PREDICTIVE MODELING OF BANK MARKETING

KRISTINE PETROSYAN
FLATIRON SCHOOL
MODULE 3 PROJECT

Who are subscribers for term deposits?

- Increase Sales/ more customers
- Reach customers and their data
- Run predictive modeling
- Efficient marketing campaign with Lowe cost
- Max customer lifetime value

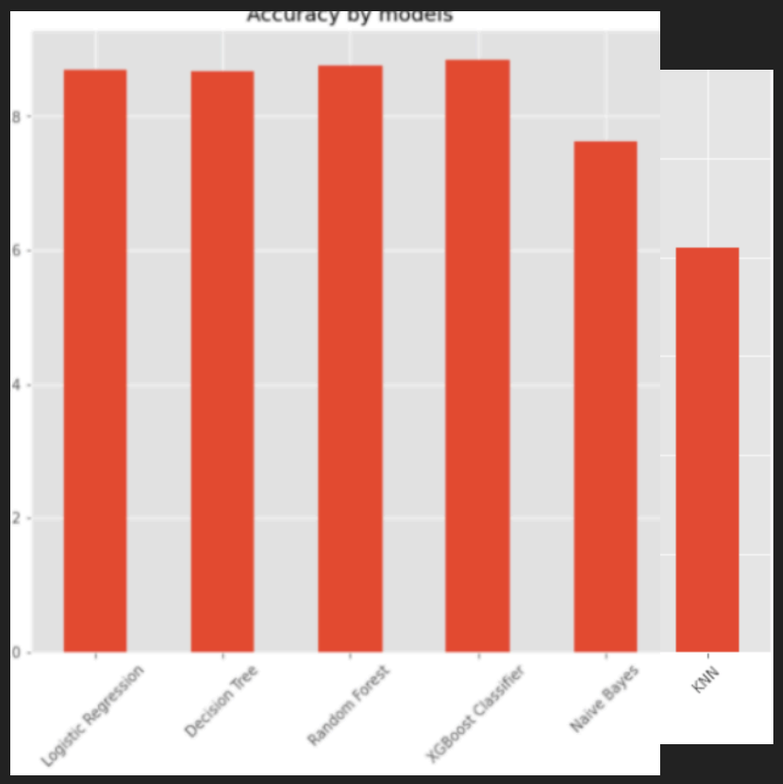
BUSINESS VALUE

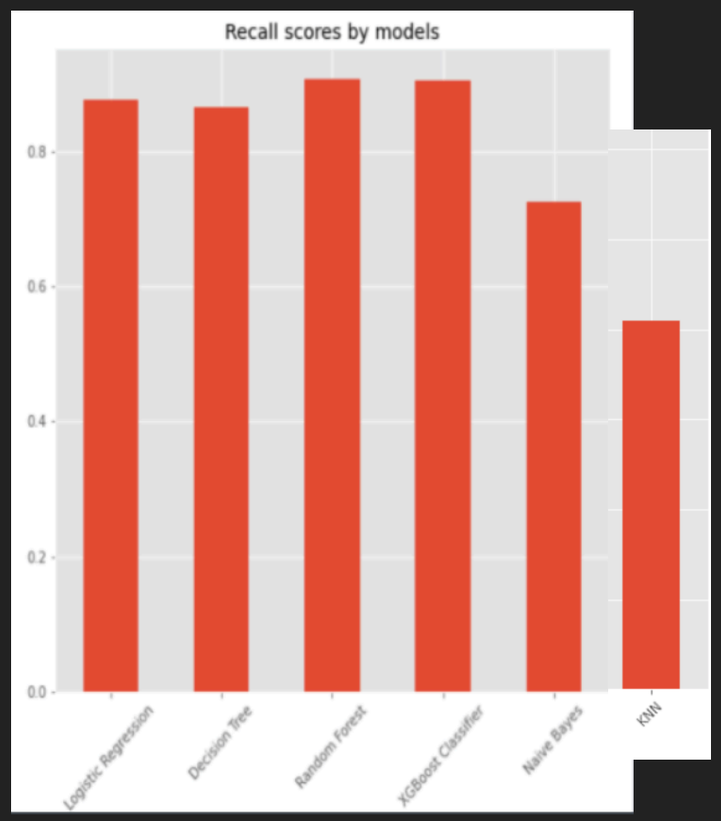
- Increase number of Subscriptions
- Use machine learning algorithms to increase the revenue
- Lower the cost of campaign

