IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁘ Impact Factor 8.102 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12626

Pre-Placement Prediction System using Machine Learning

Pranjal Khose[1], Siddhi Yeole[2], Siddhee Bagool[3], Aarti Sonawane[4]

Dr.(Mrs) S. R. Khonde[5]

Department of Computer Engineering MESCOE, Pune, India[1-5]

Abstract—The recruitment process at various organizations can be difficult, especially when there are large numbers of applicants. for less job opportunities. In this regard, the enrollment policy system has emerged as an effective tool for pre-selection appropriate space. This article focuses on the admissions policy system designed to screen students for a particular company based on their skills, where they live and their salary expectations. This system uses different algorithms to predict the quality Applicants for a particular position and shortlisted candidates are based on pre-determined criteria. The paper discusses the benefits and limitations of the input forecasting system and present some case studies where the system is available successfully implemented. The results show that the system is efficient, clear and accurate, making it more efficient. and reducing stigma in the recruitment process. Overall, this article provides insight into the power of integration policy measures to change the recruitment process in organizations and improve the quality of candidate selection.

Keywords— Machine Learning, Beautiful soup, KNN, Flask API

I. INTRODUCTION

The recruitment process at various organizations can be difficult, especially when there are large numbers of applicants for less job opportunities. In this regard, the Pre-placement system has emerged as an effective tool for pre-selection appropriate space. This article focuses on the admissions policy system designed to screen students for a particular company based on their skills, where they live and their salary expectations. This system uses different algorithms to predict the quality applicants for a particular position and shortlisted candidates are based on pre-determined criteria. The paper discusses the benefits and limitations of the input forecasting system and present some case studies where the system is available successfully implemented. The results show that the system is efficient, clear and accurate, making it more efficient. and reducing stigma in the recruitment process.

This article provides insight into the power of Pre- Placement prediction system to change the recruitment process in organizations and improve the quality of candidate selection.

II. LITERATURE REVIEW

This section describes the literature review and pre-placement prediction using machine learning and data mining. In [1] authors Jagan Mohan Reddy D researched and predicted employee behaviour for organizational growth. They created an anecdotal dataset using publicly available information and career options. The data includes characteristics such as gender, age, education, and class marks that indicate whether the user will participate in the meeting or not. The authors used several algorithms, including decision trees, forest, Gaussian Naive Bayes, and nearest neighbour, to obtain results. Gaussian Naive Bayes produced the highest accuracy, precision, and recall among all the tested algorithms, which was attributed to the low agreement between the features in the dataset.

Article [2] used a sample of data collected from the college admissions service, which has a record of more than 1000 students. The characteristics that were evaluated were credits, legal documents, whether they were applied or not, and the percentage of BTech. The authors used a decision tree and random forest algorithm, allocating 80% of the dataset to training and 20% to testing. The accuracy obtained for the decision tree was 84% and for the forest 86%, which made it clear that the forest has more accuracy than the decision tree.

In [3] the authors amazed the prophetic predictions of using the Scores the event program, CGPA and others. Automatic learning algorithms as a regular number of inventory, the forest, KNN and SVM uses malic. The SVM algorithm achieved

DOI: 10.17148/IJARCCE.2023.12626

100% accuracy, while the logistic regression gave 97.59% accuracy on the given dataset. Studies show that these algorithms are suitable for binary classification problems, achieving an accuracy of more than 95%.

Article [4] proposes a method to predict the performance status of students using Random Forest algorithm. The data collection includes scores from important admission tests. To improve the accuracy of the model, the authors suggest modifying the parameters of the algorithm. The proposed system predicts the probability that a student will be placed and displays a list of institutions that have a chance of being placed. The course can also provide a list of skills that a student will acquire to meet business needs.

Article [5] This paper presents an overview of using machine learning algorithms to predict the placement performance of undergraduate students. It compares the performance of different classifiers, including multilayer perceptron, logistic model tree, sequential minimal optimization, simple logistic, and logistic classifiers, on a student dataset. The evaluation is based on various metrics such as accuracy, error rates, precision, recall, and ROC area. The goal is to identify the algorithm that provides the most accurate predictions and offer insights for improving student placement in education.

Article [6] This paper introduces a novel approach for evaluating job applicants. A prototype system is implemented to showcase and evaluate the approach in a real-world recruitment scenario. The system extracts objective criteria from applicants' LinkedIn profiles and utilizes linguistic analysis on their blog posts to infer their personality characteristics. The system performs consistently compared to human recruiters, making it reliable for automating applicant ranking and personality mining.

Article [7] This paper addresses the importance of student placement in educational institutions and proposes a model to predict the placement chances of current students based on previous year's data. The proposed algorithm is compared with traditional classification algorithms like Decision Tree and Random Forest in terms of accuracy, precision, and recall. The results indicate that the proposed algorithm performs significantly better. The findings of this study can assist institutions in improving their placement department and enhance the overall placement success of their students.

