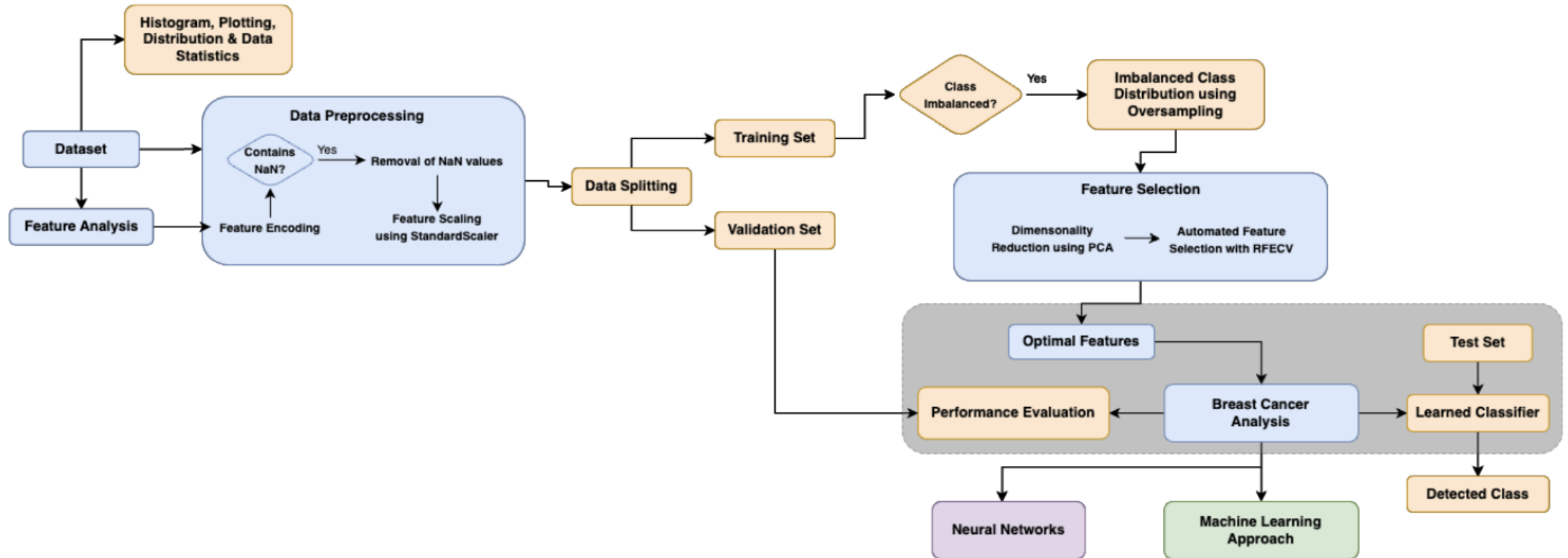# *Breast Cancer Wisconsin (Diagnostic) Data Set*

Group 5.
Team Members

Anthony Korsah
Abdullah Al Reza
Efta Khairul Bashar
Md Moinul Islam

Proposed Methodology
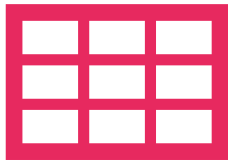
Presentation Goals

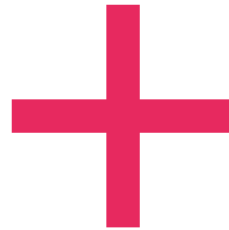Feature extraction

Model train

Prediction

Feature Extraction

We used Recursive Feature
Extraction with Cross Validation
set to extract features.
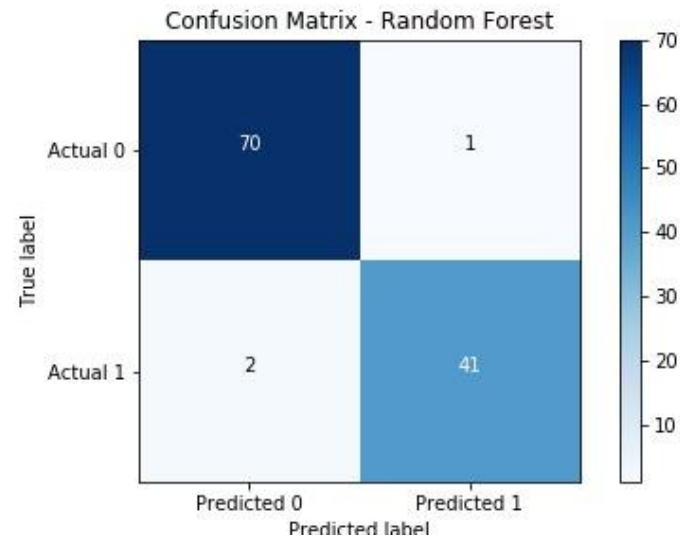
A cross-validation size of 5