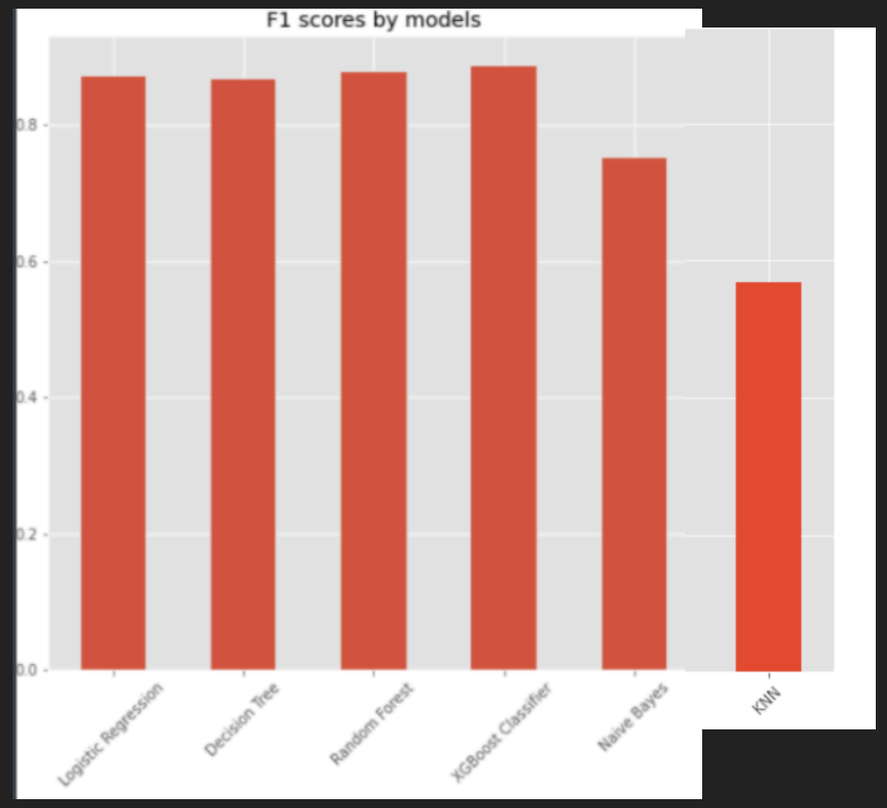
MODELING

- Cleaning and exploring
- Applying machine learning models
- Tuning hiperparameters
- Greadsearch
- Pipelines
- Final comparison of models by test/train accuracy, training time
- Predictions

