Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 19, no. 8, August 2023, pp. 536-546
DOI: 10.23940/ijpe.23.08.p6.536546

An Advanced Machine Learning Approach for Student Placement Prediction and Analysis

K. Eswara Rao[a,*], Bala Murali Pydi[b], T. Panduranga Vital[a], P. Annan Naidu[a], U. D. Prasann[a], and T. Ravikumar[a]

[a]*Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali, India*
[b]*Department of Electrical and Electronics Engineering, Aditya Institute of Technology and Management, Tekkali, India*

Abstract

As there are job opportunities worldwide, the graduates who are being produced in large numbers from various backgrounds are constantly trying to get them. Moreover, the management of graduate colleges gives proper training to the students to get those opportunities. Every student has their skills, unique creative outlook, studying, and good academic skills that help them get placed in various companies and also have a chance to get reputed positions, but most of the graduates are still failing to get the opportunity because they cannot find what skills to acquire. For this reason, in this paper, we gathered information from students who have finished their courses at different colleges. Collected information by communication and asked them about their academics, performance, families, skills, personal information, habits, etc., and what prevented them from taking the opportunity. Then, we made a dataset with all the factors that affected a student's career and used that to create a model with synthetic data. Student Placement Prediction can also benefit colleges and universities by providing valuable observations of student career outcomes. By understanding the factors influencing student job placement, colleges can conduct services and programs to help their students be better prepared for their careers. Accuracy and precision were used to evaluate the eXtreme Gradient Boost (XGBoost) machine learning model's performance compared to standard classification techniques. According to the results, the proposed algorithm is vastly superior to the alternatives.

Keywords: career prediction; traditional machine learning classifiers; crossfold Validation; boosting methods; XG Boost

1. Introduction

Placement is one of the most important parts of doing well in any graduate or postgraduate course. Every student wants to get hired by top MNCs so they can reach their goals and objectives. Colleges and universities are getting better at finding jobs for their students by giving them better tools and training through training and placement cells. One of the most crucial aspects of a modern education is helping students find solutions where they have failed to get success in their careers. However, there are so many options that it can be hard to choose the right one. Before choosing a career, it's very important to think about how interested you are, how good you are at something, how fast you think it will grow, and how long it will last. Many students have bad grades because they chose a career without thinking about their skills. Choosing the wrong professional path might cost both time and money in the long run. It has also been seen that psychological factors affect choosing the right career path. In particular, students should learn how to understand themselves so they can take part in making decisions about their jobs. But because everyone's goals and ideas are different, it is hard for students to know where they want to go after graduation. From an empirical point of view, on the other hand, you can find out a lot about a student's inner interests and where they plan to go after they graduate by looking at their behavior at school. This makes the behavior of students an important part of their career planning. Modern colleges are increasingly adding sensing, processing, and communication capabilities to their physical facilities as a result of the expansion of information technology. This means that the college information system can trace anything a student does in real time on campus. These kinds of behavioral data can show how each student's habits, skills, preferences, and state of mind are different [1]. Students can also use mining skills to learn more about themselves as they continue to collect this kind of information [2]. ML techniques have been used in recent research to look at the differences and patterns in how different types of college graduates perform [3]. Assumptions can also be used in the real world. For example, we can set up a set of teaching

* Corresponding author.
E-mail address: eswarkoppala@gmail.com

methods based on the real-life situations of college graduates and also make sure that students can get a better education based on their own situations. In this way, students can make career decisions that are relevant to their own life, which helps solve the problem of having difficulty obtaining employment [4]. It is hard to predict a student's career path based on their behavior. Even though the studies that already exist use a variety of machine learning (ML) algorithms, there are problems with low accuracy and models that don't work well. So, based on the theory of social influence [5], we look at the relationship between each student's career prediction and the behavior of students who are similar to him or her. There are a lot of different kinds of data on student behavior, like when they wake up, when they go to sleep, what they do with their friends, what they eat, what sports they play, etc. [6]. Machine learning approaches [7] are used to predict a student's career path based on how they are active. Before we train our model, we put students into groups based on their personal information which affects how we can predict their future careers, get this idea from [8]. In this paper, we specifically look at the behavior of 1020 students who have graduated from college and are now doing their own jobs.

Classification Approaches

In the domain of predicting student careers, classification is the most widely used ML method for modeling and predicting people's behaviors based on their attributes. Correctly assigning class labels to instances with known suitable factors or attribute values but unknown class values is the purpose of classification [9]. Modeling and predicting student actions from a student dataset using ML techniques is challenging since different classifiers produce inconsistent results. As a result, researchers conduct extensive tests with various models based on ML approaches, using student dataset. To evaluate the efficacy of an ML-based prediction model, they employ ten of the most common traditional classification techniques [10], including ZeroR, Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Adaptive Boosting (AdaBoost), and Logistic Regression (LR) classifiers. These methods are commonly used to foretell a student's future in the workforce, therefore they opted to apply them. There is a brief explanation of the ML classifiers: in the background and related work section. After the ML classifier-based framework has been developed, the performance of each model is evaluated by conducting experiments on real student datasets that contain information about the actions taken by each student in addition to relevant context data. Graduates from a variety of academic disciplines contributed to these datasets. Accuracy, Precision, recall, F-Score, receiver operating characteristic (ROC) value, and error rate are some of the metrics we use to evaluate these classifier-based models in our analysis.

In conclusion, our contributions are as follows:

- At college events, we observe students and record their actions. The consistency of college students' behavior is described using behavioral indices, and the behavioral differences between graduates of alumni and students pursuing their degree are examined. Finally, a feature that takes unlabeled data into account calculates how the data has changed over the years.
- We conduct a comprehensive evaluation on a synthetic dataset comprising over 1,200 alumni to ensure the credibility of our results, we conduct numerous experiments. To develop a model the behavioral information of students belonging to different categories, ML models are proposed. We validate the efficacy of our career prediction method by conducting experiments on a student behavior dataset.
- We used synthetic student datasets of alumni to test each classifier-based prediction model on unseen contextual test cases.

