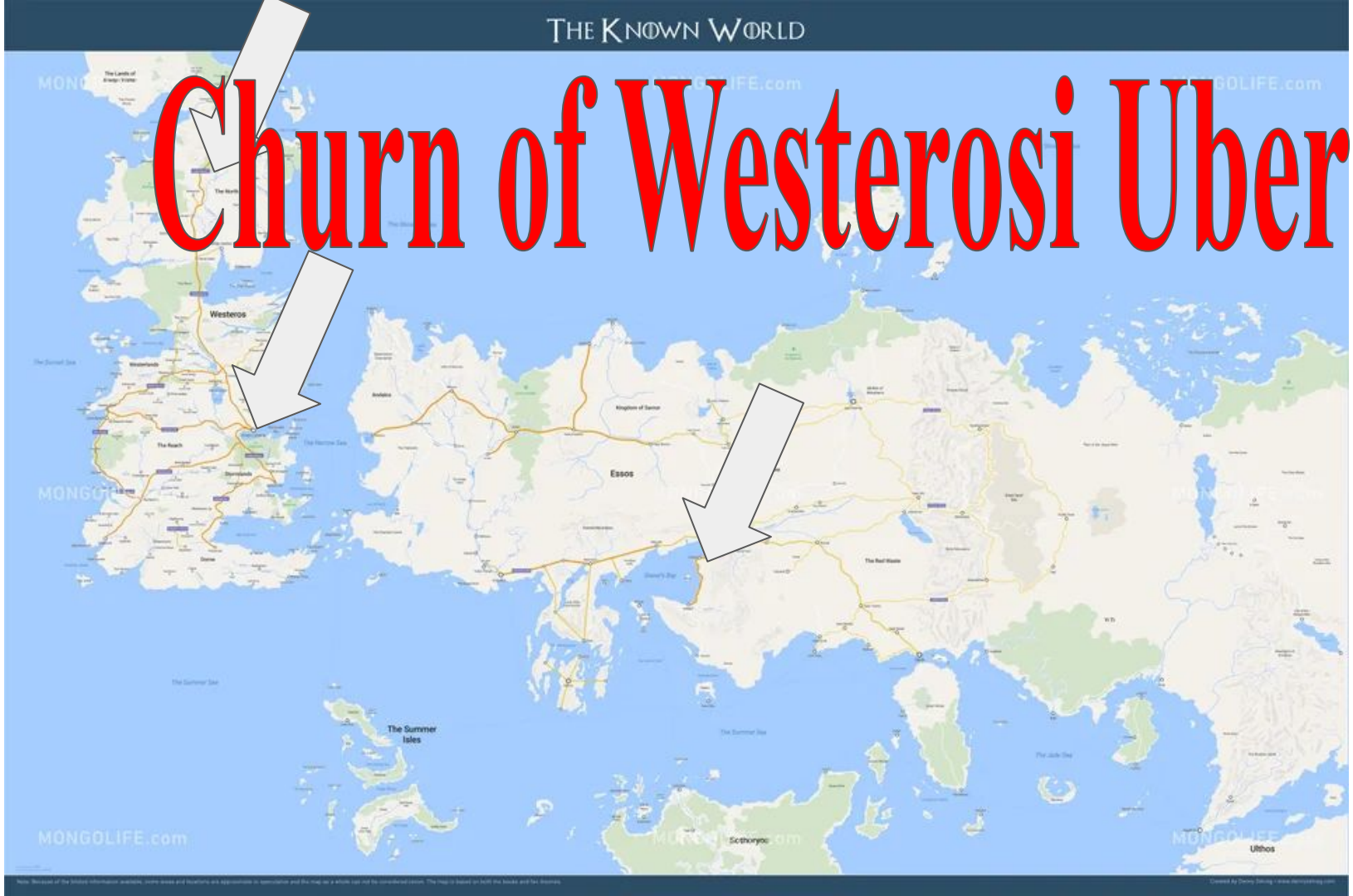
Predicting Uber Users Churn

Supervised Learning Case Study


THE KNOWN WORLD
Churn of Westerosi Uber
Westeros
Essos
The Summer Isles
Ulthos
MONGOLIFE.com

Demo Data Target Cities

- Winterfell/Wintertown
  - Cold Climate
  - Under Monarchical system of gov.
  - Small population (Small city)

- King’s Landing
  - Warm Climate
  - Many wealthy potential customers
  - Under Monarchical system of gov.
  - Large population (Urban center)
  - Port City

- Astapor
  - Warm Climate
  - Many wealthy potential customers
  - Ruled by a council of Slave Masters
  - Port City

How did we determine the target?

