



## Educational Institute Future Intake Prediction System Based on Linear SVC

G. Saminath Krisna, Dr. R. Indra Gandhi

Department of Computer Science, GKM College  
of Engineering and Technology, Chennai, Tamil Nadu, India

### ABSTRACT

All the institutions strive to find a student who is the best possible fit for their institute. They look forward to recruiting students who have the highest potential to succeed. Most of them are looking into their previous scores for making the recruiting decision. That does not always work out well for the institute, because past performance does not always prove future success. A machine learning model could solve this problem. Machine learning algorithms aim to discover hidden knowledge and patterns about student's performance. Support Vector Clustering is a relatively new learning algorithm that has the desirable characteristics like controlling the decision function, kernel method and sparsity of the solution. In this paper, we present a theoretical and empirical framework to apply the Support Vector Machines for predicting the students future performance in an educational institution. There are many factors like personality, curiosity, past academic performance, etc that are taken into account for predicting the students performance. Our results suggest that support vector clustering is a powerful tool for selecting students in the educational institution.

**Keywords:** Educational institute, student performance prediction, predictive models, predictive algorithms and training data

### I. INTRODUCTION

Artificial intelligence and machine learning algorithms are raising much interest recently and are being used in different situations of our modern life. Machine learning algorithms fed with the proper data can perform excellent predictions of the future. No algorithm can be hundred percent perfect, but they

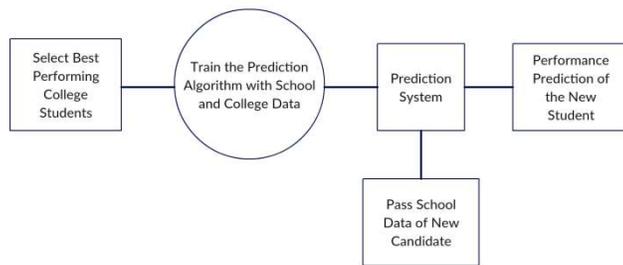
can certainly be better than humans. Machine learning algorithms are becoming better at various fields from manual to cognitive tasks. Our system provides the educational institutes with a set of comprehensive candidates for future interviews. The system analyzes different parameters of the student and also his performance in the past. Based on the parameters, it uses a classification algorithm, to find out which student could have the best future performance. The rank of the educational institute is based on the performance (both academic and extracurricular) of the current students. So they try to select the good candidates and make them great. Since just the past score doesn't indicate future performance, the institutes are often missing out the good students. Our system takes into account various factors and tries to reduce the judgemental error.

### II. REVIEW OF LITERATURE

The review of various works brings out interesting facts. There are several studies that apply data mining and analysis tools to investigate the predictability of student performance based on different criteria. The reasons for applying data mining tools are its ability to handle voluminous data and nontrivial extraction of implicit, previously unknown, and potentially useful information from the student database. The literature review points out the fact that results of prediction vary among different algorithms and different criteria. The other interesting fact is that data mining tools are found to consistently outperform other statistical approaches. The literature suggests that data mining tools are likely to predict the student performance compared to humans. The most obvious advantage of

the data mining technique is its ability to predict based on just the data, and not take into account any emotions. Different studies incorporated different criteria for predicting the student performance. However, not many studies are concerned with including the personality and aspirations of the student into consideration. Tools like Random Forest and SVM increased the efficiency of prediction from KNN and ANN. The literature review highlights that both statistical and nonstatistical measures were used to evaluate the efficiency of the data mining approaches adopted by researchers. The review of previous works showed that the earlier works undertaken are highly empirical and it is inferred that new research works in different time periods show different results. The selection of tools initially started from ANN and moved to Support Vector Machines. The SVC (Support Vector Clustering) method was not used for student performance prediction in many studies. Hence an attempt to use successful SVC and other data mining tools to evaluate the predictability of students future performance.

