Comparing a Multinomial Logit Model and Machine Learning Models for Travel Mode Choice for Work Trips. A Long Island Case Study.

Robert V. Checco

Abstract

Transportation systems are what bring the world together, by allowing for the efficient mobility of people to and from their place of employment. These systems are composed of many alternative modes of travel. There is value to urban planners, economists, and transportation engineers in understanding the factors that lead individuals to choose a particular mode of transport. The analysis in this paper uses household-level cross-sectional survey data from suburban New York areas to better understand transportation habits of the population. I used a multinomial logit model to better understand how these attributes can explain the likelihood of choosing a particular mode. Also, I contrast the findings with machine learning algorithms that will attempt to correctly understand the pattern recognition of mode choice. The results show similar findings to the current literature that machine learning algorithms are superior in prediction when compared to the conventional logit model. While the machine learning algorithms lack the interpretability of the logit models, they do offer an opportunity to correctly classify mode prediction without having to specify a model beforehand.

Keywords: mode choice, neural networks, support vector machines, multinomial logit model, transportation demand, econometrics, discrete choice.

Introduction

When choosing to commute to work, individuals are presented with many options for modes of travel. The focus of this paper is to understand what factors influence individuals to select a particular mode of travel. In order to gain insight on mode choice of commuters, a mathematical model will be used to estimate the probability of choosing one mode over a set of alternatives. The model will reflect the commuting behavior of individuals living in the suburban landscape of Nassau and Suffolk Counties of Long Island, New York. The scope of this research will strictly examine trips to and from work. The data used for the analysis was extracted from the Regional Household Travel Survey (RHTS) in 2010/2011. This paper will use a discrete choice model based on random utility maximization in order to predict mode choice. Among these models is the multinomial logit (MNL) model, first formulated by McFadden (1973), which has dominated transportation literature for mode choice. Recent research has compared conventional econometric models to data mining techniques using machine learning algorithms. This paper will follow the work of Xie (2003) by comparing the predictive performance of the conventional Multinomial Logit Model (MNL) to machine learning methods. The machine learning methods that will be used are Support Vector Machines and Artificial Neural Networks.

Transportation systems, including highways, streets, and roads are among the largest assets owned by developed nations (Moavenzadeh & Markow 2010). The systems are also comprised of other modal networks such as rail, bus, and other public transportation options. The United States (US) government allocates upward of 50 billion dollars annually to federal agencies responsible for managing the country's transportation networks. Well funded transportation infrastructure reduces the cost of moving labor, capital, and inputs across land.

However, investment in infrastructure is not flexible and a large planning stage is required. As a result, large scale infrastructure projects can take a long time from conception to completion. With such extensive timelines, urban planners are required to be able to accurately estimate demand for new projects.

Mode choice modeling gives valuable information to urban planners and transportation engineers. Both the private and public sectors have an interest in understanding which factors lead to a mode choice decision. For ride-hailing firms such as Uber and Lyft, an accurate model will be able to tell them how demand for their service may change as socioeconomic and transportation related attributes change. For public investment in infrastructure projects, governments require models that allow them to estimate which mode individuals will take as travel time, income, and other related variables change.

Literature Review

Prior to the early 1970s most researchers used gravity models to quantify transportation demand. A gravity model was used to estimate aggregate demand between two trip endpoints. Gravity models used aggregate data, such as travel time, to derive a cost function for commuters. The researchers at the time believed that longer trips would be more costly in terms of time and resources, therefore, would result in lower demand. These models had limited feasibility in the space of mode choice estimation (Evans 1973).

A major breakthrough was made by McFadden (1973) where he constructed the conditional logit model that allowed the dependent variable to take on discrete values. In the past, researchers struggled to model decision making, because continuous dependent variables allowed for infinite outcomes. By constraining the dependent variable to be discrete, researchers were able to model choice decisions that were categorically exclusive. This type of model was popularized in the Bay Area Rapid Transportation Project where researchers using gravity models predicted that around twelve percent of commuters would use the rail, where McFadden predicted that six percent of commuters would use the rail line. The actual observations showed that the percentage of ridership was very close to McFadden's estimate. This permitted the use of disaggregate data to be used for mode choice analysis. Since then researchers have derived other models from the model McFadden used.

