Traffic Flow Forecasting

Uthman Alibalogun, Alex Hubbard, Patrick Yim, Vincent Qu
W207 – 16 April 2024

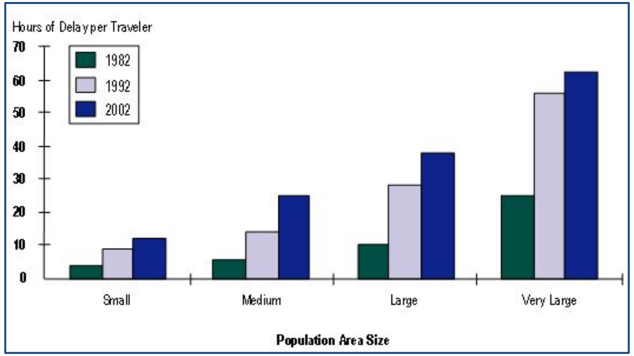
Berkeley
UNIVERSITY OF CALIFORNIA

Introduction: Background

In February 2024, total vehicle miles (240 billion miles) in the United States increased by 2.0% YoY according to trend reports from the Federal Highway Administration.

Traffic congestion has also increased causing gridlocks, delays and unpredictable travel times. Among areas with high population, hours of delay has increased by 9% between 2002 and the decade before. (Figure ES.1)

Federal Highway Administration



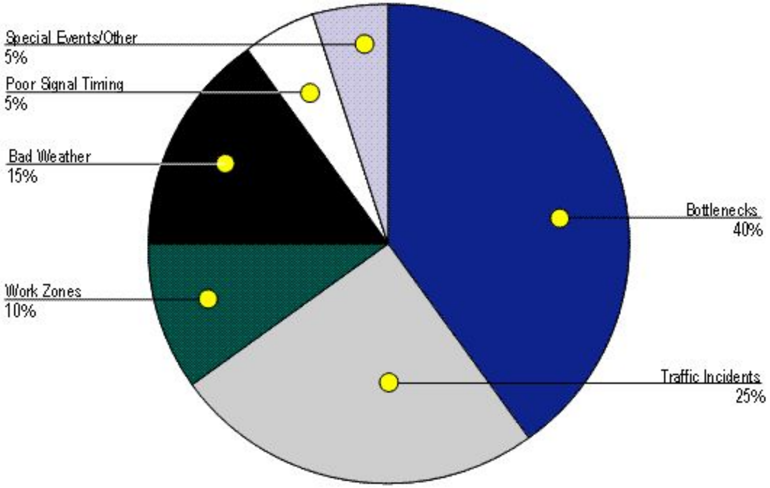
Berkeley
UNIVERSITY OF CALIFORNIA

Introduction: Motivation

Traffic congestion can lead to increased fuel consumption, environmental pollution, and decreased quality of life.

The top two causes for traffic congestion are bottlenecks (40%) and traffic incidents(25%). (Figure ES.2)

Can we leverage traffic sensors to anticipate traffic changes, make adjustments, and decrease congestion?



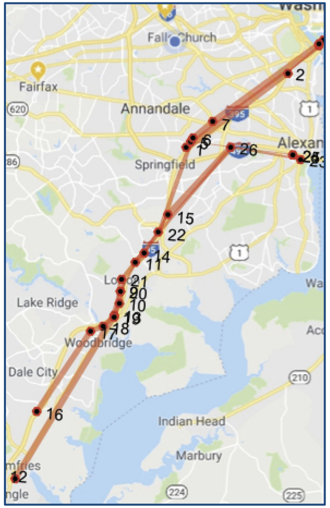
Berkeley
UNIVERSITY OF CALIFORNIA

Data Source

The data we're using is from the [UCI Machine Learning Repository] and was gathered in 15 minute increments across 36 sensor locations along two major highways in the Northern Virginia/Washington D.C. capital region.