- 50,000 users with 12 features about each user (Date joined, Last active date, etc…)
- ‘Churn’ was defined as any client that had been inactive for greater than 30 days (data was uploaded 7/1/2014).
- This was seen as good indicator that we were losing business to a competitor
- This analysis was conducted to determine how to predict, and minimize this effect.
- Factors not considered: Political upheaval, Natural Disasters, and Dragons.

Models Considered:

- Logistic Regression:
  - Accuracy: 71.4%, Precision: 73.5%, Recall: 84.9%
  - Dropped ‘Weekday pct’, (p value 0.9)
  - Accuracy: 71.3%, Precision: 73.4%, Recall: 85.0%

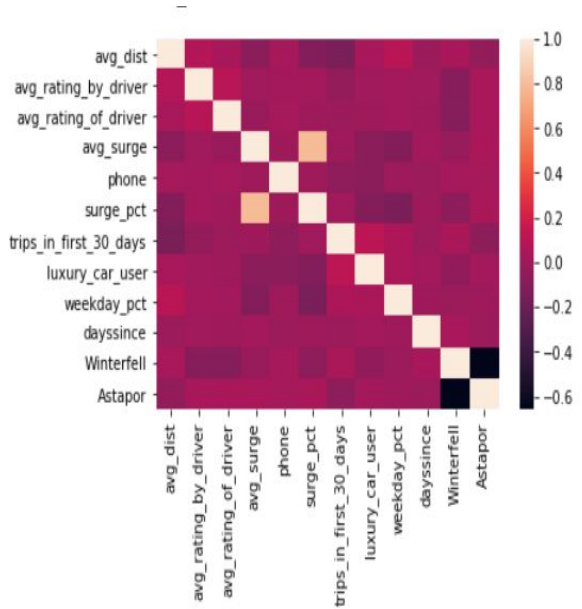
- Random Forest
  - Initial out of bag (OOB) score: 0.768