The paper is organized as follows: Section 2 provides an overview of the topic of career prediction for students, as well as an overview of the topic of categorization learning approaches in the literature. In Part 3, we outline the proposed structure and methods. Section 4 provides a description of the synthetic dataset and parameter setup. In Section 5, we discuss the outcomes of our experiments testing the suggested advanced machine learning approach on the student dataset. Section 6 concludes the paper and discusses its potential applications.

2. Literature Survey

Based on the results of an objective test, a computerized career counseling system can provide predictions about the best department for a given individual [11]. The objective of this study is to create a system for career coaching that incorporates a method for predicting a person's compatibility with a certain profession [12]. This work mimics the most popular supervised ML algorithms (DT, RF, KNN, LR, NB, Gradient Boost Tree (GBT), Multi-Linear Perceptron (MLP), and SVM) used to predict academic performance [13]. Using an ML-based hybrid recommendation system, the authors [14] propose a method for personalizing a study plan. The primary goal of this study is to provide a detailed account of the variables utilized by the National Technological University and the implementation of several automated learning approaches in order to get the metrics that allow exhibiting the best algorithm among those evaluated [15]. This study provides a model for predicting students' academic success based on how they're feeling emotionally, which has the potential to serve as an early warning system, assisting both teachers in monitoring student progress and students in conducting their own assessments of their own performance [16]. LR model, K-NN regression model, DT regression model, XG Boost regression model, GBT regression model, light GBM regression model, and random tree classifier model are all proposed to tackle the student placement prediction problem in this paper [17]. In an effort to improve the institute's training and placement activity, this study aimed to create an automatic system for predicting students' placement at the beginning of their academic careers [18]. In this research, they introduce Harmony, a Deep Learning (DL) driven ML cluster scheduler that optimizes performance by scheduling training jobs to run in isolation from one another (i.e., training completion time). Harmony is founded on a well-thought-out deep reinforcement learning (DRL) framework, enhanced with reward modeling [19]. The experts use ML and Neural Networks to determine if a user meets the criteria for a certain job posting by analyzing their responses to a series of " hyperparameters" [20]. Enhance the quality of educational processes by suggesting a decision support system that provides accurate analysis, improved decision assistance, and reporting and planning capacity to help decision-makers. In this study, we use ML classification approaches to predict the career a graduating engineering student could pursue, with the overarching goal of identifying the elements that influence students' decisions about their futures. The academic performance, athletics, and extracurricular activities of each student are analyzed in order to provide job recommendations using an ML algorithm [21]. In this study [22], they used the machine learning (ML) approach eXtreme Gradient Boosting (XG Boost) to predict students' major selection using a real-world dataset collected at a single university. Precision, recall rate, and F1 value data demonstrate that XG Boost can accurately predict students' job choices at the 89.1%, 85.4%, and 0.872 levels, respectively. C3-IoC (https://c3-ioc.co.uk) is a system introduced in this paper [23] that uses artificial intelligence (AI) to guide students toward appropriate IT career choices based on their individual qualifications. In this study [24], they use two datasets gathered from two Portuguese secondary schools to offer a data mining approach to discover important characteristics and predict student performance. At last, we develop and compare the effectiveness of classification models with SVM, NB, and MLP roots.

Data mining (DM) and ML algorithms were used to describe a model for determining student placement. "Data mining" referred to the process of using ML algorithms to sift through massive datasets in search of relevant information. The authors also made advantage of what they believed to be the superior education data mining technology. It's useful for analyzing whether a student was hired after their campus placement and for predicting how well they did there. Predictions were made using multiple linear regressions, the ML algorithm J48, NB, RF, and Random Tree from the WEKA tool. Higher education organizations can tailor their instruction to the findings [25]. The primary objective of this [26] study was to describe the applications of ML in educational settings, including how institutions can forecast students' performance and what factors should be considered when doing so. The research also evaluates the accuracy of predictions made by various ML systems. By considering a student's learning style, motivation, interest, concentration level, family background, personality type, ability to process knowledge, and testing method, the article concludes that more specific and reliable predictions about how students will do can be made. Class attitude, psychological measures, and student code metrics are used to predict programming class placement and skill ranking [27] [28]. Their qualitative study examines job placement issues of international graduate students going home, migrating abroad, or staying in the US [29]. Their data mining project investigates engineering placement student performance [30], Binary logistic regression predicted student campus placement [31], and Psychology-assisted ML to predict academic success [32]. In the past few years, a lot of work has been done to provide a full review of ML in student career prediction. ML points to possible research areas in the education field. It has specific rules that other fields don't have. ML has been used to predict how well students will do in education life by using different methods and techniques. In Table 1, some of the jobs that different researchers do are listed:

Table 1. Career Prediction using Various Approaches

S. no	***Algorithms used***	***Objective***	***Problem Type***	***Compared models***	***Performance metrics***	***Ref***
1	Voting Ensemble	Student Career Choice	Prediction	NB, K-Star, SVM	Accuracy, Training Time	[33]
2	ACCBOX	Career Choice	Prediction	DT, SVM, RF, LR, XG Boost	F1 Score, Micro Precision, Micro Recall	[34]
3	Knn	student career	prediction	CF	Accuracy, Error	[35]
4	GB, DL, MLP, RF, LR	student career	Feature Selection	-	AUC, Accuracy, Kappa, RMSE	[36]
5	XGBoost	career choice	Feature Selection	DT, SVM, RF, GBM	F1 Score, Recall,	[22]
6	CDMSE	CDD	-	-	PCA	[37]
7	ML Models	Career Prediction	Prediction	DT, RF, SVM, Adaboost	Accuracy	[38]
8	ML Models	Career Area	Prediction	SVM, RF, DT, XG boost	Accuracy	[39]
9	ML Models	Student Placement	Prediction	SVM, GNB, KNN, RF, DT, SGD, LR, NN	Precision, Recall, F1-Score, Support, Heatmap	[40]
10	KNN	Placement Predication	EDM	RF	Accuracy	[41]

The placement process is crucial to the success of any undergraduate. Students often don't know what their weaknesses and strengths are or what kind of job they would be good at [42]. Because they don't know enough, these students often end up in jobs that aren't right for them, which makes their job profiles less than ideal. This model can help these students figure out which job profiles are best for them and what they need to do to get those jobs. Students also have trouble because they don't know how much certain factors affect their placement. The results of many examinations designed to assess student knowledge reveal their academic performance. These results tell us important things about a student's learning field and how interested he or she is in that field.

