**PLACEMENT PREDICTION MODEL**

**ABSTRACT:**

Placement is an ultimate short term or long term goal of every graduate or undergraduate to maintain a healthy life style.The 20-24 age group is one of the largest groups of unemployed people, of which college graduates constitute a big portion.Getting placed in the desired domain is somewhat a hard task as it need a proper guidance, great motivation, good time management, a strong desire and a strict preparation.Campus placements for engineering students crossed the 40% mark this year (2022).The biggest challenge that institutions face today is improving students placement performance.Institutions can drastically reduce the number of unemployed graduates by introducing courses and changing the curriculum to help develop the skills that employers look for in graduates. Hence a good placement for a student deserves a better analysis on his/her skills and interests accordingly. This paper facilitates the students to predict the chance of getting placements based on parameters like the skills,experience using various classification techniques like Decision Tree, Random Forest , Support Vector Machine. To yield best results the data set is collected with factors like learning skills, speaking skills and some domain based skills.A relative analysis brings out that SVM(Support Vector Machine) is most suitable for placement prediction with accuracy 78.37%.The study revealed that the number of internships done by a student plays a key role as internships guarantees the experience in particular role which is mostly preferred by the employer. This work will be helpful to the students to analyze their capability in preparing for placements and this will also be helpful to the management of the institution to provide the best resources for the students.

**Keywords:**Pre-processing,data transformations,classification, prediction,data transformation.

1. **Introduction**

Employability plays a major role in society as in boosting up the economy as it directly related with the industrial implementation.Key concern of every student as well as university is a good opportunities to work with.However the education is becoming more employment oriented, the skills needed for the placements should be build with care.

Education institutions collects and may generate large amount of data.This data can have students personal profiles , faculty profiles and academic book records etc.This large data set must be used wisely in predicting the placement which would assist student to prepare and plan accordingly for better placements.

Algorithm that suits best for the prediction should be selected before diving into the prediction making.In this paper we have applied several classification techniques and compared them using accuracy.This paper would let you know every piece of placement prediction clearly.

1. **Related Work**

Varsha and Bhalke’s[1] paper deals with the prediction of student placement performance using various classification algorithms.The paper compared with deep learning technique-MLP and machine learning technique-SMO, simple logistic, in terms of accuracy, precision, true positive rate, FPR, ROC area, recall, incorrectly classified instances, and time taken to build model. The algorithms that gave optimal results are MLP, SMO, simple logistic, LMT with maximum accuracies 99.5%, others all gave up to 96%. The proposed model proved to be the best predicting model for solving placement chance prediction problems compared to all other algorithms.

1. Yogi reddy and Gokul Prathin Asamani[2] first trained the data and then tested with all four algorithms, and out of all, the Decision Tree Classifier gave more accuracy with 87.4%, XG Boost with 85.3%, Logistic regression with 86.4%, and Support Vector Machine with 84.9% percent accuracy. So we can conclude that the Decision Tree Classifier is the best choice for data prediction because of having its high accuracy.

Monarsh Patil and Jayant Naik[3] had been analyzed and predicted using the classification algorithms Decision Tree and the Random forest algorithm are used 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 84% and for the Random Forest is 86%. Hence, from the above said analysis and prediction its better if the Random Forest algorithm is used to predict the placement results.

Abhishek S. Rao, Aruna Kumar S V, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B, Prashanth Gouda[4] used three best-suited classification algorithms like ANN, KNN and SVM.From the experimental analysis, it was evident that the results obtained for the ANN algorithm were best compared to other two classifiers when 80% of the data was trained and 20% of the data was tested using Tanh activation function. Similarly for the KNN algorithm when 70% of the data was trained and 30% of the data was tested with 7 neighbours, best results were achieved. Whereas in the case of SVM algorithm best results were got for Radial Basis Function (RBF) kernel for 70:30 proportion. Amongst the three classification algorithms, ANN has given the best results with an accuracy of 99.02% for 80:20 proportions with Tanh activation function.Concluded that the proposed model could be implemented with the help of other classification algorithms also. The Present study focused on four-year data and was restricted to few major departments with limited features; hence, if the data set could be further strengthened, the model could be successfully utilized for other educational institution related applicationslike planning of courses and result analysis which is the biggest challenge for any educational institution.