Socio Demographics such as age and gender have been shown to influence transportation mode choice behavior. Kim & Ulfarsson (2004) found that when distance is controlled for, the elderly are much more likely to utilize public transportation than other age cohorts. While the elderly are a small percentage of the labor force, there are elderly people that do commute to work. This provides justification in including an age variable into the model. Hasnine et al. (2018) used campus based travel surveys to understand the travel behavior of college students in Toronto. The Toronto University has a robust dataset because the campus is divided between suburban and downtown campuses. Their findings were that female students who traveled to the downtown campus were more likely to use active transportation options and public transit. They also found that female students who traveled to the suburban campus were more likely to travel by Single Occupancy Vehicle (SOV). Since destination location influences mode choice behavior, this paper elects to include a destination variable because Long Island commuters are split two major locations. The split is between those who commute to work located in a suburban setting and those who commute into an urban setting such as Manhattan.

A main focus of this paper is to attempt to replicate the finds of Xie et al. (2003) where they compared the MNL to different data mining methods. Xie et al. used the Bay Area Transportation Survey to model individual mode choice for the commute to work. Their findings were that in terms of prediction, the machine learnings methods outperformed the conventional MNL. While the behavior of commuters in New York could be different than the behavior of commuters in the San Francisco Bay Area, this research will provide value to the field by doing another comparison for conventional vs machine learning mode prediction. The Long Island area is heavily suburbanized and this research can be used by others who want to do mode choice analysis in other suburban settings. It has been fifteen years since the Xie et al. paper and I have access to more machine learning resources that will allow me to more easily customize my models. This paper's use of the RHTS is significant because there is new data available and there may have been a shift in the commuting behavior. Therefore, this will provide a more recent snapshot of the comparison of the classification techniques. As far as I am aware, this paper is the first paper to do machine learning mode choice prediction using the RHTS. My results will allow researchers in the future to better understand the role of machine learning models for mode choice prediction.

Empirical Analysis

Data

The data used for this analysis was obtained from the 2010/2011 (RHTS). The survey was sponsored by two of the federally sanctioned Metropolitan Planning Organizations (MPOs) in the New York/New Jersey/Connecticut metropolitan area. The two MPOs are the New York

Metropolitan Transportation Council (NYMTC) and the North Jersey Transportation Planning Authority (NJTPA). The survey was conducted from September 2010 through November 2011 by NuStats based in Austin, Texas. NuStats is a full service survey research consultancy specialized in large-scale social research studies. At various stages of data collection NuStats was assisted by GeoStats and Parsons Brinckerhoff.

The data was collected to be used for transportation planning, research, analysis and policy. The predecessor to the RHTS was the 1997/98 RT- HIS. At the time of the RHTS, the previous travel survey was over 12 years old and contained information that no longer reflected the travel profile of the region. The data collection method for the RHTS relied on the willingness of area residents to complete diary records of their daily travel over a 24-hour period. To generate a random sample of households, a phone interview was conducted to inform the household of the survey, its purpose and the requirement to fulfill a travel diary. The data diary was created from a single day, involves observations of many households, and examines many different characteristics from each household. This data is household level cross sectional data. In total 31,156 households were recruited to participate in the survey and of these, 18,965 households completed travel diaries. All households were located in the 28-counties that constitute the New York/New Jersey/Connecticut metropolitan area.

Dependent Variables

Individuals choose a mode of transportation to travel to and from work. While some linked trips consist of multiple modes, this paper will focus on the primary mode of travel. The primary mode of travel has been used by Xie (2003) and is defined as the mode that is used for the majority of the trip. The variable is not computed, and was offered as a variable directly from the

RHTS. As suggested by Ben-Akiva et al. (1985) I combined physically different but substitutable modes of transport into categories within the choice set. The model will not be able to predict whether the individual chooses a car over a motorcycle, but will return the category of the mode of transportation. There are five categories of mode choice that individuals will choose from regarding their travel to and from work. These categories reflect the five most popular mode choices commuters utilize on Long Island.

