

Identifying Patterns and Trends in Campus Placement Data using Machine Learning

PROBLEM STATEMENT:

Identify patterns and trends in campus placement data using machine learning to improve the placement process for students.

Objectives:

1. Identify the factors that are most important for students getting placement offers.
2. Predict which students are most likely to get placement offers.
3. Identify trends in the placement process over time.
4. Develop recommendations for improving the placement process.

Data:

The data for this project could include information about the students, such as their Age, Gender, Stream, Internships, CGPA, Hostel and History Of Backlog. It could also include information about the companies that made placement offers, such as the industry they are in, the size of the company, and the location of the job.

Machine Learning Algorithms:

The machine learning algorithms that could be used for this project include clustering algorithms, classification algorithms, and regression algorithms. The specific algorithm that is used will depend on the specific objectives of the project.

Evaluation:

The performance of the machine learning model will be evaluated by its ability to identify patterns and trends in the data, and its ability to predict which students are most likely to get placement offers.

Deployment:

The machine learning model could be deployed to production to help students make informed decisions about their career paths, and to help universities improve their placement process.

PRE-REQUISITES

The prerequisites for identifying patterns and trends in campus placement data using machine learning are:

- Basic knowledge of machine learning
- Familiarity with Python programming language
- Ability to use data science libraries such as NumPy, Pandas, and Scikit-Learn
- Understanding of statistical concepts such as descriptive statistics, hypothesis testing, and regression analysis
- Ability to interpret machine learning results

In addition to these technical prerequisites, it is also helpful to have a strong understanding of the domain of campus placement. This includes knowledge of the different factors that affect placement outcomes, such as the student's academic performance, interview skills and extracurricular activities.

Here are some of the specific tasks that you need to be able to do in order to complete this project:

- Import the necessary data science libraries
- Load the campus placement data set
- Clean the data by removing null values and outliers
- Encode categorical data
- Split the data into train and test sets
- Build a machine learning model to predict placement outcomes
- Evaluate the performance of the model
- Interpret the results of the model

BUILD THE SOLUTION

- **Data collection:** The first step is to collect the data that you want to analyze. This data could include information about the students, such as their grades, their major, their extracurricular activities, and their placement offers. It could also include information about the companies that made placement offers, such as the industry they are in, the size of the company, and the location of the job.

The Dataset is collected from the Github

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

- **Data cleaning:** Once you have collected the data, you need to clean it. This means removing any errors or inconsistencies in the data. You also need to ensure that the data is in a format that can be used by machine learning algorithms.
- **Feature engineering:** Feature engineering is the process of transforming the data into features that are more meaningful for machine learning algorithms. This could involve creating new features, combining existing features, or removing irrelevant features.
- **Model selection:** There are many different machine learning algorithms that can be used to identify patterns and trends in data. The best algorithm to use will depend on the specific data set and the desired outcome.
- **Model training:** Once you have selected a machine learning algorithm, you need to train it on the data. This involves feeding the algorithm the data and allowing it to learn the patterns and trends in the data.
- **Model evaluation:** Once the model is trained, you need to evaluate its performance. This involves testing the model on a separate data set that was not used for training.
- **Model deployment:** Once the model is evaluated and deemed to be performing well, you can deploy it to production. This means making the model available to users so that they can use it to make predictions.

OUTPUT-1:

app3 - Streamlit

localhost:8501

Typing Lessons | U... Google Numpy Creating Ar... numpy online com... Online Numpy Com... Welcome To Colab... Pandas DataFrame... Pandas DataFrame... Python Class - Exer... Programming In Ja...

Placement Data

Gender-M:1,F:0

1 - +

Secondary education percentage

67.00 - +

Higher Secondary education percentage

91.00 - +

Specialization in Higher Secondary Education: Arts:0,Commerce:1,Science:2

1 - +

Degree Percentage

58.00 - +

Degree Type: Comm&Mgmt:0,Sci&Tech:1,others:2

1 - +

Work Experience: Yes:1,No:0

Placement Checking

	hsc_p	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status
0	91	Commerce	58	Sci&Tech	No	55	Mkt&HR	58.8	Placed
1	78.33	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed
2	68	Arts	64	Comm&Mgmt	No	75	Mkt&Fin	57.8	Placed
3	52	Science	52	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed
4	73.6	Commerce	73.3	Comm&Mgmt	No	96.8	Mkt&Fin	55.5	Placed