3. Methodology

The problems of career prediction for students cannot yet be fully addressed by the conventional ML approach. By mining the connections between students, we can get a better grasp on the issue of what bonds a group of students together to become alumni. Before discussing how these vast differences between students affect ML approaches, let's examine where they have failed. Luckily, we can get this ahead-of-time data from each alumni Information bridges allow us to easily link up with similar students. Classification models can be used to model student activity on the basis of synthetic data, which is useful for making predictions about student careers. In the "ML classifiers: literature study" section, we briefly discuss nine of the most common classification techniques that are widely used for predictive purposes and that we use in our analysis. These methods include LR, DT, NB, RF, KNN, SVM, GT Boost, AdaBoost, and Categorical Boosting (CAT Boost). We also consider the proposed eXtreme Gradient Boosting (XGBoost) classifier model in our research as shown in Figure 1, which is gaining traction in the field of machine learning. Here, we outline the XGBOOST method that we've developed.

XG Boost is a popular ML algorithm that excels at a wide range of classification tasks. Thankfully, it excels at solving prediction problems on massive datasets. Here we suggest XG Boost (Algorithm 1) to increase its efficiency in this regard. Like other boosting techniques, XG Boost iteratively constructs an integration model from a classification and regression tree (CART). When constructing trees, each data value starts with the same weight, which is then adjusted based on the results of the analysis. The first round of data values is used to inform the creation of a new classifying model that retains and expands upon the previous round's findings, and so on, until a reliable classifier has been established.

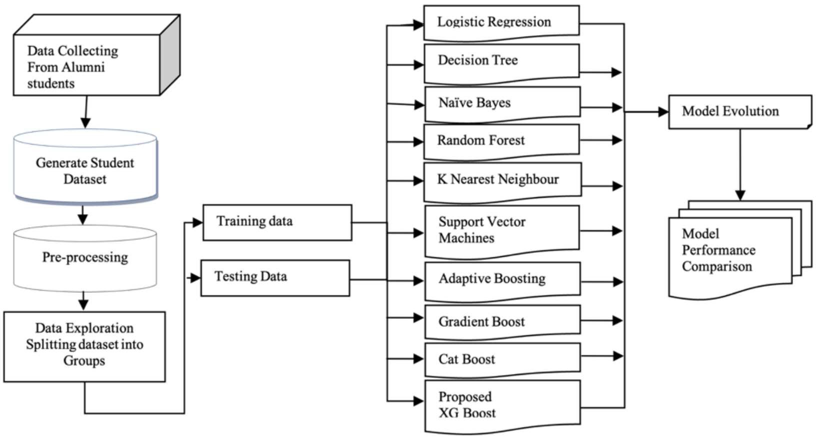


Figure 1. Proposed Research Model

Assume we have a dataset with m features and n examples, and that $D = (x_i, y_i)\ (|D| = n, x_i \in R^m, y_i \in R)$ where D is a classification model of a tree as per $\widehat{y_i} = \emptyset\,(x_i) = \sum_{k=1}^{n} f_k\,(x_i), f_k \in F$ that employs K additive functions to predict the output.

Algorithm 1. Proposed XG Boost Algorithm for Career Prediction.

1: synthetic student data D
2: $D = (x_i, y_i)\ (|D| = n, x_i \in R^m, y_i \in R)$
3: $\widehat{y_i} = \emptyset\,(x_i) = \sum_{k=1}^{n} f_k\,(x_i), f_k \in F$
Where f_k is tree structure
Define as gap $F = \{f(x) = w_q(x)\}\ (q: R^m \rightarrow T, w \in R^T)$
where T is number of leaves
q leaf weight
4: Calculate the prediction score of leaves
$L(\emptyset) = \sum_i l(\hat{y}_i, y_i) + \sum_i \partial\,(f_k)$ where $\partial(f_k)$ is measure the differences
5: Train the model $L^t = \sum_{i=1}^{n} l\left(\hat{y}_i, \hat{y}_i^{\,t-1}\right) + f_t\,(x_i) + \partial(f_k)$ where f_t improves model accuracy
6: Calculate accuracy

4. Dataset Overview and Simulation Setup

In this section, explains how to use different classification techniques to model the student dataset. It has three steps: exploring data sets, processing data and making predictions based on machine learning. Here, we'll talk briefly about each of these steps**.**

4.1 Prepare Dataset

The dataset that was utilized in this paper was produced from former students who had attended a variety of institutions and graduated with degrees after successfully completing courses in a variety of fields and being placed in a variety of companies, move on to pursue higher education, and do nothing. In order to accomplish this goal, a survey Google form was distributed to all of the alumnae of engineering colleges that are situated in the north coastal region of Andhra Pradesh (India) between April 1, 2019, and December 30, 2022, and responses were gathered from 1027 students after careful communication. This dataset is comprised of a variety of student data kinds, each of which plays an important part in a student's mental ability and physical ability. These ability skills should have an impact on a student's career at any point in time. Overall, the data set that is utilized for the purpose of student placement prediction is an essential component in the process of constructing an accurate and reliable model that has the capacity to accurately forecast the job placement of a student.

