

Personal Career Recommendation System

Dr. A Syed Mustafa¹, Prof. Anusha KV², Deepak N³, Goutham C⁴, Karthika Hiremath⁵, Lokesh SB⁶

Professor and Head of Department, Department of Information Science and Engineering¹

Assistant Professor, Department of Information Science and Engineering²

Students, Department of Information Science and Engineering^{3,4,5,6}

HKBK College of Engineering, Bangalore, Karnataka, India

Abstract: *As students are going through their academics and pursuing their interested courses, it is very important for them to assess their capabilities and identify their interests so that they will get to know in which career area their interests and capabilities are going to put them. This will help them in improving their performance and motivating their interests so that they will be directed towards their targeted career and get settled in that. Also, recruiters while recruiting candidates after assessing them in all different aspects, these kinds of career recommender systems help them in deciding which job role the candidate should be kept in based on his/her performance and other evaluations. This paper mainly concentrates on the career area prediction of computer science domain candidates.*

Keywords: Carrer, SVM, XGBoost, Machine Learning

I. INTRODUCTION

Nowadays, students are often facing a dilemma in deciding to choose a career in their life. There are several factors that influenced students when choosing their career path as their Personal aptitudes, educational achievement, and their environment. Upon completing their first degree or undergraduate students at the university, students are normally starting to consider a career path that may suit their skill and potential the best. However, many students made the wrong decision in selecting their career due to the lack of experience, help, and advice from friends and relatives, parents and lecturers, or career counseling. It is necessary to deliver career guidance in several ways like courses, training, and seminars that offer group experiences in future career planning and group or individual counseling activities. The factor that may cause students not to be successful in their careers is the wrong choice of a job that suits them. It requires a decision-making process at an early stage. So this system gives a recommendation for students about their careers based on their academic result and their abilities.

II. ALGORITHMS

2.1 Support Vector Machine

Support Vector Machine is an acronym. It is a supervised machine-learning approach that is typically applied to both classification and regression-type issues. Various classification issues are where this is primarily applied. The standard algorithmic step is to plot each data point in an n-dimensional space, where n is the number of features and a feature's value is the value of a specific coordinate. The following step is to classify by obtaining the hyper-plane that sharply divides the two classes. Practically speaking, kernels are used to implement SVM algorithms. There are three different types of SVM, and the linear SVM hyperplane is calculated or discovered by using linear algebra to the problem. The realization is that SVM can be rephrased by using the inner product of observations.

2.2 XG BOOST

Extreme Gradient Boosting is referred to as XGBoost. The implementation of gradient boosting algorithms is called XGBoost. It is accessible in a variety of tools mats, including tools, a library, etc. It primarily concentrates on computational efficiency and model performance. It significantly cuts down on the amount of time and significantly improves the model. Its implementation includes recent additions like regularisation in addition to capabilities seen in sci-kit-learn and R versions. Gradient boosting using both L1 and L2 type regularizations is referred to as regularised gradient boosting. The following are the key benefits that the algorithm's implementation offers: Automatic handling of

missing values with sparse aware implementation, block structure to facilitate parallel tree construction, and ongoing training to support further enhancing a model that has already been fitted on the new data.

2.3 Decision Tree

A popular and straight-for e-learning categorization challenge is the use of decision trees. For several sophisticated algorithms, like bagging, gradient boosting, and random forest, decision trees provided the fundamental building blocks. The improved form of the previously stated XG Boost method this broad decision tree CART, C4.5, C5, and ID3 are the three most popular decision trees. In the event that the variable is numerical, a node represents a split on the input variable (X). An output variable (y) that is essential for prediction is present in the leaf, also known as the tree's terminal nodes. Selecting a root node is the first step in the usual decision tree process. Before the split, figure out each node's information gain or entropy. All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

III. IMPLEMENTATION

3.1 Data Collection

One of the biggest and most crucial jobs of every machine learning project is data collection. as a result the input data is fed to the algorithms. Therefore, the accuracy and efficiency of the algorithms depend on how well the data is collected and how accurate it is. The result will be the same as the data. Numerous factors, including academic performance in multiple disciplines, specialties, programming, analytical skills, memory, relationship status, interests in sports, contests, hackathons, workshops, and books, among others, are needed to predict a student's career. All of these characteristics are taken into account since they are crucial in determining how far a student will advance in a certain vocational field. Many methods exist for gathering data. Some information is gathered from employees of various organizations, LinkedIn, and Colleges

