**SWOT ANALYSIS**

|  |  |
| --- | --- |
| **STRENGTHS:** | **WEAKNESS:** |
| ML algorithms have been widely used in Breast cancer diagnosis and prognosis to gain different insights from data samples. Machine learning algorithms can be used for medical oriented research, it advances the system, reduces human errors and lowers manual mistakes. This project has tried and tested several methods like Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree Classification and Random Forest Classification to predict the cancer. The results showed that Random forest Classification was the best predictor with the accuracy of 94.15%, with Logistic Regression achieving 92.39%. By using this algorithm, we can detect breast cancer at a very early stage which will help save lives of thousands of women. | One of the most common problems seen among the studies is the lack of attention paid to data size and learner validation. There are a number of studies with sloppy experimental design. For any ML exercise, a minimum requirement is of having a sufficiently large data set that can be partitioned into disjoint training and test sets. The size of a given training set has several implications pertaining to robustness, reproducibility and accuracy. Therefore, a larger and more reliable dataset is required to get the efficiency even higher and accurate.  In our model, 6% efficiency still remains and this could cause few problems. |
| **OPPORTUNITIES:** | **THREATS:** |
| These algorithms will help the real-world patients and doctors to gather as much information as they can in future as well as for easy detection of breast cancer in hospitals and clinics.  There will be reduction in number of breast cancer deaths if they detect the breast cancer at an early stage which will increase the survival rate of the patients.  It is likely that the use of machine learning classifier will become much more commonplace in many clinical and hospital settings. | The major threat in this model is the cyber security attack.  Data set could be hacked or manipulated or can be stolen due to which there will be wrong predictions  It can harm the patients too if wrong information is provided to them  So, a more secure data is needed in order to maintain privacy. |