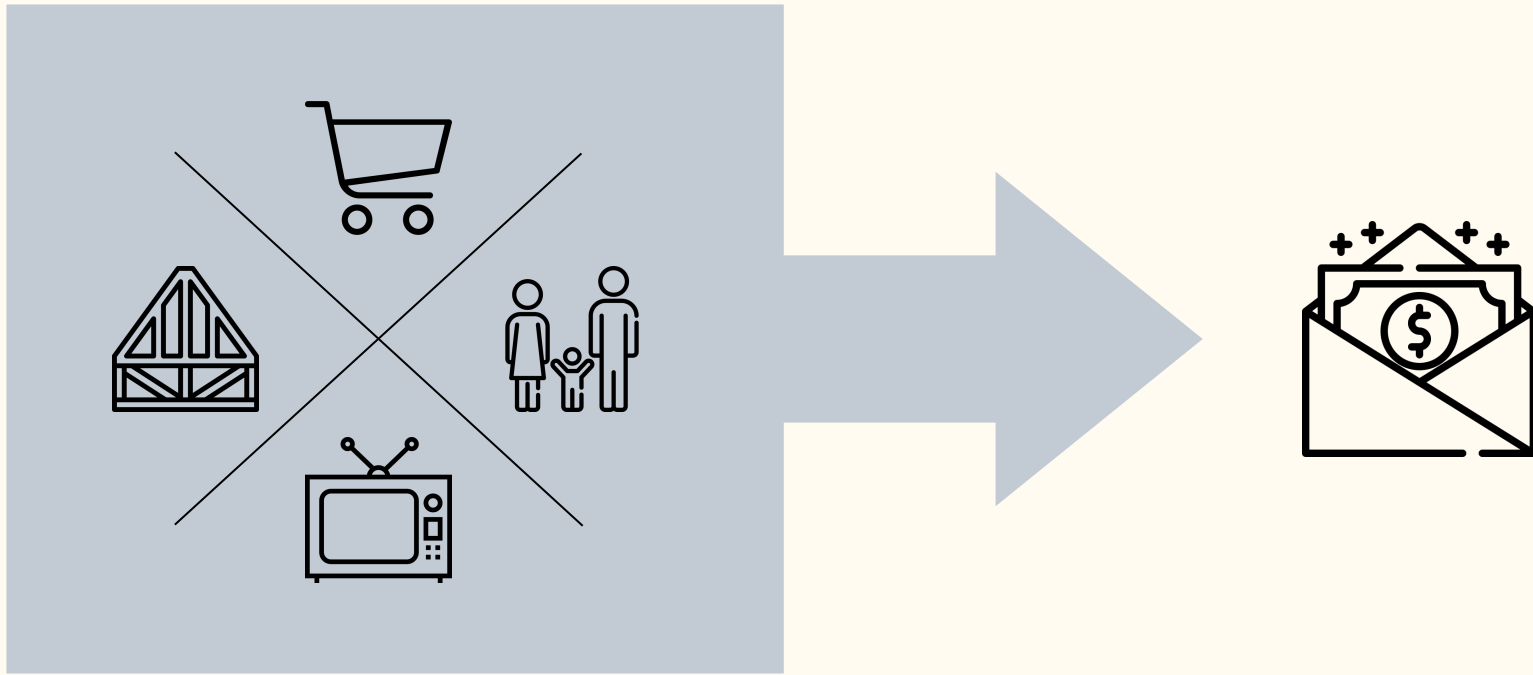# Predicting Household Income in the Philippines with Classification Modeling

