جامعة خليفة
Khalifa University

Apply machine learning to predict and avoid streetcar delays using TTC Streetcar Delay Data Deliverable 1

Adarsh Ghimire
100058927

11 APRIL 2021

Objective

- **Delay prediction of street car**
- **In production the machine learning model will predicts delay on each direction, then the direction with the minimum delay is the route the user will take to avoid delays.**

Description:

1. **Notebook → *1 Frame the problem and look at the big picture.ipynb* comprises detailed explanation about the problem scope, and way of tackling the problem.**
2. **Notebook → *2 Get the data and Explore Data.ipynb* comprises the process of data collection through API, the data cleaning, and processing of the data, the data visualizations, dependency visualization, etc. such that data is readily available for model training.**
3. **Notebook → *3 Short-list promising models.ipynb* comprises of implementation of the different models. The notebook starts with data refining, data conversion and scaling, then goes to various machine learning algorithm implementations and their results.**

Features utilized for model training

Features were cleaned, processed, and selected as per the analysis done on each of the categories on notebook 2

1. Route → Converted to categorical data
2. Time of a day → Converted to different part of day(Morning, Afternoon, Evening, and Night)
3. Day of the week → Converted to categorical data
4. Direction → 5 directional values were taken into consideration(NB, EB, WB, SB, BW)
5. Minimum gap → Continuous value, applied feature standardization.

Target

1. Given the above features above which are readily available predict the delay in minutes.

Different Models Performance

Models	R2 Score				MAE Score				RMSE Score			
	Train		Validation		Train		Validation		Train		Validation	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Linear Regression	0.641	0.063	0.524	0.793	3.867	0.400	3.902	0.347	17.74	1.803	17.02	10.65
Polynomial Regression (degree 2)	0.830		0.852		2.979		2.947		12.48		9.558	
Bayesian Regression	0.641	0.063	0.524	0.792	3.868	0.400	3.903	0.347	17.74	1.802	17.02	10.65
SVM Regression	Took so much time to train											
Random Forest	0.936	0.005	0.805	0.108	1.340	0.038	2.362	0.393	7.534	0.331	12.54	3.710
Gradient Boosting	0.873	0.010	0.844	0.087	2.157	0.053	2.264	0.35	10.587	0.468	11.16	3.550
Neural Network (100 neurons)	0.848	0.014	0.838	0.088	2.665	0.331	2.755	0.267	11.57	0.554	11.34	3.694
Neural Network (64, 32 neurons)	0.850	0.015	0.836	0.096	2.898	0.465	3.002	0.378	11.52	0.595	11.42	3.7562

List of promising models

1. Polynomial regression of degree 2
2. Random forest regression (100 estimators)
3. XGBoost regression (100 estimators)
4. Neural Network (1 hidden layer) with 100 neurons
5. Neural network (2 hidden layers) with 64,32 neurons

جامعة خليفة
Khalifa University

Thank You

ku.ac.ae