****

**Prediction on**

**Student Academic Performance**

**Using**

**Hybrid Clustering Algorithm**

**Research Methodologies and Emerging Technologies**

**Final Report**

Candidate Name: Jiten Limbu

Candidate ID: 1812720

Unit Code: PAT004-3

Date: 17 May 2019

Supervisor Name**:** Dr. Swati Sah

HOD of Patan College for Professional Studies

Contents

[**Acknowledgement** 4](#_Toc8968274)

[**Abstract** 5](#_Toc8968275)

[**1.** **INTRODUCTION** 7](#_Toc8968276)

[**1.1 Background** 7](#_Toc8968277)

[**1.1.1. Student’s Academic Performance without Data Clustering Algorithm** 8](#_Toc8968278)

[**1.1.2. Student’s Academic Performance with Data Clustering Algorithm** 8](#_Toc8968279)

[**1.1.3. Human Learning vs. Artificial Intelligence** 8](#_Toc8968280)

[**1.2. Problem Context** 9](#_Toc8968281)

[**1.3. Rational** 9](#_Toc8968282)

[**1.4. Target User** 10](#_Toc8968283)

[**1.5. Aims and Objectives** 10](#_Toc8968284)

[**1.6. Deliverables** 11](#_Toc8968285)

[**1.7. Nature of Challenges** 11](#_Toc8968286)

[**2.** **THE LITERATURE REVIEW** 12](#_Toc8968287)

[**2.1. Technical Background Introduction of Student Academic Performance** 12](#_Toc8968288)

[**2.1.1. Algorithms Used for Predicting Student Academic Performance** 12](#_Toc8968289)

[**2.1.2. Few Research on Student Academic Performance Prediction Model** 14](#_Toc8968290)

[**2.1.3. Critical Analysis of Algorithms Used for Predicting Student Performance** 15](#_Toc8968291)

[**2.1.4. Student Academic Performance (Determinant Factors/Attributes)** 16](#_Toc8968292)

[**2.1.5. Critical Analysis of Determinant Factors/Attributes** 17](#_Toc8968293)

[**2.2.** **Relevant Work Done in Student’s Academic Performance with Critical Review** 18](#_Toc8968294)

[**2.2.1.** **Analysis of Similar Systems** 18](#_Toc8968295)

[**2.2.2.** **Domain Research** 25](#_Toc8968296)

[**2.3.** **Discussion on your finding regarding similar system and existing technology** 27](#_Toc8968297)

[**3.** **Data Analysis and Requirement Validation** 28](#_Toc8968298)

[**3.1. Analysis of Collected Data** 29](#_Toc8968299)

[**3.2. Relate how the findings of the various research methods applied** 36](#_Toc8968300)

[**4.** **System Architecture** 37](#_Toc8968301)

[**4.1. SDLC** 40](#_Toc8968302)

[**4.2.** **Requirements** 40](#_Toc8968303)

[**4.2.1. Functional** 40](#_Toc8968304)

[**4.2.2. Nonfunctional** 43](#_Toc8968305)

[**4.2.3. MoSCoW Prioritization** 44](#_Toc8968306)

[**5. Design Diagrams** 45](#_Toc8968307)

[**6.** **Development & Testing** 52](#_Toc8968308)

[**6.1.** **Development** 52](#_Toc8968309)

[**6.1.1. Programming Language** 52](#_Toc8968310)

[**6.1.2. IDE and Platform** 52](#_Toc8968311)

[**6.1.3. Libraries** 52](#_Toc8968312)

[**6.1.4. Code** 53](#_Toc8968313)

[**6.1.5. User Manual** 68](#_Toc8968314)

[**6.2.** **Testing** 69](#_Toc8968315)

[**8.** **Evaluation** 75](#_Toc8968316)

[**8.1.** **Evaluators** 75](#_Toc8968317)

[**8.2.** **Result of Evaluation** 75](#_Toc8968318)

[**9.** **Critical Analysis** 78](#_Toc8968319)

[**10.** **Conclusion & Future Work** 78](#_Toc8968320)

[**REFERENCES** 79](#_Toc8968321)

[**Appendices** 82](#_Toc8968322)

# **Acknowledgement**

I would like to express my special thanks to my project supervisor Dr. Swati Sah as well as our module teacher Mr. Sagun Dhungana who gave me golden opportunity to do this research on topic **Prediction on Student Academic Performance Using Clustering Algorithm**. Each and every steps of this research project are guided by my research supervisor and module teacher. I am really grateful to them for their useful materials and assessments which has improved the quality of this research. It helped me in my Research and came to know about so many new things about my research topic. I am really thankful to them.

Also, I would like to thank my academic writing and programming teachers, my classmates as well as whole unit of Patan College for Professional Studies (PCPS) Family who helped me a lot to finish this work. Thank you again.

# **Abstract**

More and more researchers are doing research on prediction model to predict student academic performance while dealing with large scale of educational databases. The current research and there product has not been able to give maximum accuracy due to various reasons like lack of appropriate data and lack of best algorithm. If existing systems can predict student's academic progress and performance with 100% accuracy result then further research on this field is not required. But to develop the prediction model on student's academic performance with maximum accuracy rate is more challenging. These type of problems encouraged the researcher to research on this field with data mining techniques. The application of data mining technique is done in order to divide information into separate clusters or classifications so that particular dataset can be studied under a classified group. The output of this prediction model helps to identify student's educational status as well as future performances by analyzing their data. It also helps to identify particular institution’s weak students and this prediction model can provide him/her to prevent that issue by providing relevant suggestions. It will directly boost the performance of those identified weak students. So, educational institutions can easily take benefits by using this prediction model. The educational institutions can bring special package to improve those weak identified students by analyzing the student academic performance predictive model.

The main aim of this research is to develop student academic performance prediction model for the Nepalese bachelor level student in Computer Science stream using unsupervised learning with some data clustering method. Here, the Clustering Algorithm is the cluster-based unsupervised learning algorithm that is more popular in prediction model which clusters the particular dataset by analyzing its similarity. Our prediction model will be done through python programming language. One of the clustering algorithm will be implemented through this python programming language environment.

To conduct this research, each student's internal subject marks are selected as parameter. The main objective of this research is to provide more description of the Clustering Algorithm and provide how the clustering algorithm will cluster or separate student's academic internal marks and then develop the platform based on the research. A desktop based application will be designed to get prediction model which predicts the student's academic performance by implementing one of the clustering algorithm. The application will show the future performances of each particular student. It is also helpful to find out different categories of the student (i.e. weak, normal, intelligent) by analyzing their internal marks.

**Keywords**: - Clustering, Prediction Model, Data Mining Technique, AI, K-means Clustering Algorithm, Educational Data Mining, Machine Learning.

# **INTRODUCTION**

From the ancestral time of the human beings, we can say that the human Intelligence has played a key role in our survival. During stone ages, humans had used their knowledge about creating fire and crafting stones as tools and weapons and taught those information to their successor so that it would help them survive. And as human civilization began to grew, the need for passing on the knowledge to our future generations became very important so that the human civilization continues to thrive. And as of this 21st century, we still need to provide a better and quality education to our young generation. So we need to review their academic performances and see how we can effectively provide quality education to them. Education is the prime and more talkative term now days. It plays vital role to develop each and every nation. Now, all the big inventions and developments are happening day by day because of educated peoples.

Measuring student’s academic performance using clustering algorithm is the advance improvement of human being where Artificial Intelligence (AI) plays vital role to predict student’s academic performance. Now days, data mining technique is used to develop prediction model. Data mining technique can work through different classification algorithms and clustering algorithm. The overview or outline of this Report is as follows:

## **1.1 Background**

There are two types of research that has been done to predict student academic performance by analyzing their internal marks and different types of parameter like family background, socio- economic and educational factors. First one is human learning prediction and another one is using Artificial Intelligence (AI) in their research. The Human Learning Technique is first developed with its own predicting ideas that comes from studies of Mathematics, Economics, Psychology, Philosophy, Social and Political Science and others (Kraus, 2018). In Human Learning, researcher provides judgmental analysis through more common sense and decision making theory like Utility Theory which can do interdisciplinary effort for prediction from collected data after doing deep analysis. But in AI, all the decision have been done through machine.

In this research will be follow section 1.1.2 which is designing and developing a prediction model to predict student academic performance by using one of the best unsupervised clustering algorithm.

### **1.1.1. Student’s Academic Performance without Data Clustering Algorithm**

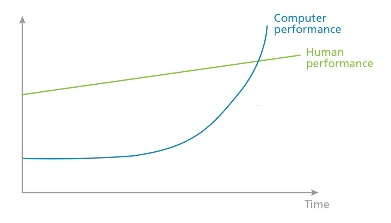
There are so many research that has been done for the aim of predicting student academic performance without using Data Mining Technique. The researcher uses their full effort to get that solution which helps to prove validity and meet required objective (Harford, 2004). The researcher don’t have any knowledge about emerging technologies. But they know clear knowledge about how to do research in traditional way.

### **1.1.2. Student’s Academic Performance with Data Clustering Algorithm**

Predicting Student Academic Performance through AI is more advance and more accurate and trustable model. With the help of data mining technique, researcher can be able to predict student’s academic performance faster by clustering through different types of clustering algorithms. In section 2.1.1, there have been described, some classification algorithms that is used by data mining. And especially in section 2.1.1.2, there have been some description of some clustering algorithms which can predict student’s academic performance that processes dataset.

### **1.1.3. Human Learning vs. Artificial Intelligence**

As a conclusion of Section 1.1.1 and Section 1.1.2, the performance of **HL** is increasing but not more than **AI**. The presence of AI can be more effective to predict student’s academic performance. Because, the accuracy rate of AI is maximum than HL. Here following diagram has been used to prove this statement.

 (4imprint, 2018)

*Fig. 1: The growth of machine learning*

Here, Computer’s performance for prediction model is far better than human performance. Therefore, Artificial Intelligence may outplace the Human Learning in prediction case (4imprint, 2018).

## **1.2. Problem Context**

After the study, I found so many research papers which is related to prediction on student academic performance. Those papers are developed through the supervised learning in linear regression algorithm, decision tree etc. but those algorithms take too much time to process big scale of data and also there is a memory problem while dealing with large scale of dataset (Hetal Bhavsar, 2012). That's way, I will test my parameter (i.e. internal marks) in one of the unsupervised learning clustering algorithm. And for further research, I will design a set of questionnaires for the specific academic institutions.

## **1.3. Rational**

This research is important to address prediction related problem. As a solution of this problem, predictive application will be designed after deep research after finding which algorithm is best in prediction model.

**Methodology**

**The Quantitative Research Methodology** will be implemented in this research. The Quantitative Research Methodology is the research method which deals with number and anything that can be measurable by their research process (Perumal, 2014).

## **1.4. Target User**

This research is mainly focused on academic students of Nepal. Primary targeted students are those who are studying bachelor. Because, this system will be trained and developed through dataset of bachelor level students. Students will get direct benefits from this system. Especially student can be able to know their educational pattern and their future education status so they can get lots of information about their future. So they can easily take right decisions for their career. In a simple way, students will get their predictive result that might be very useful for their golden life.

Teachers and Academic Institutions are this research’s next secondary targeted users. After using this prediction model, teachers and academic institutions will get judgmental decision to know either that student is intelligent, normal or weak in that education sector.

## **1.5. Aims and Objectives**

The aim of this research is to develop student academic performance prediction model for the Nepalese Bachelor Level Student in Computer Science Background. To do that task, the model used will be unsupervised learning algorithm. Clustering task will be done under unsupervised learning. By selecting best clustering data mining technique, our prediction model will reach its prime aim. The python programming language will be selected to build the system. Where, the parameter (i.e. student’s internal subject marks) are clustered through one of the best unsupervised clustering data mining algorithm which have maximum probability to process maximum dataset in a minimum timeframe with highest accuracy result.

The main aim is:

* **To discover knowledge and patterns about student’s academic performance through prediction model.**

The objective of this research is to get more knowledge about data mining technique and its clustering algorithms. This research will also help to discover and cluster student’s all the internal marks by data analysis process. And a desktop based application will be designed to get student’s academic performance. That output is the prediction model which will be developed through one of the best clustering algorithm for the purpose to find out intelligent, normal and weak students in Computer Science stream in Nepalese Education Sector. To get that related aim of this predictive model, following objectives are playing vital role:

* **To cluster the student’s academic dataset through one of the clustering algorithm.**
* **To develop the prediction model that takes the student’s dataset, process that data and show predictive model.**
* **To develop a platform that predict student academic performance.**

## **1.6. Deliverables**

Student's academic internal marks are analysed and clustered using one of the clustering algorithm for the purpose to identify and suggest further enhancement of their academic performance through this prediction model.

A desktop-based application will be designed in Python Programming language. In Python Programming language, a program will be developed for the prediction model that takes the student's dataset, process that data and show predictive result. By using python library, predictive result will be generated through best clustering algorithm.

Different Graphical Infographic diagrams will be generated through application system.

## **1.7. Nature of Challenges**

This was my first research paper. And, since I had no primal experience in research, it was very hard for me to understand the process of preparing a research paper. It was challenging for me to write the literature review and follow the process on domain research and similar system analysis. Many relevant e- books, journals and articles were used as the source of information for this topic which made the task very tedious since the information on them were simply too many to be handled.

Since research on this topic was done so many times, most of the contributions were already made by the former researchers, so it’s hard to provide any further contributions to this topic.

# **THE LITERATURE REVIEW**

Literature review provides an insight to a particular field of study through the work of past researchers on the same field. A literature review is one of the key steps in a research process. (Abdullah Ramdhani, 2014).

## **2.1. Technical Background Introduction of Student Academic Performance**

Deep Research will be performed to predict Student’s Academic Performance of each and every students who are studying bachelor level in Nepal under the stream “Computer Science”. To complete this research, multiple technical terms will be implemented in each and every sections of the research work. Mainly the title of this research can easily be able to show its technical introduction. In this work the research will take data mining clustering algorithms and dataset that is also known as factors to predict student’s academic performance. For more detail section 2.1.1 and section 2.1.2 will show those technical terms more clearly.

### **2.1.1. Algorithms Used for Predicting Student Academic Performance**

The research will be use Hybrid Algorithm (i.e. combination of K-means Algorithm and Support Vector Machine algorithm).

All the prediction task is done through supervised and unsupervised learning predictive model. Section 2.1.1.1 and section 2.1.1.2 will describe this terms more clearly.

#### **2.1.1.1. Supervised Learning**

There are several algorithms, which are under supervised learning and unsupervised learning in predictive model. The supervised learning does data classification and regression on predictive model. Supervised learning requires target value to be achieved which is supplied by the ‘supervisor’ or ‘teacher’ with some knowledge on labelled data (Flach, 2012). Following algorithms are used for supervised learning:

1. Decision Tree Algorithm (DT)
2. Support Vector Machine (SVM)
3. Naïve Bayes Classification Algorithm (NBC)
4. Neural Networks (NN)
5. K-Nearest Neighbor (KNN)
6. Linear Regression etc. (LR)

##### **2.1.1.1.1. Decision Tree Algorithm**

Most of the researchers symbolize Decision Tree Algorithm as a very popular prediction technique due to its simplicity and its suitability for small to large volume of data (Shahiri, 2015). The key benefits of this algorithm is that its easily understand by the users and do not require maximum dataset (Romero, 2008).

##### **2.1.1.1.2. Support Vector Machine**

Support Vector Machine is a supervised learning method used as most popular classification algorithm. Support Vector Machine is the alternative of another good technique for classifying dataset. If we have small dataset then it is suitable for prediction model with high chance of accuracy result. One of the research done by Geraldine Gray in 2014 with various classification algorithms, SVM had a maximum accuracy point along with other algorithms such as: DT, NN, KNN, NB and LR with same parameters (Geraldine Gray, 2014).

##### **2.1.1.1.3. Naïve Bayes Classification Algorithm**

Naïve Bayes Algorithm is also an option for researchers to predict classification technique. Most of the researchers choose naïve Bayes algorithm to predict student’s performance by making comparisons. Naïve Bayes Algorithm uses all the attributes and then, by analyzing its importance and dependency scale of each attributes and make prediction through this algorithm (Suljić, 2012).

##### **2.1.1.1.4 Neural Networks**

Neural Network is a popular technique to predict through classification model. Neural Network is a set of all interconnected input/output units and each connection has a weight to present learning ideas. In learning phase, neural network learns to predict specific output by adjusting their weights (Pal, 2011). Neural Network has ability to select all possible interactions between predictor variables (Geraldine Gray, 2014). And neural network can easily predict nonlinear complex relationship between dependent and independent predictor variables (Shahiri, 2015).

##### **2.1.1.1.5. K-Nearest Neighbors**

K-Nearest Neighbors is a classification algorithm which is used in prediction model to predict some information from dataset. KNN is very easy and effective algorithm. It works with numeric and nominal values in simple form as well as easily understandable by user. But according to the author Peter Harrington, the main problem of this algorithm is KNN doesn’t gives us known idea of the underlying structure of the data (Harrington, 2012).

##### **2.1.1.1.6. Linear Regression**

Linear Regression is not only developed to predict only nominal values for the targeted variables, it also predicts that targeted values which is continuous. Linear Regression Algorithm is very easy to implement and computationally inexpensive. But nonlinear are not addressed by this linear regression algorithm (Harrington, 2012).

#### **2.1.1.2. Unsupervised Learning**

The Unsupervised Learning is the form of predictive clustering in predictive model. Inside unsupervised learning, there are several clustering algorithms such as: K-means Clustering Algorithm, Hierarchical Clustering Algorithm.

##### **2.1.1.2.1. K-means Clustering Algorithm**

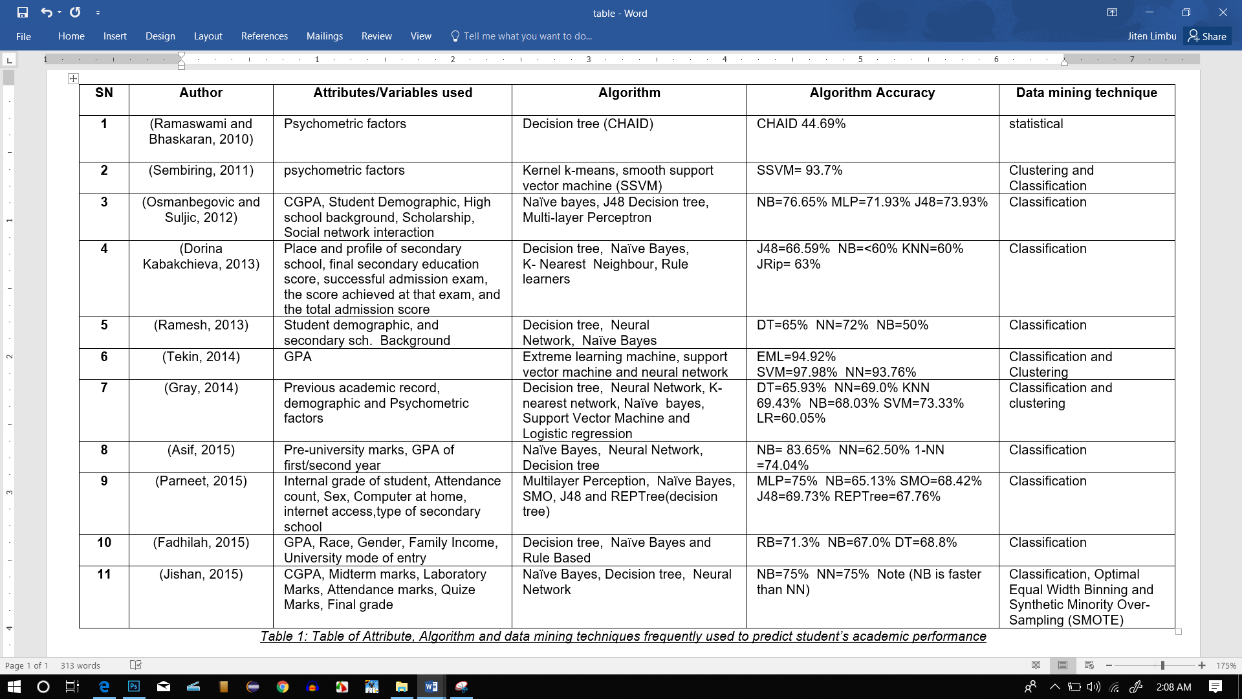
In K-means Clustering Algorithm, data clustering technique is used (Snehal Bhogan, 2017). The data points of the specific dataset is partitioned into homogenous clusters. The important task of the k-means is to solve clustering problems of unsupervised learning algorithm. The k-means clustering algorithm uses dataset and gather them into number of clusters to generate predictive results (Anasse Bari, 2014).

##### **2.1.1.2.2. Hierarchical Clustering Algorithm**

Hierarchical Clustering Algorithm is the unsupervised learning clustering algorithm which defines the similarity between a pair of clusters. The frequently used algorithms in hierarchical clustering algorithms are the single-link algorithm, complete-link algorithm and minimum-variance algorithm. According to the author, the single-link and complete-link algorithm define the similarity between a pair of clusters. In other-case, two clusters are merged into a single large cluster based on minimum distance criteria (David Hand, 2001).

### **2.1.2. Few Research on Student Academic Performance Prediction Model**

This is the few research list which has been done by famous researchers through different algorithms. The following table shows researcher’s name, the algorithms they used, datamining technique, attributes that were used while doing research and their output result.



(Abdulazeez Yusuf, 2016)

*Table 1: Attribute, Algorithm and data mining techniques frequently used to predict student’s academic performance*

### **2.1.3. Critical Analysis of Algorithms Used for Predicting Student Performance**

Section 2.1.1.1 and Section 2.1.1.2, defines the algorithms that has been used for predicting student’s academic performance. Section 2.1.1.1 represents the supervised learning which have those types of algorithms which are only able to do classification. So that this learning algorithm also called Classification Learning (Ian H. Witten, 2005). In this classification algorithm, all the algorithms are better for labeled data which is small in the case of their data volume.

