

# Predicting Students' Career by using Machine Learning Algorithms

Akanksha Pandey<sup>1</sup>, Dr. L S Maurya<sup>2</sup>

<sup>1</sup>Student of M.Tech (Computer Science & Engineering) Professor (Department of Computer Science & Engineering)

<sup>1,2</sup>Shri Ram Murti Smarak College of Engineering and Technology, 13 km, Bareilly-Nainital Road, Ram Murti Puram, Bareilly-243202, (U.P)-INDIA

*akanksha08cs@gmail.com*

**Received on:** 11 June ,2022

**Revised on:** 07 August ,2022,

**Published on:** 09 August,2022

**Abstract-** Now a days it's very important to choose a suitable career because the students' whole life depends on it and still there are so many students who don't know which path, they should choose not only to be successful but also to be happy in life. As these students are still confused what they should do and what not. Which leads to the failure because of the wrong career choices which they make in confusion. In this paper, we are predicting the career for an engineering student can select after the completion of graduation using machine learning classification techniques. To help students to make decision we will describe the machine learning techniques and we have also given algorithmic methods and machine learning algorithms are presented here. This system will help them to select the best career they should go for and brighten their future.

**Keyword:** - Artificial Intelligence, Heat map, Machine Learning, Prediction, Student Career

## I. INTRODUCTION

Engineering is pursued by so many students in our country after the completion of their schooling. As engineering opens up, there are so many lines for students to do after completing graduation and it is also considered to be the most favoured undergraduate degree. But the problem arises for the students to choose the correct path after the graduation. And opting the correct career has become so complex for students as there are multiple career choices and increasing job competition. A large part of our life is spent in accomplishing our career goals, so it is very important to make sure that right steps

were taken and correct planning was done in the early years of your life. There are very few lucky ones who are born with a clear mind and who knows what they want to do and by considering the learning style of students, their motivation and interest, concentration level, family background, personality type, information processing ability and the way they attempt the exams (Halde 2016).

This proposed work deals with the career prediction of the students as to whether they will be going for their next level of higher education from their present graduation level using machine learning concepts like DT (Decision Tree) and RF (Random Forest). Applying the concept of DT it yields a result of about 91% of accuracy and applying RF it gives 93% of accuracy level. The result of the proposed system helps the recruiters to select the only needed and proper candidates. (N. VidyaShreeram et al. 2021).

In their paper analyze show to introduce machine learning algorithms in to the process of direct volume rendering. A conceptual framework for the optical property function elicitation process is proposed and particularized for the use of attribute value classifiers (Cerquides et al. 2005).

Performed a study on performance analysis of classification algorithms for activity recognition using Micro-Doppler Feature (Lin, Yier, and Julien Le Kerne 2017).

Performed a study based on multi-label classification with weighted classifier selection and stacked ensemble (Xia, Yue long, Ke Chen, and YunYang, 2020).

In their study aims to identify the key trends among different types of supervised machine learning algorithms, and their performance and usage for disease risk prediction (Uddin et al. 2019).

In their study performed news articles classification using Random Forests and Weighted Multimodal Features (Liparas et al. 2014).

In this paper, authors analyzed and performed computation times of different classification algorithms on many datasets using parallel profiling and computing techniques. Performance analysis was based on many factors, such as the unique nature of the dataset, the size, and type of the class, the diversity of the data in the data set, and so on (Upadhyay, Navin Mani, and Ravi Shankar Singh 2018).

In this paper, we have proposed a few supervised machine learning classifiers which may be used to predict the placement of a student in the IT industry based on their academic performance in class Tenth, Twelve, Graduation, and Backlog till date in Graduation. (L.S Maurya, et al 2021).

**II -RESEARCH DESIGN AND METHODOLOGY**

1. Data Collection: We used the Google form to collect data. The form consist of ten questions. Form which the eight questions where for input and one question was for output and other for the user name. The total data collect by form was 330. They were:

- I. Students Name
- II. percentage in class 10?
- III. Percentage in class 12/Diploma?
- IV. Marks obtained in B.Tech.
- V. Aptitude skill
- VI. Communication skill
- VII. Technical skill
- VIII. Management Skill
- IX. General Knowledge
- X. Choose a career option after graduation

2. Data Processing: To get the structured data we preform the data processing by filtering the required data from the data collected.

