Car Repositioning System for Staff, Users, and Autonomous Vehicles

Project by: Nguyen Lai - Yaying Huang
Advisers: Prof. Samuel Burden
Peter Dempster (ReachNow)
Daniel Torres (ReachNow)
Dominik Gusenbauer (ReachNow)

Background

- ReachNow is operating a free-floating car sharing service in Seattle, Portland, and New York.
- In Seattle. ReachNow members can end their trips in any regular or metered on-street parking spots within the Seattle area.
- This may result in not enough cars being available to satisfy demands in certain areas at some points in time.
- There is often a mismatch between demands and supplies of ReachNow cars. Therefore, car relocation is needed.

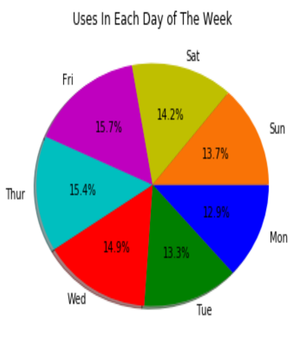
Dataset & Approach

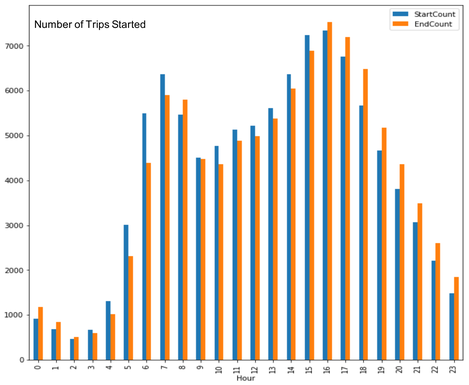
Data Analysis and Visualization:

- Datasets were given in form of historical trips record.
- The following is an example of a single historical trip record in the dataset

eventid	vehicleid	starttime	endtime	mileagestart	mileageend	fuelliterstart	fuelliterend	memberid	startlat	starting	endlat	endlng
aaaabbbbcccc	999	3/8/2018 2:00PM	3/8/2018 3:00PM	20000	20100	50	45	eeefffgg	somewhere	somewhere	somewhere	somewhere

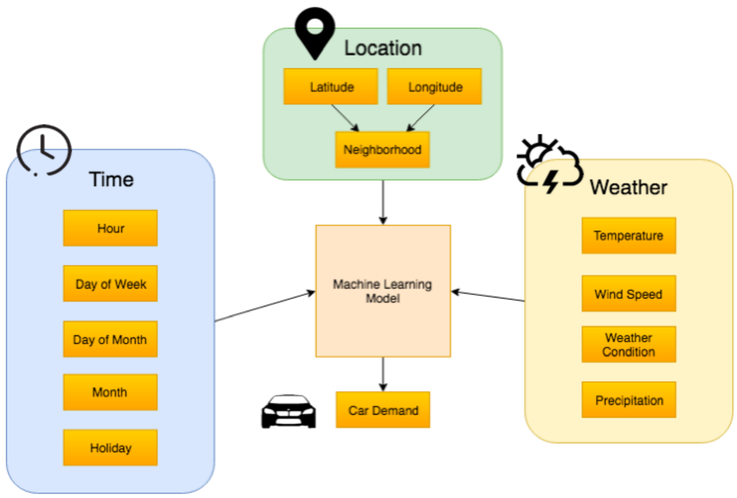
- To gain better knowledge and understanding of given datasets, the datasets were first analyzed and visualized.
- Figures below show how different days and different hours can affect car usage.





Approach:

Feature Flow Chart for Machine Learning Model:



Data Preprocessing:

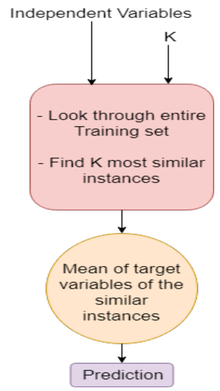
- Clean outliers and noise.
- To reserve cyclical effect of time features, we decompose hour/day of week/month into sin/cos components.
- Normalization/Standardization.

Machine Learning (ML) Algorithms

Three different algorithms were used:

K-Nearest Neighbors (K-NN):

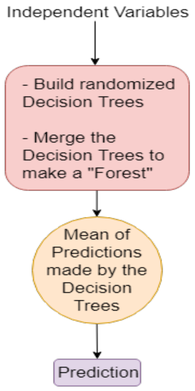
- Among one of the simplest ML algorithms.
- How prediction is made:



- Why this algorithm?
  - Cost of learning is 0.
  - Simple to implement.
  - Suitable for our problem (predict future cars usage based on similar historical usage.)

Random Forest (RF):

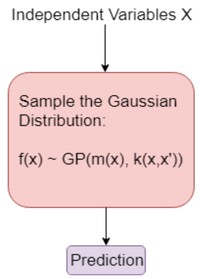
- Ensemble of Decision Trees.
- How prediction is made:



- Why this algorithm?
  - Versatility.
  - Capable of explaining complex dependencies.
  - Run efficiently on large datasets.
  - Simple yet powerful.

