Predictive analysis of Parking Tickets

Luddy School of Informatics, Computing, and Engineering
Indiana University Bloomington
Bloomington, IN 47405, USA

Rajat Kapgate
Indiana University Bloomington
rkapgate@iu.edu

Usha Jain
Indiana University Bloomington
ushajain@iu.edu

Pranav Sorte
Indiana University Bloomington
psorte@iu.edu

GitHub Repository:
https://github.iu.edu/Luddy-B565-SP24/rkapgate-ushajain-psorte/tree/main/finalproject

Abstract

Our study explores predictive modeling for parking ticket violations in New York City. Parking tickets are a significant source of municipal revenue. The aim is to develop a predictive algorithm that could enhance the efficiency of parking regulation, ensuring consistent revenue generation. We utilized data from over one million parking ticket records spanning several years, and employed decision trees and artificial neural networks (ANN) to predict violations based on various vehicle and incident-related attributes. Our methodology involves comprehensive data pre-processing to reduce class imbalance as well as noise, enhancing the quality of predictions. The decision tree we used provided a high accuracy of 98% but raised concerns about potential overfitting. On the other hand, we implemented ANN, which showed robust performance across various metrics, maintaining high accuracy with better generalization capabilities. The ANN model, which happens to be particularly effective in handling class imbalances and capturing nonlinear relationships, achieved an accuracy of 97.73% on the test set. The study highlights the potential of ML models (specifically Decision Tree and ANN) to support law enforcement agencies in preemptively identifying and managing parking violations. This could lead to optimized patrol routes and potentially influence policy decisions regarding parking regulations.

Keywords: Predictive modeling, Violations, decision tree, ANN, policy.

1 Introduction

New York city issues fines for breaches of various laws and regulations. These include health code violations, building code violations, environmental violations, parking violations and many more. Fines from these violations not only help maintain a safe place for the people to live but also acts as a source of income.

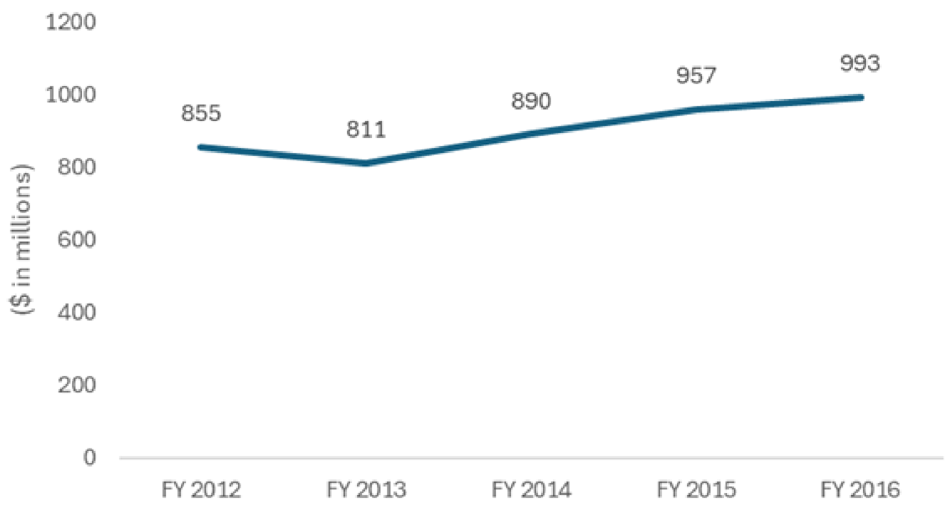


Fig. 1. NYC Revenue from fines

In 2012, 855 million dollars were collected in fines in New York City itself. This has been increasing year on year and in 2016 it went to nearly a billion dollars [1]. (State and local governments collected a combined $13 billion in revenue from fines, fees, and forfeitures in 2021 [2].) Studies show that parking violations remain the most frequent category of fines by a significant margin.

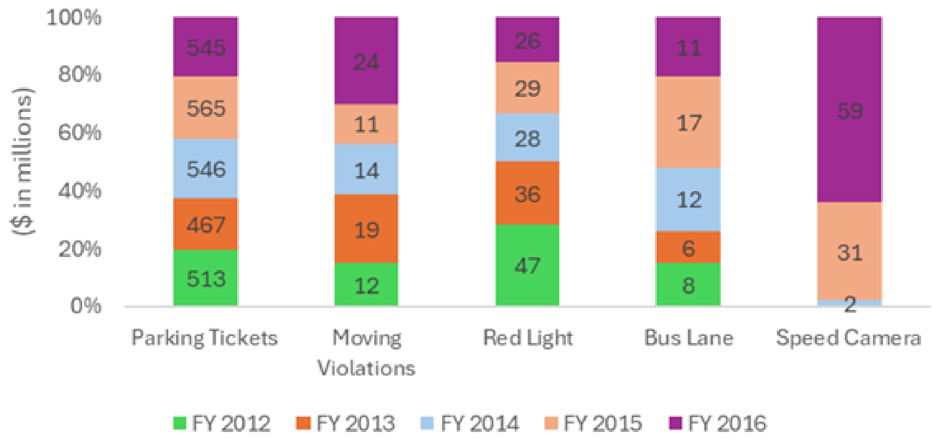


Fig. 2. Revenue from different sources of violation

If you see the trend shown in the image, you'll observe that parking violations contribute to almost 50% of the total city fines per year. These parking fines vary from 35to515, which is dependent on violation codes, areas and frequency of offenses. Despite the city collecting over half a billion dollars annually in parking fines, millions of dollars in fines remain uncollected or are written off each year.