Features gathered include:

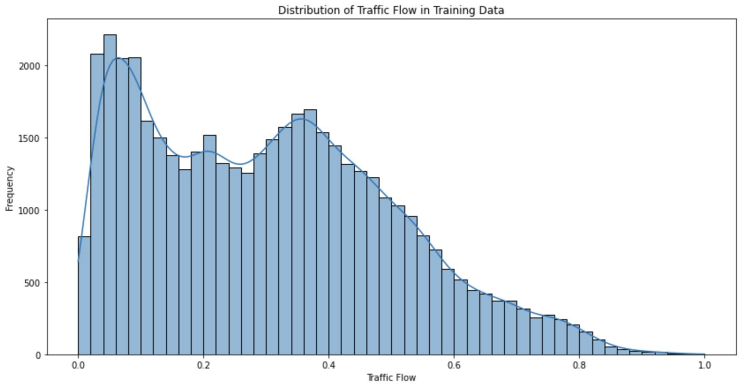
- Last 10 traffic volume measurements
- Day of Week
- Hour of Day
- Road Direction
- Number of Lanes
- Name of Road



Exploratory Data Analysis

The dataset is set up as an array of matrices with each matrix consisting of 36 sensors and their corresponding 48 features (36 x 48).

When looking at the distribution of traffic volume feature we learned that the unit being used is a range from 1.0 (heavy traffic) to 0.0 (low traffic volume) and that most of the time traffic volume is on the lower end with a median of 0.297.

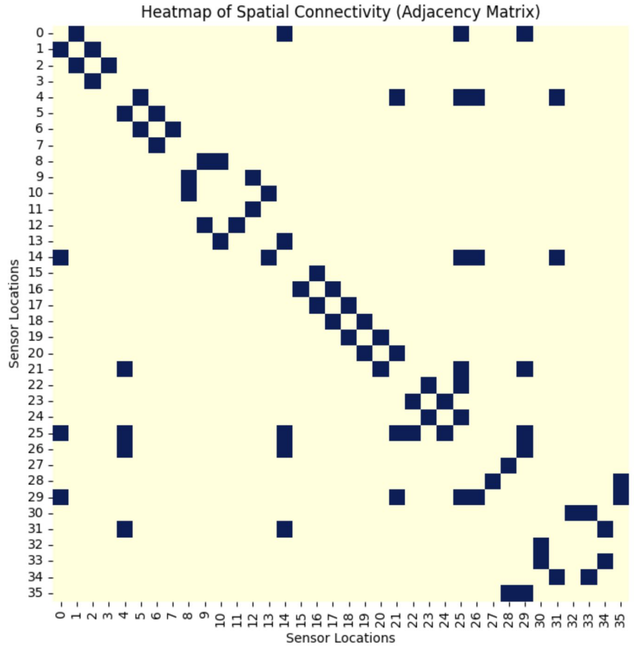


Berkeley
UNIVERSITY OF CALIFORNIA

Exploratory Data Analysis

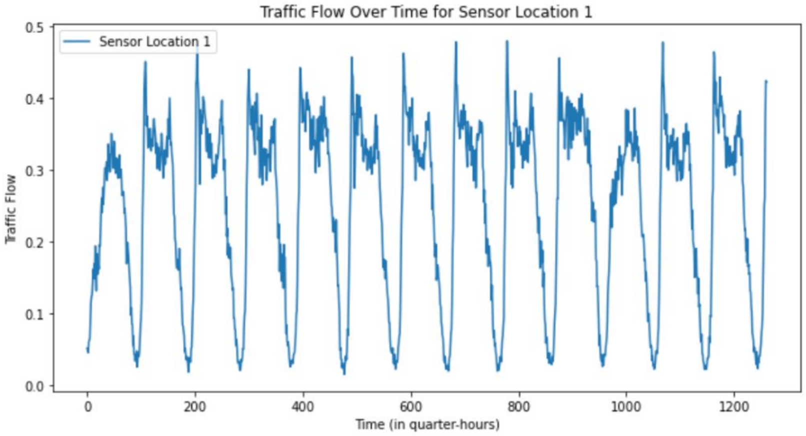
We also found that the sensors were largely connected to each other in a chain through an adjacency matrix.