Article [8] This paper addresses the issue of high unemployment rates among college graduates in India. To combat this, the authors propose a system that analyzes the skill sets of placed and non-placed students to predict job placement success. The system considers both technical and soft skills and achieves high accuracy using SVM and XGBoost models. The study highlights the importance of technical skills, projects, certified courses, and internships in securing employment. The findings suggest that implementing such a system can improve placement rates in colleges and reduce graduate unemployment.

Article [9] This article examines the concept of an admissions policy model for undergraduate engineering students. iIt emphasizes the importance of job placement for students and the need for prediction systems to help them assess their chances and plan their academic journey accordingly. The paper provides a literature survey of existing models in this area, highlighting the use of data mining and machine learning techniques. By analyzing previous year student data, these models aim to assist institutions in academic planning and improve the placement outcomes for student.

Article [10] This paper focuses on developing models for predicting student performance by considering not only past academic performance but also other important factors such as interests, attributes, and opinions. It employs machine learning and deep learning techniques and conducts exploratory data analysis to uncover correlations between performance and psychographic attributes. The goal is to create a comprehensive student performance analysis system that provides a holistic understanding of students' performance and aids schools and universities in supporting their students effectively

III. PROPOSED SYSTEM

The Proposed system is a web application created to simplify the application process and work as a powerful recruiting tool for businesses. Colleges can also use the system to get information about which students would make good candidates for recruitment campaigns. The input policy, one of the system's key components, establishes the likelihood of candidates being shortlisted based on a number of factors. These criteria cover the candidate's skill, backlog size, CGPA, and academic score. Machine learning methods—more specifically, the Random Forest Regressor algorithm—are used to forecast a candidate's chances of being shortlisted. This algorithm was picked because it can predict outcomes accurately and handle categorical and numerical data. Data from prior placement campaigns is used to train the model. This dataset, which includes data on candidate profiles and whether they were successfully placed or not, is used to train the model. The model discovers trends and correlations between the input variables and the placement likelihood by examining this historical data.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12626

Based on a candidate's profile and the given input parameters, the model can be used to forecast the likelihood that the candidate will be hired once it has been trained. A probability score that reflects a candidate's chances of being shortlisted is produced by the system. Due to their increased ability to predict future events, organizations are better equipped to evaluate candidates quickly and make wise hiring choices. Conversely, colleges can use the system to find suitable candidates to suggest for the recruitment drive, increasing the likelihood that their students will find employment.

In order to provide a thoughtful and data-driven approach to candidate assessment and recruitment, the proposed system makes use of machine learning and the Random Forest Regressor algorithm. It improves decision-making, simplifies the application process, and increases the recruitment campaign's effectiveness as a whole.

A. System Architecure

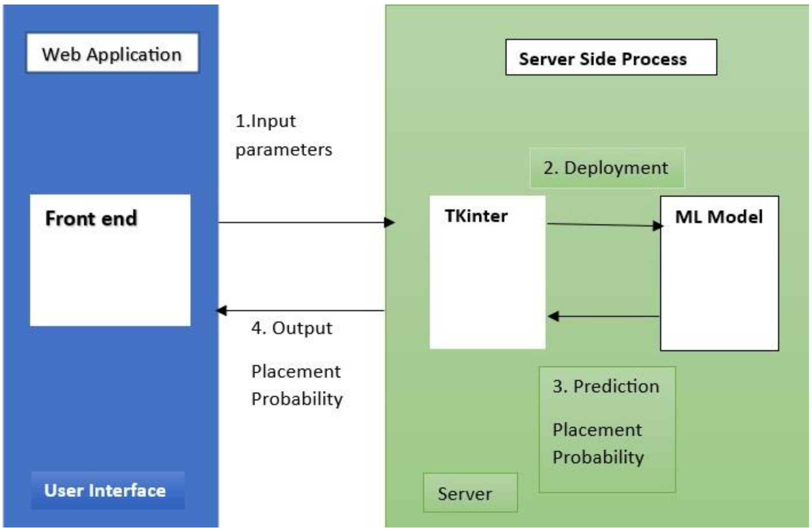

Fig. 1. System Architecture

A visually appealing graphical user interface (GUI) was created at the user interface layer using the tkinter library. The tkinter widgets used to implement the various elements of this GUI, including buttons, labels, text fields, and dropdown menus. Event handlers were also created in order to record user interactions and start the corresponding actions. The scikit-learn library was essential for the machine learning layer. This library was used to create and train machine learning models, and functions were added to preprocess the data before feeding it to the models. This required performing operations like feature selection, scaling, and encoding. The created models were then used to predict the future or carry out other desired tasks, like classification or regression. Between the user interface and the machine learning components, the integration layer made sure that communication flowed without interruption. This required developing functions for data transfer between the GUI and the machine learning model. The user experience was seamless as the model received the data inputs from the GUI and returned the results to the user interface.

