Attrition Analytics: Employee Attrition

Team:-Roorkie

Contents:

1. Problem Statement
2. Objective
3. Implementation
4. Application idea
5. Result

Problem Statement :

People are the lifeblood of an organization. Employee attrition is an extremely important and relevant business problem all organizations face. The major problem with high attrition is the cost to the organization and the time and resource investment. We invite you to propose ideas that can help in predicting and finding proactive ways of reducing attrition using technologies of your choice For eg: use available data/ and create modeling tools to predict which employees are more likely to leave given some attributes like absenteeism, stagnation in a role, active disengagement, etc.

Objective:

The objective of the present report to study factor like BusinessTravel,Department, distancefromHome, policies and procedures, recognition, appreciation, suggestions of the employee’s by which it helps to know the Attrition level in the organizations and factors relating to retain them. This study also helps to find out where the organizations are lagging in retaining.

Implementation:

1. We had used a dataset from kaggle where we had attributes where we can

Find the attrition from the employee in the organisation and by help of many

Graph chart we will able to find which group are the most attrition in the

Different field and by help of age, salary and work experience we can predict

The attrition

2. We have used decision tree modelling which is greedy algorithm, it search

Entire space for possible decision tree, so we had find optimum parameter

For stopping decision tree at same point

3. We had use hypothesis with all attribute in the data set and it will help us

To find all the attrition by comparing with the attributes attrition at what range

In the organisation employs faces issue. For example age vs attrition,

Salary vs attrition, working hour’s vs attrition, job involvement vs attrition.

4. We will be using OneHotEncoder to predict the accuracy of the data

5. We will be using Logistic Regression and random forest algorithm to finding the accuracy of the attrition of the data.

6. We will be using the library of pivot and graphic for fetching the graph

7. We had evaluated by help of kfold validation for assessing how the result of

a Model will generalize to an independent test data set.

8. We had used also confusion matrix for tabulating the number of misclassification .From the confusion matrix we have seen there are lot of misclassifications. This is due the fact that the data is highly imbalanced. Decision Tree algorithm is bias towards classes which have number of instances, here in this dataset we have more number of employees didn’t leave the company. So the decision tree algorithm tend to only predict the majority

Class data. The features of the minority class are treated as noise and are often ignored. Thus, there is a high probability of misclassification of the minority class as compared to the majority class

Result:

\*\*From the classification report:-

Precision recall f1-score support

0...84 0.93 0.90 370

0.45 0.30 0.36 7

Age/total 0.80 0.83 0.81 441

\*\*due to imbalanced data we get f1-score \*\*

Application idea:

1. From above implementation we can able to conclude about what range these all attrition occurred.
2. We will able to know what type of problem are faced by the employee in working hour in the office as well as work from home.
3. What type talented employee in the office how we can use them in other platform and improve their skill which will be beneficial for the organisation.

Thank you