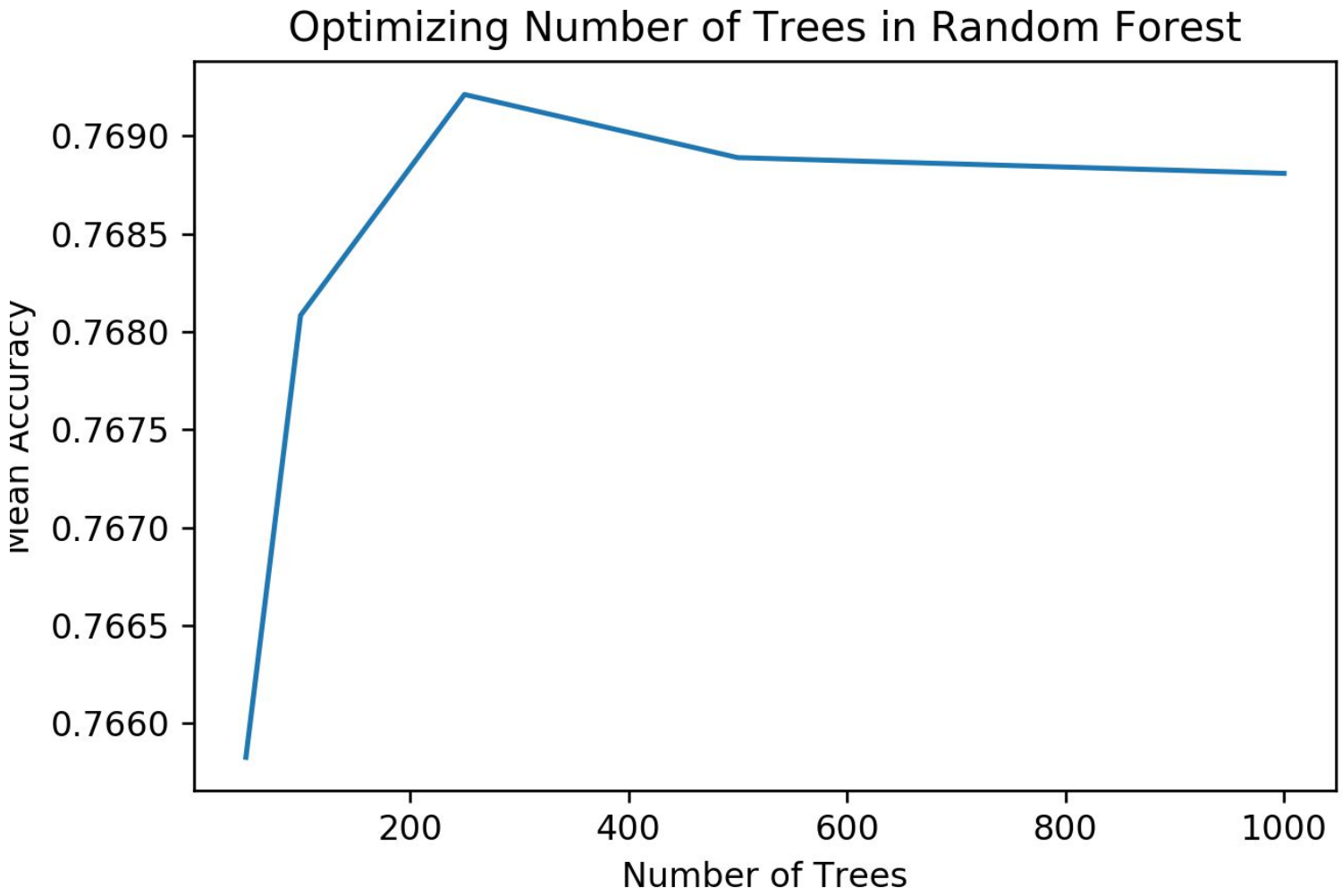
Logistic Regression: Accuracy: 71.4%, Precision: 73.5%, Recall: 84.9%

Logit Regression Results

Dep. Variable:	1	No. Observations:	39683
Model:	Logit	Df Residuals:	39671
Method:	MLE	Df Model:	11
Date:	Fri, 24 Apr 2020	Pseudo R-squ.:	0.1614
Time:	14:52:03	Log-Likelihood:	-22046.
converged:	True	LL-Null:	-26288.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
avg_dist	0.0358	0.002	15.455	0.000	0.031	0.040
avg_rating_by_driver	0.1400	0.024	5.954	0.000	0.094	0.186
avg_rating_of_driver	0.0497	0.019	2.570	0.010	0.012	0.088
avg_surge	0.1752	0.084	2.086	0.037	0.011	0.340
phone	1.1021	0.028	39.899	0.000	1.048	1.156
surge_pct	-0.0035	0.001	-3.694	0.000	-0.005	-0.002
trips_in_first_30_days	-0.1214	0.004	-31.296	0.000	-0.129	-0.114
luxury_car_user	-0.8888	0.024	-37.046	0.000	-0.936	-0.842
weekday_pct	3.983e-05	0.000	0.124	0.902	-0.001	0.001
dayssince	-0.0091	0.001	-10.886	0.000	-0.011	-0.007
Winterfell	1.1911	0.030	39.224	0.000	1.132	1.251
Astapor	1.7105	0.033	51.447	0.000	1.645	1.776