Setting up a SQLite database to store and retrieve pertinent data for the project was part of the data layer. Utilizing the SQLite library, functions for data retrieval, preprocessing, and storage were developed. On the basis of the project requirements, this layer provided data integrity, security, and suitable access controls.

A control module that oversaw the overall coordination and flow between the various system components was included in the system control layer as a final component. System initialization, handling errors, and logging all have their own functions. This layer made sure that the user interface, machine learning model, and data components were all coordinated and executed in the right order. System control, data storage, machine learning models, and user interface were all integrated into one structure by the system architecture. Users interacted with the GUI, which communicated with the machine learning layer to process data and make predictions. A seamless data exchange was made possible by the integration layer, while storage and retrieval were handled by the data layer. The data layer took care of storage and retrieval, while the integration layer facilitated easy data exchange. A reliable and user-friendly application was produced as a result of the system control layer's efficient coordination and error handling throughout the system.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁝ Impact Factor 8.102 ⁝ Peer-reviewed / Refereed journal ⁝ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12626

B. Proposed methodology

When selecting candidates, companies often consider various criteria to narrow the pool of candidates. This includes cleaning data so that only candidates who meet certain requirements are considered. The following methods are often used in the data cleansing process.

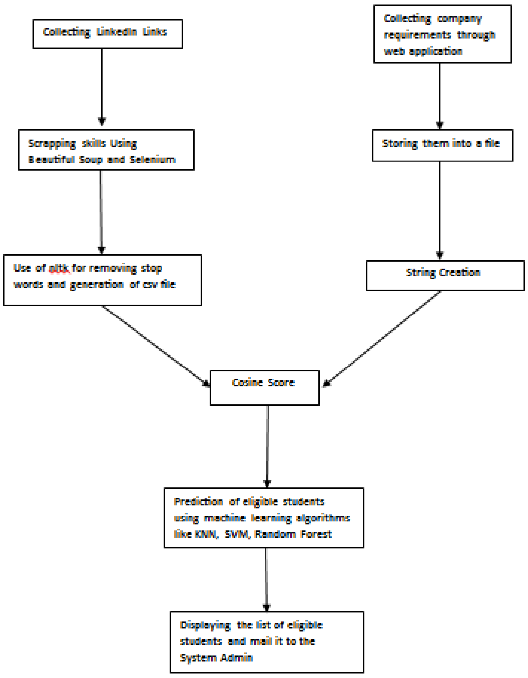

Fig. 2. Proposed Methodology.

Stage 1: collecting Linkedin link from students

The students' LinkedIn profile links were collected through research and direct media, ensuring that the process was carried out with full consent and authority. Students were given the option to voluntarily share their LinkedIn profile as part of the data collection process. The surveys were distributed among student groups and efforts were made to contact individual students to obtain their LinkedIn connections. Strict adherence to privacy principles and regulations is maintained throughout this process, ensuring the confidentiality and security of the information collected.

By using several methods of data collection, including research and direct communication, a complete data collection of LinkedIn profiles was obtained, enabling the extraction of valuable insights and skills for pre-planning processes.

Stage 2:Scrapping Skills from Likendin using Beautiful soup and selenium driver

A Python library called Beautiful Soup is used for web scraping. It enables developers to extract particular data from HTML or XML documents, simplifying the parsing and navigating of intricate website architectures. In addition to a variety of tools for searching, filtering, and modifying data, Beautiful Soup offers a set of techniques for finding tags and text within an HTML or XML document. It is frequently used to extract data from websites and has grown to be one of the most well-liked Python web scraping libraries.

Popular open-source framework Selenium is used to automate web browsers. It enables programmers to mimic user actions on web pages like clicking links, completing forms, and page scrolling. Developers can automate web-based tasks and extract data from websites using Selenium WebDriver, a Selenium component that lets them control a web browser through code.. A number of web browsers, including Chrome, Firefox, Safari, and Internet Explorer, are supported by Selenium WebDriver, which also offers a number of methods for interacting with website elements. Developers are able to complete tasks like finding and clicking on elements, filling out forms, and switching between pages with Selenium WebDriver.

Using a combination of Beautiful Soup and Selenium driver, we were able to extract the students' skills from their LinkedIn profiles. We can parse HTML and XML documents using the Python library called Beautiful Soup to get data out of them. We used it to extract the HTML from the students' LinkedIn profiles. We also used the Selenium driver, a

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ჯ Impact Factor 8.102 ჯ Peer-reviewed / Refereed journal ჯ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12626

web driver that enables us to automate web browsers, but since LinkedIn requires a login in order to access the profiles. We were able to navigate to the student profiles and extract their skills thanks to the Selenium driver, which allowed us to automatically fill in the LinkedIn login information.

We used Beautiful Soup to find and extract the students' skills after we had the HTML content of the LinkedIn profiles. We were able to parse the HTML code using Beautiful Soup and locate the precise HTML tags and classes that held the skill-related data. Then, we extracted this data and included it as a feature in our machine learning models to forecast the most qualified applicants for open positions.

