NAME: Raghav Jindal

UNIVERSITY ID: U83386066

SUBJECT: CS677 DATA SCIENCE WITH PYTHON

PROJECT SUMMARY

Employee Attrition, in layman terms can be explained as a long time period when a company fails to fulfill the vacancies. These vacancies could be a result of layoffs, resignation, retirements, illness, death or unable to rejoin from a leave of absence.

Based on the reason of the vacancy, attrition have been classified in two categories - Voluntary and Involuntary.

In the field of business management, over the years, a lot of emphasis is being put to Human Resources as importance of worker quality and skills id being factored in. In order to, feed the needs artificial intelligence has started to direct employee-related decisions with the HR department. This technological advancement is aimed to understand the factors that influence employee attrition and the reasons for such an attrition. Hence, it can be concluded that with the help of AI, behavioral data is being collected and analyzed in order to curb attrition and the contributing reasons.

GENERAL DESCRIPTION OF USECASE

Layoffs and the variables behind employee attrition are particularly important in the current situation and how the recession is anticipated since they have a significant impact on future economic situations.

We'll delve deeply into attrition and examine what factors—such as job satisfaction—drive employees to quit. Or which factors differ depending on the group of people? We'll make an effort to resolve that.

Need for predicting Attrition

* Demand for forecasting Attrition - A crucial worker quitting in the middle of a project could delay its conclusion. If a key team member were to quit, the team dynamics might be affected.
* Finding the perfect replacement for a lost employee can be challenging, and the cost of onboarding the new employee may rise.
* These unfavorable conditions could be prevented by anticipating employee turnover and providing incentives for staying to employees who have a high risk of leaving.
* The project's objective is to pinpoint the elements that influence employee attrition, including pay raises, career opportunities, the workplace, travel opportunities for business, relationships between managers and staff, appreciation and recognition, the length of time since the last promotion, etc. These factors would then be used to predict worker attrition.

KEY STEPS INVOLVED

1. Data Understanding– understanding the data and variables that can have an impact on the prediction.

2. Data Cleaning: The dataset had multiple columns with NULL values. The data cleaning

process involved filling the missing values with the median of the values.

3. Exploratory Data Analysis: In the exploratory data analysis section, we analyze the data and

try and find various patterns in the features of the dataset. Some of the key observations

are:

 The correlation values between features

 Looking at categorical and numerical variables by using different plots and diagrams to make inferences about the data

4.Data Splitting – Splitting the data set into the train (70%) and test (30%) data sets.

5. Model Building: To predict the final result of a Attrition we use various classification machine learning algorithms such as logistic regression, Decision Tree Classifier, Random Forest Classifier, and Support Vector Classifier.

6. Model Evaluation: To evaluate the classification model we calculate the confusion matrix, accuracy score, roc score, cross validation score.

KEY RESULTS

1.Random Forest Classifier algorithm gives the best result for classification of match results.

2.The accuracy achieved for the model is 97.54%.

3.The ROC score achieved for the model is 97.56%.

4.The Cross Validation Score achieved for the model is 99.08%.