

Best Customer Attributes for Marketing to New Bike Buyers



Data Driven Marketing

Project Objectives:

- Find best buyers
- Target them



Finding the Right Data -...

Data Driven Metrics

Attributes and Targets

- Demographic data
 - Started with 25 features
 - Ended with 12 features
 - City, State, Country/Region, Age, Education, Occupation, Gender, Marital status, Home Ownership, Cars Owned, Children, Income
- Targets: Buyer & Non-Buyer
- Highest probabilities to buy

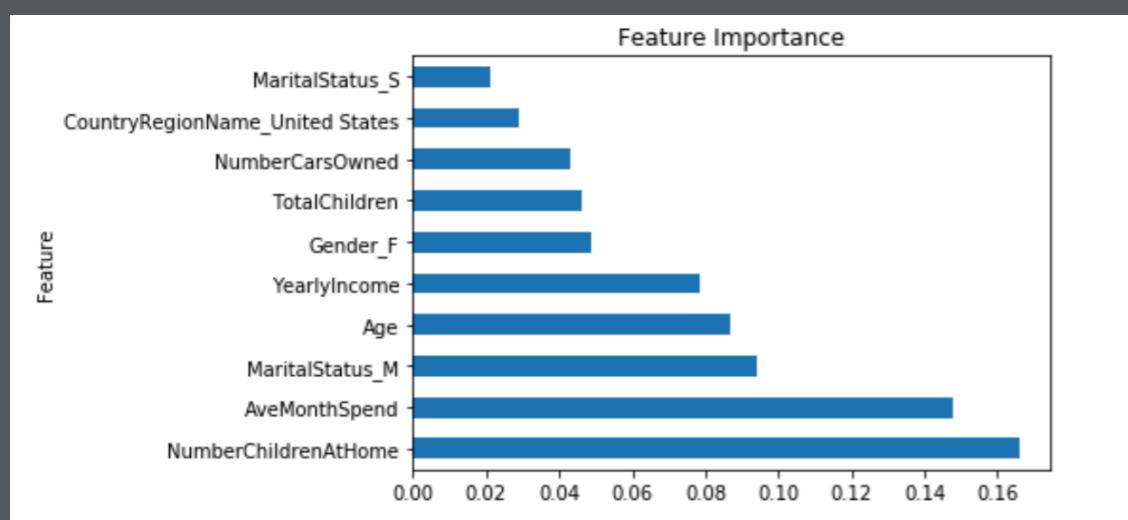


Data Driven Models

Chosen Machine Learning Model

Random Forest

Features



Accuracy: 78.2%

Pros

Remains accurate with data missing

Less variance than Decision Tree so better with large range of data items

Extremely flexible & very high accuracy

Cons

Overfitting model or the algorithm forces data to work with the data no matter what

