

Student Result Predictor Using Machine Learning Techniques

**P.H Thilakaratne
2021**



Student Result Predictor Using Machine Learning Techniques

**A Dissertation Submitted for the Degree of Master
of Business Analytics**

P.H Thilakaratne

2018/BA/035

18880358

**University of Colombo School of Computing
2021**



DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

Student Name: P.H Thilakaratne

Registration Number: 2018/BA/035

Index Number:

18880358

Signature of the Student & Date



14th September 2021

This is to certify that this thesis is based on the work of Mr. P.H Thilakaratne under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by,

Supervisor Name: D.T. Bamunuarachchi



Signature of the Supervisor & Date: **14th September 2021**

Table of Contents

Abstract.....	7
Acknowledgment	8
Chapter 1: Introduction	9
1.1 Chapter Overview	9
1.2 Problem Domain	9
1.3 Previous Work.....	10
1.4 Research Gap	11
1.5 Project Aim	11
1.6 Project Objectives	12
1.7 Proposed Solution	14
1.8 Resource Requirements	16
* Resource requirements stated below are subjected to change	16
1.9 Project Document Structure	17
Chapter 2: Literature Review	19
2.1 Chapter Overview	19
2.2 Parameter Identification on Student Result Prediction	19
2.3 Technical Background on Problem Domain.....	21
2.3.1 Prediction System.....	21
2.3.2 Key Success Factors of Prediction System	22
2.3.3 Prediction System Architecture.....	22
2.3.4 Machine Learning Algorithms usage in Student Result Predicting System	25
2.4 Chapter Summary	27
Chapter 3: Project Management.....	28
3.1 Chapter Overview	28
3.2 Project Management Methodology.....	28
3.2.1 Time Allocation.....	28
3.2.2 Constraints and Dependencies.....	29
3.2.3 Potential Risks and Mitigation Plan	30
3.3 Development Methodology	32
3.4 Research Methodology	33

3.5 Chapter Summary	33
Chapter 4: Requirement Specification	34
4.1 Chapter Overview	34
4.2 Requirement Gathering Process	34
4.3. Data Collection and Analysis	36
4.3.1 Questionnaire	36
4.3.2 Interviews	40
4.3.3 Summary of Requirement Gathering	41
4.4 Analysis of Stakeholders.....	41
4.5 Use Case Diagram of the System and Description	43
4.6 Functional Requirement.....	44
4.7 Non-Functional Requirements.....	46
4.8 Chapter Summary	46
Chapter 5: System Architecture and Design.....	47
5.1 Chapter Overview	47
5.2 High-level design of the System.....	47
5.3 Machine Learning Algorithms.....	49
5.4 ER Diagram of the application	49
5.5 Chapter Summary	51
Chapter 6: Implementation.....	52
6.1 Chapter Overview	52
6.2 Selection of Technology	52
6.2.1 Selection of Machine Learning Algorithms.....	52
6.2.2 Data Pre-Processing.....	53
6.2.3 Training the Algorithms.....	54
6.3 Code Implementation	55
6.3.1 Front End Implementation	58
6.3.2 Database Implementation.....	58
6.4 Chapter Summary	59
Chapter 7: Testing	60
7.1 Chapter Overview	60
7.2 Goals and Objectives of Testing.....	60
7.3 Procedure of Testing	60
7.3.1 Testing the Machine Learning models	61
7.4 Chapter Summary	67

Chapter 8: Evaluation	68
8.1 Chapter Overview	68
8.2 Evaluation Process	68
8.3 Chapter Summary	70
Chapter 9: Conclusion	71
9.1 Chapter Overview	71
9.2 Completion of Aims and Objectives.....	71
9.3 Utilizing of Knowledge from Master’s Program Modules	72
9.4 Usage of Existing Knowledge.....	72
9.5 Problems and Challenges	72
9.6 Future Implementations.....	73
9.7 Concluding Remarks.....	73
References.....	74
Appendix I Gantt Chart.....	76
Appendix II Turnitin Report.....	77

Abstract

Education is one of the most important aspect in human life. As human all of us spending twelve to thirteen years at schools and then passing through to higher education institutes such as universities. In universities student face exams and some of them get through it and others get stuck. This study will predict student academic performance with the help of machine learning algorithms. This study will discuss about four classification algorithms such as Decision tree classifier, Support Vector Machine Classifier, Naïve Bayes and Random Forest classifier. This study will address which would be the optimum algorithm can be used to predict student results under identified parameters.

Keywords:

Machine Learning, Student Result Predictor, Semester Result Predictor, Support Vector Machine, Decision Tree, Naïve Bayes

Acknowledgment

It took me a lot of effort and hard work to make the project a reality with support and guidance from all the people who have helped me. I take this opportunity to express my profound gratitude to everyone who helped me to make this project a success.

First and foremost I wish to express my deepest gratitude to my project supervisor Mr. Dasun Bamunuaarchchi, Lecture, University of Colombo School of Computing for great assistance and guidance on completing this project. Without his constant support and encouragement this task would have been impossible to complete.

I extend my warm gratitude to Dr.Thilina Halloluwa, Senior Lecture, University of Colombo School of Computing for giving guidance throughout the project.

I hereby express my gratitude to all my lectures and management at the University of Colombo School of Computing for providing valuable knowledge and assistance throughout the masters degree program.

I offer my thanks to everyone who provide feedback on requirements gathering and evaluation phases which helped me to conclude my project.

Finally, special thanks goes to my family, friends and office colleges for their continuous support which helped me to complete this project successfully.

Chapter 1: Introduction

1.1 Chapter Overview

This chapter give an overview of the project that author has undertaken. Chapter starts by providing a glance view student result prediction and prerequisites. Next, previous work done under same domain is evaluated along with domain limitations. After that, project research gap in briefly highlighted. Next, aim and objective of the research is highlighted along with features of the proposed solutions and resources requirements. Finally, overview of the chapter contained in this report.

1.2 Problem Domain

Education is an important issue in a person's life. It is the key to success in the future, as well as more opportunities in our lives. Education has many benefits for people. For example, it enlightens a person's mind and heart. It helps students to plan for a career or pursue higher education through a university degree. Local education helps people to think, feel, and behave in a way that contributes to their success, and not only enhances their satisfaction but also their community. In addition, education develops personality, thinking, and social skills. It also prepares people for what happens in life. It gives people special positions in their community and in all areas where they live. I believe everyone has the right to an education " from birth to the grave ". There are many benefits to having an education, such as getting a good job, a good social status, and self-confidence.

In Sri Lankan context student usually spend thirteen years to complete the school life, then after completing G.C.E Advanced Level examination some student get selected to state universities, some opting for private universities, some of them leaving the country for higher educations and others selecting diplomas, certifications etc. Reasons for above selection is, without obtaining a higher education qualification it's hard to retain in the society specially when finding a job. So the university life of the student would be the final chance and that's most of them are following to obtain their qualification in order to lay the foundation to the cooperate world.

In Sri Lanka there are number of students who are not performing well though they have been selected to the university. After completing their three or four years at university they pass out from the university by obtaining a basic degree without a class or proper qualification due to various reasons. Once they are not qualified, they won't be able to find an appropriate job in order to live a sufficient life. If there is a possibility of predicting student educational results at early stages that would be idle hence student can adopt and correct their way of studying before obtaining results at the end. So early result predictor would be idle for university students, and for whom engaged in education to predict their academic results.

1.3 Previous Work

In recent decades, predicting the performance of students in the academic field has revealed the attention by researchers for enhancing the weaknesses and provides support for future students(Almasri, Celebi and Alkhaldeh, 2019) In recent years, a lot of data has been generated about students, which can be utilized for deciding the career path of the student(Sharma, Pandey and Garg, 2020)We first collect a large number of people's profile and extract features from the descriptive information. Hand rules and clustering algorithm has been applied to help avoid the negative effect of natural language. We model people's career developments with Markov Chain, and present our approach to estimate the transition probability matrix. Finally, we solve the problem that given a person's current career path and his/her goal, what is the best best career development recommendation for him/her. As a conclusion, we will analyze the results and discuss possible improvements of our model(Lou, Ren and Zhao, no date)

Prediction of student's performance became an urgent desire in most of educational entities and institutes. That is essential in order to help at-risk students and assure their retention, providing the excellent learning resources and experience, and improving the university's ranking and reputation. However, that might be difficult to be achieved for startup to mid-sized universities, especially those which are specialized in graduate and post graduate programs, and have small students' records for analysis. So, the main aim of this project is to prove the possibility of training and modeling a small dataset size and the feasibility of creating a prediction model with credible accuracy rate. This research explores as well the possibility of identifying the key indicators in the small dataset, which will be utilized in creating the prediction model, using visualization and clustering algorithms. Best indicators were fed into multiple machine learning algorithms to evaluate them for the most accurate model. Among the selected algorithms, the results proved the ability of clustering algorithm in identifying key indicators in small datasets. The main outcomes of this study have proved the efficiency of support vector machine and learning discriminant analysis algorithms in training small dataset size and in producing an acceptable classification's accuracy and reliability test rates(Abu Zohair, 2019)

As mentioned above there are several research have been conducted in predicting student performance, results, career etc. As mentioned by above researchers they have been used various parameters, assumptions, factors when predicting in those research papers. But there are some factors which has not been addressed by the above-mentioned research papers which the author mentioned in research gap.

1.4 Research Gap

As mentioned in above research papers by researchers who has undertaken the above researchers, they have selected the parameters, factors based on their context. In Sri Lanka, still the country is third world developing country which is under several social and human factors. In Sri Lankan context parameters, which leads to student performance, results and outcome would be different due to several reasons. In order to address issues and limitations, project work will focus on creating an effective and efficient system that will predict student performance using machine learning within range of supported student result, performance factors. In detail discussion on research gap is stated in chapter two.

1.5 Project Aim

To design, develop and evaluate machine learning based student result prediction system that will predict one's semester, result based on pre-defined factors and user feedbacks, which can be used by student, teachers, lecturers and other relevant personnel.

Further elaborating on the aim, this project will produce a machine learning based result prediction system that predicts student semester and annual result based on user feedback and pre-defined risk factors data using machine learning processing. This proposed solution consists of two main sub applications which are main application and machine learning model. Main application allows managing parameters, factors and its associated data, acquiring user feedback and providing student result predictions. The machine learning module is used to get predictions based on user feedback and predefined factors.

In this study, author has selected four classification algorithms and parameters which related to student result prediction in literature review. Based on the selected parameters, authors' aim is figure out best machine learning algorithm which can be used to predict student results among selected four machine learning algorithms. Once the optimum algorithm is identified that algorithm will be used to predict the student result under selected parameters. This study is the process of fulfilling the selected project aim step by step with relevant studies and analysis.

1.6 Project Objectives

In order to successfully fulfill above aim, objectives were identified and stated below (Table 1.1), which will be followed until the project competition.

1. Prepare Terms of Reference
Prepare Terms of Reference document which include project background, aim and objectives, project deliverables, proposed solution and activity schedule.
2. Literature Review and Survey
<ul style="list-style-type: none"> • Research about current research work on student academic result prediction systems , services and projects • Research about key technologies , algorithms (including machine learning) , methodologies, protocols, and strategies that can be used in proposed system • Research, identify and select key factors associated with student result predictor. • Research and identify suitable data sets /sources that can be used to model, train and evaluate the proposed system. • Determine the best suited solutions for above options that will be used to develop the system.
3. Selection of a software development methodology
Select a software development methodology to carry out project life cycle tasks.
4. Requirements gathering
<p>Carryout in-depth analysis on the domain and gather requirements;</p> <ul style="list-style-type: none"> • By using software elicitation techniques such as observation, questionnaires, and interviews. • By reviewing feedback from domain experts.
5. Prepare Software Requirement Specification (SRS)
Prepare Software Requirement Specification based on the requirements gathering data for the system which include functional and non-functional requirements.

6. Selection of software and hardware resource
Select most suitable hardware and software resources that require to design and develop prototype.
7. Prepare software design specification
Design the prototype based on the in-depth analysis data. By the end, information such as UML diagrams (such as Class diagrams and Sequence diagrams), database designs and user interface designs for the prototype is produced.
8. Develop the prototype
Develop the prototype using selected software and hardware resources which should fulfill functional and non-functional requirements.
9. Prototype testing
Test the prototype using pre-defined test plan which include test cases and evaluate.
10. Evaluation of the project work
Critically evaluate the prototype based on test data and evaluation surveys (such as user evaluation testing, review the prototype by domain experts and supervisor and self evaluation). Also evaluation should include current features and limitation on the solution and future enhancements.
11. Documentation

Produce relevant documentations (such as interim report and final report) throughout project life-cycle within specified time-frame.

