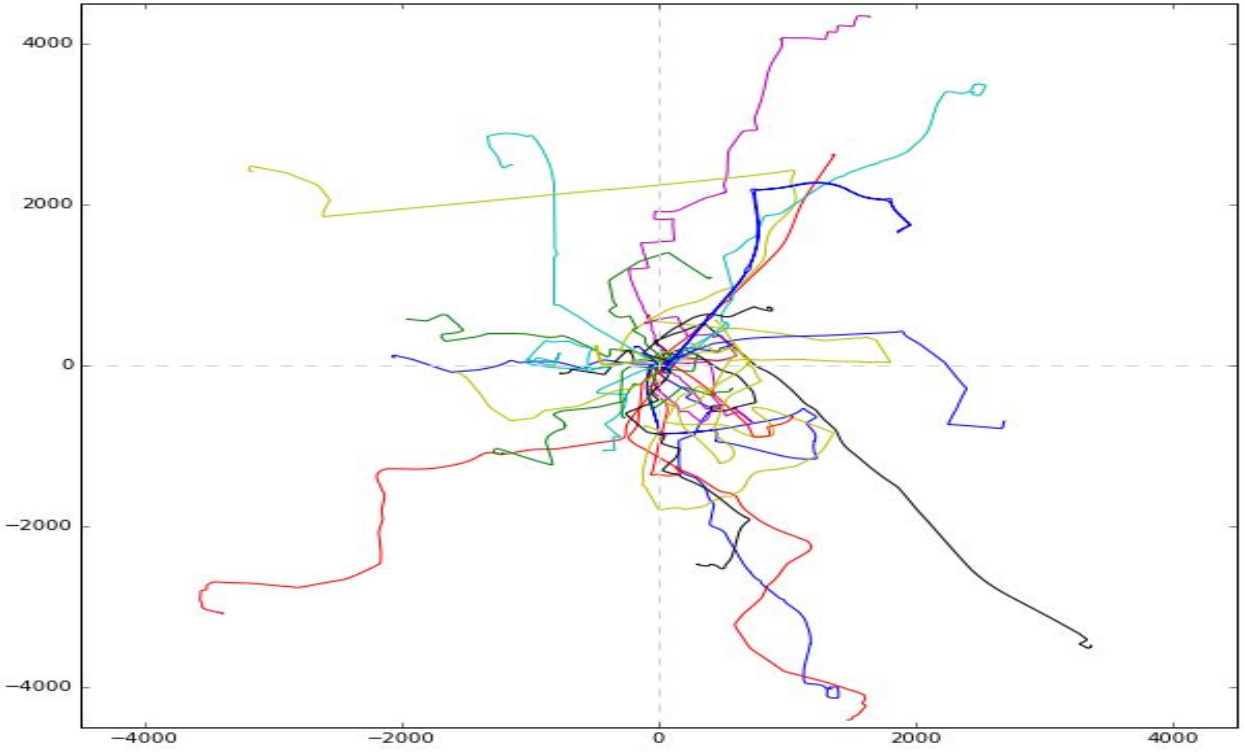
Identifying Drivers Signatures Through Unlabeled Telematic Data

Su-Young Hong, Mayank Kedia, Harry O'Reilly

Driver Trips



The Data

- 2,736 Drivers
- 200 trips per driver (547,200 total)
- Time series of (x,y) coordinates

The Data

- 2,736 Drivers
- 200 trips per driver (547,200 total)
- Time series of (x,y) coordinates
- False trips randomly inserted into each driver set

Challenge 1:

Unlabeled Data

Manufacture Labels

Driver	Trip	Label
(Driver 1, Trip 001)		?
(Driver 1, Trip 002)		?
(Driver 1, Trip 003)		?
(...)		

Manufacture Labels

Driver	Trip	Label
(Driver 1,	Trip 001)	1
(Driver 1,	Trip 002)	1
(Driver 1,	Trip 003)	1
(...)		

Manufacture Labels

Driver	Trip	Label
(Driver 1,	Trip 001)	1
(Driver 1,	Trip 002)	1
(Driver 1,	Trip 003)	1
(...)		

