**Employee-Performance-Analysis**

A trained model which can predict the employee performance based on factors as inputs. This will be used to hire employees

Recommendations to improve the employee performance based on insights from analysis

The given Employee dataset consist of 1200 rows. The features present in the data are 28 columns. The shape of the dataset is 1200x28. The 28 features are classified into quantitative and qualitative where 19 features are quantitative (11 columns consists numeric data & 8 columns consists ordinal data) and 8 features are qualitative. EmpNumber consist alphanumerical data (distinct values) which doesn't play a role as a relevant feature for performance rating.

From Correlation we can get the important aspects of the data, Correlation between features and Performance Rating. Correlation is a statistical measure that expresses the extent to which two variables are linearly related. The analysis of the project has gone through the stage of Univariate, Bivariate & Multivariate analysis, correlation analysis and analysis by each department to satisfy the project goal.

The dataset consists of Categorical data and Numerical data. The Target variable consist of ordinal data, so this is a classification problem

One of the important goals of this project is to find the important feature affecting the performance rating. The important features were predicted using the machine learning model feature importance technique.

The main technique used in the preprocessing data using the Manual & Frequency encoding method to convert the string - categorical data into numerical data, because, most of machine learning methods are based on numerical methods where strings are not supportive. The overall project was performed and achieved the goals by using the machine learning model and visualization techniques.

## 1. Requirement

The data was given from the IABAC for this project where the collected source is IABAC™. The data is based on INX Future Inc, (referred as INX ). It is one of the leading data analytics and automation solutions provider with over 15 years of global business presence. INX is consistently rated as top 20 best employers past 5 years. The data is not from the real organization. The whole project was done in Jupiter notebook with python platform.

## 2. Analysis

Data were analyzed by describing the features present in the data. the features play the bigger part in the analysis. The features tell the relation between the dependent and independent variables. Pandas also help to describe the datasets answering following questions early in our project. The data present in the dataset are divided into numerical and categorical data.

\*\*Categorical Features\*\*

EmpNumber

Gender

EducationBackground

MaritalStatus

EmpDepartment

EmpJobRole

BusinessTravelFrequency

OverTime

Attrition

\*\*Numerical Features\*\*

Age

DistanceFromHome

EmpHourlyRate

NumCompaniesWorked

EmpLastSalaryHikePercent

TotalWorkExperienceInYears

TrainingTimesLastYear

ExperienceYearsAtThisCompany

ExperienceYearsInCurrentRole

YearsSinceLastPromotion

YearsWithCurrManager

\*\*Ordinal Features\*\*

EmpEducationLevel

EmpEnvironmentSatisfaction

EmpJobInvolvement

EmpJobLevel

EmpJobSatisfaction

EmpRelationshipSatisfaction

EmpWorkLifeBalance

PerformanceRating