### III. ARCHITECTURE



- 1) Data Acquisition
- 2) Feature extraction
- 3) Predictive model
- 4) Prediction

Student performance prediction, is a new field in which machine learning is being applied. Students performance depends on a number of factors like the student's thinking, reasoning, past performance, curiosity, etc. How this model works is, first the best performing students in the college are identified, then that student's school and college details are passed on to the data mining model. The predictive model is trained based on these best performing student's data. Once the predictive model is trained, the new student's school data is passed to the model, and it predicts what the performance the student will be in college. There are four different phases in a student

performance system, namely: Data acquisition, feature extraction, predictive model and prediction.

#### A) Data Acquisition

There are two important stages required for this prediction model. The first stage is, we have to identify who are the best performing students in the college. Then we have to collect the following data from the student:

1. Personality test
2. IQ test
3. EQ test
4. Marks from 1st to 12th
5. Extracurricular activities
6. Previous Achievements
7. University GPA and CGPA
8. Participation in college events

The data can be acquired from the school and college database. Most schools and colleges in developed countries, collect this data. It is difficult to find the above mentioned data in developing countries. The data obtained is then cleared of unnecessary data in the next stage of the process.

#### 1. Personality Test

The personality of the candidate plays an important role in the prediction of future performance. This can be measured by personality tests like The Myer-Briggs Type Indicator, Disc Assessment, The Winslow Personality Profile, Process Communication Model. These are becoming standard tests in educational institutions nowadays. By comparing the personalities of the best performing students, we can get a pretty good idea of the ideal candidate. Curiosity in learning new things plays an important role in future performance. We can also find whether a student is introverted or extroverted, from these tests.

#### 2. IQ Test

The IQ test or the Intelligence quotient test is used to identify the thinking capability of the student. This includes the analytical and reasoning abilities. A good IQ score indicates that the student has the ability to think in critical situations. Some institutions rely heavily on the IQ tests to identify a good candidate. But IQ alone is not a good indicator of future performance. There are lots of other factors to consider.

### 3. EQ Test

EQ test is used to measure, how well the candidate can regulate his/her emotions. The ability to recognize, evaluate and regulate your own emotions, emotions of those around you and groups of people is referred to as emotional intelligence. The emotional intelligence of a candidate is an important factor in determining his future success, especially in this machine driven world.

### 4. Grades

Academic scores doesn't always indicate future success. Yet, it is the most commonly used criteria for selecting quality candidates. The scores indicate the determination of the student and trying his/her best to achieve their goal. It is an important factor to consider, but it is not the only option to consider. There are a lot of factors which influence the future performance of a candidate.

### 5. Extra - curricular Activities

Extracurricular activities are used as a channel to reinforce the things that they have learned from the classroom. It offers the students the opportunity to apply academic skills in a real-world context, and are thus part of a well-rounded education. The extracurricular activities increases the student's load balancing skills, self esteem, future planning and teamwork. These skills are essential for future success.

### 6. Previous Achievements

Previous achievements show that the student is capable of putting the hard work needed to achieve his/her goal. The previous achievements section includes everything from extracurricular to awards and academic achievements. This an important quality to consider, since it indicates that the student is capable of achieving his goals. But this doesn't always work, because a student performance can differ from situation to situation and how it aligns with his goals and values.

#### B) Feature Extraction

Once all the required data is obtained from the schools and colleges, the next step is to extract the required features from the data and feed it into the learning algorithm.

#### C) Predictive Model

The predictive model is trained based of the data of the best performing college students. The predictive

model is passed with the following details of the applied student:

1. Personality test
2. Psychology test
3. IQ test
4. Emotional intelligence test
5. Marks from 1st to 12th
6. Extracurricular activities
7. Previous Achievements

The model analyzes all the parameters of the new student and compares it with the best performing students in college. It can then predict how the student will perform in college. The predictive algorithm used in the proposed system is linear SVC. Different algorithms can be used to find the patterns and correlations.