In Section 2.1.1.2, clustering techniques is used to predict student academic performance. All the algorithms are used to predict student academic performance by clustering their dataset. It is called Unsupervised Learning. K-means clustering algorithm is classic and simple effective technique. Its job is to cluster unlabeled data several times until desired outcome is not displayed. By providing several iterations, k means can only meet its prediction destination. But unsupervised learning algorithms are better than supervised learning algorithms. Here, clustering algorithms can easily handle a big scale of dataset without taking more time, comparing with supervised learning algorithms. Therefore, I would like to choose unsupervised clustering algorithm over supervised classification techniques.

### **2.1.4. Student Academic Performance (Determinant Factors/Attributes)**

Determinant Factors/ Attributes refers to the predictor variables that directly affects the student’s academic performance which helps to understand and know the educational background of the particular student (Sen, 2012). Therefore, researcher’s main curiosity is to define or select relevant variables from student’s academic achievement by using data mining techniques.

In literature, there are previous research works aimed at determining the prime factors or attributes that can help to predict performance of the student and the methods that gives the maximum accurate result. Researcher Auirah Mohamed Shahiri did a systematical literature review on ‘Predicting Student’s Performance Using Data Mining Techniques’. On this work, Shahiri provided an clear overview of data mining techniques; how that techniques are used to predict student academic performance through best prediction algorithms (i.e. the supervised learning; Decision Tree Algorithm) with the help of student’s academic achievement (i.e. the predictor variables) (Shahiri, 2015). According to their research the attribute that have been used is Grade Point Average (GPA). Because the GPA attribute is more tangible value for future educational system (Abdulazeez Yusuf, 2016).The other researchers Ahmet Tekin on ‘Early Prediction of Student’s Grade Point Averages at Graduation: A Data Mining Approach’ (Tekin, 2014), Justice Stephen Tetteh Zotorvie on ‘Student’s Accommodation and Academic Performance: the Case of Ho Technical University Ghana’ (Zotorvie, 2017) and Syed Tanveer Jishan on ‘Improving Accuracy of Student’s Final Grade Prediction Model Using Optimal Equal Width Binning and Synthetic Minority Over-Sampling Technique’ (Jishan, 2015) also used GPA as a determinant factor on their research.

In prediction process, other attributes are also used to predict student’s academic performance. Some key attributes like Gender, Age, Family Background, Disability, etc. play vital role in prediction of student’s academic performance.

### **2.1.5. Critical Analysis of Determinant Factors/Attributes**

Most of the researcher are choosing Grade Point Average (GPA) as a prime factor to determine student’s academic performance. But there are also lost of parameters are still their which can be helpful for prediction model.

In my view, other factors such as Gender, Age, and Family Background are also main factors that will be used to predict student’s academic performance.

## **Relevant Work Done in Student’s Academic Performance with Critical Review**

This section 2.2 is core part of literature review, where analysis of similar systems and domain research has been done in section 2.2.1 and section 2.2.2 respectively. In similar system, application developed in past era are studied. And those applications are predictable application product which has been done through massive research. The analysis and reviewing of existing system, their output, seeking advantage and disadvantage of that product is the main goal of this similar system research. And in section 2.2.2, domain research has been done. That is the study of similar or tangible research papers which is related with student’s academic performance prediction research area.

### **Analysis of Similar Systems**

This is the part of doing analysis of similar systems that is previously done by many researchers. Therefore, similar system analysis is the part of review and study of those similar applications which is already designed and developed as an output. By analysis of similar system each and every researcher can get more clear knowledge about their system which had been developed form their prospective, and it helps to those researchers who are going to do research on that similar research topic.

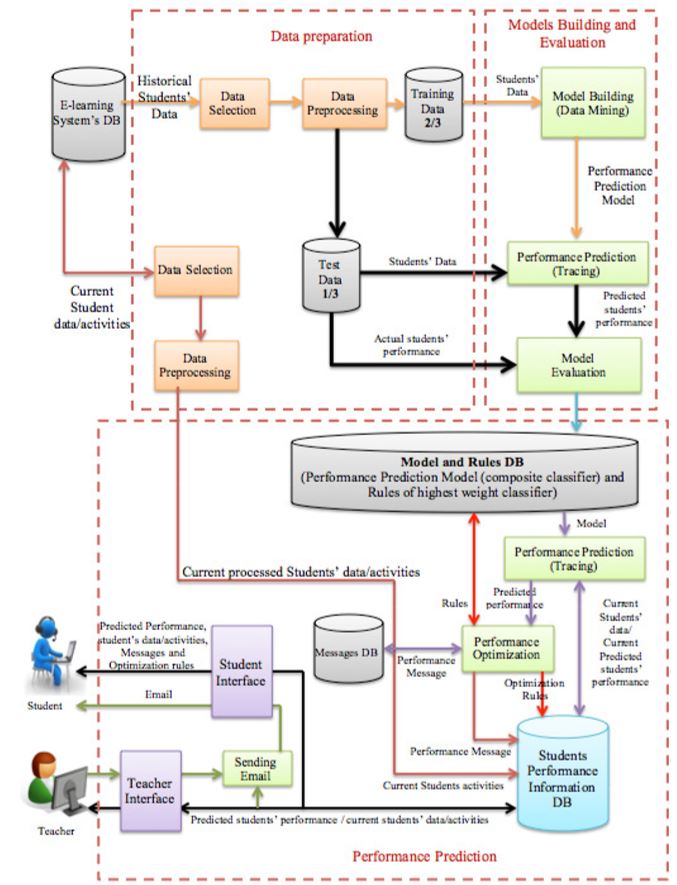
#### **2.2.1.1. Similar System 1**

Here, introducing a system which is named “STUDENT’ PERFORMANCE PREDICTION SYSTEM” from research of “Student’ Performance Prediction System Using Multi Agent Data Mining Technique” by Dr. Abdullah AL-Malaise, Dr. Areej Malibari and Mona Alkhozae (Dr. Abdullah AL-Malaise, 2014).

##### **2.2.1.1.1. Conceptual Model & Architecture**

The conceptual model and architecture is the process of designing basic major processes of system development part. All the system development components has been constructed and given required role and responsibilities to show system executions.

There is a Student’ Academic Performance Prediction Agent to predict their current progress. The whole architecture is following:



(Dr. Abdullah AL-Malaise, 2014)

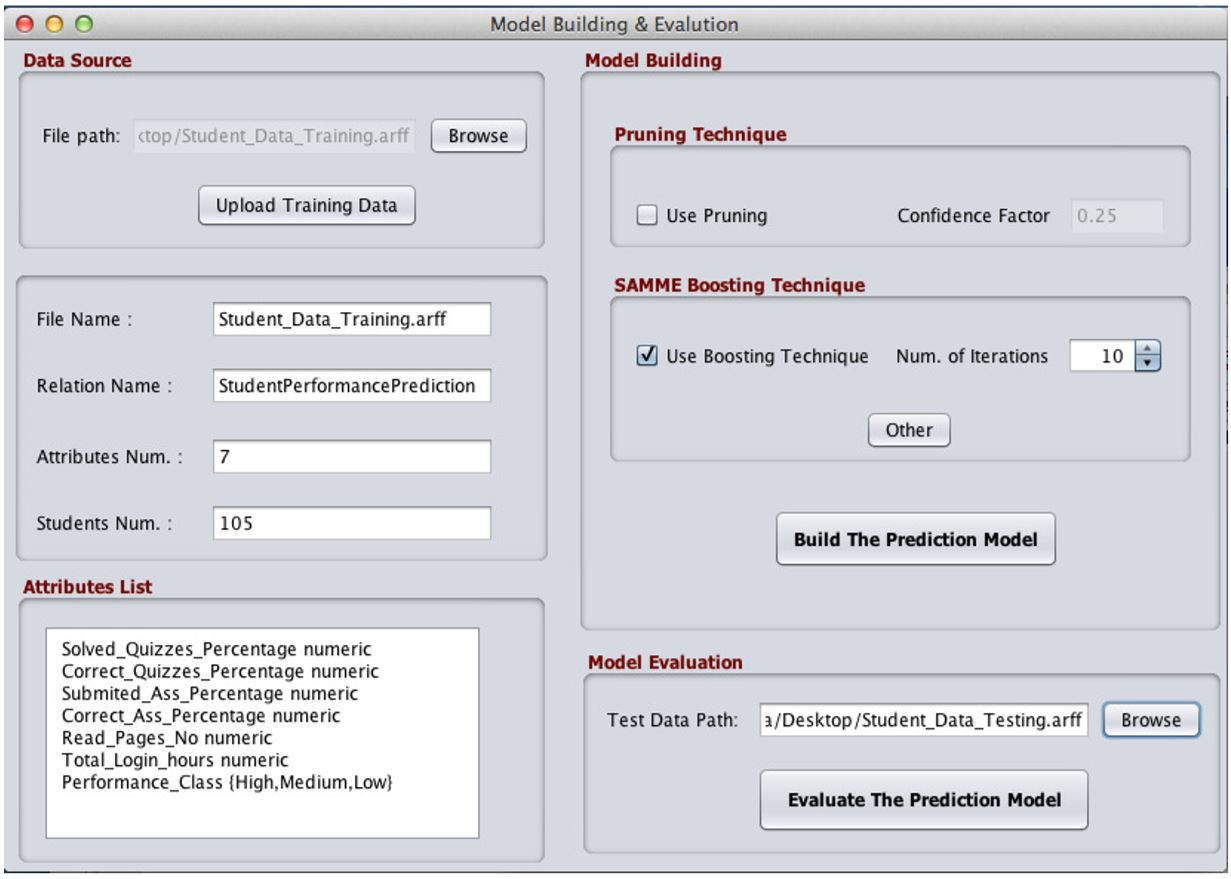
*Fig. 2: Conceptual System Architecture*

##### **2.2.1.1.2. Programming Language, Tools and Libraries**

Here in this system, Java Programming Language has been used to develop the system through Netbeans IDE. Java Agent Development Environment (JADE) is a library which is used as a middleware which can access agent paradigm in application system. And also WEKA (Waikato Environment for Knowledge Analysis) used as a library for this system. WEKA is an open source toolkit which have lots of machine learning algorithms for datamining task which deals large scale of dataset very easily (Ian H. Witten, 2005). There are lots of algorithms that are used to evaluate the system. SAMME boosting algorithm, Adaboost.M1 and LogitBoost algorithm are the prime algorithms for this system. Last, C4.5 Decision Tree Algorithm is used for base classifier for those algorithms.

##### **2.2.1.1.3. System Implementation**

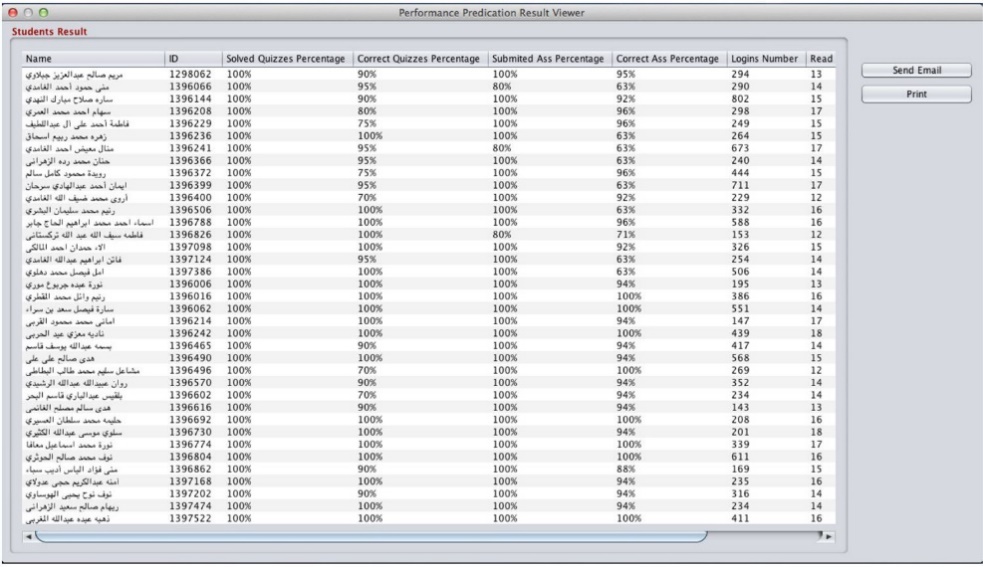
According to this system, data preparation are implemented manually for data collection. After data collection, required useful data has been selected for prediction. Only useful data are filtered through collected data. Then again re-processing of those data has been done by cleaning, transforming into meaningful data and reducing multiple similar type of data. Then after all the data has been converted into excel sheet to Attribute Relation File Format (ARFF). In further step of implementation, those data are divided into testing data and training data. Then model has been developed, evaluated them. Finally, student’s academic performance interface has been built.



(Dr. Abdullah AL-Malaise, 2014)

*Fig. 3: Student’ Performance Evaluation Interface*

After evaluation of all the data set following result has been shown:



(Dr. Abdullah AL-Malaise, 2014)

*Fig. 4: Student’ Performance Prediction Result*

After getting those prediction result. Those results can be printed and sent as an email.

##### **2.2.1.1.4. Experimental Output**

Experiment was conducted from multiple multi agent data mining algorithms (i.e. C4.5 Decision Tree Algorithm as a base classifier for SAMME Boosting Algorithm, Adaboost.M1 and LogitBoost Algorithms) with the help of EMES e-learning system dataset.

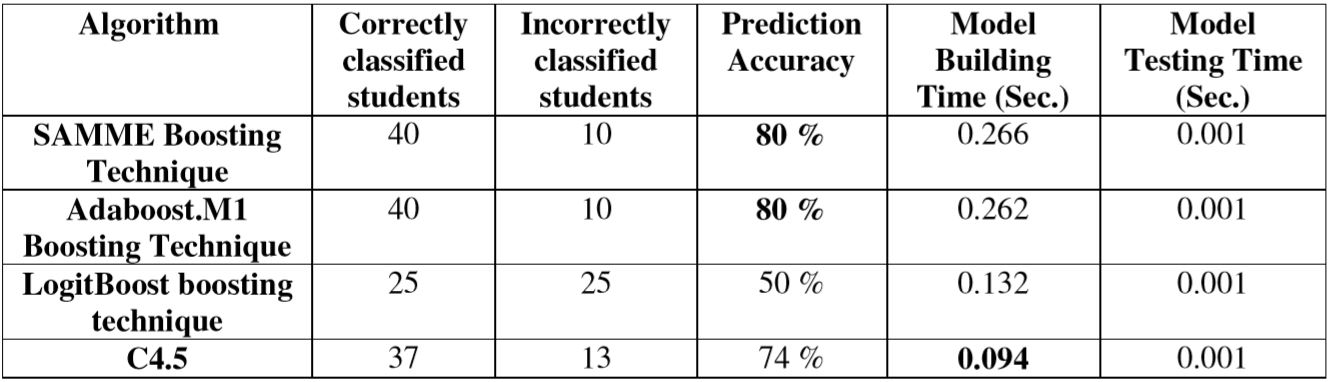


(Dr. Abdullah AL-Malaise, 2014)

*Fig. 5: Student’ Performance Prediction Model with Accuracy Result*

##### **2.2.1.1.5. Experimental Performance Comparison Result**

The comparison between all algorithms has been done to identify percentage of accuracy of each and every algorithms.



(Dr. Abdullah AL-Malaise, 2014)

*Table. 2: Algorithms with their Percentage of accuracy*

In this table, SAMME Boosting Algorithm and Adaboosting.M1 Algorithm have maximum accuracy percentage.

##### **2.2.1.1.6. Critical Analysis of Similar System 1**

This system has very perfect conceptual model architecture. And also have maximum number of dataset, which help to get maximum rate of accuracy result in this prediction model. System’s designing and developing process is also attractive. User-friendly application has been developed to solve prediction problem.

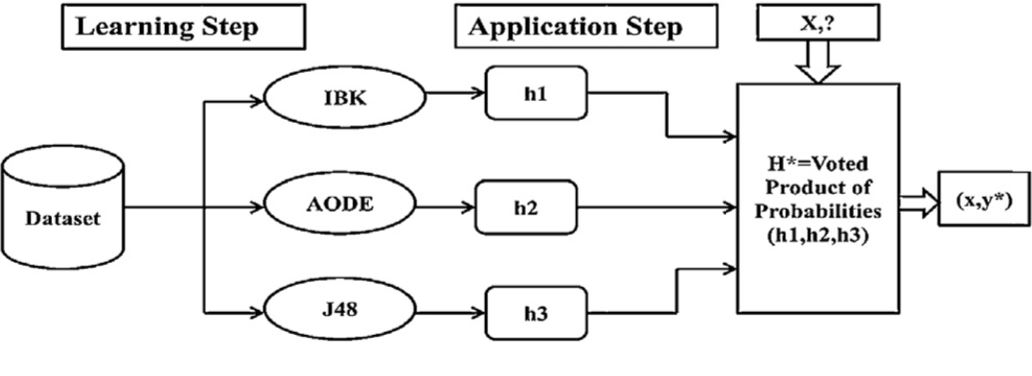
But somehow, there were still some factors which diminish their research output and system implementation. The researcher trying to implement new thing in data mining technique which is really impressive work by 9Dr. Abdullah AL-Malaise, Dr. Areej Malibari and Mona Alkhozae. But their testing data through system is too small. Therefore, the accuracy rate is only 80% at highest value. I think, the developer also needs to change their system implementation strategy from java programming language to other prediction supportive language such as Python or Matlab. Because WEKA is still running in old techniques and they still have GUI limitation which are difficulty to deal with large scale of dataset (Inf.ed.ac.uk, 2019).

**2.2.1.2. Similar System 2**

The System 2 is the output of research “Towards the Integration of Multiple Classifier Pertaining to The Student’s Performance Prediction”. To complete this research, probability combining rule is implemented with multiple algorithms. All essential algorithms are merged to introduce new algorithm (Mrinal Pandey, 2016).

##### **2.2.1.2..1 Proposed Model & Architecture**

The algorithm Decision Tree was named as a J48, K-Nearest Neighbour Algorithm was named into IBK and Aggregating One-Dependence Estimators Algorithm was abbreviated into AODE. And other thing combined of all these algorithm was named and a single composite model was proposed called KNNAD.

****

(Mrinal Pandey, 2016)

*Fig. 6: Proposed Model of KNNAD.*

##### **2.2.1.2..2. Tools and Libraries**

The Open Source Toolkit WEKA (Waikato Environment for Knowledge Analysis) was used to develop prediction model.

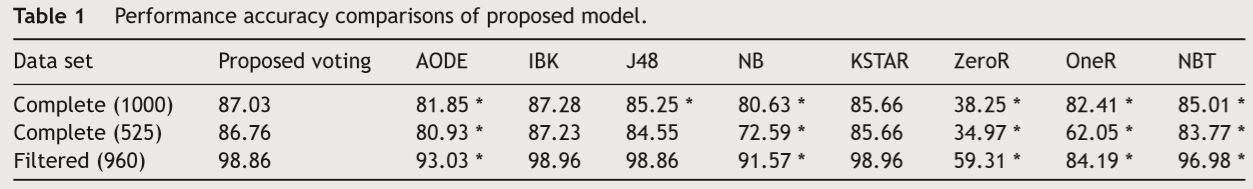
##### **2.2.1.2..3. System Implementation**

The data was taken from engineering college of India. There are two parameters in collected dataset. One is academic information and other one is demographic information. Three categorized dataset were used to predict student’s academic performance. After removing outlier from dataset, following dataset was remained to train:

1. Complete 1000
2. Complete 525 and
3. Filtered 960

##### **2.2.1.2..4. Experimental Output**

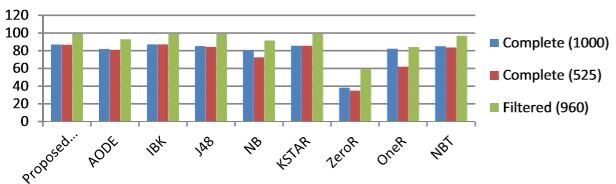
In Output, the proposed model (i.e. KNNAD) is better than other algorithms.



(Mrinal Pandey, 2016)

*Fig. 7: Accuracy Performance and Experimental Status*

In Figure 7, the performance accuracy of KSTAR and Proposed Voting is no far different. Their results are almost same. But other algorithm like J48 decision tree algorithm have less accuracy rate.



(Mrinal Pandey, 2016)

*Fig. 8: Accuracy Performance Comparison*

Here, the accuracy performance of Proposed Algorithm in Filtered Dataset is higher than others.

##### **2.2.1.2..5. Critical Analysis of Similar System 2**

In this system, the researcher proposed hybrid model KNNAD by combining three supervised learning algorithms (i.e. J48 decision tree algorithm, K-Nearest Neighbour Algorithm and Aggregating One-Dependence Estimators Algorithm) by analysing less amount of three datasets. First is 1000, second one is 525 and last one is 960 data. As an output, the maximum accuracy rate is 98.86%.

In my view, this system is perfectly developed, all the research procedures are very satisfying. But at the part of processing dataset very less amount of data has been trained by the researcher. If researcher had collected maximum amount of dataset and process it through data mining technique, then the output result would have been generated more than this current result.

### **Domain Research**

Clustering is the method process in unsupervised learning to group similar type of undefined data (Dr. Swati Sah, 2018). This is the way of cutting large scale of data into multiple small cluster by its similarity (K. M. Faraoun, 2007).

Maximum education institutions are very curious about whether their students will have positive result or not during their educational career. Before admission of new students and during their courses. Prediction of student’s academic performance through clustering algorithms are more appropriate and reliable (Kefaya, 2009). The main target of clustering is to cluster all the predictor variables into homogeneous group on the basis of their similar characteristics. Therefore, data clustering is the best process to predict future of students on the basis of their academic performance (Kefaya, 2009).

The data mining in educational sector have encouraged researchers to discover knowledge from the student data. Data mining techniques can be approached in educational sectors to maximize our understanding of learning focus on identifying, extracting and evaluating variables related to the learning process of students (Ch.Rupa, 2013).