4.2 Data Preprocessing

In this study, we modeled individual career prediction in various contexts based on a synthetic dataset compiled from students. To construct the prediction model, we first extract from the dataset the contextual information discussed above. However, the raw contextual data cannot be used with ML techniques to construct the prediction model. To make such student contextual data applicable for building the prediction model, we transform the contextual information into a meaningful category, such as continuous variables into nominal values, fill the missing values, remove duplicates entries, normalize or standardize the numerical features and etc., these datasets use to ML classifiers before building the prediction model.

4.3 Data Set Description

The data for these models come from a dataset of undergraduates. The things that were found to be most important for placing a student were looked at. Table 2 shows the same. The variables used in this study to cover all the parameters that affect student placement. In our dataset, there are 39 attributes which include a particular student's academic information like academic performance (CGPA), attendance, 10^{th} and class XII percentage, technical skills such as different programming languages(C, C++, Java, Python, etc.,) software tools, hardware knowledge and Certifications (Cloud computing, Cyber security, and web development), non-technical skills such as communication skills, teamwork, leadership, creativity, problem-solving ability, logical and analytical thinking ability and their placement information, academic performance data, basic information data of students, family details, behavior data, health reports, regular activates, existing competitive experience, participation in cultural activities, working skills, communication skills, and career prediction data. Student’s behaviors include at home and college and exits, consumption at campus locations, wakeup times, extra cultural activities (e.g., sports, dances), book reading, and academic achievements. Placement Status is the target variable that indicates whether the student was placed in a job or not.

Table 2. Attribute Information of the dataset

Sl No	Attribute Name	Type of attribute
1	Name	String
2	Age	Numeric
3	Branch	String
4	Gender	String
5	Do you have any active backlogs	Boolean
6	Attendance in %	Numeric
7	B. Tech CGPA	String
8	Inter or Diploma Percentage	Numeric
9	10th Percentage	Numeric
10	Rate your communication skills	String
11	Rate your Coding and DBMS skills	String
12	Rate your Web and Cloud Computing skills	String
13	Number of certifications you have?	Numeric
14	Have you done any Internships?	Boolean
15	Is your father a job holder?	Boolean

16	Is your mother a job holder?	Boolean
17	Is your sibling a job holder?	Boolean
18	Are you involved in any Extra Circular Activities?	Boolean
19	Do you have any chronic disorders?	Boolean
20	Have you ever qualified for any SET/NET/CET examinations? (like EAMCET, JEE, POLYCET,APRJC etc.,)	Boolean
21	Have you ever participated in technical workshops?	Boolean
22	Rate your Logical and Reasoning skills	String
23	Have you gone for any industrial tours?	Boolean
24	Institution type	String
25	Shelter type	String
26	Do you have any educational gap?	Boolean
27	Do you know any industrial skills (new tools/technologies)?	Boolean
28	Are you a hostler or Day scholar?	String
29	How many hours do you study in a day?	Numeric
30	Can you adapt in other states?	Boolean
31	Do you have social media awareness?	Boolean
32	Are you an active member in NSS or NCC?	Boolean
33	Have you published any research paper?	Boolean
34	How many languages you know other than your mother tongue and English?	Numeric
35	Family annual income	Numeric
36	Family type	String
37	Are you doing any part time job?	Boolean
38	Are you getting any government scholarships?	Boolean
39	Are you getting any placement support from your college?	Boolean
40	Have you got placed for any company in your final year?	Boolean

4.4 Configuring the Simulation Environment

The Simulation Environment, System, and Parameter Setup are all contained within this section. The experiment was carried out on a desktop computer manufactured by ACER (ASPIRE A315-58G). The operating system used was Windows–10 Pro 64-bit, and the processor used was an 11^{th} Generation Intel(R) Core(TM) i^5-1135-G7 operating system at 2.40 GHz 2.42 GHz. There are 8 GB of RAM and 1 TB of secondary storage. For data analysis, the sk-learn framework and the classification-metrics framework were used. The Pandas framework and the Numpy Python framework were used for data comparison. In Table 3, you can find information about how to set up the parameters. The preprocessed data set can then be split into 70%–30% ratio for training and testing sets, and evaluates the model's performance by hyper parameters are tuned based 10-fold cross validation. In our experiments, we use the accuracy, recall, precision, and micro-F1 as our evaluation metrics.

Table 3. Setup of Parameters

Classifiers	**Config of Variables**
LR	{max_iter=1000}
DT	{random_state=43}
NB	{random_state=3}
RF	{n_neighbors=3}
KNN	{n_neighbors=1}
SVM	{cache_size=200, tol=0.001}
ADA BOOST	{base_estimator=dt_classifier,n_estimators=300}
GBOOST	{learning_rate=0.07)
C BOOST	{iterations=250, learning_rate=0.01}
Proposed XGBOOST	{n_estimators = 500, learning_rate = 0.1)

In Figure 2, we see how each dataset is related to all the others via the correlation heatmap. Lighter colors indicate less of a connection between the two sets of data, while blue denotes a close correlation of 1.0 to 0.2. The quality of the synthetic dataset is demonstrated by the increasing correlation of samples over time, with all attributes being highly correlated with other column attributes.

5. Results Analysis

This research enables us to start figuring out the exact factors that affect a student's performance in campus recruitment. It provided an easy-to-understand idea of which considerations directly and indirectly contribute to overall successes and which are the most important things that could help them. We have compared some ML and Advanced ML algorithms such as LR, DT, NB, RF, KNN, SVM, GBoost, CBoost, AdaBoost, and proposed XGBoost. In this research, we have compared the models, by considering parameters like Accuracy, F1 Score, Precision, TPR, FPR and TNR, ROC-AUC values and find the best among them.

