JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Campus Placement Prediction Using Machine Learning

[1]Shriniwas Bhagat, [2]Sanskar Dhavale, [3]Navin Hegde, [4]Girish Jadhav, [5]Priyanka Patil
[1234]Student, [5]Professor
[12345]Department of Artificial Intelligence and Machine Learning Engineering,
[12345]Universal College of Engineering, Mumbai, India

Abstract : A campus placement prediction system is developed to calculate the possibility of a student getting jobs in a company through campus placements. The model takes many parameters which can be used to get an idea about the skill level of the student. While some data are taken from the college level like academic performance, CGPA, pointers, attendance etc others are obtained from tests conducted in the placement management system. Combining these data points, the model is to accurately predict if the student will or will not be placed in a company. Also, Data from past year students are used for training the model. We are using educational data mining by which we can get real past year's students data of that specific college. This will help the machine learning model to be more effective about the predictions of that particular college. The purpose behind this placement prediction system is to help students improve their results, academic performance and also develop another soft skill that will help them to maximize their chance of getting placed. Such a study will help the faculties of the college to train the students accordingly and improve the placement department of their institutions. This will give an idea about how students are doing and will ensure their institution can satisfy the needs of recruiters. For this supervised machine learning especially logistic regression is better, Logistic model designing plays a key role to get correct predictions. This process includes the selection of tuples for training data and their preknown outcome often known as real data.

IndexTerms - Data Mining, Campus Placement, Logistic Regression, EDM, Supervised Learning, Logistic Function, Machine Learning

I. Introduction

Placements are considered to be very important for each and every college. The basic success of the college is measured by the campus placement of the students. Every student takes admission to the colleges by seeing the percentage of placements in the college. Hence, in this regard the approach is about the prediction and analyses for the placement necessity in the colleges that helps to build the colleges as well as students to improve their placements. In Placement Prediction system predicts the probability of a undergrad students getting placed in a company by applying classification algorithms such as Decision tree and Random forest. The main objective of this model is to predict whether the student he/she gets placed or not in campus recruitment. For this the data consider is the academic history of student like overall percentage, backlogs, credits. The algorithms are applied on the previous year's data of the students.

II. Literature Survey

In this section, we reference, to the best of our knowledge, some studies related to the utilization of Artificial Intelligence (AI) and Machine Learning (ML) for enhancing financial management systems. This exploration delves into the cutting-edge technologies and their potential to revolutionize HRMS finance tools, investment analysis, and valuation prediction.

In [1], "A Comparative Study of Machine Learning Algorithms for Campus Placement Prediction" by R. Gupta, N. Sharma, and A. Singh (2020), the researchers conducted a comparative study of various machine learning algorithms for campus placement prediction. The study evaluated algorithms such as logistic regression, decision trees, k-nearest neighbors (KNN), and neural networks based on accuracy metrics and computational efficiency. While providing valuable insights into different ML algorithms, the paper identified a gap in the exploration of feature engineering techniques and the incorporation of unstructured data, which could enhance prediction accuracy.

In [2], "Campus Placement Prediction System Using Machine Learning Techniques" by Ankit Yadav, Nitin Pandey, and Priyanka Gautam (2021), the authors proposed a campus placement prediction system using machine learning algorithms such as decision trees, random forest, and k-nearest neighbors (KNN). The system utilized features like academic performance, skills, and demographics to predict placement outcomes. This work fills a gap by applying machine learning to campus placement prediction, showcasing the potential for more accurate and data-driven decision-making in student placements.

In [3], "Predicting Campus Placements using Machine Learning Techniques" by Divya Sharma and Vipin Kumar Sharma (2022), the researchers explored the use of support vector machines (SVM) and artificial neural networks (ANN) for predicting campus placements. Factors such as academic performance, technical skills, and communication skills were considered for prediction. This research contributes by comparing the performance of SVM and ANN for campus placement prediction, offering insights into the effectiveness of different machine learning approaches in this domain.

In [4], "Campus Placement Prediction using Machine Learning Algorithms" by Snehal Borse, Pranav Dixit, and Kunal Gaikwad (2023), the paper presented a comparative analysis of machine learning algorithms including logistic regression, decision trees, and gradient boosting machines (GBM) for campus placement prediction. Academic scores, aptitude test results, and interview performance were among the factors considered. The study addressed the gap by evaluating multiple machine learning algorithms specifically for campus placement prediction, highlighting the strengths and weaknesses of each algorithm.

In [5], "Predictive Analysis of Campus Placements using Machine Learning" by Rohit Kumar and Pooja Singh (2024), the paper employed a hybrid approach combining decision trees and genetic algorithms (GA) for campus placement prediction. A wide range of features such as academic records, technical skills, extracurricular activities, and personality traits were considered. This research fills a gap by introducing a hybrid machine learning model for campus placement prediction, leveraging the strengths of both decision trees and genetic algorithms to enhance prediction accuracy.

III. METHODOLOGY

3.1 Dataset

The dataset we are using is a placement.csv file, imported from Kaggle. The dataset consists of 215 rows and 15 columns. The first five rows of the dataset are as follows:

```
data.head()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
0	1	0	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	270000.0
1	2	0	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000.0
2	3	0	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	250000.0
3	4	0	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	NaN
4	5	0	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	425000.0

Fig-1: Dataset Statistics 1

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   sl_no           215 non-null    int64
 1   gender          215 non-null    int64
 2   ssc_p           215 non-null    float64
 3   ssc_b           215 non-null    object
 4   hsc_p           215 non-null    float64
 5   hsc_b           215 non-null    object
 6   hsc_s           215 non-null    object
 7   degree_p        215 non-null    float64
 8   degree_t        215 non-null    object
 9   workex          215 non-null    object
 10  etest_p         215 non-null    float64
 11  specialisation  215 non-null    object
 12  mba_p           215 non-null    float64
 13  status          215 non-null    object
 14  salary          148 non-null    float64
dtypes: float64(6), int64(2), object(7)
memory usage: 25.3+ KB
```

Fig-2 Dataset Statistics 2

```
data.describe()
```

	sl_no	gender	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
count	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	148.000000
mean	108.000000	0.353488	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
std	62.209324	0.479168	10.827205	10.897509	7.358743	13.275956	5.833385	93457.452420
min	1.000000	0.000000	40.890000	37.000000	50.000000	50.000000	51.210000	200000.000000
25%	54.500000	0.000000	60.600000	60.900000	61.000000	60.000000	57.945000	240000.000000
50%	108.000000	0.000000	67.000000	65.000000	66.000000	71.000000	62.000000	265000.000000
75%	161.500000	1.000000	75.700000	73.000000	72.000000	83.500000	66.255000	300000.000000
max	215.000000	1.000000	89.400000	97.700000	91.000000	98.000000	77.890000	940000.000000

Fig-3: Dataset Statistics 3

The status column is our target feature. It has two values: Placed and Not placed. The columns in the dataset are processed and encoded to convert from categorical to numerical data. The training-testing split is 80-20 with a random state of 42.

	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	specialisation	status
0	0	67.00	0	91.00	0	1	58.00	2	0	1	1
1	0	79.33	1	78.33	0	2	77.48	2	1	0	1
2	0	65.00	1	68.00	1	0	64.00	1	0	0	1
3	0	56.00	1	52.00	1	2	52.00	2	0	1	0
4	0	85.80	1	73.60	1	1	73.30	1	0	0	1

Fig-4: Dataset Statistic 4

3.2 Proposed System

This chapter includes a brief description of the proposed system and explores the different modules involved along with the various models through which this system is understood. We are using six models for model training. The model with the best accuracy will be chosen for predictions.

Logistic Regression:
Logistic Regression is a statistical method used for binary classification problems, modeling the probability that a given input belongs to a particular class. It employs the logistic function to compute the probability.

K-Nearest Neighbors (KNN):
K-Nearest Neighbors (KNN), on the other hand, is a non-parametric and lazy learning algorithm utilized for classification and regression tasks. It makes predictions based on the majority class among its K nearest neighbors in the feature space, without a specific formula; rather, it relies on a majority vote or averaging of values.

Support Vector Machine (SVM):
Support Vector Machine (SVM), a supervised learning algorithm, seeks to find the hyperplane that best separates classes in the feature space. For linearly separable data, the decision boundary is defined by a hyperplane, and SVM aims to maximize the margin between classes.

Decision Tree:
Decision Tree, another supervised learning algorithm, partitions the feature space into regions and predicts the target variable based on majority class or average value within each region.

Random Forest:
Random Forest, an ensemble learning method, constructs multiple decision trees during training and aggregates their predictions for classification or regression tasks, providing robustness against overfitting.

Gradient Boosting:

Lastly, Gradient Boosting builds a strong predictive model by sequentially adding weak learners and correcting errors made by previous models. It fits new models to residual errors, minimizing a loss function over the ensemble of models. These algorithms offer a diverse set of tools for various machine learning tasks, each with its own strengths and applications.

