Project Business Case Analysis:

Data Requirement:

- The data is given in an excel file and is about a data analytics firm, whose employee performance indexes are coming down, and customer satisfaction levels are declining.
- To be able to take any remedial measures, the factors affecting the poor performance and to be able to predict if an employee is poorly performing or not based on the attributes, is to be determined.
- Seeking the help of ML, we are asked to develop a model that analyses the current employees and figures out the causes/factors of poorer performances. Additionally,the model can be designed to predict if an employee is underperforming or not and the rating of performance.
- Insights expected from the project are as given:
    1. Department wise performances
    2. Top 3 Important Factors affecting employee performance
    3. A trained model which can predict employee performance based on factors as inputs. This will be used to hire employees
    4. Recommendations to improve the employee performance based on insights from analysis.

On observing the data, the target variable that shall be the basis of our model is the overall performance rating. The data is clean with no misappropriation of data types or erroneous data. We will analyze the features, observe the correlation between the features and remove features which have higher correlation with other features. A threshold of 0.6 is put for feature correlation, to filter out high correlated features.

The data has some categorical columns too. We'll appropriately encode them, and accordingly.

To ensure that the range of values of data of all columns remain the same, so that the distances shall not influence the model equations, we do scaling on the entire data.

From the model perspective,it looks like a multi class classification problem with 3 classes since the overall performance rating of employees is rated as 2,3,4.

We shall use different classification algorithms, to generate a model and based on the results and f1 score, we shall pick the right model. We'll use multi class logistic regression, KNN, decision tree, Random Forest, Gradient Boosting algorithms with the data.