In this paper, we have developed ten no ML approaches, that are developed for analyzing student's placement prediction status. The proposed method Extreme Gradient Boosting has been compared with other nine methods such as LR, DT, NB, RF, KNN, SVM, GBoost[43], CBoost, AdaBoost. Figure 3 Represents the Training and Testing Error analysis of all classifiers along with Extreme Gradient Boosting method. It is obvious to conclude that, the proposed method is having less error rate in terms of both training and testing as compared to other considered approaches. Table 4 indicates the result analysis of all the methods in terms of several performance indicators such as accuracy, TP, FP, TN, FN, TPR, FPR, Precision, TNR, F1 Score and ROC-AUC. Out of nine compared methods, the accuracy of XGBoost (86%) is high. It has more number of TP (101), TN(100) values and less number of FP(18), FN(14) values. Similarly, the TPR of the proposed method is 0.87, and TNR is 0.84 which are highest among the remaining. The FPR of the proposed method is 0.15 which least among the all. By considering all the above results, the performance of the proposed Extreme Gradient Tree Boosting method is quite encouraging for effective prediction of student placement status.

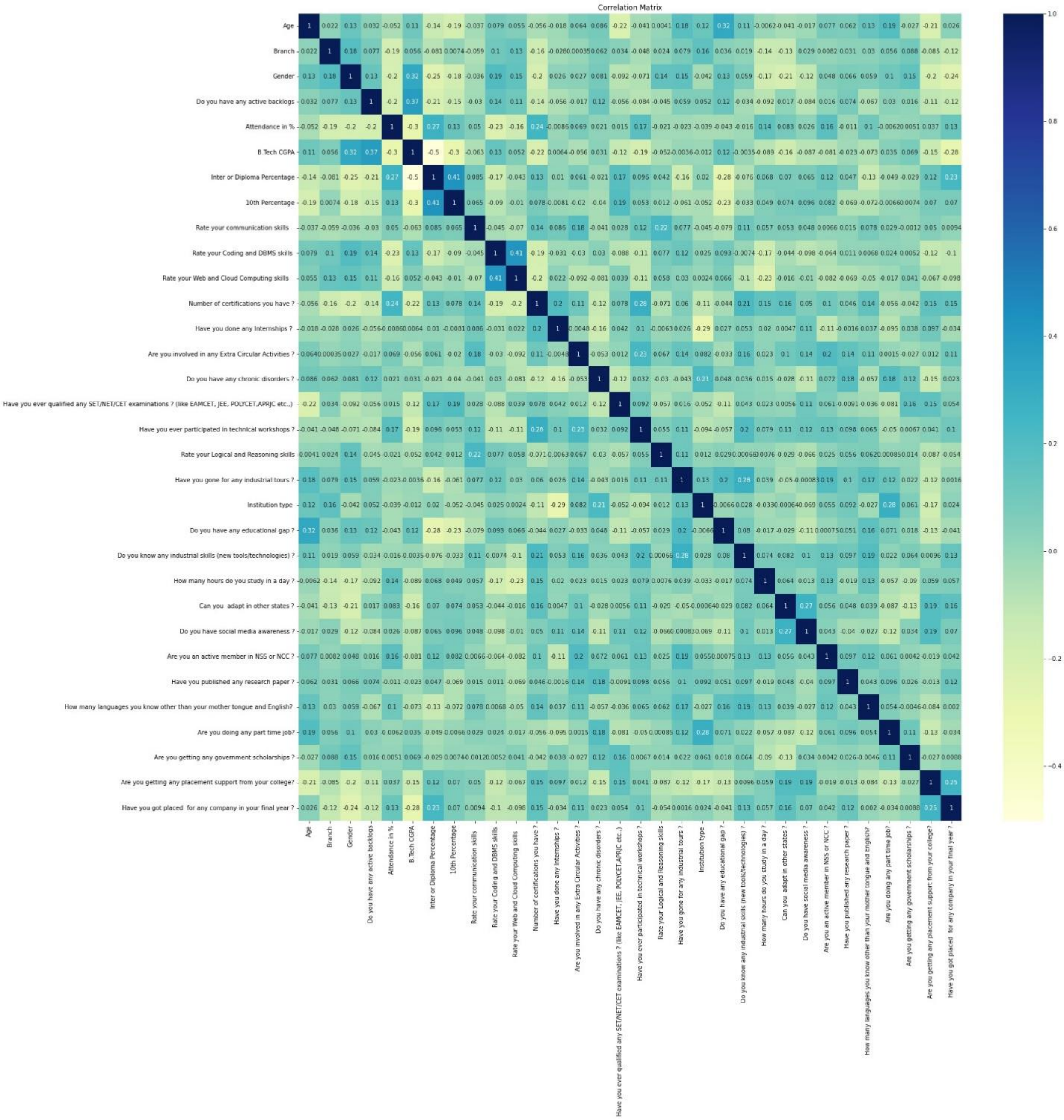


Figure 2. Correlation and Heat map

In the Figure 3 1st graph represents the training errors of all the algorithms for 10 K-Folds and the 2nd graph represents the testing errors of all the algorithms for 10 K-Folds.

