**A Synopsis of**

**Breast Cancer Detection**

**Machine Learning Python**

**Contents**

|  |  |
| --- | --- |
|  | Page No. |
| Introduction | 3 |
| Objectives | 3 |
| Background | 3 |
| System Requirements | 3 |
| Coding and Output | 4-6 |
| Future Scope | 6 |
| Conclusion | 6 |
| Bibliography and References | 6-7 |

**Introduction:**

Artificial intelligence in healthcare is an overarching term used to describe the use of machine-learning algorithms and software, or artificial intelligence (AI), to mimic human cognition in the analysis, presentation, and comprehension of complex medical and health care data, or to exceed human capabilities by providing new ways to diagnose, treat, or prevent disease. Specifically, AI is the ability of computer algorithms to approximate conclusions based solely on input data.

Breast cancer is considered a multifactorial disease and the most common cancer in women worldwide with approximately 30% of all female cancers (i.e. 1.5 million women are diagnosed with breast cancer each year, and 500,000 women die from this disease in the world). Over the past 30 years, this disease has increased, while the death rate has decreased. However, the reduction in mortality due to mammography screening is estimated at 20% and improvement in cancer treatment is estimated at 60%.

Machine learning, as a modeling approach, represents the process of extracting knowledge from data and discovering hidden relationships, widely used in healthcare in recent years to predict different diseases. Some studies only used demographic risk factors (lifestyle and laboratory data) in predicting breast cancer, and several studies predicted based on mammographic stereotypes or used data from patient biopsy. Others showed the application of genetic data in predicting breast cancer.

**(2.) Objective:**

This study aimed to predict breast cancer using different machine-learning approaches applying demographic, laboratory, and mammographic data

**(3.) Background:**

Breast cancer is considered one of the most common cancers in women caused by various clinical, lifestyle, social, and economic factors. Machine learning has the potential to predict breast cancer based on features hidden in data.

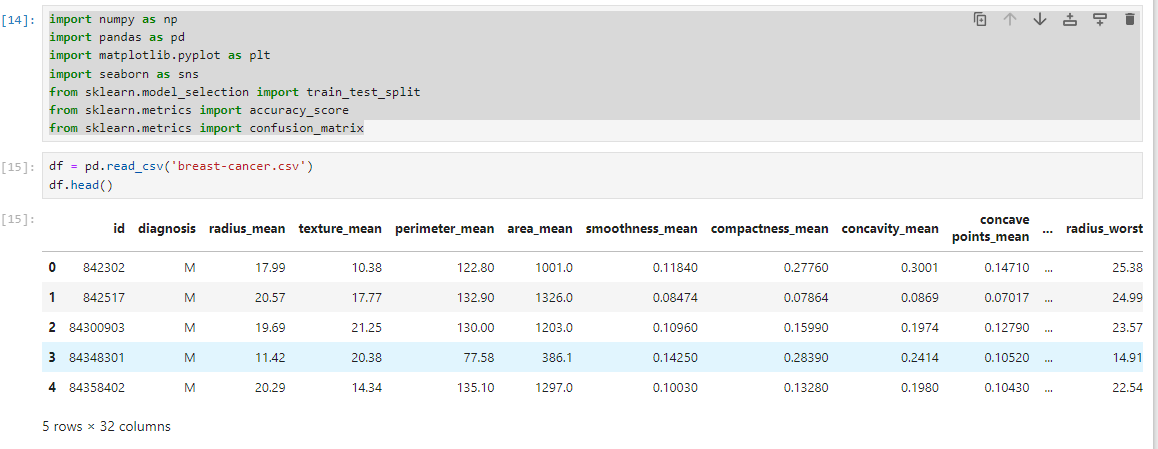
**(4.) Hardware and Software Requirements:**

|  |  |
| --- | --- |
| **HARDWARE TOOLS** | **MINIMUM REQUIREMENTS** |
| Processor | I5 or above |
| RAM | 8 GB |
| Monitor | 15.6” colored |
| Mouse | Optical |
| Keyboard | 122 keys |
| Graphics Card | Intel HD graphics |

|  |  |
| --- | --- |
| **SOFTWARE TOOLS** | **MINIMUM REQUIREMENTS** |
| Platform | Windows, Linux, MacOS |
| Operating System | Windows, Linux, MacOS |
| Technology | Machine Learning Python |
| Scripting Language | Python |
| IDE | Jupyter Notebook |

**(5.) Coding and Output:**

In this section we will go over how to load our dataset using pandas.

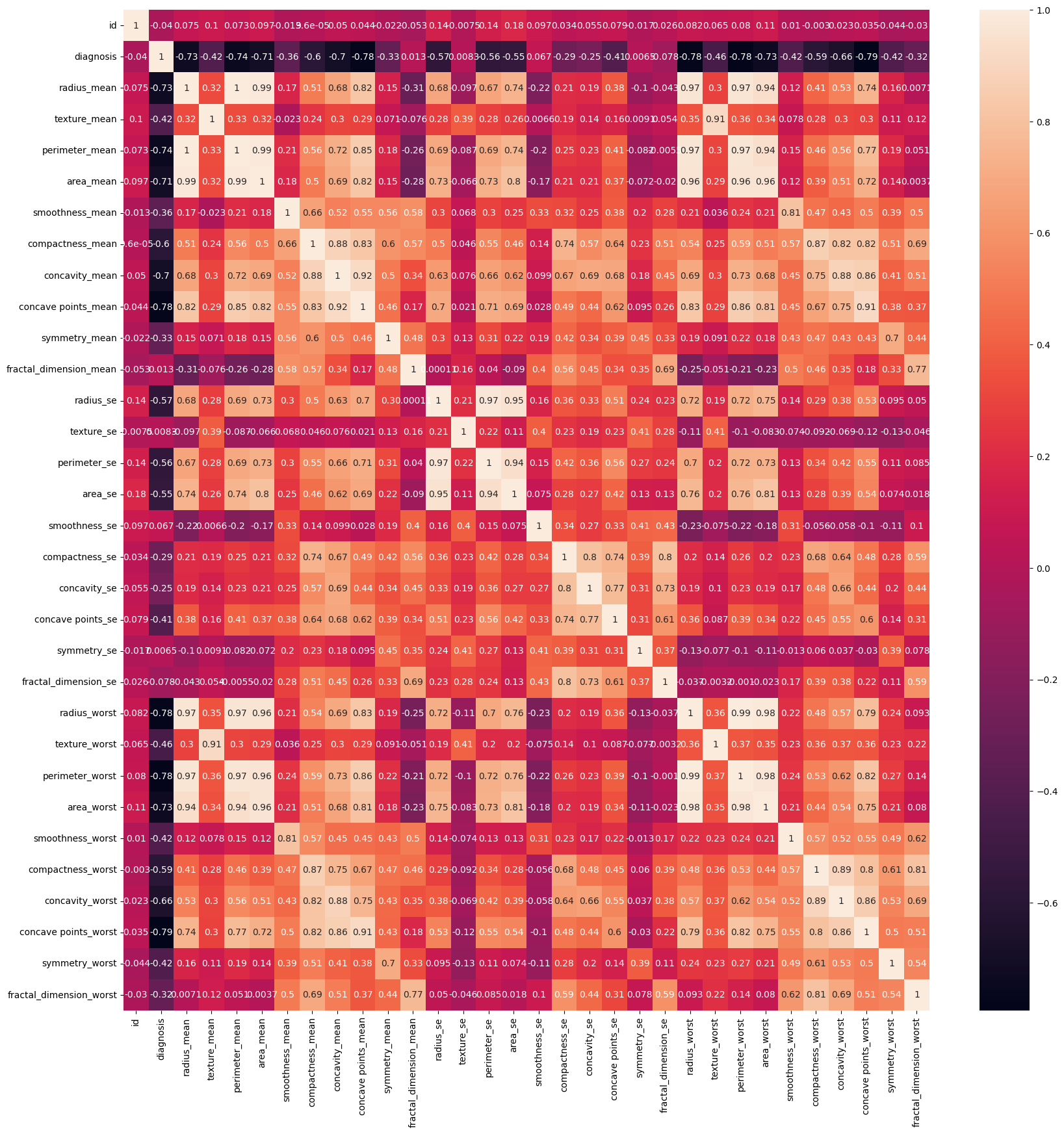


#Data Preprocessing

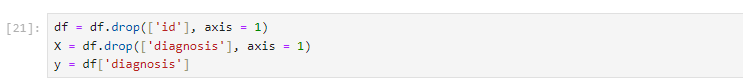
We will first convert diagnosis vlaues (B and M) to some numerical value (0,1) so that our ML models could be trained.



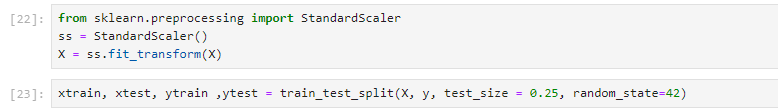
#Heatmap of the dataset



#There is no need of id in predicting breast cancer so we will drop it.



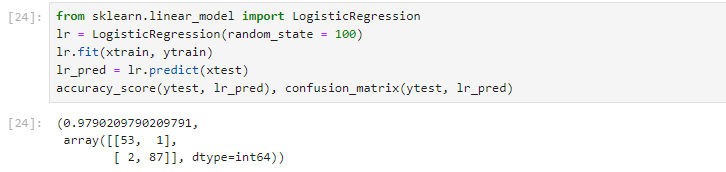
#Splitting dataset into test test and training set.



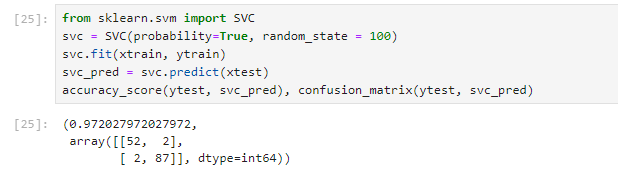
#To test our ML models for their accuracy, we will split our dataset: 25% data for testing (test set) and 75% for training our model(train set).

#Training Models

#Logistic Regression



#Support Vector Machines(SVM)



**(6.) Future Scope:**

We believe that AI has an important role to play in the healthcare offerings of the future. In the form of machine learning, it is the primary capability behind the development of precision medicine, widely agreed to be a sorely needed advance in care. Although early efforts at providing diagnosis and treatment recommendations have proven challenging, we expect that AI will ultimately master that domain as well. Given the rapid advances in AI for imaging analysis, it seems likely that most radiology and pathology images will be examined at some point by a machine. Speech and text recognition are already employed for tasks like patient communication and capture of clinical notes, and their usage will increase.

**(7.) CONCLUSION:**

Combining multiple risk factors in modeling for breast cancer prediction could help the early diagnosis of the disease with necessary care plans. Collection, storage, and management of different data and intelligent systems based on multiple factors for predicting breast cancer are effective in disease management.

The proposed machine-learning approaches could predict breast cancer as the early detection of this disease could help slow down the progress of the disease and reduce the mortality rate through appropriate therapeutic interventions at the right time. Applying different machine learning approaches, accessibility to bigger datasets from different institutions (multi-center study), and considering key features from a variety of relevant data sources could improve the performance of modeling.

**(8.) REFERENCES AND BIBLIOGRAPHY:**

1. Zhang X, Shengli SU, Hongchao WA. Intelligent diagnosis model and method of palpation imaging breast cancer based on data mining. Big Data Research . 2019

2. Chen SI, Tseng HT, Hsieh CC. Evaluating the impact of soy compounds on breast cancer using the data mining approach. Food & function

3. Aavula R, Bhramaramba R, Ramula US. A Comprehensive Study on Data Mining Techniques used in Bioinformatics for Breast Cancer Prognosis. Journal of Innovation in Computer Science and Engineering . 2019

4. Kaushik D, Kaur K. Application of Data Mining for high accuracy prediction of breast tissue biopsy results. 2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC); Moscow, Russia: IEEE

5. Mokhtar SA, Elsayad A. Predicting the severity of breast masses with data mining methods.