Sudan University of Science & Technology
College of Graduate Studies
جامعة السودان للعلوم والتكنولوجيا
كلية الدراسات العليا
كلية الدراسات العليا

Approval Page

(To be completed after the college council approval)

Name of Candidate: Mozdalefa Abdalla Ibrahim

Thesis title: Building A Classification Model to Identify the Effect of APGAR score Factor on Cesarian Operation
بناء نموذج تصنيف لمعرفة تأثير معامل APGAR على الولادة القيصرية

Degree Examined for: ……………………

Approved by:

1. External Examiner

Name: Dr. Hozeifa Adam Abd Alshafy

Signature: …………… Date: 21/4/2019

2. Internal Examiner

Name: Wafaa Faisal Mukhtar

Signature: …………… Date: 21/4/2019

3. Supervisor

Name: Albarra Abuobieda

Signature: …………… Date: 21/4/2019

Sudan University of Science and Technology

College of Graduate Studies

COLLEGE OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY

Building a Classification Model to Identify the Effect of Apgar Score Factor on Cesarean Operation

بناء نموذج تصنيف لمعرفة تأثير معامل Apgar على الولادة القيصرية

A Thesis Submitted in Partial Fulfillment of the Requirements of M.Sc. in Computer Science

Prepared by:

Mozdalefa Abdalla Ibrahim

Supervised by:

D.Albaraa Abuobieda Mohammed

January 2019

الآية

وَقُلِ اعْمَلُوا فَسَيَرَى اللَّهُ عَمَلَكُمْ وَرَسُولُهُ وَالْمُؤْمِنُونَ ۖ وَسَتُرَدُّونَ إِلَىٰ عَالِمِ الْغَيْبِ وَالشَّهَادَةِ فَيُنَبِّئُكُم بِمَا كُنْتُمْ تَعْمَلُونَ

صدق الله العظيم

سورة التوبة الآية 105

Dedication

I dedicated this humble work to my family for their support and encouragement to accomplish this research.

Acknowledgement

I thank Allah for power and guidance that help me to accomplish this research. I would like to thank my family for their encouragement and support during all the time. Also, my thanks to all the neonatologistand Obstetrician doctors that made the data available to me .I express special thanks for my supervisor,Dr.AlbaraaAbuobieda Mohammed for his assistances and suggestions throughout this research. I wish to express my considerable gratitude to my friends, for their support.

Abstract

In today's world, gigantic amount of data is available in medical domain, science, industry, business and many other areas. The data can provide valuable information which can be used by management for making important decisions.

The research is focused on comparison of various classification algorithms using WEKA tool to obtain the highest accuracy. The aim of research building a classification model to identify the effect Apgar factor on cesarean operation.

Applied five data mining classification techniques were used in the research, it included j48, IBK, SMO, NB, and MLP algorithms. The results of performance classification algorithms nearly same, but the highest accuracy 96.8 % obtained IBK algorithm using 10 cross validation. Used statistical analysis for some attributes for interesting information, such as an approximately53.74% women that gave birth natural operation and 46.26% women took cesarean operation, and another used statistical analysis to the Apgar score based on data from the mother, newborn and medical interventions. it values normal obtained approximately 80%,low of approximately12.9%, very low of approximately 7.1%. Data mining techniques are useful for effect Apgar factor on cesarean operation to obtained optimal Apgar score.

مستخلص البحــث

فى عالم اليوم ,تتوفر كمية هائلة من البيانات فى المجال الطبى والعلوم والصناعة و الاعمال والعديد من المجالات الاخرى. يمكن أن توفر البيانات معلومات قيمه والتى يمكن استخدامها من قبل الإدارة لاتخاذ قرارات مهمة.ركز البحث على مقارنه خوارزميات التصنيف المحتلفه باستخدام اداهweka للحصول على أعلى دقة. الهدف من البحث بناء نموذج تصنيف لمعرفه تأثير معامل Apgar على الولادة القيصرية . تم تطبيق خمسه تقنيات التصنيف لتنقيب البيانات التى استخدمت فى هذا البحث وشملت خوارزميات j48,IBK,SMO,NB,and MLP نتيجه اداء الخوارزميات كانت متشابه ، ولكن اعلى دقه اتحصلت عليها خوارزمية IBK وكانت 96.8% باستخدام 10 cross validation . استخدم التحليل الاحصائى بعض العوامل للحصول على معلومات مثيره للاهتمام,وعلى ذلك تقريبا كانت العمليه الطبيعيه تمثل 53.47% بينما الولاده القيصرية كانت تمثل 46.26% .واستخدم تحليل احصائى اخر معامل Apgar بناء على معلومات الام والطفل والتدخلات الطبية ،وكانت القيم طبيعيه بدرجه 80% تقريبا ،بينما كانت منخفضه تقريبا 12.9% ,ومنخفضه جدا كانت تقريبا 7.1%. تقنيات تنقيب البيانات مفيده فى تاثير معامل Apgar على الولادة القيصرية بالحصول على Apgar مثالى.

Table of Contents

الآية	I
Dedication	II
Acknowledgement	III
Abstract	IV
مستخلص البحث	V
Table of Contents	VI
List of Tables	IX
Table of Figures	X
List of Abbreviations	XII
CHAPTER ONE: INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	1
1.3 Research Significance	1
1.4 Research Questions	2
1.5 Research Objectives	2
1.6 Research Methodology	2
1.7 Research Scope	2
1.8 Research Structure	2
CHAPTER TWO: LITERATURE REVIEW	3
2.1 Introduction	3
2.2 Cesarean Operation	3
2.3 Apgar Score	4
2.4 Weka Tool	5
2.5 Data Mining in Medical Domain	5
2.6 Data Mining (DM)	5
2.7 Data Mining Techniques	6
2.8 Classification Algorithms	6
2.8.1 Decision Tree (DT)	7
2.8.1.1 J48	7
2.8.2 K-Nearest Neighbor (KNN)	8
2.8.2.1 IBK	8
2.8.3 Support Vector Machines (SVM)	8
2.8.3.1 Sequential Minimal Optimization (SMO)	9
2.8.4 Naive Bayesian (NB)	9

2.8.5 Neural Networks (NN)	10
2.8.5.1 MultiLayer Perceptron (MLP)	10
2.9 Evaluation Mode	10
2.10 Performance Measures	11
2.11 Previous Studies and Related Works	12
2.12 Summary	15
CHAPTER THREE: METHODOLOGY	16
3.1 Introduction	16
3.2 Dataset Description	16
3.3 Data Preprocessing	18
3.4 Methodology Framework	19
3.5 Summary	20
CHAPTER FOUR: RESULT ANALYSIS AND DISCUSSION	21
4.1 Introduction	21
4.2 Statistical Analysis	21
4.3 Experiments Results	23
4.3.1 First Experiment	23
4.3.2 Second Experiment	23
4.3.3 Comparison Classification Algorithms with Test Mode	24
4.4 Summary	25
CHAPTER FIVE: CONCLUSION AND RECOMMENDATION	26
5.1 Introduction	26
5.2 CONCLUSION	26
5.3 RECOMMENDATION	26
References	27
Appendix	A
A1: weka 3.8 Tool	A
A2: Apgar Score	A
A3: Statistical Analysis	B
A4: Experiments	D

List of tables

Table 2.1 Comparison between Vaginal Birth and Cesarean Birth………….. 4

Table 2.2 Confusion Matrix…………………………………………………… 11

Table 2.1 Related Work……………………………………………………….. 14

Table 3.1 Description of the DAS……………………………………………. 17

Table 4.1 Accuracy of Dataset Using Cross Validation Mode……………… 23

Table 4.2 Accuracy of Dataset Using Percentage Split Mode………………… 24

Table A2.1 Apgar Score System……………………………………………… B

Table of Figures

Figure 2.1 Apgar Score System	4
Figure2.2 Data Mining Model and Tasks	6
Figure 2.3 Classification Algorithms	7
Figure 2.4 Decision Tree Structure	7
Figure 2.5 KNN	8
Figure 2.6 SVM	9
Figure 2.7 NN	10
Figure 3.1 Newborns.csv File	17
Figure 3.2 Newborns. arff File	19
Figure 3.3 Proposed Classification Model	20
Figure 4.1 Type of Births	21
Figure 4.2 Apgar Score	22
Figure 4.3 the Reason of Cesarean Operation	22
Figure 4.4 Comparison Test Mode with Classification Algorithms	24
Figure A 3.1 Statistical Analysis for Type of Birth	B
Figure A3.2 Statistical Analysis for Apgar Score	C
Figure A3.3 Statistical Analysis for Reason of Cesarean	C
Figure A4.1 Accuracy J48 Classifier Using Cross Validation Mode	D
Figure A4.2 Accuracy of IBK Classifiers. Using Cross Validation Mode	D
Figure A4.3 Detailed Accuracy by Class Using SMO Classifier. Figure A4.4 Detailed Accuracy by Class using Naïve Bayes classifier.	D D
Figure A4.5 Detailed Accuracy by Class Using Multilayer Perceptron Classifier.	D
Figure A4.6 J48 Classifier Using Percentage Split Mode	E
Figure A4.7 IBK Classifier Using Percentage Split Mode	E
Figure A4.8 SMO Classifier Using Percentage Split Mode	E
Figure A4.9 Naïve Bayes Classifier Using Percentage Split Mode	E
Figure A4.10 Multilayer Perceptron Classifier Using Percentage Split Mode	E

List of Abbreviations

- **DM** **D**ata **M**ining
- **MLM**achine **L**earning
- **DTD**ecision **T**ree
- **KNNK-N**earest **N**eighbor
- **IBK** **I**nstance-**B**ased method based on **K** neighbors
- **SVMS**upport **V**ector **M**achines
- **SMOS**equential **M**inimal **O**ptimization
- **NB** **N**aive **B**ayesian
- **NN** **N**eural **N**etworks
- **MLPM**ultiLayer**P**erceptron
- **TPT**rue **P**ositive
- **TN** **T**rue **N**egative
- **FPF**alse **P**ositive
- **FNF**alse **N**egative
- **FPR** **F**alse **P**ositive **R**ate
- **TPRT**rue **P**ositive **R**ate
- **ROC** **R**eceiver **O**perating **C**haracteristic
- **WEKAW**aikato **E**nvironment for **K**nowledge **A**nalysis
- **GPLG**eneral **P**ublic **L**icense
- **GUIG**raphical **U**ser **I**nterfaces
- **ARFFA**ttribute-**R**elation **F**ile **F**ormat
- **CSVC**omma **S**eparated **V**alues
- **URLU**niform **R**esource **L**ocator
- **SQLS**tructured **Q**uery **L**anguage
- **JDBCJ**ava **D**ata**B**ase**C**onnectivity

CHAPTER ONE

INTRODUCTION

CHAPTER ONE: INTRODUCTION

1.1 Background

The childbirth experience has always represented a very important event in women's lives, a unique and special moment, marked by the transformation of the woman in her new role, that of being a mother [1].

Childbirth is the period of increased risk of mortality for mothers and their babies. An estimated 42% of the world's 535.900 annual maternal deaths are intrapartum-related, these deaths are closely linked to the deaths of 1.02 million babies during labor and 904.000 intrapartum related ("birth asphyxia") neonatal deaths [2].

