Assignment 1 Report - CS 482

Introduction to Machine Learning and K-NN Classifier

Using Machine Learning to Classify Tumors

Maddy Connell & Dylan Lozon

Kettering University

10/20/2023

Table of Contents

Meet the Data

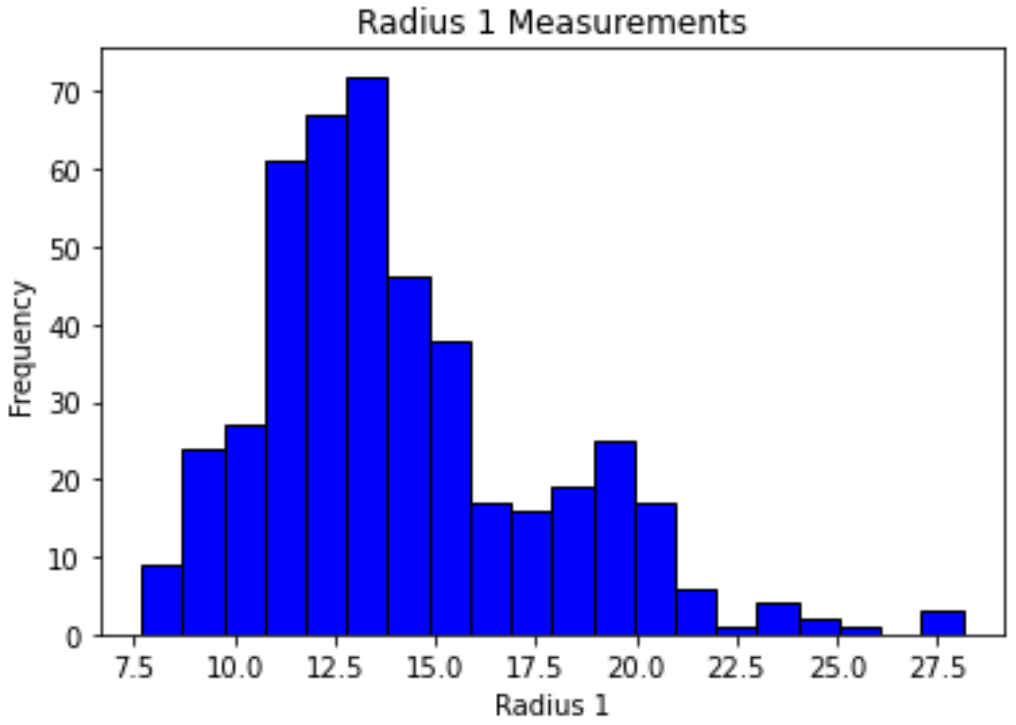
Number of Features: 30
Names of Features:
1. RADIUS1
2. TEXTURE1
3. PERIMETER1
4. AREA1
5. SMOOTHNESS1
6. COMPACTNESS1
7. CONCAVITY1
8. CONCAVE_POINTS1
9. SYMMETRY1
10. FRACTAL_DIMENSION1
11. RADIUS2
12. TEXTURE2
13. PERIMETER2
14. AREA2
15. SMOOTHNESS2
16. COMPACTNESS2
17. CONCAVITY2
18. CONCAVE_POINTS2
19. SYMMETRY2
20. FRACTAL_DIMENSION2
21. RADIUS3
22. TEXTURE3
23. PERIMETER3
24. AREA3
25. SMOOTHNESS3
26. COMPACTNESS3
27. CONCAVITY3
28. CONCAVE_POINTS3
29. SYMMETRY3
30. FRACTAL_DIMENSION3

