**SOFTWARE REQUIREMENTS SPECIFICATIONS DOCUMENT**

**Index-**

1. Introduction:
   1. Abstract
   2. Purpose
   3. Motivation
   4. Intended Audience
   5. Objective
   6. Scope
   7. Definitions and Acronyms
2. Overall Descriptions
   1. Users Need
   2. Assumptions and Dependencies
3. Systems Features and Requirements:
   1. Functional Requirements:
   2. Non-Functional Requirements
4. **INTRODUCTION**
   1. **Abstract**

Women are seriously threatened by breast cancer with high morbidity and mortality. The lack of robust prognosis models results in difficulty for doctors to prepare a treatment plan that may prolong patient survival time. Hence, the requirement of time is to develop the technique which gives minimum error to increase accuracy. Four algorithm SVM, Logistic Regression, Random Forest and KNN which predict the breast cancer outcome have been compared in the paper using different datasets. All experiments are executed within a simulation environment and conducted in JUPYTER platform. Aim of research categorises in three domains. First domain is prediction of cancer before diagnosis, second domain is prediction of diagnosis and treatment and third domain focuses on outcome during treatment. The proposed work can be used to predict the outcome of different technique and suitable technique can be used depending upon requirement. This research is carried out to predict the accuracy. The future research can be carried out to predict the other different parameters and breast cancer research can be categorises on basis of other parameters.

* 1. **Purpose**

Breast cancer (BC) is one of the most common cancers among women worldwide, representing the majority of new cancer cases and cancer-related deaths according to global statistics, making it a significant public health problem in today’s society.

The early diagnosis of BC can improve the prognosis and chance of survival significantly, as it can promote timely clinical treatment to patients. Further accurate classification of benign tumours can prevent patients undergoing unnecessary treatments. Thus, the correct diagnosis of BC and classification of patients into malignant or benign groups is the subject of much research. Because of its unique advantages in critical features detection from complex BC datasets, machine learning (ML) is widely recognized as the methodology of choice in BC pattern classification and forecast modelling.

Classification and data mining methods are an effective way to classify data. Especially in medical field, where those methods are widely used in diagnosis and analysis to make decisions

* 1. **Motivation**

Breast Cancer is the most affected disease present in women worldwide. 246,660 of women's new cases of invasive breast cancer are expected to be diagnosed in the U.S during 2016 and 40,450 of women’s death is estimated. The development in Breast Cancer and its prediction fascinated. The UCI Wisconsin Machine Learning Repository Breast Cancer Dataset attracted as large patients with multivariate attributes were taken as sample set.

* 1. **Intended Audience**
* 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. **Objective**

The objective of the model is to:

* To find the algorithm for the prediction for the particular dataset.
* To study various machine learning approaches for breast cancer diagnosis through their implementation.
* To make a comparative study of the approaches.

The main aim of the project is to predict the breast cancer using Wisconsin dataset. We have tested seven algorithms and observed which algorithm is best for the model.

* 1. **Scope:**

This model will help to predict whether the person will suffer from breast cancer. Since it uses the Wisconsin data so the accuracy is More accurate rather than a general one. This model will investigate the entire conditions for providing new and unique solution for given problems. The scope of the product is not limited. It has a wider range of perspective.

* 1. **Definitions And Acronyms**
  + df: variable to store imported dataset.
  + df.info: information of dataset
  + x: variable that stores independent attributes.
  + y: variable that stores dependent attribute.
  + sc: object that call standard scaler.
  + x\_train: x training data.
  + x\_test: x testing data.
  + y\_train: y training data.
  + y\_test: y testing data.
  + classifier: object of the classification algorithms.
  + y\_pred: predicted values.
  + cm: confusion matrix

1. **OVERALL DESCRIPTION**
   1. **Users Need:**

Requirements of users is to get a model predict breast cancer by giving the required input. The report generated by mammogram is filled as input and the type of tumour is predicted. The hospital management will also be benefitted. As the time to tell the type of tumour will be reduced. Efficiency and accuracy of the report will be increased.

* 1. **Assumptions and Dependencies**

We are not giving guarantee that the predicted result will 100% correct. There are several problems that are related to some external features which this model would not understand so be cautious about it. Problems related to hospital management, filling up the data correctly.

1. **SYSTEMS FEATURES AND REQUIREMENTS**
   1. **Functional requirements:**

* Importing libraries.
* Importing Datasets.
* Data Pre-processing
* Data Exploration
* Data Preparation
* Data Visualisation
* Categorical Data
* Feature Scaling
* Model Selection
* Prediction
  1. **Non-functional requirements:**
* Security
* Performance
* User Friendly
* Maintainability
* Durability
* Sustainability