Every day, databases of enormous size are collected. The analysis of these data may help extract interesting and useful information, by using data mining techniques [3].

Data Mining (DM) and Machine Learning (ML) make the patterns extraction more convenient, Data can be classified and prediction can be made using variety of ML algorithms. ML facilitates tools, methods and techniques that can provide diagnosis and analytical facilities in the number of medical domains. A reasonable amount of work exists in the domain of biomedical, biomedicine and diagnosis of diseases using different machine learning techniques. However, there are many areas in the medical domain that are not yet explored using ML/DM techniques [4].

1.2Problem Statement

Childbirth increased risk of mortality for mothers and their babies [2]. A cesareans rate of cesarean delivery has increased dramatically over the past decade .There is a need to find its causes and control the rate of C-section [4]. According, increased a cesareans rate based on some factors to affect mothers and newborns health. This researchutilizes the data mining tools in order to correlate between Apgar score with cesarean operation.

1.3Research Significance

The importance of this research helps the neonatologist in makingdecision on keep the newborns health (Apgar score).

1.4 Research Questions

Using formulate some questions, such as:

- What are the role of building a classification model?
- What are the risk factors of low Apgar score?
- Is there a relationship between caesarean operation and Apgar score?

1.5 Research Objectives

The objectives of this research are as follows:

- To apply classification model based on the identified factors.
- To build model to identified the effect Apgar score on a caesarean operation.
- To specify the best accuracy algorithm by using building a classification model.
- To obtain optimal Apgar score

1.6 Research Methodology

The research collected dataset and it was preprocessing and building a classification model to specify the best accuracy algorithm.

1.7 Research Scope

The research focuses on build classification model for supporting hospitals by extracting knowledge from births records data, it was obtained from medical reports.

1.8 Research Structure

This research contains five chapters. Chapter one consists of background, research problem, significance, objective, scope, and research structure. Chapter two contains classification data mining techniques and evaluation model, using weka tool, and definition of cesarean operation and Apgar score in data mining medical domain, finally previews studies and related work. Chapter three cantinas two sections, one of the section include dataset description and dataset preprocessing, and the second section related by framework or software tool used weka tool with classification algorithms. Chapter four cantinas three sections, one the section regarded dataset loaded in weka and how preprocessing in weka. Second section regarded all experiments description and applied classification algorithms with weka .third section challenges in the research faced, and the last chapter explains the research conclusion and recommendation

CHAPTER TWO

LITERATURE REVIEW

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter contains classification data mining techniques and evaluation model, using weka tool. Also, definition of cesarean operation and Apgar score in data mining medical domain, finally previews studies and related work

2.2 Cesarean Operation

Cesarean section, also known as C-section, or cesarean delivery, is the use of surgery to deliver babies. A caesarean section is often necessary when a vaginal delivery would put the baby or mother at risk. This may include obstructed labor, twin pregnancy, high blood pressure in the mother, breech birth, or problems with the placenta or umbilical cord. A caesarean delivery may be performed based upon the shape of the mother's pelvis or history of a previous C-section. The World Health Organization recommends that Caesarean section be performed only when medically necessary. Some C-sections are performed without a medical reason, upon request by someone, usually the mother. A C-section typically takes 45 minutes to an hour [1].found two kind of operations vaginal birth and cesarean birth, as seen in table 2.1 comparison between them.

Comparison	**Comparison type**	**Cesarean section**	**Natural(vaginal)section**
Risks for mothers around the time of birth	- maternal death - blood clots and stroke - injuries from surgery -longer time in hospital - infection - pain -depression	Low Low MODERATE Very high High Very high Some different in studies	Very low Very low Moderate Low Low Very high Some different in studies
Risks for babies around	-injuries	High	Moderate

the time of birth	-respiratory problems	moderate	low
Ongoing risks for mothers	-Pelvic pain -painful vaginal area	Some different in studies Very low	low very high
Ongoing risks for babies	Asthma	high	Low
Future reproductive risks for mothers	- maternal death -ectopic pregnancy -placenta abruption	Very low Moderate moderate	Very low Some different in studies Some different in studies
Risks for babies in future pregnancies	-still birth or death shortly after birth -malformation	Moderate -	- -

Table 2.1 Comparison between VaginalBirth and Cesarean Birth

In the table 2.1 found that vaginal birth involves many fewer risks than either cesarean section, a spontaneous vaginal birth is likely to be the safest way to give birth.

2.3 Apgar Score

Newborn infants should assess immediately after delivery. The Apgar score is a simple and effective Method for assessing of the neonatal health in the immediate period after birth. The Apgar score includes five components, Appearance, Pulse, Grimace, Activity and Respiration, each of the five clinical findings is assessed a value of 0 to 2.This score is the sum of the five components [5]. The measuring points are in the first minute of birth repeat in 5 minutes after birth [6].In this research using Apgar score at first minute of age was categorized into three ordinal groups; low (Apgar 0-3), intermediate (Apgar 4-6) and normal (Apgar 7-10).As seen in Figure 2.1 Apgar Score System.

APGAR SCORING SYSTEM

	0 Points	1 Point	2 Points	Points totaled
Activity (muscle tone)	Absent	Arms and legs flexed	Active movement	
Pulse	Absent	Below 100 bpm	Over 100 bpm	
Grimace (reflex irritability)	Flaccid	Some flexion of Extremities	Active motion (sneeze, cough, pull away)	
Appearance (skin color)	Blue, pale	Body pink, Extremities blue	Completely pink	
Respiration	Absent	Slow, irregular	Vigorous cry	

Severely depressed	0-3
Moderately depressed	4-6
Excellent condition	7-10

Figure 2.1 Apgar ScoreSystem

2.4 WekaTool

Weka (Waikato Environment for Knowledge Analysis) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka is free software available under the GNU (General Public License). The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces (GUI) for easy access to this functionality. Weka is a collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost any platform. The algorithms can either be applied directly to a dataset or called from your own Java code [7], some details to weka tool, look in appendix A1.

2.5 Data Mining in Medical Domain

The medical domain is one of the most important domains to apply data mining techniques .data mining apply successfully in the medical field [3]. It make the patterns extraction more convenient and can provide diagnosis and analytical facilities in the number of medical domains. The data mining help decision makers to make the right decision.

2.6 Data Mining (DM)

Data mining is a Process of extracting knowledge “mining” from large amount of data. Also it is the task of discovering interesting patterns from large amount of data. DM can be defined by many terms, all carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data (pattern) analysis, data archaeology, and data dredging. Data mining is a step in the knowledge discovery process. Knowledge Discovery from Data (KDD) a Process consists of an iterative sequence of the following steps [8]:

1. Data cleaning (to remove noise and inconsistent data).

2. Data integration (where multiple data sources may be combined).

3. Data selection (where data relevant to the analysis task are retrieved from the database).

4. Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance).

5. Data mining (an essential process where intelligent methods are applied in order to extract data patterns).

6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interesting measures.

7. Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user).

2.7 DataMining Techniques

Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks. In general, data mining tasks can be classified into two categories: descriptive and predictive. Descriptive mining tasks characterize the general properties of the data in the database such as Clustering, Association rule, etc. Predictive mining tasks perform inference on the current data in order to make predictions. Such as Classification, Regression, Prediction, etc. [8]

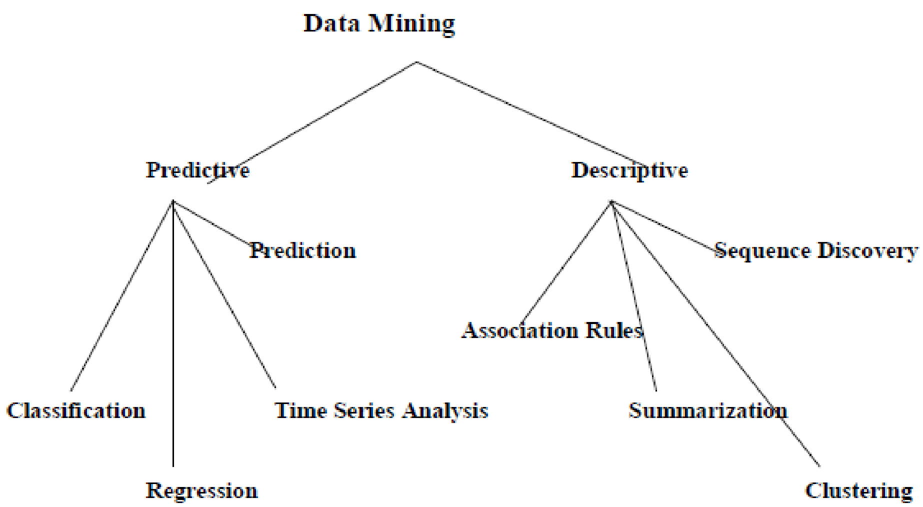


Figure2.2 Data Mining Model and Tasks

2.8Classification Algorithms

Classification is possibly the most frequently used data mining technique. Classification is the process of finding a set of models that describe and differentiate data classes and concepts, for the purpose of being able to use the model to predict the class whose label is unknown [9],and maximize the predictive accuracy obtained by the classification model. Classification task can be seen as a supervised technique where each instance belongs to a class. There are several model techniques are used for classification some of them are [10] Decision Tree, K-Nearest Neighbor, Support Vector Machines, Naive Bayesian and Neural Networks.

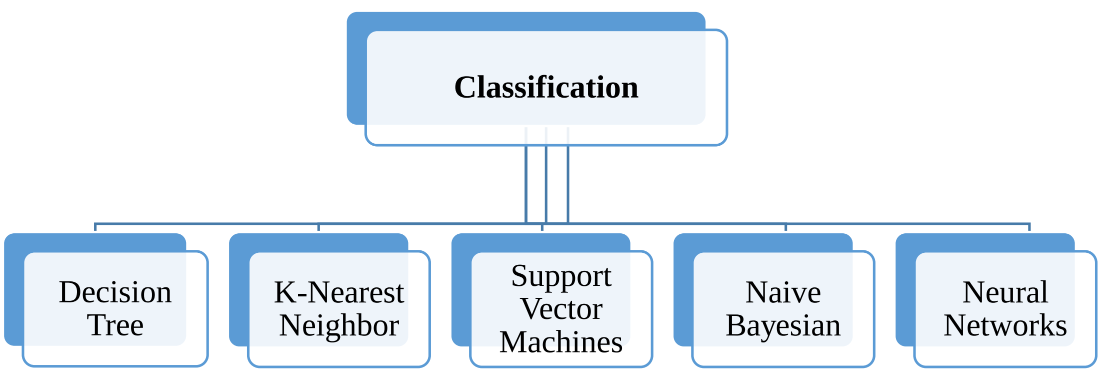


Figure 2.3Classification Algorithms

2.8.1 Decision Tree (DT)

Decision trees are a way of representing a sequence of rules that lead to a class or value. Decision Tree is a flowchart like tree structure, the decision tree consists of three fundamentals, root node, internal node and leaf node. Top most fundamental is the root node. Leaf node is the terminal fundamental of the structure and the nodes in between is called the internal node. Each internal node denotes test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label. Various decision tree algorithms are used in classification, like J48 [9].