In contrast, revenues from moving violations like speeding, seat belt, and cell phone violations issued by police officers are much lower, with an average of around only 3% of total parking ticket revenues over the past five years. The state retains a significant portion of these revenues to cover adjudication costs.

Now it is crucial to allocate proper resources for collection of the fines to maintain a safe environment in New York City. This project aims to help in that itself - predict parking ticket violation based on statistical data of each violation. Developing a predictive algorithm for this purpose would enable law enforcement agencies to optimise revenue strategies. By doing so, they could maintain consistent revenue levels while potentially reallocating resources towards addressing more serious criminal activities. Additionally, such an algorithm could assist in more efficient deployment of traffic enforcement personnel, optimise patrol routes, and perhaps even influence policy decisions regarding parking regulations and fine structures. It also opens the possibility for the use of real-time data to adapt enforcement strategies dynamically, enhancing the overall effectiveness of law enforcement efforts in traffic management and public safety. Mainly, we'll be testing our strategies with Decision Tree and Artificial Neural Network models for predicting the violation codes

2 Related Work

2.1 Review on Existing Systems

The department of Finance is responsible for collecting and processing the parking ticket. In New York City, parking tickets are traditionally detected by human enforcement officers. In New York City, parking tickets are traditionally detected by human enforcement officers. The officers patrol the streets by vehicle or sometimes on foot, checking parked vehicles for violations. They look for things like cars parked in illegal zones, or exceeding time limits on residential streets and so on.

PARKEY : Ticket-less Parking System Using License Plate Recognition Approach : The goal of the article "PARKEY: Ticket-less parking system using license plate recognition approach" [4] is to enhance the user experience and efficiency of parking systems in shopping malls through the development of a ticket less parking system. In order to solve problems like ticket loss and sluggish distribution times, the planned PARKEY system uses license plate recognition technology to replace paper parking tickets. To expedite the parking entry and leave procedures, it connects an e-wallet based payment system with a web application. Technologies like Laravel for web development and a Raspberry Pi configuration for the license plate recognition hardware were used in the construction of the system. The project's conclusion shows that it was successfully implemented, and by increasing operational effectiveness and reducing paper waste, parking systems may be modernized.

Automatic Parking Management System and Parking Fee Collection Based on Number Plate Recognition : The paper discusses the development of an automatic parking system that utilizes vehicle license plate recognition to facilitate parking management and fee collection[5]. The technology aims to enhance convenience and security in public parking areas by reducing human interaction. The system incorporates image processing techniques for recognizing number plates, and integrates these capabilities with electronic fee collection. By using algorithms for plate detection, character segmentation, and optical character recognition (OCR), the system efficiently manages vehicle entry and exit, as well as fee calculation, with minimal human oversight. The results showed the system's effectiveness in automating parking processes and improving operational efficiency.

Smart and efficient system for the detection of wrong cars parking : The paper Smart and efficient system for the detection of wrong cars parking [6], is designed to detect incorrect car parking by using a mobile monitoring setup. It utilises a combination of two cameras, DC motors and PIR sensor and Arduino microcontrollers. The primary function of this setup is to monitor parking spaces and identify any cars that are parked incorrectly within marked boundaries. The technology hinges on image processing techniques that process captured images of parked cars to determine their alignment within the parking spots. If a car is found to be parked incorrectly, the system records the licence plate and stores this information for potential notification or penalization purposes. This smart parking solution aims to enhance the efficiency of parking space management, reduce the need for human monitoring, and ensure better compliance with parking regulations.

We are looking to develop a system tailored for enforcement officers to proactively identify potential parking violations. Unlike existing systems which issue e-tickets or help drivers detect their own parking mistakes, our system will utilise details like vehicle numbers and city information. This will enable officers to anticipate violations before they occur, enhancing the efficiency of parking management and enforcement. This proactive approach aims to streamline the process of parking regulation and ensure compliance more effectively.

3 Methodology

The proposed method is a centralized system for the officers to identify the violations. It consists of the following steps:

3.1 Data Collection :

The datasets used in the research were obtained from openly accessible data on data.gov [7]. The data that is being used consist of 10,44,433 with 33 attributes, providing a comprehensive basis for analysis.