Stage 3:Use of NLTK library for text cleaning

For text processing and cleaning, the system makes use of the NLTK (Natural Language Toolkit) library. A variety of functionalities provided by NLTK help in handling text data efficiently. The elimination of stop words is one of the key features used. Articles, prepositions, and conjunctions—words that are frequently used but have little meaning in the context of text analysis—are among the stop words that are provided by NLTK.

The have reduced noise and enhanced the processed text's quality by getting rid of these stop words. The various cleaning procedures are carried out, including the elimination of non-ASCII characters, punctuation, URLs, mentions, and hashtags from the text.

.

Stage 4: collecting Requirements from the company

The front-end application's ability to save and retrieve data efficiently by storing the gathered business requirements as strings. It is simple to manipulate and process the requirements inside the application because they are presented as strings. When storing organizational requirements, such as job titles, skills, experience levels, and qualifications, as strings, there is flexibility in how different requirements can be handled. It is simple to combine, alter, and compare the requirements using the string format as necessary.

Additionally, the system integration with other parts is made simple by storing the company requirements as strings. The string representation makes it simple to transfer data between system modules or layers, enabling smooth processing and communication.

Stage 4: shortlisting candidate using classification algorithm

Students' skills are matched with the specifications provided by the employers using a classification algorithm. For this, a variety of algorithms can be used, including decision trees, neural networks, and k-Nearest Neighbors (k-NN). The database of student skills and the previously recorded business requirements are both used by the classification algorithm. The skill sets of the students are examined, and they are compared to the standards that the employers hope to see. The system can effectively identify and shortlist candidates who closely match the specified requirements by using a suitable classification algorithm. By calculating the distance between two points in a multi-dimensional space, the k-NN algorithm takes into account the similarity between a student's skill profile and the requirements of a company. Using a set of decision rules and decision trees, classification of students according to their skills is made structured.

Metrics like accuracy, precision, recall, and F1 score are used to assess the performance of the chosen algorithm. To improve the model's predictive abilities, hyperparameter tuning and other techniques may be used to fine-tune it.

Through the use of a classification algorithm during the shortlisting phase. The pre-placement prediction system is effective at identifying students who have the abilities and credentials that employers are looking for. The selection of candidates is streamlined as a result, and the likelihood of successful placements rises as a result of a better match between student abilities and business requirements.

Stage 6: Display list of shortlisted Candidate

A list of candidates who have been shortlisted is shown in the pre-placement prediction system based on a few general characteristics that determine their eligibility for the placement process. These aspects include standards like the number of classes a student fails, the cumulative grade for the academic year, the undergraduate grade, and other pertinent variables. The list of shortlisted candidates only includes students who meet or exceed the prerequisites.

This step makes sure that applicants who have shown satisfactory academic progress and satisfy the established requirements are given the chance to move forward in the placement process. The list of shortlisted candidates offers a focused and controllable pool of candidates who have satisfied the requirements for potential job placements.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12626

C. Data collection

We collected data from students using a Google form including their email address, name, department, undergraduate and postgraduate information (graduation year, cumulative total, level of education and number of fees), top three locations Likes, LinkedIn and Github links, and their resume. For company information, we collected data through an online portal, including company name, requirements, salary, year of completion, and years of experience required. The collected data is cleaned and processed to remove any errors or omissions, and appropriate mathematical techniques use machine learning techniques to create predictive models.

TABEL I
LIST OF ATTRIBUTES USED IN MODEL

Sr.no	**List of attributes/Features used in model**		
	Features	**Description**	**Type**
1	Email	Email id for sending mails	nominal
2	Name	Full name	nominal
3	Degree	B.E/BSC/B.Tech	Nominal
4	Department	Mechanical, Computer Eng.	Nominal
5	Degree status	Pursuing/Completed	Nominal
6	Year of passing UG	2023/2022/2021/2020	Ordinal
7	Aggregate score in UG	SGPA or CGPA	Ordinal
8	Year of passing PG	2023/2022/2021/2020	Ordinal
9	Aggregate score in PG	SGPA or CGPA	Ordinal
10	SSC & HSC	Marks	Ordinal
12	Educational Gap	0/1/2	ordinal
13	No of backlogs	No of dead or current backlogs	Ordinal
14	Salary expectation	Salary range expected by the candidate	Ordinal
15	Job location	Job location preferred by the candidate	nominal
16	Location Preference 1/2/3	Pune/Mumbai/Pan India	nominal
17	LinkedIn Link	Links	nominal

D. Data Preprocessing

When selecting candidates, companies often consider various criteria to narrow the pool of candidates. This includes cleaning data so that only candidates who meet certain requirements are considered. The following methods are often used in the data cleansing process.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12626

Stage 1: Attribute selection

1. Years of study: To determine if the candidate has recently graduated, some companies will indicate the period in which the candidate must have graduated. For example, we only consider applicants who graduated within the last three years. During data cleaning, applicants who graduated more than 3 years ago are removed from the database.

