Architecture Design

CAMPUS PLACEMENT

PREDICTION

Written by	TRISHIT NATH THAKUR
Version	1.4
Date	25– 06 – 2023

Document Control

Change Record:

Version	Date	Author	Comments
1.0	08-06-2023	TRISHIT	First Draft
1.1	12-06-2023	TRISHIT	Added Exception Scenarios Overall, Constraints
1.2	13-06-2023	TRISHIT	Added user I/O flowchart
1.3	19-06-2023	TRISHIT	Added dataset overview and updated user I/O flowchart.
1.4	24-06-2023	TRISHIT	Restructure and reformat LLD

Abstract

Machine Learning is a category of algorithms that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build models and employ algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. These models can be applied in different areas and trained to match the expectations of management so that accurate steps can be taken to achieve the organization's target. In this paper, The 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. That is why all the institutions, arduously, strive to strengthen their placement department so as to improve their institution on a whole. Taking various aspects of a dataset collected for placement prediction, and the methodology followed for building a predictive model, results with high levels of accuracy are generated, and these observations can be employed to predict whether the student will be recruited in campus placements or not based on the available factors in the dataset.

1. Introduction

1.1 What is Architecture Design?

The goal of Architecture Design (AD) or a low-level design document is to give the internal design of the actual program code for the `Campus placement prediction`. AD describes the class diagrams with the methods and relation between classes and program specification. It describes the modules so that the programmer can directly code the program from the document.

Big Mart Sales Prediction

1.2 Scope

Architecture Design(AD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software, architecture, source code, and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work. And the complete workflow.

1.3 Constraints

We only predict the expected placement status based on the marks, subject and stream with imbalance data.

1.4 Risks

Document specific risks that have been identified or that should be considered. 1.5 Out of Scope Delineate specific activities, capabilities, and items that are out ofscope for the project.

2. Technical Specification

2.1 Dataset

• gender - sex of the student

 • secondary education percentage-marks obtained in secondary education

• higher secondary percentage-marks obtained in higher secondary education

 • degree percentagemarks obtained in degree

• Undergraduation(Degreetype)-Field of degree education

• Work experience

• Employability- how employable a person is

• specialisation-field of study

File descriptions

• train.csv - the training set

• test.csv - the test set

• SampleSubmission.csv - a sample submission file in the correct format.

Big Mart Sales Prediction

Data of students for predicting campus placement is collected from various institutions. This data about students contains their percentage at different educational level and which stream they choose in graudation and post graduation. The dataset looks like as follow:

	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
210	211	0	80.60	Others	82.00	Others	Commerce	77.60	Comm&Mgmt	No	91.0	Mkt&Fin	74.49	Placed	400000.0
211	212	0	58.00	Others	60.00	Others	Science	72.00	Sci&Tech	No	74.0	Mkt&Fin	53.62	Placed	275000.0
212	213	0	67.00	Others	67.00	Others	Commerce	73.00	Comm&Mgmt	Yes	59.0	Mkt&Fin	69.72	Placed	295000.0
213	214	1	74.00	Others	66.00	Others	Commerce	58.00	Comm&Mgmt	No	70.0	Mkt&HR	60.23	Placed	204000.0
214	215	0	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89.0	Mkt&HR	60.22	Not Placed	NaN

The data set consists of various data types from integer to floating to object as shown in Fig.

```
train.info()

<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
```

In the raw data, there can be various types of underlying patterns which also gives an in-depth knowledge about the subject of interest and provides insights into the problem. But caution should be observed
with respect to data as it may contain null values, or redundant values, or various types of ambiguity, which also demands pre-processing of data. The dataset should therefore be explored as much as possible.
Various factors important by statistical means like mean, standard deviation, median, count of values and maximum value, etc. are shown below for numerical attributes.

