EVALUATING SUPERVISED MACHINE ALGORITHMS FOR STUDENT PLACEMENT PREDICTION

Navyashree S L[1], Gnanavi N[1], Deeksha D[1], Krupa N[1], Dr. Meenakshi Malhotra[2]

[1]Student, Department of Computer Science Engineering, Jyothy Institute of Technology, Bangalore, India.
[2]Associate Professor, Department of Computer Science Engineering, Jyothy Institute of Technology, Bangalore, India.

---***---

Abstract - *One of the criteria required by technical institutions to excel is better campus placements results. To achieve the same, the institutions need to prioritize as well as train their students to acquire the skills to perform better in interviews. Hence there is a need for student placement prediction and recommender system that can enable the placement authorities to predict the placement. With the advancement of machine learning algorithms, the placement prediction system can guide the institutions as well as the students to excel. Previously various machine learning algorithms have been implemented to predict student placement such as decision trees, k-nearest neighbors, support vector machines and many more. The objective of this paper is to compare the algorithms implemented in the past for student placement prediction. The comparison results provide guidance to choose an efficient algorithm to build the model of the system based on the training data.*

Key Words: **Placement prediction, Machine learning, Decision tree, Support vector machine, Supervised learning.**

1. INTRODUCTION

 The first and foremost aim of every student is to have the best job once they finish the degree. At present, during admission process a student will enquire about the college placement percentage. Thus, it becomes important for college authorities to improve their placements and provide better job opportunities to the students. Institutions provide a defined structure and process for the training and placement. The college provides seminars, guest lectures, hands-on workshops, soft-skill session and conferences for students to prepare them for placements. Additionally, on-campus and off-campus recruitment activities are also planned. The placement results for colleges can be improved using the student placement prediction followed by the recommendations to place them in better jobs and companies. Machine learning algorithms have provided a mechanism to predict the placement of a student in accordance to the current skill before a student is placed.

Machine learning is defined as "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E [1]". Machine learning algorithms have been existing from past few decades but have gained immense popularity in the present due to the accessibility of more resources, tools and techniques. Furthermore, the availability of cheap computational power and huge amount of data has resulted in efficient machine learning models. As a result, the learned models are used to automate various tasks in wide-ranging domains. The components required to do prediction as shown in the flow diagram given in figure 1.
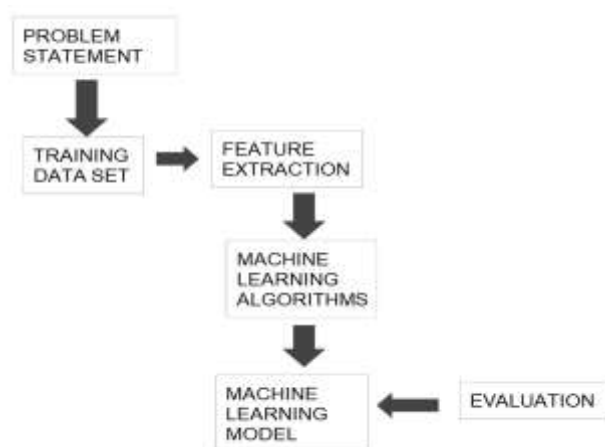

Fig -1: Structure for creating Machine learning models.

Machine Learning is a natural outgrowth of the intersection of Computer Science and Statistics. Machine learning predictions are used to solve various real-world problems. Some of them are listed below [1]:

- Automatic Recognition of Handwritten Postal Codes

- Online shopping recommendations

- Speech recognition

- Stock market prediction

- Spam email detection

- Traffic predictions can be done using GPS Navigation Services.

- Medical diagnosis

- Credit risk analysis

- Customer profiling

- Fraud detection

This paper refers to the prediction of placement of the current final year students based on the current skill-set and recommend the skills that can get a student placed in better companies. The placement result of previous year's data is taken as training data to train the machine learning algorithm. Thus, the alumni also indirectly contribute to the better employability of the current batch students. As a result of this process, the institutions can help prepare students for better placements that will achieve greater recognition for the institution in return.

