

Artificial Intelligence for Hiring

Ishan Borker¹, Ashok Veda²

AI Intern, Rubixe, Bengaluru, India
ishan[at]rubixe.com

CEO, Rubixe, Bengaluru, India
ashok[at]rubixe.com

Abstract: *Artificial Intelligence (AI) for hiring is a very important category under HR to reduce time-consuming activities like manually screening of resumes. AI in HR provides a simple way not only to reach out to possible candidates, but also screen, rank, and shortlist their resumes based on the traits most relevant to your company. This paper gives us the idea of predicting the number of students placed using machine learning algorithms. We have used various machine learning algorithms like XGBoost Classifier, Random Forest Classifier, Gradient Boosting Classifier, Decision Tree Classifier, Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Multi-Layer Perceptron (MLP) Classifier, and Extra Trees Classifier. The algorithms were used based on the number of features and for predicting students placed or not. We also compare the accuracy of different algorithms.*

Keywords: XGBoost Classifier, Random Forest Classifier, Gradient Boosting Classifier, Decision Tree Classifier, Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Multi-Layer Perceptron (MLP) Classifier, Extra Trees Classifier, Machine Learning, Artificial Intelligence (AI) and Accuracy

1. Introduction

Artificial Intelligence (AI) is one of the most important business opportunities for 2020. AI is changing our lives. This technology has the capability that is normally related to the human brain. AI for hiring is a very important category under HR to reduce time-consuming activities like manually screening the resumes. 76% of recruiters believe that AI's impact on recruiting will be somewhat significant. A recent study found that 46 percent of companies struggle with finding and attracting the right candidates for their open positions. AI programs can search for online resumes and social profiles to find the best candidates for each job based on specific traits. They can also relay personalized messages to promising candidates and do it on a scale that human recruiters could not do alone.

AI is being taught to overcome human biases during sourcing and screening. The key is teaching the program on data that presents as gender-neutral and training it to ignore other identifying information that might trigger biased decisions. An organization may end up with a pool of applicants far more diverse than if the HR team itself had sourced them. Once your AI program sources and contacts candidates, AI can lead them through the recruiting quickly and efficiently, ensuring the candidate experience goes smoothly. Recruiter chatbots can provide real-time answers to candidate questions, offer quick feedback, and suggest the next steps. They can provide links to promising job descriptions, clarify company hours and location, and schedule interviews.

AI-powered conversational tools can also give the screening process a boost. AI tools can store essential data on all applicants, saving time and effort when you're ready to reach out to them again. Companies that use AI tools have reduced their cost per screening by 75 percent. Using technology to screen talent also saves time and effort for candidates. The tool also brought a measurable improvement

in performance, training, new hire retention, and operational outcomes.

AI in HR provides a simple way not only to reach out to possible candidates, but also screen, rank, and shortlist their resumes based on the traits most relevant to your company. Then, once you have a list of people you would like to interview certain AI tools can also help you conduct a later-stage virtual interview before inviting a candidate to come in person. AI is also improving onboarding procedures by automating repetitive or tedious tasks like conducting background checks, putting together documents about benefits

Confirm the Opportunity

It is very important to sell this idea in the market as this is the most unique feature which is being created to improve the hiring activity in all the colleges. We are also assessing the primary opportunities to benefit recruitment using AI. Consider what other players are doing in recruitment using AI. Look for the business problem that can be solved using AI. Work in a collaborative environment with a dedicated team of specialists, domain expertise, and data scientists who have specialization in Deep Learning and Artificial Intelligence.

Advantages of using this opportunity of AI in recruitment are

- 1) It improves the quality and objectivity of recruitment.
- 2) It attracts the right candidates.
- 3) It saves time and fills the position faster.
- 4) It can also help in eliminating bias in the recruitment process.
- 5) It also improves the candidate experience.
- 6) It improves the quality of hire.

