iNeuron

Mushroom Classification Report

By

ANNAMALAI SUBRAMANI

Revision Number: 1.0
Last date of revision: 29/01/2024

Abstract:

This study delves into machine learning for mushroom classification, offering a time-efficient alternative to traditional methods. A dataset of labeled mushroom samples, coupled with feature extraction techniques, enables the training and evaluation of machine learning models, showcasing heightened accuracy compared to conventional approaches. The findings emphasize the potential for automated mushroom identification and quality control in the food industry.

Objective:

Develop a predictive model for Mushroom classification to discern whether a mushroom is edible or poisonous.

Benefits:

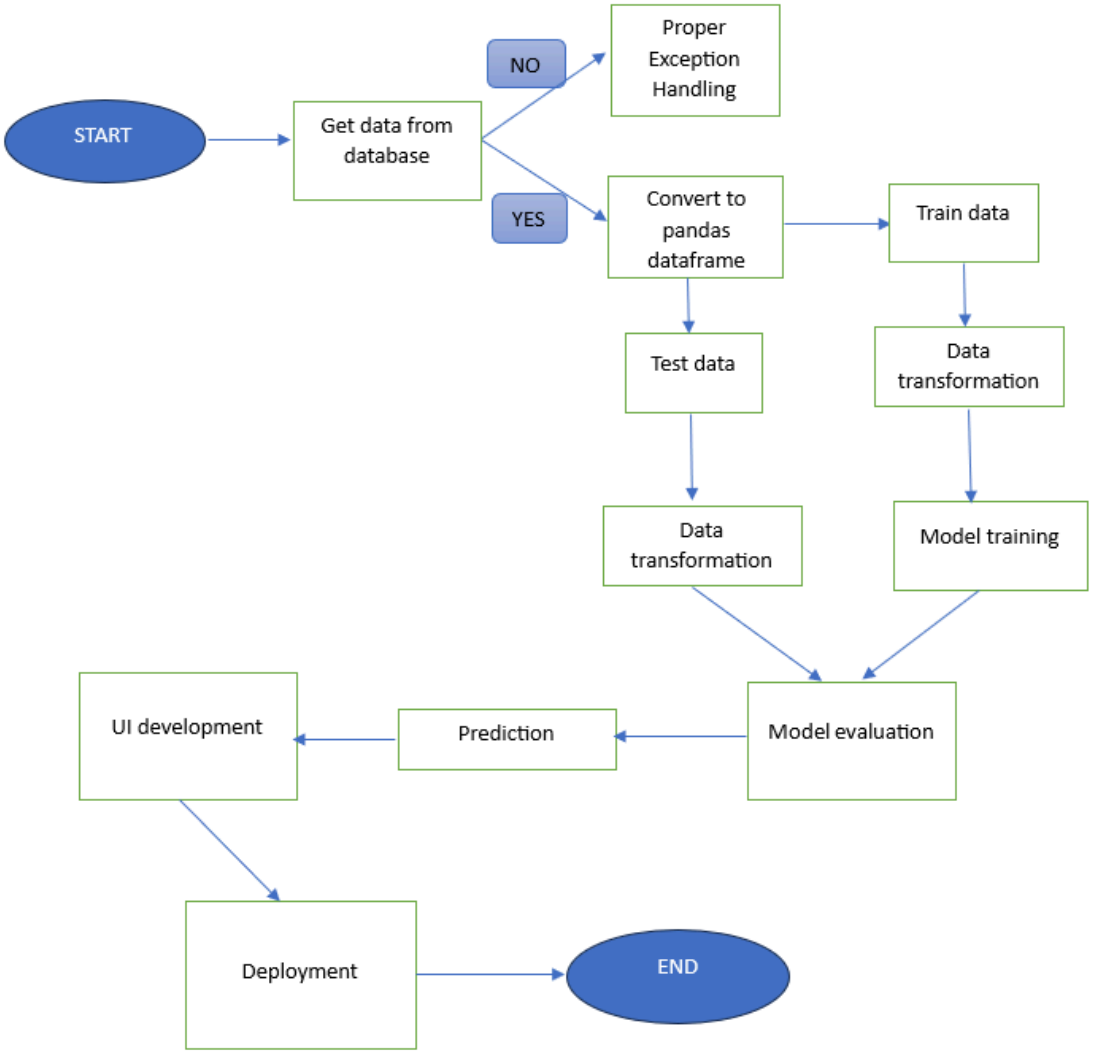
- Detection of upcoming poisonous mushrooms.
- Enhanced understanding of edible mushrooms.
- Improved anticipation of poisonous mushrooms.

Data Description:

- Cap Features:
  - Shape: bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s.
  - Surface: fibrous=f, grooves=g, scaly=y, smooth=s.
  - Color: brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y.
- Bruises and Odor:
  - Bruises: bruises=t, no=f.
  - Odor: almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s.
- Gill Features:
  - Attachment: attached=a, descending=d, free=f, notched=n.
  - Spacing: close=c, crowded=w, distant=d.
  - Size: broad=b, narrow=n.
  - Color: black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y.
- Stalk Features:
  - Shape: enlarging=e, tapering=t.
  - Root: bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=?.
  - Surface (above and below ring): fibrous=f, scaly=y, silky=k, smooth=s.
  - Color (above and below ring): brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y.
  - Veil Color: brown=n, orange=o, white=w, yellow=y.

- Ring Type: cobwebby=c, evanescent=e, flaring=f, large=l, none=n, pendant=p, sheathing=s, zone=z.
- Spore Print Color: black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y.
- Population: abundant=a, clustered=c, numerous=n, scattered=s, several=v, solitary=y.
- Habitat: grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d.

Architecture:



Model Training:

- Perform Exploratory Data Analysis (EDA) for insights into data distribution, outliers, and trends.
- Check for and impute null values.
- Encode categorical values numerically.
- Perform Standard Scalar for value scaling.

Feature Selection:

- Utilize feature selection techniques to retain relevant information, simplifying classification tasks and avoiding overfitting.

Model Selection:

- Evaluate models (Logistic Regression, Random Forest, Decision Tree) using exhaustive Random Search, AUC, and F1 score.
- Assess metrics and generate graphs for model evaluation and importance analysis.
- Identify Logistic Regression as performing well.
- Fit the Logistic Regression model with optimal tuning parameters on the entire dataset.

Prediction:

- Utilize the testing file for validation, data transformation, and insertion.
- Export accumulated data from the database in CSV format for prediction.
- Implement data pre-processing techniques.

Deployment:

- Deploy the application on AWS EC2 Instance.