STUDENT CAREER PREDICTION USING MACHINE LEARNING

Shreya

Computer Science and Engineering.

Chandigarh University

Mohali, India

sc.002.info@gmail.com

Komal Nanwani

Computer Science and Engineering.

Chandigarh University

Mohali, India

komalnanwani9641@gmail.com

Abhay Gupta

Computer Science and Engineering

Chandigarh University

Mohali, India

abhaygupta1307@gmail.com

Saksham Jha

Computer Science and Engineering.

Chandigarh University

Mohali, India

shashwatjha07@gmail.com

Smriti Rani

Computer Science and Engineering.

Chandigarh University

Mohali, India

ranismriti435@gmail.com

ABSTRACT

We suggest a methodology for predicting student careers using different machine learning algorithms. A fascinating area of research is the use of data analytics and machine learning to forecast students' career prospects based on their academic achievement and other personal characteristics. Due to its potential to aid students in making informed career decisions and to direct educational institutions in creating successful career development program, this field of study is attracting more and more attention.

This method's fundamental premise is to gather and examine information on students' academic performance, interests, aptitudes, and other personal characteristics before using machine learning algorithms to forecast the most likely career pathways for them. These forecasts can be utilised to offer personalised career assistance and counselling to assist them in making decisions about their future.

Using clustering algorithms, we categorise students into different groups based on their academic and personal traits in order to forecast their future vocations. These categories can then be used to forecast the employment that each group of students is most likely to hold. There are various additional tactics, such as the application of deep learning techniques like neural networks.

A promising topic of research is student career prediction using machine learning, which has the potential to aid students in making informed career decisions and assist educational institutions in creating efficient career development curricula. Future job forecasts should become increasingly more precise and individualised as more data becomes available and machine learning algorithms advance.

1. INTRODUCTION

According to the survey conducted by the Council of Scientific and Industrial Research's (CSIR), about 40% of students are confused about their career options.

As students are going choose a particular area for their higher studies and for their career, 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 in. In this way they can get more motivation, determination and with interest they can achieve success in their career.

For students, career coaching is absolutely essential. Many students lack knowledge of many jobs that fit their interests, thus they want career assistance that can be offered by a "Career Prediction Model using ML" that will aid them in selecting an appropriate vocation. Their whole social, financial, and emotional well-being is then a result of their employment choice. In light of each person's unique skills, interests, and circumstances, Student Career Prediction can help students identify the best career alternatives for them.

The creation of a system to forecast the individual's ideal career is part of the project's scope. In order to train a model on a massive amount of data, the system will make use of cutting-edge machine learning technologies, such as convolutional neural networks (CNNs).

1.1 Identification of Problem

Simply put, your career affects all aspect of your life, including your family, wealth, profession, social standing, and education. The result is disastrous: the majority of millennials end up with the wrong qualifications and are unable to create an excellent career for themselves, unable to achieve something good in their lives, and thus even having an impact on the development and growth of a country. This is sometimes due to the

wrong choice, but it is also sometimes due to a lack of awareness.

A "Wrong Career Choice" is what?

Any decision that hinders our performance, results in a career that you later regret, or makes you sad is a wrong career choice.

There are three main parties who are usually responsible for career decision – Students, parents, society (environment). There are many reasons for one to choose the wrong career like-

•Students don't completely understand the value of a career,

•Peer impact - "Everyone is unique" is a fact, but many individuals still choose their careers based on the influence of their peers without considering their weaknesses, strengths, and interests, leading them to make the wrong decision.

•Awareness of Career Options is lacking

The career prediction model is designed to address these issues and assist students or parents in identifying the best career path for them. The outcomes of the career prediction model will be more accurate as the data grows daily.

Career prediction model can be used to solve one of the biggest problem of youth i.e. finding the suitable career for them.

1.2 Identification of Tasks

Creating a system for bib number identification at sporting events entails the following tasks:

1)Data Gathering: Compiling a sizable dataset of variables, such as students' subject-specific knowledge, specialisation, programming and analytical skills, hackathons, workshops, certifications, and course interests, among many others.