3.2 Data Pre-processing :

While analyzing the project data, we identified 120 distinct colors in the vehicle dataset, which we streamlined to six primary colors for more focused analysis. Additionally, we conducted binning on vehicle purchase year and the duration of the day to facilitate more accurate analysis. Here are the trends we observed:

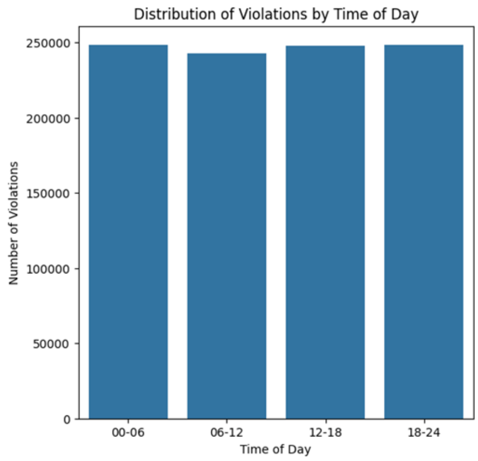


Fig. 3. Distribution of Violations by Time of the Date

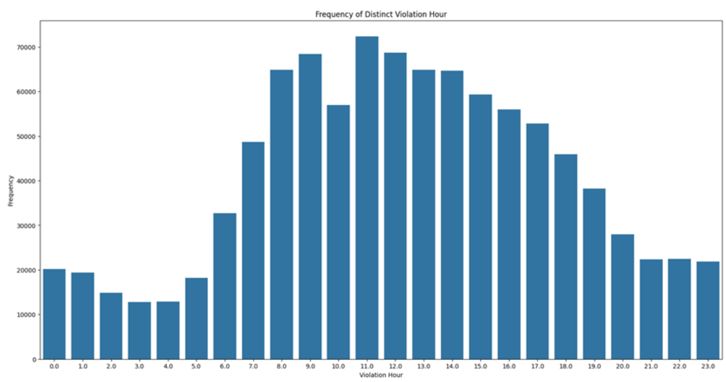


Fig. 4. Frequency of violation taking place every hour

Now turning our attention to the violation code, the dataset originally contained 100 classes. However, 5% of the data was noise as this 5% of data was contributing to around 40 out of 100 classes. So, we filtered out dataset of 10,44,433 entries to include only the data consisting the leftover 60 classes, giving us around 990k entries.

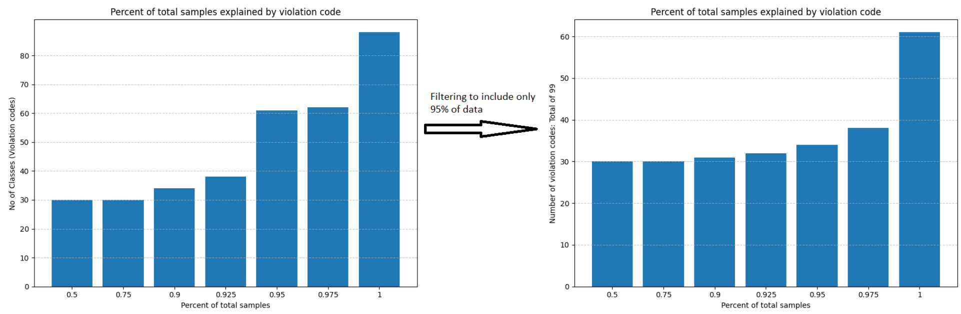


Fig. 5. Filtering based on violation code

We then executed feature thresholding as well to solve the class imbalancing problem. We eliminated rows where the frequency of feature values fell below the 90th percentile of counts within the column. This curation process aimed to mitigate the impact of less common feature values that could potentially introduce noise into our model's predictions.

Furthermore, we adopted a selective strategy for dropping values, focusing primarily on columns with fewer pairwise correlation coefficient (PCC) values. By prioritizing columns with higher correlation coefficients, we aimed to keep and retain the most informative features while minimizing redundancy in our dataset.

Moreover, we were cautious when considering the removal of columns containing a sparse distribution of unique values, such as 'VehicleYearBinned' and 'VehicleColor'. Despite their lower correlation coefficients, we recognized the potential value of these features and opted to retain them in our dataset. This decision was motivated by the desire to preserve as much relevant information as possible, thus enhancing the robustness and effectiveness of our ML model.

After this data reduction, our dataset was streamlined to 729,823 entries.

