PROJECT REPORT

ASSIGNMENT-4

CSE – 6363: MACHINE LEARNING

SUBMITTED BY:

NAME: DIVYA DARSHI

UID: 1002090905

DECISION TREE:

A decision tree is a powerful and widely used tool in machine learning and data analysis. It is a flowchart-like structure in which each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label or a decision. Decision trees are particularly useful for decision-making and classification tasks because they provide a clear and intuitive representation of the decision-making process.

OUTPUT:

Decision Tree Accuracy : 0.825112

RANDOM FOREST:

Random Forest is a powerful and versatile ensemble learning method in machine learning. It is particularly popular for both classification and regression tasks. The term "ensemble" refers to the technique of combining multiple models to create a stronger and more robust predictor.

The basic idea behind a Random Forest is to build a multitude of decision trees during training and output the mode (classification) or average prediction of the individual trees for a given input. Each tree in the forest is trained on a random subset of the training data, and a random subset of the features is considered for each split in the tree. This introduces randomness into the learning process, and it helps to reduce overfitting.

OUTPUT:

Random Forest Accuracy: 0.807175

ADABOOST :

The main idea behind AdaBoost is to combine the predictions of multiple weak learners to create a strong learner. A weak learner is a model that performs slightly better than random chance, and it is often a simple model, like a shallow decision tree. AdaBoost assigns weights to each training example and adjusts them at each iteration to give more emphasis to the misclassified samples. This way, subsequent weak learners focus more on the examples that previous models found difficult to classify correctly.

AdaBoost has several advantages, including its simplicity and the fact that it can adapt to complex datasets. However, it's important to note that AdaBoost can be sensitive to noisy data and outliers.

OUTPUT:

AdaBoost Accuracy: 0.766816

LEARNING OUTCOMES:

Decision Trees:

Interpretability: Decision trees are easy to understand and interpret. The rules created by a decision tree can be visualized, making it a useful tool for explaining the decision-making process to non-experts.

Handling Non-linearity: Decision trees can model complex, non-linear relationships in the data. They are versatile and can handle both numerical and categorical data without the need for extensive preprocessing.

Feature Importance: Decision trees can provide a measure of feature importance, indicating which features are more influential in making decisions.

Random Forest:

High Accuracy: Random Forests generally achieve high accuracy in predictions. By aggregating the results of multiple decision trees, the model reduces overfitting and provides robust predictions.

Reduction of Overfitting: The use of multiple trees in a Random Forest helps mitigate overfitting, a common issue with individual decision trees. The ensemble approach allows for better generalization of unseen data.

Feature Importance: Random Forests provide a more robust estimate of feature importance compared to individual decision trees, as it considers the average importance over all trees.

AdaBoost (Adaptive Boosting):

Improved Accuracy: AdaBoost focuses on improving the accuracy of the model by assigning higher weights to misclassified instances. This adaptability often leads to higher accuracy compared to individual weak learners.

Versatility: AdaBoost can be used with different base learners, making it versatile. It is not limited to decision trees and can be combined with various types of weak learners.

Handles Class Imbalance: AdaBoost is effective in handling class imbalance. Assigning higher weights to misclassified instances, puts more emphasis on the minority class, improving its predictive performance.