International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

Student Career Guidance System using Machine Learning

Aniket Surve[1], Amit Singh[2], Shivam Tiwari[3]

[1,2,3]*U.G. Students, Department of Information Technology, TSEC College, Mumbai, Maharashtra, India*

Abstract - *Information Technology has a lot to offer and the ones pursuing it often are without someone who can guide them in choosing the right career. Due to the poor teacher to student ratio, students miss out on quality guidance and have to oversee their decisions by themselves. An intelligent student career guidance system that could help them overcome these problems was the need of the hour. The existing systems looked only at the academic performances as the sole criteria to guide the students on what career would be most suitable for them. The system we propose looks for various insights from the collected data (a hypothetical dataset) that can explain what are all the factors (including academic performances) that could have had an impact over the student's career path. It then uses Logistic Regression to predict the most suitable career for anyone who uses it.*

Key Words: Student Career Guidance, Machine Learning, Logistic Regression, Web Technologies, MOOCs

1.INTRODUCTION

Career guidance is the guidance given to individuals to help them acquire the knowledge, information, skills, and experience necessary to identify career options, and narrow them down to make one career decision. This career decision then results in their social, financial and emotional well-being throughout.

A report published in 2019 by the Government of India showed concerning results about the teacher to student ratio in the country. The ratio was as high as 24 students to one teacher, a number twice or even higher than that of countries like Canada, Russia and Sweden [1]. The students are not necessarily enrolled in the wrong undergraduate programs, it is just that they have not been given the right guidance that can help them choose the career paths that are not a mismatch for them [2]. An automated system that could help the students in choosing the right career path after analyzing their interests, academic performances and involvement in extracurricular

We developed a website that could help the student find the right career for him based on the data that the student provides while filling the self-assessment form, the Massive Open Online Courses (MOOCs) [4] to complete and the certifications to be done. The technologies used for web development were Node.js, Express.js and MongoDB database. Exploratory Data Analysis was done using Python to find helpful insights and the models were built using Sci-kit learn library. Logistic Regression was chosen for the backend processing of the input received.

1.1 Related Work

Models, research papers and several guidance systems have previously been implemented and discussed. These were the ones that talked about replacing the actual education system wherein a student would be given their career paths based on their academic performances. The GPAs were the direct metric and none of the personal skills or other capabilities was considered. The newer ones did consider the latter two but did not make use of any historical data [3].

There are also other monitoring systems [5] that focus on providing one to one guidance to the students who have not been able to smoothly transition into the career paths that they have chosen. These systems are designed in such a way that for each unique individual there was a list of mentors that they could choose from. However, we have proposed an automated self-monitoring and self-assessing system wherein the individual, if very determined, could program his progress.

There have been discussions made over the impact of MOOCs [4]. We have provided MOOCs and certification recommendations based on what the career path the individual gets as the result after taking the self-assessment test.

2. Proposed Methodology

The entire process can be divided into two halves - the frontend process that saw us develop the website which could provide the user with an eye-pleasing interface to work with. The technologies used were, as aforementioned, Node.js, Express.js for the servers, MongoDB to store the sources for the certifications and MOOC recommendations.

The front end starts with the main page having a button that would take the individual to the self-assessment form. The student, after clicking, can go and take the self-assessment test wherein questions would be asked to him about his academic results, how he rates himself in a few things, what he wants in his life and how he sees himself in the next few years. The answers to these would then serve to us as the features for the Logistic Regression model that is built in the backend.

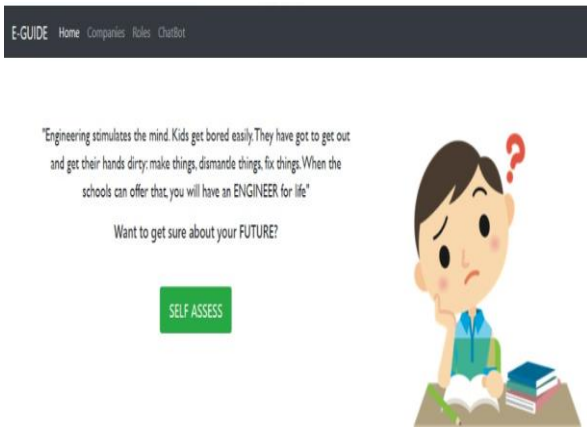

Fig 1: The homepage with the self-assess button

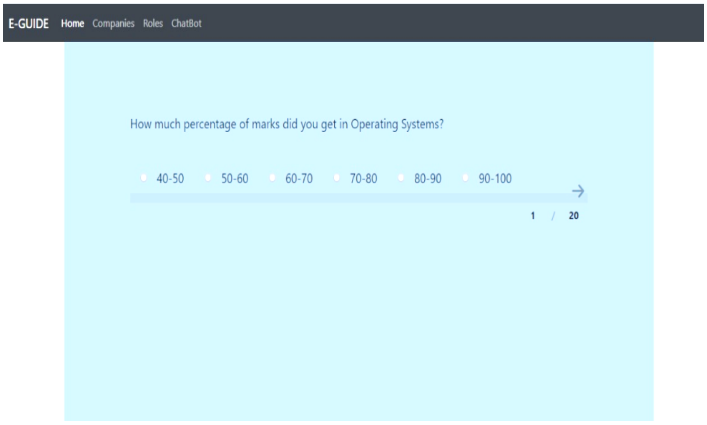

Fig 2: The self-assessment form

With the data from the self-assessment form, the Logistic Regression model is run over it and a role prediction is made. Once a role is predicted, the frontend presents the user with the sources for the certifications and MOOCs by fetching it from the MongoDB database.

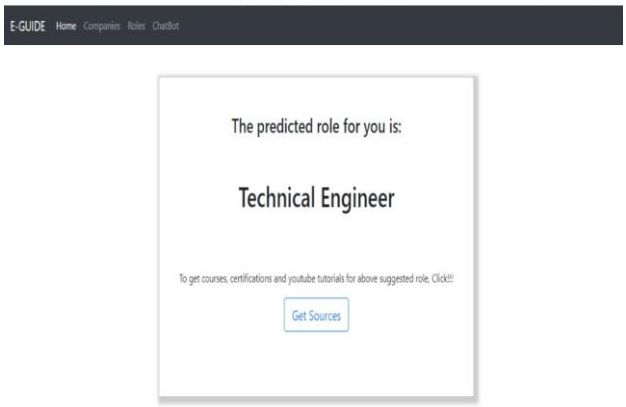

Fig 3: The career path prediction

The backend process started with data acquisition. We used a dataset obtained from GitHub user KLGLUG which had features that we identified as the best parameters for our problem. The dataset gave equal preferences to academic performances through features like Academic percentage in Operating Systems, Academic percentage in Computer Networks and many others, self-assessment by the individual through features like Public Speaking Points, can work for long hours? among many others and finally the extracurricular through features like Talent Tests Taken, Olympiads and Hackathons. The dataset also included features that took care of the individual's aspirations as an engineer and whether the student wished to study further or work before attaining a Master's degree

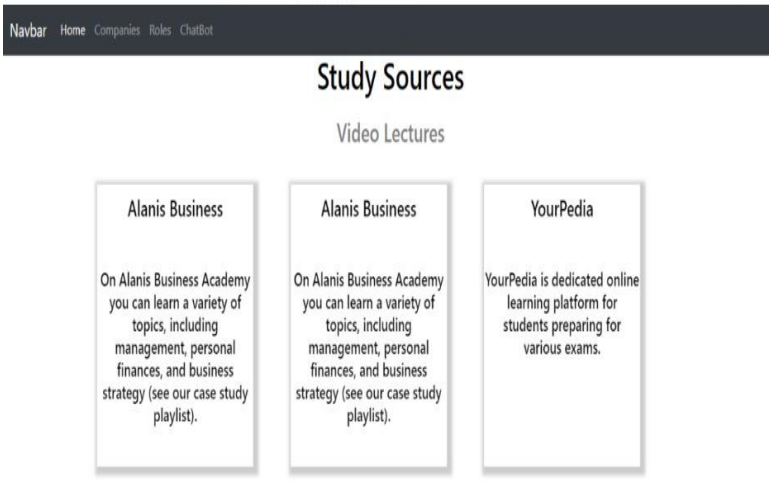

Fig 4: The sources for MOOCs relevant with the role predicted

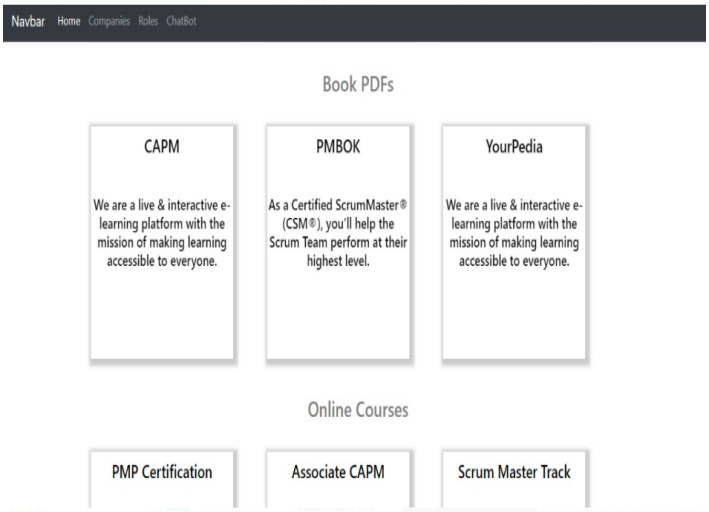

Fig 5: Some other Books and Certifications

The dataset was put under the preprocessing process with features that didn't help much removed from the dataset. There were initially too many similar roles predicted for different individuals while our focus has been on developing a model that could predict a career-path rather than a role. So, we clustered similar roles and put

them under one name to make sure there are no more than 8 technical career paths.

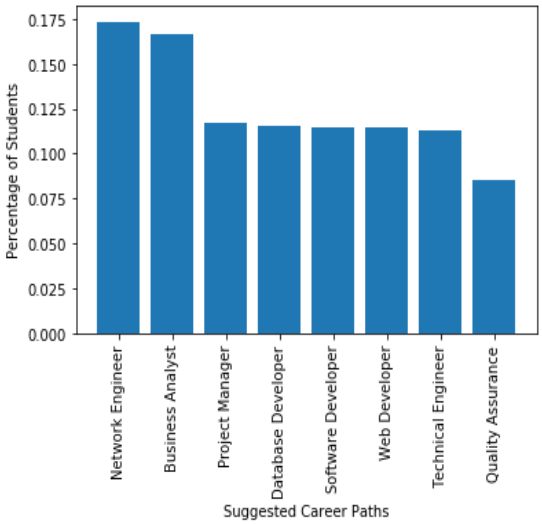

Fig 6: The career paths

Once the clustering was done, we started with Label encoding the categorical features. Since the Logistic Regression algorithm only works with numeric values, the categorical ones had to be label encoded using the Sci-kit Learn's LabelEncoder. LabelEncoder encodes labels with a value between 0 and n_classes-1 where n is the number of distinct labels. There were 20+ features to be encoded.

The data was then split into training and test sets. 33% of the data was reserved for testing while the rest was used for training the model. The train-test split was done again by using Python's Sci-kit Learn module. Multinomial Logistic Regression was fit using the training module and hyperparameters like *C, penalty* and *solver* were tuned to find the best fit model.

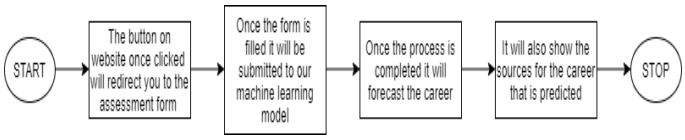

Fig 7: The Flowchart depicting of our System

Multinomial Logistic Regression is Logistic Regression but for more than two classes. For it, we could have either used the unsupervised technique i.e. SoftMax Regression (generalized Logistic Regression for more than two classes) or the supervised Logistic Regression. Logistic Regression calculates the odds ratio for the classes, thus providing users with probabilities for each class for each data point. An example of Multinomial Logistic Regression would be finding out whether a loan application should be rejected, accepted with normal terms or accepted with special terms.

The major advantage of using Logistic Regression over other classical Machine Learning techniques is down to its interpretability and providing output in terms of probabilities thus handing the user the ability of changing the threshold. In our example, it is best suited because the users would then be provided with the probabilities of different career paths they would be best suited for, thus handing them the power of choosing what they would think is best for them.

3. Result and Discussion

The Logistic Regression model after hyperparameter tuning and model tuning came up with an accuracy of 74%. Better models can be developed using advanced machine learning classification algorithms such as Decision Trees, Gradient Boosting and Support Vector Machines. However, with a large number of features used in the model, it would require very good expertise to come up with solutions better and more optimized.

The database of the sources can always be upgraded and improved upon. There will always be newer and better lectures, certifications and what not to add. The list of the companies that the individuals can target for the career path they get as results has all the necessary details on the website. It would help the user know what he can expect after using the system.

4. CONCLUSIONS

The system developed has been an upgrade over the previous ones as it uses data to make decisions, unlike the ones that took decisions based on some rule-based tests. The data used is mostly self-assessment of the user, which can have some bias and hence we also suggest a system that can rather than asking to fill a form, would in some way test the skills of the individual and then rate them accordingly by itself.

One more improvement that could possibly be done is having some kind of verification system for the academic results inserted by the individual. There has to be a check on what the individual puts in as his scores as wrong data would always mislead the individual. The other way to deal with it would again be having an online test on the website that would need the user answer questions from the subjects that are considered while asking for academic percentages.

Overall, the system does have space for further improvements but is also a better system than a system that would guide the individual based on the scores from academics without considering the interpersonal skills and other capabilities.

5. REFERENCES

[1] Higher Education in India Records Alarming Student-Teacher Ratio

[2] Too many graduates are mismatched to their jobs. What's going wrong?

[3] Automated System for matching scientific students to their appropriate career pathway based on Science Process Skill Model - by Mohammed Abdellah Alimam, Hamid Seghiouer, Mohammed Amine Alimam, Mohammad Cherkaoui.

[4] The Application of Career-Guiding MOOC Course in Undergraduate Colleges by Yuanyan Yang

[5] Implementation of Mentoring System in College for Smooth Transition to Work by Juhyun Jeon, Jaeeung Lee