Predicting Students' Career by using Machine Learning Algorithms

Akanksha Pandey[1] Dr. L S Maurya[2]

[1]*Student Of M.Tech (Computer Science & Engineering)*
Shri Ram Murti Smarak College of Engineering and Technology,
13 km, Bareilly-Nainital Road, Ram Murti Puram,Bareilly-243202,(U.P)-INDIA

[2]*Principal (Shri Ram Murti Smarak College of Engineering and Technology& Research)*
Shri Ram Murti Smarak College of Engineering and Technology,
Bareilly-Nainital Road, Ram Murti Puram,Bareilly-243202,(U.P)-INDIA

[1]*akanksha08cs@gmail.com ,*[2]*lsmaurya@gmail.com*

Abstract *– Now a day 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 open ups, the so many lines to students to do after completing graduation and it is also considered to be the most favored 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 is 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 where they see themselves in life ahead. But majority of us are not sure what we want from life and so it in very important to plan out things. Thus career planning is what gives your career and in some way your life, true meaning and purpose.(accessed 2022). Selecting a career path can help one set professional goals and develop strategies for ensuring a secure future. The most essential element of choosing an appropriate career path involves making an honest self-evaluation of ones talents, abilities and interests. While elements of ones path may vary over time, due to interests or circumstances, having a professional

International Journal of Innovations in Engineering and Science,   www.ijies.net

objective will act as a guide to make critical decisions with greater clarity. Students, these days, have options of multiple careers in multiple fields and are forced to choose their careers as soon as high school is over.(Rucha Hemant Rangnekar.et.al,2018).

In this research work the machine learning is applied to find the career of the each student. This predication plays an important role for institution. As selecting a right path can help students to set there professional goals and also develop the strategies to secure their future. And to make a right choice the students need to honest to themselves and must have self-evaluation of their ability , interest and talent. But for student choices may vary as the interest and circumstances changes over the time period. Students are having option multiple careers in multiple fields and are forced to choose their careers. Before, the one have degree would help and individual get a good and high profile job. However, in current scenario the transparency in recruitment has increased.

The challenges and competition are pert of life in this world, so to achieve one goal they should be prepared and plan thing well. But there are students who are totally clueless about the path the should choice after graduation. And after that they feels disgruntled and dissatisfied with there choice. Having a proper plan for your career is basic start after the graduation and at that level its really important to have the clear vision about the career to pursue that will help them. Section II addresses literature review. Section III represents proposed work. Section IV discusses research design and methodology. Section V describes the results and discussions of our research work. Section VI shows the conclusions of the entire research. Finally, Section VII lists the references.

II - RELATED WORK

Give an introduction to classification algorithms and the metrics that are used to quantify and visualize their performance. They first briefly explain what we mean with a classification algorithm, and, as an example, they describe in more detail the naive Bayesian classification algorithm. Using the concept of a confusion matrix, they next define the various performance metrics that can be derived from it (Korst et al. 2019).

Different classifiers algorithms namely Naive Bayes, Multi layer perceptron, Instance-based K-Nearest Neighbor (IBK), J48 Decision Tree, Simple Cart, Zero R, CV Parameter, and Filtered Classifier performance was analyzed. The diabetes datasets, nutrition datasets, E. coli protein datasets, mushrooms datasets were used for calculating the performance by using the cross validations of parameter. Finally identified classification algorithms performance was evaluated and compared in terms of the classification accuracy and execution time under different data sets (Swarupa and Jyothi 2016).

In this paper, the main aim was to describe the various ways in which the machine learning is used in educational institutes and how institutes can get prediction of students' performance, and the important features that are needed to be considered while making prediction for different things. In addition to this, the study also compares the prediction given by different machine learning algorithms. The paper concludes that the prediction of the students' performance can be made more precise and accurate 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 predication of the students as weather 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 Kernec 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 differ- ent classification algorithms on many datasets using parallel profiling and computing

International Journal of Innovations in Engineering and Science, www.ijies.net

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 super-vised 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).

III -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:

    *   Students Name

    *   percentage in class 10?

    *   Percentage in class 12/Diploma?

    *   Marks obtained in B.Tech.

    *   Aptitude skill

    *   Communication skill

    *   Technical skill

    *   Management Skill

    *   General Knowledge

    *   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.
4. **Encoding categorical features:** As we know the machine learning perform with with numerical, so we converted the data into numerical format. Such as for excellent:5, very good:4, good:3, average:2 and for poor:1. As the students have rated them self on the form.And similarly for Gov Job- 0, M.Tech/ME/MS -1, MBA - 2, Others - 3, Prvt Job - 4.
5. **Classifier:**we have used two classifier that are:
    1. K- nearest neighbor

2. Decision Tree.
6. **Training and Testing:** we have used 80-20 rule on data, for training 80% and for testing 20% of 330.

IV - PROPOSED WORK

We have number of prediction system which help the students to predict the career. But the the system which are available on different site are on factors like there personal traits. And also there are system on websites which provide the career option for the students.

