Palestine Polytechnic University

Academic Performance Prediction for Engineering Students using RBF-SVM Classification Model Case Study: PPU.

by

Amal Abuzalata

Supervisor: Dr. Mohammed Aldasht

Co-supervisor: Dr. Radwan Tahboub

A thesis submitted in partial fulfillment for the degree of Master in Informatics

in the

College of Graduate Studies and Scientific Research

July 2019

Declaration

I declare that this thesis titled, **'Academic Performance Prediction for Engineering Students using RBF-SVM Classification Model.Case Study: PPU.'** is my own original work, and all work contained within this thesis is my own independent research and has not been submitted for the award of any other degree at any institution, except where due acknowledgement is made in the text.

Amal Jaber Abuzalata

Signed: _____

Date: _____

Statement of permission to use

In presenting this thesis in partial fulfillment of the requirement for the master degree in Informatics at Palestine Polytechnic University. I agree that the library shall make it available to borrowers under rules of the library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of the source is made.

Permission for extensive quotation from, reproduction, or publication of this thesis may be granted by main supervisor, or in in his absence by the Dean of Graduate Studies and Scientific Research, when in the opinion of either the proposed use of the material is for scholarly purpose.

Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Amal Jaber Abuzalata

Signed: _____

Date: _____

Acknowledgement

I would like to express my sincere gratitude to my supervisors Dr. Mohammed Aldasht and Dr. Radwan Tahboub. They continuously provided me with help, encouragement and knowledge.

I thank my colleagues Bissan Abusharar, Feda Amro, Mahdi Atawneh, Iyad Hreni and Safa Adi, they gave me a lot of knowledge and answered all my questions.

Abstract

Predicting student performance is becoming more challenging due to large amount of data in the educational databases. Data mining tools could be used to explore academic databases with the objective of extracting knowledge about student's performance. Machine learning (ML) teaches computer to learn and make decisions, ML tools may include but not restricted: Decision Trees (DT), Artificial Neural Networks (ANN), Support Vector Machines (SVM)...etc.

The academic databases at Palestine Polytechnic University (PPU), contain millions of records of data about tens of thousands of students over the last 40 years. In this thesis, we decided to extract a subset of this huge data, and build our own data set which contains the most important features about engineering students according to expert advice at PPU. Then, we have built our model using SVM as a classifier and the Radial Basis Function (RBF) with linear kernels as a pre-processor for feature selection.

The proposed model is used to predict student performance at PPU. The data set of students in the engineering discipline is used as a benchmark. Results of our model show excellent classification accuracy and AUC for the student's performance, highest obtained accuracy and AUC is **96%** for optimized RBF-SVM model using the genetic algorithm and grid search method, these results make the model a reliable one to be generalized for the university students from other disciplines. Furthermore, it could be developed to be used by instructors, academic supervisors and students to take proactive steps to improve learning strategies.

Dedication

I dedicate this work to my family, my daughter Ritta, to all people who help and support me, Bissan Abusharar, Mahdi Atawneh, Feda Amro, Safa Adi. I appreciate their patient and love.

To my supervisor Dr. Mohamad Aldasht, co-supervisor Dr. Radwan Tahboub, also I want to dedicate Dr. Mohamad Awad and Dr. Ala Halwani for their effort and valuable comments on my thesis document.

Ø§ÙÙÙØ®Øµ

Ø³ÙØ¯Ù ÙØ°Ù Ø§ÙØ¯Ø±Ø³©.

Contents

List of Figures

List of Tables

Abbreviations

AI	**A**rtificial **I**ntelligence
ANN	**A**rtificial **N**eural **N**etworks
ASS	**Ass**igment
ATT	**Att**endance
AUC	**A**rea **U**nder **C**urve
BCA	**B**achelor of **C**omputer **A**pplications
CART	**C**lassification **A**nd **R**egression **T**rees
CSV	**C**omma **S**eparated **V**alues
CTG	**C**lass of **T**est **G**rade
EA	**E**volutionary **A**lgorithm
EDM	**E**ducational of **D**ata **M**ining
ESM	**E**nd of **S**emester **M**arks
FN	**F**alse **N**egative
FP	**F**alse **P**ositive
GA	**G**enetic **A**lgorithm
GP	**G**eneral **P**roficiency
ID3	**I**terative **D**lichotomiser 3
KDD	**K**nowledge **D**iscovery in **D**atabases
KNN	**K**-**N**earest **N**eighbour
LW	**L**ab **W**ork
ML	**M**achine **L**earning
MLP	**M**ulti **L**ayer **P**erceptron
MLR	**M**ultiple **L**inear **R**egression
NN	**N**earest **N**eighbour
PSM	**P**revious of **S**emester **M**arks
OVA	**O**ne of **V**s **A**ll
RBF	**R**adial **B**asis **F**unction
RF	**R**andom **F**orest
RL	**R**ule **L**earner
PPU	**P**alestine **P**olytechnic **U**niversity

SVM	**S**upport**V**ector**M**achine
SEM	**Se**minar of **P**erformance
TN	**T**rue **N**egative
TP	**T**rue **P**ositive

Chapter 1

Introduction

Predicting students performance is an important issue for students, instructors, and academic leaders. Depending on the results of a prediction model could be helpful for teachers to discover the level of their current students and take proactive steps in their teaching strategies. Also, it could be helpful in applying the suitable academic strategies at the academic program level, and in assessing factors of weakness level of students.
The fast improvement of technology during the last years, facilitates the collection of huge amounts of data, like students data, alumni data, courses data and much more all collected in data bases. To extract useful information from databases we need Data mining or what called 'Knowledge Discovery in Databases'.
Data mining could be used in academic manner to improve developing methods that discover academic databases and called 'Educational Data Mining' [3, 12, 13], many tools and algorithms could be used in Data mining like decision trees, artificial neural network, naive base, support vector machine, and many others.
Support vector machine (SVM) is a popular classification technique in data mining, this task involves separate data set into a training set and testing set, each set has a target value called 'class'. Our class in this thesis is GPA. Label 'A' mapped to students who have GPA equals to 90% and greater, 'B' for students who have GPA equals to 85% and greater, 'C' for students who have GPA equals to 78% and greater, 'D' for students who have GPA equals to 71% and greater,'E' for students who have GPA equals to 68% and greater,'F' for students who have GPA less than 68%, and 'O' label for students who did not have a bachelor degree in Engineering. In this thesis we used a benchmark data set collected from Palestine Polytechnic university then used to test build predicting models ones with 'O' label, and ones without.
SVM algorithm is used as binary classifier to find separating linear hyperplane between different classes, also to find a non-linear separator called kernel function like linear and radial basis function will be used in this research[14].

The aim of this thesis is to use Radial Basis Function (RBF) and linear kernels, under SVM classifier to build a prediction model that predicts students academic performance at Palestine Polytechnic University at engineering discipline.

1.1 Thesis objectives and hypothesis

Predicting academic performance for students from different disciplines is an important research topic for both students and teachers[15]. This thesis aims to:

- Predict the student's performance based on the student's GPA. The training data is collected from Palestine Polytechnic University databases over more than ten years.
- Based on the results obtained from the prediction model, teachers and instructors can take proactive decisions and steps to improve academic performance for weak students before getting fail.
- The model is also important for students, where low performance students can develop their way of learning and use different learning strategies. Results can make poor students rethink about their academic level and learning ways.

1.2 Thesis methodology

Many algorithms and techniques are employed in predictive models, including regression, classification, and clustering. This research aims to classify engineering students depending on their academic performance after five years of engineering study.
Two classification scenarios are presented and tested in this thesis using the support vector machine algorithm. Two models build using RBF-SVM and linear-SVM depending on data set that collected from Palestine Polytechnic University data base for engineering students to predict their academic performance after five years of study.
After building model of prediction, an optimization task is done using two ways: First way is using grid search method, Second using genetic algorithm in order to optimize algorithm parameters and to have a better predictive model.

1.3 Thesis contributions

- Build our own data set from PPU data base about engineering students who joined university from 1994 to 2017 that contains personal, pre-university, and university information about students. Then make some preprocessing to be useful for use.
- Create a classifier model based on a Support Vector Machine algorithm, that classifies students performance at PPU based on two kernels: radial basis function and linear to build two prediction models.
- Make a comparison between two models, then try to enhance RBF-SVM model using genetic algorithm and grid search method.

1.4 Thesis structure

This thesis is composed of five chapters. Chapter 1 gives an introduction, problem statement, methodology and contribution. Chapter 2, Background and Related Work, brief historical background and explains some fundamental concepts related. . Following, Chapter 3 introduced proposed data set, how is collected and prepossessed, then show system architecture , describes the proposed architecture of the developed system, and the implementation of the theoretical model used on the matter; methodology already addressed to the present problem, showing some existing solutions in this field of studies. Chapter 4 presents experiments and results that describe the several case studies made, showing result analysis and the solutions obtained. Finally, chapter 5 gives a conclusion which summarizes the respective work done and the following conclusions about the results obtained and the approaches used, together with several future work propositions.

Chapter 2

Background and Literature Review

This chapter has two sections, background that presents a theory explanation about academic performance, data mining and prediction algorithms, support vector machine and our data set, each topic in sub section, literature review that make a caparison between our research and previous work.

2.1 Background

2.1.1 Academic performance.

In our recent days and technology revolution, education is the first and most important step in humans life. Education develop opportunities and links between people, it also increase productivity and level of lives [16].
Academic performance is a combination of two words, academic is a well known specialized in education and research. Performance is a measurement word, that means how someone or something performs, so academic performance measures how someone do well or bad in academic manner.
Someone level in academic manner often reflected in grades and exams scores, but there are many key factors that affect students performance also [17]. In this thesis a prediction model will be build to measure engineering students performance after five years of studying by predicting students GPAs category.

2.1.2 Data mining and prediction algorithms.

Urban's book "Decisions support systems and intelligent systems" data miming is a term that used to describe knowledge discovery in databases and its a process that use

mathematical statistics to extract knowledge from large data bases [18].
One of the most popular question asked by most people about particular machine learning algorithms whether is a supervised or unsupervised algorithm. Supervised algorithms are algorithms that need a training set to learn, or labeled data set, where unsupervised algorithms do not need any labeled training data set to work properly [19].
The following are the main types of algorithms [20]:

- Classification: classification algorithms use existing data as training sets, and put a class based on some attributes for each sample in training data set.
- Regression: regression algorithms use existing data to build a mathematical model. The most important difference between classification and regression is the output type. Regression predicts numeric values output, where classification predict a class label.
- Clustering: clustering algorithms divide existing data set into groups or clusters, depending on properties similarity.
- Association: between different features or attributes there is a relation called correlation. association algorithms finds this correlation, by find items that frequently occur together.
- Sequence analysis: sequence analysis algorithms find a frequent sequence in dataset.
- Time series: output of theses algorithms is numeric values, but in an ordered series.
- Dimensional Reduction Algorithms: these algorithms identify the most important variables in datasets.

Data mining is a discipline or process that aims to extract knowledge and finds correlations from large datasets, since data growing fast and numerous.

2.1.3 Support Vector Machine algorithm.

Next step in data mining is to determine which prediction technique to use, here classification task is applied to build a prediction model.
Classification task predict categorical labels of classes, so it is a supervised learning task that consists of two points: First to build a classifier or model based on training set, and second to use classifier or model for classification [21].
Classification technique consists of many algorithms like decision trees, Bayesian classifier, Rule-based, Support Vector Machines and others. This thesis presents a prediction

model that predict the academic performance of engineering students at PPU. The model is an SVM algorithm that uses linear and RBF kernels, then finds accuracy of each kernel, then make some optimization using grid search and GA.
Support Vector Machine or SVM is a useful supervised learning algorithm used for classification or regression type.
SVM as classification algorithm attempt to classify data into target classes with the widest possible margin, while SVM as regression algorithm tries to find a continuous function where the maximum number of data points are within an epsilon-wide tube around those data elements.

This subsection will explain SVM briefly. This classification algorithm involves separating data into 'training' and 'testing' sets. Each sample in the training set has a 'target value' called 'class label' or 'label' and has also several 'attributes' or 'features'. SVM algorithm used to produce a model depends on training data set to predict classes of test set.
SVM is a binary classifier but in this research SVM used as multi class classifier. Multi class classification is a major requirement in engineering and science fields because of discrimination of objects in these fields. It is more complex than binary classification since each decision boundary for some class is known so that rest or complement of first class is second class, where in multi class classification there are several boundaries and probability of errors will increase also.

Support Vector Machine is a machine learning tool based on supervised learning, used to classify data samples into two distinct classes, that what called "binary classification". Two diff rent classes divides by hyperplane that separate samples belongs to some class from other samples belong to another class [22]. This was the case when data is linear separable but mostly we find data is non-linear separable and data set needs then to use kernels like linear, RBF, polynomial...etc.

SVM gives accuracy almost equal to neural networks (ANN). Support vector machine is used for many applications such as text categorization, pattern recognition, face recognition, handwriting analysis but especially for classification applications. Support vector machine is the most accurate algorithm compared to others [23]. Support vector machine benefits are:

1. Scales well to high dimensional data and explicitly controls trade off between complexity and error.

2. SVMs and kernel methods used for a particular problem and also applied directly to the data.

3. Used in problems that have a lot of structure of the data that loses by the feature extraction process.

The limitations of SVM are:

1. Speed, size both in training and testing [24].

2. High algorithmic complexity and extensive memory requirements.

3. How to select the kernel function parameters.

For **binary classification** problems, the idea behind SVM is to split the data set classes. Binary classification is used when we need to classify to two classes. There are many examples of binary classification like success or fail in exams, go to play outside or not. Support vector machines are designed for two binary classification problems [25] and consider two topics:

1. LINEAR SEPARATION.

2. NON-LINEAR SEPARATION.

Let us consider first case **(LINEAR SEPARATION)**. There are many linear decision boundaries that divide the data. But one of them gives maximum margin. Support vectors are the data points that lie closest to the decision boundary and they are the most difficult samples to classify [26]. Thr problem here is to find a hyperplane that margin is not too large, also not too small in order to obtain the lowest misclassification errors [22].