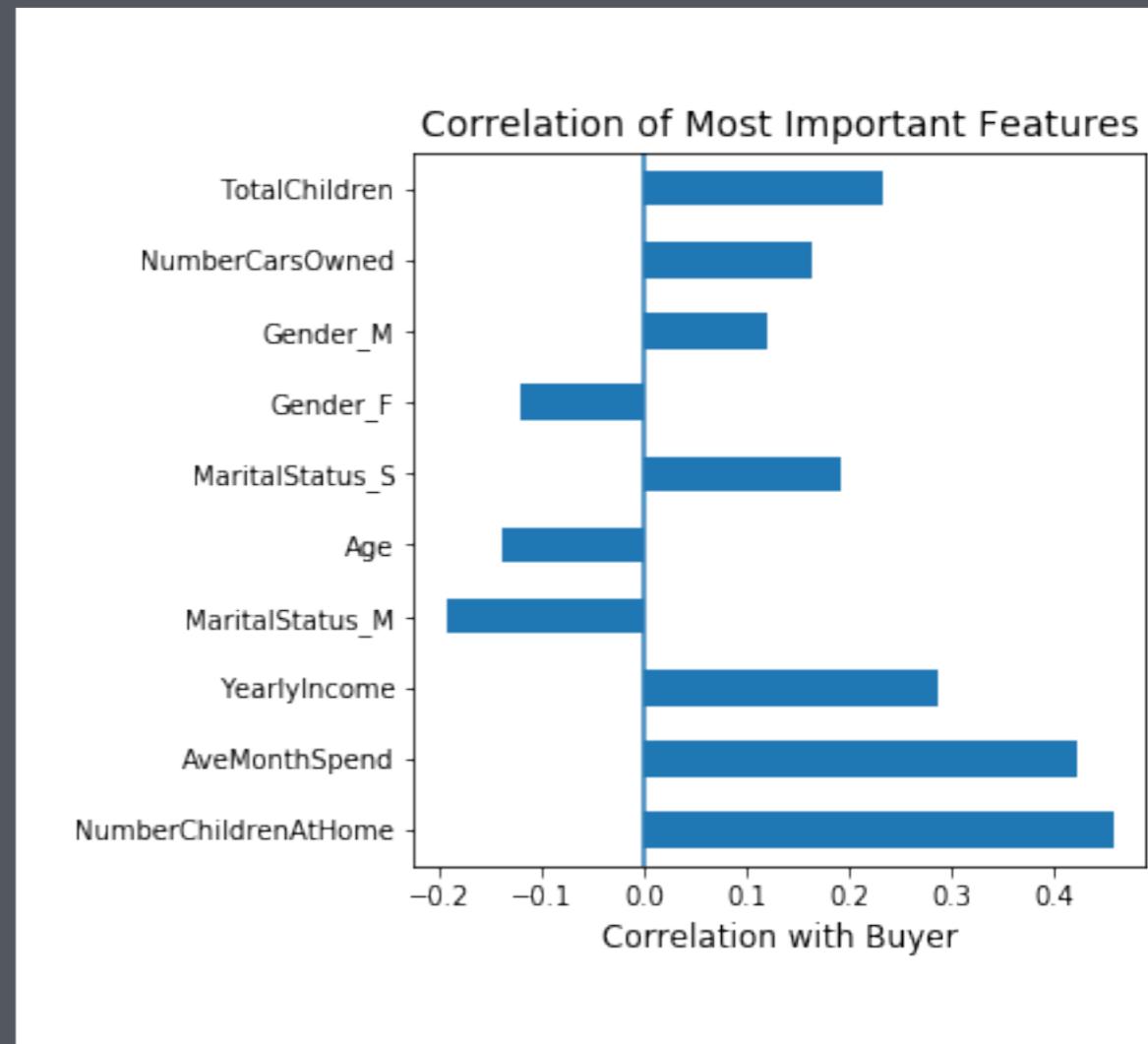
Large collections of decision trees can be difficult to grasp all the attributes relationships

ommendations

Focus Ads/Campaigns most likely to buy

Attributes to Target

1. Number of children (more = better)
2. Martial status (singles)
3. Gender (males)
4. Average monthly spend (more)
5. Age (focus on younger)
6. Education (some college or higher)



Future Work

- Expand demographics with outside data sources to expand customer base
 - Add kids data
 - Add retailers