Jahnvi Shah1 and Shivangi[5] determined that the SVM model gives an accuracy of 87% and the XGBoost model gives an accuracy of 90%.They found that the technical skills, projects, certified courses taken and the internships of the student matter the most for predicting if they will get placed or not.

Pothuganti Manvitha and Neelam Swaroopa[6] discussed that a work has been analyzed 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 84% and for the Random Forest is 86%. Hence, from the above said analysis and prediction it’s better if the Random Forest algorithm is used to predict the placement results.

Dr. Rajiv Suresh Kumar1 and Fathima Dilsha[7] claimed that from the study it is clear that the student dataset containing academic and placement details are a potential source for predicting the future placement chances and It is clear that SVM gives an accuracy of 100. This prediction can enlighten students to identify their capabilities and improve accordingly. This system also helps in the academic planning of an institution to prepare proper strategies and improve the placement statistics for the future years.

Shreyas Harinath and Aksha Prasad[8] said that proposed work contains many educational parameters to predict placement status which will be more accurate.

1. Shukla1 and A.K. Malviya [9] proposed that several clustering techniques are compared and EM technique is selected for obtaining the clusters of student instances. This cluster prediction is appended to the student dataset. This valuable information improves the processing of Naïve Bayes Techniques and as a result its accuracy increased from 95.99% to 98.75%. Hence, the combined approach of EM clustering and Naïve Bayes classification is added to modify the model with a few new stages. This modified model in the paper has improved efficiency in classifying and predicting the placement chance of students during campus selection. In the future scope, association rule mining and other machine learning languages, techniques and tools may be used to apply this model to a large scale of student data.

Tripti Mishra,Dharminder Kumar and Sangeeta Gupta[10] has identified factors affecting employability and then has applied and compared various classifiers. The effect of some of the emotional skill parameters on placement has been established where others have not shown so much effect as per expectation.

V.Ramesh, P.Parkavi and P.Yasodha[11] issued a paper that says from the results it is proven that MultiLayerPerception algorithm is most appropriate for predicting student performance. MLP gives 87% prediction which is relatively higher than other algorithms. This study is an attempt to use classification algorithms for predicting the student performance and comparing the performance of NaiveBayesSimple, MultiLayerPerception, SMO, J48, and REPTree.

Table 1.1: Analysis Of Other Papers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Name of the model | Algorithm used | Parameters considered | Accuracy |
| 1 | Student Placement Prediction System using Machine Learning | Multi Layer Perceptron | Branch,Gender,10th%,10th board ,12th%, 12th board, Diploma%, FE-I SEM%, FE-II SEM% ,SE-I SEM%, SE-II SEM% ,TE-I SEM%, TE-II SEM%, Agg.Engineering, Live backlogs, Year down, Education gap, Company selected, Placed/not placed | 99.5% |
| 2 | Student Placement Prediction | Decision Tree Classifier | Kaggle data set | 87.4% |
| 3 | FUTURE PREDICTION OF STUDENT PLACEMENT USING MACHINE LEARNING | Random Forest | Not mentioned | 86% |

1. **PROPOSED MODEL**

The proposed work is to use the two step method for predicting placement chance of students.The first step is to clean and prepare the data , apply and compare various classification algorithms on the placement data set of the students.The second step is to make a model with SVM classification technique.SVM is used because it is the maximum accuracy as compared to all other classification techniques.Jpyter Notebook is used for performing the classification and prediction.

