32 A43 3016 A61 A73 3 A94 A101 1 A123 24 A143 A152 1 A173 1 A191 A201 1
A11 36 A32 A42 2712 A61 A75 2 A93 A101 2 A122 41 A141 A152 1 A173 2 A191 A201 2
A11 8 A34 A40 731 A61 A75 4 A93 A101 4 A121 47 A143 A152 2 A172 1 A191 A201 1
A14 18 A34 A42 3780 A61 A72 3 A91 A101 2 A123 35 A143 A152 2 A174 1 A192 A201 1
A11 21 A34 A40 1602 A61 A75 4 A94 A101 3 A123 30 A143 A152 2 A173 1 A192 A201 1

A11 18 A34 A40 3966 A61 A75 1 A92 A101 4 A121 33 A141 A151 3 A173 1 A192 A201 2
A14 18 A30 A49 4165 A61 A73 2 A93 A101 2 A123 36 A142 A152 2 A173 2 A191 A201 2
A11 36 A32 A41 8335 A65 A75 3 A93 A101 4 A124 47 A143 A153 1 A173 1 A191 A201 2
A12 48 A33 A49 6681 A65 A73 4 A93 A101 4 A124 38 A143 A153 1 A173 2 A192 A201 1
A14 24 A33 A49 2375 A63 A73 4 A93 A101 2 A123 44 A143 A152 2 A173 2 A192 A201 1
A11 18 A32 A40 1216 A61 A72 4 A92 A101 3 A123 23 A143 A151 1 A173 1 A192 A201 2
A11 45 A30 A49 11816 A61 A75 2 A93 A101 4 A123 29 A143 A151 2 A173 1 A191 A201 2
A12 24 A32 A43 5084 A65 A75 2 A92 A101 4 A123 42 A143 A152 1 A173 1 A192 A201 1
A13 15 A32 A43 2327 A61 A72 2 A92 A101 3 A121 25 A143 A152 1 A172 1 A191 A201 2
A11 12 A30 A40 1082 A61 A73 4 A93 A101 4 A123 48 A141 A152 2 A173 1 A191 A201 2
A14 12 A32 A43 886 A65 A73 4 A92 A101 2 A123 21 A143 A152 1 A173 1 A191 A201 1
A14 4 A32 A42 601 A61 A72 1 A92 A101 3 A121 23 A143 A151 1 A172 2 A191 A201 1
A11 24 A34 A41 2957 A61 A75 4 A93 A101 4 A122 63 A143 A152 2 A173 1 A192 A201 1
A14 24 A34 A43 2611 A61 A75 4 A94 A102 3 A121 46 A143 A152 2 A173 1 A191 A201 1
A11 36 A32 A42 5179 A61 A74 4 A93 A101 2 A122 29 A143 A152 1 A173 1 A191 A201 2
A14 21 A33 A41 2993 A61 A73 3 A93 A101 2 A121 28 A142 A152 2 A172 1 A191 A201 1
A14 18 A32 A45 1943 A61 A72 4 A92 A101 4 A121 23 A143 A152 1 A173 1 A191 A201 2
A14 24 A31 A49 1559 A61 A74 4 A93 A101 4 A123 50 A141 A152 1 A173 1 A192 A201 1
A14 18 A32 A42 3422 A61 A75 4 A93 A101 4 A122 47 A141 A152 3 A173 2 A192 A201 1
A12 21 A32 A42 3976 A65 A74 2 A93 A101 3 A123 35 A143 A152 1 A173 1 A192 A201 1
A14 18 A32 A40 6761 A65 A73 2 A93 A101 4 A123 68 A143 A151 2 A173 1 A191 A201 2
A14 24 A32 A40 1249 A61 A72 4 A94 A101 2 A121 28 A143 A152 1 A173 1 A191 A201 1
A11 9 A32 A43 1364 A61 A74 3 A93 A101 4 A121 59 A143 A152 1 A173 1 A191 A201 1
A11 12 A32 A43 709 A61 A75 4 A93 A101 4 A121 57 A142 A152 1 A172 1 A191 A201 2
A11 20 A34 A40 2235 A61 A73 4 A94 A103 2 A122 33 A141 A151 2 A173 1 A191 A202 2
A14 24 A34 A41 4042 A65 A74 3 A93 A101 4 A122 43 A143 A152 2 A173 1 A192 A201 1
A14 15 A34 A43 1471 A61 A73 4 A93 A101 4 A124 35 A143 A153 2 A173 1 A192 A201 1
A11 18 A31 A40 1442 A61 A74 4 A93 A101 4 A124 32 A143 A153 2 A172 2 A191 A201 2
A14 36 A33 A40 10875 A61 A75 2 A93 A101 2 A123 45 A143 A152 2 A173 2 A192 A201 1
A14 24 A32 A40 1474 A62 A72 4 A94 A101 3 A121 33 A143 A152 1 A173 1 A192 A201 1
A14 10 A32 A48 894 A65 A74 4 A92 A101 3 A122 40 A143 A152 1 A173 1 A192 A201 1
A14 15 A34 A42 3343 A61 A73 4 A93 A101 2 A124 28 A143 A153 1 A173 1 A192 A201 1
A11 15 A32 A40 3959 A61 A73 3 A92 A101 2 A122 29 A143 A152 1 A173 1 A192 A201 2
A14 9 A32 A40 3577 A62 A73 1 A93 A103 2 A121 26 A143 A151 1 A173 2 A191 A202 1
A14 24 A34 A41 5804 A64 A73 4 A93 A101 2 A121 27 A143 A152 2 A173 1 A191 A201 1
A14 18 A33 A49 2169 A61 A73 4 A94 A101 2 A123 28 A143 A152 1 A173 1 A192 A201 2
A11 24 A32 A43 2439 A61 A72 4 A92 A101 4 A121 35 A143 A152 1 A173 1 A192 A201 2
A14 27 A34 A42 4526 A64 A72 4 A93 A101 2 A121 32 A142 A152 2 A172 2 A192 A201 1
A14 10 A32 A42 2210 A61 A73 2 A93 A101 2 A121 25 A141 A151 1 A172 1 A191 A201 2
A14 15 A32 A42 2221 A63 A73 2 A92 A101 4 A123 20 A143 A151 1 A173 1 A191 A201 1
A11 18 A32 A43 2389 A61 A72 4 A92 A101 1 A123 27 A142 A152 1 A173 1 A191 A201 1
A14 12 A34 A42 3331 A61 A75 2 A93 A101 4 A122 42 A142 A152 1 A173 1 A191 A201 1
A14 36 A32 A49 7409 A65 A75 3 A93 A101 2 A122 37 A143 A152 2 A173 1 A191 A201 1
A11 12 A32 A42 652 A61 A75 4 A92 A101 4 A122 24 A143 A151 1 A173 1 A191 A201 1
A14 36 A33 A42 7678 A63 A74 2 A92 A101 4 A123 40 A143 A152 2 A173 1 A192 A201 1
A13 6 A34 A40 1343 A61 A75 1 A93 A101 4 A121 46 A143 A152 2 A173 2 A191 A202 1