2. UG Branch: Some jobs may require applicants to have specific education, such as a degree in computer science or other fields. Students who do not meet the requirements are removed from the file.

3. No Backlog: Company structure may limit the number of backlogs a candidate can have, or applicants with no backlog at all will be preferred. Students who exceed or perform above the performance limit are excluded from the data set.

4. Salary Requirements: Most companies have salary requirements for the positions they hire for.Remove students whose salary requirements are outside the specified range to identify students whose salary requirements are based on the company's range.

5. Mark criteria: Institutions may establish a minimum percentage or GPA requirement for a UG degree candidate. Students who do not achieve the specified scores are excluded from the exam.

6. Location Preference: Depending on factors like proximity to clients or business requirements, the company may have a particular location preference for the position they are hiring for. In this instance, the data cleaning process would entail eliminating any applicants who are unwilling to work in the designated location.

Stage 2: Cleaning missing values

To make sure the data used for our machine learning models in the pre-placement prediction system was of high quality and reliability, we performed extensive data cleaning. In order to build a reliable and accurate system, we used a variety of techniques to handle missing values and outliers.

We carefully examined the data and used the appropriate imputation techniques based on the type and distribution of the features to address missing values. We filled in the missing values for numerical features using techniques like mean, median, or regression imputation. The mode was used to impute missing values for categorical features, or they were given a special category to indicate their absence.

We used statistical techniques like the Z-score or the interquartile range (IQR) to identify these outliers, which can have a significant impact on model performance and need to be handled. Then, in order to reduce the impact of outliers, Additionally, we normalized the data using transformations like logarithmic or square root transformations when dealing with highly skewed or long-tailed distributions.

We carried out missing value imputation and outlier handling separately for the training and testing datasets to ensure the accuracy and realism of our models. This preventative measure made sure that our models were tested and trained using real-world scenarios and prevented any information from leaking. We carefully validated the imputed values and made sure they made sense within the context of the data as we cleaned the data, paying close attention to data quality throughout. In order to ensure transparency and reproducibility, we also painstakingly recorded every step taken during the cleaning process. This improved the quality and validity of the system's outcomes overall by producing predictions that were more accurate and trustworthy.

Stage 3: one hot encoding

The process of converting categorical data into a format that machine learning algorithms can easily use is known as one hot encoding. Data with categories or labels, like color, gender, or nationality, are referred to as categorical data..

We used one-hot encoding to transform our dataset into a format suitable for machine learning algorithms. This method was used to transform categorical variables, like department and location preferences, into numerical form. For each distinct value of the categorical variable and, new columns are created using one-hot encoding gives each row in the dataset a 1 or a 0, depending on whether that value is present. We performed one-hot encoding on our data using the pandas get_dummies() function, which produced a new dataset with extra columns for each distinct value of the categorical variables. Our machine learning model now includes categorical data thanks to this process.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁘ Impact Factor 8.102 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12626

E. Extracting skills from LinkedIn

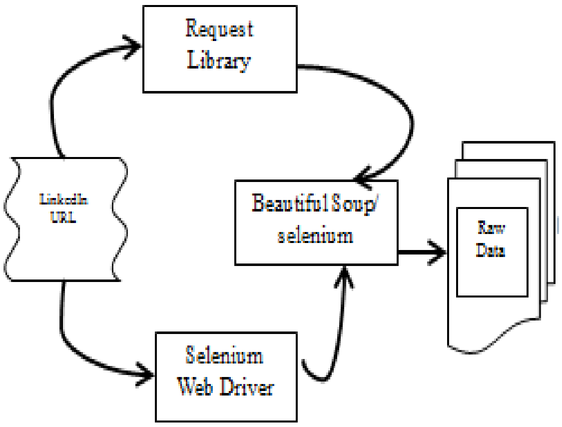

Fig. 3. .Scrapping tool structure

Using Beautiful Soup and Selenium Driver to pull skills from LinkedIn profiles involves the following steps:.