The first category is (SOV) and it consists of cars and motorcycles. I suggest that cars and motorcycles are substitutable and belong to the same category for mode option. A motorcycle can be thought of as a different style of a personal car because they both serve the same purpose. In New York State motorcycles are required to be registered with the State Department of Motor Vehicles and must be insured. Like cars, motorcycles operate on fossil fuels and require similar levels of maintenance. SOV is the most popular mode category for work related trips of Long Island residents.

The second category is the rail transportation mode. This is composed of the Long Island Railroad (LIRR) and New York City (NYC) subway. The LIRR is a commuter rail system that operates trains that terminate in NYC terminals. Many residents who commute into NYC utilize this form of public transit to travel to work. This is primarily because of the high levels of congestion present during rush hour and the lack of available on-street parking in NYC. Although the subway is not accessible from Long Island, some residents in Western Nassau County may choose a mode of transportation to get to the nearby subway, and then utilize the subway as the primary mode of travel. Riders who elect to travel this way would view the subway and LIRR as substitutes.

The third category is the auto-passenger category. This category is composed of those who use taxis, carpools, or are non-driving auto passengers. I suggest they are all members of a similar class of commuters because they rely on getting a ride from someone else. If a commuter is a passenger often, the second best alternative is a mode within the same class. Substituting modes within the class is based on habit and comfort, as it is time-consuming to research into learning a rail or bus schedules for only a day. Getting a ride from a person, means that the driver incurs a cost of driving the commuter. This cost will limit the distance that they are willing to drive the commuter, because the driver will have to make a solo trip back to the origin location. The costliness of the solo trip back to the origin creates a constraint where the commuter likely lives close to their place of employment. This requires that the friend or partner drive back to their location of trip origin. Since taxis operate on a meter based system, and friends don't drive friends very far, it would be expected for a car passenger to routinely substitute their commute with a taxi service.

The fourth category of mode alternative is walking to work. Long Island is a suburban landscape on a relatively flat land so there are not many physical barriers that prevent commuters from walking to work. However, zoning throughout the counties often prevents mixed use development which makes residential neighborhoods separated from places of employment. One limitation to this data is that the climate of Long Island does see snowfall during the winter months and it is not possible to determine the date of the respondent's survey response. The survey was conducted over a one year period, so the survey responses may have been gathered during a time of more mild weather conditions which would encourage active transportation. Alternatively, the survey responses could have been collected during the winter months and the number of walking commuters would decrease during the colder months. This category is only composed of the walking mode of commuting.

The fifth category of the local bus system. Nassau County is served by the Nassau Inter County Express (NICE) and Suffolk County is served by Suffolk County Transit Bus. The Long Island bus system is primarily for travel within the counties as a local service. While the buses do offer transfers to buses in NYC, the frequency of service is low and the bus commute from eastern Queens is time-intensive. The bus system was made into a separate category from the railroad because the buses are local while the trains offer direct service into Manhattan. A commuter who opts to take the bus likely works within the county they live in.

Table 1: Summary of Mode Choice Splits

Mode	Number	Percent (%)
SOV	4908	89.09%
Railroad	294	5.34%
Auto Passenger	198	3.59%
Bus	32	0.58%
Walk	77	1.40%
Sum	5509	100.00%

Explanatory Variables

To properly model the mode choice selection by the commuter, I carefully selected the following explanatory variables. The first variable chosen was the total number of travelers in the trip (TOTTR_R). When an individual travels with more than one person, they may have dependents to drop off somewhere before going to work. For families, having to drop a child off at daycare or school would make it more likely for them to commute to work via car. As mentioned previously, the car allows for custom routes, and traveling to a child's school and then to place of employment is a unique trip. It would be difficult to coordinate the same level of efficiency using public transportation.

The household income range (INCOM_R) allows the model to obtain information about how much the individual is able to spend to be more comfortable. An individual with lower income may not be able to afford the associated costs of owning a car or motorcycle. There are high fixed costs in purchasing a car, and regular variable costs of gas and maintenance that vary as the distance driven changes. I expect that those who have lower income may also be less likely to own their dwelling and may rent apartments or other forms of shared living. Areas with more apartments, and hence more people per square mile, are more likely to be served by public transportation.

