Milestone 3: Machine Learning Implementations

By: Albina Cako & Joshua Dalphy

Sunday January 10th, 2021

Objectives

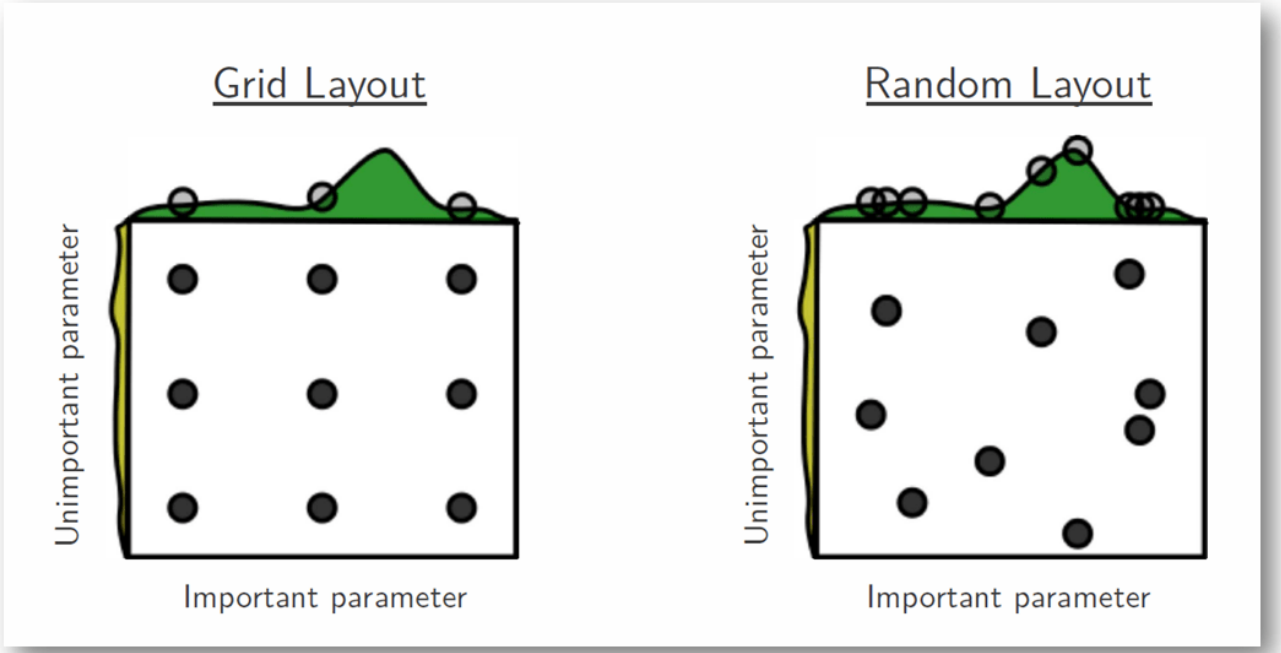
Implement 2 machine learning algorithms and analyze results using appropriate evaluation and selection criteria

Implement various ensemble learning methods and analyze results using appropriate evaluation and selection criteria

Compare the performance of each model and select the best based on accuracy and runtime

Methodology

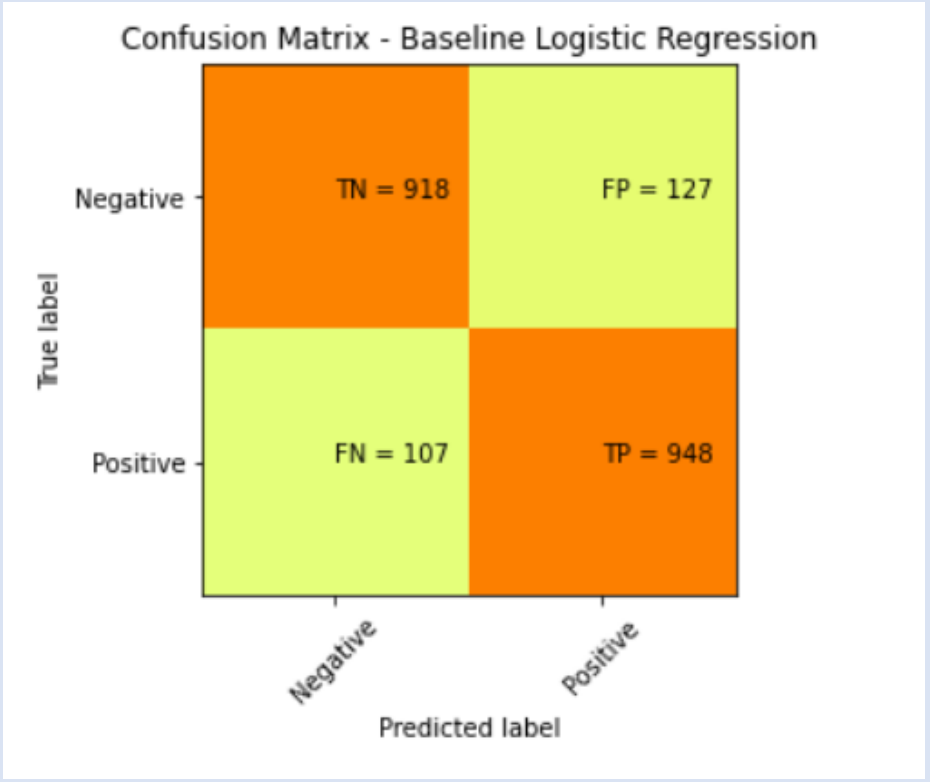
- Using the features selected from milestone 2, baseline models were produced and benchmarked (Logistic Regression, Naïve Bayes & RF).
- Hyperparameter Tuning was performed where applicable using *GridsearchCV* and *RandomizedSearchCV*.
- Ensemble learning methods were applied using the tuned models.
- The results of each modelling methods were analyzed and compared using performance and evaluation metrics.



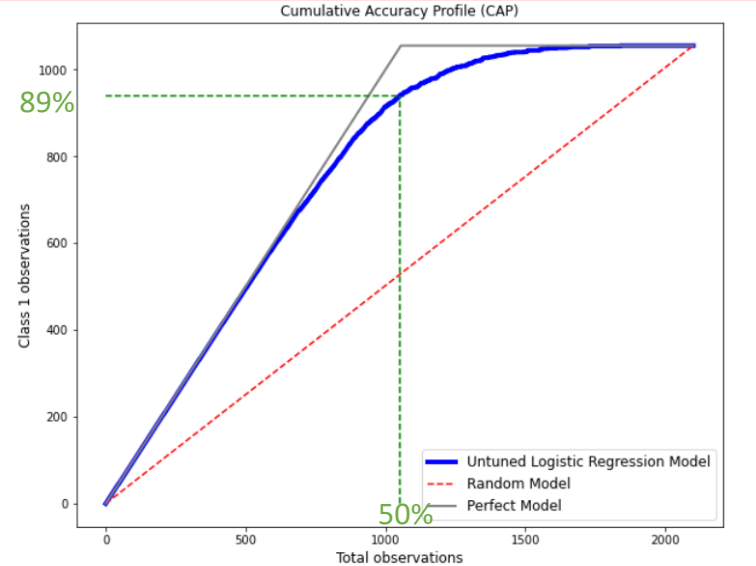
https://medium.com/@senapati.dipak97/grid-search-vs-random-search-d34c92946318

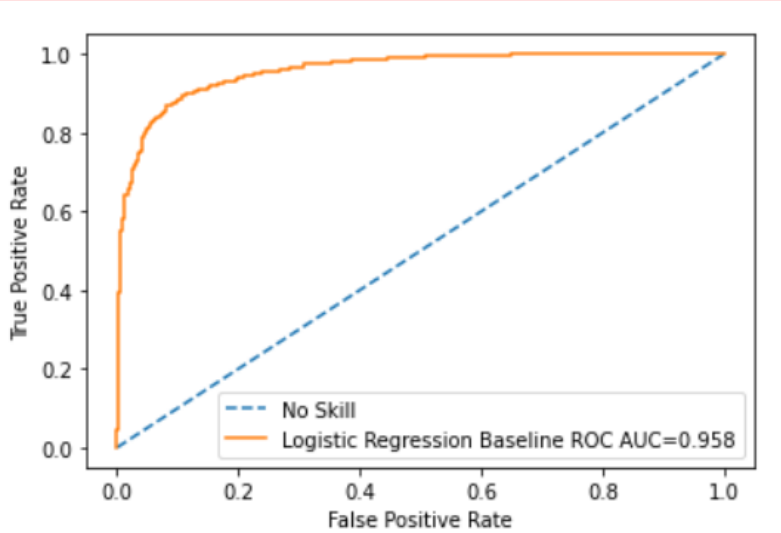
Baseline Model: Logistic Regression

```
Model Performance metrics:
------------------------------
Accuracy: 0.8886
Precision: 0.8887
Recall: 0.8886
F1 Score: 0.8886

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.88      0.90      0.89      1055
           0       0.90      0.88      0.89      1045

    accuracy                           0.89      2100
   macro avg       0.89      0.89      0.89      2100
weighted avg       0.89      0.89      0.89      2100
```



Baseline Model: Logistic Regression (cont.)