A cost parameter "C" allow flexibility, that controls trade off between errors and how much is margins by creating soft margin.

Multiclass classification means "Each training point belongs to one of N different classes. The goal is to construct a function which, given a new data point, will correctly predict the class to which the new point belongs." [27].
Numerous approaches suggested to solve the problem of multi class classification using SVMs, but **(one-versus-all or OVA classification)** is the best where each category is split out and all other categories merged to be as one category. Its just like having "n" binary problems.

Suppose that samples $(x_i, y_i), i = 1, ..., l$ where $x_i \in R^n$ and $y \in \{1, -1\}^l$, then SVM requires to solve the following optimization equation [28]

$$min_{w,b,\varepsilon} \frac{1}{2} w^T w + \sum_{i=1}^{l} \varepsilon_i \tag{2.1}$$

$$To find : y_i \left(w^T \Phi (x_i + b) \right) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0. \tag{2.2}$$

Where Φ is a function that maps training samples w_i into higher dimensional space. $C > 0$ is an error parameter, b is bias, ε_i is epsilon variable that determines the level of accuracy of the approximated function in linear SVM. $\Phi (x_i)^T \Phi (x_j)$ is called kernel function and denoted as $K (x_i, x_j)$, there are four famous or basic kernels:

- Linear:
$$K (x_i, x_j) = x_i^T x_j \tag{2.3}$$
- Polynomial:
$$K (x_i, x_j) = \left(\gamma x_i^T x_j + r \right)^j, \gamma > 0 \tag{2.4}$$
- Radial Basis Function RBF:
$$K (x_i, x_j) = exp (-\gamma \| x_i - x_j \|^2), \gamma > 0 \tag{2.5}$$
- Sigmoid:
$$K (x_i, x_j) = tanh \left(\gamma x_i^T x_j + r \right) \tag{2.6}$$

Where γ, r, d are kernel parameters. In this thesis two kernels were used, linear and RBF. Parameters optimized using grid search and Genetic Algorithm.
The procedure used in SVM classifier is explained in the following points:

1. Transform the data set into format that SVM package accept like csv format.
2. Conduct some scaling to the data set. Scaling is very important, scaling avoid dominating small numeric values from large numeric values, also to make calculations in SVM kernels easier so commonly linear scaling is recommended.
3. For some kernel, like RBF kernel=$exp (-\gamma \| x_i - x_j \|^2)$, euclidean distance is used.
4. Use cross validation to find the best values of γ and C parameters where γ is the radius of RBF and C is penalty parameter of the error term.

5. Finally test your test set, it is recommended to use blind data set to be sure that your model did not overfit your data.

For a training set (x_i, y_i) where $i = 1, ..., l$ and x_i could be vector of features or p-dimensional real vector, y_i is the class and takes two values -1 or 1. The goal of Linear SVM is to find maximum-margin hyper plane that divide samples into two groups, one class -1, and other class 1 [29].

$$w.x - b = 0. \tag{2.7}$$

Equation 2.7 determines suitable hyper plane, where w is weight vector, and b is bias. Fig 2.1, shows the case of linear kernel where the circled data points are the support vectors that are closest to the decision bounder. They determine the margin with which the two classes are separated .

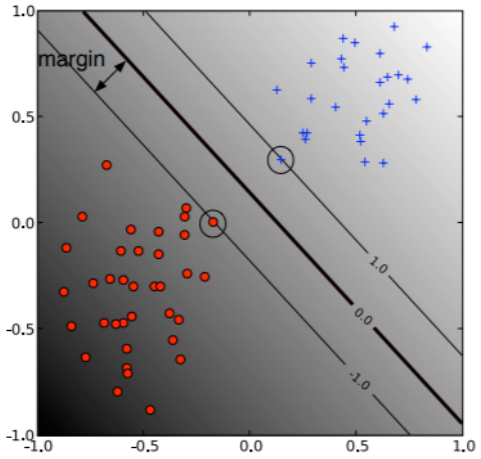

FIGURE 2.1: A linear SVM.

The other kernel used by SVM is radial basis function (RBF) kernel. This kernel maps x_i samples to higher dimensional space. It is used when the relation between samples and labels nonlinear, unlike linear kernel.

$$exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \gamma > 0 \tag{2.8}$$

RBF kernel is good choice in SVM since its numerical calculations not difficult like polynomial kernel for example, it has two parameters, C and gamma (γ). Gamma is the radius of RBF.

Figure 2.2 shows that the separated hyper plane between class is nonlinear, and shows two different values of γ.

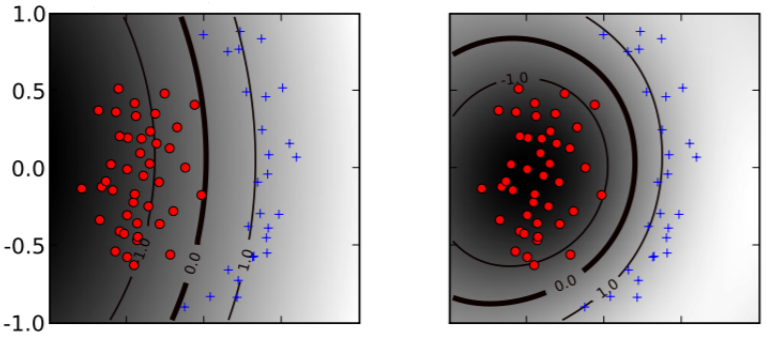

FIGURE 2.2: A Kernel SVM. Small value of γ to the left vs. large value of γ to the right.

2.1.4 Our dataset.

To build a prediction model, data must be collected to establish training data set for model. So our data set collected from PPU (Palestine Polytechnic University) data base to be used in building prediction model for academic performance for engineering students at PPU.
Collected data can be divided into three types:

1. Personal data
2. Pre university
3. University data.

This data set consists of 22 attributes, selected by an expert advise, Dr. Mohammad Aldasht the dean of registration department at PPU. There attributes listed and explained in details at table 3.1 in chapter three of this thesis document. These attributes are:

1. The country from which the student obtained a high school certificate.
2. Year of high school diploma.
3. General Secondary Branch (Scientific, Industrial, Air Conditioning, ...).
4. High school average (GPA).
5. Specialization code at university.

6. Specialization name.
7. Number of times the student received an academic warning.
8. University GPA.
9. Regularity of students in the study (regular, irregular, leaving study).
10. Has the student been dismissed from university.
11. Did student graduate from university.
12. The town or city where student lives.
13. Students mark in Physics 1.
14. Students mark in Physics 2.
15. Students mark in Calculus 1.
16. Students mark in Calculus 2.
17. Students mark in English 1.
18. Students mark in English 2.
19. The year when student start studying at PPU.
20. Number of semesters until student graduated.
21. Number of honor semesters.
22. Class of GPA.

2.2 Literature Review

In this section, there is a discussion and a comparison between previous researches about predicting students performance, and this research. First paper [3] is talking about prediction of students performance using data mining methods, it focus on two aspects of under graduate students performance. First is prediction students achievement after four years study, second combining prediction results with typical progressions. The difference here they used some courses as attributes to train their model, and used data mining tools like clustering and decision trees.
Used data set in this research is 270 students, and the following are predictors or attributes in this data set in table 2.1:

TABLE 2.1: Table of attributes [3]

Code	Country
1	student's gender.
2	nationality category.
3	first language.
4	teaching language in the university.
5	high school percentage.
6	student status depending on his/her earned credit hours.
7	living location.
8	does the students have any sponsorship.
9	any parents works in the university.
10	student discounts.
11	how the students comes to the university.
12	family size.
13	total family monthly income.
14	parents marital status.
15	fathers qualification's.
16	fathers occupation status.
17	mothers occupation status.
18	number of friends.
19	average number of hours spent with friend weekly.
20	previous semester GPA.

Three types of decision trees are used to predict students performance as four classes excellent, very good, good, pass. Used decision trees are C4.5, ID3, CART decision trees with obtained accuracy from these models are 35.19%, 33.33%, 34.07% respectively [3].

In [4] study, authors study the performance of engineering students in engineering dynamics and other challenging courses since engineering students always perform poorly or even fail in these courses. These courses requires good understanding of concepts and skills of fundamentals.

This study aims to build some mathematical models in order to predict engineering students performance in engineering dynamics. 24 predictive models were build in this study using four statistical and data mining modeling techniques and a combinations of eight predictors. Used techniques including MLR, MLP networks, RBF network, and SVM.

Predictors used in this study are the following in table 2.2:

The upper most accuracy of each model using MLR, MLP, RBF, SVM after combinations of predictor variables is 89.7%, 89.6%, 89.9%, 88.7% respectively. So the most accurate model is the one using RBF network.

Another research [5] built a model using decision trees in order to predict engineering students performance, it supposed that students score in statics course and cumulative

TABLE 2.2: Table of attributes [4]

Code	Attribute
1	cumulative GPA.
2	statics grade.
3	calculus I grade.
4	calculus II grade.
5	physics grade.
6	score of dynamics mid-term exam #1.
7	score of dynamics mid-term exam #2.
8	score of dynamics mid-term exam #3.

GPA plays a critical roll in prediction students performance in dynamics course. Predictors used in this research are in table 2.3:

TABLE 2.3: Table of attributes [5]

Code	Attribute
1	static's course mark.
2	cumulative GPA.
3	physics.
4	Calculus I.
5	Calculus II.

Results from this model shows for each semester as four classes A, B, C, DF. It applied on two semesters A and B. In semester A accuracy is 83.3% where in semester B is 85,9%.

Other research [6] is a comparative study uses data mining methodologies to study how students perform in their courses. In this research decision trees used as a classification task, many types of decision trees are used ID3, C4.5, and CART. Many predictors or attributes were collected about students to predict their performance at exams of the end of semester

Collected attributes about 48 students are in table 2.4:

TABLE 2.4: Table of attributes [6]

Code	Attribute
1	Previous semester marks.
2	Class test grade.
3	seminar performance.
4	Assignment.
5	Attendance.
6	Lab work.
7	End semester marks.

Table 2.5 shows the obtained accuracy using each type used of decision trees ID3, C4.5, and CART.

TABLE 2.5: Classifiers accuracy [6]

Algorithm	Correctly classified instances
ID3	52.0833%
C4.5	45.8333%
CART	56.25%

In [7] the research apply decision trees as a classification method also to predict students performance at final exams. So decision tree gives an approximate number of students who will pass, fail or promoted to the next year so results will be used to improve and to cover weakness points for students who are predicted to fail or promoted to next year. Three types of decision tree algorithm used in this research ID3, C4.5, and CART. 17. Students related variables collected about 90 students to be predictors or attributes for prediction model listed below in table 2.6:

TABLE 2.6: Table of attributes [7]

Code	Attribute
1	Students branch.
2	Students gender.
3	Students category.
4	Students grade in high school.
5	Students grade in senior secondary.
6	Admission type.
7	Medium of teaching.
8	Living location of student.
9	Student lives in Hostel or not.
10	Students family size.
11	Students family status.
12	Family annual income.
13	Fathers qualification.
14	Mothers qualification.
15	Fathers Occupation.
16	Mothers Occupation.
17	Result in B. Tech 1st year.

Table 2.7 shows obtained accuracy using each type used of decision trees ID3, C4.5, and CART.

Next research is [8] which studies a turning point in India which is academic performance of students in higher education. This research presents a model to predict students performance at final exams using Naive Bayes classifier. Data collected about 300 student

TABLE 2.7: Classifiers accuracy for [7]

Algorithm	Correctly classified instances
ID3	62.2222%
C4.5	67.7778%
CART	62.2222%

and 17 attributes used as predictors in obtained model. Theses features are in table 2.8.

TABLE 2.8: Table of attributes [8]

Code	Attribute
1	Students gender.
2	Students category.
3	Medium of teaching.
4	Students food habit.
5	Students other habit.
6	Living Location.
7	Student live in hostel or not .
8	students family size.
9	Students family status.
10	Students grade in Senior Secondary education.
11	Family annual income.
12	Fathers qualification.
13	Mothers qualification.
14	Fathers Occupation.
15	Mothers Occupation.
16	Grade obtained in BCA.

Another research used classification methods to analyze students performance [9]. Authors used ID3 algorithm as a decision tree method to evaluate students performance at the end of semester exams, they try to justify the capabilities of education data mining techniques by offering a prediction model for higher education students performance at a university. This research helps students and teachers to know weakness side for students so that teachers can give their students some advise or some help to improve their situation before final exams.
Data set used to build prediction model consists of eight predictors or features about 50 students, collected along four years of students, features are in table 2.9:

Obtained model from this data set is shown as IF Then rules shown in the following IF-Then rules algorithm:

if PSM ='First' AND ATT= 'Good' AND CTG='Good' or 'Average' **then**
ESM ='First'.