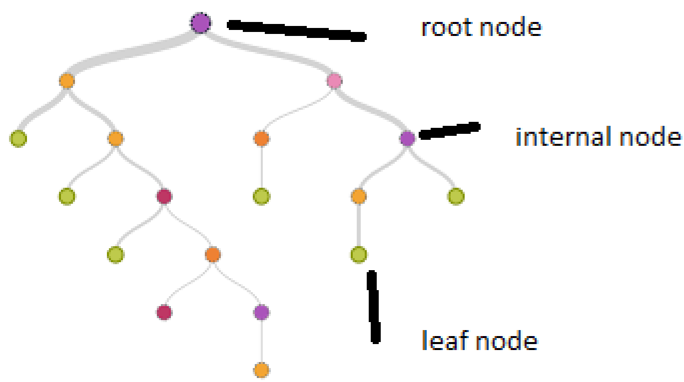


Figure 2.4 Decision Tree Structure

2.8.1.1 J48

J48 are the improved versions of C4.5 algorithms or can be called as optimized implementation of the C4.5. The output of J48 is the Decision tree [11].A predictive machine-learning model which decide the target value of a new sample based on different attribute values of the available data is J48 decision tree. the different attributes denote by the internal nodes of a decision tree, the branches between the nodes tell us the possible values that these attributes can have in the experimental samples, while the terminal nodes tell us the final value of the dependent variable[9].

2.8.2 K-Nearest Neighbor (KNN)

This classifiers are based on learning by training samples. Each sample represents a point in an n-dimensional space. All training samples are stored in an n-dimensional pattern space. When given an unknown sample, a k-nearest neighbor classifier searches the pattern space for the k training samples that are closest to the unknown sample. "Closeness" is defined in terms of

Euclidean distance, where the Euclidean distance between two points, X=(x1, x2,…,xn) and Y=(y1,y2,….,yn) is denoted by d(X, Y).

$d\ (x, y) = \sum_{i=1}^{n} \sqrt{xi^2 + yi^2}$

Nearest neighbor classifiers assign equal weight to each attribute. Nearest neighbor classifiers can also be used for prediction, that is, to return a real-valued prediction for a given unknown sample [10] Instance-based learning algorithms are lazy-learning algorithms (Mitchell, 1997), as they delay the induction or generalization process until classification is performed [12]

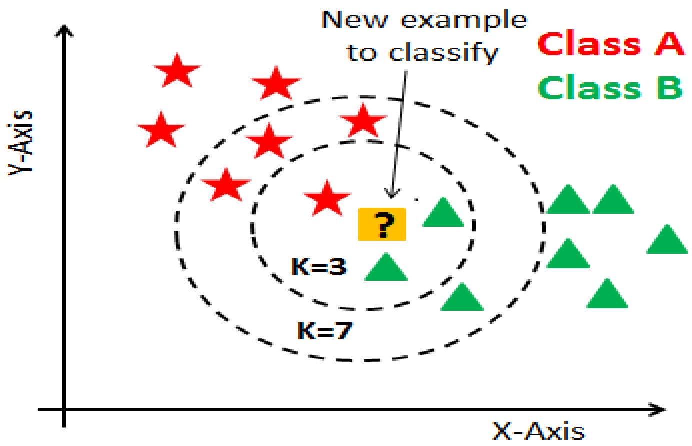


Figure 2.5 KNN

2.8.2.1 IBK

The abbreviation IBK means that this is an Instance-Based method based on k neighbors. The default value of k is 1. So, build a 1-NN model.

2.8.3 Support Vector Machines (SVM)

SVM is a very effective method for regression, classification and general pattern recognition. It is considered a good classifier because of its high generalization performance without the need to add a priori knowledge, even when the dimension of the input space is very high. For a linearly separable dataset, a linear classification function corresponds to a separating hyper plane, the hyper plane that maximize the margin between two classes. SVMs were initially developed for binary classification but it could be efficiently extended for multiclass problems [10].

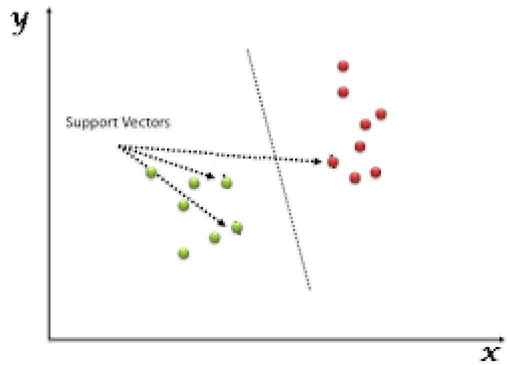


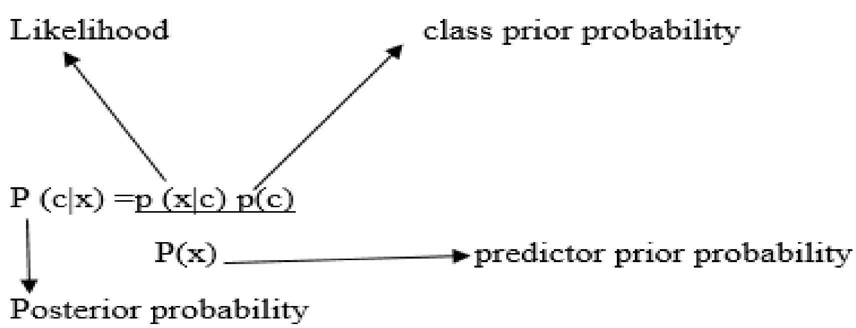
Figure 2.6 SVM

2.8.3.1 Sequential Minimal Optimization (SMO)

The Weka software implements John Platt's Sequential Minimal Optimization (SMO) algorithm for training a support vector classifier.

2.8.4 Naive Bayesian (NB)

Bayesian classifiers are statistical classifiers. They can predict class membership based on probabilities. The Naive Bayes Classifier technique is particularly suited when the dimensionality of the inputs is high. Naive Bayes can often outperform more sophisticated classification methods [10].Bayes theorem provides away of calculating the posterior probability p(c/x), from p(c),p(x) ,and p(x/c).NB classifier assumes that the effect of the value of a predictor (x)on given class(c)is independent of the values of other predictors



$$P(c|x) = p(x_1 | c) \times p(x_2 | c) \times \ldots \times p(x_n | c) \times p(c)$$

- P(c/x): is the posterior probability of class (target) given predictor (attribute).
- P(c): is the prior probability of class.
- P(x/c): is the likelihood which is the probability of predictor given class.

- P(x): is the prior probability of predictor.

2.8.5 Neural Networks (NN)

Neural Network used gradient descent method based on biological nervous system having multiple interrelated processing elements. These elements are known as neurons. Rules are extracted from the trained Neural Network to Improve interoperability of the learned network. To solve a particular problem NN used neurons which are organized processing elements [10].

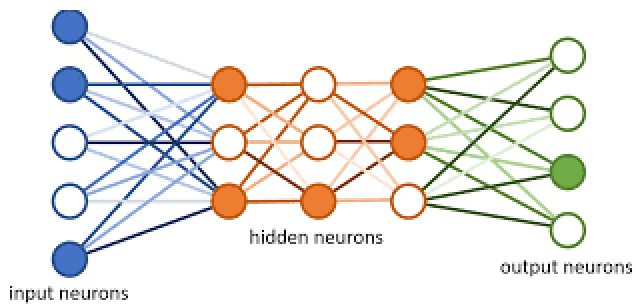


Figure 2.7 NN

Neural Network is used for classification and pattern recognition. An NN changes its structure and adjusts its weight in order to minimize the error. Adjustment of weight is based on the information that flows internally and externally through network during learning phase. In NN multiclass, problem may be addressed by using multilayer feed forward technique, in which Neurons have been employed in the output layer rather using one neuron

2.8.5.1 MultiLayerPerceptron (MLP)

Multilayer Perceptron technique was introduced by Werbos in 1974 and Rumelhart, McClelland, Hinton in 1986 also named feed forward networks. MLP is mainly used to solve non-linear problem with good quality solution. It is suitable for regression and classification In MLP the problem is converted into finite directed acyclic graph which contain n number of inputs, hidden and output nodes. Here the parameters are measured based on Weightage of each node and unlabeled patterns are estimated by using hidden layer concept. Here minimization approach is used for adjusting the weight in the hidden layer [12].

2.9 Evaluation Model

To evaluate the selected tool using the dataset, two test modes are used

1-The K-fold cross validation mode

The K-fold CV refers testing procedure where the database is randomly divided into K disjoint blocks of objects , then the data mining algorithm is trained using k- 1 blocks and the remaining blocks is used to test the Performance of the algorithm. This process is repeated k times. At the end, the recorded measures are averaged. It is common to choose k=10 or any other size depending on the size of the original dataset [13].

2- Percentage split mode

In percentage split (holdout method), the database is randomly split in to two disjoint datasets. The first set, which the data mining system tries to extract knowledge from called training set. The extracted knowledge may be tested against the second set which is called test set, it is common to randomly split a data set under the mining task in to two parts. It is common to have 66% of the objects of the original database as a training set and the rest of objects as a test set. Once the tests is carried out using the selected datasets, then using the available classification and test modes ,results are collected and an overall comparison is conducted[13].

2.10 Performance Measures

Confusion matrix is a matrix representation for the classification, The Confusion Matrix is a useful tool for analyzing how well your classifier can recognize tuples of different classes [13].It is shown in Table 2.2.

	P‘(predicted)	N‘(predicted)
P(actual)	True Positive	False Negative
N(actual)	False Positive	True Negative

Table 2.2 Confusion Matrix

There are some parameters on the basis of which we can evaluated the performance of the classifiers such as[9]depend on Confusion matrix.

- N – Total number of classified instances.

- True Positive (TP) – correctly predicted of positive classes.
- True Negative (TN) – correctly predicted of negative classes.
- False positive (FP) – wrongly predicted as positive classes.
- False Negative (FN) – total wrongly predicted as negative classes.
- False Positive Rate (FPR) – negatives in correctly classified/total negatives.
- True Positive Rate (TPR) – positives correctly classified/total positives.
- Accuracy (A): It shows the proportion of the total number of instance predictions which are correctly predicted.

  $$\text{Accuracy}=\frac{TP+TN}{N} \quad N=TP+TN+FP+FN.$$

2.11 Previous Studies and Related Works

In the recent years researchers of classification algorithms have shown interest in the medical field. These algorithms have been used for comparison, prediction and predictive analysis. Discuss a few studies exist on the use cesarean operation or Apgar score.

The study[3] on births in Bega Obstetrics and Gynecology Clinique, in Timisoara,(Moreira, October 2017) Romania, was presented by Robu and Holban in 2010. It analyzed 2086 births based on 16features such as (Month of birth, Mother's age, City of residence, Location, Gesta, Para, The number of gestation weeks ,Presentation , Apgar score , baby's gender, baby's weight , Videx,type of birth, reason of cesarean, Episiotomy, EMP).Data were analyzed both statistically and using some classification techniques from weka tool. The statistical analysis revealed interesting information, the classification models were built using Naive Bayes, J48, k-Nearest Neighbor, Random Forest, Support Vector Machines, AdaBoost, LogitBoost, JRipp, REPTree, Simple Cart algorithms. They were tested through cross validation with 10 folds. The best model LogitBoost algorithm, it allows an estimation with an 80% accuracy of the interval of the newborn's Apgar score based on data regarding the mother, baby and medical interventions.