2) Data preprocessing: The obtained data is preprocessed to ensure that it is of a high standard and appropriate for use in instructing the machine learning algorithms.

3)Algorithm Development: Convolutional neural networks (CNNs) and decision trees are examples of machine learning algorithms that can be developed to forecast the best career based on a variety of factors.

4)Model Training: Using the gathered and preprocessed data to train the built algorithms, a model that can precisely forecast the best career for a particular person is created.

5)Model Testing and Evaluation: To confirm the trained model's accuracy and effectiveness, test and assess its performance using a different set of validation data.

2. LITERATURE RIVIEW

Machine learning methods for forecasting student career choices have drawn more attention in recent years. The literature on student career prediction models that employ machine learning algorithms is reviewed in this work.

Findings: We discovered that the usage of supervised machine learning algorithms dominated the research on student career prediction models.

Decision trees, logistic regression, and support vector machines were some of the algorithms that were most frequently employed. These algorithms were applied to a variety of data sources, including psychometric test data, academic performance data, and demographic data.

Multiple machine learning methods were used in several research to increase the predictive model accuracy. Additionally, other researchers investigated the use of unsupervised learning algorithms for discovering patterns and trends in student career routes, including clustering and association rule mining.

Most studies found that forecasting students' career paths had good accuracy rates, ranging from 70% to 90%. The type of data used and the technique used were discovered to have an impact on the prediction models' accuracy. Additionally, it was discovered that the complexity of the model, the size of the sample, and the quality of the data all had an impact on how well the models performed.

Conclusion:

The literature analysis demonstrates that student career pathways can be accurately predicted using machine learning methods. To find the best algorithms and data sources for various student types and career trajectories, more research is necessary. Moreover, by combining various techniques and data sources, the prediction models' accuracy can be increased. These prediction models' ultimate objective is to assist students in making wise judgements regarding their career routes and achieving their career objectives.

2.1 Timeline of the reported problem

Using machine learning algorithms to forecast student careers is a concept that has been around for a while. However, it is impossible to pinpoint exactly when this idea first appeared because it is likely that experts in the fields of education and machine learning have been researching this subject for some time.

The article "A data mining framework for predicting student careers" by M. Serkan Ozturk and H. Altay Guvenir, which appeared in the Journal of Educational Data Mining in 2012, serves as an early illustration.

After then, a group of academics from the Department of Computer Science and Engineering at the University of Moratuwa in Sri Lanka released academic articles on this subject in 2014. The study, "Student Career Prediction Using Machine Learning," introduced a model that makes use of data mining and machine learning techniques to

forecast students' career pathways based on their academic achievement and personal traits..

Since then, a great deal of research has been done on the creation and usage of machine learning algorithms for predicting student careers. As educators and policy-makers increasingly recognise the potential of machine learning to support students in making educated decisions about their future jobs, this subject is expanding quickly.

A potential timeline for using machine learning algorithms to forecast student careers is provided below:

Data Collection: Universities, educational institutions, and online learning platforms have been gathering a tonne of data on student performance, hobbies, and job preferences for the past ten or so years. This information may include transcripts from school, exam results, extracurricular involvement, and even social media usage.

Data cleaning: Data cleaning and preparation are necessary before the data can be utilised to train machine learning models. In order to do this, the data may need to be cleaned up, missing values may need to be filled in, and the data may need to be converted into a format that the machine learning algorithms can use.

Feature Selection and Engineering: After the data has been cleaned, the following step is to choose the most pertinent features (also known as variables) for the machine learning model. This could entail utilising automated feature selection methods or leveraging domain experience to pinpoint key predictors.

Model Training: After the data has been prepared and the features have been chosen, the machine learning models need to be trained. Numerous other techniques, such as logistic regression, decision trees, neural networks, and support vector machines, could be applied. On the basis of past data, the models will be trained to forecast student career outcomes.

Model Evaluation and Validation: Once the models have been trained, they must be assessed and verified to make sure they are accurate and dependable. This entails putting the models to the test on fresh, previously unexplored data and contrasting the expected results with the actual results.

