

Career Counseling System For Graduate Student Using Machine Learning

Ajit Kumar Singh

Abstract:

Proper Career Counselling of final year undergraduate student is an important task of Higher Education system because it shapes the entire life of a student and change their life to next level. This is basically a task of a good teacher who observed their student minutely and guided them but number of such teachers are not up to the mark and this became a serious problem in higher education. Most of the student in India choose their career based on relatives' suggestions, friends' suggestions or influenced by media and internet without knowing their own potential and liking they blindly follow others suggestions. Due to lack of career guidance many of them are not happy with their job because the job they choose cannot suit their natural behaviour. We can solve this problem with the help of Artificial Intelligence. In this paper I am implementing a Machine Learning based Recommendation System that help student to choose their career among business and higher study depending upon eight parameters. I am using python to implement our recommender system using Decision Tree and Random Forest algorithms.

Keywords: *Machine Learning, Decision Tree, Random Forest, Recommender System, Education, Artificial Intelligence.*

1. Introduction

There is an important aspect of education system is to guide a student towards the future he or she actually deserves and that would must be according to his skills, ability and personality. In India, revelation came during nine-month survey conducted by CSIR (Council of scientific and Industrial Research) and NML (National Metallurgical Laboratory) (17) that about 40% of the school students are clueless about their career choice, 30% want to go for engineering, 20% say that they are preparing for medical stream and 10% go for MBA. Sometimes the students are guided towards the wrong career options due to some circumstances that are not desired at all. In other countries we can also see such incidences of mismatched career. According to an article published in The Guardian on 25th November 2018(16), huge number of mismatched graduates in UK face poorer prospects and lower earnings than their peers who enter on careers, that are a better fit for the knowledge and skills, that they have acquired through three or four years of study. According to a survey report published on 19th May, 2022 in Financial Express '59% of the Indian workforce is not happy at work' and another report published in The Economic Times on 18th January 2022 points to the fact that '71% of employees rethinking their careers' (19).

In India the student teacher ratio in higher education institutions is not very satisfying and it is not possible for the educators to guide individual student for his or her appropriate

future. The only solution to this problem is automated carrier guidance system or smart carrier guidance system which can be implemented without any human interventions. We have proposed a system that will recommend students who should continue higher education and who should be going for entrepreneurship or some business. We have thousand data of students, our system will analyse this data and direct the students towards his or her career option which will be according to their personality, skills and ability.

2. Literature Review

After studying several literatures, we found that intelligent career guidance system can be categorize into three types. The first one guide a student or a job aspirant towards the job he or she wants to have in future. Second type is work for job recruiter end to choose best candidate among database depending upon their skills, personality, academic performance etc. the third type shows the pathways to students from schools and higher education institutions for their future study or job sector. Some notable articles and their outcomes are shown in chronological order in below table 1.

Year	Article Title	Outcomes	Machine Language Used
2016	Machine Learning Approach for Future Career Planning	The complete pathway to a desired job profile is recommended here	Markov Chain Theory
	A Probability Machine Learning Approach for Eligible Candidate Selection	Used Naïve Bayes classifier algorithm of supervised machine learning approach to develop the software which will help choosing appropriate candidate for a specific type of job	Naïve Bayes Classifier
2018	Intelligent Recommendation System for Course Selection in Smart Education	Here the Computation of spares matrix is preferred as majority of the element of the matrix are zero and they have computed the hit rate and the average reciprocal Hit-Rank and compared the outcomes with the order methods.	SLIM(Sparse Linear Matrix)
	Student Career Prediction Using Advanced Machine Learning Techniques	XG boost has highest accuracy as compared to Decision Tree or SVM algorithm	SVM (Support Vector Machine), XG boost, Decision Tree
2019	Smart Career Guidance and Recommendation System	Machine learning algorithm such as decision trees Matrix factorization collaborative filtering and probabilistic graphics models have been applied to develop students' performance prediction algorithms	Decision Trees, Matrix factorization, collaborative filtering and probabilistic graphics models
	Student Future Prediction Using Machine Learning	The system will recommend suitable career options for the student based on their personality, interest and their capacity	Decision Tree, Logistic Regression
2020	Career Guidance System Using Machine Learning for	The System is based on some personal and academic	Naïve Bayes

