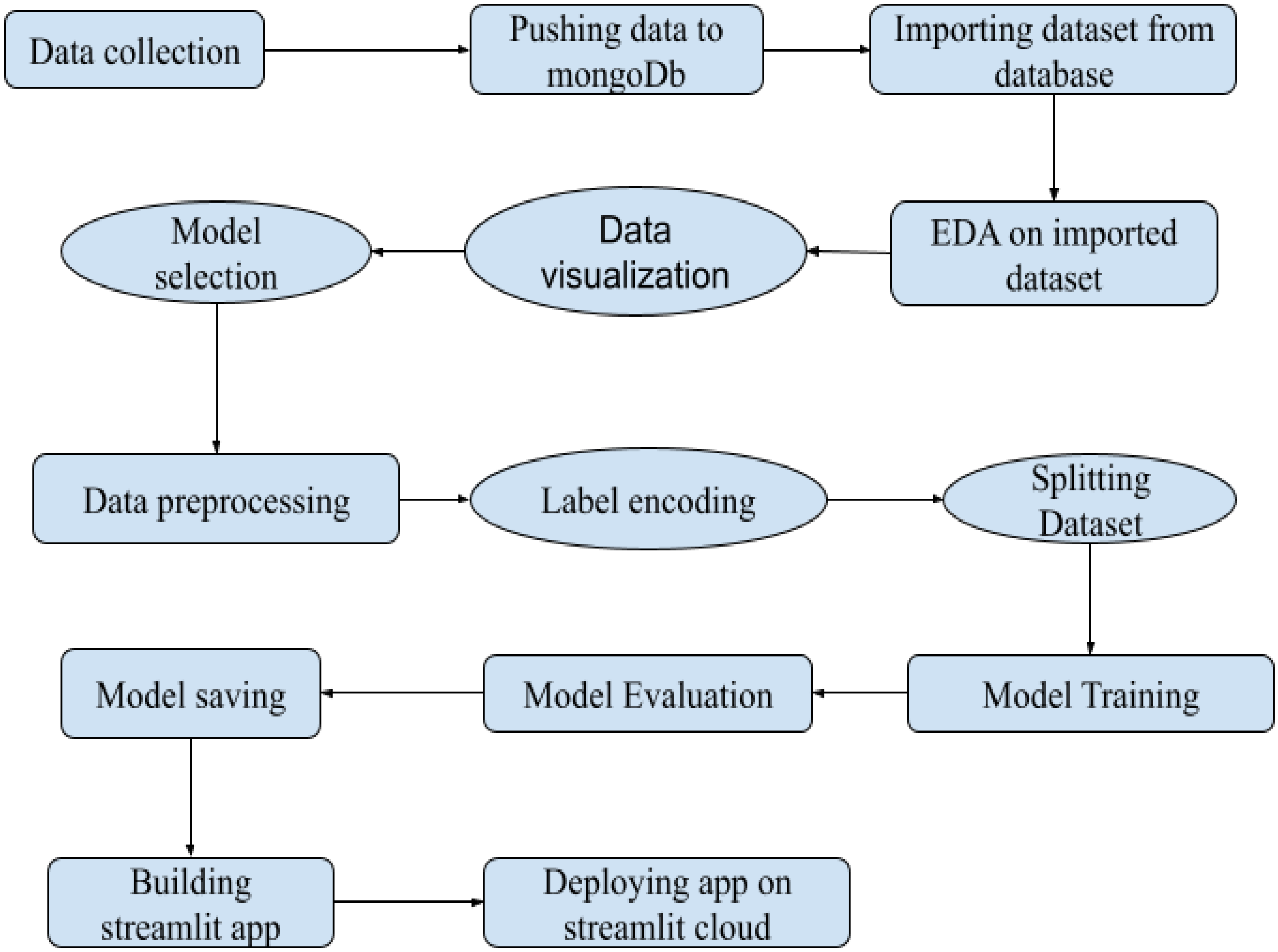
Architecture (Mushroom-Classification)

Contents

Architecture



Architecture Description

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.

Data Description :

This dataset contains descriptions of hypothetical samples matching to 23 species of gilled mushrooms from the Agaricus and Lepiota Family Mushroom, as found in The Audubon Society Field Guide to North American Mushrooms (1981). Each species is labeled as either certainly edible, definitely toxic, or maybe edible but not advised. This latter category was integrated with the toxic category.

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.

Data Preprocessing:

I'll be studying our data collection here and performing data preparation as needed. I first investigate our data collection in Jupyter Notebook to determine what pre-processing and validation I need to convert all of those to numerical values via label encoding, and then I have to develop distinct modules based on our analysis so that I can apply it for both training and prediction data.

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.