TABLE 2.9: Table of attributes [9]

Code	Country
1	Previous Semester Marks.
2	Class Test Grade.
3	Seminar Performance.
4	Assignment.
5	General Proficiency.
6	Attendance.
7	Lab Work.
8	End Semester Marks.

end if
if PSM ='First' AND GTC= 'Good' AND ATT='Good' or 'Average' **then**
ESM ='First'.
end if
if PSM ='Second' AND ATT= 'Good' AND ASS='Yes' **then**
ESM ='First'.
end if
if PSM ='Second' AND CTG= 'Average' AND LW='Yes' **then**
ESM ='Second'.
end if
if PSM ='Third' AND ASS= 'No' AND ATT='Average' **then**
PSM ='Third'.
end if
if PSM ='Failed' AND CTG= 'Poor' AND ATT='Poor' **then**
PSM ='Fail'.
end if

Another paper [10] presents results of research done on one famous Bulgarian university, using data mining techniques and based on personal, pre-university and university performance data that collected about students admitted to university in three consecutive years.

This research used these educational data and data mining algorithms as educational data mining (EDM) classification algorithms like rule learner, decision tree classifier, neural networks and nearest neighbour then build some prediction models depending on obtained data set.

Data collected in this research is three types, personal data, pre-university data and university data as explained in table 2.10:

TABLE 2.10: Table of attributes [10]

	Code	Attribute
Personal data	1	Gender.
	2	Age.
	3	Birth year.
Pre-university data	4	Place of previous education.
	5	Profile of previous education.
	6	Total score of previous education.
	7	Admission year.
	8	Admission exam.
	9	Admission exam score.
	10	Admission score.
University data	11	University specialty name.
	12	Current semester.
	13	Number of failures.
	14	Students class, takes two values strong and weak.

So generated models using classification algorithms in this research are OneR Rule Learner, Decision Tree, Neural Network and K-Nearest Neighbour and applied on their collected data about their students reveal an accuracy for each model.
Neural Network model achieved the highest accuracy of 73.59%, followed by accuracy of Decision Tree model with 72,74%, and K-NN model achieved accuracy of 70.49%. Neural Network performs better with "strong" class, where rest models make better with "poor" class.

In [11] WEKA software is used, and educational data mining principles using university data from two databases to determine students profiles and predict their performance. This research includes 10330 student, described by the following 13 parameters shown in table 2.11:

Task here is to predict the students performance at university based on personal information and pre-university data about student. The target performance will be as class, here five distinct classes:

- Excellent:(5.50-6.00).
- Very good: (4.50-5.49).
- Good:(3.50-4.49).
- Average:(3.00-3.49).
- Bad:(below 3.00).

TABLE 2.11: Table of attributes [11]

Code	Attribute
1	Gender.
2	Birth year.
3	Living place.
4	Type.
5	Profile.
6	Place and total score from previous education.
7	University admittance year.
8	Exam and achieved score.
9	Current semester.
10	Total university score.
11	Country.
12	Place of birth.

Using WEKA, researchers used some classification algorithms, C4.5 (J48) as decision tree, NaiveBayes (NB) and BayesNet (BN) as Bayesian classifiers, Nearest Neighbour algorithm (NN), OneR and JRip algorithms. Table 2.12 shows obtained accuracy results using these algorithms depending on the collected data about students.

TABLE 2.12: Results for the accuracy of the applied classifiers [11]

Results	J48	NB	BN	k-NN, k=100	k-NN, k=250	OneR	JRip
TP Rate	0.663	0.586	0.591	0.613	0.593	0.546	0.632
Precision	0.640	0.594	0.597	0.574	0.563	0.480	0.611

Another research [1] proposed an organized and timely prediction algorithm that works when accuracy of prediction is sufficient, to predict the final grade of under graduate students. Both classification and regression tasks performs using this algorithm, regression to predict the over all grade for each individual student, or as classification task to divide students into class who did well, and other class who did poorly. Predictors used in this research are achievements scores for each student in early assessment like class exams, midterm exams, home works, age, gender and previous GPA. Prediction algorithms shows that class exams is better predictor for students final grade.
Figure 2.3 shows system diagram for single student.

The purpose of [30] study is to analyze students performance using sparse dataset, students data collected from DMIT (Department of Mathematical Information Technology) at the University of JyvÂšaskylÂša in Finland between 2009 to 2013.
Their data set consists of 13640 study records sample with 21 attributes, related to the passed course and the studentâs affiliation, and of 1040 students who has attended a total of 1271 different courses, completing a total of 64905 credits. Only 64% of these

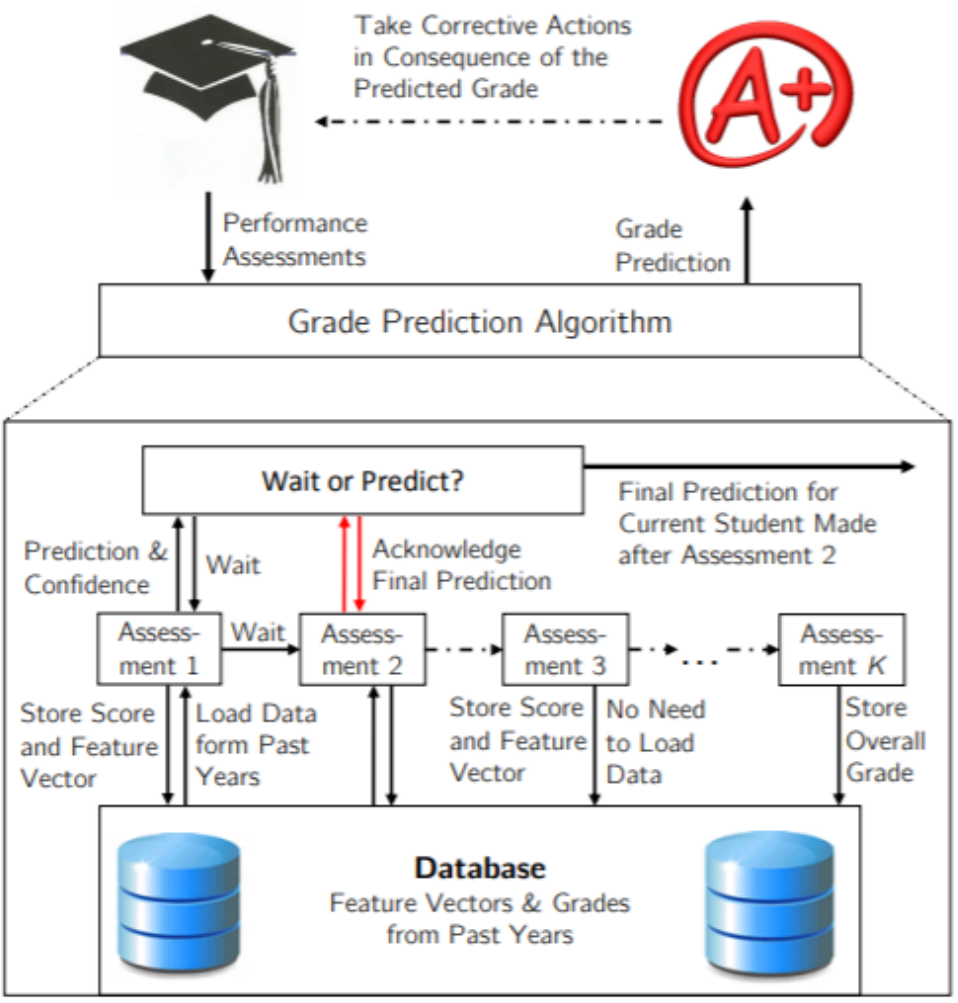

FIGURE 2.3: System diagram for a single student [1].

credit hours are the computer science students obtained from courses in their own faculty or college.

Authors applied a non standard educational data mining techniques on the passed courses for bachelor degree students, to build a non linear regression model using perceptron (MLP) neural networks with cross validation. Their study shows that students success could be predicted better than specific IT skills.

At the end of this study authors conclude that introductory courses, Data networks and Computer Structure and Architecture determined the quality of studies.

Recommend system approach in [31] used to predict students performance at next term. They build a Factorization Machines (FM) model that predict students grades in next term using simple base line and MF (Matrix Factorization) based methods for data set collected from George Mason University (GMU), this model achieved lowest prediction error.

Data used in this study collected from public university data base for 33000 student that enroll university from 2009 to 2014 for total 15 terms from 144 majors. During this period, students have 9085 different courses taught by one of 7347 instructors.
Predictors used in this study belong to students and terms, for each students collected attributes are age, race, sex, high school CEEB code and GPA, zip code, and 1600-scale SAT scores. For each term, we have the GPA from the previous term as well as the cumulative GPA, number of credit hours, academic level obtained from credit hours and relative term number. Also a variety of data for courses like particular discipline, particular course level, and aggregate GPA for each term.
Three methods based on Matrix Factorization (MF) used:

- Singular Value Decomposition (SVD).
- SVD-kNN: SVD post-processed with kNN.
- Factorization Machine (FM).

In subsequent study [32] from previous study [31] authors apply FM, random forest, multi linear regression algorithms to learn previous model with additional information about courses and instructors teaching them.
Four different regression models:

- Random Forest (RF).
- Stochastic Gradient Descent (SGD) Regression.
- k-Nearest Neighbors (kNN).
- Personalized Multi-Linear Regression (PMLR).

The following are features used to build prediction model
Student features:

- sid: Unique identifier of the student. When used in training data, the student ID is one-hot encoded to learn student bias terms.
- grdpts: [0, 4] grade the student has obtained for a particular course.
- major: Declared major during current term.
- race: Self-reported race of student; may be unspecified.
- sex: Self-reported gender; may be unspecified.

- age: Age determined from birth date in admissions records.
- zip: Zip code, or postal code for students from outside the US.
- sat: 1600-scale SAT score, if available.
- hs: High school CEEB code.
- hsgpa: High school GPA. For transfer students, this feature contains the GPA from the institution the student is transferring from.
- lterm gpa: Term GPA from the previous term.
- lterm cum gpa: Cumulative GPA as of the previous term.
- term chrs: Number of credit hours the student is enrolled in during the current term.
- total chrs: Number of credit hours student has taken (not passed) up to the previous term.
- alevel: Academic level of the student. Obtained by binning total chrs: [0,30)=0, [30,60)=1, [60,90)=2, [90,120]=3, (120+)=4.
- sterm: Chronological numbering of terms relative to this student. The studentâs first term is 0, his second is 1, and so on.

Course features:

- cid: Unique identifier of the course. When used in training data, the student ID is one-hot encoded to learn course bias terms.
- cdisc: Course discipline.
- chrs: Number of credit hours this course is worth.
- clevel: Course level [1, 7].
- termnum: Number of the term this course was offered in. These are relative to the dataset only. Term 0 is the first term for which we have data, and they are numbered chronologically onward from there.
- num enrolled: Number of students enrolled in this course for the current term, across all sections.
- total enrolled: Total number of students enrolled in this course since its first offering, including the current term.

- lterm cgpa: The aggregate [0, 4] GPA of all students who took this course during the previous term.
- lterm cum cgpa: The aggregate [0, 4] GPA of all students who have ever taken this class up to the previous term.

Finally, instructor features:

- iid: Unique identifier of the instructor. When used in training data, the student ID is one-hot encoded to learn instructor bias terms.
- iclass: Classification (Adjunct, Full time, Part time, GRA, GTA).
- irank: Rank (Instructor, Assistant Professor, Associate Professor, Eminent Scholar, University Professor).
- itenure: Tenure status (Term, Tenure-track, Tenured).

This chapter provides a fundamental concept about one-vs-all method, selection, prediction, support vector machine and used kernels especially RBF and Linear kernels, genetic algorithm and grid search methods. Focus on related work on different prediction algorithm, data sets and predictors to better understand problem and proposed solution.
In this thesis I will create my own data set, from Palestine Polytechnic University, then biuld a prediction model using SVM algorithm under two kernels: RBF and Linear kernels that predicts students performance after five years of engineering study, and this model will be enhanced using two methods: grid search and GA.

Chapter 3

Applied models

This chapter represented as five sections: first section shows how data set is collected, second section talks about pre processing of collected data, third section present how prediction model is build, then fourth section shows how built model were tested, last section shows how this model optimized. Figure 3.1 shows General block diagram of prediction model in this thesis.

