

MUSHROOM CLASSIFICATION AND MODEL COMPARISON

• • •

Presented by
Makthala Sai Teja Goud





UCI MACHINE LEARNING AND 1 COLLABORATOR · UPDATED 8 YEARS AGO

▲ 2417

New Notebook

Download



Mushroom Classification

Safe to eat or deadly poison?



DATASET

The dataset was sourced from Kaggle, has a size of 374 KB, contains 23 columns, and over 8000 rows of data

mushrooms.csv (374 kB)

Detail Compact Column 10 of 23 columns ▾

About this file + Add Suggestion

Attribute Information: (classes: edible=e, poisonous=p)

- cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
- cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
- cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r,pink=p,purple=u,red=e,white=w,yellow=y
- 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-attachment: attached=a,descending=d,free=f,notched=n
- gill-spacing: close=c,crowded=w,distant=d



**Lets Check out the
Dataset**



Supervised Machine Learning Algorithm

01

K-Nearest Neighbors
(KNN)

02

Decision Tree

03

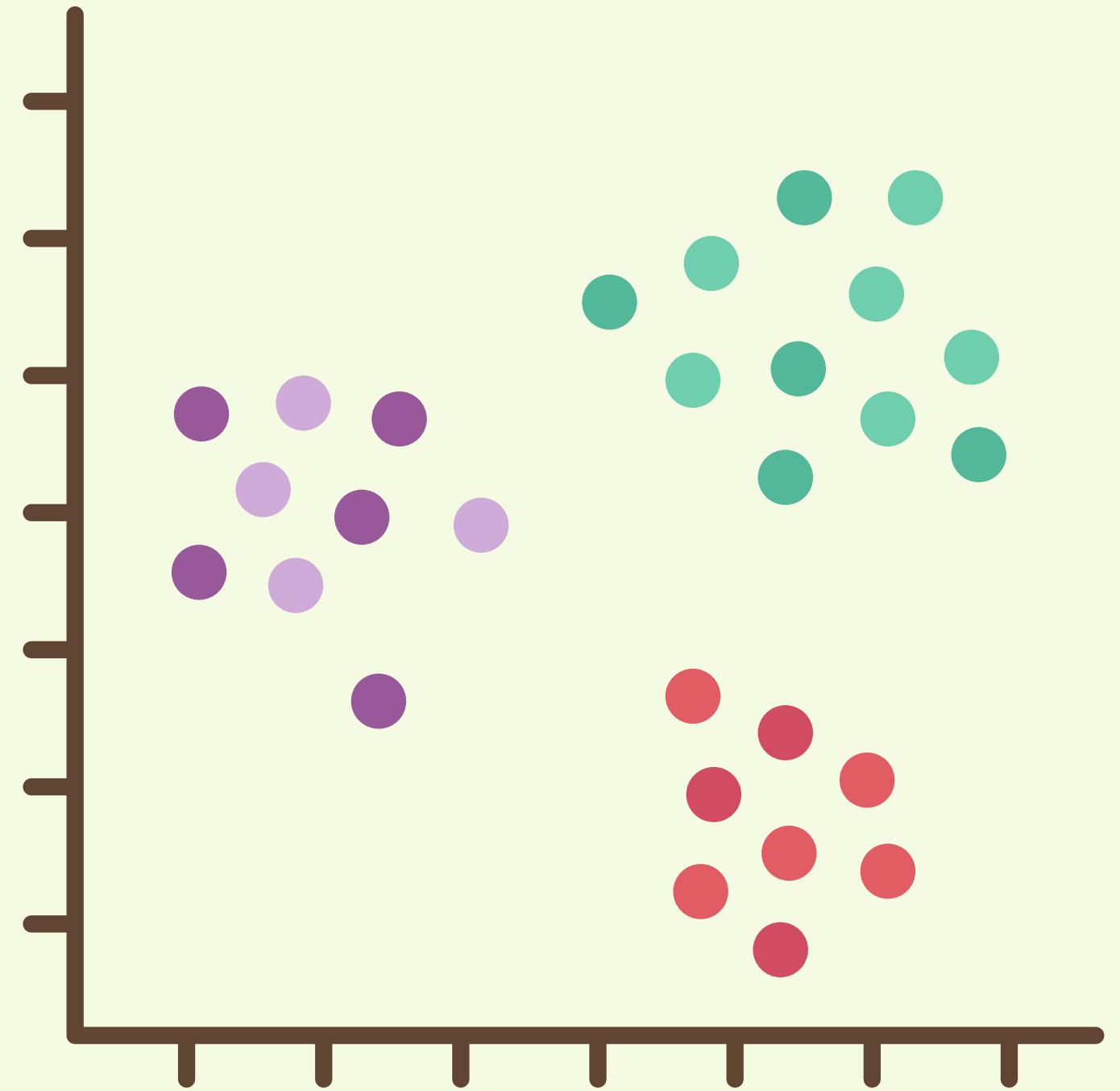
Random Forest



Un-Supervised Machine Learning Algorithm

04

K Means Clustering

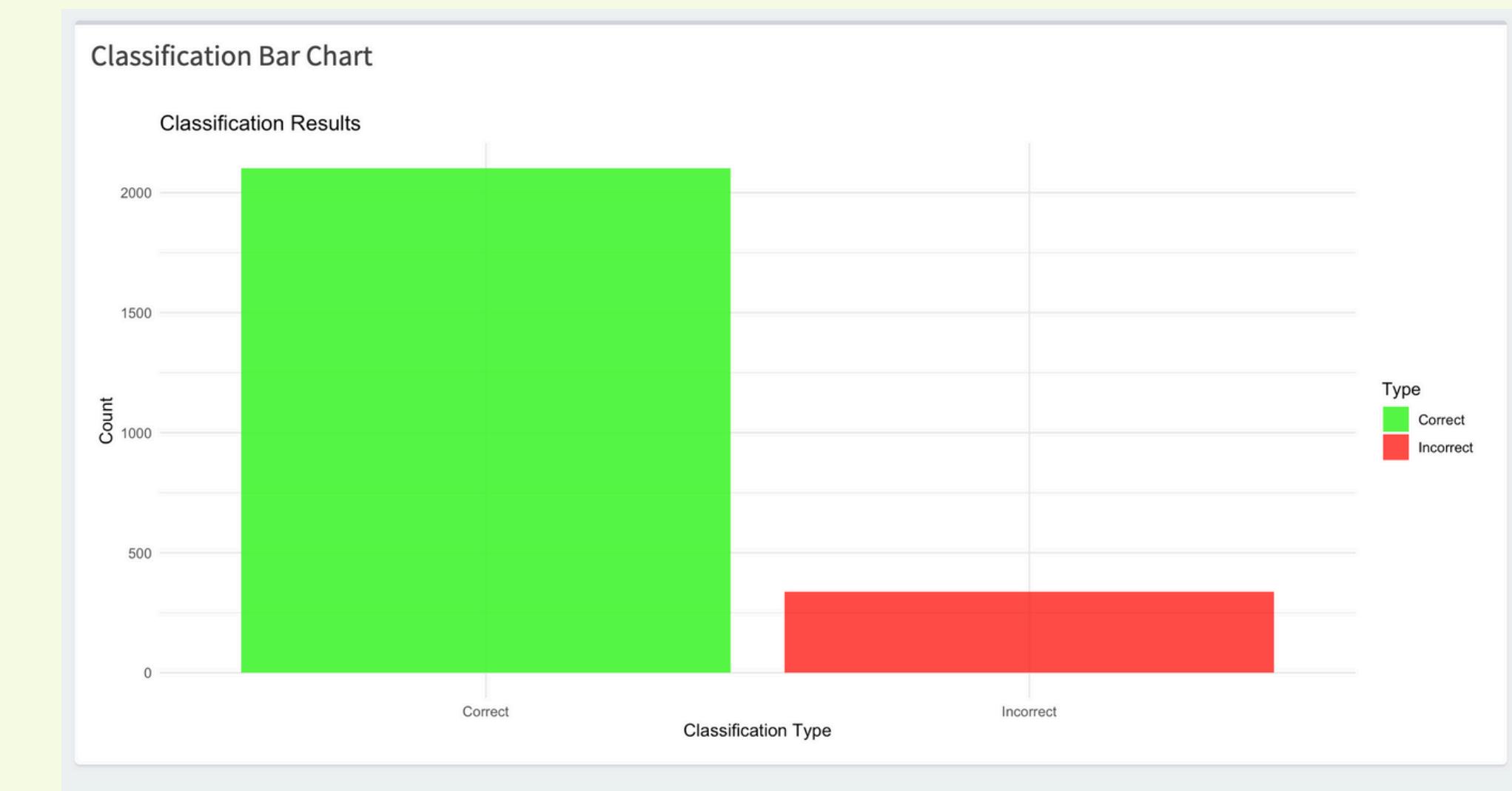


K-Nearest Neighbors (KNN)

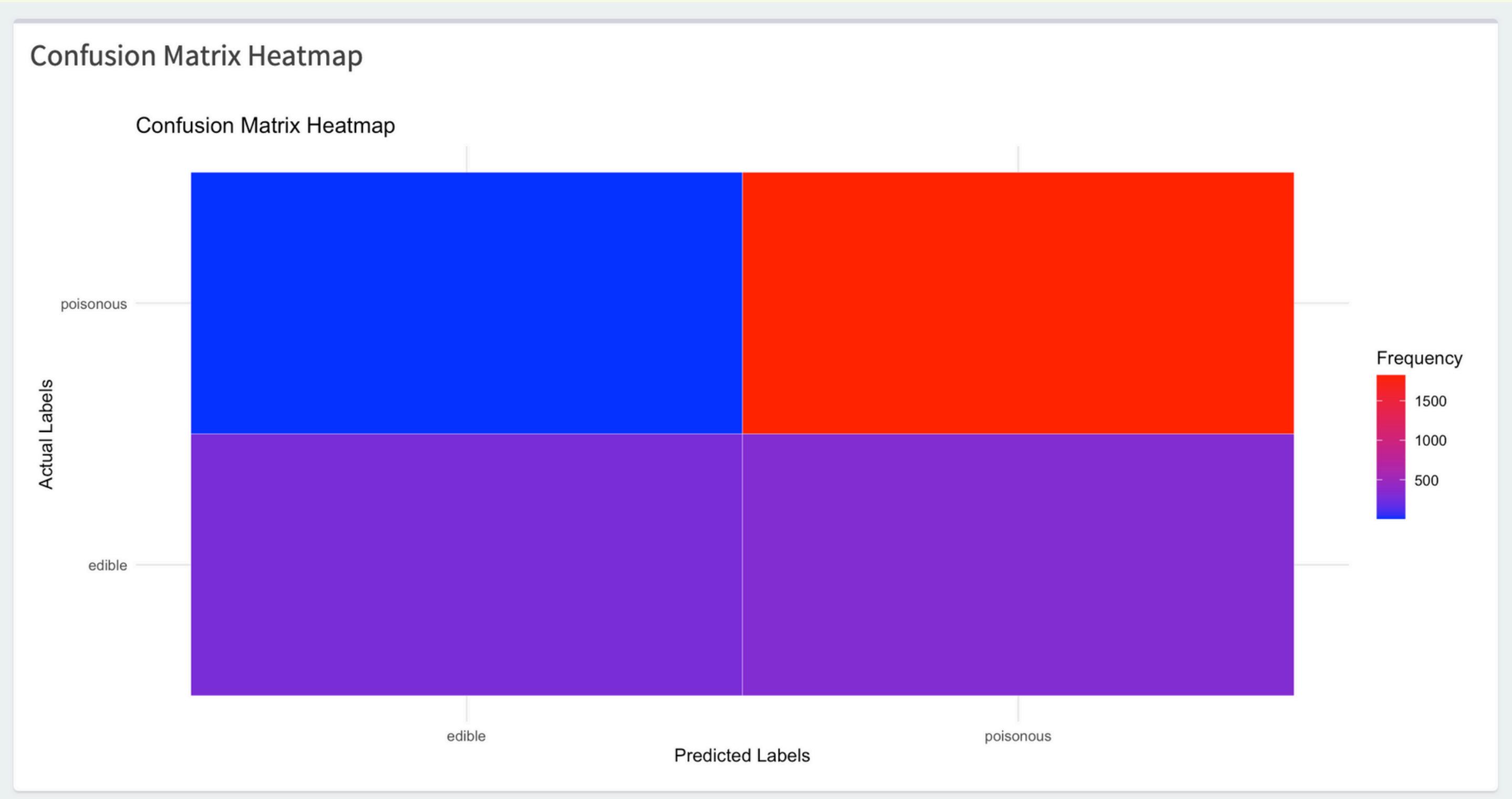


K-Nearest Neighbors (KNN) is a supervised machine learning algorithm used for classification and regression. It predicts outcomes based on the average value of the nearest data points, determined by a distance metric like Euclidean distance

