Employee Future Prediction

By K.Vamsi Krishna Prasad
-19BCE7515

Guided By:
Dr.Srinivasa Rao B

About DataSet

- This is a dataset consists of details of Employee in a Company.
- This dataset consists of 4654 rows & 9 columns.
- The features in this dataset are as follows:

  Education, JoiningYear, City, PaymentTier, Age, Gender, EverBenched, ExperienceInCurrentDomain , LeaveOrNot

- From the data we should predict weather the employee will leave the company or not.

Libraries Used:

- Numpy
- Pandas
- Matplotlib
- Seaborn
- sklearn

DATASET

	A	B	C	D	E	F	G	H	I
1	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot
2	Bachelors	2017	Bangalore	3	34	Male	No	0	0
3	Bachelors	2013	Pune	1	28	Female	No	3	1
4	Bachelors	2014	New Delhi	3	38	Female	No	2	0
5	Masters	2016	Bangalore	3	27	Male	No	5	1
6	Masters	2017	Pune	3	24	Male	Yes	2	1
7	Bachelors	2016	Bangalore	3	22	Male	No	0	0
8	Bachelors	2015	New Delhi	3	38	Male	No	0	0
9	Bachelors	2016	Bangalore	3	34	Female	No	2	1
10	Bachelors	2016	Pune	3	23	Male	No	1	0
11	Masters	2017	New Delhi	2	37	Male	No	2	0
12	Masters	2012	Bangalore	3	27	Male	No	5	1
13	Bachelors	2016	Pune	3	34	Male	No	3	0
14	Bachelors	2018	Pune	3	32	Male	Yes	5	1
15	Bachelors	2016	Bangalore	3	39	Male	No	2	0
16	Bachelors	2012	Bangalore	3	37	Male	No	4	0
17	Bachelors	2017	Bangalore	1	29	Male	No	3	0
18	Bachelors	2014	Bangalore	3	34	Female	No	2	0
19	Bachelors	2014	Pune	3	34	Male	No	4	0
20	Bachelors	2015	Pune	2	30	Female	No	0	1

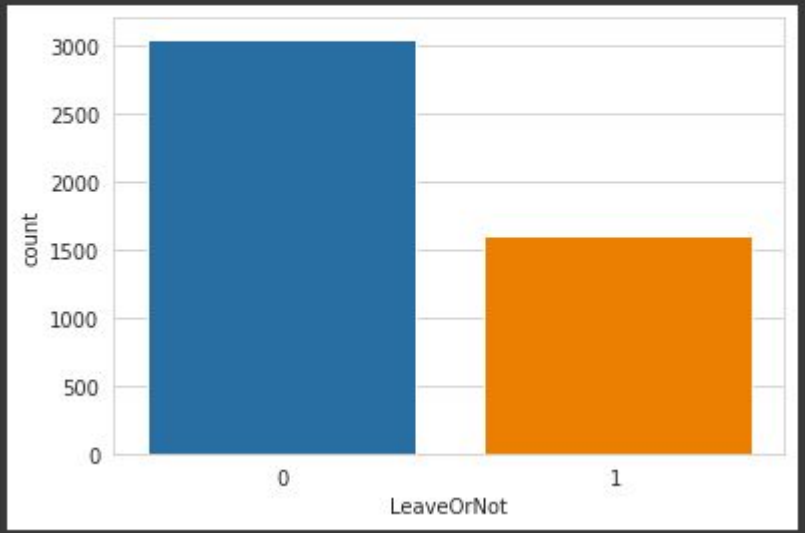
Data Reading

```
df = pd.read_csv('/content/Employee.csv')
df.head()
```

	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot
0	Bachelors	2017	Bangalore	3	34	Male	No	0	0
1	Bachelors	2013	Pune	1	28	Female	No	3	1
2	Bachelors	2014	New Delhi	3	38	Female	No	2	0
3	Masters	2016	Bangalore	3	27	Male	No	5	1
4	Masters	2017	Pune	3	24	Male	Yes	2	1

Data Preprocessing

```
[41] sns.countplot(x="LeaveOrNot", data=df)
     plt.show()
```



Test & Train data

```
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size = 0.2,random_state=0)
```

Algorithms Used:

- Logistic Regression
- KNN Algorithm
- SVM Algorithm
- Decision Tree
- Random Forest

Algorithms

Logistic Regression

```
0.7013963480128894
Test Accuracy 70.14%
confuision matrix
[[545  74]
 [204 108]]
```

KNN Algorithm

```
KNN-7 Score: 76.69%
confuision matrix
[[559  60]
 [157 155]]
```

Algorithms

SVM

```
Test Accuracy of SVM Algorithm: 66.49%
confuision matrix
[[619   0]
 [312   0]]
```

Decision Tree

```
Decision Tree Test Accuracy 81.31%
confuision matrix
[[544  75]
 [ 99 213]]
```

Algorithms

Random Forest

```
Random Forest Algorithm Accuracy Score : 82.28%
confuision matrix
[[555  64]
 [101 211]]
```

Model Comparison