First step in this thesis is to collect data, the aim is to improve academic performance of university students, so data mining concepts will help in this manner since data mining tasks like classification are useful in this case. Academic performance prediction will help students to improve their learning ways and lecturer to focus on poor students before being fail.
Engineering students in PPU were chosen to be part of university students to apply research on them, so personal, pre-university and university data about engineering students collected from PPU data base.
Next step is to do some pre processing on collected data in order to be useful and helpful in prediction, so this aims to correct the problems in data before build prediction model. Many problems may be in collected data like missing values, attributes or different ranges, all these problems cause prediction model to be not accurate.
Third step is to build model, this prediction model build using R language since developer of this programming language built great packages that make work easier. Theses packages are optimized well, and make work simple since they handle many exceptions, calling a function with right parameters is just what needed to apply many data mining algorithms simply.
Model testing step important to check whether prediction model accurate as it needed or not. Testing in this research done in two ways one by slice data set into training and

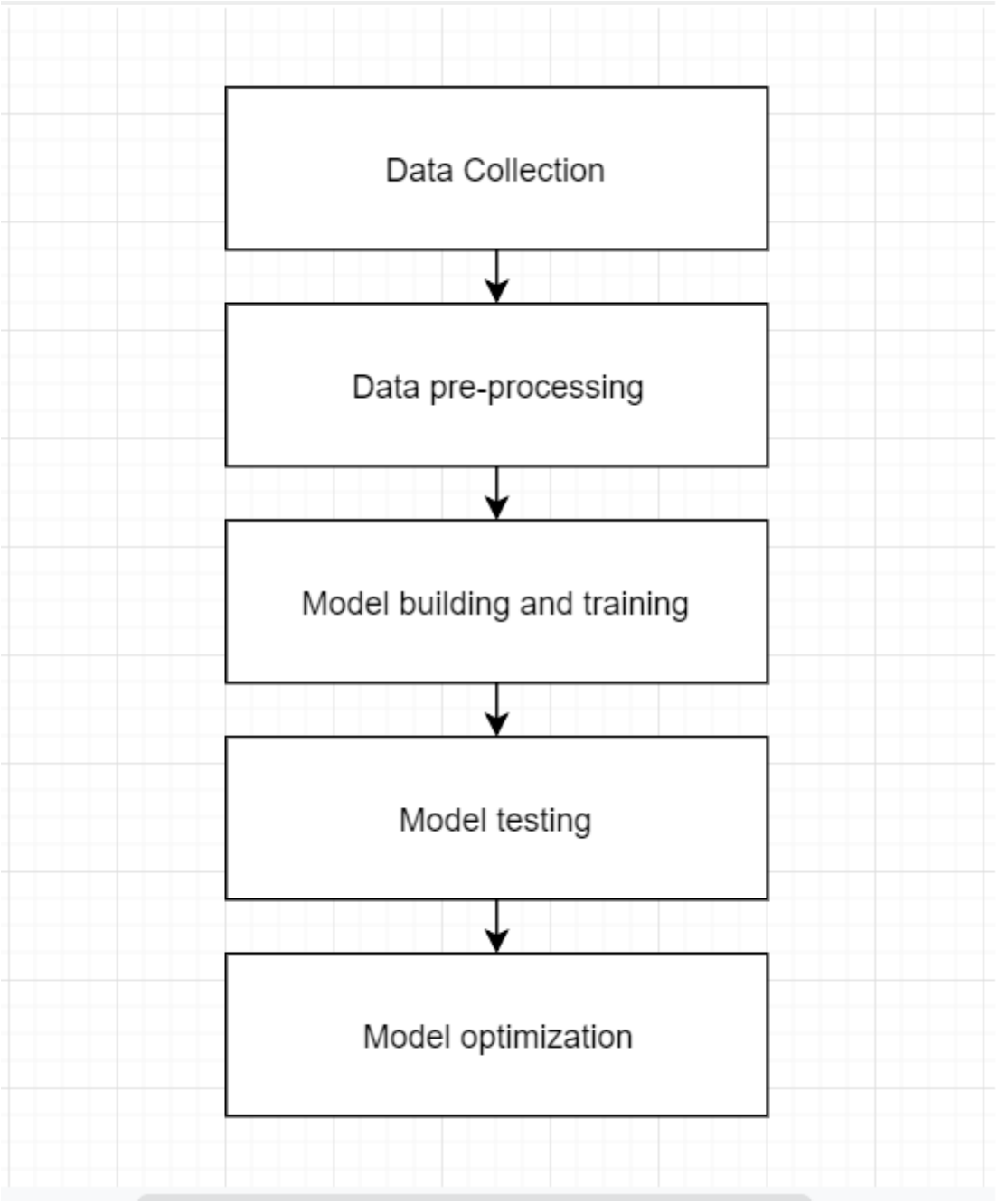

FIGURE 3.1: General block diagram of prediction model.

testing sets, and other one by divide data set into tow parts randomly and test model with randomly 30% data set as blind data.

Last step is a try to optimize model, in order to achieve more accurate model, optimization using grid search space and with apply some evolutionary algorithms like genetic algorithm.

3.1 Data collection

Engineering discipline is a critical study where scientific knowledge and mathematics is used to develop the whole range of industry, science, medicine and many other specializations. Every single day new developments are taking place, so its important to have strong students in this discipline.
Knowledge Discovery in Databases is a field of study that extract knowledge from databases. KDD and data miming used to predict engineering academic performance at Palestine Polytechnic University ho help students and instructors. Depending on prediction model instructors can improve learning of low performance students who earn less than 65 out of 100 in their GPA. Low performance students also encouraged to develop a better learning strategy depending on prediction model.
To build a prediction model, data must be collected to establish training data set for model. So data set called Engineering_df collected from PPU data base to be used in building prediction model for academic performance for engineering students at PPU.
Collected data can be divided into three types: personal data, pre university and university data. Personal data about students like birth place, birth date and pre university data like the place of tawjihi certificate, GPA. University data like students score in calculus, physics, year of starting college,... etc.

This data set is a sample of 8291 sample, and 22 attributes which are shown table 3.1.

The last attribute is the target classes for predicting model which are 'A', 'B', 'C', 'D', 'E', 'F', 'O'.

3.2 Data pre-processing

This section explained the stages that information passed until it is ready for use. The data is from one source, registration department at PPU saved as a csv file.Coding stage is begin for available values. First attribute is the country where student has his high school certificate, each country is mapped to a code number as shown in table 3.2.

For General Secondary branch, table 3.3 shows a code for each branch. Now, for the next attribute specialization code at university, table 3.4 shows the corresponding code for each engineering discipline at PPU.
Regularity in engineering study at PPU take an attribute in this data set, regular student is one that still studying, irregular is the one who is graduated or who was dismissed from the study, the last case is one who left studying. Table 3.5 shows the corresponding

Table 3.1: Table of features (data set attributes).

	Attribute
1	The country from which the student obtained a high school certificate.
2	Year of high school diploma.
3	General Secondary Branch (Scientific, Industrial, Air Conditioning, ...).
4	High school average (GPA)
5	Specialization code at university.
6	Specialization name.
7	Number of times the student received an academic warning.
8	University GPA.
9	Regularity of students in the study (regular, irregular, leaving study).
10	Has the student been separated from university.
11	Did student graduate from university.
12	The town or city where student lives.
13	Students mark in Physics 1.
14	Students mark in Physics 2.
15	Students mark in Calculus 1.
16	Students mark in Calculus 2.
17	Students mark in English 1.
18	Students mark in English 2.
19	The year when student start studying at PPU.
20	Number of semesters until student graduated.
21	Number of honor semesters.
22	Class of GPA.

Table 3.2: Table of countries

Code	**Country**
1	Palestine
2	Qatar
3	Kingdom of Saudi Arabia
4	Jordan
5	United Arab Emirates
6	Egypt
7	Oman
8	England
9	Sudan
10	Kuwait
11	Syria

code for each case of students. Dismissed students takes code number 1, others takes code number 2 as shown in table 3.6. Table 3.7 shows if student has been graduated from engineering college or not.

Before last attribute is town or city where student lives. Most of PPU students are from Hebron city and its villages so each village and each town in hebron takes a unique

Table 3.3: Table of secondary school branches

Code	Specialization
1	Scientific
2	Motor cars
3	Electricity works
4	Turning and settling
5	Industrial Electronics
6	Computer Maintenance
7	Conditioning
8	Heating and sanitary ware
9	Office machines
10	Radio and TV
11	Industrial
12	Industrial Electronics
13	Building and Area
14	Communication
15	Industrial - professional
16	Electricity of cars
17	Carpentry
18	Building, Tower and Armament
19	Blacksmith and Welding

Table 3.4: Table of Engineering Specialization

Code	Engineering Specialization
1	Building Engineering
2	General Engineering
3	Mechanical Engineering Branch of Mechanics Engineering
4	Civil Engineering Branch of Geometry Engineering
5	Mechanical Engineering Branch
6	Electrical Engineering Branch Engineering Industrial Automation
7	Electrical Engineering Branch Engineering of medical devices
8	Mechanical Engineering Branch HVAC Engineering
9	General Engineering - Preparatory Year
10	Electrical Engineering Branch Engineering Industrial Automation
11	Architectural Engineering
12	Electronics Engineering
13	Electrical Engineering Department
14	Department of Civil and Architectural Engineering
15	Mechanical Engineering Department
16	Electrical Engineering Branch Electrical Engineering Technology
17	Environmental Technology Engineering
18	Civil Engineering and Infrastructure
19	Renewable Energy Engineering

code as shown in table 3.8, but any other palestinian cities like Ramallah, Bethlehem, Jenin...etc takes a code for city and its villages, all Gaza strip takes a code as shown in

Table 3.5: Table of regularity in engineering study

Code	regularity in engineering study
1	Irregular
2	Left University
3	Regular

Table 3.6: Table of Dismissed cases

Code	Dismissed cases
1	Dismissed Student
2	Other

Table 3.7: Table of Graduated cases

Code	Graduated cases
1	Graduated Student
2	Other

table 3.9.

Table 3.8: Table of Hebron City

Code	Hebron city or village	Code	Hebron city or village
1	Halhul	2	Hebron City
4	Al-thahria	6	Se'er
8	Tarqumia	9	Al-aroub camp
10	Al-samou'	11	Al-fawwar camp
12	Dura	15	Beit Ommar
16	Beit Kahel	17	Beit Oula
19	Yatta	21	Bani N'iem
22	Idna	24	Sourief
25	Beit Awwa	29	Der Samet
30	Nouba	31	Al-shioukh
32	Raboud	33	Kharas
34	Tafouh	35	Al-berj
36	Khorsa	37	Karma
38	Al-sekka	41	Al-majd
42	Al-tabaqa	43	Al-rehia
44	Beit el-roush	46	Mo'skar
47	El-koum	52	Mreish
53	Beit Marsam		

Last attribute is the target one, each student has a commutative average where each category include some range as shown in table 3.10. Class 'O' means that student did not finish his engineering study for any reason. For attributes of marks of Physics 1 and

Table 3.9: Table of Cities

Code	City
3	Gaza strip
5	Jerusalem
7	Jenin
13	Nablus
14	Bethlehem
18	Qalqilia
20	Tulkarem
26	Salfit
27	Ramallah
50	Haifa
48	Beer Sheva
49	Rahat
40	Jericho
51	Nazareth
45	Jordan

2, English 1 and 2, Calculus 1 and 2, and Arabic language each mark has the value W, T, R1, R2, R3, R4, WF, or Not available takes the code of -1.

Table 3.10: Table of cumulative average classes

Code	Range
A	>= 90
B	>=85
C	>=78
D	>=71
E	>=68
F	>=65
O	other

3.3 Model building and training

After preparing data set, building our prediction model will start, R language is selected since it simple, and has built in algorithms like SVM and packages that used for large data sets.

Support vector machines (SVMs) is a binary classification algorithm, but many approaches used to solve multi class problems by using a combination of several binary classifiers. Best approach is one vs. all [33]

Figure 3.2 shows detailed block diagram for prediction model.

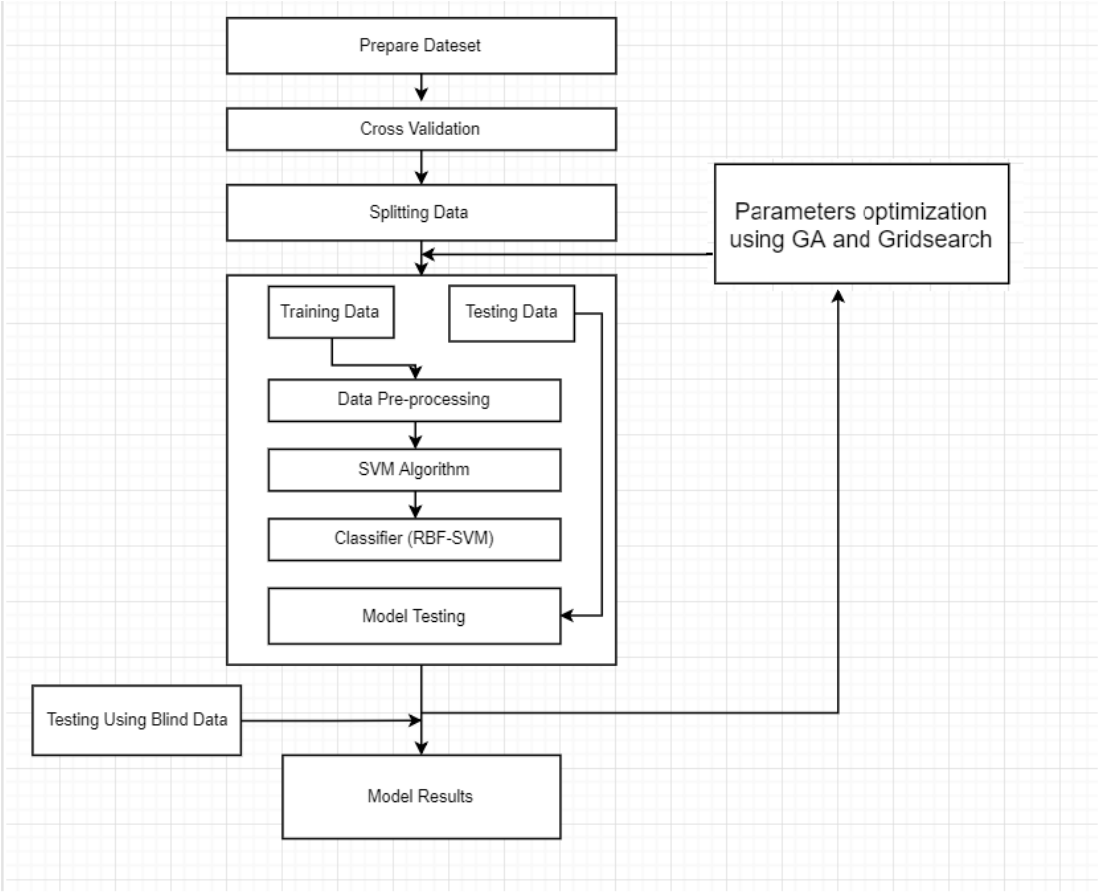

FIGURE 3.2: Block diagram of prediction model.

To implement SVM classifier algorithm in R programming language, 'caret' or 'e1071' packages are used for large data sets, this package also is well optimized and can handle exceptions. The main goal of thesis is to predict engineering students performance then to call functions using sutable parameters to each kernel (linear and RBF)
SVM classifier algorithm aims to construct separating hyper plane between data belongs to different classes and labels. Here 'caret' package is used and first step is to install caret package with all dependencies on Rstudio which mentioned at point 1 in pseudo code algorithm.

Algorithm 1 Create Support Vector Machine Classifier Pseudo Code

Input: Engineering_df data set

Output: Classification model.

Data: Training set,Testing set

1: Caret Package Installation.
2: Data Import
3: Data Slicing
4: Preprocessing Training
5: Training the SVM model using linear, RBF kernels.
6: Trained SVM model result
7: Test Set Prediction
8: How Accurately prediction model is working
9: RBF-SVM model enhancement.

Install "caret" library and its dependence's also. This package provides direct access to many functions that used in train and test prediction model. Now our data set contains 23 attributes as explains in data collection section, first 22 variables all are numeric values used to predict last attribute which is character and its our target variable.
Next step is just to import caret package. After read the CSV (comma separated values) and called "Engineering_df", now data slicing to split data set into training and testing parts, here training data set is 70% called "intrain"and test set is 30% of original data set. CreateDtataPartition is a built in method in caret package, used to split data.

To avoid overfitting, cross validation is used. Learning the parameters of a prediction function and testing it on the same data which is here Engineerig_df data set is a methodological mistake since a model that would just repeat the labels of the samples that it has just seen would have a perfect accuracy but would fail to predict anything or any sample but learning samples on unseen or blind data. This situation is called overfitting, to avoid it, it is common practice when performing a machine learning experiment to hold out part of the available data as a test set X1_test, X2_test. Figure 3.3 shows a flowchart of typical cross validation workflow in model training. The best parameters can be determined by grid search techniques [34].

CreateDataPartition() function has three arguments:

- y: parameter takes the value of variable that data partitioned according to it, here variable 23 is passed in partitioning method.

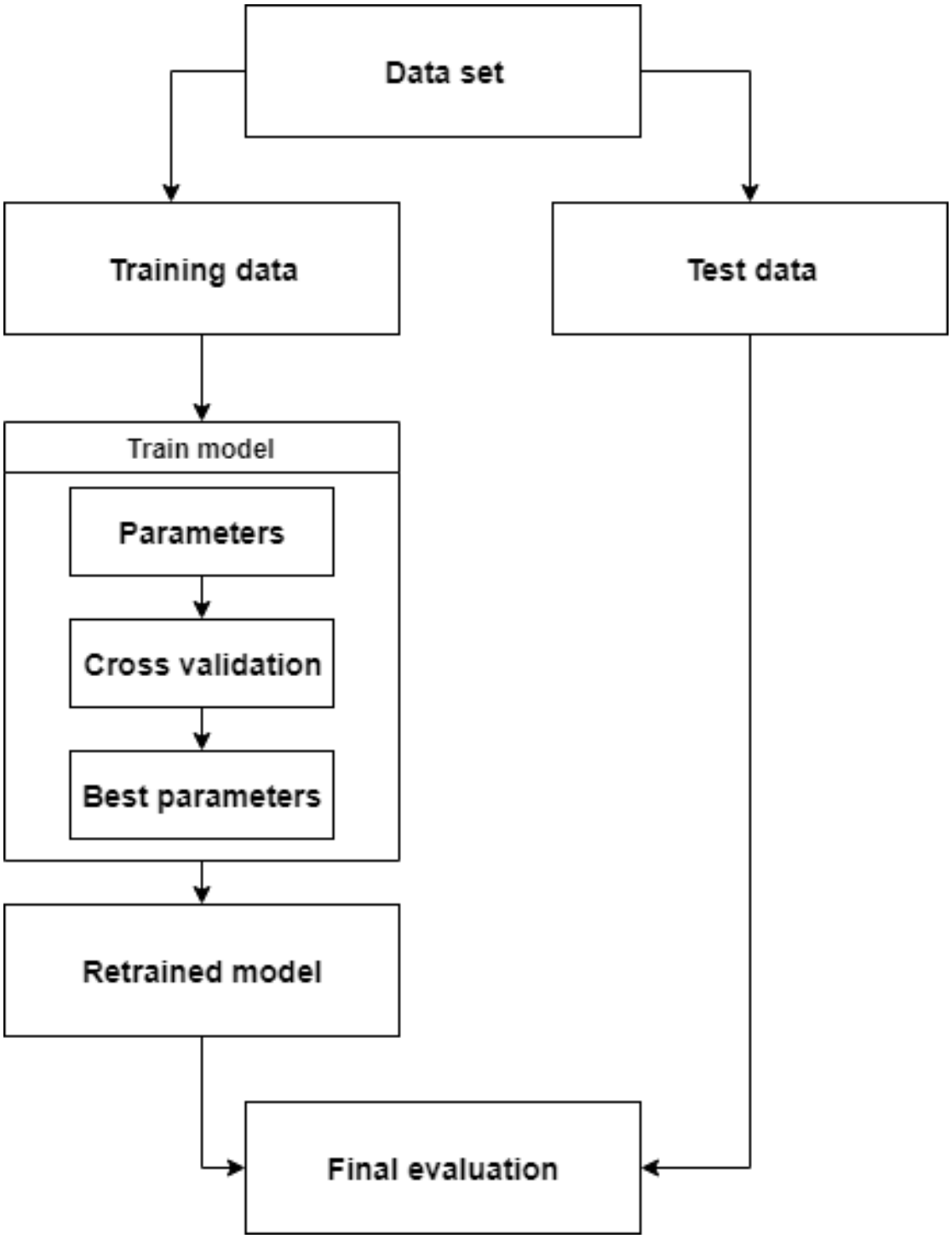

FIGURE 3.3: Cross validation flowchart.

- p: this parameter takes a decimal value from 0 to 1, here 0.7 means training part is 70% of Engineering_df and the rest 30% used for testing data.
- list: this parameter is for return value, list or matrix, in our model it list argument assigned to be 'FALSE', which mean not return list but return a matrix.

Figure 3.4 shows how a 10-folds cross validation data set is sliced at each repeated t me, here in this research this criteria repeated 3 times as shown in R code, trainControl function.

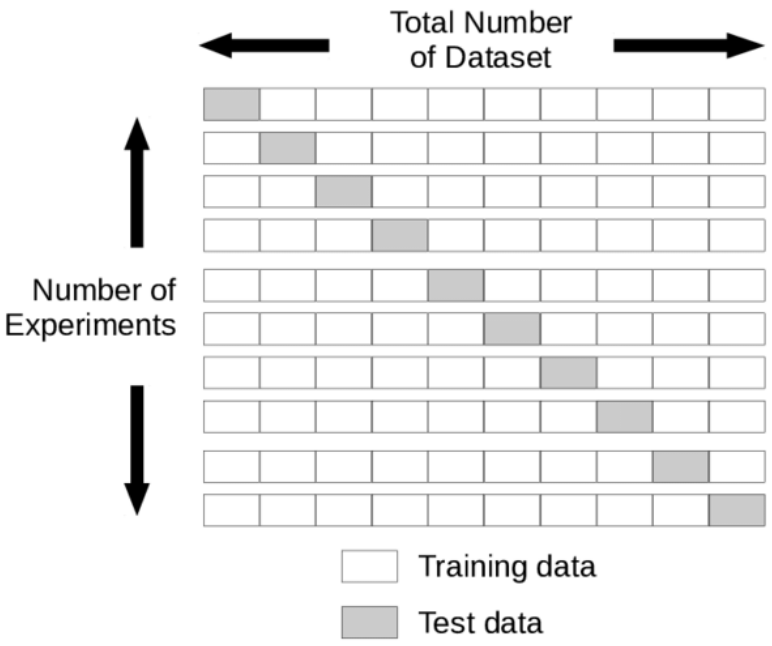

FIGURE 3.4: Example of 10-fold Cross-validation [2].

After slice data set into training and testing sets, caret provides a train control method that control next train method, by passing multiple arguments explained as following:

- method: is re sampling method for training and testing splits.
- number: number of folds or number of re sampling iterations, here it is number of folds.
- repeats:repeated k-folds for cross validation [35].

Now to train SVM classifier, train method used by passing multiple arguments:

- target value: which is here attribute number 23.
- data: data set used to train our model, its trainig data set here.

- method:method used to train SVM classifier, in this thesis two methods used, one 'svmLinear' and 'svmRadial'.
- trControl: take results from previous traincontrol method.
- preprocess: for pre processing training data, "center" and "scale" used for centering and scaling our data.
- tuneLength: an integer value for tuning algorithm [36].

3.4 Model testing

After training model using training data set, its time to test build model using testing data set. Now, our model ready to predict classes of testing data set, depending on obtained svm_Radial model, using new data which is testing data set. Final step is to find how accurate our model, confusion matrix prints many statistics about model like Accuracy, Sensitivity, Specificity, Positive Prediction Value, Negative Prediction Value, Prevalence, Detection Rate, Detection Prevalence, Balanced Accuracy.
Model modified many times until it takes its final state, the following information and figures of model obtained using RBF kernel, the most early state was when the data set Engineering_df has A, B, C, D, E and F each class explained in section 3.2, so all student even they graduated or not follow the same criteria when they labeled, just depending on their accumulative average, not on their status graduated or not.
Statistics and accuracy presented in chapter four, theses numerical values extracted from confusion matrix function in R language that used to build this model, many statistics calculated automatically like sensitivity, specificity positive prediction value, prevalence, balanced accuracy for each class and accuracy for model. All these phrases explained in chapter four, model is highly accurate predictor for class F with 97% where class B balanced accuracy is 83%. The best values of SVM parameters under RBF kernel or tuning parameters C and sigma chosen automatically by train() method, the final values of C that gives largest accuracy was 8, and sigma was 0.04662.
10 folds cross validation method is used and repeated three times, sample sizes were 5224, 5227, 5227, 5224, 5225, 5226, ... etc. Re sampling results across tuning parameters. Values of parameter C, accuracy of model for each value of automatically selected C, and Kappa value shown in table 3.11.

Next when samples of data in Engineering_df manipulated to have a new label, 'O' label or class this class for student who has not graduated for many reasons, may be they left

Table 3.11: Accuracy of prediction model before 'O' class.

C	Accuracy	Kappa
0.25	0.773	0.647
0.50	0.816	0.741
1.00	0.856	0.801
2.00	0.876	0.829
4.00	0.888	0.847
8.00	0.894	0.856
16.00	0.893	0.854
32.00	0.889	0.849
64.00	0.886	0.845
128	0.882	0.839

study, of dismissed from university, or they still study engineering until now.

The last step in our model, is testing using blind data in order to be sure that data did not make over fitting model hypothesis. Engineering_df data set sliced into two smaller data sets randomly in a ratio of 70:30 for training and testing data sets respectively. Random() function used for slicing, then model trained using training data set, then tested using testing data set, table 3.4 shows confusion matrix for testing model using blind data, diagonal shows correct classified samples, where other values are mis classified samples.

Table 3.12: Confusion matrix for optimized RBF-SVM using blind testing data.

		Reference						
		A	B	C	D	E	F	O
	A	13	5	0	0	0	0	0
	B	1	36	3	0	0	0	0
	C	0	9	231	15	1	0	0
Predicted	D	0	0	21	578	28	0	1
	E	0	0	0	14	91	11	0
	F	0	0	0	0	0	0	0
	O	0	0	0	0	0	0	1235

SVM model trained using linear kernel also to determine differences between RBF and linear kernels for our data set, table 3.13 shows confusion matrix for Linear-SVM model, where diagonal numbers are number or samples correctly classified, and other numbers are mis classified objects.

TABLE 3.13: Confusion matrix for SVM model using linear kernel data.

		Reference						
		A	B	C	D	E	F	O
	A	10	0	0	0	0	0	0
	B	1	32	7	0	0	0	0
	C	0	7	178	9	0	0	0
Predicted	D	0	0	8	434	22	0	0
	E	0	0	0	10	66	5	0
	F	0	0	0	0	0	0	0
	O	0	0	0	0	0	0	1007

3.5 Model optimization

There are many ways to optimize SVM parameters for different kernels, the case of RBF kernel there are two parameters, C and sigma. In this thesis two ways used to optimize values of C and sigma first is manually using grid search method, second using one of Evolutionary Algorithms like Genetic Algorithm. Figure 3.5 and 3.6 show block diagrams for model optimization using grid search and GA.

What is Grid search. Grid search is a method used for optimization problems [37], that builds a model for combinations of different parameters, so its a process that make a scanning for each combination to find the optimal and more accurate one. Grid search build and store a model for each combination. Just like a grid, x-axis is value of C parameter, y-axis is value of sigma parameter, each cross point is a combination from one value of C and another value of sigma. Grid search train RBF-SVM model using this parameters combination, then store this model, and check the accuracy of model using another combination. Finally grid search decide which parameters combination gives the most accurate model. [38]

Figure 3.5 shows the steps of optimizing RBF-SVM model parameters c and sigma, first step is to determine minimum and maximum values of parameters c and sigma and number of steps to jump between one value and next to make combinations between these two parameters and find which combination gives best accuracy.
Then R will build a space of combinations of c and sigma to check the best couple, training step comes next to find most accurate parameters to be used in testing phase, finally model accuracy calculated using confusion matrix and depending on obtained combination of c and sigma.

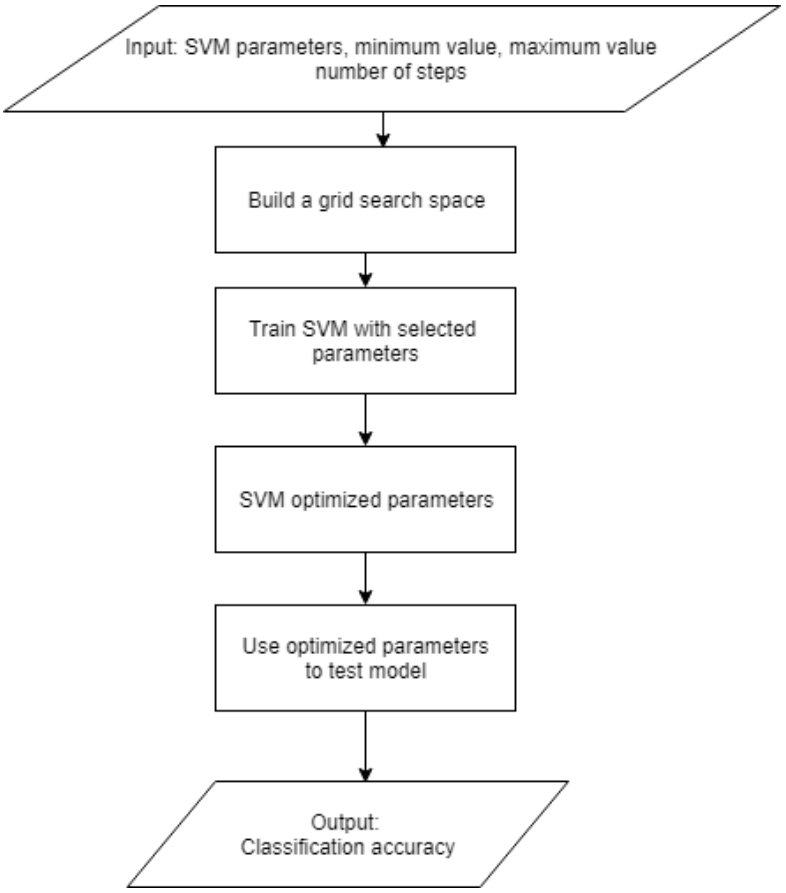

FIGURE 3.5: SVM parameters optimization using grid search.

For grid search a new argument in train function used to search for the most suitable pairs of C and sigma (gamma), that gives best accuracy.
First when Engineering_df has A, B, C, D, E, F classes the model accuracy was 90%. Grid search used to optimize RBF parameters, confusion matrix of grid search model shown in table 3.14, optimization using grid search method gives accuracy of 91%. The best C and sigma values obtained when model using grid search was C=9.9 and sigma=0.01.

TABLE 3.14: Confusion matrix for RBF-SVM using grid search.

		Reference					
		A	B	C	D	E	F
	A	30	8	0	0	0	0
	B	8	68	0	0	0	0
	C	0	15	395	22	0	0
Predicted	D	0	0	21	927	50	1
	E	0	0	0	30	259	19
	F	0	0	0	0	25	600

Second way used to optimize model is using GA, **What is Genetic Algorithm.** Genetic algorithm is a population based algorithm and like other evolutionary algorithms. figure 3.6 shows detailed flow chart of genetic algorithm steps in each generation.

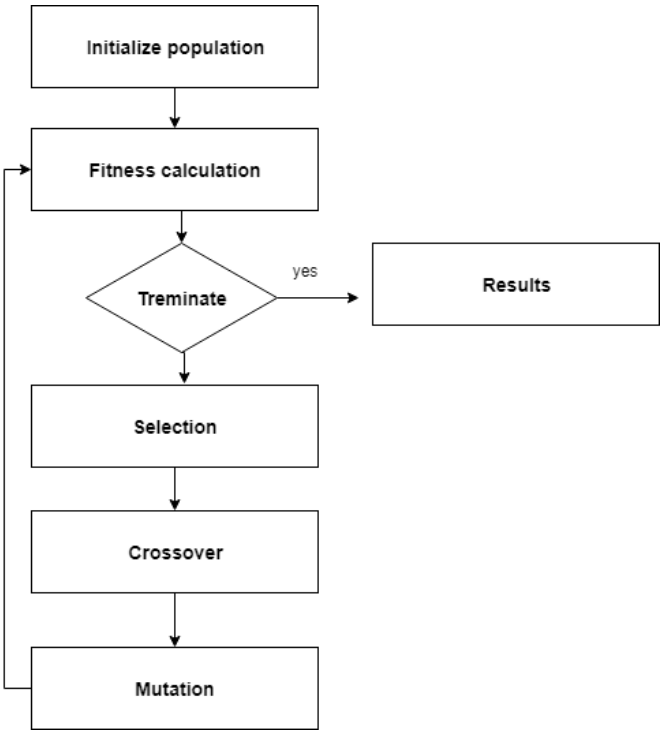

FIGURE 3.6: Genetic algorithm steps.

Each step in GA has three main rules in order to create next populations generation from recent population, are the following [39]:

1. Selection, its the rule of selecting parents.
2. Crossover, its combining two genes from two parents to create children's of next generation.
3. Mutation, some changes on recent parents, to contribute children at next generation.

In this thesis, first population were 50, each parent has two genes, numerical genes, first gene is parameter C and second one is parameter sigma. Our fitness function is SVM algorithm under RBF kernel. Figure 3.7 shows how C and sigma were optimizes using genetic algorithm.

First step in to determine minimum and maximum values of SVM-RBF parameters, then to apply genetic algorithm in order to train prediction model. Fitness function

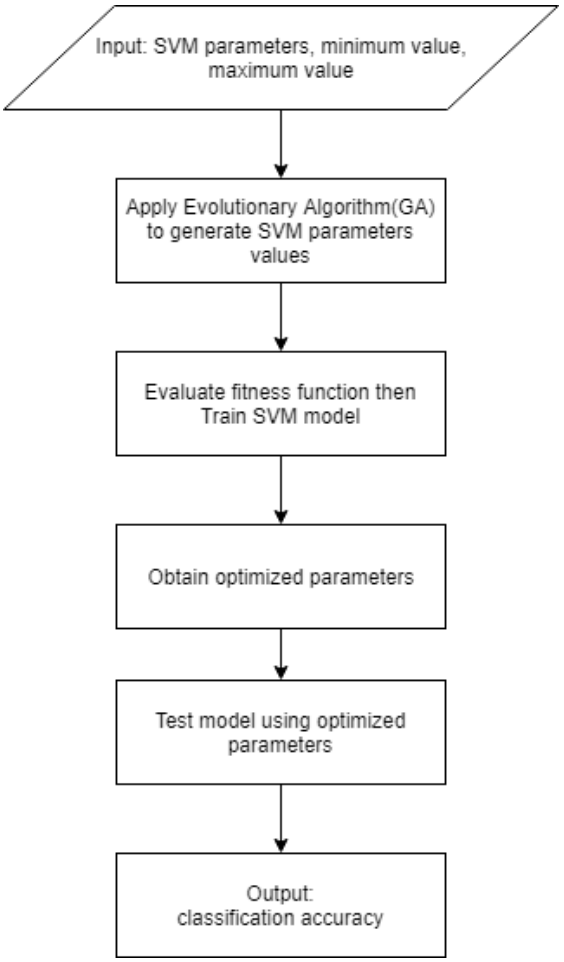

FIGURE 3.7: Parameters optimization using evolutionary algorithm (GA) .

comes next to train model and obtain optimized values of C and sigma. Last step is to determine accuracy of model using optimized parameters.

When data set has 7 classes A, B, C, D, E, F, O model accuracy was 96% as shown previous, GA used and gives accuracy of 96% also.
Block diagram of parameters optimization using GA shows steps using in prediction model to optimize values of c ans sigma using evolutionary algorithms, first step is to determine minimum and maximum values of SVM parameter c and sigma, then to apply GA to generate RBF-SVM parameters, GA fitness function is training function using RBF kernel, this function used to train model using support vector machine under RBF kernel, finally to test model using obtained parameters and find accuracy of optimized model.
Real-valued GA used, with population size of 50, 10 generations, cross over probability is 0.8, and mutation probability is 0.1, minimum values of c and gamma are 0, and maximum values are 50. Optimized value of c is 36.89 and gamma 49.6, results taken after 2.64 days of training the optimization model.

Confusion matrix for optimization model using GA is shown in table 3.15 where diagonal shows number of correctly classified objects, and other values are mis-classified.

Table 3.15: Confusion matrix for optimized RBF-SVM using GA.

		Reference						
		A	B	C	D	E	F	O
	A	14	3	0	0	0	0	0
	B	1	47	5	0	0	0	0
	C	0	4	247	8	0	0	0
Predicted	D	0	0	18	608	28	0	0
	E	0	0	0	19	100	5	0
	F	0	0	0	0	2	2	0
	O	0	0	0	0	0	0	1378

Chapter 4

Experiments and results

Experiments and results in this chapter obtained using Rstudio and free server called Kaggle. Rstudio is an open source software program, version 1.1.453, Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/538.1 (KHTML, like Gecko) rstudio Safari/538.1 Qt/5.4.1. Kaggle is an online community, owned by GOOGLE LLC, give free sessions for 6 hours for each session.
This chapter also answers the question of "how to measure the performance of classifier". The answer is to find accuracy of your classifier, accuracy means number of examples the correctly classified in ration of all samples. Classifier model takes test data set and find class for each sample in testing data set, so for each sample classifier guess whether the class is 0 or 1, then finds how many right decisions classifier makes and divide on total number of test samples in testing data set.
So classifier tries to put hypotheses and guess for each example whether its Yes (0) or No (1) and each decision can be one of the following choices:

1. True positive: positive sample classified as positive.
2. False negative: positive sample miss classified as negative.
3. True negative: negative sample classified as negative.
4. False positive: negative sample missclassified as positive.

These types of decisions can shown as 2x2 matrix where columns are true or actual classes where rows are predicted or hypothesized classes [40].
This is called also confusion matrix or contingency table. Entire cells contains number of correctly and in correctly classified samples, accuracy could calculated simply by divide sum of diagonal number on total of all matrix numbers.

$$Accuracy = \frac{Truepositive + Truenegative}{Truepositive + Truenegative + Falsepositive + Falsenegative} \quad (4.1)$$

Prediction outcome

		p	**n**	**total**
actual value	**p′**	True Positive	False Negative	P′
	n′	False Positive	True Negative	N′
	total	P	N	

More then accuracy, many statistics can be calculated from contingency table, the following are some metrics:

True positive rate.

$$Truepositiverate = TP/\left(TP + FN\right) = 1 - FalseNegativeRate. \quad (4.2)$$

False positive rate

$$Falsepositiverate = FP/\left(FP + TN\right) = 1 - TrueNegativeRate. \quad (4.3)$$

Sensitivity= True Positive Rate.
Specificity= True Negative Rate.

Positive predictive value.

$$Positivepredictivevalue = TP/\left(TP + FP\right). \quad (4.4)$$

Recall.

$$Recall = TP/\left(TP + FN\right) = TruePositiveRate. \quad (4.5)$$

Precision.

$$Precision = TP/\left(TP + FP\right). \quad (4.6)$$

F-score: is composite measure which that benefits algorithms with higher sensitivity and challenges algorithms with higher specificity[41].

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall}. \tag{4.7}$$

G-score: is geometric mean of precision and recall.

$$G = \sqrt{precision + recall}. \tag{4.8}$$

In this research confusion matrix is used to compute accuracy and other statistics like sensitivity, specificity, prevalence,...etc. Confusion matrix presents a summary for prediction results on a classification model, it gives errors done by model and what type of errors also. In R language confusion matrix is built in function just need to call it after classification model testing to show accuracy of model.
Another way to measure how accurate our prediction model is area under curve (AUC) or ROC. "A representation and interpretation of the area under a receiver operating characteristic (ROC) curve obtained by the "rating" method, or by mathematical predictions based on patient characteristics, is presented." [42]
The area under the receiver operating characteristic (ROC) curve, known as the AUC, considered to be the standard method to assess the accuracy of classification models. It avoids the supposed subjectivity in the threshold selection process [43].
AUC is, in general a better measure than accuracy for most of data mining algorithms. Many popular data mining algorithms should then be re-evaluated in terms of AUC. Experiments compared many algorithms like Naive Bayes, C4.5, C4.4, and SVM in both accuracy and AUC, results show that they have very a similar predictive accuracy. In addition, Naive Bayes, C4.4, and SVM produce similar AUC scores, and they all outperform C4.5 in AUC with significant difference. conclusions will provide important guidelines in data mining applications on real-world datasets [44].

This section is divided into three subsections, first section is: results of linear kernel SVM model that shows some results and statistics of our model using linear kernel under SVM classifier, second is: results of RBF kernel SVM model that shows results and accuracy of our built model using RBF kernel under SVM classifier, last section is: results of enhanced RBF kernel SVM model, using genetic algorithm and grid search method that shows how our model enhanced to be more accurate using two ways genetic algorithm and grid search method, last section is testing using blind data.

4.1 Results of linear kernel SVM model

Linear SVM is a fast machine learning and data mining algorithm for solving multi class classification problems from ultra large data sets that implements an original proprietary version of a cutting plane algorithm for designing a linear support vector machine.