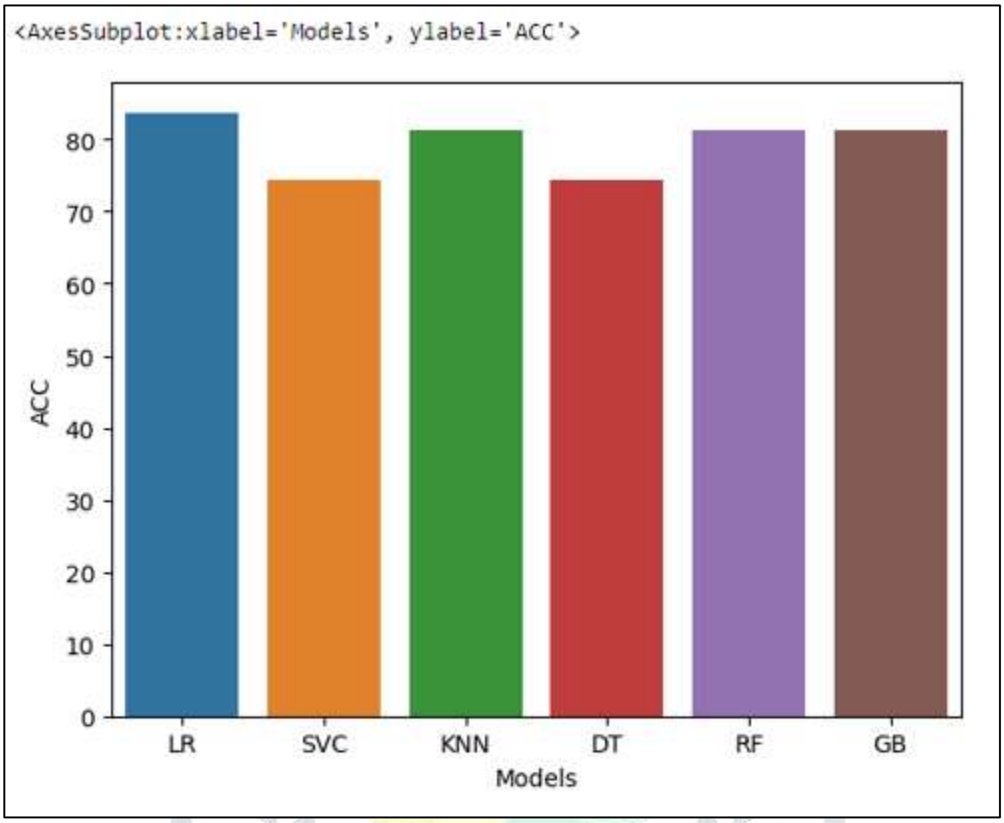

Fig- 5 Model Accuracy in Graph

As shown in above graphs, LR (Logistic Regression) has the highest accuracy. The algorithms are evaluated using the performace metrics "Accuracy". It is used for classification problem evaluation. It is imported using the following code:

```
from sklearn.metrics import accuracy_score
```

Hence, LR is chosen for final predictions. The model is trained using training data. The model is then stored using joblib as 'model_campus_placement', so that it can be imported later for predictions.

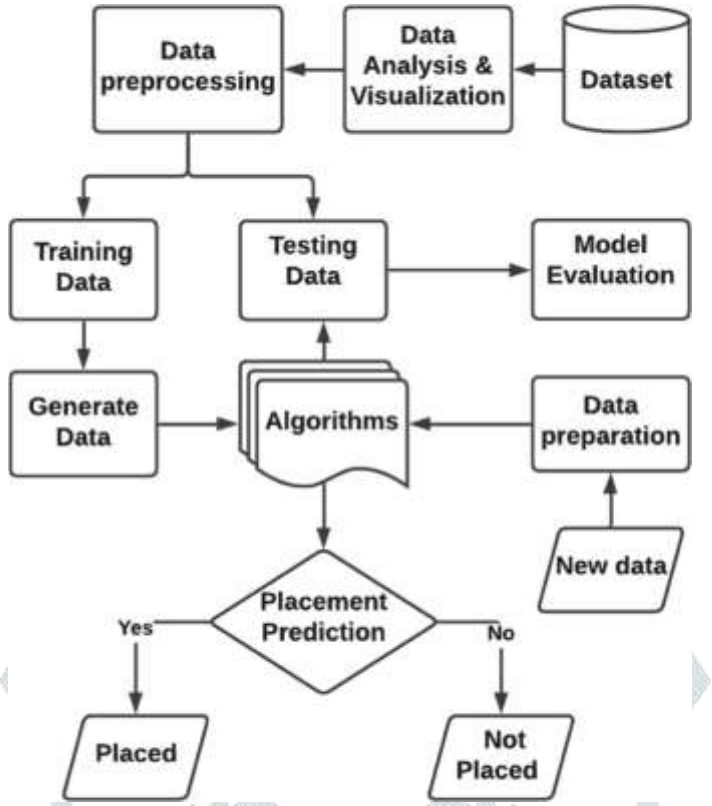

Fig-6: System Architecture

IV. Result and Discussion

This section includes the final snapshots of our project. We used the tkinter tool for creating a simple and user-friendly GUI.

