

# Student's Placement Prediction Using Support Vector Machine

Yashodeep Ingale<sup>1</sup>, Tanuja Bedse<sup>1</sup>, Shivani Khairnar<sup>1</sup>, Dhyanshawari Ghute<sup>1</sup>

<sup>1</sup>Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra, India

## ABSTRACT

### Article Info

Volume 6, Issue 5

Page Number: 55-60

Publication Issue :

September-October-2020

### Article History

Accepted : 01 Sep 2020

Published : 12 Sep 2020

The ultimate goal of any educational institution is offering the best education experience and good placement opportunity to the students. Identifying the students who need extra support and taking the appropriate actions to enhance their performance plays an important role in achieving good placement. Student's academic achievements and their placement in campus selection becomes as challenging issue in the educational system. Proposed student prediction system is most vital approach which may be used to differentiate the student data/information on the basis of the student performance. The proposed system will classify the student data with ease and will be helpful to many educational organizations. There are lots of machine learning algorithms and statistical base technique which may be taken as good assets for classify the student data set in the education field. In this paper, various machine learning algorithms like Naïve Bayes, SVM, KNN, decision tree algorithm has been applied to predict student performance which will help to identify performance of the students and also provides an opportunity to improve to performance.

**Keywords:** Machine learning, SVM, Placement Prediction, KNN, Data Analysis.

## I. INTRODUCTION

The student performance prediction is a very important issue for improving the educational process. The students' performance level may be affected by many factors related to father's job, gender, and their average marks in the previous years. The early prediction of student performance may help in improving the educational process. The performance prediction can be obtained by employing data mining techniques on educational data sets. The data classification is the most important technique in data mining research. It depends on

categorization (giving a class) of data based on the values of the predicting attributes . Classifiers are represented by different kinds of models. And the variation of algorithms is great in some times for inducing classifiers from data. Some popular classification algorithms are K -Nearest Neighbors classifier, Naïve Bayes, Decision Tree Algorithms, and Support Vector Machine (SVM).

The proposed system makes the prediction of placement of students easy. We are developing a system in which the students will register/login into the system and enter their bio-data and skillets, according to students academic details the system will identify whether the student is eligible for the

placement and recommend the courses to the students. Admin creates the courses and registers students to the respective courses. Admin can view the courses and the students along with their attributes. Admin predicts the placement status of the current students. If the student is eligible for placement mail will be sent to the student from admin and students names will be displayed on the dashboard in their colleges.

We are here using SVM Algorithm as our work goes in Type of Classification. As we have data set of students as their academic score and their placement company types or sectors. According to that our model will be trained. After model training when next time we will give academic information of new student system will give predict that in which possible class these data can be fit according to trained model and same class companies will be given as suggestion or possible placement prediction of student. For that we need such data of students with their academic score of every year from SSC to last year and Company he placed in and Domain/Designation he got. On basis of that our model will be trained and next output will be given.

## II. BASIC CONCEPT

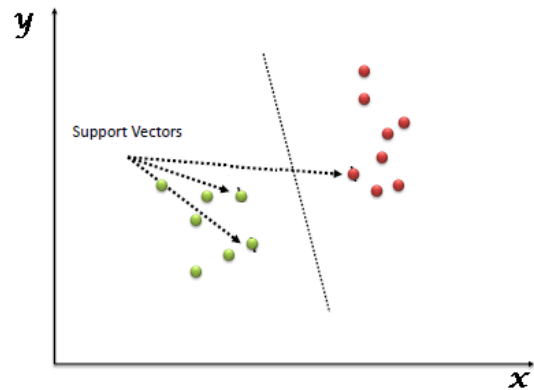
### 2.1 Machine Learning :

Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn and improve from explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

### 2.2 Support Vector Machine (SVM) :

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of

features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well).



### 2.3 Naive Bayes:

In machine learning, Naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. It is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. One of the main assumptions of the Naïve Bayes algorithm is that each feature is independent, which holds good for the problem considered, since the score of the student in each subject is independent, though it could be related with similar subjects. Due to this assumption, this classifier is very effective for this problem.

Naive Bayes Classifiers rely on the Bayes' Theorem, which is based on conditional probability or in simple terms, the likelihood that an event (A) will happen *given that* another event (B) has already happened. Essentially, the theorem allows a hypothesis to be updated each time new evidence is introduced. The equation below expresses Bayes' Theorem in the language of probability:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Let’s explain what each of these terms means.

- “P” is the symbol to denote probability.
- P(A | B) = The probability of event A (hypothesis) occurring given that B (evidence) has occurred.
- P(B | A) = The probability of the event B (evidence) occurring given that A (hypothesis) has occurred.
- P(A) = The probability of event B (hypothesis) occurring.
- P(B) = The probability of event A (evidence) occurring.

**2.4 KNN Algorithm:**

KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. To evaluate any technique we generally look at 3 important aspects:

1. Ease to interpret output
2. Calculation time
3. Predictive Power

Let us take a few examples to place KNN in the scale :

KNN algorithm fairs across all parameters of considerations. It is commonly used for its easy of interpretation and low calculation time.