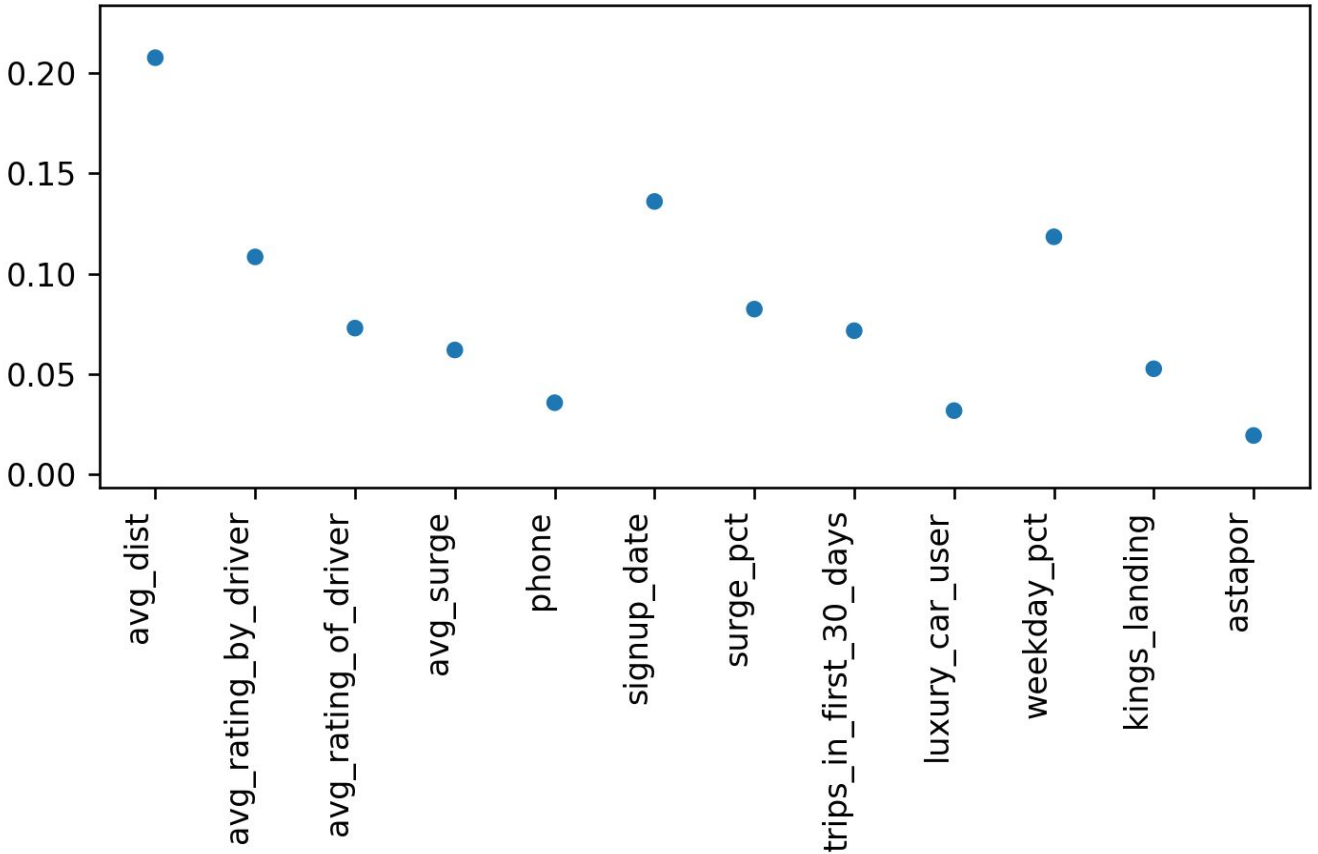



Optimizing Number of Trees in Random Forest
Mean Accuracy
0.7690
0.7685
0.7680
0.7675
0.7670
0.7665
0.7660
200
400
600
800
1000
Number of Trees

