Breast Cancer Prediction

Project Report

Sri Lanka Institute of Information Technology

IT3051 – Fundamentals of Data Mining

IT21183690	Danawardana H.M.I.K.
IT21380914	Jayawardena K.M.S.P.
IT21208294	Mudalige T.N.
IT21387562	Ekanayake E.M.A.M

Date of Submission:

22/10/2023

Abstract

This project presents a comprehensive web application designed to analyze and predict breast cancer survival rates, focusing on fundamental principles of data mining. Unlike previous manual processing systems using paperwork, this computerized breast cancer survival prediction system streamlines operations related to data analysis and prediction. By using advanced data mining techniques, the system helps users to effectively manage and understand breast cancer survival rates.

The project is structured into different modules to create an integrated system: data pre-processing, predictive modeling, result interpretation, visualization, and reporting. Through these modules, the system facilitates essential tasks such as feature selection, algorithm execution, result interpretation, and insightful visualizations. Using the power of data mining, this app ensures accurate predictions. Healthcare professionals and researchers can make informed decisions about the likelihood of breast cancer survival. Developed using state-of-the-art technologies like Python, scikit-learn, and pandas, the system ensures seamless data processing and analysis. The web-based nature of the application allows users to access the system from anywhere, anytime, via the Internet. Ensures flexibility and convenience in performing predictive analytics. By using data mining techniques, this project aims to improve our understanding of breast cancer survival rates and provide valuable insights to the field of oncology research.

Declaration

This project report is our original work and the content is not plagiarized from any other resource. References for all the content taken from external resources are correctly cited. To the best of our knowledge, this report does not contain any material published or written by third parties, except as acknowledged in the text.

Registration Number	Name	Signature
IT21208294	Mudalige T.N.	Tharushi
IT21380914	Jayawardena K.M.S.P	
IT21387562	Ekanayake E.M.A.M	
IT21183690	Danawardana H.M.I.K.	

Date : 20/10/2023

Acknowledgments

We would like to express our sincere gratitude and appreciation to all those who contributed to the successful completion of our Breast Cancer Survival Model Building Project, conducted under the Fundamentals of Data Mining (IT3051) module. This endeavor was carried out using Python, and it would not have been possible without the support, guidance, and encouragement of numerous individuals and entities.

First and foremost, we would like to express our deepest appreciation to lecturer Dr. Amithalal Caldera, and course instructors Ms. Vajira Kavindi, and Ms. Supipi Karunathilaka for their invaluable guidance, expertise, and unwavering support throughout the project's development. Their continuous feedback and mentorship played a pivotal role in shaping the project's direction and ensuring its quality.

We extend our sincere thanks to our colleagues and peers for their collaborative efforts and insightful discussions that greatly contributed to the refinement of our ideas and the overall success of the project. Your dedication and teamwork have been instrumental in this achievement.

Furthermore, we are grateful to the Department of Computing for providing us with a conducive learning environment and the necessary resources to undertake this project. We would like to express our appreciation to the academic and non-academic staff who supported us throughout our academic journey.

Lastly, we express our sincere gratitude to everyone who played a part in the realization of this breast cancer survival model-building project. Your involvement and contributions

Table of Contents

Table of Figures

Introduction

Breast cancer is one of the most common forms of cancer affecting women worldwide. Effective therapy and positive patient outcomes depend on the early detection and accurate classification of patients with breast cancer. This initiative assumes special significance as we approach October, which is observed worldwide as Breast Cancer Awareness Month.

Our principal objective is to develop a data mining and machine learning solution that predicts breast cancer patients' chances of survival. The factors used in this predictive model include age, race, marital status, tumor stage, and more. We empower healthcare professionals to make data-driven decisions regarding patient care and treatment strategies by thoroughly examining these factors.

This project's significance expands well beyond the realm of data analysis. It provides medical professionals with an effective means to evaluate the results of patients and determine whether to continue or modify treatment in a trained manner. Our solution offers dynamic updates, allowing medical staff to adapt treatments based on evolving survival predictions.

It is important to remember that, despite the possibility that healthcare professionals would find our solution useful, it is not suggested that cancer patients use it themselves. Cancer treatment choices are complicated and necessitate the knowledge of skilled medical specialists. But for doctors and other medical professionals, the insights produced by our model can be a valuable resource for enhancing treatment strategies and enhancing patient outcomes.

During Breast Cancer Awareness Month, our project stands as a testament to our commitment to fighting breast cancer. Our effort serves as a symbol of our dedication to the battle against breast cancer during Breast Cancer Awareness Month. We explore our data-driven approach's methodology, outcomes, and implications in this project with the objective of strengthening healthcare professionals' decision-making skills and advancing the battle against breast cancer.

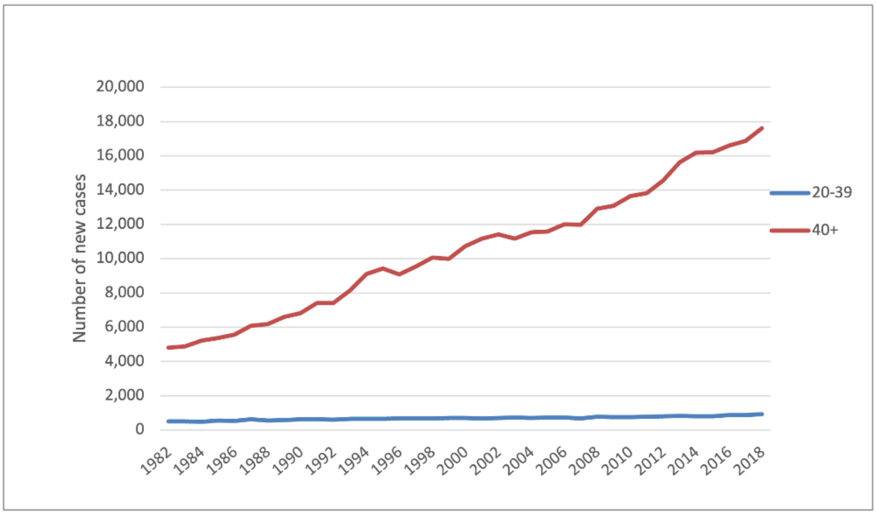


Figure 1 : Incidence of breast cancer in women, by age group, 1982 to 2018

Methodology

Dataset Description

This dataset comprises critical information about breast cancer patients, sourced from the November 2017 update of the SEER (Surveillance, Epidemiology, and End Results) Program by the National Cancer Institute (NCI). The dataset, which spans the years 2006 to 2010, focuses on female patients who have been diagnosed with infiltrating duct and lobular carcinoma breast cancer, as indicated by SEER primary source recode NOS histology codes 8522/3. The accuracy and usefulness of the data were ensured through a careful selection procedure. The dataset comprised patients who met the following requirements:

- Known tumor size.
- Examination of regional lymph nodes (LNs).
- Positive regional LNs.
- Survival months greater than or equal to 1 month.