Table 1.1 - Objectives of the project

1.7 Proposed Solution

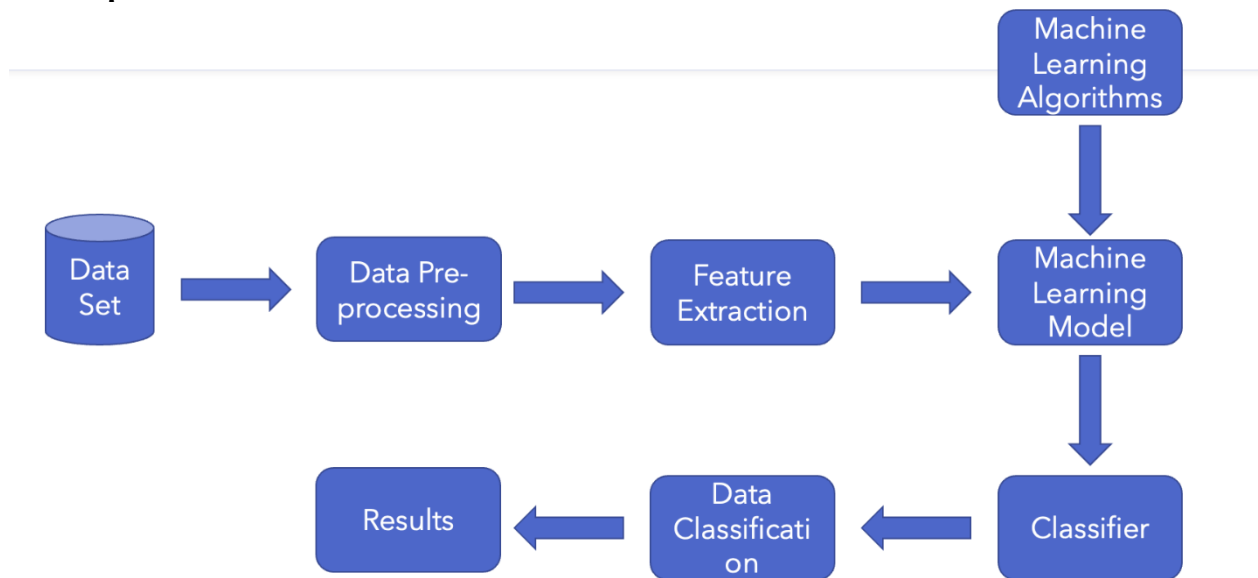


Figure 1.1 Solution Design

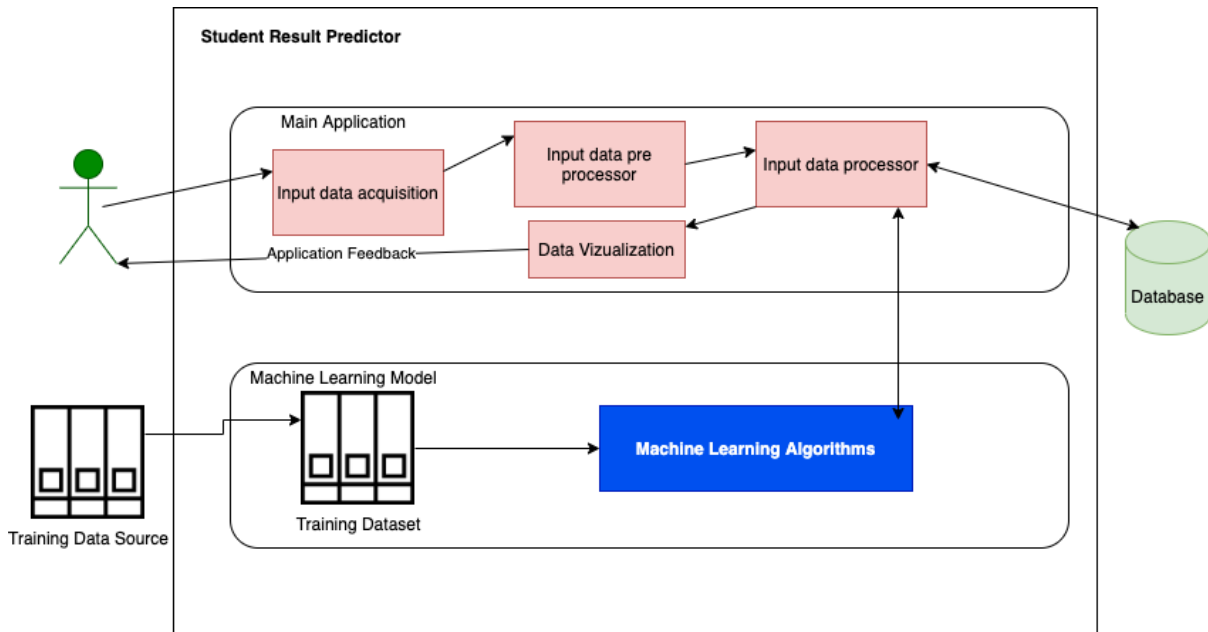


Figure 1.2 Basic Flow of the proposed system

Main Application

Main application consists of four sub modules which include;

- **Input data acquisition:** This module will be used to capture user feedback. First, pre-defined questionnaire (which will be used determine student results) will be presented to user to fill. Next, user feedback for the questionnaire will be which captured via this module.
- **Input data pre-processor:** This module will pre-process (clean and format data using pre-defined rules) data (for further processing using ML algorithms), that was acquired in “Input data acquisition” module.
- **Input data processor:** In this module, all the pre-formatted data will be passed to machine-learning module and receive the result (Student result predictor).
- **Data visualization:** In this module, the processed result is presented to user in formatted report.

Machine Learning module

Machine Learning (ML) module will include all the application supported machine-learning models and training datasets. It is needed to point out that, throughout this document, term ML will be used adequately to refer “machine learning” keyword. ML models will use, training data sets to train and evaluate to predict accurate predictions on user result behavior.

The name of proposed solution can be identified as “Student Result Predictor” and throughout this document proposed solution name will be used adequately.

1.8 Resource Requirements

Below defined resources (Table 1.2) are require to successfully complete the project.

*** Resource requirements stated below are subjected to change**

Hardware Requirements
<ul style="list-style-type: none">• PC / Laptop with Anaconda , Python installed with following configuration (laptop with below configuration , will be used to develop the proposed system) ;<ul style="list-style-type: none">○ Core i5 2.6 GHz processor○ 4GB DDR3 RAM○ 100GB HDD• Internet connectivity.
Software Requirements
<ul style="list-style-type: none">• MacOS / Windows OS• Python (for machine learning module implementation)• Anaconda (for python)• Machine Learning Library• Microsoft office package (MS Word, MS power point , MS project)• Draw.IO (for modeling diagrams)• Mendeley (for reference management)• IDE (VisualStudio Code)
Knowledge
<ul style="list-style-type: none">• Domain knowledge in Student Results behavior, Result early identification and prediction.• Machine learning algorithms• Programming knowledge in Jupyter Notebook python

Time
<ul style="list-style-type: none"> • Time to carry out literature review and survey. • Time to carry out analysis and design. • Time to carry out development and testing the prototype. • Time to carry out new learning curves (such as on new technologies and new algorithms).
Other
<ul style="list-style-type: none"> • Testers and domain experts“ feedback and evaluation on the system.

Table 1.2 - Resource requirements

1.9 Project Document Structure

Final project document will contain following chapters (table 1.3) followed by reference and appendix at the end.

Chapter 2 – Literature Review
Literature review chapter will include an in-depth review on social background and technical approaches on student result prediction covering aspects such as risk factors, parameters which identified and their usage, machine learning approach and already existing systems or services in result prediction in contrast to accuracy and performance factors.
Chapter 3 - Project management
The project management chapter will include details about the selected project management method, project plan, risk reduction plan and the selection of the appropriate software development method.
Chapter 4- Requirements Specification
This chapter will explain the project participants, the service delivery strategies used for the collection of needs, the strategies used to identify identified and identified operational and non-operational needs and requirements.
Chapter 5- System Architecture and Design
System Architecture and Design will provide details about design decisions and goals and modal diagrams covering high level architecture of the system, domain model and sequence diagrams.
Chapter 6- Implementation

Implementation chapter will contain descriptions about technology selection, implementation of functional and non-functional requirements stating sample code segments along with challenges encountered.

Chapter 7- Testing

Testing chapter will include the details about testing phase of the project covering from unit testing to final prototype requirements testing. Testing feedback acquired from tests will be recorded and analyzed.

Chapter 8- Evaluation

Evaluation chapter will provide details about project evaluation covering expert evaluation, self-evaluation and review of those.

Chapter 9 – Conclusion

Conclusion chapter will provide a analysis covering success of the project, modules contribution, skills utilization, problem encountered, limitation on the research, future enhancements and concluding remarks.

Chapter 2: Literature Review

2.1 Chapter Overview

Previous chapter include an introduction to the project covering problem domain, previous work, project aim and objective and proposed solution. This chapter will present a critical review on identified parameters and previous work of the problem domain. The chapter starts by presenting a discussion about identified parameters of the result prediction. Next, in detail discussion on technical approaches (possible techniques, technologies, algorithms and methodologies) that relate to the problem domain is present. Finally, author's reflection on adaptable key points for proposed solution is highlighted.

2.2 Parameter Identification on Student Result Prediction

As mentioned in the previous chapter result of a student will depends on certain factors, when considering those factors some are social factors and other's are subject related factors which directly depends on the subject area which the student is following. Since subject areas are differentiating heavily in this study author is focusing on social factors which will affect to student results and predictions.

Correlation among attendance and grades. As we see, attendance has very less correlation with Mid Sem Grades and final Grade. One plausible reason could be due to the bucketing of the **Attendance** attribute into 5 levels, due to which the variance in attendance is at least five times as slow as variance in the grades. We also see, **Mid Sem Grades** and final **Grade** have a much higher correlation.(Venkat *et al.*, 2018)

	Attendance	Mid Sem Grades	Grade
Attendance	1.00	0.03	0.16
Mid Sem Grades		1.00	0.71
Grade			1.00

Table 2.1 Correlation among attendance and grades (Venkat *et al.*, 2018)

According to above study which was carried out by Naveen Venkat, has proved that there

is a correlation among attendance and grades which student are obtaining specially when it comes to the end of the year. Since attendance is not limiting for specific subject area and hence it's common to all subject areas, as the first parameter attendance of the student has identified when predicting the results of the student.

From the table, it is found that the second high potential variable for students' performance is their living location, and the third high potential variable for students' performance is medium of teaching. In Uttar Pradesh the mother tongue language of students is Hindi. Hence, students tend to be more comfortable in Hindi and other languages, than in the English language.(Agarwal *et al.*, 2020)

Variable	Description	Probability
GSS	Student's Grade in Secondary Education	0.8642
LLoc	Living Location	0.7862
Med	Medium of Teaching	0.7225

Table 2.2 Study Results(Agarwal *et al.*, 2020)

According to the research which carried out by Havan Agarwal in 2020, has identified that living location and medium of teaching has a significant correlation to the final results of a student. Among those identified parameters living location has been identified as the second parameter since medium of teaching would be English most of the time in Sri Lanka hence it won't be a huge factor when considering about university students. Same study has identified the family education level matters when predicting the student results. it was found that the factors like students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance(Agarwal *et al.*, 2020)

According to the research which carried out by Havan Agarwal in 2020, has identified that living location and medium of teaching has a significant correlation to the final results of a student. Among those identified parameters living location has been identified as the second parameter since medium of teaching would be English most of the time in Sri Lanka hence it won't be a huge factor when considering about university students. Same study has identified the family education level matters when predicting the student results. it was found that the factors like students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance(Agarwal *et al.*, 2020)

2.3 Technical Background on Problem Domain

2.3.1 Prediction System

Prediction system concentrates on analyzing set of relevant data, and predicts a future implication(Kaul, Kaul and Verma, 2015). Prediction system usually make a prediction model (using historical data and supported processors with or without human intervention) and use that model to make prediction for new requests or new entries. These systems apply knowledge discovery techniques to create prediction models, which used to make predictions(Kaul, Kaul and Verma, 2015). Apart from Education, predictions systems are used in financial analytics, weather forecasting, customer relationships, supply chain management and many more covering diverse range of fields. It is worth highlighting that many of the prediction systems or prediction models come in embeds with recommendation systems. Ricci *et al.* (2010) highlight that Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user) or with expert systems (An expert system is a computer program that uses artificial intelligence (AI) technologies to simulate the judgment and behavior of a human or an organization that has expert knowledge and experience in a particular field (TechTarget, 2016)) as in module(s) or feature(s).

2.3.2 Key Success Factors of Prediction System

It is emphasized that the accuracy is considered as the main success key factor when it comes to prediction systems. Apart from accuracy, performance factor signifies high priority when creating prediction models to achieve high accuracy rate towards predictions ((Cook et al., 2016),(Pestian et al., 2012)). Apart from above key factors, O’Dea et al. (2015), AbuKhoua and Campbell (2012), Barros et al. (2016), Nguyen et al. (2016) and Dinov et al. (2016) highlight sensitivity and specificity (towards greater accuracy in predictions), scalability, reliability, robustness, ease of interpretation and generalization ability as some of the key

success factors that expect from a prediction system. It is known that other key factors may correlate towards accuracy in negative or positive way. You can’t always expect 100% fulfillment from all the key factors and it is wise to adapt the optimum level on desired key factors making sacrifices in others. The research will focus on identifying best approaches that can be adapted to make a student result prediction system considering accuracy and performance as the key success factors.

2.3.3 Prediction System Architecture

Hussain *et al.* (2015) authors proposed an architecture that can be used to predict depression using social networking sites (SNS) data. Furthermore, the proposed architecture (figure 2.1) can be adapted for different disease prediction systems by introducing relevant changes to the platform design.

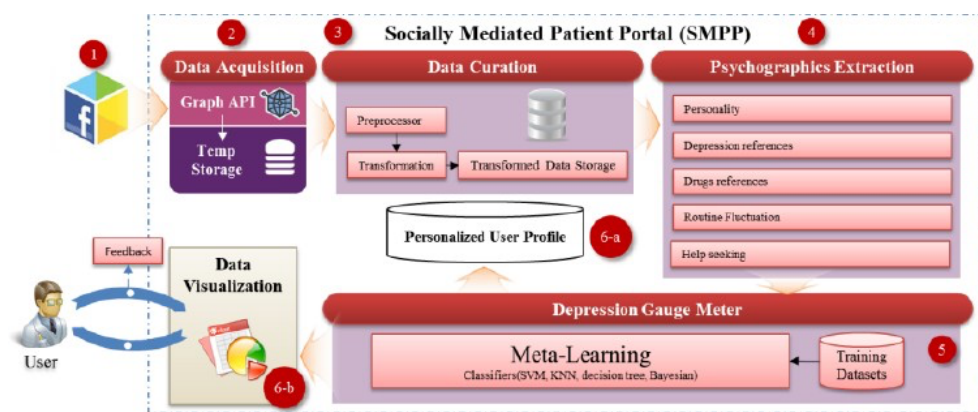


Figure 2.1 - Overview of SNS based depression prediction system- (Hussain *et al.*, 2015)

Hussain *et al.* (2015) authors highlight that in order to cross check the results of the tool, authors perform peer review of the subject on the basis of well-defined standard questionnaires. The authors of the paper didn't share their findings on prediction model(s) validity over the reviewed data. Thus, the validity of the prediction model(s) is unclear in means of accuracy and performance factors.

To address the common issues of above literature review such as class imbalance, data hi-dimensionality and classification errors, this study has proposed a model which have following phases. Figure 1 shows the main steps of proposed methodology. (Imran *et al.*, 2019).

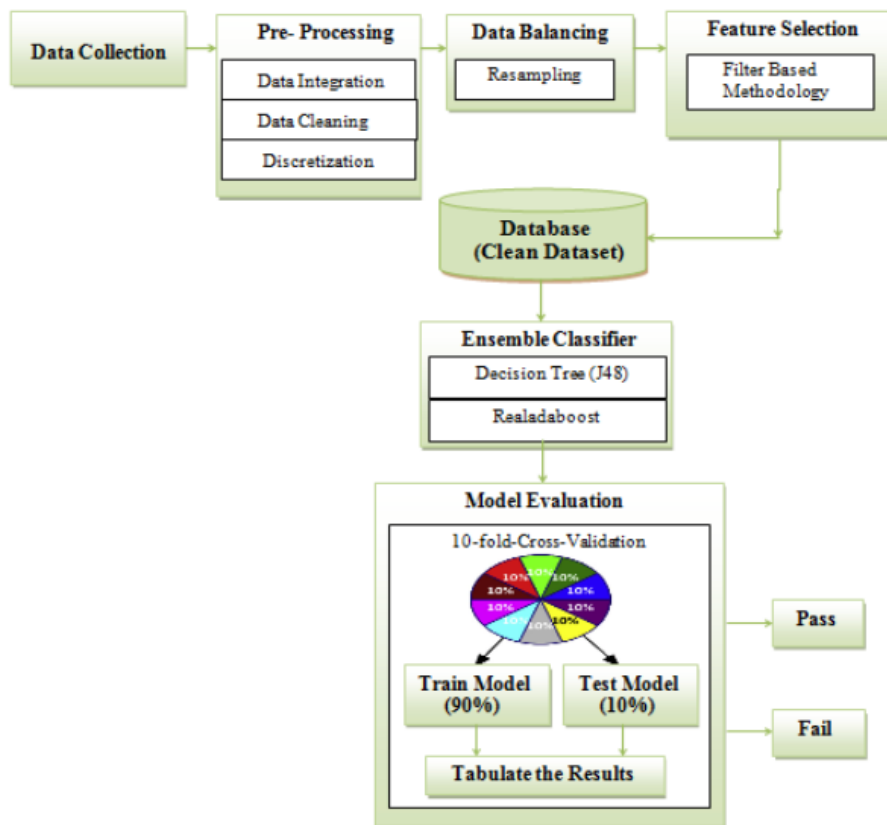


Figure 2.1 Proposed System for the Student Career Prediction(Imran *et al.*, 2019)

The training set is ready to be used for constructing the predictive models of student performance. There is a set of ML techniques from different families that could be used to construct the predictive models. These families differentiate from each other in the theoretical process used for building the model. These models can be deployed in systems for classifying future instances. Machine learning techniques construct a hypothesis from a space set of hypotheses using in the training set. (Almasri, Celebi and Alkhalwaleh, 2019)

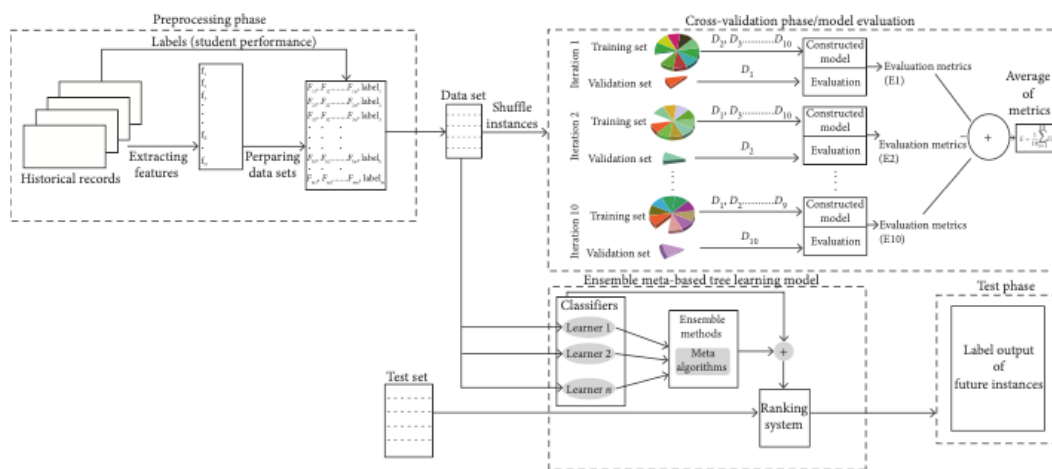


Figure 2.2 The architecture and evaluation of EMT technique.(Almasri, Celebi and Alkhalwaleh, 2019)

To achieve the project’s aims, quantitative simulation research methods were conducted as suggested in the framework phases shown in Fig. 1. In these phases the data-set will be prepared to be passed through visualization and clustering techniques, i.e. like heat map and hierarchical clustering, to extract the top correlated indicators. Then, the indicators will be used in different classification algorithms and the most accurate model will be the chosen for predicting student performance in dissertation projects and all courses grades. In between, and before the classification models’ evaluation phase, the datasets will pass through a pre-processing stage to make it ready for the analysis phase.(Abu Zohair, 2019)



Figure 2.3 Architectural Diagram(Abu Zohair, 2019)

Furthermore, to extend the above work authors are considering about adding visualize data with more diagrammatical representation as well since these visualized data express more information.

2.3.4 Machine Learning Algorithms usage in Student Result Predicting System

Decision Tree: Decision tree is a supervised learning algorithm which is mostly used for classification problems. Decision tree is used for both categorical and continuous dependent variables. In this algorithm, we divide the population into two or more similar sets. A decision tree is a graphical representation that makes use of branching methodology to represent all possible outcomes of a decision based on certain conditions. **Support Vector Machine:** It classifies the data into different classes by finding a line (hyperplane) which divides the training data set into classes. This algorithm plots each data item as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a particular coordinate. Support vectors are data points that are closer to the hyperplane. Using these support vectors, we maximize the margin of the classifier(Sharma, Pandey and Garg, 2020)

SVM denotes Support Vector Machine. It is a supervised machine learning algorithm which is generally used for both regression and classification type of problems. The main applications of this can be found in various classification problems. The typical procedure

of the algorithm is first each data item is plotted in a n-dimensional space, where n is the number of features and the value of each feature being the value of that particular coordinate. Next step is to classify by getting the hyper-plane that separates the two classes very finely. (Roy *et al.*, 2018)

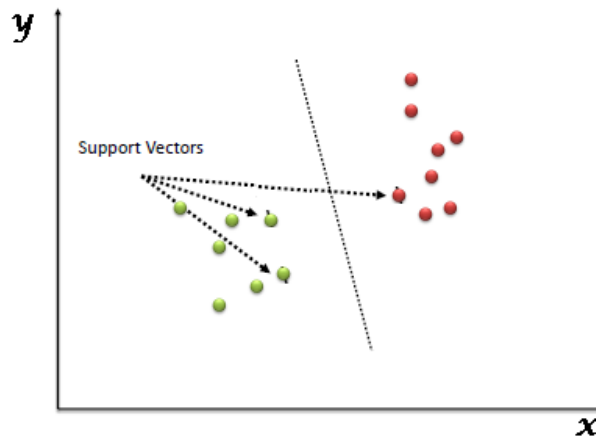


Figure 2.4 Support Vector Machine Examples(Roy *et al.*, 2018)

SVM algorithms practically are implemented using kernels. There are three types of SVM's and in linear SVM hyperplane is calculated or found by transforming the problem using linear algebra. The insight is that SVM can be rephrased by using the inner product of two observations. The sum of the multiplication of each pair of inputs is called inner product of two vectors. The equation for dot product of a input x_i and support vector x_i is:
 $f(x) = B_0 + \sum(a_i * (x, x_i)).$

Instead of using the dot-product, a polynomial kernel can be used, for example:

$$K(x, x_i) = 1 + \sum(x * x_i)^d$$

And not only that a more complex radio kernel is also there. The general equation is:

$$K(x, x_i) = \exp(-\gamma * \sum((x - x_i)^2)) \text{ (Roy } et al., 2018)$$

NAIVE BAYES:

It is a classification technique based on Naïve Bayes Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes is a simple but surprisingly powerful algorithm for predictive modeling. The model is comprised of two types of probabilities that can be calculated directly from your training data: 1) The probability of each class; and 2) The conditional probability for each class given each x value. Once calculated, the probability model can be used to make predictions for new data using Bayes Theorem. (VidyapriyaC and VishhnuvardhanRC, no date)

THEOREM : $P(c/x) = P(x/c)P(c)/P(x)$ (VidyapriyaC and VishhnuvardhanRC, no date)

2.4 Chapter Summary

This chapter critically reviews best concepts, approaches and technical feasibilities to address the problem domain. Chapter starts by identifying social parameters of student result identification and prediction. Next parameters identified under different priorities which is the base factor on student result predictions. Key parameter which identified from the literature review is student results are depends on several social factors which are not subjective and depends on the curriculum. From previous studies, author has identified Study Hours, Parents Education level, Living Location and Attendance are parameters which student results are depends on. After identifying parameters, next important identification is the machine learning algorithm which can be used to predict values based on above selected parameters. When referring to the previous studies, author has identified that Decision Tree Algorithm, Support Vector Machine and Naïve Bayes algorithm are successfully used in such studies. Based on above conclusion, for this study author is predicting student results based on four parameters such as Study hours, Student attendance during that time period, Parents education level and Living location with the help of three machine learning classification algorithms such as Decision Tree, Naïve bayes and Support Vector Machine.

Chapter 3: Project Management

3.1 Chapter Overview

While the previous chapter highlights the content of the literature on a troubled domain, this chapter outlines details of project management aspects. The chapter begins by identifying the right approach to project management and the allocation of time, issues and project dependencies. The following highlights the potential risks of the project and suggests mitigation measures against those risks. After that, details of the appropriate development methods for the project were discussed by examining several development options. Finally, there was a brief discussion of the research method.

3.2 Project Management Methodology

Effective project management is essential to any project, regardless of the type of work and its scale. The main issues of the project are the scale, time and cost. In order to produce a quality product from a project made, one has to deal with the above in a logical way. Project management methods assist the right people to design, plan, implement and achieve project objectives within a set schedule and budget. Research projects have more flexibility of needs change compared to industry projects. Therefore, it is very important to know exactly how to deal with such situations. PRINCE2 is a widely used method of project management that has proven to be very effective in various aspects of engineering. This approach is product-based and supports high-level management, control and project planning. In order to effectively and efficiently manage the project, it was decided to adopt PRINCE2 as a project management approach.

3.2.1 Time Allocation

Below table 3.1 represents an overview of the main tasks of the project (Refer Appendix A for Gantt chart). In project, literature review will be written from inception of the project to end of testing phase due to frequent changes in requirements. Similarly, the documentation of project will also carry throughout project life cycle. At the end, system prototype and documentation will be presented as final deliverables of the research.

Id	Task	Start Date	Finish Date
TA1	Initial Research	01 Aug 2020	18 Aug 2020
TA2	Proposal Preparation and Submission	18 Aug 2020	23 Aug 2020
TA3	Literature review phase / Project Dev	23 Aug 2020	01 Mar 2021
TA4	Interim Submission	01 Mar 2021	05 Mar 2021
TA5	Designing phase of the system	25 Mar 2021	05 Jun 2021
TA6	Interim Presentation	05 Jun 2021	05 Jul 2021
TA7	Testing phase of the system	09 Jun 2021	05 Jul 2017
TA8	Documentation and Thesis Submission	1 Aug 2021	14 Sep 2021

Table 3.1 - Time Allocation

3.2.2 Constraints and Dependencies

Successful completion of the project is dependent on following constraints and dependencies.

- Time constraint – As project is carried out individually, time element have critical impact on project completion, where author should complete all project tasks and achieve project goals in timely manner.
- Lack of prior knowledge – Due to the lack of knowledge on machine learning algorithms for student result prediction, development of the proposed system can be delayed.
- Financial Obligations – Financial constraints may occur when trying to accumulate datasets or any other data or service that requires to built the proposed system. Author should address or look for alternative solutions when such scenario occurs.
- Legal and Ethical Obligations – Due to the gravity and sensitivity of the problem domain, author should address any legal and ethical constraints when handling sensitive data that will be used throughout the project for different operations (such as accumulating data to be used in machine learning algorithms). When gathering

student results hence data are very sensitive there was a legal process to be followed.

3.2.3 Potential Risks and Mitigation Plan

Below Table 3.2 highlight the identified potential risks and mitigation steps for Student Result Predictor system.

Risk	Failure to keep with the up-to-date domain knowledge and technology changes		
Risk level	High	Occurrence Frequency	High
Description	Student performance prediction systems and service is a constantly evolving, actively research area. This makes it hard to keep up with latest developments of the research area. New technologies and trends emerge while research in progress. This present a risk that the system is not using up to date domain knowledge and technologies to develop the system.		
Mitigation	<ul style="list-style-type: none"> • Check for newly updates on domain and technological changes frequently. • Keep frequent communication with domain experts for any new developments. 		
<hr/>			
Risk	Constant and repeated changes to requirements		
Risk level	High	Occurrence Frequency	High
Description	Research projects are risky since the requirements can be constantly change and work need to be complete within given fixed time-frame.		
Mitigation	<ul style="list-style-type: none"> • Prioritize the changes and try to address those accordingly within project scope and time line. Extended work can be attached as future work. • Be prepared for the changes ahead of time, if the changes are to known. 		
<hr/>			
Risk	Unable complete tasks before given deadlines		
Risk level	High	Occurrence Frequency	Medium
Description	Sometimes, due to various reasons, tasks can't be completed within planned deadlines. This poses a risk when dealing with current and future work-		

	load.		
Mitigation	<ul style="list-style-type: none"> Track the remaining workload and prioritize remaining tasks according to an updated time-frame and proceed accordingly. Try to allocate slack time from future tasks and tackle and complete missed work-load. 		
Risk	Not being able to achieve the expected accuracy level for the student result predictions		
Risk level	High	Occurrence Frequency	Low
Description	Due to time constraints, changes to requirements and technological updates the developed system may not be able to achieve the expected accuracy rates for its predictions.		
Mitigation	<ul style="list-style-type: none"> Try to keep up with the latest developments of technical trends and approaches. Finish the prototype early so there is enough time to do any amendments if required. 		
Risk	Data lost due to hardware and software failures		
Risk level	High	Occurrence Frequency	Low
Description	The data of the project can be damage or lost due to hardware or software failures during lifetime of project. This poses a high risk on completion the project.		
Mitigation	<ul style="list-style-type: none"> Maintain daily backups and sync data using cloud solution such as Google Drive or Dropbox and OneDrive Test and debug the software continuously to maintain compatibility among new and old updates. 		

Table 3.2 - Identified potential risks and mitigation steps for Result Predictor system.

