



MUSHROOM CLASSIFICATION (Machine Learning)

LOW LEVEL DESIGN

INTRODUCTION

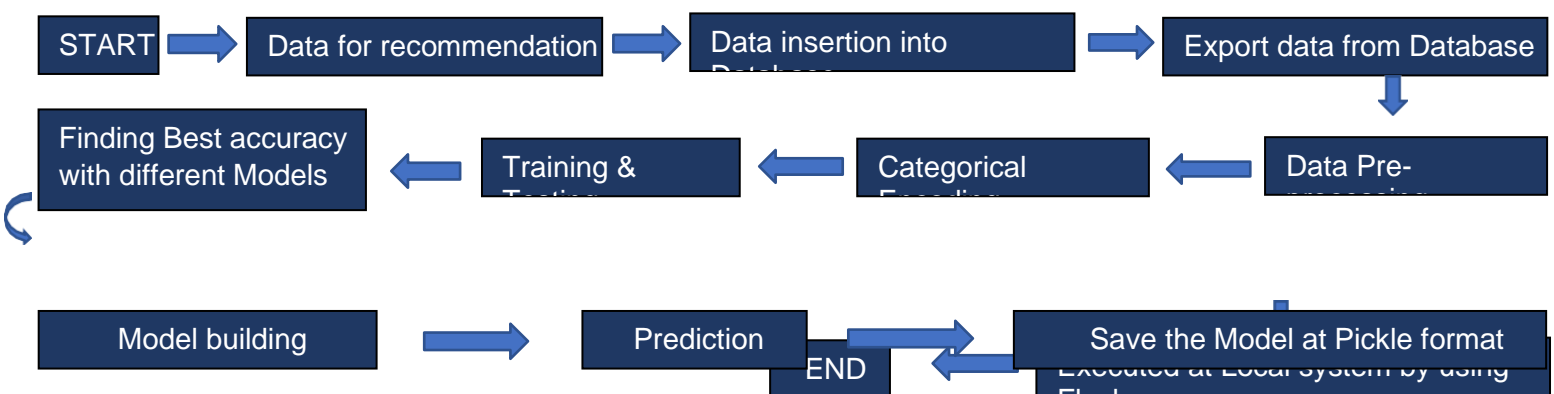
What is Low-Level design document -

The goal of LLD or a low-level design document (LLD) is to give the internal logical design of the actual program code for Mushroom classification. LLD describes the class diagrams with the methods and relation between features and level column. It describes the modules so that the programmer can directly code the program from the document.

SCOPE -

Low-level design (LLD) is a component -level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirements analysis and then refined during data design work.

ARCHITECTURE -



ARCHITECTURE DESCRIPTION -

1. DATA DESCRIPTION-

8124 rows and 23 column were present on the dataset. From Kaggle the dataset were exported it was on the CSV format.

2. DATA INSERTION INTO DATABASE -

a) Database creation and connection-Create a database with name passed. If the database is already created, open the connection to the database.

b) Table creation in the database.

c) Insertion of files in the table.

3. EXPORT DATA FROM DATABASE -

Data export from database-The data in a stored database is exported as a CSV file to be used for data Pre-Processing and model training.

4. DATA PRE-PROCESSING -

Data pre-processing steps we could use are null value handling, stop words removal, punctuation removal, Tokenization, Lemmatization, TFIDF, Imbalanced dataset handling, Handling columns with standard deviation zero or below a threshold, etc.

5. ML ALGORITHM -

All the ML algorithm is used to do classification and found the best model from that.

6. CATEGORICAL ENCODING -

All the dataset available on dataset was not on numerical, so that has been converted on numerical terms, which will be easy to do model building.

7. TRAINING AND TESTING DATASET -

As here 80 % of dataset has been trained and 20% of dataset has been tested.

8. FINDING ACCURACY WITH DIFFERENT MODEL -

All the supervised machine learning algorithm were used to classify the output such as Logistic regression, K Neighbours, SVC, Decision tree, Random forest, Gradient boosting classifier, etc. found accuracy with every models.

9. MODEL BUILDING -

After checking accuracy with different model, model building was created with the best accuracy and saved the model in pickle format.

10. WEB FRAMEWORK -

By using flask API on the local system it been tested.

Conclusion -

This is a web-based application. We have used Flask for the user interface. We can put all the feature of the mushroom as a input and can check whether the mushroom is poisonous or edible.