Classifying Edible and Poisonous Mushrooms Using Machine Learning

Nick Kornienko

September 21, 2023

Abstract

This paper presents a machine learning approach to classify mushrooms as either edible or poisonous based on various features. The project follows the CRISP-DM methodology and evaluates the performance of multiple machine learning algorithms.

1 Introduction

Mushrooms are a subject of culinary and scientific interest. However, the critical question is: Is the mushroom you found in the woods edible or poisonous? This project aims to solve this problem by applying machine learning techniques for classification.

2 Methodology

The project uses the CRISP-DM methodology, which includes the following phases:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

2.1 Business Understanding

The objective is to build a classification model with high accuracy and minimal false negatives.

2.2 Data Understanding

We use a publicly available dataset containing various features such as cap shape, cap color, and odor. Preliminary data analysis shows that the dataset is balanced and has no missing values.

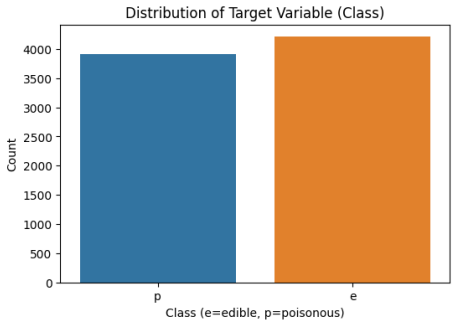

Figure 1: Distribution of Target Variable

2.3 Data Preparation

Data was preprocessed to convert categorical variables into numerical format and split into training and testing sets.

2.4 Modeling

Three machine learning algorithms were evaluated: Random Forest, Support Vector Classifier, and k-Nearest Neighbors.

2.5 Evaluation

All models showed high accuracy, with Random Forest and k-NN achieving perfect scores.

2.6 Deployment

The model can be deployed in a production environment, and continuous monitoring is recommended.

3 Conclusion

The project successfully classifies mushrooms with high accuracy, thereby serving its purpose effectively. Further research can focus on more features and different machine learning algorithms.