Characterize the Problem and Profile the Data**Problem Definition**

There are thousands of applications every year from candidates all over the country seeking admission to colleges and universities and it becomes a bottleneck for placement committee to pre-screen them by manually going through each profile. Our approach is a one-stop AI solution. To give fair interview chances to each candidate, we are building up this schematic model. This will outperform the traditional method of placement activity.

Who is facing the problem?

- 1) Placement officer from the college, employers who visit the campus for the placement activity.
- 2) Students are also not aware of how much time it takes to shortlist their profiles.

Where is the problem occurring?

Every college in the country is affected by traditional placement activity. It is difficult to figure out every college where training is given to the students.

Profiling the data

In this problem, we try to get the placement data from the placement cell of the engineering colleges and then try to use our algorithms to predict the students who are placed in a company.

2. Literature Review

In the white paper of Intel Data Centre Artificial Intelligence [1], five steps approach to generating a proof of concept (POC) was described for image recognition, natural language processing, and predictive maintenance. In students' placement prediction using machine learning [2], the authors analyzed the previous year's student's historical data and predicted the placement possibility of students for the current batch. In AICTE – CII Survey of Industry Linked Technical Institutes 2018 [3], the average performance of Tier II and Tier III engineering colleges was shown.

In AICTE Report on Management & Engineering Placements 2018[4], it was mentioned the average campus

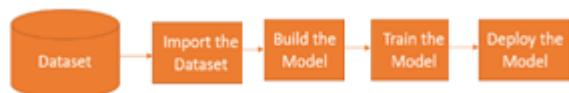
Flow diagram

Figure 1: Project Architecture

We now present our proposed application. We first describe the challenges faced in data collection, model training, and challenges for finding the students placed during recruitment activity. Then, we present extensive experiments and qualitative analysis to focus on the insights for deploying such a technology in the real world.

Steps:

- 1) Dataset: Collecting the raw data from the external sources in comma-separated value (CSV), Excel (xlsx) format. The dataset is the student information obtained

placement for the graduates from technical, management, and engineering institutes is below 60%. In India Skills Report 2019 [5], it shows that 47% of the engineering graduates are employable.

3. Methodology**Data Collection**

In this paper, the dataset is obtained from Nehru Group of Institutes as well as from the survey conducted using google survey. The description of the dataset is as follows:

- 1) Name: Name of the student
- 2) EmailID: email address of the student
- 3) Highest Education: Highest educational qualification of the student
- 4) Year of Passing: Year in which the student passed
- 5) Native Place: Hometown
- 6) SSC Marks: Aggregate marks scored by the student in the 10th standard.
- 7) HSSC Marks: Aggregate marks scored by the student in the 12th standard.
- 8) UG Marks: Aggregate marks scored by the student in the under-graduation course.
- 9) PG Marks: Aggregate marks scored by the student in the post-graduation course.
- 10) Active Backlogs: Number of active backlogs student is currently holding.
- 11) Placement Notification: Whether the placement cell of the college notifies students regarding the placement drive.
- 12) Training: Whether the training related to interviews is being provided to the students.
- 13) Communication Skills: English communication skills.
- 14) Analytical Reasoning: Ability to analyze the condition and situation in terms of key elements and various factors involved.
- 15) Logical Reasoning: Logical understanding of the concept.
- 16) Company Placed: Name of the company in which the student was placed.
- 17) Location: The place where the student was placed.
- 18) Position: Position for which the student was interviewed and placed.
- 19) CTC: Cost to company

from the college and also from the google form which is used to conduct the survey.

- 2) Import the dataset: Load the dataset in the local IDE.
- 3) Build the model: In this step, import all the necessary packages, clean the data, perform the insights, convert the categorical data into numerical using label encoder, and define input variables and the target variable.
- 4) Train the model: Use the train-test split to define the test size and then use the predict function for predicting accuracy, precision, recall, and F1 score. Use algorithms like XGBoost, Random Forest Classifier, Gradient Boosting Classifier, Decision Tree Classifier, logistic regression, support vector (SVM), K-Nearest Neighbours (KNN), Artificial Neural Networks (ANN), and Extra Trees Classifier.
- 5) Deploy the model: Once the final model is ready, we deploy the project on the server.

