# PROJECT DOCUMENTATION

**INDEX-**

* Project Name
* Methods applied
* Needs
* Solutions
* Benefits
* Financial Cost
* Conclusion

1. **Project Name**

Breast Cancer Detection Model Using Machine Learning

1. **Methods applied**

Here we used different process to find the accurate algorithm to detect the breast cancer detection and analyse the data to get the desired prediction such as,

* Data Pre-processing
* Data Exploration
* Data Preparation
* Categorical Data
* Feature Scaling
* Model Selection
* Logistic Regression
* KNN Algorithm
* SVM
* Kernel-SVM
* Naïve Bayes
* Decision Tree
* Random Forest
* Prediction

1. **Needs**

The second major cause of women's death is breast cancer (after lung cancer). It can be medically detected early during a screening examination through mammography or by portable cancer diagnostic tool. Cancerous breast tissues change with the progression of the disease, which can be directly linked to cancer staging. The stage of breast cancer (I–IV) describes how far a patient’s cancer has proliferated. The people who need this model are given below in the table:

|  |  |
| --- | --- |
| * **Women age 40–45 or older** * **Personal history of breast cancer** * **Family history of breast cancer** * **Genetic factors** | * **Childbearing and menstrual history** * **Women who menstruate for the first time at an early age (before 12)** * **Women who go through menopause late (after age 55)** * **Women who’ve never had children** |

1. **Solutions**

Breast Cancer has stood out to be one of the most feared and frequently occurring cancers at present among women. While the cure for this cancer is now available in almost all first world and some of the third world nations, the main dilemma takes place when the cancer cannot be correctly identified at the very initial stages. Machine Learning, in this field has proved to play a vital role in predicting diseases such as cancers alike. Classification and data mining methods so far have been reliant and an effective way to classify data. Especially in medical field, these methods have been used to predict and to make decisions.

In this project, we have successfully used seven classification techniques in the form of Decision Tree, K-NN, Random Forest, Logistic Regression, Naïve Bayes, Kernel -SVM and Support Vector Machine (SVM) on the Wisconsin Breast Cancer datasets. The main objective is to assess the correctness in classifying data with respect to efficiency and effectiveness of each algorithm in terms of accuracy.

1. **Benefits**
   1. Breast cancer if found at an early stage will help save lives of thousands of women or even men.
   2. Doctors will be able to classify and predict the cancer into being or malignant.
   3. This will increase the efficiency of the testing.
   4. It advances the system.
   5. It reduces human errors.
   6. It lowers manual mistakes.
   7. This will reduce the time required for testing.
2. **Financial Cost**

Since, we have applied the human work force only and we get data from Wisconsin data archive.

So, no cost has been applied for this model.

1. **Conclusion**

We implemented the methodology for predicting if the person is suffering from breast cancer. The aim of the present study is to identify the most efficient algorithm for detection. This will help the doctors to identify cancer in less time. We identified the best suited algorithm. The solution we have proposed will help the patients as well as the doctors diagnose the problem efficiently.