Modeling and Predicting the Cascading Effects of Delay in Transit Systems

Aparna Oruganti[1], Sanchita Basak[2] , Fangzhou Sun[3], Hiba Baroud[4], Abhishek Dubey[2]
[1]Analytics Associate, Presbytarian Healthcare Services
[2]Department of Electrical Engineering and Computer Science, Vanderbilt University, [3]Research Scientist, Facebook
[4]Department of Civil Engineering and Department of Earth and Environmental Science, Vanderbilt University

VANDERBILT UNIVERSITY

ABSTRACT

Research and Motivation

- Smart cities need effective transit systems planning. With the integration of AVL(automatic vehicle locators), transit agencies are able to provide arrival times estimates in real-time.
- For the study, we use data collected over several months from one such transit system and show how this data can be potentially used to learn patterns of travel time during special events such as sports games and music award ceremony.

Goal

- Study the effect of NFL football games and other factors such as traffic and weather on the transit system delay using various machine learning algorithms.
- Predict the transit travel time and visualize the cascading effect across the transportation system using heatmaps.
- Use these models to design adaptive and transient transit schedules during scheduled events.

BACKGROUND

- As a part of the smart city initiative, Nashville is involved in the implementation of Intelligent Transportation Systems (ITS).

- Traffic congestion can be caused by
  - Regular rush hour traffic
  - Inclement weather conditions
  - Incidents and special events

NASHVILLE—This is one of America's booming cities. An average of 82 people move here every day [4].

- **Cascading effects** are defined as disruptions in one system affecting the performance of other interconnected systems.

Traffic in downtown Nashville caused by a football game can cascade to delays in other interconnected road networks resulting in a gridlock.

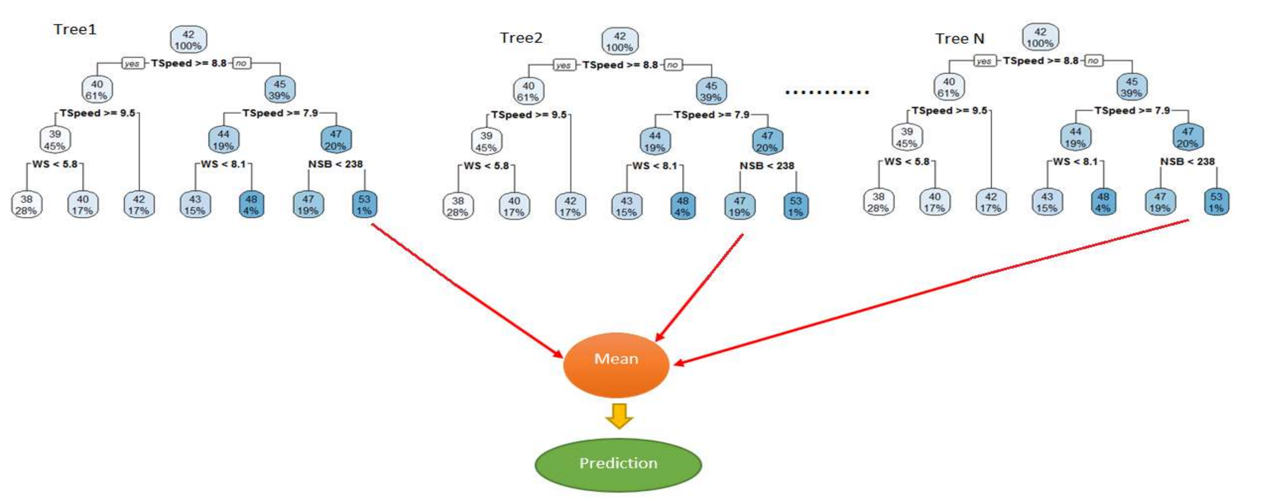
- Understanding cascading effects of delay can help provide real time and reliable updates on the performance of the transit system.
- The predicted delay and cascading effect can inform an adaptive transit schedule

METHODOLOGY

Ensemble Tree based models

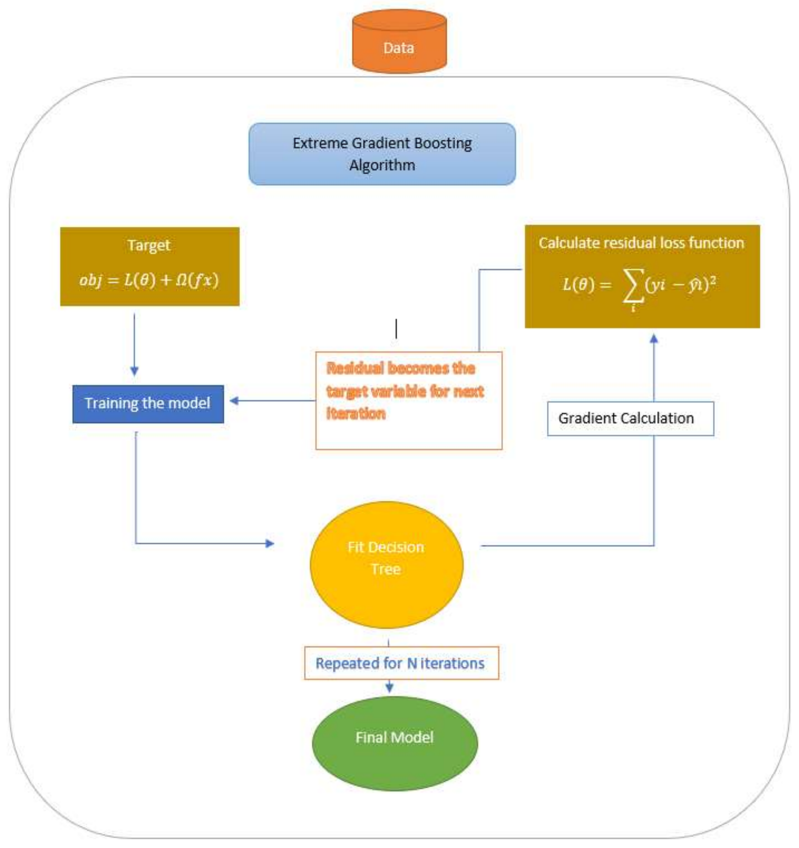
- Random Forests
- Extreme gradient boosted trees

Random Forest (RF)



Extreme Gradient Boosted Method (XGBoost)

- An ensemble method used for regression and classification.
- In gradient boosting, trees are grown sequentially using information from previously grown trees.



Advantages

- Better support for multicore processing which reduces overall training time.
- These enhancements make a big difference in speed and memory utilization.

RESULTS

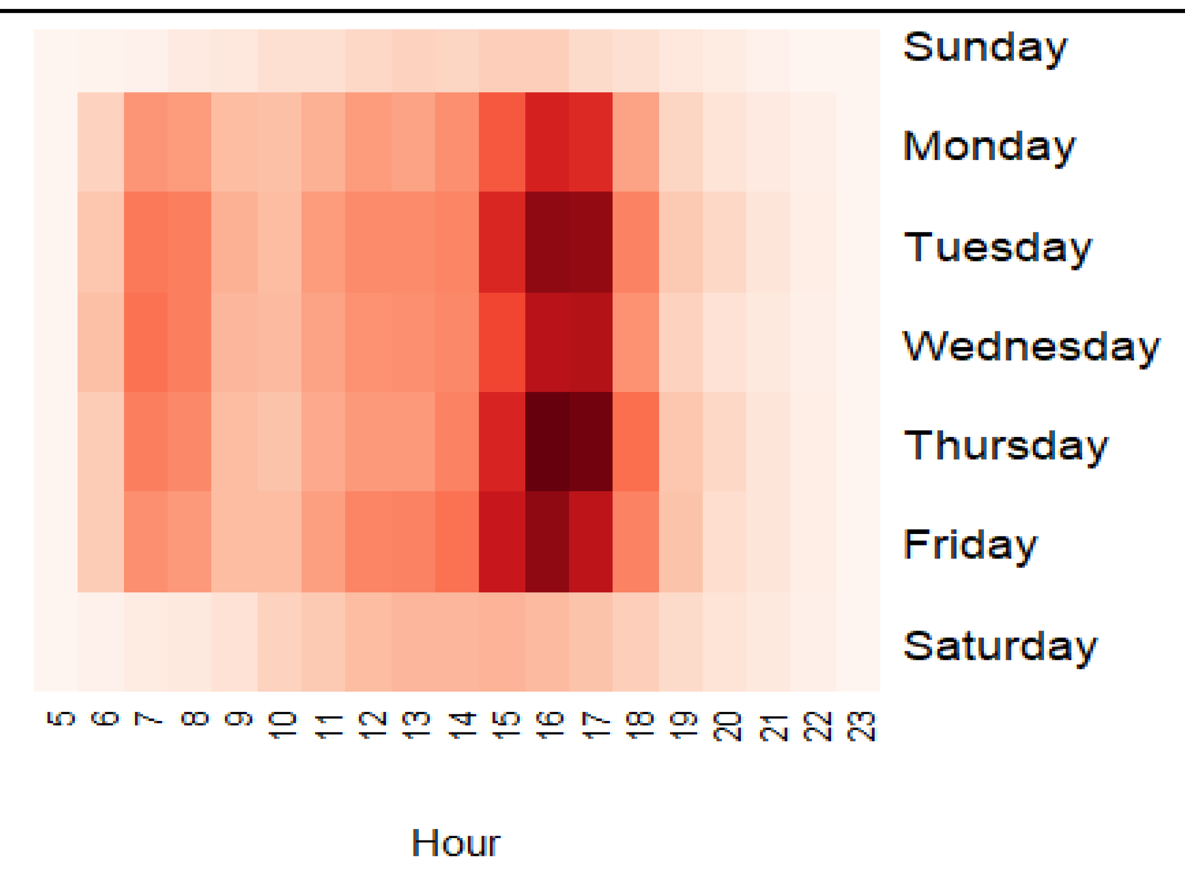


Figure 1 : Heatmap showing the correlation between travel time , day of the week and hour. A darker color indicates increased traffic.

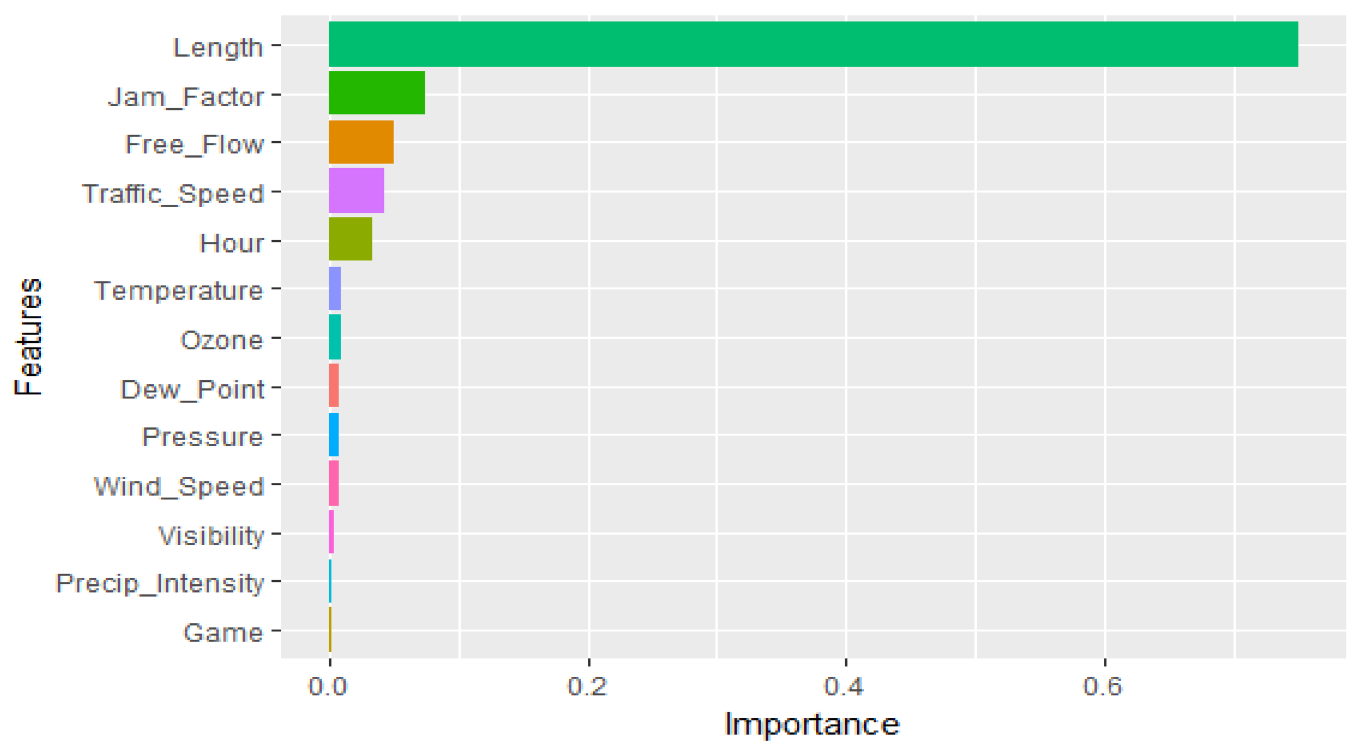


Figure 2 : Variable importance plot

Model	R^2	RMSE	Time
Extreme Gradient Boost	0.80	2.01	13.13 min
Random Forest	0.78	2.7	4 hr 25 min
General Adaptive Model	0.50	2.94	44 min
Linear Regression	0.46	3.05	0.10 nin

Figure 3: Summary of the models with their performance in terms of goodness of fit (R^2) and predictive accuracy (RMSE)

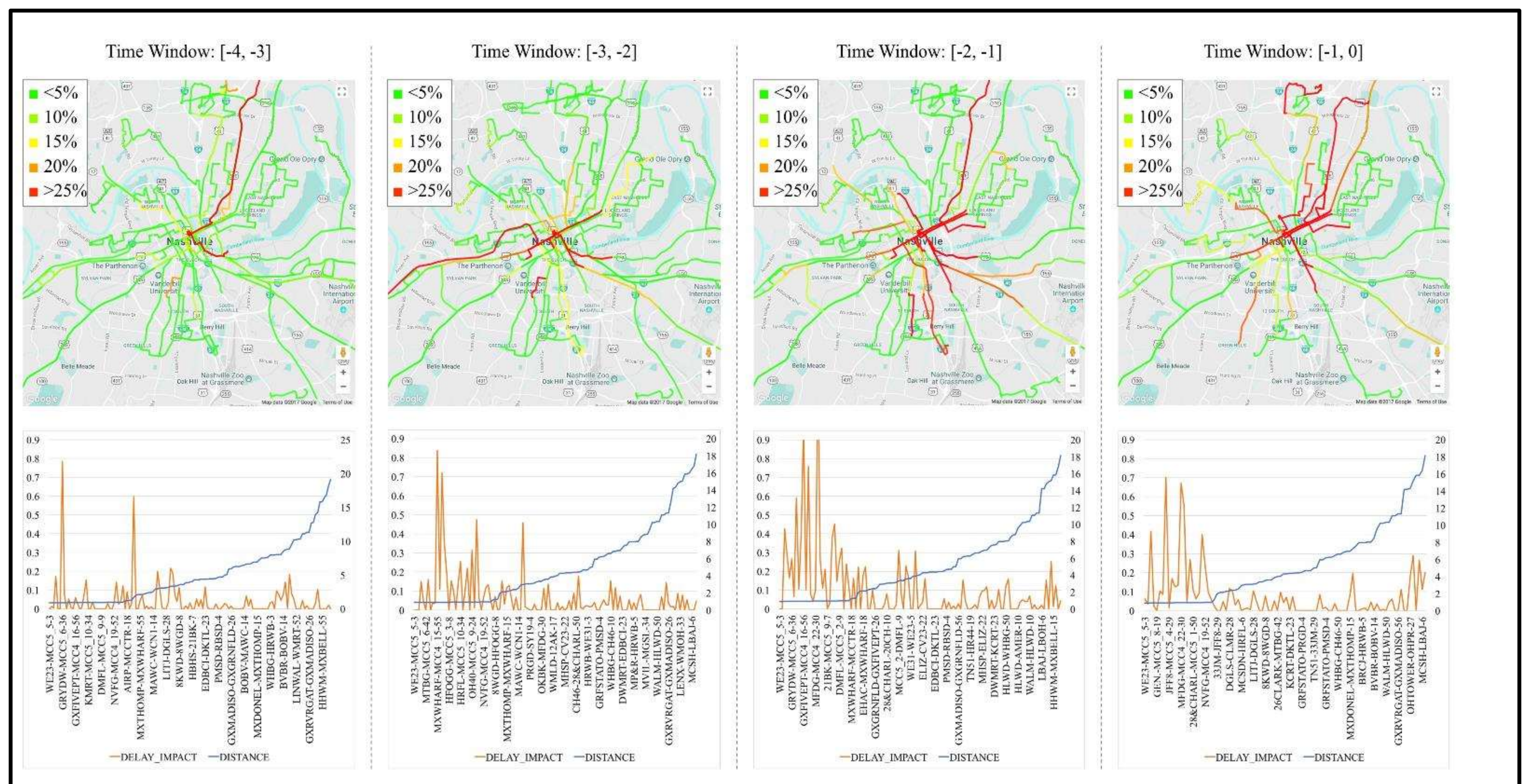


Figure 6: Impact of football games on traffic congestion based on historical average of all NFL games in four one-hour time windows before the start of the game: (a) from 4 hours to 3 hours, (b) from 3 hours to 2 hours, (c) from 2 hours to 1 hour, (d) from 1 hour to 0 hour.

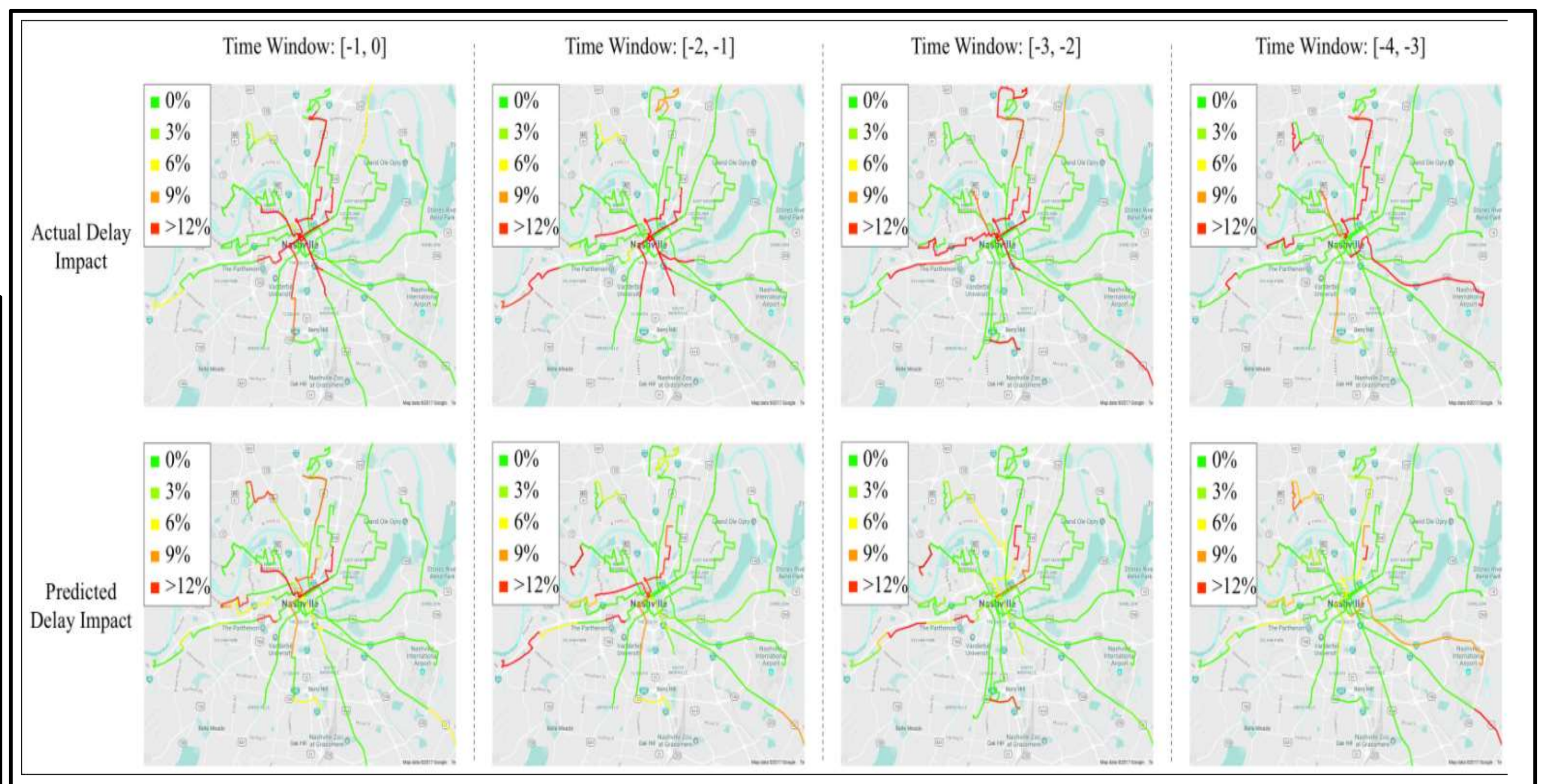


Figure 5: Predicted impact of football games on traffic congestion in four one-hour time windows before the start of the games (a) from 4 hours to 3 hours, (b) from 3 hours to 2 hours, (c) from 2 hours to 1 hour, (d) from 1 hour to 0 hour.
The colors represent the difference in delay between game days and non-game days. For example, if the average travel delay of non-game days on a route segment is 10 mins, and the average travel time on game days is 12 mins, then the difference ratio is 20% (i.e., (12-10)/10 = 20%).

CONCLUSIONS

- We are able to explain more than 80% of the variance in the bus travel time and we can make future travel predictions for each timepoint segment with an out-of-sample error of 2 minutes with information on bus schedule, traffic, weather and special events.
- Based on predicted delay, cascading effect of traffic in downtown can cover up to 6 mile radius.

Future Work

- Integrate predictive analytics with
  - (a) the decision framework for DelayRadar to help the transit agency develop a dynamic transit schedule during special events,
  - (b) the transit-hub application that provides the delay estimates to riders

ACKNOWLEDGEMENTS

This work is sponsored as part of the Vanderbilt Initiative on Smart City Operations and Research, a trans-institutional initiative funded by the Vanderbilt University. We acknowledge the support and suggestions provided by our partners from Nashville Metropolitan Transport Authority.

References

1. A. Oruganti, F. Sun, H. Baroud and A. Dubey, "DelayRadar: A multivariate predictive model for transit systems," 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, 2016, pp. 1799-1806. doi: 10.1109/BigData.2016.7840797.
2. A. Oruganti, F. Sun, H. Baroud and A. Dubey, "Modeling and Predicting the Cascading Effects of Delay in Transit Systems". (unpublished)
3. https://www.tennessean.com/story/entertainment/music/cma-music-festival/2017/06/09/sunday-nashville-beautiful-disaster/378183001/
4. https://www.theatlantic.com/business/archive/2015/10/nashville-charlotte-public-transit/412741/