Guidelines for Model:

X > 90% Overfitting

80% < X < 90% Very Good Model

70% < X < 80% Good Model

60% < X < 70% Average Model

X < 60% Poor Model

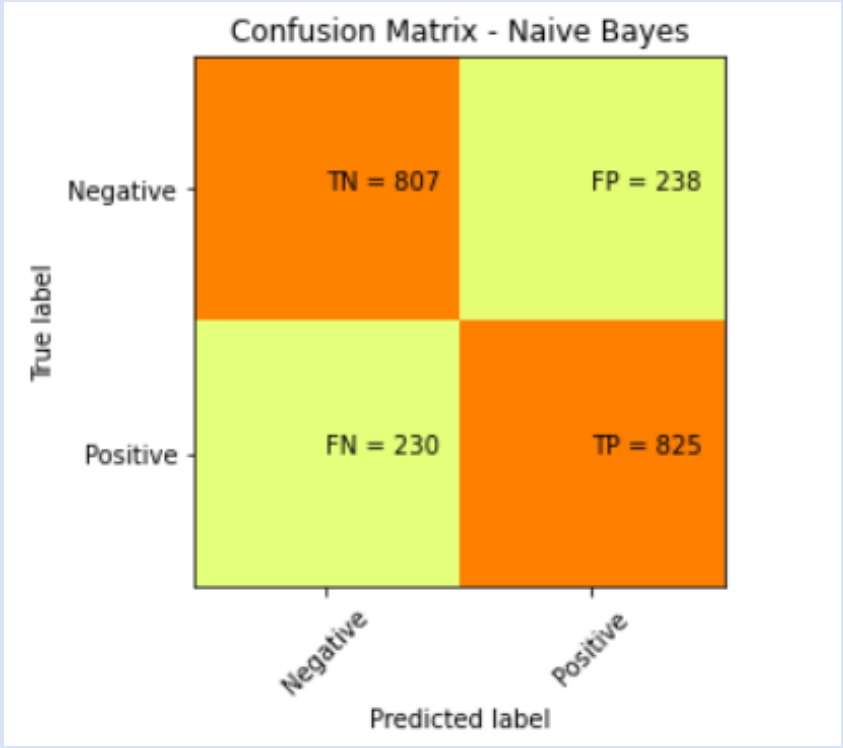
- Accuracy ratio (AULR/AUP): 0.915
- Evaluate the Model using 50% line on the CAP Curve:
  - Value at 50% line is 89% which suggests that the untuned LR Model is very good

AUP = Area under the perfect model

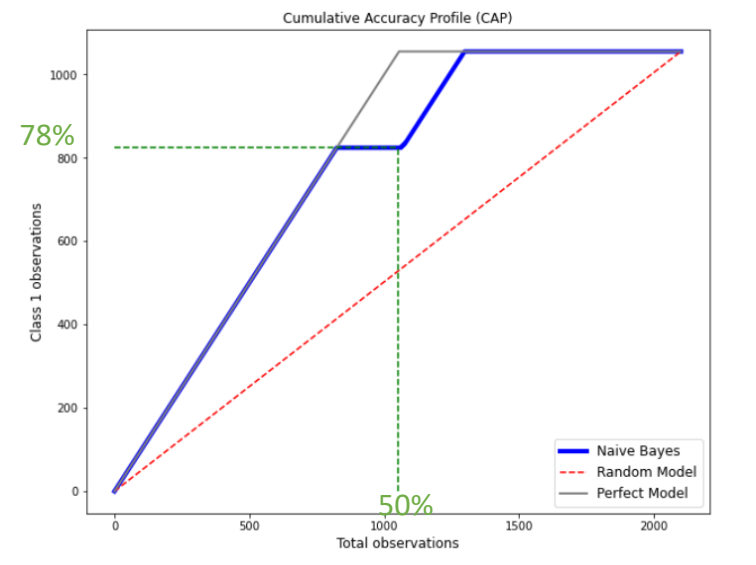
AULR = Area under the Logistic Regression model

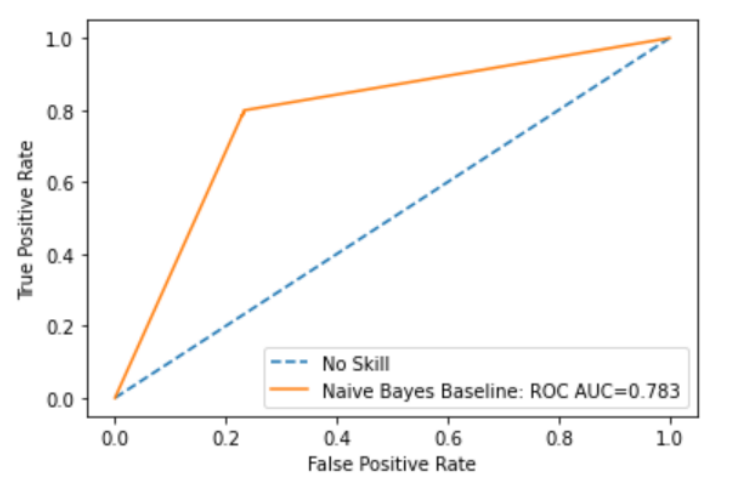
Baseline Model: Naïve Bayes

```
Model Performance metrics:
------------------------------
Accuracy: 0.7771
Precision: 0.7772
Recall: 0.7771
F1 Score: 0.7771

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.78      0.78      0.78      1055
           0       0.78      0.77      0.78      1045

    accuracy                           0.78      2100
   macro avg       0.78      0.78      0.78      2100
weighted avg       0.78      0.78      0.78      2100
```



Baseline Model: Naïve Bayes (cont.)





Guidelines for Model:

X > 90% Overfitting

80% < X < 90% Very Good Model

70% < X < 80% Good Model

60% < X < 70% Average Model

X < 60% Poor Model

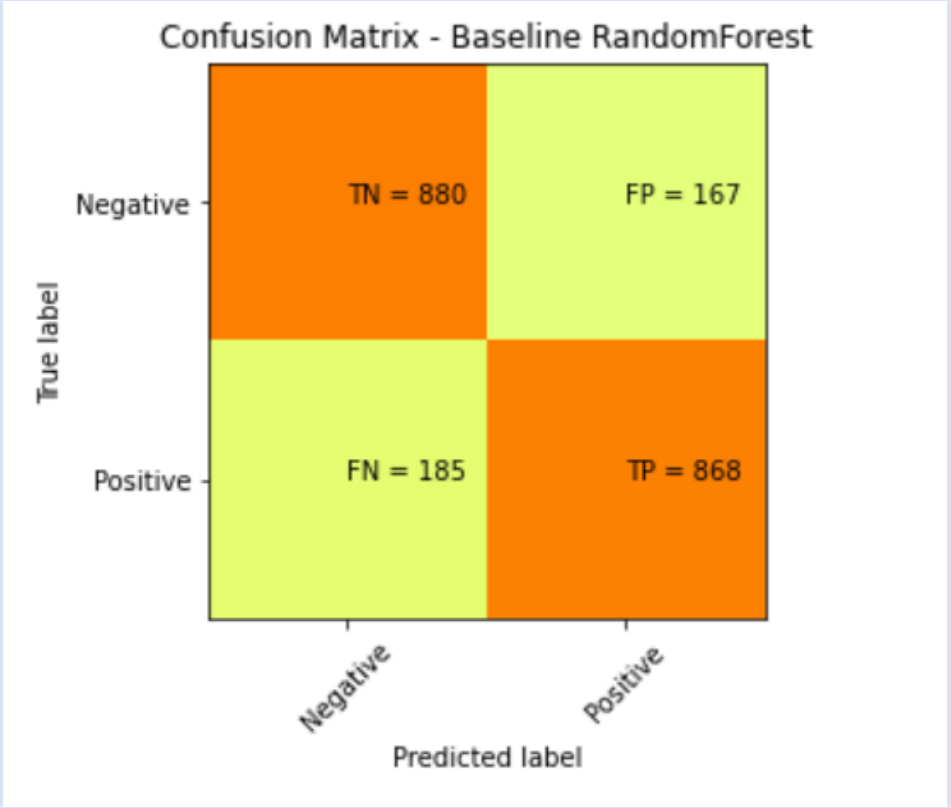
- Accuracy ratio (AUNB/AUP): 0.898
- Evaluate the Model using the 50% line on the CAP curve:
  - Value at the 50% line is 78% which suggests that the Naïve Bayes model is good

AUP = Area under the perfect model

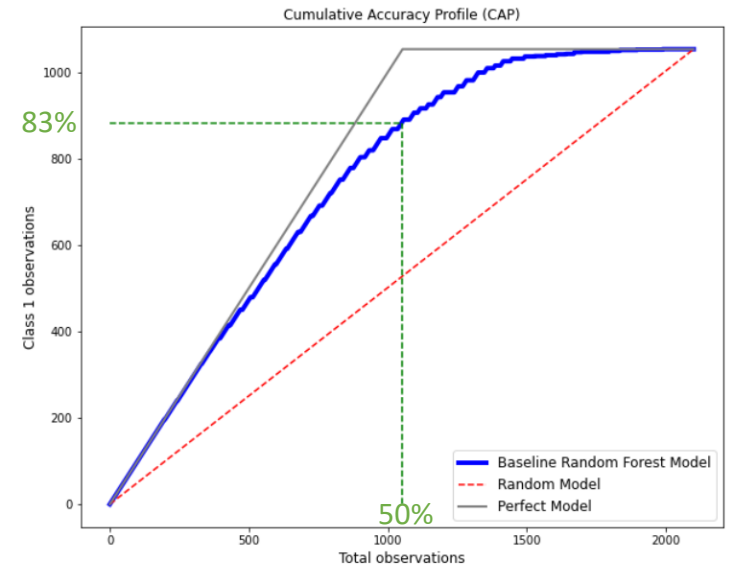
AUNB = Area under the Naïve Bayes model

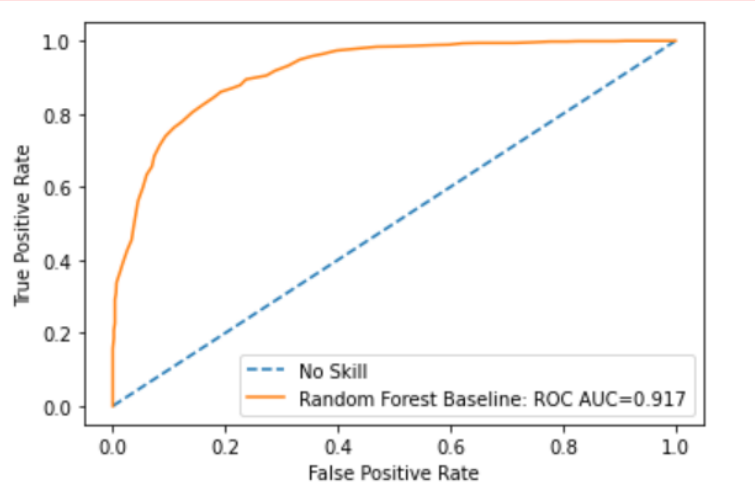
Baseline Model: Random Forest

```
Model Performance metrics:
------------------------------
Accuracy: 0.8324
Precision: 0.8325
Recall: 0.8324
F1 Score: 0.8324

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.84      0.82      0.83      1053
           0       0.83      0.84      0.83      1047

    accuracy                           0.83      2100
   macro avg       0.83      0.83      0.83      2100
weighted avg       0.83      0.83      0.83      2100
```



Baseline Model: Random Forest (cont.)





Guidelines for Model:

X > 90% Overfitting

80% < X < 90% Very Good Model

70% < X < 80% Good Model

60% < X < 70% Average Model

X < 60% Poor Model

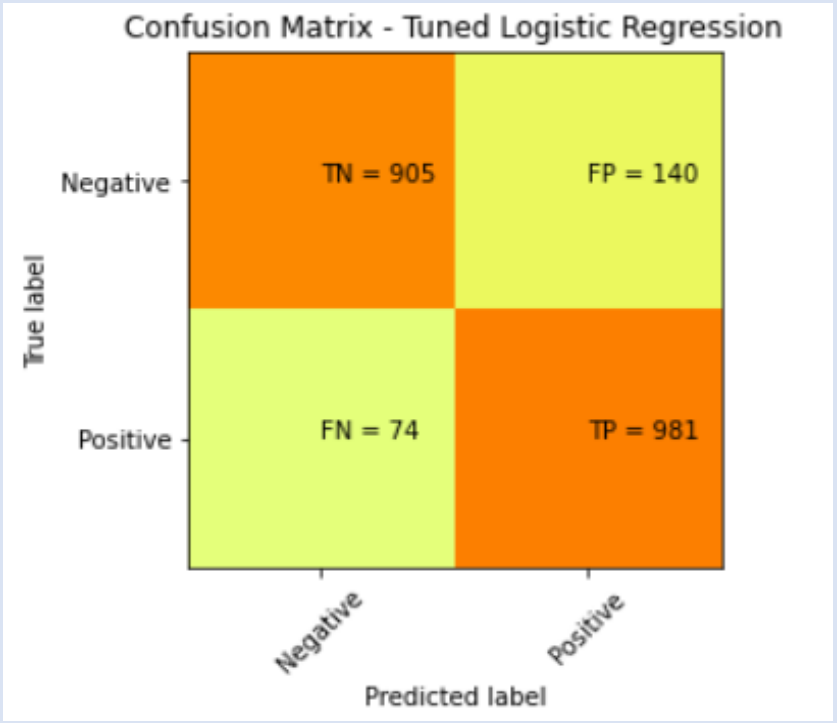
- Accuracy ratio (AURF/AUP): 0.841
- Evaluate the Model using the 50% line on the CAP curve:
  - Value at the 50% line is 83% which suggests that the Random Forest model is very good

AUP = Area under the perfect model

AURF = Area under the Random Forest Model

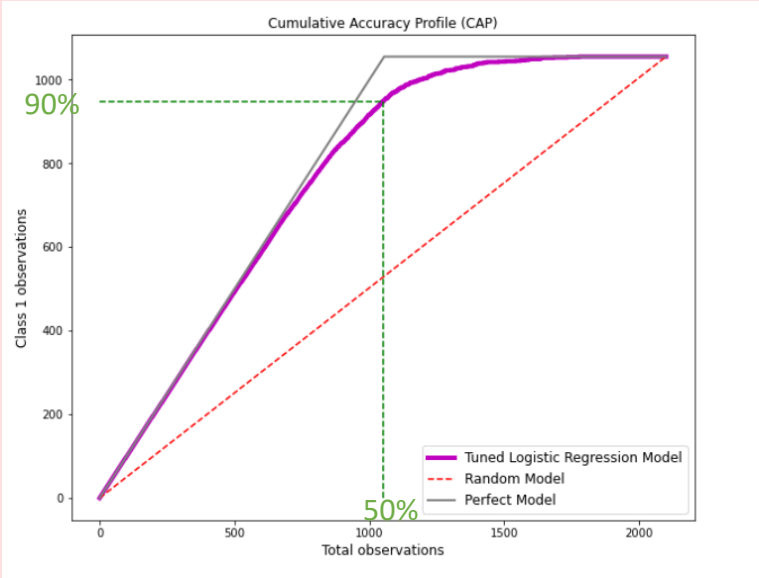
Hyperparameter tuning: Logistic Regression