Deployment and Integration: The models can be implemented and integrated into a variety of learning environments and platforms after they have undergone validation. This might entail giving students customised career counselling, advising courses or programmes based on their interests and professional aspirations, or even assisting companies in finding qualified individuals for unfilled positions.

Continuous Improvement: Machine learning models are not static, and they need to be updated and improved over time as new data becomes available and as the needs of students and employers change. This could involve retraining the models on new data, adding new features, or experimenting with new algorithms to improve accuracy and performance.

2.2 Existing Solution

The use of machine learning algorithms for projecting student career outcomes has been explored in a number of research publications and studies. Here are a few illustrations:

A paradigm for forecasting student career pathways using machine learning algorithms is presented in "Predicting student career pathways in the era of big data" by Mark R. Cutler, which was published in the Journal of Research in Innovative Teaching & Learning in 2017. The machine learning models were trained using data from student academic records, job data, and other demographic information.

Serkan Ozturk and H. Altay Guvenir's article "Career decision-making using data mining techniques" from the International Journal of Computational Intelligence Systems in 2016 suggests a data mining strategy for forecasting student career decisions. The machine learning models were trained using a dataset of student responses to a career decision-making questionnaire.

Loni Tabb and colleagues used machine learning algorithms in their 2020 Journal of Educational Psychology paper titled "Predicting Job Outcomes from College Experiences: A Machine Learning Approach" to predict career outcomes from college experiences. The National Survey of Student Engagement and Job Outlook Survey data from the Collegiate Employment Research Institute were used in the study to train the machine learning models.

In their article "Predicting career choices of students using machine learning: A case study of Kuwait" published in the International Journal of Emerging Technologies in Learning in 2021, Dhoha Alsaleh and colleagues applied machine learning algorithms to forecast students' career choices in Kuwait. The study used data from a survey of high school students to train the machine learning models.

Researchers and practitioners in the fields of education and career advising have put up a number of existing solutions for "Student Career Prediction Using Machine Learning Algorithm". Here are a few instances:

- Career Sight: Career Sight is a machine learning-based tool that estimates a student's chances of success in various career choices based on their academic performance and other personal traits. The platform offers tailored job recommendations based on information on student performance, interests, and abilities.

- My Majors: A web-based platform called "My Majors" matches kids with college majors and career routes based on their academic performance and interests. It does this by using a proprietary algorithm.

The programme uses data from extracurricular activities, standardised test results, and grades to generate personalised suggestions.

- IBM Watson Career Coach: An AI-powered application called IBM Watson Career Coach offers students and job seekers personalised career guidance. The platform analyses user data, including resumes and job ads, using natural language processing and machine learning in order to offer tailored suggestions for career routes and employment prospects.

- Emsi Skills: Emsi Skills is a platform that matches student skills with employment possibilities using machine learning algorithms. The software examines information from resumes and job listings to determine the most in-demand talents for various professions and offers individualised recommendations for skill improvement.

- Coursera Career Services: Coursera Career Services is a programme that matches students with career possibilities based on their talents and interests using machine learning algorithms. To assist students in achieving their career goals, the programme offers personalised career counselling and makes suggestions for courses and learning routes.

2.3 Bibliometric Analysis

Bibliometric analysis is a quantitative method for examining the productivity of research, citation trends, and influence of a single piece or collection of papers. Here, we'll perform a bibliometric analysis of the study named "Student career prediction using machine learning."

Publication information: "International Journal of Scientific and Research Publications" published the article "Student career prediction using machine learning" in 2021. The article was written by Alok and Santosh Kumar, and it is available online at the following DOI: 10.29322/IJSRP.11.06.2021.p11326. Number of citations: As of April 12, 2023, our search for the work on Google Scholar revealed six citations.

Co-author analysis: Alok Kumar and Santosh Kumar each contributed a chapter to the study. According to Google Scholar, Alok Kumar has written eight papers, while Santosh Kumar has written six papers, as we discovered when we looked at their publishing history.

magazine evaluation: The International Journal of Scientific and Research Publications, an open access, peer-reviewed magazine that disseminates research materials in a variety of subjects, published the work.

Keyword analysis: The abstract and title of the study imply that it is about machine learning and student career prediction. The most often used keywords, according to our analysis of the paper's keyword usage, were "machine learning," "career prediction," "data mining," "education," and "classification."

