**Employee Attrition Prediction using Machine Learning Algorithms**

***Team Members***

Sai Koushik Valluri - saikoushikvalluri@my.unt.edu ,

Venkata Durga Mani Ratna Srikar Duriseti - [maniratnasrikarvdduriseti@my.unt.edu](mailto:maniratnasrikarvdduriseti@my.unt.edu) ,

Bala Narayana Subbarao Chikkala - balachikkala@my.unt.edu ,

Naga Suchandra Tirumalasetti - nagasuchandratirumalasetti@my.unt.edu

***Workflow***

* Data Summary.
* Data Cleansing and Pre-processing.
* Exploratory Data Analysis.
* Model building with Imbalanced data.
* Model building with Balanced data using Random Oversampling and SMOTE techniques.
* Feature Selection.
* Model Evaluation.
* UI

***Abstract***

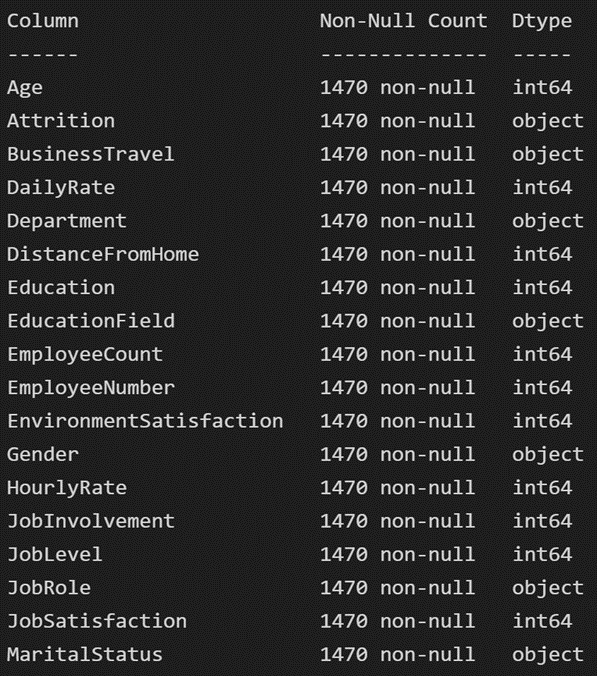
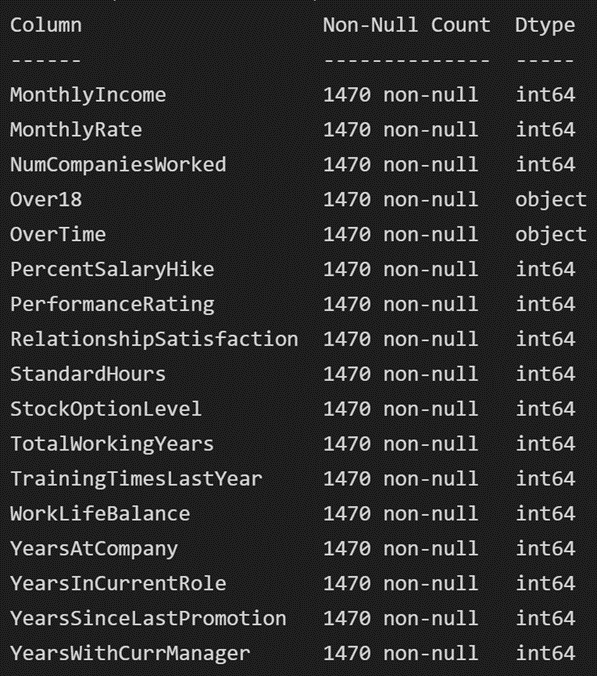
Employee attrition occurs when an employee leaves the organization through any method like resignation, retirement, layoffs etc. Employee Attrition is a major challenge to the organizations. It disrupts workflow management, decreases employee morale and destroys organization reputation. This project will provide a solution to predict employee attrition using a machine learning approach. Employee attrition is defined as the process by which employees leave the organization – for example, through resignation for personal reasons or retirement – and are not immediately replaced.

***Data Specification***

•This is a fictitious dataset created by IBM data scientists and published in Kaggle.