Training Data Stats

	ssc_p	hsc_p	degree_p	etest_p	mba_p
count	215	215	215	215	215
mean	67.3034	66.3332	66.3702	72.1006	62.2782
std	10.8272	10.8975	7.3587	13.276	5.8334
min	40.89	37	50	50	51.21
25%	60.6	60.9	61	60	57.945
50%	67	65	66	71	62
75%	75.7	73	72	83.5	66.255
max	89.4	97.7	91	98	77.89

app3 - Streamlit

localhost:8501

Typing Lessons | U... Google Numpy Creating Ar... numpy online com... Online Numpy Com... Welcome To Colab... Pandas DataFrame... Pandas DataFrame... Python Class - Exer... Programming In Ja...

Specialization in Higher Secondary Education: Arts:0,Commerce:1,Science:2

1 - +

Degree Percentage

58.00 - +

Degree Type: Comm&Mgmt:0,Sci&Tech:1,others:2

1 - +

Work Experience: Yes:1,No:0

0 - +

Employability test percentage

54.99 - +

Post Graduation(MBA)- Specialisation- Mkt&HR:1,Mkt&Fin:0

1 - +

MBA percentage

58.80 - +

Training Data

	gender	ssc_p	hsc_p	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	sta
0	1	67	91	1	58	2	0	55	1	58.8	
1	1	79.33	78.33	2	77.48	2	1	86.5	0	66.28	
2	1	65	68	0	64	0	0	75	0	57.8	
3	1	56	52	2	52	2	0	66	1	59.43	
4	1	85.8	73.6	1	73.3	0	0	96.8	0	55.5	

Placement Test Data

	gender	ssc_p	hsc_p	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p
0	1	67	91	1	58	1	0	54.99	1	58.8

Your Report:

You are Placed

Accuracy:

93.02325581395348%

OUTPUT-2:

The screenshot shows a Streamlit web application titled 'app3 - Streamlit' running on 'localhost:8501'. The left sidebar contains input fields for 'Placement Data': Gender-M,F,0 (1), Secondary education percentage (56.00), Higher Secondary education percentage (52.00), Specialization in Higher Secondary Education-Arts0,Commerce1,Science2 (2), Degree Percentage (52.00), Degree Type-Comm&Mgmt0,Sci&Tech1,others2 (1), Work Experience-Yes1,No0 (0), and Employability test percentage (65.97). The main area displays 'Placement Checking' with a table of results:

	hsc_p	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status
0	91	Commerce	58	Sci&Tech	No	55	Mkt&HR	58.8	Placed
1	78.33	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed
2	68	Arts	64	Comm&Mgmt	No	75	Mkt&Fin	57.8	Placed
3	52	Science	52	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed
4	73.6	Commerce	73.3	Comm&Mgmt	No	96.8	Mkt&Fin	55.5	Placed

Below the table, 'Training Data Stats' are shown:

	ssc_p	hsc_p	degree_p	etest_p	mba_p
count	215	215	215	215	215
mean	67.3034	66.3332	66.3702	72.1006	62.2782
std	10.8272	10.8975	7.3587	13.276	5.8334
min	40.89	37	50	50	51.21
25%	60.6	60.9	61	60	57.945
50%	67	65	66	71	62
75%	75.7	73	72	83.5	66.255
max	89.4	97.7	91	98	77.89

The screenshot shows the same Streamlit web application with different input values: Higher Secondary education percentage (52.00), Specialization in Higher Secondary Education-Arts0,Commerce1,Science2 (2), Degree Percentage (52.00), Degree Type-Comm&Mgmt0,Sci&Tech1,others2 (1), Work Experience-Yes1,No0 (0), Employability test percentage (65.97), Post Graduation(MBA)- Specialisation-Mkt&HR1,Mkt&Fin0 (1), and MBA percentage (55.50). The main area displays 'Training Data' and 'Placement Test Data'.

Training Data

	gender	ssc_p	hsc_p	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	sta
0	1	67	91	1	58	2	0	55	1	58.8	
1	1	79.33	78.33	2	77.48	2	1	86.5	0	66.28	
2	1	65	68	0	64	0	0	75	0	57.8	
3	1	56	52	2	52	2	0	66	1	59.43	
4	1	85.8	73.6	1	73.3	0	0	96.8	0	55.5	

Placement Test Data

	gender	ssc_p	hsc_p	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p
0	1	56	52	2	52	1	0	65.97	1	55.5

Your Report:

You are not Placed

Accuracy:
93.02325581395348%