Target Audience Prediction

Introduction
This document provides an overview of a classifier comparison and decision boundary visualization using various machine learning classifiers. The classifiers are evaluated on the "Social_Network_Ads" dataset, aiming to predict whether a user purchased a product based on their age and estimated salary.

Dataset
The dataset, "Social_Network_Ads.csv," contains information about users' age, estimated salary, and purchase decision. It is loaded and preprocessed for analysis.

Classifiers
The following classifiers are used for prediction and comparison:
Decision Tree Classifier (Entropy-based)
Support Vector Classifier (SVC) with Radial Basis Function (RBF) Kernel
Gaussian Naive Bayes Classifier
Random Forest Classifier
Support Vector Classifier (SVC) with Linear Kernel
k-Nearest Neighbors (KNN) Classifier
Logistic Regression

Workflow

Data Preparation: The dataset is loaded, and the features (age and estimated salary) are extracted, along with the target variable (purchase decision).

Data Splitting and Standardization: The dataset is split into training and testing sets using a 75-25 split ratio. The features are standardized using the StandardScaler to ensure consistent scaling for model training.

Classifier Comparison: Each classifier is trained on the training data and evaluated on the testing data. Accuracy scores and confusion matrices are calculated to assess classifier performance.

Decision Boundary Visualization: For each classifier, the decision boundary is visualized on the test set. Age and estimated salary are used as the x and y axes, respectively. Points are colored according to their true class label, providing insights into how well the classifier separates the classes.

Results
The performance of each classifier is evaluated based on accuracy and the confusion matrix:

Conclusion

The classifier comparison and decision boundary visualization provide insights into the performance of different machine learning classifiers on the "Social_Network_Ads" dataset. The Support Vector Classifier (SVC) with RBF Kernel and k-Nearest Neighbors (KNN) Classifier achieved the highest accuracy (0.93) in predicting whether a user purchased a product. The decision boundary visualizations enhance our understanding of how these classifiers separate the classes based on age and estimated salary. This analysis can guide the selection of an appropriate classifier for similar prediction tasks.