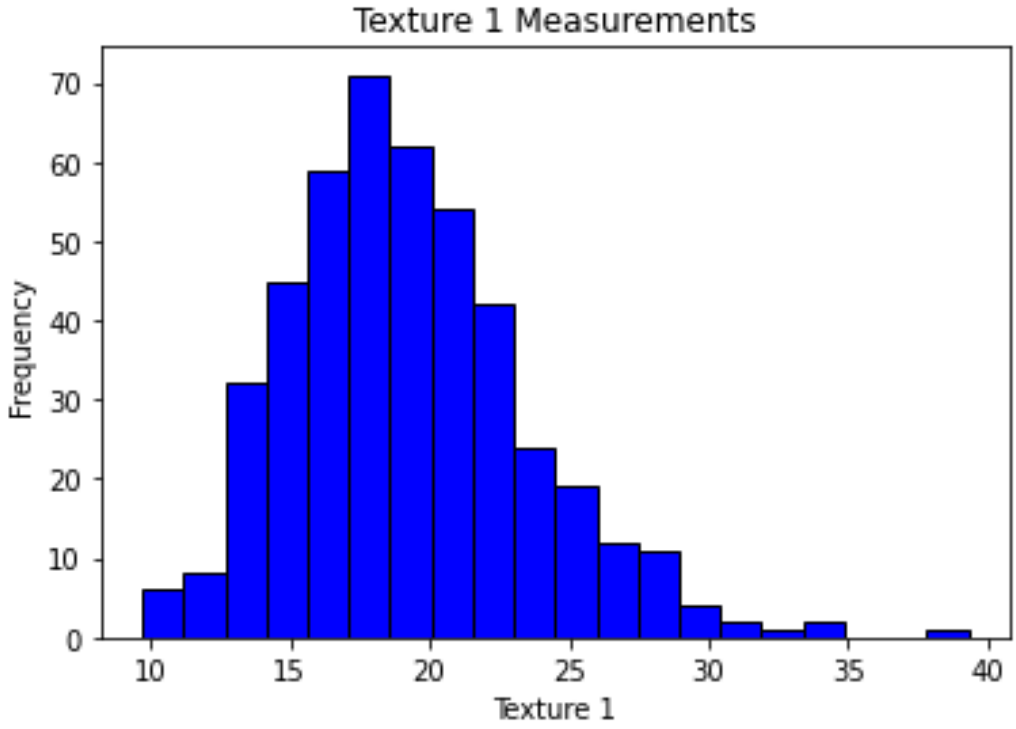
Name of Target: DIAGNOSIS
Number of Samples: 569

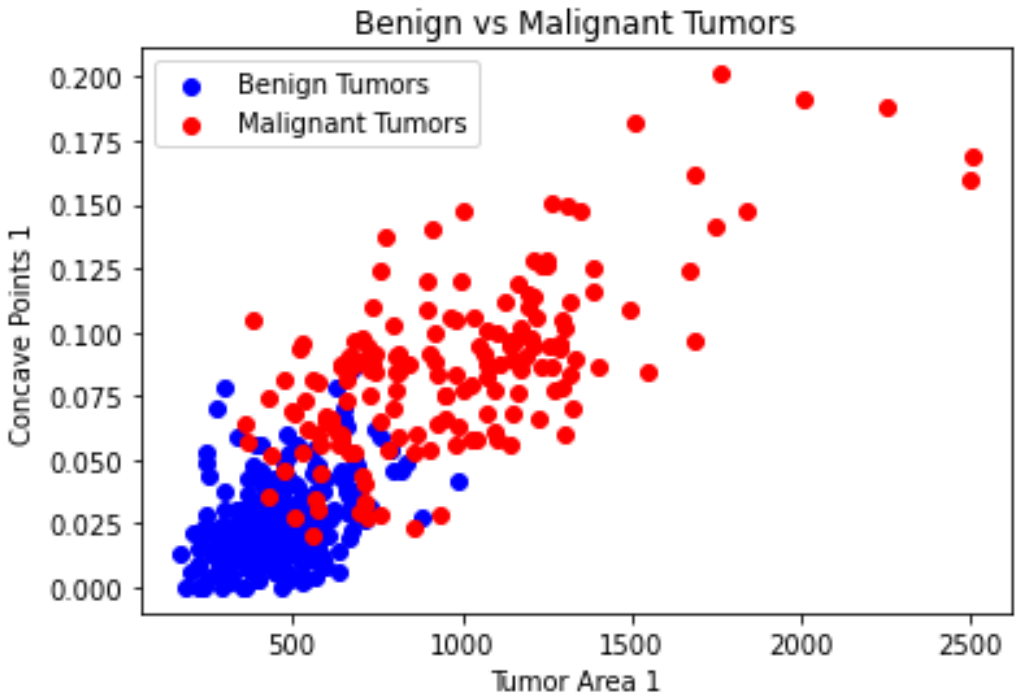
First 5 Rows of Data:

- 17.99,10.38,122.8,1001,0.1184,0.2776,0.3001,0.1471,0.2419,0.07871,1.095,0.9053,8.589,153.4,0.006399,0.04904,0.05373,0.01587,0.03003,0.006193,25.38,17.33,184.6,2019,0.1622,0.6656,0.7119,0.2654,0.4601,0.1189,1
- 20.57,17.77,132.9,1326,0.08474,0.07864,0.0869,0.07017,0.1812,0.05667,0.5435,0.7339,3.398,74.08,0.005225,0.01308,0.0186,0.0134,0.01389,0.003532,24.99,23.41,158.8,1956,0.1238,0.1866,0.2416,0.186,0.275,0.08902,1
- 19.69,21.25,130,1203,0.1096,0.1599,0.1974,0.1279,0.2069,0.05999,0.7456,0.7869,4.585,94.03,0.00615,0.04006,0.03832,0.02058,0.0225,0.004571,23.57,25.53,152.5,1709,0.1444,0.4245,0.4504,0.243,0.3613,0.08758,1
- 11.42,20.38,77.58,386.1,0.1425,0.2839,0.2414,0.1052,0.2597,0.09744,0.4956,1.156,3.445,27.23,0.00911,0.07458,0.05661,0.01867,0.05963,0.009208,14.91,26.5,98.87,567.7,0.2098,0.8663,0.6869,0.2575,0.6638,0.173,1
- 20.29,14.34,135.1,1297,0.1003,0.1328,0.198,0.1043,0.1809,0.05883,0.7572,0.7813,5.438,94.44,0.01149,0.02461,0.05688,0.01885,0.01756,0.005115,22.54,16.67,152.2,1575,0.1374,0.205,0.4,0.1625,0.2364,0.07678,1

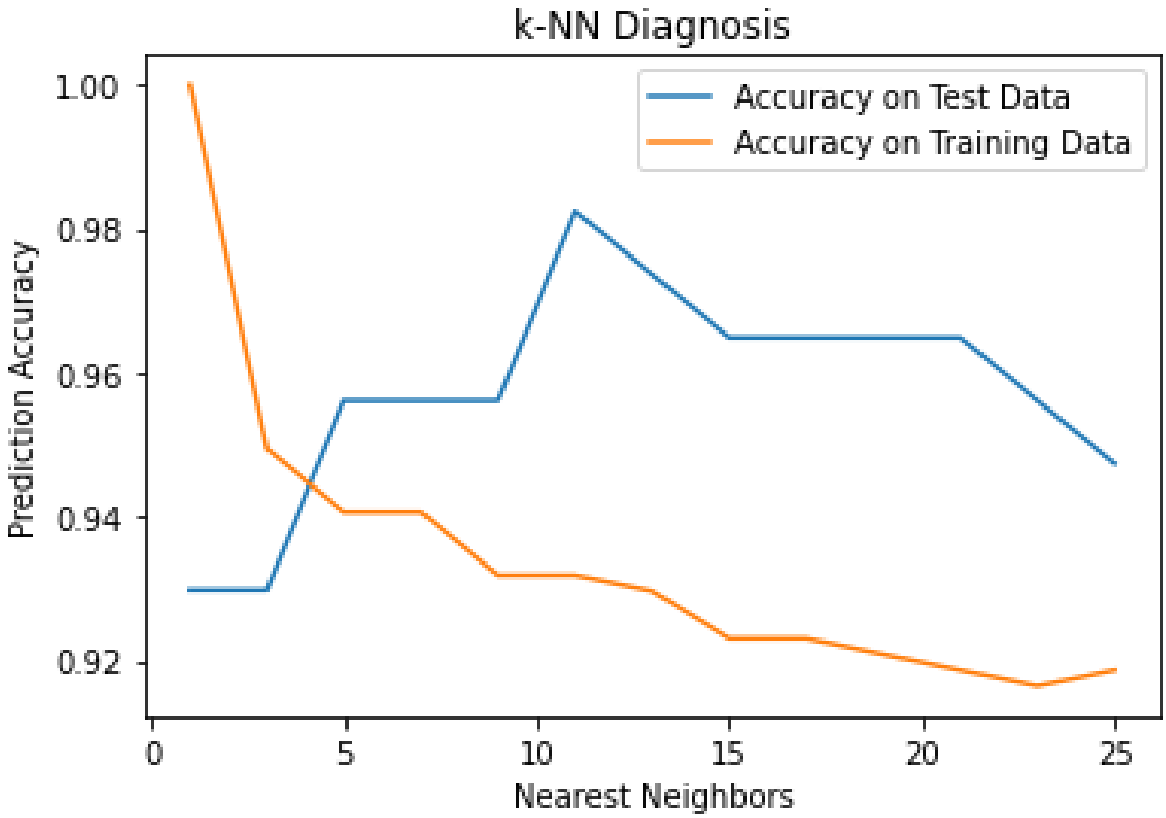
Learning from Training Data






Benign vs Malignant Tumors
Benign Tumors
Malignant Tumors
Concave Points 1
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
500
1000
1500
2000
2500
Tumor Area 1

Results of KNN Performance



The best value produced was Nearest Neighbors = 5. At this point, the Test Accuracy was 0.941, while the Test Accuracy was 0.956. This gives an overall difference of 0.015 between the two, which is the point of least difference.

Results of Cross-Validation

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
Training Accuracy	0.943	0.956	0.949	0.938	0.947	0.947
Test Accuracy	0.93	0.904	0.93	0.956	0.956	0.935