A11 24 A34 A49 1382 A62 A74 4 A93 A101 1 A121 26 A143 A152 2 A173 1 A192 A201 1
A14 15 A32 A44 874 A65 A72 4 A92 A101 1 A121 24 A143 A152 1 A173 1 A191 A201 1
A11 12 A32 A42 3590 A61 A73 2 A93 A102 2 A122 29 A143 A152 1 A172 2 A191 A201 1
A12 11 A34 A40 1322 A64 A73 4 A92 A101 4 A123 40 A143 A152 2 A173 1 A191 A201 1
A11 18 A31 A43 1940 A61 A72 3 A93 A102 4 A124 36 A141 A153 1 A174 1 A192 A201 1
A14 36 A32 A43 3595 A61 A75 4 A93 A101 2 A123 28 A143 A152 1 A173 1 A191 A201 1
A11 9 A32 A40 1422 A61 A72 3 A93 A101 2 A124 27 A143 A153 1 A174 1 A192 A201 2
A14 30 A34 A43 6742 A65 A74 2 A93 A101 3 A122 36 A143 A152 2 A173 1 A191 A201 1
A14 24 A32 A41 7814 A61 A74 3 A93 A101 3 A123 38 A143 A152 1 A174 1 A192 A201 1
A14 24 A32 A41 9277 A65 A73 2 A91 A101 4 A124 48 A143 A153 1 A173 1 A192 A201 1
A12 30 A34 A40 2181 A65 A75 4 A93 A101 4 A121 36 A143 A152 2 A173 1 A191 A201 1
A14 18 A34 A43 1098 A61 A71 4 A92 A101 4 A123 65 A143 A152 2 A171 1 A191 A201 1
A12 24 A32 A42 4057 A61 A74 3 A91 A101 3 A123 43 A143 A152 1 A173 1 A192 A201 2
A11 12 A32 A46 795 A61 A72 4 A92 A101 4 A122 53 A143 A152 1 A173 1 A191 A201 2
A12 24 A34 A49 2825 A65 A74 4 A93 A101 3 A124 34 A143 A152 2 A173 2 A192 A201 1
A12 48 A32 A49 15672 A61 A73 2 A93 A101 2 A123 23 A143 A152 1 A173 1 A192 A201 2
A14 36 A34 A40 6614 A61 A75 4 A93 A101 4 A123 34 A143 A152 2 A174 1 A192 A201 1
A14 28 A31 A41 7824 A65 A72 3 A93 A103 4 A121 40 A141 A151 2 A173 2 A192 A201 1
A11 27 A34 A49 2442 A61 A75 4 A93 A101 4 A123 43 A142 A152 4 A174 2 A192 A201 1
A14 15 A34 A43 1829 A61 A75 4 A93 A101 4 A123 46 A143 A152 2 A173 1 A192 A201 1
A11 12 A34 A40 2171 A61 A73 4 A93 A101 4 A122 38 A141 A152 2 A172 1 A191 A202 1
A12 36 A34 A41 5800 A61 A73 3 A93 A101 4 A123 34 A143 A152 2 A173 1 A192 A201 1
A14 18 A34 A43 1169 A65 A73 4 A93 A101 3 A122 29 A143 A152 2 A173 1 A192 A201 1
A14 36 A33 A41 8947 A65 A74 3 A93 A101 2 A123 31 A142 A152 1 A174 2 A192 A201 1
A11 21 A32 A43 2606 A61 A72 4 A92 A101 4 A122 28 A143 A151 1 A174 1 A192 A201 1
A14 12 A34 A42 1592 A64 A74 3 A92 A101 2 A122 35 A143 A152 1 A173 1 A191 A202 1
A14 15 A32 A42 2186 A65 A74 1 A92 A101 4 A121 33 A141 A151 1 A172 1 A191 A201 1
A11 18 A32 A42 4153 A61 A73 2 A93 A102 3 A123 42 A143 A152 1 A173 1 A191 A201 2
A11 16 A34 A40 2625 A61 A75 2 A93 A103 4 A122 43 A141 A151 1 A173 1 A192 A201 2
A14 20 A34 A40 3485 A65 A72 2 A91 A101 4 A121 44 A143 A152 2 A173 1 A192 A201 1
A14 36 A34 A41 10477 A65 A75 2 A93 A101 4 A124 42 A143 A153 2 A173 1 A191 A201 1
A14 15 A32 A43 1386 A65 A73 4 A94 A101 2 A121 40 A143 A151 1 A173 1 A192 A201 1
A14 24 A32 A43 1278 A61 A75 4 A93 A101 1 A121 36 A143 A152 1 A174 1 A192 A201 1
A11 12 A32 A43 1107 A61 A73 2 A93 A101 2 A121 20 A143 A151 1 A174 2 A192 A201 1
A11 21 A32 A40 3763 A65 A74 2 A93 A102 2 A121 24 A143 A152 1 A172 1 A191 A202 1
A12 36 A32 A46 3711 A65 A73 2 A94 A101 2 A123 27 A143 A152 1 A173 1 A191 A201 1
A14 15 A33 A41 3594 A61 A72 1 A92 A101 2 A122 46 A143 A152 2 A172 1 A191 A201 1
A12 9 A32 A40 3195 A65 A73 1 A92 A101 2 A121 33 A143 A152 1 A172 1 A191 A201 1
A14 36 A33 A43 4454 A61 A73 4 A92 A101 4 A121 34 A143 A152 2 A173 1 A191 A201 1
A12 24 A34 A42 4736 A61 A72 2 A92 A101 4 A123 25 A141 A152 1 A172 1 A191 A201 2
A12 30 A32 A43 2991 A65 A75 2 A92 A101 4 A123 25 A143 A152 1 A173 1 A191 A201 1
A14 11 A32 A49 2142 A64 A75 1 A91 A101 2 A121 28 A143 A152 1 A173 1 A192 A201 1
A11 24 A31 A49 3161 A61 A73 4 A93 A101 2 A122 31 A143 A151 1 A173 1 A192 A201 2
A12 48 A30 A410 18424 A61 A73 1 A92 A101 2 A122 32 A141 A152 1 A174 1 A192 A202 2
A14 10 A32 A41 2848 A62 A73 1 A93 A102 2 A121 32 A143 A152 1 A173 2 A191 A201 1
A11 6 A32 A40 14896 A61 A75 1 A93 A101 4 A124 68 A141 A152 1 A174 1 A192 A201 2

