CLASSIFICATION OF FORD GO BIKE TRAFFIC IN SAN FRANCISCO

-HIRANYA KRISHNA KUMAR

MOTIVATION

Bay Area Bike share is rapidly growing with over 300 stations and 3600 bikes

Problem: Imbalanced Traffic

Solution:
Build model to predict cautionary levels of traffic using classification
Identify bottlenecks and plan for future rebalancing

DATA COLLECTION

Ford GoBike

Over 500,000 trips collected for 2017. Dock capacity data collected using Ford Go station information API and station region API

Weather data collected using python-forecast.io wrapper for Dark Sky API using lat and long coordinates for San Francisco

METHODOLOGY

DATA CLEANING AND FEATURE ENGINEERING

Created Target Label:Flux
0: Normal Traffic
1: Cautionary inflow/Surplus>0.15 capacity
2: Cautionary outflow/Shortage

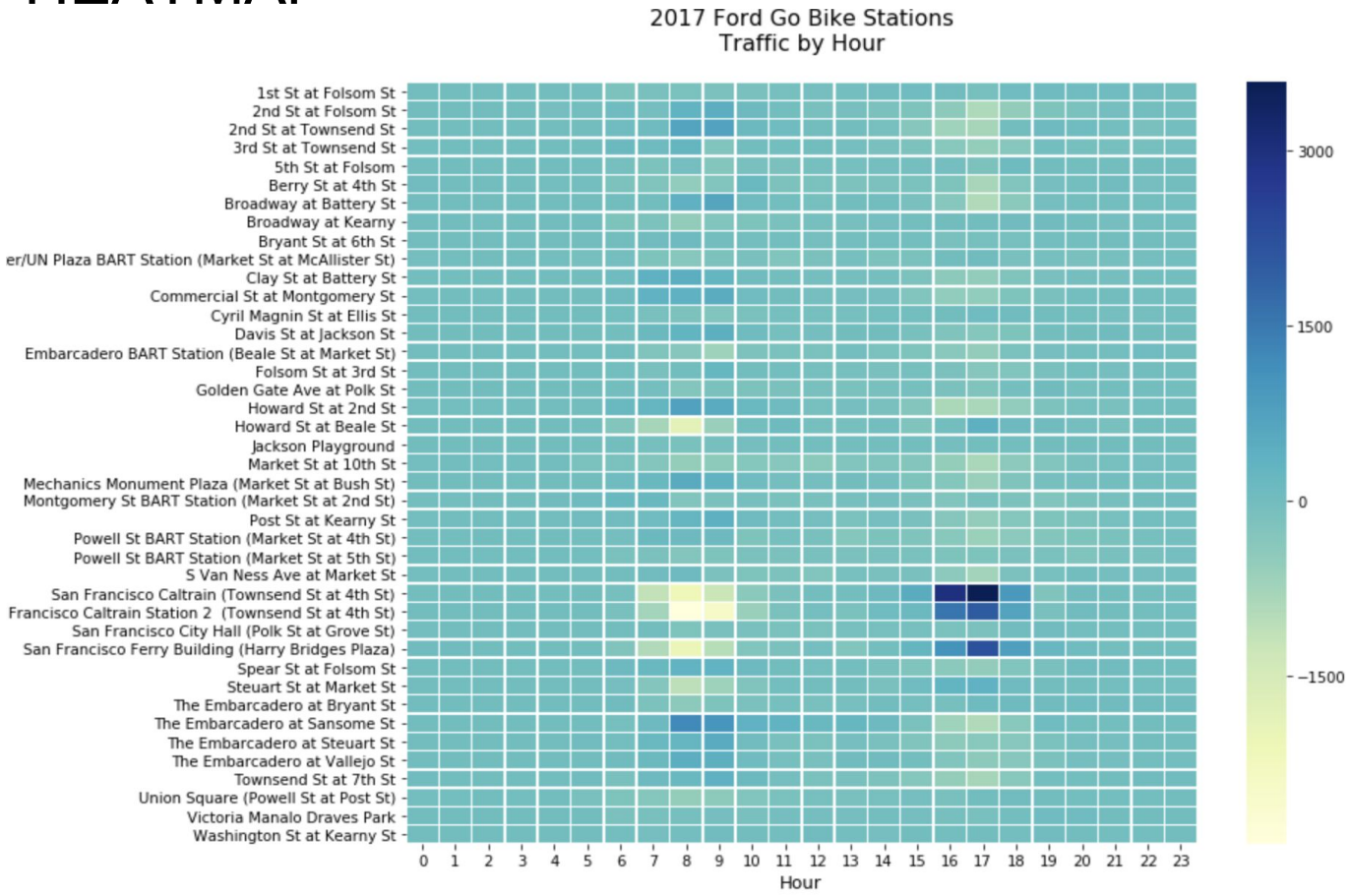
BALANCE CLASSES

Highly Imbalanced data: Applied SMOTE ENN to balance classes
(94% majority class)

BEST MODELS

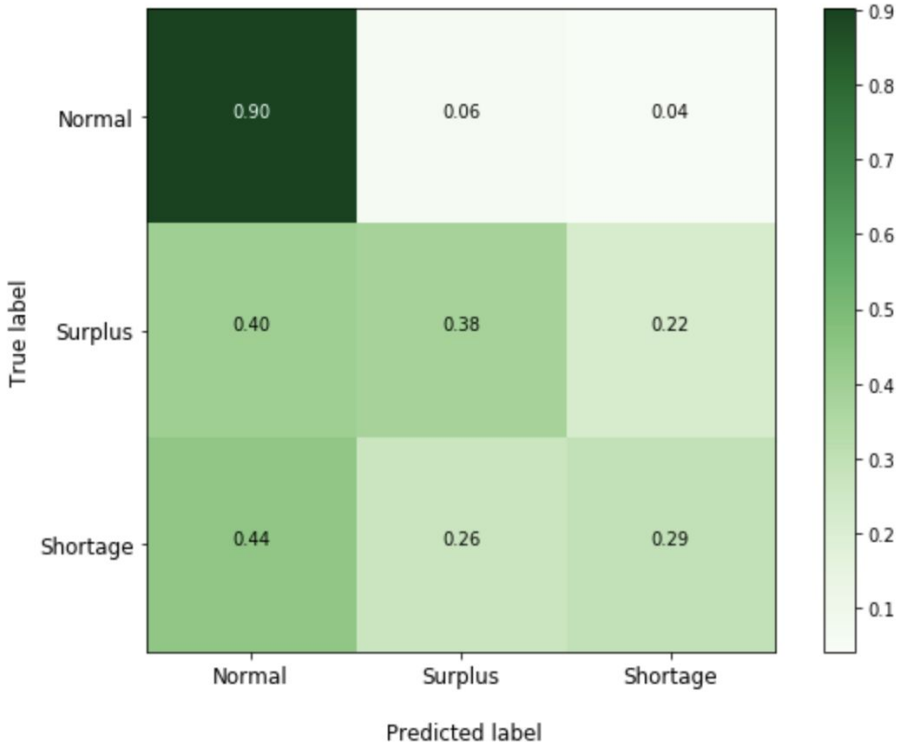
Gradient Boosting
Random Forest

HEATMAP


2017 Ford Go Bike Stations
Traffic by Hour
1st St at Folsom St
2nd St at Folsom St
2nd St at Townsend St
3rd St at Townsend St
5th St at Folsom
Berry St at 4th St