	Engineering Students CS/IT	information filled up by the student	
	Career Guide Application Using ML	Focused on Career Counselling for the school students of 10 th grade	K means clustering Algorithm
2021	Student Career Guidance System Using Machine Learning	The student's self-assessment by answering self-assessment form and then the data receive from the self-assessment format was processed and then in the back end the Logistic regression is employed to process the data and sort out the best suitable result for the students.	Logistic Regression, Decision Tree, Gradient Boosting, SVM
	Student Career Prediction Using Machine Learning Approaches	The random forest classifier IS advantageous over the other machine learning techniques DT, SVM and Ada boost. According to their study Random Forest Technology has 93.00% accuracy	Decision Tree, SVM, Ada boost, Random Forest
2022	Career Prediction Classifiers Based on Academic Performance and Skills Using Machine Learning	The Machine Learning algorithms employed here are KNN, SVM SGD, Logistics Regression, Decision Tree, and Neural Network, 80% of the collected data has been employed for training of the system and rest 20% are for testing.	KNN, SVM, SGD, Logistics Regression, Decision Tree
	Intelligent Career Guidance System Using Machine Learning	Different machine learning algorithm such as as KNN, Naive Bayes, Decision Tree, ID3, Random Forest, SVM, Regression techniques etc and then the outcome of different machine learning algorithms will be compared and finally the best suitable career path will be recommended to the students.	KNN, Naive Bayes, Decision Tree, ID3, Random Forest, Logistic Regression SVM,

Table 1: Literature Review Summary

3. System Implementation

Our system will guide a student to choose entrepreneurship or higher study after graduation depending on eight features (Marks in HS, Marks in UG, Physical Health, Communication Skill, Self Confidence, Risk Taking ability, Curiosity and Family Support) considered by our system. The steps of our proposed system are:

Step 1: Generate a dataset of thousand students and verify it see figure 1.

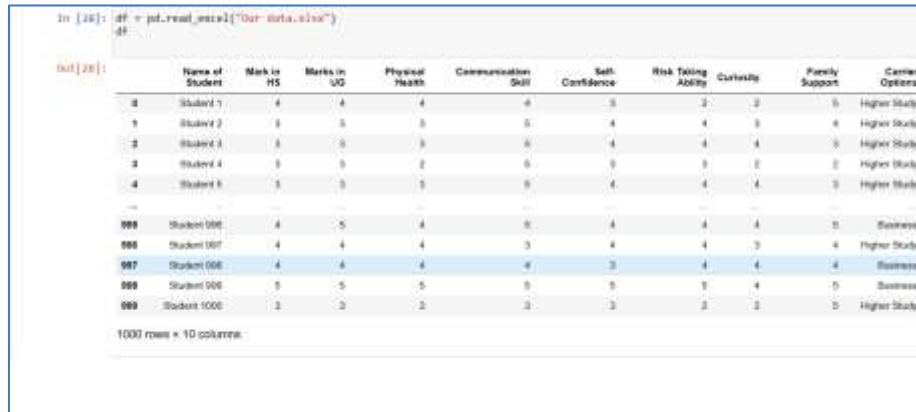


Figure 1: Proposed Data set

Step 2: Uses Python tools to find out the Correlation matrix of our dataset. Correlation matrix is representing correlation coefficient between different variables of data set. The values of correlation matrix are lies between +1 to -1 where +ve sign represents regular correlation and -ve sign indicates inverse correlation. The correlation matrix of our proposed system is shown in figure 2. The figure shows good correlation among different variables of our data set.



Figure 2: Correlation Matrix

Step 3: Before applying Machine Learning algorithm, we split our data set to training and testing. The training data set is used to create machine learning models and testing data sets are used to test the model for system accuracy. The figure 3 shows the code for data split.

```

In [51]: X_train = strat_df_train.drop('Carrier Options', axis = 1)
         y_train = strat_df_train['Carrier Options']

In [53]: X_test = strat_df_test.drop('Carrier Options', axis = 1)
         y_test = strat_df_test['Carrier Options']

In [54]: X_train.shape, y_train.shape, X_test.shape, y_test.shape
Out[54]: ((800, 4), (800,)), (200, 4), (200,))

In [ ]: X_test

In [56]: test_data1 = [3,4,5,3,4,3,5,5]
         test_data2 = [4,4,4,4,5,5,5,5]
         test_data3 = [5,5,5,5,3,4,5,5]
         test_data4 = [5,3,4,3,4,3,3,4]

```

Figure 3: Data Splitting

Step 4: Apply Decision tree algorithm to our data set. Decision tree is a supervised machine learning algorithm. Among several machine learning algorithm, we choose decision tree because it is easy to understand and mimic human thinking. A decision tree is simply taking a question and depending upon answer split the tree into sub trees. The decision tree of our data set is shown in figure 4.