The study[14] presents the GISSA framework that contains a series of software services with the purpose of assisting the process of decision making in health systems of government agencies. It analyzed 124. 876 births in the state of Cear´a northeastern region of Brazil in year 2013 .the data regarded information about mother and child with sixteen attribute features (Age, Marital

Status, Schooling, Localization, Number of live births, Number of children born dead, Gestation week, Pregnancy, Birth, Gender, Weight, Consultations, Apgar1, Apgar5, Anomaly, and color) and used eight classification algorithms (ID3, RF, BN, NB, KNN, Voted Perceptron, MLP, and PART). The work performed the cross Validation method to validate the proposed models after several experiments, the Naïve Bayes (NB) classifier to calculate the probability of the occurring of an infant death. Obtained 98.24% accuracy of live birth.

The study [2] Presents a cross-sectional study was conducted on singleton 261 live births from March to May, 2013. Data was collected from mother/newborn index using a structured and pre-tested questionnaire with some characteristics of mothers who gave birth in Gondar University Referral Hospital like (age, marital status, Occupation, Educational status, Residence, Household Income per month). It was then cleaned, coded and entered using EPI INFO version 3.4.3, then analyzed with IBM SPSS statistics versions 20.0 with some variables like Duration of labor, Birth weight, 5th Minute APGAR Score(Low (< 7)equality13.8%). Logistic regression was used to identify significant variables with low 5th minute Apgar score. The proportion of low 5th minute Apgar score in this study was 13.8%. Factors that were significantly associated with low 5th minute Apgar score were: non-vertex fetal presentation, prolonged labor, presence of me conium stained liquor, induced/augmented labor and low birth weight.

The study [4] have evaluated different machine learning techniques for birth classification (cesarean or normal). Data collected from multiple sources (interviewing patients, doctors and the hospital sources).includes 15 hospitals of Sargodha, Pakistan between 2nd February 2010 to 28th February 2010, by interviewing the medical experts in this area and the patients. The authors have identified about 50 factors that can influence the type of birth. These factors include Pre pregnancy factors like maternal age, body weight, education, drinking routine, diabetes, hypertension and various other factors, some factors are identified during pregnancy like HIV, uterine rupture, blood sugar, abnormal presentation, there are also some social factors like low education, dieting, fear of pain etc. The data filtered out the attributes that have no influence on birth type. A birth classification model is built using decision tree (j48) and artificial neural networks (MLP). Applied weka tool and it can classify the births into normal and cesarean with an average accuracy, precision and recall of 80%, 85% and 84% respectively. Association rule mining is used to extract disease patterns from the collected data.

Study	Objective	Techniques	Results	Open Issues
(Robu, R. (2015)).	Optimal Apgar score	Weka and used ten classification algorithms	The study used ten classification algorithms, The best model LogitBoost algorithm obtained accuracy 80%	Several of algorithm probable it optimal or best Apgar score like AdaBoost79.91% and JRipp 79.62% of accuracy
(Moreira, M. W. L. (October 2017)).	improve the healthcare for pregnant women as well as their newborns.	GISSA framework and used eight classification algorithms (ID3, RF, BN, NB, KNN, Voted Perceptron, MLP, and PART).	The best accuracy Naïve Bayes 98.24% And prec, recall, F-meas, and auroc obtained 0.294, 0.607 0.396 . and 0.921	In the study used 16 features to known infant live or death with some algorithms but that features depended on several information about mothers likes diseases, drug, learning... etc.
(Gudayu, T. W. (march, 2017)).	To assess proportion and factors associated with low 5th minute Apgar Apgar score among singleton newborn babies in Gondar University referral hospital; North West Ethiopia.	EPI INFO version 3.4.3, IBM SPSS statistics versions 20.0 and Logistic regression model	The proportion of low 5th minutelowApgar score<7 in this study was 13.8%. By using Logistic regression model	This study did not consider some potential risk factors for low Apgar score such as placental factors, multiple pregnancy, and used a lot of methods property, using one tool like weka
(Ayesha Sana, S. R., and Javed Ferzund. (October 2012)).	This study aims atfinding the reasons for an increased rate of cesarean section	Decision tree, ANN and Association Rule	The study used j48 and MLP to obtained Average accuracy, precision and	Possible increase factors that influence the type of birth and used other techniques on the medical data.

	and developing a prediction model for child birth		recall of 80%, 85% and 84% respectively.	

Table 2.2 Related Work

2.12 Summary

The chapter explained cesarean operation and Apgar score. Also discussed data mining definition and techniques by it discussed all the classification algorithms were used in the research .finally discussed Previous studies and related works.

CHAPTER THREE

METHODOLOGY

CHAPTER THREE: METHODOLOGY

3.1 Introduction

This chapter cantinas two sections, one of the section include dataset description and dataset preprocessing, and the second section related by framework or software tool used weka tool with classification algorithms. Finally summary.

3.2 Dataset Description

The **D**ataset of **A**pgar **S**core (DAS) was collected from medical reports.it included 730 records of births in 2018.The DAS is related mothers and their babies, it took some attributes relative by cesarean operation.

DAS for each birth:

Attribute Name	**Attribute Description**
the number of gestation weeks	Started life in baby from24 weeks and above
the reason of cesarean operation	a- Narrow pelvis b- Fetal water is low c- Previous caesarean section d- Weak pulse f- The mother had diabetes h- Fetal is sitting nf- not found(natural operation)
The baby's weight	Started from 500g, it usual measure weight by kilo Gram (KG).
type of birth	if the birth was natural or through a cesarean operation
Apgar score	Immediately after birth, even in the first 60 second after expulsion, in the delivery room, an assessment of the newborn's health state is made, evaluating the vital functions and its capacity to adapt to the extra uterine environment. Simultaneously with providing the first nursing measures, the

	neonatologist will write down the clinical state and behavior of the newborn, quantifying the vital functions with the aid of the Apgar. The Apgar score has values between 0 and 10.assumed nominal to range values depend on Apgar score system. 0-3=very low 4-6=low 7-10=normal

Table 3.1 Description of the DAS.

In seen figure 3.1, explained dataset before preprocessing in excel sheet. The dataset included five attributes, one attribute (weeks of gestation) gave numeric values from 24 to 43 weeks, two attribute (reason of cesarean) gave letters value(a,b,c,d.,h,f,nf).three attribute (weight of baby) took numeric values, four attribute (type of birth) gave letters values(n,c), final attribute (Apgar score) gave nominal values(normal,low,very low).

weeks_of_gestation	reason_of_cesarean	weight_of_babyes	type_of_birth	apgar_score
39	nf	3.1	n	normal
37	nf	2.8	n	normal
39	b	3.1	c	low
37	nf	2.8	n	normal
39	a	2.9	c	very low
38	c	3	c	very low
40	nf	3.3	n	normal
38	nf	3.8	n	normal
39	d	3.7	c	low
42	nf	2.9	n	normal
39	nf	2.9	n	normal
40	a	2.95	c	low
30	nf	1	n	very low
37	nf	2.7	n	normal
38	c	3	c	low
38	nf	3.8	n	normal
39	nf	3.2	n	normal
41	nf	3	n	normal
34	nf	2	n	normal
37	nf	2.7	n	normal
38	nf	3.2	n	normal

Figure 3.1 Newborns.csv File

3.3 Data Preprocessing

Data preprocessing is an important step in the knowledge discovery process, because quality decisions must be based on quality data. Detecting data anomalies, rectifying them early, and reducing the data to be analyzed can lead to huge payoffs for decision making [8].

Preprocessing step includes:

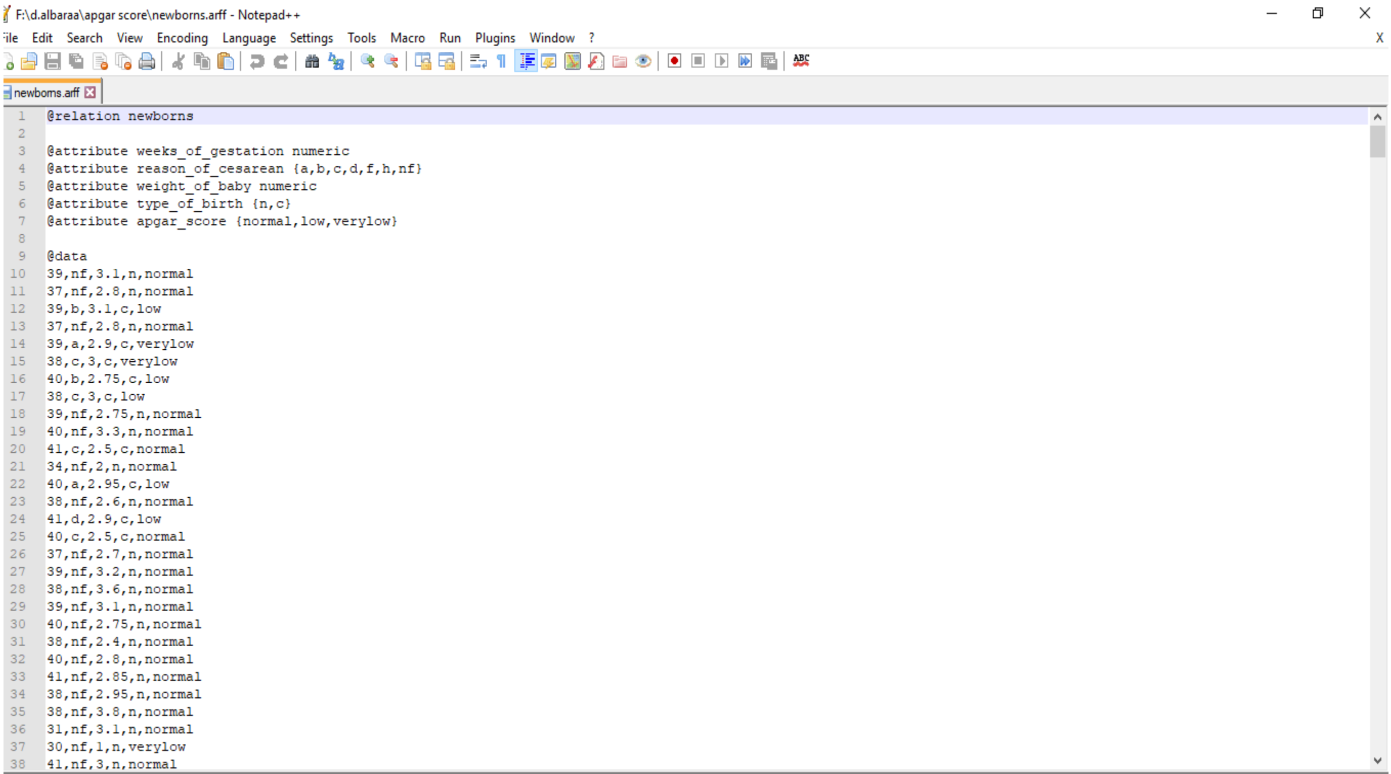
Data cleaning: applied to remove noise and correct inconsistencies in the data. The dataset contains 730 records of births but used 722 records for analysis because have some missing value.

