Breast Cancer Prediction App

Submitted By:

Simran Kachle

14 June 2023

(Task 3)

1. INTRODUCTION:

The Breast Cancer Prediction App is a groundbreaking application that aims to revolutionize breast cancer detection and contribute to early intervention and improved patient outcomes. Developed using the Django framework and employing logistic regression, the app provides users with a simple and accessible platform to assess their risk of breast cancer. By leveraging the inbuilt breast cancer dataset from scikit-learn, the app delivers accurate predictions based on user input.

2. PROBLEM STATEMENT:

Breast cancer is a major health concern for women worldwide, with early detection being critical for effective treatment and improved patient outcomes. However, the traditional screening approach of testing all women is not cost-effective, and it can result in unnecessary interventions for low risk patients. Hence, there is a need for a personalized approach to breast cancer screening that can accurately identify women at high risk of developing the disease. The aim of this project is to develop app that can predict the risk of breast cancer in women based on various risk factors, such as age, family history, reproductive history, and lifestyle factors. The model used in app will be validated using a large, independent dataset to ensure its accuracy and generalizability. This will help to ensure that the model can be used effectively in different healthcare settings and populations. Ultimately, the goal of this project is to provide healthcare providers with a reliable and effective tool to identify women who may benefit from additional screening or preventive measures, leading to earlier detection and better outcomes for breast cancer patients.

3. MARKET NEED ASSESSMENT:

- Determine target audience: Identify the demographic and geographic factors that are most likely to use the app such as age, gender, income, education level, and geographic location.

- Conduct market research: Gather data from potential users through surveys, focus groups, or other research methods. Inquire about their knowledge of breast cancer risk factors, interest in using a risk prediction app, and preferences for features or functionalities.

- Evaluate competitors: Analyze the strengths and weaknesses of existing breast cancer risk prediction apps or tools and find ways to differentiate your app from the competition.

- Assess pricing and revenue potential: Determine the pricing and revenue models that will be used, such as fees for app access, collaborations with healthcare providers or insurers, or licensing agreements with research institutions.

4. TARGET SPECIFICATIONS AND CHARACTERIZATION:

Target Audience:

The target audience for breast cancer risk prediction app includes women who are interested in knowing their risk of developing breast cancer, as well as healthcare providers who use risk prediction app to make clinical decisions regarding breast cancer screening and prevention.

- Data Collection Models:
  There are several models for collecting data for breast cancer risk prediction. These include:

  - Case-control studies: These studies compare women who have been diagnosed with breast cancer to women who have not. Data is collected on various risk factors for breast cancer, and this data can be used to develop risk prediction models.

  - Family history assessment tools: These tools collect information on the number and ages of relatives with breast cancer, as well as other factors that may be related to breast cancer risk, such as the age at which relatives were diagnosed.

5. EXTERNAL SEARCH (online information sources):

Breast Cancer Risk Assessment Tool

Reduce Your Risk | Breast Cancer UK

Breast Cancer: Symptoms, Stages, Types, and More (healthline.com)

6. BENCH MARKING ALTERNATE PRODUCTS:

- A 2019 study published in the Journal of the National Cancer Institute compared four risk assessment tools: the Gail model, the Breast Cancer Surveillance Consortium model, the International Breast Cancer Intervention Study model, and the Tyrer-Cuzick model. The study found that the Tyrer-Cuzick Model had the highest sensitivity for predicting breast cancer risk in high-risk women.

- A 2015 study published in the Journal of Clinical Oncology compared the Myriad myRisk Hereditary Cancer Test to traditional genetic testing for BRCA1 and BRCA2 mutations. The study found that the myRisk test had higher sensitivity and specificity for detecting these mutations.

- A 2008 study published in the Journal of the American Medical Association compared the Breast Cancer Risk Assessment Tool to other models, including the Gail Model, the Rosner-Colditz Model, and the CARE Model. The study found that the Breast Cancer Risk Assessment Tool was more accurate than the other models for certain populations.

7. APPLICABLE CONSTRAINTS:

- Need high-quality data on clinical and lifestyle factors.
- Data may be incomplete, inconsistent, or difficult to obtain.
- Model performance may be affected by overfitting, missing data, and bias.
- Collecting and storing sensitive health data requires careful attention to privacy and security considerations.
- Need data encryption, user authentication, and compliance with regulatory requirements.

8. BUSINESS MODEL (Monetization Idea) :

A business model for an application of breast cancer prediction could involve a combination of revenue streams, such as:

- Paid downloads or subscriptions: Users could pay to download the app or subscribe to use it on a recurring basis.

- In-app purchases: The app could offer additional features or services for purchase within the app, such as personalized risk assessments or recommendations for screening and prevention.

- Advertising: The app could display targeted advertising based on user demographics and health profiles.

- Partnerships: The app could form partnerships with healthcare providers, insurance companies, or other stakeholders to offer value-added services or generate referral revenue.

- Data licensing: The app could license user data to researchers, healthcare organizations, or other third-party entities for research or commercial purposes.

- The business model should not compromise the privacy and security of user data or conflict with ethical considerations. The app should also provide value to users and contribute to improving breast cancer prevention and management.