Final Random Forest Model - Training

- Number of Trees = 250
- Out of Bag Score (OOB) = 0.7710937290003494

Final Random Forest Model - Feature Importance



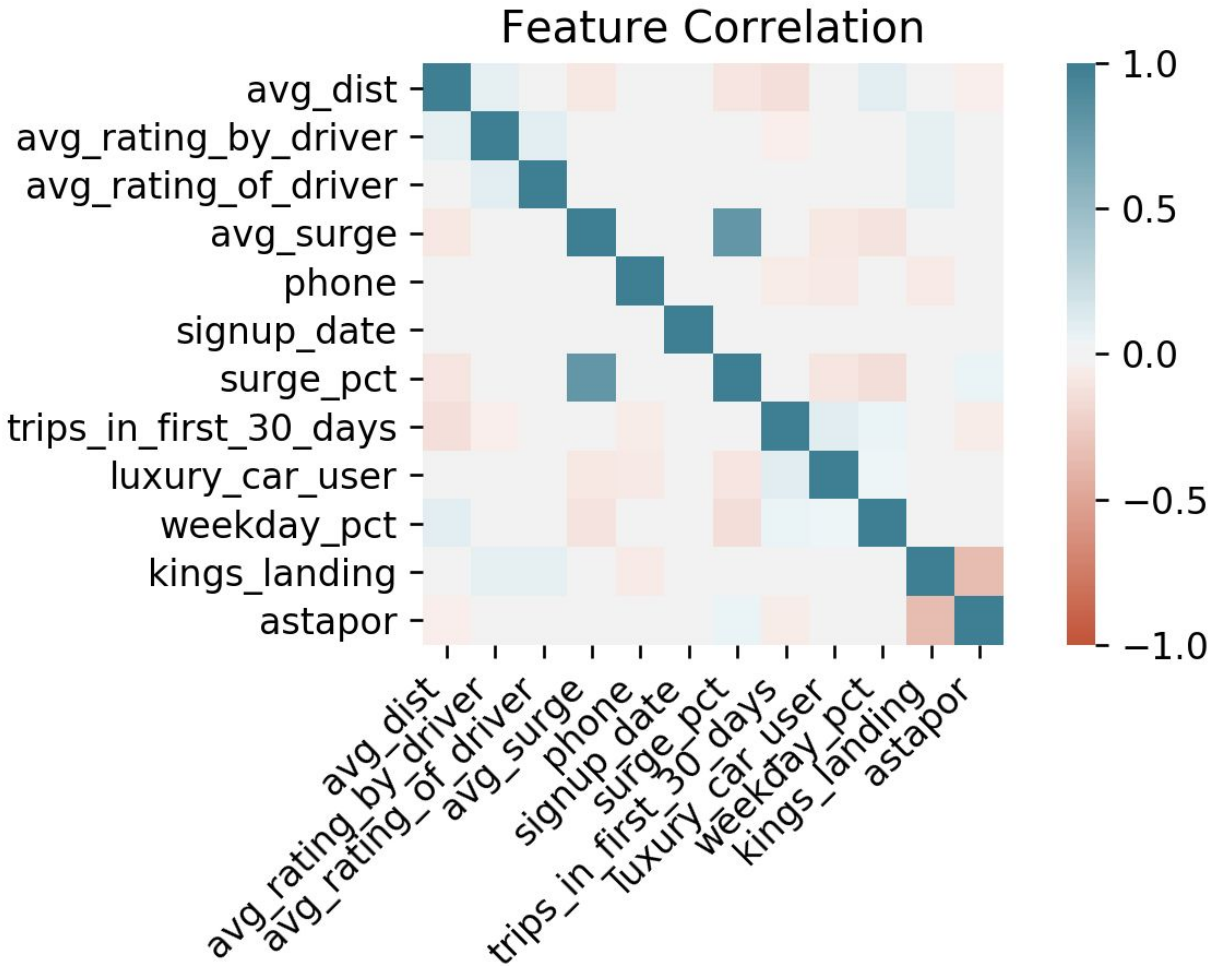
Final Random Forest Model - Test

- Final Accuracy Score = 0.943
- Based on this score, our model is ready for deployment.

Action Items:

- Use our random forest model to predict users who are at risk of churning
- Target these users for promotional offers:
  - High average Distance
  - Long term users
  - High weekday Surge users
- Create a cost benefit matrix to determine the appropriate dollar-amount for promotions
- Update the model with new data added over the next 30 days and re-assess.
- Our firm is currently designing and training and implementing a Neural network For better prediction ability


QUESTIONS?
WINTER IS ~~COMING~~
Here


Feature Correlation
avg_dist
avg_rating_by_driver
avg_rating_of_driver
avg_surge
phone
signup_date
surge_pct
trips_in_first_30_days
luxury_car_user
weekday_pct
kings_landing
astapor
1.0
0.5
0.0
−0.5
−1.0