**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

A.Y. 2023 - 24

**Course: Machine Learning**

**Project Report**

|  |  |  |
| --- | --- | --- |
| Program | BTECH AI | |
| Semester | Semester IV | |
| Name of the Project: | Breast cancer Prediction Model | |
|  | | |
| Details of Project Members |  |  |
| Batch | Roll No. | Name |
| 2 | I039 | Mann Agarwal |
| 2 | 1044 | Purav Talwar |
|  |  |  |
| Date of Submission: 4/3/24 | | |

**Contribution of each project Members:**

|  |  |  |
| --- | --- | --- |
| Roll No. | Name: | Contribution |
| I039 | Mann Agarwal | Coding, streamlit, PowerPoint Presentation and Report |
| I044 | Purav Talwar | Algorithm selection and coding of those Algorithms, PowerPoint Presentation |

**Github link of your project:**

[https://github.com/mann2705/Breast-Cancer-Prediction](https://github.com/mann2705/Breast-Cancer-Prediction" \t "_blank)

<https://github.com/purav12345678/Breast-Cancer-Predict-Analysis>

**Note:**

1. Create a readme file if you have multiple files
2. All files must be properly named (I004\_MLProject)
3. Submit all relevant files of your work
4. **Plagiarism is highly discouraged (Your report will be checked for plagiarism)**

**Rubrics for the Project evaluation:**

|  |
| --- |
| * Evaluation of project will be based on following rubrics * Domain knowledge and literature review in the selected topic (5 marks) * EDA, Implementation and performance metrics used (10 marks) * Beyond classroom knowledge gained and implemented (5 marks) |

**Project Report**

**Selected Topic**

**by**

**Mann Agarwal, I039**

**Purav Talwar, I044**

**Course: Machine Learning**

**AY: 2023-24**

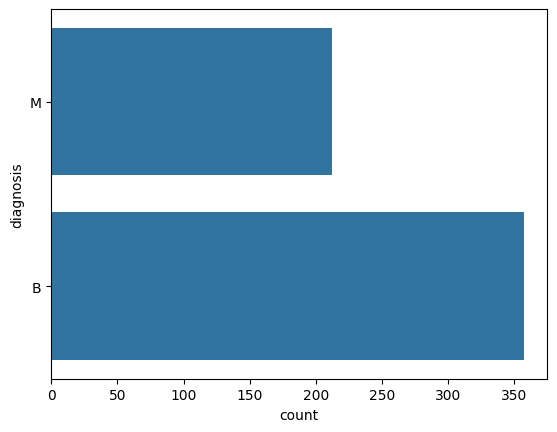
**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr no.** | **Topic** | **Page no.** |
| **1** | Storyline or Applications of Project |  |
| **2** | Literature Review |  |
| **3** | Data Preprocessing and Exploratory data Analysis with Visualization |  |
| **4** | Machine learning models with hyper parameter tuning |  |
| **5** | Performance Evaluation |  |
| **6** | Comparison of different techniques used |  |
| **7** | Deployment/GUI/ Learning beyond classroom |  |
| **8** | Learnings and challenges you faced while doing the Project |  |
| **9** | Conclusion |  |

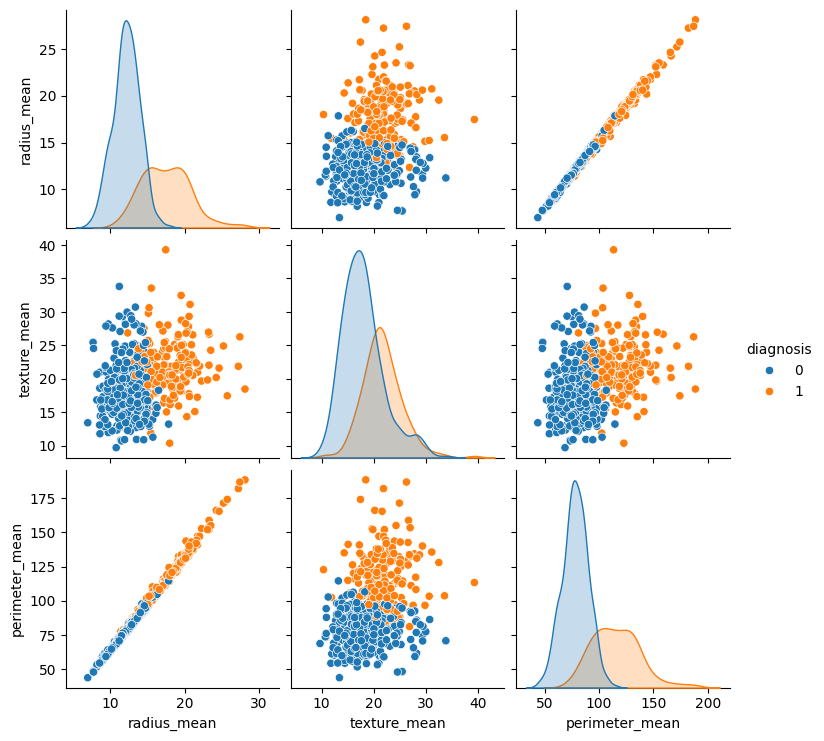
**Storyline or Applications of Project**

Our Project, Breast cancer prediction is used in the medical domain, which can be used by cancer specialists, and radiologists as a reference to predict whether a tumor is benign or malignant.

**Data Preprocessing and Exploratory data Analysis with Visualization**



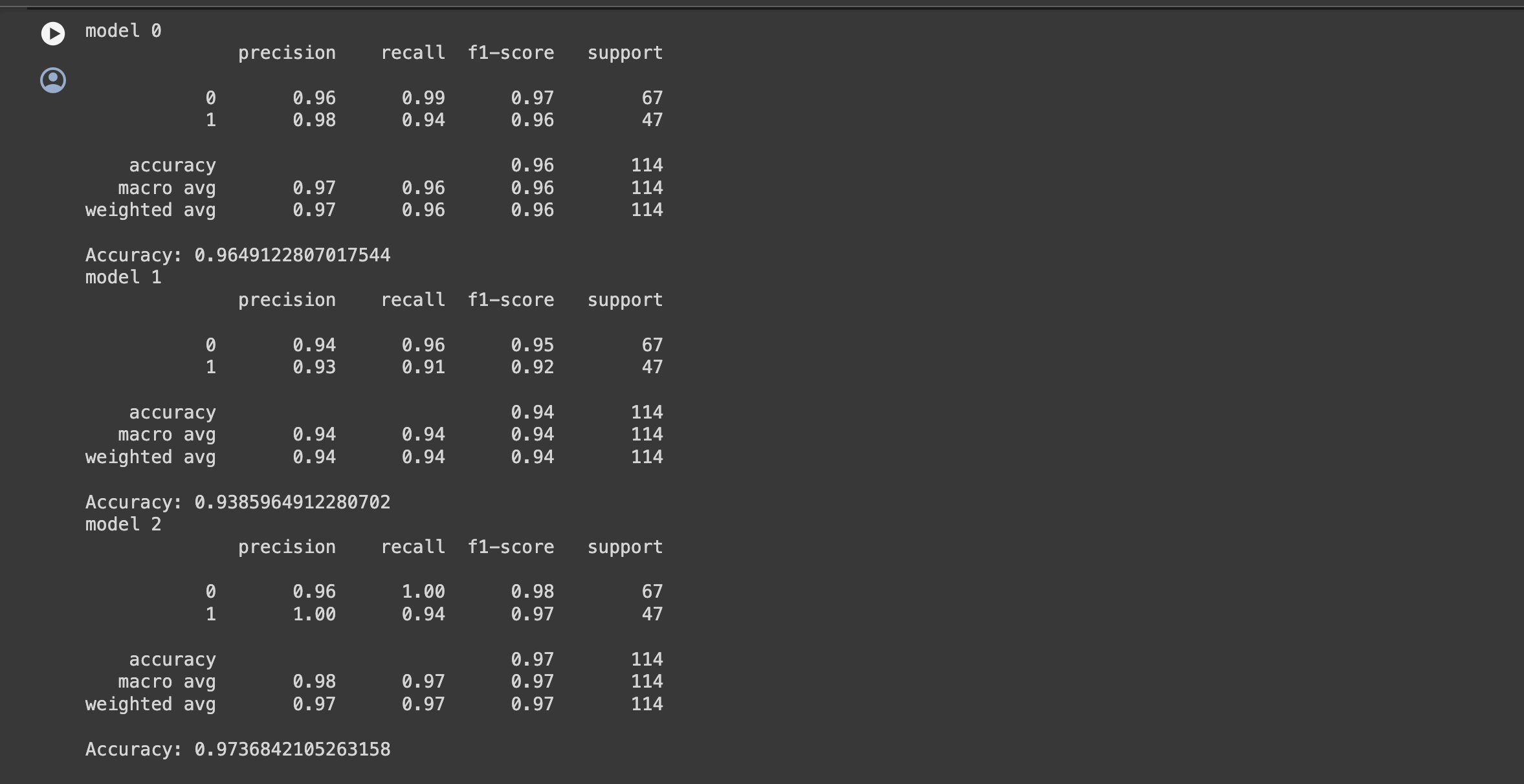
**Count of how many patients have benign or malignant tumor**



A colorful squares with white text

Description automatically generated with medium confidence

**Performance Evaluation**

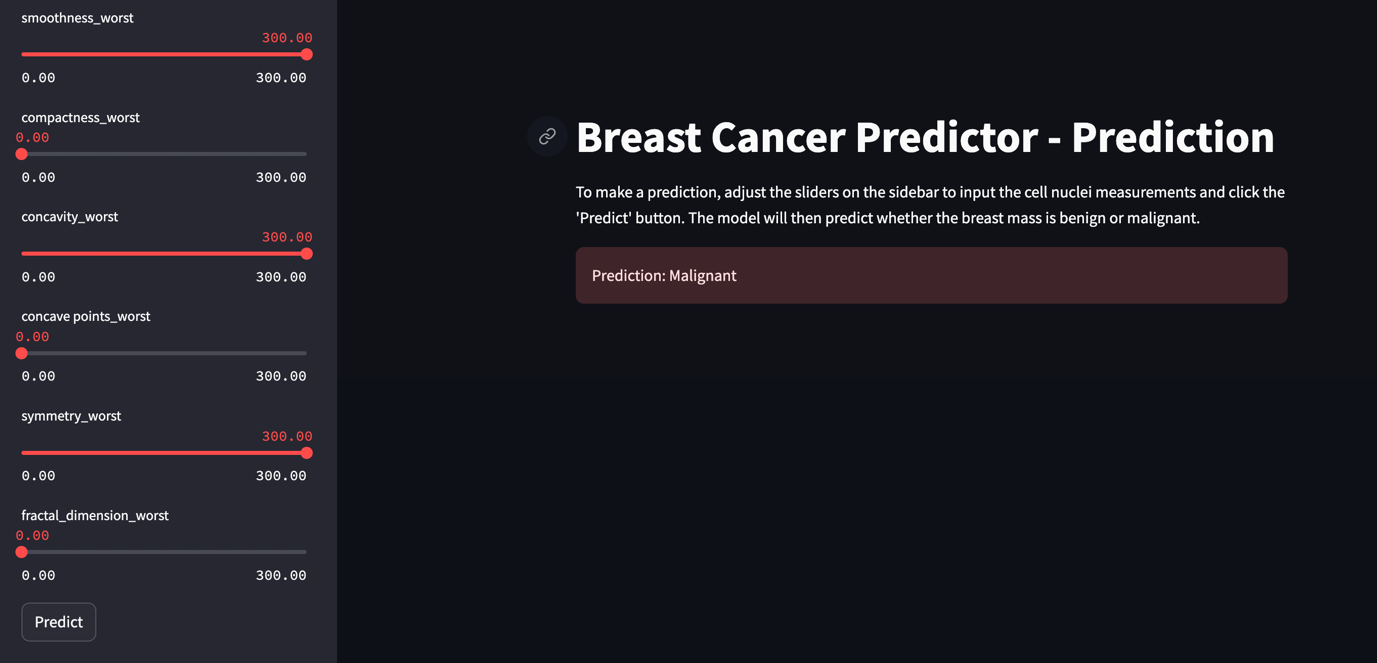
****

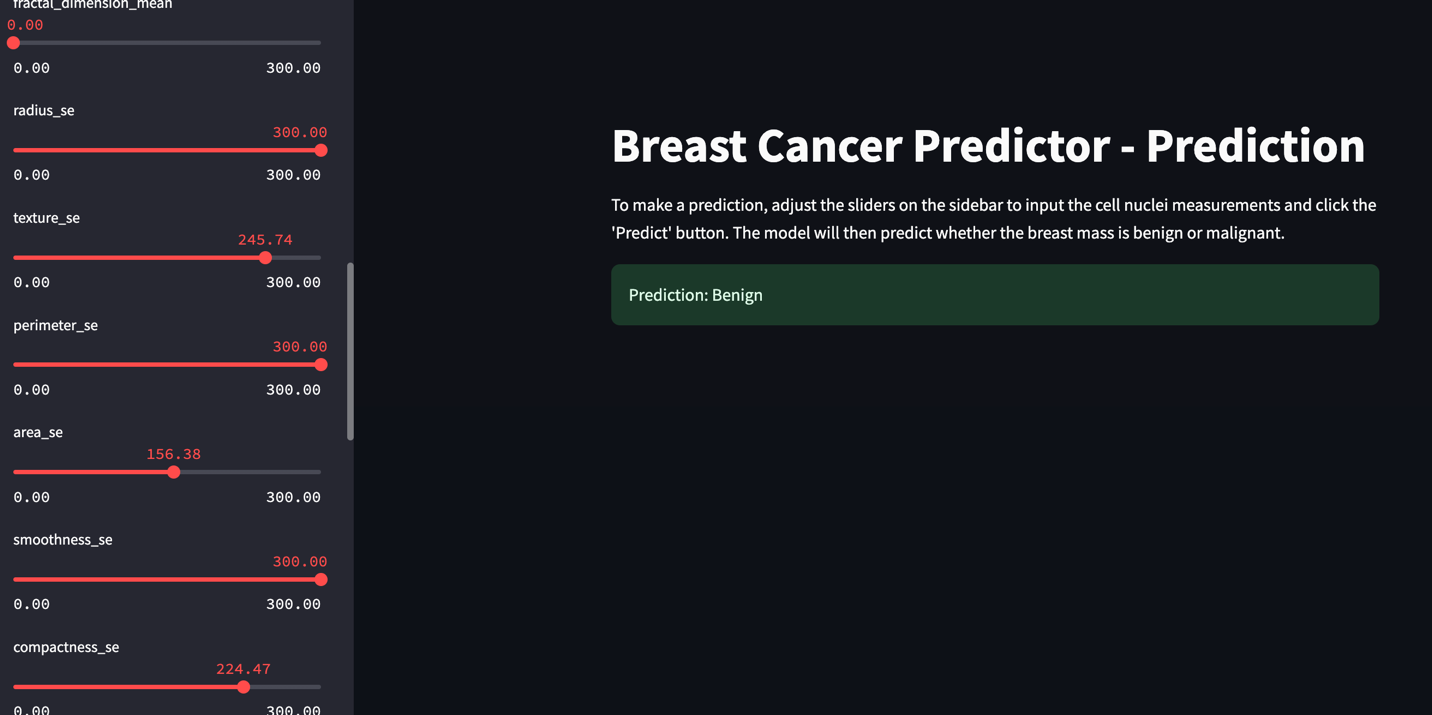
**Comparison of different techniques used**

| **Aspect** | **Decision Tree** | **Random Forest** | **Logistic Regression** |
| --- | --- | --- | --- |
| Algorithm Type | Non-linear decision tree | Ensemble of decision trees | Linear classification |
| Interpretability | Highly interpretable | Less interpretable (ensemble) | Interpretable (coefficients) |
| Performance | Prone to overfitting | Reduces overfitting | Good for linear relationships |
| Handling Non-linearity | Can model non-linearities | Handles non-linearity well | Assumes linear relationships |
| Training Time | Faster for small datasets | Slower due to ensemble | Faster than Random Forest |
| Missing Values | Can handle missing values | Can handle missing values | Requires handling before |

**Deployment/GUI/ Learning beyond classroom**

Streamlit has been used





**Learnings and challenges you faced while doing the Project**

:

**Challenges Faced:**

1. **Algorithm Integration:** One challenge was integrating the Random Forest and Decision Tree algorithms into the Streamlit app. It required careful handling of model initialization, training, and prediction within the app's structure.
2. **Data Preprocessing:** Ensuring proper data preprocessing for different algorithms was challenging, especially handling missing values and categorical variables appropriately.
3. **Performance Optimization:** Optimizing the app's performance, especially when dealing with large datasets or computationally intensive algorithms, required efficient coding practices and resource management.

**Conclusion**

The project provided valuable insights into the nuanced world of machine learning algorithms and their practical applications. Firstly, understanding the trade-offs between model interpretability and predictive performance was a significant takeaway. While Decision Trees offer transparency in decision-making processes, they can be prone to overfitting. On the other hand, ensemble methods like Random Forest strike a balance by aggregating multiple decision trees, resulting in improved generalization and robustness. This knowledge is pivotal when choosing the right algorithm based on specific project requirements, whether prioritizing interpretability for explanatory purposes or focusing on predictive accuracy.

Secondly, the experience of developing a Streamlit web application added a layer of practicality to the project. It highlighted the importance of creating user-friendly interfaces for showcasing machine learning models. From handling data preprocessing tasks such as missing value imputation and feature scaling to integrating multiple algorithms seamlessly within the app, the project emphasized the significance of efficient coding practices and resource management. Overall, the project's key takeaways encompassed algorithmic understanding, model interpretability versus performance considerations, and the practical aspects of deploying machine learning solutions in real-world applications through web interfaces.

Top of Form

Bottom of Form