Project Title: Decision Tree Algorithm for Classification of Breast Cancer (Wisconsin Dataset)

Project Overview

This project demonstrates the application of the Decision Tree algorithm to classify breast cancer tumors from the Wisconsin dataset. The primary goal is to accurately distinguish between malignant and benign tumors using various parameters and evaluation metrics.

Dataset

The dataset used in this project is the Wisconsin Breast Cancer Dataset, which includes features extracted from digitized images of fine needle aspirate (FNA) of breast masses. The target variable consists of two classes:

*Malignant

*Benign

Algorithm Used

The algorithm implemented for classification is the Decision Tree Classifier, utilizing the Gini impurity criterion for splitting nodes.

Methodology

The project workflow includes the following key steps:

Data Preparation: Loading and preprocessing the dataset.

Train-Test Split: Dividing the dataset into training and testing sets to evaluate model performance.

Model Training: Training the Decision Tree model on the training set.

Evaluation: Assessing the model's accuracy and precision on the test set.

Results

The Decision Tree model achieved an impressive precision score (accuracy) of 94.2%, indicating a high level of accuracy in classifying breast cancer tumors as malignant or benign.

Hyperparameters

The model was configured with the following hyperparameters:

Criterion: 'gini'

Max Depth: 4

Min Samples Split: 2

Min Samples Leaf: 1

CCP Alpha: 0.0

Class Weight: None

Splitter: 'best'

Future Work:

To enhance the model's performance and robustness, the following future work is proposed:

Conduct Pruning: Implement pruning techniques to reduce overfitting and improve model generalization.

Evaluate Accuracy: Continuously evaluate the model's accuracy with updated datasets and techniques.

Compare with Random Tree Algorithm: Investigate the performance differences between the Decision Tree and Random Tree algorithms.

Optimize Additional Parameters: Explore the effect of optimizing various parameters, including:

max_features

max_leaf_nodes

min_impurity_decrease

min_weight_fraction_leaf

Conclusion

This project showcases the effectiveness of the Decision Tree algorithm in classifying breast cancer tumors, achieving high accuracy and precision. The insights gained can inform future research and applications in the field of medical diagnosis and machine learning.

Acknowledgments

Special thanks to the contributors of the Wisconsin Breast Cancer Dataset and the community for providing valuable resources in machine learning.