Pre-placement prediction system using machine learning

Dr. (Mrs) S.R.Khonde[1], Siddhee Bagool[2], Siddhi Yeole[3], Pranjal Khose[4], Aarti sonawane[5]

Dept of Computer Enginering, Modern Education Society's College of Engineering, Pune, Maharashtra[1]

Dept of Computer Enginering, Modern Education Society's College of Engineering, Pune, Maharashtra[2-5]

Abstract: In this work we present a approach for shortlisting candidates for a particular job position. Every company invest a lot of money and time for recruiting people to fill some specific positions. They invest a lot of resources finding a appropriate candidate but the total invest become loss if the candidate do not join the company after completing the entire process.

The main aim of this research is to predict joining efficient candidate with minimum cost and time by using the attributes like number of hackathons given, skills, internship, salary expectation and preferred job location using the machine learning algorithm to build a model

Key Words: Machine Learning, decision tree, K-nearest neighbors, Random forest

1.INTRODUCTION

Recruiting the right people is the major aspect for every organization because the company growth is mainly depend on the employee performance. Every company is facing the problem for recruiting the appropriate candidate for the particular job profile. For example if xyz company wants the candidate who have knowledge of java and spring framework and they want the candidate who can join them PAN India but the existing system consider only the attributes like aggregate score, aptitude score, no of backlog so sometime the candidate who don't have much knowledge of java qualify the criteria because he/she is good in academics. Another candidate only want the location in Pune so he/she also not join the company.so are going add all this features to shortlist a candidate who satisfy all the criteria given by the company.

The rest of the paper organized as follows: The detailed literature review of different machine learning model are presented in Section II.

2. LITERTURE REVIEW

2.1 Prediction using K Nearest Neighbour, GBN

In 2020 Jagan Mohan Reddy D published work on predicting employee is important for the growth of the company..They have created a mock data set for this research using the combination of publicly available recruitment information and job selection factors based on several working analysis.The data collection includes diversified features such as gender, age, Marital Status, Years of Experience, Time in Current Organization, No of Companies Worked, Last Salary hike, Education, Job description, compensation, Years of promotion, Distance from home and a class label is whether employee is going to join or not in the organization. For classification task, all the nominal data converted into ordinal data in Python using one-hot encode technique. they obtained the result using the algorithm listed here decision tree, Random forest, Gaussian Naive Bayes, K Nearest Neighbour[1].They have compare the algorithms viz. accuracy, precision and recall. GBN is given highest accuracy, precision and recall among all other algorithms. Since, the dataset is don't have much interdependency among features which results Naive Bayes produces highest values.

2.2 Prediction using Random Forest and Decision tree

The paper[2] has sample data has been collected from their college placement department which consists of all the records of previous years students. The dataset collected by Pothuganti Manvitha, Neelam Swaroopa consist of over 1000 instances of students. The main attributes used for this study are credit , backlogs , whether placed or not, B.Tech , per. A general rule of the thumb is to assign 80per of the dataset to training set and therefore the remaining 20per to test set. a work has been analysed and predicted using the classification algorithms Decision Tree and the Random forest algorithm to validate the approaches. The algorithms are applied on the data set and attributes used to build the model. The accuracy

obtained after analysis for Decision tree is 84per and for the Random Forest is 86per. Hence, from the above analysis and prediction they have concluded that the Random Forest algorithm giving more accuracy compare to decision tree.

2.3 Prediction using SVM, KNN, Random Forest

The parameters in the dataset which are considered for the prediction are Quantitative scores, Logical Reasoning scores, Verbal scores, Programming scores, CGPA, No. of hackathons attended, No. of certifications and current backlogs number. The placement prediction is done by machine learning using Logical Regression, Random Forest, KNN, SVM. The data frame for the machine learning algorithm is created using pandas library based on the above sample dataset.. Based on the respective algorithm it predicts the placement of each student and accuracy can be viewed from the confusion matrix. They analyse the accuracy of different algorithms . It is clear that SVM gives an accuracy of 100. Logistic Regression is also good which gives an accuracy of 97.59 based on the given dataset. The accuracy of Machine learning algorithms may differ according to the dataset. From the result from our analysis it is clear that SVM, Logistic Regression, Random Forest, KNN are good for binary classification problems since they all give accuracy of above 95. Some recruiters consider GATE scores and history of backlogs which we didn't include in [3].