Gaussian Process (GP):

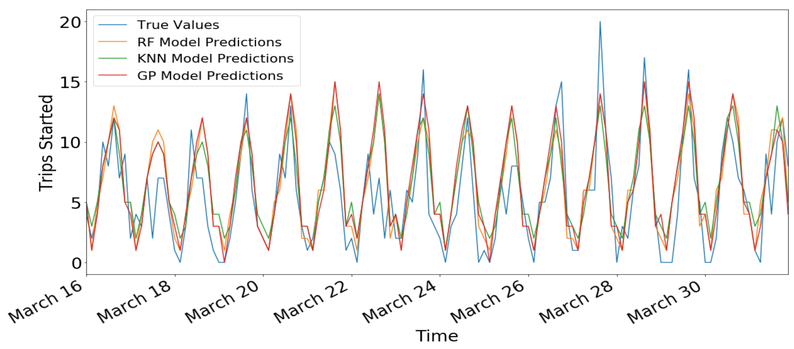
- Describes a distribution over functions.
- A Gaussian Process is defined by a mean function m(x) and a covariance function k(x,x').
- How prediction is made:



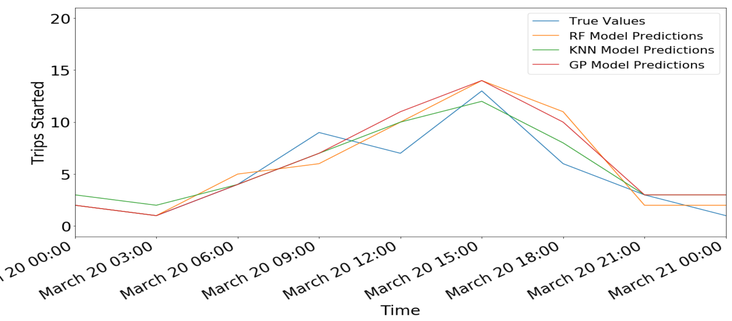
- Why this algorithm?
  - Ability to apply different kernels based on knowledge of dataset.
  - Built-in optimizer for tuning hyperparameters.

Results

- Predictions of car usage from March 16th to March 31st in the University District, Seattle:

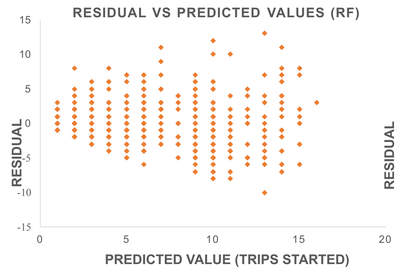


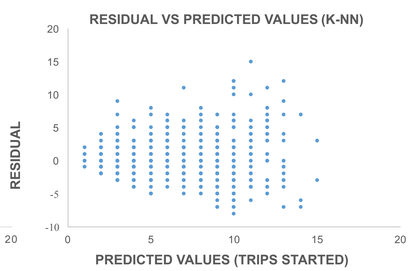
- Predictions of car usage in one day (March 20th):



Because of limited space, only one neighborhood is shown here.

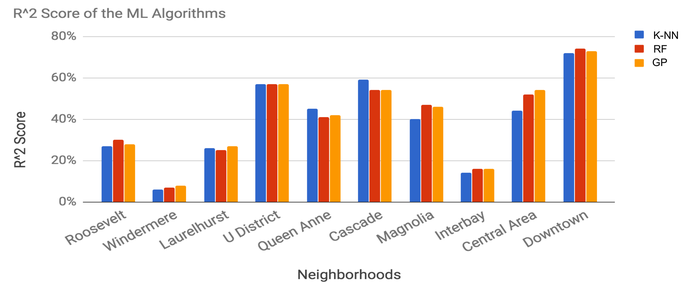
- The two figures below shows the residual versus predicted values plot for RF and K-NN models. GP model is not shown because of its similarity to RF model.
- Residual is the difference between observed values and predicting values.



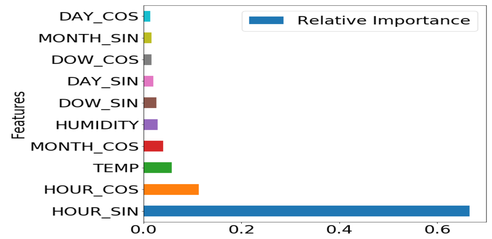


Discussion & Conclusion

- As can be seen from the results, Random Forest and Gaussian Process performed slightly better than K-NN. K-NN algorithm failed to predict the dips in car usage.
- All three of the algorithms are not capable to predict some of the highest peaks in car usage. This is because we currently don't have features to explain sudden spikes in demand, such as events like football games, block parties, etc.
- Neighborhoods play a crucial role in the predictability of car usage, as can be seen in the figure below.



- Hour is also an important feature in predicting car usage. The chart below illustrates features importance:



- Random Forest and Gaussian Process produced a very similar result. However, Gaussian Process Regression requires very long training time.
- Therefore, Random Forest will be selected as the main algorithm moving forward.
- **Overall, the models performed well. The results resembled the real values and closely followed the same pattern.**

Future Work

- Deploy models on Amazon Web Services and operationalize the models.
- Design Front-End User Interface.
- Incorporate event feature to explain spikes in demands.
- Try additional model: Deep Learning – Neural Network for better performance.

ELECTRICAL ENGINEERING
UNIVERSITY of WASHINGTON