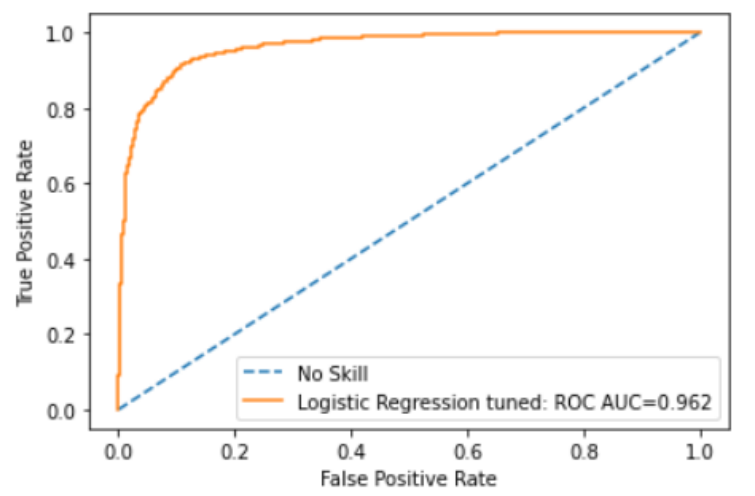
```
Model Performance metrics:
------------------------------
Accuracy: 0.8981
Precision: 0.8996
Recall: 0.8981
F1 Score: 0.898

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.88      0.93      0.90      1055
           0       0.92      0.87      0.89      1045

    accuracy                           0.90      2100
   macro avg       0.90      0.90      0.90      2100
weighted avg       0.90      0.90      0.90      2100
```



Hyperparameter tuning was performed using python's GridsearchCV function

Hyperpar-ameter tuning: Logistic Regression





Guidelines for Model:

X > 90% Overfitting

80% < X < 90% Very Good Model

70% < X < 80% Good Model

60% < X < 70% Average Model

X < 60% Poor Model

- Accuracy ratio (AUTLR/AUP): 0.924
- Evaluate the Model using the 50% line on the CAP curve:
  - Value at the 50% line is 90% which suggests that the tuned Logistic Regression model is very good

AUP = Area under the perfect model

AUTLR = Area under the tuned Logistic Regression model

```
Best Params:  {'C': 100, 'class_weight': {1: 0.4, 0: 0.6}, 'penalty': 'l2', 'solver': 'liblinear'}
```

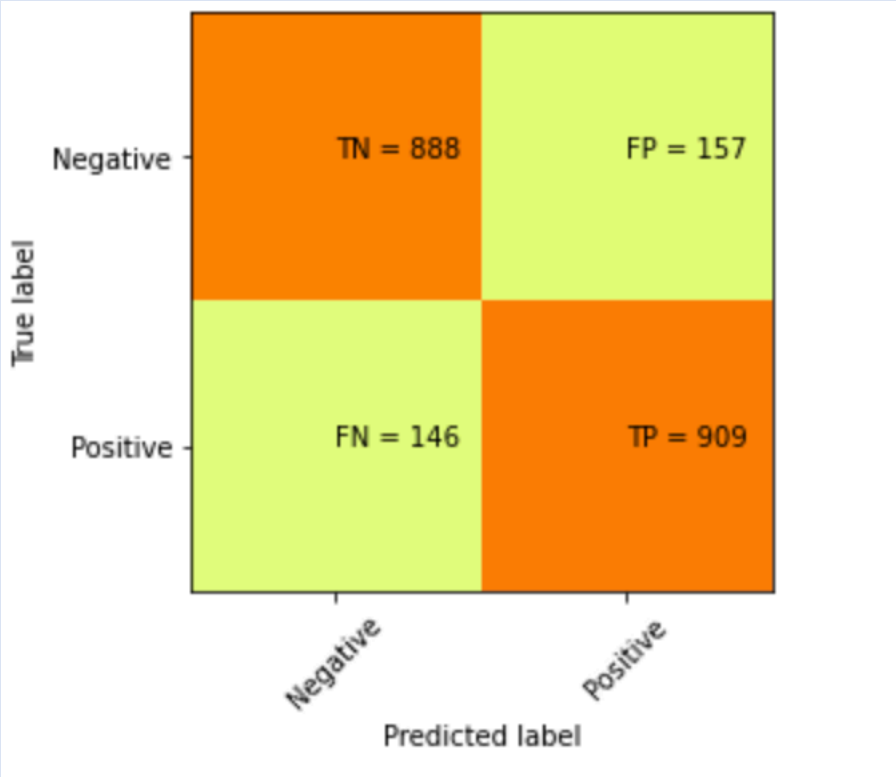
Hyperparameter tuning: Naïve Bayes

- The Naïve Bayes model does not have any parameters which can be tuned therefore hyperparameter tuning was not conducted

Hyperparameter tuning: Random Forest

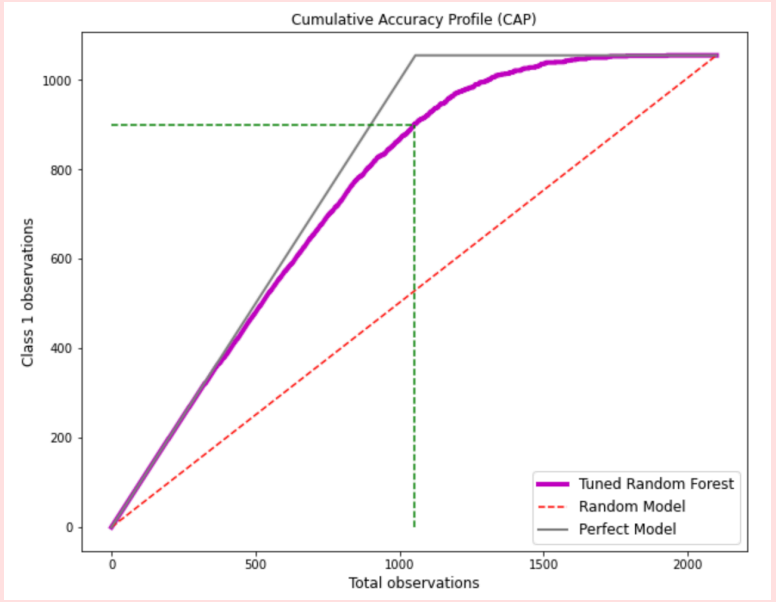
```
Model Performance metrics:
------------------------------
Accuracy: 0.8557
Precision: 0.8557
Recall: 0.8557
F1 Score: 0.8557

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.85      0.86      0.86      1055
           0       0.86      0.85      0.85      1045

    accuracy                           0.86      2100
   macro avg       0.86      0.86      0.86      2100
weighted avg       0.86      0.86      0.86      2100
```

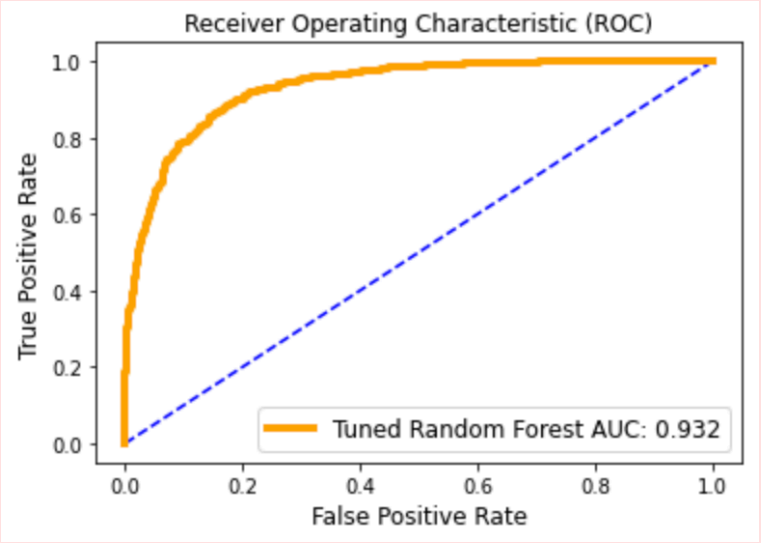
Confusion matrix – Tuned RF



Hyperparameter tuning was performed using python's RandomizedSearchCV function

Hyperpar-ameter tuning: Random Forest





Guidelines for Model:

X > 90% Overfitting

80% < X < 90% Very Good Model

70% < X < 80% Good Model

60% < X < 70% Average Model

X < 60% Poor Model

- Accuracy ratio (AUTLR/AUP): 0.864
- Evaluate the Model using the 50% line on the CAP curve:
  - Value at the 50% line is 85.3% which suggests that the tuned Random Forest model is very good

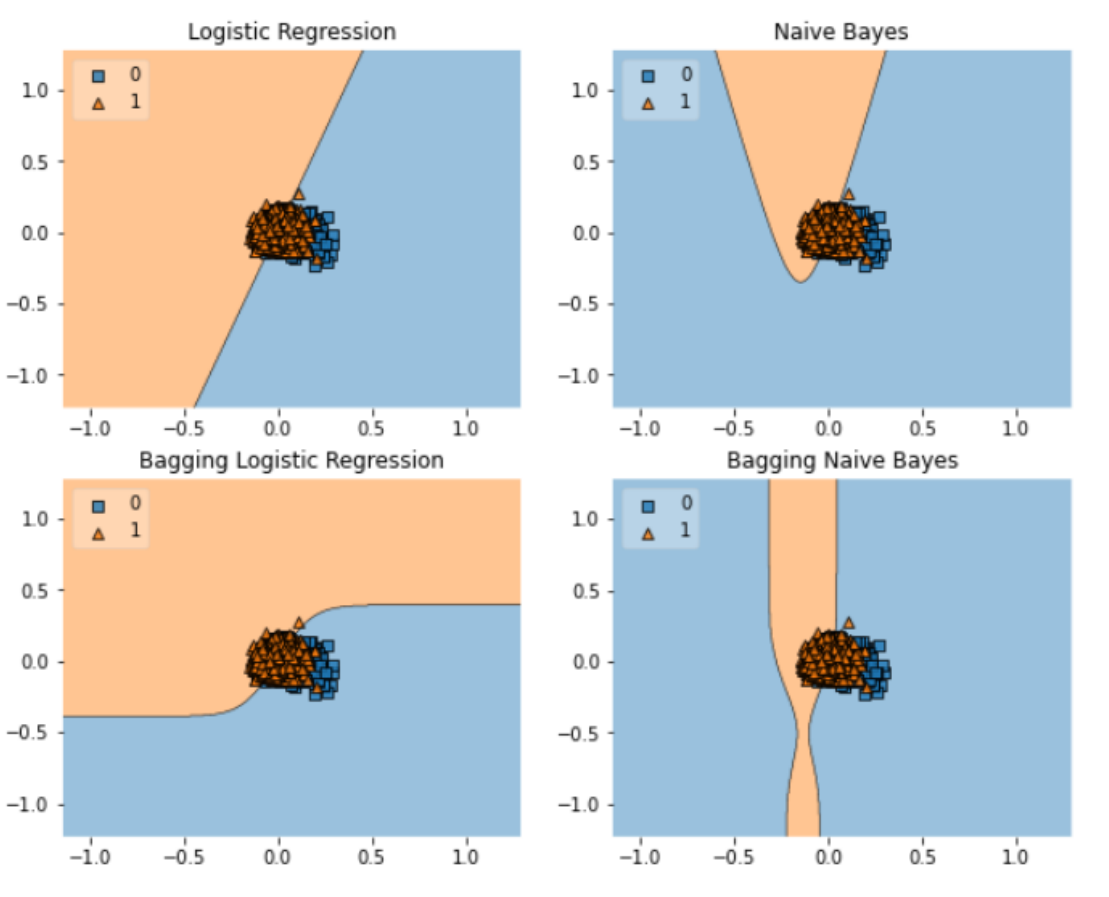
AUP = Area under the perfect model

AUTLR = Area under the tuned Random Forest model

```
Best Params:  {'n_estimators': 600, 'min_samples_split': 10, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': 40, 'bootstrap': False}
```

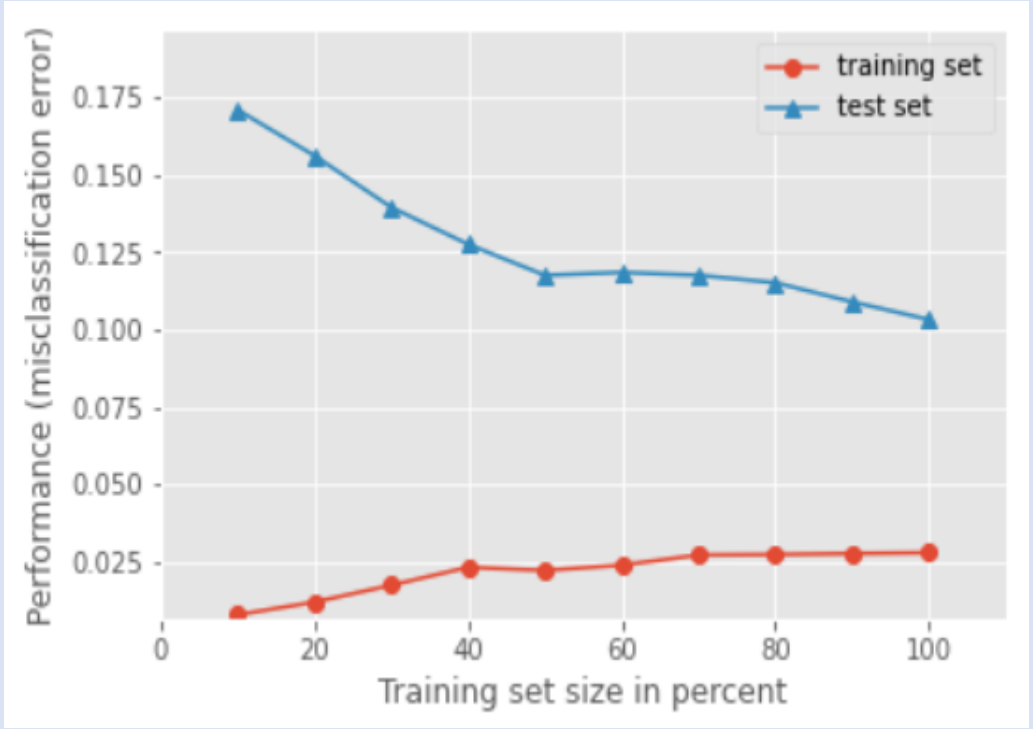
Ensemble Learning: Bagging

Bagging was implemented using the tuned logistic regression and Naïve Bayes models