A11 24 A32 A42 2359 A62 A71 1 A91 A101 1 A122 33 A143 A152 1 A173 1 A191 A201 2
A11 24 A32 A42 3345 A61 A75 4 A93 A101 2 A122 39 A143 A151 1 A174 1 A192 A201 2
A14 18 A34 A42 1817 A61 A73 4 A92 A101 2 A124 28 A143 A152 2 A173 1 A191 A201 1
A14 48 A33 A43 12749 A63 A74 4 A93 A101 1 A123 37 A143 A152 1 A174 1 A192 A201 1
A11 9 A32 A43 1366 A61 A72 3 A92 A101 4 A122 22 A143 A151 1 A173 1 A191 A201 2
A12 12 A32 A40 2002 A61 A74 3 A93 A101 4 A122 30 A143 A151 1 A173 2 A192 A201 1
A11 24 A31 A42 6872 A61 A72 2 A91 A101 1 A122 55 A141 A152 1 A173 1 A192 A201 2
A11 12 A31 A40 697 A61 A72 4 A93 A101 2 A123 46 A141 A152 2 A173 1 A192 A201 2
A11 18 A34 A42 1049 A61 A72 4 A92 A101 4 A122 21 A143 A151 1 A173 1 A191 A201 1
A11 48 A32 A41 10297 A61 A74 4 A93 A101 4 A124 39 A142 A153 3 A173 2 A192 A201 2
A14 30 A32 A43 1867 A65 A75 4 A93 A101 4 A123 58 A143 A152 1 A173 1 A192 A201 1
A11 12 A33 A40 1344 A61 A73 4 A93 A101 2 A121 43 A143 A152 2 A172 2 A191 A201 1
A11 24 A32 A42 1747 A61 A72 4 A93 A102 1 A122 24 A143 A152 1 A172 1 A191 A202 1
A12 9 A32 A43 1670 A61 A72 4 A92 A101 2 A123 22 A143 A152 1 A173 1 A192 A201 2
A14 9 A34 A40 1224 A61 A73 3 A93 A101 1 A121 30 A143 A152 2 A173 1 A191 A201 1
A14 12 A34 A43 522 A63 A75 4 A93 A101 4 A122 42 A143 A152 2 A173 2 A192 A201 1
A11 12 A32 A43 1498 A61 A73 4 A92 A101 1 A123 23 A141 A152 1 A173 1 A191 A201 1
A12 30 A33 A43 1919 A62 A72 4 A93 A101 3 A124 30 A142 A152 2 A174 1 A191 A201 2
A13 9 A32 A43 745 A61 A73 3 A92 A101 2 A121 28 A143 A152 1 A172 1 A191 A201 2
A12 6 A32 A43 2063 A61 A72 4 A94 A101 3 A123 30 A143 A151 1 A174 1 A192 A201 1
A12 60 A32 A46 6288 A61 A73 4 A93 A101 4 A124 42 A143 A153 1 A173 1 A191 A201 2
A14 24 A34 A41 6842 A65 A73 2 A93 A101 4 A122 46 A143 A152 2 A174 2 A192 A201 1
A14 12 A32 A40 3527 A65 A72 2 A93 A101 3 A122 45 A143 A152 1 A174 2 A192 A201 1
A14 10 A32 A40 1546 A61 A73 3 A93 A101 2 A121 31 A143 A152 1 A172 2 A191 A202 1
A14 24 A32 A42 929 A65 A74 4 A93 A101 2 A123 31 A142 A152 1 A173 1 A192 A201 1
A14 4 A34 A40 1455 A61 A74 2 A93 A101 1 A121 42 A143 A152 3 A172 2 A191 A201 1
A11 15 A32 A42 1845 A61 A72 4 A92 A103 1 A122 46 A143 A151 1 A173 1 A191 A201 1
A12 48 A30 A40 8358 A63 A72 1 A92 A101 1 A123 30 A143 A152 2 A173 1 A191 A201 1
A11 24 A31 A42 3349 A63 A72 4 A93 A101 4 A124 30 A143 A153 1 A173 2 A192 A201 2
A14 12 A32 A40 2859 A65 A71 4 A93 A101 4 A124 38 A143 A152 1 A174 1 A192 A201 1
A14 18 A32 A42 1533 A61 A72 4 A94 A102 1 A122 43 A143 A152 1 A172 2 A191 A201 2
A14 24 A32 A43 3621 A62 A75 2 A93 A101 4 A123 31 A143 A152 2 A173 1 A191 A201 2
A12 18 A34 A49 3590 A61 A71 3 A94 A101 3 A123 40 A143 A152 3 A171 2 A192 A201 1
A11 36 A33 A49 2145 A61 A74 2 A93 A101 1 A123 24 A143 A152 2 A173 1 A192 A201 2
A12 24 A32 A41 4113 A63 A72 3 A92 A101 4 A123 28 A143 A151 1 A173 1 A191 A201 2
A14 36 A32 A42 10974 A61 A71 4 A92 A101 2 A123 26 A143 A152 2 A174 1 A192 A201 2
A11 12 A32 A40 1893 A61 A73 4 A92 A103 4 A122 29 A143 A152 1 A173 1 A192 A201 1
A11 24 A34 A43 1231 A64 A75 4 A92 A101 4 A122 57 A143 A151 2 A174 1 A192 A201 1
A13 30 A34 A43 3656 A65 A75 4 A93 A101 4 A122 49 A142 A152 2 A172 1 A191 A201 1
A12 9 A34 A43 1154 A61 A75 2 A93 A101 4 A121 37 A143 A152 3 A172 1 A191 A201 1
A11 28 A32 A40 4006 A61 A73 3 A93 A101 2 A123 45 A143 A152 1 A172 1 A191 A201 2
A12 24 A32 A42 3069 A62 A75 4 A93 A101 4 A124 30 A143 A153 1 A173 1 A191 A201 1
A14 6 A34 A43 1740 A61 A75 2 A94 A101 2 A121 30 A143 A151 2 A173 1 A191 A201 1
A12 21 A33 A40 2353 A61 A73 1 A91 A101 4 A122 47 A143 A152 2 A173 1 A191 A201 1
A14 15 A32 A40 3556 A65 A73 3 A93 A101 2 A124 29 A143 A152 1 A173 1 A191 A201 1
A14 24 A32 A43 2397 A63 A75 3 A93 A101 2 A123 35 A141 A152 2 A173 1 A192 A201 2

