Proposal for Iris Flowers Classification

Introduction

The classification of the Iris Flowers dataset is a widely recognized classification problem in machine learning. Collected by the British biologist Ronald Fisher, this dataset includes 150 samples from three Iris species: Iris setosa, Iris versicolor, and Iris virginica. Each flower is represented by four features: sepal length, sepal width, petal length, and petal width. By applying classification techniques, we can predict the species of a flower based on these attributes.

Problem Statement

Accurate classification of Iris flower species based on measurable features remains a fundamental problem in machine learning. Despite the simplicity of this dataset, it allows the exploration of fundamental classification techniques that are also applicable in more complex real-world scenarios. This project aims to develop a model that can classify the species of Iris flowers with high accuracy by learning patterns from the dataset. The model will use features extracted from flowers and determine the best algorithm for classification.

Goals

The primary goals of this project are:

1- Data Preprocessing: Clean, analyze, and prepare the dataset for modeling.
2- Model Selection and Training: Implement logistic regression, decision trees, random forest, Xtreme Gradient Boosting, naive bayes and neural network.
3- Evaluation: Assess the models using metrics such as accuracy, precision, recall, F1-score and Confusion Matrix.
4- Optimization: Fine-tune the best-performing model to improve classification accuracy.
5- Documentation and Presentation: Summarize findings, visualize results, and provide a clear explanation of the model's performance.

Related Work

Numerous studies have utilized the Iris dataset for machine learning experiments. Some notable approaches include:

- Traditional Machine Learning:
    - Application of classical algorithms like Decision Trees, Random Forest, SVM, KNN, and Naive Bayes to achieve high classification accuracy.
- Deep Learning:
    - Exploration of deep neural networks, especially Convolutional Neural Networks (CNNs), to extract complex features from the data and improve classification performance.
- Ensemble Methods:
    - Combination of multiple models, such as bagging and boosting, to enhance overall accuracy and robustness.