The collection thus includes 4,024 patients, each of whom is represented by an extensive range of attributes. Among the key features are:

- **Age:** The age of the patient at the time of diagnosis.
- **Marital Status:** The patient's marital status, reflecting their social context.
- **Tumor Stage (T Stage):** The extent of tumor growth, categorized into stages T1, T2, T3, and T4.
- **Regional Lymph Node Stage (N Stage):** The severity of lymph node involvement, classified into stages N1, N2, and N3.
- **6th Edition Cancer Staging (6th Stage):** The cancer stage as per the 6th edition of cancer staging, encompassing stages IIA, IIB, IIIA, IIIB, and IIIC.
- **Histological Differentiation (Differentiate):** The degree of differentiation of cancer cells, categorized as poorly differentiated, moderately differentiated, well-differentiated, or undifferentiated.
- **Tumor Grade (Grade):** The aggressiveness of the tumor, graded from 1 (least aggressive) to 4 (most aggressive).
- **AJCC Stage (A Stage):** The cancer stage as per the American Joint Committee on Cancer (AJCC) criteria, distinguishing between regional and distant stages.
- **Tumor Size:** The size of the tumor, provides critical insights into disease progression.

To guarantee data quality and consistency, this dataset underwent comprehensive data preprocessing, cleaning, and integration. It is notable that categorical variables have been encoded for use in future analysis. The dataset serves as a valuable resource for data mining and machine learning projects aimed at predicting breast cancer patient outcomes and aiding healthcare professionals in optimizing treatment plans.

Model Selection

We started a thorough model selection process in the quest to create a robust and trustworthy predictive system for breast cancer patient outcomes. Our main goal was to determine which machine learning algorithms would work best for addressing the particular problems that this important healthcare domain presents.

Algorithm Exploration: To start building our model, we investigated a variety of machine learning algorithms, each of which had unique capabilities and advantages. These algorithms included:

1. **Artificial Neural Networks (ANN):** Leveraging the power of deep learning, ANNs are adept at capturing intricate patterns within complex datasets, making them an invaluable tool for predictive modeling.

2. **Random Forest:** Known for their ensemble learning prowess, Random Forests excel in handling high-dimensional data and mitigating overfitting, enhancing their suitability for healthcare predictions.

3. **Decision Trees:** Renowned for their interpretability, Decision Trees allowed us to construct clear and intuitive decision-making paths based on patient attributes.

4. **Logistic Regression:** A fundamental algorithm in classification tasks, Logistic Regression provided valuable insights into the likelihood of specific patient outcomes.

5. **Linear Regression:** This algorithm facilitated the examination of linear relationships between patient attributes and survival times, helping uncover key predictors.

Model Evaluation: We used a rigorous evaluation methodology and a variety of industry-accepted indicators to evaluate the performance of each algorithm.

- **Accuracy:** Gauging the overall correctness of predictions.

- **Precision:** Measuring the proportion of true positive predictions among all positive predictions.

- **Recall:** Quantifying the proportion of true positives correctly identified by the model.

- **F1-Score:** Balancing precision and recall to provide a harmonized performance measure.

- **R2-Score:** Assessing the goodness of fit of the model in capturing variance in survival times.
- **Silhouette Score (for Clustering):** Evaluating the quality of patient clusters in unsupervised learning.

Outcome Optimization: Our main objective was to find the method that best matched the complexity of the patient data for breast cancer, providing the maximum accuracy, reliability, and variance explanation. Healthcare providers could use the chosen model as a useful tool to make sensible decisions about patient care and treatment strategies.

Thorough experimentation, meticulous examination, and a dedication to providing a solution that could have a substantial impact on breast cancer treatment techniques were the hallmarks of the model selection process. With this methodical approach, we hoped to fully utilize machine learning to improve patient treatment and enhance our awareness of this challenging medical condition.

Decision Tree Regression Model for Predicting Breast Cancer Survival Percentage

A Decision Tree Regression model was methodically created and optimized in the endeavor to offer insightful information about the outcomes of breast cancer patients. This model's main objective is to forecast patient survival rates using an extensive record that includes important characteristics including age, tumor stage, lymph node stage, tumor differentiation, and more.

Data preprocessing

The process begins with data preprocessing, which is crucial for ensuring the accuracy and reliability of the data. In order to prepare the dataset for modeling, we handled a number of issues at this step, such as column renaming, data type conversion, and encoding of categorical features. We gave the model the ability to understand and interpret the data effectively by turning categorical variables into numerical representations.

Model Training

The Decision Tree Regression method is the brains behind this project. This model was picked because it could make the data's intricate, nonlinear relationships clear. The dataset was used to train the Decision Tree Regression model, which learned to approximatively predict the survival percentages of breast cancer patients with a predefined maximum depth of 5.

Model Evaluation

To assess the performance and dependability of the model, a thorough evaluation method is essential. To evaluate the effectiveness of the model, a number of crucial evaluation measures were used:

- **Training R-squared (R²) Score:** The training dataset was used to compute the R-squared score, which indicates the proportion of the variance in the survival percentages explained by the model. The value obtained was approximately 0.47, indicating a moderate fit to the training data.

- **Testing R-squared (R²) Score:** Extending the evaluation to the testing dataset, the model achieved a similar R-squared score of approximately 0.29. While the model demonstrates some predictive capability, there's room for improvement.

- **Mean Absolute Error (MAE):** The MAE, calculated as 0.16 for the testing dataset, measures the average absolute difference between actual and predicted survival percentages.

- **Mean Squared Error (MSE):** The MSE, which yielded a value of 0.09, quantifies the average squared difference between actual and predicted values.

- **Root Mean Squared Error (RMSE):** With an RMSE of approximately 0.30, this metric provides insight into the magnitude of errors present in the model's predictions.

- **Explained Variance Score (EVS):** The EVS, standing at around 0.29, indicates how well the model accounts for the variance in survival percentages.

Model performance Visualization

Graph 1 : This graph displays the observed vs. predicted status for the testing sample data.

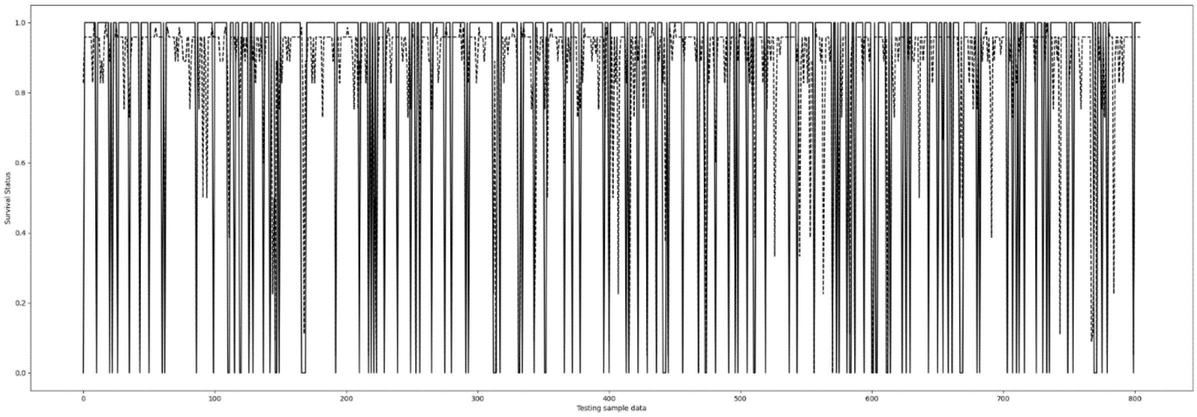
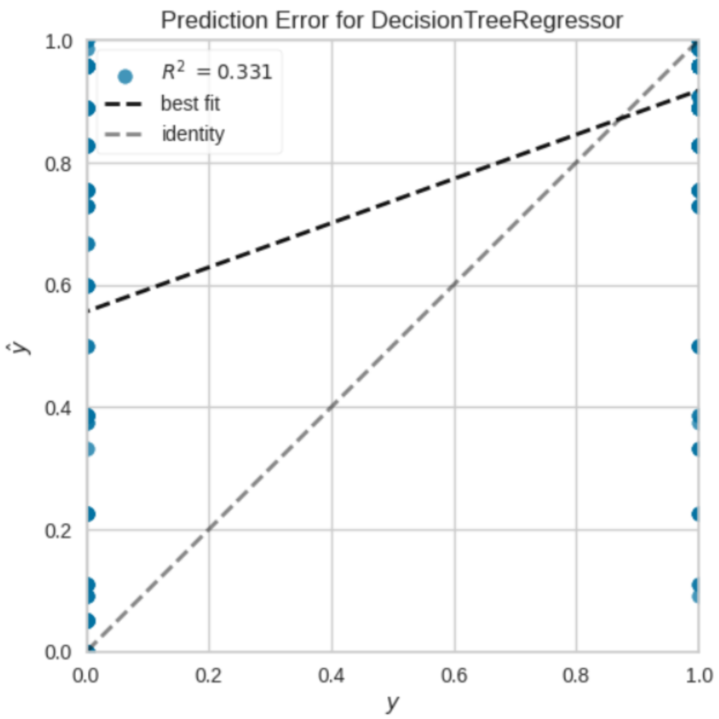


Figure 2 : Observed vs. Predicted Status

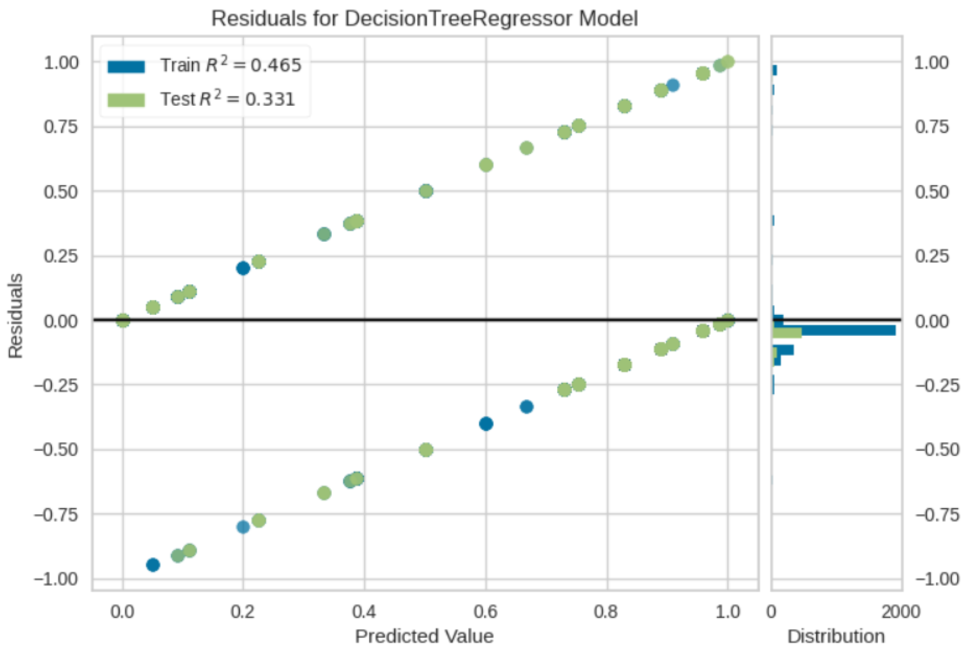
Graph 2 : This graph visualizes the prediction errors of the model using the Prediction Error visualizer.



```
<Axes: title={'center': 'Prediction Error for DecisionTreeRegressor'}, xlabel='$y$', ylabel='$\\hat{y}$'>
```

Figure 3 : Prediction Error

Graph 3 : The residual plot illustrates the distribution of residuals, providing insights into the model's performance.



```
<Axes: title={'center': 'Residuals for DecisionTreeRegressor Model'}, xlabel='Predicted Value', ylabel='Residuals'>
```

Figure 4 : Residual Plot

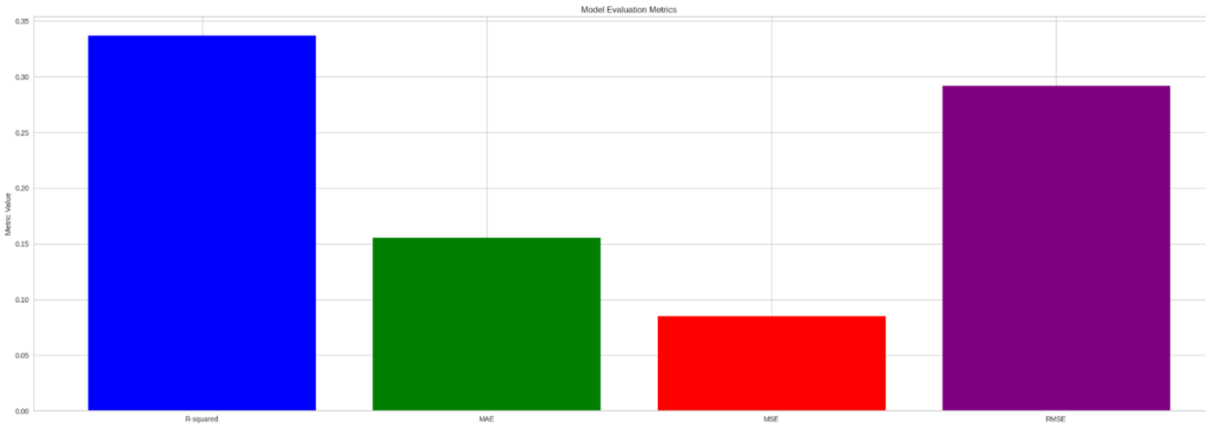


Figure 5 : Model Evaluation Metrics

Cross-Validation

K-fold cross-validation was used to increase the accuracy of the model's performance evaluation. A mean squared error of 0.08562 for the training dataset was found, showing the model's capacity to generalize to unseen data. The mean squared error for the testing dataset was 0.09354, indicating reliable performance.

Model Insights and Limitations

Although the Decision Tree Regression model exhibits promise in forecasting the percentage of breast cancer patients who survive, it is important to recognize that it has certain limitations. The model might not fully capture the nuances in the dataset, according to the comparatively low R-squared scores and the extent of errors (RMSE and MAE). Further research is necessary because factors that are not considered in the current model could have a major impact on patient outcomes.

Additionally, due to the model's simplicity, complex interactions between variables or non-linear relationships could be unaccounted for. Therefore, to improve predicted accuracy, future studies might investigate more complicated algorithms, feature engineering, and the integration of extra pertinent clinical data.

Conclusion

The Decision Tree Regression model is an excellent beginning in forecasting the percentage of breast cancer patients that survive. Although it offers helpful insights, there is an opportunity for greater improvement and refinement to enable healthcare-related decisions that are better informed. The effort to use cutting-edge machine learning methods and larger datasets to make more precise forecasts about the course of breast cancer continues.

Github : https://github.com/IT21387562/Decision-Tree-Regression-Model-

Random Forest Classification Model for Predicting Breast Cancer Survival Status

In the purpose of providing valuable insights into breast cancer patient outcomes, a random forest classification model was thoughtfully developed and refined. The primary objective of this model is to predict patient survival status based on an extensive dataset encompassing crucial attributes: age, race, marital status, T stage, N stage, 6th stage, differentiation, grade, A stage, tumor size, estrogen status, progesterone status, regional node examined, regional node positive, and survival months.

Data Preprocessing

Data preparation is the first step in the process and is crucial for ensuring the consistency and quality of the data. In order to prepare the dataset for modeling, a number of issues were resolved with at this step, including column renaming, data type conversion, and the encoding of categorical variables. Categorical variables were transformed into numerical representations, which gave the model the ability to comprehend and evaluate the data in a more efficient manner.

Model Training

The core of this endeavor lies in the Random Forest Classification algorithm. This model was chosen for its capability to handle complex, nonlinear relationships within the data. The Random Forest model was trained using the dataset to learn to predict the survival status of breast cancer patients, utilizing 600 decision trees in the ensemble.

Model Evaluation

A comprehensive evaluation process is essential to determine the model's performance and reliability. Several key evaluation metrics were employed to assess the model's efficacy:

- **Accuracy Score:** The model achieved an accuracy score of approximately 91.55%, indicating its ability to correctly predict the survival status of patients in the majority of cases.

- **Recall Score:** With a recall score of about 97.85%, the model excels at identifying patients who are actually alive, exhibiting a low rate of false negatives.

- **Jaccard Score:** The Jaccard score was approximately 90.96%, signifying the model's capability to accurately classify patients into the correct categories.

- **F1 Score:** The F1 score, around 95.26%, demonstrates a balance between accurate predictions and avoiding false negatives.

- **Precision Score:** The precision score was approximately 92.81%, indicating the model's ability to accurately predict the survival status of patients who are predicted to be alive.

Model Performance Visualization:

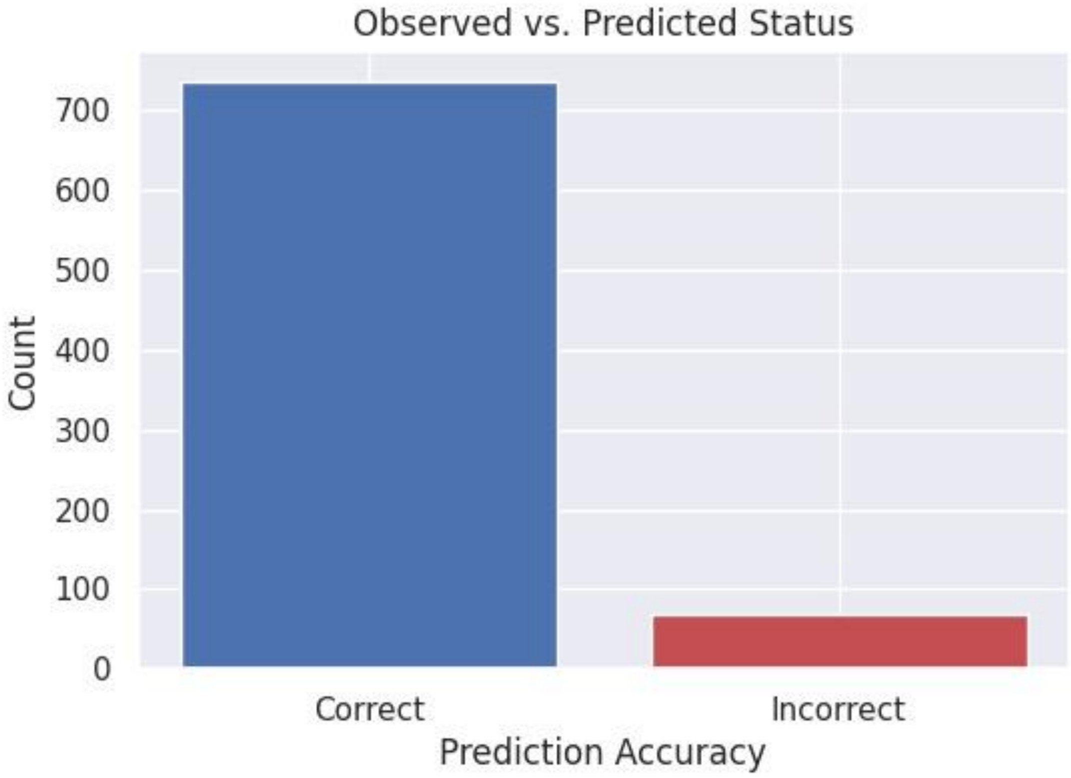
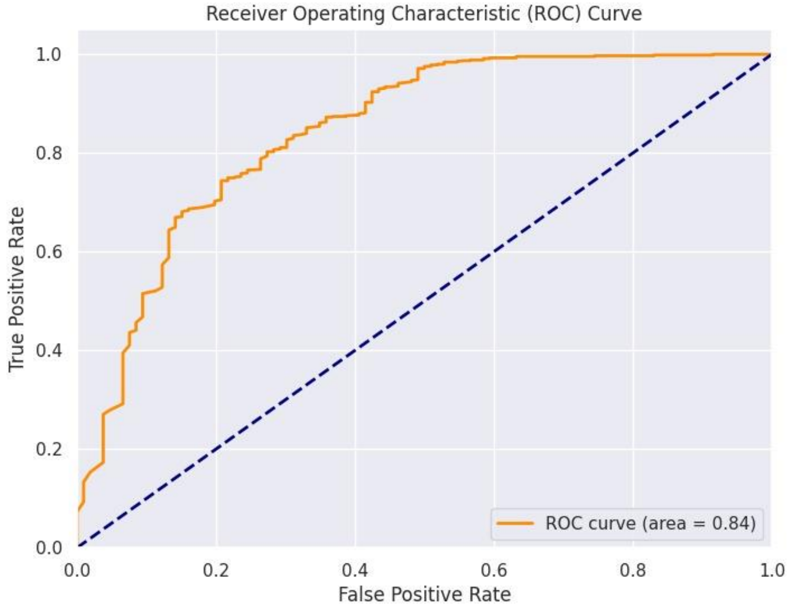


Figure 6 : Observed vs Predicted Status



Cross-Validation

To enhance the reliability of the model's performance assessment, k-fold cross-validation was applied. The average accuracy was approximately 90.48%, demonstrating the model's generalization performance.

Model Insights and Limitations

While the Random Forest Classification model holds promise in predicting breast cancer patient survival status, it is important to be aware of its limitations. This model may not account for all the factors that can significantly influence patient outcomes, warranting further investigation. To enhance its performance, one could explore strategies such as carefully selecting relevant features, fine-tuning hyperparameters, and addressing issues related to imbalanced data.

Conclusion

Random Forest Classifier-based breast cancer survival prediction model offers valuable insights for early diagnosis and treatment planning. However, it's important to acknowledge that transforming a binary outcome variable, such as "dead" and "alive," into a continuous one assumes a meaningful order between these categories. When dealing with binary outcomes lacking inherent order, alternative modeling approaches like classification may be more appropriate. Careful consideration of the data's nature and the chosen modeling approach is vital to ensure the analysis is both accurate and biologically meaningful.

Logistic Regression model for predicting breast Cancer Survival percentage

The logistic regression model was built with the aim of predicting cancer survival percentages. The key aspects such as the T stage, N stage, Estrogen level, etc... were used to come up with the predictions the process of building the model is summarized as follows:

Data Collection and Preprocessing.

- Data Collection: After research done to find the most suitable data set to come up with the model, we discovered a suitable data set from the Kaggle.com website.

- Preprocessing: During pre-processing removal of white space, encoding categorical variables, handling missing values and dropping unwanted columns was done.

Model building and Training.

- After the preprocessing the columns in the data set were split as independent variable and dependent variable after which the data was divided into training and testing set.

- The model was trained on training data, utilizing logistic regression function to predict the probability of survival.

Model evaluation.

The model was first evaluated using the testing data set.

Then a prediction was made.

Using both above-mentioned results the accuracy, f1 score, recall, precision, Jaccard score was calculated.

The following results were displayed at testing:

- **Accuracy** :0.9055
  This measures how well a model predicts both Death and Alive classes.

- **F1 score** :0.9464
  Is a combination of precision and recall especially useful when using an imbalanced real-life dataset.

- **Recall** :0.9810
  Calculates measure percentage of correctly predicted death instances.

- **Precision** :0.9142
  Calculates the percentage of correctly predicted Dead instances out of all dead instances.

- **Jaccard score** :0.8983
  Calculates how well the predicted values are related to the actual values.

- **K-fold cross validation**
  When using a value of 5 for the K value the mean accuracy of the model: 0.90.
  K-fold cross-validation helps ensure that the model's predictive capabilities are reliable and not overly influenced by the specific data split.

Model Performance Visualization

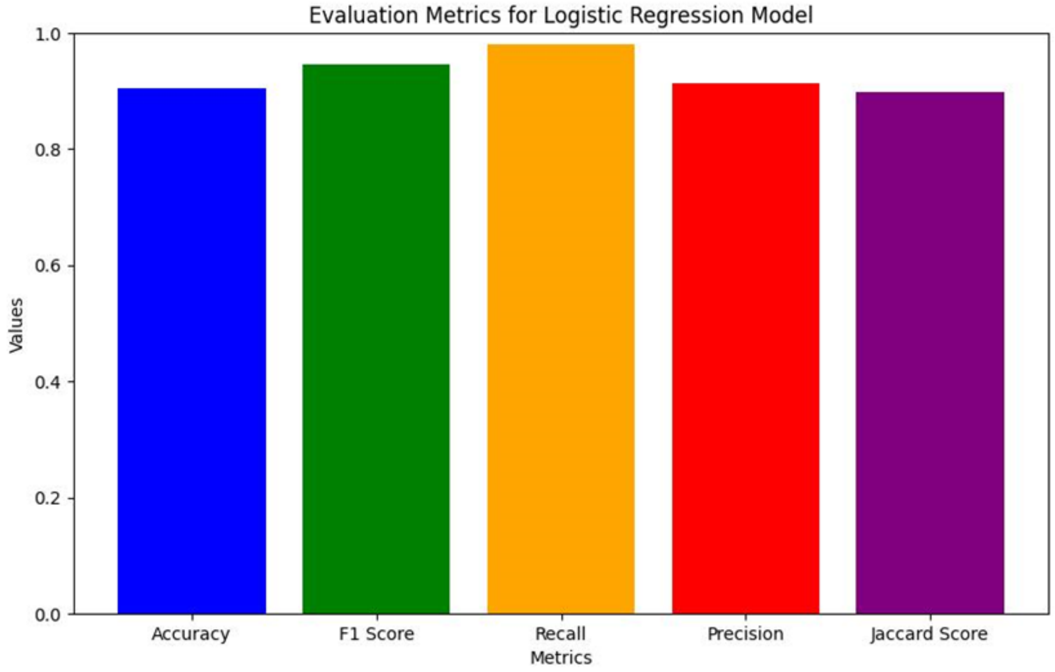


Figure 7 : Evaluation Metrics for Logistic Regression Model

Limitations of model

The logistic regression model has limitations due to factors like Linearity assumption, sensitivity for outliers and imbalanced data and so on how ever as displayed by the above scores the model matches the data set with higher accuracy.

Conclusion

The Logistic Regression model, while displaying notable advantages, also reveals its limitations when applied to our objective. Taking into consideration that logistic regression is typically used for classification tasks, and converting the predicted probabilities to percentage values would assume that the binary outcome corresponds to 0% for "dead" and 100% for "alive." This may lead to in accuracies of prediction therefore will not be the primary predictive tool to achieve our objective. However, it serves as a valuable starting point, shedding light on the intricacies of predicting breast cancer patient survival percentages.

<u>Artificial Neural Network Model for Predicting Breast Cancer Survival Percentage</u>

In tireless pursuit of the intricacies of breast cancer promotion, finely tuned artificial neural network (ANN) models have been systematically designed and refined. This advanced ANN model predicts patient survival rates for breast cancer patients using a dataset of key features like T stage, N stage, Estrogen level, etc....

Data Collection and Preprocessing

- Data Collection: After research done to find the most suitable data set to come up with the model, we discovered a suitable data set from the Kaggle.com website. 15

- Preprocessing: During pre-processing removal of white space, encoding categorical variables, handling missing values and dropping unwanted columns was done.

Model building and Training

Some columns in the preprocessed dataset were designated as independent and dependent variables. Following this split, the data was split into a training set and a test set. A model leveraging the power of artificial neural networks was trained using these datasets. Using complex algorithms, the network learned how to predict survival probabilities. This iterative process enabled the ANN to capture complex patterns, allowing accurate predictions and valuable insights into survival dynamics.

Model evaluation

The model was first evaluated using the testing data set. Then a prediction was made. Using both above-mentioned results the accuracy, f1 score, recall, precision, Jaccard score was calculated.

- **Accuracy (84.97%):** The model demonstrated a high level of accuracy, correctly predicting outcomes in 84.97% of cases. This metric signifies the overall correctness of the model's predictions.

- **F1 Score (91.62%):** The F1 score, a balanced measure of precision and recall, was impressive at 91.62%. This indicates that the model achieved a harmonious blend of accurate positive predictions and minimal false positives and false negatives.

- **Recall (97.78%):** With a recall score of 97.78%, the model excelled in identifying a vast majority of relevant instances. A high recall suggests the model's effectiveness in capturing almost all positive cases, minimizing false negatives.

- **Precision (86.18%):** The precision score of 86.18% showcases the model's ability to accurately predict positive instances. It indicates the proportion of correctly predicted positive observations among the instances the model labeled as positive.

- **Jaccard Score (84.53%):** The Jaccard score, measuring the similarity between predicted and actual sets of labels, stood at 84.53%. This metric signifies the model's effectiveness in classifying instances, considering the intersection over union of predicted and actual labels.

Model performance Visualization

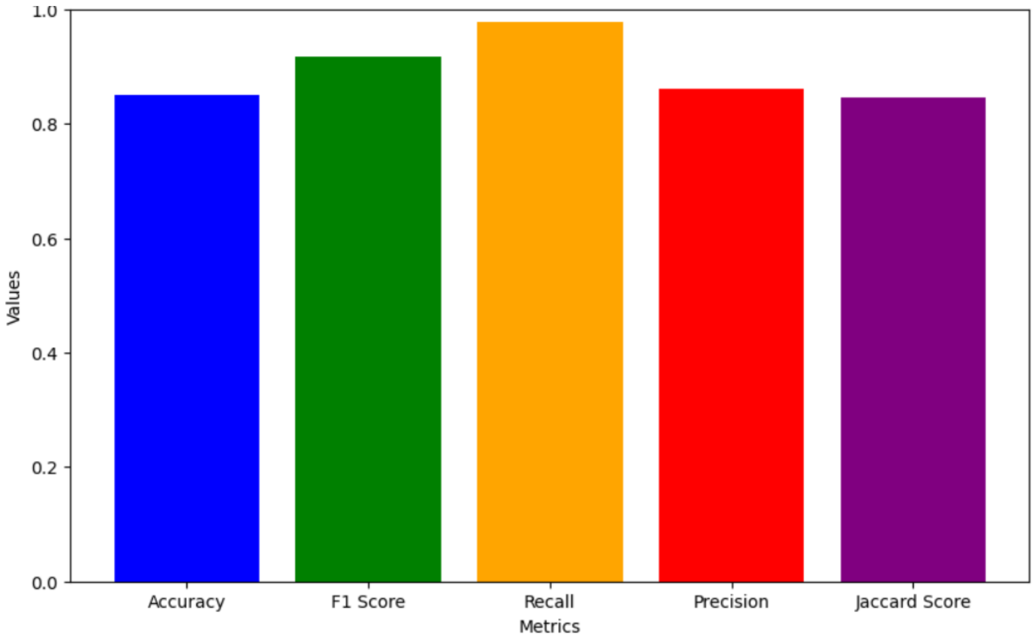


Figure 8 : Evaluation Metrics for ANN model

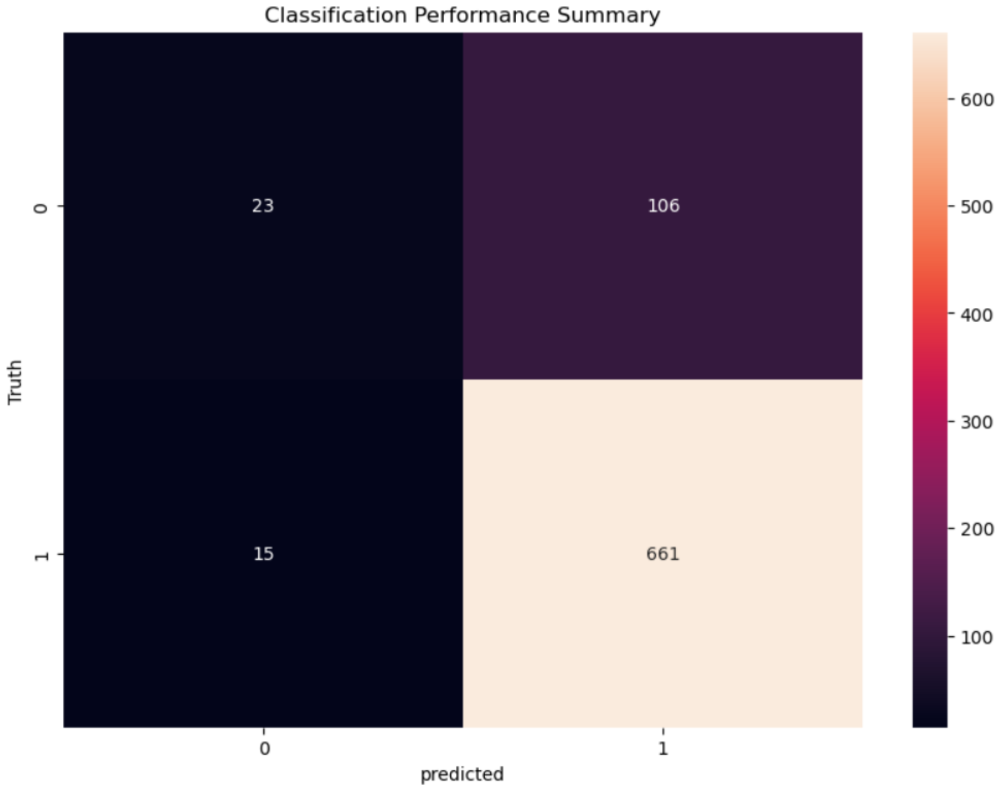


Figure 9 : Performance Summary

Conclusion

Artificial neural network (ANN) models have shown promising results in predicting survival of breast cancer patients. Its advanced computational capabilities enable accurate predictions and reveal complex patterns within datasets. However, like other models, ANNs have limitations, especially in terms of interpretability. While ANNs represent an important advance, further improvements are essential. Continued research and development efforts are essential to fully exploit the potential of this technology. As we move into the complexity of healthcare analytics, continued improvements will be needed to transform ANNs into reliable and essential predictive tools for understanding and predicting survival outcomes for breast cancer patients.

Web Development

Our Breast Cancer Survival Prediction project's web development aspect is essential to opening our predictive model to medical professionals and the public. To develop an intuitive and educational web application, we made use of several various technologies and frameworks. The core technologies employed in the development process include Flask, a micro web framework for Python, and standard web technologies such as HTML and CSS.

Flask Framework: We selected Flask as the backbone of our web application for several reasons. Flask is a very versatile and lightweight web framework, which makes it the perfect choice for our project. It allowed us to swiftly develop and deploy a web application with minimal overhead. Because of Flask's broad ecosystem of extensions and its ease of use, we were able to concentrate on the essential features of our application rather than getting bogged down in the complexities of web development.

Python: The main programming language used for our web development was Python. Python's ease of use and adaptability make it a great choice for projects involving both front-end and back-end development, as demonstrated by the Flask web framework. The large libraries and modules that Python offers are also helpful for managing data and integrating models.

HTML: The structure and content of our web pages were created using HTML (Hypertext Markup Language). The different components of the user interface, such as input forms, labels, buttons, and result displays, were defined using HTML. We made sure the online application displayed an orderly and understandable user interface using HTML, which made it easier for users to engage with it.

CSS: CSS (Cascading Style Sheets) was employed to enhance the visual appeal and user experience of our web application. We customized the styling of HTML elements using CSS to create a cohesive and aesthetically pleasing design. CSS allowed us to format text, control layout, set background colours, and apply animations to our web pages.

Development Workflow:

The workflow for our web development process was very organized. Using Flask and Python, we first developed the web application's essential features. This involved integrating the machine learning model for breast cancer survival prediction, designing the layout of input forms, and setting up pathways.

The online application's frontend was subsequently constructed using HTML, and we created an intuitive and simple-to-use user interface. HTML forms were implemented to gather user input, and result pages were created to display the prediction outcomes in a user-friendly manner.

We used CSS styling to make sure our online application was both visually appealing and simple to use. Determining colour palettes, fonts, and layout designs was part of this process to produce a visually appealing and engaging user interface.

In summary, the web development phase of our project was driven by the need to provide a user-friendly platform for breast cancer survival prediction. Flask, Python, HTML, and CSS collectively enabled us to deliver a web application that not only performs accurate predictions but also offers an excellent user experience for healthcare professionals and users interested in understanding breast cancer survival probabilities.

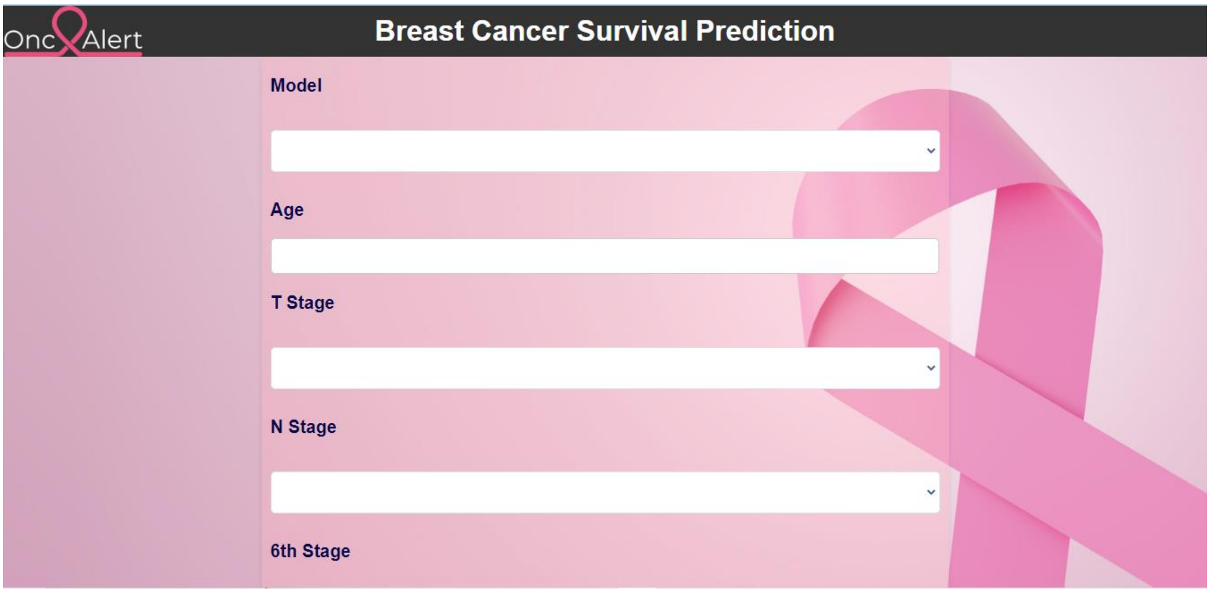
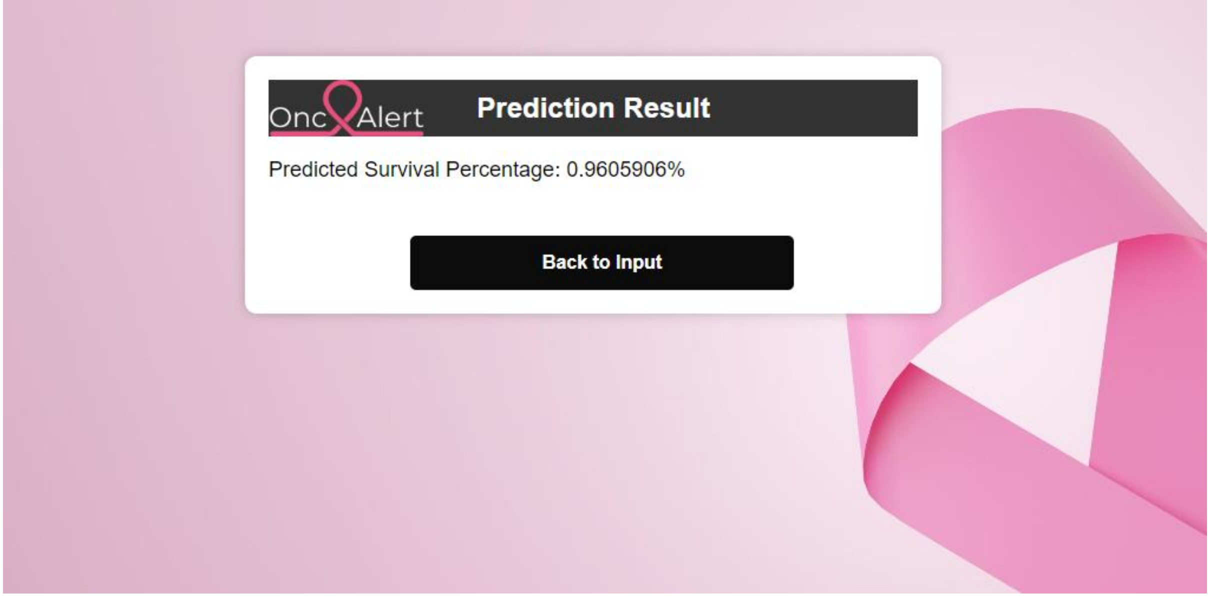


Figure 10 : Home Page of our web application



Prediction Result

Predicted Survival Percentage: 0.9605906%

Figure 11 : Results of our web application

Conclusion

The purpose of the project was to create a user-friendly tool to assist healthcare professionals in assessing breast cancer survival probabilities based on relevant medical features to help make informed decisions to help patients with their recovery.

Objectives Achieved

1. Selecting the most accurate model:
   - We decided to try four different models ANN, Decision Tree regressor, Random Forest and Logistic Regression and chose the models with high accuracy. Upon completion we realized that all models have accuracy above 80% therefore we decided to give the user the choice to select what model they would like to use.
2. Web Application development:
   - The models were developed using python and keras which was later saved as .h5 or. joblib files.
   - The application was developed in PyCharm using flask framework.
   - Using the above technologies, we have achieved a user-friendly web application.

Future Enhancements

While the project has indeed met its initial objectives, it is essential to acknowledge the intricate nature of predicting cancer survival percentages. We recognize that the complexity of an illness like cancer cannot be fully captured by assessing a predefined set of attributes from the dataset alone. However, we are committed to further elevating the precision and accuracy of our application through future enhancements such a :

- Continuous Model Improvement: Regularly updating the models with new patient data can enhance prediction accuracy over time.
- Visual Enhancements: Expanding the visualization capabilities of the web application with more interactive charts and graphs can improve the user experience and the understanding of prediction results.
- User Authentication: Implementing user authentication and access controls is crucial to protect sensitive patient data and ensure compliance with privacy regulations.

In conclusion, we envision our web application as a profound stride towards advancing the landscape of healthcare. Its potential impact extends beyond the realm of technology, as it promises to serve as an invaluable aid to both healthcare professionals and patients alike, empowering them to make enlightened decisions concerning their well-being.

References

[1] “Keras-Layers API,” [Online]. Available: https://keras.io/api/layers/.

[2] “Flask Documentation,” [Online]. Available: https://flask.palletsprojects.com/en/3.0.x/.

[3] “Kaggle-Breast Cancer Dataset,” [Online]. Available: https://www.kaggle.com/datasets/reihanenamdari/breast-cancer.

[4] “Keras-API Reference,” [Online]. Available: https://keras.io/api/models/.