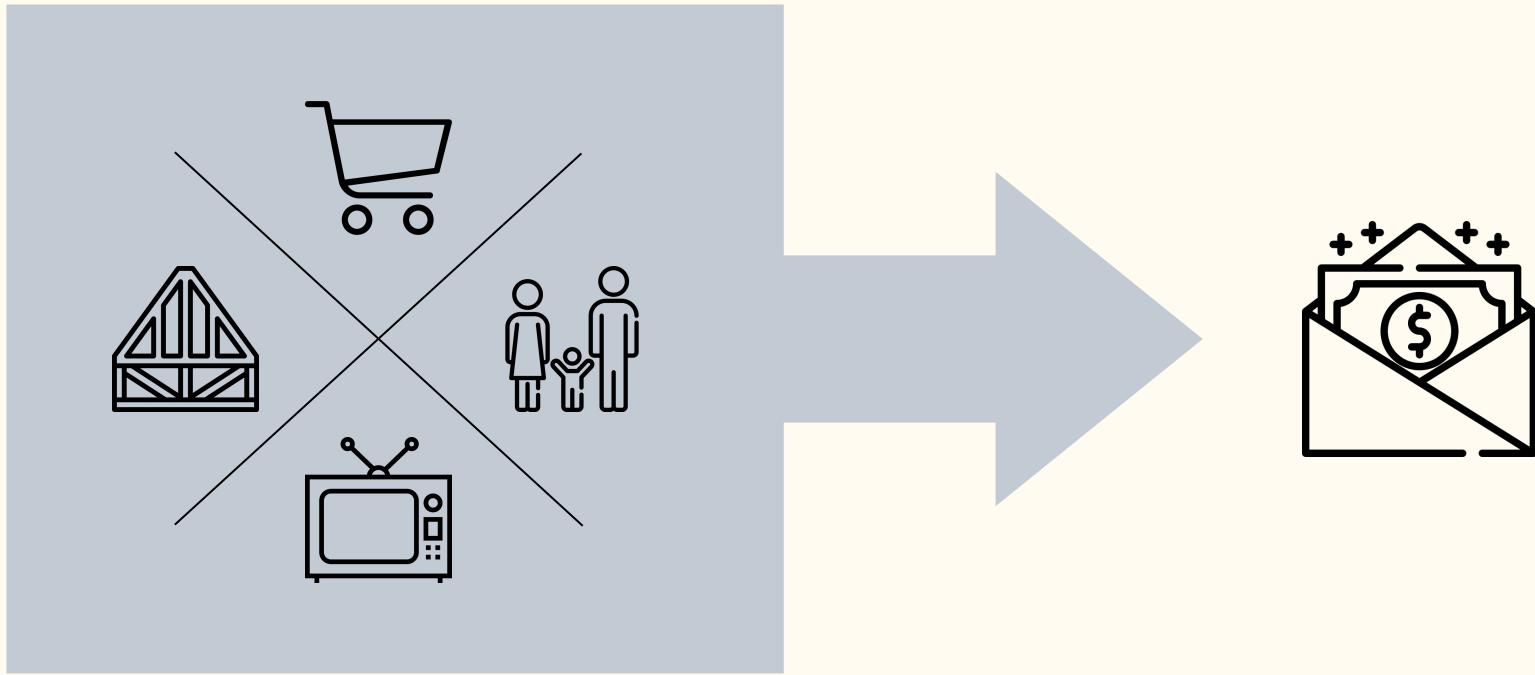# Predicting Household Income in the Philippines with Classification Modeling

