

# MUSHROOM CLASSIFICATION

-- Sachin Shinde & Riyaz Khorasi

# Introduction

In recent years, the popularity of mushrooms as a super food and the understanding of their vast health benefits has surged.

What used to be typically seen as just a traditional food, mushrooms are now being widely consumed and acknowledged for their healing and health abilities. There are thousands of species of Mushrooms in the world; they are edible and non-edible being poisonous. It is difficult for non-expertise people to Identify poisonous and edible mushrooms of all species manually. So, a computer aided system with software or algorithm is required to classify poisonous and nonpoisonous mushrooms.

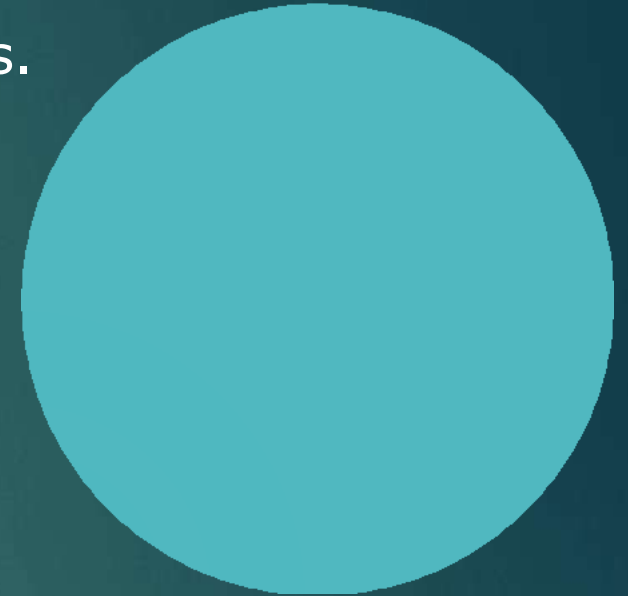
This project is presented on classification of poisonous and nonpoisonous mushrooms. Most of the research works to classify the type of mushroom have applied, machine learning techniques like Naïve Bayes, K-Neural Network, Support vector Machine (SVM), Decision Tree techniques.



# Objective

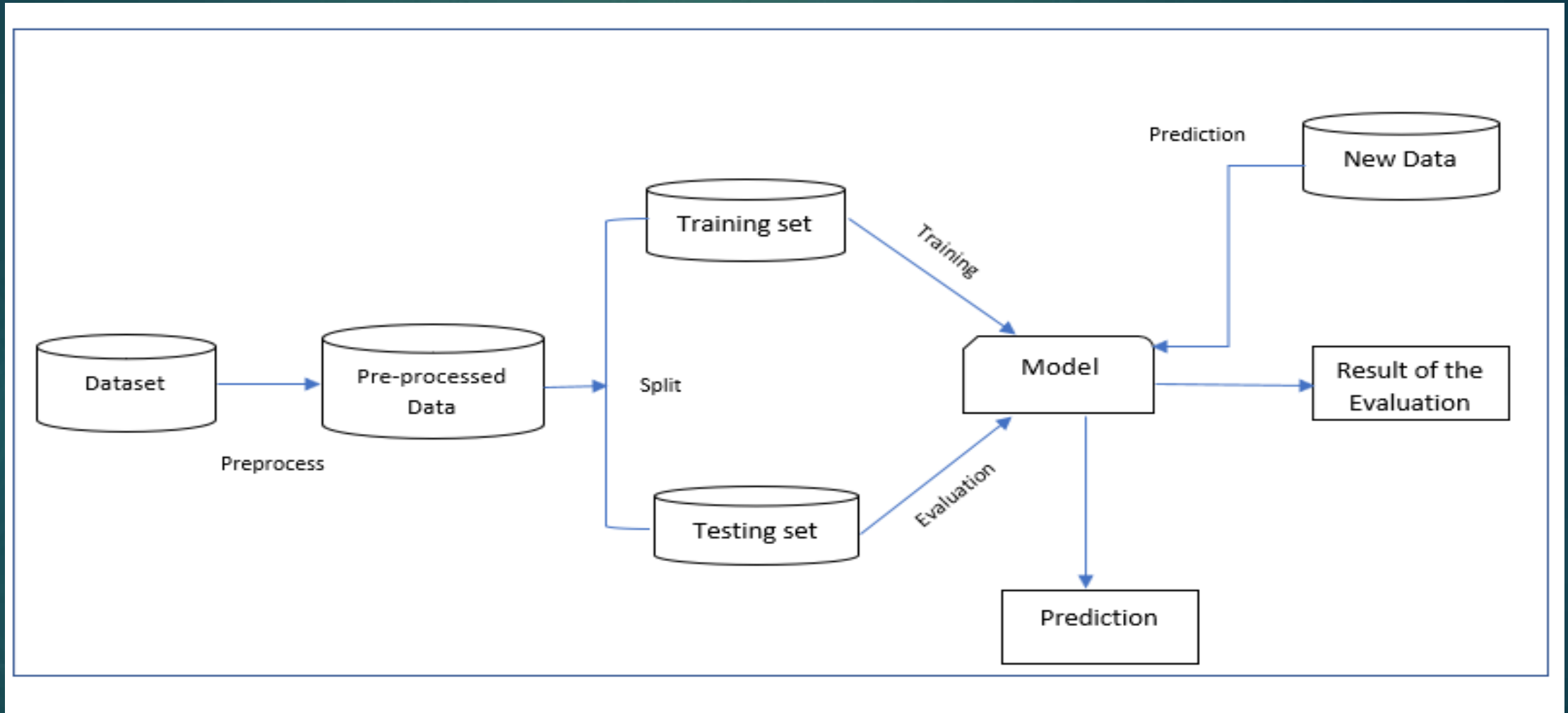
The Mushroom Classification is a machine learning based model which will help us to predict the Mushroom are edible or poisonous.

The main goal of this project is to perform exploratory data analysis and later predict the Mushrooms are edible or poisonous.



# Architecture

4



# Dataset

5



# Data Analysis

6



## DATA COLLECTION

In step 1, we collect data which is generally present in a database or on internet.



## DATA PREPROCESSING

In step 2, we preprocess the data which involves data cleaning by handling outliers, null values etc.



## EXPLORATORY DATA ANALYSIS

In step 3, we explore the data by performing univariate and bivariate analysis on the features.



## FEATURE SELECTION

In step 4, we use feature selection techniques to filter out the most important features to perform model creation



## MODEL CREATION AND EVALUATION

In step 5, we finally build models on our dataset and choose the model which gives the best accuracy.



# Random Forest Model

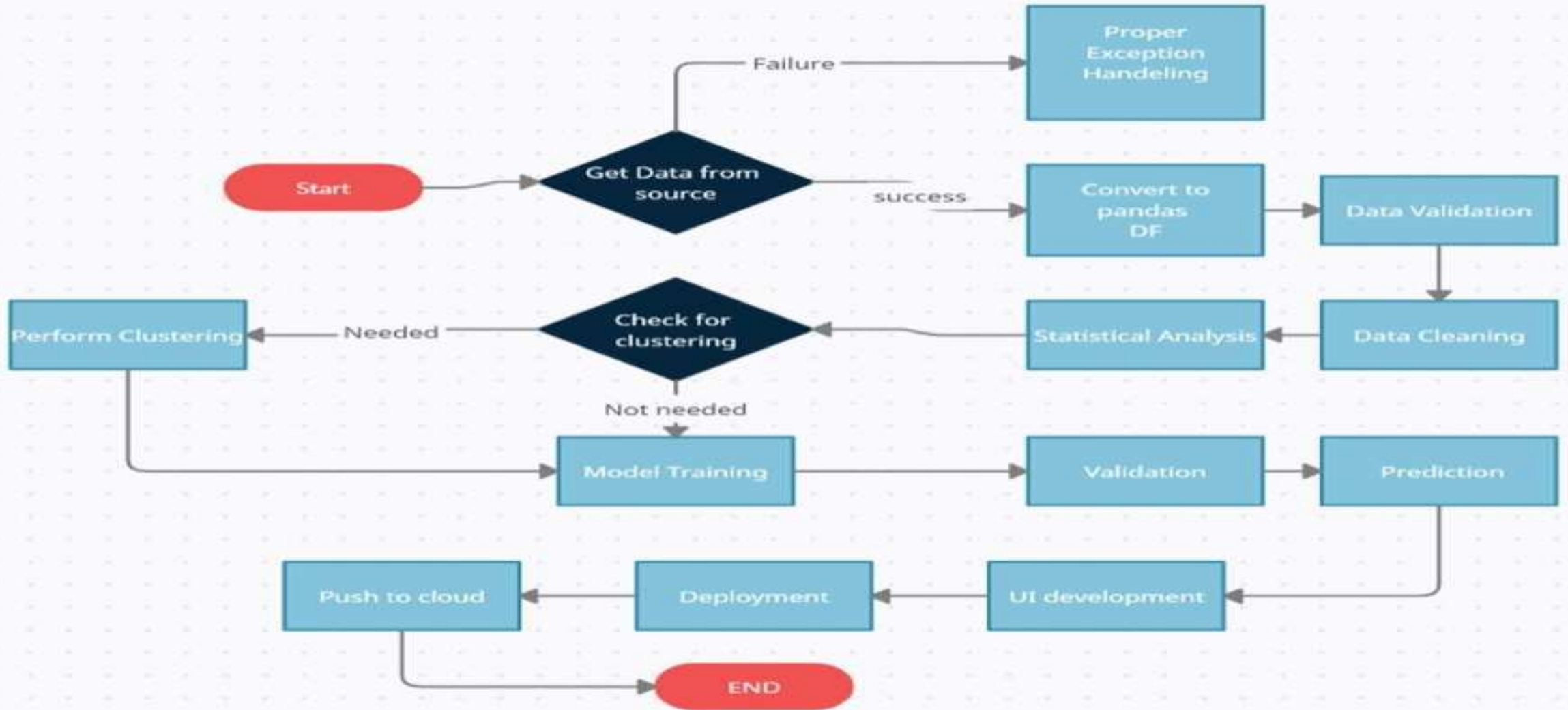
## INTRODUCTION

- ▶ The random forest classifier is a supervised learning algorithm which we can use for regression and classification problems. It is among the most popular machine learning algorithms due to its high flexibility and ease of implementation.
- ▶ It is called Random Forest because it consists of multiple decision trees just as a forest has many trees. On top of that, it uses randomness to enhance its accuracy and combat overfitting, which can be a huge issue for such a sophisticated algorithm. These algorithms make decision trees based on a random selection of data samples and get predictions from every tree. After that, they select the best viable solution through votes.
- ▶ Random Forest Classifier being ensembled algorithm tends to give more accurate result. This is because it works on the principle i.e. number of weak estimators when combined forms strong estimator. Even if one or few decision trees are prone to noise, overall results would tend to be correct.

It gives us high accuracy as 99%.

# MODEL TRAINING AND VALIDATION WORKFLOW

8





# MODEL TRAINING & VALIDATION WORKFLOW

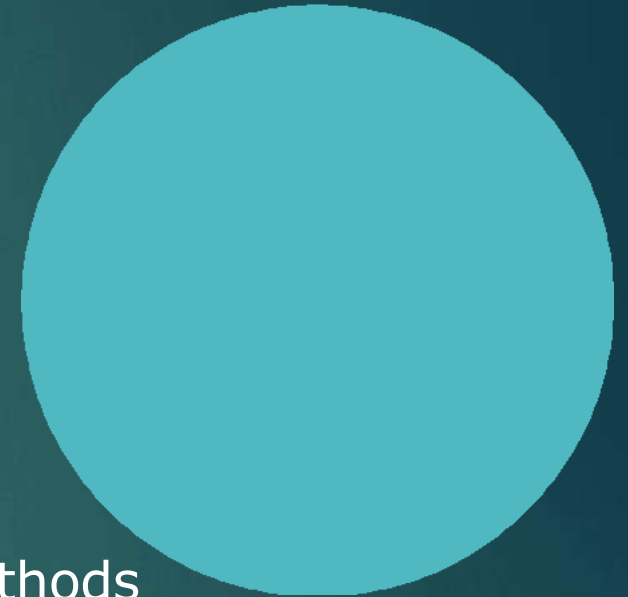
9

## Data Collection

- ▶ Mushroom Classification Data Set from Kaggle.

## Data Pre-processing

- ▶ Missing values handling by Simple imputation (mean, median and mode strategy)
- ▶ Outliers' detection and removal by boxplot and percentile methods
- ▶ Categorical features handling by ordinal encoding and label encoding
- ▶ Feature scaling done by Standard Scalar method

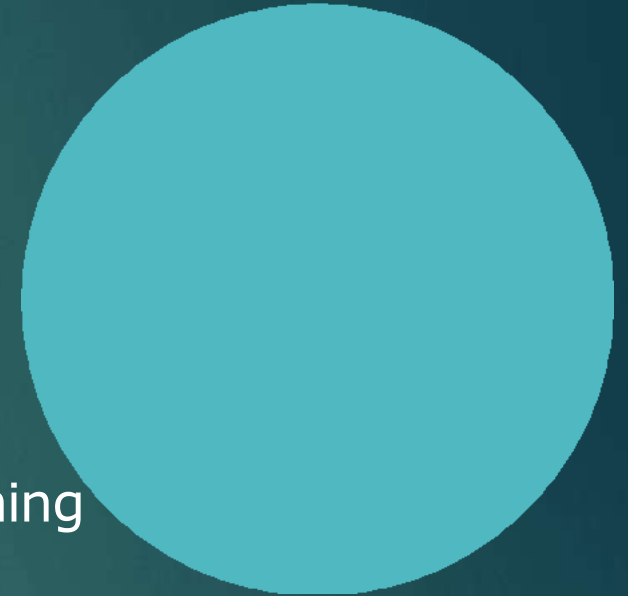


# MODEL TRAINING & VALIDATION WORKFLOW

10

## Model Creation and Evaluation

- ▶ Various classification algorithms like Logistic Regression, Random Forest, Decision Tree tested.
- ▶ Random Forest, Decision Tree and Logistic regression were given better results. Random Forest was chosen for the final model training and testing.
- ▶ Hyper parameter tuning was performed.
- ▶ Model performance evaluated based on accuracy, confusion matrix, classification report.



# MODEL DEPLOYMENT

## Model Deployment

- ▶ The final model is deployed using on Heroku using Flask framework



# Thank You