* 1. **METHODOLOGY:**

The methodology used in this involves steps namely Data Collection, Data Pre - processing, Data classification and Data Interpretation.

|  |
| --- |
| END  Not Placed  Placed  RESULT  START  Data transformations  Data cleaning  Data visualization  PRE PROCESSING  DATA Collection  Model Evaluation  Model Selection  KNN  SVM  NavieBayes  Logistic regression  Rndom Forest  Decision Tree  Classification |

* 1. **Data Collection**

A simple data set is collected form the college students of final and pre final year.

* 1. **Data Pre-processing:** Preprocessing yields the definition of preparing the datasets for model building.Again this involves in various steps like

**3.3.1 Data Cleaning** :

In some cases datasets might contain some missing or null values.All these values should be handled carefully for correct prediction results.Usually we do remove the missing values.But sometimes it leads to data loss and the major issues with data removal is finding whether we are removing incorrect or correct data?Later,we need to choose the optimized ways to clean the data.

**3.3.2 Data Transformation :**

Method of converting the format of data is called Data transformation.For better analysis and easy calculations, predictions and classifications we require whole data to be of same type.Understanding the data set becomes easy and it further leads to accurate predictions/results.

**3.3.3 Data Visualization:**

Pictorial representation of data is called as Data Visualization.To detect data relations and distributions the data set is plotted using ‘matplot.lib.pypolot’.Various plots like scatter,pie chart and bar chart are analyzed.Attribute selection is done using the visualization results.

* 1. **Classification:**

As the model deals with predicting whether a student gets placement or not several classification methods are used.

Used methodologies:

* Decision Tree Classifier

A hierarchical classification techniques to classify the nominal data.At each and every stage multiple decisions made regarding the features of the data.This yields the classification results based on the individual features of data records.

Accuracy with decision tree classifier: 63.82

* Random Forest

Integrated version of decision tree classifier.A classification algorithm that labels the data records based on the labels made by multiple decision trees.The final output can be calculated using the mean,median or any statistical measure of the outputs of all decision tree classifiers.

Accuracy with Random Forest : 43.53%

* K-Nearest Neighbour Classifier

The labels that are assigned using KNN classifier would highly depends on the neighbourhood measure.It assumes that data of similar kind typically have similar features.Distance between the data items is the prioritized measure in KNN Classifier.

Accuracy with KNN Classifier : 76.59

* Support Vector Machine Classifier

One of the most popular supervised learning algorithm used for both regression and classification.It aims to create the best line or decision boundary that segregate n-dimensional space into categories.The best suited boundary is called as ‘hyperplane’.SVM chooses extreme points to build the hyperplane.Hence the extreme points are called as vectors.It can work efficiently on multi media data.

Accuracy with SVM : 82.97

* Naive Bayes Classifier

Naive Bayes is an efficient algorithm for many real time data applications.This algorithm works based on the Bayes Theorem with an assumption of independence among predictors. The creative Naive Bayesian technique is based on the conditional probability and the maximum likelihood incidence. The formula for Baye’s theorem is given as:

P(A/B)=(P(B/A)\*P(A))/P(B)

Where P(A/B) is posterior probability, P(B/A) is likelihood probability, P(A) is prior probability, and P(B) is a marginal probability.

* 1. **Model Selection**

Model that gives the best accuracy is taken for further predictions.For placement prediction model that trained with SVM is considered.

* 1. **Mode Evaluation**

Selected model is evaluated using the testing and training data along with some accuracy measures.

**4 Data set**

Table 1 : List of student attributes and their respective values based on data collection.

