CAMPUS PLACEMENT PREDICTION SYSTEM USING BAGGING APPROACH.

Deva Rajashekar, Assistant Professor, CSE.

Kotha Amitabh, Kopanathi Sai Kumar, Singarapu Sahithi.

BE VIII SEM, CSE, Methodist college of Engineering and Technology, Abids, Hyderabad,500001

Abstract: Campus placement performs an essential position in each academic group in supporting college students to achieve their goals. Managing placement and education information in any large corporation is tough because the scholar quantity is high; in such a situation differentiation and classification on different categories will become tedious. All college students dream to get employment to provide by their hands earlier than they depart their college. A placement probability predictor allows college students to possess a thought regarding wherever they stand and what to be performed to obtain a decent placement. The foremost objective of this paper is to investigate the preceding year's student's historic data and predict placement opportunities of present day college students and aids in the growth of the placement percent of the institutions. Here we use machine learning ensemble algorithms, namely bagging algorithms. These algorithms independently predict the results, and we then examine the efficiency of the algorithms, which is predicated on the dataset. The result acquired from this can assist the students to better understand their weak regions to work upon. Working in those regions will permit college students to obtain a better number of placements in a college institution.

Keywords - Machine Learning, Prediction, Training, Ensemble learning, Bagging, Educational data mining.

I. INTRODUCTION

Every university puts its efforts into Placements that may be very important for every student. The success of the university is recognized through the number of college students obtaining placed in their Institutions. Nowadays academic institutes are developing in high numbers. Each better academic institute aims to get their college students a well-paid job via their placements. One of the biggest demanding situations that better learning institutions face in recent times is to reinforce students' placement overall performance. Placements are one of the most important tasks faced by a student in their lifetime. It is the responsibility of the colleges to offer the most placement chances to their students. In this context, consequently, the approach is to assume and compare the want for placement in institutions, that enables to construct institutions and students to enhance their placements. In this feature method, the chance that undergraduates might be put in an organization through applying machine learning ensemble algorithms predicts. The fundamental cause of this approach is to predict whether the student gets placed in campus recruitment or not. For this purpose, the statistics taken into consideration are student educational history as the entire number, backlogs, and credits. The algorithms are primarily based totally on student statistics from the previous year.

Ensemble learning

Ensemble systems are effective and versatile in solving real-world problems. The core part of ensemble learning is the strategy used to train individual classifiers. The starting idea is to use a set of classifiers to obtain a greater precision than each one of them achieves individually. Combining the output of several classifiers is useful only if there is disagreement between them. Ensemble learning is mainly implemented as two procedures: training weak component learners and selectively combine the member learners into a stronger learner. This article focuses on classification related applications of ensemble learning, however, all principal ideas described below can be easily generalized to function approximation or prediction type problems as well.

II. METHODOLOGY

Figure 1 Methodology

A. Data collection:

The data sets of college institutions are collected based on previous placement records; this model predicts the data sets of upcoming student batches with relative performances. Course grade point average (CGPA) for 4 consecutive years was gathered from the academic section.

Data associated with technical events participated, internships undergone, and certification courses completed were gathered from individual departments. As input for model prediction, data such as abilities required for students in getting placed in MRCs, Dream companies was gathered from the placement department for helping in placement prediction in the best way.

B. Data preprocessing and Feature Selection:

In this phase, redundant and missing values are recognized and corrected. The features that are significant in prediction are selected by using relevance measures like covariance. For Example, the student's technical knowledge feature is more important than the marks received in higher education. Data preparation is a step in a data analysis process in which data from one or more sources is cleaned, transformed, and enriched to improve the quality of data prior to its use. The gathered data were then pre-processed to fill the missing data and made compatible for in addition processing.

C. Training different modes and prediction:

The primary attributes taken into consideration in constructing this predictive model are marks scored in various tests of college students. The student's know-how on programming abilities, Numerical ability, reasoning, presentation skills are taken into consideration by taking an exam that results in the prediction. The classification of college students has recognized the use of ensemble learning methods namely bagging algorithms. The accuracy of consequences of each of the algorithms on education and check information units are compared.

III. LITERATURE SURVEY

In this paper, the author emphasized the mapping of students' overall performance patterns using a data mining approach with the help of a case study in Indonesia. K-means clustering turned into used for revealing the hidden patterns from the information. An overall of 306 valid instances has been used for the study. Various attributes selected have been GPA, Lecture grades, and Lab grades. The student's performance was classified into three classes i.e., low, average, and smart [1].

