IMPLEMENTATION OF RECOMMENDATION SYSTEM FOR CAREER COUNSELING OF HIGHER EDUCATION STUDENTS USING MACHINE LEARNING

Ajit Kumar Singh[1], Moumita Kumar Roy[2]

[1]Department of Computer Science and Applications, Netaji Mahavidyalaya, Arambagh, Hooghly, West Bengal

[2]Department of Electronics, Netaji Mahavidyalaya, Arambagh, Hooghly, West Bengal

Abstract

Choosing right career path is most critical task for a student who is in his or her terminal education. Most of the students in India choose their career based on suggestions from peers/ relatives or influenced by media and internet without knowing their own potential and passion they blindly follow the trends. Due to lack of proper guidance a good number of employees are not happy with their job profile. In India it is not possible for educational institutions to guide every student properly about their career because student teacher ratio is not up to the mark. This problem can be solved with the help of Artificial Intelligence. In this article we are implementing a Machine Learning based Recommendation System that will help students to choose their career among business and higher study, depending upon eight parameters. We have employed python to implement our recommender system using Decision Tree and Random Forest algorithms.

Keywords: Education, Artificial Intelligence, Machine Learning, Recommendation System Decision Tree, Random Forest algorithms

1. Introduction

We are leaving in such a time where we are habituated with technological automation in almost every aspect of life. Today we are depending upon technologies for everything 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. The educational sector especially the higher education sector is not an exception. In last one decade use of technology in education system has increased remarkably.   A student having fair internet connection and a device like smartphone or laptop or tablet will be able to learn from internet about any subject of their choice. There is another 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 to 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].

The student teacher ratio in higher education institutions is not very satisfying in India 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 reviewed articles on automated career guidance employing different machine learning algorithms and then depending on it 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

We have gone through a number of literatures before working out on our system. If we categorize them then we get mainly three types of smart career guidance system. The first one guides a student or a job aspirant towards the job he or she wants to have in future. The path is shown in steps so that they can achieve their goal [12].

Table 1: Comparative study of literature

Sl No.	Paper Title	Year	Outcomes	ML Approch
1	Machine Learning Approach for Future Career Planning	2016	The complete pathway to a desired job profile is recommended here.	ML algorithm developed on the basis of Markov Chain theory
2	A Probability Machine Learning Approach for Eligible Candidate Selection	2016	Used Naive Bayes classifier algorithm of Supervised Machine Learning approach to develop the software which will help choosing appropriate candidate for a specific type of job.	Naive Bayes classifier
3	Intelligent Recommendation System for Course Selection in Smart Education	2018	Here the computation of sparse matrix is preferred as majority of the elements of the matrix are zero, and they have computed the hit rate (HR) and the average reciprocal Hit-Rank (ARHR), and compared the outcomes with the other methods.	ML algorithm developed based on SLIM (Sparse Linear Matrix)
4	Student Career Prediction Using Advanced Machine Learning Techniques	2018	Machine Learning algorithms like SVM (Support Vector Machine) XG boost and Decision Tree for the training and testing of the system and then it is found that XG boost has highest accuracy as compared to Decision Tree or SVM algorithms.	SVM (Support Vector Machine),XG boost,Decision Tree
5	Smart Career Guidance and Recommendation System	2019	Machine Learning algorithms such as decision trees Matrix factorization collaborative filtering and probabilistic graphical models have been applied to develop students' performance prediction algorithms	Decision Tree, Matrix Factorization Collaborative Filtering, Probabilistic Graphical Models
6	'Student Future Prediction Using Machine Learning'	2019	The system will recommend suitable career options for the students based on their personality, interest and their capacity	Decision Tree, Logistic Regression
7	Career Guidance System Using Machine Learning for Engineering Students CS / IT	2020	The system is based on some personal and academic information filled up by the student.	Naive Bayes
8	Career Guide Application Using ML	2020	Focused on career counselling for the school students of 10th grade	K Means Clustering Algorithm
9	Students Career Guidance System Using Machine Learning (2021),	2021	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
10	Student Career Prediction Using Machine Learning Approaches	2021	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
11	Career Prediction Classifiers Based on Academic Performance and Skills Using Machine Learning	2022	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
12	Intelligent Career Guidance System Using Machine Learning	2022	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, SVM, Logistic Regression