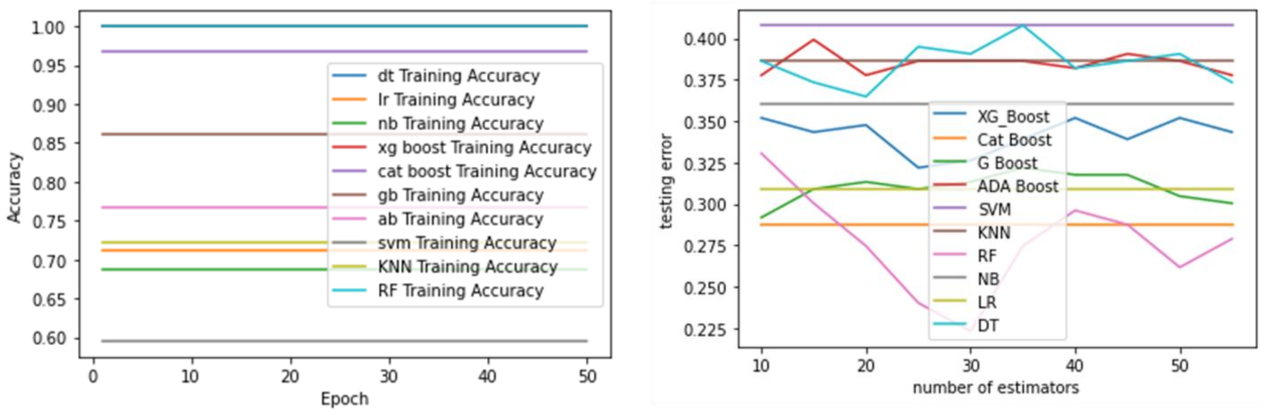
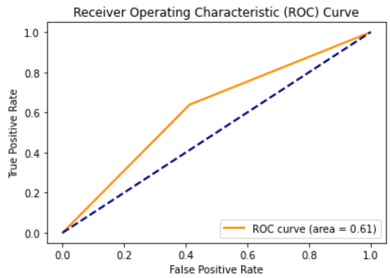


Figure 3. Training and Testing Error analysis of all classifiers

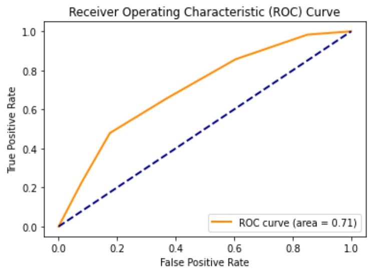
Table 4. Results of all the methods along with Proposed XGBoost

Research Models	Metrics for Performance										
	Accuracy	**TP**	**FP**	**TN**	**FN**	**TPR**	**FPR**	**Precision**	**TNR**	**F1 - Score**	**ROC-AUC**
LR	0.78	89	30	94	20	0.81	0.24	0.74	0.75	0.78	0.73
DT	0.72	85	34	84	30	0.73	0.28	0.71	0.71	0.72	0.67
NB	0.77	94	25	86	28	0.77	0.22	0.78	0.77	0.78	0.76
RF	0.72	80	39	90	24	0.76	0.30	0.67	0.69	0.71	0.77
KNN	0.82	82	17	90	24	0.77	0.15	0.82	0.84	0.79	0.66
SVM	0.76	91	18	87	27	0.43	0.77	0.83	0.82	0.80	0.75
ADA BOOST	0.76	84	35	93	21	0.80	0.27	0.70	0.72	0.74	0.67
GBOOST	0.79	91	28	94	20	0.81	0.22	0.76	0.77	0.79	0.75
CAT BOOST	0.75	80	39	95	19	0.80	0.29	0.67	0.70	0.73	0.79
Proposed XG BOOST	**0.86**	**101**	**18**	**100**	**14**	**0.87**	**0.15**	**0.84**	**0.84**	**0.86**	**0.77**

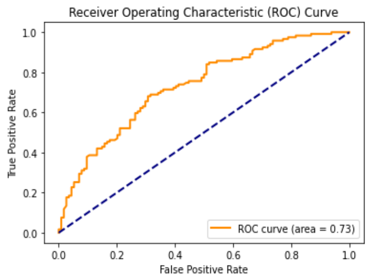
Figure 4 represents the ROC curves of all the methods i.e. a)LR, b) DT c) NB, d) RF e) KNN f) SVM g) GBOOST h) ADA BOOST i) CBOOST and j) PROPOSED XGBOOST. Here the ROC-AUC values of every algorithm are greater than 0.5, which justifies that every algorithm has high TP Rate and less FP Rate. The training error and testing error of the proposed method are 0.00 and 0.14 respectively (Shown in Table 5). By considering all the required parameters, we came with a conclusion that Extreme Gradient Tree Boosting is giving the best results for our dataset.



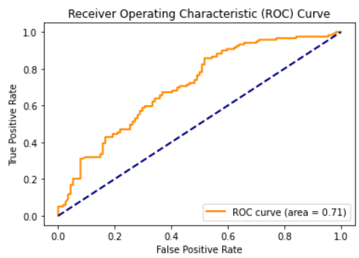
(a) DT



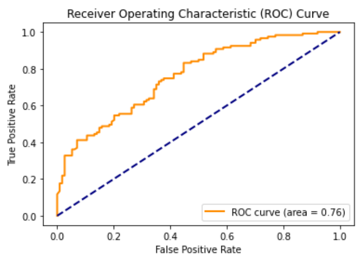
(b) KNN



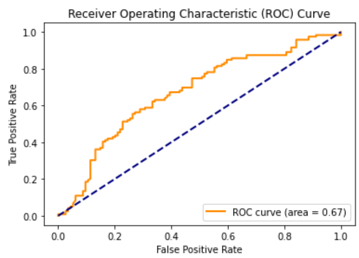
(c) ADABOOST



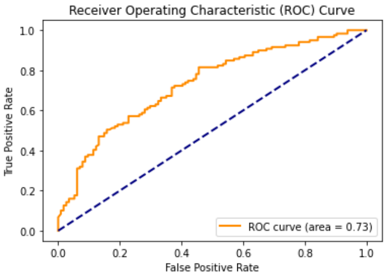
(d) LR



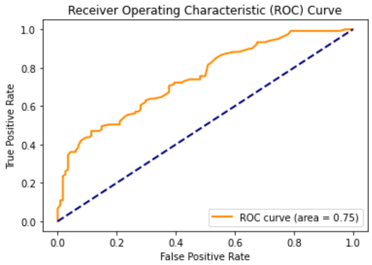
(e) CBoost



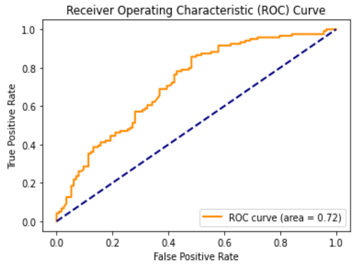
(f) NB



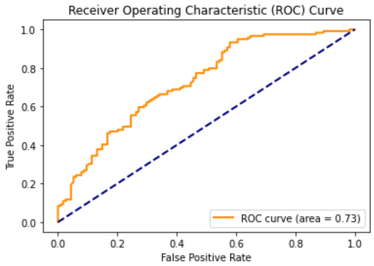
(g) GBoosting



(h) RF



(i) SVM



(j) XGBOOST

Figure 4. ROC Curve Analysis of a) DT, b) KNN, c) ADA Boost, d) LR, e) CBoost, f) NB, g) Boosting, h) RF, i) SVM and j) proposed XGBoosting