The number of vehicles owned by the household (HHVEH_R) gives the model a way to understand how costly it is for a household to take a car to work for a shift. A household that has fewer cars is less likely to drive to work because the other members of the household may need to use the car during the day. I would expect that households with fewer cars is more likely to be an auto-passenger because the car would be a scarce resource to the household and is costly to leave parked at work all day.

The binary variable that represents whether the destination of the trip is in New York City (DNYC) either takes on a 1 or 0. Both New York City and Long Island have a lot of congestion during rush hour times of the day. Since most people work during the day, and they commute during peak times, I expect that those who work in the city are likely to use the railroad mode. The railroad mode also provides the commuter with a consistent departure and arrival time and is only delayed in the event of severe weather or problems with the train tracks.

The household size (HHSIZE) variable allows the model to understand how many people live in the household. This is useful for pattern recognition when combined with the HHVEH_R variable to understand the proportion of vehicles to members in the household. The suburban landscape of Long Island and the highly rated schools draw families to live there. Larger household sizes likely correlate to more dependents for the commuter. A commuter who has many dependents living at home would value working closer to home in order to be accessible in the event of a child's illness, or to provide for an elderly parent.

The trip distance measured in miles allows the model to understand how far the commute to work is. A longer commute would discourage commuters to walk to work. An average walking speed of three miles-per-hour would mean that a distance of over three miles would take over one hour to complete. The trip distance is a computed variable by using the New York Best Practice Model (NYBM) highway network model. This model uses software to estimate the travel distance between two locations, taking main thoroughfares, rather than a straight line distance through residential areas with low speed limits. I used this computation of distance because Long Island has an overwhelming residential population and driving requires navigating around neighborhoods and staying on main roads.

The trip duration variable is measured in minutes and is the amount of time it takes the individual to reach their destination. The duration is important because as duration increases I would expect that commuters would be more likely to choose more comfortable transportation options. One example of this can be if the duration was high, it would be less likely for a commuter to choose to be an auto-passenger. A longer trip means more time on a taxi meter and a more costly taxi ride. The combination of the trip duration and the trip distance allows the model to understand how long it takes to travel a given distance. Trips that are long in distance but short in time may suggest that the commuter is utilizing the railroad. If a commuter travels a small distance in a long amount of time it may suggest that they are walking to work. The machine learning pattern recognition would likely notice the connection between these variables.

Table 2: Definitions of Explanatory Variables

	Variable	Definition	Values
1	TOTTR_R	Total Number of travelers in the trip	1: Single Occupant; 2: 2 Persons; 3: 3 Persons; 4: 4+ Persons;
2	INCOM_R	Household Income Range [Computed]	1: Below $30k; 2: $30k-$74.9k; 3: $75k-$99.9k; 4: $100k +
3	HHVEH_R	Number of vehicles owned by household [Computed]	0: 0; 1: 1; 2: 2; 3: 3+
4	DNYC	Is Destination New York City	1: Yes; 0: No;
5	HHSIZE	Number of members living in the Household	Continuous
6	TRPDUR	Duration of trip in minutes	Continuous
7	TRPDIST_HN	Computed Trip Distance in Miles	Continuous

Model

The dataset was divided into a training and a test set. The models were trained on a dataset that was composed of a random selection of 80% of the entire dataset, known as the training set.