4. Results and Discussion

Correlation Matrix

The following is the correlation matrix for the number of features present in the student's data.

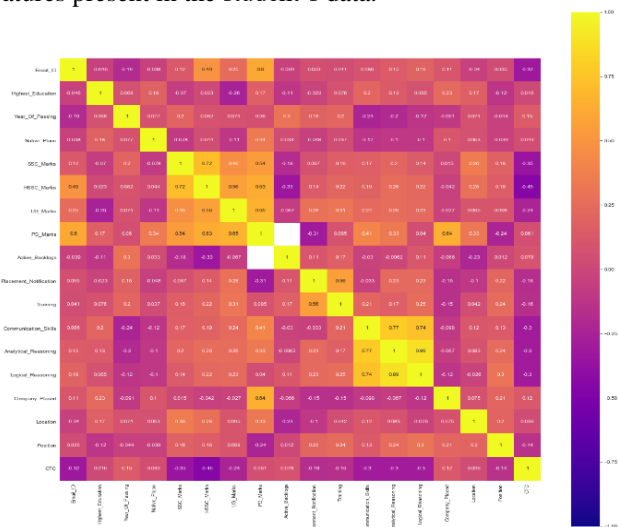


Figure 2: Correlation Matrix

Exploratory Data Analysis:

Exploratory Data Analysis (EDA) is a method to analyze the data sets using visualization. A detailed analysis of each feature was carried out here.

Bar plot for different features:

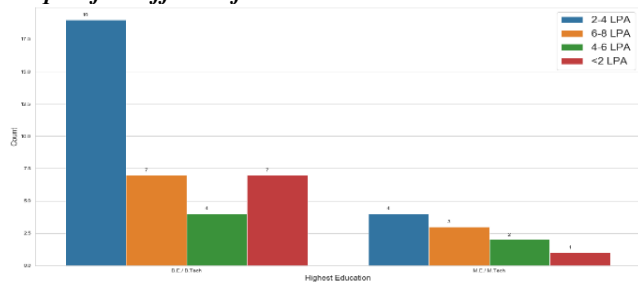


Figure 3: Count of students who received different CTC's received concerning the highest education

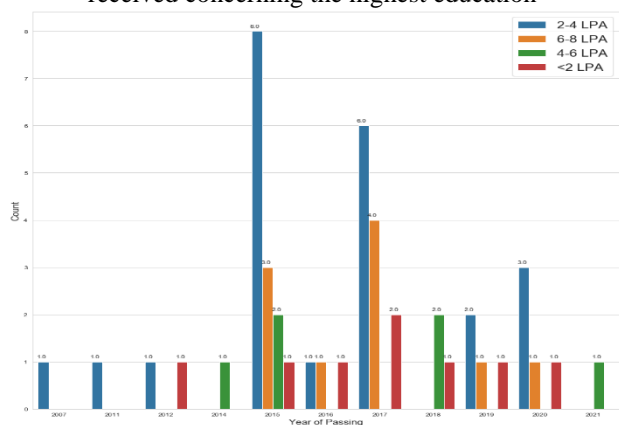


Figure 4: Count of students who received different CTC's received concerning the year of passing

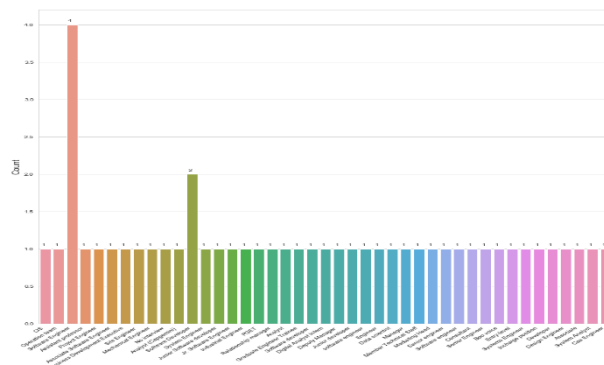


Figure 5: Count of students placed for different job roles in the organization

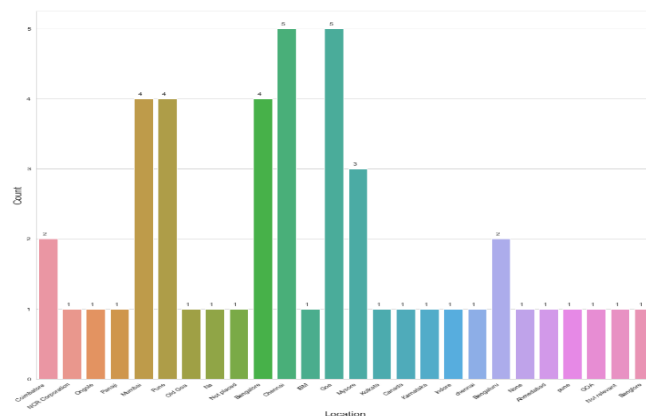


Figure 6: Count of students placed in various job locations

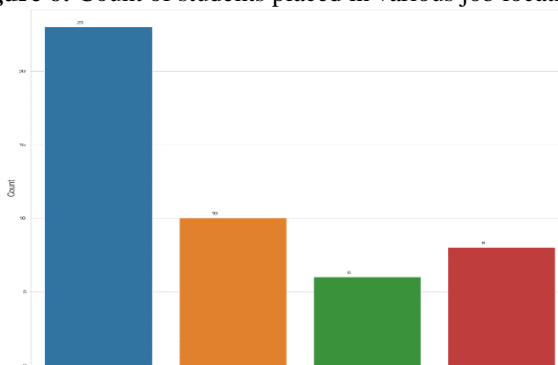


Figure 7: Count of students receiving different CTC's

Machine Learning Algorithms

XGBoost Classifier:

It is an implementation of gradient boosted decision trees designed for speed and performance.

Random Forest Classifier:

It is a supervised learning algorithm used for classification and regression. It creates decision trees on randomly selected samples of the data and performs the prediction on each tree. Then the best solution is selected using voting.

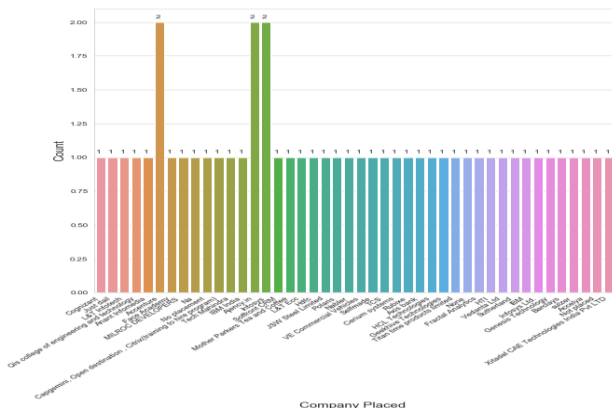


Figure 8: Count of students placed across companies

Gradient Boosting Classifier:

It is a group of machine learning algorithms which combines the weak machine learning models to create a strong predictive model. This can be used to handle complex datasets.

Decision Tree Classifier:

It is a supervised learning algorithm used to solve regression and classification problems. We need to have continuous, categorical input, and target features.

Logistic Regression:

It is a supervised learning algorithm for predicting the binary form of a target variable.

Support Vector Machine:

It is a supervised learning algorithm used in both classifications as Support Vector Classifier (SVC) and regression as Support Vector Regressor (SVR). This classifier separates data points using a hyperplane with a margin. Support vectors are the data points that are close to the hyperplane.

K-Nearest Neighbour:

It is an instance-based learning algorithm to approximate the function locally. It is used when the data is large and having non-linear decision boundaries between classes.

Multi-Layer Perceptron (MLP) Classifier:

It relies on a neural network to perform the task of classification.