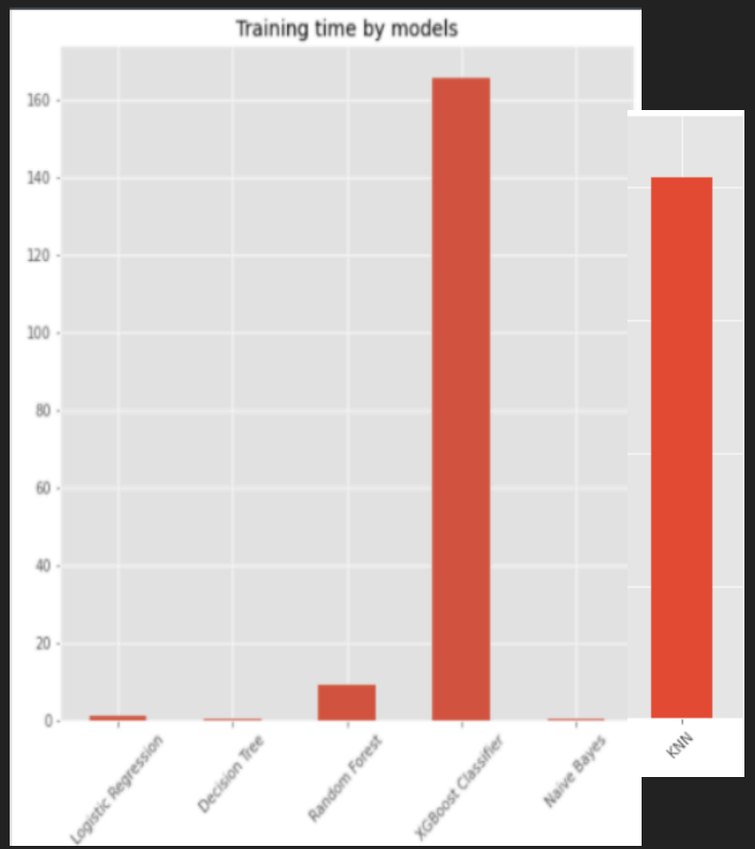
MODEL COMPARISON





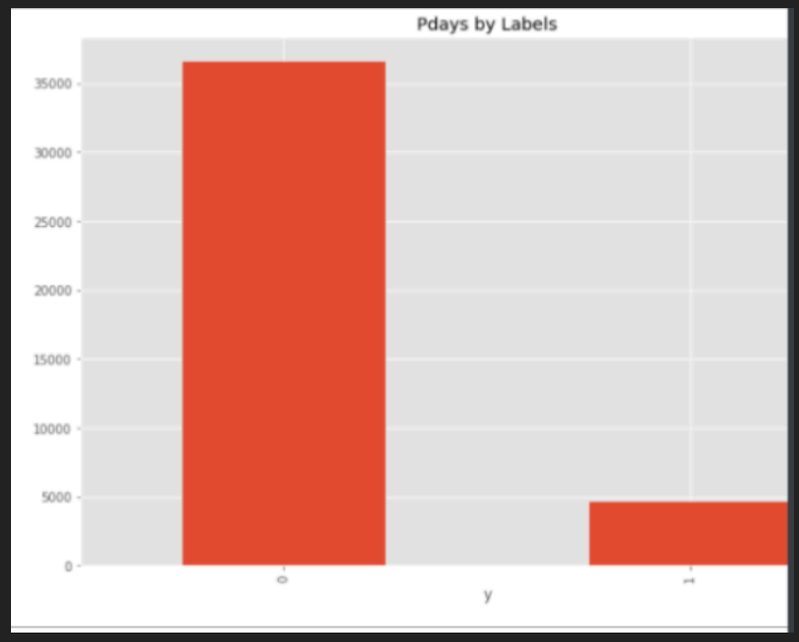


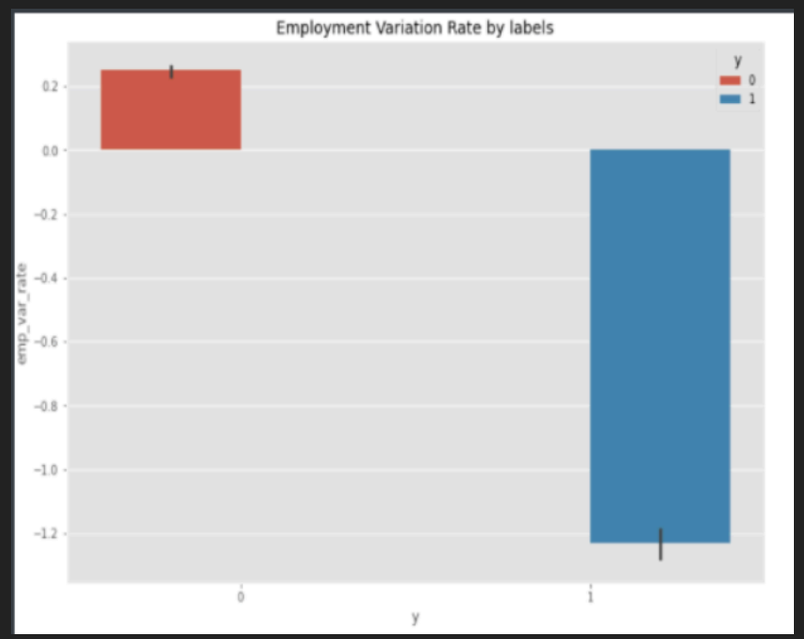
	Training time	Training score	Test score	f1 Score	Recall Score
Logistic Regression	1.089384	0.886490	0.870018	0.870655	0.874945
Decision Tree	0.565084	0.890301	0.866623	0.866286	0.864104
Random Forest	9.143574	0.899511	0.875000	0.878867	0.906921
XGBoost Classifier	165.490196	0.909870	0.884691	0.886884	0.904074
Naive Bayes	0.365406	0.759392	0.761662	0.752347	0.724047
KNN	2.010656	0.916157	0.892396	0.350528	0.256652



