Mushroom Classification Using Machine Learning

Using the UCI Mushroom Data Set

Angela Lombard

July 3, 2024

Project Overview

- Objective: Develop a model to classify mushrooms as edible or poisonous using the UCI Mushroom Data Set.
- Tools Used: Python, Pandas, Scikit-learn, Matplotlib, Seaborn.

Data Collection

- Dataset: UCI Mushroom Data Set
- Features: Various characteristics of mushrooms such as cap shape, cap surface, cap color, odor, gill size, etc.
- Target: Class (edible or poisonous)

Data Preprocessing

- Encoding categorical variables
- Handling missing values
- Splitting dataset into training and testing sets

Exploratory Data Analysis (EDA)

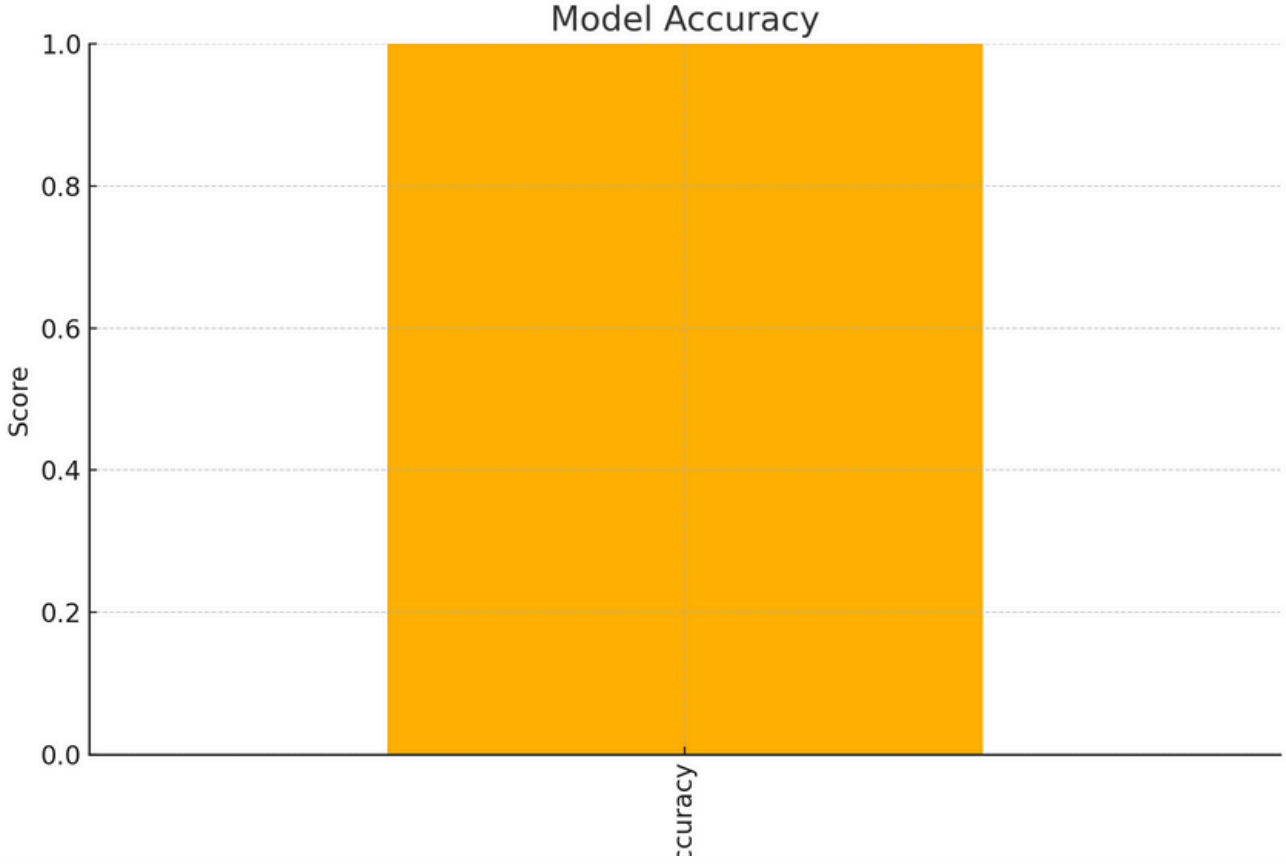
- Understanding the distribution of classes
- Identifying key features such as odor and gill size contributing to classification

Class Distribution

Predictive Modeling

- Goal: Classify mushrooms as edible or poisonous
- Model Used: RandomForestClassifier
- Features: Various characteristics of mushrooms

Model Performance



Classification Report

Accuracy: 1.00

	precision	recall	f1-score	support
0	1.0	1.0	1.0	782.0
1	1.0	1.0	1.0	843.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	1625.0
weighted avg	1.0	1.0	1.0	1625.0

Conclusion

- Summary: Developed a model to classify mushrooms as edible or poisonous with high accuracy.
- Future Work: Explore additional classification algorithms and feature engineering techniques.