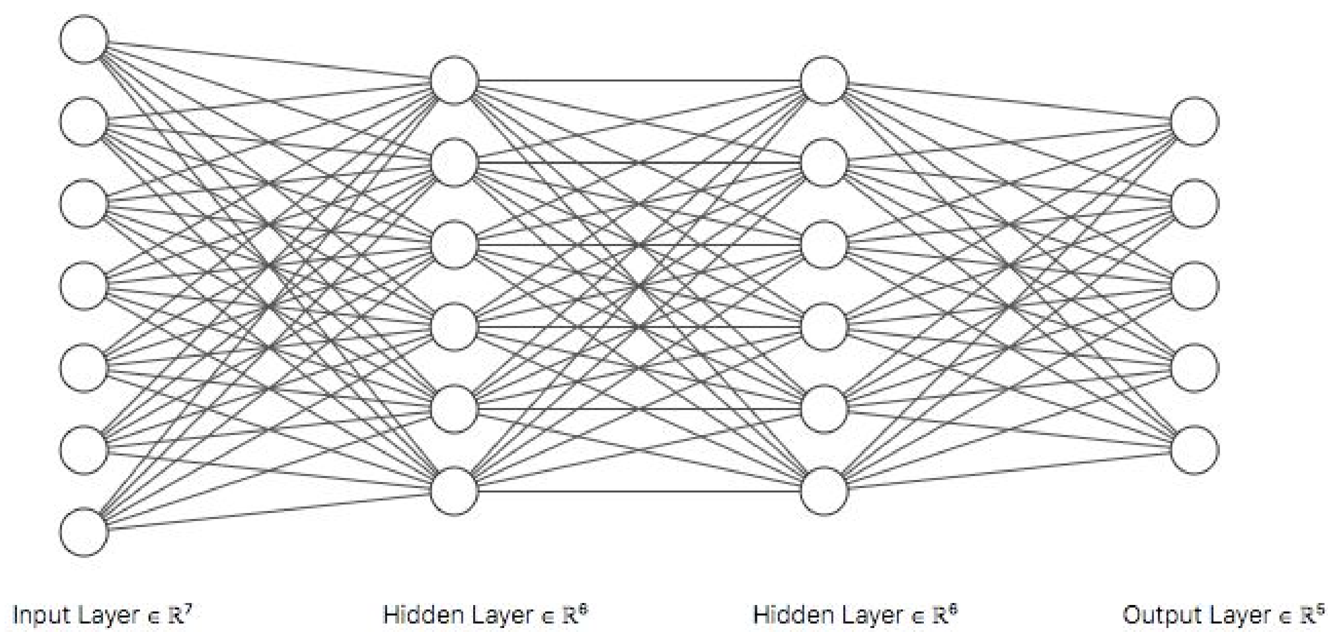
During this learning phase, the models began to understand which variables are more likely to influence mode decision. After the models were able to learn the patterns and factors of mode decision making, the model was given a new dataset and was programmed to make predictions based on the new data points. By using a separate test set, I was able to see whether the model was overfitting. An overfitting model would show a strong prediction during the training phase and a weak prediction of the new observations.

The models used in this paper were the MNL model, the support vector machine (SVM), and the artificial neural network (ANN). Below is the equation for the MNL.

$$P(i|z, C, \beta) = \frac{e^{z_{i}, \beta}}{\sum\limits_{j \, \epsilon \, C} e^{z_{j}, \beta}}$$

The MNL can be thought of as a type of single layer feedforward perceptron. The MNL consists of a few parts. The input layer contains neurons for each of the explanatory variables. Then, each of the explanatory variables is passed into the utility function where beta coefficients exist and act as the weights of how significant the individual explanatory variables contribute to the utility. Next, the likelihood function takes the utilities and evaluates the utility divided by the sum of the utilities and uses the information to indicate the most likely option for mode choice. MNL and the SVM were developed using Python Version 3.6 and sklearn. Sklearn is an open-source library that has features to improve the accessibility of implementing data analysis algorithms. Sklearn offers a wide variety of classifiers and allows for algorithm tuning by allowing researchers to alter portions of the functions. For the ANN, I used Python 3.6 and the Keras library built on a Tensorflow backend. Keras is an open source Application Programming Interface (API), that focuses on making it more user friendly to build a deep learning model in less time. Tensorflow is an open source library that was created by the Google Brain team and specializes in complex computations capable of running on CPU, GPU and TPU (Abadi et al. 2016).

The ANN consists of 4 layers: an input later, two hidden layers, and an output layer. The input dimension contains one neuron for each of the explanatory variables and the output layer has one dimension for each of the five transportation modes. The activation function for the initial layers uses the relu function and uses the softmax function for the output layer. The loss function selected for the ANN is the categorical cross-entropy which can be thought of as a log-loss function for an outcome of more than two possible outcomes.



