Learning Techniques based model for User Interest and Career Prediction

Ms. Mansi Sonawane, Ms. Rutuja Raut, Mr. Rushikesh Sonawane, Mr. Shreyash Naik, Ms. S. B. Bhonde
Department of Computer Engineering
Amrutvahini College of Engineering, Sangamner, Maharashtra, India

Abstract: *Choosing a fitting career is one of the most significant choices and with the expansion in the quantity of profession ways and openings, settling on choice have gotten very hard for the understudies. Numerous understudies are befuddled about their vocation choices. This may prompt wrong vocation determination and afterward working in a field which was not implied for them, along these lines diminishing the efficiency of human asset. As the students are belonging through their academics, they need to realize their capabilities and check their areas of interest so that they can decide which career option is best suited for them in future. This system will help nowadays youth to decide which career path is best for their future that brings out the best results if they choose that prescribed career. Also, this will help to improve the performance of the student and motivate them in their area of interest so that they will be focused on their targeted career. When one decides a career, this choice can shape one's life entirely. Recently, more and more people have begun to reconsider their career options and change careers at a later time in their life. This can be prevented by proper counselling of young teenagers before they begin their graduate studies .This system is based on a test that a student has to perform and depend on the answers that are provided by the student, it will generate a summarized result.*

Keywords: Career guidance, student counselling, professional development, education, practices, abilities, machine learning

I. INTRODUCTION

We all know that nowadays the competition goes on increasing day by day, choosing a career is also one of the important tasks of today's generation as the world is getting more technical. This issues mostly happens with the students that what their interest lies in. Every parent wants to see their children engineer or doctor but nobody asks what the child interested for, also the parents are worried about the future of their child. So, this software helps both student and their parents. It helps the student toknow where his/her actual interest lies in, which subject he has to choose that provides the best result of his future. We are trying to guide the student by taking him through a series of test that will give them the idea about how to start the career by telling them what subject they will choose so that this will be beneficial for them and what they will do further to get the best result. So, the website starts with conducting the aptitude test as it is better to polish their mind. The student will have to register himself on the app. The main process will start after the registration of the student into our website, here we are providing a platform to the student where he or she has to perform the aptitude test. So basically, when the student gets to log in to our website, there we ask him to select his standard, as there are different levels of test for different standards. After that, they will proceed for the test. On proceeding further, a student is asked about the field in which he is interested inand based on that the test questions will come. There are different fields given as Engineering and Technology, Arts, Commerce, Law, Humanities, Hotel Management, social science. The questions are based on these fields. When the student successfully submitted the quiz, the student will receive the detailed result along with the perfect advice from the expert. The expert will tell which career is best for the student. After getting summarized result from the system expert, the student will get to know about in which subject he is interested in, which proves to be a better decision for his future. Existing system does not provide a generic solution for a larger population, as these systems are customized to a particular geographicalarea, where a student with different cultures and background are looking through. There are several solutions developed like using data mining, advance statistical models, and algorithms. So, computing a system using these

techniques will result in highly complex systems which are difficult to maintain. Many information is present on the internet and getting relevant data from a large source is a difficult task.

II. LITERATURE SURVEY

In this section, we review some papers in related areas. Vaidu et al implemented ML techniques to predict employability skills, based on students' performance. They implemented KNN and Naïve Bayes algorithms for classification into several classes. Both the algorithms are used to predict of the employability of students. The results obtained for KNN is 95.33% and for the Naïve Bayes, it is 67.67% accuracy. [2] Iqbal et alhas discuss different ML techniques to predict grades in different subjects. He used Factorization, Regression and Classification models such as collective filtering and Restricted Boltzmann Machine (RBM) techniques for analysing the data collected from Information Technology University (ITU), Pakistan.They evaluated the performance of the students who got admission in the different Program, using ML techniques. The RBM technique was found to be the best among various other ML techniques he used. [3] Byung-Hak et al have proposed a Deep Leaning based algorithm GritNet for which forecasts the future performance of students. GritNet has given much better results as compared to the standard Logistic Regression according to this paper. It takes student data from Udacity Nanodegree Programs. [4] Jieet al have proposed a ML approach for predicting the student performance during their Bachelors degree. This study is based on past and present performance. The proposed system uses a bi- layered structure comprising of multiple base predictor and a data driven approach, based on latent factor models for constructing an efficient base predict. This paper shows that the proposed method gives better performance and accuracy as compared to the original benchmark approches [5] Pojon Murat investigates ML algorithms to analyze and predict students performance. He uses three algorithms, viz. and Naïve Bayes, Linear Regression and Decision Trees Classification on twodifferent datasets: row version and feature engineered version. The optimal is Naïve Bayes for the first dataset with 98% accuracy and Decision Tree is found for second dataset with accuracy of 78% [6] Singh et al discuss ML approches to predict the academic score of engineering students subject wise He predict subject scores in ongoing courses by analysing the performance in previous semester. For this purpose, they use two classification techniques, C4.5 Decision Tree classifier and Naïve Bayes. The result shows that C4.5 Decision Tree has better accuracy than Naïve Bayes. [7] BendengnuKsung et al propose a Deep Neural Network model for prediction of student performance and classify them into various category. The paper compares the accuracy of Deep Neural network with existing different Machine Learning algorithms like Decision Tree classifier [J48], ANN and NaiveBayes. This model reaches upto accuracy of 84.3% and is far better than other Machine Learning Algorithms. [8] Pushpa S et al predicted final results of students whether the student will pass or fail in the exam based on students' performance in the previous semester and exams using Machine Learning Algorithms. This paper used four algorithms: Naïve Bayes, SVM, Gradient Boosting and Random Forest . The accuracy of Random Forest was found to 89.06% which was greater than other three algorithms. [9] Gerritsen L. et al predicted student performance from Learning Management System (LMS) data in the domain of data mining in education field using Neural Networks. The dataset used for this paper was a Moodle log file which contains 4601 students' information. This paper compared the performance of Neural Networks against six other classifiers. These algorithms were Logistic Regression, SVM, Random forest, Decision Tree classifier, KNN and Naive Bayes. This paper concluded that the accuracy of the Neural network was highest than six other classifiers. [10] Martín S. et al studies the performance of four Machine Learning algorithms with different objectives, to predict the dropouts in university students. Four algorithms, Neural netwirks, Logistic Regression, SVM, Random Forest are used. The dataset uses those students who took admission in degree program at the Instituto Tecnológico de Costa Rica (ITCR) between the years 2011- 2016.The Random Forest (RF) algorithm with two variables was found best for predicting dropouts.

III. PROPOSED SYSTEM

The system of the software is built in such a format that there is a link between software and the student. As the system is based on Machine Learning and the information we embedded into the system, so based on that it will declare the final result. This system is basically for the people who are confused about what career they want in future and what career they will choose so that it would be beneficial for their future. So our product has two main components one is for recommendation and the other is for displaying the summarized result. The first set will contain the questions from all the

domain of the preferred area of interest of the user. Once the user will answer these questions, then the domains will be narrowed and this process will be several times until the final set of questions after which an appropriate result can be given to the user. So this test is particularly based on levels wherein first level we will ask a general question based on which we will categorise the interest of the student. Foreach level, we will dynamically change the questions based on the performance of the user. Advantages:- 1.The system will help the user by answering their career- related questions. 2.The built-in ML system will carry outall the processing to provide suitable answers to the user.3.As the software is easy to use so anyone can ask queries without having any foundation. 4.We will try to enhance the system where it could get more interactive for the user. The working of this software is based on different modules:- a) Maintenance module: This module is maintained by the administrator and is responsible for addition, deletion and updating of items. b) Evaluation module: Once a student answers the question, this module is responsible for the evaluation of the questions answered by the user. c) Online testing module: This module provides a platform for the students to take a test. d) Test generation module: Helps in implementing the system to generate a test for the user. e) Report generation module: This module will generate the summarize report after evaluating the final result.

IV. METHODOLOGY

A. Data Collection Data collection is one of the important tasks of Machine Learning projects. Because the input that we provide to the algorithms is the trained data. The accuracy and efficiency of the algorithms are based on the training data that we collected because the output depends on the data. There are multiple ways to collect data. some of the data is collected from the employees working in different organizations, some amount of data is collected from the social sites API, some of the data can be randomly generated and other is from the database and forms.

B. Data pre-processing: Data preprocessing is a technique used to transform the raw data into the useful and efficient format which will be understood by the computer system to perform the operations. Steps used for data pre-processing:

1. Tokenization : Tokenization of data is a technique by which the raw data is divides into small chunks which is called as tokens, tokens can be words, characters or sub words etc. The tokenization in python is done by using NLTK.
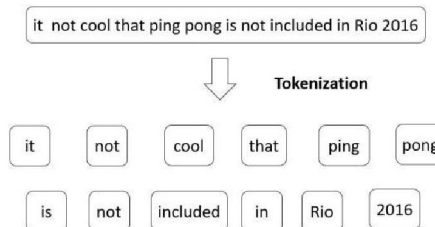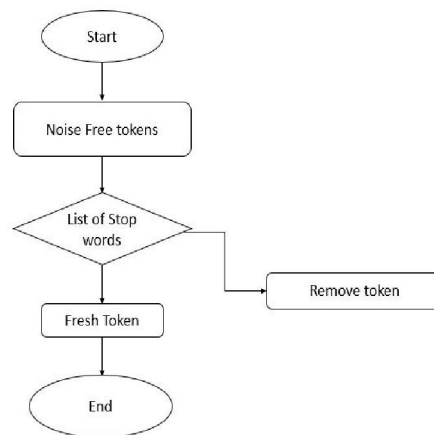

Fig 1 : Tokenization

Fig 2 : Stop word Removal

2. Stop-word removal : Removal of stop words is also major step in data pre-processing. Stop words area set of words which may not play a huge role in results or which contains very little information. So removal of such words from our data will

not affect on result. Stop words are used in NLP and text analysis to eliminate words that are used so commonly used and very less information associated with it. Example "a", "the", "is", "are"etc. To remove stop words we can use NLTK, spaCy, Gensim etc.
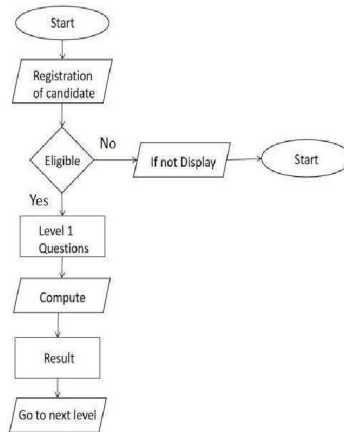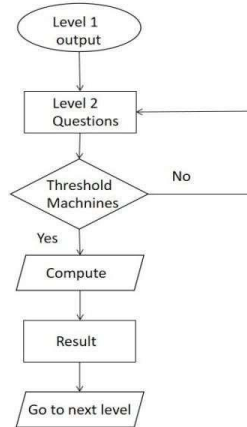
Flow Charts

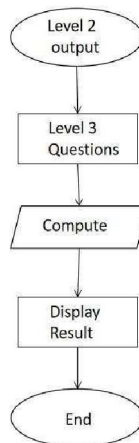
Fig 3: Level 1 flowchart

Fig 4: Level 2 flowchart

Fig 5: Level 3 flowchart

V. RESULT AND ANALYSIS

The main concept behind this system is to figure out student's area of interests.

This app will help the students to select an appropriate engineering stream after 12th.

The app covers different engineering streams and careers for every student and we also provide another stream if they don't like the first stream.

The system provides results in the graphical formatbased on the predicted results .

