1. Problem Description

Human resources has been using analytics for years. However, the collection, processing and analysis of data has been largely manual, and given the nature of human resources dynamics and HR KPIs, the approach has been constraining HR. The goal is to try to use predictive and descriptive analytics in identifying the employees most likely to get promoted.

1. Project Pipeline
   1. Load dataset and required libraries
   2. Basic dataset statistics + conclusions
      1. Shape of the dataset
      2. Viewing some rows from the dataset
      3. Description of the numerical attributes
      4. Description of the categorical attributes
   3. Checking for null values in columns
   4. Checking for columns with only one value
   5. Checking for duplicate rows
   6. Dropping the 'employee\_id' column as it's not important in the following steps
   7. Data Visualization
      1. Uni-variate
         1. BoxPlots (unsuccessful)
         2. Histogram distributions for numerical columns
         3. Bar Plots for categorical columns
         4. Bar plot for the Target column that shows there is a clear imbalancing in classes
      2. Multi-variate:
         1. Correlation heatmap between variables
         2. Crosstab plot between each feature and the target to show its effect on the target
   8. Descriptive Analysis
      1. Association Rule Mining
   9. Data Preprocessing
      1. Handling of Categorical data using One Hot Encoding
      2. Handling Imbalancing in classes Using SMOTE
      3. Splitting data into train, validation and test
      4. Data Normalization
   10. Training Models
       1. Tried out many models to find the most suitable for our dataset
2. Analysis and Solution of the problem
3. Results and Evaluation
4. Unsuccessful Trials
5. Enhancements and Future Work