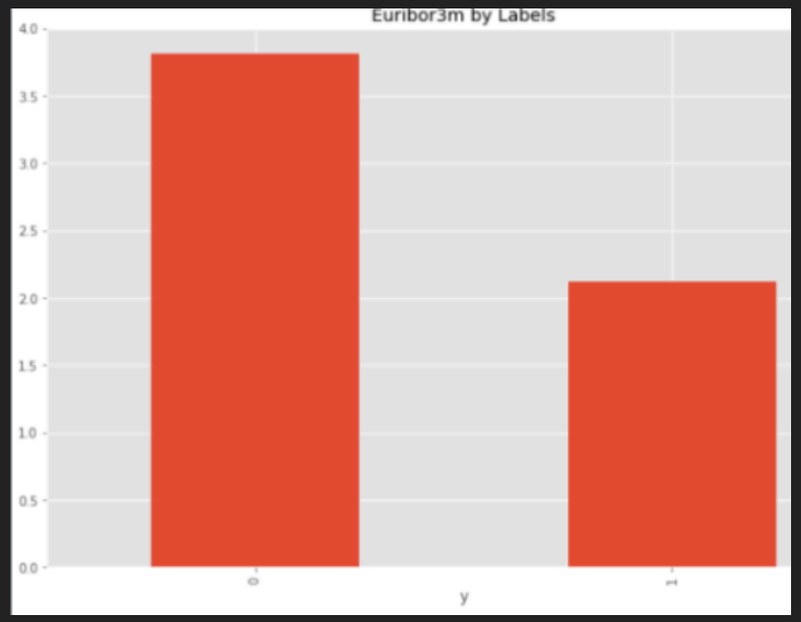
INSIGHTS

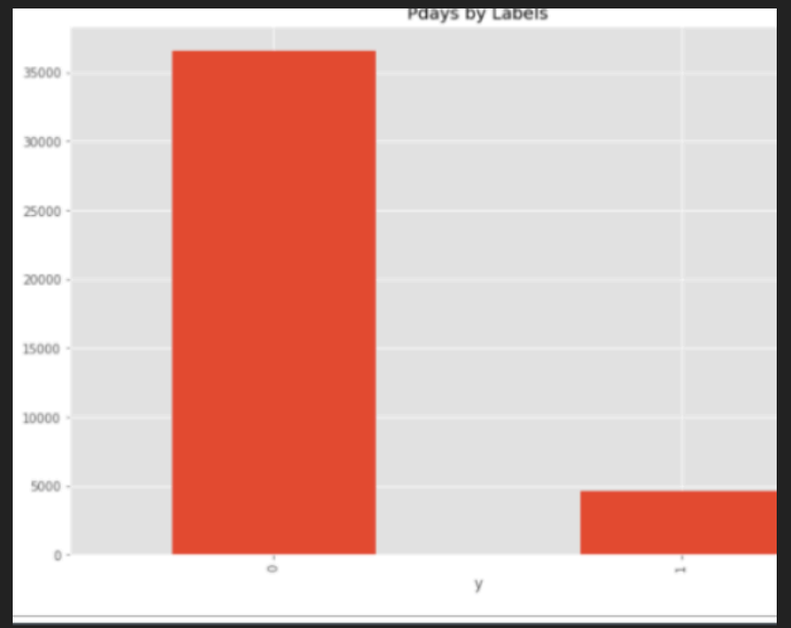
- Customers contacted after high PDAYS are more likely to refuse term deposit.
- After 10 campaigns likelihood that customer will reject the offer is higher.
- Lower Euribor3m rates lead to positive customer response.
- Lower Emp_var_rate lead to positive customer response.
- Customers likely to accept her deposits ranging from 4963.6 to 5099.1 and more likely to reject range of 5191 to 5228.1