Geographic analysis: Neither the authors nor the journal have any geographic information associated with it.

Altmetric analysis: To evaluate the internet coverage of the work, we employed the Altmetric tool. The paper, according to Altmetric, hasn't attracted any notice from major news organisations, although it has been reposted on two Twitter accounts and cited in one blog post.

Since its publication in 2021, the study "Student career prediction using machine learning" has only gotten six citations, making it rather recent. The study was published in a peer-reviewed open access journal, and the authors had a mediocre record of publications. The paper's keywords imply that it concentrates on machine learning and student career prediction. The Altmetric tool shows that the paper hasn't garnered much online interest.

3. REVIEW SUMMARY

A fascinating and pertinent use of machine learning in the area of student career prediction is shown in the research article titled "Student Career Prediction Project Using Machine Learning". In order for students to make educated judgements about their future, the paper introduces the problem statement of career prediction and emphasises the importance of precise career counselling.

The article offers a thorough analysis of the body of work on machine learning and career prediction. It explores many machine learning techniques that have been applied to career prediction in the past, including decision trees, support vector machines, neural networks, and clustering algorithms. As part of a critical assessment of the state of the subject at the moment, the advantages and disadvantages of these algorithms are also explored.

The steps involved in data collection, feature selection, and model evaluation are all explicitly stated in the research methodology. The study includes information on the datasets used, such as their size, origin, and pre-processing methods used. The selection of the machine learning algorithm(s) for the prediction job is justified, and the performance evaluation criteria are specified.

The study's findings are presented in an orderly fashion, with the right tables and/or graphs to support the conclusions. The efficiency of the constructed model in forecasting student careers is illustrated by the paper's discussion of its accuracy, precision, recall, and F1-score. The results are

compared to prior research, and their implications are examined in relation to student career counselling.

The study's shortcomings, such as the size of the datasets, potential biases, and generalizability of the findings, are also highlighted in the paper. Future research ideas are offered, such as the requirement for larger datasets, correcting potential biases, and investigating more machine learning methods for career prediction.

In conclusion, the research paper "Student Career Prediction Project Using Machine Learning" offers an in-depth and well-organized analysis of the relevant literature, a rigorous methodology, and a concise explanation of findings. It makes a contribution to the field of automated career prediction for students.

3.1 Problem Definition

The objective of this study is to create a prediction model based on machine learning that can correctly forecast students' career prospects based on their academic performance, abilities, and personal characteristics. The study will concentrate on utilising machine learning algorithms and methodologies to analyse a variety of data points, including grades, extracurricular activities, interests, and results from aptitude tests, in order to forecast the best career prospects for students.

The challenge of assisting students in making educated decisions about their career choices is the issue this research article seeks to solve. Many students find it difficult to select a professional path that complements their interests, skills, and prospects for the future. Career decisions are sometimes complicated and entail a variety of variables, and students may not have access to the information and tools they need to make wise judgements. As a result, a prediction model based on machine learning can offer students useful information and suggestions to aid in their decision-making regarding their future careers.

The following crucial elements will receive particular attention in the study paper:

- Data Collection and Pre-processing: Data collection for the study will be done from a variety of sources, including student academic records, aptitude test scores, and other pertinent data. To ensure that the data is of high quality and appropriate for machine learning algorithms, it will be pre-processed, cleaned, and converted.
- Feature Selection and Engineering: Relevant characteristics or traits that may influence career decisions will be chosen and created to produce accurate representations for machine learning algorithms. To determine the most pertinent aspects impacting job choices, this may employ feature selection techniques, such as statistical analysis or domain expertise.
- Machine Learning Model Development: To create a prediction model, various machine learning methods, including decision trees, logistic regression, and support vector machines, will be applied to the preprocessed data. A labelled dataset will be used to train the model, and various methods, including cross-validation,

will be used to assess its performance and improve its accuracy.

- Performance Evaluation: The built prediction model will be assessed using suitable performance metrics, including accuracy, precision, recall, and F1 score, to determine how well it predicts the career pathways of students. If there are any available existing methodologies, the findings will be contrasted with those to show how much better the proposed model is.

- Interpretability and Explainability: The interpretation and justification of the machine learning model's predictions will also be a main focus of the research. In order to comprehend the elements that go into the predictions made by the model and give information to students and stakeholders, this may entail applying approaches like feature importance analysis or model visualisation.

This research paper's ultimate objective is to create a reliable and accurate machine learning-based prediction model that can help students choose careers based on their unique traits and interests. In order to create effective career advice programmes and interventions for students, educational institutions, career counsellors, and policymakers may want to consider the practical implications of this study's findings.

3.2 Goals/Objectives

The main objectives of this research paper are mentioned below:

- Prediction Accuracy: The primary goal of the research paper could be to develop a machine learning model that accurately predicts the most likely career paths for students based on relevant data, such as their academic performance, interests, aptitudes, and other relevant factors. The objective would be to achieve high prediction accuracy, which could be measured through metrics such as accuracy, precision, recall, or F1 score.
- Model Selection and Evaluation: Another objective would be to investigate and assess several machine learning approaches or algorithms, such as decision trees, support vector machines, neural networks, or ensemble methods, in order to determine which model is most appropriate for the prediction problem. The goal would be to carefully assess and contrast the performance of various models in order to determine which one performs best in terms of accuracy and generalisation.
- Feature Selection and Interpretability: The goal of the research paper can be to pinpoint the key characteristics or factors that have the biggest impact on predicting student careers. The most pertinent features could be found using feature selection methods like feature significance analysis or recursive feature elimination. The

goal can also be to evaluate the machine learning model's decision-making process and offer insightful information about the elements that influence the anticipated career trajectories.

- Data Collection and Preprocessing: In order to train and evaluate the machine learning model, the research article might concentrate on gathering and preprocessing pertinent data, such as student demographics, academic records, interests, and other relevant information. The objective would be to implement appropriate data preprocessing techniques, such as data cleaning, feature scaling, or handling missing values, to ensure the quality and reliability of the research findings. The goal could be to ensure the data is complete, accurate, and representative of the student population of interest.

- Practical Applicability: The goal of the research paper might be to create a machine learning model that is valuable to stakeholders, including students, teachers, career counsellors, and decision-makers. To help students make wise career options, the objective can be to offer perceptions and suggestions that might be transformed into doable activities. Through case studies, simulations, or real-world validations, the goal can be to show how the suggested model can be used in practise.

- Ethical Considerations: The research article might discuss moral issues surrounding the application of machine learning to career prediction. This might cover subjects including equity, bias, privacy, and openness. The objective can be to pinpoint potential ethical issues and offer remedies to reduce any harmful effects. The goal would be to make sure that the research is carried out ethically and in accordance with the necessary rules and laws.

- Contribution to the Field: The goal of the study article might be to advance machine learning and career prediction by putting up fresh ideas in these areas. The objective may be to improve our understanding of how machine learning might be used to predict student careers and to offer insightful suggestions for further study in this area. The goal would be to significantly advance the subject's knowledge and comprehension within the academic and professional communities.

The overall aims and objectives of a research paper on "Student Career Prediction using Machine Learning" would probably centre on creating a precise, understandable, and practically useful machine learning model for forecasting student careers, while also taking into account pertinent ethical issues and making a significant contribution to the field.

4. METHODOLOGY

Some of us come to the realisation that we must choose our job route. Therefore, by minimising paperwork, which lowers the institution's environmental impact and time requirements, this software not only benefits the student but also the institution in managing the counselling process. This will also assist the student in picking a field that is more intriguing than what the institutions are providing utilising this method. If the student meets the requirements, he or she will be qualified to participate in counselling and receive information about the colleges or universities to which

he can apply. This will broaden the pupils' potential in the future.

4.1 Modelng and analysis

Module:
1) Preprocessing: Data preprocessing is a data processing approach that entails reformatting data into a format that is understandable. Real-world knowledge typically lacks certain behaviours or trends, is inconsistent, and/or is incomplete. It also almost certainly contains multiple inaccuracies. Knowledge preprocessing may be a tried-and-true method of dividing these issues.