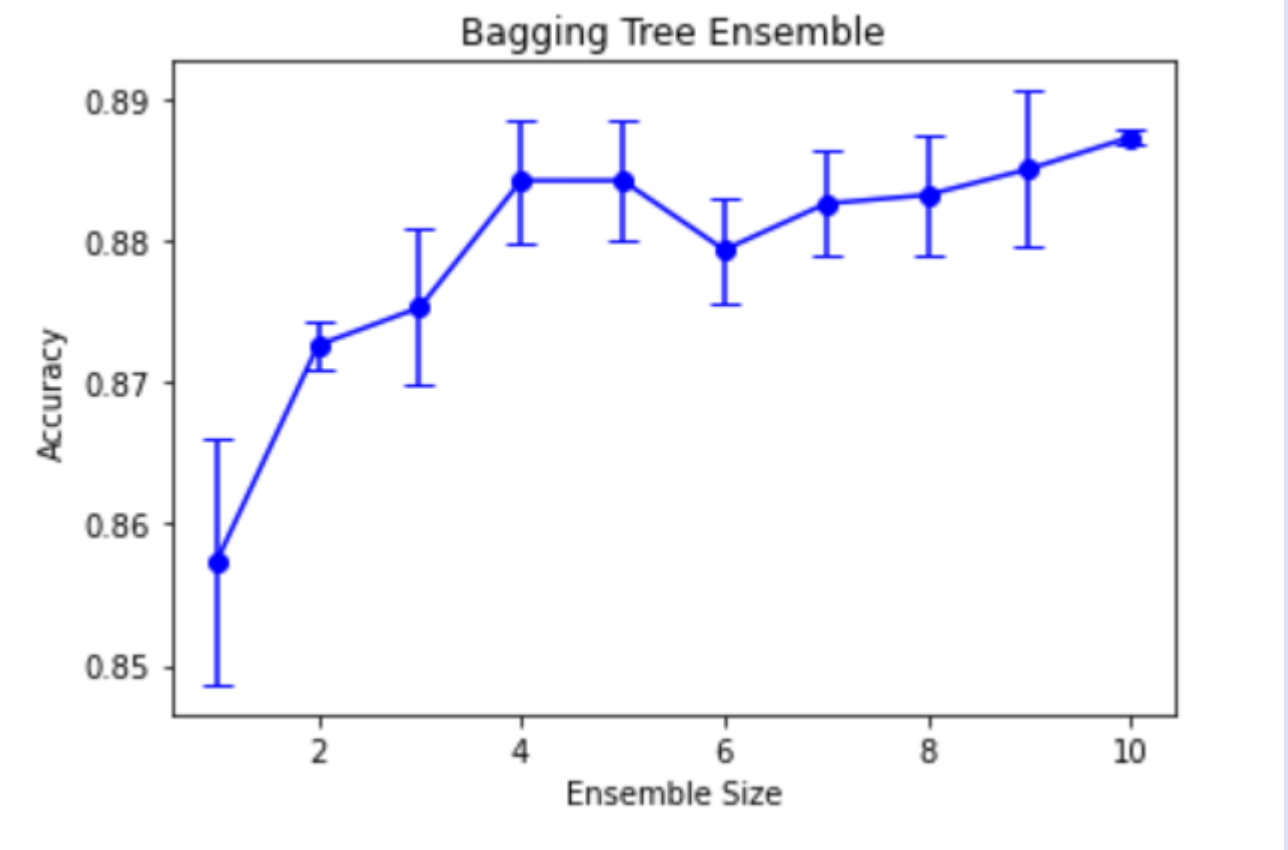

```
Accuracy: 0.89 (+/- 0.01) [Logistic Regression]
Accuracy: 0.76 (+/- 0.01) [Naive Bayes]
Accuracy: 0.88 (+/- 0.00) [Bagging Logistic Regression]
Accuracy: 0.78 (+/- 0.01) [Bagging Naive Bayes]
```

Learning curve for Bagging LR classifier



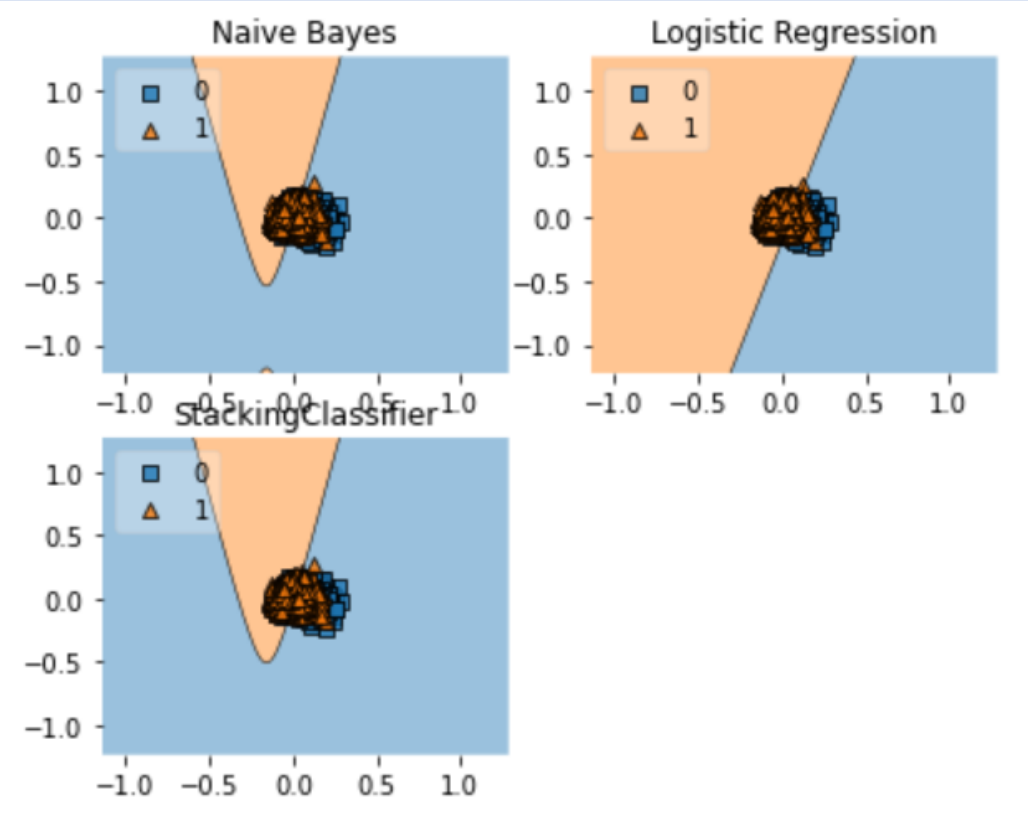
Ensemble Learning: Bagging (cont.)

Classifier Accuracy



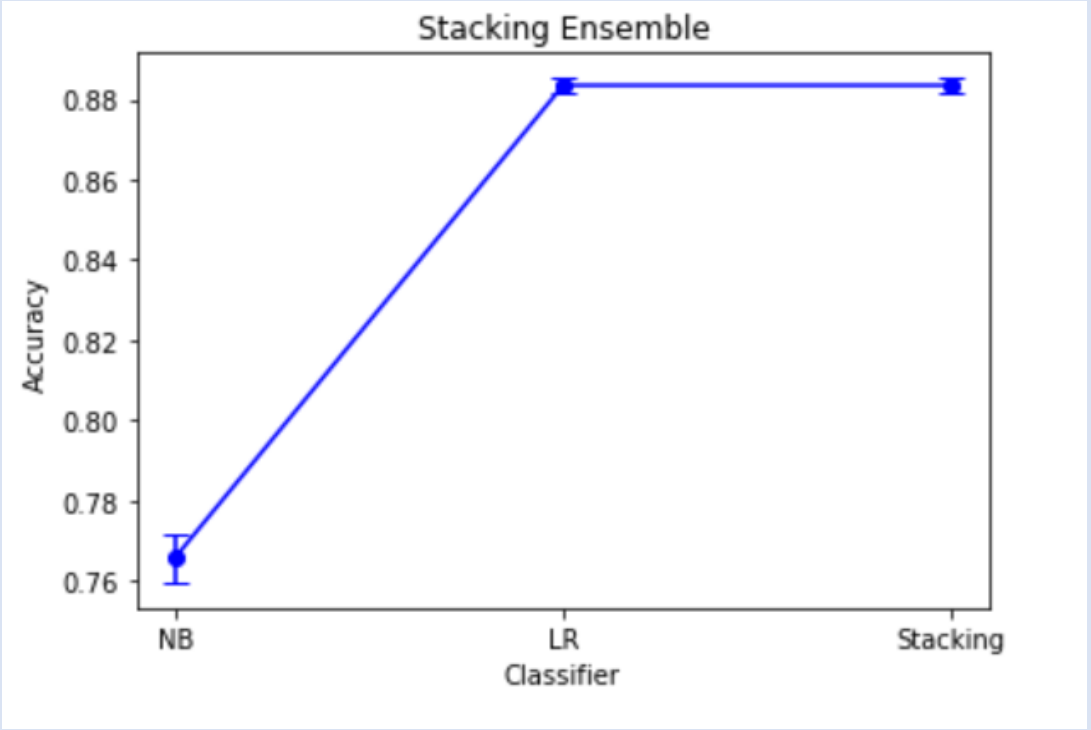
Ensemble Learning: Stacking

Stacking was implemented using the tuned logistic regression and Naïve Bayes models