- Some information unimportant and also a privacy reasons for the study were eliminated, such as the serial number, the name of the mother, mother's age, the city of residence, etc.
- There were 8 instances removed because of missing values in attribute reason of cesarean operation a value is blank and so removed.
- The weight registered for four babies was of 285, 235, 26, and 295.that is mistake values, we added comma after first number for became 2.85, 2.35, 2.6 and 2.95 KG.

Data integration: merges data from multiple sources into a coherent data store, such as a data warehouse. This dataset was taken from a single source; accordingly we do not need to apply this step

Data transformations: it is to transform the data in given format to required format for data mining. Such as normalization, aggregation, etc. Need to transform specific columns so that they would be more suitable for analyze the result. Like Apgar score taken values from 0 to 10 transformed 0-3(very low), 4-6(low), and 7-10(normal).changed reason of cesarean from string to letters such as Narrow pelvis to A characters, etc. and we changed type of birth from string to character like natural to N, cesarean to C. data transformations may be improve the accuracy and efficiency of mining algorithms involving distance measurements.

Data reduction: reduce the data size by aggregating, eliminating redundant features, or clustering for instance. The dataset was reduced from 730to 722 records, and reduced any irrelevant from attributes was became five attributes for increase the performance.



```
@relation newborns

@attribute weeks_of_gestation numeric
@attribute reason_of_cesarean {a,b,c,d,f,h,nf}
@attribute weight_of_baby numeric
@attribute type_of_birth {n,c}
@attribute apgar_score {normal,low,verylow}

@data
39,nf,3.1,n,normal
37,nf,2.8,n,normal
39,b,3.1,c,low
37,nf,2.8,n,normal
39,a,2.9,c,verylow
38,c,3,c,verylow
40,b,2.75,c,low
38,c,3,c,low
39,nf,2.75,n,normal
40,nf,3.3,n,normal
41,c,2.5,c,normal
34,nf,2,n,normal
40,a,2.95,c,low
38,nf,2.6,n,normal
41,d,2.9,c,low
40,c,2.5,c,normal
37,nf,2.7,n,normal
39,nf,3.2,n,normal
38,nf,3.6,n,normal
39,nf,3.1,n,normal
40,nf,2.75,n,normal
38,nf,2.4,n,normal
40,nf,2.8,n,normal
41,nf,2.85,n,normal
38,nf,2.95,n,normal
38,nf,3.8,n,normal
31,nf,3.1,n,normal
30,nf,1,n,verylow
41,nf,3,n,normal
```

Figure 3.2 Newborns. arff File

In seen figure 3.2 the data preprocessing after implementation weka tool.

3.4 Methodology Framework

The Proposed classification model is includes data of births records (DAS) is inputs of framework, classification algorithms with weka tool to build model based on identified factors depend on dataset.as mention in figure 3.1 explained all algorithms used in the research .The result is output of framework evaluation model represented in accuracy algorithm, the calculate accuracy previous chapter in section 2.10.

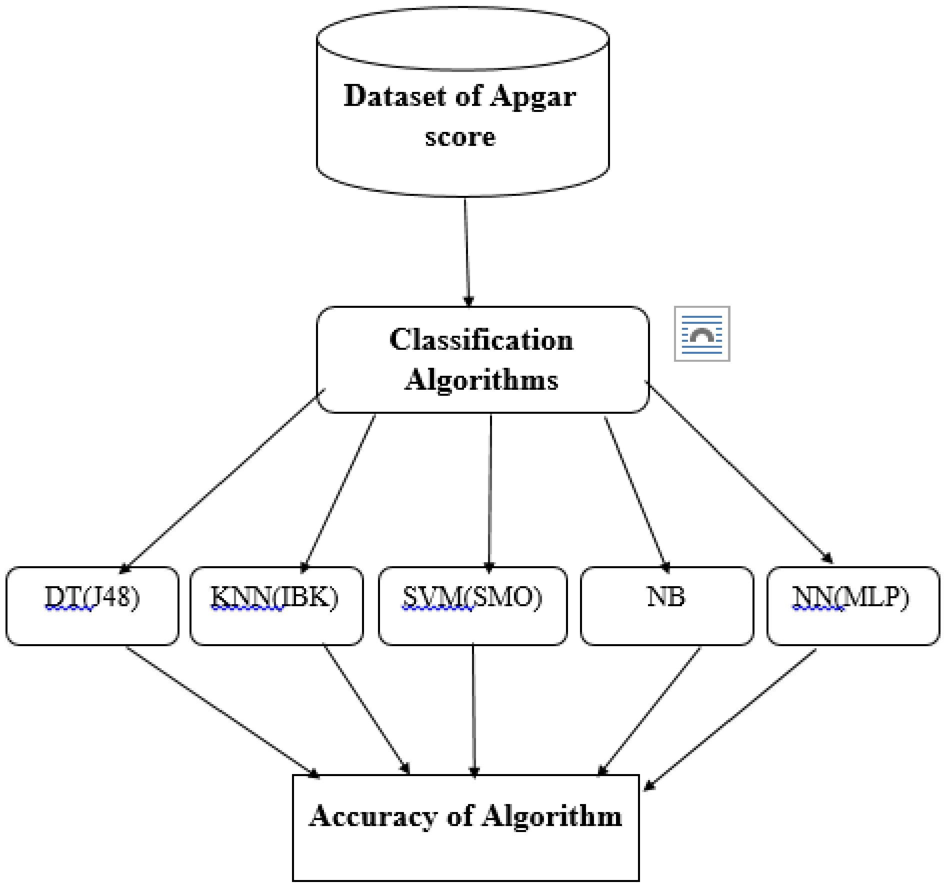


Figure 3.3Proposed Classification Model

3.5 Summary

The chapter explained dataset description and preprocessing, also discussed how work proposed classification model with weka tool in methodology framework.

CHAPTER FOUR

RESULT ANALYSIS AND DISCUSSION

CHAPTER FOUR: RESULT ANALYSIS AND DISCUSSION

4.1 Introduction

The chapter explained statistical analysis for useful and interesting information. Also Experiments description discussed all results experiments in used the research.

4.2 Statistical Analysis

Statistical analysis for useful and interesting information. In figure 4.1 shown the type of birth, the natural operation obtained 388 birthsand cesarean operation 334 births in 2018 approximately53.74% women that gave birth natural operation and 46.26% women took cesarean operation that is high range of cesarean operation.

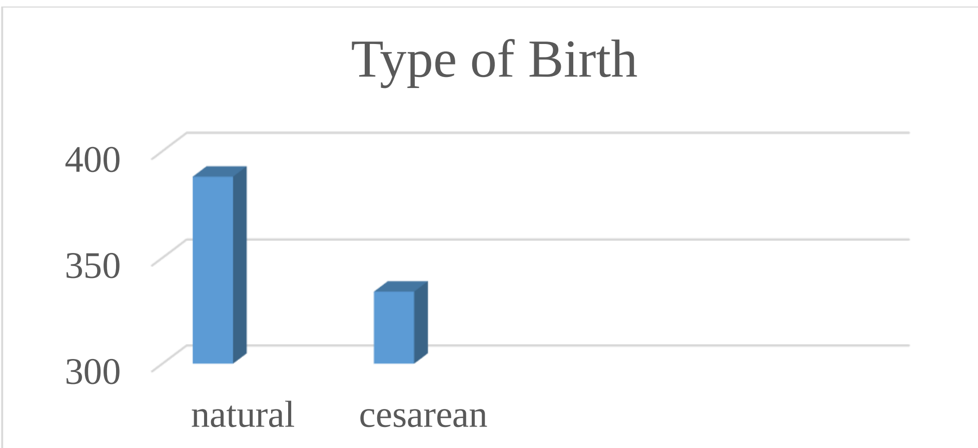


Figure 4.1 Type of Births

in figure 4.2Statistical analysis for Apgar score, shown the values of Apgar score obtained normal 578 births of approximately80%, low 93 births of approximately 12.9%, it sometime need medical interventions accordingly status of baby, and very low 51 births of approximately7.1%, it to need medical interventions by nursing and neonatologist. The risk of newborns while very low Apgar score.So, high attention for newborns babies health

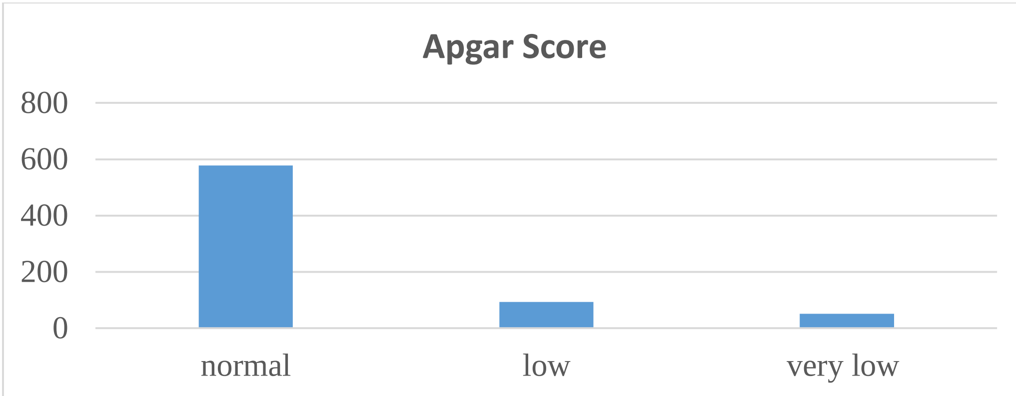


Figure 4.2 Apgar Score

In Figure 4.3 Statistical analysis for reason of cesarean operation, explain a=(Narrow pelvis)obtained on 74 births ,b=(Fetal water is low)obtained on 67 births,c=(Previous caesarean section)obtained on 151 births,d=(Weak pulse)obtained on37 births ,f=(The mother had diabetes)obtained on 4 births ,h=(Fetal is sitting) obtained on 1 birth,nf=(not found(natural operation))obtained on 388 births .Observed in natural birth obtained 388 births and normal Apgar score. The remained cesarean birth 334 births between normal, low, and very low at Apgar score. So found relationship between reasons cesarean operation and Apgarscore. Whenever increased reasons cesarean operation and increased low or very low Apgar score. That is explained effect of cesarean operation on Apgar score.