—

Kyle Aguilar

The Dataset: Features and Labels
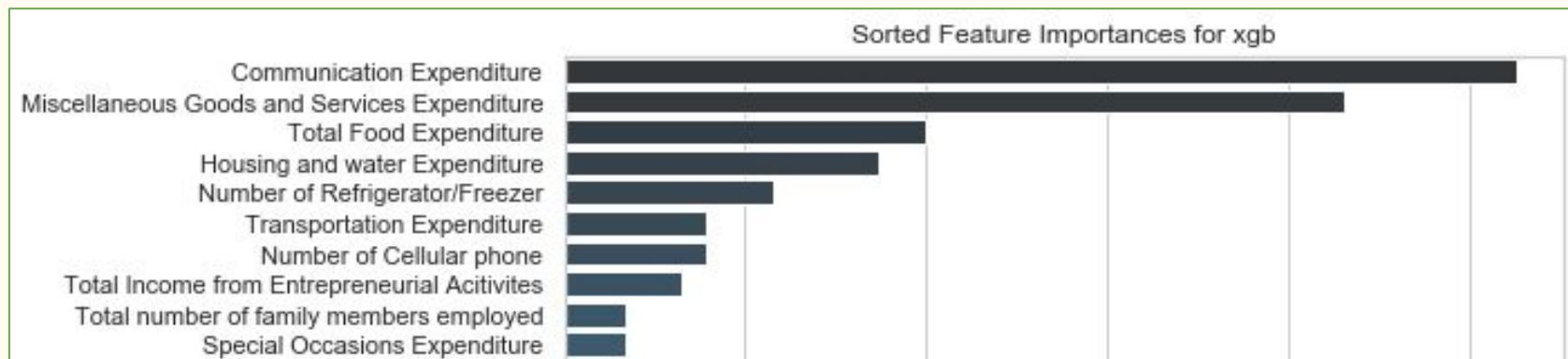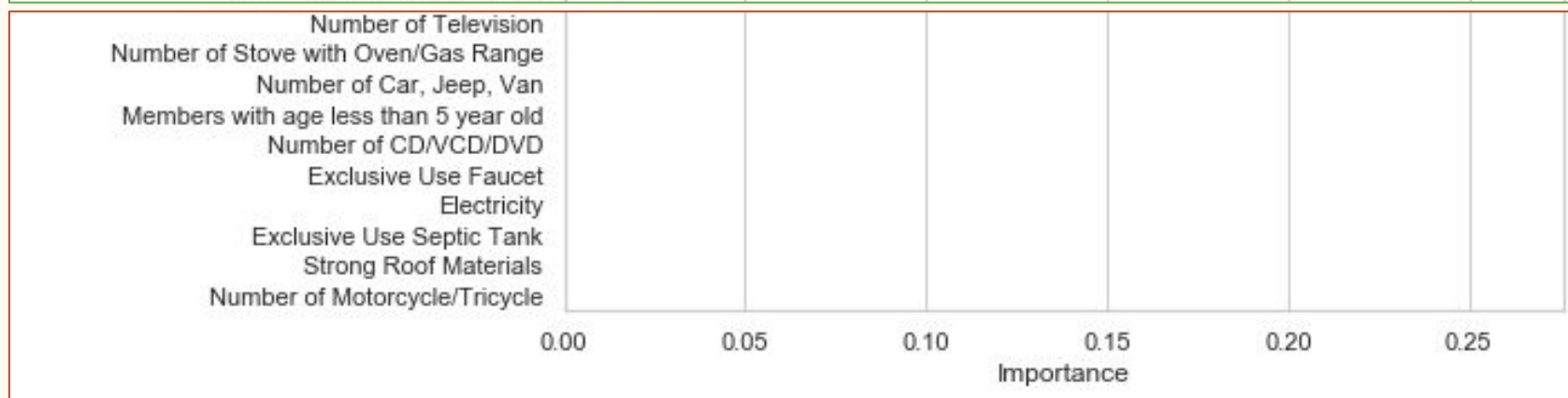
Modeling

Model_Name	F1_Score	Accuracy_Score	Precision_Score	Recall_Score
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

Importance: 0.00 0.05 0.10 0.15 0.20 0.25

XGBoost - Confusion Matrix (Precision/Recall)