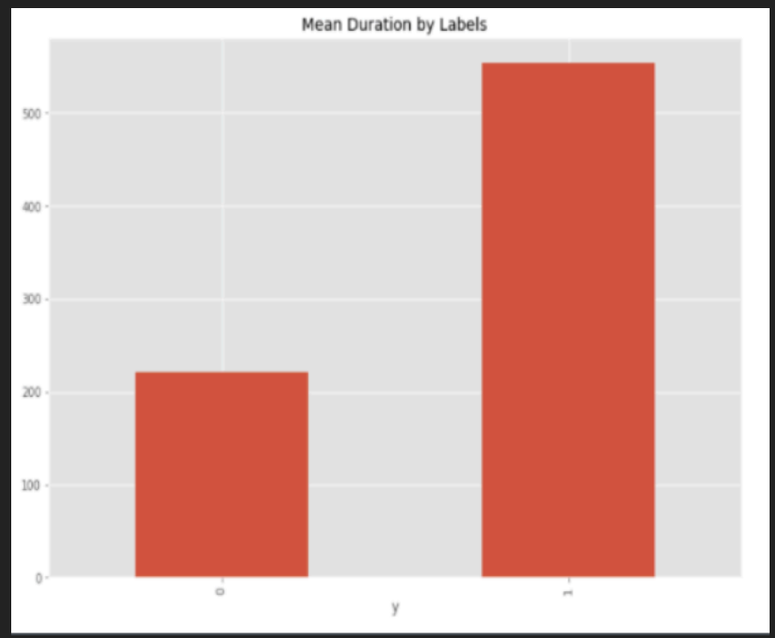




FINDINGS/ RECOMMENDATIONS







- Customers are more likely to sign up for deposits under good economic conditions.
- Customers contacted after low plays are more likely to accept deposit offer.
- Telemarketing campaigns with too much focus may affect customers negatively. Campaigns should not exceed 10.
- Longer duration shows customer’s positive response about term deposit.

FUTURE WORK

XGBOOST PERFORMANCE MODELS WORK THE BEST WITH LARGE DATA

WITH FURTHER ANALYSIS ON CUSTOMER BEHAVIOR MORE INSIGHTS COULD BE GAINED

DIFFERENT REGRESSION MODELS CAN BE APPLIED TO PREDICT DURATION

MACHINE LEARNING DATASET RELIES ON EXISTING DATA BUT IN REAL LIFE IT WILL BE DIFFERENT

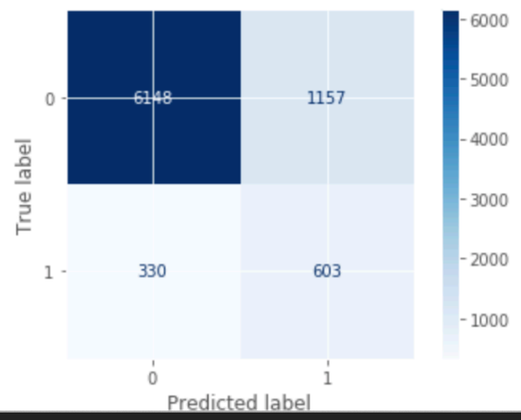
THANK YOU

Kristine Petrosyan
KRISTINELPETROSYAN@GMAIL.COM

GridSearch for Logistic Regression

```
Training Accuracy: 74.33%
Validation Accuracy: 81.95%
F1 Score: 0.4478%


Classification Report:
----------------------
              precision    recall  f1-score   support

           0       0.95      0.84      0.89      7305
           1       0.34      0.65      0.45       933

    accuracy                           0.82      8238
   macro avg       0.65      0.74      0.67      8238
weighted avg       0.88      0.82      0.84      8238
```

