



# GA Think Tank

By: David Lee, Chris Johnson, & Aidan Curley



# Agenda

- Overview
- Problem Statement
- Process
- Modeling
- Conclusions
- Next Steps
- Q & A



## **Problem Statement**

**Do Certain Features Have an Impact  
on Whether Someone Makes Over  
\$50,000 Per Year?**



# Process

- Data Cleaning
- Removing Outliers
- Exploratory Data Analysis
- Modeling



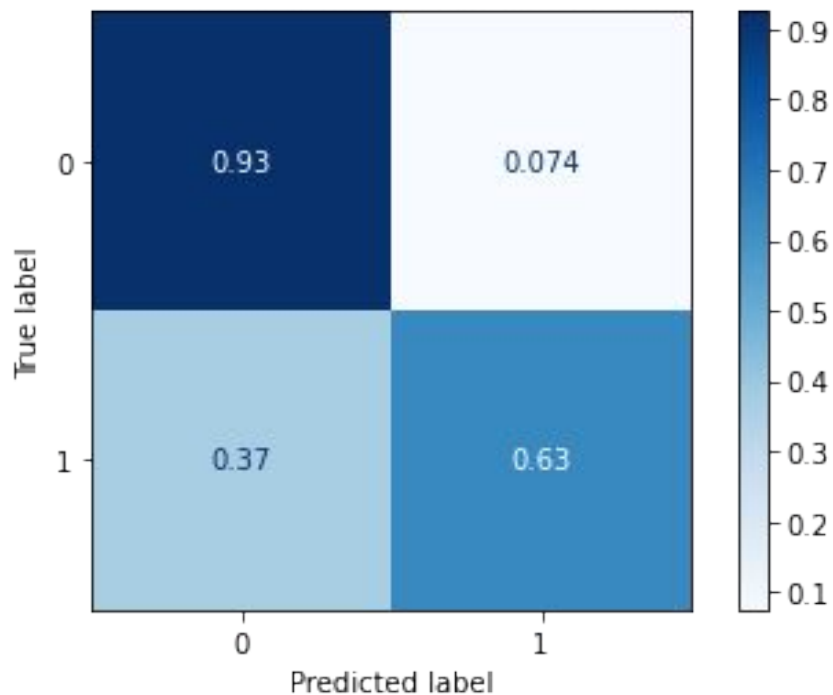
## Baseline Accuracy

76% accuracy score in predicting someone to make less than \$50,000 per year.



The Accuracy score is: 85.6%  
The Missclassification rate is: 14.45%  
The Sensitivity is: 63.5%  
The Specificity is: 92.60000000000001%  
The Precision is 73.0%

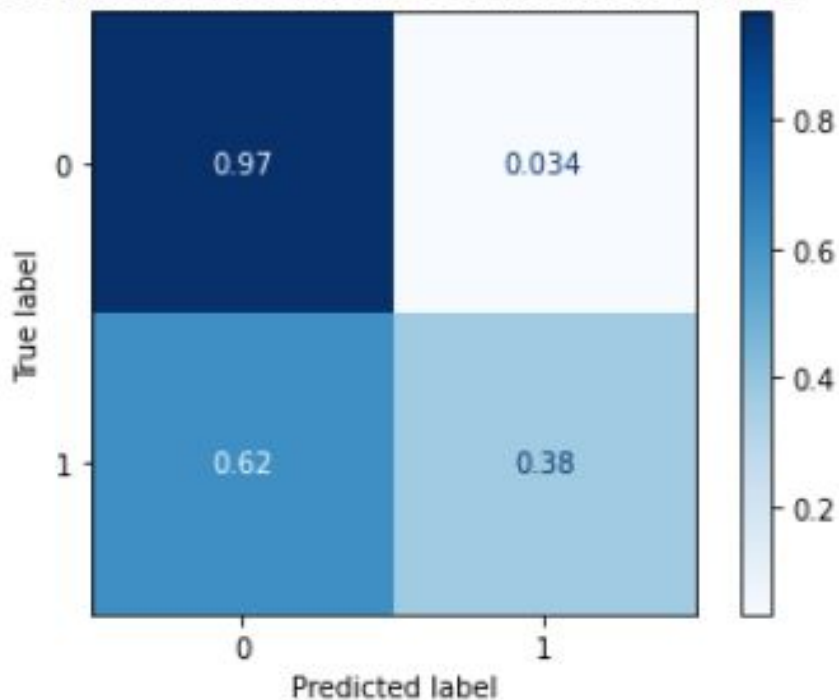
Confusion Matrix for Benchmark Model





The Accuracy score is: 83.6%  
The Missclassification rate is: 16.41%  
The Sensitivity is: 37.6%  
The Specificity is: 96.7%  
The Precision is 76.34%

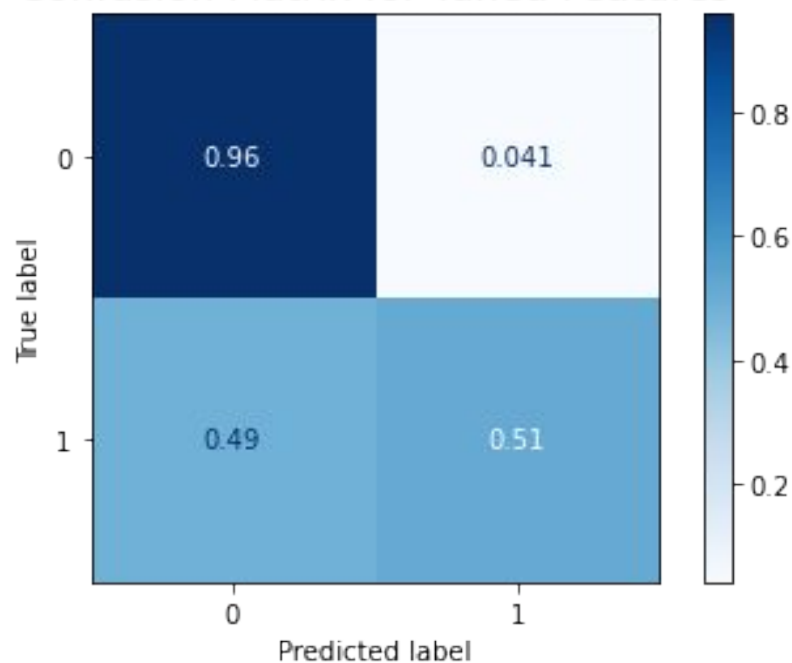
Confusion Matrix for Continuous Model





The Accuracy score is: 86.0%  
The Missclassification rate is: 13.96%  
The Sensitivity is: 51.300000000000004%  
The Specificity is: 95.89999999999999%  
The Precision is 78.10000000000001%

Confusion Matrix for Tuned Features







## Why Native Country and Working Class Does Not Play a Role in Someone Making over \$50,000

Tuned Model With All Features	Tuned Model Without Native Country and Working Class
Train Score: 85.75% Test Score: 85.7%	Train Score: 86.65% Test Score: 86.04%



# Conclusions

Working class and native country make no difference in determining if someone makes over \$50,000 per year



## Next Steps

- In future steps we plan to create different classification models such as logistic regression, k-nearest neighbors, support vector machines, and a neural network.
- We would like to collect more data and continue to improve on our models.

---

# Questions?