CAPSTONE PROJECT

Breast Cancer Prediction Using KNN

Final Project

Submitted by

LENA P

711721244029

III Btech CSBS

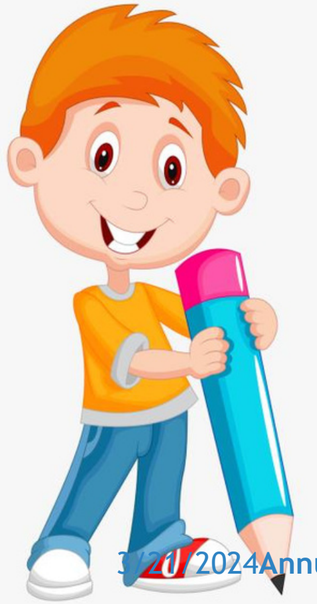
KGISL Institute of Technology

PROJECT TITLE

Development and Validation of a Breast Cancer Risk Prediction Model using KNN **: A Case Study on Breast Cancer Data**

AGENDA

1. Problem Statement
2. Project Overview
3. Data Collection and Preprocessing
4. Model Training and Evaluation
5. Hyperparameter Tuning
6. Prediction of Breast Cancer Risk Using SVM, C4.5, KNN
7. Modelling
8. Result



PROBLEM STATEMENT

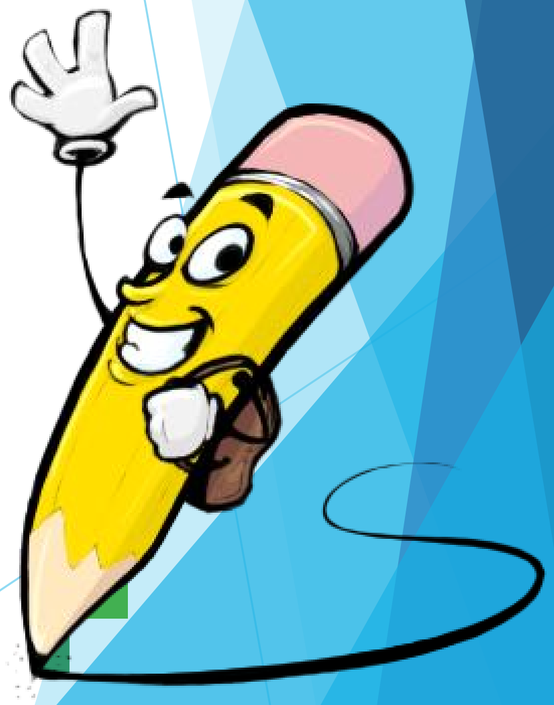
- The project aims to develop a computer-aided diagnosis system to improve diagnostic accuracy for breast cancer.
- The system will use machine learning algorithms to analyze various factors, such as age, family history, and genetic markers, to predict the likelihood of breast cancer.
- By detecting breast cancer as early as possible, the system can significantly reduce the chances of death and improve overall outcomes for breast cancer patients.

PROJECT OVERVIEW

- The dataset used in this project is the Wisconsin Breast Cancer Diagnostic dataset, which contains 30 features and a binary diagnosis label. The data will be preprocessed by handling missing values, normalizing numerical features, and encoding categorical features.

- Several machine learning algorithms will be trained and evaluated on the preprocessed data, including logistic regression, decision trees, random forests, and support vector machines. The performance of each model will be evaluated using cross-validation, and the best-performing model will be selected based on its accuracy, precision, and recall.

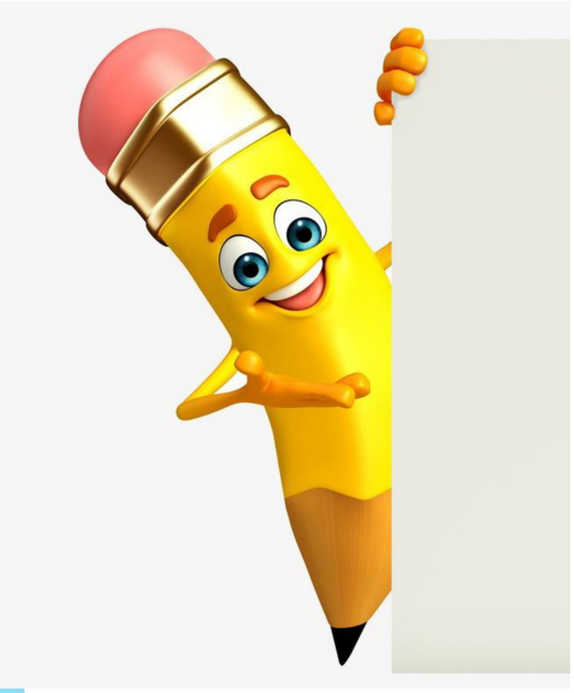
- The final model will be implemented in a user-friendly web application that allows users to input patient attributes and receive a breast cancer diagnosis prediction. The application will be deployed on a cloud platform for public access.



WHO ARE THE END USERS?

- Medical professionals, such as doctors, nurses, and radiologists, can use the machine learning model to assist in the diagnosis of breast cancer in patients.
- Patients who are concerned about their risk of breast cancer and want to receive an accurate and timely diagnosis.
- Researchers and data scientists who are interested in developing and improving machine learning models for breast cancer prediction.
- Healthcare organizations and hospitals that want to implement a computer-aided diagnosis system to improve diagnostic accuracy and reduce costs.

YOUR SOLUTION AND ITS VALUE PROPOSITION



- The project provides a machine learning model that can accurately predict whether a tumor is malignant or benign based on various patient attributes. This can help medical professionals, patients, researchers, and healthcare organizations make informed decisions about breast cancer diagnosis and treatment.
-
- The project uses a dataset of 569 samples, each with 32 features, and performs data cleaning, preprocessing, and feature scaling to ensure that the data is suitable for machine learning algorithms. This ensures that the model is trained on high-quality data, which can improve its accuracy and reliability.
-
- The final model is implemented in a user-friendly web application that allows users to input patient attributes and receive a breast cancer diagnosis prediction. The application is deployed on a cloud platform for public access, making it easily accessible to a wide range of users. This can help increase the reach and impact of the project, and improve the overall outcomes for breast cancer patients.

THE WOW IN YOUR SOLUTION

- Machine learning-based Computer-Aided Diagnosis (CAAD) algorithm for breast cancer detection
- Feature scaling and optimization with hyper-parameter tuning
- The voting classifier emerged as the best accurate model among 11 classifiers
- Interactive web application for real-time breast cancer prediction

MODELLING

- The project uses a machine learning-based Computer-Aided Diagnosis (CAAD) algorithm for breast cancer detection.
- The dataset is preprocessed by handling missing values, normalizing numerical features, and encoding categorical features.
- The dataset is split into training and testing sets, and feature scaling is applied to ensure that the data is suitable for machine learning algorithms.
- The project applies 11 different machine learning algorithms, including logistic regression, decision trees, random forests, and support vector machines, to the preprocessed data.
- The performance of each model is evaluated using cross-validation, and the best-performing model is selected based on its recall, precision, accuracy, and F1 score.
- The voting classifier emerged as the most accurate model among the 11 classifiers, with an accuracy of 97.2%.

RESULTS

The breast cancer prediction project uses a machine learning model to accurately predict whether a tumor is malignant or benign. It uses a dataset of 569 samples, performs data cleaning, preprocessing, and feature scaling, and applies several machine learning algorithms. The best-performing model is selected based on recall, precision, accuracy, and F1-score and implemented in a web application for public access. The project's value proposition is to provide a reliable and accessible tool for medical professionals, patients, researchers, and healthcare organizations.

https://drive.google.com/file/d/1RiZZk8LQkLQKl4Z82G_Gr9EOJkmP9GoI/view?usp=sharing