Title : EDI / D&I / Major Project Registration & Progress Review **FF No. 180**

Department: E&TC	**Academic Year: 2020-21**
Semester : 4	**Roll No. : 60**
Project Title: Placement prediction Model using Machine Learning.	
Project Area: Machine learning	

Group Members Details:

Sr. No.	Class & Div.	Roll No.	G.R.No.	Name of Student	Contact No.	Email ID
1	ET-A	60	11910759	Deotale Chinmay Jayendra	8530854411	chinmay.deotale19@vit.edu

Name of Internal Guide: **Prof.(Dr.) Ashwini Barbadekar Ma'am**
Contact No. & Email ID: 9970174958 / Ashwini.barbadekar@vit.edu

Project approved / Not approved

Guide Coordinator Head of Department

Project Synopsis

Introduction: Placements hold great importance for students to build a strong foundation for the professional career ahead as well as a good placement record gives a competitive edge to a college/university in the education market.

Objective : 1)To develop a placement predictor as a part of making a placement management system at college level which predicts the chances of students getting placed and helps in uplifting their skills before the recruitment process starts .

2)A placement predictor is to be designed to calculate the possibility of a student being placed in a company, subject to the criterion of the company. The placement predictor takes many parameters which can be used to assess the skill level of the student. While some parameters are taken from the university level, others are obtained from tests conducted in the placement management system itself. Combining these data points, the predictor is to accurately predict if the student will or will not be placed in a company. Data from past students are used for training the predictor.

Tools / Platform used :- JUPYTER NOTEBOOK
Libraries used : PANDAS , MATPLOTLIB ,SEABORN , SKLEARN ,FLASK , PICKLE

Methodology : *1) Data gathering :* The sample data has been collected from KAGGLE.com which consists of all the records of student placement through various sources.

2) Pre-Processing Data : Pre processing is a technique that is used to convert raw data into a clean dataset. The data is gathered from different sources is in raw format which is not feasible for the analysis.

a) Attribute selection
b) Cleaning missing values
c) Training and Test data
d) Feature Scaling

3) Processing : Processing is applying different algorithms to the data to find the best results

a) Logistic Regression
b) Naïve Bayes Classifier
c) SVM

Conclusion : The data is first trained and then tested with all Three algorithms and out of all SVM gave more accuracy with 89.75, Logistic regression with 87.18 percent accuracy and Naïve bayes with accuracy of 85.6.

- We conclude that Logistic Regression works better with better accuracy but difference in scores is highest among three
- Gaussian Naive Bayes was less accurate but the difference in known and unknown data was lesser.
- But, SVM gave better accuracy with least difference in score. So, Our final model would use SVM for Student Placement Prediction.

FF No 180

Group No.	Roll No. 60		
Activity	Review Schedule	Progress Review Report submitted	Signature of Guide
Review 1	Mid Sem. Semester	Yes / No	
Review 2	End of Semester	Yes / No	

Format of Progress Review Report:

Review No.: 1	**Roll No.: 60**	**Date: 08-04-2021**
Progress Review Report :Done with 75% of the project 1) Done with performing exploratory Data analysis a)Label Encoding Data b)Checking for outliers c)Checking for correlation		

2) Done with data visualization and analysis
3) Done with training and evaluating using Logistic Regression model (accuracy – 87 % model score -91%)
4) Done with training and evaluating using Naïve Bayes Classifier (accuracy – 85 % model score -87%)
5) Done with training and evaluating using SVM model (accuracy – 90 % model score -91%)

All the model is able to predict on both Seen and Unseen Data

Remaing part :
1) training and evaluating using Artificial Neutral Network (ANN) model .
2) Deploying model with highest accuracy and model score .

Review No.: 2 **Roll No.: 60** **Date: 12/06/2021**

Progress Review Report : Done with 100% of the project

1) Done with performing exploratory Data analysis
   a)Label Encoding Data
   b)Checking for outliers
   c)Checking for correlation
2) Done with performing exploratory Data analysis
   a)Label Encoding Data
   b)Checking for outliers
   c)Checking for correlation
2) Done with data visualization and analysis
3) Done with training and evaluating using Logistic Regression model (accuracy – 87 % model score -91%)
4) Done with training and evaluating using Naïve Bayes Classifier (accuracy – 85 % model score -87%)
5) Done with training and evaluating using SVM model (accuracy – 90 % model score -91%)\
6) Successfully Deployed the Model on Localhost using Pickle and Flask .

All the model is able to predict on both Seen and Unseen Data

Signature of Guide: