**INTRODUCTION**

When a student joins a college or any institute the first thing that they have in mind is that they will get a good placement from this college after graduation. Campus placements help the students to get a platform for themselves and they don’t have to struggle themselves in the search for a job. Hence it is important to properly analyze the whole process of placements. In order to perform thorough data analysis, we chose a data set that consists of Placement data of students on the campus. It includes secondary and higher secondary school percentages and specialization. It also includes degree specialization, type and Work experience and salary offers to the placed students. This analysis will help students understand the process and will help them in getting better results in the future.

**PROBLEM STATEMENT**

To understand the process of placements in a college, how the system works, and what are the major factors that affect the placement statistics and influence the candidates participating in it.

**FUNCTIONAL REQUIREMENTS**

Along with analysis and visualization of data, our project will also provide a prediction model which will help students understand their chances of getting placed or not.The applicability of this feature will help students to understand what are the factors they need to improve in order to get a good placement package.

**DATA ABSTRACTION**

**DATASET TYPE:**

The data set used for this project is a table which consists of Placement data of students in a XYZ campus. It includes secondary and higher secondary school percentages with specialization. It also includes degree specialization, type and Work experience and salary offers to the placed students.

**DATA TYPES:**

The data types used here are attributes and items.

The dataset table of campus placements contains 15 attributes and 215 items.

**ATTRIBUTE SEMANTICS:**

1. **Sl\_no:** Serial Number.
2. **gender:** The gender of the student. Male=”M”, Female=”F”
3. **ssc\_p:**Secondary Education percentage of students in 10th Grade
4. **ssc\_b:** Board of Education for 10th Grade - Central/ Others
5. **hsc\_p:** Higher Secondary Education percentage of students in 12th Grade
6. **hsc\_b:** Board of Education for 12th grade - Central/ Others
7. **hsc\_s:**Specialization in Higher Secondary Education
8. **degree\_p:** Degree Percentage scored by student
9. **degree\_t:** Under Graduation(Degree type)- Field of degree education
10. **workex:** Work Experience
11. **etest\_p:** Employability test percentage ( conducted by college)
12. **specialisation:** Post Graduation(MBA)- Specialization
13. **mba\_p:** MBA percentage
14. **status:** Status of placement- Placed/Not placed
15. **salary:** Salary offered by corporate to candidates

The attributes can be classified as categorical and ordered as follows:

|  |  |
| --- | --- |
| **Categorical** | **Ordered** |
| Gender : Male- M , Female- F  ssc\_b: Central / Other  hsc\_b: Central / Other  hsc\_s: Science/ Commerce/ Arts  degree\_t: Sci&Tech / Comm&Mgmt  workex: Yes / No  specialisation: Mkt&HR/ Mkt&Fin  status: Placed / Not placed | Sl\_no  ssc\_p  hsc\_p  degree\_p  etest\_p  mba\_p |

**TARGET IDENTIFICATION:**

We have focused on the following objectives or targets for our project:

* Does the candidate's **gender** ( male or female ) have any role in placement?
* Which **factors** influenced a candidate in getting placed?
* Does **10th and 12th percentage** matter for one to get placed?
* Which **degree specializatio**n is much demanded by corporate?
* Determine the **average salary** offered during placements and its factors.
* Play with the data conducting all **statistical tests**.

**Tools/Algorithm/Languages Used**

· Dataset Link :

<https://www.kaggle.com/benroshan/factors-affecting-campus-placement>

· Language Used: Python

· Libraries:

1. pandas

2. numpy

3. matplotlib.pyplot

4. seaborn

5. sklearn

Linear Regression: linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables: One variable, denoted x, is regarded as the predictor, explanatory, or independent variable.

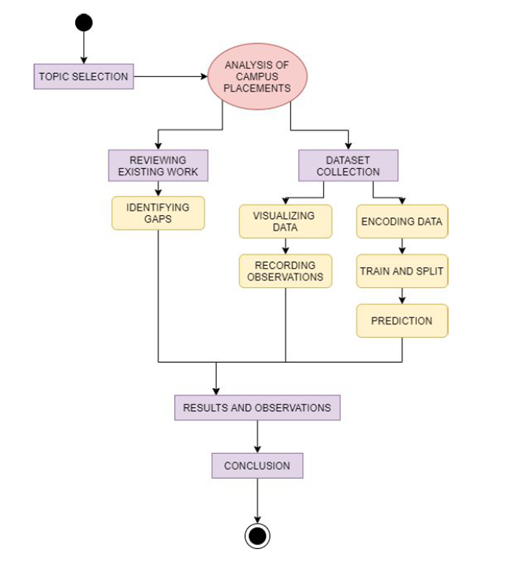
In this project, with the help of linear regression we will be able to tell whether the person will be able to get recruited by any company or not. If not, he will work on the aspects where he is lacking.

· Classification Algorithms Used:

1. Logistic Regression: Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

2. Decision Tree: Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes.

**DESIGN OF THE PROPOSED SYSTEM:**

· 

**Reviewing Existing Work and Identifying the gaps:**

We conduct a literature survey and identify the solution addresses as well as the gaps in their results.

**Dataset Collection:**

We use a dataset from kaggle – “Factors affecting campus placements” for our data.

**Visualizing data:**

Here we find the best graphical representation of our data. i.e. bar charts, pie charts, box plot etc.

**Recording Observations:**

We record the observations from the data.

**Encoding Data:**

Here we translate the data into a visual element on the plots we are making.

**Train and split:**

We divide the training sessions by body regions into two: one for training and one for testing.

**Prediction:**

Here we Predict whether the candidate will be placed or not based on some

predictors.

**Results and Observations:**

From the plots and charts, we get the information we need and make observations for patterns, averages etc.

**Conclusion:**

From our observations we conclude how each student gets placement and their salary package based on gender, their marks, the courses they chose etc.