The SVM was built using sklearn Support Vector Classifier and using the kernel trick. The SVM is an algorithm that tries to maximize the distance between two nearby vectors of different classes. In the case of non-linear separable data, a kernel function is used to map the data to a higher dimensional space and construct a hyperplane to separate the data.The SVM was initially derived to classify data that was a binary outcome. While traditional SVM can be used to compare one-to-one classification and create (n * [n-1]) / 2 classifiers to solved multi-class problems, I followed the method of Watkins & Weston (1998) who used a one-against-all strategy. In order to use the SVM in a multi-class problem, it was necessary to pass the "One vs Many" parameter to the function call in order to compare one class to the remaining classes.

Machine learning models offer advantages over conventional econometric models. One advantage is not having to strictly specify the model in advance. This provides flexibility for the model to learn patterns in the data, and spend less time preparing a model before being able to interpret the results. However, one pitfall of machine learning models is that they are difficult for many people to understand, because there is very complex mathematics behind them. This causes a barrier of understanding between those who implement the models and those who are able to build them from scratch. Also, machine learning models lack the interpretability that econometrics models offer. This means that while they excel at classifying data, they do not give meaningful insights about how the change of inputs will affect the output. Machine learning algorithms are basically functions that operate on a loss function and continuously update the weights of combinations of features. Then, based on the estimate, the error is fed back into the model and it is able to try and estimate the output again, now know the error of the previous predictions.

Results

One common trend across all of the models was that the modes that had the most observations were the most accurate. It is tough to decipher if the models classified these modes correctly because it had more available data to train. The SOV and the Rail had the most observations at 987 and 59 respectively.

The ANN performed well in the categories for the SOV and Rail at 98.8% and 78% respectively. The SVM performed well in SOV and Rail at 98.8% and 83.1% each. However, one interesting observation for the machine learning models was that it had difficulty distinguishing between the SOV and the auto-passenger classes. Of the 30 observations that were actually auto-passenger, the machine learning models incorrectly guessed the mode "SOV" 27 times using SVM and 28 times using ANN. Xie et al. (2003) ran into this same problem and suggested that it may be due to the two modes sharing similar observed attributes and that the data may not have an explanatory variable that can distinguish between the two modes.

To interpret the results from the models, I will use a confusion matrix to represent the predicted values compared to the actual values. Two scores will be calculated to understand the prediction rate for the models. The first score is the individual match rate: defined as the correct predictions divided by the actual number of observations of that mode choice. For example, if a model correctly predicts that 9 people will eat ice cream and the survey reveals that 10 people ate ice cream, it would have a 90 percent individual match rate. The second calculated rate is the aggregate match rate which is defined as the total predicted (both correct and incorrect) divided by the actual amount of people who chose that given mode. The confusion matrices are only using the actual and predicted values from the test set. Reference the Confusion Matrices at the end of this paper in the Appendix in order to better visualize the findings of this paper.

One limitation to this paper is that the models were unable to learn the distinct differences between some of the classes. This is likely because of the limited amount of observations for the less popular travel modes. One way I think this can be solved is if I were to apply a balancing function and generate more heterogeneity between the modes. The travel behavior of the Long Island area does reflect a strong attraction to SOV. It would be beneficial for the model to have a training set with more of the uncommon modes such as bus, walk, and auto-passenger, and less of the most popular modes such as SOV.

To introduce more data to this paper, I could have expanded the geographic zone to include other suburban areas around NYC. These could have included area in Connecticut and New Jersey that have similar commuting patterns to Long Island. Creating a model for the entire metro area would also be a good project to use to compare the MNL and machine learning algorithms. Further research could try implementing decision tree classification that can create main splits based on geography and then continue with the algorithms to separate the data based on splitting of the levels of the features.

