Employee Attrition Prediction Project Description

This project aims to predict employee attrition using a range of features that describe employee engagement and satisfaction within a company. Predictive modeling in this context assists HR departments in developing strategies to retain talent and reduce turnover rates, which is critical for maintaining organizational stability and growth.

The problem addressed in this project is the prediction of whether an employee will leave the company based on their work history and satisfaction levels. Accurately predicting attrition helps in proactive management of employee retention policies.

The project follows a structured data science workflow:

*1. **Data Preprocessing:** Key preprocessing steps include handling missing data, encoding categorical variables, and scaling numerical features to prepare data for modeling.*

*2. **Exploratory Data Analysis (EDA):** Initial data analysis involves generating visualizations to understand feature distributions and interactions, particularly focusing on factors influencing employee attrition.*

*3. **Model Selection and Training:** The Support Vector Machine (SVM) model was selected for its capability to model complex nonlinear relationships in the data. The model is trained using a pipeline that ensures sequential execution of preprocessing and learning tasks.*

*4. **Model Evaluation:** Beyond traditional ROC curves, precision-recall curves and F1-scores are*

utilized to provide a balanced view of model performance, especially useful in the context of imbalanced classes typical in attrition datasets.

*5. **Pipeline Integration:** Incorporating all preprocessing and modeling steps into a single pipeline ensures that procedures are reproducible and less prone to errors, making the model robust and scalable.*

The choice of SVM is driven by its efficiency in handling high-dimensional data and its flexibility through the kernel trick to adapt to non-linear decision boundaries. The evaluation metrics were chosen to provide a comprehensive understanding of the model's predictive power and reliability, critical for HR decision-making processes.

This enhanced predictive model provides an in-depth tool for HR professionals to forecast and mitigate potential employee turnover, thus allowing for more strategic planning in talent management and retention efforts.