A12 6 A32 A45 454 A61 A72 3 A94 A101 1 A122 22 A143 A152 1 A172 1 A191 A201 1
A12 30 A32 A43 1715 A65 A73 4 A92 A101 1 A123 26 A143 A152 1 A173 1 A191 A201 1
A12 27 A34 A43 2520 A63 A73 4 A93 A101 2 A122 23 A143 A152 2 A172 1 A191 A201 2
A14 15 A32 A43 3568 A61 A75 4 A92 A101 2 A123 54 A141 A151 1 A174 1 A192 A201 1
A14 42 A32 A43 7166 A65 A74 2 A94 A101 4 A122 29 A143 A151 1 A173 1 A192 A201 1
A11 11 A34 A40 3939 A61 A73 1 A93 A101 2 A121 40 A143 A152 2 A172 2 A191 A201 1
A12 15 A32 A45 1514 A62 A73 4 A93 A103 2 A121 22 A143 A152 1 A173 1 A191 A201 1
A14 24 A32 A40 7393 A61 A73 1 A93 A101 4 A122 43 A143 A152 1 A172 2 A191 A201 1
A11 24 A31 A40 1193 A61 A71 1 A92 A102 4 A124 29 A143 A151 2 A171 1 A191 A201 2
A11 60 A32 A49 7297 A61 A75 4 A93 A102 4 A124 36 A143 A151 1 A173 1 A191 A201 2
A14 30 A34 A43 2831 A61 A73 4 A92 A101 2 A123 33 A143 A152 1 A173 1 A192 A201 1
A13 24 A32 A43 1258 A63 A73 3 A92 A101 3 A123 57 A143 A152 1 A172 1 A191 A201 1
A12 6 A32 A43 753 A61 A73 2 A92 A103 3 A121 64 A143 A152 1 A173 1 A191 A201 1
A12 18 A33 A49 2427 A65 A75 4 A93 A101 2 A122 42 A143 A152 2 A173 1 A191 A201 1
A14 24 A33 A40 2538 A61 A75 4 A93 A101 4 A123 47 A143 A152 2 A172 2 A191 A201 2
A12 15 A31 A40 1264 A62 A73 2 A94 A101 2 A122 25 A143 A151 1 A173 1 A191 A201 2
A12 30 A34 A42 8386 A61 A74 2 A93 A101 2 A122 49 A143 A152 1 A173 1 A191 A201 2
A14 48 A32 A49 4844 A61 A71 3 A93 A101 2 A123 33 A141 A151 1 A174 1 A192 A201 2
A13 21 A32 A40 2923 A62 A73 1 A92 A101 1 A123 28 A141 A152 1 A174 1 A192 A201 1
A11 36 A32 A41 8229 A61 A73 2 A93 A101 2 A122 26 A143 A152 1 A173 2 A191 A201 2
A14 24 A34 A42 2028 A61 A74 2 A93 A101 2 A122 30 A143 A152 2 A172 1 A191 A201 1
A11 15 A34 A42 1433 A61 A73 4 A92 A101 3 A122 25 A143 A151 2 A173 1 A191 A201 1
A13 42 A30 A49 6289 A61 A72 2 A91 A101 1 A122 33 A143 A152 2 A173 1 A191 A201 1
A14 13 A32 A43 1409 A62 A71 2 A92 A101 4 A121 64 A143 A152 1 A173 1 A191 A201 1
A11 24 A32 A41 6579 A61 A71 4 A93 A101 2 A124 29 A143 A153 1 A174 1 A192 A201 1
A12 24 A34 A43 1743 A61 A75 4 A93 A101 2 A122 48 A143 A152 2 A172 1 A191 A201 1
A14 12 A34 A46 3565 A65 A72 2 A93 A101 1 A122 37 A143 A152 2 A172 2 A191 A201 1
A14 15 A31 A43 1569 A62 A75 4 A93 A101 4 A123 34 A141 A152 1 A172 2 A191 A201 1
A11 18 A32 A43 1936 A65 A74 2 A94 A101 4 A123 23 A143 A151 2 A172 1 A191 A201 1
A11 36 A32 A42 3959 A61 A71 4 A93 A101 3 A122 30 A143 A152 1 A174 1 A192 A201 1
A14 12 A32 A40 2390 A65 A75 4 A93 A101 3 A123 50 A143 A152 1 A173 1 A192 A201 1
A14 12 A32 A42 1736 A61 A74 3 A92 A101 4 A121 31 A143 A152 1 A172 1 A191 A201 1
A11 30 A32 A41 3857 A61 A73 4 A91 A101 4 A122 40 A143 A152 1 A174 1 A192 A201 1
A14 12 A32 A43 804 A61 A75 4 A93 A101 4 A123 38 A143 A152 1 A173 1 A191 A201 1
A11 45 A32 A43 1845 A61 A73 4 A93 A101 4 A124 23 A143 A153 1 A173 1 A192 A201 2
A12 45 A34 A41 4576 A62 A71 3 A93 A101 4 A123 27 A143 A152 1 A173 1 A191 A201 1

A.2 An overview of the normalized data structure

Response	A2	A5	A8	A11	A13	A16	A18	A1_1	A1_2	A1_3	A3_1	A3_2	A3_3	A4_1	A4_2	A
0	6	1169	4	4	67	2	1	0	0	0	0	0	0	1	0	0
1	48	5951	2	2	22	1	1	1	0	0	0	0	1	1	0	0
0	12	2096	2	3	49	1	2	0	0	1	0	0	0	0	1	0
0	42	7882	2	4	45	1	2	0	0	0	0	0	1	1	0	0
1	24	4870	3	4	53	2	2	0	0	0	0	0	0	1	0	0
0	36	9055	2	4	35	1	2	0	0	1	0	0	1	0	1	0

0	24	2835	3	4	53	1	1	0	0	1	0	0	1	1	0	0
0	36	6948	2	2	35	1	1	1	0	0	0	0	1	1	0	0
0	12	3059	2	4	61	1	1	0	0	1	0	0	1	1	0	0
1	30	5234	4	2	28	2	1	1	0	0	0	0	0	1	0	0
1	12	1295	3	1	25	1	1	1	0	0	0	0	1	1	0	0
1	48	4308	3	4	24	1	1	0	0	0	0	0	1	0	0	0
0	12	1567	1	1	22	1	1	1	0	0	0	0	1	1	0	0
1	24	1199	4	4	60	2	1	0	0	0	0	0	0	1	0	0
0	15	1403	2	4	28	1	1	0	0	0	0	0	1	1	0	0
1	24	1282	4	2	32	1	1	0	0	0	0	0	1	1	0	1
0	24	2424	4	4	53	2	1	0	0	1	0	0	0	1	0	0
0	30	8072	2	3	25	3	1	0	0	0	1	0	0	0	0	0
1	24	####	4	2	44	1	1	1	0	0	0	0	1	1	0	0
0	24	3430	3	2	31	1	2	0	0	1	0	0	1	1	0	0
0	9	2134	4	4	48	3	1	0	0	1	0	0	0	1	0	0
0	6	2647	2	3	44	1	2	0	0	0	0	0	1	1	0	0
0	10	2241	1	3	48	2	2	0	0	0	0	0	0	1	0	0
0	12	1804	3	4	44	1	1	1	0	0	0	0	0	1	0	1
0	10	2069	2	1	26	2	1	0	0	1	0	0	0	1	0	0
0	6	1374	1	2	36	1	1	0	0	0	0	0	1	1	0	0
0	6	426	4	4	39	1	1	0	0	1	1	0	0	1	0	0
0	12	409	3	3	42	2	1	0	1	0	0	1	0	1	0	0
0	7	2415	3	2	34	1	1	1	0	0	0	0	1	1	0	0
1	60	6836	3	4	63	2	1	0	0	0	0	0	0	0	0	0
0	18	1913	3	3	36	1	1	1	0	0	0	0	1	0	0	0
0	24	4020	2	2	27	1	1	0	0	0	0	0	1	1	0	0
0	18	5866	2	2	30	2	1	1	0	0	0	0	1	1	0	1
0	12	1264	4	4	57	1	1	0	0	1	0	0	0	0	0	0
0	12	1474	4	1	33	1	1	0	1	0	0	0	1	1	0	0
1	45	4746	4	2	25	2	1	1	0	0	0	0	0	1	0	0
0	48	6110	1	3	31	1	1	0	0	1	0	0	0	0	1	0
1	18	2100	4	2	37	1	1	0	1	0	0	0	1	1	0	0
0	10	1225	2	2	37	1	1	0	1	0	0	0	1	1	0	0
0	9	458	4	3	24	1	1	1	0	0	0	0	1	1	0	0
0	30	2333	4	2	30	1	1	0	0	1	0	0	1	1	0	0
0	12	1158	3	1	26	1	1	1	0	0	0	0	1	1	0	0
0	18	6204	2	4	44	1	2	1	0	0	0	0	0	1	0	0
0	30	6187	1	4	24	2	1	0	0	0	0	0	0	1	0	1
1	48	6143	4	4	58	2	1	0	0	0	0	0	0	1	0	0
0	11	1393	4	4	35	2	1	0	0	1	0	0	0	1	0	0
0	36	2299	4	4	39	1	1	0	0	1	0	0	1	1	0	0
0	6	1352	1	2	23	1	1	0	0	0	0	0	1	1	0	0
0	11	7228	1	4	39	2	1	0	0	1	0	0	0	1	0	0
0	12	2073	4	2	28	1	1	0	0	1	0	0	1	1	0	1
0	24	2333	4	2	29	1	1	1	0	0	0	0	0	1	0	0
0	27	5965	1	2	30	2	1	1	0	0	0	0	0	1	0	0

0	12	1262	3	2	25	1	1	0	0	1	0	0	1	1	0	0
0	18	3378	2	1	31	1	1	0	0	1	0	0	1	1	0	0
1	36	2225	4	4	57	2	1	1	0	0	0	0	0	1	0	0
0	6	783	1	2	26	1	2	0	0	1	0	1	0	1	0	0
1	12	6468	2	1	52	1	1	1	0	0	0	0	1	1	0	0
0	36	9566	2	2	31	2	1	0	0	1	0	0	0	1	0	0
0	18	1961	3	2	23	1	1	0	1	0	0	0	1	1	0	0
1	36	6229	4	4	23	2	1	0	0	0	0	0	0	1	0	0
0	9	1391	2	1	27	1	1	1	0	0	0	0	1	0	0	0
0	15	1537	4	4	50	2	1	1	0	0	0	0	0	1	0	0
1	36	1953	4	4	61	1	1	1	0	0	1	0	0	0	0	0
1	48	####	2	2	25	1	1	1	0	0	1	0	0	0	0	0
0	24	3181	4	4	26	1	1	0	0	1	0	0	1	1	0	0
0	27	5190	4	4	48	4	2	0	0	1	0	0	1	1	0	0
0	12	2171	2	2	29	1	1	0	0	1	0	0	1	1	0	0
0	12	1007	4	1	22	1	1	1	0	0	0	0	1	1	0	0
1	36	1819	4	4	37	1	1	0	0	1	0	0	1	0	1	0
0	36	2394	4	4	25	1	1	0	0	1	0	0	1	1	0	0
0	36	8133	1	2	30	1	1	0	0	1	0	0	1	1	0	0
0	7	730	4	2	46	2	1	0	0	1	0	0	0	1	0	0
0	8	1164	3	4	51	2	2	0	0	0	0	0	0	0	0	0
0	42	5954	2	1	41	2	1	1	0	0	0	0	0	0	0	0
1	36	1977	4	4	40	1	1	0	0	0	0	0	1	0	1	0
0	12	1526	4	4	66	2	1	0	0	0	0	0	0	1	0	0
1	42	3965	4	3	34	1	1	0	0	0	0	0	1	1	0	0
0	11	4771	2	4	51	1	1	1	0	0	0	0	0	1	0	0
0	54	9436	2	2	39	1	2	0	0	1	1	0	0	1	0	0
0	30	3832	2	1	22	1	1	1	0	0	0	0	1	1	0	0
1	24	5943	1	1	44	2	1	0	0	1	0	0	1	1	0	0
0	15	1213	4	3	47	1	1	0	0	1	0	0	1	1	0	0
0	18	1568	3	4	24	1	1	0	0	1	0	0	1	0	0	1
0	24	1755	4	4	58	1	1	0	0	0	0	0	1	0	0	0
0	10	2315	3	4	52	1	1	0	0	0	0	0	1	1	0	0
0	12	1412	4	2	29	2	1	0	0	1	0	0	0	0	0	0
0	18	1295	4	1	27	2	1	1	0	0	0	0	0	1	0	0
1	36	####	1	4	47	1	2	1	0	0	0	0	1	0	1	1
0	18	2249	4	3	30	1	2	0	0	0	0	0	1	1	0	1
1	12	1108	4	3	28	2	1	0	0	0	1	0	0	1	0	0
0	12	618	4	4	56	1	1	0	0	1	0	0	0	1	0	0
0	12	1409	4	3	54	1	1	0	0	0	0	0	0	1	0	0
1	12	797	4	3	33	1	2	0	0	1	0	0	0	1	0	0
0	24	3617	4	4	20	2	1	0	1	0	0	0	0	1	0	0
0	12	1318	4	4	54	1	1	1	0	0	0	0	1	1	0	0
1	54	####	3	4	58	1	1	1	0	0	1	0	0	0	0	0
0	12	2012	4	2	61	1	1	0	0	1	0	0	0	0	1	0
0	18	2622	4	4	34	1	1	1	0	0	0	0	1	0	0	1