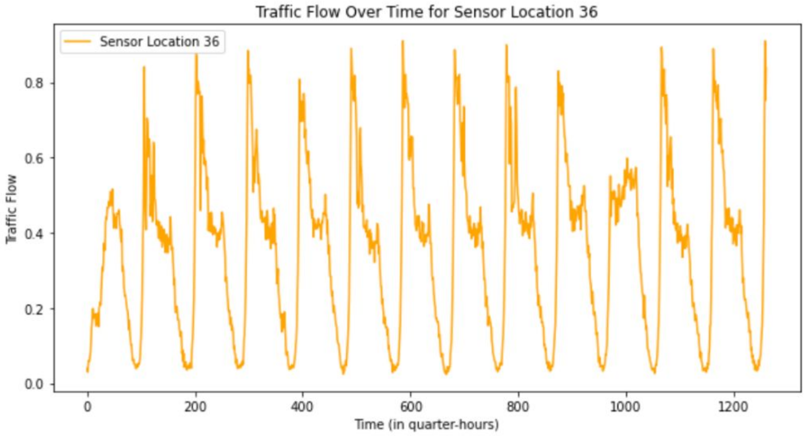
A few sensors breaking the chain, potentially sensors placed near off ramps or connecting highways.



Exploratory Data Analysis

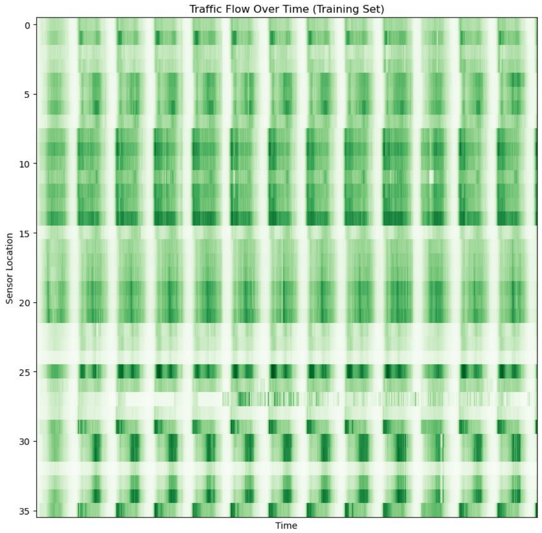
We also found that traffic volume is cyclical and has a pattern throughout the week when mapped over hour of day.





Exploratory Data Analysis

We mapped the traffic flow data for each sensor to visualize the variability between sensors and we can see that some sensors have higher traffic volume than others but follow a similar pattern.



Berkeley
UNIVERSITY OF CALIFORNIA

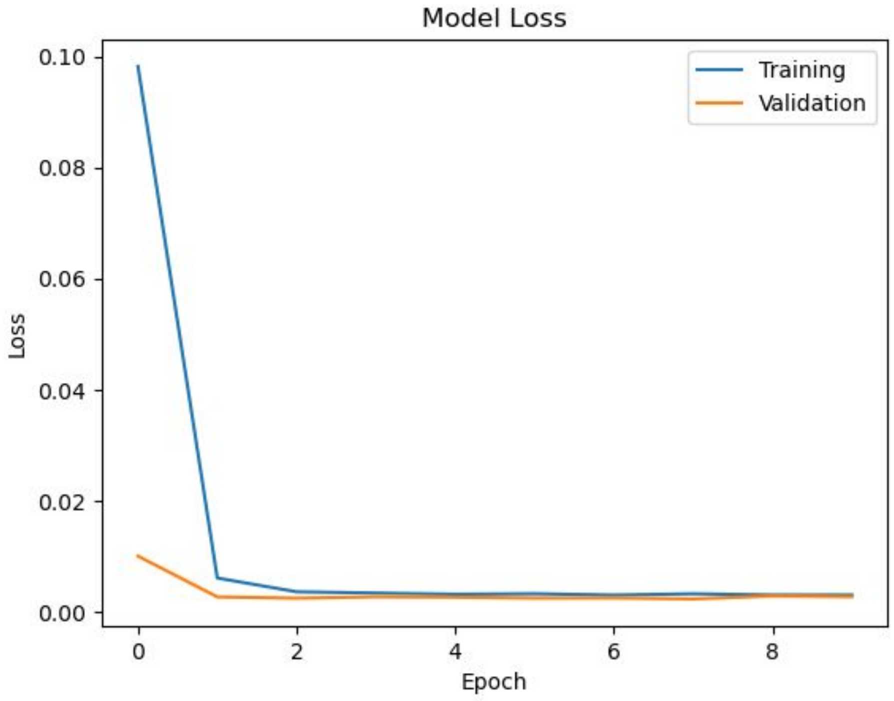
Modeling - Baseline and Final Model Performance