2.4 Prediction using Random Forest

This paper [4] proposes a method for predicting the employability status of the student using Random forest algorithm. The dataset for the work consists of scores collected from the students by conducting a test for them in the areas concentrated for recruitment process. Random Forest is an ensemble prediction method by aggregating the outcomes of the individual decision trees [4]. The accuracy of the model trained using this algorithm can be improved by tuning the algorithm parameters such as number of trees and number of the attributes that is selected randomly. A model is created using Random forest that can predict the likelihood of a student to be placed in a company. From the trained model the system can display the name of companies a student have chances to be placed based on their obtained scores. The system can also display a list of company seeking skills to be incurred by the students who are attending the placement process.

2.5 Prediction using ID3 Algorithm
This paper [5] proposes a model that predicts the probability of placement of a student in a company using ID3 decision tree algorithm. This system analyses the given dataset to identify the most relevant parameters required for placement prediction from the student dataset. Entropy and Information gain values of all parameters in the dataset is measured and the parameter with suitable measurement value is selected as split variable while building the decision tree. The Weka Tool generates an optimized decision tree with leaves representing the placement prediction chance of the student. The dataset comprises of marks obtained in secondary examinations, graduation grade points, arrear history and department type, details of various skills such as programming skill and communication skill, internships attended prediction. Using the sum of difference method, a reference value is computed corresponding to the selected attributes. If a student scores above this value indicates that student will get placed in the recruitment.

3. SOFTWARE REQUIREMENT SPECIFICATION

3.1   Purpose
     This software requirements specification provides a description of all the functions and constraints of the placement requirement system, developed for various colleges' placementcel, company.

          The pre-placement prediction system is for the college and companies which maintains the database for the students where all the students' records are entered including their academic, no. Of hackthons participate, internship in particular domain details and their personal details.

3.2 Scope of Product

The model would take all the academic as well as personal details of the students who wish to be placed.
From the literature survey we understand that we use below features for prediction which are increase the accuracy and this are more relevant.

- •   linkdeIn(skills,influncers)
- •   Github(project domain,no of projets)
- •   Internship
- •   Experience work in particular domain
- •   Salary expectation

- No of hackathon
- Location
- Blog(personality trait)

    Education stream be/btech/bsc/msc etc.

3.3 Product Perspective

In various colleges, training and placement officers have to manage the students' profiles and the documents of students for their training and placement manually.

The Pre-Placement Prediction system would maintain a huge data for the complete details of the students as well as the Companies in the Placement process which would help to save time and effort.

4. EXTERNAL INTERFACE REQUIREMENTS

4.1 User Interfaces
We are assuming that the user should have some basic knowledge of Computer.
Jobseeker should be from any fields.

4.2 Hardware Interfaces
The Project itself does not use any Hardware Interface but still it requires a system to run itself. As it is a Machine Learning on Web Based Implementation, this project can be run on any device with internet support. Supported device types include Mobile Phones, Smart Television, Desktops, Laptops etc. Communication protocols to be used is HTTPS as the web application will be hosted on a Web Server. The program will communicate with hard drive through the jupyter(filesystem). The user can communicate through form using keyboard and a display, displayed on user's screen.

4.3 Software Interfaces

The software interface uses a Anaconda(Linux, Jupyter,Python) set-up.
Operating system: Linux 4.0
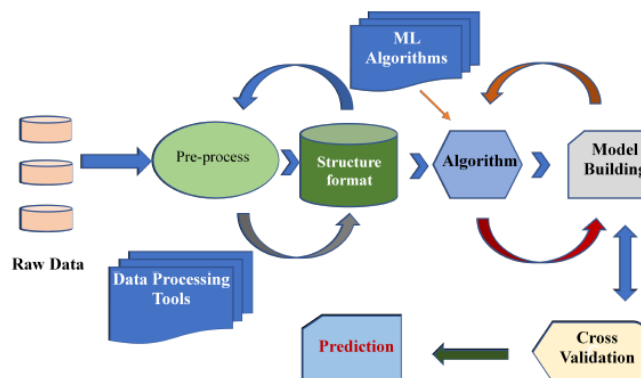Web Server: Anaconda3
Web Browser: Mozilla Firefox, chrome

4.4 comunication Interfaces
Endpoint enables the application to send data from the user to the model.
The prediction output from the model is then received back which is based upon the data inputted from the user.