0	36	2337	4	4	36	1	1	1	0	0	0	0	0	1	0	0
0	20	7057	3	4	36	2	2	1	0	0	0	0	0	1	0	0
0	24	1469	4	4	41	1	1	0	0	1	0	0	1	1	0	1
0	36	2323	4	4	24	1	1	1	0	0	0	0	1	1	0	0
0	6	932	3	2	24	1	1	0	0	1	0	0	0	1	0	0
0	9	1919	4	3	35	1	1	1	0	0	0	0	0	1	0	0
0	12	2445	2	4	26	1	1	0	0	1	0	0	1	1	0	0
1	24	####	2	3	39	2	2	1	0	0	0	0	0	0	0	0
1	18	6458	2	4	39	2	2	0	0	1	0	1	0	1	0	0
0	12	6078	2	2	32	1	1	1	0	0	0	0	1	1	0	0
0	24	7721	1	2	30	1	1	0	0	0	0	0	1	1	0	0
0	14	1410	1	2	35	1	1	1	0	0	0	0	1	0	0	0
0	6	1449	1	2	31	2	2	1	0	0	0	0	0	0	0	1
0	15	392	4	4	23	1	1	0	1	0	0	0	1	0	1	0
0	18	6260	3	3	28	1	1	1	0	0	0	0	1	1	0	0
1	36	7855	4	2	25	2	1	0	0	1	0	0	0	1	0	0
0	12	1680	3	1	35	1	1	0	0	0	0	0	1	1	0	0
0	48	3578	4	1	47	1	1	0	0	1	0	0	0	1	0	0
1	42	7174	4	3	30	1	1	0	0	0	0	0	1	1	0	0
0	10	2132	2	3	27	2	1	0	0	0	0	0	0	1	0	0
1	33	4281	1	4	23	2	1	0	0	0	0	0	0	1	0	0
0	12	2366	3	3	36	1	1	1	0	0	0	0	0	1	0	0
1	21	1835	3	2	25	2	1	0	0	0	0	0	1	1	0	0
0	24	3868	4	2	41	2	1	0	0	1	0	0	0	1	0	0
0	12	1768	3	2	24	1	1	0	0	1	0	0	1	1	0	0
0	10	781	4	4	63	2	1	0	1	0	0	0	0	1	0	0
1	18	1924	4	3	27	1	1	1	0	0	0	0	1	1	0	0
0	12	2121	4	2	30	2	1	0	0	0	0	0	0	1	0	0
0	12	701	4	2	40	1	1	0	0	0	0	0	1	1	0	0
1	12	639	4	2	30	1	1	1	0	0	0	0	1	1	0	0
0	12	1860	4	2	34	2	1	1	0	0	0	0	0	1	0	0
1	12	3499	3	2	29	2	1	0	0	0	0	0	0	1	0	0
0	48	8487	1	2	24	1	1	1	0	0	0	0	1	1	0	0
1	36	6887	4	3	29	1	1	0	0	0	0	0	0	0	1	0
0	15	2708	2	3	27	2	1	0	0	1	0	0	1	1	0	0
0	18	1984	4	4	47	2	1	0	0	1	0	0	1	1	0	0
0	60	####	2	4	21	1	1	0	0	1	0	0	1	1	0	1
0	12	1240	4	2	38	2	1	0	0	1	0	0	0	1	0	0
0	27	8613	2	2	27	2	1	0	0	1	0	0	0	1	0	0
1	12	766	4	3	66	1	1	1	0	0	0	0	1	1	0	0
0	15	2728	4	2	35	3	1	1	0	0	0	0	0	1	0	0
0	12	1881	2	2	44	1	1	0	1	0	0	0	1	1	0	0
0	6	709	2	2	27	1	1	0	1	0	0	0	1	1	0	0
0	36	4795	4	1	30	1	1	1	0	0	0	0	1	1	0	0
0	27	3416	3	2	27	1	1	0	0	0	0	0	1	1	0	0
1	18	2462	2	2	22	1	1	0	0	0	0	0	1	1	0	0

0	21	2288	4	4	23	1	1	0	0	1	0	0	0	1	0	0
0	48	3566	4	2	30	1	1	1	0	0	0	1	0	0	0	1
0	6	860	1	4	39	2	1	0	0	0	0	0	0	1	0	0
0	12	682	4	3	51	2	1	0	0	1	0	0	0	1	0	1
0	36	5371	3	2	28	2	1	0	0	0	0	0	0	1	0	0
0	18	1582	4	4	46	2	1	0	0	1	0	0	0	1	0	0
0	6	1346	2	4	42	1	2	0	0	1	0	0	1	1	0	1
0	10	1924	1	4	38	1	1	0	0	1	0	0	1	1	0	0
0	36	5848	4	1	24	1	1	0	1	0	0	0	1	1	0	0
0	24	7758	2	4	29	1	1	1	0	0	0	0	0	1	0	0
0	24	6967	4	4	36	1	1	1	0	0	0	0	0	0	0	1
1	12	1282	2	4	20	1	1	0	0	0	0	0	1	1	0	0
0	9	1288	3	4	48	2	2	0	0	0	0	0	0	1	0	1
0	12	339	4	1	45	1	1	0	0	0	0	1	0	0	1	0
0	24	3512	2	3	38	2	1	1	0	0	0	0	1	1	0	1
0	6	1898	1	2	34	2	2	0	0	1	0	0	0	1	0	0
0	24	2872	3	4	36	1	2	0	0	1	0	0	0	1	0	1
0	18	1055	4	1	30	2	1	0	0	1	0	0	0	1	0	0
0	15	1262	4	3	36	2	1	0	0	1	0	0	1	1	0	0
0	10	7308	2	4	70	1	1	1	0	0	0	0	1	1	0	0
0	36	909	4	4	36	1	1	0	0	1	0	0	1	1	0	0
0	6	2978	1	2	32	1	1	0	0	1	0	0	1	1	0	0
1	18	1131	4	2	33	1	1	0	0	0	0	0	1	1	0	0
0	11	1577	4	1	20	1	1	1	0	0	0	0	1	1	0	0
0	24	3972	2	4	25	1	1	0	0	1	0	0	1	1	0	0
1	24	1935	4	4	31	2	1	1	0	0	0	0	0	0	0	0
1	15	950	4	3	33	2	2	0	0	0	1	0	0	1	0	0
0	12	763	4	1	26	1	1	0	0	1	0	0	1	1	0	0
1	24	2064	3	2	34	1	1	1	0	0	0	0	0	1	0	0
0	8	1414	4	2	33	1	1	1	0	0	0	0	1	1	0	0
1	21	3414	2	1	26	2	1	0	0	0	0	0	0	0	1	0
1	30	7485	4	1	53	1	1	0	0	1	0	1	0	1	0	0
0	12	2577	2	1	42	1	1	0	0	0	0	0	1	1	0	0
0	6	338	4	4	52	2	1	0	0	0	0	0	0	1	0	0
0	12	1963	4	2	31	2	2	0	0	1	0	0	1	1	0	0
0	21	571	4	4	65	2	1	0	0	0	0	0	0	1	0	0
1	36	9572	1	1	28	2	1	0	0	1	0	0	0	0	0	0
1	36	4455	2	2	30	2	1	1	0	0	0	0	0	0	0	0
1	21	1647	4	2	40	2	2	0	0	0	0	1	0	1	0	0
0	24	3777	4	4	50	1	1	0	0	1	0	0	0	1	0	0
1	18	884	4	4	36	1	2	1	0	0	0	0	0	1	0	0
0	15	1360	4	2	31	2	1	0	0	1	0	0	0	1	0	0
1	9	5129	2	4	74	1	2	1	0	0	0	1	0	1	0	0
0	16	1175	2	3	68	3	1	1	0	0	0	0	0	1	0	0
1	12	674	4	1	20	1	1	0	0	0	0	0	1	1	0	1
0	18	3244	1	4	33	2	1	1	0	0	1	0	0	1	0	0