KNN Results Table	
edible	poisonous
277	333
4	1824

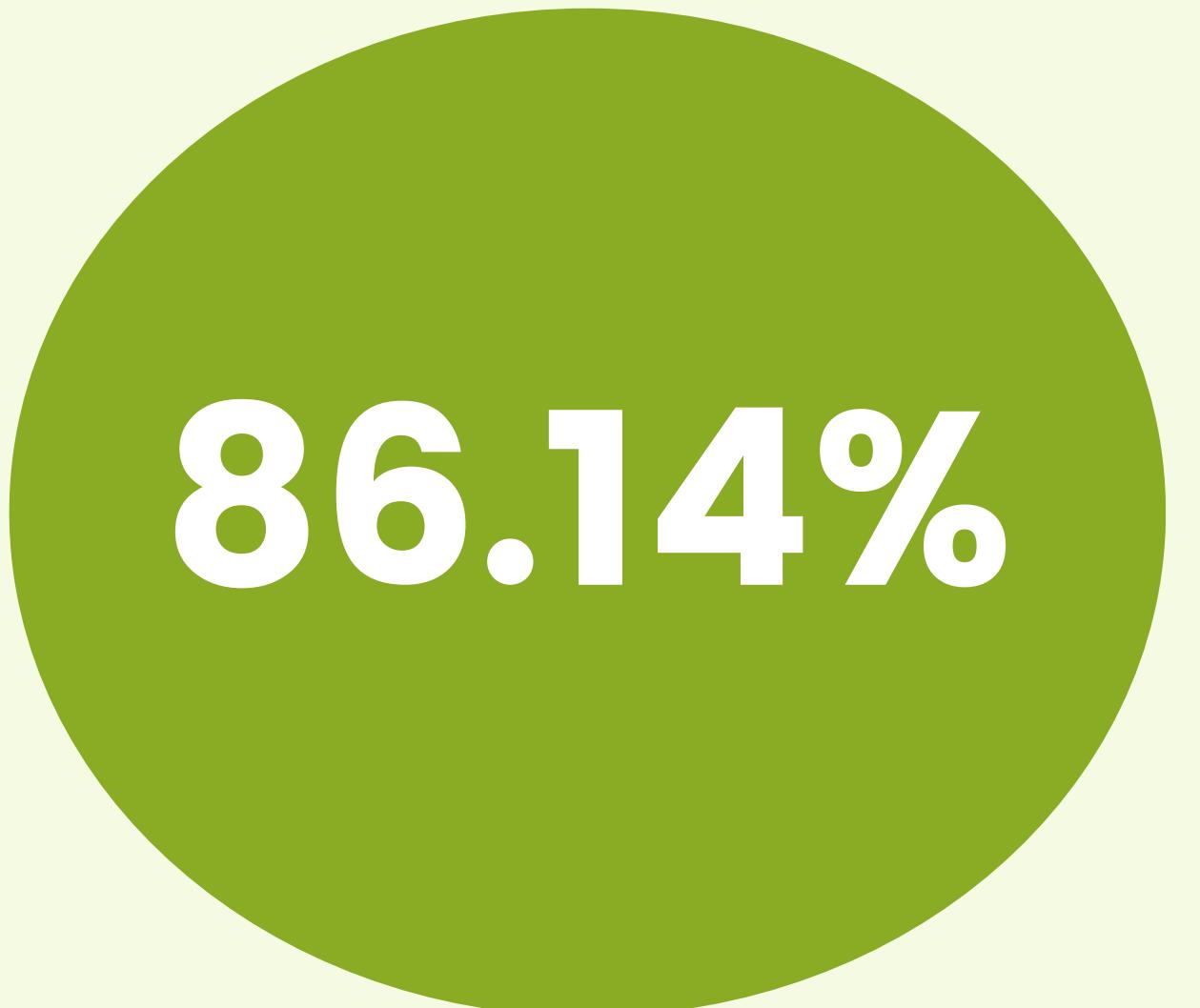


Confusion Matrix Heatmap





Accuracy of the Model

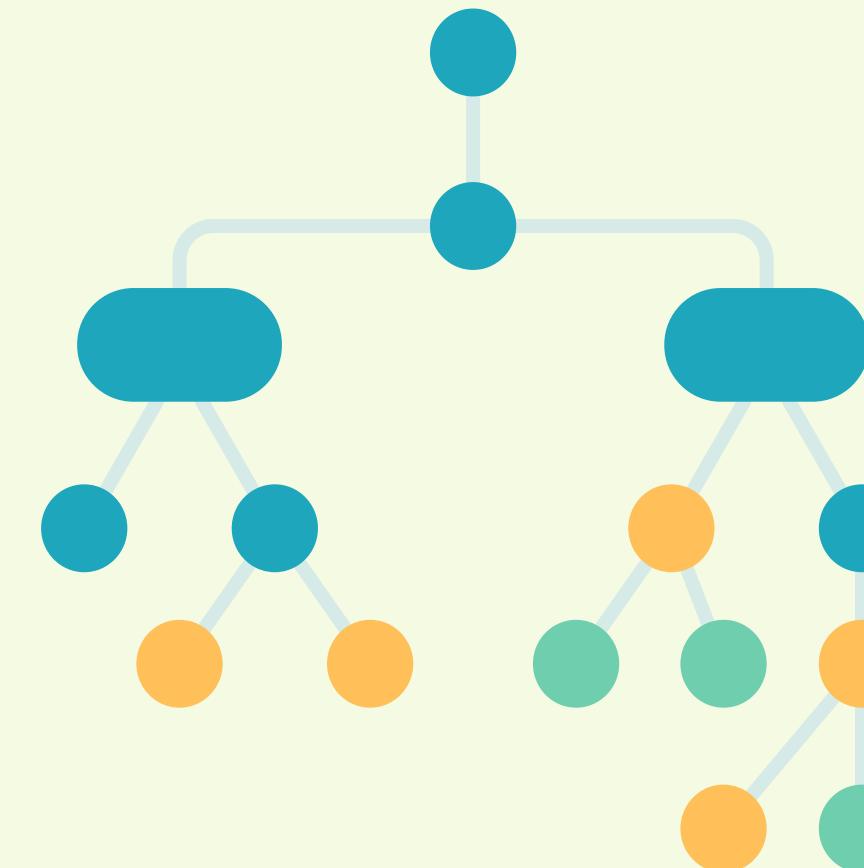


Decision Tree

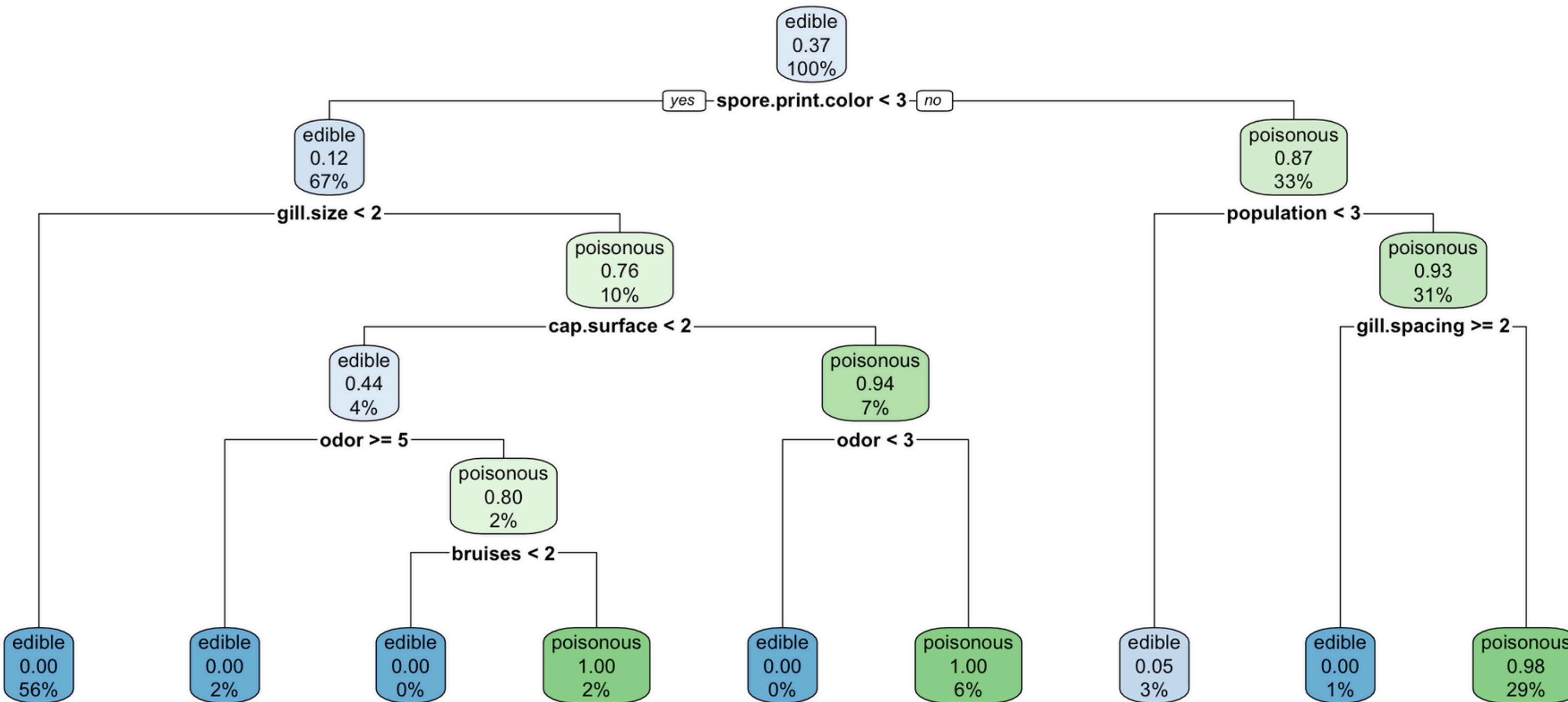


A Decision Tree is a supervised algorithm for classification and regression, using a tree-like structure to split data based on features, with nodes representing decisions and leaves providing outputs.

Confusion Matrix	
edible	poisonous
473	137
45	1783



Decision Tree Plot





Accuracy of the Model

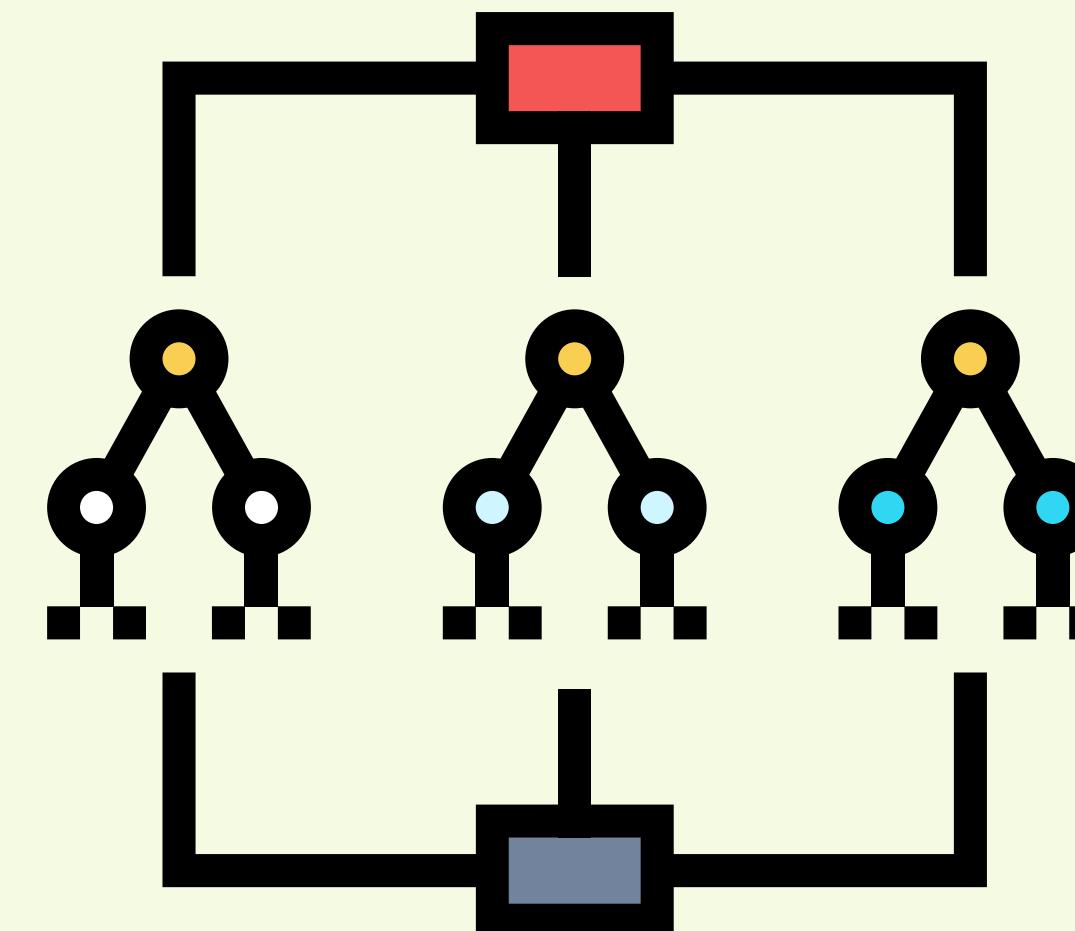


Random Forest



Random Forest is an supervised learning algorithm that builds multiple decision trees and combines their predictions through voting (for classification) or averaging (for regression) to improve accuracy and reduce overfitting.

Prediction	Reference	Freq
edible	edible	466
poisonous	edible	144
edible	poisonous	0
poisonous	poisonous	1828





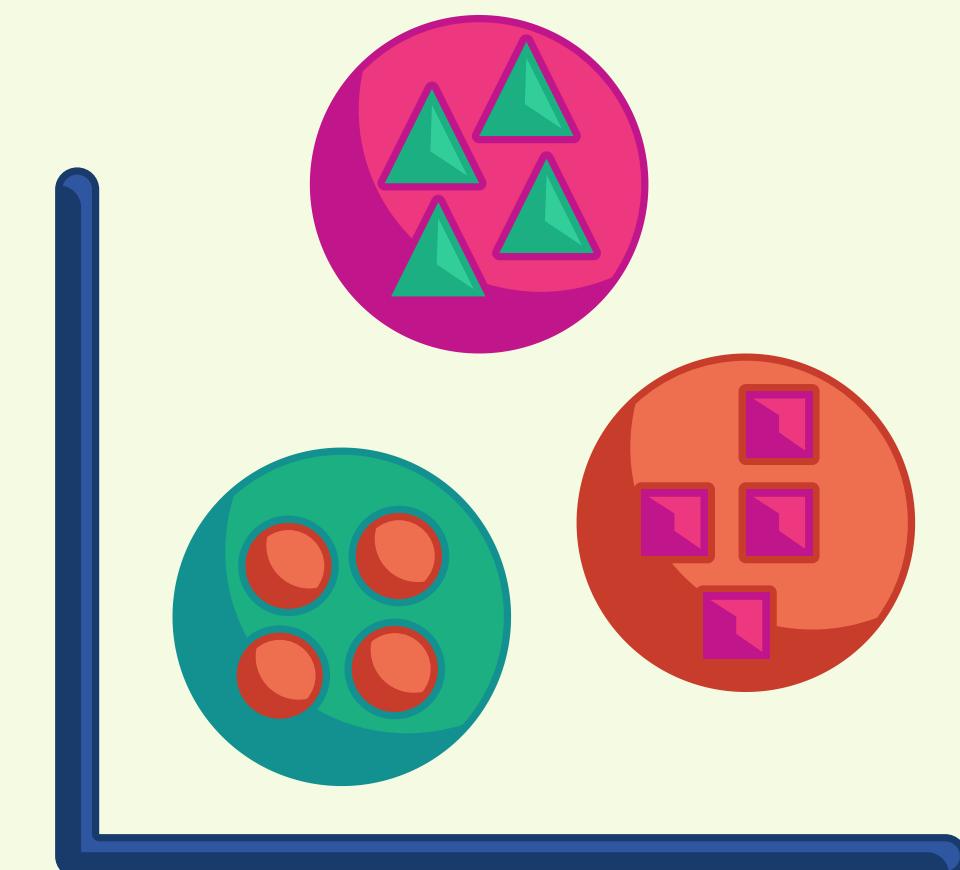
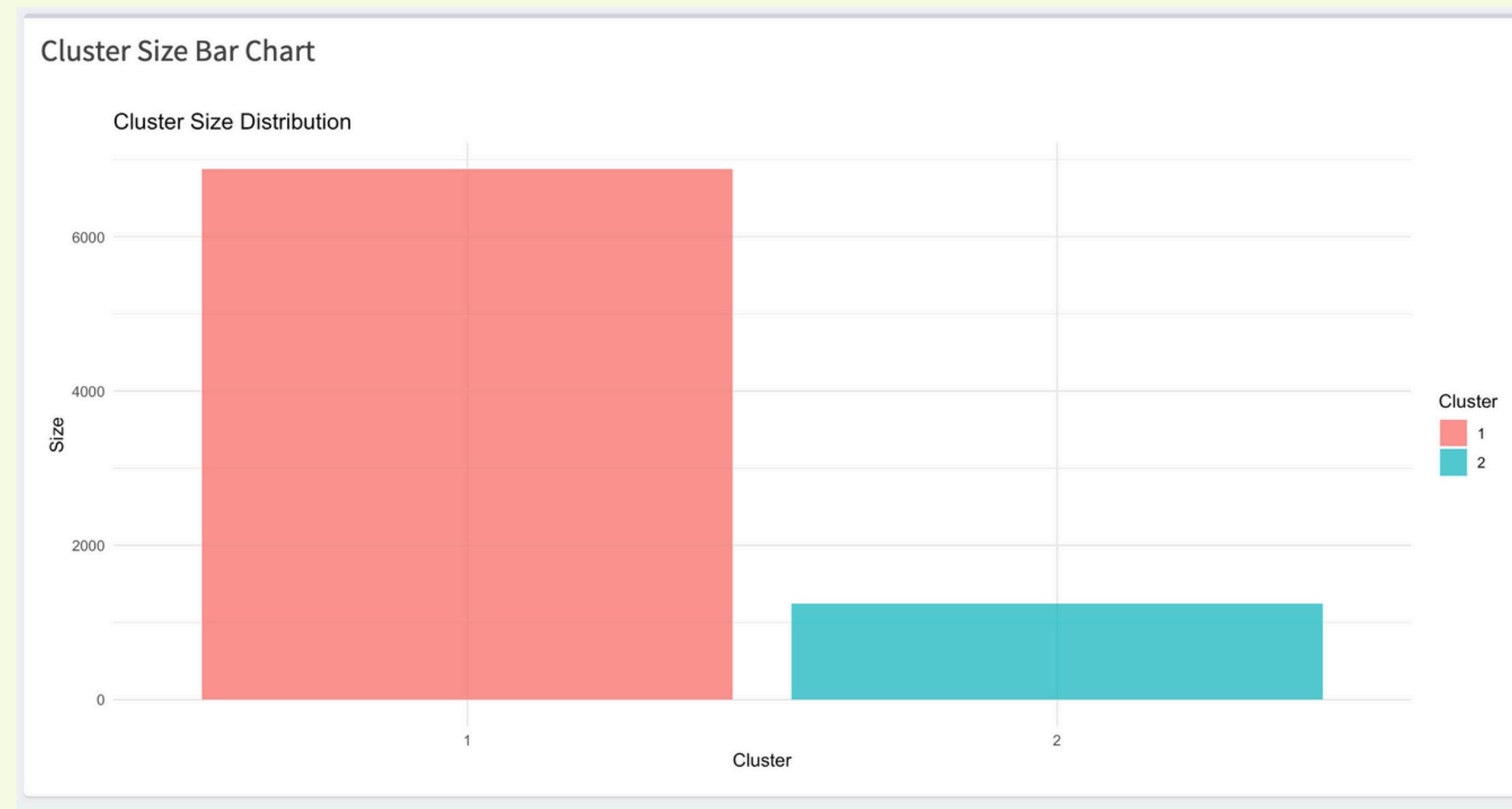
Accuracy of the Model



K Means Clustering

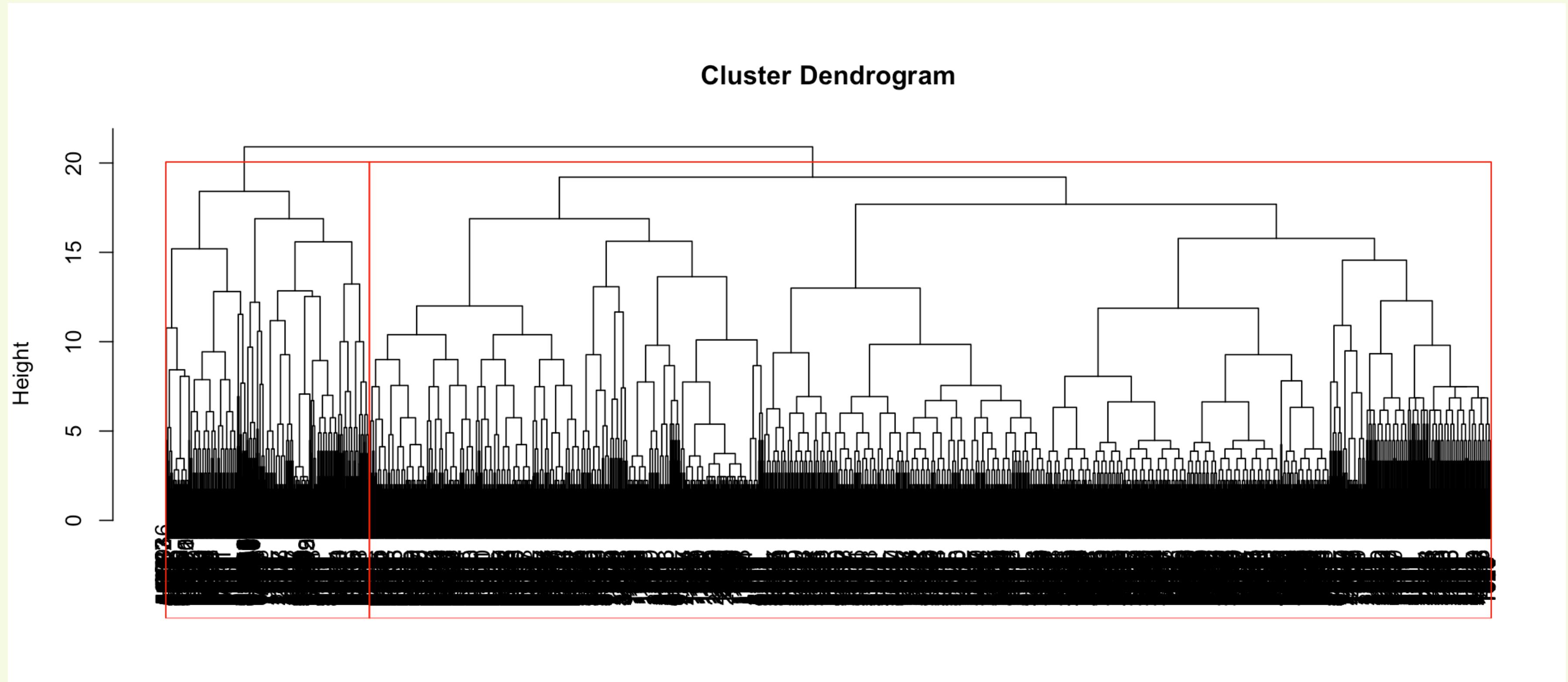


K-Means Clustering is an unsupervised machine learning algorithm used for partitioning a dataset into K distinct clusters. It assigns data points to clusters based on their similarity, measured by distance (usually Euclidean). The algorithm iteratively refines cluster centroids by minimizing the within-cluster variance.





Cluster Dendrogram

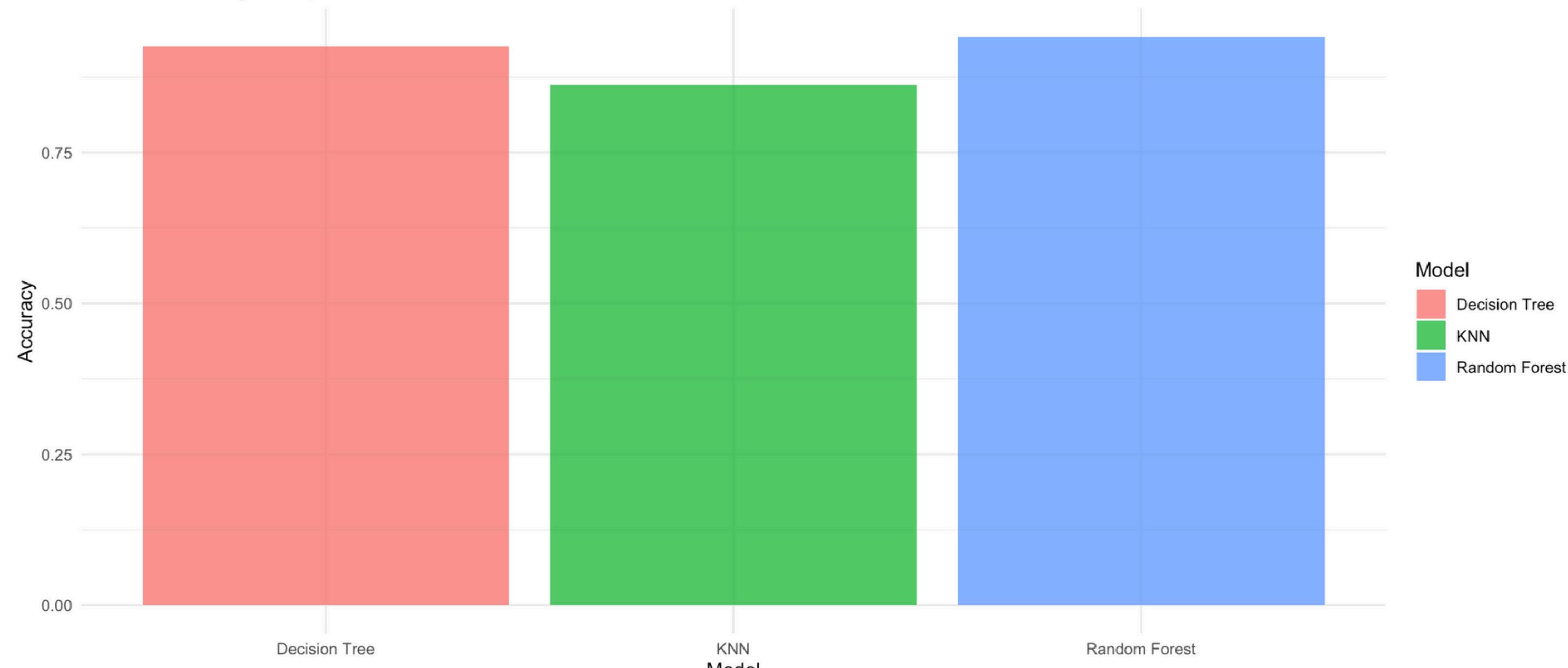


Now The Model Accuracy Comparision Time



Model Accuracy Comparison

Model Accuracy Comparison





And Finally its
Dashboards Turn

out
this
check

http://127.0.0.1:6139 | Open in Browser | C Publish ▾

Mushroom Analysis

- KNN Results
- Decision Tree
- K-Means Clustering
- Random Forest Model
- Model Comparison

Decision Tree Plot

A decision tree diagram for mushroom classification. The root node splits on 'spore.print.color < 3'. The 'yes' branch leads to a node 'gill.size < 2', which further splits on 'cap.surface < 2'. The 'no' branch leads to a node 'population < 3', which splits on 'gill.spacing >= 2'. Leaf nodes represent class proportions: 'edible' (0.37, 100%), 'poisonous' (0.87, 33%), 'edible' (0.12, 67%), 'poisonous' (0.76, 10%), 'edible' (0.44, 4%), 'poisonous' (0.94, 7%), 'edible' (0.00, 56%), 'edible' (0.00, 2%), 'poisonous' (0.80, 2%), 'edible' (0.00, 0%), 'poisonous' (1.00, 2%), 'edible' (0.00, 0%), 'poisonous' (1.00, 6%), 'edible' (0.05, 3%), 'edible' (0.00, 1%), and 'poisonous' (0.98, 29%).

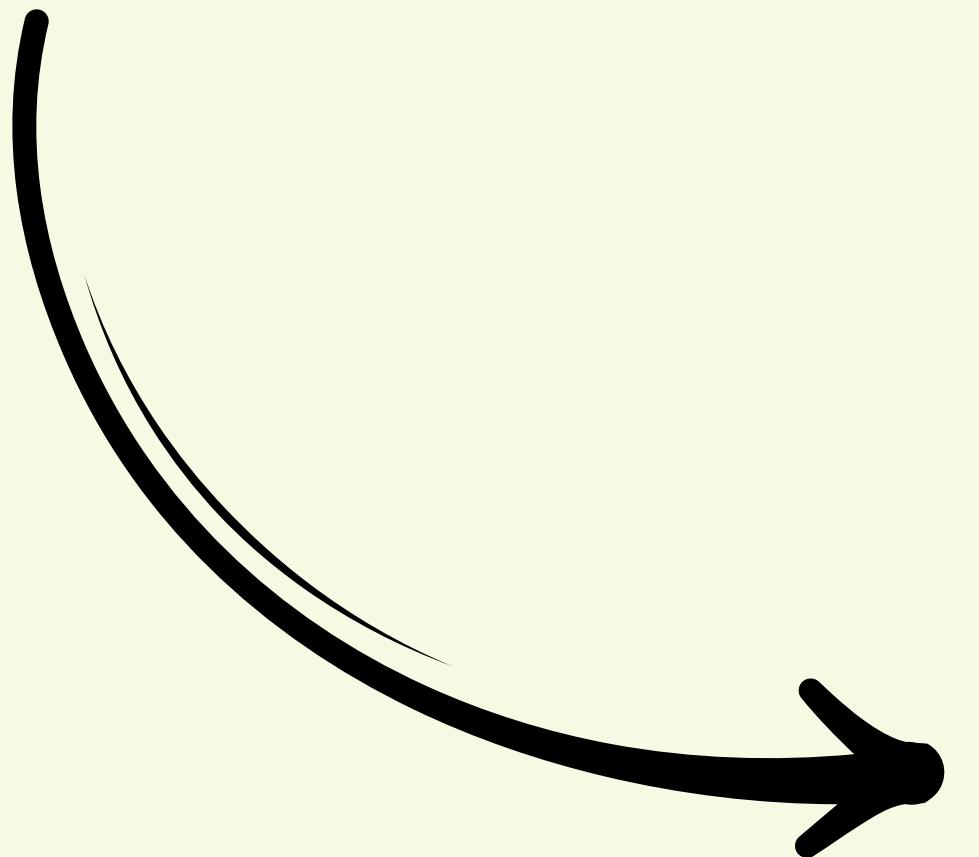
```
graph TD; Root[spore.print.color < 3] -- yes --> Node1[gill.size < 2]; Root -- no --> Node2[population < 3]; Node1 -- cap.surface < 2 --> Node3[odor >= 5]; Node1 -- cap.surface > 2 --> Node4[odor < 3]; Node3 -- bruises < 2 --> Node5[edible 0.00 56%]; Node3 -- bruises >= 2 --> Node6[poisonous 0.80 2%]; Node4 -- gill.spacing >= 2 --> Node7[edible 0.00 1%]; Node4 -- gill.spacing < 2 --> Node8[poisonous 0.98 29%]; Node2 -- gill.spacing >= 2 --> Node9[poisonous 0.93 31%]; Node2 -- gill.spacing < 2 --> Node10[edible 0.05 3%]; Node2 -- gill.spacing < 1 --> Node11[edible 0.00 0%]; Node2 -- gill.spacing < 0.5 --> Node12[poisonous 0.76 10%]; Node2 -- gill.spacing < 0.2 --> Node13[edible 0.44 4%]; Node2 -- gill.spacing < 0.1 --> Node14[poisonous 0.94 7%]; Node2 -- gill.spacing < 0.05 --> Node15[edible 0.37 100%];
```

Confusion Matrix

edible	poisonous
473	137
45	1783

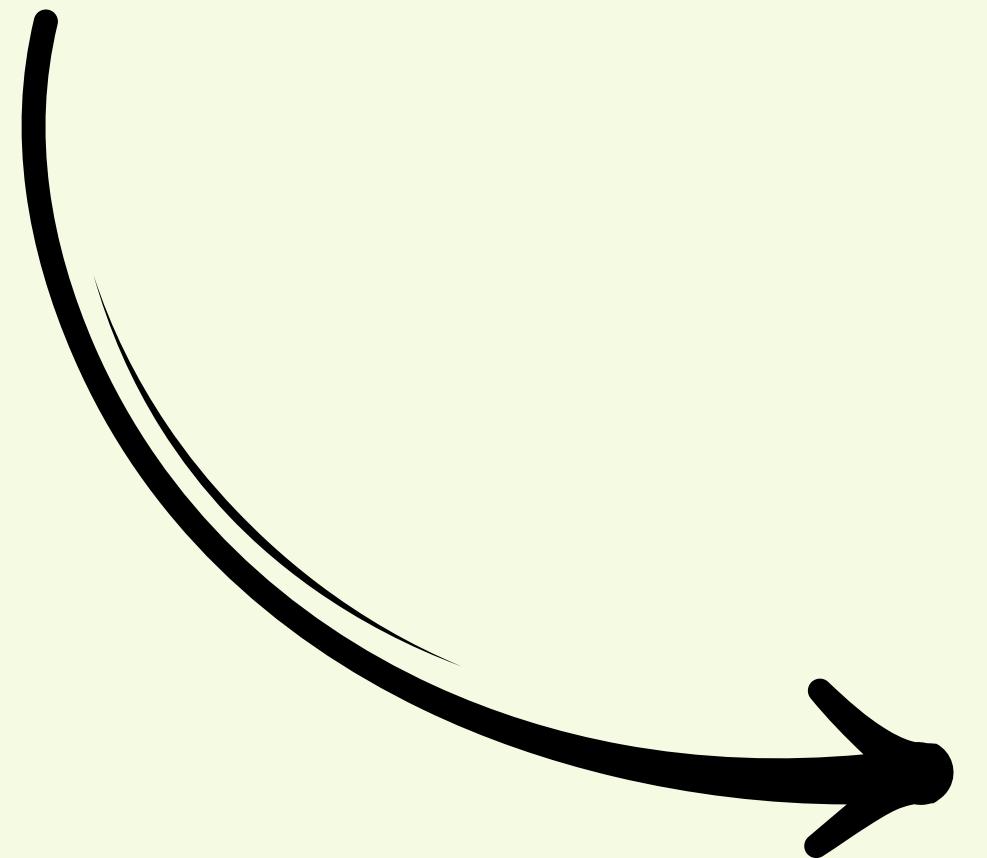


Scan here to get the
Code and dataset





**Don't forget to connect
on LinkedIn as well**



THANK YOU