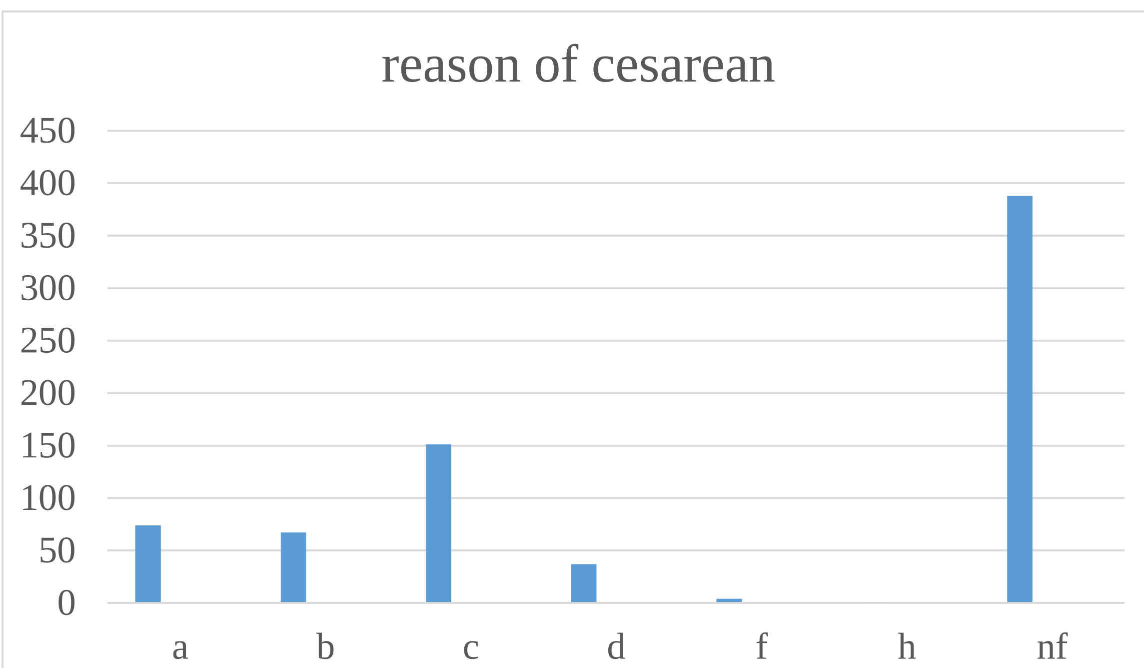


Figure 4.3 the Reason of Cesarean Operation

4.3 Experiments Results

In this experiments contains two experiments, first experiment applying all classification algorithms on full dataset and used 10 fold cross validation to evaluated test set using wekatool. Second experiment applying all classification algorithms on percentage split mode, using 60% of dataset (training set) and 40% from data to evaluated test set using weka tool.

4.3.1 First Experiment

In these experimentapplyingall classification algorithms on full dataset and used 10 fold cross validation to evaluated test set using wekatool.

Algorithms	Time taken to build model	Accuracy
J48	0.03 seconds	95.9 %
IBK	0 seconds	96.8 %
MLP	3.03 seconds	92.1 %
SMO	0.33 seconds	83.6 %
NB	0.01 seconds	81.3 %

Table 4.1 Accuracy of Dataset Using Cross Validation Mode

In table 4.1 Comparing accuracy algorithms with Time build model using cross validation mode, the best IBK algorithm obtained 0 second to build model and 96.8% accuracy.

4.3.2 Second Experiment

In these experiment applying all classification algorithms on percentage split mode, using 60% of dataset (training set) and 40% from data to evaluated test set. the results obtained by the classification algorithms on these data are presented in table 4.2.The best classification model obtained has a 95.2% accuracy and was built using the IBKalgorithm but took time in built

0.05seconds compared reminded j48 and Multilayer perceptron algorithm took 0 second and obtained 39.7% j48 and 90% MLP accuracy.

Algorithm	Time Taken to Build and Test Model	Accuracy
J48	0 seconds	93.7%
IBK	0.05 seconds	95.2%
SMO	0.05seconds	82.6%
Naïve Bayes	0.05 seconds	78.9%
Multilayer perceptron	0 seconds	90.0%

Table 4.2 Accuracy of Dataset Using Percentage Split Mode

4.3.3Comparison Classification Algorithms with Test Mode

In figure 4.4 compared algorithms and the best accuracy using test mode, from observed tables4.1, 4.2 found the best algorithm IBK obtained 96.8% using K-fold cross validation mode. That is explained the role in building classification model for facilitating of decision making process in effect of cesarean operationonApgar score factor to obtained optimal Apgar score.

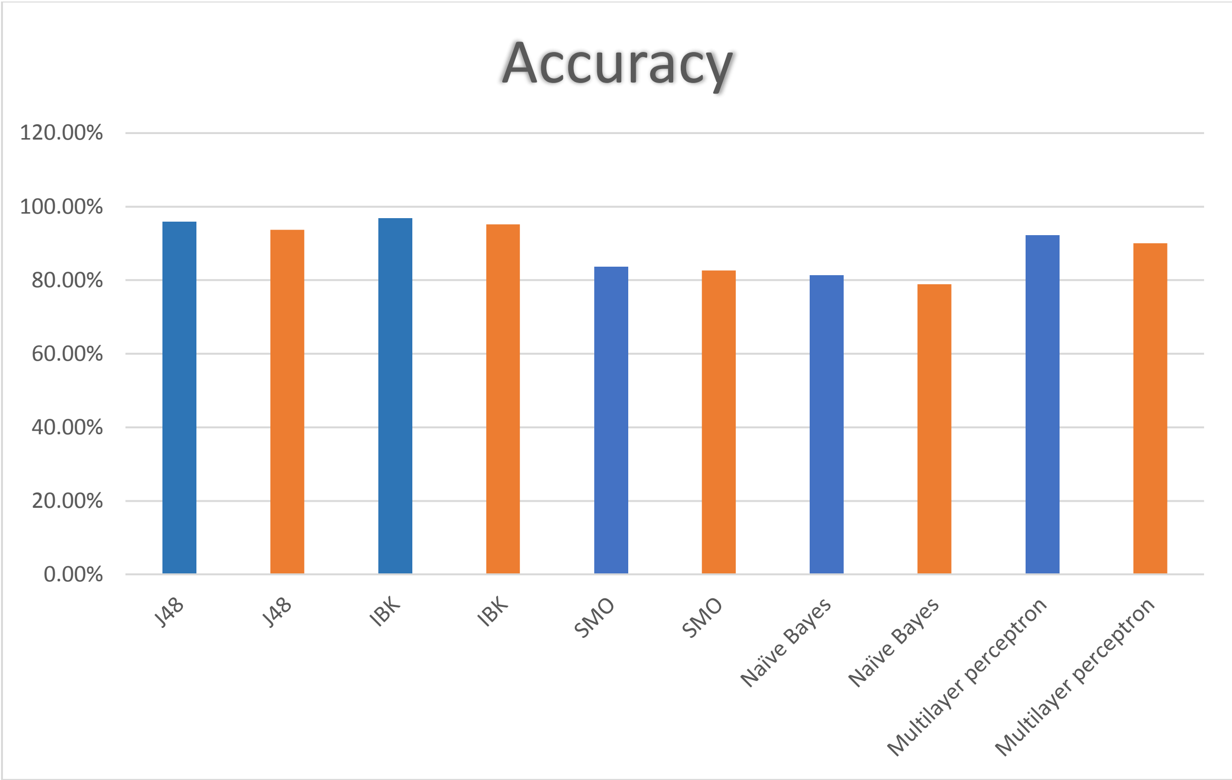


Figure 4.4 Comparison Test Mode with Classification Algorithms

4.4 Summary

The chapter discussed all experiments description and applied classification algorithms with weka by used statistical analysis for useful information and test mode for evaluation model is cross validation mode and percentage split mode.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATION

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

5.1 Introduction

The chapter discussed the research conclusion, challenges in faced research, recommendation.

5.2 CONCLUSION

The aim of the research is to study the effect of cesarean operation on Apgar factor using building classification model to obtained optimal Apgar score with healthy cesarean operation. There were a lot of difficulties faced the research. Obtaining the data was one of these difficultly. Also the hospital refused to give their data, they consider if as classified data.in many hospitals there is no good documentation.

The dataset of the research was obtained frommedical reports722 records of births after preprocessing. Applied five classification models were used in the research, the results of performance classification algorithms nearly same, but the highest accuracy 96.8 % obtained IBK algorithm using 10 cross validation and took less time to construct the model 0 seconds .used statistical analysis for some attributes to interesting information ,such as the fact that approximately53.74% women that gave birth natural operation and 46.26% women took cesarean operation .the Apgar score based on data from the mother, newborn and medical interventions. It values normal obtained approximately 80% of births, low approximately12.9% of births, very low approximately 7.1% of births. Low and very low Apgar score as found relationship between reasons cesarean operation and Apgar score after statistical analysis for reason of cesarean operation, building classification model are useful for effect cesarean operation on Apgar factor to obtained optimal Apgar score.

5.3 RECOMMENDATION

There are many recommendations:

- The early medical diagnosis for cesarean operation may avoid these problems by related mothers and their babies, like low Apgar score.
- Recommended researchers is collection data in Sudan.
- Increased used for data mining technequieses

References

[1] Velho. M. B. et al. (2012). "Experience with vaginal birth versus cesarean childbirth." 21(2) (Text Context Nursing, Florianopolis, 2012 Apr-Jun): 459.

[2] Gudayu, T. W. (March, 2017). "Proportion and factors associated with low fifth minute Apgar score among singleton newborn babies in Gondar University referral hospital; North West Ethiopia." African Health Sciences 17(1).

[3] Robu, R. et al. (2015). "The Analysis and Classification of Birth Data." ActaPolytechnicaHungarica 12 No.4

[4] Sana, A; Razzaq, R. and Ferzund, J (October 2012). "Automated Diagnosis and Cause Analysis of Cesarean Section Using Machine Learning Techniques." International Journal of Machine Learning and Computing 2, No. 5.

[5]Rahmanian, k. et al. (2014). "ASSOCIATION OF APGAR SCORE WITH DELIVERY MODE IN THE NON DISTRESS NEWBORNS." Online Journal of Biological Sciences 14(1): 21-25

[6] Sittidech, P. et al (April 2015). "Birth Asphyxia Classification Using AdaBoost Ensemble Method." Journal of Medical and Bioengineering 4, No 2: 126-129.

[7]Sudhir .B and Kodge. B, (2013). "Census Data Mining and Data Analysis using WEKA." International Conference in “Emerging Trends in Science, Technology and Management-2013, Singapore (ICETSTM – 2013):35-40.

[8] Han, J.andKamber, M. (2006). Data Mining Concepts and Techniques. . S.2th Edition. University of Illinois at Urbana-Champaign, Diane Cerra.

[9] Sewaiwar, P. et al. (October 2015). "Comparative Study of Various Decision Tree Classification Algorithm Using WEKA" International Journal of Emerging Research in Management &Technology 4(10): 87-91.

[10] Sudhir, M. (April 2017). "A Study of Some Data Mining Classification Techniques." International Research Journal of Engineering and Technology 4(4): 3112-3115.

[11]Sahoo, G. (2012). "Analysis of Bayes, Neural Network and Tree Classifier of Classification Technique in Data Mining using WEKA." (© CS & IT-CSCP 2012): 359-369.

[12] VidhuPriya, P.et al. (January 2017). "Classification Algorithms in Data Mining – A Survey." International Journal of Advanced Research in Computer Engineering & Technology 6(1).

[13] Sharma, T. C. (April 2013). "WEKA Approach for Comparative Study of Classification Algorithm." International Journal of Advanced Research in Computer and communication Engineering **2**(4): 1925-1931.

[14] Moreira, M. W. L. (October 2017). "Using Predictive Classifiers to Prevent Infant Mortality in the Brazilian Northeast." (IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)).

[15] Suman et al (2014). "A Comparative Performance Analysis of Classification Algorithms Using Weka Tool Of Data Mining Techniques." International Journal of Computer Science and Information Technologies ((IJCSIT)) **5**(3): 3448-3453.

