
Is your family
prepared to escape
natural disasters?

Optimizing Evacuation Routes using Real-Time Traffic Information

Ingrid, Dennis, Mac

Problem statement

During emergencies and disasters, we want to provide users with real time information that can aid in their evacuation and/or reduce time required to reach their desired destination.

Time and efficiency is of the essence when people's lives are on the line!

Methodology

We will be focusing on twitter posts from reputable traffic news providers and also from all users within a certain range of our location.

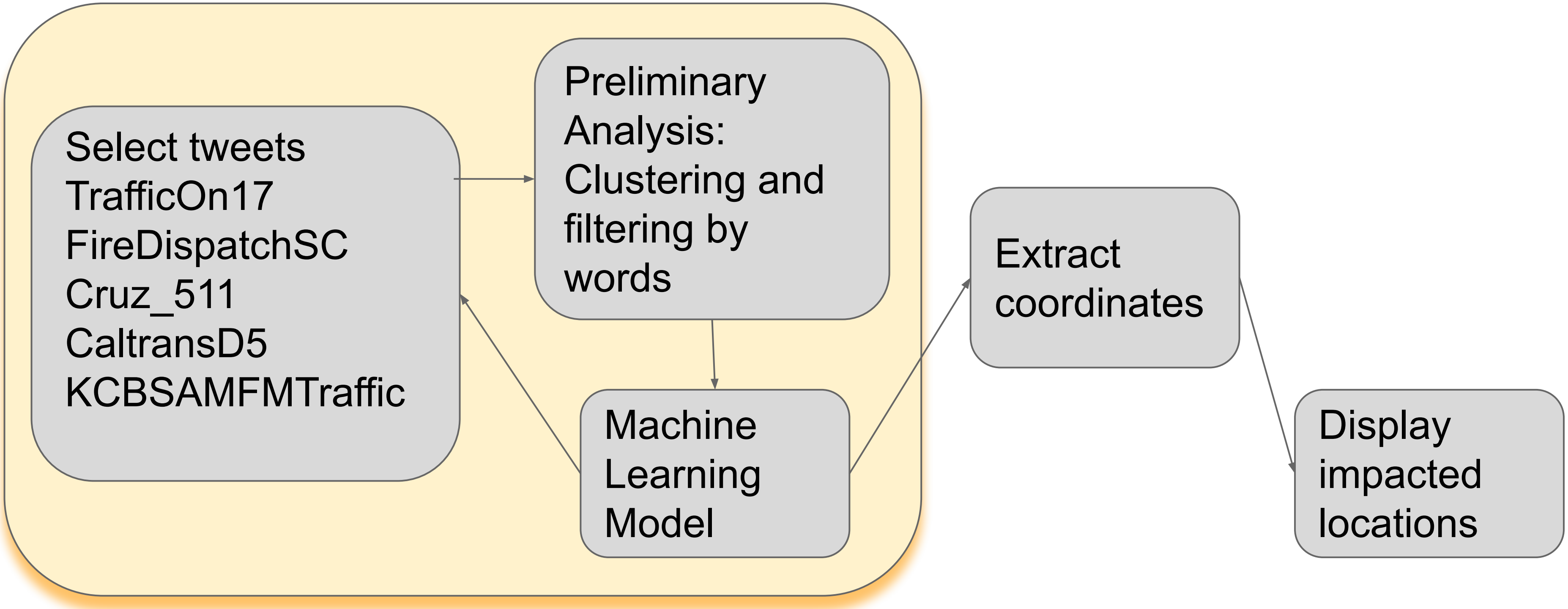
We believe information from both authorities AND nearby citizens are likely to be more accurate.

Why KCBSAMFMTraffic?

Our classification plan

- Classify tweets
- Model
- Find and keep best model
- Mapping locations

Workflow


Select tweets
TrafficOn17
FireDispatchSC
Cruz_511
CaltransD5
KCBSAMFMTraffic
Preliminary Analysis: Clustering and filtering by words
Machine Learning Model
Extract coordinates
Display impacted locations

Preliminary exploratory data analysis

KMean(CV):
n-cluster = 2

Silhouette_Score = 0.180

Not Traffic 59%

Traffic Blocks 41%

DBScan(CV):
eps = 3.5,
min_samples = 100
Silhouette_score = 0.177

Not Traffic 57%

Traffic Blocks 43%

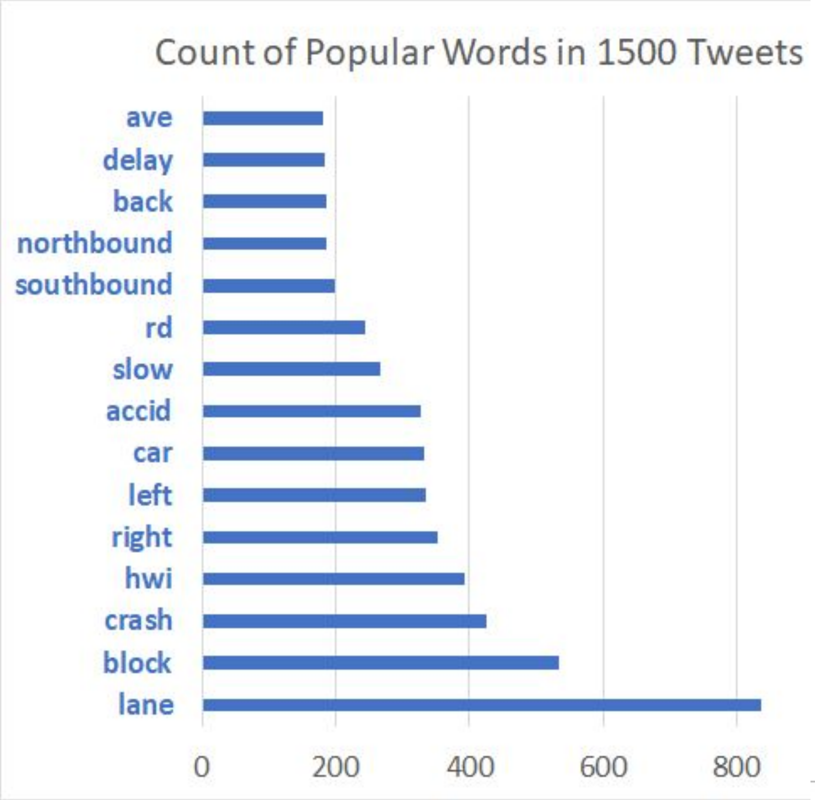
Using Traffic Words

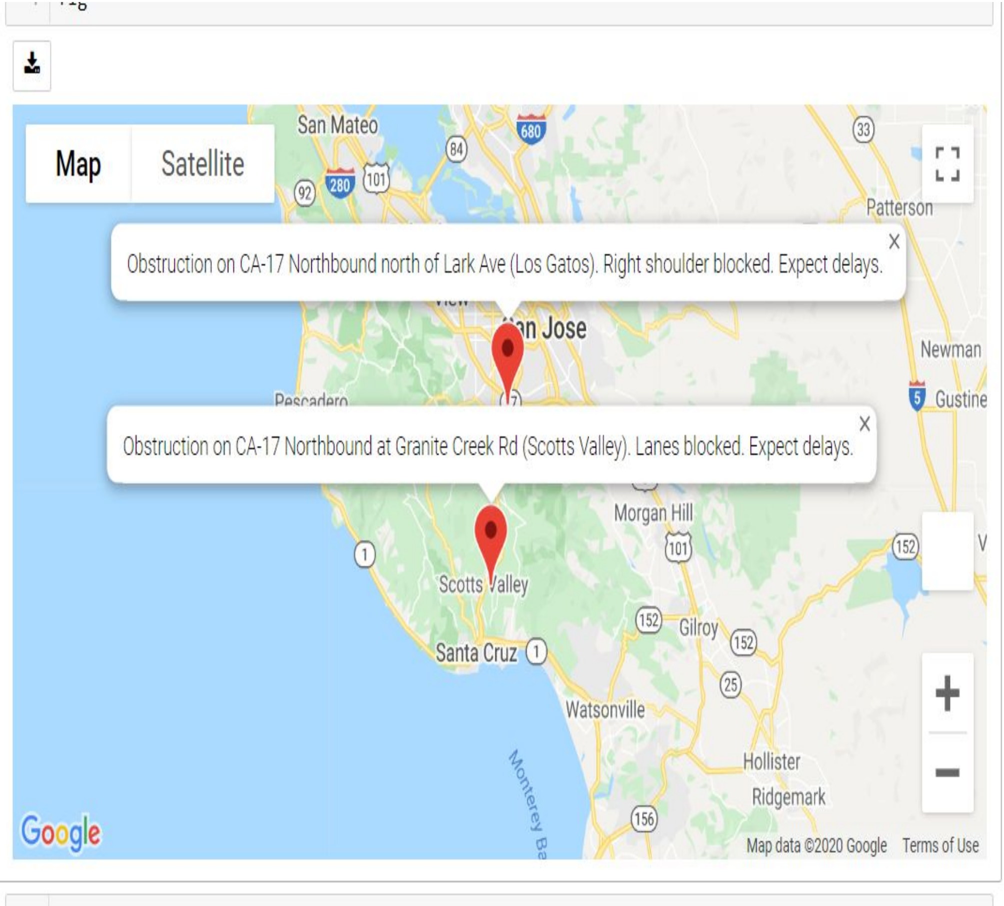
OUR CHOICE

Not Traffic 42%

Traffic Blocks 58%

Preliminary EDA





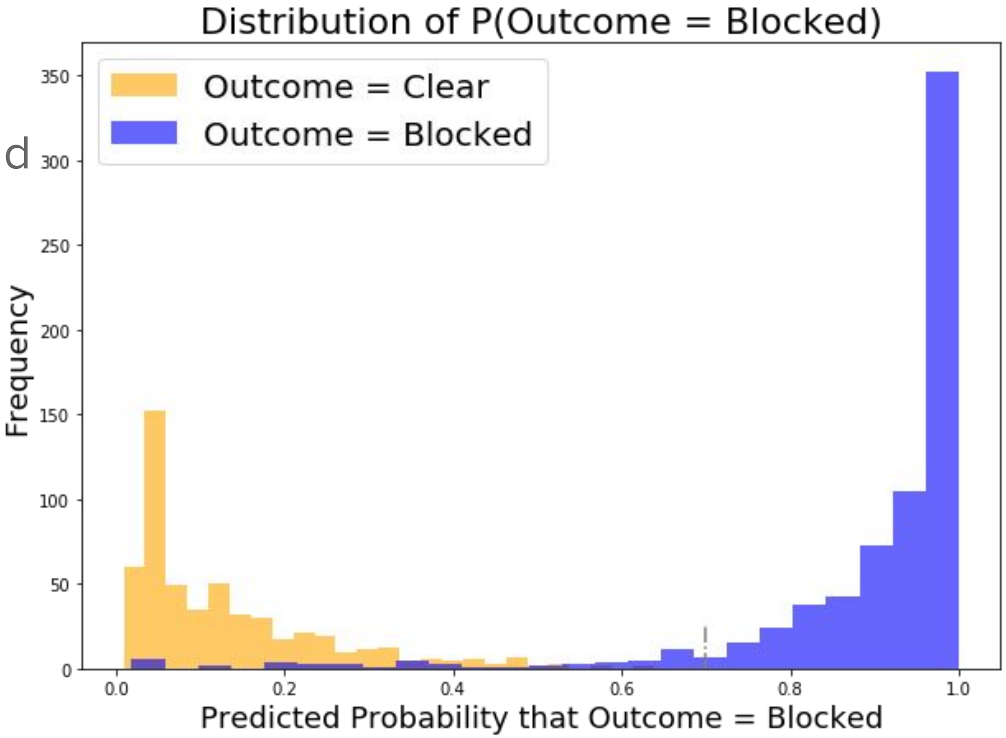
#1 Logistic Regression

CountVectorizer

Features with higher frequency of 10 and under 80% occurrence

Maximum features at 500

N-gram range of 1



#2 Random Forests

CountVectorizer

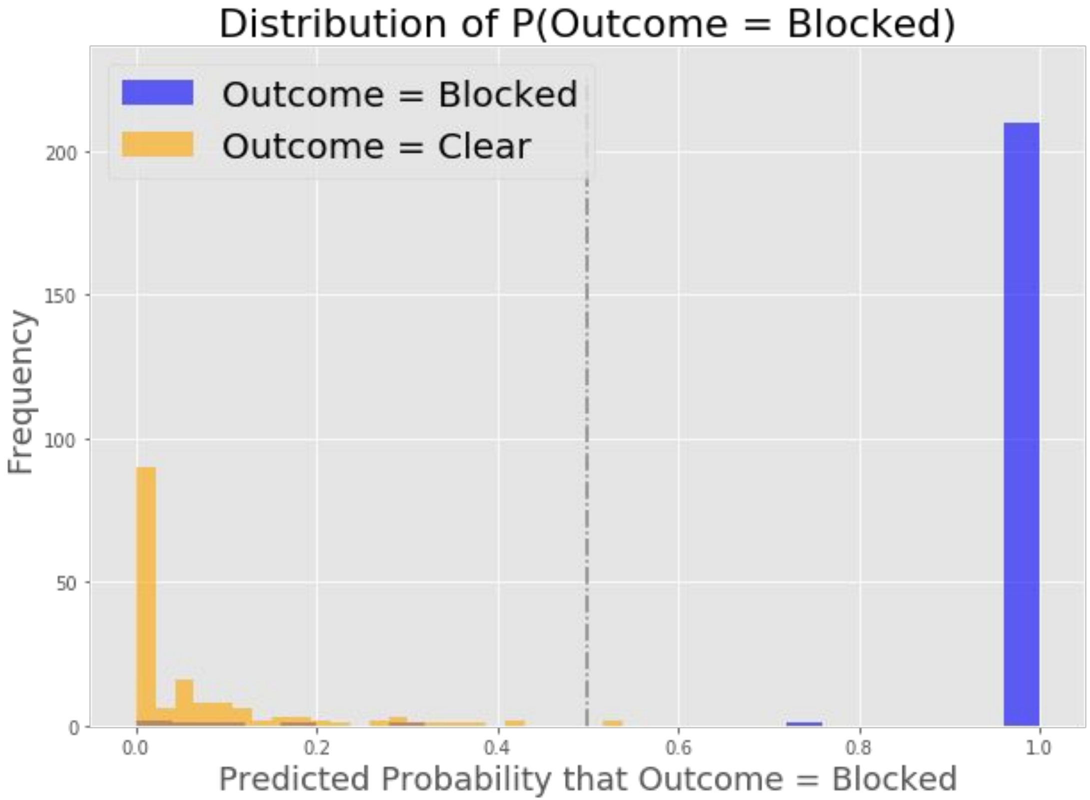
Minimum samples leaves at 5

Minimum samples split at 2

100 estimators

No max depth

No max features



#3 Multinomial Naive Bayes

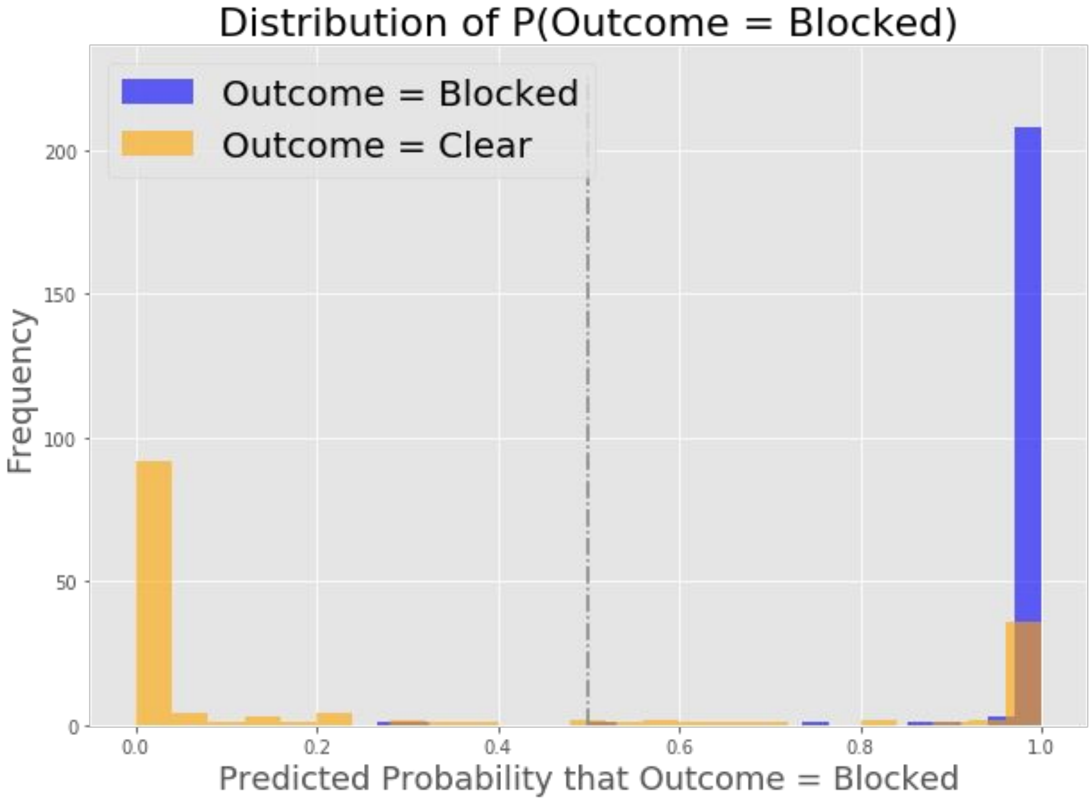
CountVectorizer

Maximum features at 1000

Min df at 100

N-gram range of 1

stop words is ‘English’



#3 K Nearest Neighbors

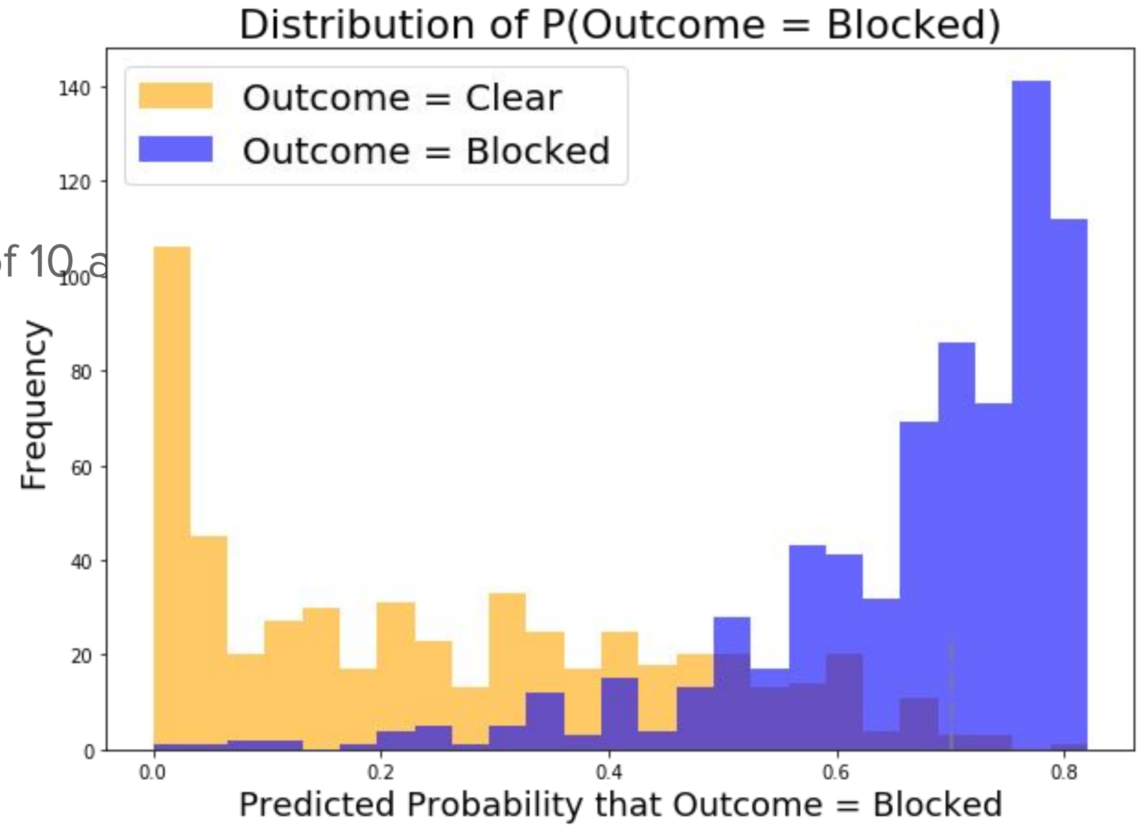
TfidfVectorizer

Nearest 50 neighbors

Features with higher frequency of 10 a
under 80% occurrence

Maximum features of 1000

N-gram range of 2



#4 Support Vector Machines

CountVectorizer

Nearest 50 neighbors

Features with higher frequency of 100 and under 80% occurrence

Maximum features of 500

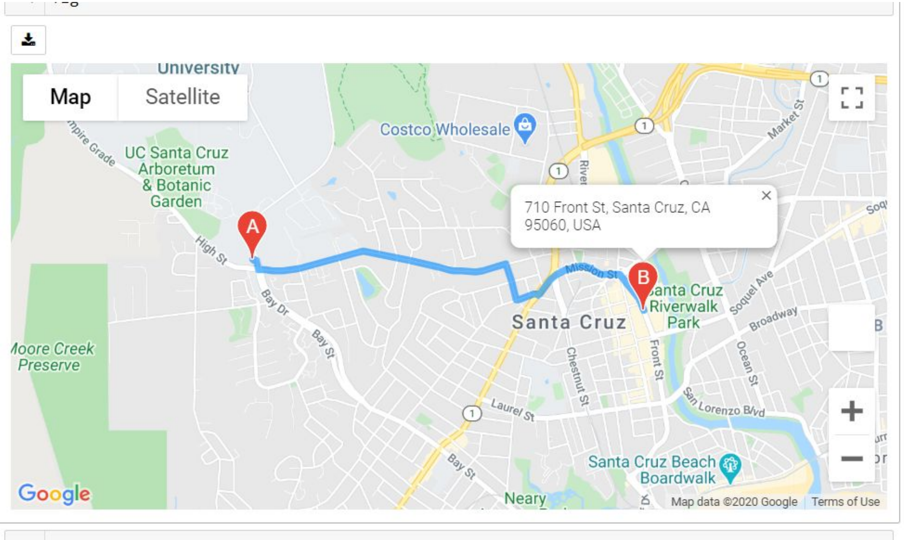