3.3 Development Methodology

System development methodology is used to structure, plan, and control the process of developing intended solution. When developing the system, correct development methodology should be identified in order to tackle cost, schedule overruns and to mitigate frequent requirement changes of the project. In order to identify the suitable development methodology to adapt on the project, several number of software development methodologies were evaluated and the details are stated below.

Waterfall development method is suited for projects that have fixed set of requirements, where research projects requirements are fluctuation for changes. Since this method not support to revisit or revise previous phase(s) work, traditional waterfall approach is not suited for research based project development. When considering the Rapid Application Development methodology, one must have certain level of prior knowledge on final product functionalities to create the prototype, where in a research project, prior knowledge of final solution is not know in higher degree of certainty at the beginning. Also, due to flexibility constraints on different project scopes adaptation, it is unwise to adapt this methodology for current research project. Similarly, Prototype Development methodology also requires having certain level of prior knowledge on final product functionalities to create the prototype. Another problem with this method is that, frequent communication with project stakeholders can lead time overruns and add over complexity to the project scope by trying to tackle requirements that are out of project scope. Agile methodology use both iterative and incremental development process heavily throughout development life cycle which may lead to time overruns. Furthermore, agile methodology is recommended to use in projects which have large and complex scope with huge number of stakeholders. Agile method is not suited for current research project due to the time constraints and there isn't any large number of stakeholders involved nor have large scope to address.

Spiral methodology allow to tackle fluctuation requirements on the software solution by allowing to start preliminary prototype with very little requirements and allow to redefine

prototype represents until the final desired product is created. In each prototype evaluation, a risk assessment is made to decide whether the project should be continue with current prototype or need a redefine to continue. Early risk identification helps to reduce the project risk level at early stages, which enables to deliver required output in timely manner. This suggests that the Spiral methodology would be the most suitable development methodology to adapt for current research project.

3.4 Research Methodology

A research approach can divide into two categories known as inductive research and deductive research. Deductive research approach focus on aiming and proving a theory while, inductive research approach focus on generating a new theory based on available data. The research project belong to deductive approach where aim of the project is to prove that the Student Result Predictor system can address the limitation define in the introduction chapter. The project will primarily proceed using more of a quantitative and less of qualitative research methods. Different quantitative & qualitative methods such as interviews, questionnaires with end-users, self evaluation will be performed throughout the project life-cycle activities.

3.5 Chapter Summary

At the initial stage of the chapter, importance of having a project management methodology when conducting a research project has mentioned. In this project PRINCE2 project management methodology has being applied hence project is conducted without any active stakeholders. Then the identified risks are mentioned and also the ways to tackle those risked are mentioned under risk mitigation. At the end of the chapter, Research methodology was discussed and why it's applied in the project discussed. In the next chapter author will discuss about System Requirement specification of the project which will discuss about the requirement engineering process of the project.

Chapter 4: Requirement Specification

4.1 Chapter Overview

In previous chapter under Literature review, discussed about literature which regarding to the selected domain, Student Result Prediction. In this chapter it discussed about the requirement gathering from the stakeholders which will need when constructing the project. Then this chapter will discuss how to analyze the gathered requirement from stakeholders and authorities. Finally, this chapter will discuss about categorizing functional and non-functional requirement which related to the project.

4.2 Requirement Gathering Process

In order to fulfill the task which carried out in the project multiple requirements gathering techniques are being used. These approaches helped author to gather information effectively and efficiently and also by above processes it helped to tackle limitations and mitigate those identified limitations. Following Table 4.1 provides an detailed overview of selected requirement gathering process and techniques.

Technique / Process	Literature Review
In Literature review author has identified what are the similar kind of studies which has been carried out so far and the techniques which used in those studies as in key technologies, methodologies, models, algorithms, architectures and strategies	
Advantages	<ul style="list-style-type: none">• It enables to identify about technologies which applicable in selected domain.• It helped to identify the parameters which need to consider in the study which helps in preparing questionnaires.
Disadvantages	<ul style="list-style-type: none">• This process consumed so much of time since previous works, papers, articles need to be identified and go through individually.• Since technology is evolving minute by minute most updated research papers might not be addressed by the time.
Technique / Process	Questionnaire

<p>Providing questionnaires to the target group is essential when carrying out research like this. From questionnaires there is a possibility of gathering real data which is being engaged with the end users. This method will increase the accuracy when implementing the system. During the research author has provided a questionnaire to the end users and stakeholders to gather necessary information</p>	
Advantages	<ul style="list-style-type: none"> • Questionnaires can gather target data effectively and efficiently. • Open ended questions can get most accurate answers without having any unwanted data. • Questionnaires will arrange the end results in organized manner which will helpful for easy analysis.
Disadvantages	<ul style="list-style-type: none"> • Once the questionnaire is given, end users / stakeholders answers might be limited only for that hence in order to get additional data, another data gathering process need to follow.
Technique / Process	Interviews
<p>During the requirement gathering process, several number of structured interviews being carried out by the author with target groups and end users. During the interviews identified parameters are being proved and the dependency of these parameters are verified.</p>	
Advantages	<ul style="list-style-type: none"> • This allows to clarify the doubts which related to the matter hence author can get direct answers from the end users perspective.
Disadvantages	<ul style="list-style-type: none"> • This method is not practical when addressing large group of people to mitigate that selected sample can be interview from the population.
Technique / Process	Data Gathering
<p>Test data and actual data are major important factor in data science driven projects. Gathering real data is significantly important in order to get an accurate result. In this project author needed to gather real data of students results from institution. Author has collected these data from a leading university in Sri Lanka where name cannot be specified under legal obligations.</p>	
Advantages	<ul style="list-style-type: none"> • Can be used to train machine learning model to do predictions. • Since these are real values, data can be used to check the accuracy of the model.

Disadvantages	Since collected data contain confidential information, collected data can be used under several restrictions.
----------------------	---

Table 4.1 – Overview of Data Gathering Process

4.3. Data Collection and Analysis

4.3.1 Questionnaire

As a part of data collection, an questionnaire was prepared and shared among target users, specifically undergraduate students. Questionnaire was shared by using emails and social media to gather adequate information from target users. From the questionnaire author focused to capture data which relates to the identified parameters under literature review. By sharing questionnaire among target group author was capable of capturing that information which has been used to train the model.

Limitations of the Questionnaire

Preparing a questionnaire and sharing it won't solve the problem of the researcher. There should be certain credibility of provided data. Since this questionnaire capturing anonymous inputs, people can fill this up with fake and irrelevant data which might leads to false prediction. So as the major limitation credibility of the feedbacks and responses can be identified.

Getting limited responses from the target users might be another problem. This might leads to bayes prediction since there are no adequate data for proper prediction. Getting limited responses from the users would be a major and common problem which researchers are being face in modern world. To avoid and overcome such issues there are some techniques to be applied.

Analysis of the Questionnaire

This questionnaire was created by the author using MS Office forms and distributed it to the end users such whom undergraduates in a leading university. Questionnaire has been taken by 304 individuals. Out of 304 responses 304 responses can consider as successful

responses where it contains relevant information. When considering about the unsuccessful responses there were student who didn't complete their semester exam due to various reasons. So those responses were taken down when training the model which will discuss at the end in the report .As a successful response rate 100% can be consider. Figure 4.1 illustrate the successful response rate.

$$\begin{aligned}
 \text{Successful Response Rate} &= \left(\frac{\text{Number of Completed Responses}}{\text{Number of Survery Responses}} \right) * 100\% \\
 &= \left(\frac{304}{304} \right) * 100\% \\
 &= 100\%
 \end{aligned}$$

Figure 4.1 Questionnaire Response Rate

Following table (Table 4.2) provides the feedback and findings of the questionnaire.

General Outcome	Feedback distribution of the questionnaire.
According to the survey which was conducted by the author, participants were undergraduate students who are following computing related degree programs in their university. When collecting data through the survey, both male and female students were participated and contributed to the questionnaire. When considering about the degree programs of the participants, they belongs to computer science, computer network degree programs.	
Finding 1	Average Study Hours Per Day

<p>This question is directly related to the identified parameter in the literature review stage. In literature review author has identified that student study hour as a direct parameter for the prediction. Out of all the responses 56% of participant study 3 to 4 hours per day in their undergraduate life, where 19% of them studying less than 2 hours, 14% of the participants are studying five hours per day and others study more than 5 hours per day. As a percentage 11%. Figure 4.2 is illustrating the responses for this question.</p>	
Finding 3	Residential Location
<p>Location where the student are staying has a direct impact for their examination results are being found out by the literature review hence this input was taken by the author by the target group. Out of all responses 58% of student are located at their permeant residence which means their home under parents' observation. While 33% of responders are staying at boarding places or local places and others at university hostel 9% as the exact percentage. Figure 4.3 is illustrating the responses for this question.</p>	
Finding 4	Parents Educational Qualification
<p>As pointed out in literature review parents education level has an impact to the student results hence this question is included to check how much impact it makes to the prediction. From the responses 60% of their parents obtaining no qualification at all as a higher educational qualification. Where 29% of them contained bachelors degree, 9% of their parents obtained masters and 2% of their parents obtain PhD qualifications. Figure 4.4 is illustrating the findings for this question.</p>	
Finding 5	Student Attendance
<p>Another highlighted factor from the literature review was the attendance of the student, which has a direct impact to the student results. Since it does make a direct impact, as the next question questionnaire includes that. Out of all responses 81% has more than 75% attendance, 16% of them having 60 % to 75% attendance 2% has more than 50% and less than 60% and 1% has less than 50% attendance. Figure 4.5 is illustrating the findings for this question.</p>	

Table 4.2 Feedbacks of the Questionnaire

Figure 4.2, Figure 4.3, Figure 4.4 and Figure 4.5 is illustrating the distribution of responses to the above mentioned questions in the questionnaire.

1. As an average time how much you have spent for studies per day

● Less than 2 hours per day	58
● 3 to 4 hours per day	170
● 5 hours per day	44
● more than 5 hours per day	32



Figure 4.2 Feedbacks of the Question 01

2. Residential Location in 1st year

● At home	176
● At Boarding Place	100
● At University Hostel	28

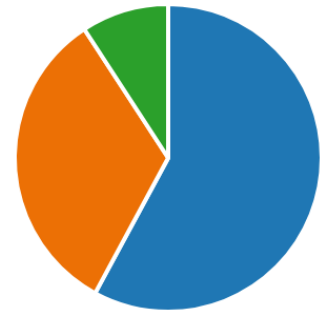


Figure 4.3 Feedbacks of the Question 02

3. Does your parents obtained higher education qualifications

● No	182
● Bachelors Degree	89
● Masters Degree	26
● Doctor of Philosophy	7

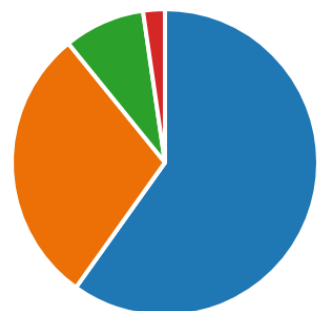


Figure 4.4 Feedbacks of the Question 03

4. How was your attendance for 1st year lectures?

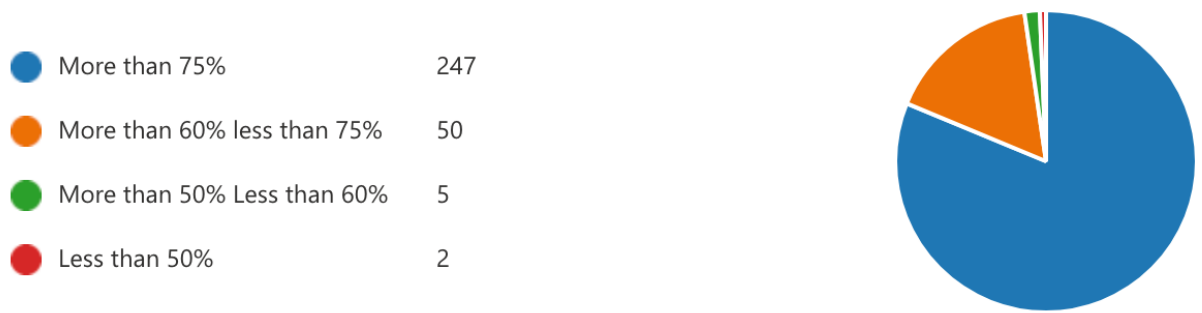


Figure 4.5 Feedbacks of the Question 04

4.3.2 Interviews

During the period which the author carried out the research, number of structured interviews are being conducted with end users and stakeholders in order to get their direct opinion regarding about questionnaire as well as to prove the selected parameters has significant impact. During the interviews author was able to understand some risk factors which were represented in previous chapter in detail.

Limitations of Interview

Below Table 4.3 illustrate the limitation of interview which conducted during the period of research.

Limitation	Credibility of Data
n	Since interviews are being conducted face to face, there were some students who didn't share actual data and information due certain facts.
Limitation	Lack of Participants
n	During the interviews since they share the information with person who conducting the interview, interviewees has less tendency for participating such interviews.

Table 4.3 Limitations of Interview

4.3.3 Summary of Requirement Gathering

Following table (Table 4.4) illustrate the findings during requirement gathering process as a summary.

Findings	Literature Review	Interview	Questionnaire
Usage of relevant machine learning algorithm in order to predict student performance and results.	✓		
Approaches to the project as input medium	✓	✓	✓
Identification of parameters in order to predict student result.	✓	✓	
Identification of questionnaire which need to be distribute among users	✓	✓	
Data set which need to be use to train the machine learning model		✓	✓
Identification of technologies to implement the system	✓	✓	
Identification of user need and their behavior patterns when using the application		✓	

Table 4.4 Summary of Requirement Gathering

4.4 Analysis of Stakeholders

When considering about stakeholders and end users of the application, there are several people can be identified as users. Following figure (Figure 4.6) represents the onion model of the proposed system.