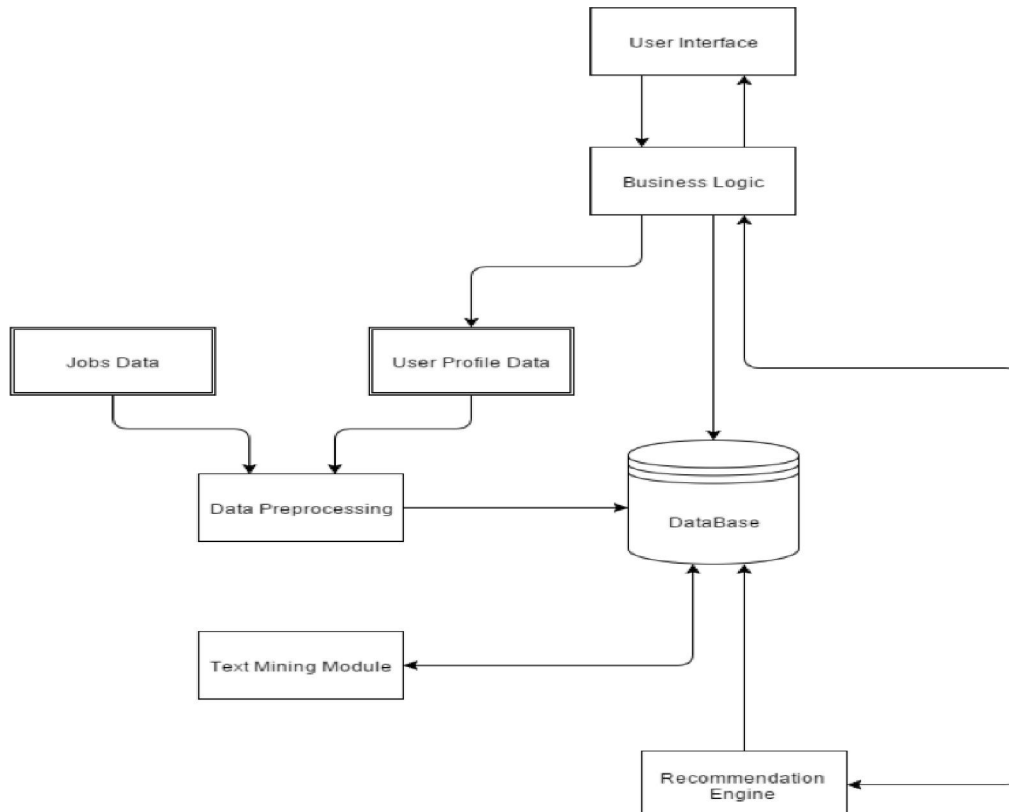
3.2 Data Pre-Processing

Making the data meaningful is a crucial effort that goes hand in hand with gathering the data. There may be many null values, invalid data values, and undesirable data in the data obtained through various methods because it will be in a disorganized manner. removing all of this data and substituting them with more accurate or suitable data. The fundamental processes in preprocessing data include identifying null and missing data, eliminating it, and replacing it with some predefined alternate values. Even acquired data could have entirely useless values. It might not be in the precise shape or manner in which it was intended. To make the meaning of the data intelligible and helpful for future processing, all such cases must be validated and replaced with alternative values. The storage of data must be organized.

3.3 One Hot Encoding

OneHot Encoding is a method for providing categorical values found in the collected data to machines by converting them to numerical or other ordinal formats. Improving prediction outcomes via learning algorithms. Categorical values are transformed using the OneHot encoding technique into a format that is most suitable for feeding into different machine learning algorithms. Nearly any machine learning algorithm is compatible with this algorithm. Only the a few algorithms, like random forest, effectively handle categorical values. In these circumstances, one hot encoding is not necessary. Although the OneHot encoding process may appear challenging, most contemporary machine learning algorithms handle that. This article clearly explains the procedure: In a data set, for example, if there are yes and no values, integers respectively 1 and 0.

IV. ARCHITECTURE DIAGRAM



V. LITERATURE SURVEY

5.1 Personal Factor

Personal factors including student behavior, such as feelings, activities, as well as the finances of the pupils, are important components of understudy academic performance. age and sexual orientation, which impact how they perform. The most often employed parameters for prediction are age and gender since they are considered internal causes of variability and are simple to define and assess. The analysts looked at how frequently psychometric factors might affect how well students performed.