However, it is not evaluated that the job will be suitable for a particular person or not this is the major drawback of this kind of career recommendation system. The second type of career smart career guidance system works forms the job recruiter end. This type of system works on job aspirant's database a recommends the suitable candidates from the database for a particular type of job, analyzing their skills, personality, academic performance etc. The other section of

career recommendation system shows the pathway to students from schools & higher education institutes for their future study or job sector. The summary is shown above table. After going through all these papers, we summarize the result of these survey into following points:

- The country like India where the student teacher ratio is not at all satisfactory to guide every student for their suitable career, automated recommendation system may be the ultimate solution.
- We have so many Machine Learning algorithms (Supervise, Unsupervised and Reinforcement learning) and Data Mining techniques for creating an efficient Recommendation System (RS) but among all of these the Random Forest Algorithm is the best suitable option for career counselling.
- Every RS has its own pros and cons.
- The major challenge to develop an efficient RS for career counselling is to evaluate human capability and skills with the help of algorithm.
- The roles of the users are also important for elevated accuracy through RS because the system will recommend according to input from the user so the users have to be responsible enough while provide the input.
- The size of sample data also plays an important role towards the accuracy of RS, the more data is available the better the accuracy we can expect from the system.
- Feedback from the user and needed modifications accordingly, will lead towards higher accuracy of the RS system.

The main databases used in above recommendation systems and the major ML algorithms used are shown in tabular form below (Table 2):

Table 2: Major database and ML algorithms used in automated career recommendation system

Popular database employed for data collection	SQL database
	Kaggle
	Institutional Students Records
	MongoDB database
	LinkedIn profiles
	SQL database
Popular ML algorithms employed for data processing	KNN
	Decision Tree
	SVM
	Adaboost
	Random Forest
	Gradient Boosting
	Naïve Bayes
	ID3
	Neural Network
	Logistic regression
	XG Boost
	K Means Clustering Algorithm

3. Machine Learning Algorithm

The term machine learning was coined in 1959 by Arthur Samuel. Machine Learning is a branch of Artificial Intelligent that focuses on the use of data and algorithms to imitate the way that human learn and gradually improving the accuracy [20]. The machine learning algorithm is classified into three categories which are supervised, unsupervised and reinforcement learning. Supervised learning refers to the Machine learning in which the machine needs external supervision to learn. But in unsupervised learning the machine does not need any external supervision to learn from the data. In case of reinforcement learning an agent interact with its environment by producing some actions and learn with the help of feedback.

The goal of this paper is to implement a recommendation system that recommend the students to choose their career either higher study or go for some business depending upon data that they provided. For implementing above system, we have so many machine learning algorithms among them we have chosen Decision Tree and Random Forest algorithms for our system. Decision Tree machine learning algorithm is a tree like structure where each node represents a binary condition and depending on true and false value tree is subdivided into branches each branch corresponds to attributes and each leaf node represent prediction. Random Forest algorithm was developed by Leo Bremen. It is a powerful and widely used algorithm that is used for classification as well as regression. For huge data set Random Forest works efficiently and accurately. Random subset of training data creates trees. The test data passed through the forest after the forest have been trained. Each tree generates an output and majorities of outputs

considered as final output. In our proposed system Random Forest algorithm predict the career options of students depending on our dataset.

