Utilizing Generative Pre-trained Transformers for Predictive Modeling of Traffic Casualties in Taiwan

An-Syu Li
Department of Computer Science
National Chung Hsing University
yessir0621@gmail.com

Hao-Quan Liu
Department of Computer Science
National Chung Hsing University
contact@haoquan.me

January 12, 2025

Abstract

The increasing trend of traffic fatalities and injuries in Taiwan highlights the urgent need for effective forecasting systems to support policy intervention. In this study, we present a novel approach utilizing the generative pre-trained transformer (GPT) model for predictive modeling of traffic casualties. Traditional methods, including ensemble-based machine learning and recurrent neural networks, often struggle with error accumulation in long-term predictions. By leveraging the capabilities of the GPT architecture, which excels at capturing long-term dependencies, our proposed system addresses these limitations.

Experimental results demonstrate the effectiveness of the GPT model, which consistently outperformed baseline models, including Gradient Boost and Random Forest, across datasets from various cities in Taiwan. Specifically, the GPT model achieved the lowest average Dynamic Time Warping (DTW) score, indicating superior accuracy in predicting traffic casualty trajectories. These findings validate the potential of transformer-based models for traffic injury forecasting and provide a foundation for developing data-driven solutions to enhance road safety.

1. Introduction

Traffic accidents have been a critical issue in Taiwan, with 2023 recording the highest number of traffic-related fatalities since 2013. [1] As shown in Figure 1, over the last decade, the trend exhibits no decreasing or flat patterns, highlighting an urgent need for intervention.

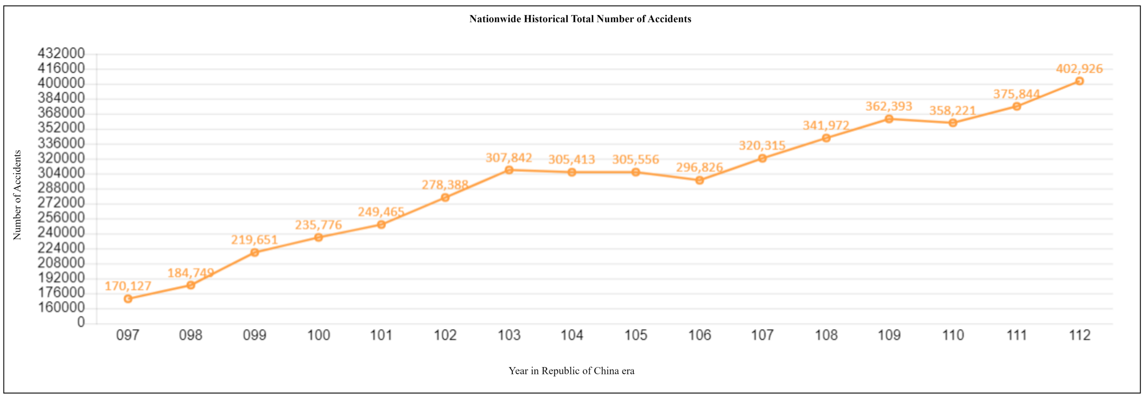

Figure 1. Traffic accident trends in Taiwan based on the Republic of China (ROC) calendar. ROC years are translated as follows: 98 (2009), 99 (2010), ..., 112 (2023). [2]

Therefore, inspired by the statistics, we aim to build a forecasting system that predicts traffic accident trends. This system might provide policymakers or agencies with insights to take action.

In forecasting system modeling, researchers have increasingly turned to traditional machine learning techniques to extract insights and patterns from historical data for making predictions. Previously, scholars relied heavily on statistical models for such systems. However, Jamal et al. [3], recognizing the inherent biases in these models, adopted an ensemble method—XGBoost—which demonstrated superior performance compared to other approaches. Despite their effectiveness, machine learning models exhibit limitations in specific scenarios. Addressing these challenges, Aniekan Essien et al. [4] employed a deep bidirectional LSTM, achieving significantly higher accuracy than baseline models, such as K-Nearest Neighbors and XGBoost, particularly in urban traffic datasets.

