Mushroom Classification

Detailed Project Report

Abstract

Mushrooms are an essential nutrient-dense element in our diet. Because most mushrooms are toxic (inedible), we must distinguish between poisonous and edible mushrooms. Machine learning (ML) techniques such as logistic regression, decision trees, Random forest, and others were used to categorize mushroom attributes as edible or not. There has been little study on mushroom classification; existing research has concentrated on applying ML approaches separately, with some algorithms outperforming others in terms of accuracy. This study developed an integrated model that combines the most accurate technique's judgements into a single decision rather than addressing them separately. Kaggle mushroom dataset downloaded. The results demonstrate that the xgboost regressor outperforms other strategies by 100% accuracy score.

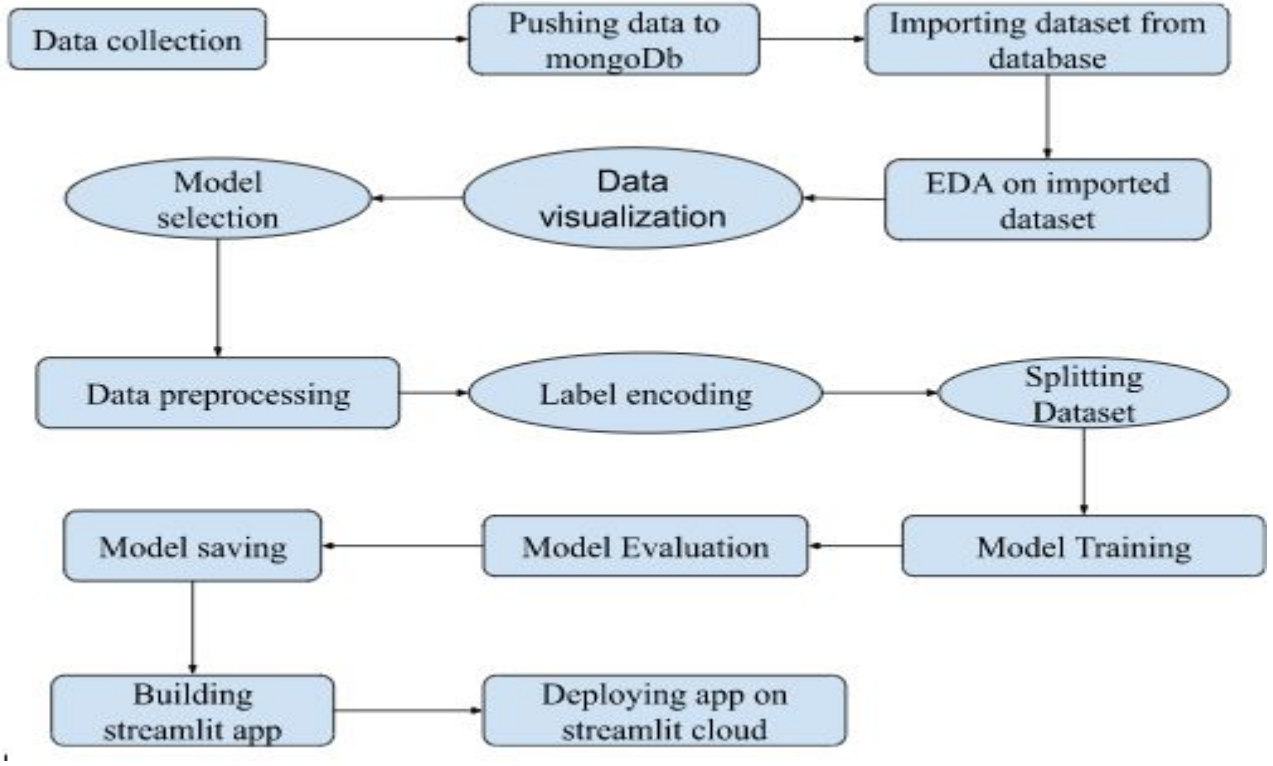
OBJECTIVE

The main objective is to figure out which mushrooms are toxic and which are edible.

Problem Statement

The Audubon Society Field Guide to North American Mushrooms includes descriptions of hypothetical samples belonging to 23 Agaricus and Lepiota Family Mushroom species (1981). Each species is classified as certainly edible, definitely toxic, or maybe edible but not advised. This last category has been combined with the poisonous category. The Guide claims explicitly that there is no straightforward criterion for determining the edibility of a mushroom, such as "leaflets three, leave it alone" for Poisonous Oak and Ivy. The main objective is to figure out which mushrooms are harmful and which are edible.

ARCHITECTURE



DATASET COLLECTION

The data for this project was gathered from the Kaggle Dataset, the URL for which is provided below

Dataset: https://www.kaggle.com/datasets/uciml/mushroom-classification.

Exploratory Data Analysis

The dataset is in the form of a csv file. This data set has 8124 rows and 23 columns. The columns are all of the categorical variety. In our goal column, there are two classes: 'p' - poisonous and 'e' - edible. Also, in our data, I have approximately equal counts for deadly and edible classifications. As a result, I can state that our data is balanced.

Model Training and Model Evaluation

I selected XGBClassifier as a model for model training since it was quite quick compared to the other models and delivered 100% accuracy on both train and test data, which was ideal for our project.

Model Deployment

I constructed a web app using the streamlit app and user interface, integrated our model with the streamlit app and UI, and tried it on my local PC first. Then I used streamlit cloud to deploy my model. I used many input combinations to anticipate the outcome, and the findings were correct. There were no problems discovered with the app.

Conclusion

This project is about pre-processing procedures, processes to discover essential features that aid in the categorization of edible and deadly mushrooms, and a comparison of attribute selection methods to determine whether both approaches yield the same outcome.