Broadway at Battery St
Broadway at Kearny
Bryant St at 6th St
er/UN Plaza BART Station (Market St at McAllister St)
Clay St at Battery St
Commercial St at Montgomery St
Cyril Magnin St at Ellis St
Davis St at Jackson St
Embarcadero BART Station (Beale St at Market St)
Folsom St at 3rd St
Golden Gate Ave at Polk St
Howard St at 2nd St
Howard St at Beale St
Jackson Playground
Market St at 10th St
Mechanics Monument Plaza (Market St at Bush St)
Montgomery St BART Station (Market St at 2nd St)
Post St at Kearny St
Powell St BART Station (Market St at 4th St)
Powell St BART Station (Market St at 5th St)
S Van Ness Ave at Market St
San Francisco Caltrain (Townsend St at 4th St)
Francisco Caltrain Station 2 (Townsend St at 4th St)
San Francisco City Hall (Polk St at Grove St)
San Francisco Ferry Building (Harry Bridges Plaza)
Spear St at Folsom St
Steuart St at Market St
The Embarcadero at Bryant St
The Embarcadero at Sansome St
The Embarcadero at Steuart St
The Embarcadero at Vallejo St
Townsend St at 7th St
Union Square (Powell St at Post St)
Victoria Manalo Draves Park
Washington St at Kearny St
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour
3000
1500
0
−1500

