High Level Design (HLD)

CAMPUS PLACEMENT PREDICTION

REVISION NUMBER: 2.0
DATE: 29/04/2022

Document Version Control

Date Issued	Version	Description	Author
20/04/2022	1	Initial HLD-V1.0	Yashika A Pawar
29/04/2022	2	Updated HLD-V1.1	Yashika A Pawar

Contents

ABSTRACT:

Placement of students is one of the most important objective of an educational institution. Reputation and yearly admissions of an institution invariably depend on the placements it provides it students with. Institutions make great efforts to achieve placements for their students' .This will always be helpful to the institution. The objective is to predict the students getting placed for the current year by analyzing the data collected from previous year's students. This model is proposed with an algorithm to predict the same. The data has been collected by the institution for which prediction is going to be done and by applying suitable data pre-processing techniques. This model is prepared by using Support Vector Machine [SVM] algorithm. This algorithm independently predict the results and we then compare the efficiency of the algorithm, which is based on the dataset. This model will helps the placement cell to focus on the potential students and help them to improve their technical and social skills.

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
  - Security
  - Reliability
  - Maintainability
  - Portability
  - Reusability
  - Application compatibility
  - Resource utilization
  - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2 General Description

2.1 Product Perspective

The campus placement prediction system is a machine learning based model which uses support vector machine[SVM] algorithm which predicts whether the student gets placed or not.

2.2 Problem statement

The placement prediction model considers only academic performances of the students so that the prediction of the student getting placed or not can be done. We cannot consider the placement of students just by their academic performances because some students may be good at aptitude, technical and communication skills due to their low score in their academic that may tend to be their drawback. For predicting the placement of a Student needs parameters like cgpa, logical and technical skills Academic performances may be important but the model is design to predict the placements based on the parameters of the students.

2.3 PROPOSED SOLUTION

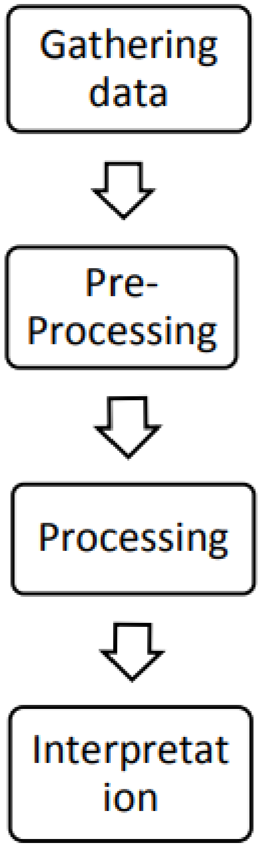
We have designed a models for checking prediction of student getting placed or not in a campus drive this model will help both student and institutions for preparing well in advance for campus recruitment. The main objective of this model is to know the capability of the student and where he stands by predicting the probability of getting placed. This also help the students and institutions to improve performance of the potential students. This model will consider the academic history of the student such as percentage as well as their domains and specializations which are tested by companies. We used SVM algorithm on the student's data gathered from the institution of previous year. This model take scores of student in secondary education along with academic CGPA of both UG and PG.in the technical education till date and also some parameters which adds weightage to kick start the career.

2.4 FURTHER IMPROVEMENTS

Machine learning tools help placement team members by **tracking a candidate's journey throughout the interview process and helping out to speed up the process of getting feedback**. After evaluating, a feedback will be given to candidates in order to improve their competency level or to specify the area of improvement.

2.5 METHODOLOGY

The whole approach is depicted by the following flowchart.



Flow chart of the technique

2.5.1 Data gathering the sample data has been collected from our college placement department which consists of all the records of previous year's students. The dataset collected consist of over 1000 instances of students.

2.5.2 Preprocessing Data preprocessing 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. Pre-processing for this approach takes 4 simple yet effective steps.

2.5.3 Attribute selection some of the attributes in the initial dataset that was not pertinent (relevant) to the experiment goal were ignored. The attributes name, roll no, credits, backlogs, whether placed or not, b.tech %, gender are not used. The main attributes used for this study are credit, back-logs, whether placed or not, b.tech %.

2.5.4 Cleaning missing values in some cases the dataset contain missing values. We need to be equipped to handle the problem when we come across them. Obviously you could remove the entire line of data but what if you're inadvertently removing crucial information? After all we might not need to try to do that. One in every of the foremost common plan to handle the matter is to require a mean of all the values of the same column and have it to replace the missing data. The library used for the task is called Scikit Learn preprocessing. It contains a class called Imputer which will help us take care of the missing data.

2.5.5 Training and Test data splitting the Dataset into Training set and Test Set Now the next step is to split our dataset into two. Training set and a Test set. We will train our machine learning models on our training set, i.e. our machine learning models will try to understand any correlations in our training set and then we will test the models on our test set to examine how accurately it will predict. A general rule of the thumb is to assign 80% of the dataset to training set and therefore the remaining 20% to test set.

