# Identifying Prospects based on User Data

*Sera Diebel*

# Model Details

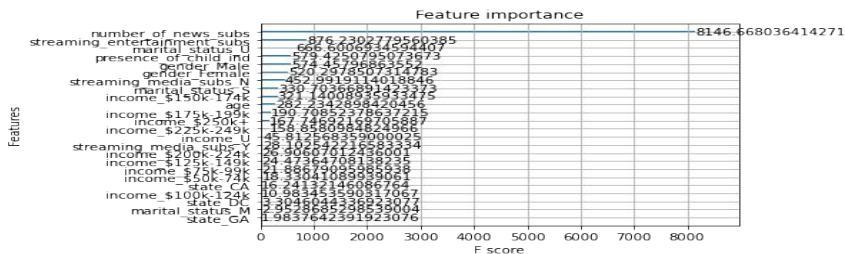| | |
|---|---|
| **Model Objective** | Predict the probability that a user with a mobile app will listen to podcasts |
| **Data Inputs** | <ul><li>Age</li><li>Gender</li><li>Marital status</li><li>Income band</li><li>State</li><li>Presence of a Child in the Household (Y/N)</li><li>Number of subscriptions to news services (digital and non-digital)</li><li>Streaming media subscription status (Y/N)</li><li>Streaming entertainment listening status (Y/N)</li></ul> |
| **Model Output** | 1 / 0 predictions (1 if the model predicts the user will listen to podcasts, 0 otherwise) |
| **Methodology** | ➔ Both traditional (logistic) and Machine Learning (XGBoost) approaches were explored |

# Results

| | |
|---|---|
| **Model Performance (XGBoost vs. Logistic)** | ➔ XGBoost model has superior performance (AUC 98% vs. 96%) [1 being the perfect classification]<br>➔ XGBoost predicts the outcome:<br>    ◆ more accurately (96% vs. 93% accuracy)<br>    ◆ more precisely (95% vs. 90% precision) than logistic regression |
| **Key Drivers** | XGBoost feature importance as measured by improvement in accuracy brought by a variable (gain metric)<br><br><br>Feature importance |
| **Insights** | ➢ Lower income band information is not as informative in predicting the outcome<br>➢ Among states, residing in California contributes most to the predictive power<br>➢ Odds of Female users listening to podcasts is higher than that of Male users, all else equal<br>➢ Compared to Unknown marital status, Married people are more likely to listen to podcasts, and Single users have even higher odds of listening than married users |