1.Selenium Driver Setup: To interact with the LinkedIn website, set up a Selenium Driver, such as ChromeDriver. Establish the required dependencies, including the WebDriver executable, and confirm that your browser version is supported.
2.Authentication : Use Selenium to automate the LinkedIn login process. Provide the required login credentials to the program, allowing Selenium to scan and access LinkedIn profiles.
3.Get Profile: Use Selenium to get the LinkedIn profile of any student by following the provided LinkedIn link. This includes opening URLs in programs and web browsers and tracking user behavior.
4. HTML Analysis: Extract HTML content from profile pages using Selenium. By accessing the source page or specific elements, retrieve the HTML code below that contains the profile's art section.
5.Beautiful Soup Integration: Add Beautiful Soup to your Python script to render rendered HTML content. Use the fancy soup function to navigate and search within the HTML structure of the profile page.
6. Extracting skills: Using Beautiful Soup, extract the HTML elements or sections that contain the information about the skills. This may involve looking for pertinent HTML tags or class names that contain the necessary skill-related information.
7.Extracting Data and storage: Using Beautiful Soup, separate the individual skills from the section that has been identified. Obtain the text or attributes related to the skill elements to collect the skill data. Put the extracted skills in a suitable database or data structure and organize them there for later use.
8.Clean Up Your Data: On the skills that were extracted, take the necessary data cleaning measures. To maintain consistency in the data, this may entail eliminating duplicates, standardizing skill names, changing text to lowercase, or handling formatting variations.
9. Database Integration: Connect the pre-placement prediction system to the database of student skills. Create a connection to the database, and then create a suitable schema to store the information on student skills, linking each skill to the corresponding student.
10. Matching Skills: When assessing candidates for a particular company, compare the skills listed in the student skills database to the company's requirements. Use database queries or other filtering methods to find students whose skills match those needed by the business, making it easier to narrow the pool of candidates.

F. Algorithm used for building model.
a) Prediction using Support Vector Machine

For classification and regression analysis, the Support Vector Machine (SVM) is a potent supervised machine learning algorithm. SVM is especially helpful when working with high-dimensional data or when there are more features than samples. The fundamental goal of SVM is to identify the hyperplane that best divides the data into two classes. The margin, or the separation between the hyperplane and the nearest data points from each class, is maximized by selecting a hyperplane that does just that. The decision boundary of the SVM model is defined by these nearest data points, which are referred to as support vectors.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12626

SVM can be applied to a variety of kernel functions, including linear, polynomial, radial basis function (RBF), and sigmoid. The decision boundary's shape and how well the SVM model can fit the data are both governed by the kernel function. Because it can handle non-linear decision boundaries and is less prone to overfitting, SVM is frequently chosen over other machine learning algorithms.

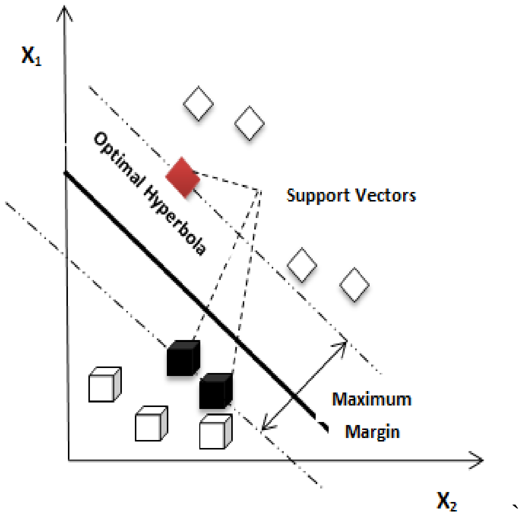

Fig. 4. .Prediction Using SVM

In our prediction system, SVM is used to rank candidates based on their characteristics and companies requirements. SVM finds a hyperplane separating the two classes with the largest margin and defines the decision boundary using support vectors. The margin is the distance between the decision boundary and the support vectors. SVM uses several kernel functions, such as linear, polynomial, RBF, and sigmoid, to determine the shape of the decision boundary and how well the SVM model can fit the data.

b) Prediction using Random Forest Algorithm

Random Forest is an ensemble machine learning algorithm that combines various decision trees to produce more precise predictions. To make a final prediction, the algorithm builds a large number of decision trees (the forest). Utilizing a randomly chosen subset of the training data and each decision tree in the forest is built an arbitrary subset of features. This lessens overfitting and enhances the model's ability to generalize. Every tree in the forest predicts something during the prediction phase, and the outcome is decided by adding up all of the predictions. Taking the average or the mode of the predictions is just one example of how the aggregation can be done. In comparison to other machine learning algorithms, Random Forest has several benefits, including the ability to handle missing values, nonlinear data, and high-dimensional data. It can be used for both classification and regression problems. Furthermore, it can withstand erratic data and outliers.

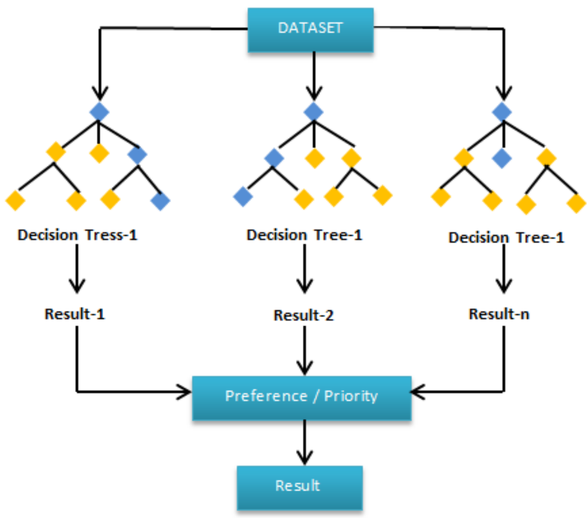

Fig. 5. .Prediction Using Random Forest

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁘ Impact Factor 8.102 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12626

To determine whether students are qualified for a job position, we used Random Forest as one of the machine learning algorithms. In order to produce a final decision, Random Forest builds multiple decision trees using different random subsets of data. The Random Forest algorithm's hyperparameters have been tuned to achieve the best performance. Adjustments to the number of decision trees, the depth of each decision tree, and the number of features used in each split are all part of this process.

c) Prediction using KNN

KNN is a straightforward and efficient machine learning algorithm that can be used to categorize new data points based on how similar they are to existing data points. It operates by figuring out how far each new candidate's input features are from those of each candidate already present in the dataset. Then, using their labels, it chooses the k nearest neighbors based on the calculated distances.