"Data Mining Approach for Predicting Student and Institution's Placement Percentage", Professor. Ashok M Assistant Professor Apoorva A ,2016 International Conference on Computational Systems and Information Systems for Sustainable Solutions. In this paper, the writer has used the data mining method for the prediction of the student's placement. For the prediction of college student's placement creator has divided the data into the two segments, the primary section is the training the section that's historic records of passed-out college students. Another section includes modern-day records of college students, primarily based totally on the historic records writer has designed the set of rules for calculating the position chances. The creator has used the diverse data mining algorithms which include decision tree, Naive Bayes, neural network, and the proposed set of rules were applied, and choices are made with the assist of a confusion matrix [2].

"Class Result Prediction using Machine Learning", Pushpa S K, Associate Professor, Manjunath T N, Professor and Head, Mrunal T V, Amartya Singh, C Suhas, International Conference on Smart Technology for Smart Nation, 2017. In this paper, the class results are predicted using machine learning. Performance of students in the past semester along with scores of internal examinations of the current semester is taken into consideration to predict whether the scholar passes or fails in the current semester before trying the final examination. The writer makes use of SVM, Naive Bayes, Random Forest Classifier and Gradient Boosting to compute the result. Boosting is an ensemble learning set of rules which learning set of rules to achieve higher predictive performance [3].

"Student Placement Analyzer: A Recommendation System Using Machine Learning", Apoorva Rao R, Deeksha K C, Vishal Prajwal R, Vrushak K, Nandini, IJARIIE-ISSN-(O)-2395-4396" challenges, the idea of machine learning and various algorithms are explored to are expecting the result of class students. For this purpose, the training data set is historical data of past college students and that is used to train the model. This software system predicts placement status in five categories to obtain dream company, core company, mass recruiter, now no longer eligible and not interested in placements. This system is also beneficial to weaker students. Institutions can offer more care toward weaker students so that it will enhance their performance. By use Naïve Bayes algorithm, all the data might be monitor and suitable decisions might be provided [4]

In this paper, the writer designed a data model for better management of student's performance, the use of a senior student's dataset. J48 algorithm was used, and the accuracy of the model was in comparison with other data mining algorithms. High accuracy of 97.27% was achieved by the J48 algorithm while in comparison to Naïve Bayes and multilayer perceptron i.e., 85.92% and 94.94% respectively. Selected features for the observation were CGPA, arrears, attendance, PUC marks, Engineering cut-off, Medium of education, and type of board. Performance classified notations given were good average and best [5].

EXISTING SYSTEM:

An educational institute includes student records that are a prosperity of data however is just too big for one person to understand in its entirety. Finding necessary traits from this data is an important mission in educational research. Finding the placement reputation of every student.

In the institute it is a hard challenge. Hence, the limitation of the system consists of time consumption, less efficiency, and less user satisfaction. Also, this system is a manual process that provides a limitation. some student placement prediction models consider the best academic performances of the students so that the prediction of the student getting positioned or not may be done. We cannot consider the placement of students just through their academic performances because some students may be good at aptitude, technical and communication skills due to the low score of their academics placed on may be done.

PROPOSED SYSTEM

Campus placement of students performs a very important function in educational institutions. Student placement prediction is the technique where organizations meet colleges and pick out college students who are talented and qualified before they complete

their graduation. So, this system makes the work of prediction of placement of students easy. We are developing a system in which the college students will Register/Login into the machine and input their biodata and skill sets, according to students' educational details the system will discover whether the student is eligible for the placement and recommend the courses to the students. Admin creates the publications and registers college students to the respective publications. Admin can view the publications and college students along their attributes. Admin predicts placement status of current students. If the student is eligible for placement mail will be dispatched to the student from admin and college students' names will be displayed on the dashboard in their colleges.

IV. IMPLEMENTATION AND WORKING PROCESS

SYSTEM ARCHITECTURE

The system architecture for this application requires a Database and a Backend to access the database. The Backend provides an interface for the web application and the database. This is also responsible for the prediction of results of the student placement test, that is the Machine Learning model can also be accessed only through the backend.
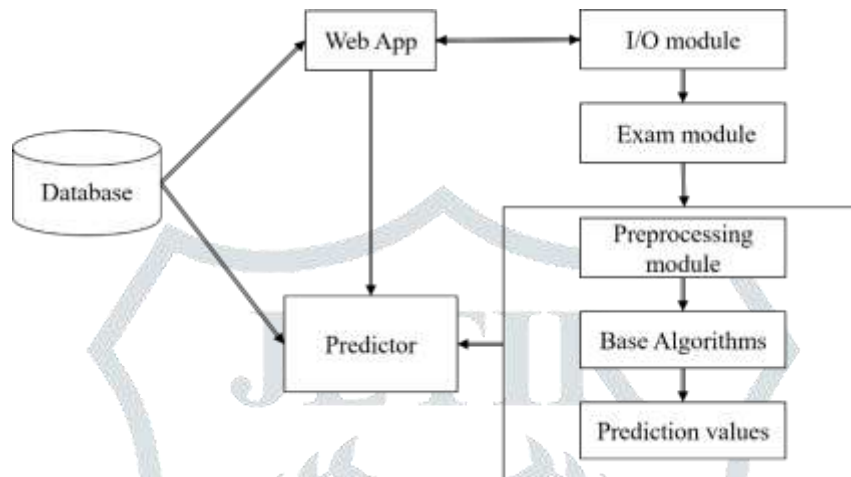