1	24	4591	2	3	54	3	1	0	0	1	0	0	1	0	0	0
1	48	3844	4	4	34	1	2	1	0	0	1	0	0	0	0	1
1	27	3915	4	2	36	1	2	1	0	0	0	0	1	0	0	0
0	6	2108	2	2	29	1	1	0	0	1	0	0	1	1	0	0
1	45	3031	4	4	21	1	1	1	0	0	0	0	1	1	0	1
1	9	1501	2	3	34	2	1	1	0	0	0	0	0	0	1	0
0	6	1382	1	1	28	2	1	0	0	1	0	0	0	1	0	0
1	12	951	4	4	27	4	1	1	0	0	0	0	1	1	0	1
0	24	2760	4	4	36	1	1	1	0	0	0	0	1	1	0	0
1	18	4297	4	3	40	1	1	1	0	0	0	0	0	1	0	0
0	9	936	4	2	52	2	1	0	0	1	0	0	0	0	1	0
0	12	1168	4	3	27	1	1	0	0	0	0	0	1	1	0	0
0	27	5117	3	4	26	2	1	0	0	1	0	0	0	0	0	0
1	12	902	4	4	21	1	1	0	0	0	0	0	1	0	1	0
0	12	1495	4	1	38	2	2	0	0	1	0	0	0	1	0	0
0	30	####	3	4	38	3	2	0	0	0	0	0	0	1	0	0
0	12	1935	4	4	43	3	1	0	0	1	0	0	0	1	0	0
0	12	1424	4	3	26	1	1	1	0	0	0	0	0	1	0	0
0	24	6568	2	2	21	1	1	0	0	0	0	0	1	0	0	0
0	12	1413	3	2	55	1	1	0	0	1	0	0	1	1	0	0
0	9	3074	1	2	33	2	2	0	0	1	0	0	0	1	0	0
0	36	3835	2	4	45	1	1	0	0	1	0	0	1	1	0	0
1	27	5293	2	4	50	2	1	0	0	0	1	0	0	0	0	0
1	30	1908	4	4	66	1	1	0	1	0	0	0	0	0	0	0
0	36	3342	4	2	51	1	1	0	0	1	0	0	0	1	0	0
0	6	932	1	3	39	2	1	1	0	0	0	0	0	0	1	0
0	18	3104	3	1	31	1	1	0	0	0	1	0	0	0	0	0
0	36	3913	2	2	23	1	1	0	1	0	0	0	1	1	0	0
0	24	3021	2	2	24	1	1	0	0	0	0	0	1	1	0	0
0	10	1364	2	4	64	1	1	0	0	1	0	0	1	1	0	0
0	12	625	4	1	26	1	1	1	0	0	0	0	1	1	0	0
0	12	1200	4	4	23	1	1	0	0	0	0	0	1	0	1	0
0	12	707	4	2	30	2	1	0	0	1	0	0	1	1	0	0
0	24	2978	4	4	32	2	2	0	0	1	0	0	0	0	0	0
0	15	4657	3	2	30	1	1	0	0	1	0	0	1	1	0	0
0	36	2613	4	2	27	2	1	0	0	1	1	0	0	1	0	0
1	48	####	1	2	27	2	1	1	0	0	0	0	1	1	0	0
1	12	7865	4	4	53	1	1	0	0	0	0	0	1	1	0	0
1	9	1478	4	2	22	1	1	0	0	1	0	0	1	1	0	0
0	24	3149	4	1	22	1	1	0	0	0	0	0	1	1	0	0
1	36	4210	4	2	26	1	1	0	1	0	0	0	1	1	0	0
0	9	2507	2	4	51	1	1	0	0	1	0	0	1	1	0	0
0	12	2141	3	1	35	1	1	0	0	1	0	0	1	1	0	1
0	18	866	4	2	25	1	1	1	0	0	0	0	1	1	0	0
0	4	1544	2	1	42	3	2	0	0	1	0	0	0	1	0	0
1	24	1823	4	2	30	1	2	0	0	0	0	0	1	1	0	0

1	6	####	1	2	23	1	1	1	0	0	0	0	1	1	0	0
1	21	2767	4	2	61	2	1	1	0	0	0	0	1	0	0	1
0	12	1291	4	2	35	2	1	0	0	1	0	0	0	1	0	0
0	30	2522	1	3	39	1	2	0	0	0	0	0	1	1	0	0
1	24	915	4	2	29	1	1	0	0	0	0	0	1	1	0	0
0	6	1595	3	2	51	1	2	0	0	1	0	0	1	1	0	0
1	48	4605	3	4	24	2	2	0	0	0	1	0	0	1	0	0
0	12	1185	3	2	27	2	1	0	0	1	0	0	0	0	0	0
0	12	3447	4	3	35	1	2	0	0	1	0	1	0	0	1	0
0	24	1258	4	1	25	1	1	0	0	1	0	0	1	0	0	0
0	12	717	4	4	52	3	1	0	0	1	0	0	0	1	0	0
0	6	1204	4	1	35	1	1	0	0	1	1	0	0	1	0	1
0	24	1925	2	2	26	1	1	0	1	0	0	0	1	1	0	0
1	18	433	3	4	22	1	1	0	0	1	0	0	1	1	0	0
0	6	666	3	4	39	2	1	0	0	0	0	0	0	1	0	0
0	12	2251	1	2	46	1	1	0	1	0	0	0	1	1	0	0
1	30	2150	4	2	24	1	1	1	0	0	0	0	1	1	0	0
0	24	4151	2	3	35	2	1	0	0	1	0	0	0	1	0	1
0	9	2030	2	1	24	1	1	1	0	0	0	0	1	1	0	0
0	60	7418	1	1	27	1	1	1	0	0	0	0	0	1	0	0
0	24	2684	4	2	35	2	1	0	0	1	0	0	0	1	0	0
1	12	2149	4	1	29	1	1	0	0	0	0	1	0	1	0	0
0	15	3812	1	4	23	1	1	0	0	1	0	0	1	1	0	1
0	11	1154	4	4	57	3	1	0	0	1	0	0	0	1	0	1
0	12	1657	2	2	27	1	1	0	0	0	0	0	1	1	0	0
0	24	1603	4	4	55	1	1	0	0	0	0	0	1	1	0	0
0	18	5302	2	4	36	3	1	0	0	0	0	0	0	1	0	0
0	12	2748	2	4	57	3	1	0	0	1	0	0	0	0	1	0
0	10	1231	3	4	32	2	2	0	0	1	0	0	0	1	0	0
1	15	802	4	3	37	1	2	1	0	0	0	0	1	1	0	0
0	36	6304	4	4	36	2	1	0	0	1	0	0	0	0	0	0
0	24	1533	4	3	38	1	1	0	0	1	0	0	1	1	0	0
1	14	8978	1	4	45	1	1	0	0	0	0	0	1	1	0	0
0	24	999	4	2	25	2	1	0	0	1	0	0	1	1	0	0
0	18	2662	4	3	32	1	1	0	0	1	0	0	1	1	0	0
0	12	1402	3	4	37	1	1	0	0	1	0	0	0	1	0	0
0	48	####	4	4	36	1	1	1	0	0	0	1	0	1	0	0
1	48	3060	4	4	28	2	1	1	0	0	0	0	1	1	0	0
1	30	####	1	1	34	1	1	0	0	0	0	0	1	1	0	0
0	9	2697	1	2	32	1	2	0	0	1	0	0	1	1	0	0
0	18	2404	2	2	26	2	1	0	0	1	0	0	0	1	0	0
0	12	1262	2	4	49	1	1	0	0	0	0	0	1	1	0	0
1	6	4611	1	4	32	1	1	0	0	1	0	0	1	1	0	0
0	24	1901	4	4	29	1	1	0	0	1	0	0	1	1	0	1
0	15	3368	3	4	23	2	1	0	0	1	0	0	0	1	0	0
0	12	1574	4	2	50	1	1	0	0	1	0	0	1	1	0	0