VI. CONCLUSION

Career counsellor system is an interesting idea. The opportunities that we are providing to the students through an online medium can make the use of this software for choosing the career based on the appropriate skills. There is a need of career counseling or career guidance in today's world of increasing career paths day by day. This research paper proposes a career counseling system which integrates career guidance with the artificial intelligence. This system comprises of an intelligent engine which delivers best career choices on the data of industry leading professional and with the help of ML applications system will perform accurately and at high performance. This system will be helpful to very individual seeking career guidance. This research paper concludes that this career prediction system is a boon for upcoming and today's world.

REFERENCES

[1]. Mccrea, N., "An Introduction to Machine Learning Theory and Its Applications: A Visual Tutorial with Examples",https://www.toptal.com/machine-learning/machine-learning-theory-an introductory-primer

[2].Vaidu, G., and Sornalakshmi, K., "Applying Machine Learning Algorithms for student employability prediction using R," International Journal of Pharmaceutical Sciences Review and Research, pp. 38- 41, 05, March 2017. [Online].Available: http://globalresearchonline.net/journalcontents/v43- 1/11.pdf

[3]. Iqbal, Z, Qadir, J., and Kamiran, F., "Machine Learning based student grade prediction: A case study," 17 Aug 2017. [Online]. Available: https://arxiv.org/pdf/1708.08744.pdf

[4]. Kim. B, Vizitei, E., and Ganapathi, V., "GritNet:Student performance prediction with Deep learning,"

19 Apr   2018 [Online]. Available:https://arxiv.org/abs/1804.07405

[5]. Xu, J., Horoon, K., and Schaar, V., "A Machine Learning Approach for Tracking and predicting studentperformance in degree program," IEEE Journal of Selected Topics in Signal Processing, Vol 11, pp. 742- 753, Aug. 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7894238/

[6]. Pojon Murat, "Machine Learning to predictStudentperformance,"2017 [7]. Singh, M., and Singh, J., "Machine Learning Techniques for prediction of subject scores:comparative study", International Journal of Computer Science and Network, Vol 2, issue 4, pp. 77-80, August2013. [Online]. Available: https://pdfs.semanticscholar.org/2368/3634d0999020d6a 90bf79fa605 ceebe90891.pdf

[8]. BendengnuKsung, and Prabhu, P., "Students performance prediction using Deep Neural Network," International Journal of Applied Engineering Research, Vol 13, Number 2, pp. 1171-1176, 2018. Available: https://www.ripublication.com/ijaer18/ijaerv13n2_46.pd f

[9]. Pushpa, S., Manjunath, T., Mrunal, T., Singh, A., and Suhas, C., "Class result prediction using machine learning," 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, 2017, pp. 1208-1212.

[10]. Gerritsen L. and Conijn R., "Predicting student performance with Neural Networks," dissertation, Dept. Humanities, Tilburg University, The Netherlands, May 2017.

[11]. Solis, M., Moreira, T., Gonzalez, R., Fernandez, T.,and Hernandez, M., "Perspectives to Predict Dropout in University Students with Machine Learning," 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), San Carlos, 2018, pp. 1-6.

[12]. https://www.dezyre.com/article/top-10-machine-learning-algorithms/2 02

[13]. Essaid EL HAJI, Abdellah AZMANI, Mohamed EL HARZLI(2014),"Multi-expert system design for educational and career guidance: an approach based on a multi-agent system andontology," Department of Computer Science, LIST Laboratory,Faculty of Science and Technology. 3. S.Saraswathi, "Design of an onlineexpert system for career guidance", Department of Computer Science, LIST Laboratory,Faculty of ScienceandTechnology.

[14]. Yulius Lie, Bens Pardamean, " Information System Model of Succession Planning and Career Path", Information Management and Technology(ICIMTech), International Conference on. IEEE, 2016. 15. Python Introduction,

[15]. .https://www.w3schools.com/python/python_intro.as p