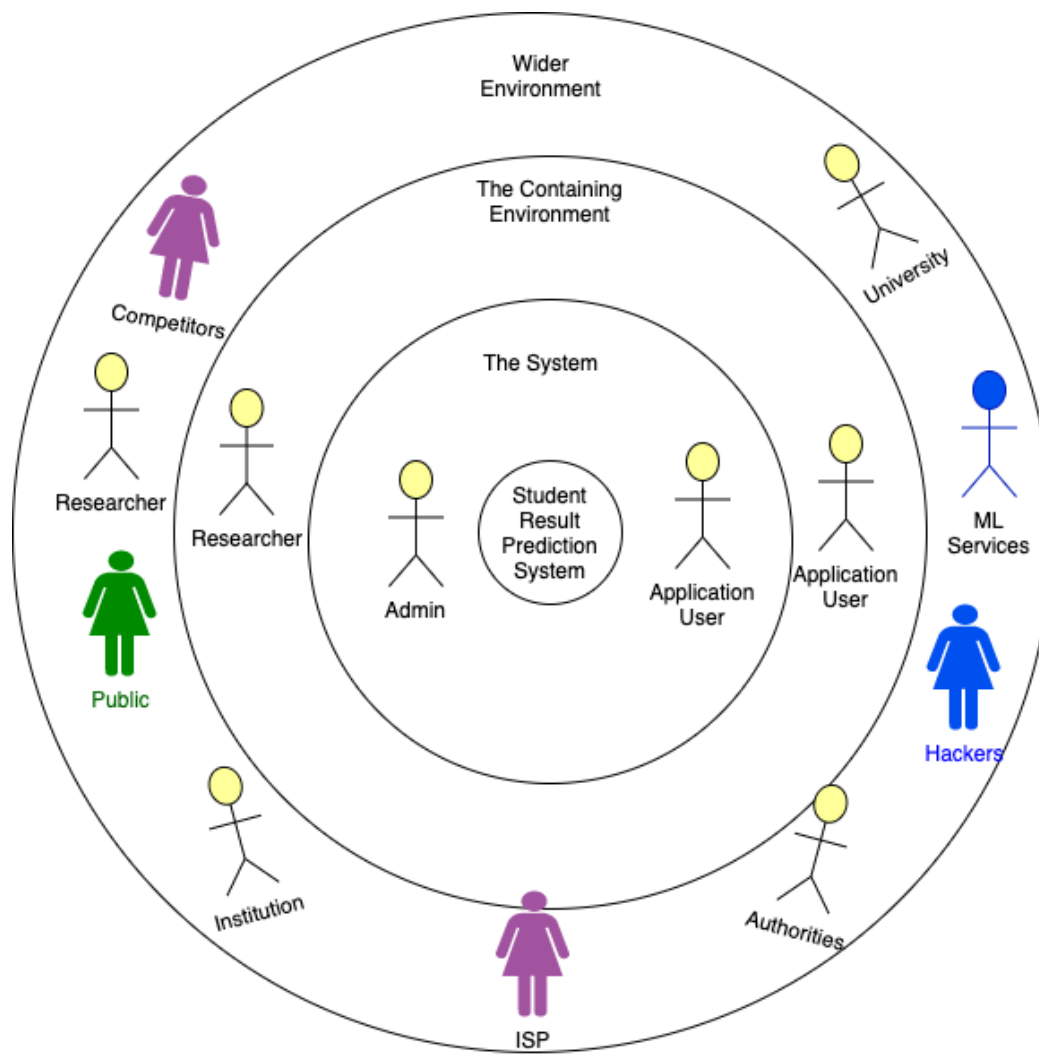


Figure 4.6 Onion model of the proposed result prediction system

In above figure 4.6 expressed the user which involved with the system using an onion model. Below table (Table 4.5) represents their individual contribution and task which will carried out in the system.

User / Stakeholder	Responsibility	Contribution
Admin	Overall administration of the system	Admin is responsible for manage user profile inside the application Admin is responsible for managing the questionnaire / quiz inside the application

Application User / User	Operator of the system	User will fill out the given questionnaire and view results based on predicted values.
Competitors	Negative Stakeholder	Try to understand the drawbacks and negative features and functionalities in the system
Hackers	Negative Stakeholder	User group will perform unethical operations inside the application.
Public Users	Negative / Positive	Public users will provide negative and positive feedbacks regarding the system based on their user experience
Authorities	Authorization	This user group will provide authorities and regulations which related to the proposed system including legal and ethical considerations
Institutions	Beneficiary	These user group can use the system and experience it inside their institute to do predictions
Universities	Beneficiary	Universities would become a benefited user group where they can use the system in order to do predictions inside the organization.
Researchers	Positive / Negative Stakeholder	Researchers would consider this as a positive or negative project based on their opinion hence this user group become natural stakeholder.

Table 4.5 Summary of Stakeholders and Responsibilities

4.5 Use Case Diagram of the System and Description

In this section, it describes about the use case diagram of the proposed system and how the flow of application is happened until user get an prediction. In order to get a predicted values first user should get registered with the system by entering basic information. Once user get registered with the system user can get logged in with the system in order to face the given questionnaire. Questionnaire will gather information from user which basically focused on identified parameters in the research stage. Once user completed the quiz user will be able to see the predicted values along with the data visualization which provides better understanding regarding the outcome. Following figure (Figure 4.7) illustrate the use case diagram which involved with the system.

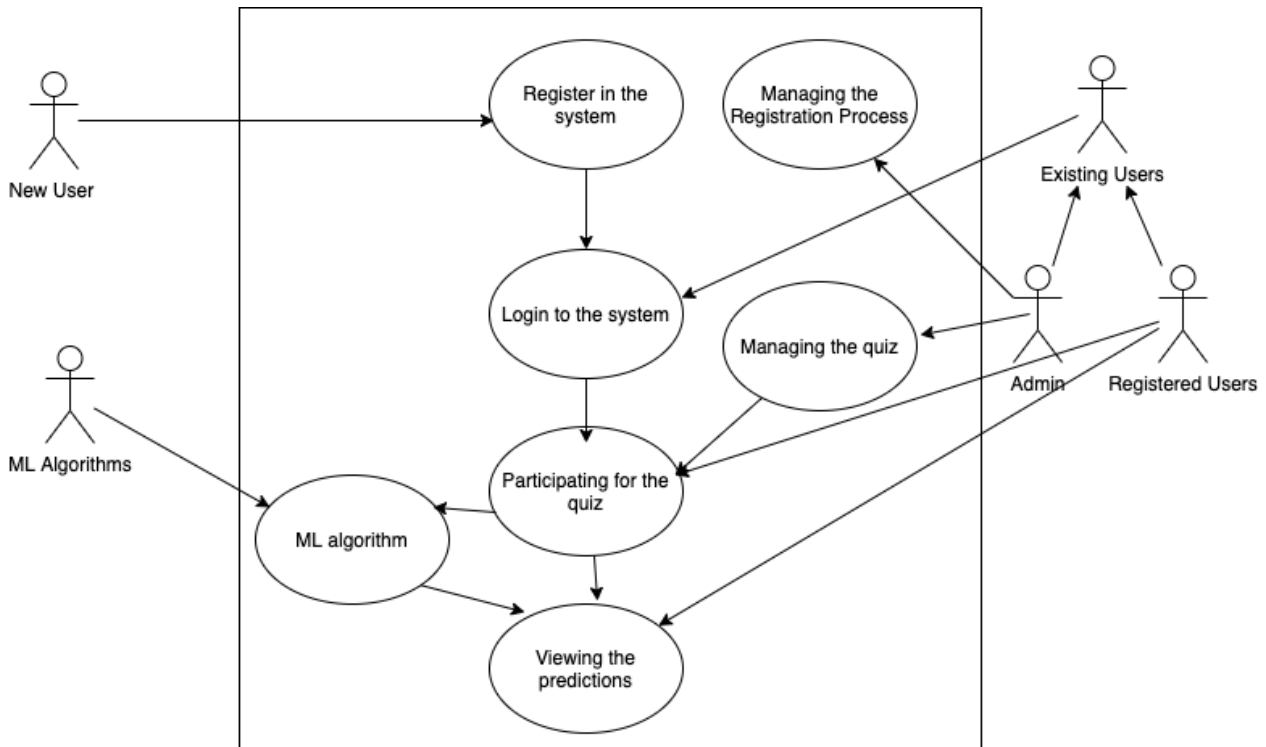


Figure 4.7 Use Case Diagram of Student Result Prediction System

4.6 Functional Requirement

Based on the functionalities of the system following functional requirements are being identified and also stated the priorities of the identified functional requirements.

- New users should get registered with the system in order to do predictions using the system. This functional requirement identified as critical priority hence without registering the user in the system, user won't be able to use the system. This functional requirement is mandatory.
- Once user get registered and existing users should login to the system in order to get the predictions. In this functional requirement it identifies as a critical priority hence without registering in the system user won't be able to get the predictions. So this a mandatory process.
- Registered users should fill the quiz. This is a critically identified functional requirement since without filling the quiz users won't get the prediction results.

- View predicted values. Based on the quiz input data, user would be able to see the predictions and this is a critical functional requirement where this process is mandatory during the operation.
- Data visualization when representing predicted values. This functional requirement again would become a critical functional requirement where all users should get this as a mandatory process.
- Machine Learning algorithms should process the data which is inserted by the user in order to do prediction, which is identified as a mandatory functional requirement hence it is listed under critical prioritize functions
- Admin should be able to manage registered users and their information. Which is identified as a functional requirement which is in an important level where this functional requirement is not a mandatory process hence it falls under medium priority.
- Admin should be able to manage the quiz and questionnaire, and this is not a mandatory functional requirement hence this falls under important priority section.
- When presenting the output to the user results should contain multimedia components such as images, videos etc. This is identified as a functional requirement but this falls under optional category hence this is not mandatory when providing results.
- Admin should be able to add questions to the questionnaire which again a functional requirement falls under Important category where it's not mandatory during the process.
- Admin should be able to edit questions to the questionnaire which again a functional requirement falls under Important category where it's not mandatory during the process.
- Admin should monitor the system time to time in order to make sure that the system provides relevant and efficient information and predictions which falls under critical priority list.

4.7 Non-Functional Requirements

Based on the proposed system following nonfunctional requirements were identified.

- Accuracy of the system is a non-functional requirement which identified in the proposed system. If the accuracy level is high users can get more accurate predictions from the system.
- Performance of the system is another nonfunctional requirement which identified during the process. How system perform during the execution would be a key factor if it's running without exceptions, system would provide better user experience.
- Usability of the system is another identified nonfunctional requirement. In the system there is a quiz a to be filled up by the user. When user attempting the quiz, usability is a significant factor to be consider as non-functional requirement.

4.8 Chapter Summary

This chapters refers how requirement gathering process has been carried out during the project under various techniques and constraints. Mainly author has conducted interviews, questionnaires and gather information from institution get the values to be predict. Since system need real data for train the model, author focused on gathering real data from target group and train the model using them. Data collection in lined with certain rules, limitations and legal boundaries hence in this project it addresses confidential data. At the end of the chapter it discussed about analysis of gathered requirements and functional and non-functional requirements engaged with the proposed system. Next chapter will discuss the system architecture and design in detail.

Chapter 5: System Architecture and Design

5.1 Chapter Overview

Previous chapter highlighted about requirement gathering and analyzing process of the requirement. This chapter focus on how the architecture of the system is built based on gathered requirement and design of the system. During this chapter techniques and technologies which used to implement the system are being addressed by the author.

5.2 High-level design of the System

This topic discuss about the system design diagrams and the architecture of the system. When considering about the proposed system, this system is a standalone system which running inside the computer without using any online material or source such as online database, online service, or any API. System integrated with a database which contain all user data and registration information. As mentioned in above chapter new users' needs to get registered with the system in order to use the features and facilities of the system. When user registering in the system all user information are taken by the database which integrated with the system, and those data will use when user get login with the system next time. Also inside the local database quiz materials are being located. Such as questions, responses which provided by the users etc. Those data can be used for future predictions and to test the accuracy of the system in future implementations. When user insert data to the system all the data will be passed to the integrated database and those information will be stored inside the database for future predictions. Also system integrated with front end application where users can insert data through the application and also users will be able to see predicted values through the user interface. Users will be able to see all the predictions by using the front-end application with visualized data. system will be used by users and admins where users can insert data and do the predictions. Production values are being calculated by the machine learning algorithm based on user input data. Inside the application the user input data will be processed by the machine learning algorithm under defined parameters. once the algorithm feed with proper parameters and values machine learning algorithm is capable of providing the predicted

values. In this project author has been used three classification algorithms in order to do the prediction. Inside the project author will be comparing these three algorithms and output will be provided by using the optimum algorithm based on the score which provided at the implementation level. As another option in the application if users want to see the predicted value based on different algorithms users are capable of selecting desired algorithm from the application.

when considering about the admin users admin users are capable of managing the entire application such as editing the questions, managing feedbacks, managing users etc. which mentioned in above chapter. in the following figure mentioned the architectural diagram of this proposed system.

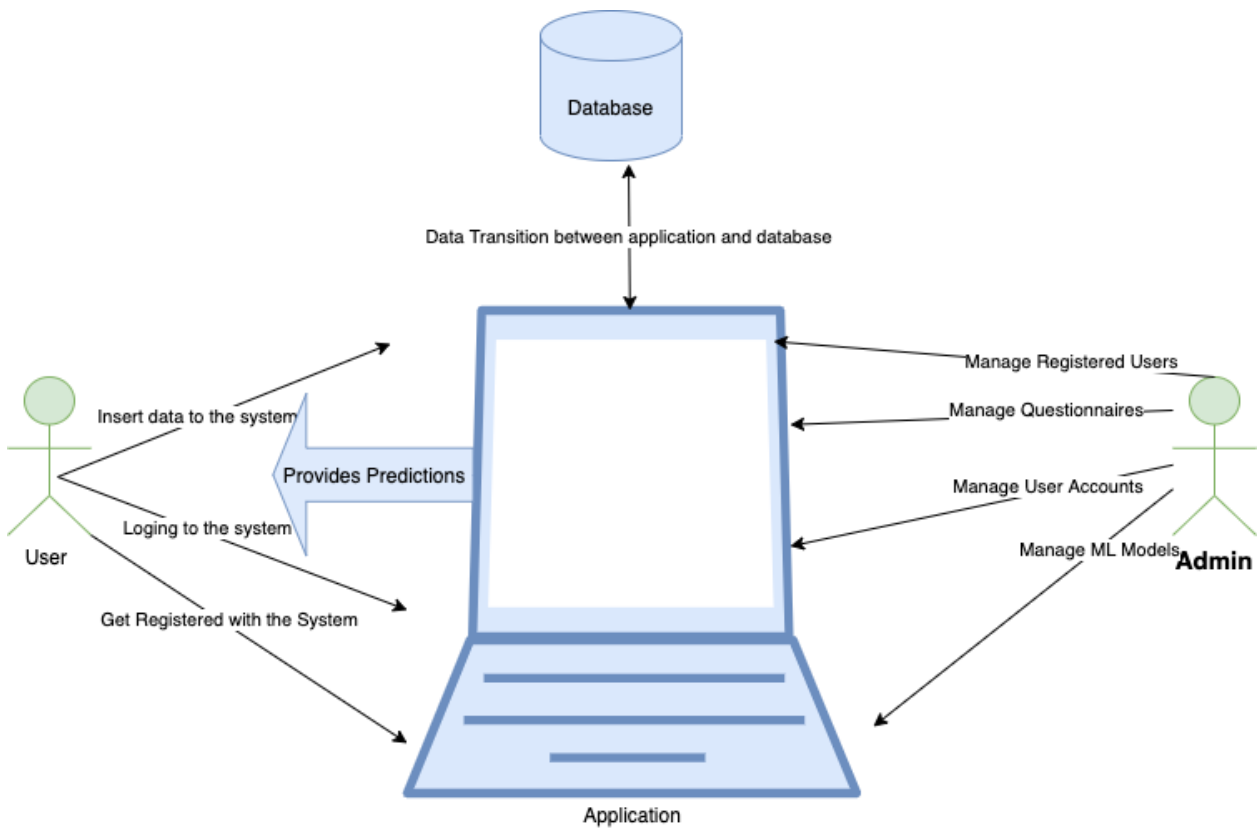


Figure 5.1 System Architecture of the Proposed System

5.3 Machine Learning Algorithms

In this project There are four machine learning algorithms use in order to the predictions. These three algorithms are being identified During the literature review when going through the similar kind of works. Motioned and used algorithms inside the projects are Decision tree classification algorithm, Naïve bayes classification algorithm, Support vector machine algorithm and Random Forest classifier. All mentioned algorithms are classification algorithm where the based is equal when predicting the values. Inside the project author has compared the algorithms and their accuracy when predicting student results based on identified parameters. For future predictions author will be using highly scored algorithm and implement it as the primary algorithm of the system. Rest of other algorithms will be integrated with the project for comparisons in predicted values.

