Title : Aviva CancerCare

Breast Cancer Categorization

ISSUE / PROBLEM

Design a machine learning model that can classify breast cancer cells as benign or malignant based on the cell nuclei scan under microscope.

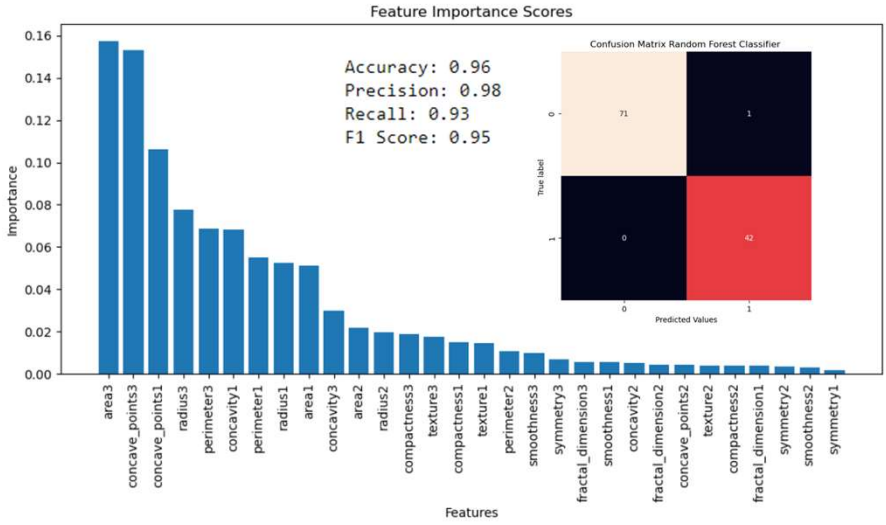
RESPONSE

- We analyze the data and identify the data distribution and correlation between data.
- We settle on what possible models could be good fit and test them.
- We tried Logistic Regression and Random Forest Classifier models.

IMPACT

- This model will allow the hospital to preemptively classify breast cancer as benign or malignant.
- This can help provide hospital to manage their resources properly and also allow patients to plan their healthcare accordingly.

Logistic Regression Model Performance Evaluation. We see high accuracy and high recall which are important for this project.



Random Forest Classifier model performance and corresponding feature importance.

KEY INSIGHTS

- Logistic regression model performs optimally. We can use it to classify the cancer cell category. Random Forest Classifier performs well too but not as good.
- Top three determining factors are “area3”, “concave_point3” and “concave_point1” based on random forest classifier.
- Training on larger dataset may give us a more robust model. We can check with **Aviva** data engineering team if a larger dataset is available.