Name – Anubhav Dwivedi Roll No. – 10

**Project - HR-Employee**

**Introduction**

Exploratory Data Analysis (EDA) is the process of visualizing and analyzing data to extract insights from it.

I am working individually on HR Employee Dataset.

I wanted to do EDA on HR Employees to understand business problem. In this dataset we are having 1470 rows × 35 columns.

**Why I have chosen this dataset**

To determine whether a predictive model is a feasible analytical tool for business challenges or not.

It will not only help to explore more in EDA but this project is highly beneficial in my CV.

I came to know about data cleaning, preperation tools like univariate, bivariate etc, statistical tools.

My project is particularly related to BUSINESS domain.

The aim is to predict the employee analysis.

**COLUMNS:**

Age: age of the candidate

Education: To describe the level of education example diploma holder, graduate or 10th pass

Education Field: mechanical, technical etc course taken.

Employee

Number: their unique ids

Job Level: ranking according to their work in office

**Relation**

The notebook has been divided into three parts:

EDA

Feature Engineering

Machine Learning Models

**Conclusion**

Here we can compare the accuracy obtained by different Classification Models with different strategy

Here we can compare the accuracy obtained by different Classification Models with different strategy

Accuracy with all features means the all features of data were used for prediction of will employee left or not? This accuracy is obtained on the test data which was not used in training.

Accuracy with important features means the same as above but here only 5 most important features were used. The importance of features we got by using Random Forest Classifier.