Project Proposal

For

Breast Cancer Prediction

Submitted by

MUHAMMAD AHMAD KALEEM, FA16-BSE-131

Department of Computer Science, CU, Islamabad.

July 26th, 2020

**Table of Contents**

1 Introduction 1

2 Proposed System 1

3 Advantages/Benefits of Proposed System 1

4 Scope 1

5 Modules 1

5.1 Module 1: Model for Logistic Regression 1

5.2 Module 2: Graphical User Interface 1

6 Dataset 1

7 System Limitations/Constraints 1

8 Tools and Technologies 1

9 Conclusion 2

10 References 2

Project Category:

* **A-**Desktop Application/Information System

# Abstract

This Projects aims to solve a binary classification problem that is whether a person has breast cancer or not.

# Introduction

This project aims to provide a solution to quickly diagnose a person with breast cancer.

# Proposed System

The system will use a trained model which will categorize a person’s diagnosis into either benign or malignant. This will be solved by using logistic regression as this is a binary classification problem.

# Advantages/Benefits of Proposed System

The advantage of this system is that it will provide an immediate answer to whether a person has breast cancer or not.

# Scope

The system will first train a model by the provided dataset, after training the program will require parameters to be put into Graphical User Interface of the program then the system will predict whether the tumor is malignant or benign.

# Modules

## 5.1 Module 1: Model for Logistic Regression

This module will contain training a model for predicting whether the tumor is malignant or benign using a dataset

## 5.2 Module 2: Graphical User Interface

This module will contain the building of Graphical User Interface for the application which will contain a textbox where parameters are entered, submit button for classifying tumor into malignant or benign

# Dataset

The dataset is publicly available on the following [link](https://www.kaggle.com/uciml/breast-cancer-wisconsin-data). The only preprocessing required on the dataset will be to remove special characters.

# System Limitations/Constraints

* The model is trained by a dataset which will not be able to correctly predict all the time.

# Tools and Technologies

Table 1: Tools and Technologies

|  |  |  |  |
| --- | --- | --- | --- |
| **Tools**  **And**  **Technologies** | **Tools** | **Version** | **Rationale** |
| PyCharm | 2015 | IDE |
| MS Word | 2015 | Documentation |
| **Technology** | **Version** | **Rationale** |
| Python | 3.8.3 | Programming language |
| Html | 5 | Web Development |

# Conclusion

This project will provide an immediate solution for breast cancer prediction.

# References

Dataset: https://www.kaggle.com/mchirico/montcoalert