N-gram range of 1

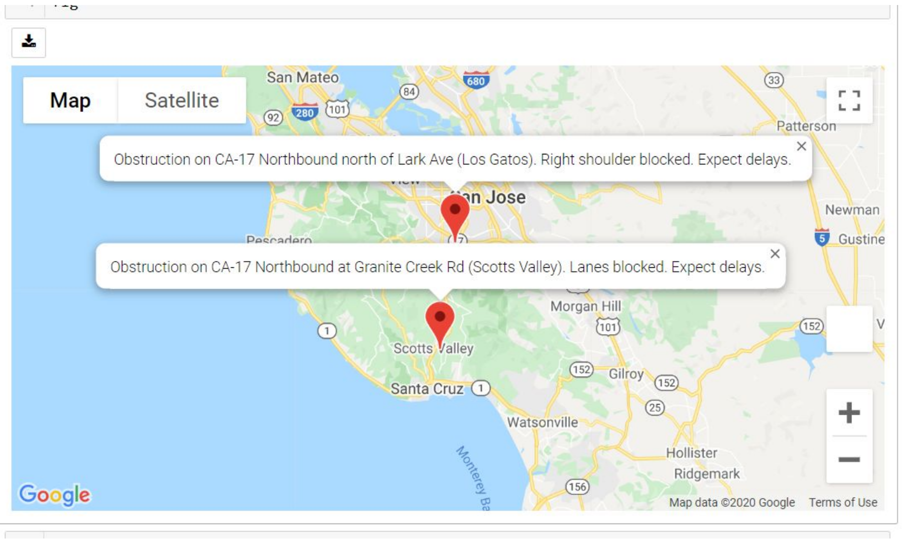
	Training Score	Testing Score	Sensitivity
Logistic Regression(CV)	**.971**	**.970**	**.972**
Logistic Regression(TV)	.961	.965	.969
Random Forest	.973	.978	.972
Multinomial Naive Bayes	.904	.866	.990
Gaussian Naive Bayes	.936	.770	.866
K Nearest Neighbors	.881	.869	.933
Support Vector Machines	.974	.976	.969