Despite these advancements, traffic forecasting presents unique challenges due to its auto-regressive nature, where predictions are based on previously forecasted values. Both LSTM and ensemble methods, while effective for short-term forecasting, suffer from error accumulation as the prediction horizon extends. This compounding of errors diminishes their performance for long-term forecasting tasks, making it challenging to predict unknown future trends accurately. Therefore, we turn to the generative pre-trained transformers (GPT) architecture [5], which excels at capturing long-term dependencies and understanding broader context. While transformer-based models were originally and predominantly used for language-related tasks, researchers are now increasingly exploring their capabilities to address challenges in other fields. The implementation of the system is publicly available on GitHub at https://github.com/nchu-machine-learning/Taiwan-Traffic-Safety-Forecast-System.

2. Methods

2.1. Dataset Description

The dataset utilized in this study was obtained from the Taiwan Government Open Data Platform [6], managed by the National Police Agency, Ministry of the Interior. This open-access dataset provides detailed records of traffic accidents, categorized as A1 (fatal accidents involving deaths within 24 hours) and A2 (accidents resulting in injuries or deaths beyond 24 hours). The data spans multiple years and is available in semi-annual CSV files, allowing researchers to analyze traffic trends across time and regions in Taiwan. Due to lack of information and incomplete features over certain periods, for this study, we specifically focus on the number of injuries aggregated across all available cities in Taiwan between January 2018 and January 2020. This subset of the dataset captures the temporal trends in accident severity during this period. Each record consists of the following key attributes:

- **Time of occurrence:** The specific date and time when the traffic accident occurred.
- **Location of occurrence:** The geographical location or administrative region where the accident took place.
- **Number of fatalities and injuries:** The aggregated count of people who were injured or died as a result of the accident during the selected timeframe.
- **Vehicle types:** The types of vehicles involved in the accident, such as motorcycles, cars, or trucks.
- **Environmental factors:** Additional details such as weather conditions, road type, and presence of traffic signals at the accident site.

As shown in Figure 2, the patterns of fatalities and injuries over the selected period are evident. In practice, however, our analysis will emphasize city-wide trends by segmenting the dataset based on the respective cities and developing individual models for each.

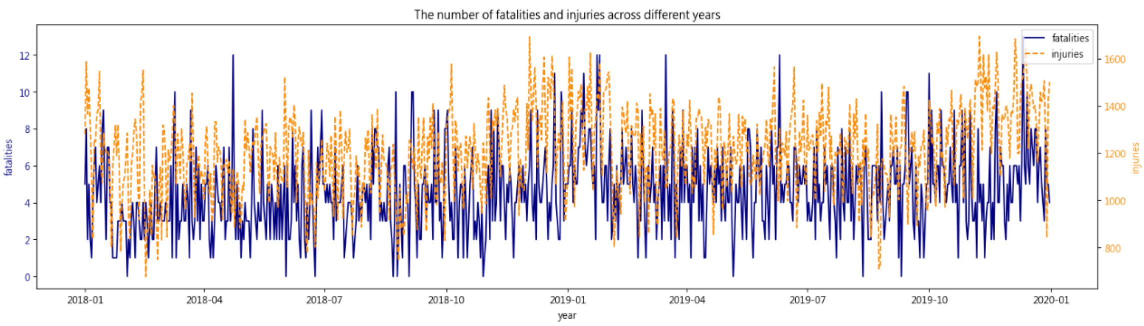

Figure 2. Overview of the traffic fatalities and injuries over the selected period

2.2. Workflow

The workflow includes two major processes: (1) online process and (2) offline process, as shown in Figure 3.