2.5.6 Feature scaling is the final step of data preprocessing is feature scaling. But what is it? It is a method used to standardize the range of independent variables or features of data. But why is it necessary? A lot of machine learning models are based on Euclidean distance. If, for example, the values in one column (x) is much higher than the value in another column (y), (x2-x1) squared will give a far greater value than (y2-y1) squared. So clearly, one square distinction dominates over the other square distinction. In the machine learning equation.

2.6 Data Requirements

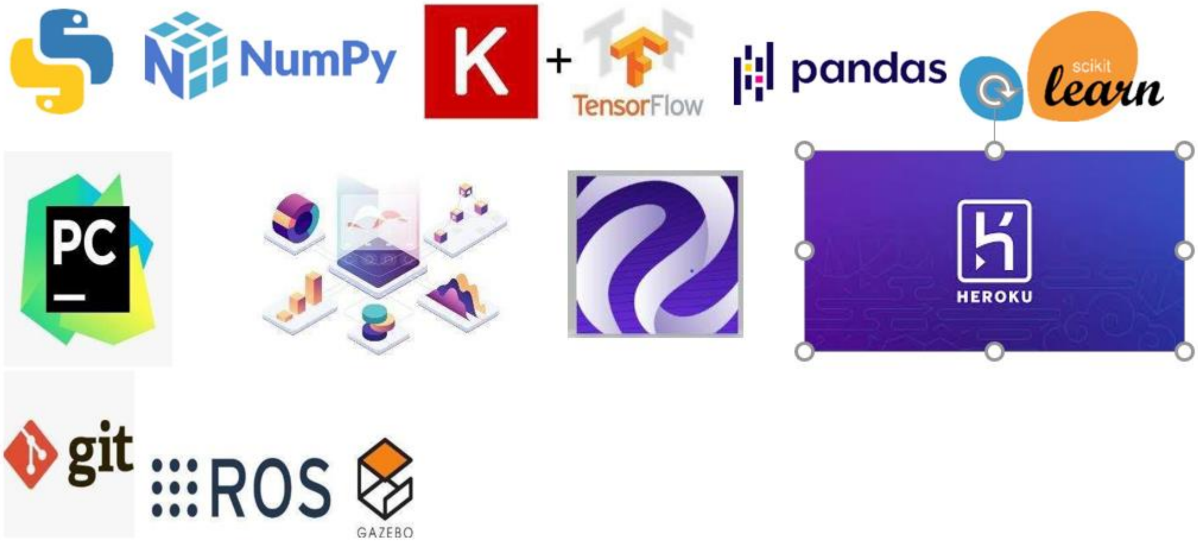
File descriptions

- train.csv - the training set
- test.csv - the test set
- SampleSubmission.csv - a sample submission file in the correct format

Data fields

- gender - sex of the student
- secondary education percentage-marks obtained in secondary education
- higher secondary percentage-marks obtained in higher secondary education
- degree percentage-marks obtained in degree
- Under-graduation(Degree-type)-Field of degree education
- Work-experience
- Employability-test-package
- specialisation-field of study

2.7 Tools used



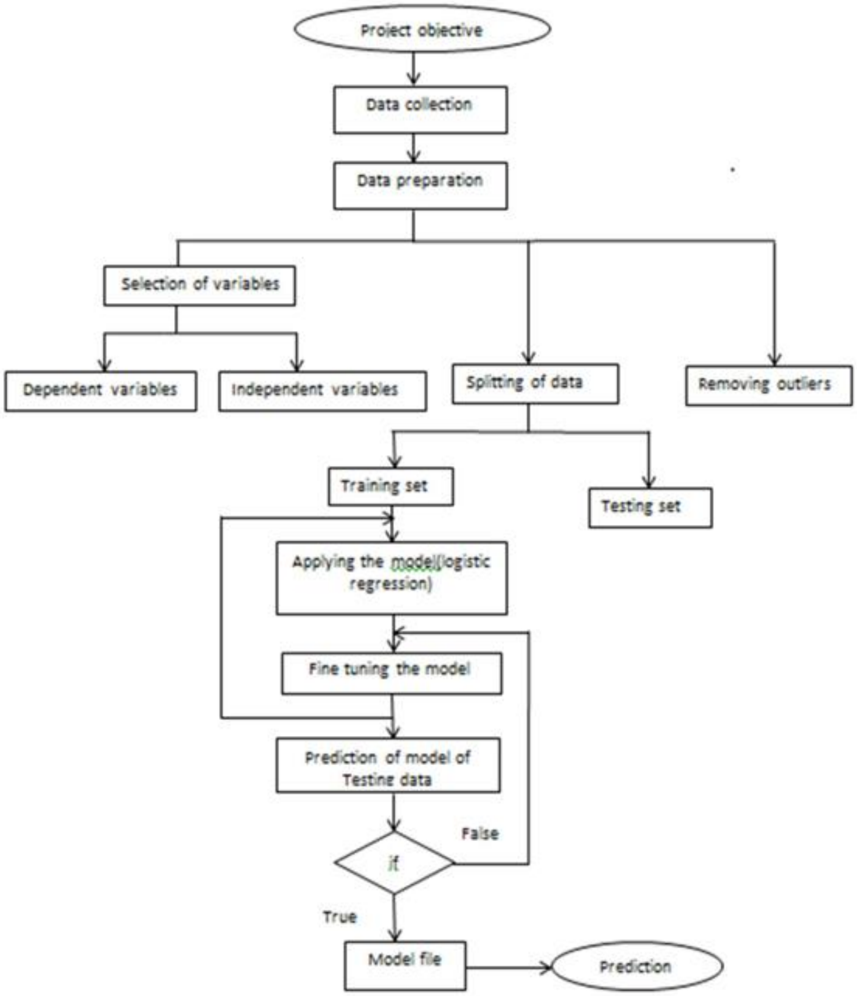
- Python Programming language and frameworks such as NumPy, Pandas, Sckit-learn, TensorFlow are used to build the whole model.
- NumPy-Numpy is a python library used for working with arrays.
- Pandas-Pandas is used for working with machine learning and data analysis/data driven tasks.
- Sckit-learn-a set of modules for machine learning and data mining.
- Git-a series of controlled versions.
- Heroku-a cloud platform for deployment of project.