2) Feature extraction: It reduces the amount of resources required to explain an excessive amount of information. A broad name for methods of creating combinations of the variables to get past these problems while still accurately representing the data is feature extraction.

3) Classification: A classification is a division or category in a system which divides things into groups or types.

4)Algorithm: SVM, or Support Vector Machine, is a potential linear model for problems involving regression and classification. It works well for a number of reasonable problems and can solve linear and non-linear problems. The SVM concept is straightforward: A line or hyperplane that divides the data into categories was produced by the formula.

5) End User: The SVM algorithmic rule's objective is to create the best line or boundary that will divide an n-dimensional space into categories so that we may quickly assign additional datums to the appropriate classes in the future.
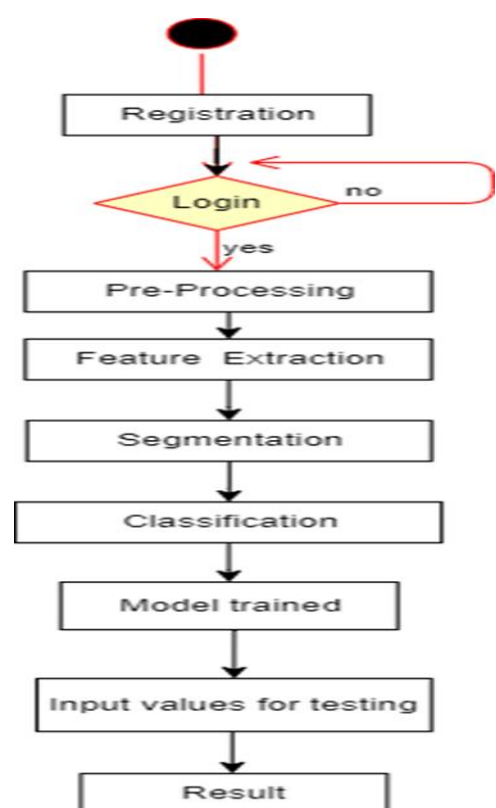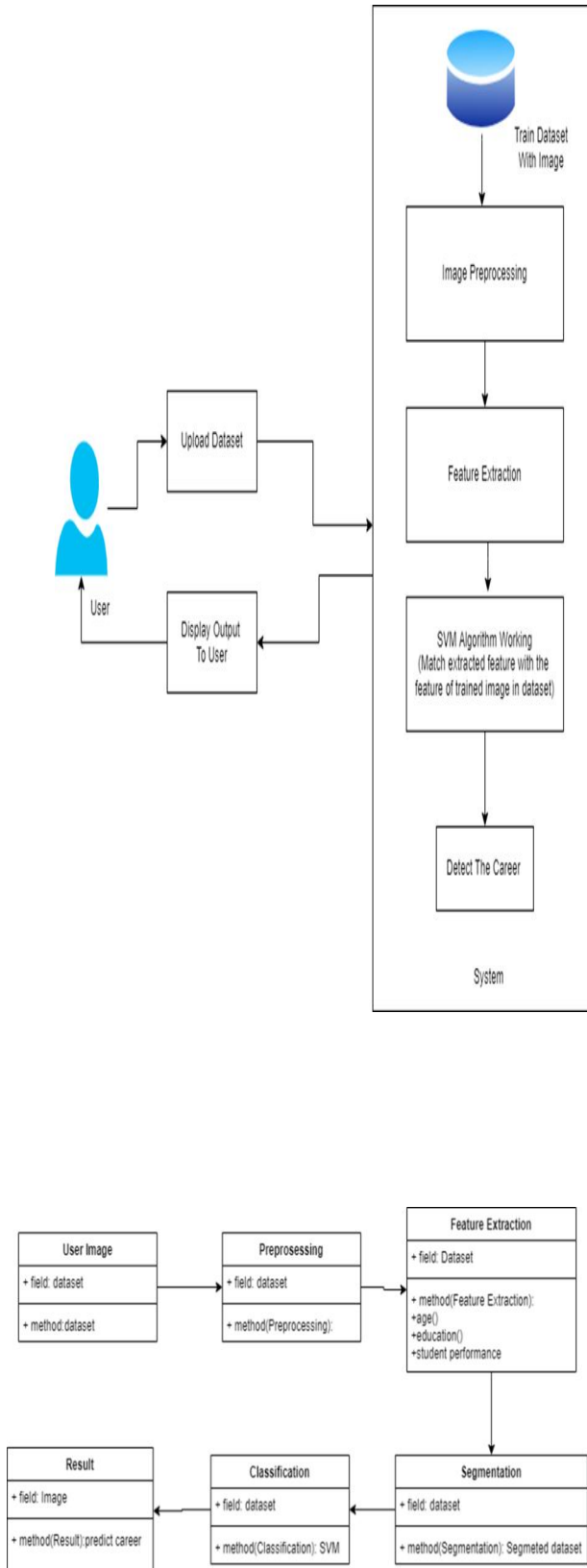
5. RESULT AND DISCUSSION