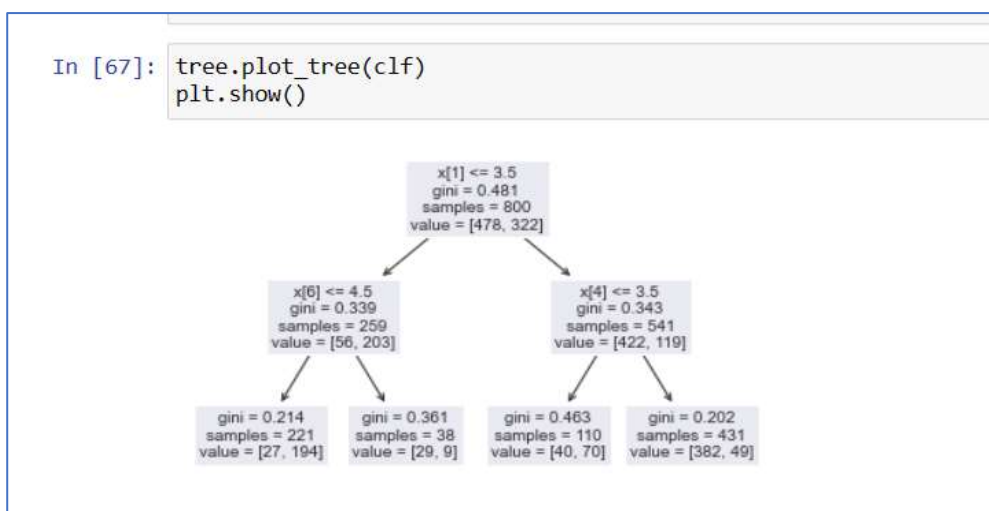


Figure 4: Decision Tree

The above decision tree taking decision for student that he / she should go for business or higher study. In figure:4 $X[i]$ represent a condition. Gini refers to the quantity of split. It is calculated according to the following formula

$$\text{Gini} = 1 - (x/n)^2 - (y/n)^2$$

Where x represents business and y represent higher study. Its value is always lie between 0.0-.05. Sample represent total number of students left at this point of decision. Value represents out of sample how many goes for business and how many go for higher study.

Step 5: Apply Random Forest algorithm to our data set. Random Forest is most popular supervised machine learning algorithm. It is based on the concept of combining multiple classifiers to solve the complex problem with high accuracy. It takes less training time as compared to other machine learning algorithm. We create data model with training data set and applied test data set to check the model is working properly as well as accurately. The figure 5 shows how training and testing data sets are selected and outputs of test data set.

```

In [38]: X_train = strat_df_train.drop('Carrier Option', axis = 1)
         y_train = strat_df_train['Carrier Option']

In [39]: X_test = strat_df_test.drop('Carrier Option', axis = 1)
         y_test = strat_df_test['Carrier Option']

In [40]: X_train.shape, y_train.shape, X_test.shape, y_test.shape
Out[40]: ((100, 8), (100, 1), (100, 8), (100, 1))

In [41]: from sklearn.ensemble import RandomForestClassifier
         from sklearn.model_selection import train_test_split

         clf = RandomForestClassifier()
         clf.fit(X_train.values, y_train)

Out[41]: RandomForestClassifier()
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [42]: test_data1 = [1,4,3,1,4,1,5,5]
         test_data2 = [1,4,4,4,5,5,5,5]
         test_data3 = [1,5,1,1,1,4,5,5]
         test_data4 = [1,5,1,4,1,4,1,5,2]

In [44]: print("TEST DATA 1")
         predict_scores(test_data1)

TEST DATA 1
- Park in US : 1
- Parks in US : 4
- Physical Health : 5
- Communication Skill : 1
- Self-Confidence : 4
- Risk Taking Ability : 1
- Curiosity : 5
- Family Support : 5

MEAN:
- 4.0

PREDICTION : ['Business']

In [45]: print("TEST DATA 2")
         predict_scores(test_data2)

TEST DATA 2
- Park in US : 1
- Parks in US : 4
- Physical Health : 2
- Communication Skill : 4
- Self-Confidence : 5
- Risk Taking Ability : 5
- Curiosity : 5
- Family Support : 5

MEAN:
- 4.5

PREDICTION : ['Business']

```

Figure 5: Random Forest

Step 6: Both the algorithms recommend a carrier option for student but we choose recommendation of Random Forest Algorithm because The Decision Tree algorithm gave accuracy 88.7 percent and random forest gave 97.5 percent.