5.2 Academic Factors

Academic factors are descriptors used to describe variables that explain a student's performance on the academic track at a university. Given its enormous influence on the future of training, the cumulative grade point average (CGPA) is the primary factor that has been used most frequently student author took into account the student GPA. The most significant factors in forecasting a student's future academic performance, according to the authors, are their past academic performance and their parents' educational backgrounds. Others investigated how past academic success affected how well pupils will do going forward.

5.3 Financial Factors

Financial aspects Related elements imply that the parents' financial capacity to support their children's education and guide their future careers. A few specialists looked at the relationship between educational programming, parental instruction, and salary.

5.4 Family Factors

Family variables have an impact on an individual's educational background, ability to help their children with their education, and ability to create an environment that is conducive to learning. Results showed that regardless of whether students' school has an impact on students' performance, parents' efforts have a crucial role in predicting grades. The

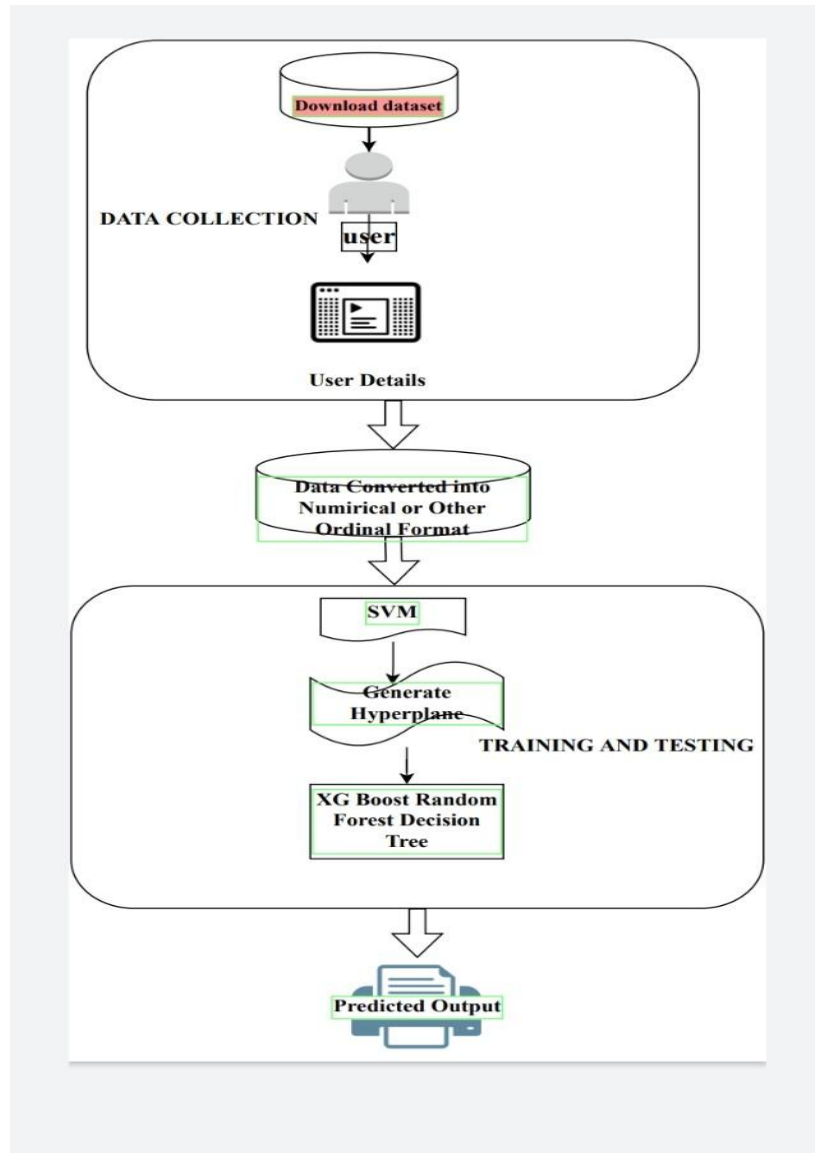
most important factors in predicting an understudy's future scholastic presentation in the art hour were determined to be their former academic performance and their parents' teaching background. Studies often examine how parents have an impact. education background, income, and the impact of students' interpretation on academic success.

5.5 Institution Factor

The factors that affect this classification have to do with the academic program and resources that the foundation provides for the highest academic performance of its trainees. The authors looked at how psychometric factors frequently affect how candidates present themselves.