Driver 2

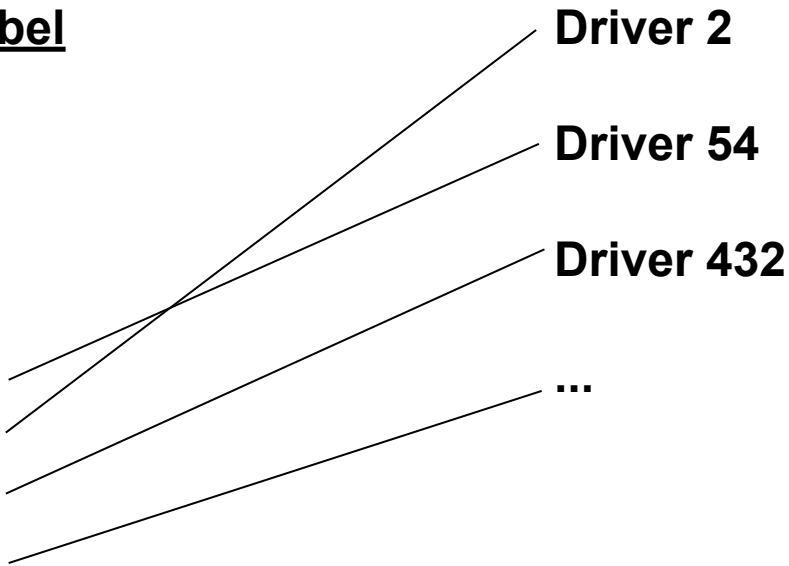
Driver 54

Driver 432

...

Manufacture Labels

Driver	Trip	Label
(Driver 1, Trip 001)		1
(Driver 1, Trip 002)		1
(Driver 1, Trip 003)		1
(...)		
(Driver 1, Trip 201)		0
(Driver 1, Trip 202)		0
(Driver 1, Trip 203)		0
(...)		



Challenge 2:

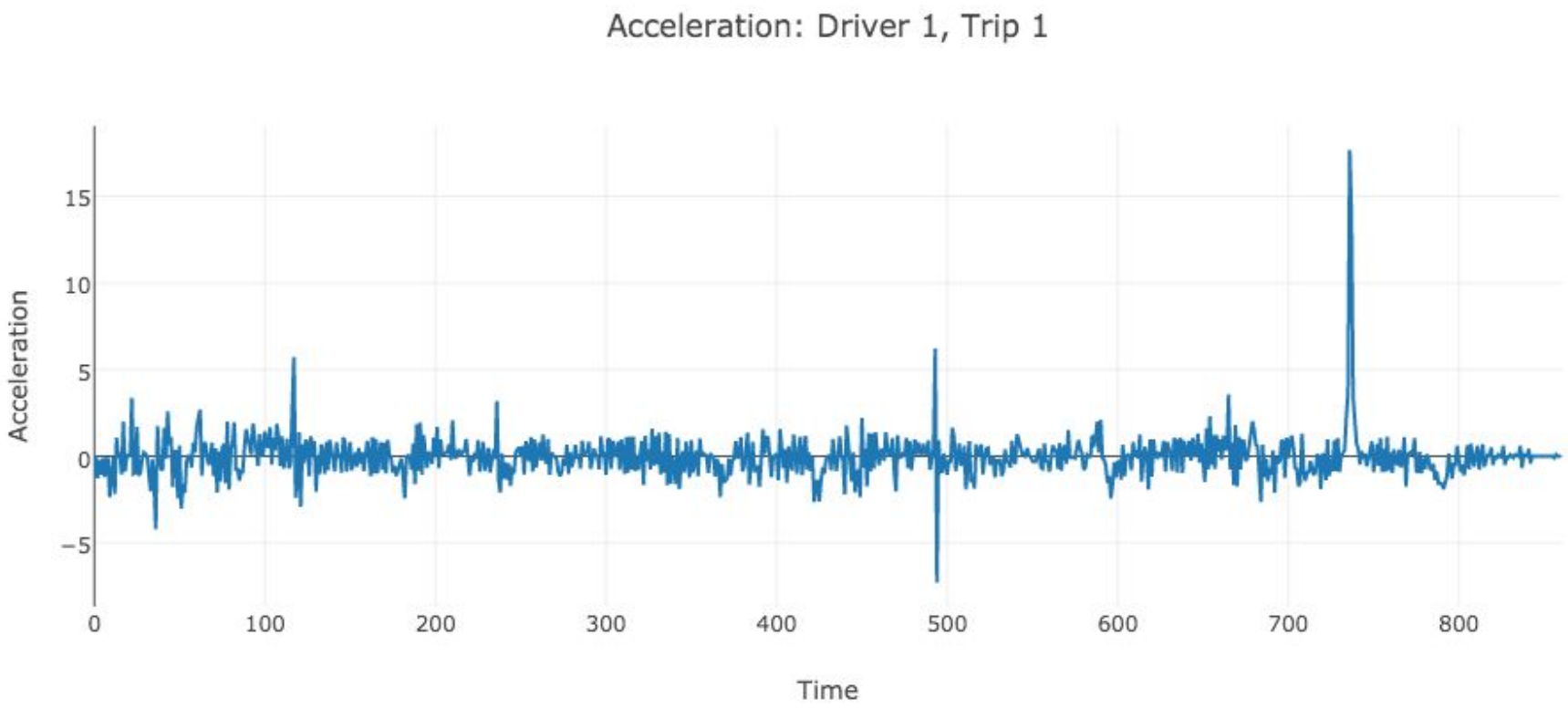
Feature Engineering

Step Level Features

- Start with time series of Cartesian Coordinates
- Generate velocity, acceleration, jerk

Step Level Features

- Start with time series of Cartesian Coordinates
- Generate velocity, acceleration, jerk
- Convert to polar, generate centripetal acceleration, angular velocity, tangential acceleration
- Smooth


Acceleration: Driver 1, Trip 1
Acceleration
15
10
5
0
−5
0
100
200
300
400
500
600
700
800
Time

Step Level Features

- Start with time series of Cartesian Coordinates
- Generate velocity, acceleration, jerk
- Convert to polar, generate centripetal acceleration, angular velocity, tangential acceleration
- Smooth

Trip Level Features

- Aggregate step level features
- Max and min
- Mean and Standard Deviation
- Quantiles

Challenge 3:

Model Building

Models

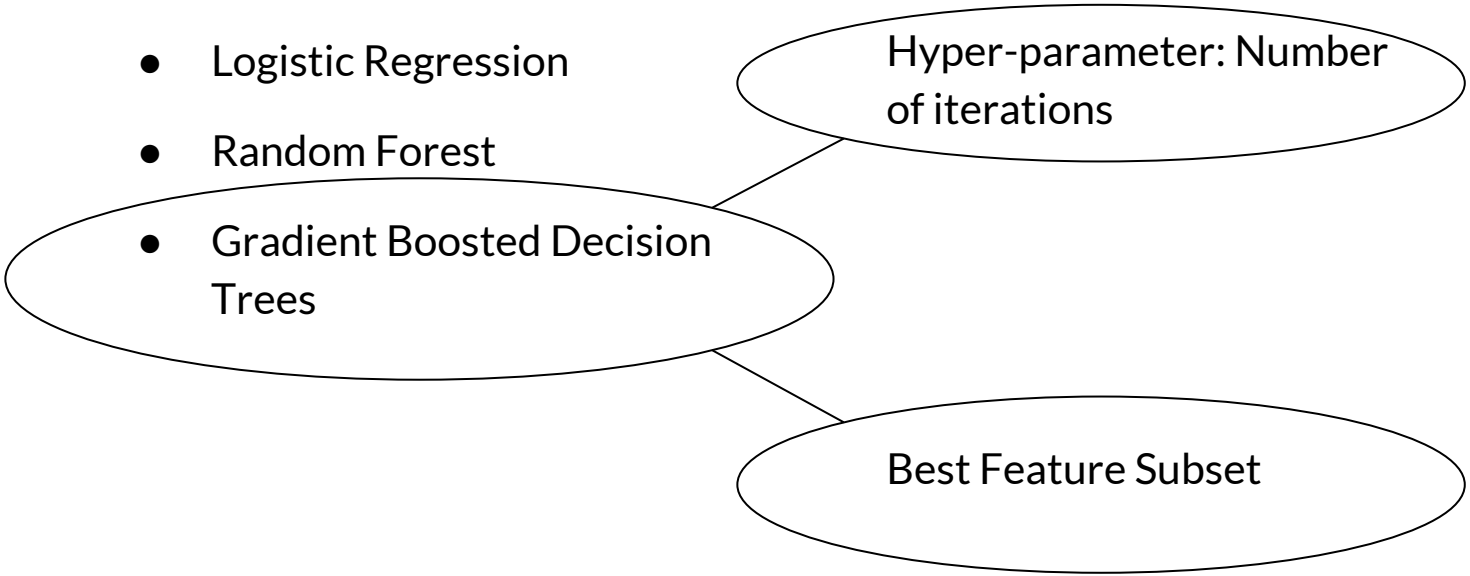
- Logistic Regression
- Random Forest
- Gradient Boosted Decision Trees

