**A PROJECT REPORT ON**

**CAREER GUIDANCE SYSTEM USING ML TECHNIQUES**

**BACHELOR OF TECHNOLOGY IN**

**INFORMATION TECHNOLOGY**

Submitted By

**Yousuf Ali 1604-19-737-106**

**Osman Ali 1604-19-737-109**

**Syed Rafey Ahmed 1604-19-737-113**

****

**DEPARTMENT OF INFORMATION**

**TECHNOLOGY**

**MUFFAKHAM JAH COLLEGE OF ENGINEERING AND TECHNOLOG**

(Affiliated to Osmania University, Hyderabad**)**

**Mount Pleasant, 8-2-249, Road No.3, Banjara Hills, Hyderabad-34**

2022-2023

**CERTIFICATE**

It is certified that the work contained in the project titled “**Career Guidance System using ML Techniques,**” by

Yousuf Ali  (1604-19-737-106)

Osman Ali  (1604-19-737-109)

Syed Rafey Ahmed  (1604-19-737-113)

has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree”

Project Guide Signature of Head

Mr. MD. Asrar Ahmed Dr.MOUSMI AJAY CHAURASIA

Information Technology Dept. Information Technology Dept.

MJ College of Engineering & Tech. MJ College of Engineering & Tech.

Hyderabad-500034 Hyderabad-500034

**Career Guidance System using ML Techniques**

**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 in. 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 the candidates after assessing them in all different aspects, these kind of career recommender systems help them in deciding in 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***: *Student Career Prediction, Decision Tree, Machine Learning, SVM, OneHot Encoder, XGBoost*

**CONTENTS**

1. **INTRODUCTION………………………………………… 1-5**

1.1Introduction………………………………….. 1

1.2Problem Statement…………………………… 5

1.3Objective…………………………………….. 5

1. **LITERATURE SURVEY…………………………………. 6-8**
2. **SYSTEM ANALYSIS…………………………………… 9-14**

3.1 Existing System……………………………… 9

3.2 Problems with Existing System……………… 10

3.3 Proposed System……………………………… 11

**4. PROJECT PLAN…………………………………………. 15**

**5. REFERENCES…………………………………………… 16-17**

1. **INTRODUCTION**

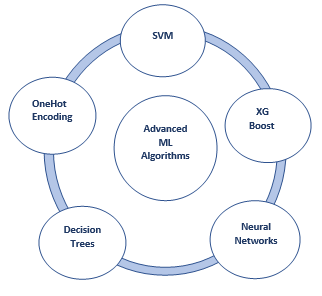
While graduation signifies the end of college many students can be left wondering, 'what's next?' Many students face a difficult time choosing a career path in college, especially in fields like Engineering where there are vast domains to choose from. Thus it is very important for students to assess their capabilities and identify their interests while pursuing their studies so that they are clear about their dreams and the career path leading towards it.

Thus, we have proposed an idea of an **ML-based “Career Guidance System”** which helps the students (esp. In their pre-final and final year) to decide the job role the candidate should undertake based on his/her performance and other evaluations.

**1.1 Introduction**

Competition in today’s society is heavily multiplying day by day. Especially it is too heavy in present day’s technical world. So as to compete and reach the goal students need to be planned and organized from initial stages of their education. So it is very important to constantly evaluate their performance, identify their interests and evaluate how close they are to their goal and asses whether they are in the right path that directs towards their target. This helps them in improving themselves, motivating themselves to a better career path if their capabilities are not up to the mark to reach their goal and pre-evaluate themselves before going to the career peek point.