A second limitation to this paper is that there was a time constraint and I certainly could have been able to better tune the machine learning models to predict at a higher rate. I could have experimented with different activation functions, number of hidden layers, training epochs, and batch sizes for the ANN. For the SVM, I could have experimented with different kernel functions that may have yielded more or less accurate predictions. Outside the scope of the two machine learning models used in this paper, there are many other algorithms used for classification that have had high success in multi-class classification problems. Further research can try and use other algorithms such as decision trees, random forest, and gradient boosting.

Conclusion

Although the MNL needs to be specified in advance and requires strict assumptions regarding IIA, it does allow for more interpretation than the machine learning models. Looking forward, there will be a lot more data coming from cell phones, smart watches, and other GPS connected devices. The data emitted from these devices will be extremely detailed and with so much data it will be difficult to specify a model in advance. Machine learning models will be able to provide rapid insight into transportation mode choice from GPS information from these devices. The cost of collecting data from these devices should be much less expensive than the travel surveys which has been the traditional way to obtain data for transportation research. Since the price will be lower for data collection, we should expect to encounter a lot more data in the future. I argue that machine learning models are here to stay, and that they will have a role in the future of mode choice understanding. They can be used for feature extraction and dimensionality reduction in order to extract the more critical data that can then be used in models better suited for interpretability such as the MNL.

Appendix (Actual Values refer to the Test Set)

Multinomial Logit Model Confusion Matrix

Predicted

		Rail (**66**)	Bus (**0**)	Auto-passenger (**9**)	SOV (**1033**)	Walk (**0**)	Individual	Aggregate
	Rail (59)	**45**	0	0	14	0	76.27%	111.86%
	Bus (8)	1	**0**	0	7	0	0	0.00%
Actual	Auto-passenger (30)	2	0	**3**	25	0	10.00%	30.00%
	SOV (987)	12	0	6	**969**	0	98.18%	104.66%
	Walk (18)	0	0	0	18	**0**	0.00%	0.00%

Support Vector Machine Confusion Matrix

Predicted

		Rail(**62**)	Bus(**2**)	Auto-passenger(**2**)	SOV(**1035**)	Walk(**1**)	Individual Match Rate	Aggregate Match Rate

Artificial Neural Network Confusion Matrix

		Predicted						
		Rail (**62**)	Bus (**0**)	Auto-passenger (**0**)	SOV (**1038**)	Walk (**0**)	Individual Match Rate	Aggregate Match Rate
	Rail (59)	**46**	0	0	13	0	77.97%	108.47%
	Bus (8)	4	**2**	0	4	0	25.00%	0.00%
Actual	Auto-passenger (30)	2	0	**0**	28	0	0.00%	0.00%
	SOV (987)	12	0	0	**975**	0	98.78%	105.17%
	Walk (18)	0	0	0	18	**0**	0.00%	0.00%

Works Cited

Abadi, Martín, et al. "Tensorflow: a system for large-scale machine learning." *OSDI*. Vol. 16. 2016.

Ben-Akiva, Moshe E., Steven R. Lerman, and Steven R. Lerman. *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press, 1985.

Evans, Suzanne P. "A relationship between the gravity model for trip distribution and the transportation problem in linear programming." *Transportation Research* 7.1 (1973): 39-61.

Hasnine, Md Sami, et al. "Determinants of travel mode choices of post-secondary students in a large metropolitan area: The case of the city of Toronto." *Journal of Transport Geography* 70 (2018): 161-171.

Kim, Sungyop, and Gudmundur Ulfarsson. "Travel mode choice of the elderly: effects of personal, household, neighborhood, and trip characteristics." *Transportation Research Record: Journal of the Transportation Research Board* 1894 (2004): 117-126.

McFadden, Daniel. "Conditional logit analysis of qualitative choice behavior." (1973).

Moavenzadeh, Fred, and Michael J. Markow. *Moving millions: transport strategies for sustainable development in megacities*. Vol. 14. Springer Science & Business Media, 2010.

Weston, Jason, and Chris Watkins. *Multi-class support vector machines*. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, 1998.

Xie, Chi, Jinyang Lu, and Emily Parkany. "Work travel mode choice modeling with data mining: decision trees and neural networks." *Transportation Research Record: Journal of the Transportation Research Board* 1854 (2003): 50-61.