- Want to predict traffic flow for 15 minutes in the future

Model	Validation Loss	Test Loss
Baseline	1.07	1.13
Linear Regression	0.0017	0.0041
Single-layer NN	0.0027	0.0043
Multi-layer NN	0.0017	0.0026

Single Layer Neural Net

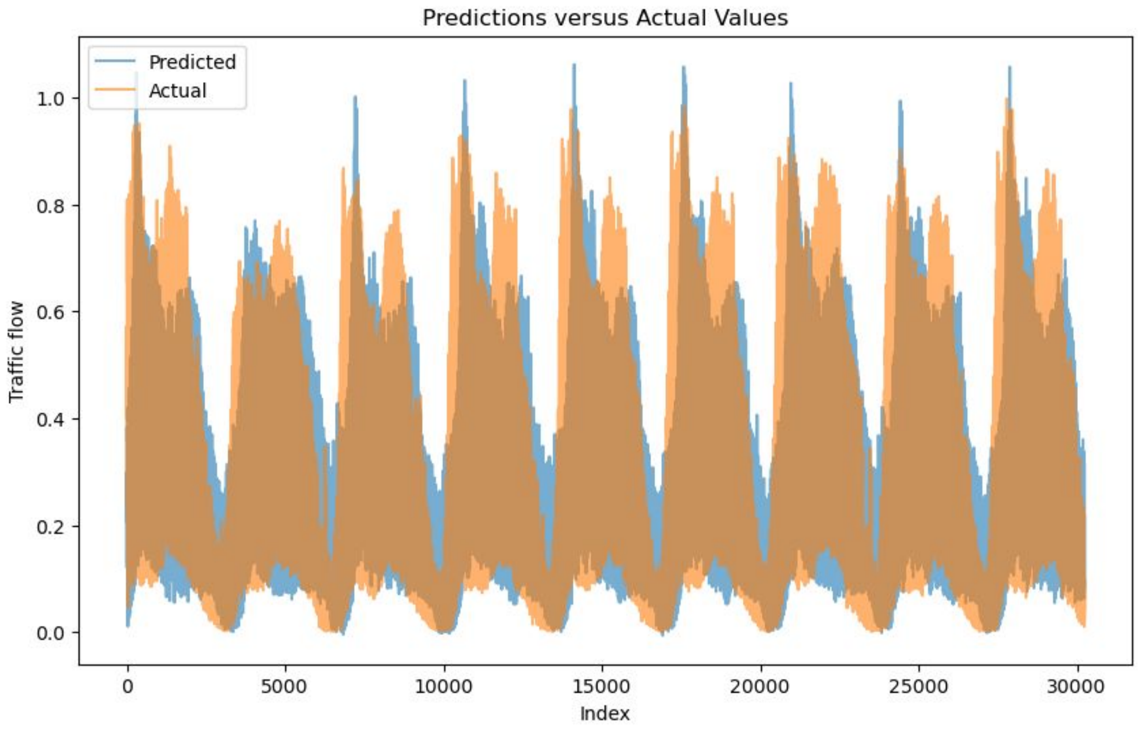
- Steep Drop after 1st epoch
- Convergence
- Overfitting



Berkeley
UNIVERSITY OF CALIFORNIA

Predicted vs Actual Values

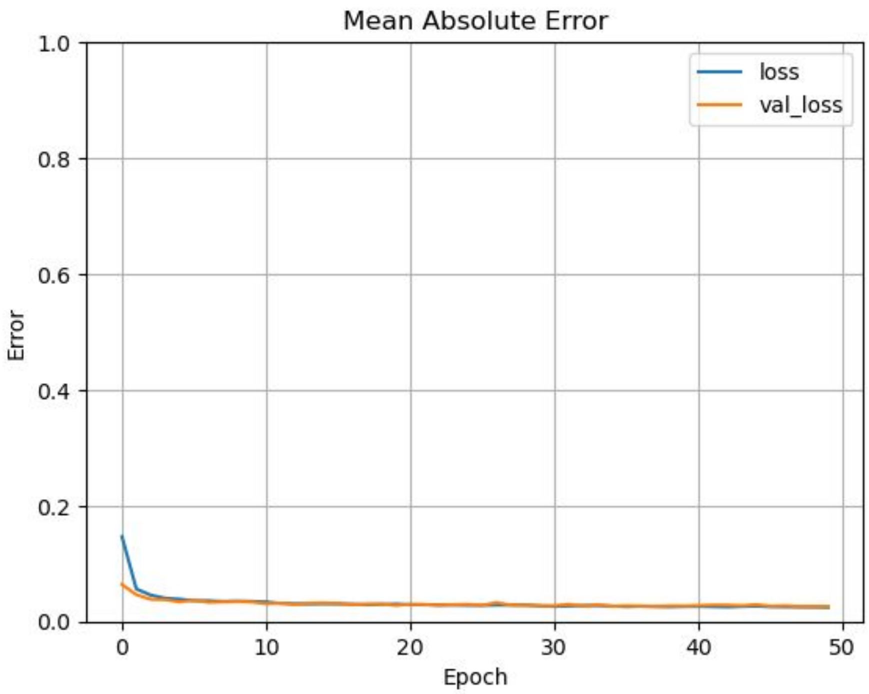
- Peaks and Valleys
- Comparison

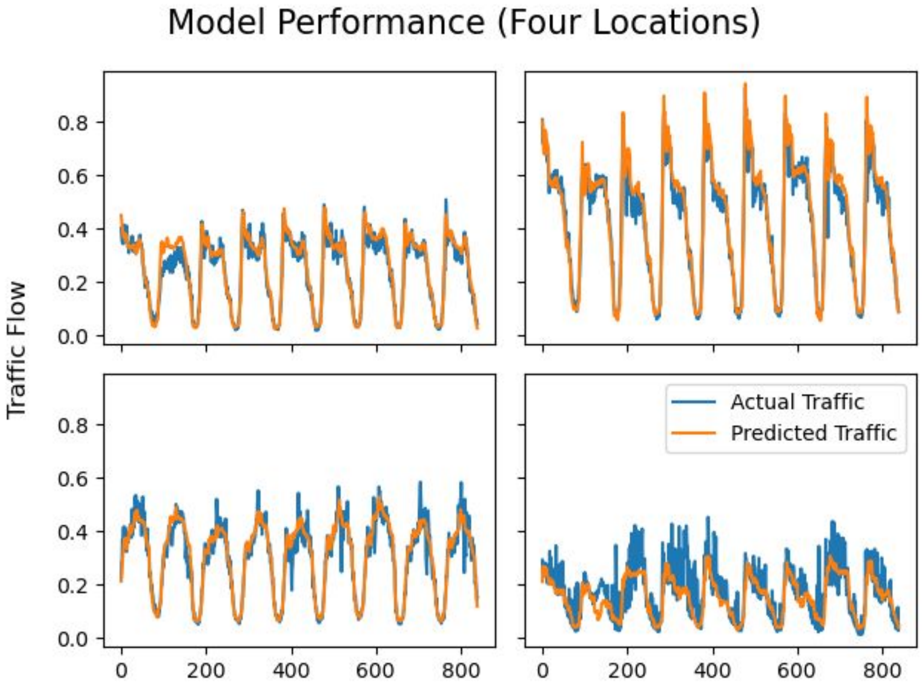


Experiments - Hyperparameter tuning

HIDDEN SIZES	ACTIVATION	OPTIMIZER	LEARNING RATE	#PARAMETERS	TEST ACCURACY
[]	relu	SGD	0.01	62244	0.320
[]	tanh	SGD	0.01	62244	0.223
[]	relu	Adam	0.01	62244	0.146
[128]	relu	Adam	0.01	225956	0.227
[256,128]	relu	Adam	0.01	480164	0.573
[256,128, 64]	relu	Adam	0.01	486116	0.590
[256,128, 64, 32]	relu	Adam	0.01	487044	0.556

Final Model Performance





Conclusion

Multi-layer NN model has the best performance but is not performing exponentially better than the simple linear regression. We would feel comfortable that the model could identify abnormal traffic flow leading to congestion.

After adding multiple layers, hyperparameter tuning was not very impactful

Although when considering real life application, 15-min window may be too short of time for traffic control to make any actual change but could at least provide insight for future city work planning or bottleneck discovery. Can also be useful for navigation tools to make on the fly adjustments to avoid delays.

Ethical Considerations

1. For all authors...
   - (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? **Yes**
   - (b) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes**
   - (c) Did you discuss any potential negative societal impacts of your work? **Yes**
   - (d) Did you describe the limitations of your work? **Yes**
2. If you are including theoretical results...
   - (a) Did you state the full set of assumptions of all theoretical results? **Yes**
   - (b) Did you include complete proofs of all theoretical results? **Yes**
3. If you ran experiments...
   - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **Yes**
   - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **Yes**
   - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **Yes**
   - (d) Did you include the amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **Yes**
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets... **Not Using**
5. If you used crowdsourcing or conducted research with human subjects... **Not Using**

Limitations

Context for features missing:

1. No coordinates to identify specific location of road. Need additional information and location knowledge to connect sensors to actual locations.
2. Traffic volume units is limited to a 0 to 1 scale that is applied across all sensors without accounting for past traffic density.
3. Additional information regarding historical data could help scale 0 to 1 depending on traffic density.

Other applications and further research

Civil Use Case

The model can be used to run simulations through understanding spatial variability in traffic flow. We can attempt to simulate traffic congestion to understand how bottlenecks in one section could affect overall traffic flow of highway. Results of the simulation can be used to help governments make traffic design decisions.

Commercial Use Case

The model using live sensor data can be used to add more contextual information for navigational apps such as Google Maps or Waze.

Thank you!

Contributions

- **Motivation:**
  a. Uthman Alibalogun
  b. Patrick Yim
  c. Vincent Qu
- **Data:**
  a. Uthman Alibalogun
  b. Alex Hubbard
  c. Patrick Yim
  d. Vincent Qu
- **Modeling:**
  a. Uthman Alibalogun
  b. Alex Hubbard
  c. Patrick Yim
- **Experiments:**
  a. Uthman Alibalogun
  b. Alex Hubbard
  c. Patrick Yim
- **Conclusions:**
  a. Uthman Alibalogun
  b. Patrick Yim
  c. Vincent Qu

Citing

- Liang Zhao, Olga Gkountouna, and Dieter Pfoser. 2019. Spatial Auto-regressive Dependency Interpretable Learning Based on Spatial Topological Constraints. ACM Trans. Spatial Algorithms Syst. 5, 3, Article 19 (August 2019), 28 pages. DOI:https://doi.org/10.1145/3339823

- Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation, Federal Highway Administration, 2005, https://ops.fhwa.dot.gov/congestion_report/executive_summary.htm

- Traffic Volume Trends, Federal Highway Administration, Feb 2024, https://www.fhwa.dot.gov/policyinformation/travel_monitoring/24febtvt/24febtvt.pdf

Berkeley
UNIVERSITY OF CALIFORNIA