Models

- Logistic Regression
- Random Forest
- Gradient Boosted Decision Trees

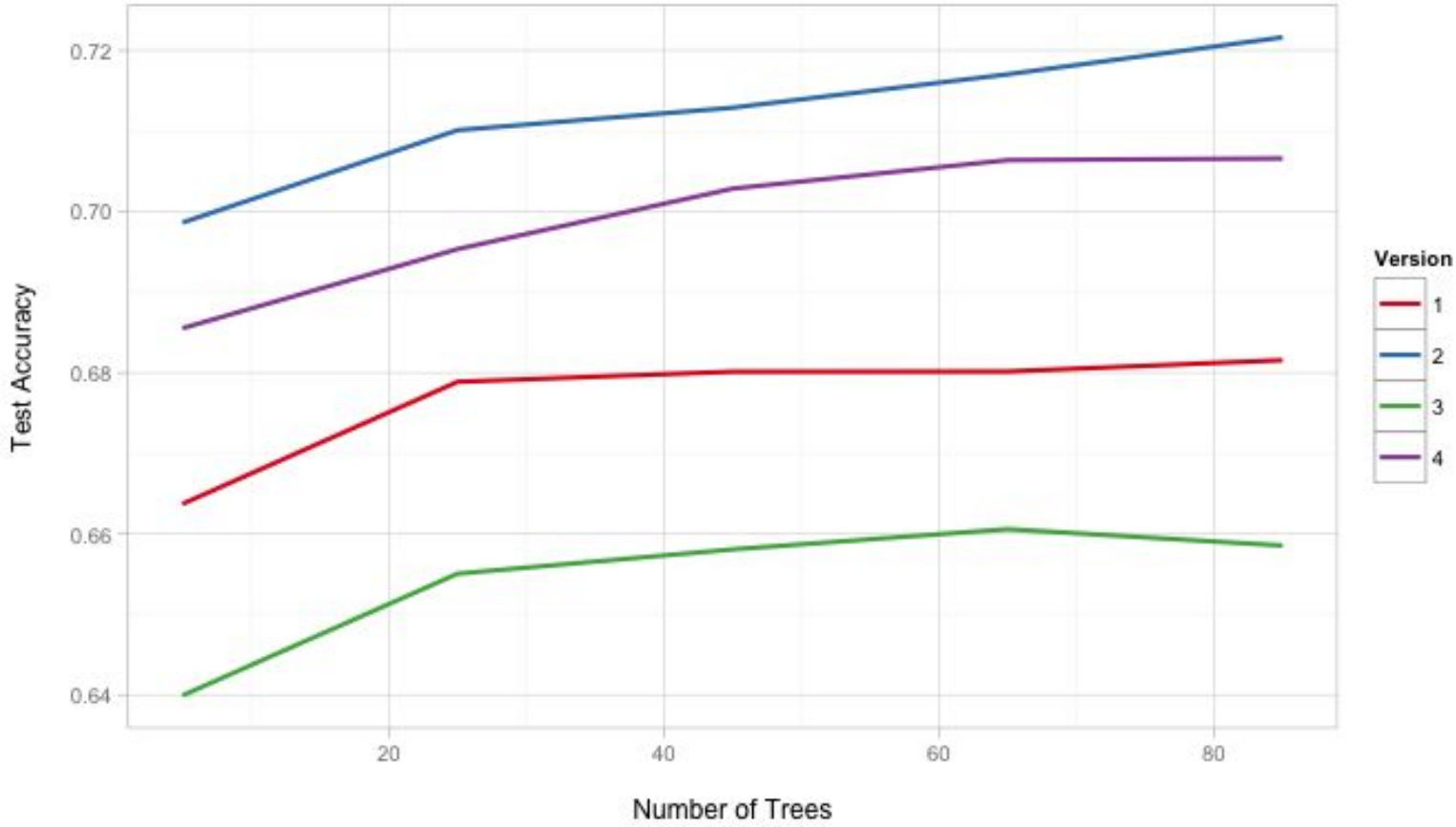
Models

- Logistic Regression
- Random Forest
- Gradient Boosted Decision Trees



Results:

Test Accuracy: GBT With 4 Feature Sets



GBT Version 2 Performance Metrics

Accuracy	Precision	Sensitivity	Specificity	Miss Rate
0.72	0.68	0.68	0.71	0.32

Version 2: All features without smoothing.

GBT Version 2 Performance Metrics

Accuracy	Precision	Sensitivity	Specificity	Miss Rate
0.72	0.68	0.68	0.71	0.32

Version 2: All features without smoothing.

Fin

suyoung.hong@gmail.com

horeilly2@dons.usfca.edu

mkkedia@dons.usfca.edu