4. Proposed System

We are trying to implement a recommender system that will guide a student to choose career options among business and higher study. Our system recommends career depending upon eight features provided by students which are Marks in HS, Marks in UG, Physical Health, Communication Skill, Self Confidence, Risk Taking ability, Curiosity and Family Support. Initially we have downloaded a data set from Kaggle and apply several tools on python for data verification then we have generated a data set of one thousand students and apply Decision Tree and Random Forest algorithm to predict the career options for students. The steps of our proposed system are shown in figure 1.
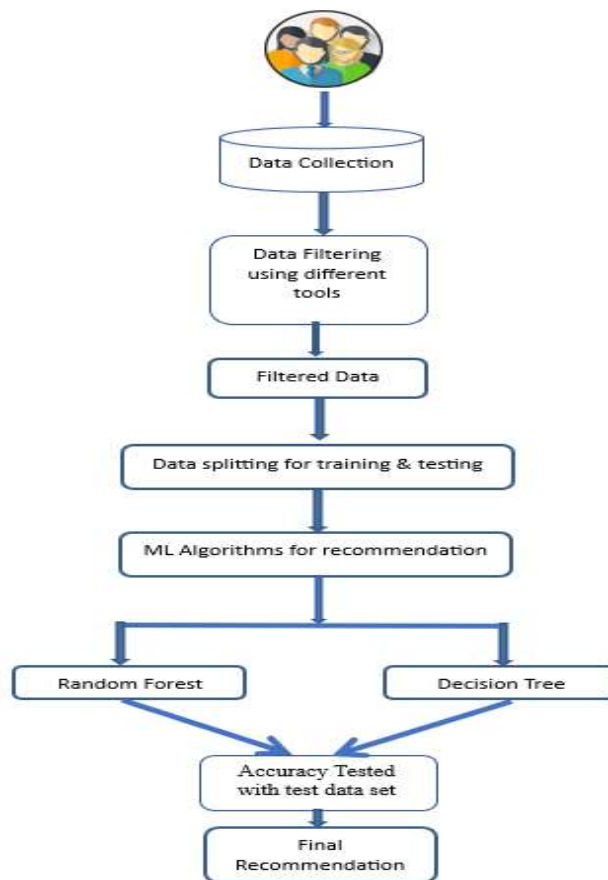

Figure 1: System Architecture

Our proposed data set for students depending on eight attributes is shown in figure 2.

```
In [28]: df = pd.read_excel("Our data.xlsx")
         df
```

Out[28]:

	Name of Student	Mark in HS	Marks in UG	Physical Health	Communication Skill	Self-Confidence	Risk Taking Ability	Curiosity	Family Support	Carrier Options
0	Student 1	4	4	4	4	3	2	2	5	Higher Study
1	Student 2	3	3	3	5	4	4	3	4	Higher Study
2	Student 3	3	3	3	5	4	4	4	3	Higher Study
3	Student 4	3	3	2	5	3	3	2	2	Higher Study
4	Student 5	3	3	3	5	4	4	4	3	Higher Study
...	...	...	...	...	...	...	...	...	...	...
995	Student 996	4	5	4	5	4	4	4	5	Business
996	Student 997	4	4	4	3	4	4	3	4	Higher Study
997	Student 998	4	4	4	4	3	4	4	4	Business
998	Student 999	5	5	5	5	5	5	4	5	Business
999	Student 1000	3	3	3	3	3	2	2	5	Higher Study

1000 rows × 10 columns

Figure 2: Proposed Data set

Matrix is nothing but a collection of rows and columns, 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 3. The figure shows good correlation among different variables of our data set.

Figure 3: Correlation Matrix

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 4 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, 8), (800,), (200, 8), (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 4: Data Splitting

4.1 Algorithms: To implement our proposed system, we are using two most popular and efficient machine learning algorithm which are Decision Tree and Random Forest Algorithm.

     4.1.1 Decision Tree: 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. We apply decision tree machine learning algorithm to our data set and we get the tree shown in figure 5.


```
In [67]:  tree.plot_tree(clf)
          plt.show()
```
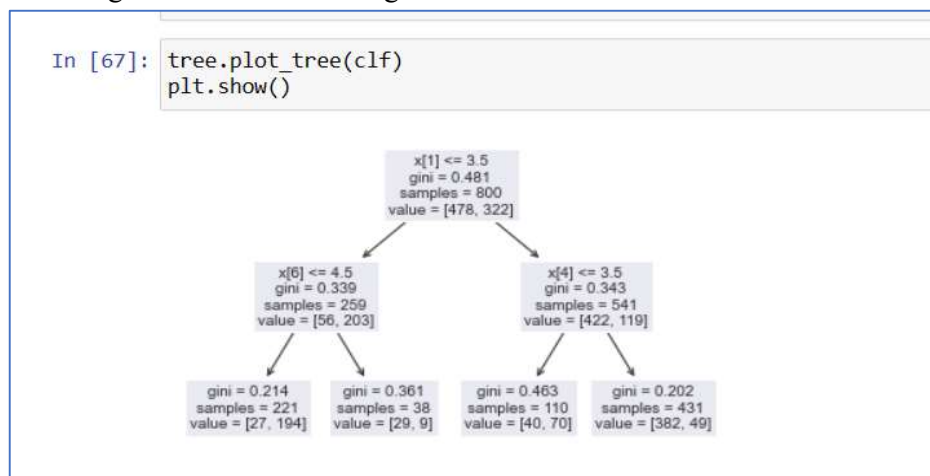
Figure 5: Decision Tree