Out[54]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fa5348fe3c8>

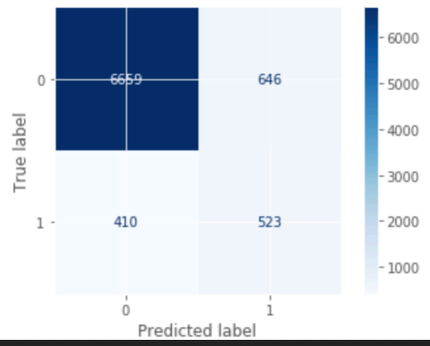


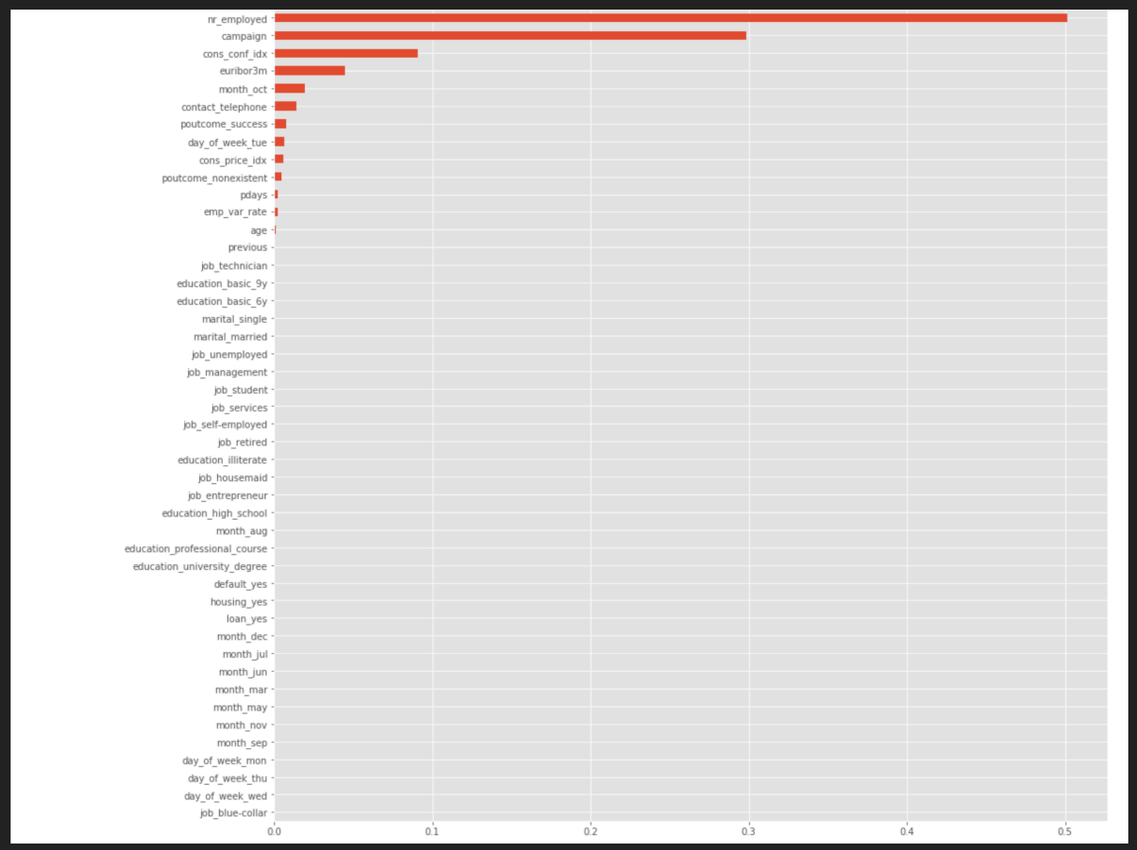
APPENDIX

- GridSearch for Decision Tree

```
Training Accuracy: 80.06%
Validation Accuracy: 87.18%
F1 Score: 0.4976%


Classification Report:
----------------------
              precision    recall  f1-score   support

           0       0.94      0.91      0.93      7305
           1       0.45      0.56      0.50       933

    accuracy                           0.87      8238
   macro avg       0.69      0.74      0.71      8238
weighted avg       0.89      0.87      0.88      8238
```

Out[88]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fa4e3a0c1d0>

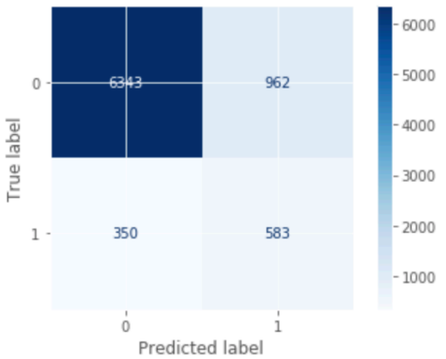


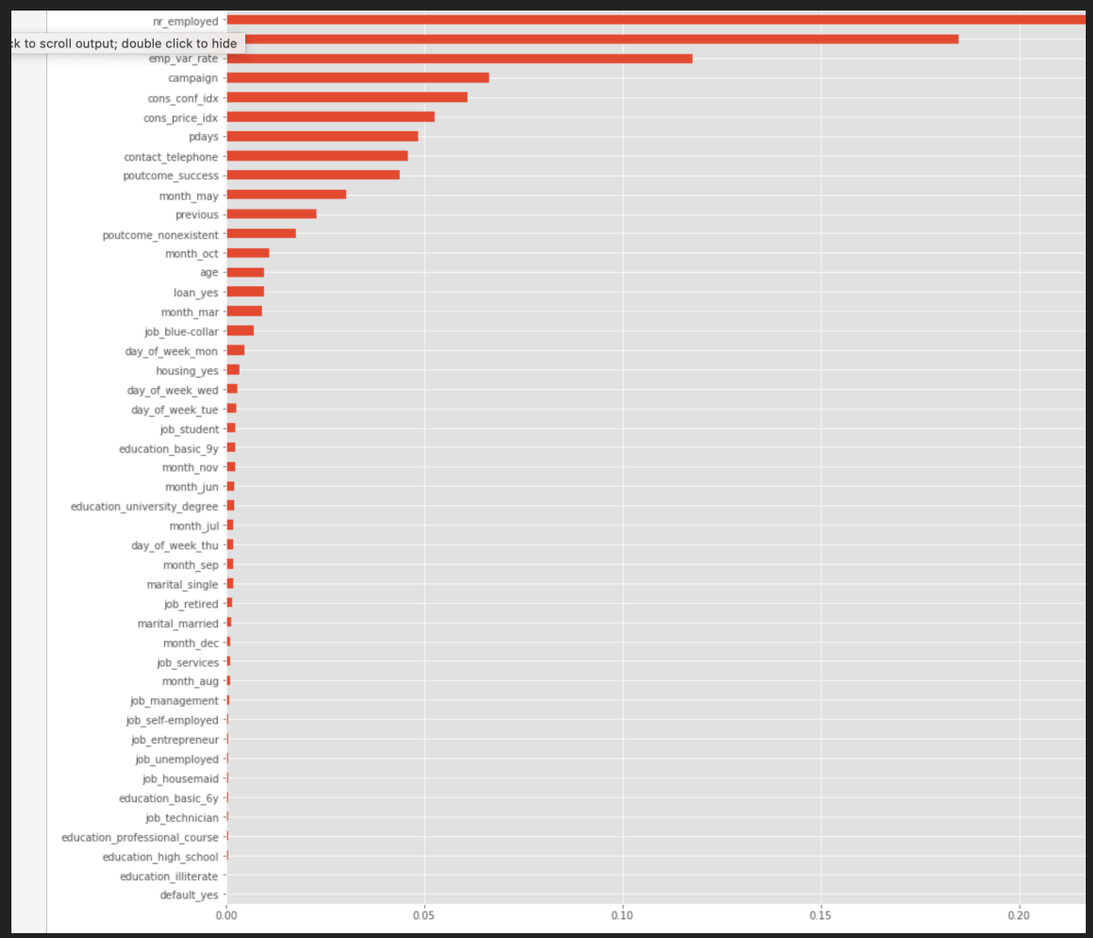


APPENDIX

- GridSearch for Random Forest

```
Training Accuracy: 75.95%
Validation Accuracy: 84.07%
F1 Score: 0.4705%


Classification Report:
----------------------
              precision    recall  f1-score   support

           0       0.95      0.87      0.91      7305
           1       0.38      0.62      0.47       933

    accuracy                           0.84      8238
   macro avg       0.66      0.75      0.69      8238
weighted avg       0.88      0.84      0.86      8238


Out[92]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fa520adb048>
```