JOURNAL NAME & YEAR	AUTHOR	TITLE& METHOD	REMARKS
Hindawi.2022.	Qingwan, Linye	Career Recommendations for students based on Deep Learning and Machine Learning. A hybrid CNN model is proposed for employment recommendation of college students.	By improving the Activation function, pooling strategy, and loss function in the algorithm, the quality of model prediction is greatly improved.
International Journal of Advanced Research in Computer and Communication Engineering. 2021.	Sushma Koushik N, Chandana M S, Lavanya V, Suhas Y, HarshithaV	Educational Career Recommendation System Using Machine Learning. The problems of cold start, trust, and privacy are solved in this approach	The recommendation system is built in Python because it is easy and efficient to put into effect algorithms on exclusive operating systems.
Multidisciplinary Digital Publishing Institute. 2020.	MinNie, Zhaohui Xiong Ruiyang Zhong, Wei Deng and Guowu Yang	Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students Proposed a Prototypical cluster center generation approach to use the priori information from each college	A novel regularization Item was introduced by To bridge the gap between the real-world examples and prototypical cluster centers.
Research gate.2017	Bharat Patel, Varun Kakuste, Magdalini Eirinaki	A career path Recommendation on framework A novel career path framework for personalized job and skills Recommend youngs, focusing on students and young professionals is presented.	The system can run on cloud infrastructure using apache spark or Hadoop cluster which can handle more number of jobs description and more users. To increase the system capability at any time, new machines can be added to the existing cluster without affecting the running application.
International Journal of Health Science Journal of Health Sciences. 2022.	Dr. Sadasivam R, Paramasaivam S, Prakash raj N, Saravanam M	Student carrer prediction Student suggestion was proposed to help student career prediction in concluding there abilities in which they are solid and frail.	RF contains various choice threes different subsets of the given datasets and taken the normal to work on the precision of that datasets. It tends to be utilized for both Arrangements and Regression issue in ML

VI. METHODOLOGY



The system was created using Python. The system produces after receiving input from the datasets. As a result The steps that make up the system-building process are done in order.

Obtaining the dataset

Cleaning the Dataset

Selecting for feature dataset

Build a model

Predict outcomes using the model

VII. RESULT

All three algorithms are tested and trained on the data, and SVM topped all others in terms of accuracy, scoring 90.3 percent, followed by XG Boost at 88.33 percent. As SVM All subsequent data predictions are picked to be followed by SVM because it provided the highest accuracy. Thus, a web application is created to provide the student's input parameters and the end A prediction is produced and shown. SVM is indeed the background method in use, and fresh predictions are continuously added to the dataset to increase accuracy.

VIII. CONCLUSION

We looked at how college students choose their careers based on their professional expertise, behavior consistency, and other associated behaviors. The study has also provided a number of crucial insights for enhancing the model.

We have suggested a prototype approach for creating cluster centers that takes advantage of the past knowledge from each college. In order to bridge the gap between real-world instances and a prototypical number of clusters, we introduced a unique normalization item, which is motivated by the cluster assumption that samples in the same cluster should have the same label. Multiple experiments' findings show that our method is superior to other methods for predicting professional choice.

REFERENCES

- [1]. <https://towardsdatascience.com/building-a-job-recommender-for-non-technical-business-roles-via-nlp-and-machine-learning-626c4039931e>
- [2]. https://www.researchgate.net/publication/358791179_Career_Recommendation_for_College_Students_Based_on_Deep_Learning_and_Machine_Learning
- [3]. https://www.researchgate.net/publication/269272063_Career_path_recommendation_system_for_UiTM_Perlis_students_using_fuzzy_logic
- [4]. <https://ieeexplore.ieee.org/document/6869553>
- [5]. <https://www.semanticscholar.org/paper/Student-Career-Prediction-Using-Decision-Tree-and-VidyaShreeram-Muthukumaravel/28490efaca5b4cd1c215fa79dec03f14f58c17c4>
- [6]. <https://ieeexplore.ieee.org/document/9137992> <https://www.ijraset.com/research-paper/career-recommendation-using-ann>
- [7]. P.KaviPriya, "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering.
- [8]. Ali Daud, Naif Radi Aljohani, "Predicting Student Performance using Advanced Learning Analytics", 2017 International World Wide Web Conference Committee (IW3C2).
- [9]. Nikita Gorad, Ishani Zalte, "Career Counselling Using Data Mining", International Journal of Innovative Research in Computer and Communication Engineering.