

# Predicting bus time to arrival at stops using GPS data only

Author: Olga Chernytska

Email: olha.chernytska@gmail.com

## Goal

Develop an algorithm that predicts time to arrival on stop using GPS location as an input.

## Approach

- 1) Preprocessing to transform data from data frame 'time - vehicle id - latitude - longitude' to 'drive id - starting stop - ending stop - travel time in seconds'.
- 2) Linear Regression as baseline model and Gradient Boosting Regressor as improved.

Median absolute error for baseline model: 30.86 seconds  
Median absolute error for GBR model: 21.77 seconds

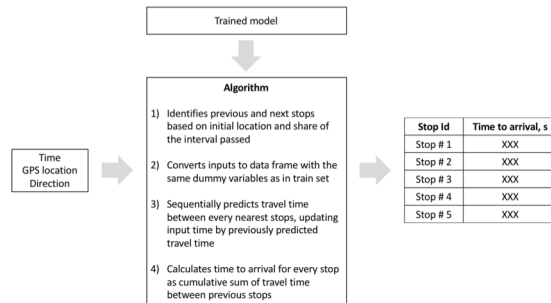
## Result

A working demo that accepts location, direction and time as inputs, and returns data frame that contains seconds to arrival and arrival time for the next 5 stops.

## Challenges

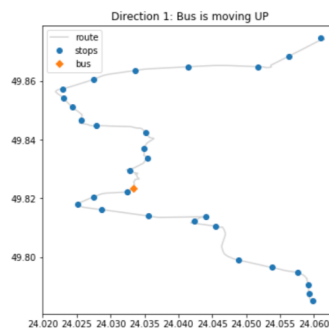
Location data required a lot of transformation before model training. Data was collected every minute, it sparse so artificial data as middle points were imputed. Data is dirty: stops locations do not corresponds to true stops where bus moves.

## Algorithm Schema



## Demonstration

Time: 2017-09-26T04:36:37.060401750  
Direction: 1  
Bus Location: [49.82356652125, 24.03342645875]



Bus is departed from stop #13: id=39164 and passed share=0.23 of the interval to stop #14: id=36559.

Next closest stops on the route: ['stop #14: id=36559' 'stop #15: id=36589' 'stop #16: id=36631' 'stop #17: id=36685' 'stop #18: id=36711']

	next_stop	seconds_to_arrival	arrival_time
0	stop #14: id=36559	155.847378	2017-09-26 04:39:12.907779750
1	stop #15: id=36589	298.304957	2017-09-26 04:41:35.365358750
2	stop #16: id=36631	433.488589	2017-09-26 04:43:50.548990750
3	stop #17: id=36685	628.974608	2017-09-26 04:47:06.035010750
4	stop #18: id=36711	795.359942	2017-09-26 04:49:52.420344750

## Practical Value

Model was developed using route of Lviv bus A53 (route id = 1723724), but most of the code is standardized, so it can be easily used for other routes and cities. Demo works fast, so it can be used for predicting in real time.

- The results of the project can be used for developing electronic scoreboards on the bus stops which are in many European cities but not so widely used in Ukraine.
- Results may be used by EasyWay to improve their time to arrival predictions.