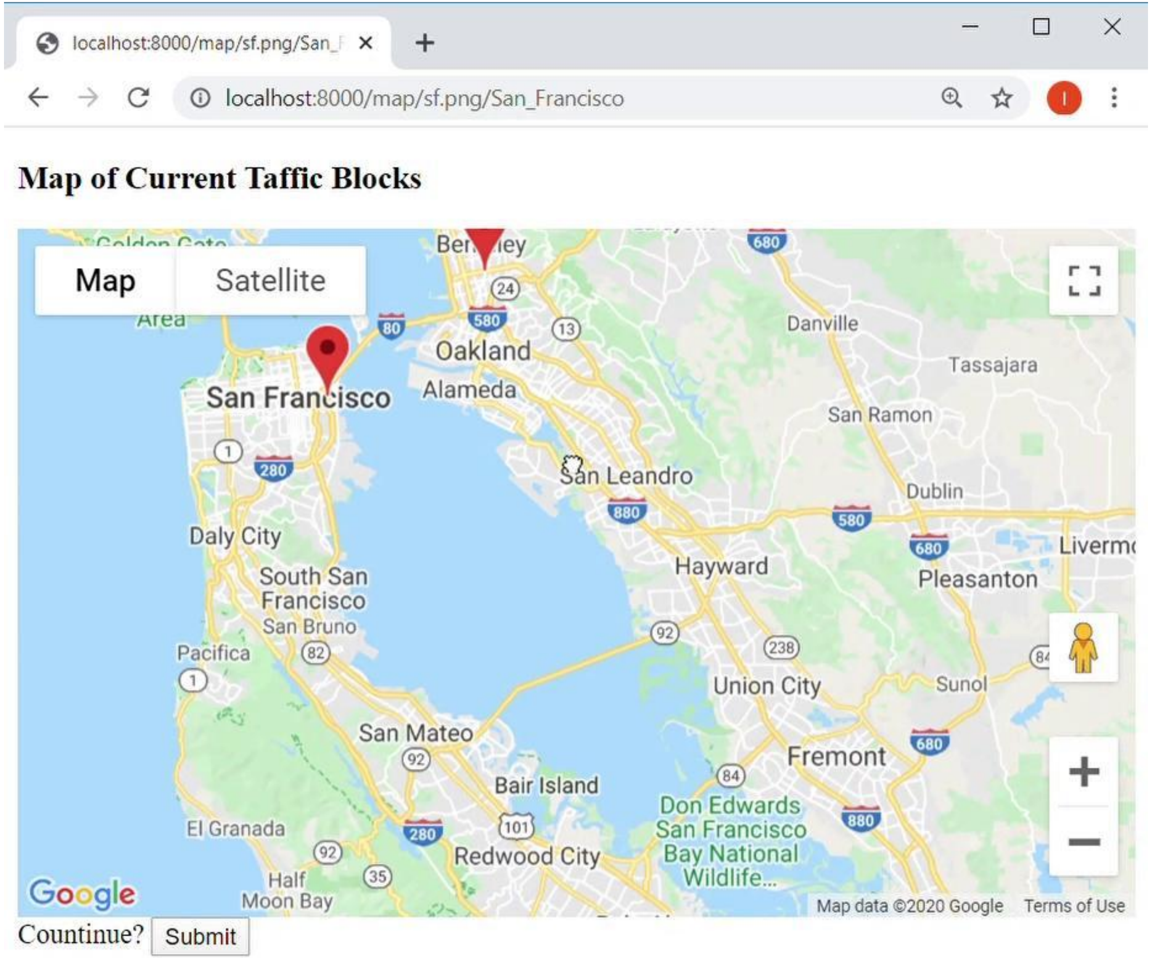
Findings

Logistic Regression!

Conclusions / Results






localhost:8000/map/sf.png/San_
localhost:8000/map/sf.png/San_Francisco
Map of Current Taffic Blocks
Map
Satellite
Oakland
Alameda
San Francisco
San Leandro
Daly City
South San Francisco
San Bruno
Pacifica
Hayward
Danville
Tassajara
San Ramon
Dublin
Pleasanton
Union City
Sunol
Fremont
San Mateo
Bair Island
Redwood City
El Granada
Half Moon Bay
Don Edwards San Francisco Bay National Wildlife...
Map data ©2020 Google
Terms of Use
Countinue?
Submit

Next steps

- Clustering continuation
- Funding for Google API
- Funding for AWS
- Word2Vec implementation
- Route optimization via Google map or Here.com
- Location agnostic w/ real-time updates
- More sources!