TUT05-09

ENGG2112

PROJECT PRESENTATION

Diabetes and Cardiovascular Disease
Stacked-Ensemble Model Predictions

Problem Statement

"The rising incidence of diabetes and its strong correlation with Cardiovascular Disease (CVD) calls for significant improvements to diagnostic intervention. To develop ensemble-stacking models to improve early detection and diagnosis for both diseases become imperative."

Background & Objectives

20.5M

CVD is the leading cause of mortality in 2023 (WHF).

422 M

Individuals are diagnosed with diabetes worldwide (WHO).

11.8%

Total domestic health expenditure for diabetes and CVD treatments (NIH).

- **Health Impact**: CVD and diabetes highlight a critical intersection between two major health crises.
- **Diagnostic Gaps**: Traditional models fail to capture complex factors influencing these diseases, leading to decreased accuracy in patient diagnosis.
- **Ensemble Models**: Stacking models enhance accuracy by leveraging the strength of multiple classifiers.
- **Implementation Challenges**: Integrating to healthcare systems and high computational costs.

Our Datasets

Both datasets were found on Kaggle and both contain patient medical records that are relevant to their respective disease.

Diabetes

Obtained from a hospital in Frankfurt, Germany.

2000 records and 9 features, including:
- Age
- Glucose
- Insulin levels
- BMI

Cardiovascular Disease

Obtained from ongoing study on resident in Framingham, Massachusetts.

4000 records and 14 features, including:
- Age
- Blood Pressure
- Heart Rate
- Smoking Status

Stacking model process flow chart

STEP 1
Input training data
Load dataset csv file into the juypter notebook

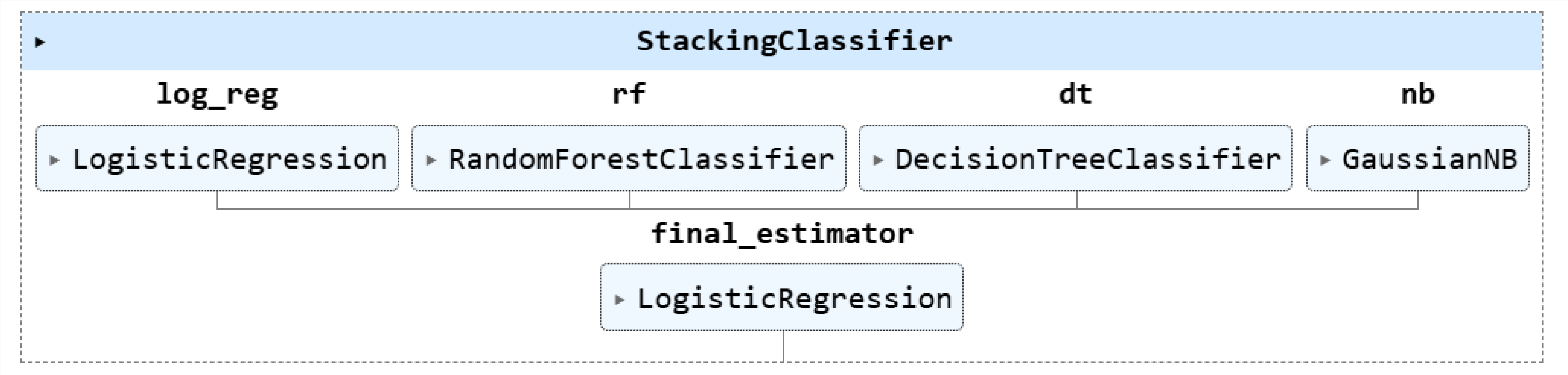
STEP 2
Base Classifiers
Train the individual base classifiers and select the top 4 performing classifiers

STEP 3
Meta Classifier
Using the outputs from the base classifier, train the meta classifier

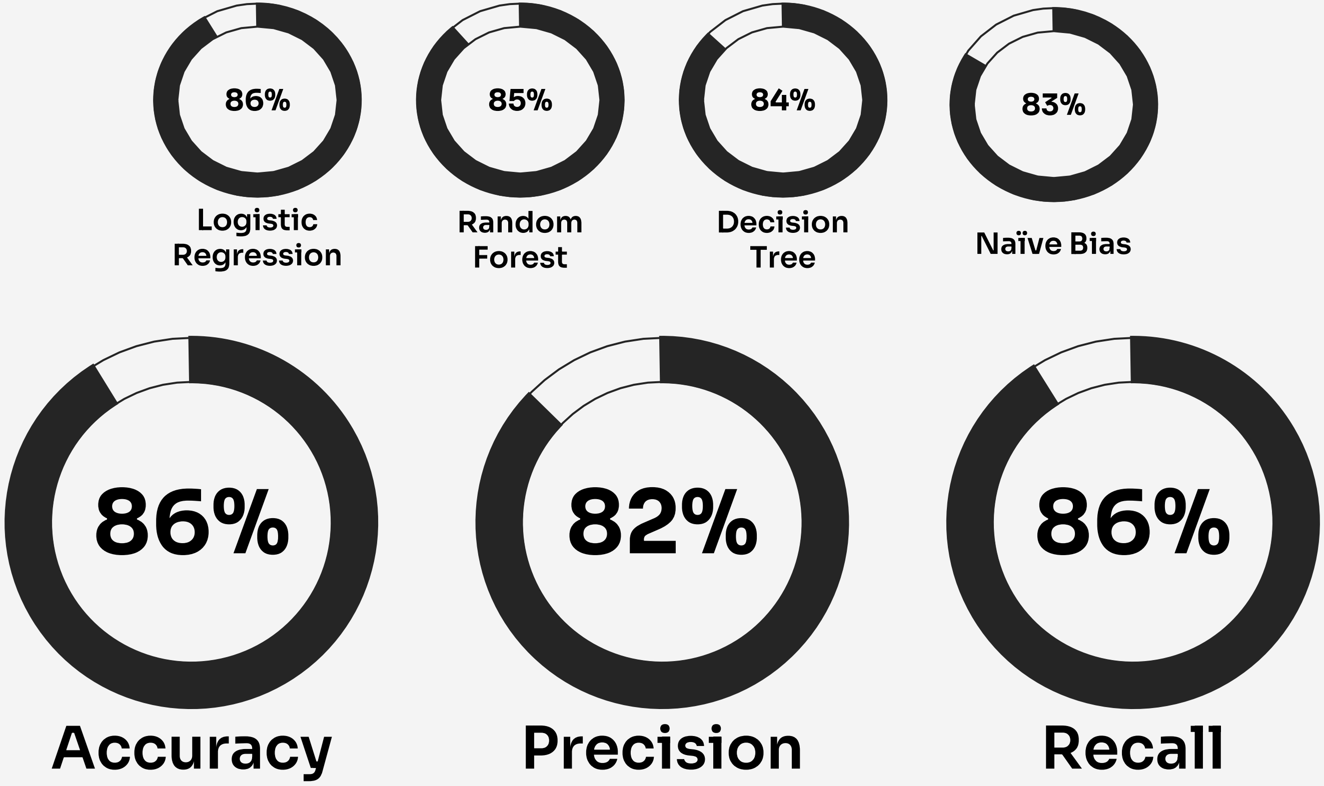
STEP 4
Evaluate Results

STEP 5
Final Results
Encouraging continued integration of machine learning models in healthcare to address complex health challenges

The CVD Model



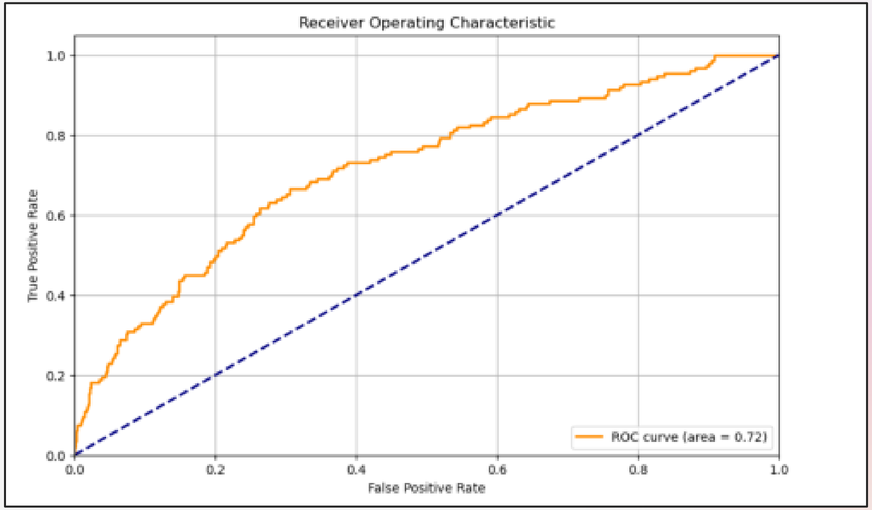
Classification Report: CVD model

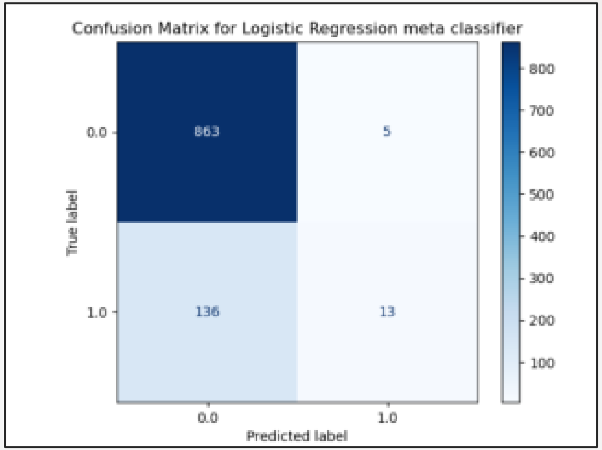


CVD – ROC Curve and Confusion Matrix

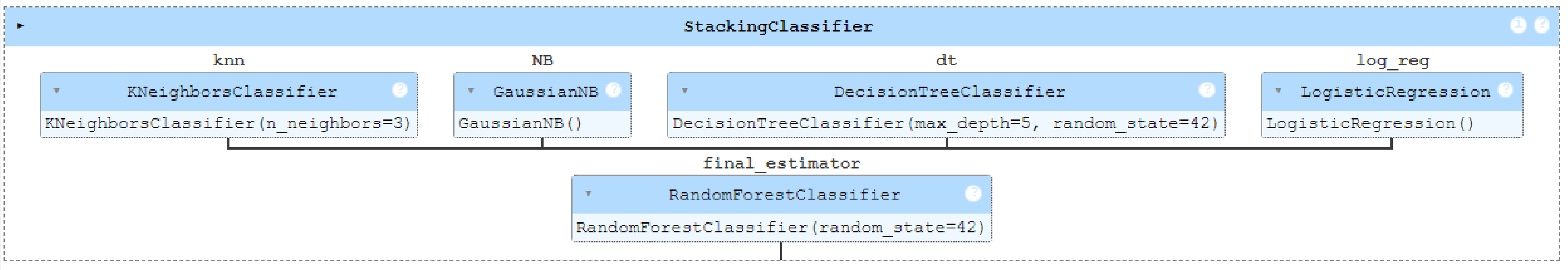
ROC curve indicates that it can fairly separate positive and negative cases

Confusion matrix shows that we have a high true negative rate, but a low true positive rate.

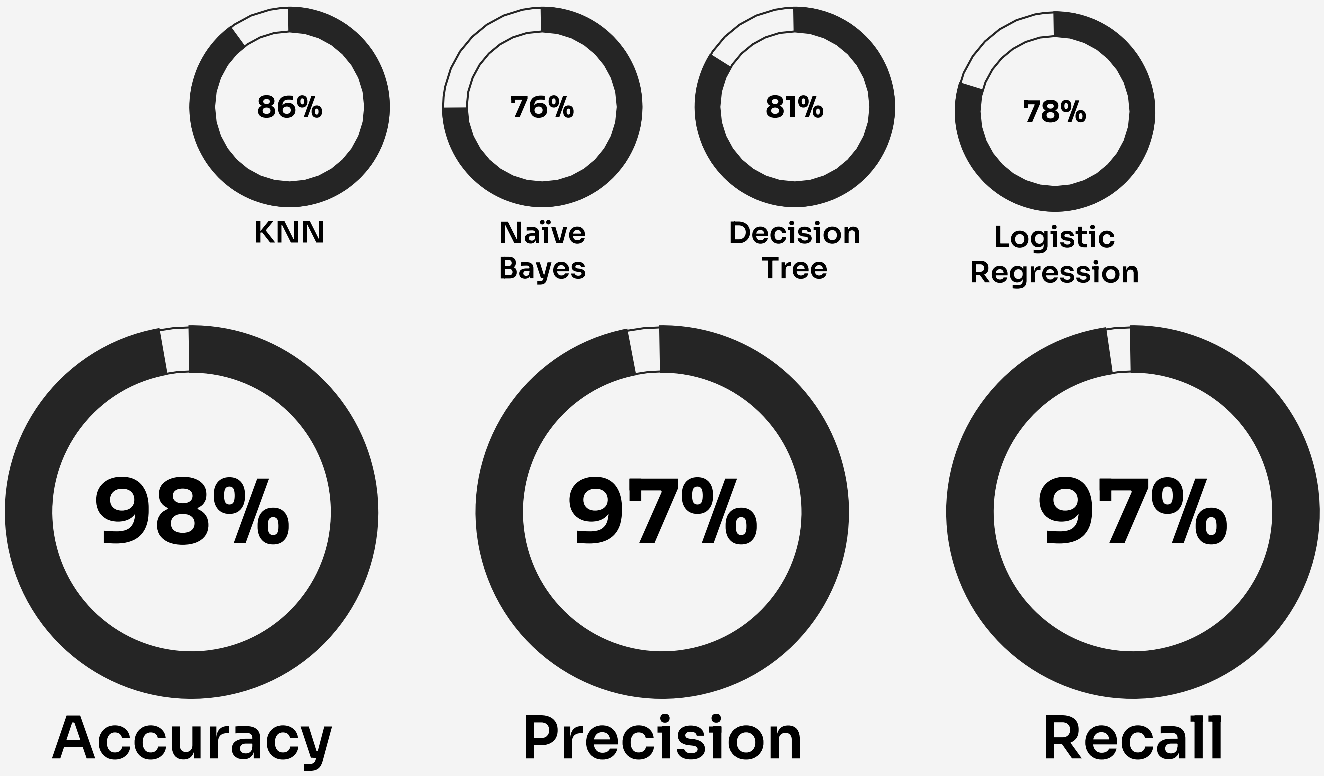




The Diabetes Model



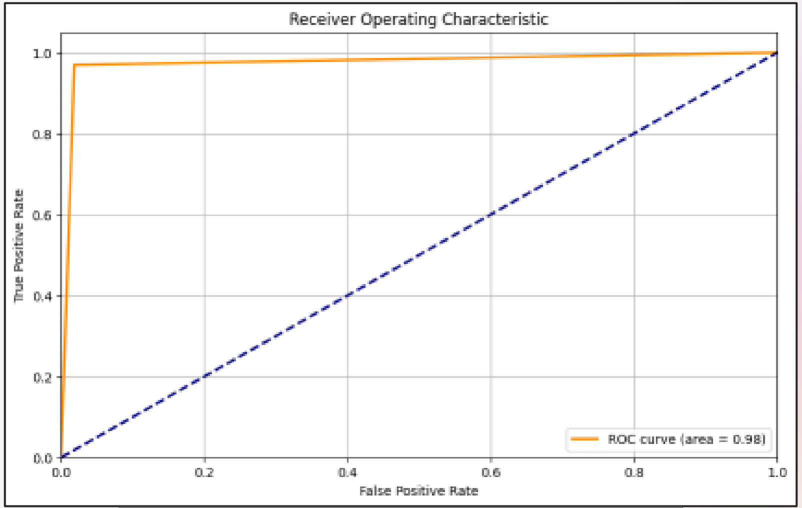
Classification Report: Diabetes model

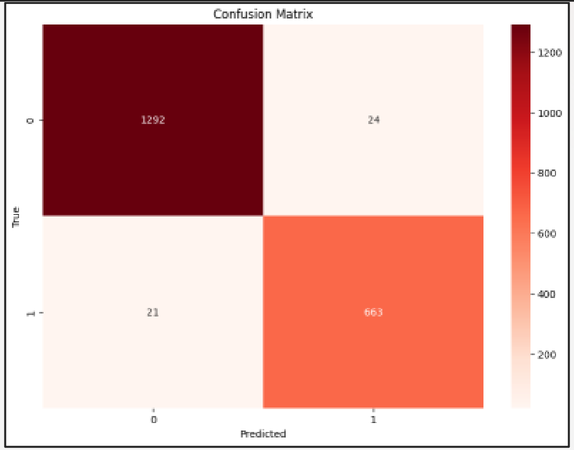


Diabetes – ROC Curve and Confusion Matrix

ROC curve indicates that our model is excellent at distinguishing between positive and negative cases.

Confusion matrix shows that we have both a high true positive and negative rate with our stacked model.





Reccomendations for client

Implementation

Embed models into CDSS to boost diagnostic accuracy, enable real-time risk assessments, and improve patient outcomes through enhanced decision making.

Action Items

Reactive ⇒ proactive

Transition from reactive to proactive care by regularly screening high-risk patients and offering preventative guidance to at-risk demographics.

Further work

Continuously enhance model performance by updating with new data, evaluating performance regularly, and refining algorithms based on ongoing research and clinical feedback.

Future Potential

Healthcare industry

To revitalise preventative healthcare in Australia's healthcare industry, by enabling early detection and intervention to chronic diseases. To address improvements to patient health outcomes, reducing long-term healthcare costs and burden on public health resources.

Client

Empower healthcare players with precise, real-time diagnostic tools, improving patient care through personalised treatment plans and enhancing patient trust and satisfaction with medical services.

Thank you!