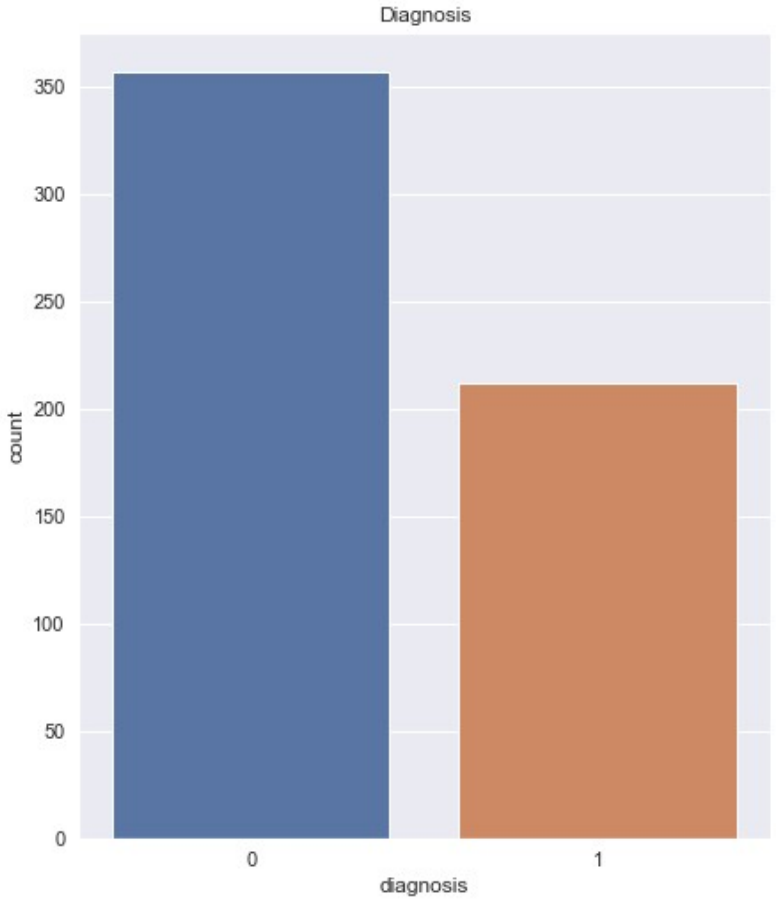
Capstone 3 Project:
Breast Cancel Prediction

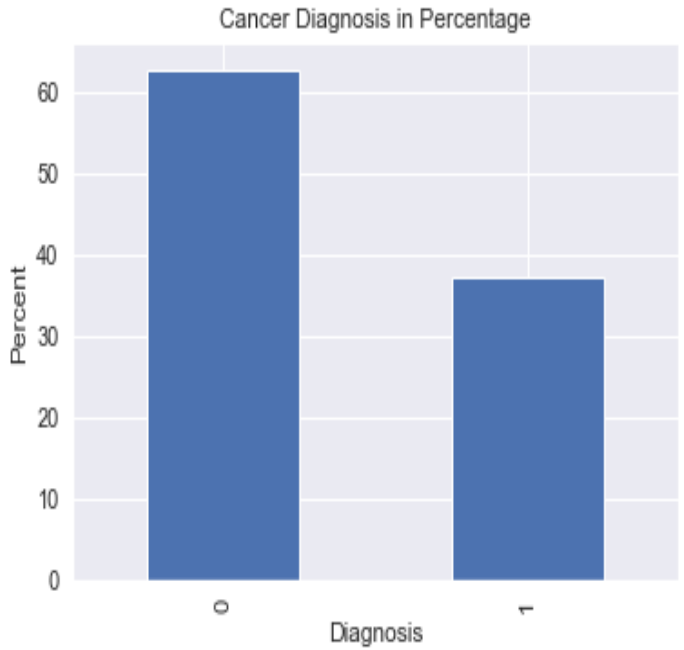
Dev Joshi[1]
(1) Spring Board Bootcamp

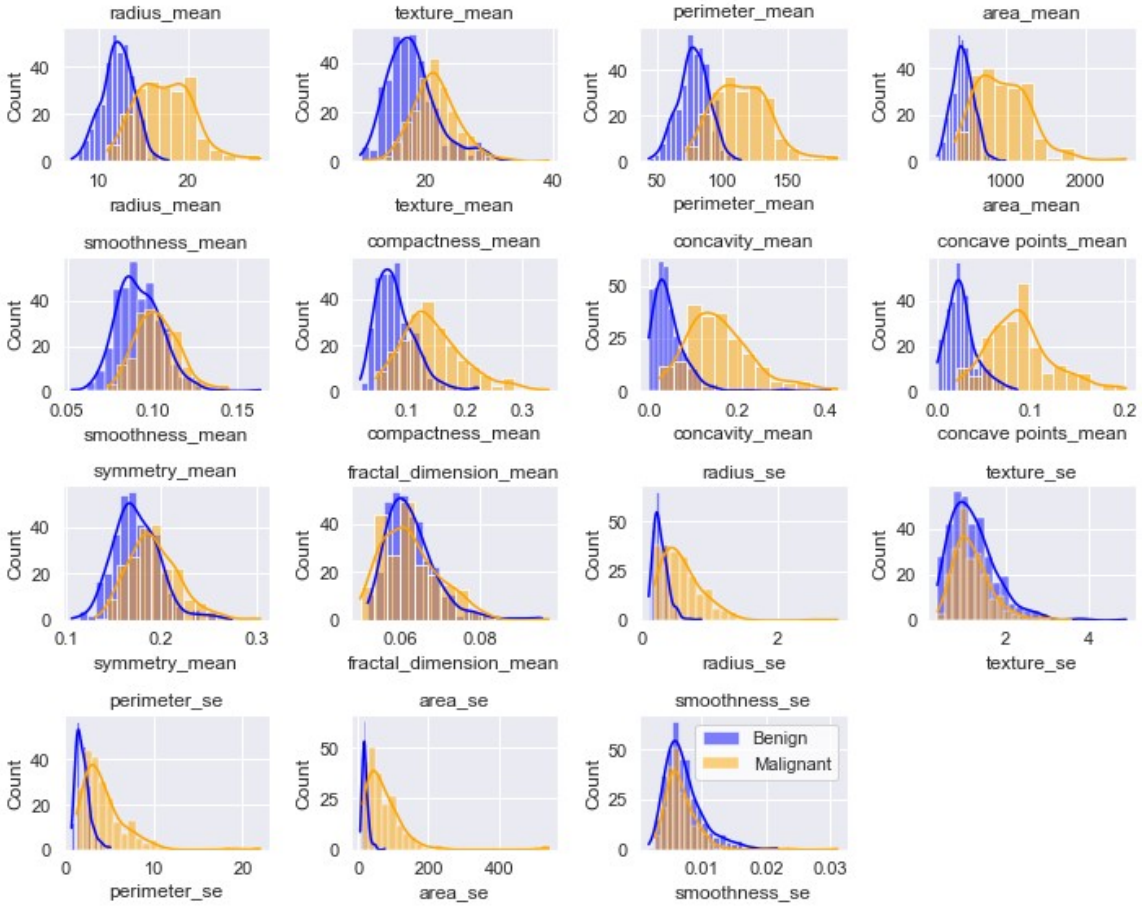
Problem Statement

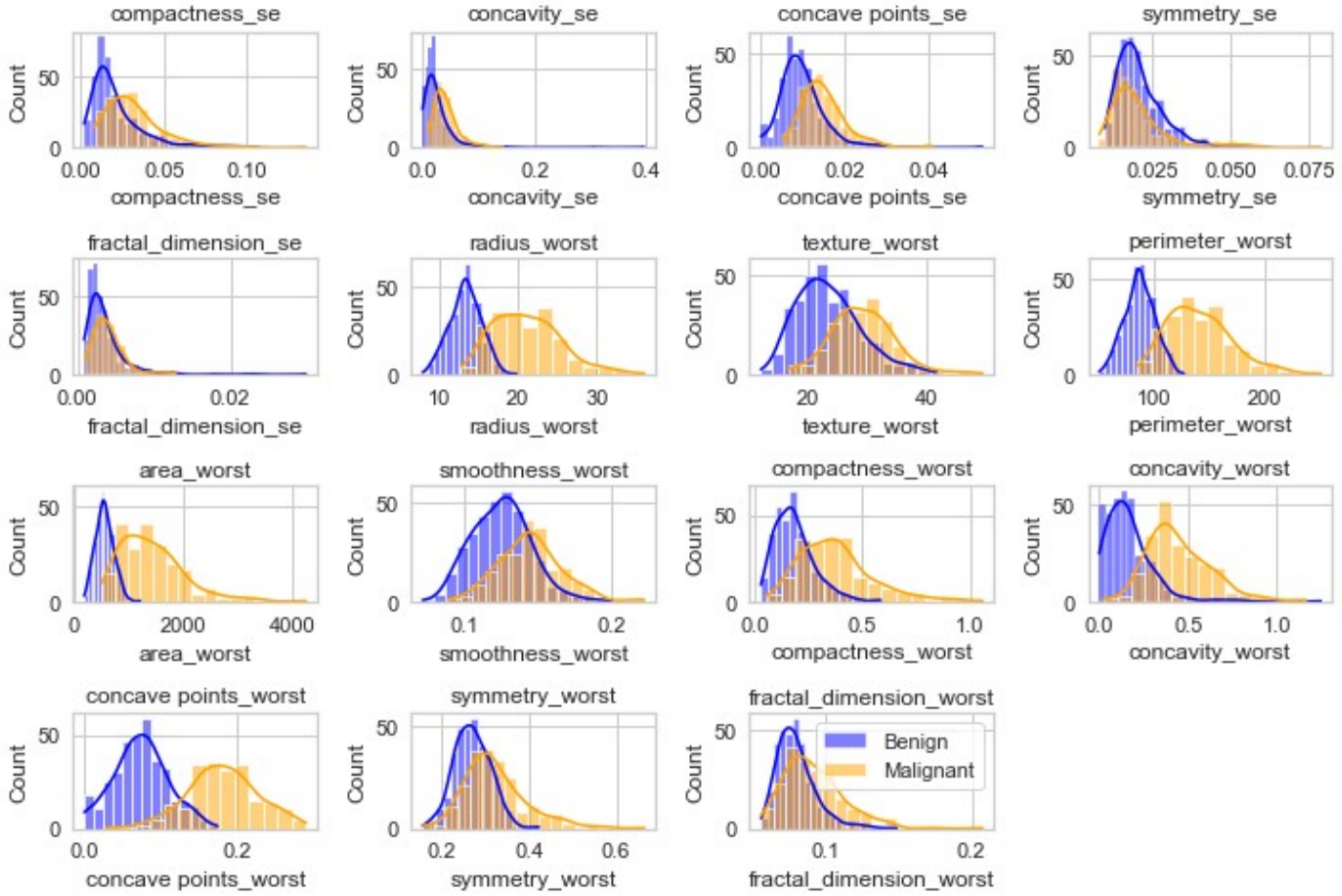
The dataset consisting of digitized imaging of fine needle aspirate (FNA) of a breast tumor cell mass can be used to distinguish between a benign and malignant tumor to aid in clinical diagnosis.

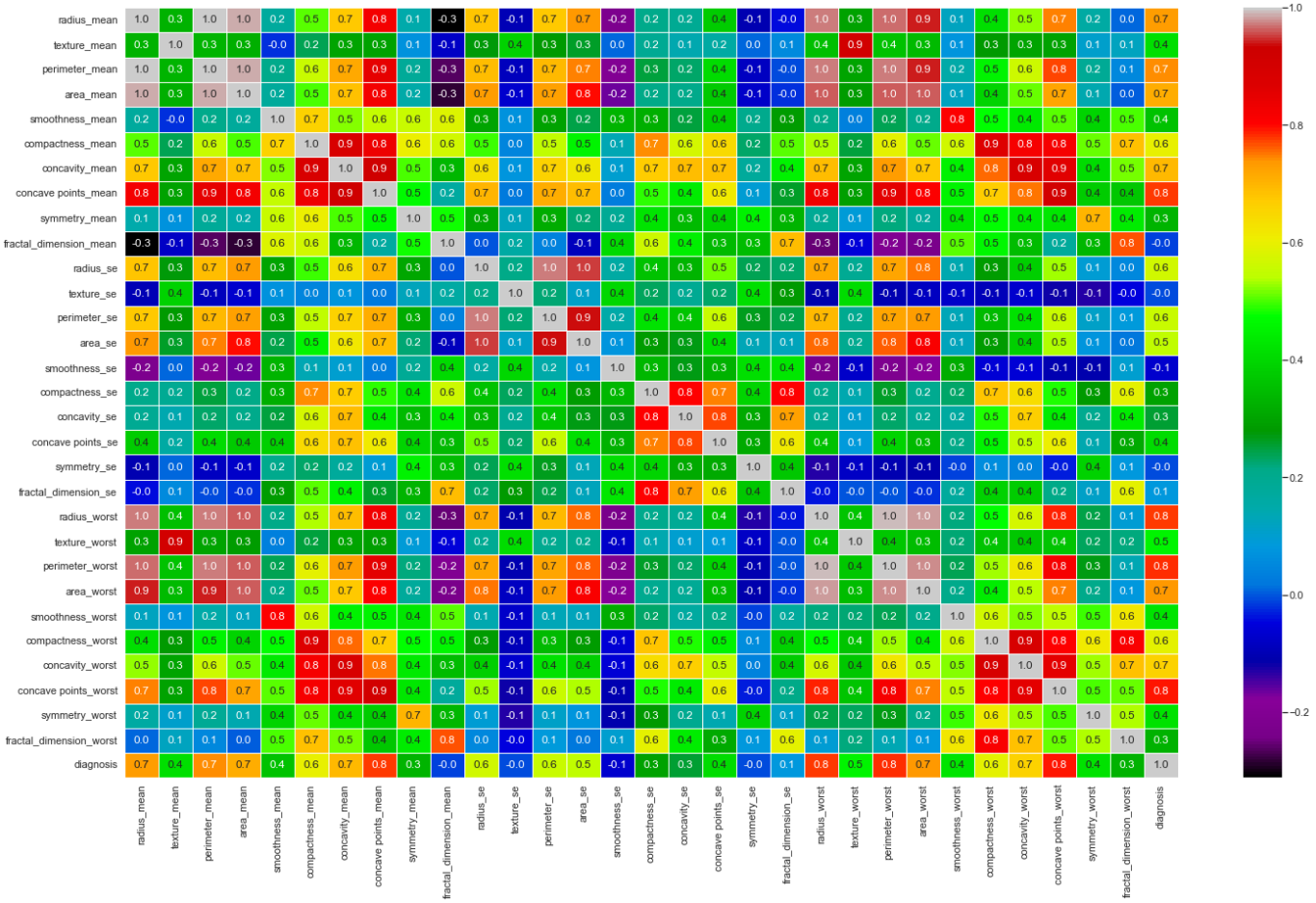
Features ?

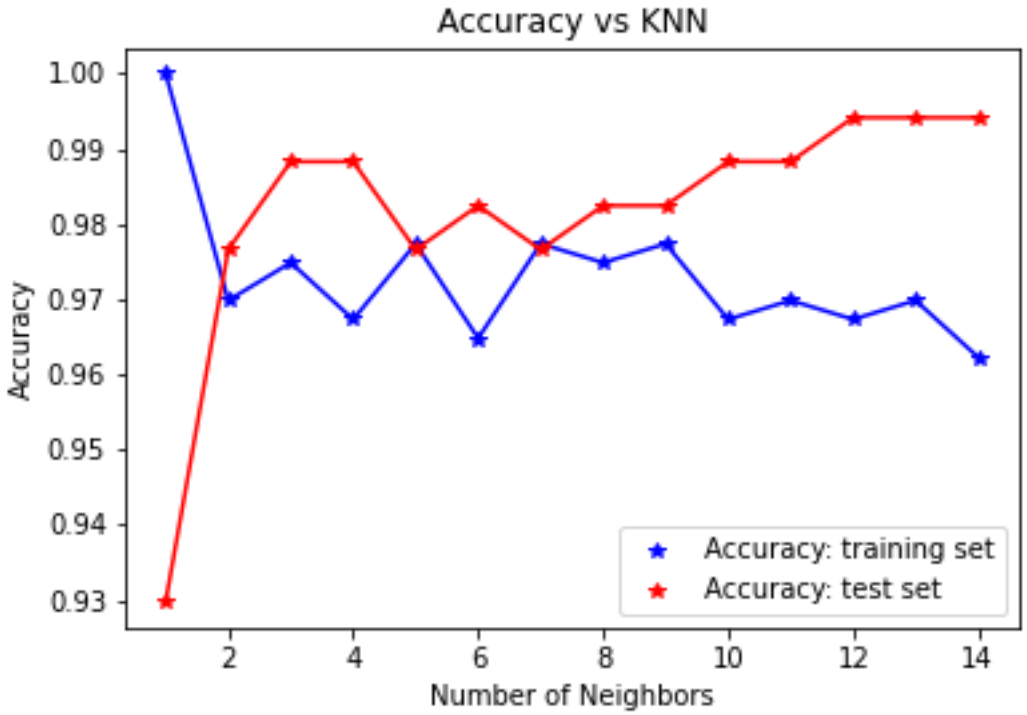

Diagnosis
count
350
300
250
200
150
100
50
0
0
1
diagnosis

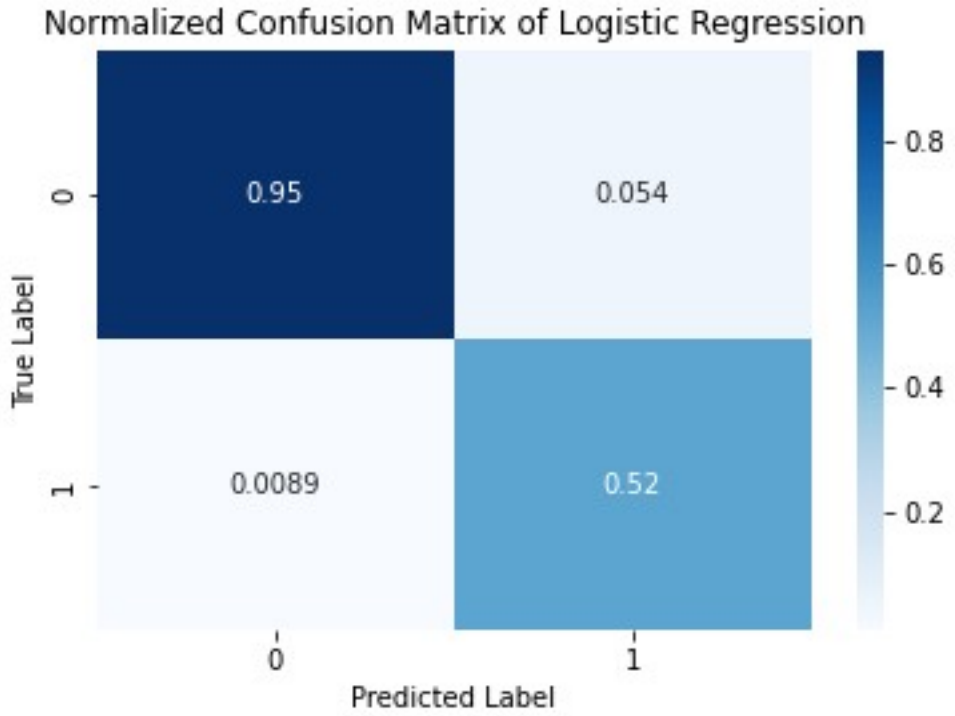

Cancer Diagnosis in Percentage
Percent
60
50
40
30
20
10
0
0
1
Diagnosis

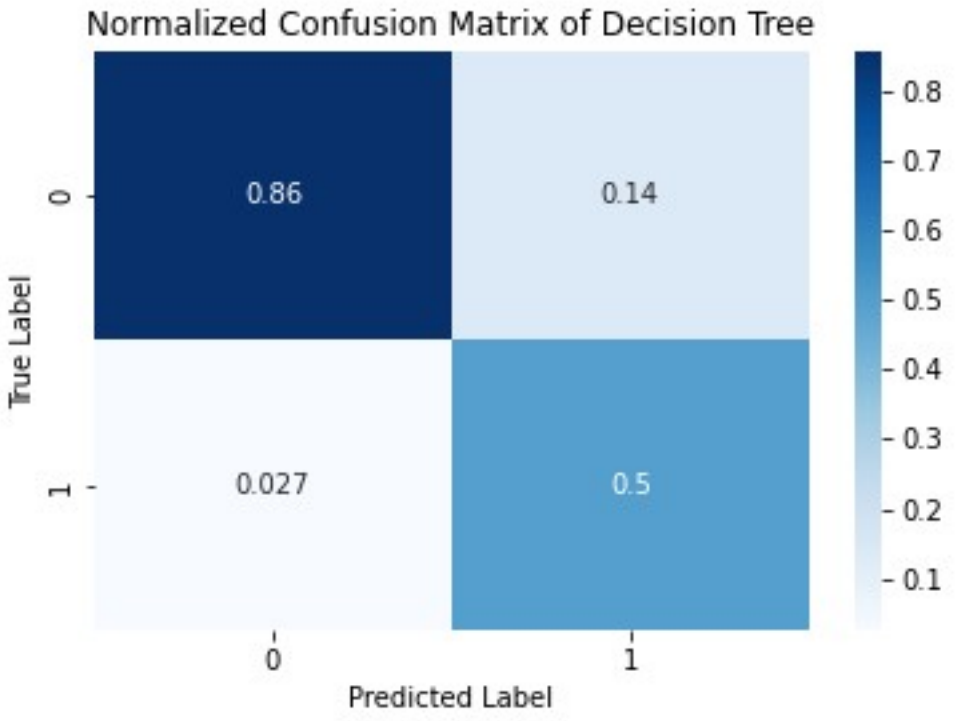

radius_mean
texture_mean
perimeter_mean
area_mean
smoothness_mean
compactness_mean
concavity_mean
concave points_mean
symmetry_mean
fractal_dimension_mean
radius_se
texture_se
perimeter_se
area_se
smoothness_se
Count
Benign
Malignant

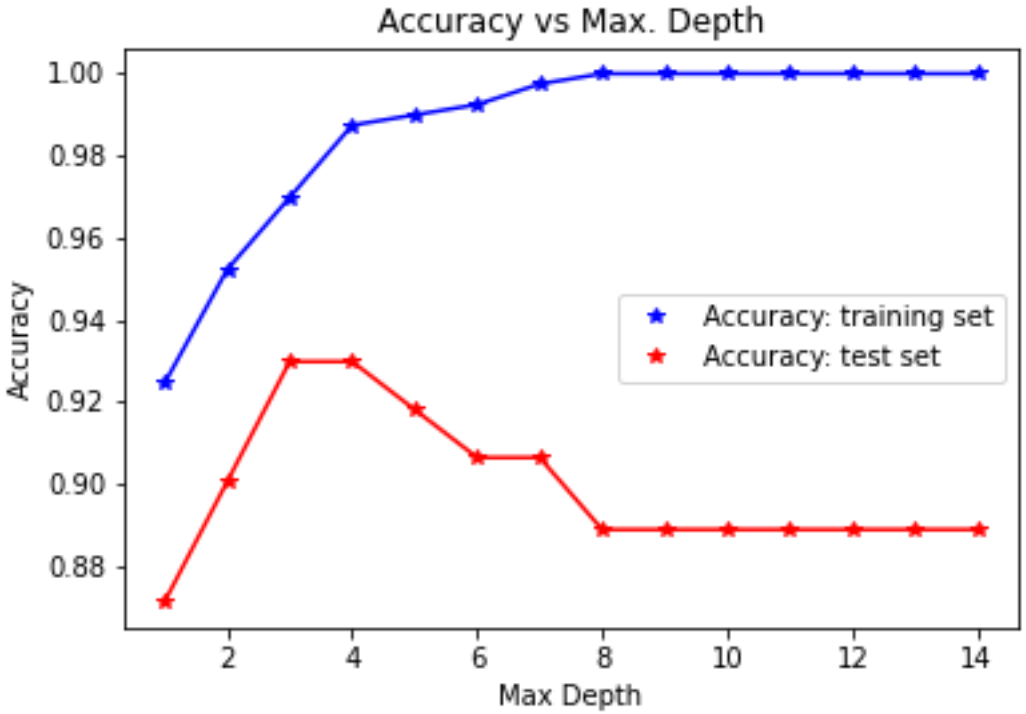

compactness_se
concavity_se
concave points_se
symmetry_se
fractal_dimension_se
radius_worst
texture_worst
perimeter_worst
area_worst
smoothness_worst
compactness_worst
concavity_worst
concave points_worst
symmetry_worst
fractal_dimension_worst
Count
Benign
Malignant

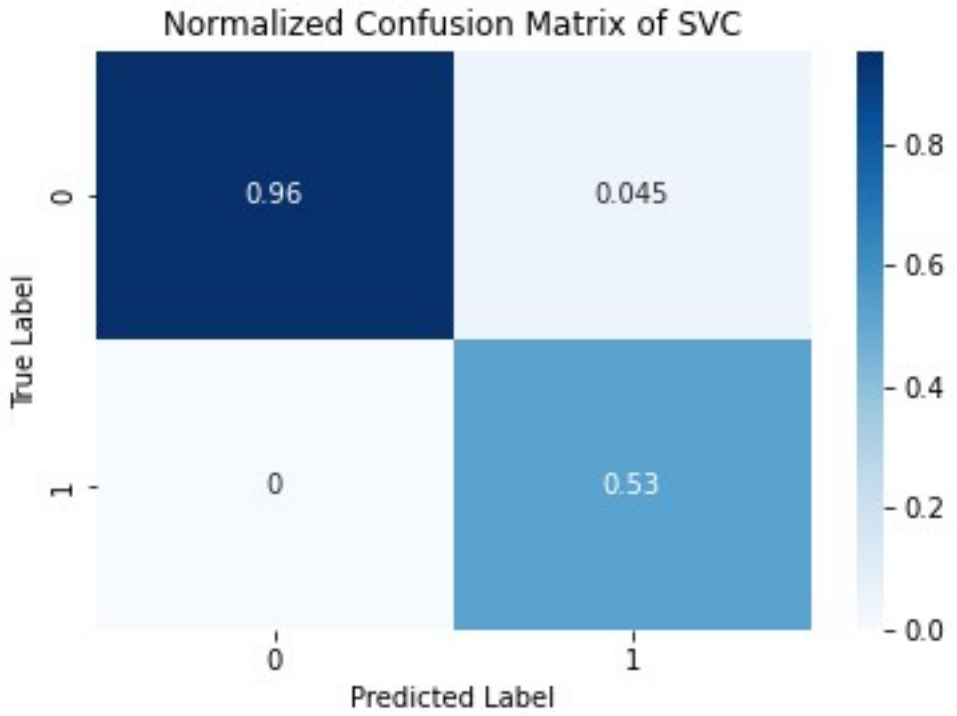

radius_mean
texture_mean
perimeter_mean
area_mean
smoothness_mean
compactness_mean
concavity_mean
concave points_mean
symmetry_mean
fractal_dimension_mean
radius_se
texture_se
perimeter_se
area_se
smoothness_se
compactness_se
concavity_se
concave points_se
symmetry_se
fractal_dimension_se
radius_worst
texture_worst
perimeter_worst
area_worst
smoothness_worst
compactness_worst
concavity_worst
concave points_worst
symmetry_worst
fractal_dimension_worst
diagnosis
1.0
0.8
0.6
0.4
0.2
0.0
−0.2


Accuracy vs KNN
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
Accuracy
2
4
6
8
10
12
14
Number of Neighbors
Accuracy: training set
Accuracy: test set


Normalized Confusion Matrix of Logistic Regression
0.95
0.054
0.0089
0.52
0
1
True Label
0
1
Predicted Label
0.8
0.6
0.4
0.2


Normalized Confusion Matrix of Decision Tree
0.86
0.14
0.027
0.5
True Label
0
1
Predicted Label
0
1
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1


Accuracy vs Max. Depth
Accuracy
Max Depth
1.00
0.98
0.96
0.94
0.92
0.90
0.88
2
4
6
8
10
12
14
Accuracy: training set
Accuracy: test set


Normalized Confusion Matrix of SVC
0.96
0.045
0
0.53
True Label
0
1
0
1
Predicted Label
0.8
0.6
0.4
0.2
0.0

Conclusions

Thank You !