The structure of the paper is as follows: Section II includes discussion regarding the existing placement prediction system. Section III provides evaluation on the comparison various machine learning methods used for placement predictions. Then the conclusions are given in the last section, section IV.

2. TECHNOLOGY USED

Different types of machine learning algorithms are: Supervised learning, Unsupervised learning, Reinforcement learning.

Supervised learning algorithms are predictive and has labeled data try to model relationships and dependencies between the target prediction output and the input features such that we can predict the output values for new data based on those relationships which it learned from the previous data sets [1].

Unsupervised learning - The computer is trained with unlabeled data. They are the family of machine learning algorithms which are mainly used in pattern detection and descriptive modeling. However, there are no output categories or labels here based on which the algorithm can try to model relationships. These algorithms try to use techniques on the input data to mine for rules, detect patterns, and summarize and group the data points which help in deriving meaningful insights and describe the data better to the users [1].

Reinforcement learning algorithm (called the agent) continuously learns from the environment in an iterative fashion. In the process, the agent learns from its experiences of the environment until it explores the full range of possible states and it allows machines and software agents to automatically determine the ideal behavior within a specific context, in order to maximize its performance. Simple reward-feedback is required for the agent to learn its behavior; this is known as the reinforcement signal [1].

Decision Trees, the main objectives are [5]:

- to classify correctly as much of the training sample as possible;

- generalize beyond the training sample so that unseen samples could be classified with as high of an accuracy as possible;

- be easy to update as more training sample becomes available;

- and have as simple structure as possible;

K-NN: K-nn algorithm is basically used for classification and regression problems. It is also called as non-parametric lazy learning algorithm. Its main purpose is to predict the classification of a new sample using database where data points are separated into several classes. Non-parametric means that it does not make assumptions on the data distribution which means that the model

structure is determined from data. Therefore k-nn is one of the first choices for classification problems when there is no prior knowledge about data distribution.

Support Vector Machine (SVM): It is also called Support Vector Networks. Basically, it has its roots in Statistical Learning Theory (SLT) and optimization method. It helps to solve the problems in machine learning by converting machine learning problems to optimization problem. SVM algorithms are also used to solve linear programming, quadratic programming, semi definite programming [9].

The Applications of SVM include text classification, image classification and used to recognize the handwritten characters.

3. LITERATURE SURVEY

Bayesian Belief Network (BBN) machine learning techniques are used to predict the student placement using their competence and skills. Slim, Heileman, Kozlick and Abdullah took the dataset of 400 from different branches. The implementation of BBN obtained as marginal error of 0.16[7].

C4.5 that is a Successor of ID3, Classification and Regression Tree (CART) and CHAID. The above-mentioned decision tree algorithms work with supervised data and provide a decision tree for the classification problem of placement prediction. The prediction involved training the various algorithms with dataset that included data for 1342 students. The ID3 algorithm provided the maximum accuracy of 95.33 % [2].

Decision tree learning, Sci-Kit learn and Logistic Regression are the techniques that has been used. Decision tree works with the supervised data and uses logistic regression to find the probability of the student getting placed belonging to different departments. Thangavel, Bkaratki and Sankar have used decision tree approach to train the model dataset of size of 2205 tuples as training data and test data of size 289. The implementation provided accuracy of 71.66% with tested real life data [3].

Data mining and Classification Association Rules (CARM) are the two techniques that have been used by Badr, Algobail, Hanadi and Manal. Data mining is used in higher education to understand the student learning activities. Whereas CARM is used to create classifier using the rules called Classification Association Rules (CARs). Thus, instead of using dataset to train the model the authors used the courses and divided each course into four rules. The accuracy of Data mining and CARM is 67.33% [4].