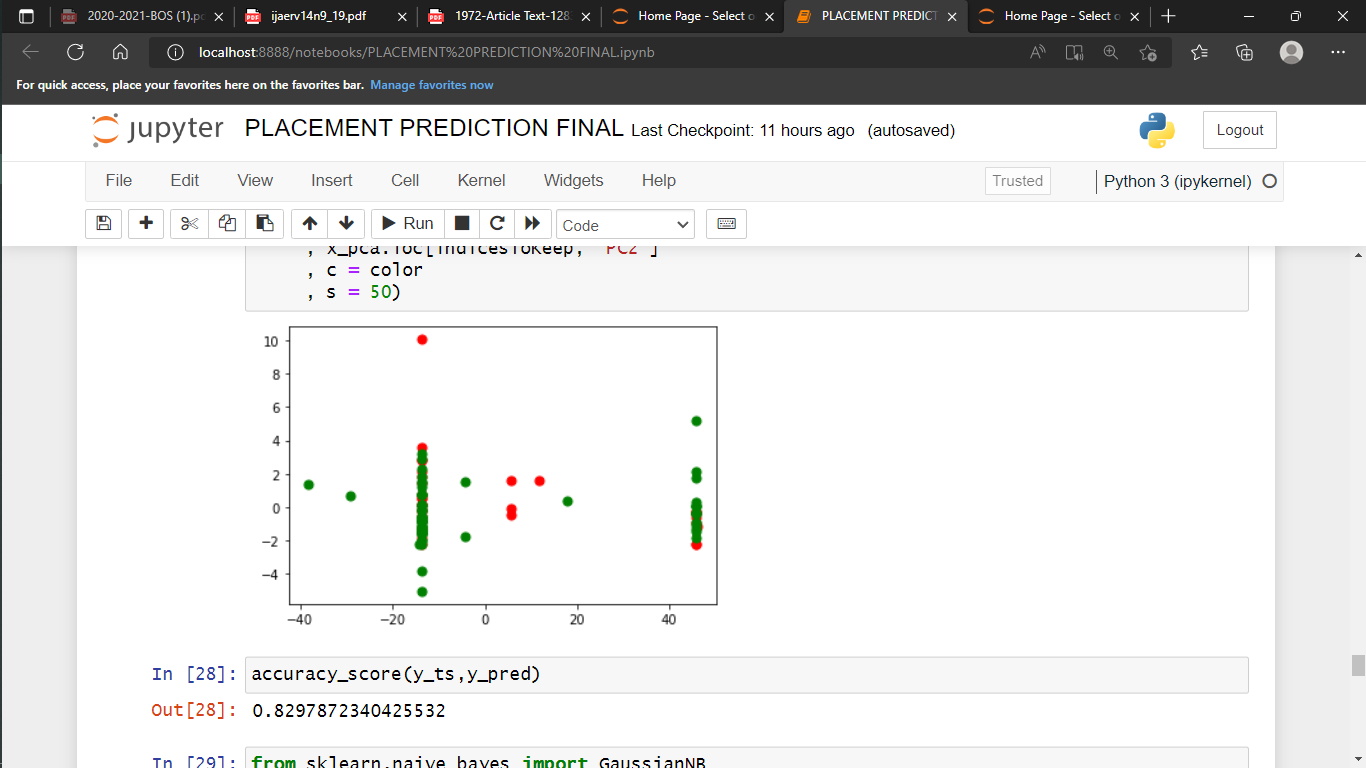
|  |  |
| --- | --- |
| Attribute Name | Attribute Value |
| JOB | Yes , no |
| CGPA | Numerical value |
| INTERVIEW COUNT | Numerical value |
| JOB COUNT | Numerical value |
| INTERNSHIP COUNT | Numerical value |
| Mode of Internship | Online , Offline |
| Stipend For Internship | Numerical value |
| Mode of Interview | Online,Offline |
| Communication skills | Numerical value |
| Presentation skills | Numerical value |
| Study time (in hours per day) | Numerical value |
| COMAPANY NAME | Name of the company |
| Mode of the entrance exam | Online , Offline |
| No of rounds in selection process | Online , Offline |
| Skills required for the job | Set of skills |
| Related skills you have for the job | Set of skills |
| Company open to all | Yes,no |