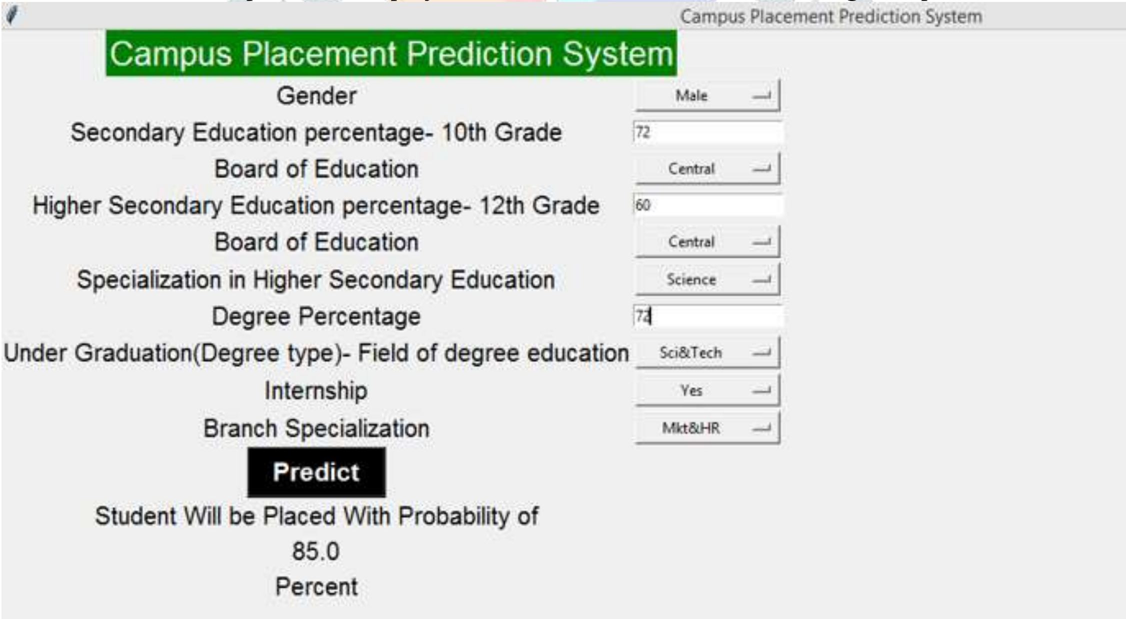

Fig 7- Training and validation loss for investment analysis

V. Acknowledgment

We take this opportunity to express our deep sense of gratitude to our project guide and project coordinator, Mrs Priyanka Patil, for her continuous guidance and encouragement throughout the duration of our major project work. It is because of his experience and wonderful knowledge. we can fulfil the requirement of completing the major project within the stipulated time. We would also like to thank Mr. John Kenny, Head of AIML engineering department for his encouragement, whole-hearted cooperation and support. We would also like to thank our Principal, Dr. J. B. Patil and the management of Universal College of Engineering, Vasai, Mumbai for providing us all the facilities and the work friendly environment. We acknowledge with thanks, the assistance provided by departmental staff, library and lab attendants.

REFERENCES

[1] R. Gupta, N. Sharma, A. Singh. "A Comparative Study of Machine Learning Algorithms for Campus Placement Prediction." International Journal of Computer Applications, 2020.

[2] Ankit Yadav, Nitin Pandey, Priyanka Gautam. "Campus Placement Prediction System Using Machine Learning Techniques." International Journal of Scientific & Technology Research, 2021.

[3] Divya Sharma, Vipin Kumar Sharma. "Predicting Campus Placements using Machine Learning Techniques." International Journal of Scientific Research in Computer Science, Engineering, and Information Technology, 2022.

[4] Snehal Borse, Pranav Dixit, Kunal Gaikwad. "Campus Placement Prediction using Machine Learning Algorithms." International Journal of Advance Research, Ideas, and Innovations in Technology, 2023.

[5] Rohit Kumar, Pooja Singh. "Predictive Analysis of Campus Placements using Machine Learning." International Journal of Computer Science and Information Technology Research, 2024.