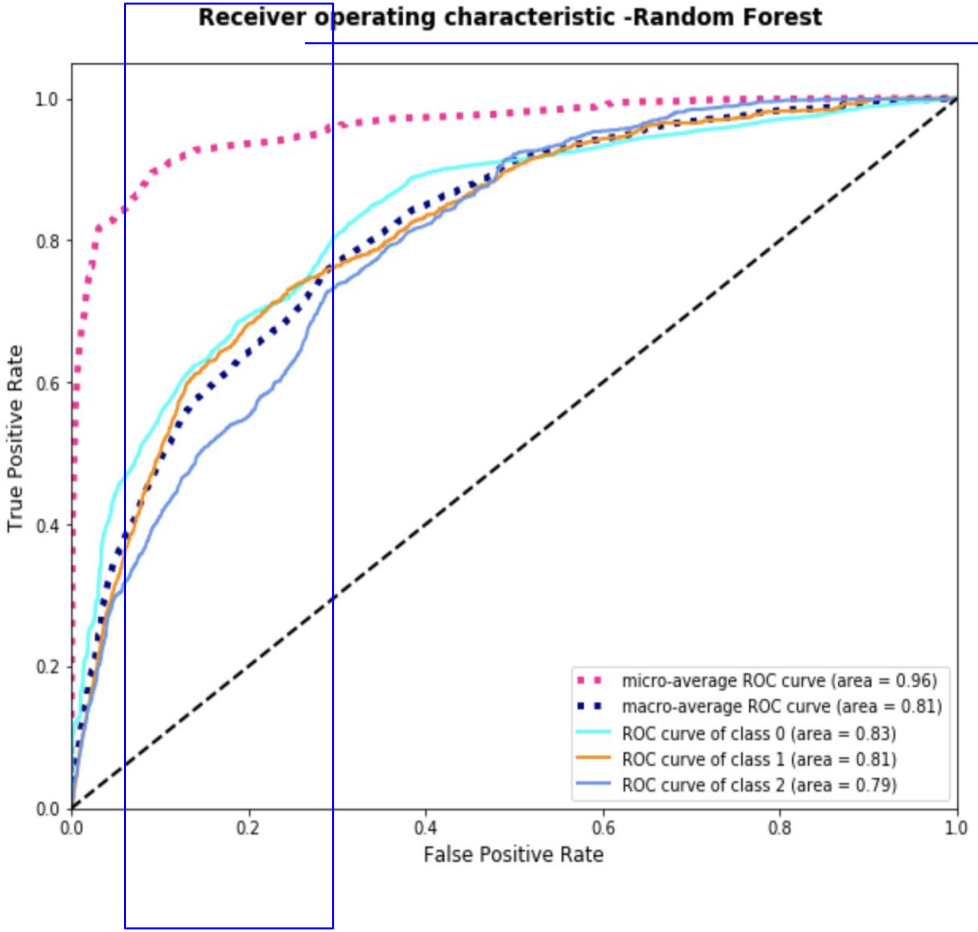
RESULTS

FINAL MODEL:RANDOM FOREST,STRATIFIED 3 FOLD CV,RANDOMIZED SEARCH



CLASS	PRECISION	RECALL	F1
Normal	0.96	0.90	0.93
Cautionary Surplus	0.13	0.38	0.19
Cautionary Shortage	0.23	0.29	0.26

