

# Prediction on Student Academic Performance Using Hybrid Clustering Algorithm

Jiten Limbu<sup>1</sup> and Dr. Swati Sah<sup>2</sup>

<sup>1</sup>Patan College for Professional Studies, Kathmandu, Nepal

<sup>2</sup>HOD, School of Computing Patan College for Professional Studies in collaboration with University of Bedfordshire (U.K) Kathmandu, Nepal

## 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 their products have 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 then further research on this field would not have been required. But to develop the prediction model on student's academic performance with maximum accuracy rate is a very challenging work. These types of problems encourage 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 his/her future performances by analyzing their data. It also helps to identify particular institution's weak students and this prediction model can provide him/her with solutions by providing relevant suggestions to avoid the future failure. 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 provide 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 which clusters the particular dataset by analyzing its similarity. Our prediction model will be done through python programming language. One of the clustering algorithms 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 algorithms. 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.

## I. INRODUCTION

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 that information to their successor so that it would help them survive. And as

human civilization began to grow, 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 21<sup>st</sup> 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 a vital role to develop each and every nation. Now, all the big inventions and developments are happening day by day because of educated peoples.

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

## 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). 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 (Rosenfeld, 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 decisions are made through machine.

In this research, we will 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.

### *Student's Academic Performance without Data Clustering Algorithm*

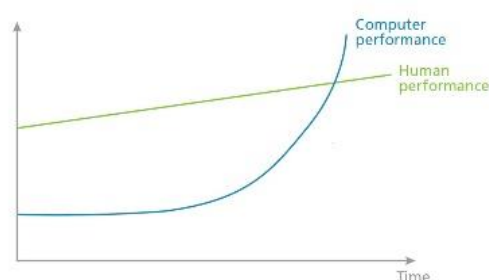
There are so many research that has been done for the aim of predicting student academic performance without using Data Mining Technique. The researchers gave their full effort to get that solution which helps to prove validity and meet required objective (Manning,1962). Those researchers didn't have any knowledge about emerging technologies. But they had clear knowledge about how to do research in traditional way.

### *Student's Academic Performance with Data Clustering Algorithm*

Predicting Student Academic Performance through AI is more advanced, more accurate and a trustable model. With the help of data mining technique, researcher can predict student's academic performance faster by clustering through different types of clustering algorithms.

### *Human Learning vs. Artificial Intelligence*

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 higher 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 outpace the Human Learning in prediction case (4imprint, 2018).

### **Problem Context**

After searching, I found so many research papers which is related to prediction on student academic performance. Those papers were 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 because 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.

### **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.

## **II. 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 is measurable by their research process (Kothari, 2014).

### **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.

### **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 play 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.

### **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.

### **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.

## **III. 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).

### **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.

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

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

#### ***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:

- a. Decision Tree Algorithm (DT)
- b. Support Vector Machine (SVM)
- c. Naïve Bayes Classification Algorithm (NBC)
- d. Neural Networks (NN)
- e. K-Nearest Neighbor (KNN)

- f. Linear Regression etc. (LR)

### *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 it's easily understood by the users and do not require maximum dataset (Romero, 2008).

### *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, 2014).

### *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 (Osmanbegovic, 2012).

### *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 (Baradwaj, 2012). Neural Network has ability to select all possible interactions between predictor variables (Geraldine, 2014). And neural network can easily predict nonlinear complex relationship between dependent and independent predictor variables (Shahiri, 2015).

### *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 the 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).

### *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).

### *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.

### *K-means Clustering Algorithm*

In K-means Clustering Algorithm, data clustering technique is used (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).

*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, 2001).

*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, data mining technique, attributes that were used while doing research and their output result.