Today, prediction on student academic performance has been a hot topic to each and every researchers interested in Educational Data Mining (EDM). The main objective of predicting student’s academic performance through EDM is to provide an appropriate suggestions to the teachers to enhance or update their teaching/learning strategy (Baker, 2013).

The main purpose for using data mining techniques is to estimate unexpected relationships between student’s different parameters (i.e. predictor factors). Thus, the data mining techniques can extract a maximum knowledge from large datasets (Nikolovski, 2015).

The Educational Data Mining process transforms raw data coming from educational institutions into meaningful predictive information that could potentially have a great impact on predicting student’s future learning behavior (Parneet Kaur, 2015).

Educational Data Mining Techniques cover the predictive model which is usually used in predicting academic performance (Abdulazeez Yusuf, 2016). The Educational Data Mining is the process of extracting meaningful predictive information and knowledge from large scale of dataset. EDM not only has the power to collect and manage data, but also has the power to analyze dataset for the predictive tasks (Ahmad, 2015).

The educational institutions try to measure the percentage of successful students through several subject-wise physical examinations and data mining techniques (Abdulazeez Yusuf, 2016). According to the researchers, Abdulazeez Yusuf and Dr. Abdulahab Lawan, lack of system which can analyze and monitor the student’s performance is the main problem of their country Nigeria; and student academic performance evaluation is less examined by the higher educational institutions. This paper takes the literature review of Yusuf and Lawan’s research which is aimed to select specific determinant factors which can be relevant with academic performance. Student’s data which are available at educational institutions can be used to predict the academic performance of the student using data mining technique (Abdulazeez Yusuf, 2016).

#### **2.2.1.2..1. Critical Analysis of Domain Research**

Here, domain research has been done to get clear knowledge about research area. Most of the research paper is similar and have clear definition of data mining technique. Most of the selected papers are researched through data mining algorithms but that work has been done through supervised classification algorithms. Very less research is found that has been done in cluster based algorithms.

I think when maximum researcher choose unsupervised learning algorithms then the revolution will come in student’s academic performance prediction model. Because, day to day the world is going to collect maximum datasets rapidly. Therefore, unsupervised learning algorithm will be more applicable than supervised learning algorithms.

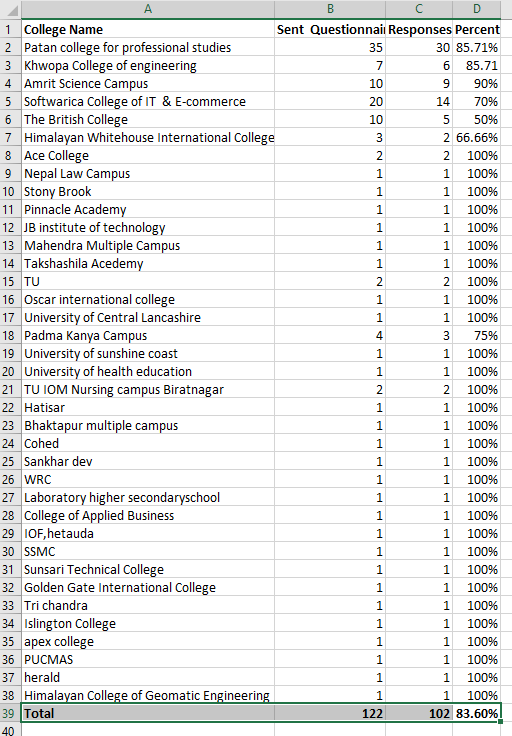
## **Discussion on your finding regarding similar system and existing technology**

I found many similar systems and existing technology. In **Section 2.2.1.1** and **Section 2.2.1.2,** there are two different similar system which are using WEKA library toolkit and its inner algorithms. The system has multiple data mining algorithms such as SAMME boosting algorithm, Adaboost.M1, LogitBoost Algorithm, C4.5 Decision Tree Algorithm, Decision Tree J48 Algorithm, K-Nearest Neighbour Algorithm and Aggregating One-Dependence Estimators Algorithm and combined multiple algorithm etc.

In System designing part, Java Programming Language has been used for GUI. But in my view Python and Matlab programming languages are the best for GUI development. Because these programming languages have lots of features that will be implemented in prediction model.

# **Data Analysis and Requirement Validation**

The data which is come from survey where analysed and find out their correlation with the research (Johnson, 2011). Total 102 data are collected through survey. The total responses rate form survey is given in table 3. Here total response rate is 83.60% for this research survey. This is catagorised through student’s college name.



*Table 3: Total Survey Responses Data*

## **3.1. Analysis of Collected Data**

Collecting Data Using Questionnaire. There are multiple sets of questions that was designed to ask to the students for getting feedbacks as a responses. Following questions are asked to them with following respective reasons.

**Survey Number 1. Gender**

*Fig. 9: Gender*

The total number of participants is 102 of which 56% and 46% of participants were male and female respectively. Asking their gender to each and every student is objective because, there is still possible to have that chance where the student’s performance is taking direct impact while being girls in our society. Because the response rate is low as comparing with boys. But not too much lower than my expectations.

**Survey Number 2. Age**

*Fig. 10: Age*

3% of the total participants were above 28 years of age. 6% of the participants were between 26 to 28 years. 21% of them were between 17 to 19 years. 35% of them were between 23 to 25 years and remaining 35% of them were between 20 to 22 years.

The age group also supportive area for the predictive model. That data also can meet their objective to predict student’s academic performance.

**Survey Number 3. Your last year’s result (academic)**

*Fig. 11: Last year academic result*

The last year’s result of the participants were classified as honours, outstanding, good, pass and fail. And based on the opinions of the participants the above pie chart shows how their last year’s academic result occur to be. It shows that most of the students are average students according to their result output. Fail and honours ratio is equal here. It means that, 1% of student will hardly pass and other 1% students are hardly get honuours degree in their academic life.

**Survey Number 4. Your Family’s highest level of education**

*Fig. 12: Family’s Highest Level of Education*

The above pie chart shows the highest level of education attained by any family member of any individual participant. From the above data, most of them have graduate studies as family’s highest level of education. So, family education background is also factor that directly imparts each and every student’s academic performance.

**Survey Number 5. Have you ever had a class top result?**

*Fig. 13: Class Top Result*

45% of the participants has topped their class result at least once. This question can also predict student’s academic performance by measuring their historical education career. This is the one of the key concept to predict someone’s academic performance.

**Survey Number 6. I study because I’m interested in learning.**

*Fig. 14: Statement Opinion*

This chart shows, if the participants were really interested in learning or not. Through this process, we can easily predict their academic performance conditions. If student have maximum interest in learning than there is maximum chance to have good academic performance. If some students are not have interest in learning then obviously their academic performance is too weak.

**Survey Number 7. I am happy to invest my time on study**

*Fig. 15: Happy to invest time on study or not*

This chart shows, if the participants were happy to put their time on studying or not. If they are really happy to study then their academic performance will be better. If not then their academic performance will be weak significantly.

**Survey Number 8. My academic performance depends on the effort I make**

*Fig. 16: academic performance depends on the effort*

This bar chart shows, the effort for their academic performance of student. If at that statement student is totally agree, it means that his/her academic performance is very good. The option agree is also good for academic performance. But, other options are denotes weak academic performance.

**Survey Number 9. Normally, my state of mind is positive and I feel good**

*Fig. 17: Feelings*

Here, this data shows, the user’s feelings about their positive and negative mind. In this data, maximum 53.9% of student have agree with this statement and other 27.5% are totally agree with it. It means that normally while they are studying, their mind is positive and they feel good at that time. This positive data also indicates the maximum positive responses about student’s academic performance.

**Survey Number 10. I know the objective of the course subjects**

*Fig. 18: The objectives of the course subjects*

In this survey question, 63% of student know the objective of the course subjects while they are studying. This data is maximum positive rate for this question. If student’s knowledge is more for particular course subject then their academic performance in that education sector is strong. From that data predictor will be able to predict positively.

**Survey Number 11. I create a suitable atmosphere to study in to be productive**

*Fig. 19: atmosphere to study in to be productive*

Here, 56.90% of students are agree with the statement i.e. student will create suitable atmosphere to study in to be productive. And second highest response is 21.60% in totally agree with that statement. Respectively other data are 13.70% on undecided and 7.80% on disagree with lower then totally agree and agree statement. As a conclusion, maximum student’s academic performance is better.

**Survey Number 12. I pick suitable classmates for teamwork**

*Fig. 20: Suitable Classmates for Teamwork*

In this statement 56% and other 26% of students are very serious about their education. So we can be analyse that, their academic performance is better in their future.

**Survey Number 13. What I learn in some course subject can be used in others and also in my future profession**

*Fig. 21: Subject Relation with Future*

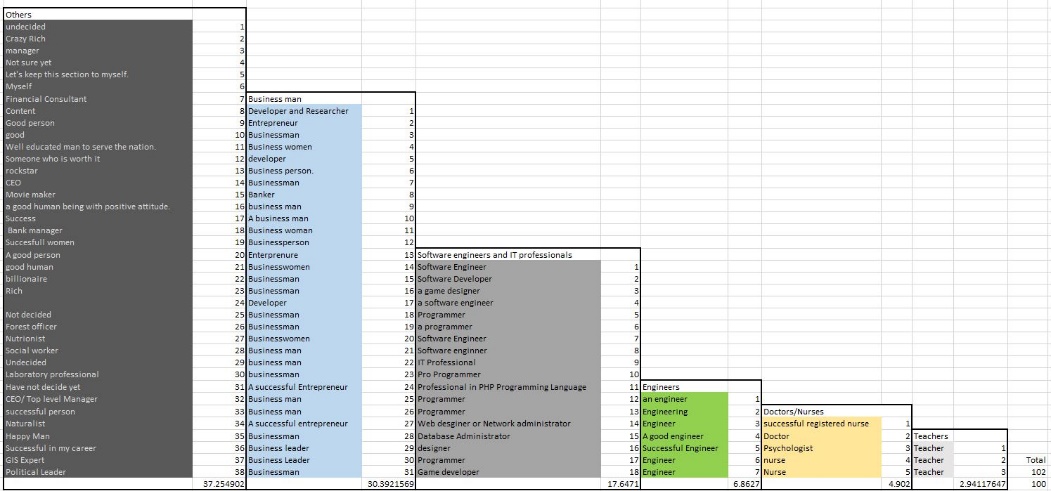
In this bar diagram, 38.20% of students and 51% of other students are agree with the statement that says their course content will useful in their future life. And other 9.8% of student are undecided with that statement and 1% of students are totally disagree with dis statement. Those person who select totally agree and agree option, their academic performance is good.

**Survey Number 14. What makes me comfortable while I study.**

*Fig. 22: Tools*

Above chart shows the tools that were thought to be needed for the participants for helping in their study. Of this tools, 32% chose book stand as their preferred tool while studying, 25% chose computer armset, 4% chose vibrating neck massager, 19% chose LED booklight, 6% chose applications like Noisli, 2% chose music. And remaining 12% chose other opinions like free environment, etc.

**Survey Number 15. In the future, I want to become...**

**

*Table 4: Categorizing answers into excel table.*

*Fig. 23: Bar graph of future choice*

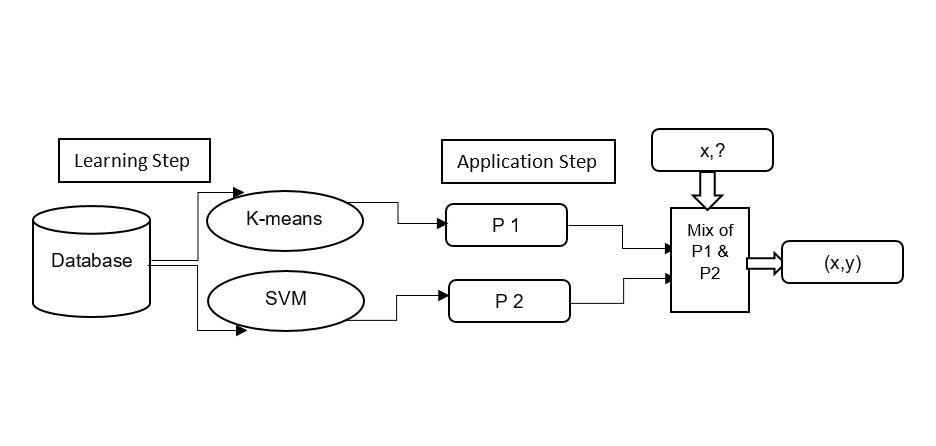
This question possibly had more answers because the participants had free choice for answer. But, their answers were divided among six categories as shown in above chart (fig 23). Their answers were manually asserted into one of those categories as shown in table4 (excel sheet).

## **3.2. Relate how the findings of the various research methods applied**

By processing those data, we can be able to judgment some of decisions which can directly maters with the objective of the research. Those questionnaires are designed for students for the purpose of collecting predictive answers form them.

# **System Architecture**

The system of this project is very simple, prediction model is worked through clustering and classification model.

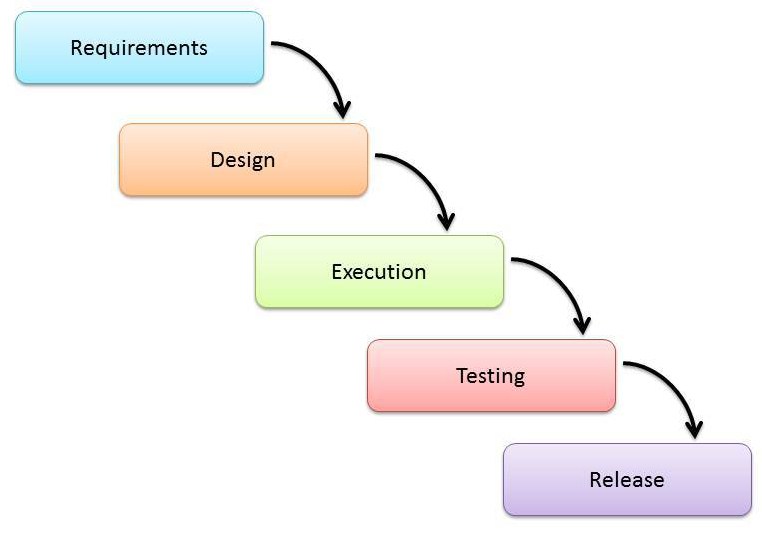


*Fig 24: Final Proposed Model*

Here, this model will be worked as a Hybrid Model with the combination of K-means Algorithm and Super Vector Machine Algorithm. These two algorithms are the combination of supervised and unsupervised algorithm.

## **4.1. SDLC**

Under Software Development Life Cycle (SDLC), the developed system followed the Waterfall Model for deliver better software project.



(Mohd Herry Mohd Nasir, 2019)

*Fig 26: Waterfall Model*

## **Requirements**

The requirements of the project are catagorised according to section 4.2.1. and 4.2.2.

### **4.2.1. Functional**

|  |  |
| --- | --- |
| **ID** | FR1 |
| **Title** | Register and Login |
| **Description** | In database, admin has user id 1 and other are registered as a student/ normal user of the system. |
| **Rational** | - Only on id holder of database table can be admin.  - Blank username and password can’t be logged in database for login. |

|  |  |
| --- | --- |
| **ID** | FR2 |
| **Title** | Admin Dashboard |
| **Description** | Admin can insert student details, their internal marks, modify and remove detail as well as can able to cluster data and predict their performance. |
| **Rational** | - Only admin can have permission to insert student details. |

|  |  |
| --- | --- |
| **ID** | FR3 |
| **Title** | Student Dashboard |
| **Description** | Different Panel for students with limited accessibility. |
| **Rational** | - Student doesn’t have full access. |

|  |  |
| --- | --- |
| **ID** | FR4 |
| **Title** | Add Student for Login Access. |
| **Description** | Admin can able to create student login profile. |
| **Rational** | - Incorrect student login details will not accepted by the system.  - Student login data will has its own directory in database file. |

|  |  |
| --- | --- |
| **ID** | FR5 |
| **Title** | Remove Student from Login Access. |
| **Description** | Admin has full authority to remove student as a user login. |
| **Rational** | -Incorrect student login details will not accepted by the system.  - Login data also removed from the database file. |

|  |  |
| --- | --- |
| **ID** | FR6 |
| **Title** | Kmeans Clustering |
| **Description** | Dataset is cluster through kmeans clustering algorithm. |
| **Rational** | - Admin and Student also can able to use this feature. |

|  |  |
| --- | --- |
| **ID** | FR7 |
| **Title** | Performance Prediction |
| **Description** | Student’s internal mark can be predicted. |
| **Rational** | - Admin and Student also can able to use this feature. |

|  |  |
| --- | --- |
| **ID** | FR8 |
| **Title** | Insert, Update, Delete, Search Student Data and Marks |
| **Description** | Features that insert student’s details with marks as well as update remove and search them. |
| **Rational** | - Only admin can do that. |

### **4.2.2. Nonfunctional**

|  |  |
| --- | --- |
| **ID** | NFR1 |
| **Requirement** | Security |
| **Description** | In initial phase, this system doesn’t need any additional strong security feature inside. General login is enough for this time. |

|  |  |
| --- | --- |
| **ID** | NFR2 |
| **Requirement** | Portability |
| **Description** | This is also nonfunctional requirement of this application. |

|  |  |
| --- | --- |
| **ID** | NFR3 |
| **Requirement** | Certification |
| **Description** | This feature also no need to implement at this phase of development. |

|  |  |
| --- | --- |
| **ID** | NFR4 |
| **Requirement** | Scalability |
| **Description** | For further scalable case like, database upgrade, manpower expansion also nonfunctional at this this. |

|  |  |
| --- | --- |
| **ID** | NFR5 |
| **Requirement** | Optimization |
| **Description** | This is just an initial phase of development, so code optimization is not required. |

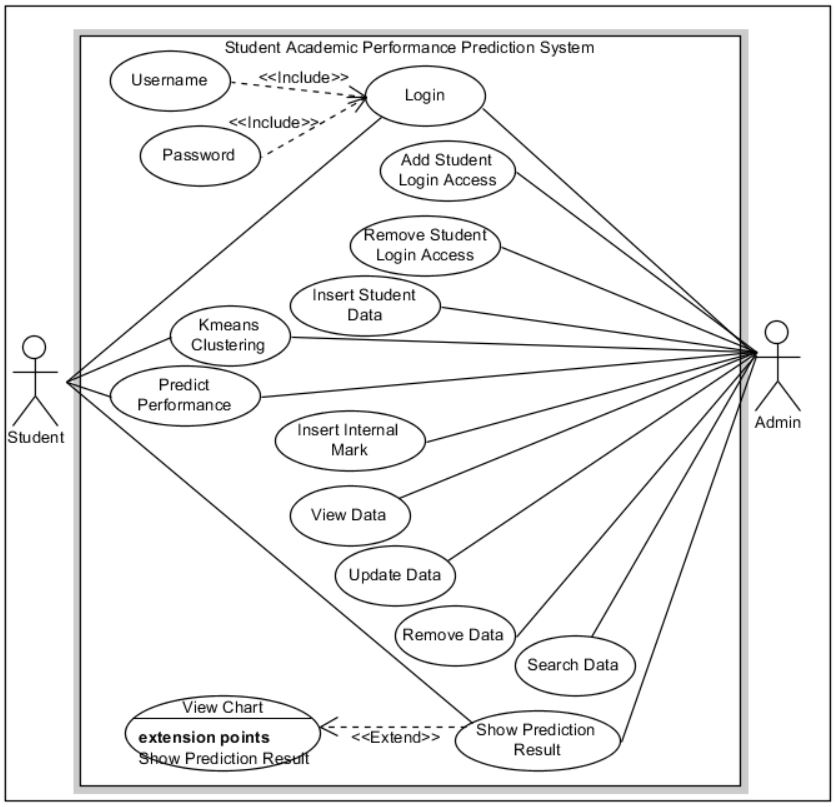
### **4.2.3. MoSCoW Prioritization**

|  |  |
| --- | --- |
| ID | MoSCoW Prioritization |
| FR1 | Must Have |
| FR2 | Must Have |
| FR3 | Must Have |
| FR4 | Must Have |
| FR5 | Must Have |
| FR6 | Must Have |
| FR7 | Must Have |
| FR8 | Must Have |

# **5. Design Diagrams**

All the designed diagrams are related with Student Academic Performance Prediction System.

**Use Case Diagram**

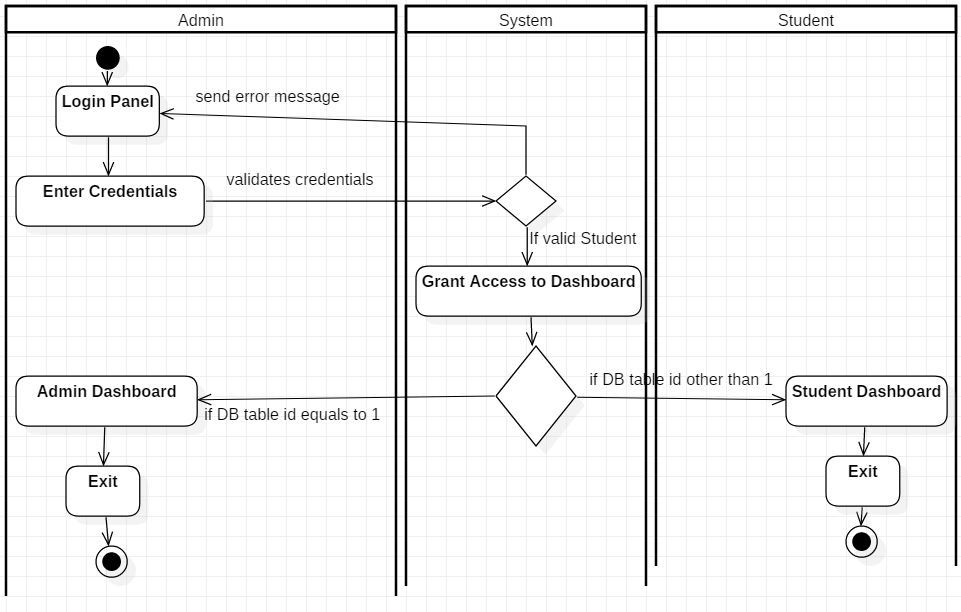


Here, admin have superior permission than student. Admin can able to login, creates students login details, insert, update, delete student’s details and their internal mark. And main thing admin have also have permission to do prediction and data cluster.