ROC CURVE



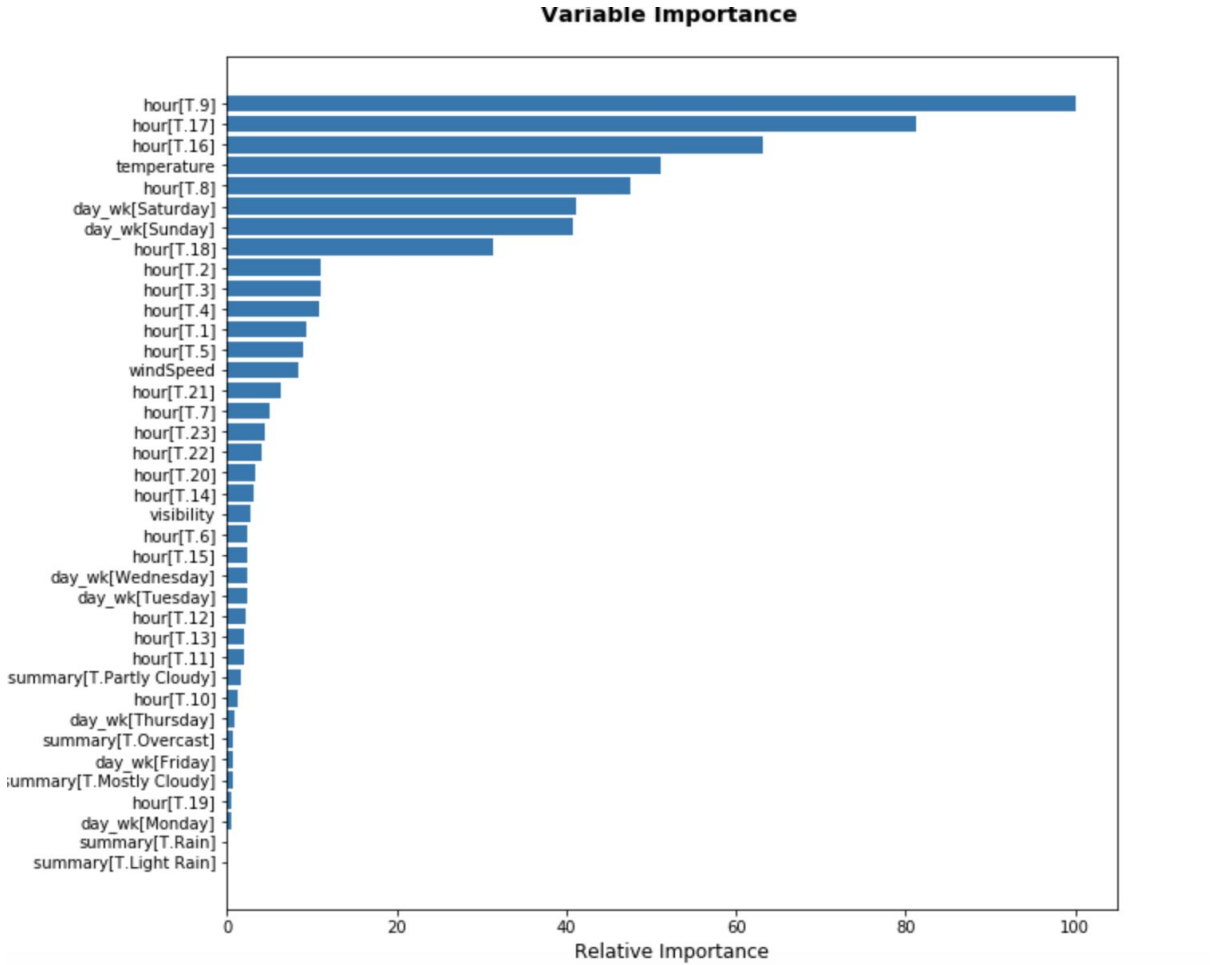
Minimize FP for Class 0: Normal traffic
Maximize Recall for all classes,Minimize FN on Class 1 and 2

FUTURE WORK

- Gather realtime-data for bike availability to calculate flux
- Use Poisson regression to predict bike counts
- Further tune hyperparameters for random forest , gradient boosting model
- Use only Features with high importance and add more relevant features
- Build a flask app to classify hourly bike traffic for every station
- Try other models like adaptive gradient boosting, Xgboost

THANK YOU!

APPENDIX



PARAMS:min_samples_leaf=9,n_estimators=150, max_depth=5

GRADIENT BOOSTING

	precision	recall	f1-score
0.0	0.97	0.84	0.90
1.0	0.11	0.36	0.17
2.0	0.15	0.35	0.21
micro avg	0.81	0.81	0.81
macro avg	0.41	0.52	0.42