SN	Author	Attributes/Variables used	Algorithm	Algorithm Accuracy	Data mining technique
1	(Ramaswami and Bhaskaran, 2010)	Psychometric factors	Decision tree (CHAID)	CHAID 44.69%	statistical
2	(Sembiring, 2011)	psychometric factors	Kernel k-means, smooth support vector machine (SSVM)	SSVM= 93.7%	Clustering and Classification
3	(Osmanbegovic and Sujic, 2012)	CGPA, Student Demographic, High school background, Scholarship, Social network interaction	Naive bayes, J48 Decision tree, Multi-layer Perceptron	NB=76.65% MLP=71.93% J48=73.93%	Classification
4	(Dorina Kabakchieva, 2013)	Place and profile of secondary school, final secondary education score, successful admission exam, the score achieved at that exam, and the total admission score	Decision tree, Naive Bayes, K- Nearest Neighbour, Rule learners	J48=68.59% NB=<60% KNN=60% JRip= 63%	Classification
5	(Ramesh, 2013)	Student demographic, and secondary sch. Background	Decision tree, Neural Network, Naive Bayes	DT=65% NN=72% NB=50%	Classification
6	(Tekin, 2014)	GPA	Extreme learning machine, support vector machine and neural network	EML=94.92% SVM=97.98% NN=93.76%	Classification and Clustering
7	(Gray, 2014)	Previous academic record, demographic and Psychometric factors	Decision tree, Neural Network, K-nearest network, Naive bayes, Support Vector Machine and Logistic regression	DT=65.93% NN=69.0% KNN 69.43% NB=68.03% SVM=73.33% LR=60.05%	Classification and clustering
8	(Asif, 2015)	Pre-university marks, GPA of first/second year	Naive Bayes, Neural Network, Decision tree	NB= 83.65% NN=62.50% 1-NN =74.04%	Classification
9	(Pameet, 2015)	Internal grade of student, Attendance count, Sex, Computer at home, internet access,type of secondary school	Multilayer Perception, Naive Bayes, SMO, J48 and REPTree(decision tree)	MLP=75% NB=65.13% SMO=68.42% J48=69.73% REPTree=67.76%	Classification
10	(Fadhilah, 2015)	GPA, Race, Gender, Family Income, University mode of entry	Decision tree, Naive Bayes and Rule Based	RB=71.3% NB=67.0% DT=68.8%	Classification
11	(Jishan, 2015)	CGPA, Midterm marks, Laboratory Marks, Attendance marks, Quize Marks, Final grade	Naive Bayes, Decision tree, Neural Network	NB=75% NN=75% Note (NB is faster than NN)	Classification, Optimal Equal Width Binning and Synthetic Minority Over-Sampling (SMOTE)

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

(Yusuf, 2018)

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

*Critical Analysis of Algorithms Used for Predicting Student 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.

*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 Airah 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(Yusuf,2018).The other researchers Ahmet Tekin on 'Early Prediction of Student's Grade Point Averages at Graduation: A Data Mining Approach' (Tekin, 2014), Justice Stephen TettehZotorvie 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.

#### *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 can be used to predict student's academic performance.

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

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.

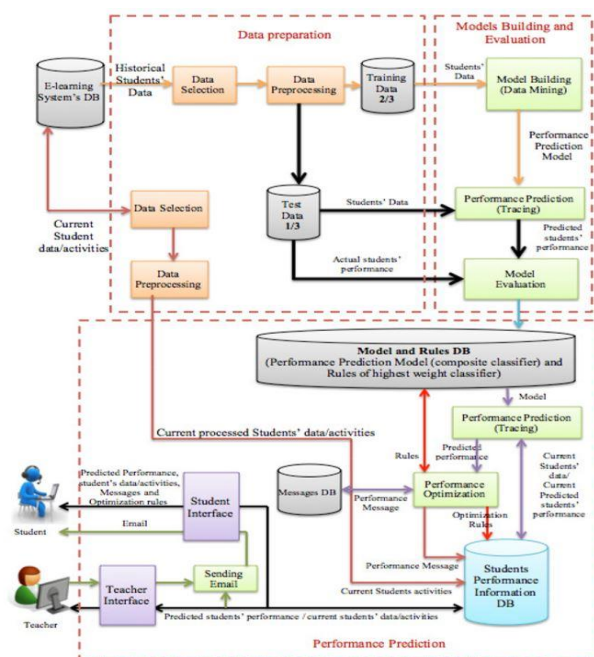
#### *Similar System 1*

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. AreejMalibari and Mona Alkhozae (Abdullah 2014).

#### *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:



(Abdullah, 2014)

Fig. 2: Conceptual System Architecture

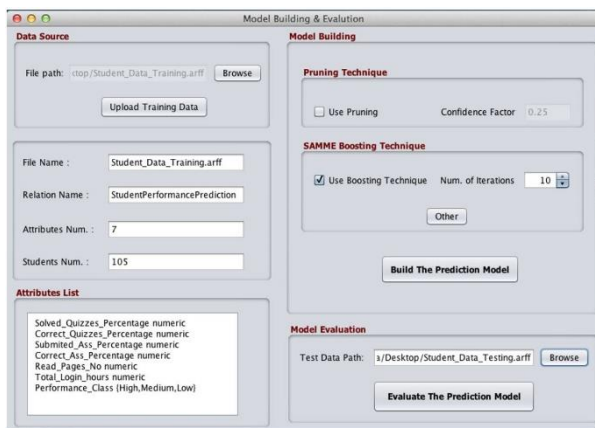
*Programming Language, Tools and Libraries*

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, WEKA (Waikato Environment for Knowledge Analysis) is also 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.

*System Implementation*

According to this system, data preparation are implemented manually for data collection. After data collection, required useful data are 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 are developed and evaluated. Finally, student's academic performance interface is built.





