

Assignment

Bank Marketing

In this assignment, you will work with the Bank Marketing dataset to predict whether or not a customer will subscribe to a term deposit. You will use decision trees, random forests, and XGBoost to build classification models and evaluate their performance.

Instructions:

1. Use bank.csv dataset
2. Load the dataset into a Jupyter notebook or another Python environment of your choice.
3. Preprocess the dataset by encoding categorical features and handling missing values, if any. Use the following steps:
 - a. Convert the categorical features to numerical using one-hot encoding or label encoding, depending on the nature of the feature.
 - b. Handle missing values in the dataset. You can either drop the rows with missing values or impute them using techniques like mean or median imputation.
4. Split the dataset into training and testing sets using a 70/30 split.
5. Build a decision tree classifier using scikit-learn's DecisionTreeClassifier and train it on the training set.
6. Evaluate the performance of the decision tree classifier on the testing set using accuracy, precision, recall, and F1-score.
7. Build a random forest classifier using scikit-learn's RandomForestClassifier and train it on the training set.
8. Evaluate the performance of the tuned random forest classifier on the testing set using accuracy, precision, recall, and F1-score.
9. Build an XGBoost classifier using the xgboost library and train it on the training set.
10. Evaluate the performance of the tuned XGBoost classifier on the testing set using accuracy, precision, recall, and F1-score.
11. Create a visualization of all the three algorithms accuracy with respect to there

accuracy

Deliverables:

Your deliverables for this assignment should include:

1. A Jupyter notebook or another Python script with your code.
2. A report summarizing your results, including the performance of the decision tree, random forest, and XGBoost classifiers, as well as a comparison of their performance.
3. A description of your preprocessing steps and how they impacted the performance of the classifiers

DataSet-<https://drive.google.com/file/d/1XkoLTkJMy2pheCzkDEio03FrWaYvHw5I/view?usp=sharing> .