IBM HR Analytics Employee Attrition Modeling Project Report

**Executive Summary:**

The goal of this project is to build a predictive model to identify employees who are most likely to leave the company, and to understand the factors that contribute to attrition. The model will be used by HR to proactively address retention risk and improve employee satisfaction.

**Data Collection and Preprocessing:**

The data was collected from IBM's HR system and included information on demographics, job satisfaction, and performance metrics. The data was preprocessed to remove missing and irrelevant values, and to ensure data quality.

**Modeling Approach:**

A Random Forest Classifier was selected as the model due to its ability to handle complex relationships between variables and its robust performance in binary classification problems. The model was trained on a 70/30 train/test split of the data and evaluated using F1 score, precision, and recall.

**Feature Importance:**

The model identified the following features as the most important predictors of attrition:

Age

Total working years

Distance from home

Job level

Stock option level

**Results:**

The model achieved an F1 score of 0.83, precision of 0.84, and recall of 0.83. These results indicate that the model is effective at identifying employees who are likely to leave, with a low rate of false negatives (missed predictions) and false positives (incorrect predictions).

**Conclusion:**

The IBM HR Analytics Employee Attrition Modeling project was successful in identifying key factors that contribute to attrition and in building a predictive model to identify at-risk employees. This model will be useful for HR in proactively addressing retention risk and improving employee satisfaction. Further improvements to the model may be possible by incorporating additional data sources and using more advanced modeling techniques.