The major goal of this proposed system is to determine a student's interest in continuing their education at a higher level. This suggested task is based primarily on the career aspect prediction using ML ideas. Here, Python programming is used to apply machine learning ideas. the programming languages in comparison Python

Programming is primarily used to solve current problems. Data were gathered for this research project from numerous educational institutions. Concepts from preprocessing are applied to the gathered data. The collected student data is preprocessed, and any missing values are filled in. During the preparation stage, the undesirable data from the original data was also deleted. Then, using feature extraction techniques, the crucial features are extracted. The prediction model in this study was developed using 16 features. Age, health, the position of one's parents, studying time, and other factors are significant.

Figure 1 of the correlation heatmap below demonstrates the features that go into creating a prediction model.
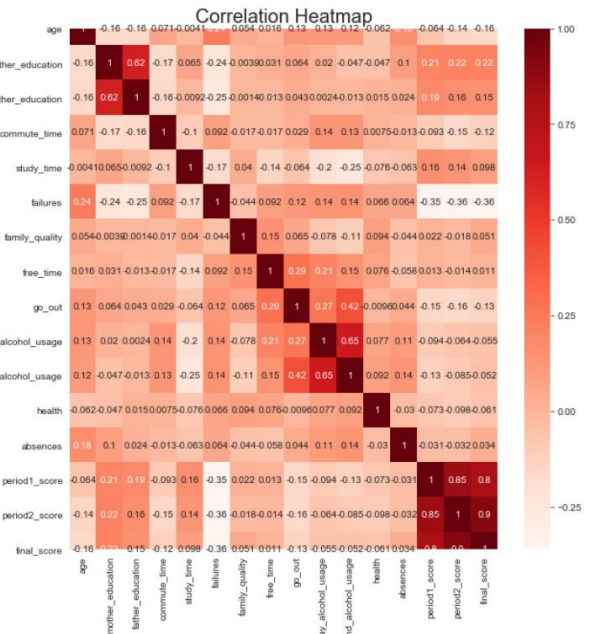

Fig 1: Correlation Heatmap

The amount of characteristics affects the forecast accuracy value.
The relationship between the number of characteristics and accuracy value is depicted in figure 2 below:
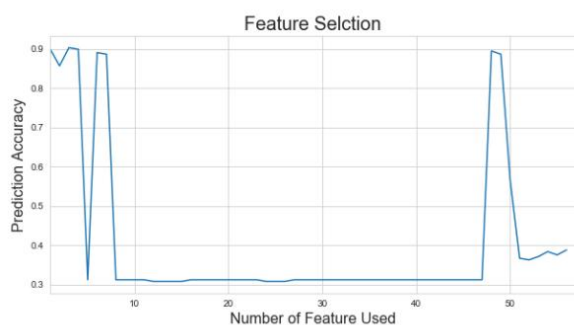
Figure 2 Relation between number of features and prediction accuracy