(Abdullah, 2014)

Fig. 3: Student' Performance Evaluation Interface

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

Name	ID	Solved Quizzes Percentage	Correct Quizzes Percentage	Submitted Ass Percentage	Correct Ass Percentage	Logins Number	Read
محمد خالد محمد العبدالله	1391010	100%	95%	100%	95%	234	12
ياسر محمد احمد العبدالله	1396006	100%	95%	80%	61%	290	14
سليمان محمد عبد العبدالله	1396184	100%	90%	100%	92%	802	15
ياسر محمد احمد العبدالله	1396208	100%	80%	100%	96%	298	17
محمد احمد محمد العبدالله	1396224	100%	75%	100%	96%	249	15
ياسر محمد احمد العبدالله	1396230	100%	100%	100%	61%	204	15
ياسر محمد احمد العبدالله	1396244	100%	95%	100%	61%	673	17
ياسر محمد احمد العبدالله	1396300	100%	95%	100%	61%	240	14
ياسر محمد احمد العبدالله	1396312	100%	75%	100%	96%	444	15
ياسر محمد احمد العبدالله	1396399	100%	95%	100%	61%	711	17
ياسر محمد احمد العبدالله	1396400	100%	70%	100%	92%	229	12
ياسر محمد احمد العبدالله	1396405	100%	100%	100%	61%	832	16
ياسر محمد احمد العبدالله	1396748	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396758	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396768	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396778	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396788	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396798	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396808	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396818	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396828	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396838	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396848	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396858	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396868	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396878	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396888	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396898	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396908	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396918	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396928	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396938	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396948	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396958	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396968	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396978	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396988	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1396998	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1397008	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1397018	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1397028	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1397038	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1397048	100%	100%	100%	96%	588	16
ياسر محمد احمد العبدالله	1397058	100%	100%	100%	96%	588	16

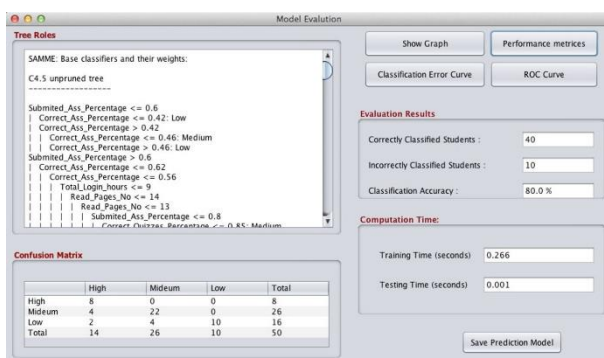
(Abdullah, 2014)

Fig. 4: Student' Performance Prediction Result

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

**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.



(Abdullah, 2014)

Fig. 5: Student' Performance Prediction Model with Accuracy Result

**Experimental Performance Comparison Result**

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

Algorithm	Correctly classified students	Incorrectly classified students	Prediction Accuracy	Model Building Time (Sec.)	Model Testing Time (Sec.)
SAMME Boosting Technique	40	10	80 %	0.266	0.001
Adaboost.M1 Boosting Technique	40	10	80 %	0.262	0.001
LogitBoost boosting technique	25	25	50 %	0.132	0.001
C4.5	37	13	74 %	0.094	0.001

(Abdullah, 2014)

Table. 2: Algorithms with their Percentage of accuracy

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

*Critical Analysis of Similar System 1*

This system has very perfect conceptual model architecture. And also have maximum number of dataset, which helps 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.

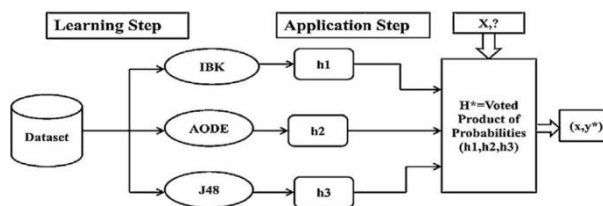
But somehow, there were still some factors which diminish their research output and system implementation. The researcher trying to implement new thing in data mining technique which is really impressive work by Dr. Abdullah AL-Malaise, Dr. AreejMalibari and Mona Alkhozae. But the tested data going through 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 causes difficulties to deal with large scale of dataset (Inf.ed.ac.uk, 2019).

*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 (Pandey, 2016).

*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.



(Pandey, 2016)

Fig. 6: Proposed Model of KNNAD.

*Tools and Libraries*

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

*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 were remained to train:

- a. Complete 1000
- b. Complete 525 and
- c. Filtered 960

*Experimental Output*

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

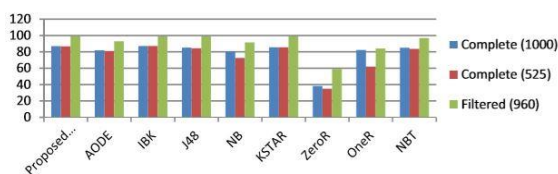
**Table 1** Performance accuracy comparisons of proposed model.