	Logistic Regression	CART	Random Forest	KNN
1. Ease to interpret output	2	3	1	3
2. Calculation time	3	2	1	3
3. Predictive Power	2	2	3	2

**III. LITERATURE SURVEY**

M. F. Lee, N.F. Mat Nawi& C.S. Lai [1] proposes a methodology for job performance of an individual which can be predicted through personal highlighted. Capable of predicting job performance is benefiting from all perspectives. Job performance in this study refer to the abilities of students in completing the

task given by lecturer. Through this study, student’s job performance through Adversity Quotient (AQ) and career interest were determined. A total of 370 Bachelor’s students from eight faculties in University Tun Hussein Onn Malaysia

(UTHM) has been selected as respondents by the sampling strata. The research was a survey study and the instruments were a questionnaire of AQ Related Personality Traits (ARP) Stoltz, Inventory Self-Directed Search Form Easy (SDSS) Holland and Job Performance Questionnaire. Data were analyzed using sum, mean, frequency, percentage and Multiple Linear Regression test. The findings show that the majority of respondents are in high level of AQ, the dominant of career interest of students is the Social personality. The Multiple Linear Regression analysis shows job performance is predictable through the variables Adversity Quotient (AQ) and two types of personality career interest (Social and Entrepreneur), that is Job Performance = 35.21 +7.28 (S) + 3.98 (E) + 0.11 (AQ). This model is expected to be applied to Bachelor’s students of UTHM to predict future job performance based on scores of AQ, Social, and Entrepreneur.

S. Indhu Priya, Dr. P. Devaki [2] Presents that in an education system, predicting student’s performance in placement has become more challenging due to the large volume of data and imprecise data with fuzziness in educational databases. Large volume of data is processed using big data analytics methods. Processing of data with different factors and with different parameters is difficult in traditional system, where big data analytics can help organizations to better understand the information contained within the data. It also helps them to identify the data that is most important for the prediction and future decision making. The aim of evaluating student’s performance is to help them to develop individual student’s professionalism, to encourage self-improvement, to maintain

achievements and also to give them prior idea about their level of skills in placements. It also plays a vital role in increasing placements. In this paper some of the existing methodologies and their drawback for the student analysis have been discussed.

Ashok M.V, Professor Apoorva A [3] Placement of students is one of the very important activities in educational institutions. Admission and reputation of institutions mainly depends on placements. Hence all institutions strive to strengthen placement department. In this study, the objective is to analyze previous year's student's historical data and predict placement chance of the current students and the percentage placement chance of the institution. A model is proposed along with an algorithm to predict the placement chance of students. Data pertaining to the study were collected from the same institution for which the placement chance prediction and percentage placement need to be found from 2006 to 2015. Data collected is divided into historic data from 2006 to 2015 and test data i.e, 2016; 2016 data is considered as current data. Suitable data pre-processing methods are applied. Students having better chance of placement are characterized as good if not bad. This proposed model is compared with other classification algorithms such as Naive bayes, Decision tree, and Neural network with respect to accuracy, precision and recall. From the results obtained it is found that the proposed algorithm predicts better in comparison with other algorithms.

Animesh Giri, M Vignesh V Bhagavath, Bysani Pruthvi , Naini Dubey, [4] Presents a Placement Prediction System which predicts the probability of a undergrad student getting placed in an IT company by applying the machine learning model of k-nearest neighbors's classification. We also compare the results of the same against the results obtained from other models like Logistic Regression and SVM. To do so we consider the academic history of the student as well as their skill set like, programming skills,

communication skills, analytical skills and team work, which are tested by the hiring companies during the recruitment process. The data that is used for this purpose is the Placement Statistics of PES Institute of Technology, Bangalore South Campus for the previous two academic batches.

Pushpa S K, Manjunath T N, Mrunal T V, Amartya Singh, C Suhas [5] . More than 2.5 quintillion bytes of data is being generated across the globe. In fact this data is as much as 90% of the data in the world today, and has been created in the last two years alone. Big data describes the large volume of data that inundates the business on a day to day basis. Huge amount of data is being generated by everything around us at all times and is produced by every digital process and social media exchange through systems, sensors, mobile devices, etc. Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. The author's aim is to develop a Class Result

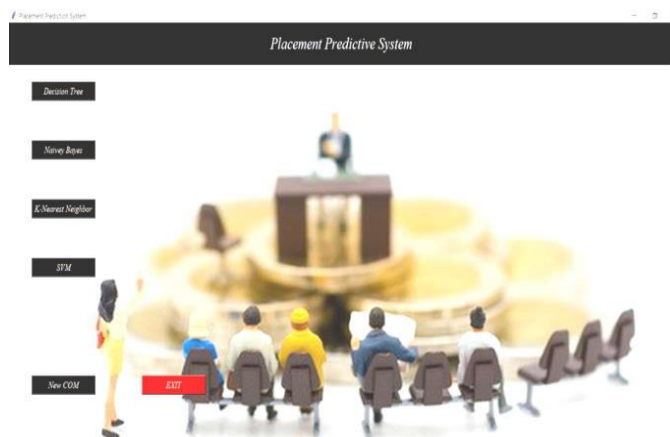
Prediction using Machine Learning. To extract meaningful value from big data one needs optimal processing power analytics capabilities and skills. Using the concept of machine learning, a number of algorithms are explored in order to predict the result of class students. Based on the performance of students in previous semester, whether the student passes or fails the current semester is computed before the final examinations actually takes place.