0	18	1445	4	4	49	1	1	0	1	0	0	1	0	1	0	0
0	15	1520	4	4	63	1	1	0	0	1	0	0	0	1	0	0
0	24	3878	4	2	37	1	1	1	0	0	0	0	0	1	0	1
0	47	####	1	1	35	1	1	0	0	0	0	0	1	1	0	0
0	48	4788	4	3	26	1	2	0	0	0	0	0	1	1	0	0
0	48	7582	2	4	31	1	1	1	0	0	0	0	0	0	0	1
0	12	1092	4	4	49	2	1	1	0	0	0	0	1	1	0	0
1	24	1024	4	4	48	1	1	0	0	0	0	0	0	1	0	0
0	12	1076	2	2	26	1	1	0	0	1	0	0	1	0	0	0
1	36	9398	1	4	28	1	1	1	0	0	0	0	1	1	0	0
0	24	6419	2	4	44	2	2	0	0	0	0	0	0	1	0	0
0	42	4796	4	4	56	1	1	0	1	0	0	0	0	1	0	0
0	48	7629	4	2	46	2	2	0	0	1	0	0	0	0	0	0
1	48	9960	1	2	26	1	1	1	0	0	0	0	1	1	0	0
0	12	4675	1	4	20	1	1	0	0	1	0	0	1	1	0	0
0	10	1287	4	2	45	1	1	0	0	1	0	0	1	1	0	0
0	18	2515	3	4	43	1	1	0	0	1	0	0	1	1	0	0

A

APPENDIX B

R codes snippets

1. R code for ranking the German bank credit data in order of importance with respect to the response target variable as part of data pre-processing process applied to the dataset.

```
##### Find the Variables Importance
library(RSNNS)
x <- data1[, c('A2', 'A5', 'A8', 'A11', 'A13', 'A16',
'A18', 'A1_1', 'A1_2', 'A1_3', 'A3_1', 'A3_2', 'A3_3',
'A4_1',
'A4_2', 'A6_1', 'A6_2', 'A6_3', 'A7_1', 'A7_2',
'A7_3', 'A9_1',
'A9_2', 'A10_1', 'A10_2', 'A12_1', 'A12_2', 'A12_3', 'A14_1', 'A14_2',
'A15_1', 'A15_2', 'A17_1', 'A17_2', 'A17_3', 'A19_1', 'A20_1')]

y <- data1[, 'Response']
mod <- mlp(x, y, size = 5)
library(NeuralNetTools)
garson(mod)

###
rand.vars=data.frame(A2, A5, A8, A11, A13,
A16, A18, A1_1, A1_2, A1_3, A3_1, A3_2, A3_3,
A4_1, A4_2, A6_1, A6_2, A6_3, A7_1, A7_2, A7_3,
A9_1, A9_2, A10_1, A10_2, A12_1, A12_2, A12_3, A14_1,
A14_2, A15_1, A15_2, A17_1, A17_2, A17_3, A19_1, A20_1)
y=data.frame(Response)
mod1<-nnet(rand.vars,y,size=8,linout=T)

#import 'gar.fun' from Github
library(devtools)
source_url('https://gist.githubusercontent.com/fawda123/6206737/
raw/d6f365c283a8cae23fb20892dc223bc5764d50c7/gar_fun.r')
```

```

#create a color vector for the bar plot
#num.vars<- 37
#cols<-colorRampPalette(c('green','blue'))(num.vars)
#use the function on the model created
#par(mar=c(6,8,1,1),family='serif')

gar.fun('y',mod1)
rel.imp<-gar.fun('Y',mod1,bar.plot=F)$rel.imp
rel.imp=data.frame(rel.imp)
write.csv(rel.imp,file = "C:\\Users\\documents\\rel.imp.csv")

```

2. R code for the voting/average model aggregation.

```

## aggregate the created base models by voting/averaging
predict.ensemble.basic <- function(models, data, predf=predict)
{
  bp <- data.frame(lapply(models, function(h) predf(h, data)))
  combf <- if (is.numeric(bp[,1])) mean else modal # combination scheme
  cp <- sapply(1:nrow(bp), function(i)
    combf(as.vector(as.matrix(bp[i,]))))
  if (is.factor(bp[,1]))
    factor(cp, levels=levels(bp[,1]))
  else
    cp
}

```

3. Misclassification error code

```

werr <- function(pred.y, true.y, w=rep(1, length(true.y)))
{ weighted.mean(pred.y!=true.y, w) }

# double weight for the least frequent class
s.w2test <- ifelse(s.test$Class=="german_credit_data", 2, 1)
werr(predict(s.tree, s.test, type="c"), s.test$Class,s.w2test)

```

APPENDIX C

LIST OF PUBLICATIONS

1. **Discriminating input variables for fraud detection using radial basis function network.** *Circulation in Computer Science (2018)*, 3(1), 1-9.
2. **Financial fraud detection using radial basis function network.** *Circulation in Computer Science (2018)*, 3(1), 10-21.
3. **Detecting Fraud Transactions Using Radial Basis Function-Artificial Neural Network.** *National Mathematics Society 45th conference proceeding*, May 3 – 6, 2016.
4. **Towards Modeling Financial Fraud Detection using Radial Basis Function-Artificial Neural Network.** *2015 ISTEAMS book of proceedings (series 7)*, pp. 733 – 742.