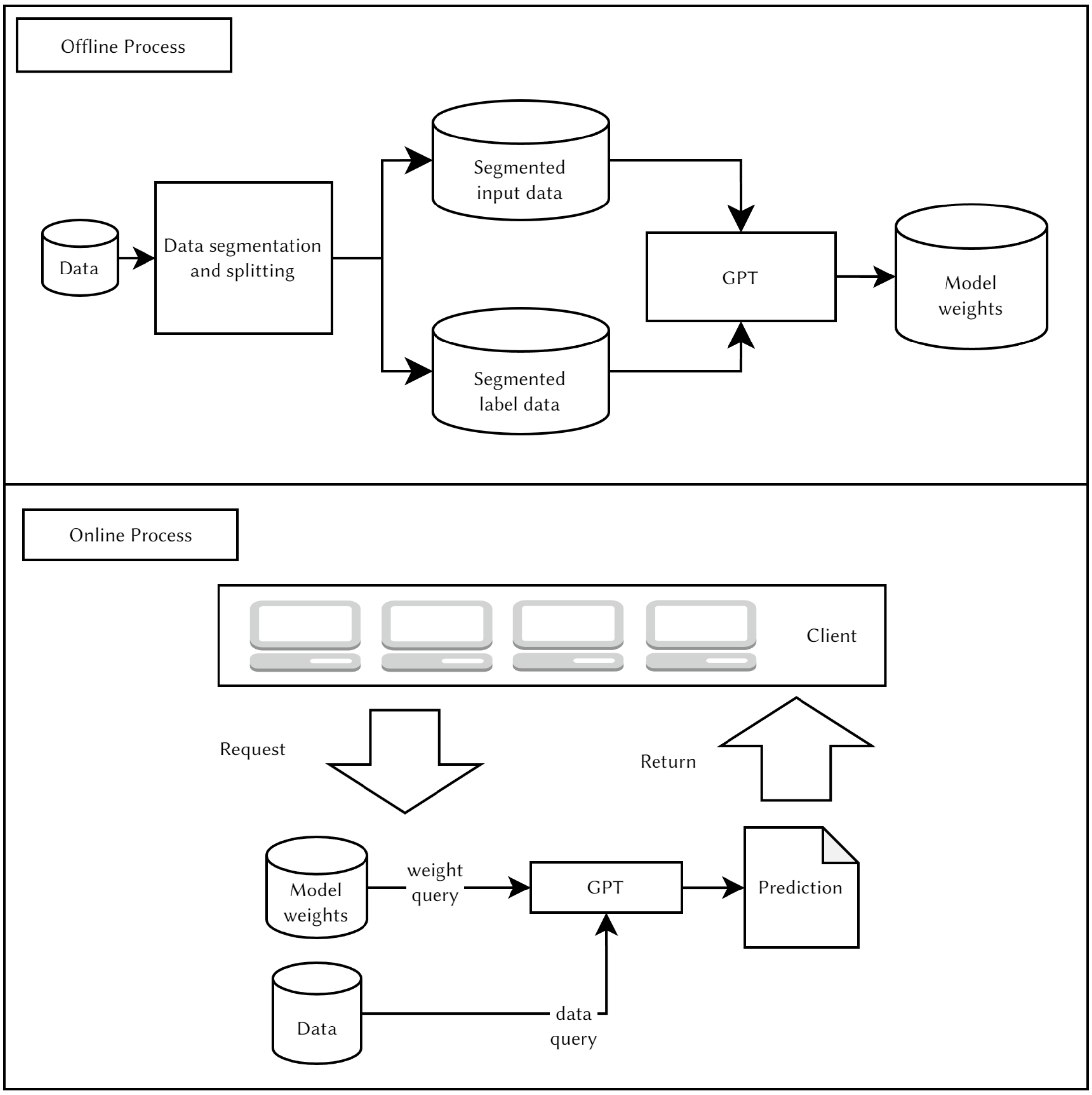

Figure 3. Workflow of construction of the traffic casualty forecasting system

2.2.1 Offline Process

The upper part of 3 exhibits the workflow of the offline process, where the models are trained and data is pre-processed. It consists of two steps: (1) Data segmentation and splitting (2) Model training and storage.

(1) Data segmentation and Splitting. Since we are dealing with data by city, the data is first segmented based on its corresponding city to ensure that patterns unique to each region are preserved. This allows the model to account for city-specific variations in traffic accidents. Additionally, due to the nature of the decoder-only transformer that GPT relies on, the input and output data need to be formatted as sequences, even though the data is fundamentally different from text. To achieve this, we format the data such that each sequence represents a time window of traffic accident statistics, and the corresponding output represents the subsequent time period. For example, if the input sequence is defined as $x = [i, i+1, \ldots, i+n]$, the output sequence is defined as $y = [i+1, \ldots, i+n+1]$, where:

- i represents the starting time step of the sequence (e.g., a specific month or day).

- n is the length of the input sequence (e.g., the number of months or days in the window).
- x represents the input features, such as the number of fatalities and injuries for each time step in the sequence.
- y represents the output features, i.e., the number of fatalities and injuries predicted for the subsequent time steps.

For instance, if $x = [10, 15, 12]$ represents the number of fatalities in three consecutive months, the model learns to predict $y = [15, 12, 18]$, where 18 represents the forecasted number of fatalities in the month following the input sequence. This approach ensures that the data is appropriately formatted for the autoregressive nature of the transformer model, allowing it to effectively learn temporal dependencies in traffic accident trends.