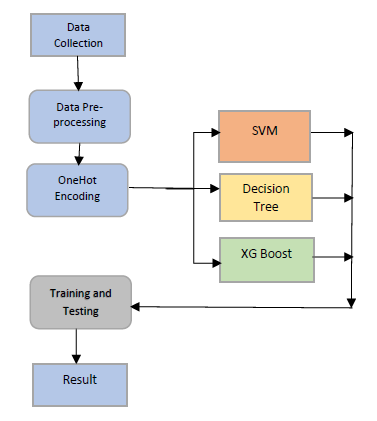
Not only that recruiters while recruiting people into their companies evaluate candidates on different parameters and draw a final conclusion to select an employee or not and if selected, finds a best suited role and career area for him. There are many types of roles like Database administrator, Business Process Analyst, Developer, Testing Manager, Networks Manager, Data scientist and so on. All these roles require some prerequisite knowledge in them to be placed in them. So, recruiters analyze these skills, talents and interests and place the candidate in the right job role suited for them. These kind of prediction systems make their recruitment tasks very easy because as the inputs are given, recommendation is done based on inputs. Already these type of various career recommendation systems and job role recommendation, prediction systems are being used in various third party performance evaluation portals like Co-Cubes, AMCAT. They only take factors like technical abilities and psychometry of students into consideration. These portals asses the students technically and suggest the students and companies job roles suited on their performance. But here various factors including abilities of students in sports, academics and their hobbies, interests, competitions, skills and knowledge are also taken into consideration. Considering all the factors the total number of parameters that were taken into consideration as inputs are 36. And the final job roles are fixed to 15 in number. As the input parameters and final classes of output are large in number typical programming and normal algorithms cannot give the best possible output classification and prediction. So advanced machine learning algorithms like SVM, Random Forest decision tree, OneHot encoding, XG boost are used.



**Figure 1**: Overview of various Advanced Machine Learning Algorithms

Machine Learning is a technique where the machines are trained in such a way that it gains the ability to respond to a particular input or scenario based on the previous inputs it has learnt. Simply it the giving computers the ability to learn by using statistical techniques. Machine learning helps the computers to act without explicitly being programmed. This aims at reducing the human intervention in the machine dependable problems and scenarios. This helps in solving very complex tasks and problems very easily and without involving much human labor. Various applications of machine learning include NLP, classification, prediction, image recognition, medical diagnosis, algorithm building, self-driving cars and much more. In this paper classification and prediction are being done. Let us see what is classification and prediction.

Majority of problems in machine learning can be solved using supervised and unsupervised learning. If the final class labels are previously know and all the other data items are to be assigned with one of the available class labels, then it is call supervised. And if the final output classes and sets are not known and it is done by identifying the similarity between data point and their characteristics and finally they are made into groups based on these characteristics then it is called unsupervised. Classification falls under supervised. Input parameters are given and based on their properties a predefined class label is assigned. There are other alternatives like clustering and regression. Based on the type of problem the apt model is chosen.



**Figure 2**: Process Flow Diagram of proposed system.

**1.2 Problem Statement**

Unavailable career guidance has resulted in many students making wrong career choices and then regretting for the rest of their lives.

**1.3 Objective**

The aim of this project is to help pre-tertiary students make a better career choice, with the following objectives.

1. To explore the problems encountered by the existing manual system.

ii. To design a career guidance system that will improve upon the existing manual/human career guide.

iii. To help young ones get a good understanding of themselves and advise them on career path that best suits them. And also serve as a complementary tool for career guide and counselors.

1. **LITERATURE SURVEY**

Career selection is playing a major role in the life student’s life. Traditionally specialized profession people using questionnaires to identify the important factors affecting career paths. But it is very difficult to predict the career path because of complication of every student aim and dreams. A recent fact provides the suggestion using student’s data based on their behavioral aspects to forecast the career path. Min Nie et al., proposed a novel model known as ACCBOX (Approach Cluster Centers Based On XGBOOST) to forecast student’s choices in their career. The final result clearly states that the current method is better than other predicting methods. In this model uses 13 behavioral data collected from 4000 students [1].

Mining education data is also one of the important tasks in education field. In the beginning days data mining methods are used in education field by using less number of arguments, because low record maintenance in concern institutions. Recently the large volume of data can be stored on student. In India 0.3 % people only move from their PG level to research level. This prediction task concentrates performance of the student’s by using various arguments and the students are classified as low, high and medium type. To execute the above process the authors K. B. Eashwar et al., combines SVM and K-means method. A SVM concept is used for classification purpose and K-mean technique is mainly used for cluster the student’s data [2].