In the pre-placement prediction system project, we classify students into those who are suitable for a job position and those who are not based on their characteristics and the needs of the employer. The parameter k, which establishes how many nearest neighbors will be taken into account when making a prediction, must first be set to an appropriate value.

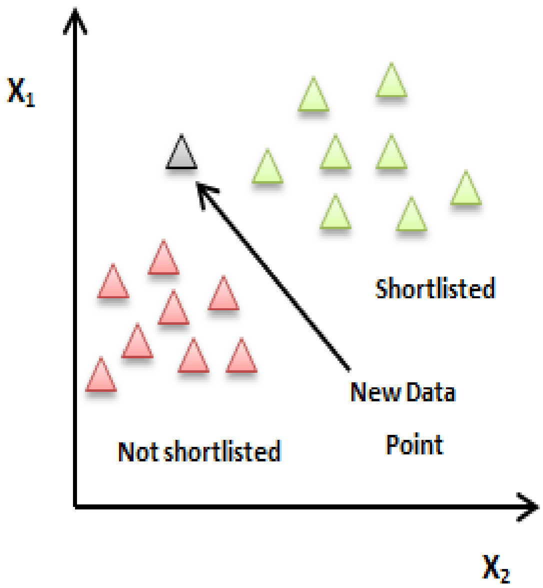

Fig. 6. .Prediction Using KNN

The students' SGPA, participation in hackathons, and LinkedIn profiles are among the features used as input to the KNN model, along with the qualifications needed by the employer for the job position. To calculate how far apart the new candidate is from all the other candidates in the dataset, these features are used. Since KNN stores all of the current data points in the dataset and uses them for classification, one of its benefits is that it does not require training. KNN can, however, be computationally expensive, particularly for large datasets.

G. Testing Dataset

We have use a variety of metrics, including accuracy, precision, recall, and F1 score, to assess the effectiveness of the machine learning algorithms. These metrics can offer information about how well the algorithms predict the suitability of the applicants for the open positions. Additionally, we can perform cross-validation, in which the dataset is split up into different subsets, each of which is used as a testing dataset while the other subsets are used for training. By doing so, the danger of overfitting may be diminished and the machine learning algorithms may be evaluated more accurately.

H. Email Functionality for Notifying Shortlisted Students to TPO

We have included an email functionality in our pre-placement prediction system to inform the Training and Placement Officer (TPO) about the list of students who have been shortlisted for the pre-placement program. The ability to effectively communicate and work together between the system and the TPO is made possible by this feature. Following the shortlisting procedure, the system automatically creates an email to the TPO with thorough information about the selected students. The email contains important details, including the students' names, email addresses, and cumulative

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12626

grade point averages (CGPAs). These specifics are necessary for further communication, evaluation, and subsequent placement procedures.

IV. RESULT AND DISCUSSION

The results of using different machine learning algorithms and our candidate placement prediction findings are discussed in this section. We used decision trees, random forests, support vector machines, and K Nearest Neighbor to obtain our results. The three metrics in Fig. are the focus of this paper. System design is used to validate algorithms. viz. accuracy, precision, and recall. The dataset is initially split into two subsets for training and testing, each comprising 70% and 30% of the total. A machine learning model is created after the data has been partitioned, and its accuracy in classifying test datasets is then verified. In order to make the training set more robust, cross validation was applied later. An unknown or fresh sample is then provided to forecast whether an candidate is capable to shortlist or not after the results have been validated. To evaluate the effectiveness of an algorithm, various machine learning models are constructed and their accuracy is calculated. Out of the entire sample space, precision predicts how likely an Candidate will get shortlisted, and recall will show how many actually shortlisted for the company after recruitment. SVM, Random forest, and KNN all have accuracy rates of 96.5 percent, 98.3 percent, 93 percent, and respectively. All algorithms' levels of precision are respectively 75%, 10%, and d44%. In a similar vein, the recall rates are 72 percent, 17 percent, and 23 percent, respectively.

TABEL II

RESULTS

Models	Accuracy	Precision	Recall
SVM	98	97.5	98
Random Forest	98	98	98
KNN	99	99	99

B. Performance Evaluation

Figure 8. shows the performance of top 3 algorithms whose accuracies are calculated using training set and test set testing options.