**Partial Data set with first 10 attributes**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| JOB | CGPA | INTERVIEW COUNT | JOB COUNT | INTERNSHIP COUNT | Mode of Internship | Stipend for internship() | Mode of Interview | Rate your communication skills (out of 10 like 8.9,2.9) |
| No | 9.1 | None | None | None |  | None |  | 5 |
| Yes | 8.6 | 5 | 5 | 2 | Online | 0 | Online | 10 |
| No | 8.27 | 2 | 0 | 1 | Online | 0 | Online | 8 |
| Yes | 9 | 1 | 1 | 1 | Online |  | Online | 8 |
| Yes | 8 | 15 | 3 | 2 | Online | 2,000,050,000 | Online | 9 |
| Yes | 84 | ⁴ | ² | 1 | Online | 0 | Online | 8 |
| Yes | 8.6 | 10 | 5 | 1 | Online | 0 | Online | 8 |
| Yes | 9.2 | 5 | 4 | 2 | Online | 0 | Online | 8.5 |
| Yes | 8.4 | 1 | 1 | 0 | Offline | 15000 | Online | 9 |

**REULTS AND CONCLUSION**

**SVM REULTS**



Collected data set contains over 1000 records and 18 attributes.The data set is in excel format.This file given to jupyter notebook for further porcessing.Data pre-processing is the first step in the evaluation of this project. For this project, choose Jupyter Notebook interfaces for the classifying model. Here, choosen the SVM classifier.

The algorithms that gave optimal results are SVM, and Naive Bayes Classifier with maximum accuracies 82.9%.

CONCLUSION:

This paper deals with the placement prediction of students.This study paper helps the institution to know the placement prediction of the students in advance.The proposed model proved to be the best predicting model for solving placement chance prediction problems with 82.9%accuracy compared to all other algorithms.

**References :**

1. Varsha and Bhalke’s **”**Student Placement Prediction System using Machine Learning”.
2. M.Yogi reddy and Gokul Prathin Asamani “Student Placement Prediction”.
3. Monarsh Patil and Jayant Naik “ FUTURE PREDICTION OF STUDENT PLACEMENT USING MACHINE LEARNING”.
4. Abhishek S. Rao, Aruna Kumar S V, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B, Prashanth Gouda“Student Placement Prediction Model: A Data Mining Perspective for Outcome-Based Education System”.
5. “Campus Placement Prediction” Jahnvi Shah1 , Shivangi Kochrekar2 , Neha Kale3 , Sakshi Patil4 , and Anand Godbole5 Computer Engineering Department, Sardar Patel Institute of Technology, Mumbai, India.
6. Pothuganti Manvitha and Neelam Swaroopa ”Campus Placement Prediction Using Supervised Machine Learning Techniques”.
7. “Student Placement Prediction Using Support Vector machine Algorithm“ Dr. Rajiv Suresh Kumar1 , Fathima Dilsha2 , Shilpa A N3 , Sumayya A A4 Department of Computer Science and Engineering, JCT College of Engineering and Technology, Coimbatore, TamilNadu, India1-4.
8. “STUDENT PLACEMENT PREDICTION USING MACHINE LEARNING”

Shreyas Harinath 1, Aksha Prasad2, Suma H S3, Suraksha A4, Tojo Mathew 5 1,2,3,4BE, Department of CSE, NIE Mysore, Karnataka, India 5Assistant Professor, Department of CSE, NIE Mysore, Karnataka, India.

1. “Modified Classification and Prediction Model for Improving Accuracy of Student Placement Prediction” by M. Shukla1 and A.K. Malviya.
2. “Students’ Employability Prediction Model through Data Mining”by Tripti Mishra,Dharminder Kumar,Sangeeta Gupta.
3. “Performance Analysis of Data Mining Techniques for Placement Chance Prediction” by V.Ramesh, P.Parkavi, P.Yasodha