Table 5. Training error versus testing error

Models	**Metrics**	
	Train Error	***Test Error***
LR	0.27	0.22
DT	0.00	0.28
NB	0.32	0.23
RF	0.00	0.28
KNN	0.24	0.18
SVM	0.39	0.24
ADA BOOST	0.00	0.24
GBOOST	0.16	0.21
C BOOST	0.01	0.25
Proposed XG BOOST	**±0.00 - ±0.00**	**±0.10 – ±0.20**

The experimental results show that our method is superior to its state-of-the-art alternatives, demonstrating the positive impact that the XG Boost technique has on the overall performance of the model.

6. Conclusion

In this research, we examine how professional competencies, behavioral consistency, and other factors influence the career paths of college students. The research has also provided numerous useful suggestions for enhancing the model. This research aims to bridge the gap between traditional machine learning concepts and advanced machine learning algorithms and the experimental categorization of students for career prediction. We have proposed an XG Boost method that incorporates each institution's a priori knowledge. Assumptions that items in the same category should all share the same label inspired this. Several trials have shown that compared to other methods of career prediction, ours produces significantly more accurate outcomes. There are interesting avenues for research to take in the future. There is a more precise way to discover initial classification models. Our approach can also be expanded to incorporate additional types of data, such as academic performance and survey responses, in addition to behavioral observations. Additionally, it makes sense to enhance our model to not only anticipate job choices but also advise on career planning, such as recommending the necessary courses.

References

1. Albion, M.J. and Fogarty, G.J. Factors Influencing Career Decision Making in Adolescents and Adults. Journal of Career Assessment, vol. 10, no. 1, pp. 91-126, 2002.
2. Festinger, L. A Theory of Social Comparison Processes. Human relations, vol. 7, no. 2, pp. 117-140, 1954.
3. Nie, M., Yang, L., Sun, J., Su, H., Xia, H., Lian, D., and Yan, K. Advanced Forecasting of Career Choices for College Students based on Campus Big Data. Frontiers of Computer Science, vol. 12, pp. 494-503, 2018.
4. Hunt, D.E. Matching Models in Education: The Coordination of Teaching Methods with Student Characteristics. Ontario Institute for Studies in Education, Monograph, 1971.
5. Yao, H., Nie, M., Su, H., Xia, H., and Lian, D. Predicting Academic Performance via Semi-Supervised Learning with Constructed Campus Social Network. In Database Systems for Advanced Applications: 22nd International Conference, DASFAA 2017, Suzhou, China, March 27-30, 2017, Proceedings, Part II, Springer International Publishing, pp. 597-609, 2017.
6. Thompson, M.N. and Subich, L.M. The Relation of Social Status to the Career Decision-Making Process. Journal of Vocational Behavior, vol. 69, no. 2, pp.289-301, 2006.
7. Kuhn, M. and Johnson, K. Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC, 2019.
8. Boer, E.R. Behavioral Entropy as an Index of Workload. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Sage CA: Los Angeles, CA: SAGE Publications, vol. 44, no. 17, pp. 125-128, 2000.
9. Han, J., Pei, J., and Tong, H. Data mining: concepts and techniques. Morgan kaufmann, 2022.
10. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., and Data, M. Practical Machine Learning Tools and Techniques. In Data Mining. Amsterdam, The Netherlands: Elsevier, vol. 2, no. 4, pp. 403-413, 2005.
11. Vignesh, S., Priyanka, C.S., Manju, H.S., and Mythili, K. An intelligent career guidance system using machine learning. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, vol. 1, pp. 987-990, 2021
12. Bannaka, B.E., Dhanasekara, D.G., Sheena, M.K., Karunasena, A., and Pemadasa, N. Machine Learning Approach for Predicting Career Suitability, Career Progression and Attrition of IT Graduates. In 2021 21st International Conference on Advances in ICT for Emerging Regions (ICter), IEEE, pp. 42-48, 2021.
13. Chakrapani, P. and Chitradevi, D. Simulation of Machine Learning Techniques to Predict Academic Performance. In 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC), IEEE, pp. 329-334, 2022.
14. Sarangarajan, S., Chitra, C.K., and Shivakumar, S. AUtomation of Competency & Training Management using Machine Learning Models. In 2021 Grace Hopper Celebration India (GHCI), IEEE, pp. 1-6, 2021.
15. Delgado, E.J., Morales, L.R., and Araya, Y.C. Implementation of Machine Learning Algorithms to Classify University Academic Success. In 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), IEEE, pp. 1-4, 2022.
16. Laurens-Arredondo, L. and Hernández-García, R. Academic Performance Predicting Model based on Machine Learning and Keller's Motivation Measure, 2022.
17. Aravind, T., Reddy, B.S., Avinash, S., and Jeyakumar, G. A Comparative Study on Machine Learning Algorithms for Predicting the Placement Information of under Graduate Students. In 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, pp. 542-546, 2019.
18. Bhoite, S., Patil, C.H., Thatte, S., Magar, V.J., and Nikam, P. A Data-Driven Probabilistic Machine Learning Study for Placement Prediction. In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), IEEE, pp. 402-408, 2023.
19. Bao, Y., Peng, Y., and Wu, C. Deep Learning-Based Job Placement in Distributed Machine Learning Clusters. In IEEE INFOCOM 2019-IEEE conference on computer communications, IEEE, pp. 505-513, 2019.
20. Hiremath, A., Vinutha, K., and Yogisha, H.K. Job Role Prediction System. In 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), IEEE, vol. 1, pp. 1-4, 2021.
21. Sadasivam, R., Paramasivam, S., Prakash raj, N., and Saravanan, M. (2022). Students Career Prediction. International Journal of Health Sciences, vol. 6, no. S5, pp. 1357-1365, 2022.
22. Wang, Y., Yang, L., Wu, J., Song, Z., and Shi, L. Mining Campus Big Data: Prediction of Career Choice using Interpretable Machine Learning Method. Mathematics, vol. 10, no. 8, pp. 1289, 2022.
23. José-García, A., Sneyd, A., Melro, A., Ollagnier, A., Tarling, G., Zhang, H., Stevenson, M., Everson, R., and Arthur, R. C3-IoC: A Career Guidance System for Assessing Student Skills using Machine Learning and Network Visualisation. International Journal of Artificial Intelligence in Education, pp. 1-28, 2022.
24. Ng, H., bin Mohd Azha, A.A., Yap, T.T.V., and Goh, V.T. A Machine Learning Approach to Predictive Modelling of Student Performance. F1000Research, vol. 10, 2021.
25. Rao, K.S., Swapna, N., and Kumar, P.P. Educational Data Mining for Student Placement Prediction using Machine Learning Algorithms. International Journal of Engineering Science Technologies, vol. 7, no. 1.2, pp. 43-46, 2018.
26. Halde, R.R., Deshpande, A., and Mahajan, A. Psychology Assisted Prediction of Academic Performance using Machine Learning. In 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, pp. 431-435, 2016.
27. Elayidom, S., Idikkula, S.M., and Alexander, J. A Generalized Data Mining Framework for Placement Chance Prediction Problems. International Journal of Computer Applications, vol. 31, no. 3, pp. 0975-8887, 2011.
28. Ishizue, R., Sakamoto, K., Washizaki, H., and Fukazawa, Y. Student Placement and Skill Ranking Predictors for Programming Classes using Class Attitude, Psychological Scales, and Code Metrics. Research and Practice in Technology Enhanced Learning, vol. 13, pp. 1-20, 2018.
29. Shen, Y.J. and Herr, E.L. Career Placement Concerns of International Graduate Students: A Qualitative Study. Journal of Career