In education domain predicting the student performance level is one of the important tasks. Data mining concepts are used to predict the student’s performance by using various types of tests. AnkitaKadambande et al., uses semantic rules and SVM concepts to do predictions. Semantic rules are used to improves the educational content quality and convey education action to every student. Here the authors helps the students by providing better suggestion and issues recommendations for improving student’s performance level in forthcoming exams. This system will provide the helps to low level and high level students and also to increase the student’s interest about their education. The main purpose of this research work is to increase the quality of learning measures and support the students by forecasting their academic level and help the students [3].

The authors S.A. Oloruntoba et al., identifies the association among initial academic profile level and the last academic profile level. For this research work Federal Polytechnic student database can be used. The initial academic profile is represented as O level. The performance of the students in academic is defined by using GPA (Grade Point Average). Current research work mainly focused on develops a new model for predicting performance of the student with the help of data mining approaches. Preprocessing task is used to delete the unwanted data. Here student’s performance can be predicted by using SVM concept. The result of this classifier is compared with other machine learning concepts like linear regression, KNN and decision tree. The accuracy level of SVM is better than other machine learning concepts [4].

Student’s performance prediction is important in higher educational institutions. The prediction result is used to spot and increase the performance level of the students. Various factors are influenced to improve the performance level. AhmedSharafElDen et al., uses classification concept to increase the quality of advanced learning system. Here the authors uses Adaboost technique with genetic technique is known as Ada-GA to increase the performance of the classifiers. Ada-Ga technique is useful to identify the student’s risk level in earlier manner with large amount of data. This output is used by the tutor to issue the proper advice to the concern students [5].

Student’s data are increased day by day. Among the various prediction methods machine learning concept is one of the outstanding model. Meimei Han et al., proposed a Adaboost model to predict the students level. The result of the Adaboost classifier is compared with other machine learning concepts like neural network, decision tree, SVM and random forest. Initially association policy and correlation study are used to find the characteristics of the model. In next level various prediction models are used to predict the data. Finally compare the accuracy level of the prediction models. In term of accuracy level AdaBoost is higher but the time and cost is high compared with other models. Association rule mining is used to assist the students locate their issues from the origin of the issue to assist them to resolve the issues [6].

Level of student’s performance is one of the major significant values of the all type of educational organizations. To improve the value of the institutions, need to forecast the performance of the student’s. Special type of treatment is needed for low level performer. FaridJauhari, et al., proposes three various boosting approaches to construct the classifier for forecasting the level of student’s. In this research 1UCI dataset is used for developing mode [7].

1. **SYSTEM ANALYSIS**

**3.1 EXISTING SYSTEM :**

Currently, engineering students mainly rely on seniors, mentors, professors, and parents for career guidance. Some may also seek guidance from counselors.There is no unified Online Career Guidance System for carer guidance.In the absence of adult guidance, they often find it difficult to find relevant information about the career they want.Generally students make poorly researched decisions which they have to struggle with.

**3.2 Problems with Existing System :**

● Sometimes Counselors have limited Knowledge to tackle the situation or give a solution to a problem.

● There might be uncertainties. It is not sure or guaranteed for the desired outcome.

● The outcome can be influenced by the counsellor’s personal interests.

● It is difficult to find qualified and trustworthy career counselors and some people cannot afford costs for counselling.

**3.3 PROPOSED SYSTEM**

**Our Solution :**

Our proposed idea of the Career recommendation system takes into consideration the students’ abilities in academics, technicality, hobbies, interests, psychometry, skills, and knowledge.

Taking these inputs machine learning algorithms will be applied and suitable job roles will be suggested.

ML Algorithms used:

● Support Vector Machine(SVM)

● XG Boost

● Decision Tree

**Implementation**

**1 Data Collection:**

Collection of data is one of the major and most important tasks of any machine learning projects. Because the input we feed to the algorithms is data. So, the algorithms efficiency and accuracy depends upon the correctness and quality of data collected. So as the data same will be the output. For student career prediction many parameters are required like students academic scores in various subjects, specializations, programming and analytical capabilities, memory, personal details like relationship, interests, sports, competitions, hackathons, workshops, certifications, books interested and many more. As all these factors play vital role in deciding student’s progress towards a career area, all these are taken in to consideration. Data is collected in many ways. Some data is collected from employees working in different organizations, some amount of data is collected through LinkedIn api, some amount of data is randomly generated and other from college alumni database. Totally nearly 20 thousand records with 36 columns of data is collected.