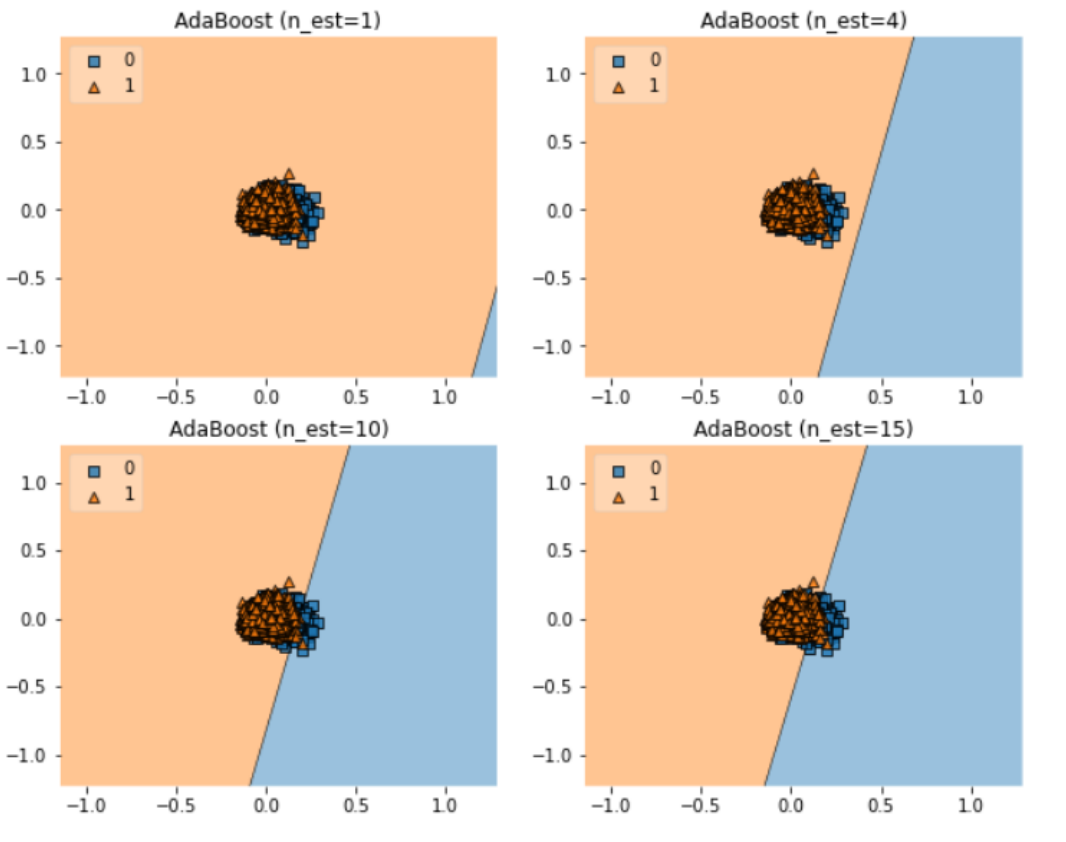

```
Accuracy: 0.77 (+/- 0.01) [Naive Bayes]
Accuracy: 0.88 (+/- 0.00) [Logistic Regression]
Accuracy: 0.88 (+/- 0.00) [StackingClassifier]
```

Classifier Accuracy

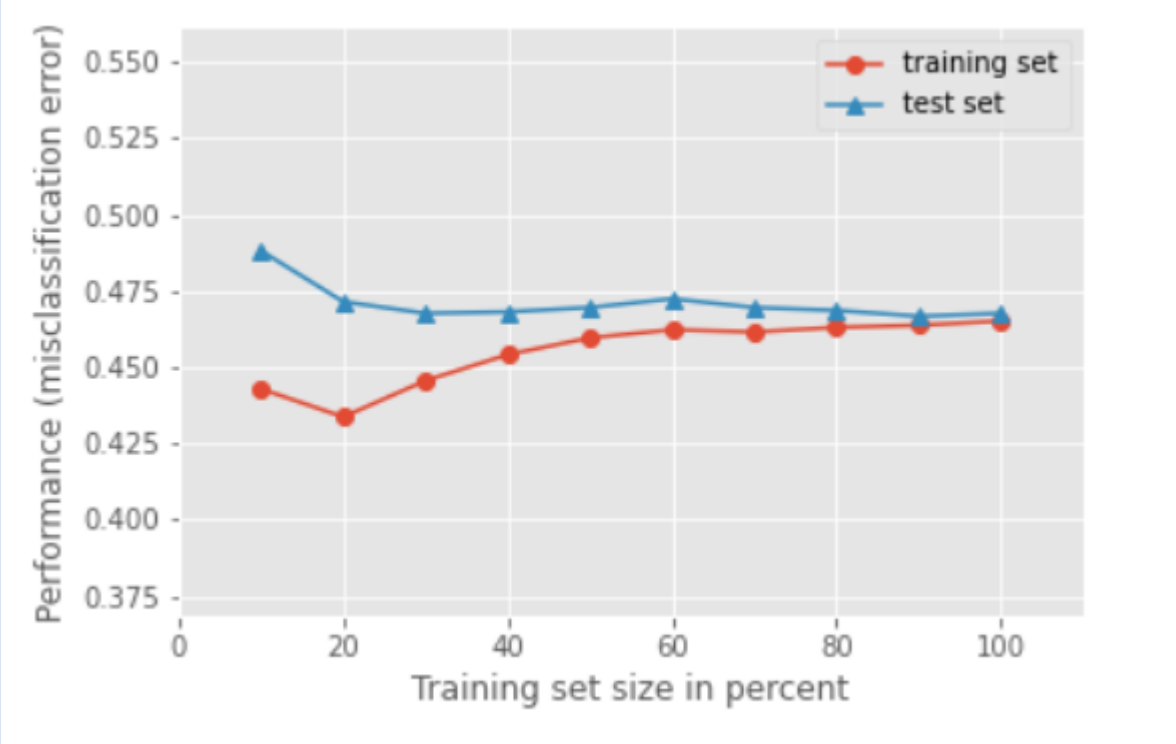


Ensemble Learning: Boosting

Boosting was implemented using AdaBoost. N_est = 10 and n_est = 15 seemed to perform best.

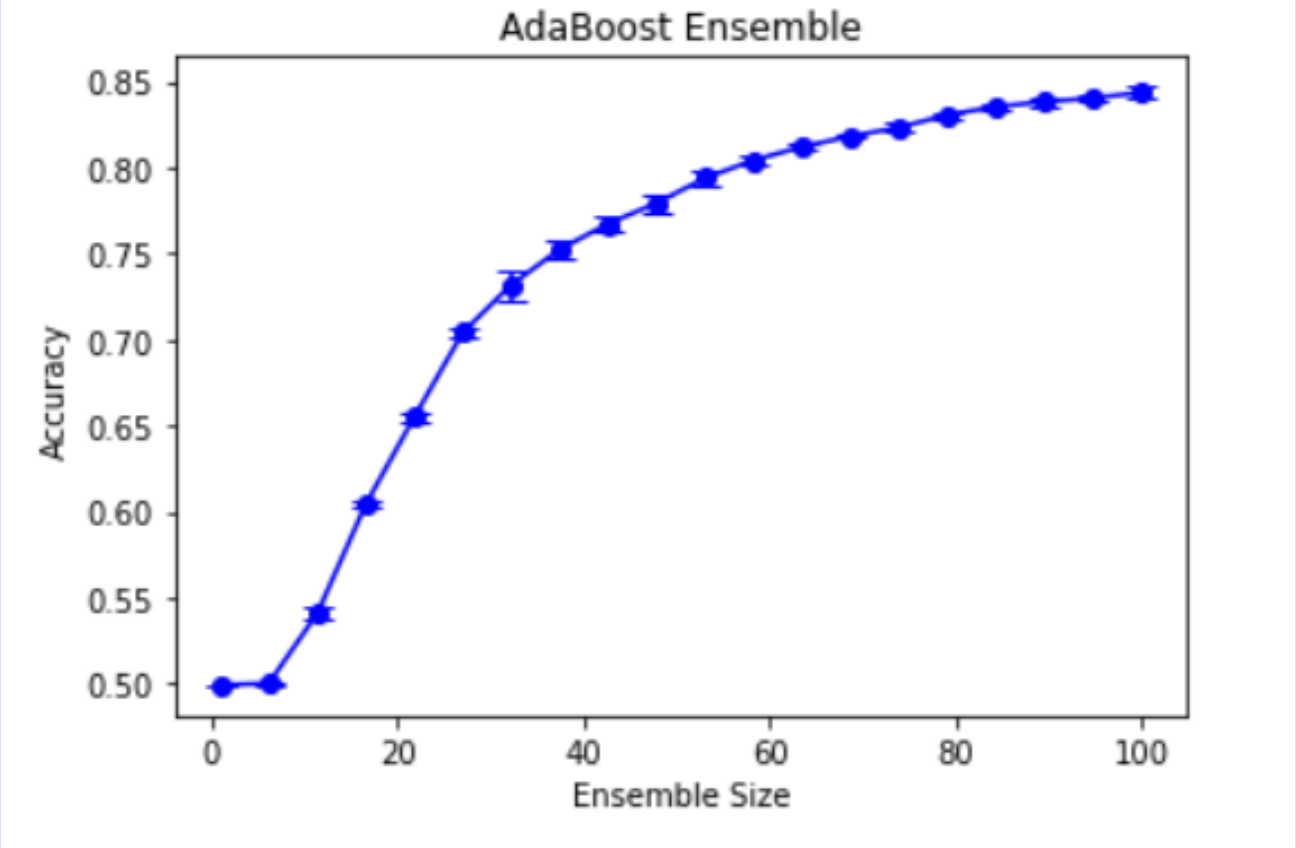


Learning Curve for n_est = 10



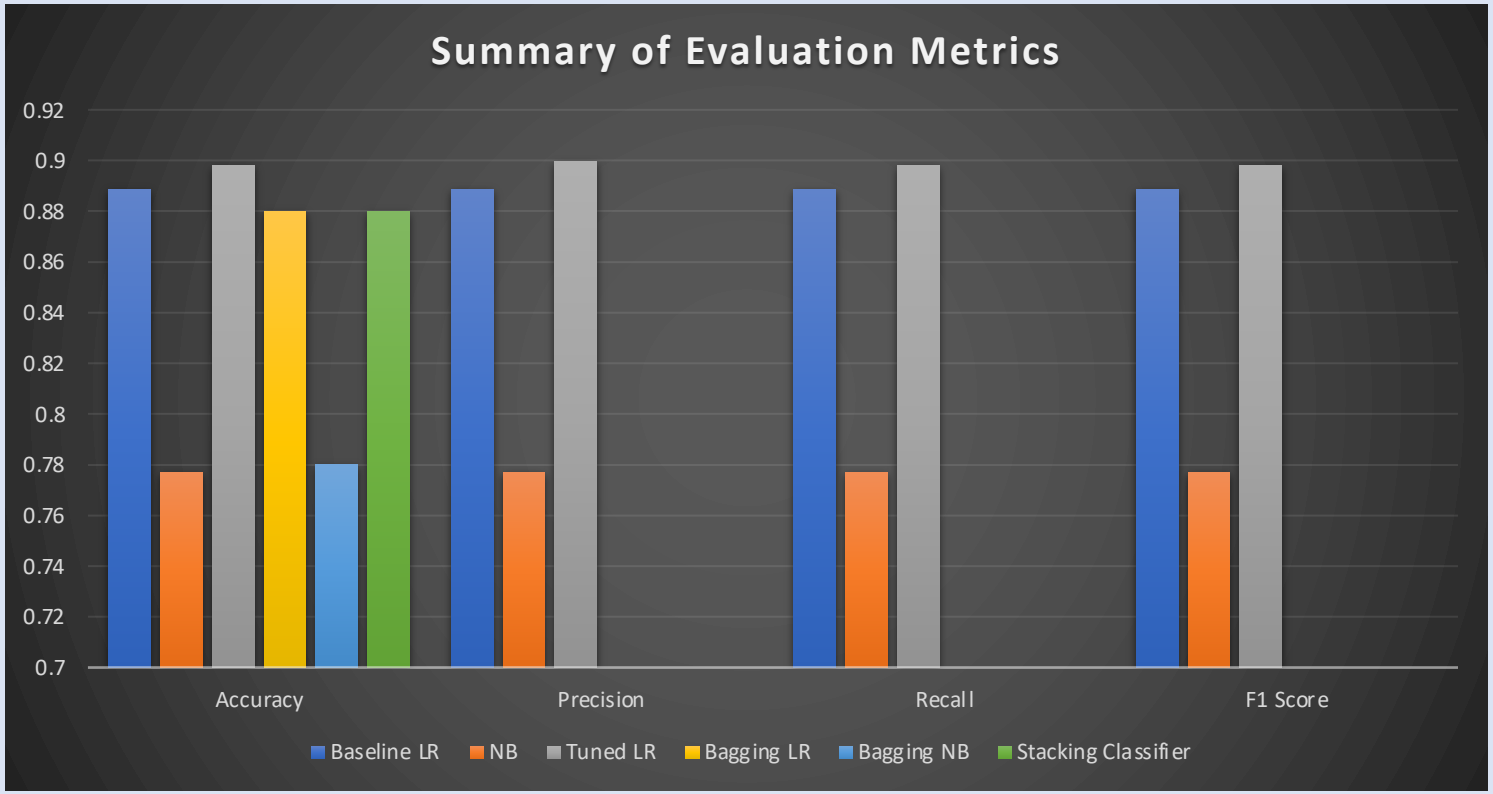
Ensemble Learning: Boosting (cont.)

Classifier Accuracy

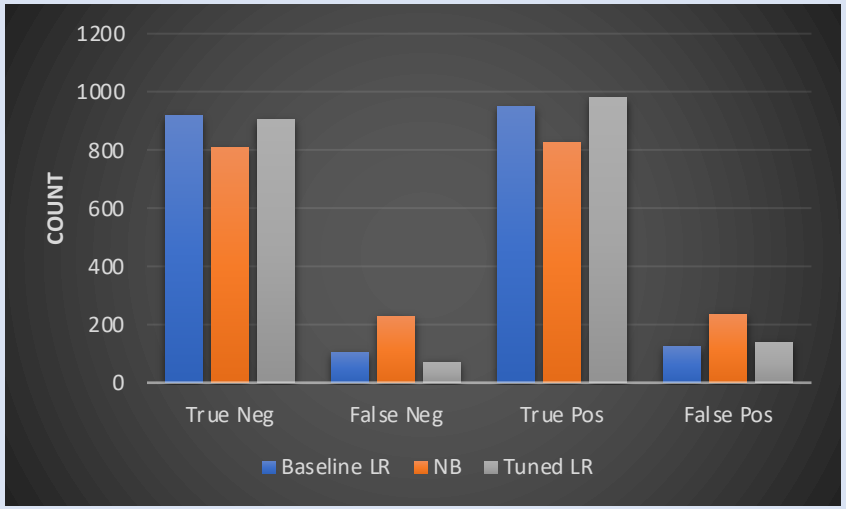


Comparison of Results

Evaluation Metrics for all models



False Positive and False Negative for Logistic Regression and Naïve Bayes



Model	Baseline LR	NB	Tuned LR	Bagging LR	Bagging NB	Stacking Classifier
Accuracy	0.89	0.78	0.90	0.88	0.78	0.88

Comparison of Results and Conclusions

- The tuned Logistic Regression model had the best accuracy compared to the ensemble methods, the Naïve Bayes model and random forest model
- The tuned Logistic Regression model's accuracy was slightly better than the ensemble methods (+1%)
  - The ensemble methods runtimes were consistently an order of magnitude larger than that of the tuned Logistic Regression model.
- Moving towards milestone 4, the tuned Logistic Regression model will be used, and feature engineering will be performed.

Thank you
Questions?

https://blog.tcea.org/instructional-coaching-questions/