4. Conclusion

We are depending upon technologies for everything we want to do, that may be buying a new appliance for home or watching a movie or travelling to somewhere, whatever may be the need we first look for recommendation from the internet and most of the time we rely on the recommendation we have received. Here I have implemented a recommendation systems that recommend the students to choose their carrier either higher study or go for some business depending upon data that we provided. Though I have use created dataset but the accuracy of the system is very high. I plan to develop a web application to collect real dataset of undergraduate student and run the system on real dataset . In near future on the basis of suggestions from the students and academicians' additional features will be added to this basic career recommendation system.

7.References

1. Lin, J. (2018). Intelligent Recommendation System for Course Selection in Smart Education, *Procedia Computer Science* ,129, 449–453.

2. Prasanna, L. (2019). Smart Career Guidance and Recommendation System. IJEDR, 7(3), 633-638.
3. Daharwal, A. (2020). Career Guidance System using Machine Learning For Engineering Students (CS/IT), 7(6),3417-3420.
4. Patel, H. (1986). Career Guidance System Using Machine Learning. IJCRT, 9(5), a57-a60.
5. Surve, A.(2021). Student Career Guidance System using Machine Learning. IRJET, 8(8), 3543-3546.
6. Sneha, H. S. (2022). A Survey On Intelligent Career Guidance System Using Machine Learning. IRJMETS, 4(4), 830-834.
7. Mourya, V.(2020). Career Guide Application using ML. IRJET, 7(9), 1170-1172.
8. Pandey, A. (2021). Career Prediction Classifiers based on Academic Performance and Skills using Machine Learning IJCSE, 9(3), 5-20.
9. Roy, K.S. (2018). Student Career Prediction Using Advanced Machine Learning Techniques, International Journal of Engineering & Technology, 7 (2.20), 26-29.
10. VidyaShreeram, (2021). Student Career Prediction Using Machine Learning Approaches. I3CAC, 7(8), DOI 10.4108/eai.7-6-2021.2308642
11. Chaudhary, D. (2019). Student Future Prediction Using Machine Learning. IJSRCSEIT, 5(2), 1104-1108. DOI : <https://doi.org/10.32628/CSEIT1952300>
12. Lou,Y. (2010). A Machine Learning Approach for Future Career Planning.
13. Jannat, E. M. (2016). A Probabilistic Machine Learning Approach for Eligible Candidate Selection, International Journal of Computer Applications,144(10),1-4.
14. Sneha, H. S. (2022). Intelligent Career Guidance System using Machine Learning. IJARCCCE, 11(7), 117-120.
15. Das, D. (2017). A Survey on Recommendation System. International Journal of Computer Applications, 160(7), 6-10.
16. www.ibm.com/topics/machine-learning
17. Available at: <https://www.theguardian.com/higher-education-network/2018/jan/25/too-many-graduates-are-mismatched-to-their-jobs-whats-going-wrong#:~:text=These%20mismatched%20graduates%20face%20poorer.careers%20advice%20isn't%20working.>
18. Available at <https://timesofindia.indiatimes.com/home/education/news/survey-shows-40-students-confused-about-career/articleshow/12888793.cms>
19. Available at: <https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1725914#:~:text=Pupil%20Teacher%20Ratio%20in%20Higher%20Education%20in%202019%2D20%20is%2026.>
20. Available at:<https://economictimes.indiatimes.com/jobs/71-of-employees-rethinking-their-careers-report/articleshow/88976325.cms>
21. Kusumaningrum, D.P. (2017). Recommendation System for Major University Determination Based on Student's Profile and Interest. Journal of Applied Intelligent System, 2(1), 21 – 28.
22. Goldberg, D. (1992) Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, 35, 61-70.
23. <https://www.javatpoint.com/machine-learning-random-forest-algorithm>
24. <https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm>