Take the Bait: Predicting Coupon Redemption on a Sharing Economy Site

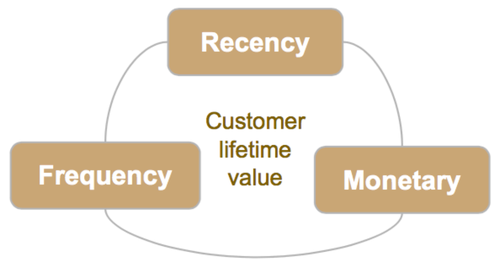
(Final Project of SI699: Big Data Analytics)

Mei Fu[1], Phoebe(Hui) Liang[1]
[1] School of Information, University of Michigan

Problem: Did you know that majority of the promotional coupons are unused by customers?

Motivation

- Coupons are the most simple and common way to increase customer conversions and keep customers coming back.
- Few studies examine the usage rate of promotional coupons and the effect of different factors on coupon usage.
- Sending coupons to the right users not only saves cost in marketing but also helps companies to grow new customers while keeping existing customers.
- We use machine learning models to predict the probability of a customer redeeming a coupon given their prior order and redemption history.



Feature Construction

Data Source: Air Kitchen

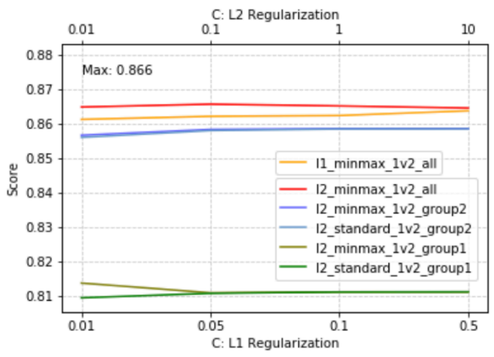
Two months of coupon redemption and order history (7 million records)

All the features only contains information before the coupon starting time

User Features	Sex	Age	City	Last App Version	Kitchen Entropy	Distance	Longitude	Latitude
Coupon Features	Type	Received Reason	Money	Max Money	Effective Days	Covers Weekday		
RFM Features	Count Order and Used Coupon	Coupon Usage Rate (and over Order, Type)	% Order and Coupon Used at Weekend	Median Interval Order, Usage	Coupon Usage of Last Order	Interval after Last Usage	Worth Money	
	Frequency			Recency			Monetary	

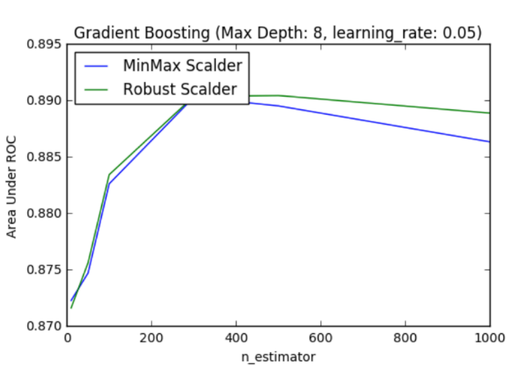
Training Process

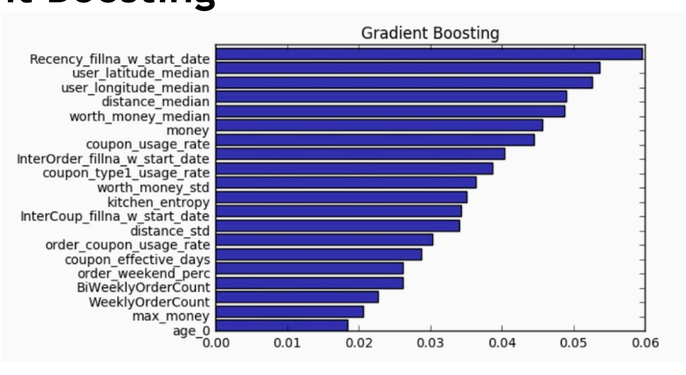
1 Logistic Regression Baseline



Keep N	Continuous	Categorical
~10	OrderCount x 2, Max_money	City x 5, Reason x 2
~25	'distance_median', 'money', 'WeeklyCouponUsedCount', 'Recency'	'AppVerLast_2.1', 'covers_tue', 'type1', 'RecallUserDaily', 'jiangzhecai', 'muqinjie', 'preuser', 'shareuser', ...
~40	'kitchen_entropy', 'effective_days', 'BiWeeklyCouponUsedCount', 'coupon_usage_rate', 'order_coupon_usage_rate', 'worth_money_median', 'InterCoup'	'age_0', 'AppVerLast_2.3', 'covers_mon', 'covers_thu', 'covers_sun', 'type1', 'type6', 'Complaints', 'shareuser'

2 Random Forest & Gradient Boosting

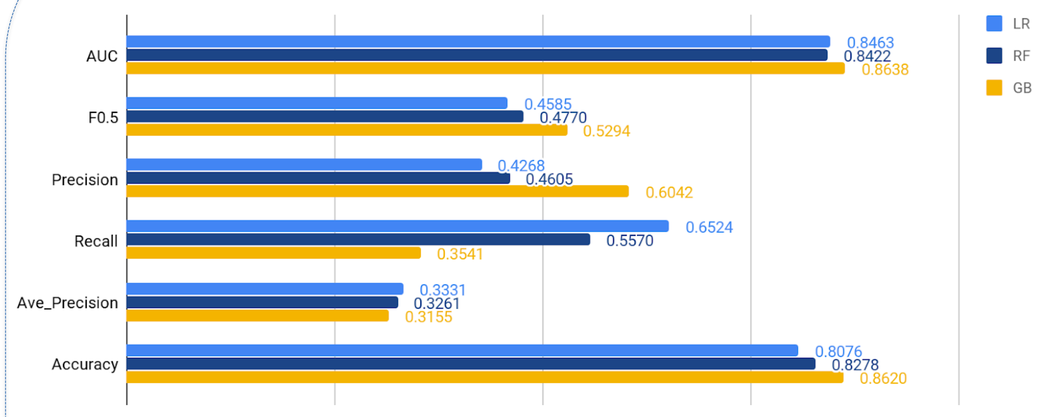




3 Feature Selection

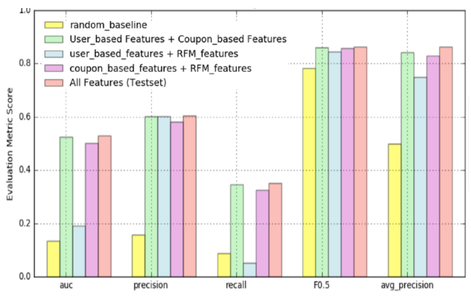
- Linear Models: L1, AIC, RFE (Recursive Feature Elimination)
- Nonlinear Models: Gini Importance

Model Performance

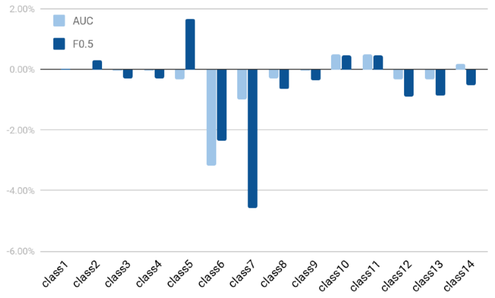


Error Analysis: Performance Changes…

After Removing a Group of Features



After Removing a Subgroup of Features



***Key Features**: Received Reason, Covers Weekday, Coupon Type

This is a final project presentation of SI 699: Big Data Analytics. We thank Professor Mei and his PhD student Teng Ye for the mentorship.

Contact:
meif@umich.edu, lianghui@umich.edu