Career Recommendation System

[1] Shivam Mathwad, [2] Rushikesh Pede, [3] Aniket Chaudhari, [4] Prashant Patil, [5] Mr. Nikhil Dhavase

[1][2][3][4] Department of Information Technology, MMCOE, Pune, India
[5] Assistant Professor, Department of Information Technology, MMCOE, Pune, India

Abstract---- recently, more and more people have begun to re-evaluate their career decisions and change careers at a later stage in life. A study conducted by GTI Media survey found that 18% of students say that they regret their choice of degree and 1,805 respondents cite a lack of initial research as the main cause of their disappointment. Also many people are confused as of which career path to choose. This can be prevented by proper counselling of young teenagers before they begin their graduate studies. In India, there are 350 million students, the biggest student population in the world. So for them to find a suitable career we need at least 1.4 million counsellors to maintain a globally acceptable student-to-school-counsellor ratio. But the number of counsellors is not the only issue there is, more often than not counsellors charge upwards of 2500-4000 for career counselling, which may not be financially viable to everyone. Thus, there is a need to develop a scientific career counselling software to tackle this issue.

We aim to create an app for this purpose. Here, in the proposed system, we evaluate the aptitude and personality of a person based on the user's academic level using carefully curated psychometric and aptitude tests. The user will have to give personality test and all aptitude tests in the app to get their career recommendation. The user test scores (personality & aptitude) will be fed into a machine learning algorithm, which will then generate a model to predict career streams based on the user's scores. Thus, the recommendation will be very close to accurate as all significant data related to user ability, skills and personality will be taken into consideration. Users will be given in-depth analysis of their test results and career recommendation.

Keywords— Machine learning, Restful APIs, Big5 personality test, Recommender System, Career prediction, Android app

I. INTRODUCTION

Traditional career counselling is done at a career centre wherein the person has to give many aptitude tests, which are then scored manually. These scores are then reviewed by a career counsellor who then assigns you a career stream. This whole process is long, arduous and error-prone. Our system automates this process, resulting in a better error-free career recommendation system, since our recommendation algorithm also takes into account the user's personality, which is a very important factor when deciding someone's career. Aptitude & Academic/Technical skills alone are not the only deciding factor when it comes to a career decision, Personality and Aptitude both are important. Another advantage of our system is that the user can give his/her tests anytime on the app and receive a career recommendation through app and email, all from the comfort of their own home. This app is mostly directed towards 9th to 12th standard students, since deciding a career stream is very hard at a young age. The user will first have to create an account on our app and then all the tests will be available to them.

The Personality and Aptitude tests are taken from standardized and legitimate sources, and scoring of these tests is also done using their guidelines. The aptitude tests are taken from the Tamanna Aptitude test by CBSE for 9th to 12th standard students(Website link: https://ncert.nic.in/tamanna/). As a collaborative work, piloting of the aptitude test was done by the CBSE with 17,500 students studying in classes IX and X through its affiliated schools across different parts of the country. The app contains various aptitude tests from Tamanna such as: Numerical, Verbal, Perceptual, Abstract Reasoning and Spatial aptitude tests, all of which measure different dimensions of the user's abilities. The personality test is a standard Big 5 personality test taken from IPIP(International Personality Item Pool) to ensure that the test is up to the mark (Website link: https://ipip.ori.org/New_IPIP-50-item-scale.htm). The Big 5 Personality test measures five broad dimensions to describe the human personality, temperament and psyche. They are: Openness to experience, Conscientiousness, Extraversion, Agreeableness & Neuroticism, also sometimes called OCEAN. This system has various applications such as: it can be used by employers to automate the recruitment process in an organization, it can be used for student career counselling, and it can also be used for psychological analysis to determine individual personality traits like leadership, charisma, creativity etc. The intuitive & attractive UI and the unique app test-taking experience will surely attract customers to try our career recommendation software.

II. LITERATURE SURVEY

A Recommendation System, in simple language, is a classification and information filtering system which

shows the user(s) material and information tailored around the profile, history and current data of the user(s).

In [1], the author talks about a job recommendation system. Here machine learning (Collaborative filtering & Vector space model) and data mining techniques were applied to a RESTful web server application that bridges the gap between the frontend(android app) and the Backend(MongoDB instance) using APIs. The data communicated through APIs is fed into the database and the Recommendation System uses that data to synthesize the results.

In [2], the recommendation system is created using the concept of AIS(Artificial Immune System). Artificial Immune System is a novel computational intelligence technique inspired by immunology that has appeared in the recent few years and takes inspiration from the immune system in order to develop new computational mechanisms to solve problems in a broad range of domain areas. This paper presents a problem-oriented approach to design an immunizing solution for job recommendation problem.

In [3], the author uses the data of existing students, where the personalities, aptitude and general information of the student is mapped with their careers. The created Intuitive Career System uses a variety of questions that students have to answer to test their aptitude as well as questions on student's background. The answers are then entered into the created model to predict the career that matches the students aptitude and personality.

In [7], this website gives very helpful information on Recommender Systems in practice. It talks about the various types of recommender systems and the different algorithms needed to implement it.

III.   METHODOLOGY

A) Feature selection -

The Recommendation System needs various features to be able to recommend a career stream. Here, in our system, we recommend 3 career streams to user. For this purpose, 3 different datasets and models are created. Each of these models recommend a career stream based on either personality or aptitude scores.

O_score	C_score	E_score	A_score	Job_Category1
0.593	0.29	0.403	0.48	Literature
0.713	0.3	0.35	0.507	BF
0.763	0.3	0.763	0.633	Literature

Fig 3.1 Dataset - personality1

O_score	C_score	E_score	A_score	Job_Category2
0.593	0.29	0.403	0.48	Engineering
0.713	0.3	0.35	0.507	Literature
0.763	0.3	0.763	0.633	MM

Fig 3.2 Dataset – personality2

Numerical Aptitude	Spatial Aptitude	Perceptual Aptitude	Abstract Reasoning	Verbal Reasoning	Job Category
0.75	0.5	0.75	0.75	0.25	0
0.5	0.25	0.5	0.75	0.5	1
0.5	0.25	0.5	0.5	0.75	2
0.5	0.75	0.25	0.5	0.25	3

Fig 3.3 Aptitude dataset

The personality dataset contains personality trait scores as features. They are: O_score, C_score, E_score and A_score. The aptitude dataset contains all aptitude test types and their scores as features. They are: Numerical Aptitude, Spatial Aptitude, Perceptual Aptitude, Abstract Reasoning and Verbal Reasoning. The final predicted column for both the datasets is: Job_Category.

B) Algorithm -

We used K-nearest neighbours algorithm in this project which predicts the user's career stream based on the test score input. The k-nearest neighbors (KNN) algorithm is a supervised machine learning algorithm that can be used to solve both classification and regression problems. The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. The KNN algorithm hinges on this assumption being true enough for the algorithm to be useful. KNN captures the idea of similarity (sometimes called distance, proximity, or closeness) with some mathematics - calculating the distance between points on a graph.
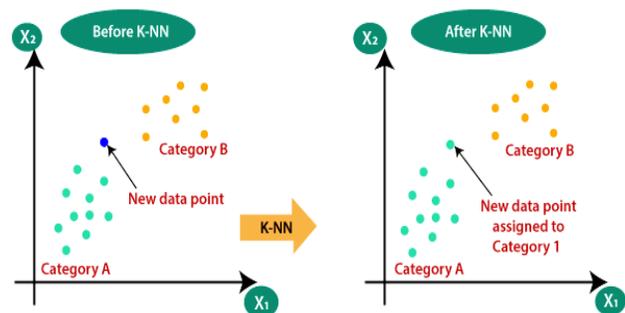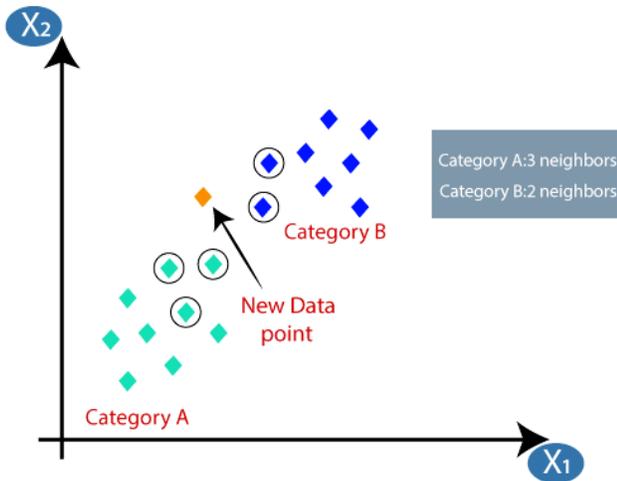
Fig 3.4 KNN Working

KNN algorithm –

1. Load the data

2. Initialize K to the chosen number of neighbors

3. For each example in the data

3.1 Calculate the distance between the query example and the current example from the data.

3.2 Add the distance and the index of the example to an ordered collection

4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances

5. Pick the first K entries from the sorted collection

6. Get the labels of the selected K entries

7. If regression, return the mean of the K labels

8. If classification, return the mode of the K labels

After the aptitude test scores are calculated, before feeding them into the recommendation model, the scores are converted into 3 categories: High(H), Medium(M) and Low(L) corresponding to 0.75, 0.5, 0.25 respectively in the dataset, for ease in calculation. Personality trait scores are not changed and put as it is in the dataset. Neuroticism personality trait score is calculated but never put in the dataset, as it is required in all the career streams and hence can't be used as a distinguishing factor to recommend a career stream.

IV.   IMPLEMENTATION
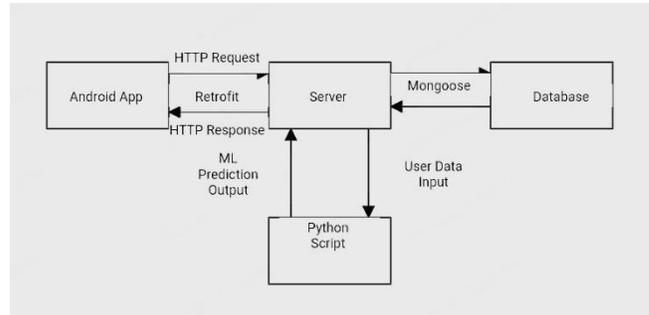
Fig IV.1 System Architecture

This system is based on Client-Server model, client being the user of android app. It contains 3 important modules/parts, they are:

Frontend -

1) Android app is the frontend of our system.

2) Here, Java programming language is used to code Native Android apps.

3) The user has to first create an account on this app, and then he/she can access the tests followed by detailed results and career recommendation.

4) Retrofit library is used to make HTTP requests to server. GET/POST requests are made to the server and an object is given back in response.

5) Glide library is used for it's efficient image loading and caching capability.

6) Android.util.Base64 & Javax.crypto libraries are used for hashing of user passwords.

Backend -

1) Node.js is used for server side programming i.e in Backend of our project.

2) The backend Node.js and ML Python code are going to be deployed on a cloud server.

3) User test scores are given as input to our machine learning algorithm, and it in turn gives a response in form of a career prediction.

4) Express package is used to create an Express server and nodemailer package is used for all mailing services.

5) Mongoose is a npm package, which is a MongoDB object modeling tool designed to work in an asynchronous environment. Mongoose basically helps and greatly simplifies MongoDB database connection with Nodejs server. It also helps in performing database operations like Create, Read, Update, Delete(CRUD) easily on any collection of the database.

Database -

1) MongoDB database is used in our system. It is a NoSQL database.

2) All Test Data(questions and answers of each test) and User Data(user email, password, username, user test results) are stored in MongoDB Atlas.

3) MongoDB Atlas is a cloud database i.e any device located anywhere can access the database anytime.

4) Various types of CRUD queries are used for performing specific operations on collections.

After creating an account on the app, the user receives a signup mail. Now, all the tests are available to them. After completing the checklist of all the tests required for the career recommendation, the user receives a career recommendation in the form of maximum 3 career streams from our predefined list of career streams. In-depth information about these career streams is given to user such as: description of that career stream, jobs available in that stream, best universities that the students should aim for if they choose to go into that field, and the skills required to be successful in that field. After career recommendation, an email is sent with all those details to the user, so that he/she can view the results anytime. In case, the user forgets his password, he can reset it through the "Forgot Password" option on the login page, after which a mail regarding the same will be received by that user.

V. RESULTS

The app was tested by 30+ people who created an account on our app, and gave their aptitude & personality tests and received their career recommendation. KNN algorithm was chosen for this use-case, as it is a supervised classification algorithm and it gave the highest accuracy when compared to other supervised classification algorithms. 3 different datasets were created to predict 3 different career streams, and as such 3 different KNN models were generated. The KNN model trained on the personality_1 dataset had an accuracy of 75.67%, whereas personality_2 KNN model had an accuracy of 72.23%. The KNN model trained on the aptitude dataset had an accuracy of 79.78%. The dataset contains 1200+ records. This dataset was created with the help of a career counsellor, who helped us in making a marking scheme which will map the given aptitude and personality test score ranges to our predefined list of career streams. Lots of research also went into making this dataset. The Recommendation System results were mapped and found

to be satisfactory.

REST API Client (Postman) was setup to test the calls and the Recommendation System. The modules were tested again with more calls and data and successful records were recorded. The whole workflow is impeccable and works seamlessly. The average call response time recorded with MongoDB was 120ms whereas average call response time for SQL and Hibernate Framework was 900ms. All the calls were tested on the Postman REST Client and are now shifted to the Android Interface. The Retrofit Library in android enables the application to make calls from the main thread itself which is not allowed by normal HTTP calls. Frontend can make API calls and receive them without any occurrence of Exception or the Application undergoing a forced stop. The app is functional 24x7 as it is deployed on Heroku server, and can be accessed by anyone anywhere. The Heroku server is of Hobby dyno type. It is hosted in the United States and will continue to run for another 2 months free of cost. As more and more people start using our app, we intend to use their test scores and recommendations to increase the length of the dataset and the efficiency of the model.

VI. CONCLUSION

The proposed framework for Career Recommendation System is aimed at bridging the gap between students and career counsellors. Here, we have developed a software tool to evaluate the aptitude and personality of a person based on his/her academic level using carefully curated personality and aptitude tests. This tool will help you determine your aptitude and personality traits, and will eventually help you in choosing your own career path. The application has the potential to stand financially in the market but needs to be improved a bit and then presented to Investors looking to invest in such projects. Further studies into the topic can help enhance the precision of the system.

REFERENCES

[1] Harsh Jain, Misha Kakkar, "Job Recommendation System based on Machine Learning and Data Mining Techniques using RESTful API and Android IDE", IEEE 9th International Conference on Cloud Computing, Data Science & Engineering, 2019

[2] Shaha T. Al-Otaibi, Mourad Ykhlef, "An Artificial Immune System for Job Recommendation", IEEE International Work Conference on Bio-inspired Intelligence (IWOBI), 2014

[3] Rucha Rangnekar, Khyati Suratwala, Sanjana Krishna, Dr. Sudhir Dhage, "Career Prediction Model

Using Data Mining And Linear Classification", IEEE Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018

[4] M.Fatih Akay, Ebru Cetin, Imdat Yarim, Ozge Bozkurt, M.Mikail Ozciloglu, "Using Machine Learning Techniques to Earlier Predict Student's Performance", IEEE 2018

[5] S.K.Pushpa, T.N.Manjunath, T.V.Mrunal, Amartya Singh, C Suhas, "Career Prediction Model Using Data Mining And Linear Classification", IEEE 2018

[6] Anupam Varshney, Shalini Puri, "A Survey on human personality identification on the basis of Handwriting using ANN ", IEEE 2017

[7] https://towardsdatascience.com/recommender-systems-in-practice-cef9033bb23a.