#### 1. Support Vector Machines

Support Vector Machine is a supervised machine learning model with associated learning algorithms that are used for analyzing and predicting probabilities, using the given data. It used to perform classification and regression analysis on the given structured and unstructured data. SVM is a non-probabilistic linear classifier. Given a labeled set of training examples, each marked as belonging to one or the other. The SVM training algorithm builds a model that assigns new examples to one category or the other. It is a representation of the examples as points in space, mapped so that the data points are separated clearly. Additional new data plots are then incorporated into that same space. Based on where the point falls on, it is added to specific categories. Support Vector Machines can also perform non linear classification using the kernel trick.

#### 2. Support Vector Clustering

Clustering is to partition a data set into different groups according to some criterion in an attempt to organize data into a more meaningful form. There are many ways of achieving this form. Clustering may proceed according to some parametric model or by grouping data points according to some distance or similarity measure as in hierarchical clustering. It usually adds cluster boundaries within regions of the data space where there is insufficient data in the probability distribution area. This is the path taken in support vector clustering, which is based on the support vector approach. In SVC data points are mapped from data space to a high dimensional feature

space using the kernel function. In feature space we look for the smallest sphere that encloses the image of the data points using the Support vector domain description algorithm (DDA). This sphere, when mapped back into the data space, will form a set of contours which can enclose the data points. We interpret these contours as cluster boundaries, and points enclosed by each contour are associated by support vector clustering to the same cluster.

#### D) Prediction

Based on the best performing college student's school and college data, the prediction algorithm finds the correlation with the new student and predicts his/her future performance.

The prediction is based on a confidence rating which shows how confident are the algorithm is with its prediction. Usually predictions with the highest confidence rating are displayed to the user, but it can be configured and customized.

#### CONCLUSION

This paper proposes a SVM-based student performance prediction system using which, the educational institutions can understand the potential of the student. This reduces the partiality, emotions and other factors from the human judgement. The system takes into account a good feature subset, which contains features that are highly correlated with the output, yet uncorrelated with each other. The selected features are evaluated carefully and prioritized. The feature selection and feature evaluation are filtered by correlation-based SVM. It reduces dimension and noise of student data as well as provides analyzed set of candidates for the institutions to make their decision. In the proposed system, the setting of parameters have a critical impact on the performance of the resulting system. We need to investigate to develop a structured method of selecting an optimal value for the parameters in the proposed prediction system for the best results.

#### REFERENCES

1. M. M. A. Tair, A. M. El-halees, "Mining Educational Data to Improve Student's Performance: A Case Study", *Int. J. Inf. Commun. Technol. Res.* , vol. 2, no. 2, pp. 140-146, 2012.
2. P. Kaur, M. Singh, G. Singh, "Classification and prediction based data mining algorithms to predict slow learners in education sector", *Procedia - Procedia Comput.Sci.* , vol. 57, pp. 500-508, 2015
3. S. Shakuntala Devi, B. Raju, V. Rama, SURVEY ON CLASSIFICATION , vol. 5, no. 3, pp. 169-172, 2016.
4. Ogunde. Ajibade, "A Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm Ogunde A. O I. and Ajibade D. A1", *Comput. Sci. Inf Technol.* , vol. 2, no. 1, pp. 21-46, 2014.
5. F. Ahmad, N. H. Ismail, A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques", *Appl. Math. Sci.* , vol. 9, no. 129, pp. 6415-6426, 2015.
6. H. Hamsa, S. Indiradevi, J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm", *Procedia Technology.* , vol. 25, pp. 326-332, 2016
7. M. Tomlinson, "Graduate employability: A review of conceptual and empirical themes", *Higher Education Policy*, vol. 25, no. 4, pp. 407-431, 2012.
8. J. Straub, S. Kerlin, T. Stokke, "How we're changing computer science education and how you can help", *Presented at Proceedings of the 2014 Midwest Instructional Computing Symposium*, 2014