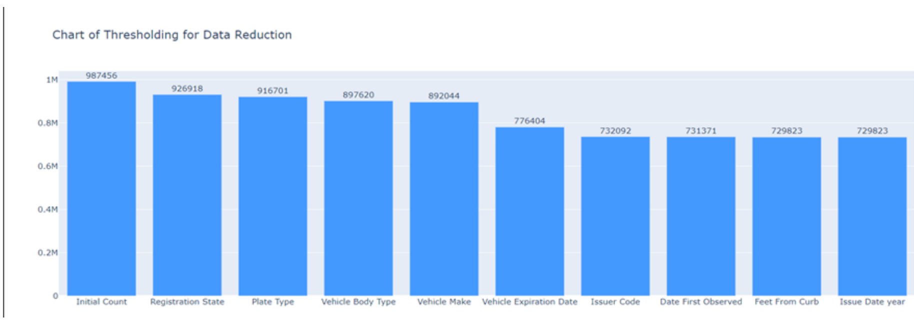


Fig. 6. Performing feature thresholding for data reduction

3.3 Correlations :

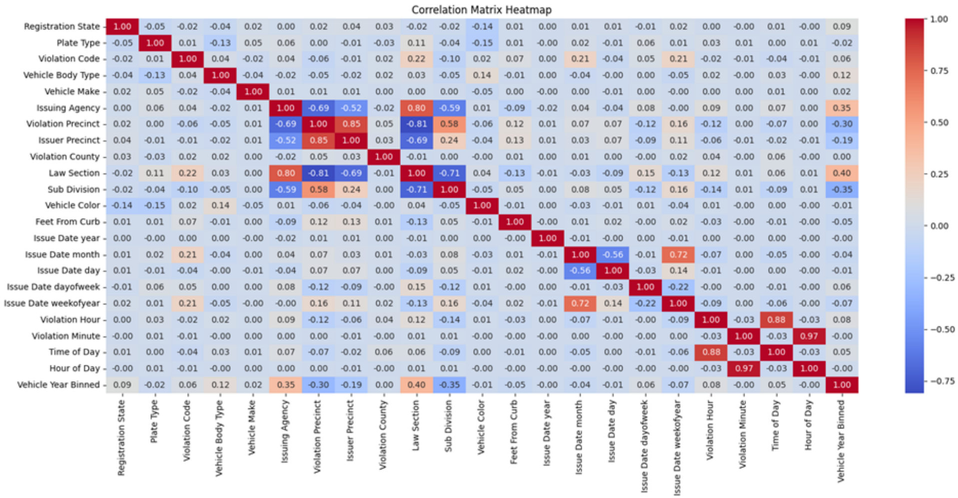


Fig. 7. Pearson Correlation Coefficient

We observe a strong correlation between the violation code and the law section, it is because the law section is derived from the violation code itself. We thus have to discard the law section from the attributes. Similarly, Issuer precinct, subdivision, and issuing agency tend to be derived from other attributes that may not be relevant in the context of violation code prediction.

3.4 Machine Learning Model

Decision Tree

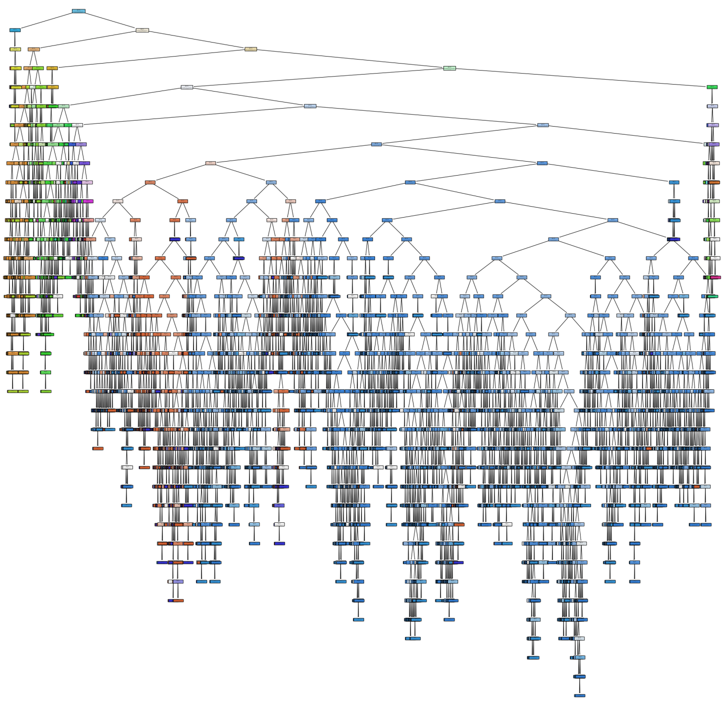


Fig. 8. Visualization of the decision tree with default hyperparamters

We use the decision tree algorithm over 29 input features which are as follows 'Registration State', 'Plate Type', 'Violation Code', 'Vehicle Body Type', 'Vehicle Make', 'Issuing Agency', 'Street Code1', 'Street Code2', 'Street Code3', 'Vehicle Expiration Date', 'Violation Precinct', 'Issuer Precinct', 'Issuer Code', 'Violation County', 'Street Name', 'Date First Observed', 'Law Section', 'Sub Division', 'Vehicle Color', 'Vehicle Year', 'Feet From Curb', 'Issue Date year', 'Issue Date month', 'Issue Date day', 'Issue Date dayofweek', 'Issue Date weekofyear', 'Violation Hour','Violation Minute', 'Time of Day', 'Hour of Day','Vehicle Year Binned'. Decision tree yields an accuracy of 99.48% on the test set, but this comes with a risk of overfitting the train data. The classification report however, shows a fair balance for precision and recall scores for underrepresented samples