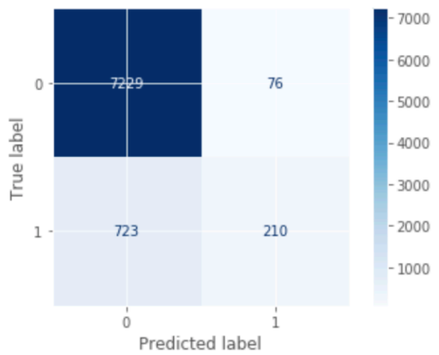


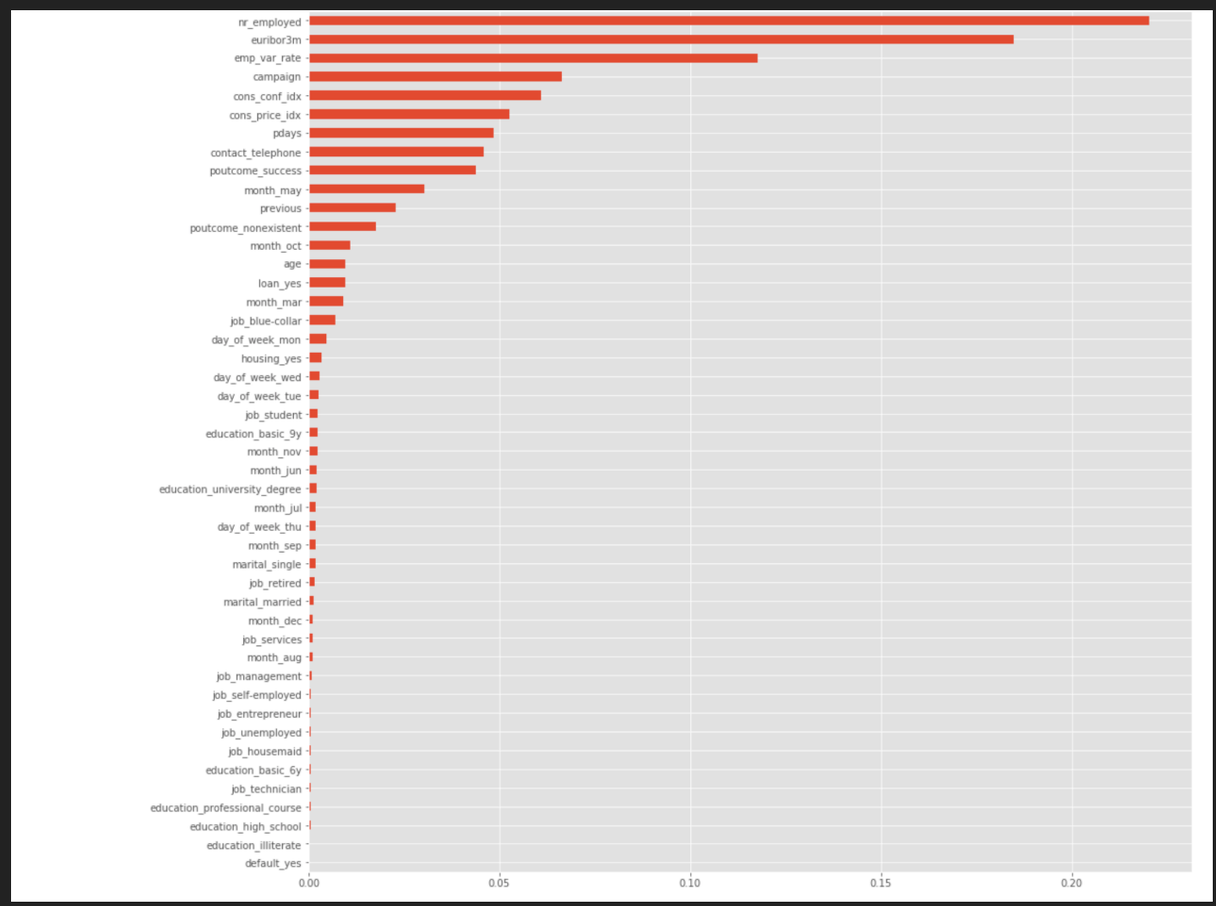
APPENDIX

- GridSearch for XGBoost

```
Training Accuracy: 59.9%
Validation Accuracy: 90.3%
F1 Score: 0.3445%


Classification Report:
----------------------
              precision    recall  f1-score   support

           0       0.91      0.99      0.95      7305
           1       0.73      0.23      0.34       933

    accuracy                           0.90      8238
   macro avg       0.82      0.61      0.65      8238
weighted avg       0.89      0.90      0.88      8238
```

Out[75]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fa5418609e8>

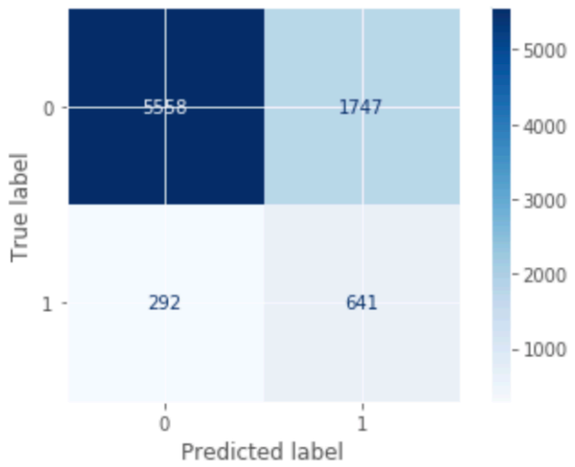




GridSearch for Naive Bayes Classifier

```
Training Accuracy: 71.45%
Validation Accuracy: 75.25%
F1 Score: 0.386%


Classification Report:
----------------------
              precision    recall  f1-score   support

           0       0.95      0.76      0.85      7305
           1       0.27      0.69      0.39       933

    accuracy                           0.75      8238
   macro avg       0.61      0.72      0.62      8238
weighted avg       0.87      0.75      0.79      8238
```

Out[69]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fa512f8d4e0>



GridSearch for KNN

```
Training Accuracy: 91.95%
Validation Accuracy: 78.45%
F1 Score: 0.3543%


Classification Report:
----------------------
              precision    recall  f1-score   support

           0       0.93      0.82      0.87      7305
           1       0.27      0.52      0.35       933

    accuracy                           0.78      8238
   macro avg       0.60      0.67      0.61      8238
weighted avg       0.86      0.78      0.81      8238
```

Out[72]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7fa5206f9198>