Linear Kernel is used when the data is linearly separable. It can be separated using a single hyperplane or more for multi class problems. It is one of the most common kernels to be used. It is mostly used when there is a Large number of features in a particular data set. One of the examples where there are a lot of features, is text classification, as each alphabet is a new feature in this research we have 23 features also. So we mostly use Linear Kernel in Text Classification.
Advantages of using Linear Kernel:

1. Training a SVM with a Linear Kernel is Faster than with any other Kernel.
2. When training a SVM with a Linear Kernel, only the optimisation of the C (cost parameter).

Regularisation parameter is required. On the other hand, when training with other kernels, there is a need to optimise the λ parameter which means that performing a grid search will usually take more time.

For linear kernel under SVM model, Table 4.1 shows statistics and results of our prediction model about Linear-SVM that gives accuracy of 95% .Figure 4.1 shows AUC of this model equals to 95% also. Note that NAN means Not a Number, its equal to some number over zero.

4.2 Results of RBF kernel SVM model

First step done after the building model is testing, to measure accuracy of model. Here accuracy presented in three digits. First obtained accuracy and some statistics shown in table 4.2 for prediction model using RBF-SVM algorithm depending on data set without 'O' class for students who have not graduated from engineering college for some reasons, accuracy measured using confusion matrix is 90% where AUC is 90% also as shown in figure 4.2.

TABLE 4.1: Linear-SVM model statistics and accuracy.

	Class A	Class B	Class C	Class D	Class E	Class F	Class O
Sensitivity	0.900	0.820	0.922	0.958	0.750	0.000	1.000
Specificity	0.998	0.997	0.994	0.977	0.991	1.000	1.000
Pos Pred Value	1.000	0.800	0.917	0.935	0.814	NaN	1.000
Neg Pred Value	0.999	0.996	0.990	0.985	0.987	0.997	1.000
Prevalence	0.006	0.021	0.107	0.255	0.049	0.002	0.560
Detection Rate	0.005	0.017	0.099	0.241	0.036	0.000	0.560
Detection Prevalence	0.005	0.022	0.108	0.258	0.045	0.000	0.560
Balanced Accuracy	0.954	0.907	0.956	0.967	0.870	0.500	1.000

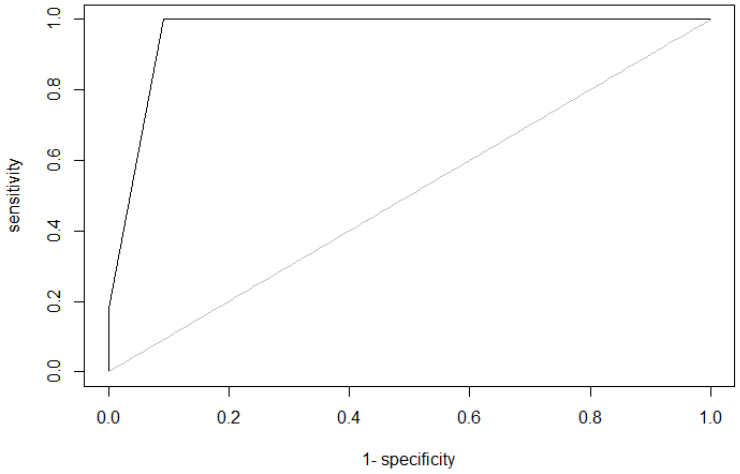

FIGURE 4.1: AUC for Linear-SVM model.

After testing our model depending on Engineering_df data set without "O" label, some changes done on our data set to have a new class, "O" label or class is an abbreviation for others. Others are students who did not had engineering degree from PPU for many reasons like being dismissed, or leaving college. Accuracy of RBF-SVM model depending on data set with 'O' class is 96% as shown in table 4.3 and AUC is 95% as shown in figure 4.3.

TABLE 4.2: Accuracy of prediction model before 'O' label.

	Class A	Class B	Class C	Class D	Class E	Class F
Sensitivity	0.789	0.681	0.914	0.935	0.739	0.959
Specificity	0.994	0.992	0.981	0.942	0.971	0.983
Pos Pred Value	0.697	0.775	0.910	0.913	0.799	0.952
Neg Pred Value	0.996	0.987	0.982	0.957	0.960	0.986
Prevalence	0.015	0.036	0.170	0.394	0.134	0.249
Detection Rate	0.012	0.024	0.155	0.368	0.099	0.239
Detection Prevalence	0.017	0.032	0.171	0.403	0.124	0.251
Balanced Accuracy	0.892	0.836	0.948	0.938	0.855	0.971

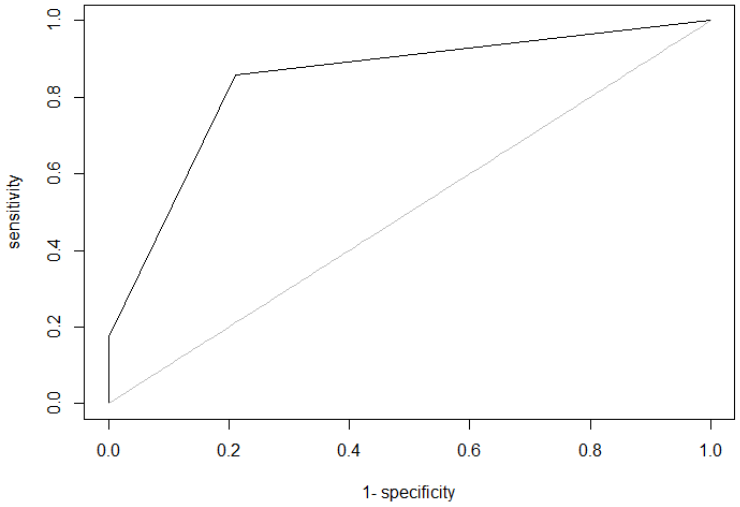

FIGURE 4.2: AUC for RBF-SVM model without 'O' label.)

4.3 Results of enhanced RBF kernel SVM model, using genetic algorithm and GridSearch method

After biulding RBF-SVM model, grid search method is used to optimize algorithm parameters C and sigma, accuracy depending on data set with 'O' class is 96% as shown in table 4.4.

AUC used to measure accuracy for model optimization using grid search method, AUC is 95% as shown in figure 4.4.

TABLE 4.3: RBF-SVM model statistics and accuracy using 'O' label data set.

	Class A	Class B	Class C	Class D	Class E	Class F	Class O
Sensitivity	0.909	0.846	0.896	0.940	0.727	0.000	0.999
Specificity	0.999	0.995	0.988	0.990	0.998	0.999	0.998
Pos Pred Value	0.909	0.804	0.905	0.922	0.790	0.000	0.999
Neg Pred Value	0.999	0.996	0.987	0.979	0.986	0.997	0.998
Prevalence	0.006	0.021	0.107	0.252	0.049	0.002	0.560
Detection Rate	0.005	0.018	0.096	0.237	0.045	0.001	0.560
Detection Prevalence	0.006	0.022	0.106	0.257	0.049	0.001	0.554
Balanced Accuracy	0.966	0.953	0.954	0.967	0.949	0.624	0.999

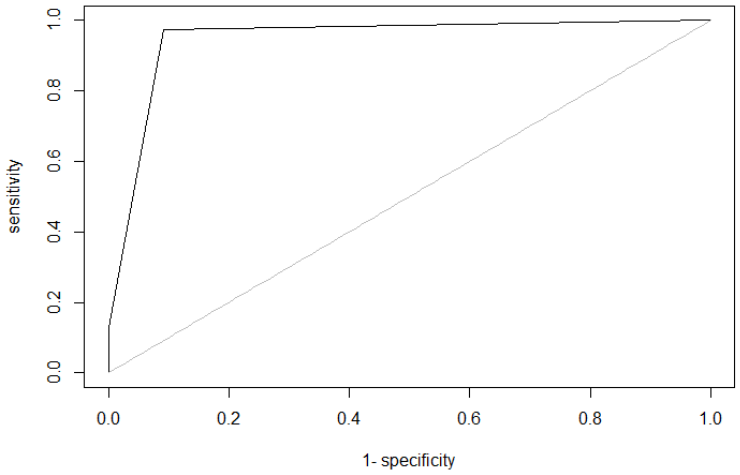

FIGURE 4.3: AUC for RBF-SVM model with 'O' label.)

Accuracy and statistics by class for model optimization using genetic algorithm shown in table 4.5, accuracy of this model is 96% and AUC is 95%.

4.4 Testing using blind data

To be sure that our model did not overfitt our data, a blind data data set is used to test RBF-SVM model. Table 4.6 shows accuracy and statistics by class obtained after test RBF-SVM model using blind data, accuracy obtained using blind data is 95%, where

TABLE 4.4: Enhanced RBF-SVM model statistics and accuracy using grid search.

	Class A	Class B	Class C	Class D	Class E	Class F
Sensitivity	0.789	0.747	0.933	0.946	0.775	0.967
Specificity	0.996	0.993	0.982	0.952	0.977	0.986
Pos Pred Value	0.789	0.819	0.914	0.927	0.840	0.960
Neg Pred Value	0.996	0.990	0.986	0.965	0.965	0.989
Prevalence	0.015	0.036	0.170	0.394	0.134	0.249
Detection Rate	0.012	0.027	0.159	0.373	0.104	0.241
Detection Prevalence	0.015	0.033	0.173	0.402	0.123	0.251
Balanced Accuracy	0.893	0.870	0.957	0.949	0.876	0.977

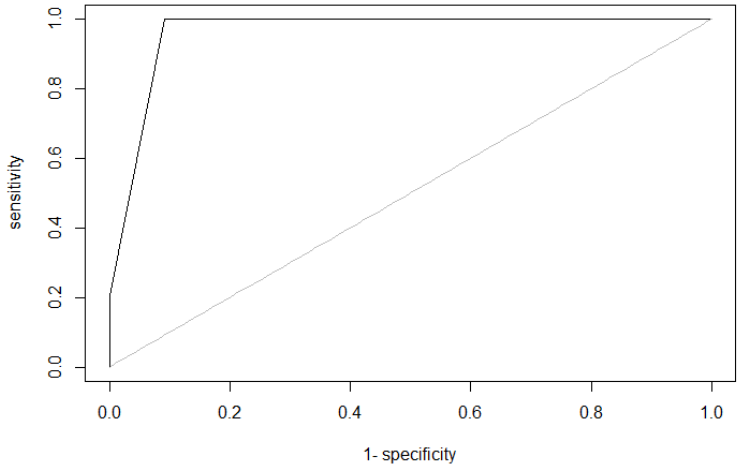

FIGURE 4.4: AUC for enhanced RBF-SVM model with grid search optimization.

model accuracy was 96%.

Random samples selected from testing data set, and find if this sample is a correct or in correct classified using our RBF-SVM model. As mentioned in table 3.6, predictors used in training and testing data set are: The country from which the student obtained a high school certificate, Year of high school diploma, General Secondary Branch (Scientific, Industrial, Air Conditioning, ...), High school average (GPA), Specialization code at university, Specialization name, Number of times the student received an academic warning, University GPA, Regularity of students in the study (regular, irregular, leaving study), Has the student been separated from university, Did student graduate from university, The town or city where student lives, Students mark in Physics 1 and 2, Students

Table 4.5: Enhanced RBF-SVM model statistics and accuracy using GA.

	Class A	Class B	Class C	Class D	Class E	Class F	Class O
Sensitivity	0.933	0.870	0.914	0.957	0.06	0.250	1.000
Specificity	0.998	0.997	0.994	0.977	0.989	0.999	1.000
Pos Pred Value	0.823	0.886	0.953	0.936	0.806	0.500	1.000
Neg Pred Value	0.999	0.997	0.989	0.985	0.989	0.997	1.000
Prevalence	0.006	0.021	0.108	0.255	0.049	0.003	0.554
Detection Rate	0.005	0.018	0.099	0.244	0.040	0.000	0.554
Detection Prevalence	0.006	0.021	0.104	0.261	0.049	0.001	0.554
Balanced Accuracy	0.966	0.933	0.954	0.967	0.898	0.624	1.000

Table 4.6: RBF-SVM model statistics and accuracy using blind testing data.

	Class A	Class B	Class C	Class D	Class E	Class F	Class O
Sensitivity	0.928	0.720	0.905	0.952	0.785	0.000	0.999
Specificity	0.997	0.998	0.987	0.970	0.988	1.000	1.000
Pos Pred Value	0.722	0.900	0.902	0.920	0.784	NaN	1.000
Neg Pred Value	0.999	0.993	0.988	0.982	0.986	0.995	0.999
Prevalence	0.006	0.021	0.111	0.264	0.052	0.004	0.539
Detection Rate	0.005	0.015	0.100	0.252	0.039	0.000	0.538
Detection Prevalence	0.007	0.017	0.111	0.273	0.050	0.000	0.538
Balanced Accuracy	0.963	0.859	0.946	0.961	0.873	0.500	0.999

mark in Calculus 1 and 2, Students mark in English 1 and 2 and Arabic language, The year when student start studying at PPU, Number of semesters until student graduated, Number of honor semesters. First, some miss classified cases listed below:

1. 1, 2010, 1, 72, 20401, 1, 0, 70, 1, 2, 1, 14, 61, -1, 65, -1, -1, -1, -1, 2012, 13, 0, E.
2. 1, 2010, 1, 98, 20401, 1, 0, 91, 1, 2, 1, 8, 90, 90, 93, 84, 88, 85, 85, 2010, 13, 7, A.

