

BREAST CANCER PREDICTION SYSTEM

By : Shruti Srivatsan

1. INTRODUCTION

1.1 OVERVIEW

This system has been developed for predicting the occurrence of breast cancer in women. It has been developed using AutoAI, an automatic end-end development tool by IBM. The UI has been developed using NodeRed. About 96% accuracy is obtained from the model.

1.2 PURPOSE

Breast Cancer is on the rise throughout the country. There is no cure yet for cancer other than early diagnosis. This system has been developed to detect early diagnosis of cancer after which the patient could take necessary actions. This will save a lot of lives.

2 LITERATURE SURVEY

2.1 EXISTING PROBLEM

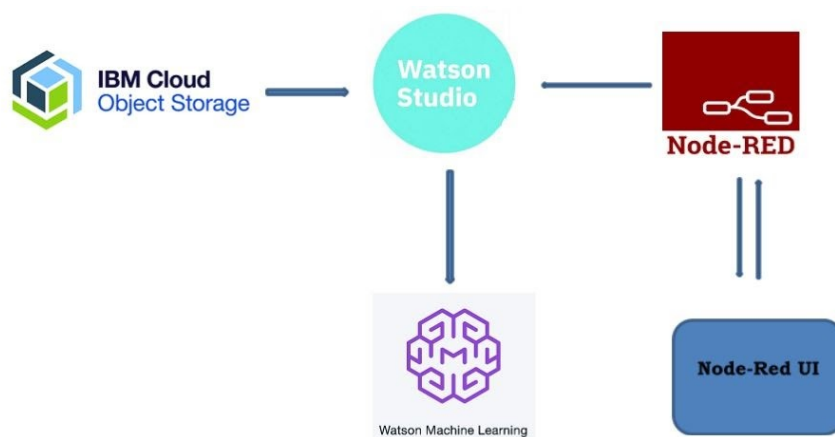
There is no efficient system for breast cancer prediction which provides accurate results. The results also seem to give a lot of false positive and false negative results which are extremely fatal in the medical field.

2.2 PROPOSED SOLUTION

The proposed solution is selected by the AutoAI feature of IBM studio. It has selected LightGBM as the best model in terms of accuracy. Without AutoAI, K Nearest Neighbours was chosen, which gave 95% accuracy. About 95% recall and precision are obtained which are really good for a medical domain application.

3 THEORITICAL ANALYSIS

3.1 BLOCK DIAGRAM



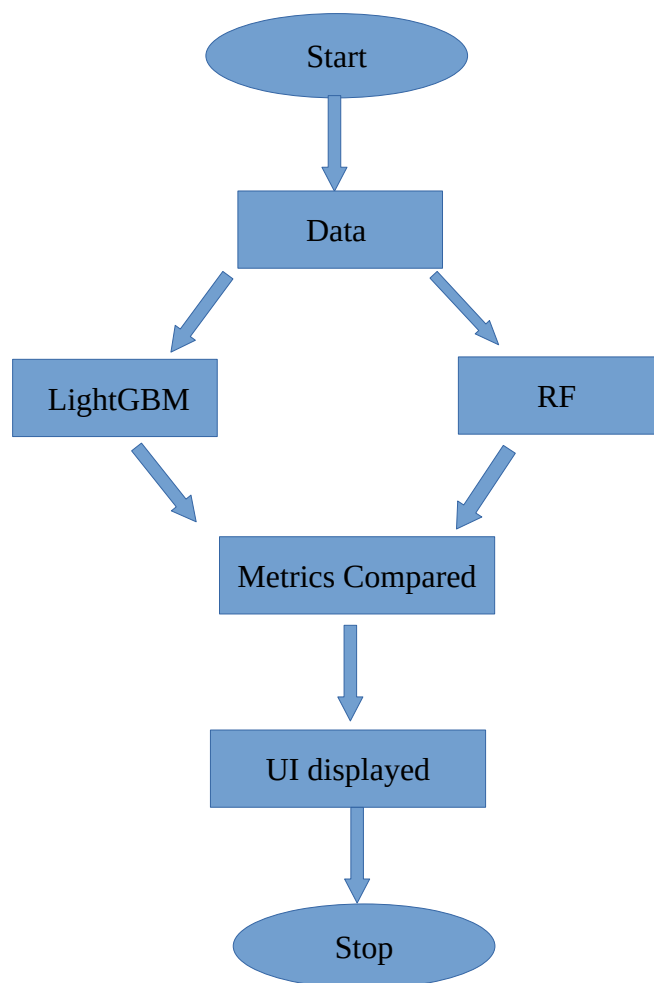
3.2 HARDWARE/SOFTWARE DESIGNING

The model was developed using AutoAI, an automated machine learning feature in IBM Watson Studio. User Interface was developed using Node Red . Deployment was done using Cloud storage provided by IBM.

4 EXPERIMENTAL INVESTIGATIONS

The model developed by the AutoAI feature uses all features fitted in LightGBM, which is the latest boosting model proven to work 7 times faster than other boosting techniques and Random Forest, a Bagging model. Hyperparameter tuning is done automatically along with preprocessing. The model is chosen on the data fed.

5 FLOWCHART



6 RESULT

Metrics	Accuracy	Precision	Recall	ROC AUC
Best Model	LGBM	LGBM	RF	LGBM

7 ADVANTAGES AND DISADVANTAGES

- ADVANTAGES
 - Better metrics
 - Efficiency increased due to hyper parameter tuning
 - Quicker results
- DISADVANTAGES
 - Specific to breast cancer
 - Large number of attributes
 - Increased memory space

8 APPLICATIONS

This system can be used in hospitals to aid doctors for diagnosing breast cancer.

9 CONCLUSION

An end-end system has been developed using AutoAI feature of IBM Watson and Node Red. This system seems to provide efficient results in terms of precision and recall.

10 FUTURE SCOPE

The model could use feature selection to select only the most important features. Other automated hyper parameter tuning techniques could be used.

11 BIBLIOGRAPHY

I would like to thank my mentor for guiding me in this project.