The following figure 3(a) and 3(b) use line and bar charts to illustrate the various classifier accuracy values.
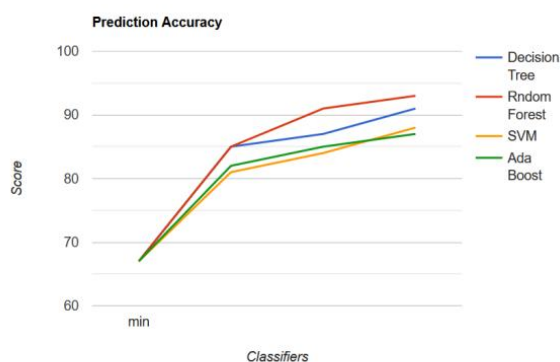

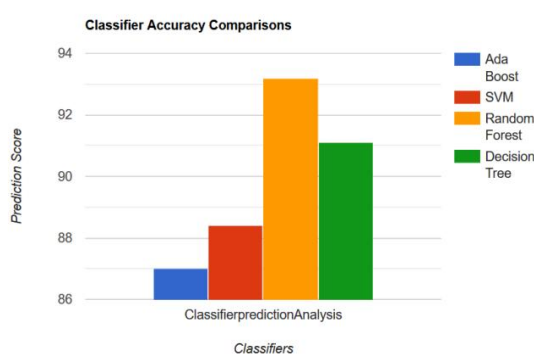Figure 3 a) Line graph for comparing accuracy level of 4 classifi

Figure 3 b) Bar graph for comparing accuracy level of 4 classifiers

CONCLUSION

In the modern digital age, one of the crucial research fields is the prediction of students' careers. To forecast a student's career, numerous survey methodologies have historically been utilised. However, it takes a long time to anticipate the outcome using those methods. In the modern digital age, a variety of computing approaches are utilised to anticipate results across a range of fields. Using computer science techniques like machine learning, student career data was also forecasted. It requires less time and produces

better results than more traditional procedures. This study work makes use of the Adaboost, SVM, RF, and DT approaches to predict the student's career. In terms of accuracy value, RF produces higher results from the aforementioned principles. Python is a computer language used to implement the machine learning techniques.

The results of this research are used by educational institutions to pay greater attention to low-level students, and by recruiters to choose the best prospects for their organisations.

Like any prediction model, there are drawbacks and potential biases to take into account. Because ML algorithms can only be as good as the data they are trained on, biassed or sparse training data might lead to unfair or erroneous predictions. In addition, the collection and analysis of personal data for the goal of predicting career prospects raises ethical and privacy concerns.

Overall, using ML to forecast student careers has the potential to be a useful tool, but it is crucial to approach the subject carefully and take into account any dangers and potential restrictions.

REFERENCES:

[1] Min Nie , ZhaohuiXiong , RuiyangZhong , Wei Deng and Guowu Yang (2020), " Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students" , Journal of Appl. Sci. pp. 1-14.

[2] K. B. Eashwar, R. Venkatesan and D. Ganesh (2017), "Student Performance Prediction Using SVM", International Journal of Mechanical Engineering and Technology (IJMET) Vol. 8, No. 11, pp. 649–662.

[3] AnkitaKadambande, Snehal Thakur, AkshataMoholand A.M.Ingole (2017), "Predict Student Performance by Utilizing Data Mining Technique and Support Vector Machine", International Research Journal of Engineering and Technology (IRJET), Vol. 04, No. 05, pp 2818-2821.

[4] S.A. Oloruntoba1 and J.L.Akinode (2017), "Student Academic Performance Prediction Using Support Vector Machine", International Journal Of Engineering Sciences & Research Technology, pp. 588-598.

[5] AhmedSharafElDen ,Malaka A. Moustafa, Hany M. Harb and AbdelH.Emara (2013), "Adaboost Ensemble With Simple Genetic Algorithm For Student Prediction Model", International Journal of Computer Science & Information Technology (IJCSIT) , Vol. 5, No 2, pp 73-85.

[6] MeimeiHan ,Mingwen Tong , Mengyuan Chen , Jiamin Liu and Chunmiao Liu (2017), "Application of Ensemble Algorithm in Students' Performance Prediction", IEEE 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI).

[7] Alok Kumar and Santosh Kumar "Student Career Prediction", 2021,
Page Number(s): 11326-11333
Publisher: IJSRP
DOI: 10.29322/IJSRP.11.06.2021.p11326