5.4 ER Diagram of the application

This application consists with a local SQL database which contained four tables where it contained user information's, survey questions, User feedbacks and predictions. Following figure will illustrate the ER diagram which used inside the system.

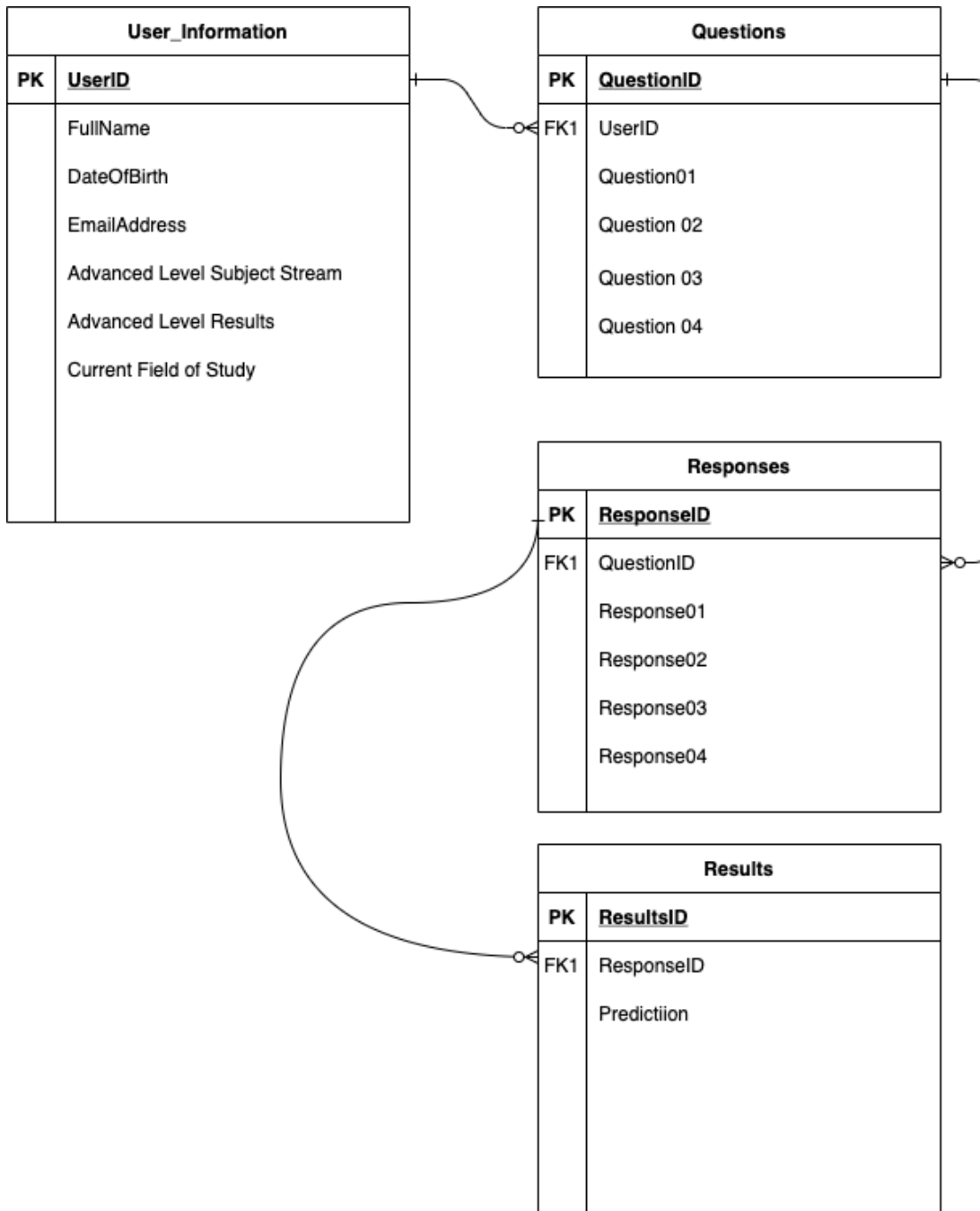


Figure 5.2 ER Diagram of the Proposed System

5.5 Chapter Summary

This chapter discussed about the design and architecture of the proposed system by mentioning the techniques, technologies and approaches which has been used to implement the system. Also chapter includes the architectural diagrams which illustrate a clear opinion regarding the proposed system. In this chapter, the ER diagram of the system is presented which provides a clear understanding how data flow happens inside the application. In next chapter author will discuss about the implementation of the system in detail.

Chapter 6: Implementation

6.1 Chapter Overview

Previous chapter explained how the architecture of the project is laid out and what type of design and architectural approaches are being used inside the application. This chapter focuses on the system implementation along with the relevant and used technologies inside the application. Also this chapter will discuss the relevant code segments which are used inside the application.

6.2 Selection of Technology

When doing a research project, selecting proper technology is very important to come up with effective output. Hence technology selection was very critical in this project. The author has gone through several research papers, articles, publications and similar studies to finalize the technologies which need to be used inside the project. All referred articles, documents and papers are mentioned under the literature review chapter with adequate explanation.

6.2.1 Selection of Machine Learning Algorithms

During the literature review there were three algorithms being identified which are Decision Tree classification algorithm, Naïve Bayes classification algorithm and Support Vector Machine (SVM) algorithm. During the self-studies the author has identified Random Forest classifier where it can be used to do prediction based on classifications. This project has concluded by using the above mentioned algorithms and in addition to that the author tested results with regression algorithms to compare results to identify which algorithm is optimum in order to do the predictions. During this study the author has identified that these four algorithms could be used in order to do student result prediction. To implement these four algorithms the author has used Jupiter notebook as an IDE and Python programming language. In Python there is an inbuilt library called scikit-learn during this study the author has been using this particular library as a framework to implement the system. In this

framework there are a number of machine learning algorithms available and those algorithms can be used to predict and for the implementations. Among these available machine learning algorithms author has used decision tree classifier, support vector machine, naïve bayes and random forest classifier to do the prediction inside this study. During implementation all machine learning algorithms are being trained by training data set which gathered during requirement gathering process. In order to train these algorithms by using training data set first dataset need to be arranged in certain format. This process is known as cleansing data. In following topic author will discuss how the data preparation process went through during the project.

6.2.2 Data Pre-Processing

From requirement gathering process author was capable to gather adequate data in order to train the machine learning algorithms. As mentioned in above requirement analysis chapter data was collected by using several methods. when collecting these data from several methods, data is not being in the correct format. When doing data science related project the data preprocessing is a necessary topic and necessary process to be done before the implementation. When collecting these data using questionnaires, interviews and also from relevant repositories data might be in different formats. During this study author has faced that difficulty. The collected data contained number of null values, unwanted information, also data was not in a proper format in order to train machine learning algorithms. as the very first step all the data arranged in one specific format where that can be used to train machine learning algorithms as the initial step. As Data preprocessing methodology author has addressed all missing values and values which are not in the correct format to train algorithms. after completing the process, all collected data from different sources are located in one excel CSV file where that file can be used to train machine learning algorithms. Also collected data could be use to test the accuracy off the developed model and identify which algorithm provides the better outcome.

6.2.3 Training the Algorithms

After completing data preprocessing step processed data are being used to train all four machine learning algorithms. As mentioned previously since Python language is being used to this particular study there is an inbuilt library which is located in scikit learn framework. during the implementation this particular framework is being used to train machine learning algorithms. As initial step collected data are divided into two segments such as training and testing datasets. By using training datasets all for machine learning algorithms are trained. Then by using testing data set the accuracy of these machine learning algorithms was checked. when dividing these data into two segments inbuilt library in scikit learn framework is been used. this library will divide the entire data set randomly in to training and testing segments. That data are being used to train machine learning algorithms. In Python there is inbuilt library called pandas

6.3 Code Implementation

In this topic author will discuss the implementation of machine learning algorithms in Python programming language. Each and every machine learning algorithm which has been used in this particular study are listed in following figures.

```
In [300]: import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score

Mydata_New = pd.read_csv('/Users/pramudyathilakaratne/Desktop/MSc Project/Individual Project/Data Set/Final Data Sets/F
Mydata_New

Mydata_New = Mydata_New.loc[:, ~Mydata_New.columns.str.contains('^Unnamed')]
Mydata_New

x_New = Mydata_New.drop(columns=['Grades', 'Total', 'Average'])
x_New

y_New = Mydata_New['Grades']

x_train, x_test, y_train, y_test = train_test_split(x_New, y_New, test_size=0.1)
x_New
y_New

model_New = DecisionTreeClassifier()
model_New.fit(x_train, y_train)
Prediction_New = model_New.predict(x_test)

Prediction_New

score = accuracy_score(y_test, Prediction_New)
score
```

Figure 6.1 Implementation of Decision Tree Classifier

In figure 6.1 did mention the implementation of decision tree classifier, data segmentation and training the model and predicting the output. This classification model is selected during the literature review based on proved factors from previous studies.

```

In [297]: import pandas as pd
from sklearn import svm
from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score
Mydata_New = pd.read_csv('/Users/pramudyathilakaratne/Desktop/MSc Project/Individual Project/Data Set/Final Data Sets/F
Mydata_New

Mydata_New = Mydata_New.loc[:, ~Mydata_New.columns.str.contains('^Unnamed')]
Mydata_New

x_New = Mydata_New.drop(columns=['Grades', 'Total', 'Average'])
x_New

y_New = Mydata_New['Grades']

x_train, x_test, y_train, y_test = train_test_split(x_New,y_New,test_size=0.1)
x_New
y_New

model_New = svm.SVC()
model_New.fit(x_train,y_train)
Prediction_New = model_New.predict(x_test)

Prediction_New

score = accuracy_score(y_test,Prediction_New)
score

```

Figure 6.2 Implementation of SVM Classifier

Figure 6.2 describes the implementation of support vector machine classification algorithm, data segmentation, training the machine learning model and doing the predictions.

```

In [191]: import pandas as pd
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score
Mydata_New = pd.read_csv('/Users/pramudyathilakaratne/Desktop/MSc Project/Individual Project/Data Set/Final Data Sets/F
Mydata_New

Mydata_New = Mydata_New.loc[:, ~Mydata_New.columns.str.contains('^Unnamed')]
Mydata_New

x_New = Mydata_New.drop(columns=['Grades', 'Total', 'Average'])
x_New

y_New = Mydata_New['Grades']

x_train, x_test, y_train, y_test = train_test_split(x_New,y_New,test_size=0.1)
x_New
y_New

model_New = GaussianNB()
model_New.fit(x_train,y_train)
Prediction_New = model_New.predict(x_test)

Prediction_New

score = accuracy_score(y_test,Prediction_New)

score

```

Figure 6.3 Implementation of Naïve Bayes Classifier

figure 6.3 describes the implementation of naive bayes classification algorithm, data segmentation, training the model and doing the predictions.

```
In [296]: import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score
Mydata_New = pd.read_csv('/Users/pramudyathilakaratne/Desktop/MSc Project/Individual Project/Data Set/Final Data Sets/F
Mydata_New

Mydata_New = Mydata_New.loc[:, ~Mydata_New.columns.str.contains('^Unnamed')]
Mydata_New

x_New = Mydata_New.drop(columns=['Grades', 'Total', 'Average'])
x_New

y_New = Mydata_New['Grades']

x_train, x_test, y_train, y_test = train_test_split(x_New, y_New, test_size=0.1)
x_New
y_New

model_New = RandomForestClassifier(n_estimators=5, max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fra
model_New.fit(x_train, y_train)
Prediction_New = model_New.predict(x_test)

Prediction_New

score = accuracy_score(y_test, Prediction_New)

score
```

Figure 6.4 Implementation of Random Forest Classifier

figure 6.4 describes the implementation of random forest classifier, data segmentation, training the algorithm and doing the predictions.

Apart from above mentioned four classification algorithms author has used regression algorithms in order to compare and contrast the predictions and finding the optimum algorithm in order to predict student results. As regression algorithms author has implemented decision tree regression, linear regression in order to compare accuracies. implementation of these regression algorithms are listed under appendices at the end of the project report.

6.3.1 Front End Implementation

front end of this application is developed by using simple web development tools for the user interfaces. By using the front-end application users, admins can get registered and use the system. This front-end application will help users to get better user experience rather than using some online arguments in Python programming language. Author of this study wanted to provide a better user experience to the users who are involved in with this study and the project. Once the values are being predicted by back-end machine learning algorithm all these predictions will be visible to users by using this front-end application. Data visualization, selection of different algorithms are possible by using this front end application where user can have better and effective user experience.