Big Mart Sales Prediction

```
train.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

Preprocessing of this dataset includes doing analysis on the independent variables like checking for null values in each column and then replacing or filling them with supported appropriate data types so that analysis and model fitting is not hindered from their way to accuracy. Shown above are some of the representations obtained by using Pandas tools which tell about variable count for numerical columns and model values for categorical columns. Maximum and minimum values in numerical columns, along with their percentile values for median, play an important factor in deciding which value to be chosen at priority for further exploration tasks and analysis. Data types of different columns are used further in label processing and a one-hot encoding scheme during the model building.

2.2 Logging

We should be able to log every activity done by the user
- The system identifies at which step logging require.
- The system should be able to log each and every system flow.
- Developers can choose logging methods. Also can choose database logging.
- The system should be not be hung even after using so much logging. Logging just because we can easily debug issuing so logging is mandatory to do.

2.3 DataBase

The system needs to store every request into the database and we need to store it in such a way that it is easy to retain and look into the records.The system should capture every data that any user gave and the prediction that has been made by that input.

2.4 Deployment

For the hosting of the project, we will use heroku

Big Mart Sales Prediction

3. Technology Stack

| Front End	HTML/JavaScript
Backend	Python
Deployment	Heroku

4. Proposed Solution

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 ofstudents 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 student..
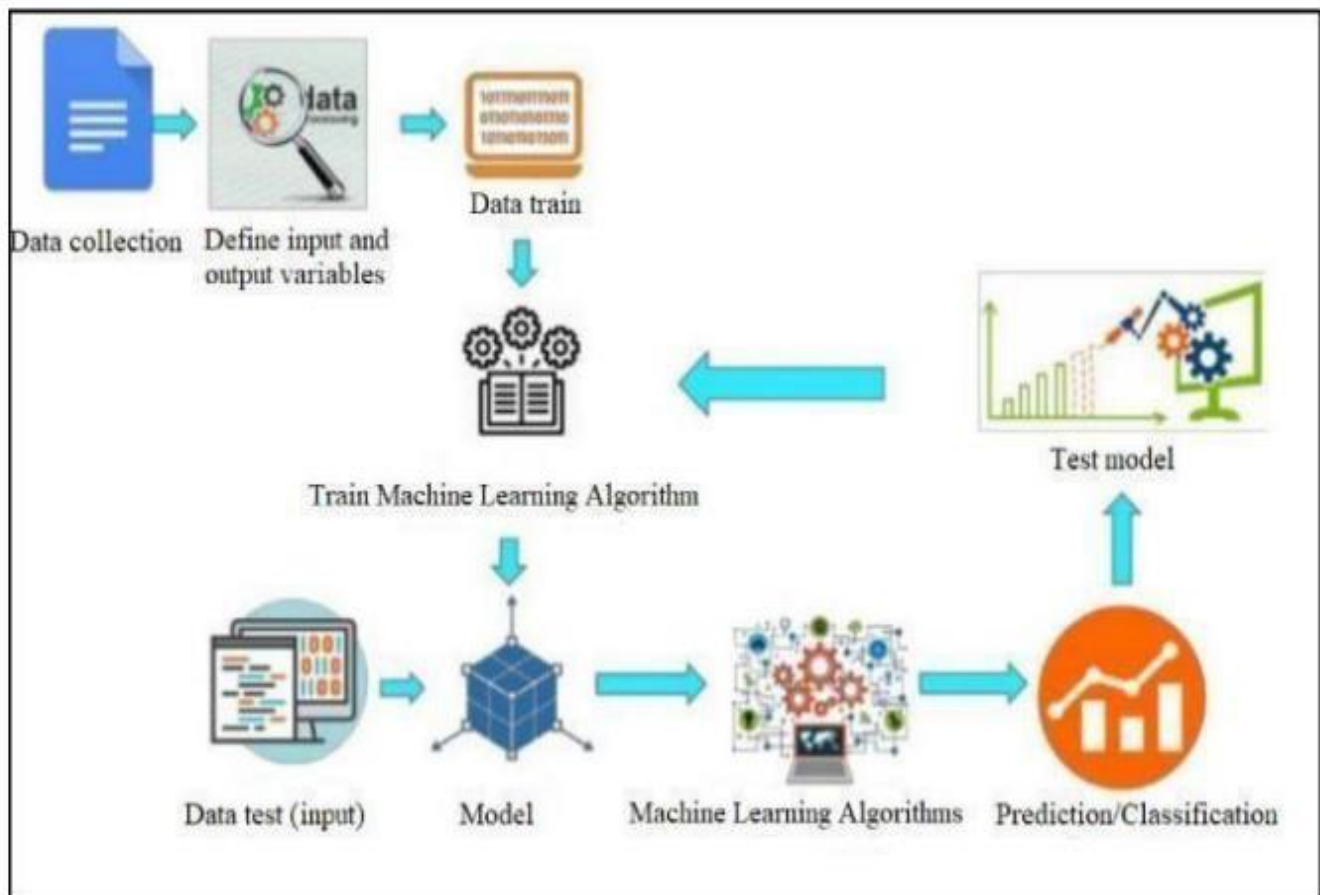
5 Architecture detail

Big Mart Sales Prediction
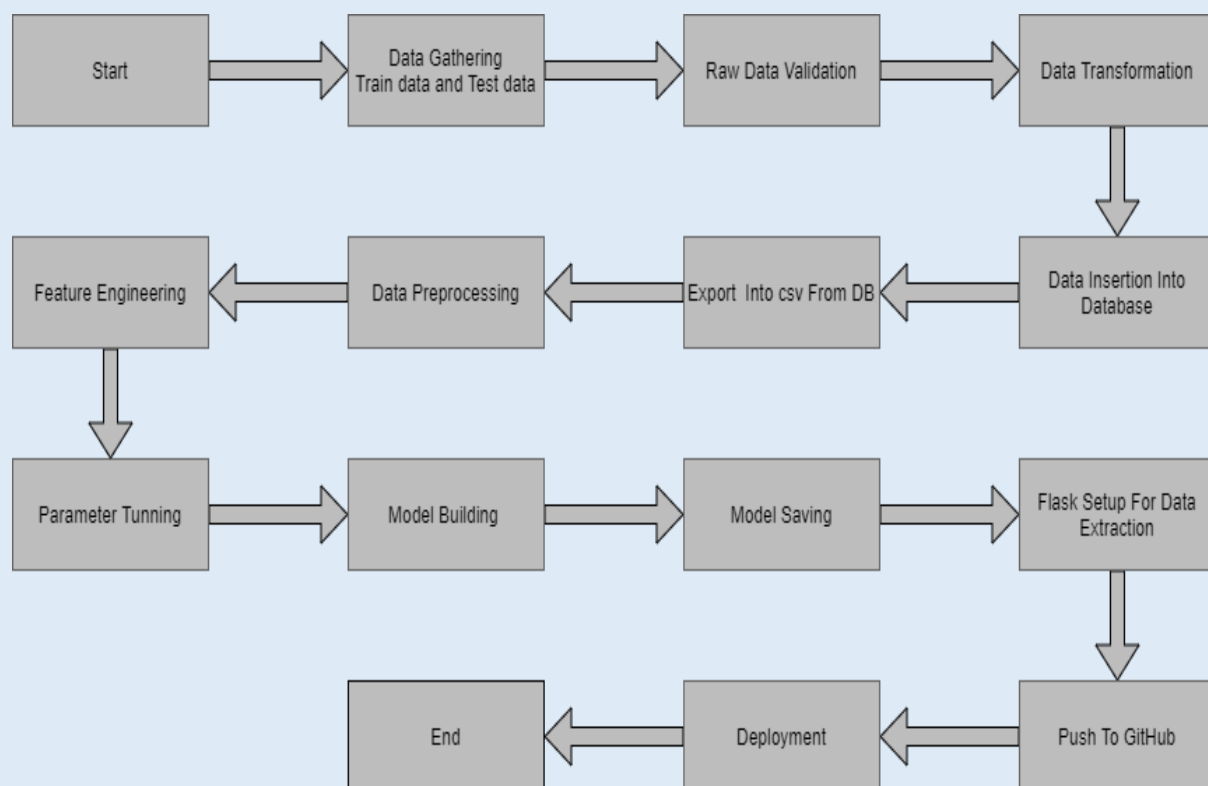
5 Model training/validation workflow

Big Mart Sales Prediction
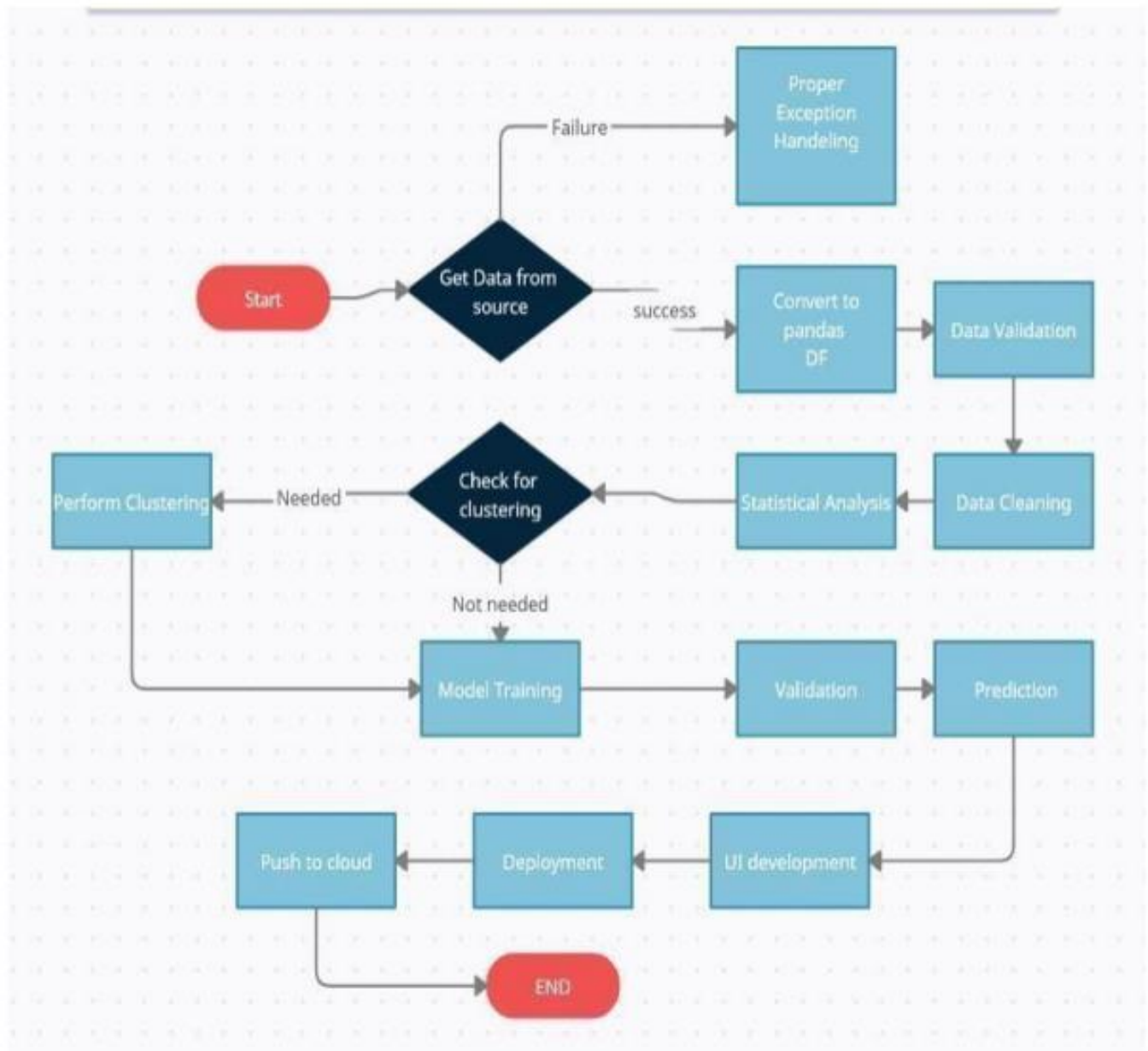
6 User I/O workflow

7. Conclusion

In this project, basics of machine learning and the associated data processing and modeling algorithms have been described, followed by their application for the task of placement prediction in Educational instituions. On implementation, the prediction results show what are the educational records and marks they to get job placement, Any assistance in this particular area will have a positive impact on an institution's ability to place its students. This will always be helpful to both the students, as well as the institution.

Big Mart Sales Prediction