—

Kyle Aguilar

The Dataset: Features and Labels
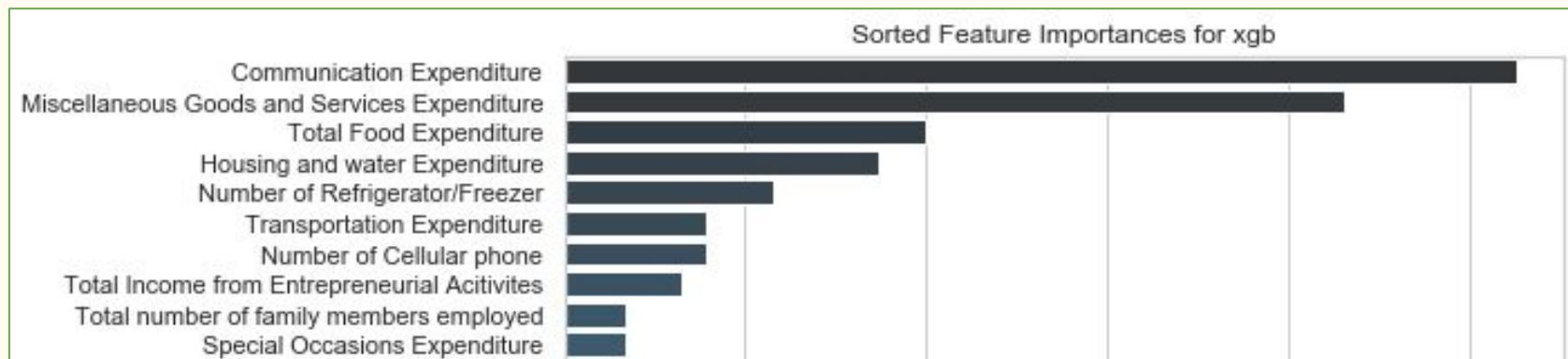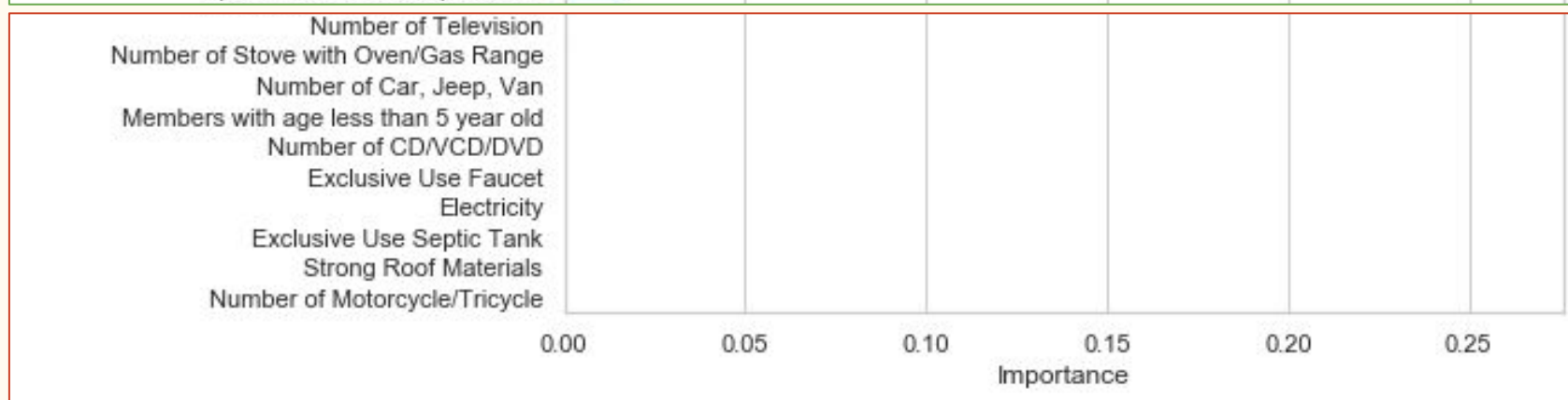
Modeling

Model_Name	F1_Score	Accuracy_Score	Precision_Score	Recall_Score
LogReg_Baseline	0.8614	0.8737	0.9596	0.7815
LogReg_FeatureSelection	0.8621	0.8743	0.9602	0.7821
KNN_Baseline	0.8395	0.8546	0.9425	0.7567
DecisionTree	0.8612	0.8602	0.8589	0.8634
RF_Baseline	0.9012	0.9024	0.9169	0.8861
RF_GridSearch	0.9064	0.9066	0.9124	0.9005
XGBoost_Baseline	0.911	0.9121	0.9271	0.8954
XGBoost_GridSearch	0.9158	0.9166	0.9296	0.9024

XGBoost - Feature Importance

Sorted Feature Importances for xgb

Top 10
- Communication Expenditure
- Miscellaneous Goods and Services Expenditure
- Total Food Expenditure
- Housing and water Expenditure
- Number of Refrigerator/Freezer
- Transportation Expenditure
- Number of Cellular phone
- Total Income from Entrepreneurial Acitivites
- Total number of family members employed
- Special Occasions Expenditure

Bottom 10
- Number of Television
- Number of Stove with Oven/Gas Range
- Number of Car, Jeep, Van
- Members with age less than 5 year old
- Number of CD/VCD/DVD
- Exclusive Use Faucet
- Electricity
- Exclusive Use Septic Tank
- Strong Roof Materials
- Number of Motorcycle/Tricycle

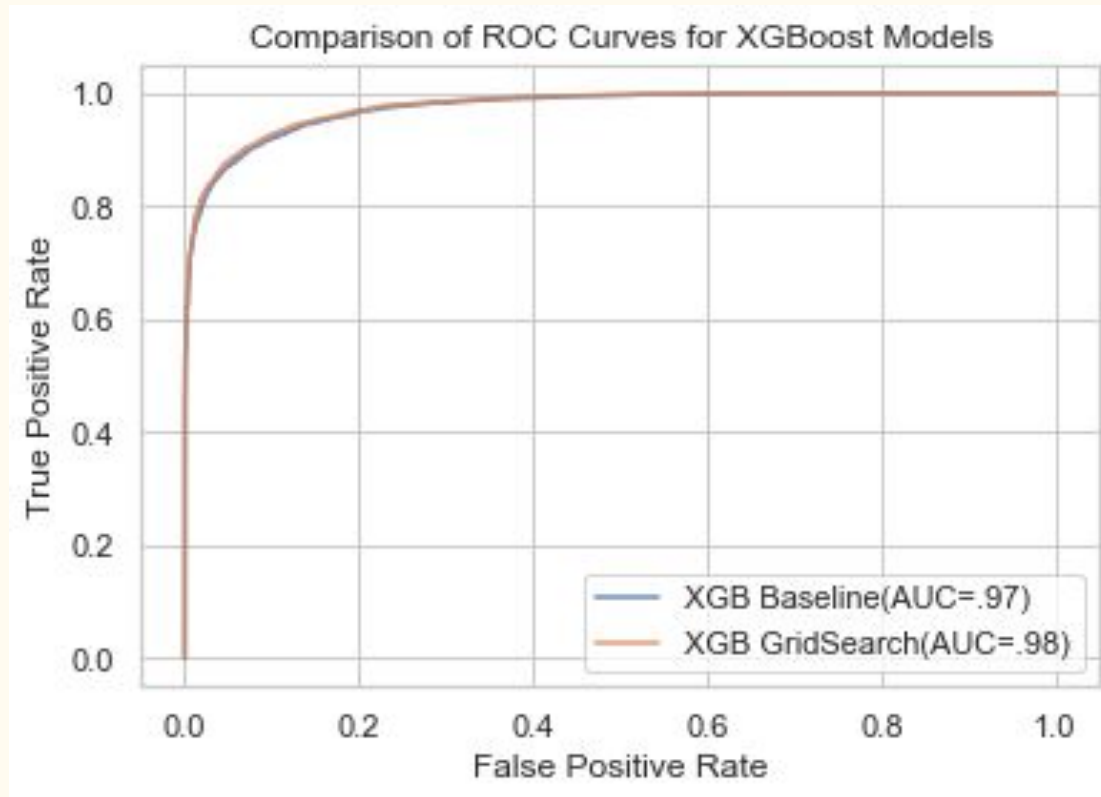# XGBoost - Confusion Matrix (Precision/Recall)
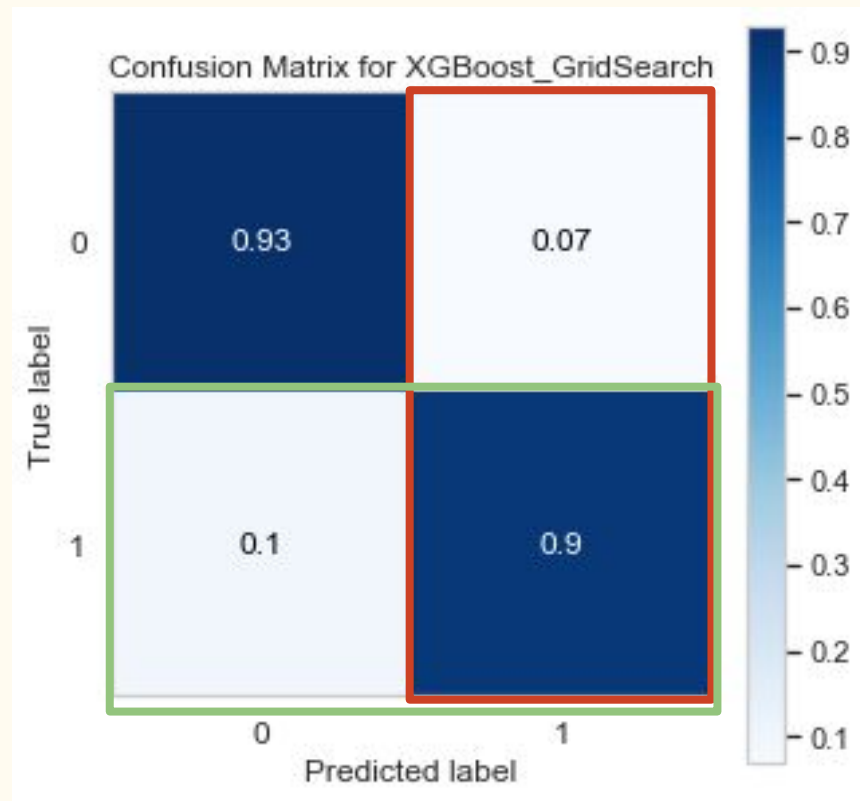
XGBoost - ROC Curve

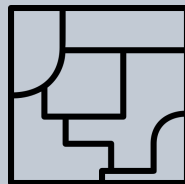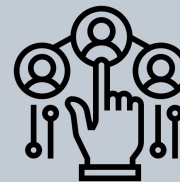# Future Work

Changes
Over Time

Regional
Specificity

Feature
Engineering

Thank you.