6.3.2 Database Implementation

database of this application is being developed by using is structured query language as it mentioned in architectural diagram chapter. Based on the provided ER diagram database has been created and connected to the front-end application where all using information are being transferred to the database and store for the future implementations and predictions. Having a real time database is very effective in modern day application development hence parameters and factors are being changed often. If there is a modification to be done inside the application admin has to change the relevant information which is located inside the database which will affect in front end application. With this feature admin can have a better application management when it comes to this particular application. Also, all use information, their questionnaire information are being located inside the database and also respective prediction values also stored inside the database in order to calculate the accuracy and efficiency of the system. If there wasn't a back-end database each and every process which mentioned above have to be done manually by the admin of the system.

6.4 Chapter Summary

this chapter discussed about the implementation of the system in detailed manner.

Basically, application consist with machine learning algorithms which running behind the front end application. Front end application is being connected to a back-end database and also to the machine learning algorithms which has been used to do the predictions. In this chapter detailed explanation is being provided of the implementation of each and every machine learning algorithm, data processing, database creation and front-end application development. in next chapter author will discuss how the system is being tested once it implemented

Chapter 7: Testing

7.1 Chapter Overview

In previous chapter it highlighted about the implementation of the system which carrying through this study. Once the system is implemented the next important thing is to do the testing and figure out whether the expected values are coming through the system. In this study as mentioned in previous chapters author is using four classification algorithms with identified parameters in order to predict student results and come up with the optimum output. Since there are four classification algorithms are being used in order to finalize the optimum algorithm certain amount of testing needs to be done. In this chapter author will discuss how the testing process has been done throughout the implementation of this study.

7.2 Goals and Objectives of Testing

When testing a certain system there are certain objectives and goals which expected by the developer of the system. In previous chapters author has mentioned functional and non-functional requirement of the application also objectives of the application. In testing interim the most important thing is to check whether above mentioned functional and non-functional requirements are being covered inside the developed system. And also by using system testing methodology, the quality of develop system can be improved based on the testing feedbacks.

7.3 Procedure of Testing

procedure of testing can be divided into two main segments such as functional testing and non functional requirement testing. When considering about the functional requirement testing all mentioned functional requirements need to be addressed in the implemented system. as a testing mechanism each mentioned functional requirement are being tested individually and proved that each functional requirement it's properly working inside the application. Next testing category is mentioned non functional requirements, all mentioned non functional requirements are being tested individually and proved that each and every

non functional requirement which addressed in requirement gathering process chapter are being fulfilled by the application.

7.3.1 Testing the Machine Learning models

The most important part in this particular study is to identify the optimum machine learning algorithm to use when predicting student results based on selected parameters. As mentioned in literature review the identified parameters are study hours, living location, parents education qualification and attendance of this student. By using these parameters machine learning algorithms such as decision tree classifier, naive bayes classifier, support vector machine classifier and random forest classifier are being used inside this study. Out of these mentioned machine learning algorithms each and every algorithm is capable of predicting student result based on the given parameters. From these algorithms it predicts the student grade whether the student has obtained A +, A or A- based on provided parameters.

When testing this machine learning algorithms different datasets are being used since the algorithm should be checked in different conditions. As mentioned in implementation chapter the data have been segmented into different categories by using the Python libraries. When running the models different datasets are being used to trained the machine learning algorithms and also different datasets are being used to do the testing. Since the testing and training are being done by different datasets the results or the score might get fluctuated time to time.

In following figure it mentioned the results which taken when testing the system by using different datasets. As mentioned earlier five datasets are being used to train and test the system. In following diagram express each and every algorithms response for those five datasets.

Decision Tree Classifier

Decision tree classifier algorithm is been tested by using five different datasets and taken the output by using machine learning score generator. This score generator is been located inside scikit learn framework. By using this particular framework there is a possibility of calculating the accuracy of the prediction model based on input data set and the trained data set. In following figure (Figure 7.1) it illustrate the score which obtained by decision tree classifier in five instances.

Test Attempt	Score
1	0.57812
2	0.6123
3	0.5989
4	0.61221
5	0.5981

Table 7.1 scores for Test Attempts in Decision Tree Classifier

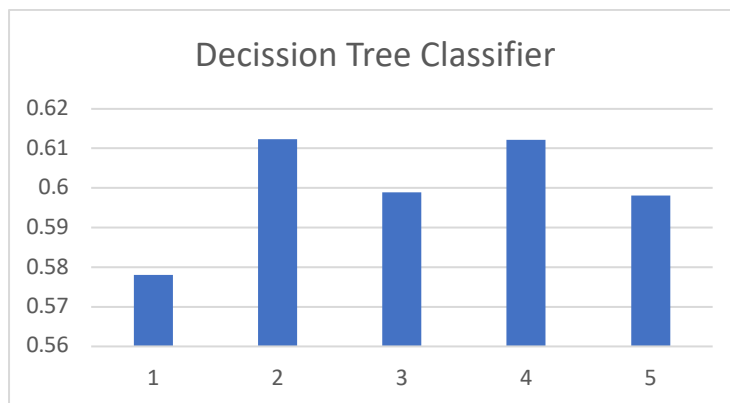


Figure 7.1 scores for Test Attempts in Decision Tree Classifier

Support Vector Machine

as the second algorithm off the system SVM classifier has been used to do the predictions inside the application. SVM machine learning algorithm Was identified in literature review as a algorithm which can be used to predict student results based on certain parameters. In this study SVM algorithm has been implemented and tested under test data said which has been used in previous testing scenario. Same as decision tree classifier SVM also trained by using test data set and the score has been taken in five instances as mentioned below. In figure 7.2 graphical lustrated the score values which obtained by this SVM classifier

Test Attempt	Score
1	0.5123
2	0.52789
3	0.49891
4	0.51232
5	0.52331

Table 7.2 scores for Test Attempts in SVM Classifier

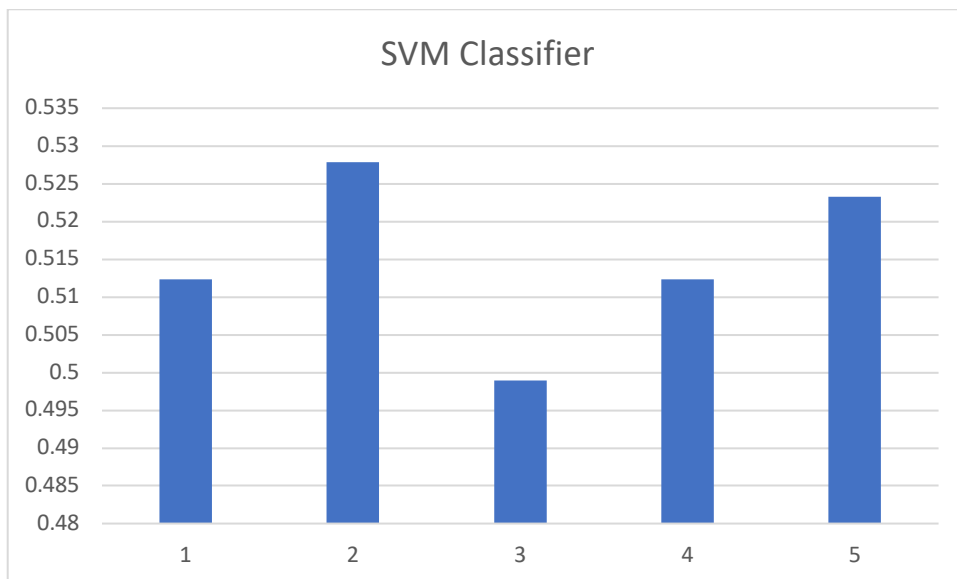


Figure 7.2 scores for Test Attempts in SVM Classifier

Naive Bayes classification algorithm

same as previous instances naive bayes algorithm also being tested by using five datasets which has been used to train the model as well. In following table and the figure illustrate the distribution of the score which obtained by Naïve bayes algorithm at the implementation.

Test Attempt	Score
1	0.20121
2	0.2113
3	0.2112
4	0.1898
5	0.19871

Table 7.3 scores for Test Attempts in Naïve Bayes Classifier

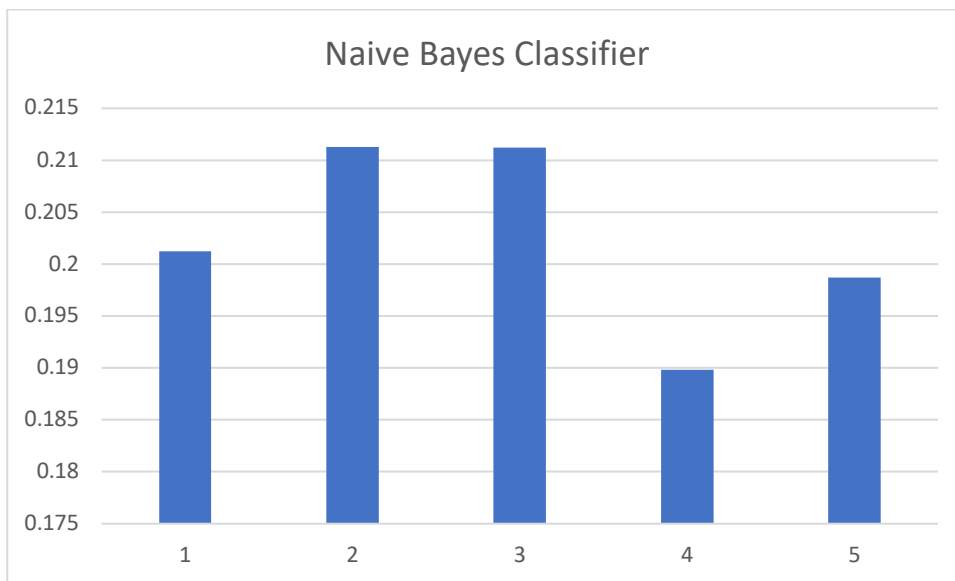


Figure 7.3 scores for Test Attempts in Naïve Bayes Classifier

Random forest classifier

random forest classifier is the 4th classification algorithm which has been used to do the predictions. As previous occurrence this algorithm is also trained by same data set and the distribution of the score has being taken in five instances. Following table and the figure illustrate the distribution of the marks in those five instances.

Test Attempt	Score
1	0.5671
2	0.5891
3	0.578954
4	0.5219
5	0.51432

Table 7.4 scores for Test Attempts in Random Forest Classifier

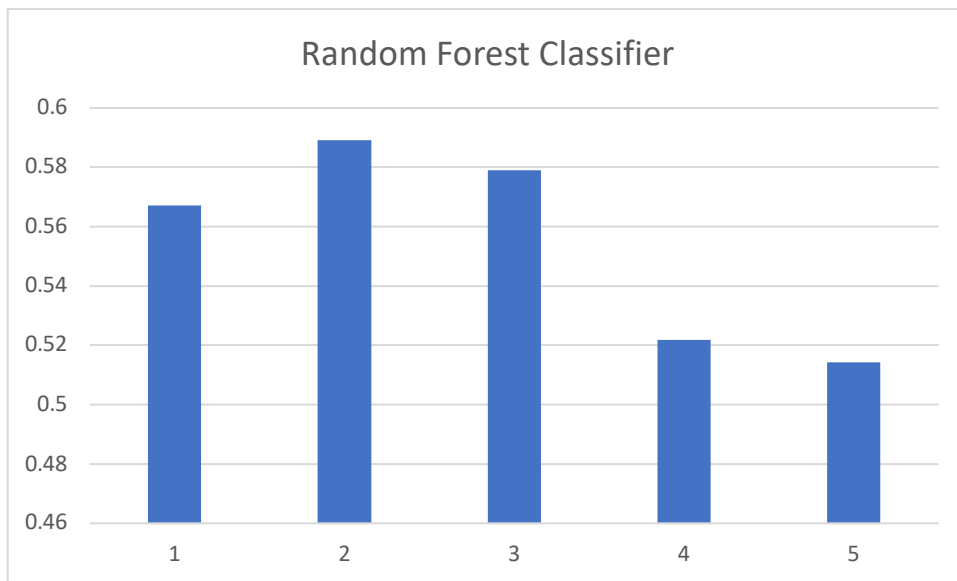


Figure 7.4 scores for Test Attempts in Random Forest Classifier

when comparing above four algorithms scores, decision tree classifier and random forest classifier are providing best scores out of these four algorithms. naive bayes algorithm his algorithm which provides the least scores among all four algorithms. For easy understanding in below table and figure explains score distribution of all four algorithms in five instances.

Test Attempts	Decision Tree	SVM	Naïve Bayes	Random Forest
1	0.57812	0.5123	0.20121	0.5671
2	0.6123	0.52789	0.2113	0.5891
3	0.5989	0.49891	0.2112	0.578954
4	0.61221	0.51232	0.1898	0.5219
5	0.5981	0.52331	0.19871	0.51432

Table 7.5 Score for Distributions of Classification Algorithms

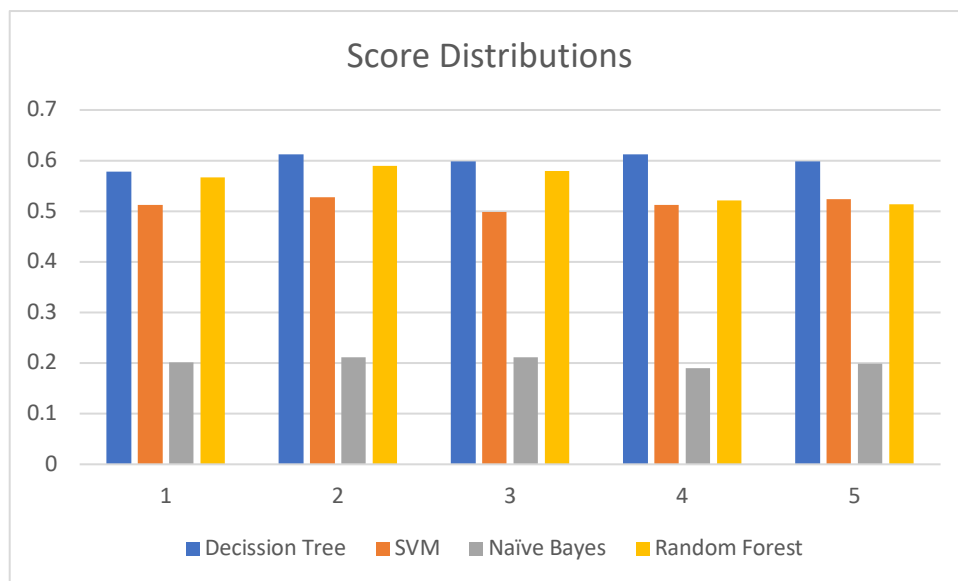


Figure 7.5 Score for Distributions of Classification Algorithms

By observing above score distribution decision tree algorithm and random forest algorithm are the algorithms which are providing best scores in all five instances. Based on that information decision tree algorithm and the random forest algorithm could be used in order to do the prediction of student results under selected parameters. These algorithms can be fine tuned to get more accurate results and that is being discussed under evaluation chapter in detail.

