Decision Tree Documentation

1. Introduction

A Decision Tree is a supervised machine learning algorithm used for classification and regression tasks. It uses a tree-like structure to model decisions based on feature values.

2. Objective

- Develop an interactive web application using Streamlit for Decision Tree classification.
- Provide users with the ability to upload a dataset, select features, train the model, and evaluate its performance.

3. Dataset Description

- Dataset: **A suitable dataset**
- Features: Describe the input variables
- Target: Describe the variable to predict
- Number of records and attributes

4. Implementation Details

- **Frontend**: Developed using Streamlit for user interaction.
- **Backend**: Decision Tree implemented using `scikit-learn`.
- **Steps**:
  1. Upload the dataset in CSV format.
  2. Select features and target variables.
  3. Configure hyperparameters like depth, criterion (Gini or Entropy), and minimum samples per leaf.
  4. Train the Decision Tree model.

5. Evaluate the model using accuracy, confusion matrix, and classification report.

5. Results and Analysis

- Provide model accuracy and other evaluation metrics.
- Visualize the decision tree using `plot_tree()`.
- Discuss how different hyperparameters impact model performance.

6. Challenges and Solutions

- Managed overfitting by pruning and limiting tree depth.
- Addressed imbalanced datasets using appropriate class weighting.
- Improved model interpretability using visualization.

7. Conclusion

Decision Trees are intuitive and interpretable models suitable for both classification and regression tasks. The web application allows users to build, visualize, and analyze Decision Trees effectively.

8. References

- Scikit-learn Documentation
- Streamlit Documentation
- Dataset Source