Case number one, is a miss classified sample using RBF-SVM model, this sample is an 'E' class GPA, but classified as 'F', this student sample from Bethlehem, its high school certificate is from Palestine, taken in 2010, with scientific branch and 72 GPA. Samples specialization at university is Building Engineering with 20401 code, received no warnings, also no honor semesters, graduated with 70 GPA and -1, 61, -1, 65, -1, -1

marks at physics 1,physics 2, calculus 1 ,calculus 2, English 1 and English 2 respectively. Model was wrong in its prediction, may be the reason behind that are marks of this sample in physics 1 , calculus 1 around 60's its an 'F' level, also marks of physics 2, calculus 2, English language 1 and 2 and Arabic language are -1 and these definitely 'Failed Case'.

Case number two is also a missclassified sample using RBF-SVM model. This sample is an 'A' class GPA, but classified as 'B', this student sample from Nablus, its high school certificate is from Palestine, taken in 2010, with scientific branch and 98 GPA. Samples specialization at university is Building Engineering with 20401 code, received no warnings, seven honor semesters, graduated with 91 GPA and 90, 90, 93, 84, 88, 85, 85 marks at physics 1,physics 2, calculus 1 ,calculus 2, English 1 and English 2 and Arabic language respectively.

Model was wrong in its prediction, may be the reason behind that are marks of this sample in calculus 2, English language 1 and 2 and Arabic language around 80's its an 'B' level so model predict the wrong class.

Second, some correctly classified cases listed below:

1. 1, 2012, 1, 97, 20201, 5, 0, 90, 1, 2, 1, 7, 98, 88, 94, 96, 98, 98, 92, 2012, 13, 7, A.
2. 1, 2012, 1, 98, 20403, 11, 0, 86, 2, 2, 2, 2, 82, -1, 94, 82, 82, 71, 84, 2012, 2, 1, O.

Case number one, is a correctly classified sample using RBF-SVM model, this sample is an 'A' class GPA, and classified as 'A', this student sample from Jenin, its high school certificate is from Palestine, taken in 2012, with scientific branch and 97 GPA. Samples specialization at university is Mechanical Engineering Branch with 20201 code, received no warnings, seven honor semesters, graduated with 90 GPA and 98, 88, 94, 96, 98, 98, 92 marks at physics 1,physics 2, calculus 1 ,calculus 2, English 1 and English 2 respectively. Model was correct in its prediction, may be the reason behind that are marks high marks of this sample, most marks around 90's and this is an 'A' class, also there is some honor semester its 7 semesters here.

Case number two, is a correctly classified sample using RBF-SVM model, this sample is an 'O' class GPA, and classified as 'O', this student sample from Hebron city, its high school certificate is from Palestine, taken in 2012, with scientific branch and 98 GPA. Samples specialization at university is Architectural Engineering Branch with 20403 code, received no warnings, one honor semesters, its GPA is 86 and 82, -1, 94, 82, 82, 71, 84 marks at physics 1,physics 2, calculus 1 ,calculus 2, English 1, English 2 and Arabic language respectively.

Model was correct in its prediction, may be the reason behind that are regularity in

engineering study code is 2 which is for student who left university, also for graduation attribute code is 2, this code means that student did not graduate from engineering discipline and so its an 'O' class.
Experiments and Results chapter presents how build model is tested and optimized using grid search and Genetic Algorithm for SVM-RBF and linear models. Accuracy and Area Under Curve calculated and plotted for each case and for a blind testing data set. Finally some students samples are presented and discussed as correctly and miss classified cases and reasons behind each manner of classification.

4.5 Results discussion

From previous experiments and results, we can observe the SVM classifiers gives an excellent prediction accuracy, RBF kernel more accurate than linear kernel since high dimensional dataset, or data set with large number of attributes will be trained well using RBF kernel more than Linear kernel. Optimization of RBF-SVM prediction model using grid search and genetic algorithm gives nearly the same accuracy and AUC.

Chapter 5

Conclusion and future work

5.1 Conclusion

This work proposes a prediction model for engineering performance after five years of study at Palestine Polytechnic University, this model built depending on a data set collected from registration department at Palestine Polytechnic University about 8291 students who applied for enrollment to engineering college from 1994 to 2017.
Two models were build to predict how will students perform or students GPA after five years of engineering study, using support vector machines, two kernels RBF and Linear were used under SVM classifier, then two models RBF-SVM and Linear-SVM parameters were optimized to obtain better prediction accuracy. Accuracy measured using confusion matrix and area under curve for better accuracy. Finally both models were tested using blind data set to be sure that prediction not over fitting.
Data set collected from PPU data base, and has three categories:

- pre university data like year of high school diploma, general secondary branch and high school average.
- university data like university GPA, regularity of students in the study, did student graduate from university, students mark in Physics 1 and 2, Students mark in calculus 1 and 2 and number of honor semesters.
- personal data like the town or city where student lives.

Results of accuracy and area under curve of both RBF and Linear SVM shows that proposed prediction model provided an improvement for classification task for academic performance prediction. Results shows that data set without label 'O' was less accurate in prediction than data set with label 'O' as shown in chapter 4.

Classification accuracy and AUC for proposed prediction models before and after optimization using grid search method and genetic algorithm are nearly the same with a very small different as shown in chapter 4.

So as a summary for our research, a prediction model for engineering students at PPU, moved through the following steps:

1. Create out own data set, two data sets used to obtain final prediction model, first one without 'O' label, and another one with 'O'label.
2. Apply linear kernel under SVM classifier depending on both data sets, with and without 'O' label.
3. Apply radial kernel (RBF) kernel under SVM depending on 'O' label data set.
4. Optimize our RBF-SVM prediction model using grid search method, and genetic algorithm to enhance RBF-SVM parameters C and sigma.
5. Test our enhanced RBF-SVM model using blind dataset, to avoid over fitting.

A summary for accuracy and AUC for proposed models, depending on two data sets, one with 'O' label or class, and other one without 'O' label. Accuracy of RBF-SVM model without 'O' label is 90%, where accuracy of the same prediction model with 'O' label is 95%. Optimized RBF-SVM model using grid search and GA is 96%. Accuracy of proposed model under Linear kernel is 95%. AUC is 90%, 94%, 95%, 95% respectively.

From results we can conclude that optimized RBF-SVM model using grid search gives the most accurate prediction result.

5.2 Future work

Depending on results of accuracy and AUC for proposed model, our future work will to be:

- To cover other university disciplines more than engineering, it may include information technology, information systems, diploma disciplines
- Also collect data from other universities in Hebron, not only PPU but also hebron university, alqus open university and others in country and whole Palestine.

- Build desktop application or mobile application for academic planning that could be used by students, instructors, or any stakeholders. Each stakeholder can submit their information and take their future academic prediction. Poor students could take proactive steps to make better in their study, instructors could develop their teaching strategies for example, some stakeholder like the dean of college could plane enhance academic level in college.

Bibliography

[1] Yannick Meier, Jie Xu, Onur Atan, and Mihaela Van der Schaar. Predicting grades. *IEEE Transactions on Signal Processing*, 64(4):959–972, 2016.

[2] Ali Talpur. *Congestion Detection in Software Defined Networks using Machine Learning*. PhD thesis, 02 2017.

[3] Amjad Abu Saa. Educational data mining & studentsâ performance prediction. *International Journal of Advanced Computer Science and Applications*, 7(5):212–220, 2016.

[4] Shaobo Huang. Predictive modeling and analysis of student academic performance in an engineering dynamics course. 2011.

[5] Ning Fang and Jingui Lu. Work in progress-a decision tree approach to predicting student performance in a high-enrollment, high-impact, and core engineering course. In *2009 39th IEEE Frontiers in Education Conference*, pages 1–3. IEEE, 2009.

[6] Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal. Data mining applications: A comparative study for predicting student's performance. *arXiv preprint arXiv:1202.4815*, 2012.

[7] Surjeet Kumar Yadav and Saurabh Pal. Data mining: A prediction for performance improvement of engineering students using classification. *arXiv preprint arXiv:1203.3832*, 2012.

[8] Brijesh Kumar Bhardwaj and Saurabh Pal. Data mining: A prediction for performance improvement using classification. *arXiv preprint arXiv:1201.3418*, 2012.

[9] Brijesh Kumar Baradwaj and Saurabh Pal. Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417*, 2012.

[10] Dorina Kabakchieva. Student performance prediction by using data mining classification algorithms. *International Journal of Computer Science and Management Research*, 1(4):686–690, 2012.

[11] Dorina Kabakchieva, Kamelia Stefanova, and Valentin Kisimov. Analyzing university data for determining student profiles and predicting performance. In *Educational Data Mining 2011*, 2010.

[12] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. 2003.

[13] Raheela Asif, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider. Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113:177–194, 2017.

[14] Theodoros Evgeniou and Massimiliano Pontil. Support vector machines: Theory and applications. volume 2049, pages 249–257, 01 2001. doi: 10.1007/3-540-44673-7_12.

[15] Siu-Man Raymond Ting and R Man. Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables. *International Journal of Engineering Education*, 17(1):75–80, 2001.

[16] Muhammad Shahid Farooq, Azizul Haque Chaudhry, Mohammad Shafiq, and Girma Berhanu. Factors affecting studentsâ quality of academic performance: a case of secondary school level. *Journal of quality and technology management*, 7(2):1–14, 2011.

[17] Fotios Misopoulos, Maria Argyropoulou, and Dionisia Tzavara. Exploring the factors affecting student academic performance in online programs: A literature review. In *On the Line*, pages 235–250. Springer, 2018.

[18] Eka Sugiyarti, Kamarul Azmi Jasmi, Bushrah Basiron, Miftachul Huda, K Shankar, and Andino Maseleno. Decision support system of scholarship grantee selection using data mining. *International Journal of Pure and Applied Mathematics*, 119 (15):2239–2249, 2018.

[19] Pavel Laskov, Patrick Düssel, Christin Schäfer, and Konrad Rieck. Learning intrusion detection: supervised or unsupervised? In *International Conference on Image Analysis and Processing*, pages 50–57. Springer, 2005.

[20] Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, S Yu Philip, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.

[21] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[22] Laura Auria and Rouslan A Moro. Support vector machines (svm) as a technique for solvency analysis. 2008.

[23] Vikramaditya Jakkula. Tutorial on support vector machine (svm). *School of EECS, Washington State University*, 37, 2006.

[24] *Disadvantages of Support Vector Machines*, (accessed May 11, 2019). URL http://www.svms.org/disadvantages.html.

[25] Ajay Mathur and Giles M Foody. Multiclass and binary svm classification: Implications for training and classification users. *IEEE Geoscience and remote sensing letters*, 5(2):241–245, 2008.

[26] Donald Knuth. *Support vector machines: The linearly separable case*, (accessed May 1, 2019). URL https://nlp.stanford.edu/IR-book/html/htmledition/support-vector-machines-the-linearly-separable-case-1.html.

[27] Ryan Rifkin. *Multiclass Classification*, (accessed May 20, 2019). URL http://www.mit.edu/~9.520/spring09/Classes/multiclass.pdf.

[28] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[29] Asa Ben-Hur and Jason Weston. A userâs guide to support vector machines. In *Data mining techniques for the life sciences*, pages 223–239. Springer, 2010.

[30] Mirka Saarela and Tommi Kärkkäinen. Analysing student performance using sparse data of core bachelor courses. *Journal of educational data mining*, 7(1), 2015.

[31] Mack Sweeney, Jaime Lester, and Huzefa Rangwala. Next-term student grade prediction. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 970–975. IEEE, 2015.

[32] Mack Sweeney, Huzefa Rangwala, Jaime Lester, and Aditya Johri. Next-term student performance prediction: A recommender systems approach. *arXiv preprint arXiv:1604.01840*, 2016.

[33] Kai-Bo Duan and S Sathiya Keerthi. Which is the best multiclass svm method? an empirical study. In *International workshop on multiple classifier systems*, pages 278–285. Springer, 2005.

[34] *Cross-validation: evaluating estimator performance*, (accessed May 05, 2019). URL https://scikit-learn.org/stable/modules/cross_validation.html.

[35] Max Kuhn. *trainControl Control Parameters For Train*, (accessed February 28, 2019). URL https://www.rdocumentation.org/packages/caret/versions/6.0-81/topics/trainControl.

[36] Max Kuhn. *train Fit Predictive Models Over Different Tuning Parameters*, (accessed February 28, 2019). URL https://www.rdocumentation.org/packages/caret/versions/4.47/topics/train.

[37] K Muralitharan, Rathinasamy Sakthivel, and R Vishnuvarthan. Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing*, 273:199–208, 2018.

[38] M Ataei and M Osanloo. Using a combination of genetic algorithm and the grid search method to determine optimum cutoff grades of multiple metal deposits. *International Journal of Surface Mining, Reclamation and Environment*, 18(1):60–78, 2004.

[39] Darrell Whitley. A genetic algorithm tutorial. *Statistics and computing*, 4(2):65–85, 1994.

[40] TOM FAWCETT. *The Basics of Classifier Evaluation Part 1*, 2015 (accessed April 3, 2019). URL https://www.svds.com/the-basics-of-classifier-evaluation-part-1/.

[41] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pages 1015–1021. Springer, 2006.

[42] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[43] Jorge M Lobo, Alberto Jiménez-Valverde, and Raimundo Real. Auc: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography*, 17(2):145–151, 2008.

[44] Jin Huang, Jingjing Lu, and Charles X Ling. Comparing naive bayes, decision trees, and svm with auc and accuracy. In *Third IEEE International Conference on Data Mining*, pages 553–556. IEEE, 2003.