7.4 Chapter Summary

This chapter focuses on testing phases of the developed system under different conditions. When considering about the testing most important section is to address the coverage of functional and non functional parameters which has been mentioned under requirement gathering topic. This project was capable of covering up all provided functional and non functional requirements at the beginning of the project as well as a beginning of the dissertation. Since this application involved with machine learning algorithms accuracy of this machine learning algorithm is another important factor which need to be considered under testing chapter. As mentioned in detail these four algorithms are performing and providing outputs based on the input values. The accuracy of these output values need to be considered when considering about the efficiency of this application. As mentioned in above diagrams and figures four algorithms are providing different sort of values under provided parameters. The next chapter it discusses about evaluation process on various aspect of the project.

Chapter 8: Evaluation

8.1 Chapter Overview

Previous chapter focused on testing phases of the developed system under various parameters. Testing has been done by using number of datasets while training the algorithm at the mean time. In previous chapter it discuss about the testing of the application in detail. This chapter is focusing on evaluation of the system by validating functional and non functional requirements which has been mentioned at the beginning of the project. At the end of this chapter a self evaluation is being included regarding the student result prediction system.

8.2 Evaluation Process

When considering about an application it should take some input value and process it inside the application and provide an output based on certain parameters. In student result prediction system student study hours, attendance, parents education level and living location has been considered as the parameters where the prediction is being done by using four machine learning algorithms such as decision tree classifier, support vector machine, naïve bayes and random forest classifier. When evaluating the application as evaluation criteria it needs to be make sure that's all given requirements which listed under functional and non functional requirements are being covered or not.

when considering about the functional requirement there should be an application to be logged in based on user input username and password and once user has logged into the system the system should be capable of taking user input values as parameters in order to do the prediction. If the user is not a registered user there should be a way to register with the system once user get registered with the system user can use all the features and facilities which provided by the system itself. In order to manage user behaviors and technical management there should be a person with full authorities which known as the admin. This system contained within admin who can do the modification to the application when needed, also admin can manage all the users who registered with the system. The most important part of this study is implementation of the machine learning algorithms

which has been used to do the prediction inside the application. Machine learning algorithms are being implemented by using Python programming language with the help of Jupiter notebook IDE. implemented algorithms are being trained and tested by using collected data in requirement gathering process. Data which has been collected in the requirement gathering process are being successfully used to train and to test the algorithm in order to do the prediction by the system itself. Once user run the application user need to enter the answers for the given questions by the system itself and once user provided these answers system itself is capable of predicting the results of that particular student under given conditions.

This study is being carried out as a dissertation project to fulfill the requirement of the masters degree program where it needs to be implemented data science related application as a research. In this particular study author has decided to find the optimum algorithm in order to predict student results based on identified parameters. Parameters identification an algorithm identification has been done inside the literature review where author had to go through multiple research articles, research papers, previous studies in order to identify optimum parameters and algorithms for this particular study and the project. When identifying parameters author had to go through certain limitations and boundaries since this project is related with confidential data. Since the application is working with confidential data the project is being restricted into a certain level. For an example when analyzing student results and also when collecting student results author had to go through certain legal considerations with the data collected organization. When it comes to the structured interviews again some sort of difficulty has been raised since end users and stakeholders are not willing to share the accurate and the correct information hence there is an issue with the credibility of data which has been gathered. Those can be considered as the obstacles and the difficulties beach author has faced during this study and in order to mitigate those mentioned risk and difficulties author has taken necessary steps.

When considering about the functional and non functional requirements which mentioned in chapter five are been concluded by the author at the end of the project. The main issue which the author has faced during the project was the accuracy of the developed algorithms. In order to increase the scores Algorithm can be trained by using large number of data set then the scores will improve significantly. As another step which can be taken to

improve the score is feature engineering where machine learning models can be fine tuned by using data engineering technologies. In next chapter under conclusion author has represent key technologies where the scores can be improved.

8.3 Chapter Summary

This chapter discuss about the evaluation process and the importance of evaluation once the project is completed. evaluation can be done in two ways where the project can be evaluated by third party evaluator, or the project can be evaluated by the author itself. When evaluating the project the main consideration needs to be whether the given requirements are being fulfilled during the project implementation. In next chapter it includes the conclusion of the project, remarks, future implementation of this study

Chapter 9: Conclusion

9.1 Chapter Overview

In previous chapter it represented the evaluation of student result prediction system. This chapter focuses on the final concluding remarks of the project. In this chapter it will discuss about the aims and the objectives of the project and also the completion of those mentioned aims and the objectives. At the end of this chapter concluding remarks and future implementation of this study is explained in detail.

9.2 Completion of Aims and Objectives

When considering about the main aim and the objective of the project was to develop a system which can predict the student results under defined parameters. The reason why the author has selected this topic is there are so many students who are dropping out from the universities without having their expectations and proper qualification. The reason which has identified for this issue is lack of proper guidance, lack of understanding regarding the subject matters and also some social factors. author wanted to address this particular issue by using a scientific approach hence author has decided to use a machine learning predicting system where users can predict their exam results once they have inserted certain parameters. In this study author wanted to compare the suitable and optimum algorithm which can be used in such scenario.

At the end of the project above mentioned aims and objectives was able to cover by the author and also created a system which can be used to predict the student results once the input data has been provided to the system.

9.3 Utilizing of Knowledge from Master's Program Modules

In order to complete this particular study knowledge has been absorbed by certain modules which has been conducted in masters of business analytics program which conducted by university of Colombo school of computing. Modules such as individual project, data programming, academic writing were helpful when conducting this particular study. When analyzing data and training the algorithms the knowledge which shared in practical sessions also being helpful.

9.4 Usage of Existing Knowledge.

When completing the project the knowledge regarding software development database development and UML and design diagrams were helpful which can be considered as existing knowledge. When creating databases and designing this software those knowledge has been really helpful since it needs to be used in this implementation.

9.5 Problems and Challenges

When considering about problems and challenges which author has faced during implementation was the scope of the project. When submitting the project proposal author had a broader scope such as predicting the student carrier based on their university results. The identified scope of the project is broader since there are some constraints which cannot be addressed during the project implementation. And also when identifying the required parameters again it's a problem because the identified scope of the project is huge. In order to predict a carrier of a student there are number of parameters which need to be consider. As the guidance of the supervisor the project scope was narrowed down to predict student results at the end of the academic year. When identifying the parameters again author has faced certain problems. Again sudden results might depends on number of parameters such as social and subjective point of views.

the second problem which faced during this study is lack of literature materials available under this subject domain. There are a few researches are being carried out under this subject matter and those researchers are high level researchers where they carried out those research for their doctor philosophes and post doctoral studies.

The main problem and the challenge which fails during this study is time constraint. Since this project needs to be carried out in entire one academic year the time management was difficult since author has carried out this project along with other responsibilities and also new technologies need to be learned throughout the project.

In this project main prediction was done by using confidential data such as student results because of that reason author has faced certain legal obligations.

9.6 Future Implementations

As the future implementation of this particular project author is focusing on to improve the score from machine learning algorithms where users can obtain more accurate result. And also more parameters can be considered when doing the predictions so that more accurate result can be taken as an output in practical situations. In this study author has only use classification algorithms to do the predictions but in future enhancement user is planning to use regression algorithms where system is capable of predicting value as a numerical value. As another future enhancements author is planning to develop a mobile application in order to use for predictions.

9.7 Concluding Remarks

Student result predictor is the first application which developed under identified parameters and factors which is capable of predicting student results. This presented solution can be use by number of users who has been engaged with academic and education field. This project is carried out to identify the optimum solution and algorithm which can be used to do the prediction when it comes to academic background. This provided solution can be enhanced and improve by modifying the used technologies and methodologies which

applies inside the project. As mentioned in future enhancement author is planning to improve the current system to get a better prediction result in future.

References

Abu Zohair, L.M. (2019) "Prediction of Student's performance by modelling small dataset size," *International Journal of Educational Technology in Higher Education*, 16(1). doi:10.1186/s41239-019-0160-3.

Almasri, A., Celebi, E. and Alkhaldeh, R.S. (2019) "EMT: Ensemble meta-based tree model for predicting student performance," *Scientific Programming*, 2019. doi:10.1155/2019/3610248.

Imran, M. *et al.* (2019) "Student academic performance prediction using supervised learning techniques," *International Journal of Emerging Technologies in Learning*, 14(14), pp. 92–104. doi:10.3991/ijet.v14i14.10310.

Lou, Y., Ren, R. and Zhao, Y. (no date) *A Machine Learning Approach for Future Career Planning*.

Roy, K.S. *et al.* (2018) *Student Career Prediction Using Advanced Machine Learning Techniques*, *International Journal of Engineering & Technology*.

Sharma, S., Pandey, S.K. and Garg, K. (2020) "Machine Learning for Predictions in Academics," *International Journal of Recent Technology and Engineering*, 8(5), pp. 4624–4627. doi:10.35940/ijrte.e6965.018520.

VidyapriyaC and VishhnuvardhanRC (no date) *STUDENT CAREER PREDICTION*.

Nazareno, A. L. *et al.* (2019) 'An artificial neural network approach in predicting career strand of incoming senior high school students', *Journal of Physics: Conference Series*, 1245(1). doi: 10.1088/1742-6596/1245/1/012005

Nie, M., Yang, L., Sun, J. *et al.* Advanced forecasting of career choices for college students based on campus big data. *Front. Comput. Sci.* 12, 494–503 (2018). <https://doi.org/10.1007/s11704-017-6498-6>

Selingo, J. (2016). *There Is Life After College: What Parents and Students Should Know About Navigating School to Prepare for the Jobs of Tomorrow*. New York, NY: HarperCollins.

²Cai, Y. (2013). Graduate employability: a conceptual framework for understanding employers' perceptions. *Higher Education*, 65(4), 457-469.

National Association of Colleges and Employers. (2017). The Key Variables Employers Seek on Students' Resumes. Retrieved from <https://www.naceweb.org/about-us/press/2017/the-key-variables-employers-seek-on-students-resumes/>.

⁴Nemanick, R. C., & Clark, E. M. (2002). The differential effects of extracurricular activities on attributions in resume evaluation. *International Journal of Selection and Assessment*, 10, 206–217.

⁵Rothman, M. & Sisman, R. (2016). Internship Impact on Career Considerations Among Business Students. *Education & Training*, 58(9), 1003-1013.

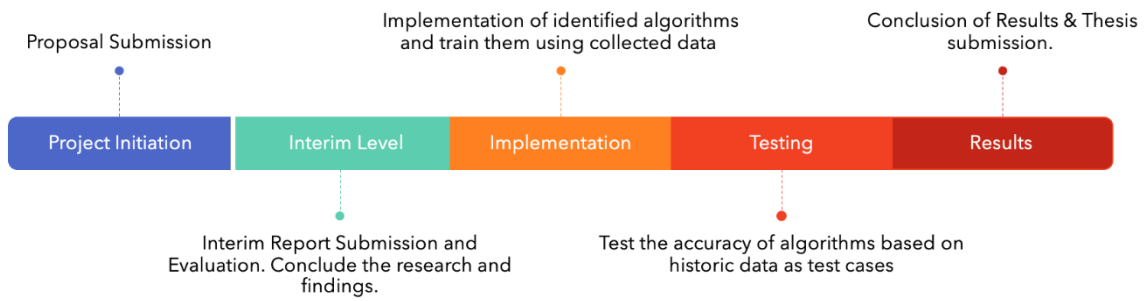
Meizar Raka Rimadana, Sri Suning Kusumawardani, Paulus Insap Santosa, Maximillian Sheldy Ferdinand Erwianda, "Predicting Student Academic Performance using Machine Learning and Time Management Skill Data", *Research of Information Technology and Intelligent Systems (ISRITI) 2019 International Seminar on*, pp. 511-515, 2019

Ammar Almasri, Erbug Celebi, Rami S. Alkhawaldeh, "EMT: Ensemble Meta-Based Tree Model for Predicting Student Performance", *Scientific Programming*, vol. 2019, Article ID 3610248, 13 pages, 2019. <https://doi.org/10.1155/2019/3610248>

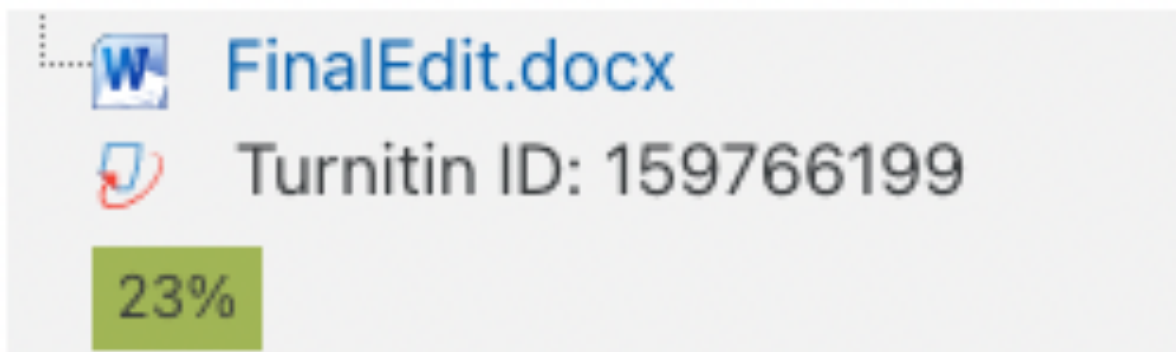
Roy, K., K. Roopkanth, V. U. Teja, V. Bhavana and J. Priyanka. "Student Career Prediction Using Advanced Machine Learning Techniques." *International journal of engineering and technology* 7 (2018): 26.

Appendix I Gantt Chart

Gantt Chart



Appendix II Turnitin Report



Summary of Changes

#	Change Requested by the supervisor	How you addressed the supervisor request
01	Explain the project aim	Explained the project aim in detailed.