9. CONCEPT GENERATION:

Breast cancer prediction application is an idea that emerged due to the need for improved early detection and prevention of breast cancer. Breast cancer affects millions of women worldwide and is the leading cause of cancer related deaths among women. Traditional breast cancer screening and prevention programs have limitations, and many women at high risk for breast cancer are not identified until the disease has progressed. With the use of machine learning and artificial intelligence, applications can provide personalized risk assessments and recommendations for prevention. These applications can improve early detection and prevention of breast cancer and provide a more patient-centered and collaborative approach to breast cancer management.

10. CONCEPT DEVELOPMENT:

The application can use advanced machine learning algorithms and artificial intelligence to provide a more accurate and personalized risk assessment. The application can analyze large datasets to identify patterns and factors that can contribute to the development of breast cancer, thereby improving the accuracy of the risk assessment. The application may also use cloud services such as Amazon Web Services (AWS) or Microsoft Azure for storage and computation of user data.

11. PROTOTYPE DEVELOPMENT:

GITHUB LINK: https://github.com/SimranKachle/FeyNN-Labs-Final-Task

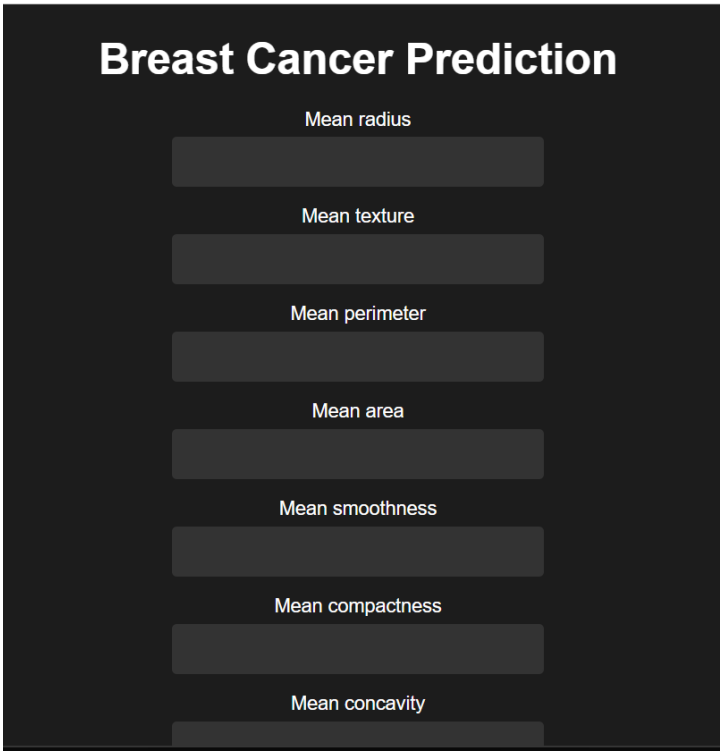
12. Product Details:

12.1. How the Breast Cancer Prediction App Works

The Breast Cancer Prediction App is designed to provide users with a straightforward and accurate assessment of their breast cancer risk. By leveraging machine learning techniques, specifically logistic regression, the app offers a user-friendly platform for users to input their relevant information and receive an estimation of their likelihood of having breast cancer. App operates in following manner:

12.1.1. Data Collection:

Users are prompted to provide important information such as previous medical conditions, and results from breast cancer screening tests. This data is essential in evaluating the individual's risk profile.



12.1.2. Preprocessing and Feature Extraction:

Once the user submits their information, the app performs necessary data preprocessing steps. This includes ensuring data quality, handling any missing values, and transforming the data into a format suitable for analysis. Relevant features are extracted from the input data to facilitate accurate predictions.

12.1.3. Logistic Regression Analysis:

The preprocessed data is then fed into a logistic regression model. Logistic regression is a widely-used machine learning algorithm that is suitable for binary classification tasks, such as predicting the likelihood of breast cancer (positive or negative). The model is trained using a pre-existing breast cancer dataset to learn patterns and make accurate predictions.

12.1.4. Model Training and Validation:

Prior to generating predictions, the logistic regression model undergoes training on the available breast cancer dataset. This training process allows the model to learn from known cases and understand the relationships between the input features and the likelihood of breast cancer. The trained model is then validated to assess its performance and reliability.

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2)

model = LogisticRegression()
model.fit(X_train, Y_train)
```

12.1.5. Prediction Generation:

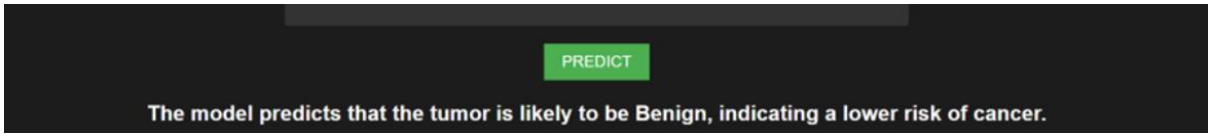
With a trained and validated logistic regression model, the app utilizes this model to generate predictions based on the user's input. The model analyzes the provided data, evaluates relevant risk factors, and provides an estimation of the individual's likelihood of having breast cancer.

