Project Demonstration: Lazy Classification

Introduction

This project focuses on lazy classification, which involves training and evaluating various classification models on a given dataset. The goal is to predict whether a student is lazy or not based on their performance in subjects such as Maths, Physics, and Chemistry.

Data Analysis and Preprocessing

- The dataset is read from a CSV file and stored in a Pandas DataFrame.
- Initial exploratory data analysis is performed, including shape, info, and descriptive statistics of the dataset.
- Null values in the dataset are checked and removed if present.
- Numerical variables are scaled using the StandardScaler.

Data Visualization

- Distribution plots are created to visualize the distribution of marks in Maths, Physics, and Chemistry subjects.
- Box plots are generated to identify outliers in the dataset.

Model Training and Evaluation

The following classification models are trained and evaluated on the dataset:

1. Logistic Regression:
    - The model is trained using the training set.
    - Predictions are made on the test set.
    - Confusion matrix and accuracy score are calculated.
2. Naive Bayes:
    - The Gaussian Naive Bayes model is trained using the training set.
    - Predictions are made on the test set.
    - Confusion matrix and accuracy score are calculated.
3. Support Vector Machines (SVM):
    - The SVM model is trained using the training set.
    - Predictions are made on the test set.
    - Confusion matrix and accuracy score are calculated.
4. K-Nearest Neighbors (KNN):
    - The KNN model with k=5, metric='minkowski', and p=2 is trained using the training set.
    - Predictions are made on the test set.
    - Confusion matrix and accuracy score are calculated.

5. Decision Tree:
    - The Decision Tree model is trained using the training set.
    - Predictions are made on the test set.
    - Confusion matrix, accuracy score, and the tree visualization are displayed.
6. Decision Tree with Cost Complexity Pruning:
    - The Decision Tree model with cost complexity pruning is trained using the training set.
    - The optimal alpha value is determined using accuracy vs. alpha plots.
    - Predictions are made on the test set using the pruned tree.
    - Confusion matrix, accuracy score, and the pruned tree visualization are displayed.

Model Deployment

- The trained Decision Tree model is saved using the `pickle` module for future use.
- The saved model is loaded back into memory to make predictions on new data.

Conclusion

The project demonstrates the process of lazy classification by training and evaluating various classification models on a student performance dataset. The Decision Tree model, with and without cost complexity pruning, achieves the highest accuracy on the test set. The pruned Decision Tree model provides a more generalized and interpretable solution.

My Project Link :
https://github.com/kavimiya/Project/blob/main/Laszy%20predection%20project.ipynb