Figure 2 System Architecture

BAGGING ALGORITHM

The simplest method of combining classifiers is known as bagging, which stands for bootstrap aggregating, the statistical description of the method. A bootstrap sample is a sample taken from the original dataset with replacement, so that we may get some data several times and others not at all. The bootstrap sample is the same size as the original, and lots and lots of these samples are taken: B of them, where B is at least 50, and could even be in the thousands.

The name bootstrap is more popular in computer science than anywhere else, since there is also a bootstrap loader, which is the first program to run when a computer is turned on. It comes from the nonsensical idea of 'picking yourself up by your bootstraps,' which means lifting yourself up by your shoelaces, and is meant to imply starting from nothing. Bootstrap sampling seems like a very strange thing to do. We have taken a perfectly good dataset, mucked it up by sampling from it, which might be good if we had made a smaller dataset (since it would be faster), but we still ended up with a dataset the same size. Surely this is just a way to burn up computer time without gaining anything.

The benefit of it is that we will get lots of learners that perform slightly differently, which is exactly what we want for an ensemble method. Another benefit is that estimates of the accuracy of the classification function can be made without complicated analytic work, by throwing computer resources at the problem.

MODULES INCORPORATED DURING IMPLEMENRATION

Python : Python It is a language which will change the way you program and hence look at a problem. It is currently being used in diverse application domains. These include software development, web development, Desktop GUI development, education, and scientific applications.

Numpy : NumPy is a Python library that provides a multidimensional array object, and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

Sci-Kit Learn : Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Flask : Flask is a Python web framework built with a small core and easy-to-extend philosophy. Flask is considered more Pythonic than the Django web framework because in common situations the equivalent Flask web application is more explicit.

Pandas : Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, realworld data analysis in Python.

WORKING

The Application works based on Bagging Algorithms, these can be found in Sklearn package for python. The basic flow of the application is that the input data is taken from the user(student) and is then processed into the backend through the flask framework. The student data which is collected is stored in the database and the predictions are made based on the same data.

Figure 3 Home Page

The application flow begins with the registration of the student to the system. This allows for abstract working of the application from one student to another. Once the student is registered into the system, the student can take up the placement eligibility examination. The HTML GUI enables the student to login into system.
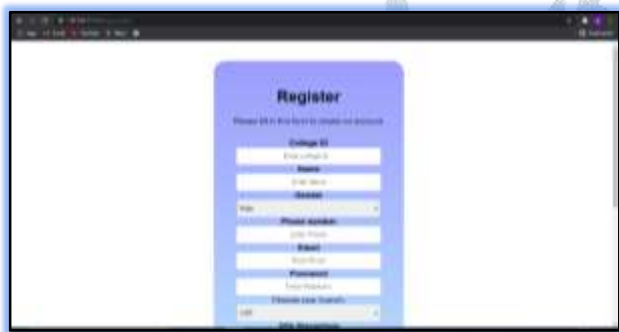

Figure 4 Registration Page

Figure 5 Login Page

To Provide the students with a dynamic web page, the Flask framework is used for this, which enables the web pages to be changed dynamically for each student. The examination is also generated dynamically each time for every student using the same. This examination is the primary input to the prediction model. The examination has 10 categories of topics each having their set of questions randomly being picked from the database.

The students are judged based on the questions they answer, and this will determine if the student can be placed or not. The system can store the results of the students and provide them with the necessary feedback they will need to clear the placements. The random forest algorithm provides the prediction of examination. The algorithm is running in the backend for the application and the Flask web framework provides a dynamic result each time the student takes the examination.
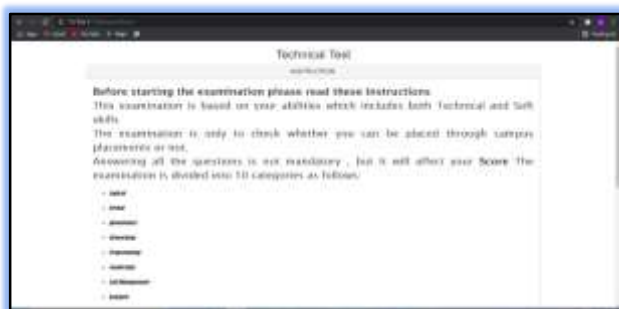