[16] Rogers, B. a. Joynt, C (February, 2015). "APGAR Scoring System". www.pedcases.com/podcasts

Appendix

A1: weka 3.8 Tool

Weka is platform independent software. These tools and software provide a set of methods and algorithms that help in better utilization of data and information available to users, including methods and algorithms for data analysis, cluster analysis, Genetic algorithms, Nearest neighbor, data visualization, regression analysis, Decision trees, Predictive analytics, Text mining, etc.
Weka GUI chooser consists five buttons (applications) these are [15]:

- Explorer: An environment for exploring data.
- Experimenter: An environment for performing experiments and conducting statistical tests between learning schemes.
- Knowledge Flow: This environment supports essentially the same functions as the Explorer, but with a drag and- drop interface. One advantage is that it supports incremental learning.
- Workbench: The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces (GUI) for easy access to this functionality.
- Simple CLI: Provides a simple command-line interface that allows direct execution of WEKA commands from operating systems that do not provide their own command line interface.

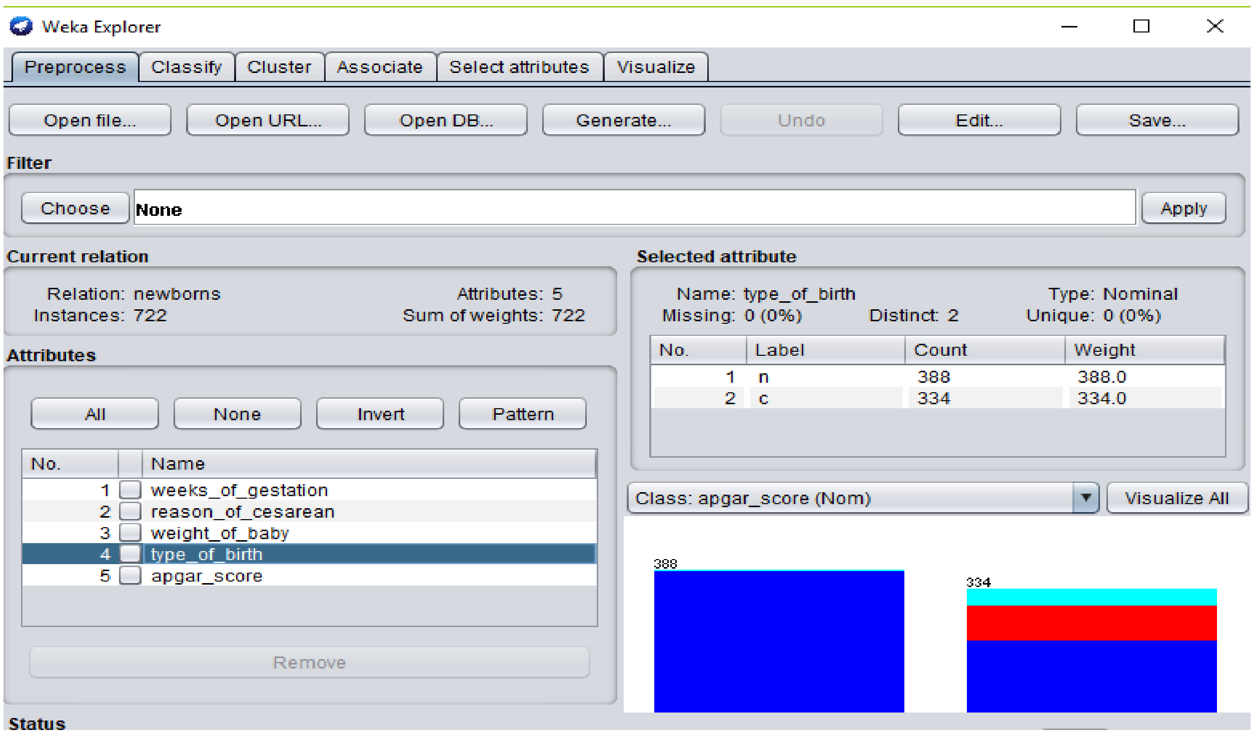
A2: Apgar Score

The Apgar score has been used for over fifty years to evaluate the overall condition of newborns in the first minutes of neonatal life. Apgar score is a method to quickly summarize the health of newborn children.Dr. Virginia Apgar, an anesthesiologist at New York–Presbyterian Hospital, developed the score in 1952 in order to quantify the effects of obstetric anesthesia on babies
Reference table A2.1, Newborn Resuscitation Algorithm 2010 American Heart Association [16]. It explained 5 assessment categories of the APGAR scoring system and criteria for the 0-2 scoring in each category. Observed intableA2.1 Apgar score in 1 min equal 4 scores that is very low, so recall 5 min equal 9score is normal and 10 min took 10 scores that is normal baby.

APGAR Score	0	1	2	1 min	5 min	10 min
Heart Rate	Absent	Under 100	Over 100	1	2	2
Respirations	Absent	Slow irregular	Good cry	1	2	2
Muscle Tone	Limp	Some flexion	Active	1	2	2
Reflex	None	Grimace	Cough/Sneeze	1	2	2
Color	Blue/Pale	Body pink limbs blue	Pink	0	1	2
Total /10				4	9	10