Data set	Proposed voting	AODE	IBK	J48	NB	KSTAR	ZeroR	OneR	NBT
Complete (1000)	87.03	81.85 *	87.28	85.25 *	80.63 *	85.66	38.25 *	82.41 *	85.01 *
Complete (525)	86.76	80.93 *	87.23	84.55	72.59 *	85.66	34.97 *	62.05 *	83.77 *
Filtered (960)	98.86	93.03 *	98.96	98.86	91.57 *	98.96	59.31 *	84.19 *	96.98 *

(Pandey, 2016)

Fig. 7: Accuracy Performance and Experimental Status

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



(Pandey, 2016)

Fig. 8: Accuracy Performance Comparison

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

*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 the current result.

*Domain Research*

Clustering is the process in unsupervised learning to group similar type of undefined data (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 (Kaiya, 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 (Qaddoum, 2009).

The data mining in educational sector have encouraged researchers to gather 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 technique is to estimate unexpected relationships between student's different parameters (i.e. predictor factors). Thus, the data mining techniques can extract a maximum information 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 (Kaur, 2015).

Educational Data Mining Techniques cover the predictive model which is usually used in predicting academic performance (Yusuf, 2018). 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 (Yusuf, 2018). 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 (Yusuf, 2018).

### *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 are found to be 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**

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 opinion 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.

## IV. DATA ANALYSIS AND REQUIREMENT VALIDATION

The data which come from survey were analysed and their correlation with the research were found (Johnson, 2011). Total 102 data were 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.

### Analysis of Collected Data

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

#### Survey Number 1. Gender

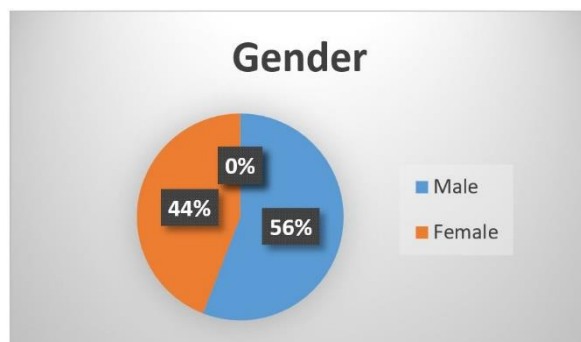


Fig. 9: Gender

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

#### Survey Number 2. Age

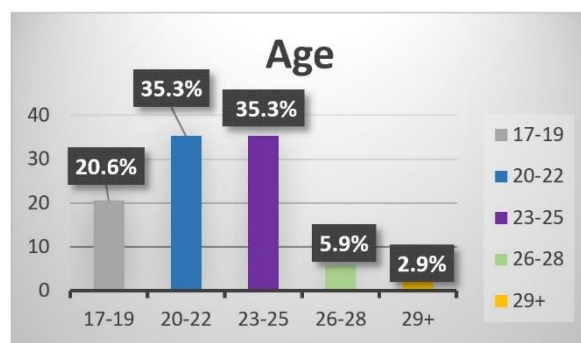


Fig. 10: Age

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

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

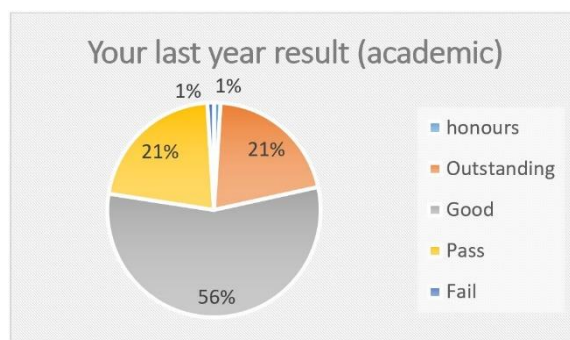


Fig. 11: Last year academic result

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 were average students according to their result output. Fail and honours ratio was equal here. It means that, 1% of student will hardly pass and other 1% students would hardly get honours degree in their academic life.

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

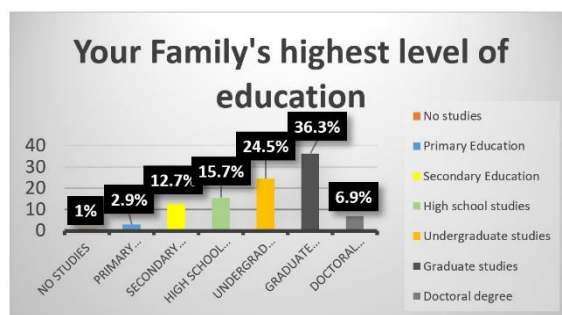


Fig. 12: Family’s Highest Level of Education

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 was also factor that directly imparts each and every student’s academic performance.