Figure 6 Exam Instructions

Figure 7 Exam Subjects

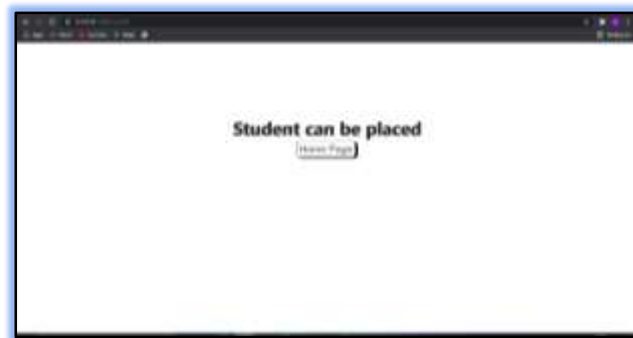Figure 8 Prediction - Cannot be placed   Figure 9 Prediction - Can be Placed

Figure 10 Previous Exam Results

The main objective of our system is to simplify the process of taking the placement examinations and help the students work on the fields which they are not good enough. The examination is a prediction of the student's current performance and helps them understand the aspects they need to improve for them to get placed in through the campus placements.

V. CONCLUSION AND FUTURE SCOPE

Campus recruitment activity is important from an institution point of view in addition to a student point of view. In this regard to enhance the student's performance, work has been analyzed and predicted using ensemble techniques namely bagging algorithms to validate the approaches. The algorithms are applied to the data set and attributes used to build the model. here we test the eligibility of students based on percentage & prediction test results. This will help college students to discover the category of organization in which they may be eligible and put together consequently in an experienced manner.

The future improvements of the project are to focus on adding a few extra parameters to predict more efficient placement status. We can also improve the task by predicting some solutions or suggestions for the output generated by the system. We can use extra optimized algorithms for better predictions. We can also combine online courses and offerings for college students to enhance their abilities and knowledge. The system can also be used to predict appropriate courses for better studies. Currently, this approach may be used as a complementary approach to get insights for placements of college students in the upcoming sessions. As the data of each coming batch can be stored and cleaned to be fed to the ensemble methods and obtain more examples for the prediction.

These new training examples carry out new characteristics that allow you to help the algorithm to examine different aspects for generating the skills that are demanded in the industry. Eventually, all the efforts put in solution to these insights will increase the placement record by a massive factor. Resources may be used on essential matters and on the students who are in want of it, that allows you to enhance the overall productivity. Also, this model can predict the lagging skills of every student. a high probability of student's placement, which is expressed in the form of newly learned parameters. These new parameters when applied to the classifier on the test dataset, ultimately will increase the accuracy of the version. Later, a few new functions can be added to the training data that prove to be an excellent predictor. After repeated modifications in the model and its data; more data can be fed to the algorithm, it can become the most powerful tool to get insights of students placed in the future with much more accuracy so that improvement in students' performance can be achieved timely, to improve the possibilities of getting placed

REFERENCES

[1] Karan Pruthi, Dr. Parteek Bhatia - "Application of Data Mining in Predicting Placement of Students".

[2] Ravi Kumar Rathore and J. Jayanthi2 "STUDENT PREDICTION SYSTEM FOR PLACEMENT TRAINING USING FUZZY INFERENCE SYSTEM".

[3] Mansi Gera, Shivani Goel- "A Model for Predicting the Eligibility for Placement of Students Using Data Mining Technique".

[4]"Class Result Prediction using Machine Learning", Pushpa S K, Associate Professor, Manjunath T N, Professor and Head, Mrunal T V, Amartya Singh, C Suhas, International Conference on Smart Technology for Smart Nation, 2017.

[5] "Data Mining Approach for Predicting Student and Institution's Placement Percentage", Professor. Ashok M Assistant Professor Apoorva A ,2016 International Conference on Computational Systems and Information Systems for Sustainable Solutions.

[6] Siddhi Parekh, Ankit Parekh, Ameya Nadkarni, Riya Mehta, "Results and Placement Analysis and Prediction using Data Mining and Dashboard".

[7] Getaneh Berie Tarekegn, Dr. Vuda Sreenivasa Rao- "Application of Data Mining Techniques to Predict Students Placement into Departments".

[8] Ankita A Nichat, Dr. Anjali B Raut - "Predicting and Analysis of Student Performance Using Decision Tree Technique".

[9] "A Placement Prediction System Using K-Nearest Neighbors Classifier", Animesh Giri, M Vignesh V Bhagavath, Bysani Pruthvi, Naini Dubey, Second International Conference on Cognitive Computing and Information Processing (CCIP), 20164.

[10] Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar, "PPS - Placement Prediction System using Logistic Regression" in 2014 IEEE International Conference on MOOC, Innovation and Technology in Education