Prediction of Youth employment using Machine Learning

Team Members:-

1)Sudarshan Thakur (sthakur12@hawk.iit.edu) (Cwid= A20522408)

2)Bandhavi Parvathaneni (bparvathaneni@hawk.iit.edu) (A20516844)

3)Urva Surti (usurti1@hawk.iit.edu) (A20505142)

Problem Description:-

The problem statement of this project revolves around predicting the employability i.e. whether a particular candidate participating in the placement process in order to seek employment manages to bag in a job offer or not and provide a better performing prediction model which will outperform all the conventional & previously employed models by the researchers. The covid-19 pandemic struck humanity as an unprecedented event. Along with loss of lives there was also a huge loss of jobs. There was a massive job loss around the world. The students who were just about to start their career were in a lot of stress regarding their placements. This project addresses this issue of the campus placements. Students with a good college gpa's, 10-12th marks, projects and work experience are usually considered to have a strong foothold. This project focuses on placement of a college student considering the parameters/features such as 10th -12th percentage marks, graduation gpa, salary expectation, work experience, stream the student belongs to, the board in which he/she studied etc. The machine learning model inputs the features from the user and gives out a result in the form of placed and unplaced. In this project we are going to test various machine learning algorithms like Gaussian Naive Bayes, Random Forest, Support Vector Machine and K Nearest Neighbours. The best performing algorithms will be chosen for stacking and creating a new model which will be deployed. On the basis of the result the student can decide about how he/she has to go around for the preparation of campus placements.

Description of the data used in the project:-

So the dataset we are using consists of 15 attributes out of which 13 we can be used to predict the youth employment. The 13 attributes include SSC percentage which range from 0-100 (10^{th} Standard Percentage),Board from which they pass out(eg. It can be Central board) , High School percentage which will range from 0-100 , Highschool Board(Central or Other) , Undergrad Degree(Science, engineering, commerce, Undergrad Percentage which ranges between 0-100 , Work experience (if they have or not) , Their specialisation, MBA percentage , Status (whether placed or not), and their salary. Our dataset consists of total 215 entries. We are planning to split the data into 70% and 30% where 70% of the data will be used for the training purpose whereas the 30% data will be used for the testing purpose. So here the above mentioned variables or the attributes will be in the passed as an input to the model, based on the input our target variable will placed or not placed.

Things done so far:-

We have searched for various algorithm which can be used to predict the employment of the student provided given the above inputs(attributes) and the Output employed or not. The various machine learning algorithms that can be used in our model are Gaussian Naive Bayes, Random Forest, Support Vector Machine and K Nearest Neighbours. We are currently in the stage of reading the dataset into the system, pre-processing the data. Currently we are pre-processing data we are handling null values within the dataset, feature elimination and encoding features into suitable form for the

process of model building. We will try to implement a hybrid model which will consist of the algorithms with highest accuracy.

What remains to be done:-

Implement the above mentioned algorithms. Creation of a model, calculating, testing the model calculating the accuracy based on different algorithms. Choosing the best algorithm and using it for the final model. The deployment of the model created, for which we might be using flash to deploy a web application. A snapshot of our dataset which we will be used for the prediction purpose

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
1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55	Mkt&HR	58.8	Placed	270000
2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000
3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75	Mkt&Fin	57.8	Placed	250000
4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed	
5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.5	Placed	425000
6	M	55.00	Others	49.80	Others	Science	67.25	Sci&Tech	Yes	55	Mkt&Fin	51.58	Not Placed	
7	F	46.00	Others	49.20	Others	Commerce	79.00	Comm&Mgmt	No	74.28	Mkt&Fin	53.29	Not Placed	
8	M	82.00	Central	64.00	Central	Science	66.00	Sci&Tech	Yes	67	Mkt&Fin	62.14	Placed	252000
9	M	73.00	Central	79.00	Central	Commerce	72.00	Comm&Mgmt	No	91.34	Mkt&Fin	61.29	Placed	231000