Extra Trees Classifier:

It is an ensemble machine learning algorithm that combines the predictions from many decision trees. It has better performance than the random forest algorithm.

Table 1: Accuracy values

Algorithms	Accuracy	Precision
XGBoost Classifier	98.0%	98.04%
Random Forest Classifier	98.5%	98.52%
Gradient Boosting Classifier	98.5%	98.49%
Decision Tree Classifier	98.5%	98.49%
Logistic Regression	99.0%	99.01%
Support Vector Machine	98.0%	98.04%
K-Nearest Neighbour	96.0%	96.18%
Multi-Layer Perceptron Classifier	94.5%	94.46%
Extra Trees Classifier	91.5%	92.31%

Table 1 shows that the Logistic Regression algorithm gives 99% accuracy and precision which is the highest.

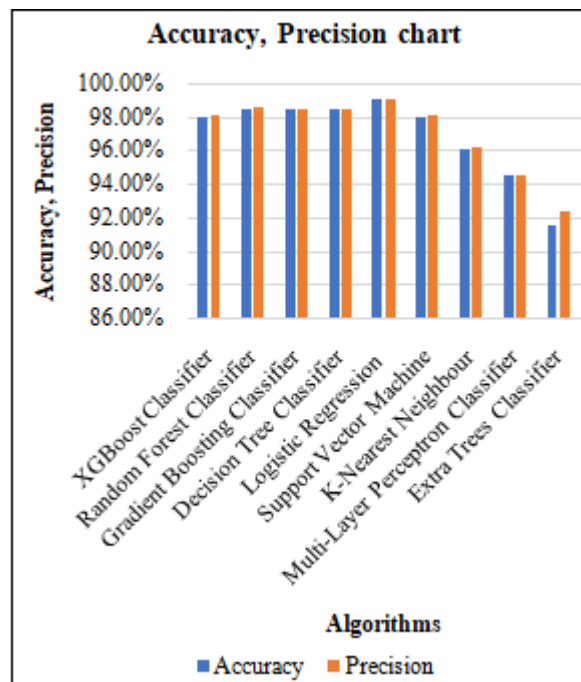


Figure 9: Comparing the values of accuracy and precision concerning different machine learning algorithms

5. Conclusion and Future Work

We can conclude that the accuracy for predicting the number of students placed is about 99% and the extra trees classifier gives the lowest accuracy and precision.

Future Scope: The data set which we worked on is very small. Therefore, we can add more characteristics of the dataset in the future. The features could be training is given, aptitude test marks, coding test marks, etc. We would also handle imbalanced datasets to make the predictions better. The features like how the candidate is performing in the company, appraisals can also be included to track the performance. We can build an AI program to smartly analyze candidates’ facial expressions, tone of voice, mannerisms, and word choice.

6. Acknowledgment

I have completed this project under the supervision of Mr. Ashok Veda (CEO), Rubixe, and Mr. Deepak D. (Program Manager), Rubixe. I would like to express my gratitude to both the mentors for allowing me to work as a Data Science Intern at Rubixe and encouraging me to write this paper. I would also like to thank Miss. Garima Jain (AI Intern), Rubixe for the assistance.

References

- [1] 5 Steps to an AI Proof of Concept, <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/ai-poc-whitepaper.pdf>
- [2] Shreyas Harinath, Aksha Prasad, Suma H S, Suraksha A, Tojo Mathew, (April 2019): Student Placement

Prediction Using Machine Learning, International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, Volume: 06, Issue: 04.

- [3] AICTE – CII Survey of Industry Linked Technical Institutes 2018, https://www.aicte-india.org/sites/default/files/AICTE-CII-Survey-Report-2018_1.pdf
- [4] <https://www.collegedekho.com/news/campus-placements-management-engineering-students-statistics-15132/>
- [5] <https://www.peoplestrong.com/india-skills-report-2019-47-engineering-graduates-employable-ap-leads-with-highest-employability-rate/>