(2) Model Training and Storage. To facilitate quick access while conserving storage space, only parameters—weights and biases—are stored in the hash table for the models trained on the dataset for each city.

2.2.2 Online Process

The lower section of Figure 3 illustrates the system's workflow, which consists of the following operations:

1. **User Access:** Users send a request to the server to access a model trained for the desired city.
2. **Weight Query:** The system retrieves the corresponding model weights from the hash table and loads the model.
3. **Data Query:** To generate predictions, the system fetches data from the designated city to create a prompt suitable for the GPT model.
4. **Prediction and Response:** The GPT model processes the input and produces the predicted results, which are then returned to the client.

This workflow ensures efficient handling of user requests by dynamically loading model parameters and leveraging city-specific data for accurate predictions. The flowchart visually represents these steps for better comprehension.

3. Experiment

3.1. Experiment Setup

To account for both temporal and spatial distances between sequences, dynamic time warping (DTW) is employed instead of relying on Euclidean distance to evaluate model accuracy. DTW minimizes discrepancies across timestamps [7]. The estimation of distances in DTW is explained as follows:

$$D(X, Y) = \min_{\pi} \sum_{(i,j)\in\pi} \mathrm{cost}(x_i, y_j)$$

$$\mathrm{cost}(x_i, y_j) = (x_i - y_j)^2$$

The cumulative cost $d(i, j)$ is computed using the recurrence relation:

$$d(i, j) = \mathrm{cost}(x_i, y_j) + \min \begin{cases} d(i-1, j) & \text{(Insertion)} \\ d(i, j-1) & \text{(Deletion)} \\ d(i-1, j-1) & \text{(Match)} \end{cases}$$

With the boundary conditions:

$$d(0, 0) = 0, \quad d(i, 0) = \infty, \quad d(0, j) = \infty \quad \text{for } i, j > 0.$$

The final DTW distance between sequences $X = [x_1, x_2, \ldots, x_n]$ and $Y = [y_1, y_2, \ldots, y_m]$ is:

$$D(X, Y) = d(n, m),$$

where n and m are the lengths of X and Y, respectively.

3.2. Performance Assessment

To fairly evaluate the performance of the GPT model, baseline models such as Gradient Boost, Random Forest, and AdaBoost are used. The models are compared across various metrics, including accuracy, precision, recall, and F1-score.

As shown in Table 1, the GPT model achieves the best performance among the models based on the DTW metric. However, it is important to note that the high DTW scores across all models are due to the long sequence lengths. Since DTW aggregates the distances over the entire sequence, longer sequences naturally result in higher scores. This highlights the challenges posed by the autoregressive task, where predicting extended sequences amplifies the cumulative differences.

Table 1. Average DTW scores of GPT and baseline models across all available Taiwan's cities

Model	**DTW**
Gradient Boost	135.49
GPT Model	**121.44**
Random Forest	139.74
AdaBoost	134.14

To better understand the differences in model performance across individual cities, we refer to Figure 4. This figure presents the DTW (Dynamic Time Warping) scores for various models and reveals that the GPT's DTW scores are generally lower than baseline models in many cities' datasets.

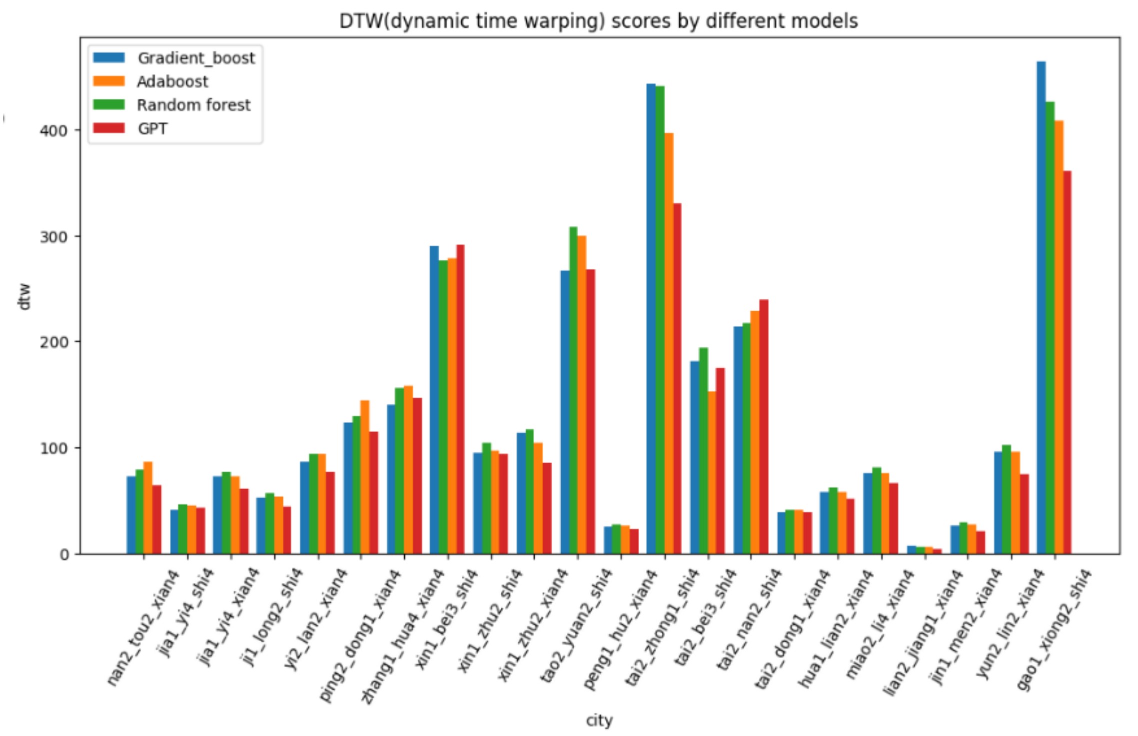

Figure 4. Models' DTW scores from all available cities.

3.3. Case Study

Figure 5 illustrates the predictions of the GPT model alongside those of baseline models compared to the ground truth. The GPT model effectively captures patterns based on historical data, closely aligning its predictions with the ground truth. In contrast, the baseline models exhibit a declining trend in their predictive accuracy over time, as evidenced by the downward trajectories of their prediction curves.

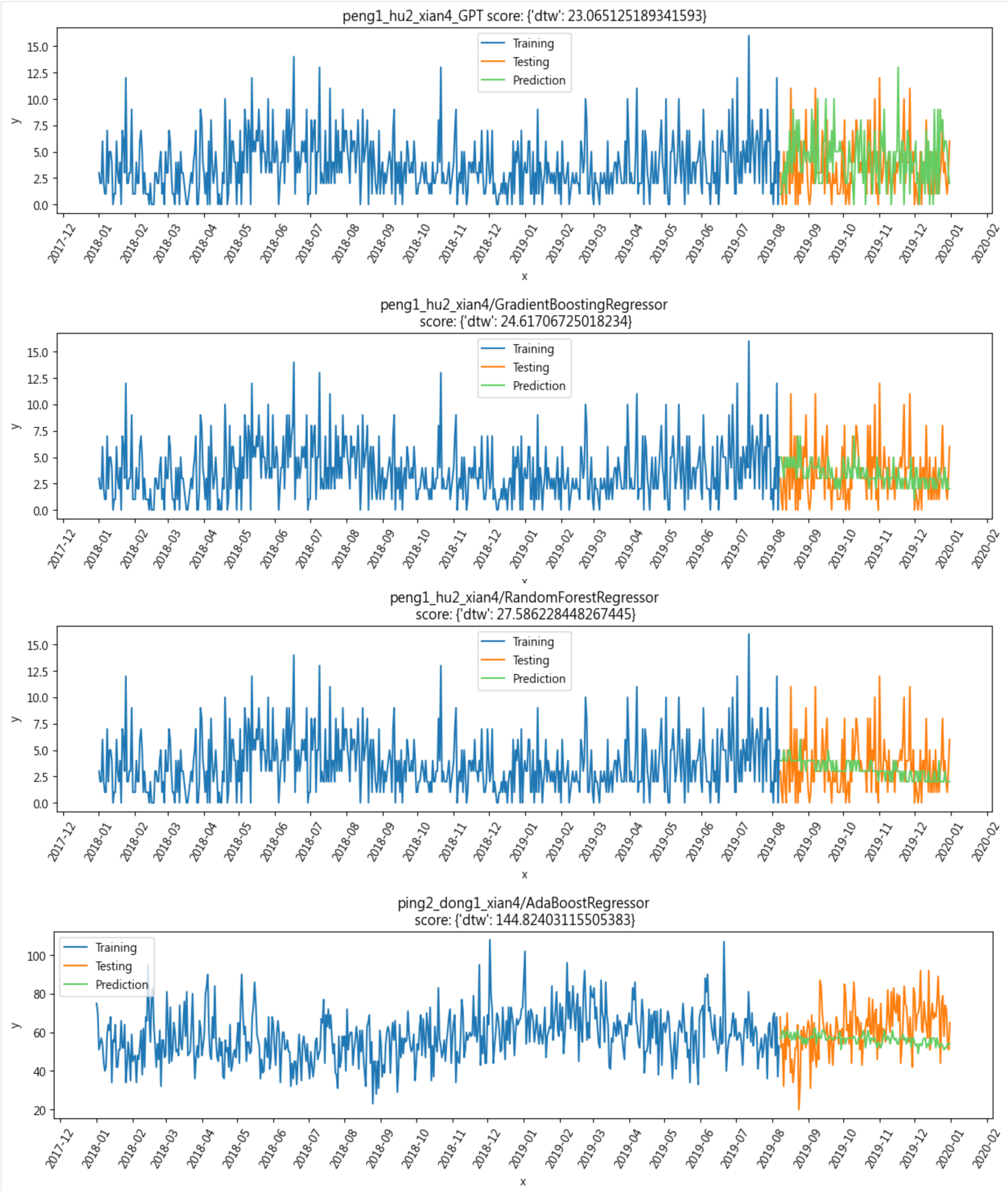

Figure 5. Models' predictions v.s. ground truth from a specific city's dataset

4. Conclusion

With the rising trend of traffic fatalities and injuries in Taiwan, it is crucial for the government to implement effective measures. Motivated by this pressing issue, we developed a forecasting system that accurately captures the trajectories of traffic injuries. While previous research predominantly employed conventional machine learning methods or recurrent neural networks, these approaches often face challenges like accumulative error, which undermines their effectiveness in long-term predictive tasks.

To address this, we were inspired by the generative pre-trained transformer (GPT) model and leveraged its capabilities to design a robust forecasting system. Our experimental results demonstrate that, compared to ensemble-based machine learning models, the GPT model consistently outperforms across datasets from various cities, achieving the lowest average Dynamic Time Warping (DTW) score. These findings validate the effectiveness of using the GPT model for traffic injury forecasting in this context.

References

[1] P. Charlier, “Taiwan traffic deaths in 2022 up year on year and highest since 2013,” *Taiwan English News*, March 25 2023, accessed: January 2025. [Online]. Available: https://taiwanenglishnews.com/taiwan-traffic-deaths-in-2022-up-year-on-year-and-highest-since-2013/

[2] “roadsafety,” Ministry of Transport and Communications, Taiwan, 2020, accessed: Jan. 12, 2025. [Online]. Available: https://roadsafety.tw/Dashboard/Custom?type=%E7%B5%B1%E8%A8%88%E5%BF%AB%E8%A6%BD%E5%9C%96%E8%A1%A8

[3] A. Jamal, M. Zahid, M. T. Rahman, H. M. Al-Ahmadi, M. Almoshaogeh, D. Farooq, and M. Ahmad, “Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study,” *International Journal of Injury Control and Safety Promotion*, vol. 28, no. 4, pp. 408–427, 2021, pMID: 34060410. [Online]. Available: https://doi.org/10.1080/17457300.2021.1928233

[4] A. Essien, I. Petrounias, P. Sampaio, and E. Karapanos, “A deep-learning model for urban traffic flow prediction with traffic events mined from twitter,” *World Wide Web*, vol. 24, no. 5, pp. 1345–1368, 2021. [Online]. Available: https://doi.org/10.1007/s11280-020-00800-3

[5] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” *CoRR*, vol. abs/2005.14165, 2020. [Online]. Available: https://arxiv.org/abs/2005.14165

[6] M. o. t. I. National Police Agency, “Historical traffic accident data,” https://data.gov.tw/dataset/12197, 2015, accessed: Dec. 14, 2024.

[7] S. Chu, E. Keogh, D. Hart, and M. Pazzani, *Iterative Deepening Dynamic Time Warping for Time Series*. Society for Industrial and Applied Mathematics (SIAM), 2002, pp. 195–212. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611972726.12