We provide the prediction on the bases of the students academic performance and skills. So, lets have look into the classification algorithm which will help in prediction:

i.   **K-nearest neighbor:-** It is a supervised learning method which is used for classification and regression. To classify it uses the majority of votes of K amount of neighbors. It follows three steps:
a)   It choose the K
b)   Then identifies the K nearest neighbor
c)   It provides the mode of K nearest neighbor for the classification and mean for regression

Note: there is no any particular way or formula to choose K , it is selected randomly.

ii. **Decision Tree:-**It is also the supervised learning algorithm in machine learning. As its name shows the flow or we can say structure like tree which works on condition. It is also known for effective and good algorithm to perform prediction. In this algorithm the decision on root node is final and because as the algorithm starts the root node is compared with every other leaf nodes till the end.

We have developed two classification models by using above algorithms to predict the career of students. These two models help the student to know the right path and what choices they should make to get success in their life. Not only they will be able to know their career options they should opt but also analyses their skills and academic performance. With the help of model and after knowing the prediction student can get focused on one path and they will not be in dilemma and can easily start working on it , so that they can achieve their goals.The students who had

decided to opt for any of the career option will also get assured if their decision is right or wrong and their skills are sufficient to achieve their goal.

V - RESULT & DISCUSSION

This part contain the main research work which is done. It consist of result which is obtained by different demonstration.

International Journal of Innovations in Engineering and Science,   www.ijies.net

Table 1 : *It contain the algorithm used , hyper parameter, random state, execution time ,and accuracy for the classifier development*

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*

S.no	Developed classifier	Confusion Matrix
1	K-nearest neighbor	[[11 0 0 0 11] [0 0 0 0 2] [1 0 0 0 2] [4 0 0 0 0] [4 0 0 0 31]]
2	Decision Tree	[[6 0 0 0 14] [0 1 0 0 4] [0 0 0 0 2] [2 1 0 0 2] [2 0 0 0 32]]

Heat-map:

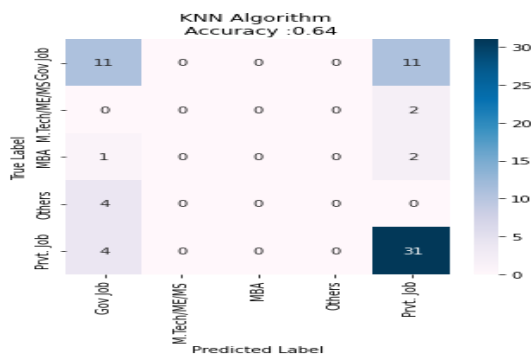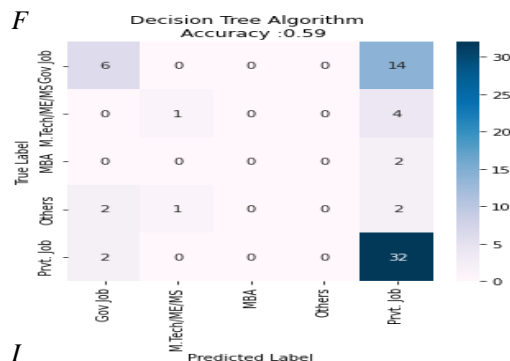*Fig. 1 - K nearest neighbor*

Fig- 2 -Decision Tree

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	recall f1-score	support
		Govt 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
		Pvt. Job	0.67	0.89 0.77	35
		Accuracy		0.64	66
		macro avg	0 .24	0.28 0.26	66
		weighted avg	0.54	0.64 0.58	66
2	Decision tree		precision	recall f1-score	support
		Govt. 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
		Pvt. Job	0.59	0.94 0.73	34
		accuracy		0.59	66
		macro avg	0.34	0.29 0.28	66
		weighted avg	0.52	0.59 0.52	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

ROC Curve- It is graph which represents and shows us the performance of the classification model at every thresholds.

1.   K nearest neighbor-

2.   Decision Tree-

International Journal of Innovations in Engineering and Science,   www.ijies.net
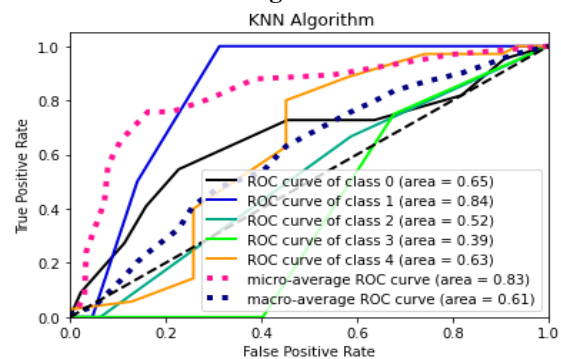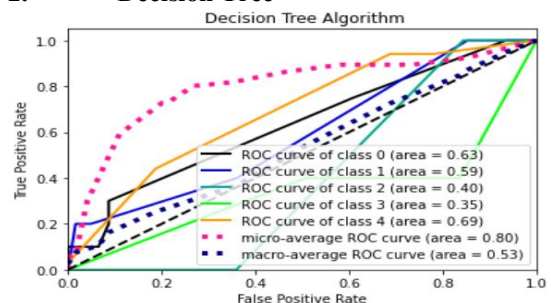
VI -CONCLUSION

This system will be helpful for each and every students 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 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.

REFRENCES

[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 (ICACDOT), 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 (ICACDOT), 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 Communicaton 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.*