Apoorva Rao , Deeksha K C , Vishal Prajwal R, Vrushak K, Nandini M S [6] The Institutions today face a challenge of placements and to improve the same. It is a very complicated process to predict the placement of students manually. Educational institutions today strive to improvise the procedures and strategies that support decision making capabilities that improvise the students' placements. This can be addressed by making use of machine learning technique to predict the placement of the

students. We make use of the historical data of the past students, this data is considered as the training data set and is used to train the model. The system then predicts the placement status of the student to one of the five categories or statuses, viz., Dream Company, Core Company, Mass Recruiters, Not Eligible and Not Interested in Placements. This model helps the placement cell of the organization to identify the weaker students and provide extra care towards them so that they improve their performance henceforth. Furthermore, the students in the final as well as prefinal years of B. E or B. Tech course can also make use of this system to know their placement status that they are likely to achieve. By this knowledge they can put in the necessary efforts to achieve their goals and to get placed in better companies.

#### IV. RESULT AND DISCUSSION

Students Placement system Predict the probability of undergrad students getting placed in the company by applying a machine learning algorithm like SVM, Knn, Naïve Bayes, and which machine learning algorithm gives more accuracy for prediction. Admin has access to add the upcoming placement drives in the system and send emails to students.



main web page

Results between SVM classifier and KNN classifier:

Classifier	KNN	SVM
Precision	0.69	0.83
Recall	0.42	0.78
F1 Score	0.44	0.79
Support	20	20
Accuracy	45%	80%

As observed from above comparison, the SVM gives the higher accuracy in placement prediction than KNN classifier.

#### V. CONCLUSION

This paper focuses on non intellectual parameters of students which affect their study and academic growth. Utilization of data mining in the field of education can prove to be a boon for the society. Psychometric analyses of students' behavior in respect of learning help in enhancing their academic performance and placement activity. Various mining techniques like decision tree, KNN, naïve bayes and SVM have been applied to the educational data covering psychological. Our proposed model uses Support vector machine, KNN, Naïve Bayes for classification of data and predict learners' placement opportunities

#### VI. REFERENCES

- [1] Ranjan and R. Ranjan, "Modelling Key parameters in Higher Education using Logistic Regression: an Indian case based Data Analysis", 4th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), pp. 365-369, 2015, IEEE, 2015.
- [2] H. Goker, and H. I. Bulbul, "Improving an early warning system to prediction of student

- examination achievement”, In Machine Learning and Applications (ICMLA), 13th International Conference on pp. 568-573, IEEE, December 2014.
- [3] P. Kaur, M. Singh, and G. S. Josan, “Classification and prediction based data mining algorithms to predict slow learners in education sector”, *Procedia Computer Science*, 57, pp. 500- 508, 2015.
- [4] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, “Student academic performance prediction model using decision tree and fuzzy genetic algorithm”, *Procedia Technology*, 25, pp. 326- 332, 2016.
- [5] A. M. Shahiri, and W. Husain, “A review on predicting student's performance using data mining techniques”, *Procedia Computer Science*, 72, pp. 414-422, 2015.
- [6] G. Gray, C. McGuinness and P. Owende, “An application of classification models to predict learner progression in tertiary education, in: Advance Computing Conference (IACC)”, *IEEE International*, pp. 549–554, 2014.
- [7] T. Mishra, D. Kumar and S. Gupta, “Mining students’ data for prediction performance”, *Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, ACCT '14*, IEEE Computer Society, Washington, DC, USA, 2014, pp. 255–262. doi:10.1109/ACCT.2014.105.
- [8] S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana and E. Wani, “Prediction of student academic performance by an application of data mining techniques”, *International Conference on Management and Artificial Intelligence IPEDR*, Vol. 6, pp. 110–114, 2011.
- [9] M. Richardson, C. Abraham, and R. Bond, “Psychological correlates of university students' academic performance: A systematic review and meta-analysis”, *Psychological bulletin*, 138(2), pp. 353, 2012.
- [10] I. Burman, S. Som and S. A. Hossain, “Meta Analysis of Psychometric Measures and Prediction of Student’s Learning Behaviour using Regression Analysis and SVM”, *Jour of Adv Research in Dynamical and Control systems*, Vol. 10, 02- Special issue, pp. 291-298, 2018.

**Cite this article as :**

Yashodeep Ingale, Tanuja Bedse, Shivani Khairnar, Dhyaneshawari Ghute, "Student's Placement Prediction Using Support Vector Machine ", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 6 Issue 5, pp. 55-60, September-October 2020. Available at doi : <https://doi.org/10.32628/CSEIT206511> Journal URL : <http://ijsrcseit.com/CSEIT206511>