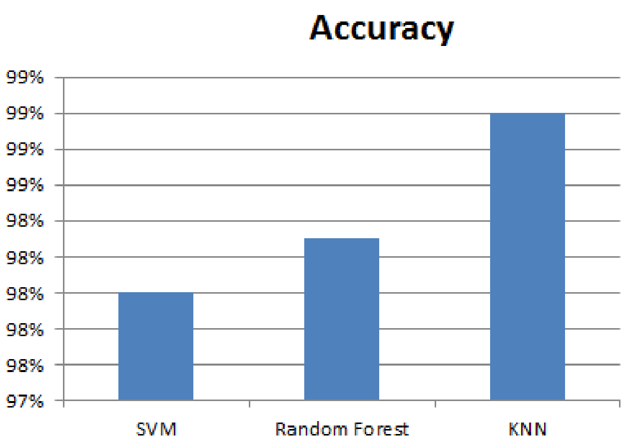

Fig. 7. Performance

Figure 9. shows the graphical representation of top 3 classification algorithms where error rates are calculated and represented.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⋇ Impact Factor 8.102 ⋇ Peer-reviewed / Refereed journal ⋇ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12626

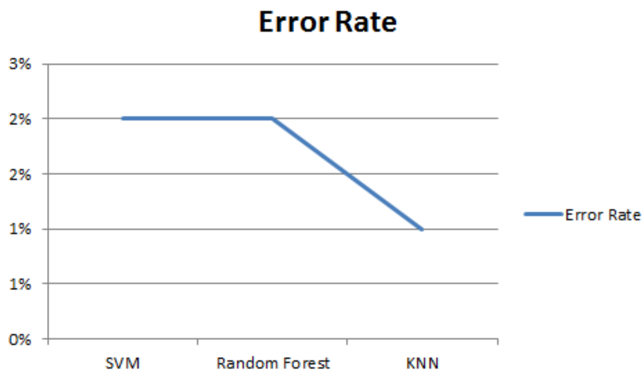

Fig. 8. Error rate

V. CONCLUSION

The pre-placement prediction system uses machine learning to assist employers in choosing qualified job candidates. According to the particular needs of the project, various machine learning algorithms can be used, including SVM, Random Forest, and KNN. The steps required to guarantee the accuracy and efficacy of the machine learning models include preprocessing the data, choosing pertinent features, and testing the algorithms. It is essential to choose the features that are most important in determining a candidate's suitability for the job. Testing is necessary to assess the effectiveness of the machine learning algorithms and pinpoint areas for development.

REFERENCES

[1] D. Jagan Mohan Reddy, Sirisha Regella, and Srinivasa Reddy Seelam. Machine Learning-Based Recruitment Prediction. 5th International Conference on Computing, Communication, and Security (ICCCS 2020), Patna, India, October 14–16, 2020. IEEE 2020, pages 1-4.

[2] Pothuganti Manvitha and [2]Neelam Swaroopa, "Campus Placement Prediction Using Supervised Machine Learning Techniques," International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 14, Number 9 (2019), pp. 2188-219 .

[3] Irene Treesa Jose, Daibin Raju, Jeebu Abraham Aniyankunju, Joel James, and Mereen Thomas Vadakkel. Placement Prediction Using Various Machine Learning Models and Their Efficiency Comparison Volume 5, Issue 5, May - 2020,

[4] K. Sreenivasa Rao and N. Swappa, P. Research Paper, International Research Journal of Engineering and Technology (IRJIET), 2018, Praveen Kumar Educational data mining for student placement prediction using machine learning algorithms.

[5] H. S. Bhatt and L. Mehta. R. D'mello, "Use of ID3 Decision Tree Algorithm for Placement Prediction," International Journal of Computer Science and Information Technologies (IJCSIT), vol. 6, pp. 4785-4789, 2018.

[6] Yogesh Bharambe, "Assessing Employability of Students Using Data Mining Techniques," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2017.

[7] J. M. Kirimi as well as C. A. Application of data mining classification in predicting employee performance by Moturi, International Journal of Computer Applications, vol. 146, no. 7, pp. 28–35, 2016.

[8] Thangavel, P. S.Bkaratki, and 8. Sankar, "Student placement analyzer: A recommendation system based on machine learning," International Conference on Advances in Computing and Communication Systems (ICACCS-2017). IEEE, 2017.

[9] Abhijith Shankar, Divya Bharathi P, and Senthil Kumar Thangavel. Student Placement Analyzer: A Machine Learning-Based Recommendation System, International Conference on Advanced Computing and Communication Systems (ICACCS-2017), January 6–7, 2017, Coimbatore, India.

[10] Professor Mangasuli Sheetal B. Savita Bakare, "Prediction of. Campus Placement Using Data Mining, Fuzzy Logic, and K Nearest Neighbors International. Advanced Computer Research Journal.Technology For Communication Vol. Issue No. 5, June 2016.