Table A2.1 Apgar Score System

A3: Statistical Analysis



FigureA3.1 Statistical Analysis for Type of Birth

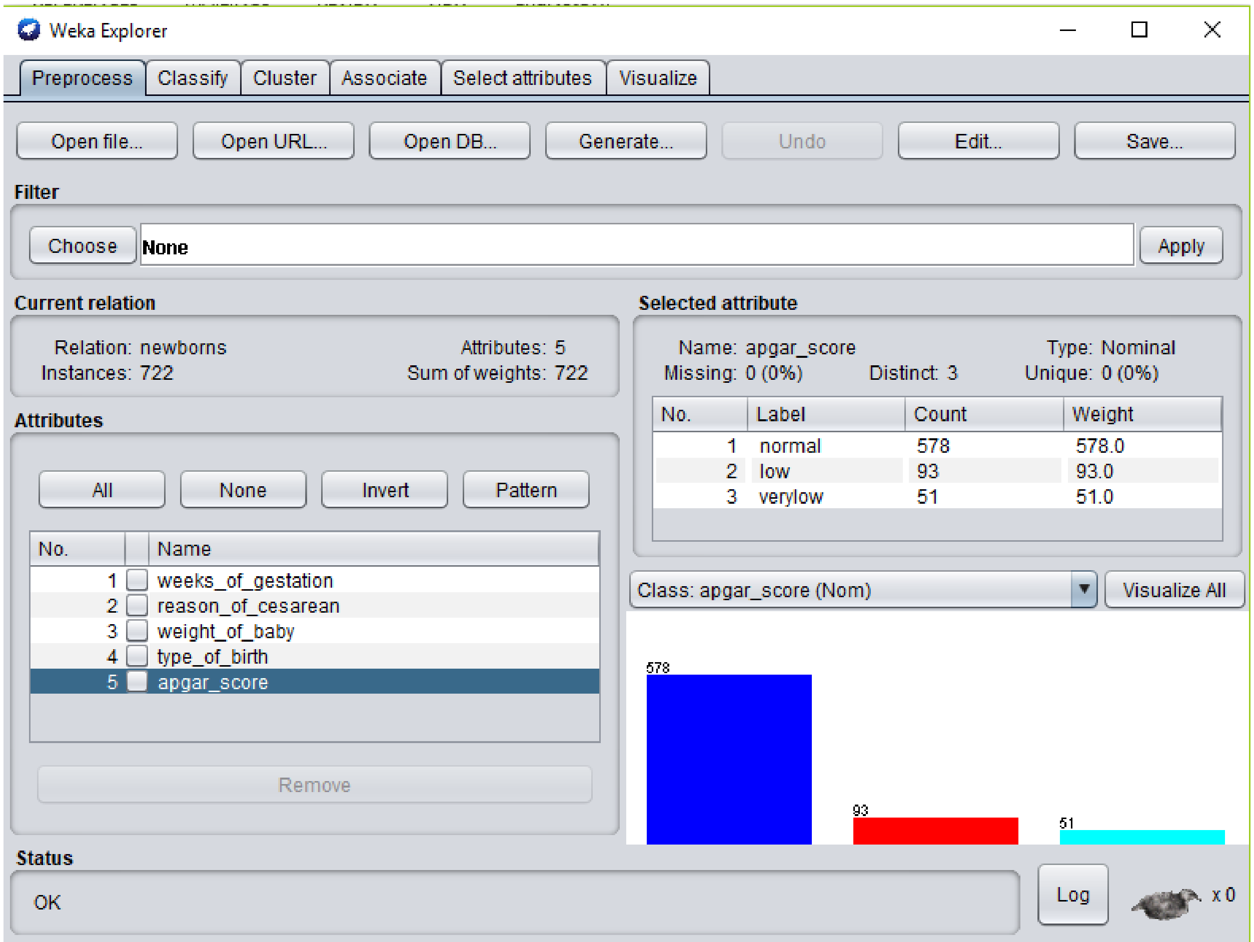


Figure A3.2 Statistical Analysis for Apgar Score

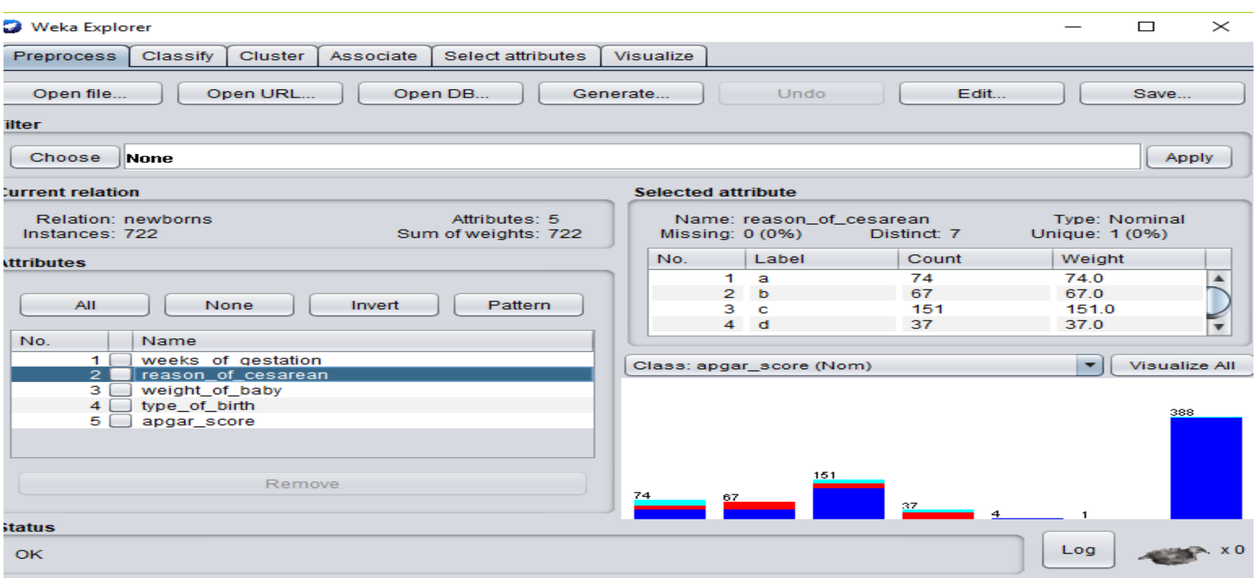


Figure A3.3 Statistical Analysis for Reason of Cesarean

A4:Experiments

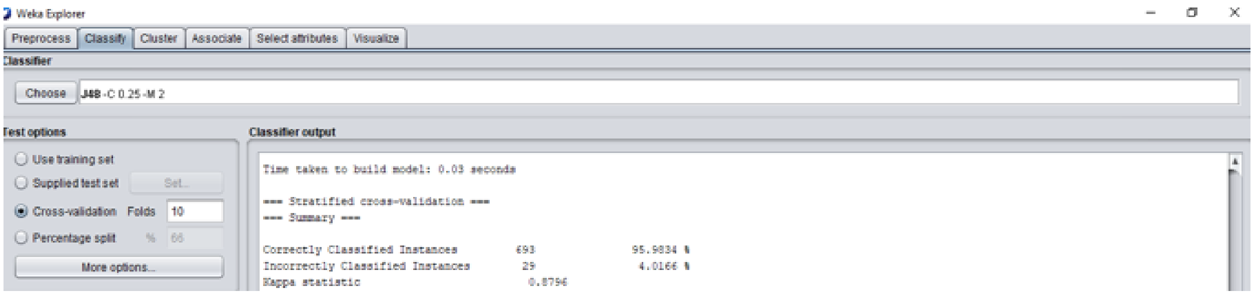


Figure A4.1 Accuracy J48 Classifier UsingCross Validation Mode

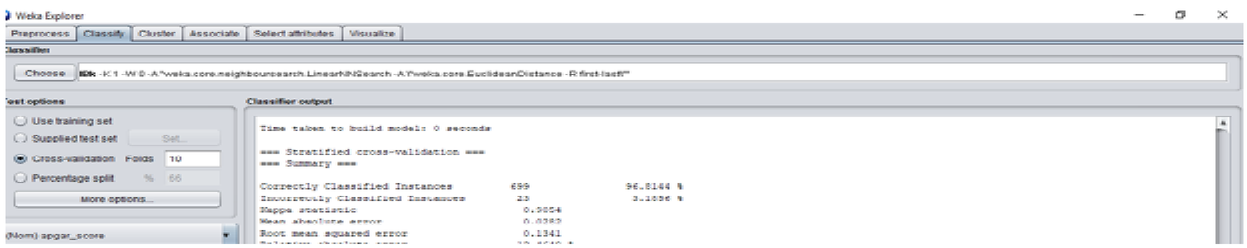


Figure A4.2 Accuracy of IBK Classifiers. Using Cross Validation Mode

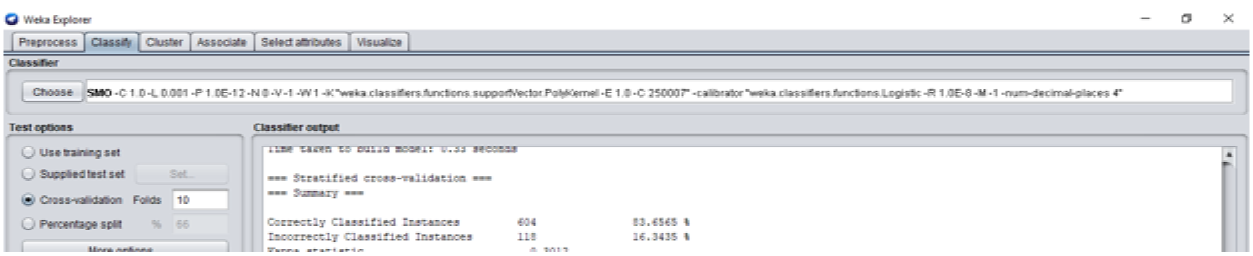


Figure A4.3 Detailed Accuracy by Class Using SMO Classifier.

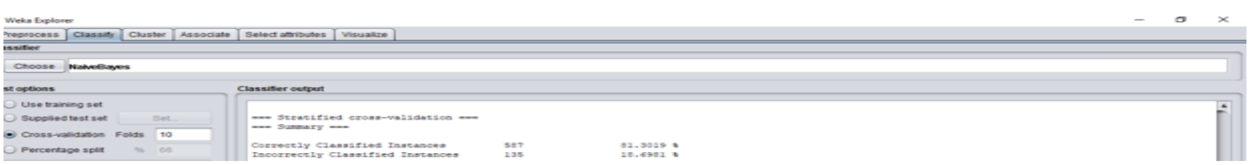


Figure A4.4 Detailed Accuracy by Class using Naïve Bayes classifier.

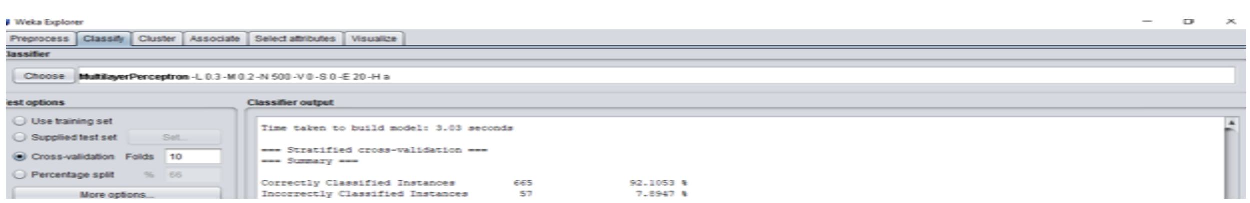


Figure A4.5 Detailed Accuracy by Class Using Multilayer Perceptron Classifier.

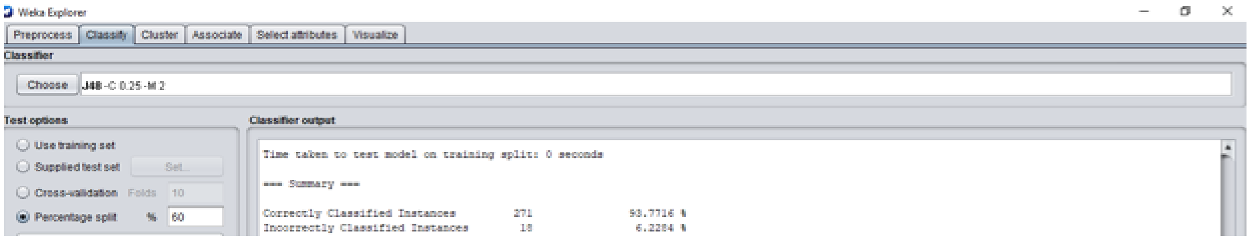


Figure A4.6 J48 Classifier Using Percentage Split Mode

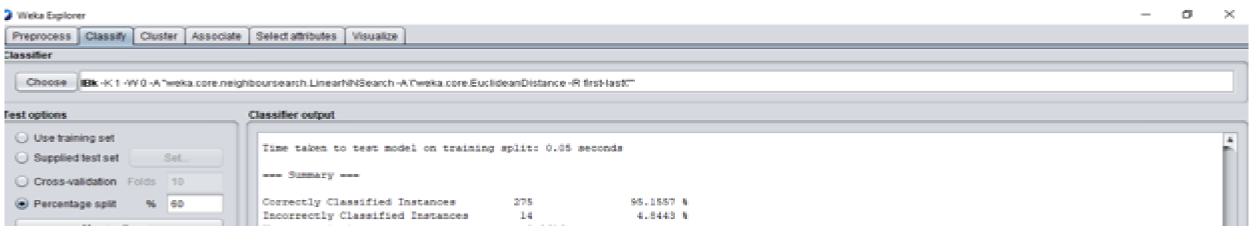


Figure A4.7 IBK Classifier Using Percentage Split Mode

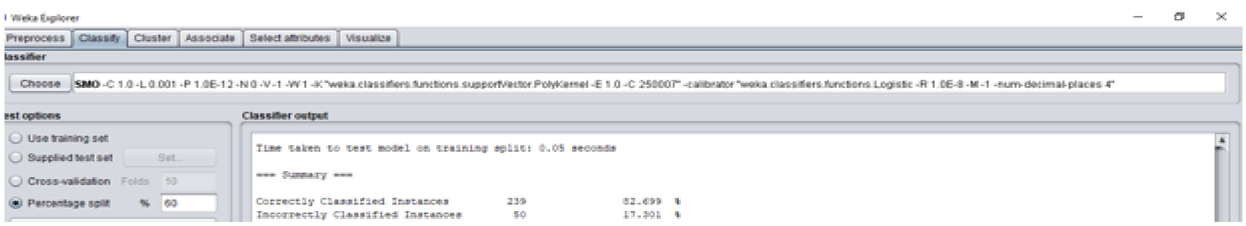


Figure A4.8 SMO Classifier Using Percentage Split Mode

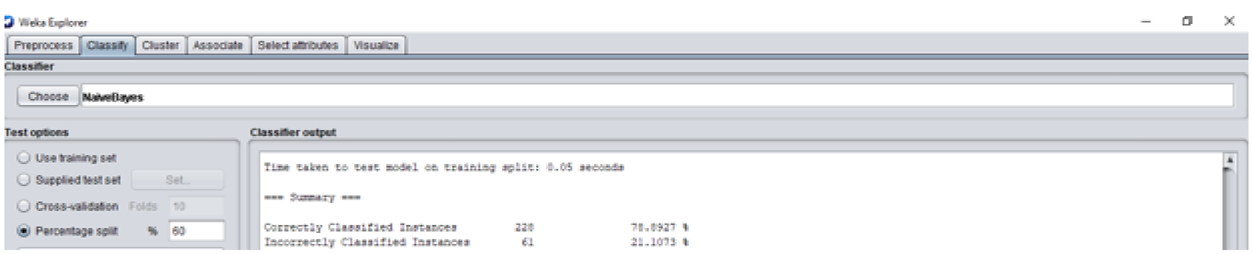


Figure A4.9 Naïve Bayes Classifier Using Percentage Split Mode

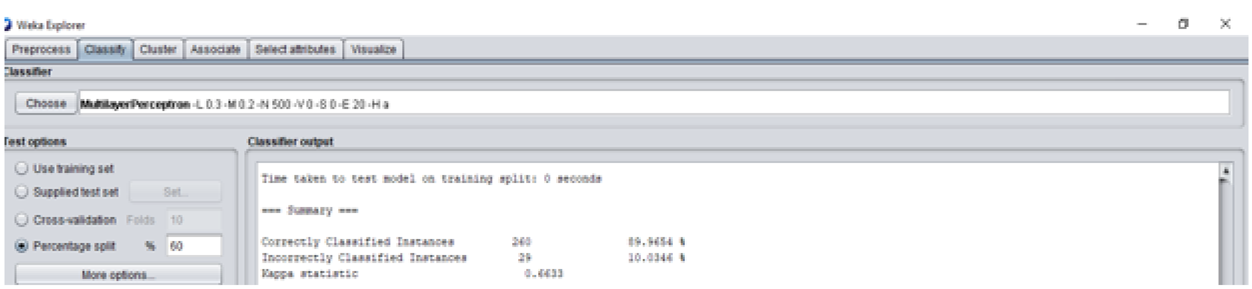


Figure A4.10 Multilayer Perceptron Classifier Using Percentage Split Mode