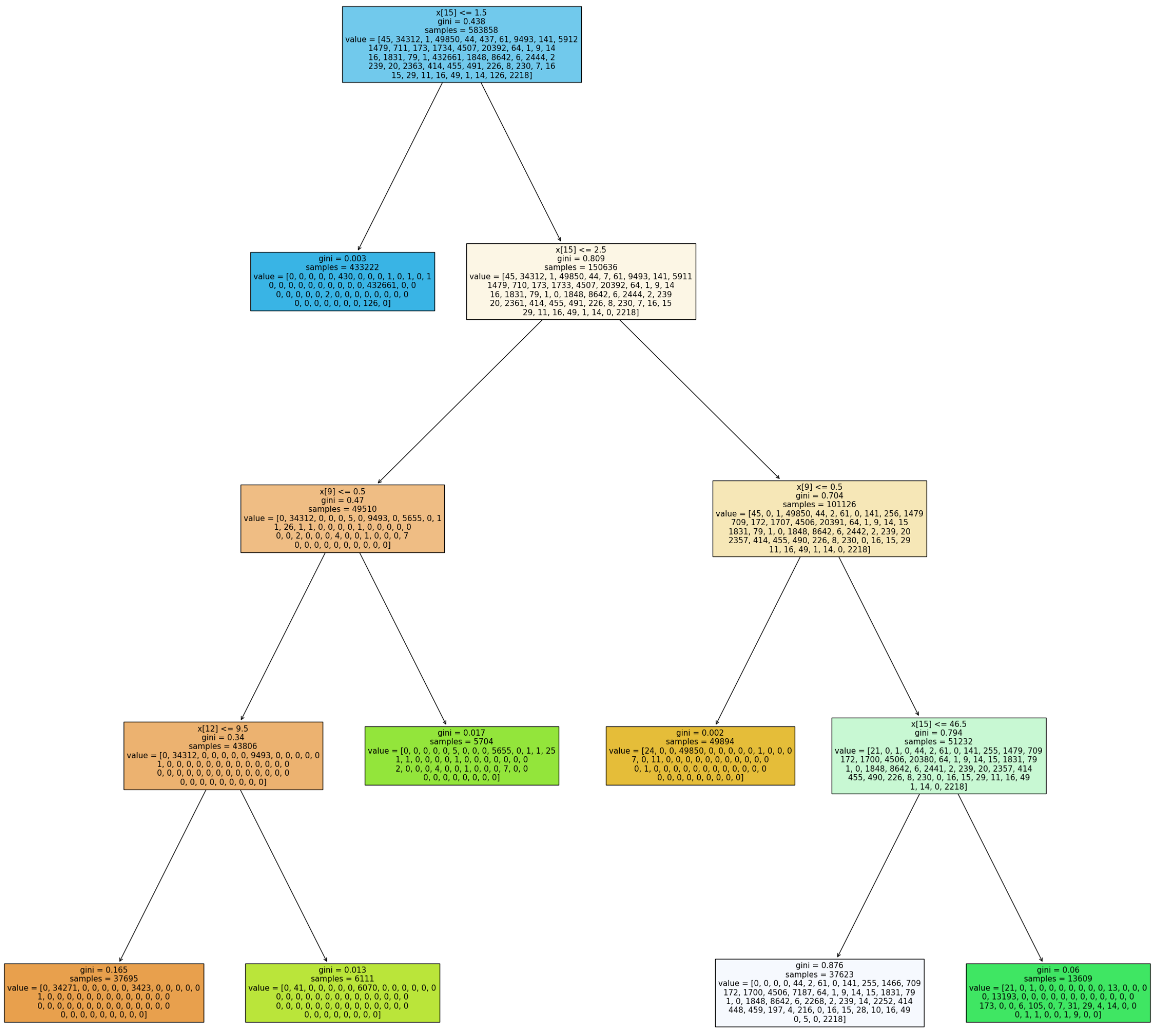


Fig. 9. X Simplified version of the decision tree with ccp_alpha hyperparameter set to 0.01. This creates a pruning effect on the decision tree

Decision trees offer an ease of understanding the underlying trends in the data and work well with a low number of samples. The decision tree uses the CART algorithm which recursively partitions the data into subsets using two splits based on thresholds to reduce the gini impurity for classification.

However, in our case, we have more than 750,000 samples. This offers the opportunity to try more sophisticated algorithms like artificial neural networks.

Artificial Neural Networks : To make better use of the large amount of data we have at hand we experimented with an Artificial Neural Network, this network consists of an input layer, 2 hidden layers and an output layer with softmax function. The neural network takes label encoded values of 29 features as the input and outputs and vector with length 60 for the 60 classes with probability of each class.

We use the he initialization for weights of the layers of the neural network, this helps to mitigate the vanishing/exploding gradient problem with neural networks and also helps achieve quicker generalization of the model.

The detailed architecture of the neural network is as follows:

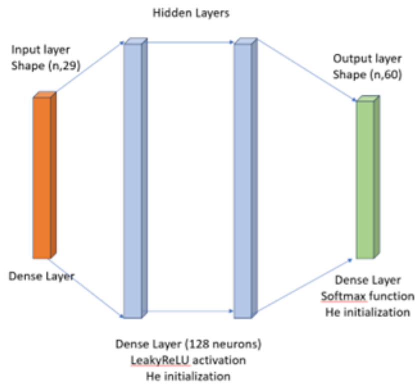


Fig. 10. Visualization of the decision tree with default hyperparamters

4 Results:

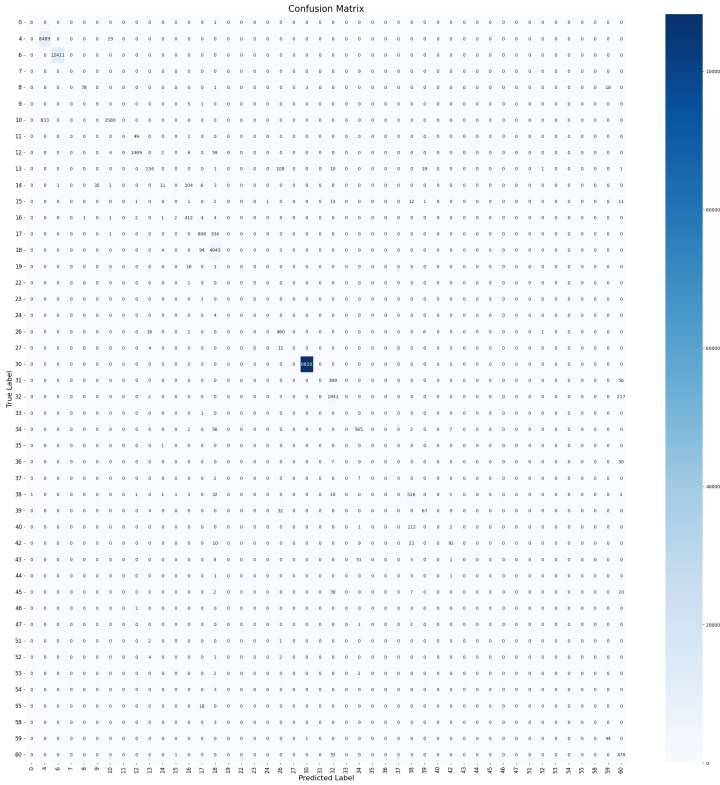


Fig. 11. Confusion matrix for decision tree

From the confusion matrix, it is shown that while some classes are predicted well like violation code 30. However, there are some violation codes that display misclassifications, indicated by non-diagonal counts. This is particularly clear for classes with fewer instances where the model seems to struggle more noticeably.

	precision	recall	f1-score	support
0	0.89	0.89	0.89	9
4	0.91	1.00	0.95	8518
6	1.00	1.00	1.00	12411
7	0.00	0.00	0.00	9
8	0.99	0.78	0.87	100
9	0.20	0.60	0.31	15
10	0.98	0.65	0.78	2413
11	0.00	0.00	0.00	50
12	0.96	0.95	0.96	1540
13	0.88	0.63	0.73	374
14	0.55	0.07	0.12	161
15	0.00	0.00	0.00	42
16	0.75	0.96	0.84	427
17	0.86	0.70	0.77	1150
18	0.90	0.98	0.94	5044
19	0.00	0.00	0.00	17
22	0.00	0.00	0.00	1
23	0.00	0.00	0.00	3
24	0.00	0.00	0.00	4
26	0.74	0.95	0.83	484
27	0.00	0.00	0.00	15
30	1.00	1.00	1.00	108207
31	0.00	0.00	0.00	455
32	0.79	0.90	0.84	2163
33	0.00	0.00	0.00	1
34	0.88	0.90	0.89	631
35	0.00	0.00	0.00	1
36	0.00	0.00	0.00	57
37	0.00	0.00	0.00	9
38	0.76	0.90	0.83	572
39	0.72	0.65	0.68	103
40	0.00	0.00	0.00	115
42	0.85	0.70	0.77	132
43	0.00	0.00	0.00	63
44	0.00	0.00	0.00	2
45	0.00	0.00	0.00	68
46	0.00	0.00	0.00	1
47	0.00	0.00	0.00	3
51	0.00	0.00	0.00	3
52	0.00	0.00	0.00	7
53	0.00	0.00	0.00	4
54	0.00	0.00	0.00	3
55	0.00	0.00	0.00	18
58	0.00	0.00	0.00	3
59	0.71	0.98	0.82	45
60	0.57	0.93	0.71	512
accuracy			0.98	145965
macro avg	0.37	0.37	0.36	145965
weighted avg	0.97	0.98	0.97	145965

Fig. 12. Classification report for Decision Tree

The model produces an overall accuracy of 98% and performs well across all classes. However, the macro averages are on the lower side suggesting that the model may struggle to make predictions over classes that have fewer instances in the train set.

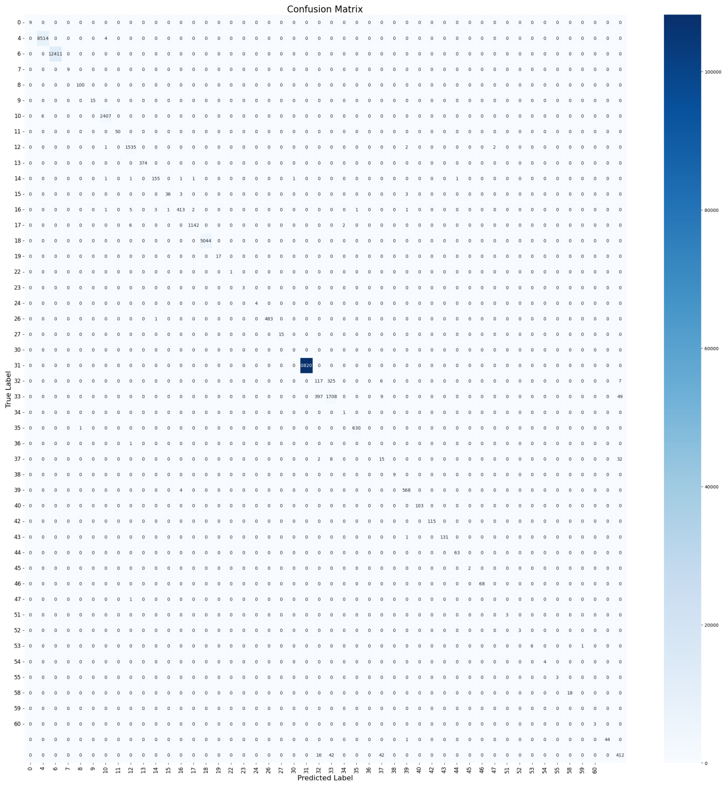


Fig. 13. Confusion Matrix for Artificial Neural Network

Comparing the confusion matrix to the Decision Tree, the Artificial Neural Network produces more uniform prediction across all classes. This is mainly due to ANN's effective ability to capture nonlinear relationships and make robust generalizations on the train data.

	precision	recall	f1-score	support
0	1.00	0.50	0.67	10
4	0.92	1.00	0.96	8682
6	1.00	1.00	1.00	12325
7	0.00	0.00	0.00	12
8	1.00	0.89	0.94	105
9	0.00	0.00	0.00	18
10	0.95	0.67	0.79	2316
11	0.00	0.00	0.00	40
12	0.97	0.93	0.95	1527
13	0.83	0.35	0.49	338
14	0.59	0.40	0.48	169
15	0.00	0.00	0.00	49
16	0.77	0.80	0.78	464
17	0.80	0.78	0.79	1148
18	0.92	0.96	0.94	5142
19	0.14	0.05	0.08	19
22	0.00	0.00	0.00	2
23	0.00	0.00	0.00	3
24	0.00	0.00	0.00	5
26	0.75	0.94	0.83	436
27	0.00	0.00	0.00	19
29	0.00	0.00	0.00	1
30	1.00	1.00	1.00	108135
31	0.00	0.00	0.00	470
32	0.75	0.90	0.82	2157
33	0.00	0.00	0.00	1
34	0.79	0.91	0.85	660
36	0.00	0.00	0.00	56
37	0.00	0.00	0.00	4
38	0.66	0.95	0.78	582
39	0.64	0.70	0.67	97
40	0.00	0.00	0.00	126
42	0.50	0.01	0.02	123
43	0.00	0.00	0.00	61
45	0.00	0.00	0.00	58
47	0.00	0.00	0.00	1
51	0.00	0.00	0.00	2
52	0.00	0.00	0.00	6
54	0.00	0.00	0.00	5
55	0.00	0.00	0.00	13
58	0.00	0.00	0.00	4
59	0.97	0.97	0.97	40
60	0.60	0.83	0.70	534
accuracy			0.98	145965
macro avg	0.39	0.36	0.36	145965
weighted avg	0.97	0.98	0.97	145965

Fig. 14. Classification report for Artificial neural network

The classification report further strengthens the narrative that the ANN is more effective than decision trees in handling the accuracy issues due to the class imbalance.

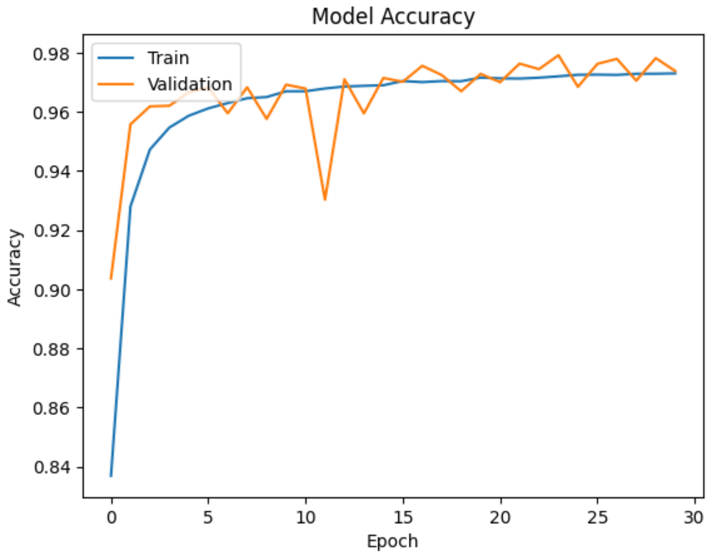


Fig. 15. Accuracy curve for ANN

The Accuracy Curve demonstrates high and stable accuracy, with the validation curve closely aligning with the training curve after initial epochs. This suggests good generalization without significant overfitting.