Guleria and M. Sood implemented prediction system using Data mining and Naive Bayes Algorithm [6]. Data mining is used to analyze educational data and extract useful information from large amount of data. Naive Bayes Algorithm is preferred when the dimensionality of the input is high and requires small amount of training data. The dataset of students were divided into tables for easy computation [6].

Nguyen implemented techniques that are used are Collaborative filtering and Matrix Factorization. Collaborative filtering is used on assumption that similar users like similar things. Matrix Factorization is used as it is one of the most successful methods for predicting. This was implementing using dataset of size 1,416,473 entries [8].

4. EVALUATION

Table 1 gives the evaluation with pros and cons for different Placement prediction systems.

Table -1: Evaluation of different Placement prediction systems

SL.NO	TECHNIQUE	PRONS	CONS
1	C4.5 that is a Successor of ID3, Classification and Regression Tree (CART) and CHAID [2]	Using different Decision tree algorithms. It helps in comparing with other algorithms and helps in getting better results	Aptitude skills and technical skills are present in data set. Prediction of placement of final year student are not done by applying classification algorithms.
2	Decision tree learning, Scikit-learn and Logistic-Regression [3]	This system helps in improving placement and in turn builds reputation of the institution.	Applying clustering technique for getting clusters of student-data is difficult.

3	Data mining and Classification Association Rules [4]	This helps in reducing dropout rates by predicting their performance in programming courses	Data classified into courses, applying algorithms for these courses is difficult.
4	Data mining and Naive Bayes Algorithm [6]	It helps in developing visualization techniques for generating results.	Choosing the target attribute is difficult. Posterior probability must be calculated properly for best results.
5	Bayesian Belief Network machine learning technique [7]	Curriculum graph is used to predict the placement of student	It does not contain multiple variable in BBN model in addition course variable
6	Collaborative filtering and Matrix Factorization [8]	The models work output is measured in binary thus easier to measure error and reduce it.	Dealing with cold-start problem when using matrix factorization is tough.

5. CONCLUSION

The evaluation results depict that Bayesian Belief Network and SVM are the best algorithm that can be used for prediction of student's placement in engineering college. The results will also help to identify those students who need attention and guidance to increase their chances of placement. The generated knowledge will be useful for management of placement process to develop policies and strategies for better planning and implementation of educational program like improving soft skill and infrastructure to increase the placed student rate in college.

REFERENCES

[1] Tom Mitchell, Machine Learning, McGraw Hill, 1997.

[2] Jeevalatha, T & N.Ananthi, N.Ananthi & Kumar, Saravana. (2014). "Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms". International Journal of Computer Applications. 108. 27-31. 10.5120/18988-0436.

[3] S.K.Thangavel, P.D.Bkaratki and A. Sankar, "Student placement analyzer: A recommendation system using machine learning," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 2017,1-5.doi:10.1109/ICACCS.2017.8014632.

[4] Ghada Badr, Afnan Algobail, Hanadi Almutairi, Manal Almutery,"Predicting Students Performance in University Courses: A Case Study and Tool in KSU Mathematics Department", Procedia Computer Science, Volume 82, 2016, Pages 80-89, ISSN 1877-0509.

[5] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, May-June 1991.doi: 10.1109/21.97458.

[6] P. Guleria and M. Sood, "Predicting student placements using Bayesian classification," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 109-112.doi: 10.1109/ICIIP.2015.7414749.

[7] A. Slim, G. L. Heileman, J. Kozlick and C. T. Abdallah, "Predicting student success based on prior performance," 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Orlando, FL, 2014, pp. 410-415. doi: 10.1109/CIDM.2014.7008697.

[8] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, Lars Schmidt-Thieme,"Recommender system for predicting student performance", Procedia Computer Science, Volume 1, Issue 2,2010,Pages2811-2819,ISSN 1877-0509.

[9] Tian, Yingjie & Shi, Yong & Liu, Xiaohui (2012).Recent advances on support vector machines research.Technological and Economic Development of   Economy.18.10.3846/20294913.2012.661205.