Assignment Title: Employee Attrition Prediction Using Machine Learning Decision Trees

Objective:
In this assignment, you will apply various machine-learning techniques to predict employee attrition. You will perform exploratory data analysis (EDA), feature selection, and importance analysis, build a decision model, and evaluate its performance on a provided dataset of employee information.

Dataset:
You will be provided with a dataset containing various attributes related to employees in a company, including demographic information, job-related factors, and historical attrition data. The dataset will be provided in a CSV format. dataset

Tasks
1. Exploratory Data Analysis (EDA)
- Load the dataset and perform initial data exploration. - Summarize key statistics and visualizations for the dataset. - Identify missing values and decide on an appropriate strategy to handle them. - Explore the distribution of the target variable ('Attrition').
2. Feature Selection and Importance Analysis - Preprocess the data, including encoding categorical variables and scaling numerical features. - Apply feature selection techniques (e.g., correlation analysis, feature importance from tree-based models) to identify the most relevant features for attrition prediction. - Justify your selection of features based on their importance.

3. Model Building - Split the dataset into training and testing sets (e.g., 80% training, 20% testing). - Choose an appropriate machine learning algorithm for classification (e.g., Decision Trees, Random Forest, Logistic Regression, etc.). - Train a classification model on the training dataset using the selected features. - Fine-tune the hyperparameters of the model (e.g., using cross-validation) to optimize its performance.

4. Model Training, Testing, and Evaluation - Use the trained model to make predictions on the test dataset. - Evaluate the model's performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score, ROC-AUC). - Visualize the model's performance using confusion matrices or ROC curves. - Discuss the results and provide insights into the model's strengths and weaknesses.

5. Conclusion and Report - Write a comprehensive report summarizing your findings and methodology. - Include visualizations, tables, and graphs to support your analysis. - Provide recommendations for the company based on your model's insights.

Submission Guidelines: - Submit your assignment as a Jupyter Notebook or a well-documented Python script. - Include comments and explanations for each code block. - Submit the report as a separate document (PDF or Word).

Grading Criteria:
Your assignment will be evaluated based on the following criteria:

- Proper data preprocessing and handling of missing values.
- Effective feature selection and justification.
- Appropriate model selection and hyperparameter tuning.
- Thorough evaluation of the model's performance.
- Clarity and completeness of the report.
- Overall quality of code and documentation.

Important Dates:

- Assignment Release Date: 22/09/2023
- Assignment Due Date: 28/09/2023