3 Design Details

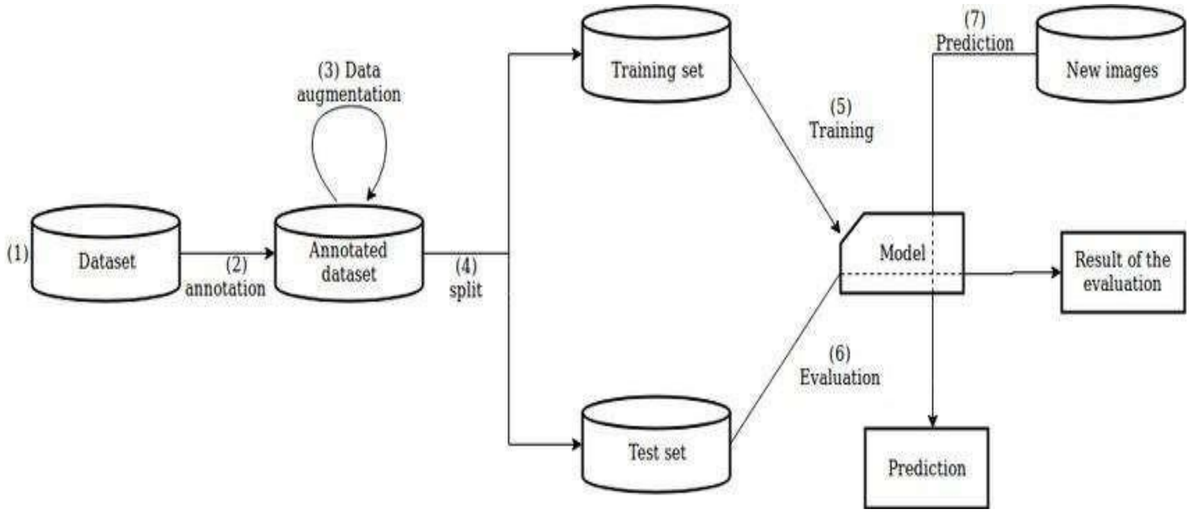
3.1 Process Flow

identifying the different types of anomalies, we will use a deep learning base model. Below is the process flow diagram is as shown below.

Proposed methodology



MODEL TRAINING AND EVALUATION



3.2 Event log

The system should log every event so that the user will know what process is running internally.

INITIAL STEP BY STEP PROCEDURE:

The system identifies at what step logging is required.

The system should be able to log each and every system flow

Developer can choose logging method. You can choose database logging/file logging as well.

System should not hang even after using so many logins. We can easily debug issues so logging is mandatory to do.

3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

4 Performance

The proposed system results in the probability of the students getting placed in the campus drives. The 'STUDENTS PLACEMENT PREDICTION USING MACHINE LEARNING' provides the help for both students and the institution .The institution can focus on the potential students by knowing the prediction of this model. The technique we used is the SVM algorithm which gives the accuracy of students getting placed. The project mainly gives the information about the students probability of getting placed in the campus drives which benefits both the students and the institution.

4.1 Reusability

The code written and the components used should have the ability to be reused with no problems.

4.2 Application Compatibility

The different components for this project will be using python as an interface between them. Each component will have its own task to perform, and it is the job of the python to ensure proper transfer of information.

4.3 Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

6 Conclusion

The campus placement activity is incredibly a lot of vital as institution point of view as well as student point of view. In this regard to improve the student's performance, a work has been analyzed and predicted using the classification algorithms like KNN algorithm, Random Forest, Naïve Bayes, Logistic Regression, Decision Tree and the SVM 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 KNN algorithm is 74%, Random Forest is 82%, Naïve Bayes is 84%, Logistic Regression 87%, Decision tree is 84% and for the SVM is 89.7%. Hence, from the above said analysis and prediction its better if the SVM algorithm is used to predict the placement results.