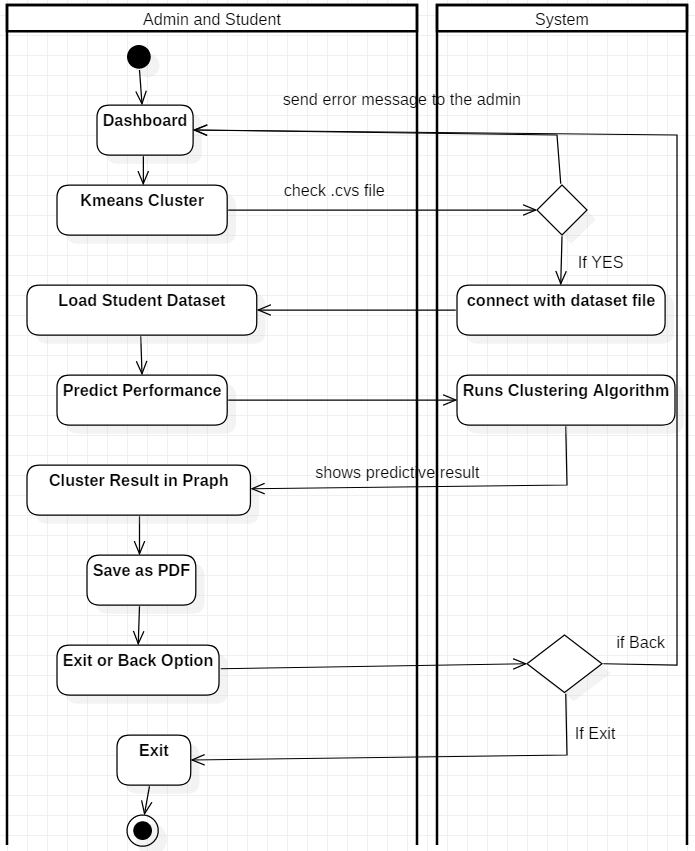
Student have very limited features in this system. Student can’t able to register new account. Admin has that right. But student can able to predict their academic performance. Also student can able to save that data.

**Activity Diagram**

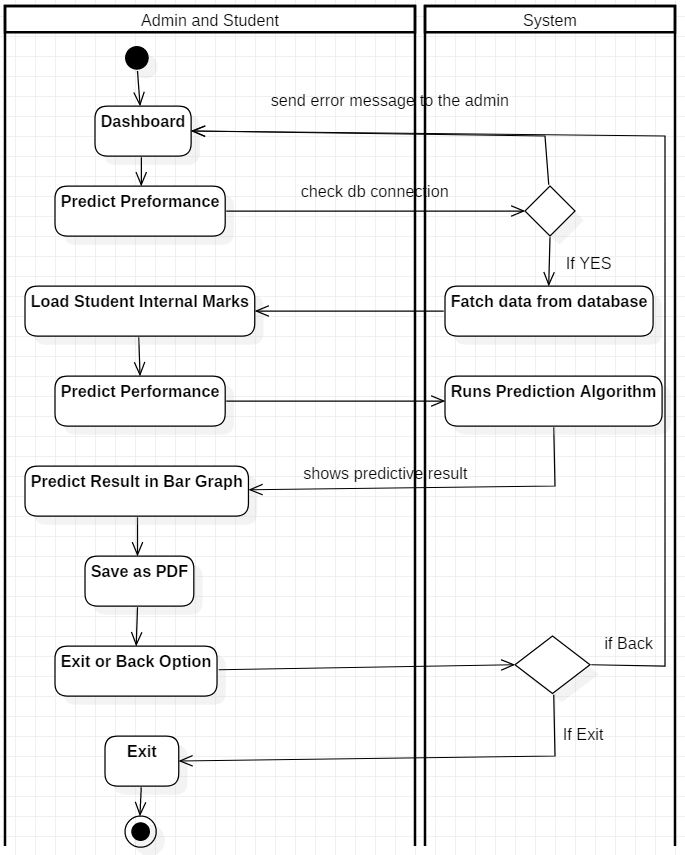
Activity1: Login for Admin and Student



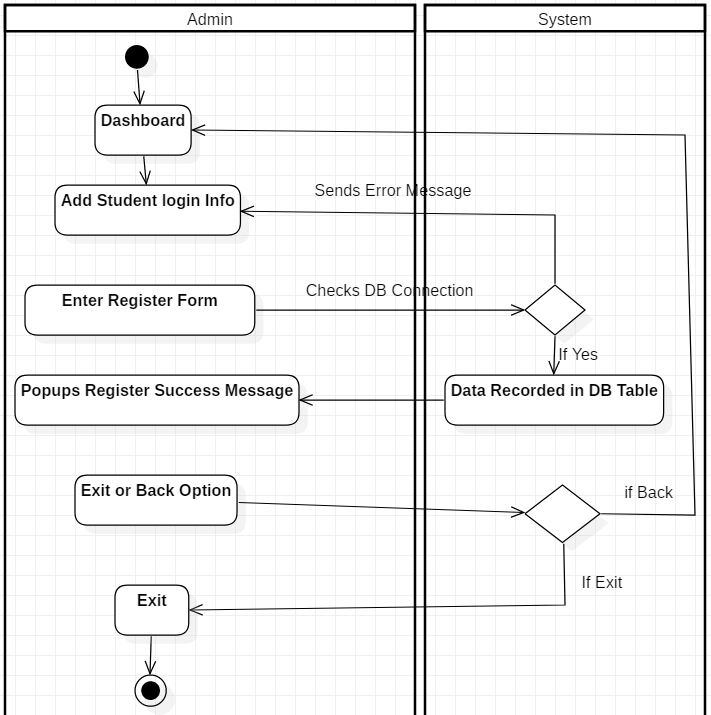
Activity2: Kmeans Clustering for Admin and Student



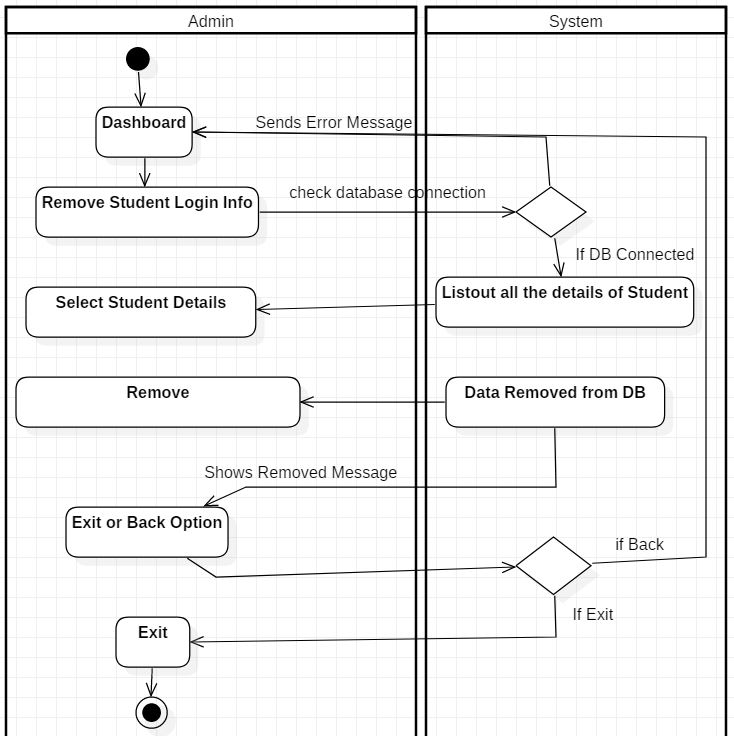
Activity3: Prediction Performance for Admin and Student



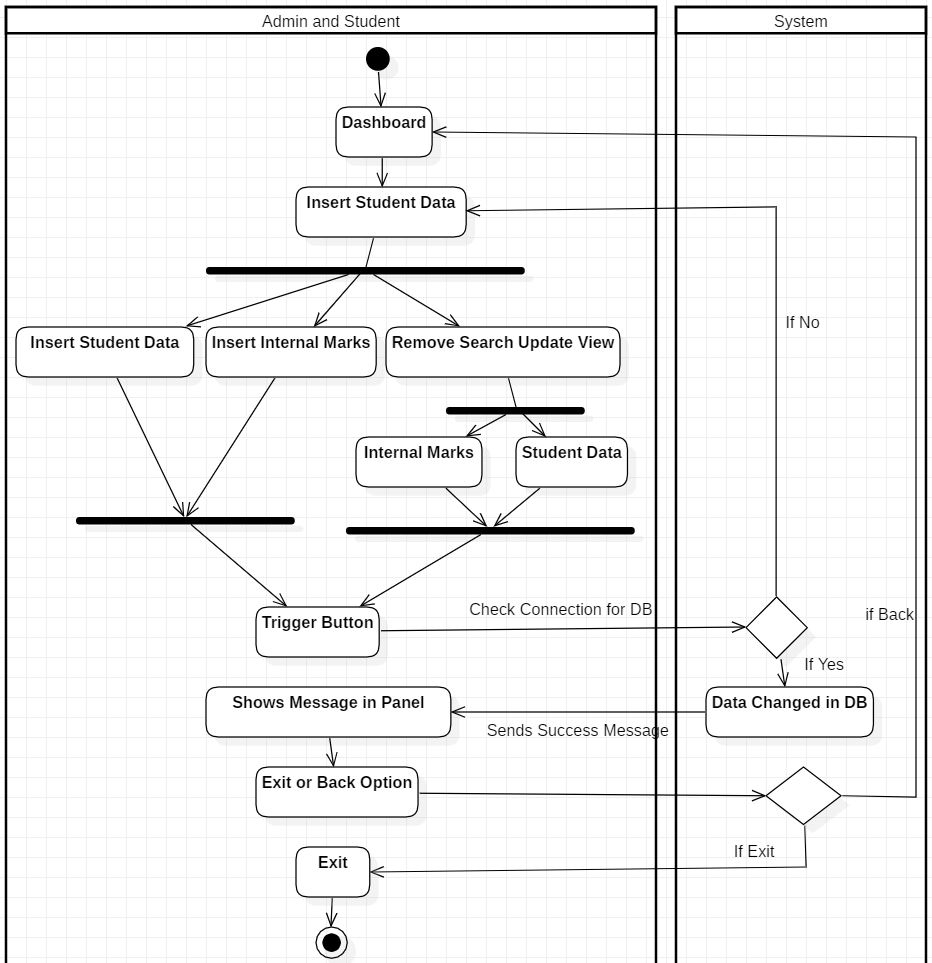
Activity4: Login Access for new Students through Admin’s Panel



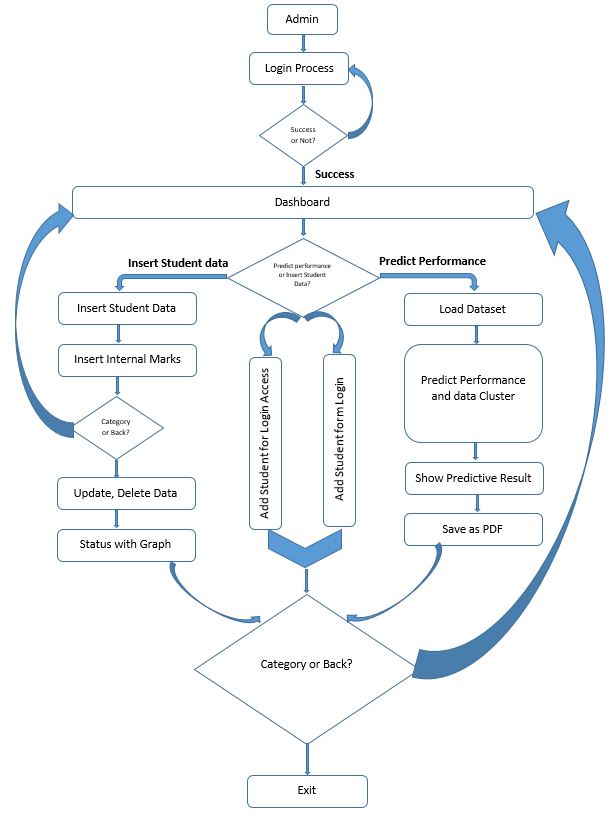
Activity5: Remove Login Access for Students through Admin’s Panel



Activity6: Other Insert, Update, Delete, Search Activities by Admin



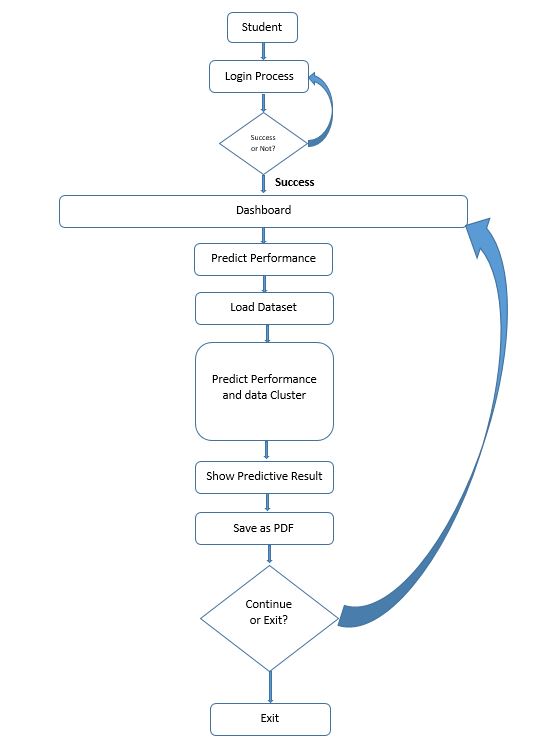
**Data Flow Diagram for Admin**



*Fig: Dataflow Diagram for Admin*

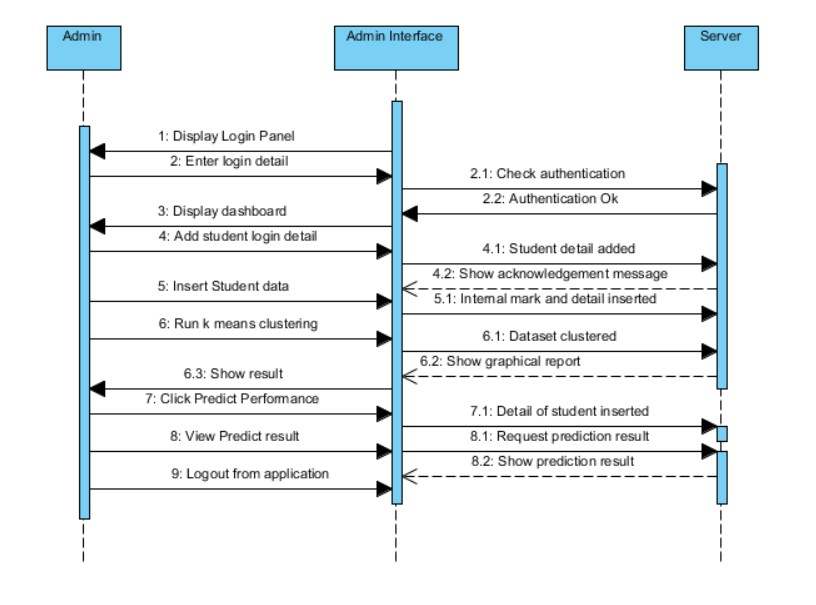
Here, this dataflow diagram is constructed for admin user.

**Dataflow Diagram for Student**



*Fig: Dataflow Diagram for Student*

Sequence Diagram for Admin



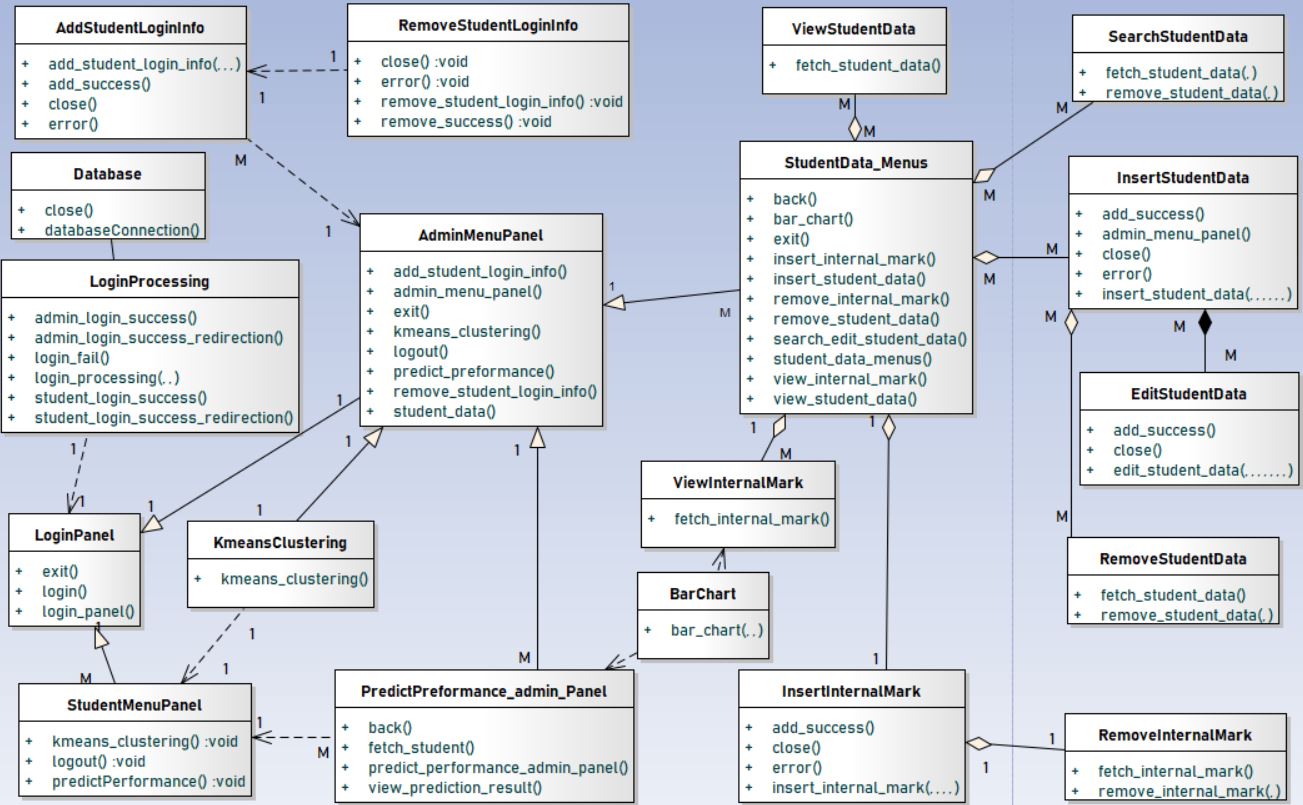
*Fig: Sequence Diagram for Admin*

Sequence Diagram for Student



*Fig: Sequence Diagram for Student*

Class Diagram



*Fig: Class Diagram*

# **Development & Testing**

This prediction model is developed through Python Programming Language and tested through Unit Testing, Integration Testing and Acceptance Test.

## **Development**

### **6.1.1. Programming Language**

The system has been developed through Python Programming Language. It has python version 3.7.0. The Python Programming Language is very advance programming language for AI. Python Programming Language is the high level language which can be very useful for open source language for predictive application design and development. And also python programming has lots of libraries that are mostly used to develop prediction model (Halterman, 2011). Therefore, python programming language is best to predict student academic performance.

### **6.1.2. IDE and Platform**

In Windows 10 pro 64x-bit Operating System, the PyCharm 2018.3.5 Community Edition was used to code this application. XAMPP version 3.2.2 for MySQL databse.

### **6.1.3. Libraries**

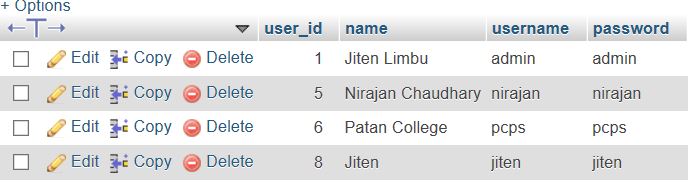
Some standard python libraries are needed to run application. Following Python Libraries are used to develop the application in PyCharm:

1. Pandas
2. Numpy
3. Matplotlib
4. Sql-connector
5. Kivisolver
6. Sklearn
7. Scikit-learn

### **6.1.4. Code**

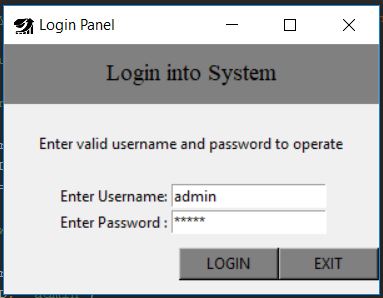
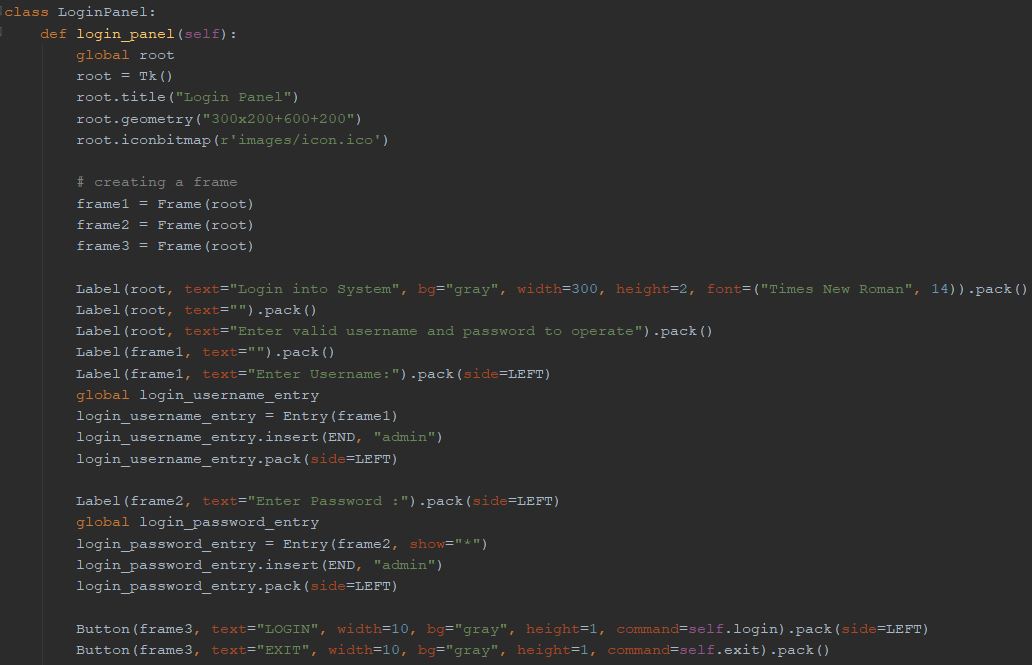
Login: Admin

An Admin can login with their default username and their password which is mentioned in database table of users.



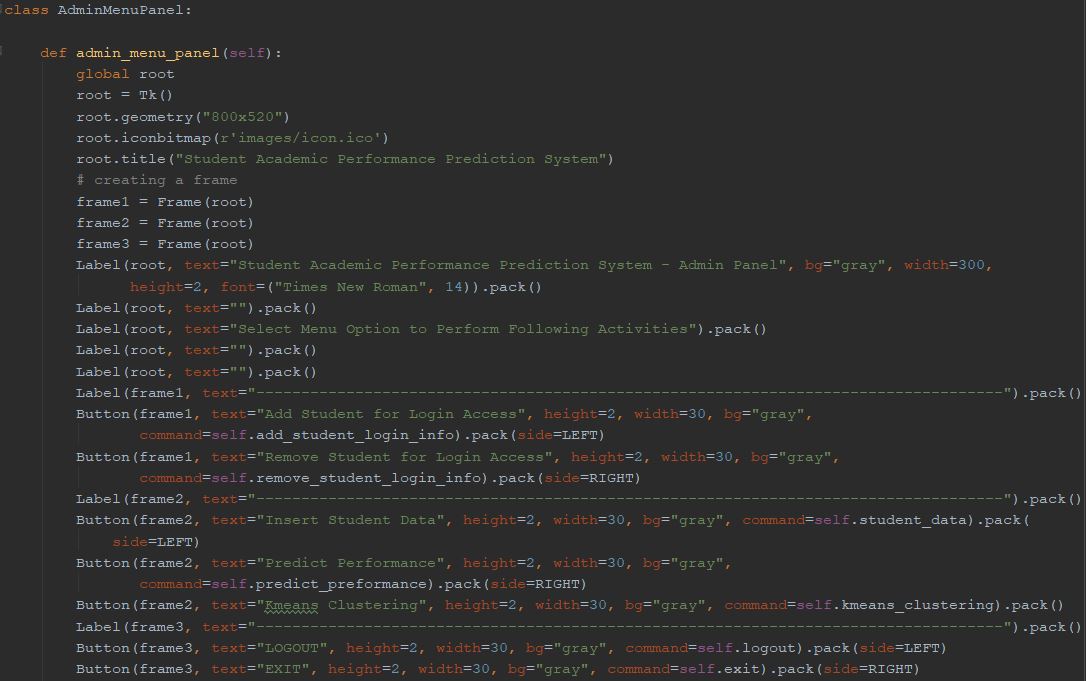
*Fig: Database Table*

Here, user\_id 1 is admin user and other are student. Admin can dynamically able to handle other student through admin’s dashboard.

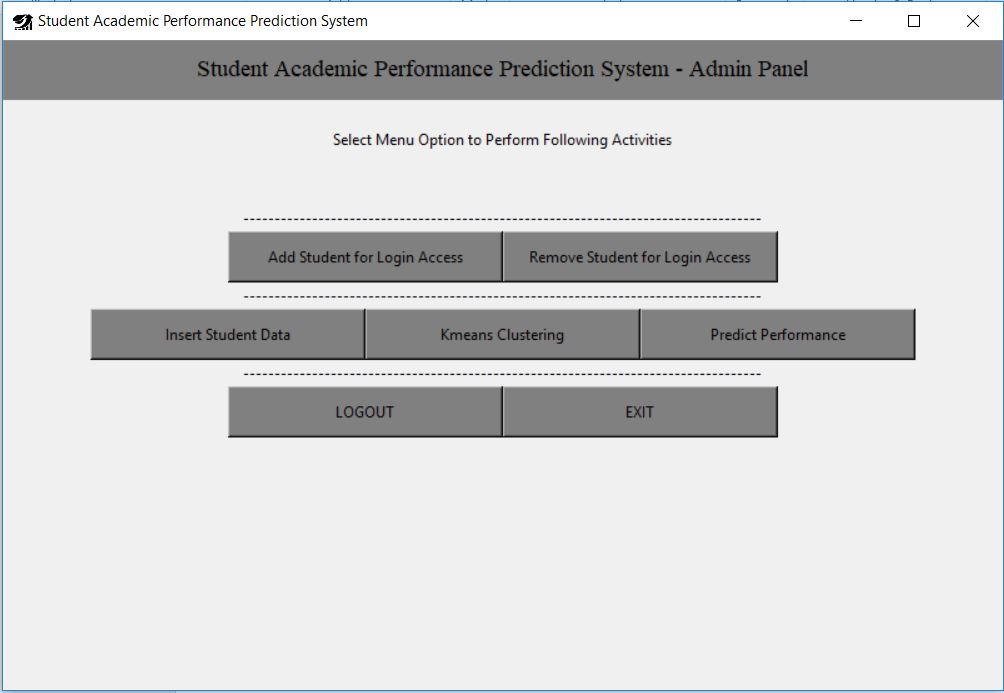


*Fig: Login Panel*

Admin Panel

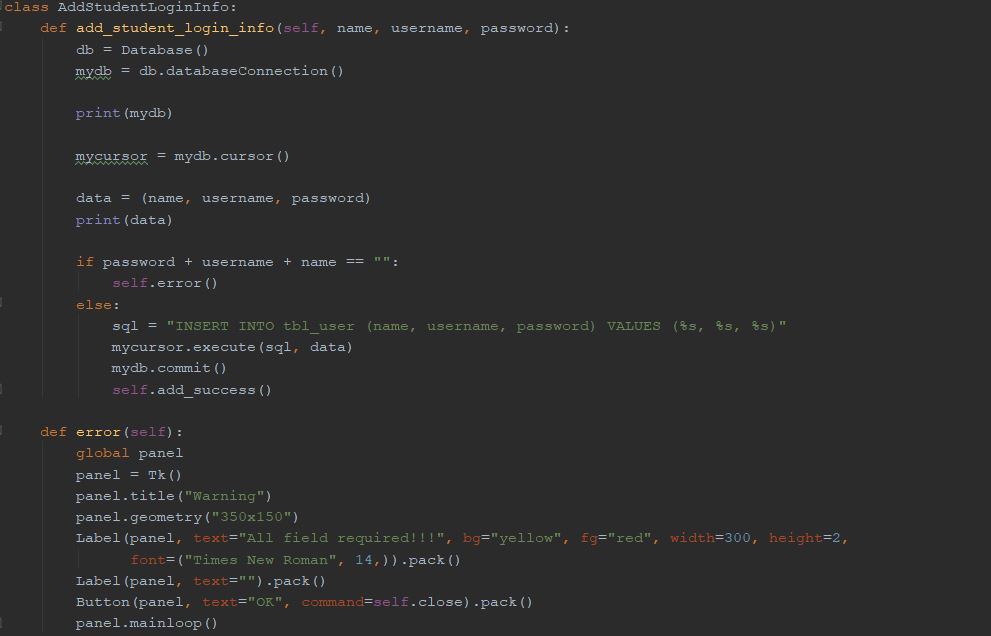


*Fig: Code of Admin Panel*

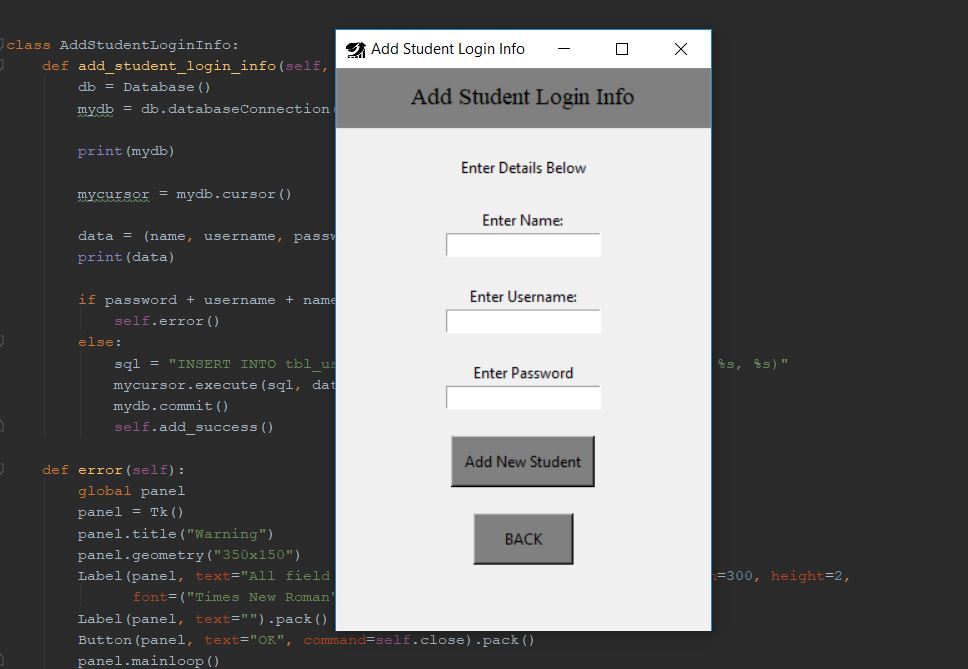


*Fig: Admin Dashboard*

Add Student Login Access



*Fig: Add Student Login Access Code*

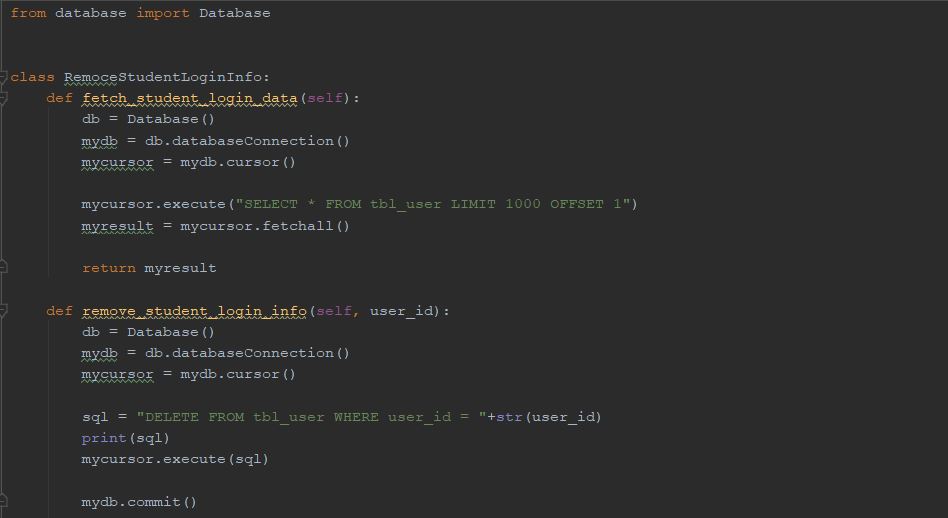


*Fig: Add Student Login Access Code*

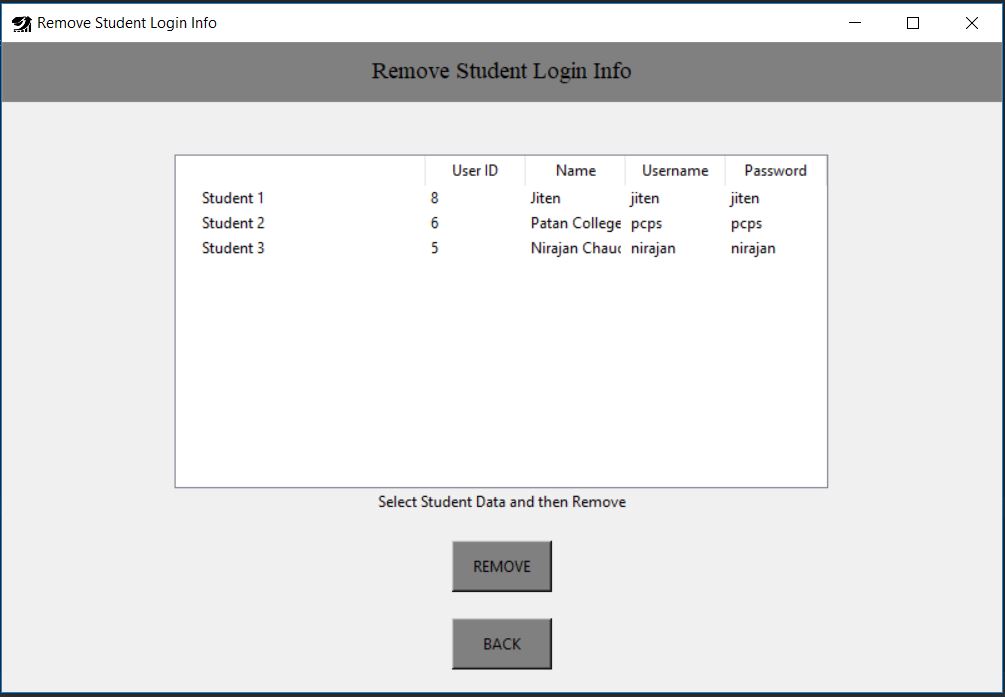
Remove Student Login info



*Fig: Remove Student Login Access Panel*

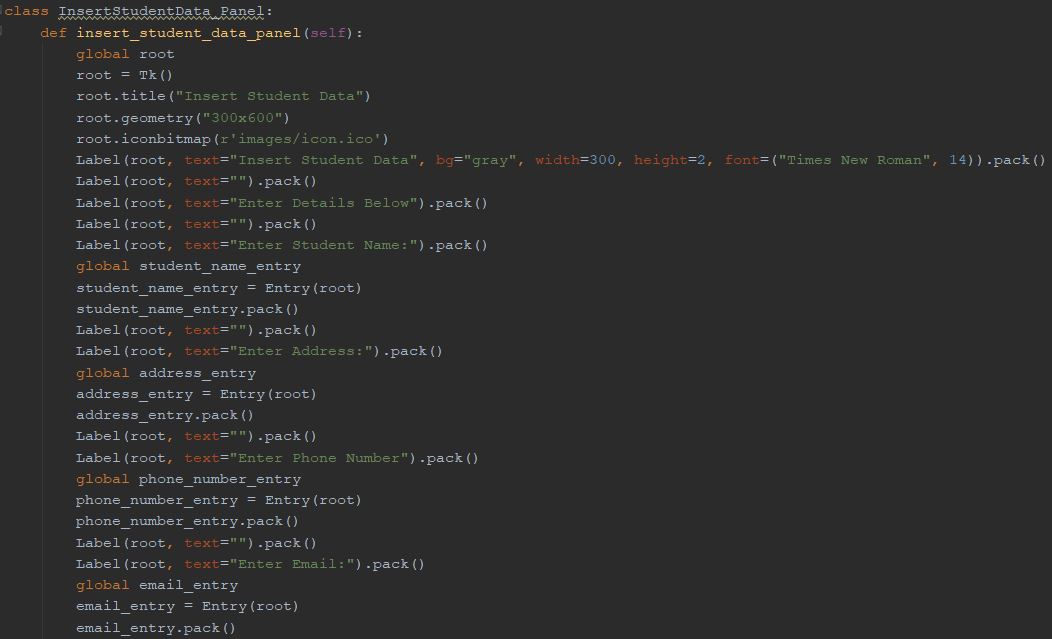


*Fig: Remove Student Login Access Code*

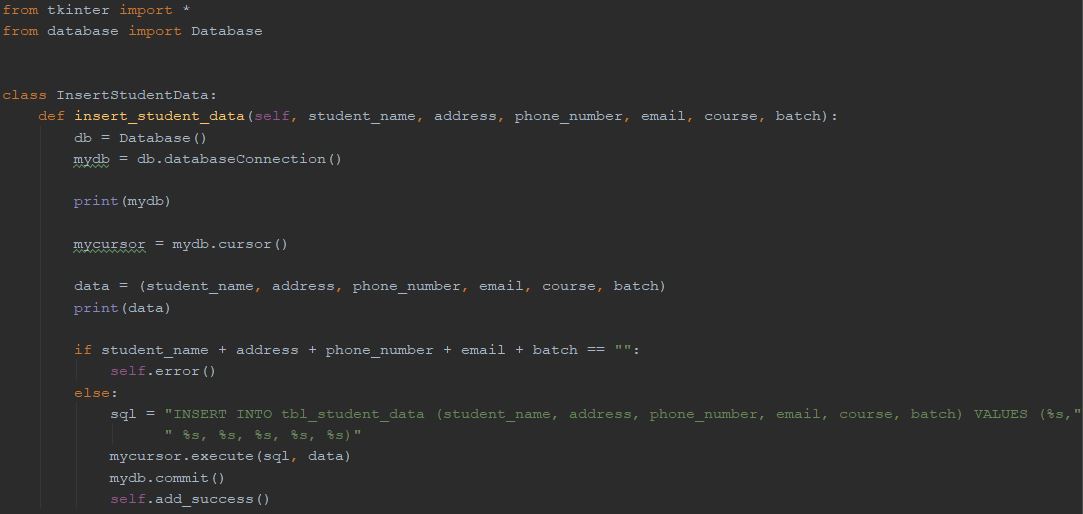


*Fig: Remove Student Login Access GUI*

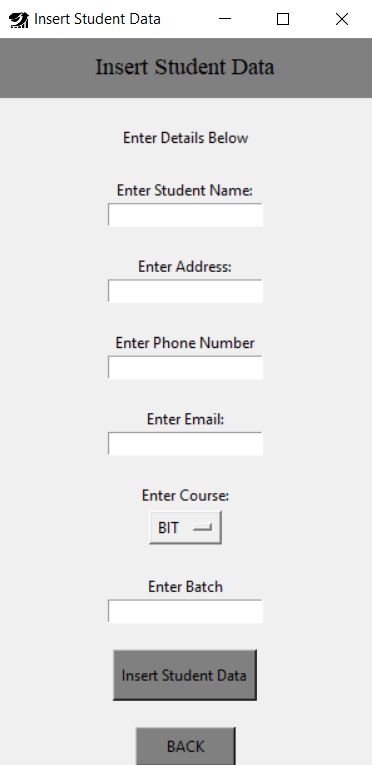
Insert Student Data



*Fig: Insert Student Data Panel*

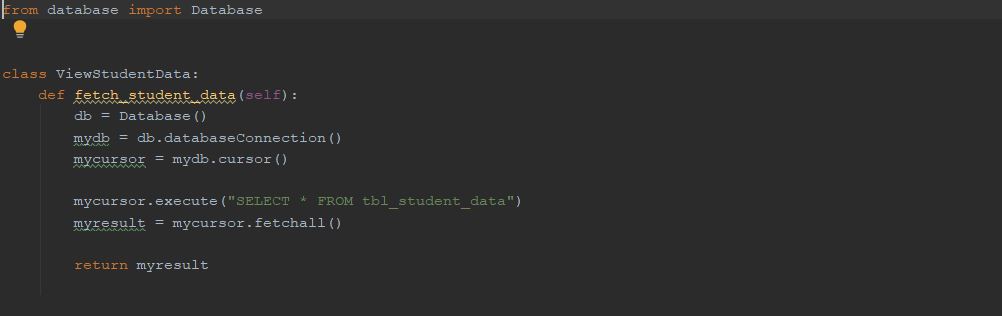


*Fig: Insert Student Data Code*

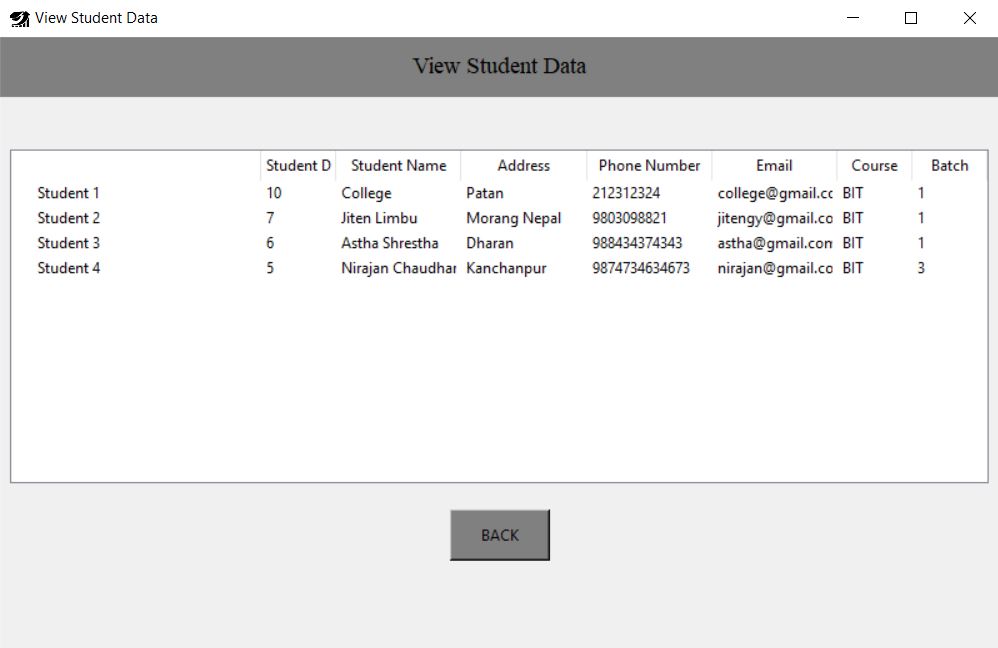


*Fig: Insert Student Data GUI*

View Student Data

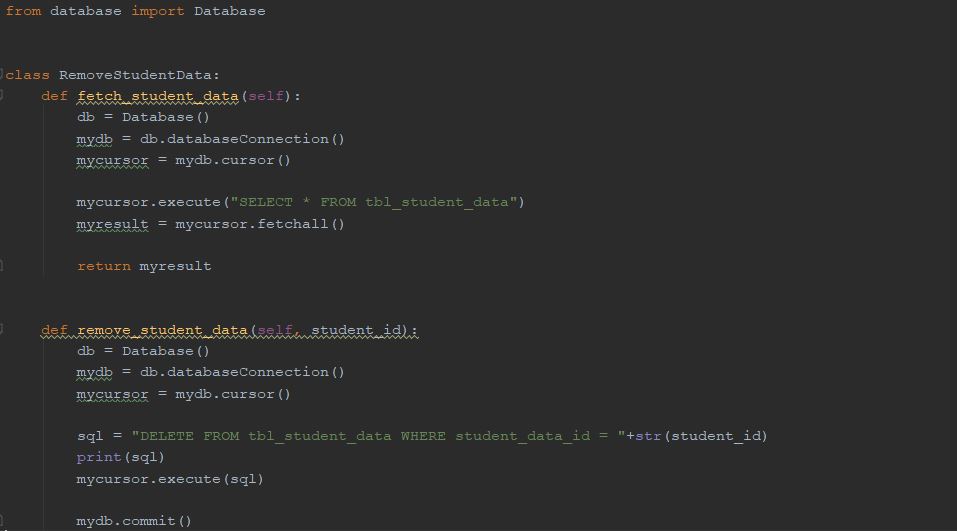


*Fig: View Student Data*

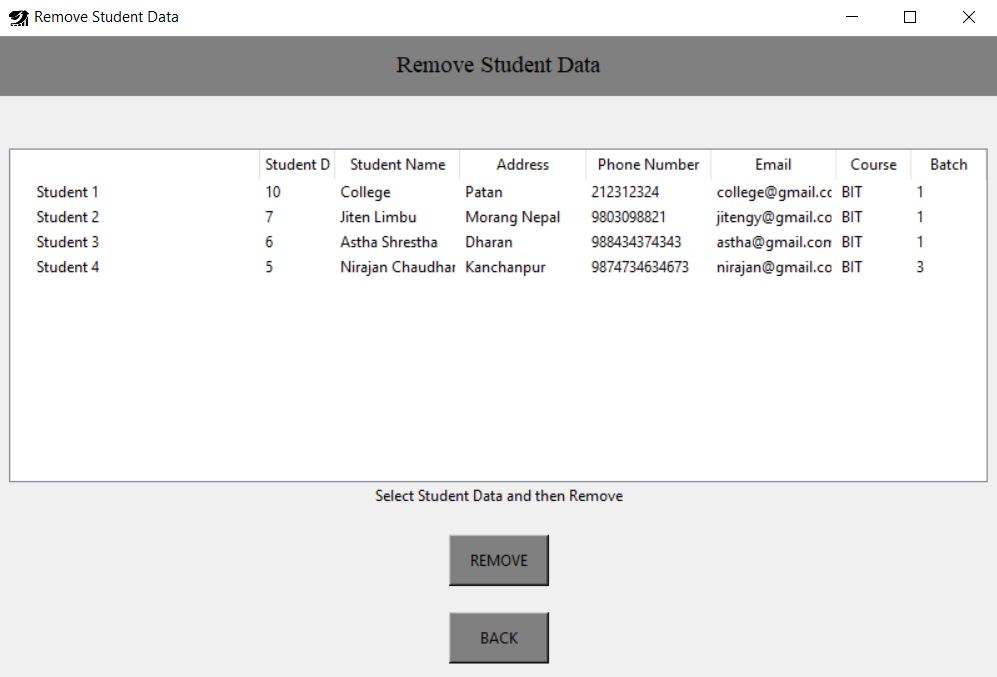


*Fig: View Student Data GUI*

Remove Student Data

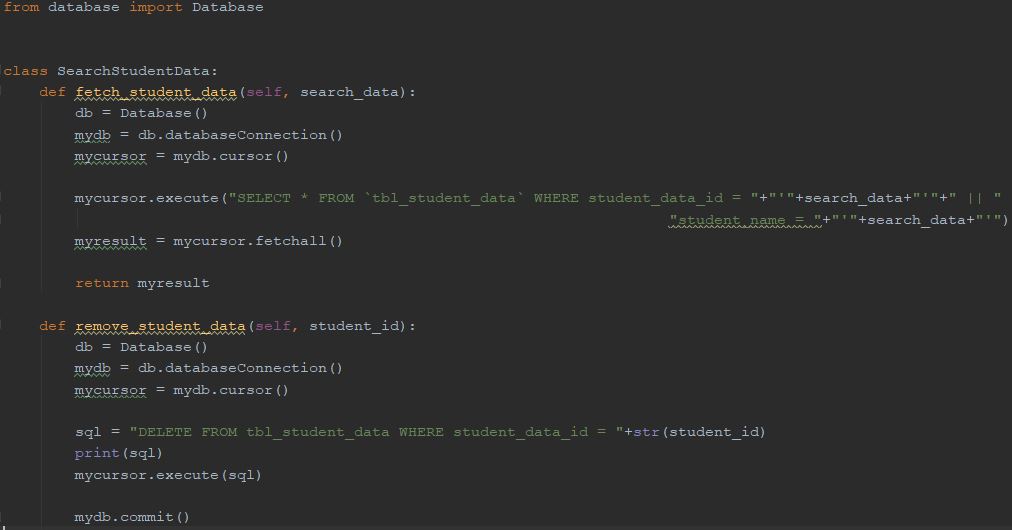


*Fig: Remove Student Data*



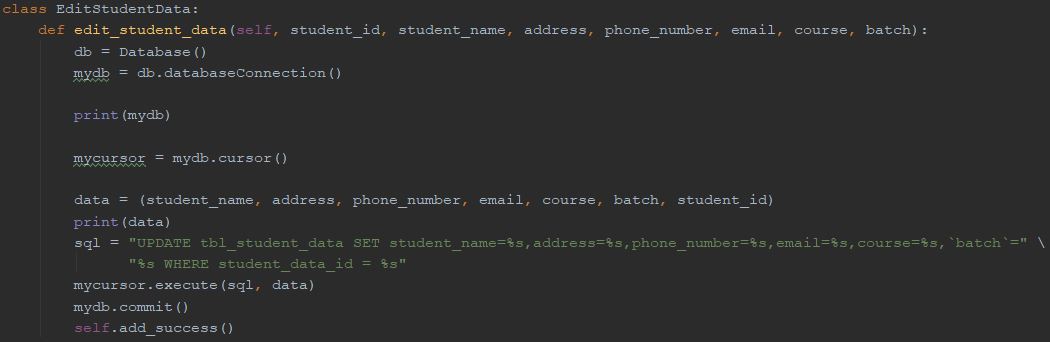
*Fig: Remove Student Data GUI*

Search Student Data



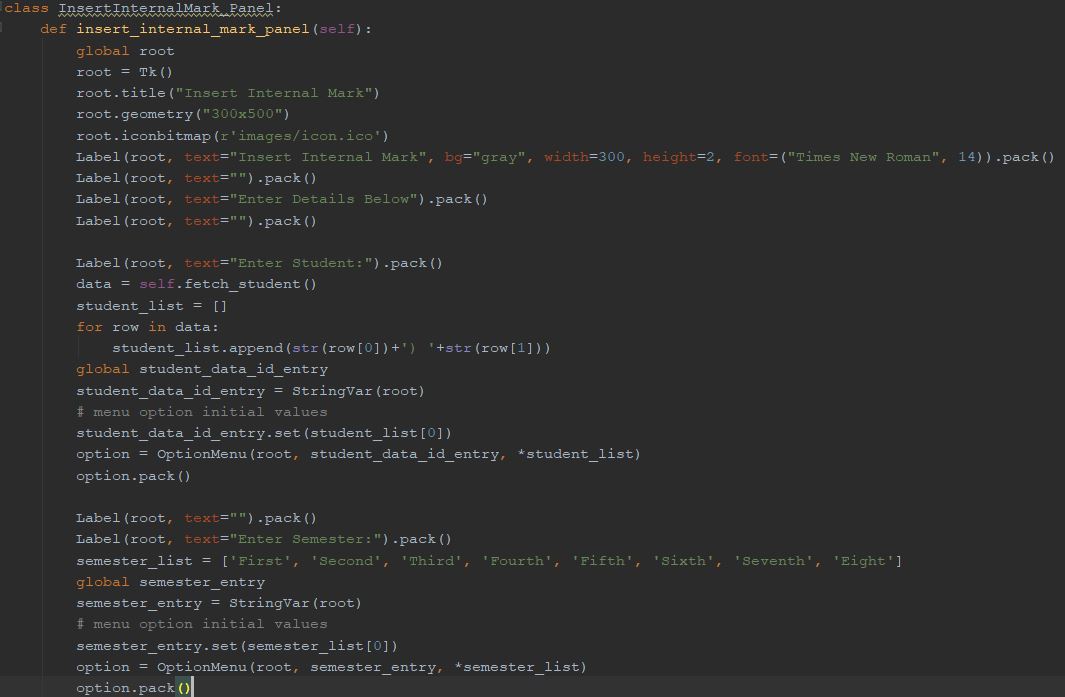
*Fig: Search Student Data*

Update Student Data

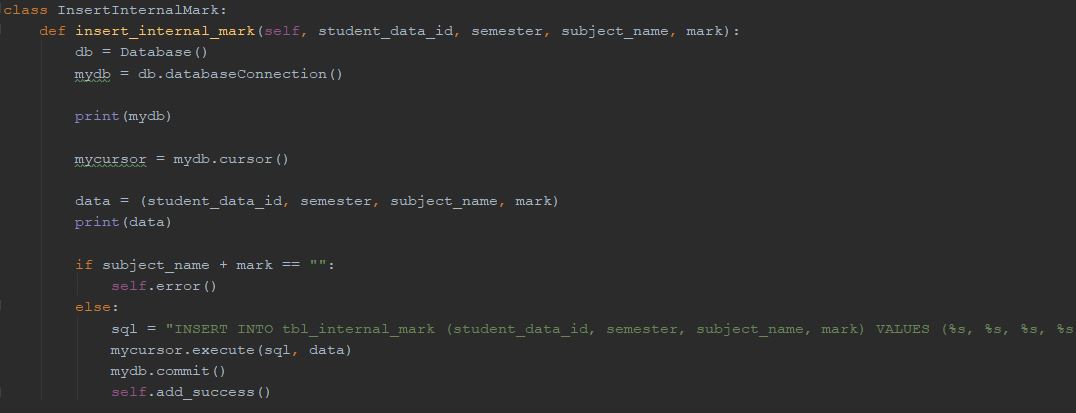


*Fig: Update Student Data*

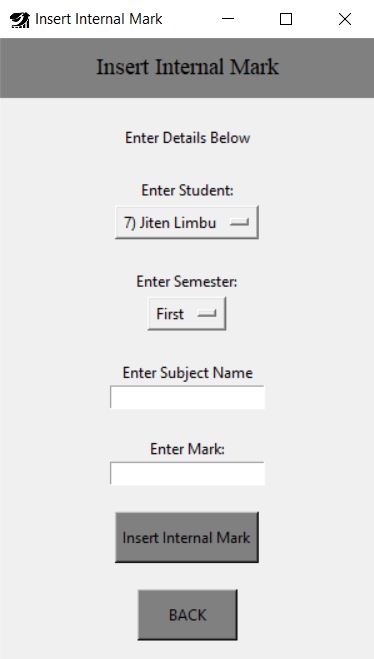
Insert Internal Mark



*Fig: Insert Internal Mark Panel*

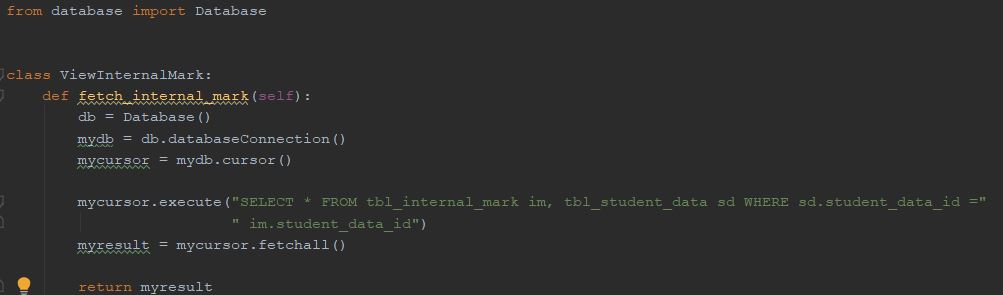


*Fig: Insert Internal Mark Code*



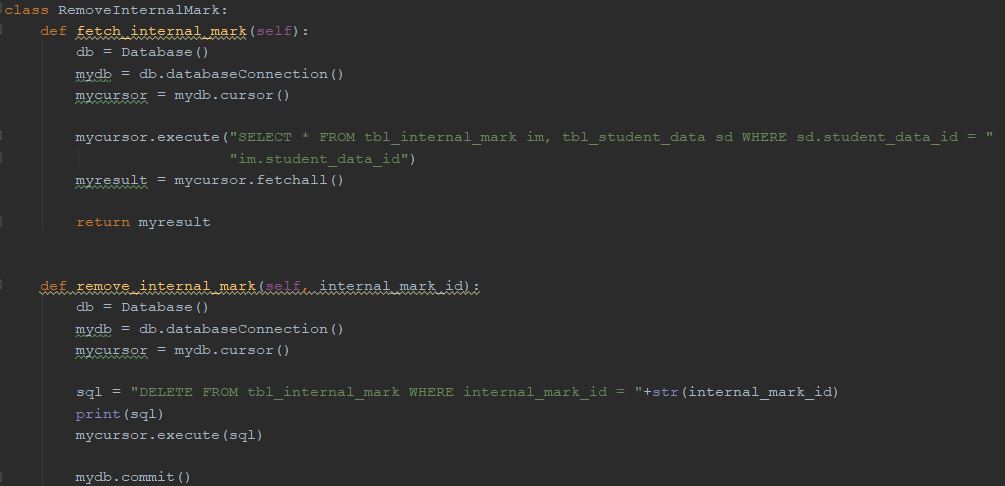
*Fig: Insert Internal Mark GUI*

View Internal Mark



*Fig: View Internal Mark*

Remove Internal Mark

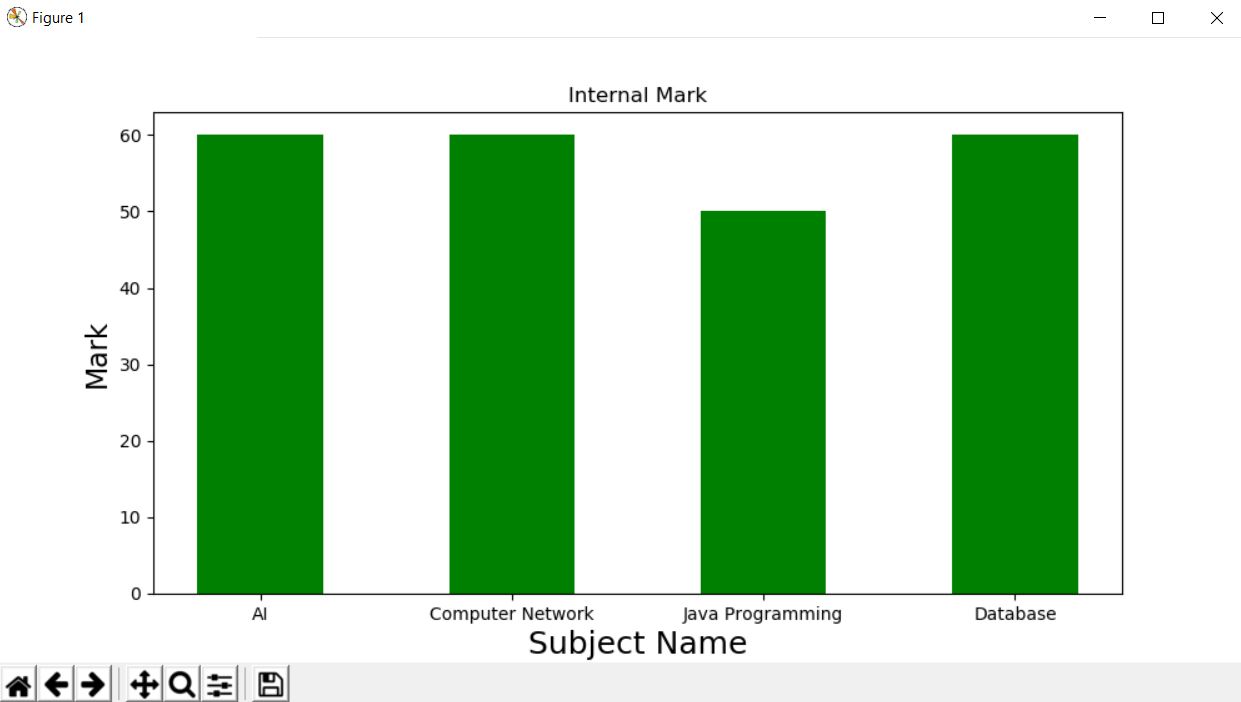


*Fig: Remove Internal Mark*

Bar Chart for Internal Mark

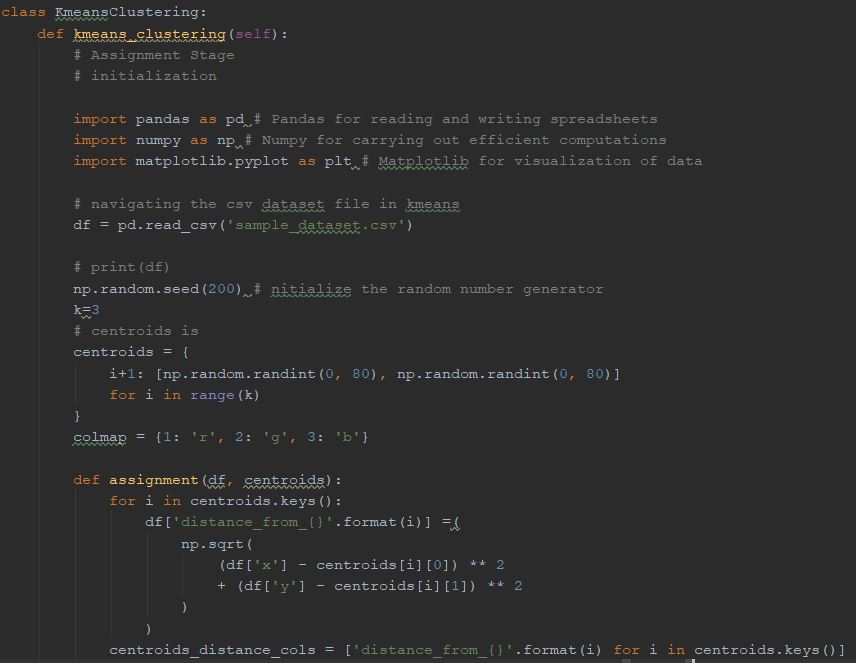


*Fig: Bar Chart for Internal Mark*

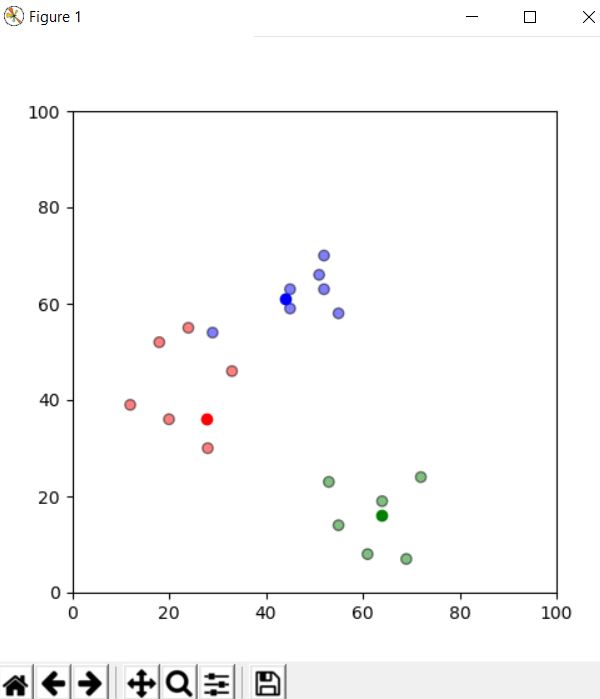


*Fig: Bar Chart*

Kmeans Clustering

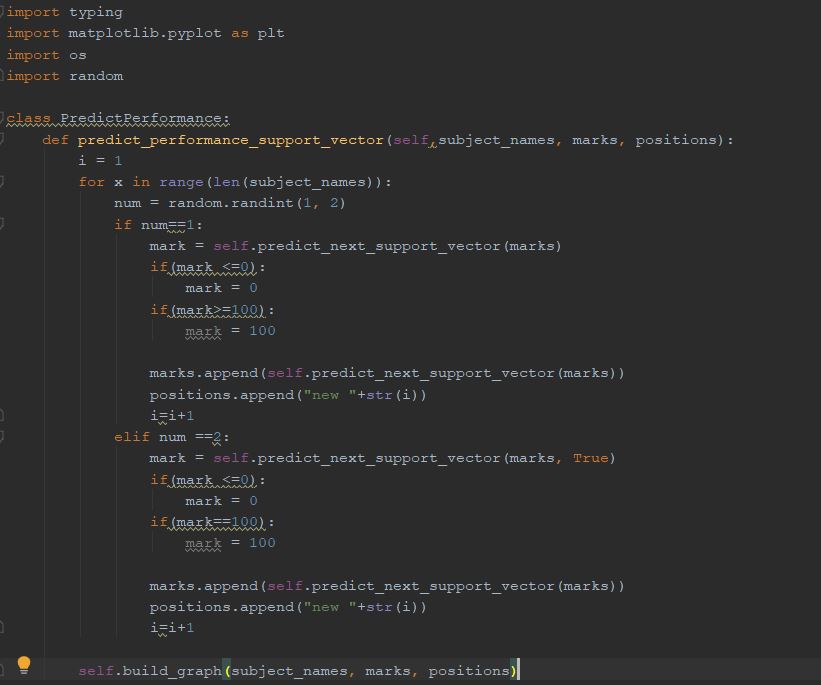


*Fig: Kmeans Clustering*

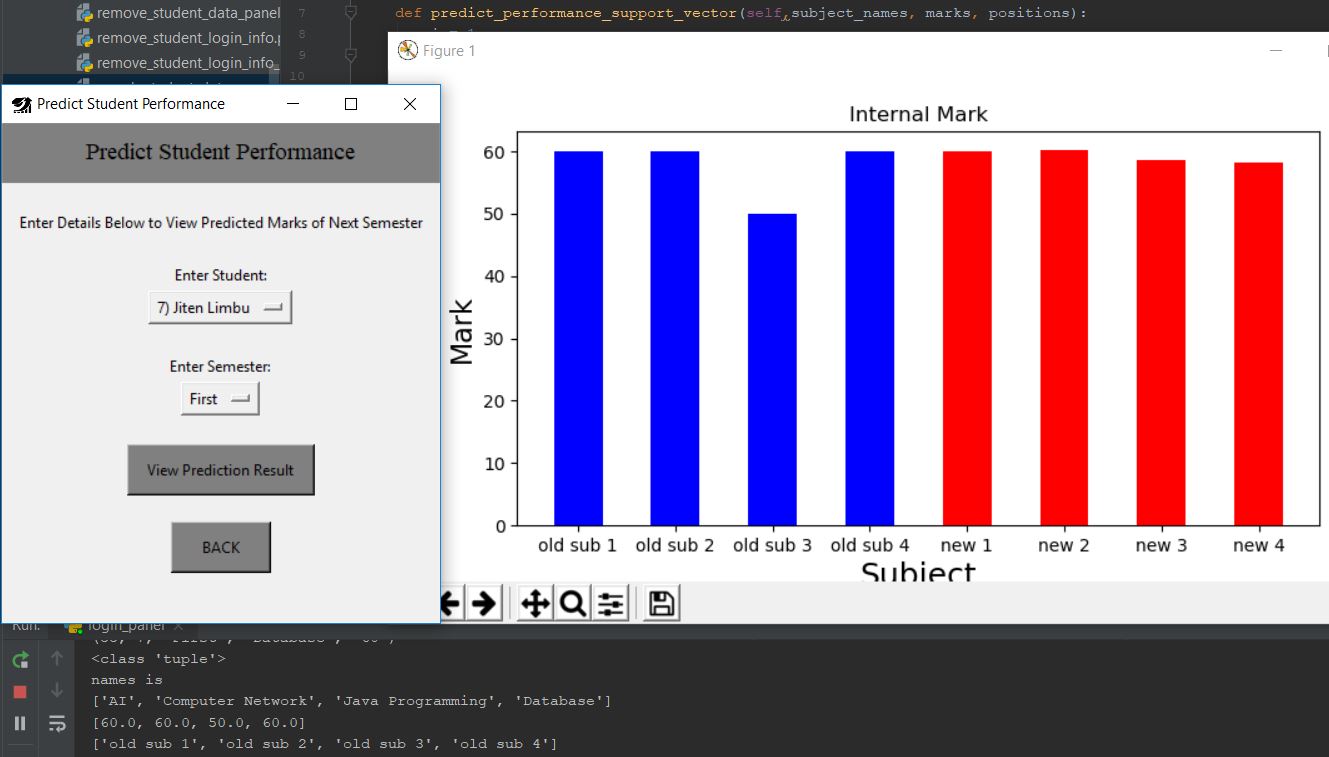


*Fig: Clustered Data*

Prediction Performance

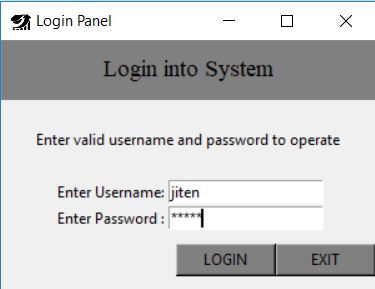


*Fig: Predict performance*

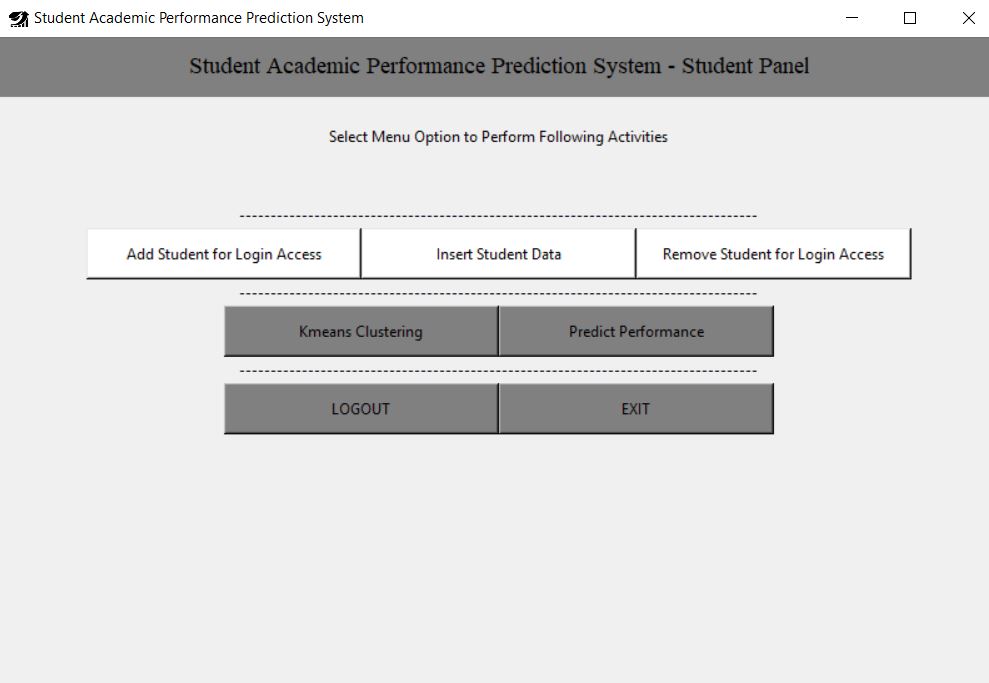


*Fig: Predictive Output*

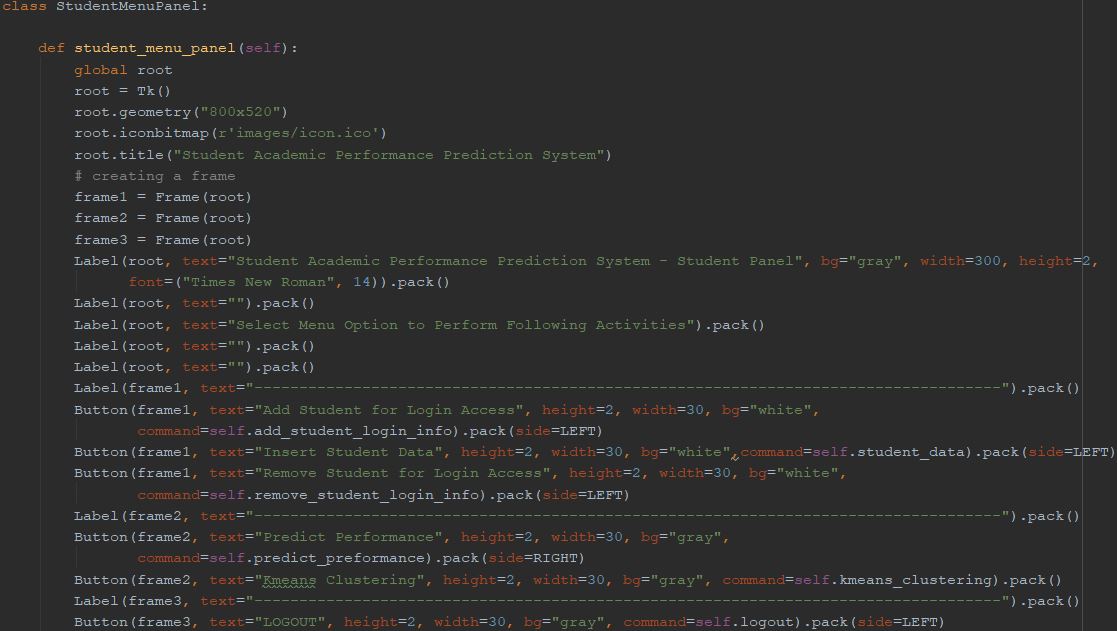
Login: Student (As a User)



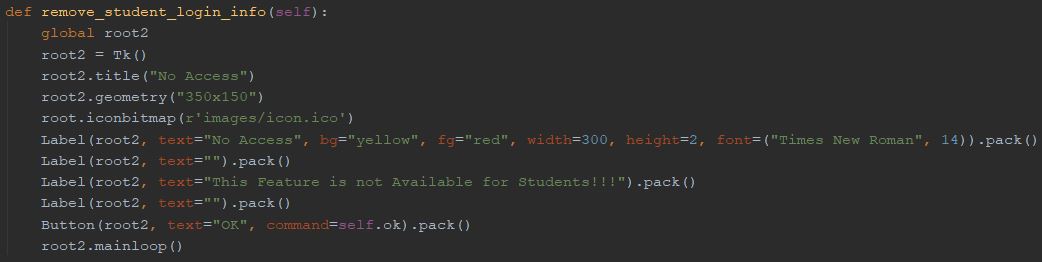
Student Dashboard



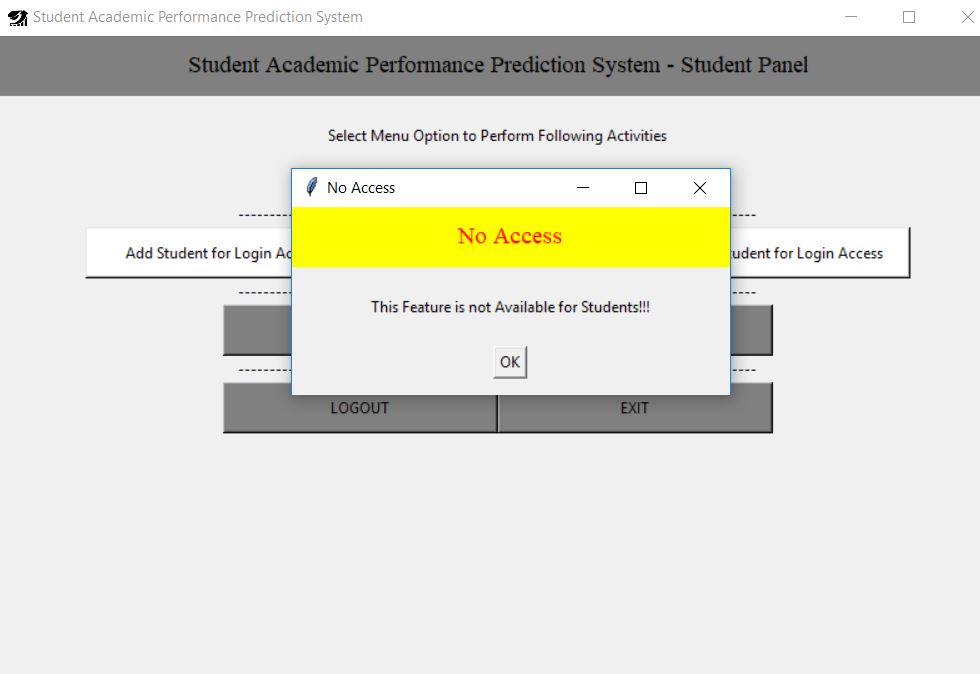
*Fig: Student Panel Dashboard*



*Fig: Student Panel*



*Fig: Removing Access*



*Fig: No Access Message*

### **6.1.5. User Manual**

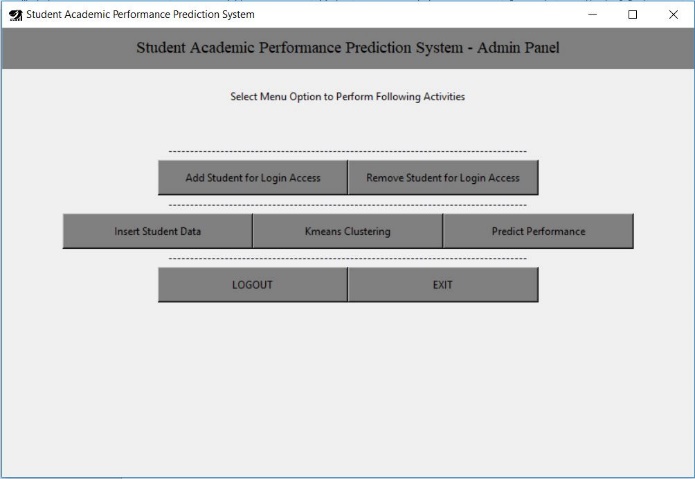
This is the general instructions to the new users of this system.

**For Admin**

Username: admin

Password: admin

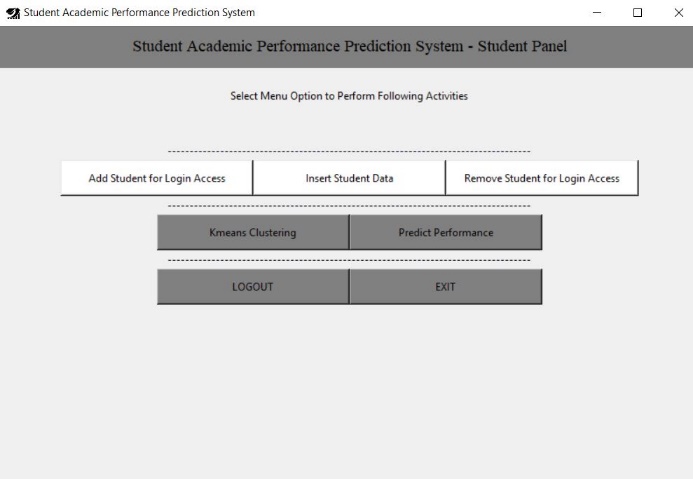
1. This is admin’s dashboard



1. Admin can add new student for login access.
2. Admin also able to remove student from their login access
3. Admin has right to insert student’s data with data alteration right.
4. Admin has also right to insert unlimited student’s academic record.
5. Admin can predict Student’s Academic Performance,
6. Admin also able to cluster dataset through kmeans clustering algorithm.

**For Student**

1. This is Student’s dashboard



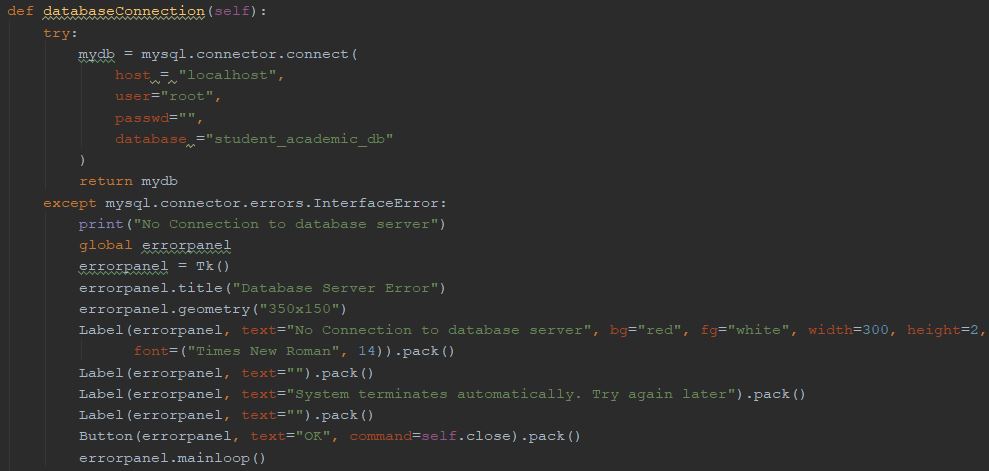
1. Student doesn’t have permission to insert their detail.
2. Student doesn’t have permission to insert their internal marks.
3. Student doesn’t have permission to make new account for login and remove that account.
4. Student has right to insert unlimited student’s academic record.
5. Student can predict Student’s Academic Performance,
6. Student also able to cluster dataset through kmeans clustering algorithm.