5. PROPOSED SYSTEM

SYSTEM ARCHITECTURE

The above Fig 1 demonstrates mainly working of `1internal process. In this scenario overall system is designed in n tiers separately. The raw data taken as input, then pre-process the raw data by using data pre-processing tools. Therefore, the structured format of raw data is obtained by removing missing values, null values and outliers. After that, a one hot ending

technique is applied to enlarge the dataset during the classification process. Our study includes various machine learning algorithms to build models. The results were obtained and predicted the unknown test sample. The measure the perforce of our proposed method, the accuracy and precision is calculated. The given process is continuous until we will get better results from system.

Table -1:

List of attributes/features used in model		
Features	**description**	**type**
University PRN	Candidate university prn	ordinal
Age	Age	Ordinal
Education	B.S.C/B.TECH/B.E	nominal
Aggregate score	Aggregate sgpa or percentage can be consider	Ordinal
Year of experience	How many years candidate for any other organization	ordinal
No of backlogs	No of dead/current backlogs	Ordinal
Salary expectiation	Salary range expected by the candidate	ordinal
Intersnship	Internship done by candidate in academic	ordinal
Job location	Job location preferred by candidate	nominal

5.1 Data Collection

A mock data set for this research is created using the combination of publicly available recruitment information and job selection factors based on several working analysis.. The data collection includes diversified features such Education, No of hackathons given, Internship done during academics, language known, salary expectation, Job location. The details of attributes are listed in Table I.

5.2 Data Preprocessing

Collection of data from each and every Candidate who are applying for that particular position. Cleaning missing and null values in valid data from unwanted data. For classification task, all the nominal data converted into ordinal data in Python using one-hot encode technique. With this an unstructured data into structured format as input dataset to machine learning task. However, some of the features can't use directly for classification due to large variation of numbers which need to be converted into bins such as Age, Salary expectation. For this paper the binning size taken as 10.

6. CONCLUSION

Pre-placement Prediction System is a system which predicts the binary outcome for a candidate getting shortlisted for a company or not. For the data Analysis and prediction different machine learning algorithms are used in python environment. By applying K-Nearest Neighbors, Logistic Regression, Random Forest, and SVM(Support Vector Machine) algorithms we are going to compare the efficiency of the same with the proposed model. We have considered the number of attributes like Internships, Job location, salary expectations. And applied the KNN algorithm on it. By applying KNN algorithm we got 69 percent accuracy.

REFERENCES

[1] Jagan Mohan Reddy D , Sirisha Regella ,Srinivasa Reddy Seelam," Recruitment Prediction using Machine Learning",2020

[2] Pothuganti Manvitha,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 Raju2, Jeebu Abraham Aniyankunju3 , Joel James4, Mereen Thomas Vadakkel "Placement Prediction using Various Machine Learning Models and their Efficiency Comparison"Placement Prediction using Various Machine Learning Models and their Efficiency Comparison Volume 5, Issue 5, May – 2020

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

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

[6] Bharambe, Yogesh, "Assessing employability of students using data mining techniques", Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on. IEEE, 2017.

[7] Bharambe, Yogesh, "Assessing employability of students using data mining techniques", Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on. IEEE, 2017.

[8] . Senthil Kumar Thangavel, Divya Bharathi P and Abhijith Shankar. Student Placement Analyzer: A recommendation System Using Machine Learning, International Conference on advanced computing and Communication systems (ICACCS-2017), Jan 06- 07,2017, Coimbatore, INDIA.

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

[10] J. M. Kirimi and C. A. Moturi, "Application of data mining classification in employee performance prediction," International Journal of Computer Applications, vol. 146, no. 7, pp. 28–35, 2016