Samuel Slomowitz

A Dive into Machine Learning and Data Visualization on Breast Cancer Data

Introduction

The incidence of breast cancer in women is the second highest behind skin cancer. The median age of diagnosis is 62. On average, 1 in 8 women will develop breast cancer in the USA. Death by breast cancer ranks second in women behind lung cancer. Further, 3.8 million patients overcome breast cancer as survivors.

Breast Cancer Dataset from Kaggle:

https://www.kaggle.com/datasets/sarahvch/breast-cancer-wisconsin-prognostic-data-set

Inspiration:

```
In [13]: #visualize the count
         sns.countplot(df1['diagnosis'],label= 'count')

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0xd26c090088>
```

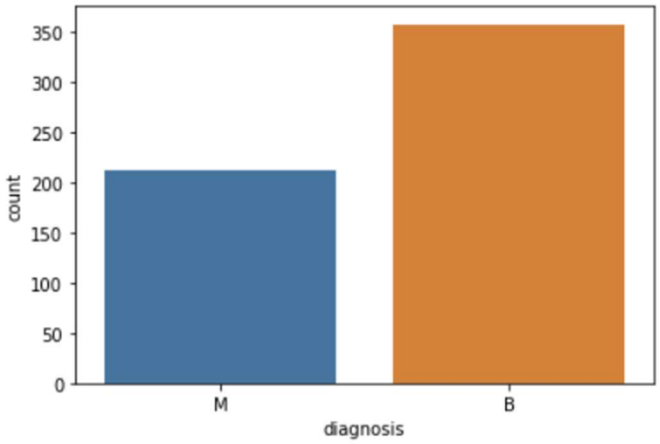
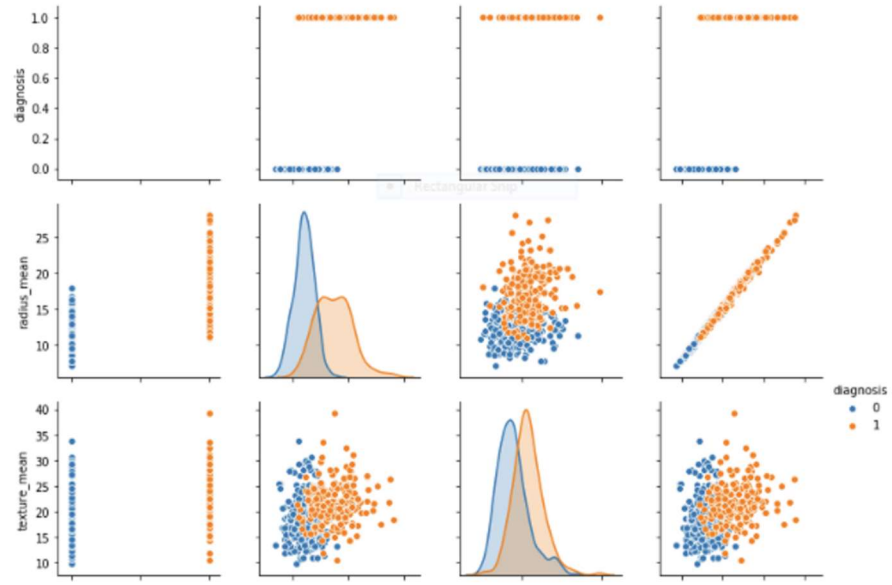
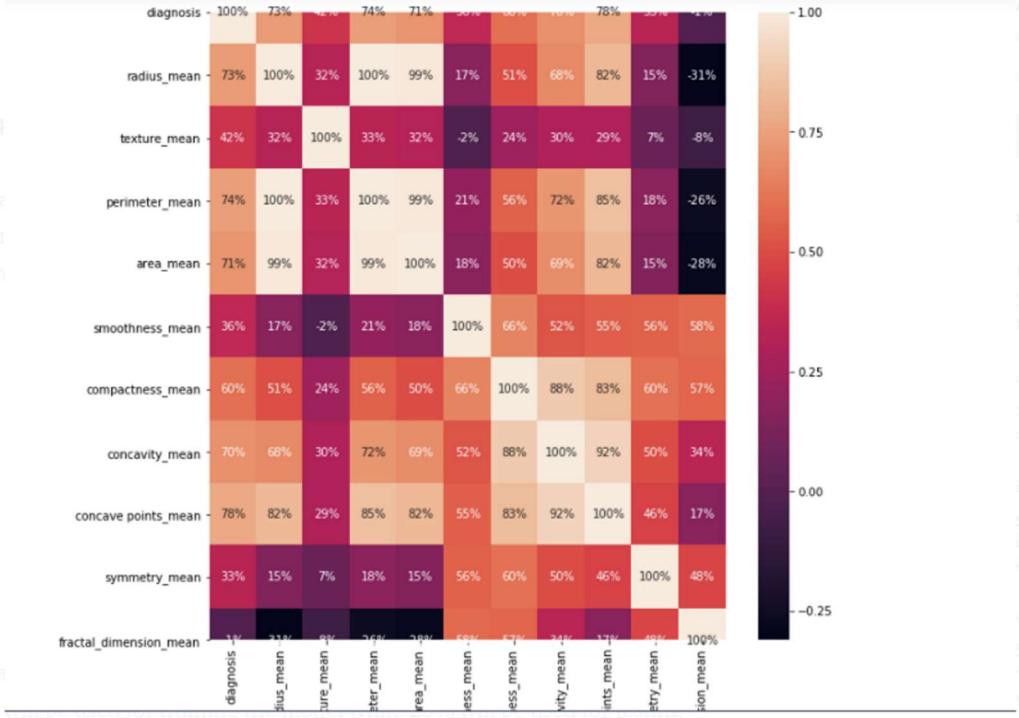


Fig : visualization of number of benign and malignant

```
Out[14]: <seaborn.axisgrid.PairGrid at 0x2e58308e88>
```





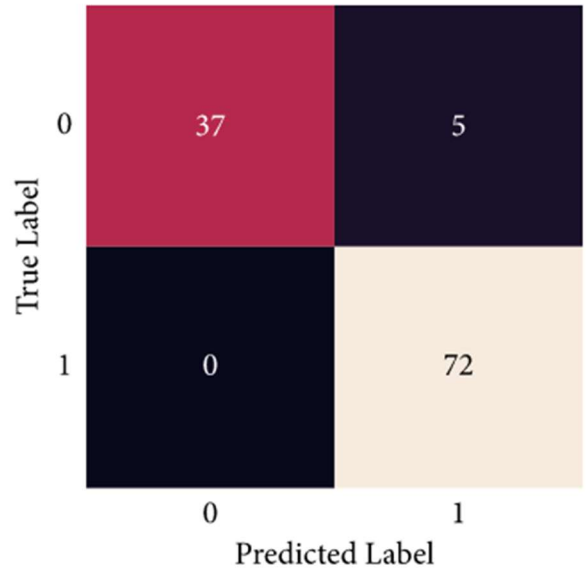


Figure 15

Random forest confusion matrix.

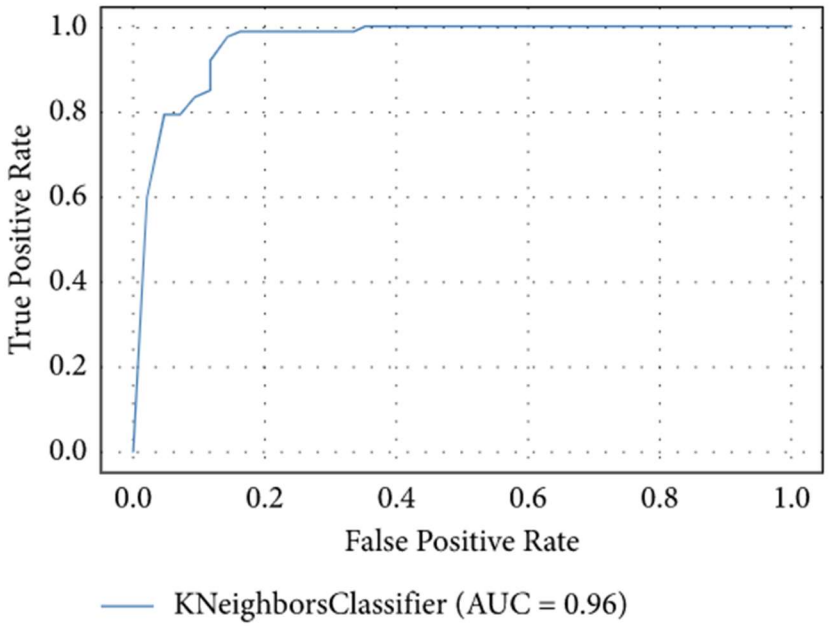
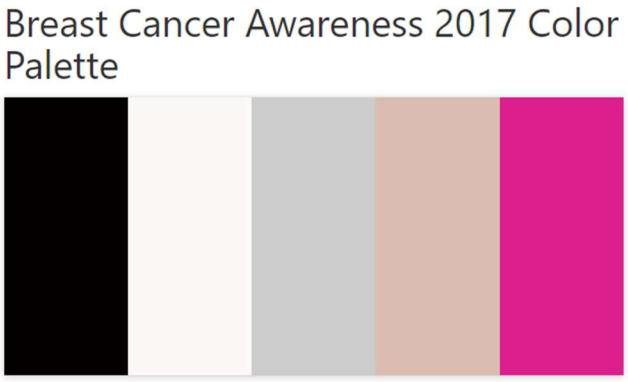


Figure 20

KNN AUC curve.

Colors:



Colors in Palette

Color	Hex	RGB
	#050000	(5,0,0)
	#fcf8f8	(252,248,248)
	#cccccc	(204,204,204)
	#e1bdb1	(225,189,177)
	#ed008c	(237,0,140)

Predictive Analytics:

I will be predicting prognosis of breast cancer, i.e. malignant or begin tumor in biopsies.

Basic design concepts for dashboard:

Menu

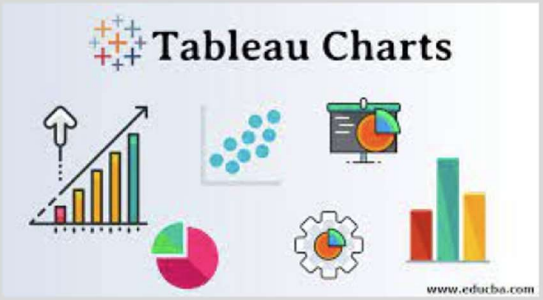
Predictive Analytics on Breast Cancer Dataset

Objective/Summary

"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum."

Filter for Interactive Dashboard

Tableau Powered, Imbeded Graph/Data Story



Machine Learning Parameters for Predictive Analytics

GO!

Output: Benign or Malignant Breast Cancer

References

https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html

https://medium.com/swlh/simple-machine-learning-model-on-breast-cancer-dataset-c1a013b594ad