```
pred = model.predict(np.array([[var1, var2, var3, var4, var5, var6, var7, var8, var9, var10,
                                var11, var12, var13, var14, var15, var16, var17, var18, var19, var20,
                                var21, var22, var23, var24, var25, var26, var27, var28, var29, var30]]))


result = ""
if pred == 0:
    result = "The model predicts that the tumor is likely to be Benign, indicating a lower risk of cancer."
else:
    result = "The model predicts that the tumor is likely to be Malignant, indicating a higher risk of cancer
```

12.1.6. Presentation of Results:

The app presents the prediction results in a clear and understandable manner to the user. The user is informed of their risk category, such as low or high risk.



12.2. Algorithm:

The Breast Cancer Prediction App utilizes the logistic regression algorithm as its core machine learning technique for predicting breast cancer. Logistic regression is a commonly used supervised learning algorithm that is particularly well-suited for binary classification tasks.In the context of the app, logistic regression is applied to analyze the input data provided by the user and estimate the probability of the individual having breast cancer. The algorithm models the relationship between the input features and the likelihood of breast cancer. During the development of the app, a logistic regression model was trained using a labeled dataset. This dataset contained a diverse set of breast cancer cases, with known outcomes. The model was trained to learn patterns and

associations between the input features and the presence or absence of breast cancer. Once the logistic regression model is trained, it can be used to make predictions on new, unseen cases. When a user inputs their information into the app, the logistic regression model applies the learned relationships to estimate the probability of breast cancer based on the provided data.

12.3. Frameworks and Software:

The development of the Breast Cancer Prediction App involved the utilization of various frameworks and software tools to ensure efficient and effective implementation. The following frameworks and software were employed in the app's development process:

12.3.1. Django Framework:

The app's web application was built using the Django framework, a powerful Python-based framework for web development. Django offers a robust set of tools and features that streamline the development process, including handling user input, data processing, and rendering the prediction results. It provides a scalable and maintainable architecture for building web applications with ease.

12.3.2. scikit-learn Library:

The scikit-learn library, a popular machine learning library in Python, played a vital role in the implementation of the breast cancer prediction functionality. It provided the necessary machine learning algorithms, including logistic regression, as well as utilities for data preprocessing, feature extraction, and model training. The well-documented API and extensive set of functionalities offered by scikit-learn simplified the implementation of the machine learning components in the app.

12.3.3. Python Programming Language: Python, a versatile and widely-used programming language, served as the primary language for developing the app. Its rich ecosystem of libraries and frameworks, such as Django and scikit-learn, made it an ideal choice for implementing the various functionalities of the app. Python's readability, ease of use, and vast community support were instrumental in accelerating the development process.

12.3.4. HTML, CSS, and JavaScript: To create an intuitive and visually appealing user interface, standard web technologies such as HTML, CSS, and JavaScript were utilized. HTML provided the structure of the web pages, CSS was used for styling and layout, and JavaScript enabled dynamic interactions and client-side validation. These front-end technologies combined to create a seamless and user-friendly experience for app users.

13. Financial Modeling (equation)

- Market Identification:
- The breast cancer prediction app falls within the healthcare industry, specifically targeting the segment related to breast cancer detection and prevention. While the app does not have direct sales and pricing, we can adapt the concept of a financial model to assess its potential impact and revenue generation.

- Data Collection and Market Analysis:
- To create a financial model, it is important to gather data and statistics about the healthcare industry, breast cancer prevalence, early detection rates, and the potential market size. Reliable data sources include research papers, healthcare organizations, government reports, and market research firms.

- Forecasting and Predictions:
- Although the breast cancer prediction app does not directly generate sales, we can focus on forecasting user adoption and potential revenue generation. This can be accomplished through regression models or time series forecasting, leveraging historical data or industry benchmarks.

- Financial Equation Design:

To estimate the revenue potential of the app, we can design a financial model based on factors such as user adoption, monetization strategies, and the potential market size. While traditional linear or exponential equations may not be directly applicable in this context, we can use the following equation:

Total Revenue = Number of Users * Revenue per User

In this equation, the "Number of Users" represents the estimated number of users who would adopt the breast cancer prediction app. The "Revenue per User" component signifies the revenue generated per user through various monetization strategies, including in-app purchases, subscriptions, partnerships, or advertising.

14. Conclusion:

The creation of the Breast Cancer Prediction App represents a significant advancement in the realm of healthcare technology. By harnessing the power of machine learning and logistic regression, the app offers an efficient and user-friendly solution for early breast cancer detection. Its integration with the Django framework enables seamless functionality and a visually appealing interface. The app has the potential to empower individuals to proactively assess their risk, leading to timely medical intervention and increased chances of successful treatment. Moving forward, continuous improvement and expansion of the app's features, such as incorporating additional datasets and collaborating with healthcare professionals, will further enhance its accuracy and impact. The Breast Cancer Prediction App stands as a valuable tool in the fight against breast cancer, underscoring the importance of technology in promoting early detection and saving lives.

Github Link: SimranKachle/FeyNN-Labs-Final-Task (github.com)