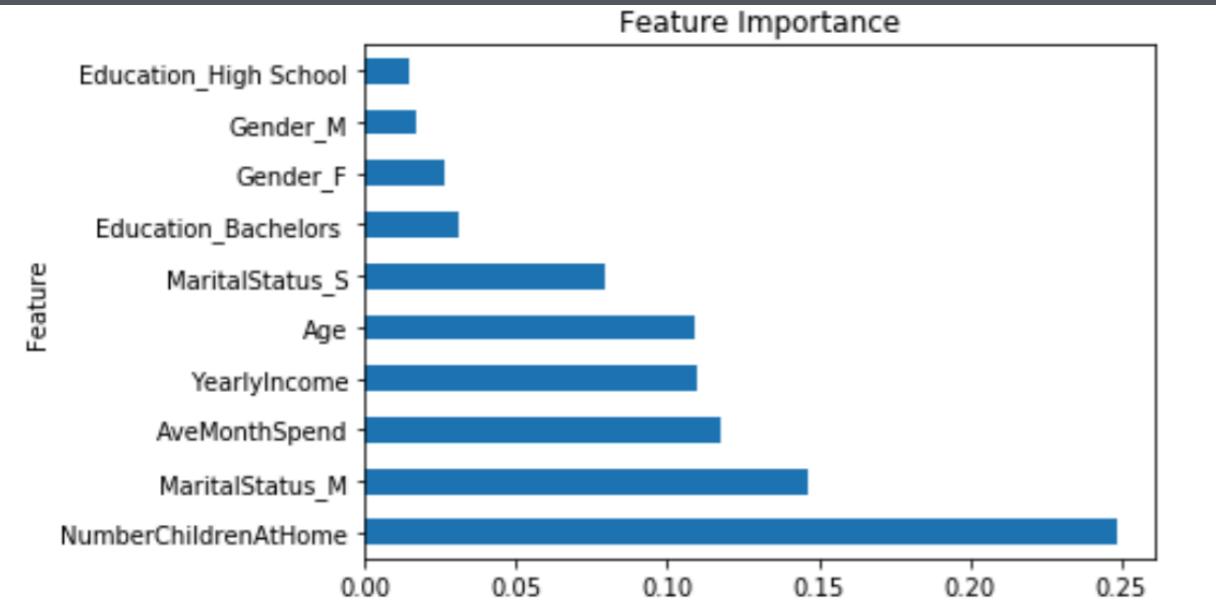


Thanks!

Data Driven Models

Multiple Machine Learning Models

1. Decision Tree



Accuracy: 75.1%

Pros

Manages irrelevant attributes well

Less changes to original data might mean better predictions (non normalized data)

Works well a lot of data

Cons

Greedy: sometimes misses best trees

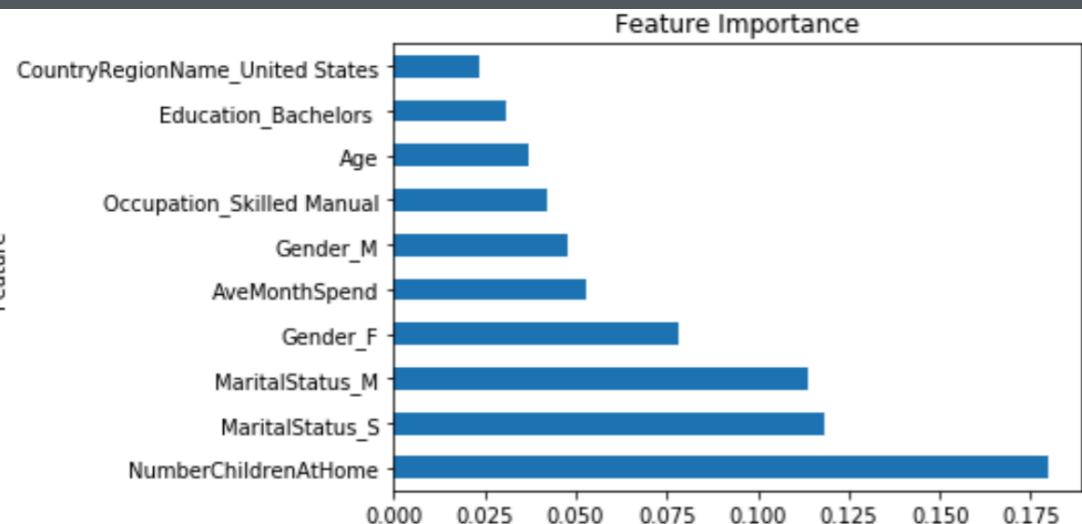
Small changes to data can cause a large change to structures of the decision tree

Wants bigger datasets

Data Driven Models

Multiple Machine Learning Models

1. XGBoost



Accuracy: 80.2%

Pros

One of the most accurate

Like others, good at automatically managing missing data

Cons

"Harder to get right." It's hyper-parameters are difficult to tune

Just like Random Forests can be difficult to grasp all the attribute relationships

Sensitive to noisy data and outliers