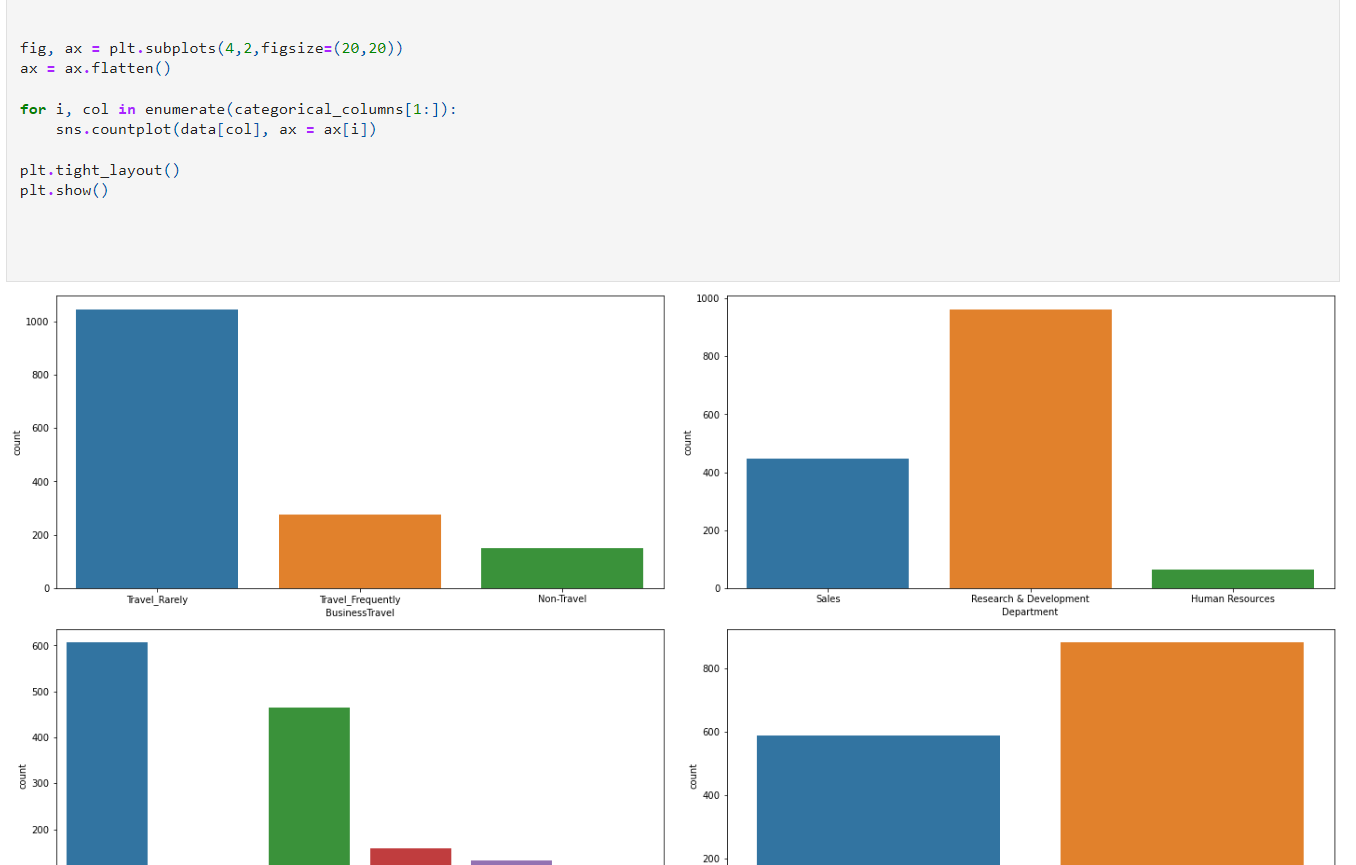
•Dataset contains 35 attributes like Education, Environment Satisfaction, Job Involvement, Job Satisfaction, Performance Rating, Relationship Satisfaction, etc.

•Attrition is the target variable which is categorical (Binary Class) in nature. This is eventually a classification problem which is supervised machine learning.

•This Dataset contains 26 integer columns, 6 string columns and 3 Boolean columns. 

***Project Design:***

We used Python for building ML algorithms and used R Programming languages to build UI. Since our target variable is a binary class, we used Logistic Regression, Decision Tree Classifier and Random Forest Classifier with different balanced and imbalanced approaches. We performed EDA with less lines of code using a for loop which saved a lot of time and effort. Please find the codes in below screenshots.

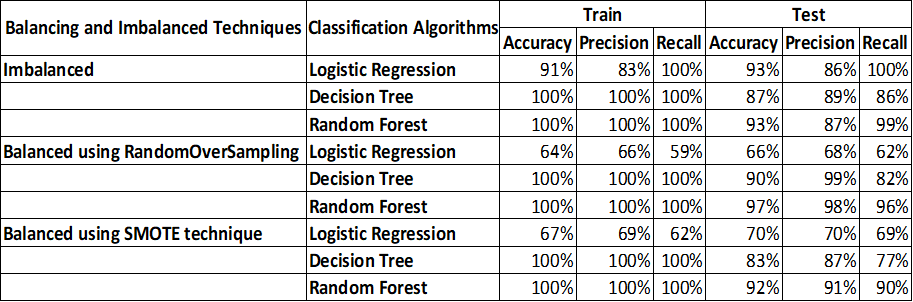
  
***Project Milestones:***

* Data Summary.
* Data Cleansing and Pre-processing.
* Exploratory Data Analysis.
* Model building.
* Features Selection.
* Model Evaluation.
* UI development.

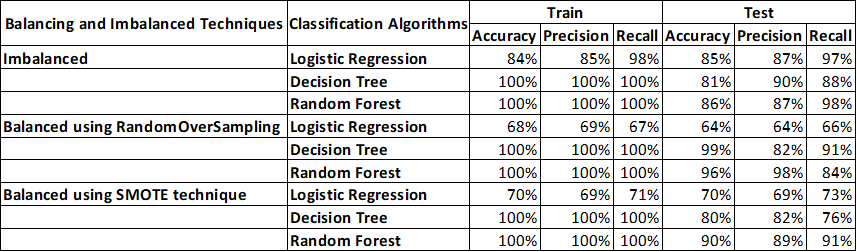
***Project Results:***

 We transformed Categorical Variables from String to Numerical data type using Label Encoder since ML models do not accept string data types.

We built models using three different approaches considering all the features and the performance metrics are listed below in the table.



We did feature selection based on EDA and below are the results after feature selection.



Based on the results obtained random forest algorithm without feature selection is the best performing machine learning algorithm to predict employee attrition.

***Repository / Archive:***

[**https://github.com/SrikarDuriseti/Employee-attrition-prediction.git**](https://github.com/SrikarDuriseti/Employee-attrition-prediction.git)

***References***

Reference1:

https://www.researchgate.net/publication/361522993\_Predicting\_Employee\_Attrition\_Us

ing\_Machine\_Learning\_Approaches

Reference 2:

https://ieeexplore.ieee.org/document/9825342

Kaggle:

https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/code