## **Testing**

Testing helps to find out drawbacks of the system after design and development of application system. One more thing is that, this is the scientific process to check whether system actually executes as expected or not?

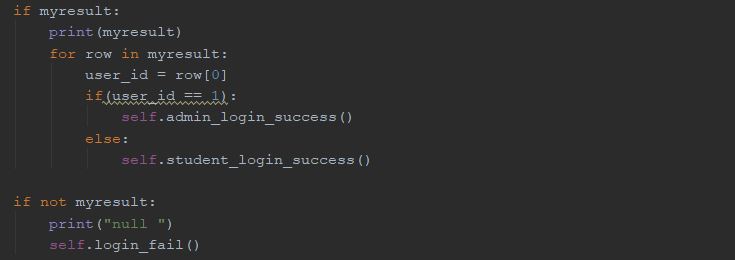
**Unit Testing**

1. Database Connecion



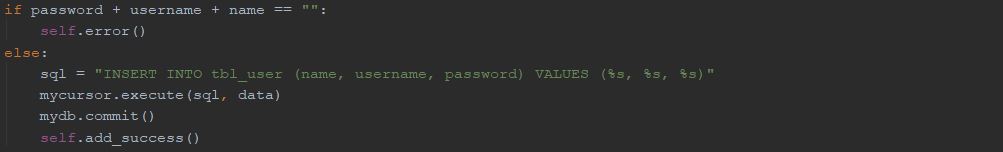
*Fig: database connection check*

1. Login Identification for user



*Fig: login condition check*

1. Data Insert Validation



*Fig: conditions for error handling*

1. Data Update

C:\Users\g_one\Desktop\unit testing\update.JPG

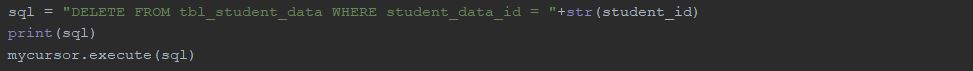
*Fig: Update Database Content Code.*

1. Data Search

C:\Users\g_one\Desktop\unit testing\search.JPG

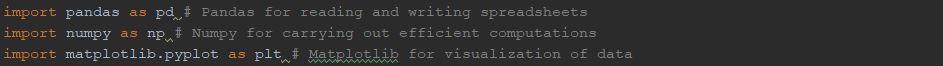
*Fig: Search database content*

1. Delete Data



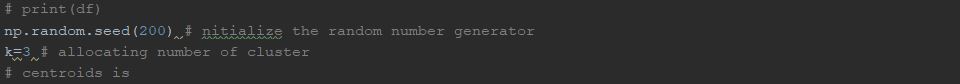
*Fig: Delete database content*

1. Imports Libraries



*Fig: Importing Python Libraries*

1. Initialize Random Number Generator



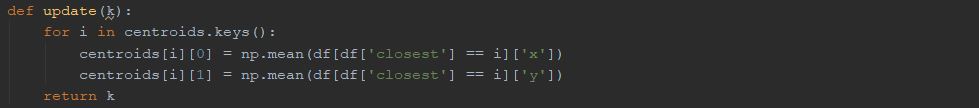
*Fig: Initialize Random Number Generator*

1. Defining Assignment



*Fig: defining Assignment*

1. Update Cluster



*Fig: Update Cluster*

1. Build Graph

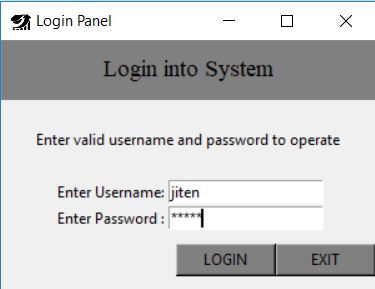
C:\Users\g_one\Desktop\unit testing\build graph.JPG

*Fig: Build Graph*

**Integration Testing**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test ID | Test Description | Expected Output | Obtained Output | Remark |
| 1 | Login with database connectivity | Must connect database while login as an admin and student. | Expected output is totally same as obtained output. | Successful |

**Output:**

****

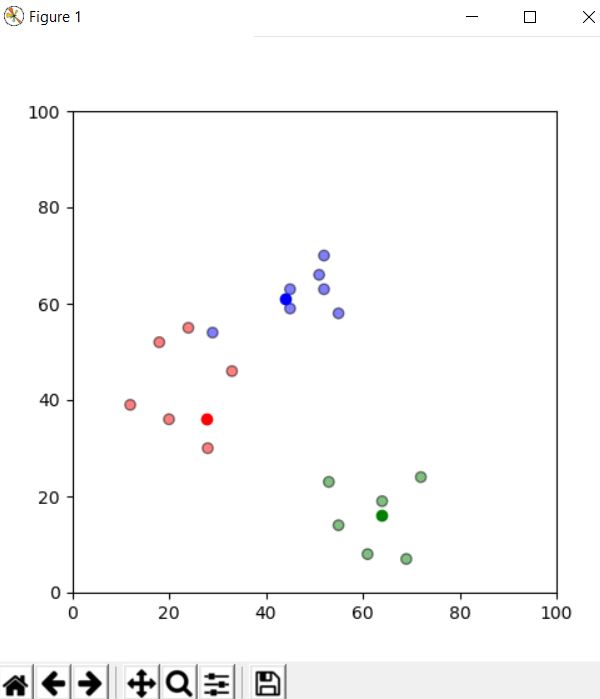
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test ID | Test Description | Expected Output | Obtained Output | Remark |
| 2 | Admin Dashboard worked as per requirement or not? | Must have workable with all buttons and functions. | Expected output is totally same as obtained output. | Successful |

**Output:**

****

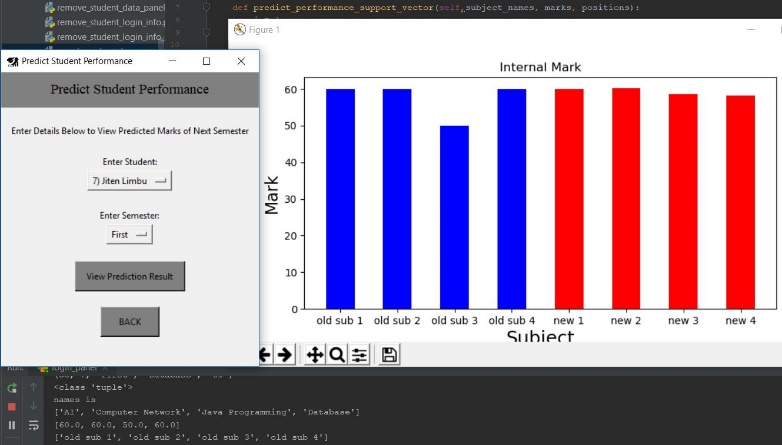
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test ID | Test Description | Expected Output | Obtained Output | Remark |
| 3 | Cluster Dataset | Cluster graph | Expected output is totally same as obtained output. | Successful |

**Output:**

**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test ID | Test Description | Expected Output | | Obtained Output | Remark |
| 4 | Predict Performance | Prediction of individual Student’s academic performance in graph | Expected output is totally same as obtained output. | | Successful |

**Output:**

****

**Acceptance Test**

Acceptance Testing is done to assure that, functional requirements of this project is working as per required or not.

|  |  |  |  |
| --- | --- | --- | --- |
| Test ID | Acceptance Requirement | Test Result | Remarks |
| FR1 | Register and Login | Pass | In database, admin has user id 1 and other are registered as a student. |
| FR2 | Admin Dashboard | Pass | Admin can insert student details, their internal marks, modify and remove detail as well as can able to cluster data and predict their performance. |
| FR3 | Student Dashboard | Pass | Different Panel for students with limited accessibility. |
| FR4 | Different Panel for students with limited accessibility. | Pass | Admin can able to create student login profile. |
| FR5 | Remove Student from Login Access. | Pass | Admin has full authority to remove student as a user login. |
| FR6 | Kmeans Clustering Features | Pass | Dataset is cluster through kmeans clustering algorithm. |
| FR7 | Performance Prediction | Pass | Student’s internal mark can be predicted. |
| FR8 | Insert, Update, Delete, Search Student Data and Marks | Pass | Features that insert student’s details with marks as well as update remove and search them. |

1. **Result**

Overall, the process of coding and their testing process is completed in very systematic form.

# **Evaluation**

The evaluation has to been done after testing those system and before launching in market. The Evaluation strategy is the process to get positive or negative feedbacks form targeted users. The main purpose of this step is to find out developed system is useful or not in present market place.

In evaluation phase, the system will be given to multiple students to get feedback from them. The collected responses and feedbacks are the supportive motivation to launch that prediction model. This evaluation will be provided to the students that is classified with specific age group, gender, colleges. And after analyzing those feedbacks, if feedback is good then system will go to the market for launch. But if feedback is bad then the application will directly go to the development phase to re-correct that errors at bug fixing.

## **Evaluators**

Evaluators age is 20-30 years old, any gender of students are selected of user evaluation. Mostly, the students of Patan College for Professional College’s students are participated for this evaluation phase. Finally the number of evaluators of this prediction model was 22 students. Among them 55% students are female and 45% students are male.

## **Result of Evaluation**

In this evaluation process, following responses are collected from survey:

**Evaluation Number 1. How is the application software?**

Here, no one evaluate their bad responses for this application. ‘Very Good’ evaluation and ‘Good’ evaluation percentage is 55% and 45% respectively. It means that, the system was perfect to lunch.

**Evaluation Number 2. Did you feel easy to use this application?**

According to this response, only 5% of evaluator felt so hard to use this application. Other 95% of evaluator has no issue.

**Evaluation Number 3. Is this application reliable for you?**

At the case of reliability, maximum 73% of evaluators said this system is reliable for them. Other 18% are confused and remaining 9% are not satisfied from this system.

**Evaluation Number 4. Do you think this application is success to perform the allocated task?**

Only 4.5% of evaluators are disagree with this question. But still 36.4% of evaluators are still facing confusion with it. 59.1% of evaluators are agree but by considering those numbers of confusion rate the system still need to be improve.

**Evaluation Number 5. Was the application helpful in predicting student academic performance?**

91% of evaluators are agree to say that this system can able to predict student academic performance.

**Evaluation Number 6. Overall, please rate this application.**

Overall rating of this system given by evaluators is satisfactory. Average rating point of this system is 4 which is too good.

**Evaluation Number 7. What extra features do you want in future in this application?**

This question was the last evaluation for the system where evaluators are free to give their feedbacks for this system. Most of the evaluator are demanded email notification features and proper validation on this system. Evaluator mentioned that the email validation and password length validation will the next correction of this system.

# **Critical Analysis**

Overall, the development phase of this system is very achievable. After contextual report there is some issue occurred in development phase. Clustering algorithm is used for this application but support vector machine algorithm is very hard to implement in this system. An issue is still occurring to implement support vector machine in this project.

# **Conclusion & Future Work**

Finally, the development phase of this project also finished in time. But by considering those issues and feedbacks of user evaluations; this is not the final product. Implantation of proper Algorithm task as well as cover nonfunctional requirement still remaining task for the development team. This is the future work for development.

# **REFERENCES**

4imprint, L., 2018. *Artificial Intelligence l 4imprint Learning Center.* [Online]   
Available at: https://info.4imprint.com/blue-paper/artificial-intelligence/2/

Abdulazeez Yusuf, D. A. L., 2016. PREDICTION OF STUDENTS’ ACADEMIC PERFORMANCE USING EDUCATIONAL DATAMINING TECHNIQUE. *Faculty of Computer Science and Information Technology, Bayero University, Kano. ,* pp. 11-12.

Abdullah Ramdhani, M. A. R. ,. A. S. A., 2014. Writing a Literature Review Research Paper: A step-by-step approach. *Internatioanl Journal of Basics and Applied Science,* 3(1), pp. 47-56.

Ahmad, F., 2015. The Prediction of Students’ Academic Performance Using Classification Data Mining Techniques. *Applied Mathematical Sciences,* 9(129), pp. 6415 - 6426 .

Anasse Bari, P. M. C. a. T. J., 2014. *Predictive Analytics For Dummies.* 111 River Street., Hoboken: John Wiley & Sons, Inc..

Baker, P. H. W. R. S., 2013. The Potentials Of Educational Data Mining For Researching Meta-cognition, Motivation And Self-Regulated Learning. *Journal Of Educational Data Mining,* 5(1).

Béla, M., 2010. *Geoinformation management 2., Project planning.* s.l.:Nyugat-magyarországi Egyetem .

Ch.Rupa, G. S. S., 2013. Data Mining: Performance Improvement In Education Sector Using Classification And Clustering Algorithm. *International Journal Of Innovative Research & Development,* 2(7), pp. 103-106.

David Hand, H. M. a. P. S., 2001. *Principles of Data Mining.* Cambridge, Massachusetts London England: Massachusetts Institute of Technology.

Dr. Abdullah AL-Malaise, D. A. M. a. M. A., 2014. STUDENTS’ PERFORMANCE PREDICTION SYSTEM USING MULTI AGENT DATA MINING TECHNIQUE. *International Journal of Data Mining & Knowledge Management Process (IJDKP),* I(2014), pp. 9-18.

Dr. Swati Sah, A. G. P. S., 2018. Evaluating Pattern Classification Techniques of Neural Network Using k-Means Clustering Algorithm. *Next-Generation Networks,* Volume 638, pp. 563-588 .

Flach, P., 2012. *MACHINE LEARNING -The Art and Science of Algorithms that Make Sense of Data.* 1 ed. The Edinburgh Building, Cambridge CB2 8RU, UK: Cambridge University Press.

Geraldine Gray, C. M. P. O., 2014. *An application of classification models to predict learner progression in tertiary education.* Gurgaon, India, IEEE.

Halterman, R. L., 2011. *LEARNING TO PROGRAM WITH PYTHON.* s.l.:Richard L. Halterman.

Harford, S. M. a. M., 2004. RESEARCH ON HUMAN LEARNING.

Harrington, P., 2012. *Machine Learning in Action.* 1 ed. 20 Balding Road , Shelter Island: Manning Publications Co..

Hetal Bhavsar, A. G., 2012. A Comparative Study of Training Algorithms for Supervised Machine Learning. *International Journal of Soft Computing and Engineering (IJSCE),* 2(4), p. 75.

Ian H. Witten, E. F., 2005. *Data Mining Practical Machine Learning Tools and Techniques.* 2nd ed. the United States of America: Diane Cerra, Morgan Kaufmann Publishers is an imprint of Elsevier..

Inf.ed.ac.uk, 2019. *Software for the data mining course.* [Online]   
Available at: http://www.inf.ed.ac.uk/teaching/courses/dme/html/software2.html  
[Accessed 25 11 2018].

Jishan, S. T., 2015. Improving accuracy of students’ final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *A Springer Open Journal,* Volume 2, pp. 1-15.

Johnson, D. R., 2011. A Quantitative Study of Teacher Perceptions of Professional Learning Communities' Context, Process, and Content. *Seton Hall University Dissertations and Theses,* p. 80.

K. M. Faraoun, A. B., 2007. Neural Networks Learning Improvement using the K-Means Clustering Algorithm to Detect Network Intrusions. *International Journal of Computer and Information Engineering,* 1(10), p. 3152.

Kefaya, Q., 2009. Mining student evaluation using associative classification and clustering. *Communications of the IBIMA,* Volume 9, p. 19437765.

KOTHARI, C., 2004. *Research Methodology methods & techniques.* 2nd ed. 4835/24, Ansari Road, Daryaganj, New Delhi - 110002: NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS.

Kraus, A. R. a. S., 2018. *Predicting Human Decision-Making From Prediction to Action.* Peter Stone ,University of Texas at Austin: A Publication in the Morgan & Claypool Publishers series.

Kumar, R., 2011. *RESEARCH METHODOLOGY a step-by-step guide for beginners.* 3rd ed. 33 Pekin Street #02-01 Far East Square Singapore 048763: SAGE Publications Asia-Pacific Pte Ltd.

Mohd Herry Mohd Nasir, Y. Y. A. T., 2019. *MASMe: Developing Mobile Accounting Software for Micro Entrepreneurs.* [Online]   
Available at: https://www.researchgate.net/figure/Waterfall-Methodology\_fig1\_326412659  
[Accessed 2019].

Mrinal Pandey, S. T., 2016. *Towards the Integration of Multiple Classifier Pertaining to The Student’s Performance Prediction.* Banasthali University, jaipur, Rajasthan, India: Elsevier.

Nikolovski, V., 2015. Educational Data Mining: Case Study for Predicting Student Dropout in Higher Education.

Pal, B. K. B. a. S., 2011. Mining Educational Data to Analyze Students‟ Performance. *(IJACSA) International Journal of Advanced Computer Science and Applications,* 2(6), pp. 63-69.

Parneet Kaur, M. S. a. G. S. J., 2015. Classification and prediction based data mining algorithms to predict slow learners in education sector. *3rd International Conference on Recent Trends in Computing,* pp. 500-508.

Perumal, T., 2014. *Research Methodology.* Malaysia: Open University Malaysia .

R, N. M. P. T., 2012. A Comprehensive Overview of Clustering Algorithms in Pattern Recognition. *IOSR Journal of Computer Engineering (IOSRJCE),* 4(6), pp. 23-30.

Romero, C. S. V. P. G. E. a. C. H., 2008. Data Mining Algorithms to Classify Students. *In Educational Data Mining,* I(12), pp. 8-17.

Sen, B., 2012. Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications ,* I( 2012 ), pp. 9468-9476.

Shahiri, A. M., 2015. A Review on Predicting Student’s Performance using Data Mining Techniques. *The Third Information Systems International Conference,* I(2015), p. 414 – 422 .

Snehal Bhogan, K. S. P. N. R. S. O. D. S. D., 2017. PREDICTING STUDENT PERFORMANCE BASED ON CLUSTERING AND CLASSIFICATION. *(IOSR) Journal of Computer Engineering (IOSR-JCE) ,* 19(3), pp. 49-52 .

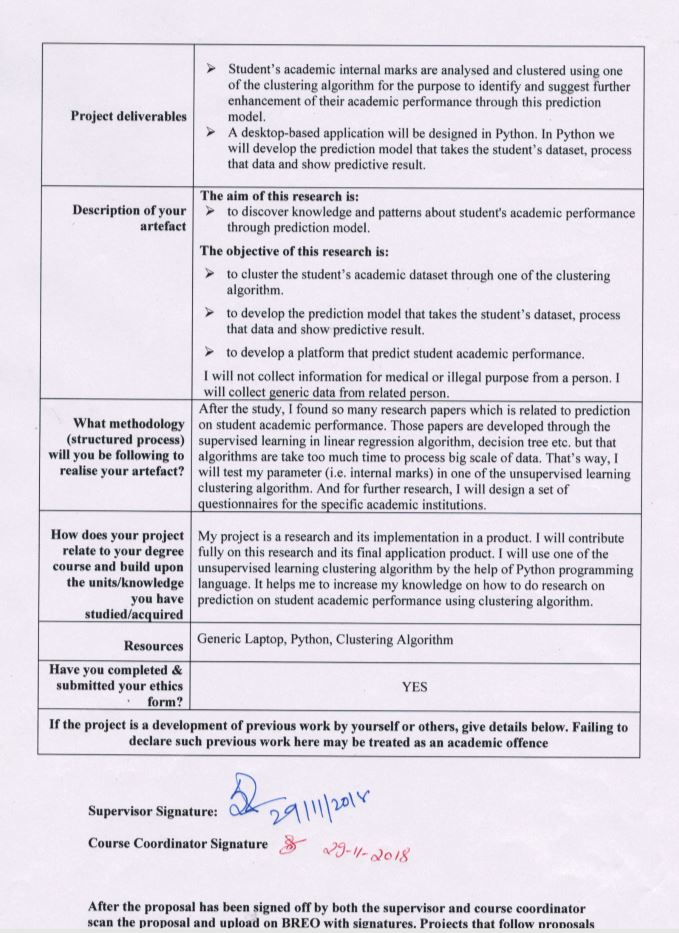
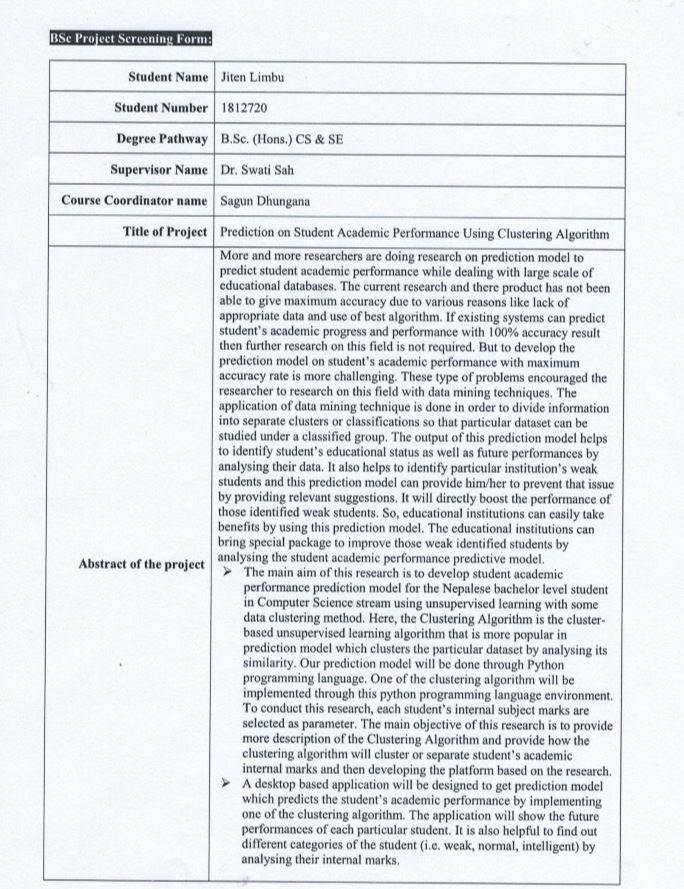
Suljić, E. O. a. M., 2012. Data Mining Approach For Predicting Student Performance. *Economic Review – Journal of Economics and Business,* X(1), pp. 4-12.

Tekin, A., 2014. Early Prediction of Students’ Grade Point Averages at Graduation: A Data Mining Approach. *Eurasian Journal of Educational Research,* I(54), pp. 207-226.

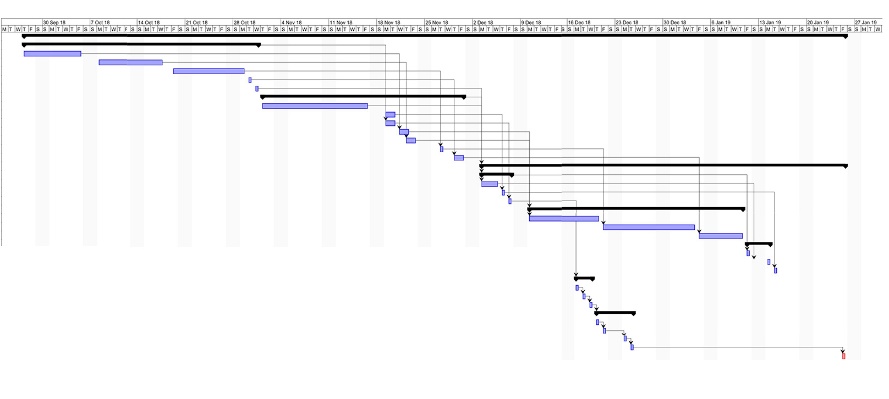
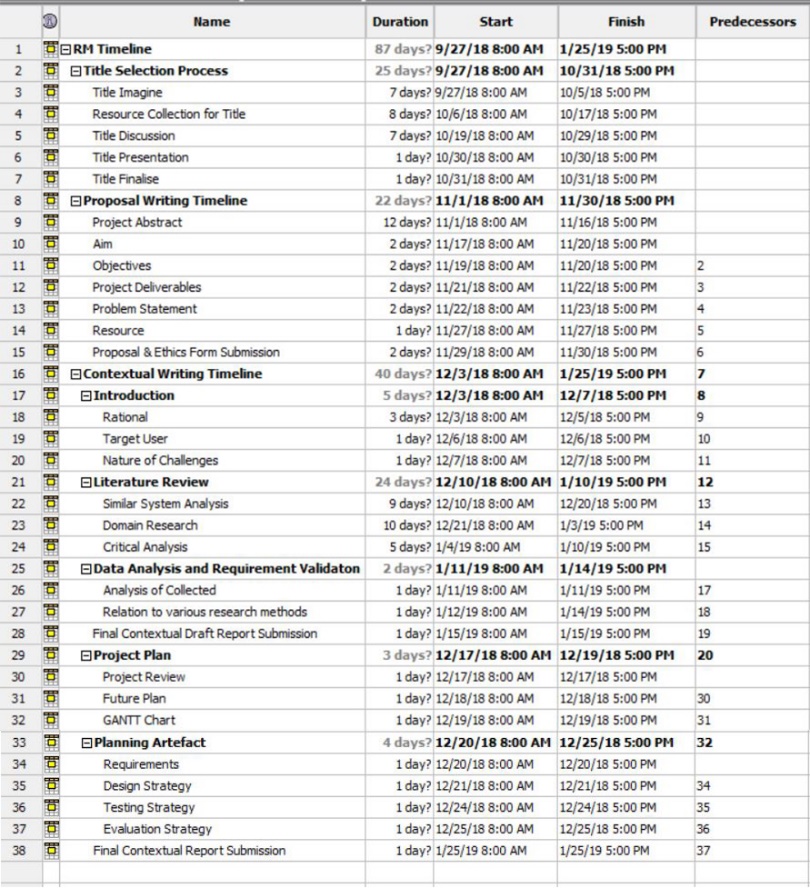
Zotorvie, J. S. T., 2017. Students’ Accommodation and Academic Performance: The Case of Ho Technical University, Ghana. *European Scientific Journal ,* 13(13), p. 1857 – 7881.

# **Appendices**

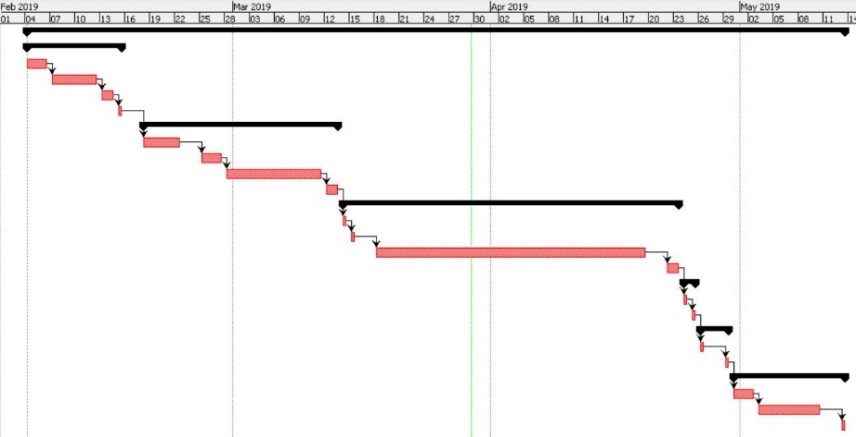
Proposal of this Project



Gantt chart (Research)



Gantt chart of this Project (Development)



Poster of this Project