Development, vol. 31, pp. 15-29, 2004.

30. Agarwal, K., Maheshwari, E., Roy, C., Pandey, M., and Rautray, S.S. Analyzing Student Performance in Engineering Placement using Data Mining. In Proceedings of International Conference on Computational Intelligence and Data Engineering: Proceedings of ICCIDE 2018, Springer Singapore, pp. 171-181, 2019.
31. Kumar, D.S., Siri, Z., Rao, D.S., and Anusha, S. Predicting Student's Campus Placement Probability using Binary Logistic Regression. International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9, pp. 2633-2635, 2019.
32. Halde, R.R., Deshpande, A., and Mahajan, A. Psychology Assisted Prediction of Academic Performance using Machine Learning. In 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, pp. 431-435, 2016.
33. Ade, R. and Deshmukh, P.R. An Incremental Ensemble of Classifiers as a Technique for Prediction of Student's Career Choice. In 2014 First International Conference on Networks & Soft Computing (ICNSC2014), IEEE, pp. 384-387, 2014.
34. Nie, M., Xiong, Z., Zhong, R., Deng, W., and Yang, G. Career Choice Prediction based on Campus Big Data—Mining the Potential Behavior of College Students. Applied Sciences, vol. 10, no. 8, pp. 2841, 2020.
35. Situmorang, Z. Analysis Optimization K-Nearest Neighbor Algorithm with Certainty Factor in Determining Student Career. In 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), IEEE, pp. 306-310, 2020.
36. Mandalapu, V. and Gong, J. Studying Factors Influencing the Prediction of Student STEM and Non-STEM Career Choice. In Educational Data Mining, 2019.
37. Zeineddine, H., Braendle, U., and Farah, A. Enhancing Prediction of Student Success: Automated Machine Learning Approach. Computers & Electrical Engineering, vol. 89, pp. 106903, 2021.
38. VidyaShreeram, N. and Muthukumaravel, A. Student Career Prediction using Machine Learning Approaches. In First International Conference on Computing, Communication and Control System, pp. 444, 2021.
39. Roy, K.S., Roopkanth, K., Teja, V.U., Bhavana, V., and Priyanka, J. Student Career Prediction using Advanced Machine Learning Techniques. *International Journal of Engineering & Technology*, vol. 7, no. 2.20, pp. 26-29, 2018.
40. Maurya, L.S., Hussain, M.S., and Singh, S. Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction based on Academic Performance. *Applied Artificial Intelligence*, vol. 35, no. 6, pp. 403-420, 2021.
41. Agarwal, K., Maheshwari, E., Roy, C., Pandey, M., and Rautray, S.S. Analyzing Student Performance in Engineering Placement using Data Mining. In *Proceedings of International Conference on Computational Intelligence and Data Engineering: Proceedings of ICCIDE 2018*, Springer Singapore, pp. 171-181, 2019.
42. Ramesh, V.A.M.A.N.A.N., Parkavi, P., and Ramar, K. Predicting Student Performance: A Statistical and Data Mining Approach. *International journal of computer applications*, vol. 63, no. 8, 2013.
43. K. Eswara Rao, G.Appa Rao, and S. Anuradha. Gradient Tree Boosting Approach for Software Defect Prediction. *International Journal of Advanced Science and Technology*, vol. 28, no. 20, pp. 750, 2019.