

Mushroom Classification



By Md Imran

Objective

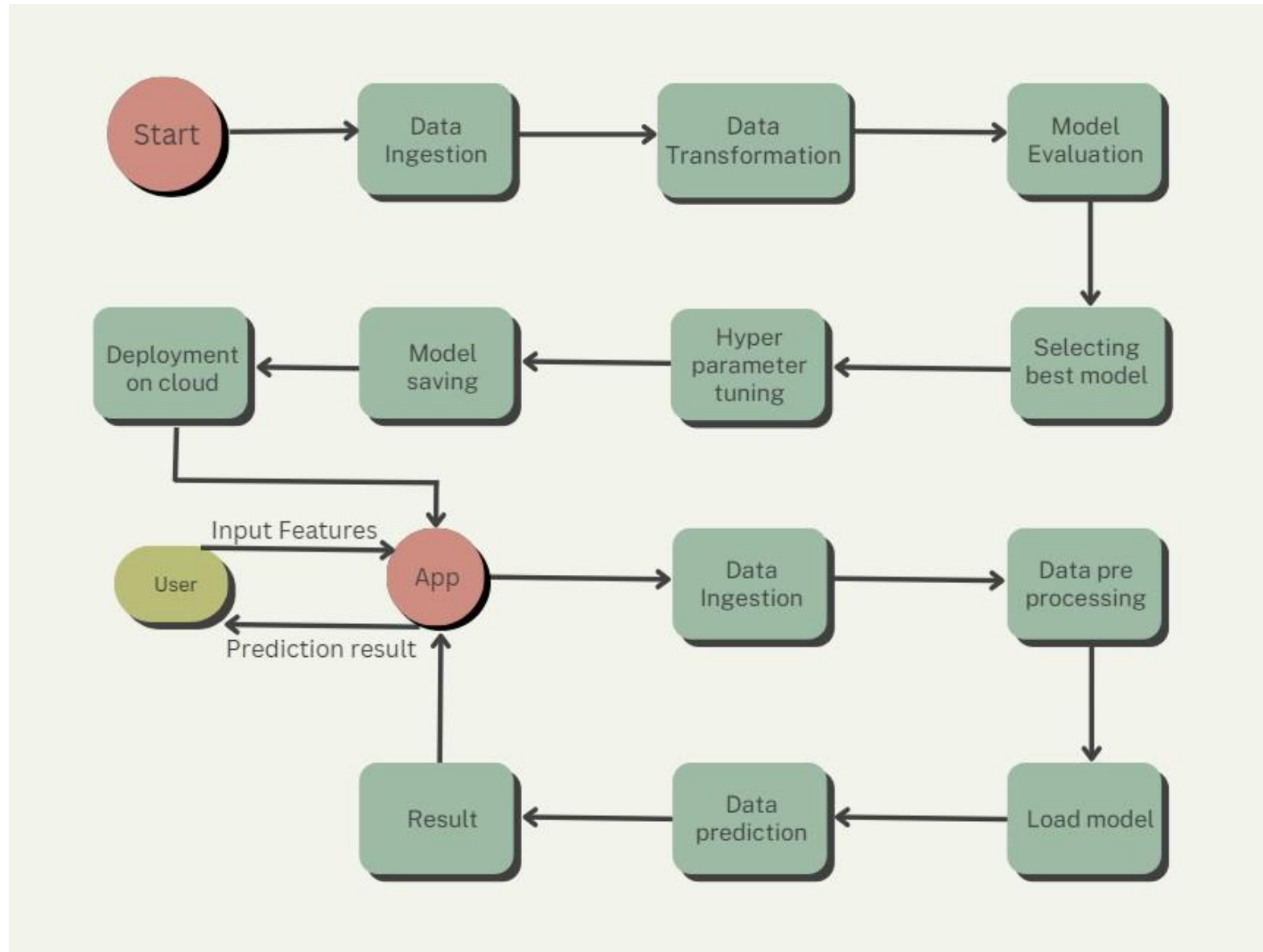
The main objective of the project is to help people who are not familiar with the mushroom types by classifying the mushroom is edible or poisonous

Benefits:

- Identify mushroom edibility
- Preventing health issues
- Enriching nutrition intake
- Promoting good mushrooms



Architecture



Data Ingestion

Data Link: <https://www.kaggle.com/uciml/mushroom-classification>

- Data is downloaded in csv format from this link.
- Data is read from that csv and converted and stored in a data frame for further process.
- Data is split in training and testing data and they are individually stored in different csv.

Data Validation and Data Transformation

- Data type of columns – All features of this data is in object format. They are converted to numerical format through one hot encoding.
- Null values in columns – This data does not contain null values but there are some missing values which are filled through imputation.
- Name of columns - The name of the columns are as per the columns name of the dataset
- Target column is separated for training and testing data and preprocessing is done on it.
- Preprocessing object is stored in pickle format for further use.

Model Training

- The transformed data is trained through models and then the accuracy of its prediction is tested
- These models are tested in the data using
 1. RandomForestClassifier
 2. XGBoostClassifier
 3. GradientBoostClassifier
- Out of these models Support Vector Classifier gave the best results and thus it was stored in pickle format for further use.

Prediction

- Input: User enters the input data through UI.
- Preprocessing: The data is stored and transformation is applied on it by loading pickle file.
- Model Prediction: Output is then predicted by using our model stored in pickle format.
- Results: The output received after model prediction is presented to user on UI.

Thank You

- The application source code is uploaded on GitHub.
- Application is hosted through AWS.