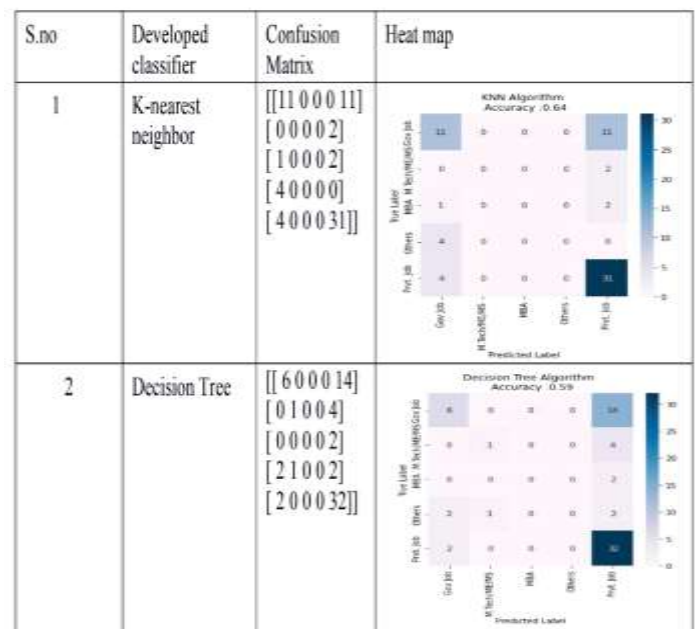
3. Feature diminution: we removed the column which were not required. So the number o column remaining are 9.

S. No.	Classification Algorithms	Hyper parameter	Random State	Execution Time	Accuracy score
1	K-nearest neighbor	n_neighbors = 18	937	0.00524	63.64%
2	Decision tree	max_depth = 5, random_state=200	317	0.00489	59.09%

Confusion Matrix and Heat map:

The confusion matrix shows us the performance of the algorithm. And heat map provide us the color presentation of the confusion matrix.

Table 2 - It contains the developed classifier and its generated confusion matrix.



**Classification Report: It measures the quality of prediction for the classifiers.**

**Table 3-** It shows the algorithm and classification report which is generated by the developed classifier.

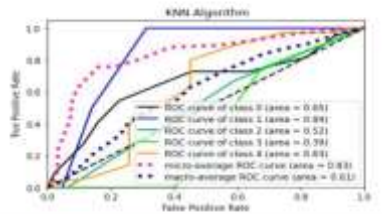
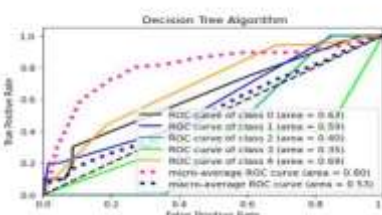
S.no	Developed Classifier	Classification report
1	K-nearest Neighbor	precision r ecall f1-scoe support Gov Job 0.55 0.50 0.52 22 M.Tech/ME/MS 0.00 0.00 0.00 2

		MBA 0.00 0.00 0.00 3 Others 0.00 0.00 0.00 4 Prvt. Job 0.67 0.89 0.77 35 Accuracy 0.64 66 weighted avg 0.54 0.64 0.58 66
2	Decision tree	pr ecision r ecall f1-scor e support Gov Job 0.60 0.30 0.40 20 M.Tech/ME/MS 0.50 0.20 0.29 5 MBA 0.00 0.00 0.00 2 Others 0.00 0.00 0.00 5 Prvt. Job 0.59 0.94 0.73 34 accuracy 0.59 66

Table 4. Mean Square Error, R2 value, Log Loss

S. No.	Algorithm	Mean Squared Error	R2 Value	Log Loss
1	K - Nearest Neighbor	4.64	-0.38	2.07
2	Decision Tree	4.91	-0.51	3.57

Table 5- It consist of developed classifier and its ROC Curve.

