# Research Proposal – Transportation Safety

Traffic accidents are a major public safety concern, causing significant loss of life and injury each year. In 2021 alone, 1.3 million people were killed in road traffic accidents around the world. The ability to predict traffic accidents and identify the factors that contribute to their occurrence would be a valuable tool in reducing the number of accidents on the roads. This project aims to forecast traffic accidents based on a variety of factors such as enforcement, road conditions, and infrastructure upgrades. By analyzing historical data, we aim to identify patterns and trends that can be used to predict future accidents. The goal of this research is to develop a model that can accurately predict traffic accidents, and in turn, aid in the development of strategies to reduce the number of accidents on the roads and make our roads safer for all users.

Vision Zero (VZ) is a road traffic safety concept that aims to achieve a transportation system with no fatalities or serious injuries. It is based on the idea that traffic crashes are not accidents, but rather the result of human error and system failures that can be prevented through the design of safer roads, vehicles, and behaviors. New York City took VZ pledge in 2014 and since then, more than 1.6B$ were invested in it. NYC's VZ agenda focuses on enforcement, engineering, legislation, and education. In order to examine the effectiveness of VZ we decided to try to answer the following hypothesis:

*"What is the probability for a traffic accident based on temporal and spatial analysis of infrastructure improvements and traffic law enforcement in NYC?"*

We hope to be able to benefit city engineers and transportation researchers in understanding how do enforcement and infrastructure changes effect road safety. This will be reflected in several aspects:

1. Improved road design: identify areas where accidents are more likely to occur, such as intersections or certain types of roads. This information can be used to design safer roadways, such as adding speed humps, traffic signals or other safety features.
2. Improved enforcement: Accident prediction can help identify areas where accidents are more likely to occur and the factors that contribute to them, such as speeding or distracted driving. This information can be used to target enforcement efforts and increase the effectiveness of traffic safety campaigns.

3. City planner simulations: Accident prediction can be used to fit into existing simulations that currently helping city planners to predict the traffic flow in new roads. The upgraded simulations will be able to take into consideration the potential impact of different infrastructure projects or traffic control measures on safety. This can help city planners make more informed decisions about how to allocate resources and design safer streets.

Over the past years several researches were trying to examine factors for traffic accidents in order to build models that predict accidents. We will further describe a scientific paper that used state-of-the-art method for that purpose.

Yu et al. (2021)[1] tried to predict traffic accident In Beijing based on road networks, traffic accidents, taxi GPS, meteorological data and POI's. They design a deep spatial-temporal graph convolutional network (DSTGCN). Their proposed model is composed of three components: a spatial layer which performs operations on spatial information, a spatial-temporal layer which utilizes graph and standard convolutions to capture the dynamic variations in both spatial and temporal perspective, and an embedding layer which aims to obtain meaningful and semantic representations of external information. They compared this model with a few classical ML models such as SVM and decision tree and even two DL models and the DSTGCN outperform them all by all tried metrics.

There is a limited number of studies that have conducted data-driven research on the effectiveness of VZ policies. One notable example is the paper by Ferenchak et al. (2022)[2], which examines the impact of VZ on pedestrian and bicyclist collisions in 18 cities across the United States. The authors used a regression model approach, fitting linear, quadratic, and cubic models to the data, to determine whether a decreasing trend in fatalities was observed. They then employed a T-test to assess the statistical significance of the decreases. The study found statistically significant decreases in total fatalities in only two cities, New York City and Chicago, since implementing a VZ pledge.

After conducting a comprehensive examination of the relevant literature, including the papers previously discussed, we will now unveil our solution. Our objective is to develop a model that can evaluate the engineering and enforcement aspects of the VZ program in NYC. Specifically, we aim to design and implement a predictive model that can predict the likelihood of traffic accidents at specific times and locations. To accomplish this, we will utilize a variety of data sets sourced from NYC Open Data [3]. By leveraging the diversity of these data sets, we aim to create a highly sophisticated model.

| name | start date | end date | samples amount | features amount |
|---|---|---|---|---|
| collisions | 1/7/2012 | present | 1.96M+ | 29 |
| traffic violations | 1/1/2022 | 30/09/2022 | 437K+ | 14 |
| historic violations | 1/1/2018 | 31/12/2021 | 3.08M | 13 |
| open parking and camera violations | 1/1/2000 | present | 91M+ | 19 |
| speed humps | 1/8/1996 | present | 3.5K+ | 6 |
| speed limits | 7/11/2014 | 7/11/2014 | 137K | 6 |
| intersections upgrade | 3/6/2009 | 28/9/2022 | 282 | 8 |
| turn traffic calming | 4/5/2016 | 7/12/2022 | 785 | 7 |
| leading pedestrian signals | 28/6/1978 | 29/11/2022 | 5.8K | 8 |
| traffic lights and signals | 1/1/2010 | present | 1.6M+ | 50 |
| traffic volume counts | 8/1/2012 | 5/9/2021 | 42.8K | 31 |
| points of intrests | 15/3/2008 | 29/12/2022 | 20.5K | 16 |
| street centerline | | | 121K | 30 |

This predictive model can be implemented by several statistical, machine learning (ML), and deep learning (DL) models. However, DL models have a significant advantage when dealing with large datasets and highly complex tasks composed of multiple components with unclear relationships. One additional challenge we face is effectively handling spatial and temporal data and integrating it with 'regular' tabular data. Therefore, we have determined that the DSTGCN model, as demonstrated in the study by Yu et al. [1], presents the most suitable solution for our problem. In their research, they compared several DL and ML models and found that DSTGCN outperformed them all across all examined metrics.

**Table 5**
Comparisons with different methods.

| Models | | *RMSE* | *PCC* | *Precision* | *Recall* | *F1-Score* | *AUC* |
|---|---|---|---|---|---|---|---|
| Classical models | LR | 0.4713 ± 0 | 0.5639 ± 0 | 0.7376 ± 0 | 0.8625 ± 0 | 0.7952 ± 0 | 0.7779 ± 0 |
| | LASSO | 0.4206 ± 0 | 0.5472 ± 0 | 0.7230 ± 0 | 0.8785 ± 0 | 0.7932 ± 0 | 0.7709 ± 0 |
| | SVM | 0.4541 ± 0 | 0.5985 ± 0 | 0.7471 ± 0 | 0.8884 ± 0 | 0.8116 ± 0 | 0.7938 ± 0 |
| | DT | 0.4120 ± 0.0075 | 0.6625 ± 0.0127 | 0.8059 ± 0.0059 | 0.8699 ± 0.0107 | 0.8367 ± 0.0063 | 0.8302 ± 0.0062 |
| State-of- the-art | SdAE | 0.3920 ± 0.0177 | 0.6644 ± 0.0229 | 0.7724 ± 0.0195 | 0.8890 ± 0.0246 | 0.8263 ± 0.0122 | 0.8130 ± 0.0139 |
| | TARPML | 0.3687 ± 0.0078 | 0.7127 ± 0.0120 | 0.8029 ± 0.0132 | 0.8942 ± 0.0134 | 0.8460 ± 0.0044 | 0.8372 ± 0.0059 |
| Our model | DSTGCN | **0.3439 ± 0.0106** | **0.7445 ± 0.0169** | **0.8213 ± 0.0122** | **0.8968 ± 0.0166** | **0.8573 ± 0.0114** | **0.8508 ± 0.0117** |

Our proposed research will bring novel contributions to the field by:

1. Analyzing continuous temporal data combined with before/after data. Considering continuous enforcement and multiple infrastructure changes overtime.

2. Investigating the correlation between enforcement and collisions over time. To the best of our knowledge, there have been no prior studies that have empirically examined the impact of enforcement on traffic accidents

3. Examining the multifactorial effects of infrastructure upgrades implemented as part of the VZ initiative. We will be the first to study the interactions between different infrastructure upgrades in relation to VZ.

To gain a deeper understanding of graph neural networks (GNNs), which are currently unfamiliar to us, we plan to study relevant material through the Stanford CS224W course [4] and the Huji 67912 course [5]. Additionally, we will learn Geo-Pandas [6] to effectively explore spatial data.

References:

[1] Le Yu, Bowen Du, Xiao Hu, Leilei Sun, Liangzhe Han, Weifeng Lv, "Deep spatio-temporal graph convolutional network for traffic accident prediction",Neurocomputing, Volume 423, 2021, Pages 135-147,ISSN 0925-2312

[2] Nicholas N. Ferenchak, U.S. Vision Zero Cities: "Modal fatality trends and strategy effectiveness", Transportation Letters, 2022

[3] New York City Open Data, NYC Office of Technology and Innovation (OTI). City of New York. 2022 All Rights Reserved, https://opendata.cityofnewyork.us/

[4] Stanford CS224W: Machine Learning with Graphs, 2021, https://www.youtube.com/playlist?list=PLoROMvodv4rPLKxIpqhjhPgdQy7imNkDn

[5] HUJI 67912: Advanced Course in Machine Learning, 2023, https://shnaton.huji.ac.il/index.php/NewSyl/67912/2/

[6] GeoPandas, https://geopandas.org/en/stable/getting_started/introduction.html