Confusion Matrix for XGBoost_GridSearch

XGBoost - ROC Curve

Comparison of ROC Curves for XGBoost Models
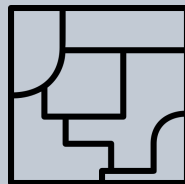
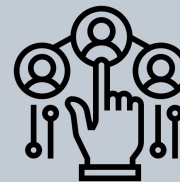- XGB Baseline(AUC=.97)
- XGB GridSearch(AUC=.98)
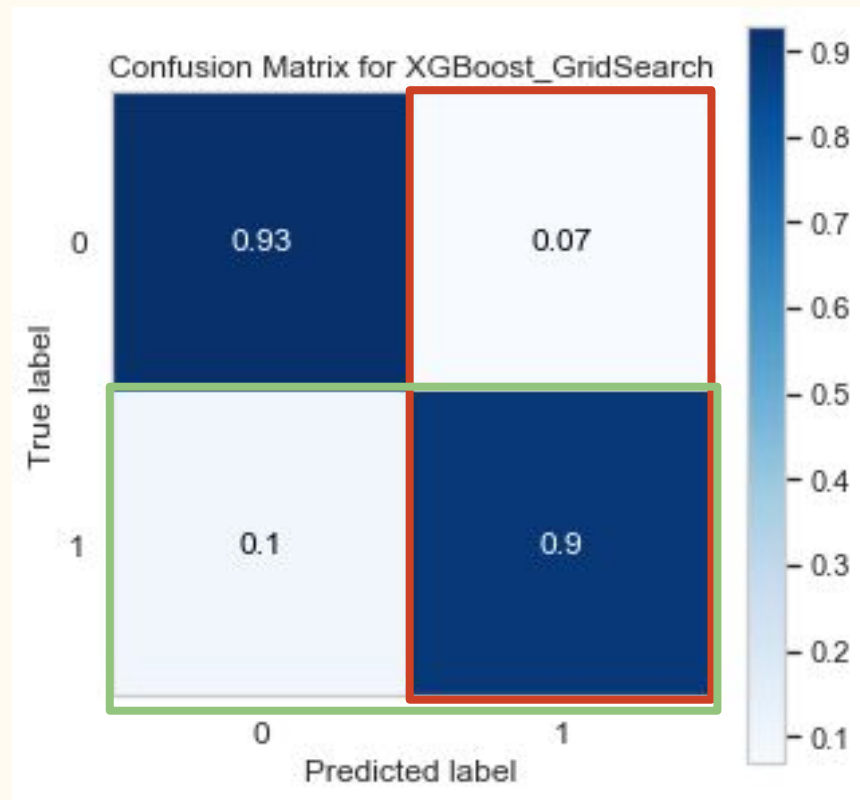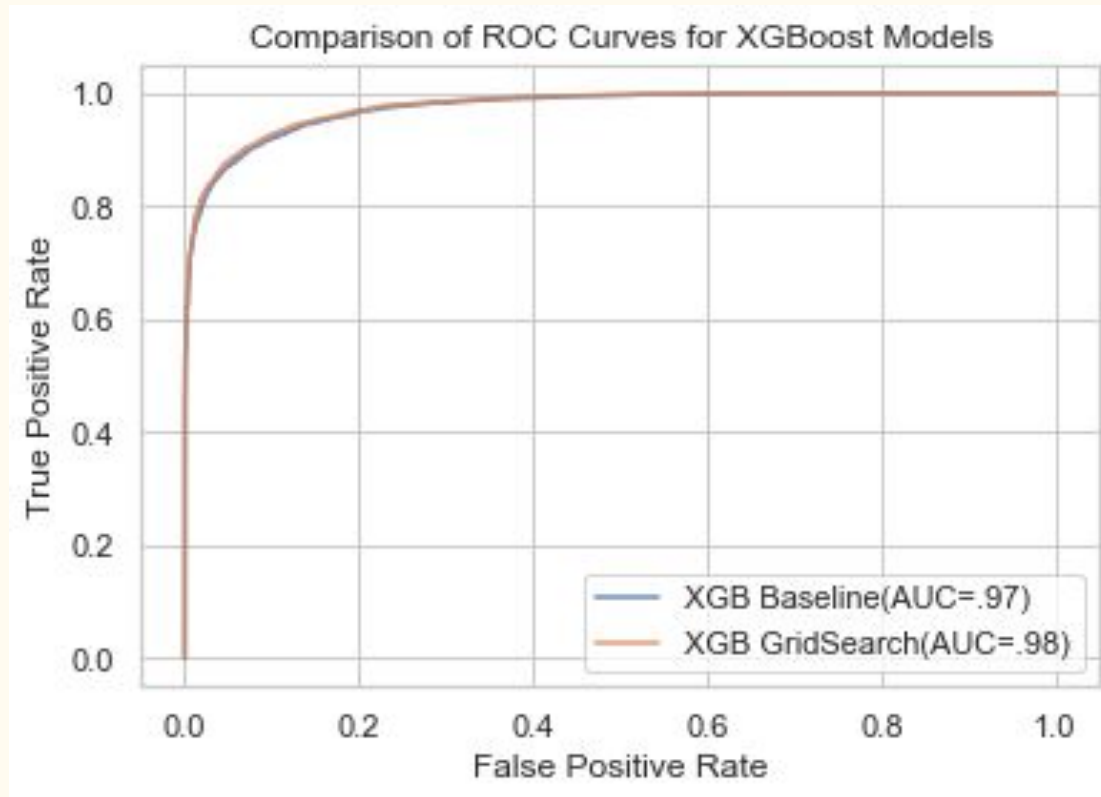
Future Work

Changes
Over Time

Regional
Specificity

Feature
Engineering

Thank you.