We got 21 optimal features
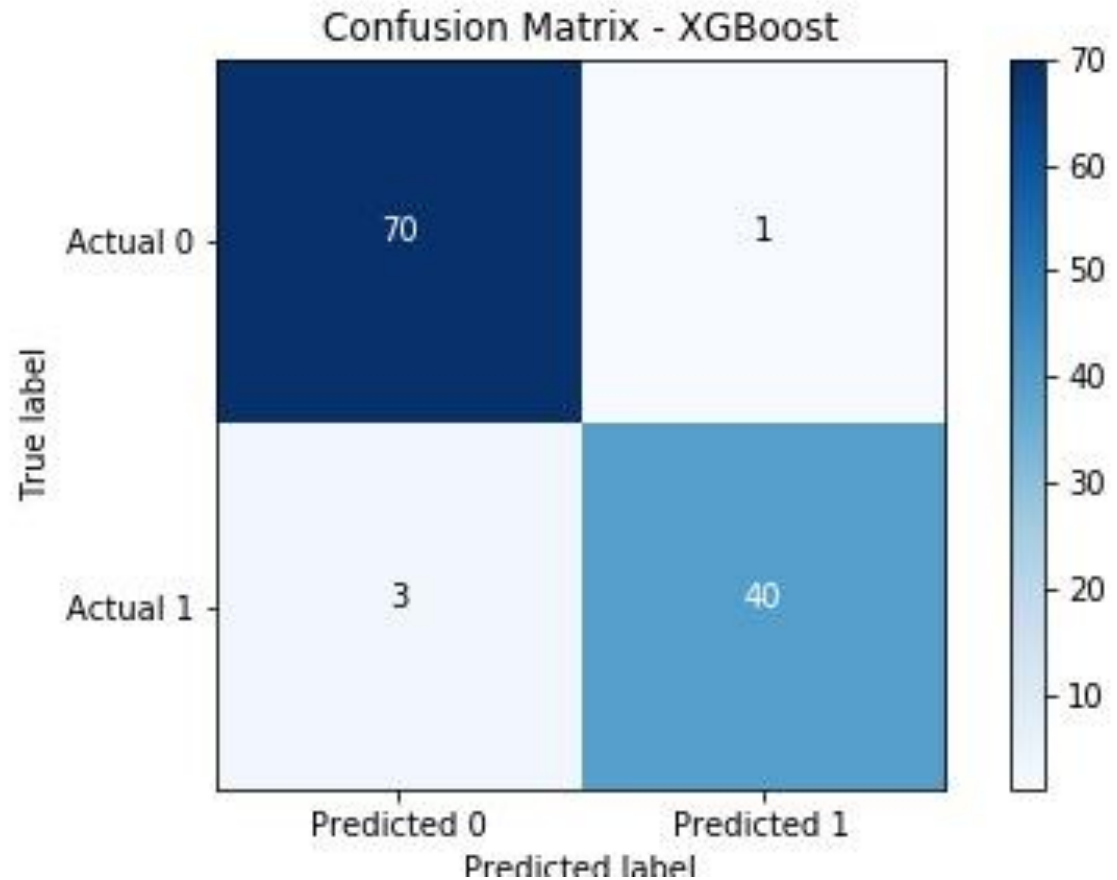
Model Training

Random Forest confusion matrix:

Confusion Matrix - Random Forest
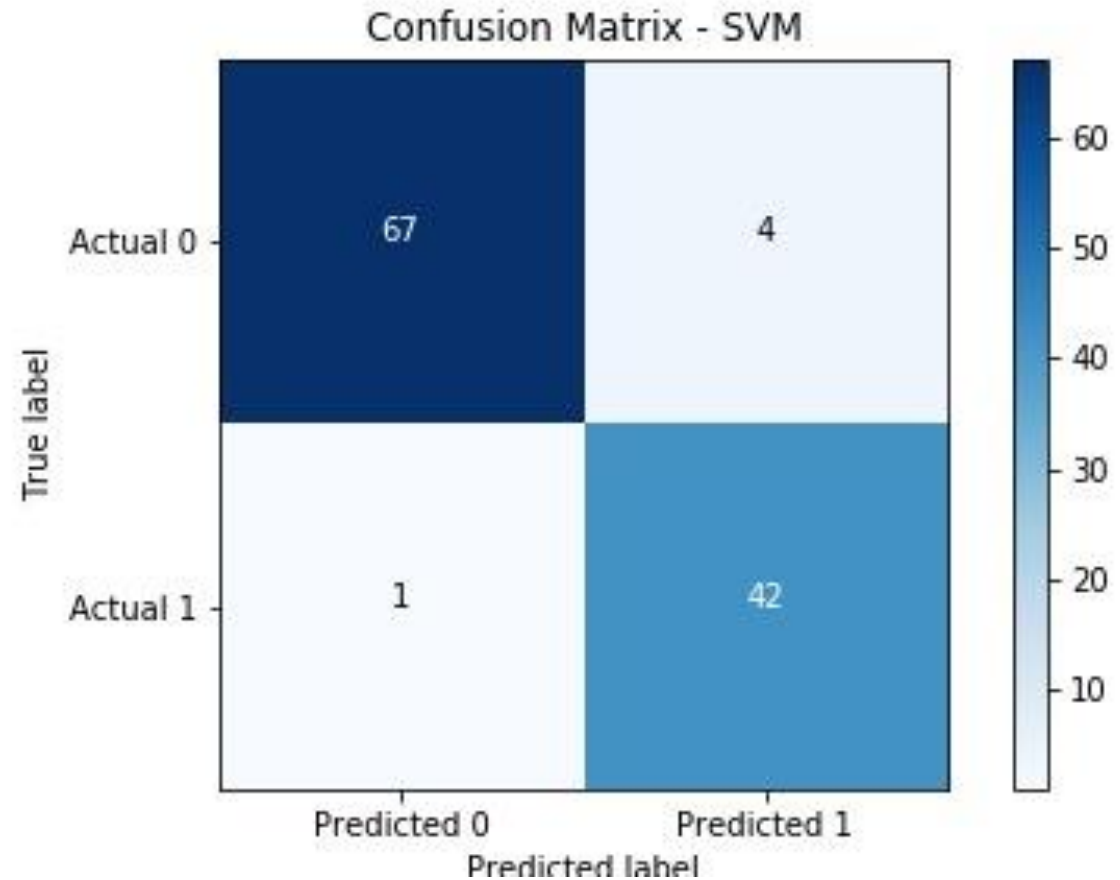
- Accuracy: 97%
- Precision: 0.976
- Recall: 0.953
- F1 Score: 0.96

XGBoost Confusion Matrix:

- Accuracy: 96.4%
- Precision: 0.975
- Recall: 0.930
- F1 Score: 0.952

Confusion Matrix - XGBoost

Neural Network Confusion Matrix

- Accuracy: 97.3%

- Precision: 0.976

- Recall: 0.953

- F1 Score: 0.964

Confusion Matrix - Neural Network
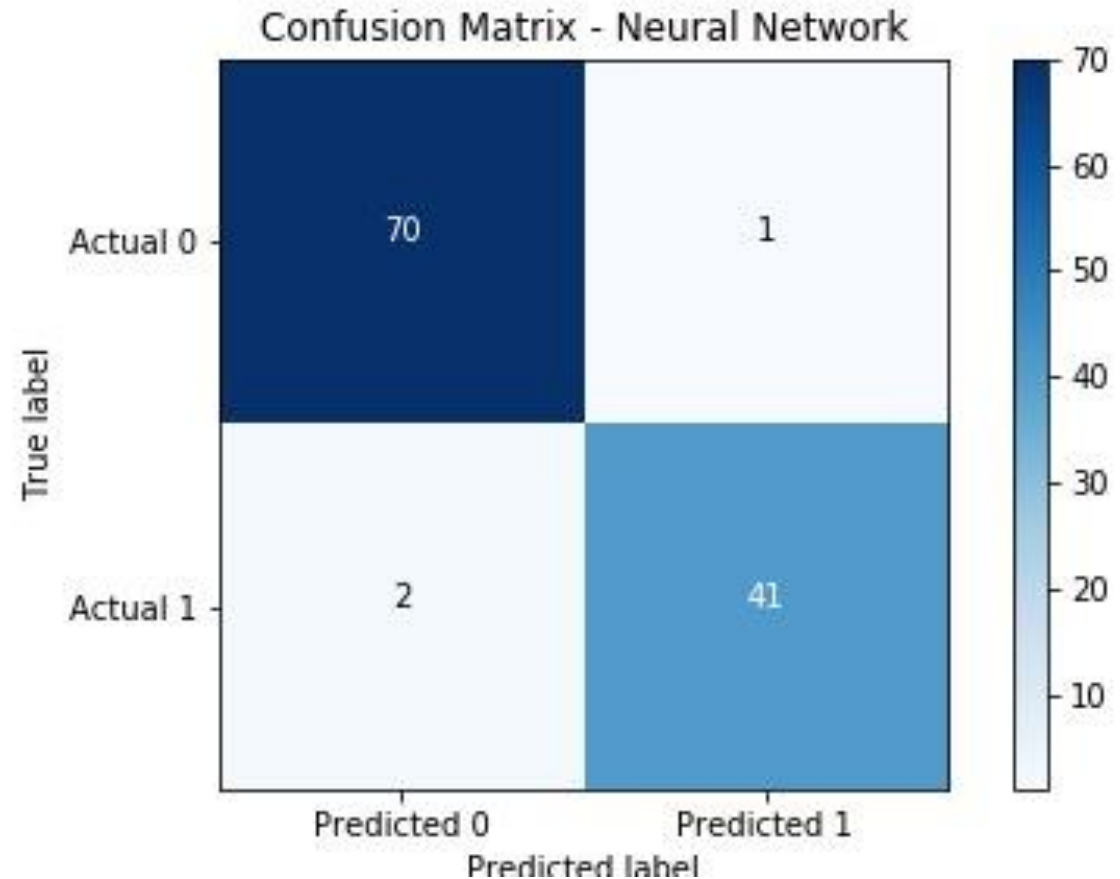
SVM
Confusion Matrix

- Accuracy: 95.6%
- Precision: 0.913
- Recall: 0.976
- F1 Score: 0.943

Features and model training

To get a different perspective,
We again extracted 20 features with the help of PCA

We again trained some models with these features.

Model: Random Forest (PCA)Confusion Matrix.

Accuracy: 37.7%

Precision: 0.377

Recall: 1.0

F1 Score: 0.547

Model: XGBoost (PCA) Confusion Matrix.

Accuracy: 37.7%

Precision: 0.375

Recall: .976

F1 Score: 0.541

Features and model training

To get a different perspective,
We again extracted 20 features with the help of PCA

We again trained some models with these features.

**Model: Neural Network (PCA)
Confusion Matrix**

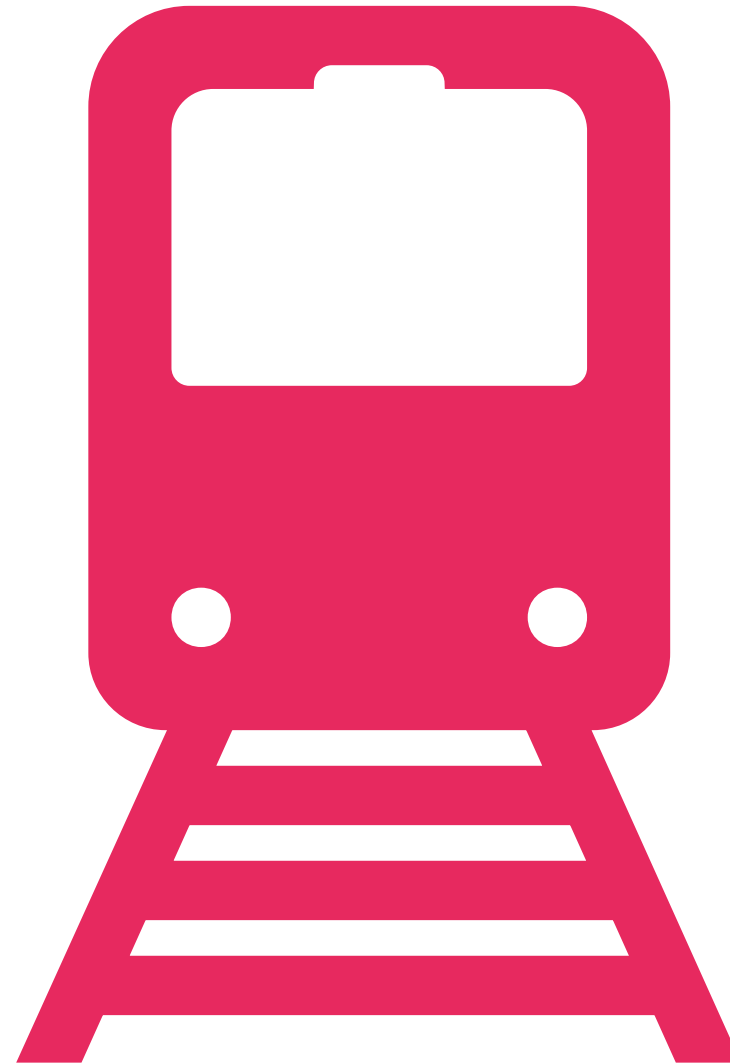Accuracy: 53.5%

Precision: 0.439

Recall: .837

F1 Score: 0.576

**Model: SVM (PCA)
Confusion Matrix**

Accuracy: 41.3%

Precision: 0.372

Recall: .813

F1 Score: 0.510

Future work

- All the models till now have been trained with outliners. We can exclude the outliners and train the models. With this approach we might get better results.

Thank you

Any questions ?