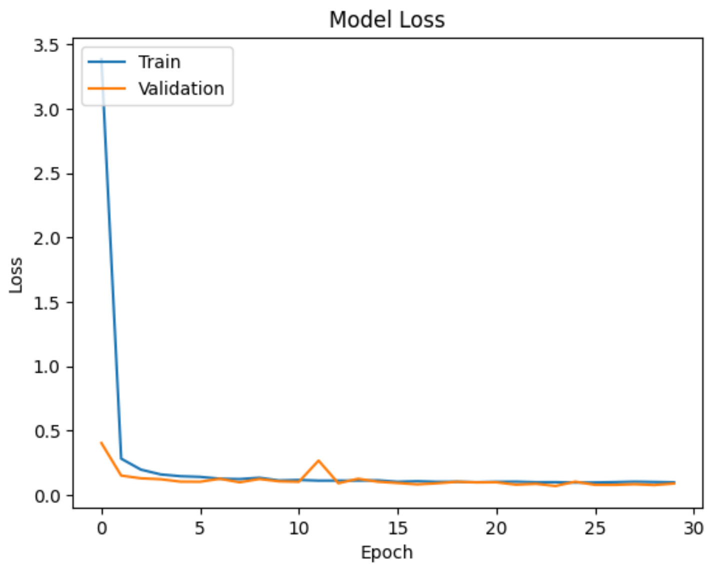
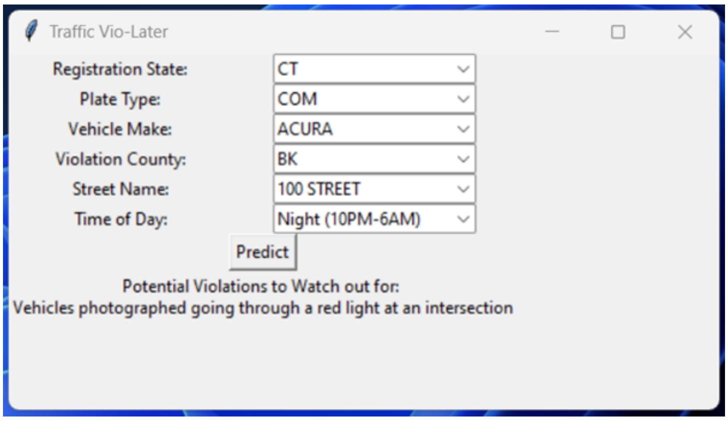


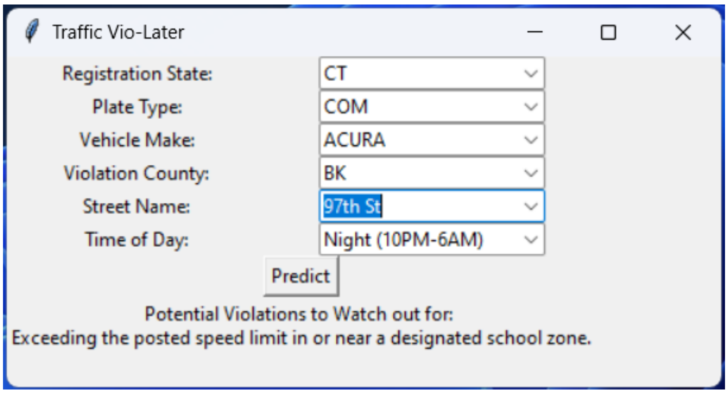
Fig. 16. Loss curve for ANN

The loss curve indicates a healthy learning process with the training loss decreasing sharply and stabilizing. The overlap of validation loss to training loss further proves the model's ability to generalize well without overfitting

The ANN yields a result of 97.16% on train set, 97.56% on validation set and 97.73% on the test set.

We also implement an application using tkinter interface in python to demonstrate a possible real world application of our study.





User can select from different attributes, in this case we choose from 6 notable attributes like the registration state, type of the plate, make of the vehicle, county where the violation information is required, street name and the time of the day. Depending on this input parameters, law enforcement can be warned of potential violations to watch out for during the time of the day at that particular place.

5 Conclusion :

The findings of this project and research into traffic violation systems reveal a substantial potential for law enforcement to better ensure safety and for general public awareness. While the dataset showed a skewness in the class distribution, further work can be done to mitigate this challenge by enriching the dataset with more samples from underrepresented classes.

Furthermore, the applications of this project can be expanded to several scenarios:

- Traffic violation prediction systems can be integrated into broader smart city projects to enhance traffic management and urban planning.
- Using real-time data, these systems could alert law enforcement to potential violations before they occur, allowing for preventative measures.
- With the rise of connected and autonomous vehicles, traffic violation prediction systems could directly interface with vehicle systems to warn drivers of potential violations or even take control of the vehicle to prevent a violation.

References

1. https://comptroller.nyc.gov/reports/new-york-city-fine-revenues-update/
2. All revenue data is from the Office of the City of New York Comptroller, Comprehensive Annual Financial Report and the New York City Financial Management System. Additional data comes from the New York City Mayor's Office of Operations, Mayor's Management Report; New York City Office of Administrative Trials and Hearings, Environmental Control Board Tribunal Data.
3. New York Department of Finance: Violation Codes Listing
   https://www.nyc.gov/site/finance/vehicles/services-violation-codes.page
4. M A M B Kamaruzaman1 and Naziffa Raha Md Nasir
   PARKEY : ticket-less parking system using license plate recognition approach
   https://iopscience.iop.org/article/10.1088/1742-6596/1860/1/012006/pdf
5. M. M. Rashid, A. Musa, M. Ataur Rahman, and N. Farahana, A. Farhana
   Automatic Parking Management System and Parking Fee Collection Based on Number Plate Recognition
6. A. H. M. Almawgani, Turki Alsuwian, Adam R. H. Alhawari, A. N. Alhuthari, M. A.Alhezabr, M. S. Alharethi, F. H. Alqahtani
   Smart and efficient system for the detection of wrong cars parking
   https://beei.org/index.php/EEI/article/view/2634/2242
7. Parking Violations Issued - Fiscal Year 2024
   https://catalog.data.gov/dataset/parking-violations-issued-fiscal-year-2023
8. https://www.kaggle.com/code/donyoe/exploring-42-3m-nyc-parking-tickets
9. ACKERMAN, Samuel S., and Rida E. MOUSTAFA - "Red Zone, Blue Zone: Discovering Parking Ticket Trends in New York City." (2011).