Employee Attrition Classfication Model

Overview

Employee attrition is defined as the natural process by which employees leave the workforce – for example, through resignation for personal reasons or retirement.

problem statement attrition proves to be costly and time-consuming problem for the organization, and it also leads to loss of productivity.

The objective of this project is to classify the employee attrition for each employee.

Dataset

- **Dataset Source** from Kaggle named "IBM HR Analytics Employee Attrition & Performance "

- **The Data Contains** records of 1470 employee with 35 features.

- **It Has Information** about employees' current employment status, the total number of companies worked for in the past. Total number of years at the current company and the current role, the education level ,distance from home monthly income, etc.

EXPLORATORY DATA ANALYSIS

No Null Values

No Missing Values

No duplicated Rows and columns
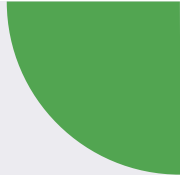
Data Processing

dropping non-relevant features

Converting categorical features into numeric form

Imbalanced Data

Convert the target variable for Yes and No to 0 and 1

Models

- Baseline model (KNN and Logistic Regression).
- Logistic Regression with Down sampling and dummy variables.
- Decision Tree Model.
- Random Forest Classifier Model.

Baseline model (KNN and Logistic Regression).

KNN

Training score: 0.83

Testing score:0.86

Cross_val:0.83

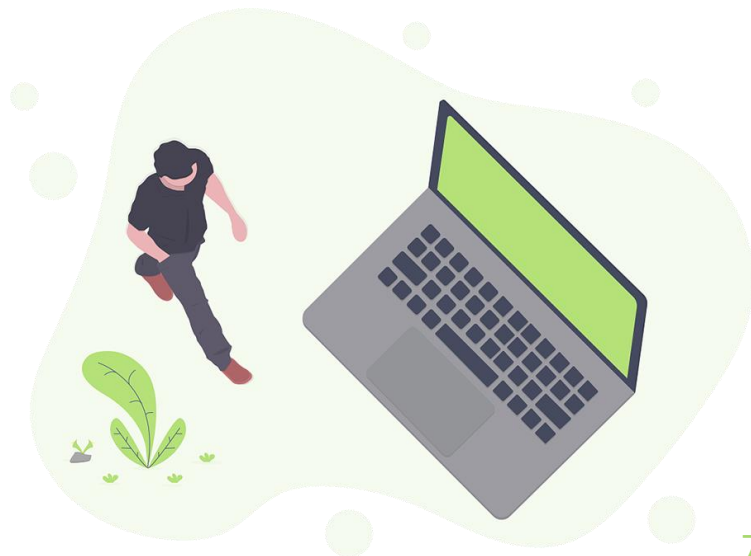**Logistic Regression**
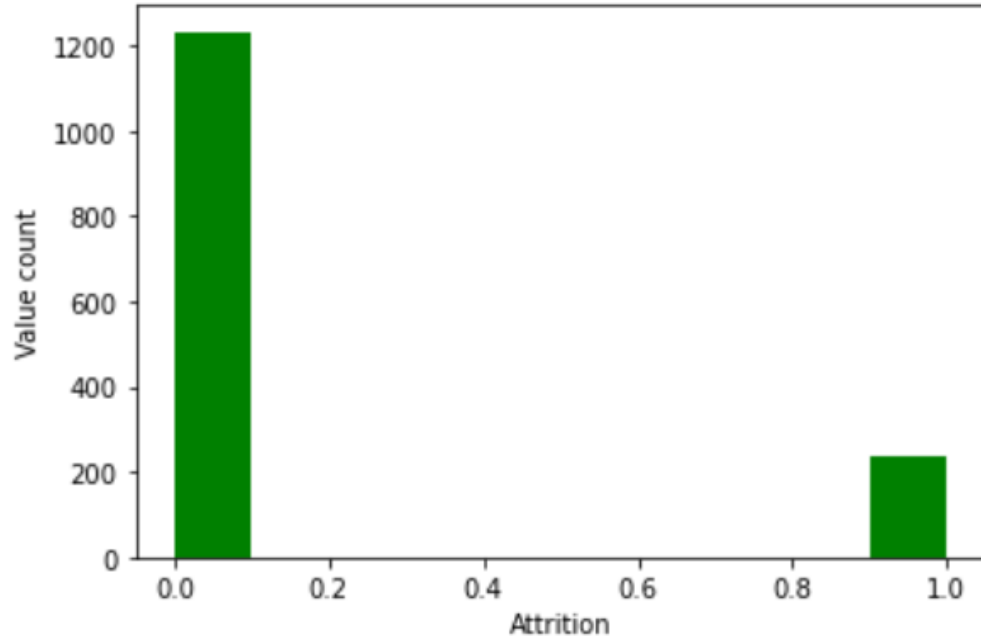
Training score: 0.83

Testing score:0.87

Cross_val:0.84

```
          precision    recall   f1-score

0            0.87        0.98       0.92
1            0.00        0.00       0.00
```

Class Imbalance

Logistic Regression with Down sampling and dummy variables

Down Sample :

Training score: 0.62

Testing score:0.66

Cross_val:0.84

Dummy variables :

Training score: 0.64

Testing score:0.62

Cross_val:0.84

```
              precision    recall  f1-score

           0       0.66      0.68      0.67
           1       0.67      0.65      0.66

    accuracy                           0.66
```

```
              precision    recall  f1-score

           0       0.65      0.51      0.57
           1       0.60      0.73      0.66

    accuracy                           0.62
```

Decision Tree Model.

1) Without pruning

Train score: 1.0

Test score:0.67

Cross Valdation :0.62

```
            precision    recall   f1-score

        0       0.66       0.71      0.68
        1       0.69       0.63      0.66

accuracy                             0.67
```

2) With pruning

Training score: 75

Testing score: 66

Cross validation : 60

```
            precision    recall   f1-score

        0       0.67       0.63      0.65
        1       0.66       0.70      0.68

accuracy                             0.66
```

Random Forest Classifier Model

Training Score : 0.83

Test Score : 0.71

Cross Validation : 0.74

```
              precision     recall    f1-score

      0          0.72         0.66        0.69
      1          0.69         0.75        0.72

accuracy                                  0.71
```

Conclution

the model will help organization to find ways to prevent attrition or to plan the hiring of new candidate.

Thanks!

Any questions?