S.no	Developed Classifier	ROC Curve
1	K nearest neighbor	
2	Decision Tree	

### III- CONCLUSION

This system will be helpful for each and every student and also the teachers of the institutes to guide the students in correct path and pursue the best career. It will also helpful in decreasing the rate of failure. As students can easily analysis themselves and be prepared if needed. As we compare the accuracy score, execution time KNN shows the better accuracy shown in table 1. Then find the mean square value, log loss and R2 value to check the performance of both the algorithm, we found that KNN performance is good as its error is less in comparison to decision tree.

We can clearly see from result that KNN (K-nearest neighbor) classifier give us good result then Decision tree. So we can go use the KNN classifier to predict students' career

### REFERENCES

- [1] Xia, Y., K. Chen, and Y. Yang. 2020. Multi-label classification with weighted classifier selection and stacked ensemble. *Information Sciences*. doi:10.1016/j.ins.2020.06.017.
- [2] Swarupa, R. A., and S. Jyothi. "Performance analysis of classification algorithms under different datasets." In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp.1584–1589. New Delhi, India: IEEE, 2016.
- [3] Halde, R. R. "Application of machine learning algorithms for betterment in education system." In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICADOT)*, pp.1110–14. Bangalore, India: IEEE, 2016.
- [4] Halde, R. R. "Application of machine learning algorithms for betterment in education system." In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICADOT)*, pp.1110–14. Bangalore, India: IEEE, 2016.
- [5] Uddin, S., A. Khan, M. Hossain, and M. Ali Moni. 2019. Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making* 19 (no. 1):1–16. doi:10.1186/s12911-019-1004-8.
- [6] Kabra, R. R., and R. S. Bichkar. 2011. Performance prediction of engineering students using decision trees. *International Journal of Computer Applications* 36 (no. 11):8–12.
- [7] Thangavel, S. K., P. DivyaBkaratki, and S. Abijitk. "Student placement analyzer: A recommendation system using machine learning." In *2017 4th International Conference on Advanced Computing and Communication*

*Systems (ICACCS)*, pp.1–5. Coimbatore, India: IEEE, 2017.

- [8] Agarwal, K., M. Ekansh, R. Chandrima, P. Manjusha, and S. Siddharth. "Analyzing student performance in engineering placement using data mining." In *Proceedings of International Conference on Computational Intelligence and Data Engineering*, pp.171–81. Springer, Singapore, 2019.
- [9] Cerquides, J., M. López-Sánchez, S. Ontañón, E. Puertas, A. Puig, O. Pujol, and D. Tost. "Classification algorithms for biomedical volume datasets." In *Conference of the Spanish Association for Artificial Intelligence*, pp.143–52. Springer, Berlin, Heidelberg, 2005.
- [10] Thangavel, S. K., P. DivyaBkaratki, and S. Abijitk. "Student placement analyzer: A recommendation system using machine learning." In *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp.1–5. Coimbatore, India: IEEE, 2017.
- [11] Halde, R. R., A. Deshpande, and A. Mahajan. "Psychology assisted prediction of academic performance using machine learning." In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pp.431–35. Bangalore, India: IEEE, 2016.
- [12] Upadhyay, N. M., and R. S. Singh. "Performance evaluation of classification algorithm in weka using parallel performance profiling and computing technique." In *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp.522–27. Solan, India: IEEE, 2018.
- [13] Brodley, C., and P. Smyth. 1997. *Applying classification algorithms in practice*. *Statistics and Computing* 7.
- [14] Cerquides, J., M. López-Sánchez, S. Ontañón, E. Puertas, A. Puig, O. Pujol, and D. Tost. "Classification algorithms for biomedical volume datasets." In *Conference of the Spanish Association for Artificial Intelligence*, pp.143–52. Springer, Berlin, Heidelberg, 2005.
- [15] Kumar, D., Z. Satish, R. D. S., and A. S. 2019. Predicting student's campus placement probability using binary logistic regression. *International Journal of Innovative Technology and Exploring Engineering* 8 (no. 9):2633–35.