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

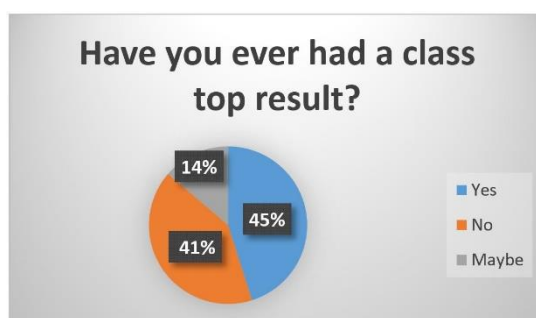


Fig. 13: Class Top Result

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.**

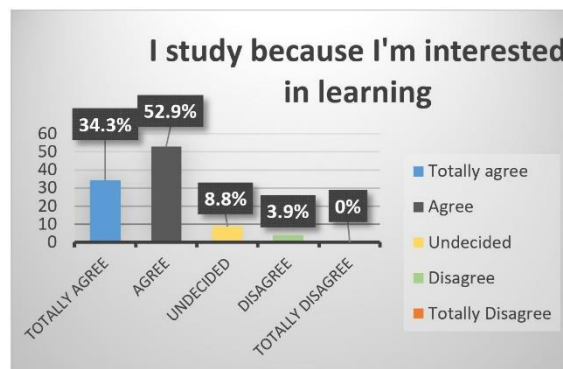


Fig. 14: Statement Opinion

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 then there is maximum chance to have good academic performance. If some students do 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**

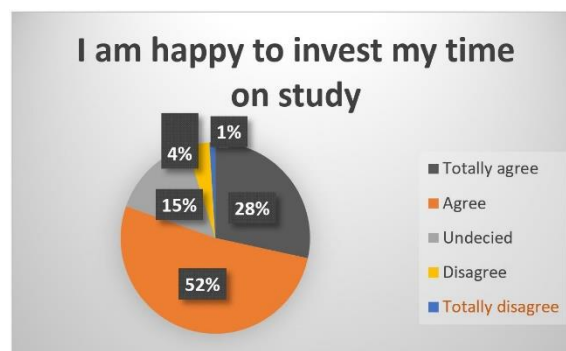


Fig. 15: Happy to invest time on study or not

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

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

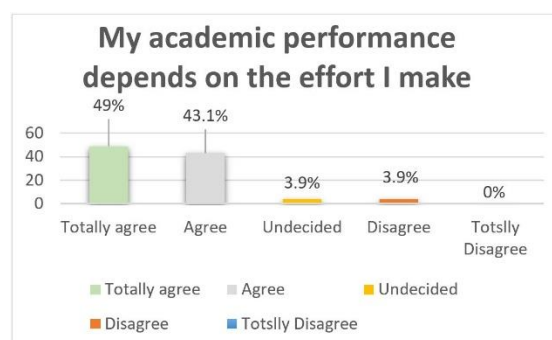


Fig. 16: academic performance depends on the effort

This bar chart shows, the effort of student brings positive result in their academic performance. The students who chose “totally agree” had his/her academic performance very good. The students who chose “agree” also had his/her academic performance very good. But, the students who chose other options had weak academic performance.



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

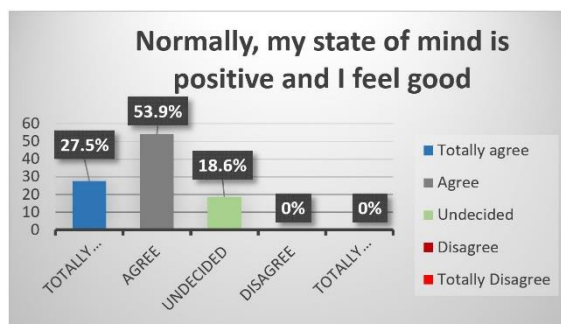


Fig. 17: Feelings

Here, this data shows that, the impact of user’s feelings such as positive and negative thoughts, etc. on their academic performance. In this data, maximum 53.9% of student had agreed with this statement and other 27.5% had totally agreed with it. It shows that normally while they are studying, their state of mind can contribute to their academic performance.

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

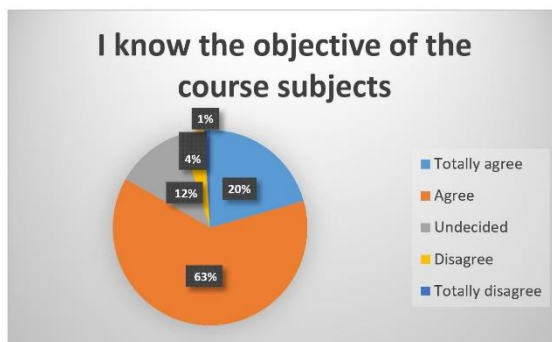


Fig. 18: The objectives of the course subjects

In this survey question, 63% of student knew the objective of the course subjects while they were studying. If students knew about their course of subject then their academic performance would get better.

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

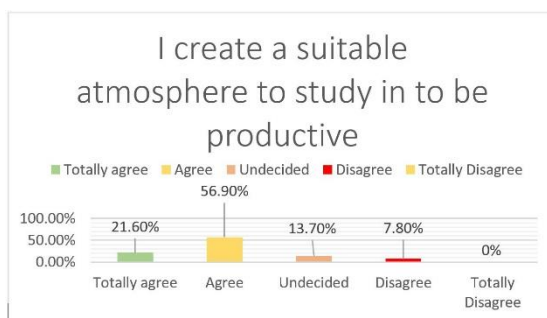


Fig. 19: Atmosphere to study in to be productive

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



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

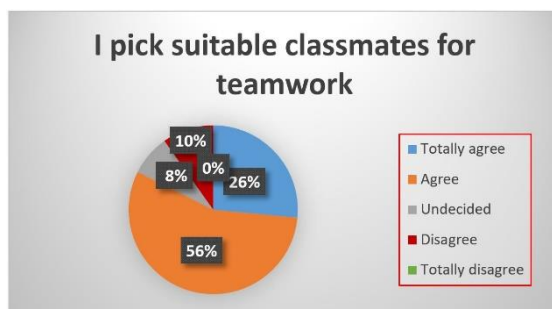


Fig. 20: Suitable Classmates for Teamwork

In this statement 56% and other 26% of students seem to be very serious about their education. So it shows that their academic performance will get 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 agree with the statement that says their course content will be useful in their future life. And other 9.8% of students are undecided with that statement and 1% of students totally disagree with this statement. Those persons who selected totally agree and agree option had good academic performance.

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

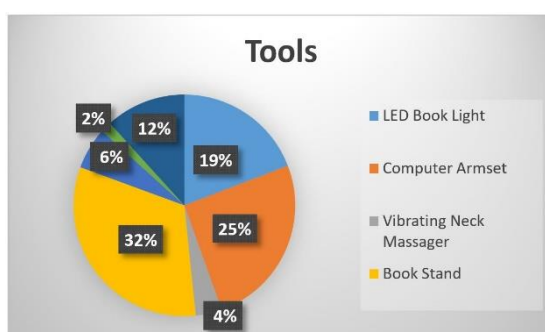


Fig. 22: Tools

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 options like free environment, etc.

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

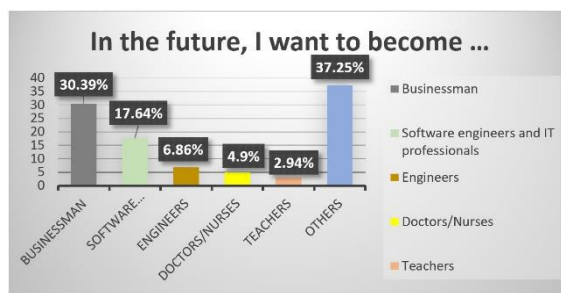


Fig. 23: Bar graph of future choice

This question possibly had more variety of answers because the participants had free choice for answer. But, their answers were divided among six categories.

**Project Plan**

Project plan is an initial work to create a product or service which is designed and developed through its own discipline (Wheelwright, 1992). Here, following proposed model is implemented in upcoming future plan in application development phase:

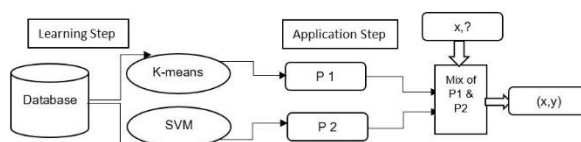


Fig 24: Final Proposed Model

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.

**Project Review**

Through this research, an application system will be designed through python programming language. It has a features to have connection between large amounts of dataset. And prediction model will be developed through Hybrid Algorithm System. The combination of K-means Algorithm and Support Vector Classification Algorithm will be used as this hybrid system.

**Future Plan**

The future plan of this project is to develop a prediction model that uses K-means Clustering Algorithm and Support Vector Machine Classification Algorithm in a Hybrid Application.

**Gantt chart for the Future Plan**

The Gantt chart is the work schedule where researcher has planned to do some specific task in limited or described timeframe (Kumar, 2011).

This is the part of project proposal and contextual report of research methodology. Where, project timeline have 87days. The title of research has been selected in 25 days, research proposal has been written in 22 days and whole contextual report is developed in 40 days timeframe.

### **Planning your artefact**

Planning will be to develop the system application in python programming language. The proposed application can be able to predict student academic performance using hybrid classification and clustering algorithm. To complete this part, proposed application will process their parameters in **K-means Clustering algorithm and Support Vector Machine Classification Algorithm**.

### **Requirements**

Requirements are the basic needs of tools and technologies that uses by upcoming application system. Following are the basic requirements for this system:

- Prediction model.
- Developed in Python Programming language.
- Clustering Algorithm i.e. K-means Clustering Algorithm.
- Classification Algorithm i.e. Support Vector Machine.
- MySQL Database.

### **Design Strategy**

To develop best solution of student academic performance after doing deep research using data mining clustering algorithm is the main target of this solution. This is necessary because predicting future performance through their past data is tougher and more challengeable. To get that particular solution regarding the student's academic performance, this research carried out prediction model which measures their academic performance in academic field. AI is used to develop this prediction model (i.e. data mining clustering algorithm). To do this research student's primary data is required for the purpose to gather student's opinions in this matter. This type of data collection process has been done through mailed questionnaire for system implementation. Those collected data are analysed with their objective of the system. Another dataset i.e. student's academic result is required for data train and test in clustering algorithm (Kothari, 2004).

Then after completion of research, the system will be developed through programming language by applying best clustering algorithm which can be easily able to cluster given dataset easily without taking too much time. To do that type of task, following designing strategy is suitable for this system:

- An application that will be designed through Python Programming Language.
- In python, by using appropriate library function of it, collected dataset (i.e. student's internal marks) will be clustered through data mining technique and clustering algorithm.
- In designing process python programming language will the student's dataset and process them through **K-means Clustering Algorithm with Support Vector Machine Algorithm** and shows predictive result which is beneficial for students and those academic institutions who really are concerned with students.
- Overall, a platform will be designed through python programming language with PyCharm IDE. The whole code of python programming language will be written through PyCharm IDE.
- Mysql Database will be implemented to store dataset in the proposed system.

### **Why Python Programming Language?**

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 (Lutz, 2013). Therefore, python programming language is best to predict student academic performance.

### **Why Clustering Algorithm?**

Clustering algorithm is the fast executable algorithm among all Unsupervised Clustering Algorithms which is simple to understand clustering algorithm. And, it is used to generate cluster data form specific parameters. Here, parameter represents the internal marks of student (Steinbach, 2000).

### **Testing Strategy**

Generally, testing strategy helps to find out drawbacks of the system. One more thing is that, this is the scientific process to check whether system actually executes as expected or not?

After design and development of application system, there are two types of testing process that will be implemented and they are blackbox testing and whitebox testing. Firstly, Black-Box Testing will be implemented and then White-Box Testing.

### **Evaluation Strategy**

The evaluation has to be done after testing those system and before launching in market. The Evaluation strategy is the process to get positive or negative feedbacks from 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.

## **V. CONCLUSION**

Here lies the important contents which I have found from my first step of the research.

### **Critical Analysis of your findings in the research**

After going through all the available information and journals of the previous researchers, it is evident that the use of unsupervised machine learning algorithm (i.e. K-means Clustering Algorithm) has one of the maximum output accuracy but the use of hybrid model gives even higher output accuracy than any other single machine learning algorithm.

### **Conclusion of Research**

So, it can be concluded that a hybrid model should be used for the prediction for the student's academic performance. This hybrid model is to be obtained by combining unsupervised machine learning algorithm (K-means clustering algorithm) and a supervised machine learning algorithm (Support Vector Machine Classification algorithm). It is shown descriptively in section 4.

This hybrid model will act as a building block for the application which will be built as the product on the next portion of my research. This application will be built by using Python programming language and with the use of MySQL database.

## **REFERENCES**

1. 4imprint, L. (2018). Artificial Intelligence | 4imprint Learning Center. [Online]. Available at: <https://info.4imprint.com/blue-paper/artificial-intelligence/2>
2. Yusuf, A. and Lawan, 2018, PREDICTION OF STUDENTS' ACADEMIC PERFORMANCE USING EDUCATIONAL DATAMINING TECHNIQUE. Faculty of Computer Science and Information Technology, Bayero University, Kano, pp. 11-12.
3. Abdulazeez, Y. and Abdulwahab, L., 2018. Application of classification models to predict students' academic performance using classifiers ensemble and synthetic minority over sampling techniques. *Bayero Journal of Pure and Applied Sciences*, 11(2), pp.142-148.
4. Ramdhani, A., Ramdhani, M.A. and Amin, A.S., 2014. Writing a Literature Review Research Paper: A step-by-step approach. *International Journal of Basic and Applied Science*, 3(1), pp.47-56.
5. Ahmad, F., Ismail, N.H. and Aziz, A.A., 2015. The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129), pp.6415-6426.

6. Bari, A., Chaouchi, M. and Jung, T., 2016. *Predictive analytics for dummies*. John Wiley & Sons.
7. Winne, P.H. and Baker, R.S., 2013. The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *JEDM| Journal of Educational Data Mining*, 5(1), pp.1-8.
8. Wheelwright, S.C. and Clark, K.B., 1992. *Creating project plans to focus product development* (pp. 70-82). Harvard Business School Pub..
9. Sree, G.S. and Rupa, C., 2013. Data Mining: Performance Improvement In Education Sector Using Classification And Clustering Algorithm. *International Journal of Innovative Research and Development (ISSN 2278-0211)*, 2(7).
10. David Hand, H. M. a. P. S., 2001. *Principles of Data Mining*. Cambridge, Massachusetts London England: Massachusetts Institute of Technology.
11. Abdullah, A.L., Malibari, A. and Alkhozae, M., 2014. STUDENTS' PERFORMANCE PREDICTION SYSTEM USING MULTI AGENT DATA MINING TECHNIQUE. *International Journal of Data Mining & Knowledge Management Process*, 4(5), p.1.
12. Sah S., Gaur A., Singh M.P. (2018) Evaluating Pattern Classification Techniques of Neural Network Using *k*-Means Clustering Algorithm. In: Lobiyal D., Mansotra V., Singh U. (eds) *Next-Generation Networks. Advances in Intelligent Systems and Computing*, vol 638. Springer, Singapore
13. Flach, P., 2012. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press.
14. Geraldine Gray, C. M. P. O., 2014. An application of classification models to predict learner progression in tertiary education. Gurgaon, India, IEEE.
15. Lutz, M., 2013. *Learning python: Powerful object-oriented programming*. "O'Reilly Media, Inc."
16. Steinbach, M., Karypis, G. and Kumar, V., 2000, August. A comparison of document clustering techniques. In *KDD workshop on text mining* (Vol. 400, No. 1, pp. 525-526).
17. Manning, W.H. and DuBois, P.H., 1962. Correlational methods in research on human learning. *Perceptual and Motor Skills*, 15(2), pp.287-321.
18. Harrington, P., 2012. *Machine Learning in Action*. 1 ed. 20 Balding Road, Shelter Island: Manning Publications Co..
19. Bhavsar, H. and Ganatra, A., 2012. A comparative study of training algorithms for supervised machine learning. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(4), pp.2231-2307.
20. Witten, I.H., Frank, E. and Hall, M.A., 2005. *Data Mining: Practical Machine Learning Tools and Techniques*.
21. 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].
22. Jishan, S.T., Rashu, R.I., Haque, N. and Rahman, R.M., 2015. Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decision Analytics*, 2(1), p.1.
23. Johnson, D.R., 2011. A quantitative study of teacher perceptions of professional learning communities' context, process, and content.
24. Faraoun, K.M. and Boukelif, A., 2006. Neural networks learning improvement using the K-means clustering algorithm to detect network intrusions. *INFOCOMP Journal of Computer Science*, 5(3), pp.28-36.
25. Qaddoum, K.S., 2009. *Mining Student Evolution Using Associative Classification and Clustering*.
26. Kothari, C.R., 2004. *Research methodology: Methods and techniques*. New Age International.

27. Rosenfeld, A. and Kraus, S., 2018. Predicting human decision-making: From prediction to action. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 12(1), pp.1-150.
28. Kumar, R., 2019. *Research methodology: A step-by-step guide for beginners*. Sage Publications Limited.
29. Abdullah, A.L., Malibari, A. and Alkhozai, M., 2014. STUDENTS' PERFORMANCE PREDICTION SYSTEM USING MULTI AGENT DATA MINING TECHNIQUE. *International Journal of Data Mining & Knowledge Management Process*, 4(5), p.1.
30. Pandey, M. and Taruna, S., 2016. Towards the integration of multiple classifier pertaining to the Student's performance prediction. *Perspectives in Science*, 8, pp.364-366.
31. Nikolovski, V., Stojanov, R., Mishkovski, I., Chorbev, I. and Madjarov, G., 2015. Educational Data Mining: Case Study for Predicting Student Dropout in Higher Education. In *Proceedings of the 12th International Conference on Informatics and Information Technologies*.
32. Baradwaj, B.K. and Pal, S., 2012. Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417*.
33. Kaur, P., Singh, M. and Josan, G.S., 2015. Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science*, 57, pp.500-508.
34. Namratha, M. and Prajwala, T.R., 2012. A comprehensive overview of clustering algorithms in pattern recognition. *IOR Journal of Computer Engineering*, 4(6).
35. Romero, C., Ventura, S., Espejo, P.G. and Hervás, C., 2008, June. Data mining algorithms to classify students. In *Educational data mining 2008*.
36. Şen, B., Uçar, E. and Delen, D., 2012. Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 39(10), pp.9468-9476.
37. Shahiri, A.M. and Husain, W., 2015. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, pp.414-422.
38. Bhogan, S., Sawant, K., Naik, P., Shaikh, R., Diukar, O. and Dessai, S., 2017 PREDICTING STUDENT PERFORMANCE BASED ON CLUSTERING AND CLASSIFICATION. *IOSR Journal of Computer Engineering (IOSR-JCE)*
39. Osmanbegovic, E. and Suljic, M., 2012. Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business*, 10(1), pp.3-12.
40. Tekin, A., 2014. Early prediction of students' grade point averages at graduation: A data mining approach. *Eurasian Journal of Educational Research*, 54, pp.207-226.
41. Zotorvie, J., 2017. Students' Accommodation and Academic Performance: The Case of Ho Technical University, Ghana. *European Scientific Journal*, 13(13), pp.290-302.
42. Fröhlich, B. and Plate, J., 2000, April. The cubic mouse: a new device for three-dimensional input. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 526-531). ACM.