     The above decision tree taking decision for student that he / she should go for business or higher study. In figure:3 X[i] represent a condition. Gini refers to the quantity of split. It is calculated according to the following formula

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.

     4.1.2 Random Forest Algorithm: 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. The working principle of Random Forest Algorithm is summarized in following steps:

     Step 1: For a given data set select random samples for training and testing.

Step 2: Construct decision tree for each sample.

Step 3: Collect all the output of the decision tree.

Step 4: The majority output is considered as the final output of the algorithm.

During our implementation of recommendation system for career counseling of higher education student we split data set into two-part training and testing. We create data model with training data set and apply test data set to check the model is working properly as well as accuracy. The figure 6 shows how training and testing data sets are selected and outputs of test data set.

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

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

In [40]: X_train.shape, y_train.shape, X_test.shape, y_test.shape

Out[40]: ((800, 8), (800,), (200, 8), (200,))

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 = [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]
```

```
In [44]: print('TEST DATA 1\n')
         predict_scores(test_data1)

         TEST DATA 1

            - Mark in HS : 3
            - Marks in UG : 4
            - Physical Health : 5
            - Communication Skill : 3
            - Self- Confidence : 4
            - Risk Taking Ability : 3
            - Curiosity : 5
            - Family Support : 5

         MEAN:
            - 4.0

         PREDICTION : ['Business']

In [45]: print('TEST DATA 2\n')
         predict_scores(test_data2)

         TEST DATA 2

            - Mark in HS : 4
            - Marks in UG : 4
            - Physical Health : 4
            - Communication Skill : 4
            - Self- Confidence : 5
            - Risk Taking Ability : 5
            - Curiosity : 5
            - Family Support : 5

         MEAN:
            - 4.5

         PREDICTION : ['Business']
```

Figure 6: Random Forest

5. Result And Discussion

The proposed system is helpful for higher education students to choose their career among business and higher study. Our system predicts career options for students based on decision tree and random forest algorithms. We are not using real data set we have created a data set, initially we filtered data set so that it fit into our machine learning algorithm. The Decision Tree algorithm gave accuracy 88.7 percent and random forest gave 97.5 percent. As Random Forest algorithm gave better result so all further development is done with Random Forest algorithm.

6. Conclusion

Authors have framed out basic career guidance system for our undergraduate students which will guide them initially if they can be an entrepreneur or pursue their higher studies in future. Here the database we have used is data set that we have generated. The system will be implemented on actual data gathered from the students after this. On the basis of suggestions from the students and academicians' additional features will be added to this basic career recommendation system and try to recommend more specific career options for the students. In near future we will develop an application for the students based on this framework and web-based user interface in our website so that students can be benefited from this recommendation system.

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. IJARCCE, 11(7), 117-120.
15. Das, D. (2017). A Survey on Recommendation System. International Journal of Computer Applications, 160(7), 6-10.
16. 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.
17. Goldberg, D. (1992) Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, 35, 61-70.
18. **Online Resources:**

- 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.

- https://timesofindia.indiatimes.com/home/education/news/survey-shows-40-students-confused-about-career/articleshow/12888793.cms

- https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1725914#:~:text=Pupil%20Teacher%20Ratio%20in%20Higher%20Education%20in%202019%2D20%20is%2026.

- https://economictimes.indiatimes.com/jobs/71-of-employees-rethinking-their-careers-report/articleshow/88976325.cms

- www.ibm.com/topics/machine-learning
- https://www.javatpoint.com/machine-learning-random-forest-algorithm
- https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm