CAPSTONE PROJECT

Breast Cancer Prediction Using KNN

Final Project

Submitted by

SUSITH K

711721244058

III Btech CSBS

KGISL Institute of Technology

PROJECT TITLE

Breast Cancer Prediction Using KNN aims to develop a machine learning model to predict breast cancer : A Case Study on Breast Cancer

AGENDA

1. Problem statement
2. Project Overview
3. End Users
4. Solution and Value Proposition
5. The Wow Factor in Your Solution
6. Modelling
7. Result

PROBLEM STATEMENT

- Breast cancer is a prevalent disease, and early detection is crucial for improving treatment outcomes and survival rates.
- Utilizing machine learning techniques, the project aims to develop a model capable of accurately predicting breast cancer based on patient data and tumor characteristics.
- The model will rely on non-invasive tests such as imaging scans and biopsies to gather data, reducing the need for invasive procedures and improving patient comfort.
- By creating a reliable tool for early detection and diagnosis, the project seeks to enhance patient care by facilitating timely interventions and potentially reducing the burden of the disease.

PROJECT OVERVIEW

- The "Breast Cancer Prediction Using KNN" project aims to develop a machine-learning model for the early detection of breast cancer. It involves analyzing patient data and tumor characteristics to predict cancer risk. By leveraging non-invasive tests and advanced analytics, the project seeks to improve patient outcomes through timely intervention and diagnosis.
- Handling missing values, normalizing data, and encoding categorical variables.
- Visualizing the data to understand the relationship between different features and the target variable.
- Implementing various machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, and SVM to predict whether a tumor is malignant or benign.
- Evaluating the performance of each model using metrics such as accuracy, precision, recall, and F1 score.
- Optimizing the hyperparameters of the best-performing model using Grid Search and Cross-Validation.

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

- Our model provides accurate predictions, aiding healthcare professionals in early diagnosis and treatment planning.
- By relying on non-invasive tests like imaging scans, our solution minimizes patient discomfort and eliminates the need for invasive procedures.
- Early detection facilitated by our model enables timely interventions, potentially improving patient outcomes and survival rates.
- Our solution streamlines the diagnostic process, saving time and resources for healthcare providers while enhancing patient care.
- By contributing to early cancer detection, our model has the potential to reduce the burden of the disease on patients, families, and healthcare systems, ultimately saving lives.

THE WOW IN YOUR SOLUTION

- Our model achieves remarkable accuracy while maintaining a user-friendly interface, making it accessible to healthcare professionals of varying expertise levels.
- With its rapid processing capabilities, our solution provides instant predictions, empowering healthcare providers to make informed decisions promptly.
- Our model is designed to adapt to evolving datasets and healthcare environments, ensuring its relevance and effectiveness in diverse clinical settings
- 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

MODELLING

- Begin by thoroughly exploring and understanding the dataset, identifying key features and potential patterns related to breast cancer diagnosis.
- Employ techniques such as correlation analysis, feature importance ranking, or dimensionality reduction to identify the most relevant features for breast cancer prediction.
- Evaluate various machine learning algorithms suitable for classification tasks, including logistic regression, decision trees, random forests, support vector machines, and neural networks.
- Train multiple models on the dataset, utilizing techniques such as cross-validation to ensure robustness and prevent overfitting.
- Fine-tune the hyperparameters of selected models to optimize performance, employing methods like grid search, random search, or Bayesian optimization.
- Assess the performance of each model using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).
- Validate the final model(s) using an independent test dataset and interpret the results to ensure the model's reliability and generalizability.

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.

Demo Link