**2 Data Pre-processing:**

Collecting the data is one task and making that data useful is another vital task. Data collected from various means will be in an unorganized format and there may be lot of null values, in valid data values and unwanted data. Cleaning all these data and replacing them with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in pre-processing of data. Even data collected may contain completely garbage values. It may not be in exact format or way that is meant to be. All such cases must be verified and replaced with alternate values to make data meaning meaningful and useful for further processing. Data must be kept in a organized format.

**3 OneHot Encoding:**

OneHot Encoding is a technique by which categorial values present in the data collected are converted into numerical or other ordinal format so that they can be provided to machine learning algorithms and get better results of prediction. Simply OneHot encoding transforms categorial values into a form that best fits as input to feed to various machine learning algorithms. This algorithm works fine with almost all machine learning algorithms. Few algorithms like random forest handle categorical values very well. In such cases OneHot encoding is

not required. Process of OneHot encoding may seem difficult but most modern day machine learning algorithms take care of that. The process is easily explained here: For example in a data if there are values like yes and no., integer encoder assigns values to them like 1 and 0.This process can be followed as long as we continue the fixed values for yes as 1 and no as 0. As long as we assign or allocate these fixed numbers to these particular labels this is called as integer encoding. But here consistency is very important because if we invert the encoding later, we should get back the labels correctly from those integer values especially in the case of prediction. Next step is creating a vector for each integer value. Let us suppose this vector is binary and has a

length of 2 for the two possible integer values. The ‘yes’ label encoded as 1 will then be represented with vector [1,1] where the zeroth index is given the value 1.Similarly ‘no’ label encoded as ‘0’ will be represented like [0,0] which represents the first index is represented with value 0.For example [pillow, rat, fight, rat] becomes [0,1,2,1].

This is here imparting an ordinal property to the variable ,

i.e. pillow < ret < fight. As this is ordinally characteristic and is usually not required and desired and so OneHot encoding is required for correct representation of distinct elements of a variable. It makes representation of categorial variables to be more expressive.

**Advantages of proposed system:**

* This system will provide career guidance based on the students skills
* Students do not have to attend career guidance seminars. They get proper guidance by accessing this application online
* This system can be used by many students who are looking for further studies

1. **PROJECT PLAN**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Activity** | **Plan Start (no. of weeks)** | **Plan Duration (no. of weeks)** | **Actual Start**  **(Week no)** | **Actual Duration (no. of weeks)** | **Percent Complete** |
| **Problem Definition** | **1** | **2** | **1** | **1** | **100%** |
| **Abstract** | **3** | **1** | **4** | **1** | **100%** |
| **Literature Survey** | **3** | **8** | **7** | **3** | **100%** |
| **System Analysis** | **9** | **5** | **8** | **3** | **100%** |
| **Review 1** | **16** | **1** | **10** | **1** | **100%** |
| **System Design** | **17** | **2** |  |  |  |
| **Implementation** | **19** | **10** |  |  |  |
| **Testing** | **30** | **2** |  |  |  |
| **Documentation** | **3** | **32** |  |  |  |
| **Review II** | **32** | **1** |  |  |  |

**5**. **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. V. Prasannakumari B. Bavani, S. Nirmala SugirthaRajini, M.S. Josephine (2019), “Heart Disease Prediction System based on Decision Tree Classifier”, Jour of Adv Research in Dynamical & Control Systems, Vol. 11, 10-Special Issue, 2019, Volume 11, Issue 10.

[8] FaridJauhari and Ahmad AfifSupianto (2019), “Building student’s performance decision tree classifier using boosting algorithm”, Indonesian Journal of Electrical Engineering and Computer Science Vol. 14, No. 3, June 20