European Transport \ Trasporti Europei (2023) Issue 92, Paper n° 3, ISSN 1825-3997

https://doi.org/10.48295/ET.2023.92.3

A Comparison of KNN Algorithm and MNL Model for Mode Choice Modelling

Gopika Vinayakumar[1], Anu P. Alex [2*], Manju V. S.[3]

[1] *M. Tech, Student, Department of civil engineering, College of Engineering, Trivandrum 695 016*
[2] *Associate Professor, Dept. of Civil Engineering, College of Engineering, Trivandrum 695 016*
[3] *Professor, Dept. of Civil Engineering, College of Engineering, Trivandrum 695 016*

Abstract

Mode choice modelling helps to identify potential users of traffic and plays an important role in policy and decision-making by the government. With the advancement of artificial intelligence and machine learning techniques, several studies were carried out to analyse the performance of mode choice models in which the backpropagation algorithm was used. However, for faster convergence of parameters, it would be interesting to explore other efficient algorithms of machine learning as the conjugated gradient search in spite of the backpropagation algorithm. The present study adds to the literature about the performance of the K-Nearest Neighbour (KNN) algorithm in mode choice modelling and compared the KNN model with the traditional MNL model. It was unveiled that the variables, which are found significant and important in both models are the same. It is also found that the KNN model is outperforming MNL with a prediction accuracy of 73.84%.

Keywords: work trip; MNL; KNN.

1. Introduction

Mode choice modelling is vital for the effective design of a transportation system. It helps the planners to determine what mode of transport will be used by the commuters and the resulting modal share. The three broad categories of factors that influence the mode choice can be trip characteristics, transportation system characteristics and trip maker's socioeconomic characteristics (Ashalatha et al., 2013). It had been an area of interest for the researchers to improve upon the techniques used for mode choice modelling. There are varieties of techniques applied in mode choice modelling, from traditional probit and logit models to artificial intelligence-based (AI) models.

Traditional models were initially used with aggregate data, later it was moved to a disaggregate level due to the advantage of capturing the taste heterogeneity of individuals. These models were widely used to find out the factors affecting mode choice owing to

* Corresponding author: Anu P. Alex (anualex@cet.ac.in)

the statistical theory involved in it. These models rely on utility maximization theory, in which it is assumed that a decision-maker chooses an alternative from a choice set whose utility is maximized. The probit model was popular among traditional models. Laborious calculation of the probit model led to the use of logit models. Multinomial logit model (Ashalatha et al., 2013; Chanda et al., 2016; Krishnapriya and Soosan, 2020), the nested logit model (Almasri and Alraee 2013), and mixed multinomial logit models (de Jong et al., 2003; de Luca, 2005) are the commonly used logit models in mode choice. Multinomial logit models are still being used by researchers due to their ability for easy interpretation and less intense calculation. Independent of Irrelevant Alternatives (IIA) property of these models led to the use of Nested Logit models (NL). The constraint of the NL model is that the unobserved attributes can be connected to one of the choice dimensions only (Ratrout and Gazder, 2014). The mixed logit model has got wide acceptance among logit models since it does not have the drawback of IIA. Logit models mainly predict outcomes based on independent variables, and hence if there is any inaccuracy in the independent variable it affects the predictive value. Continuous outcomes cannot be predicted by logit models. Data should be independent to avoid the over weightage in the significance of variables. Moreover, logit models can appear to have more predictive power than they do, if there is any sampling bias. Hence researchers are still going on to improve the traditional logit models.

Due to the performance of Artificial Intelligence and machine learning techniques in big data handling and nonlinear modelling, it has got wide applications in various science and engineering fields; specifically in mode choice modelling. ANN (Karlaftis and Vlahogianni, 2011) and Fuzzy Logic (Rahman and Ratrout, 2009; Pulugurta et al., 2013) are the commonly used AI techniques in transportation modelling. ANN is found suitable for transportation-related problems, provided there is a large set of training data. ANN has generic abilities and hence it is widely used in transportation problems. Its output has no limitation and no chance of losing data since they are stored in the network and it helps to get efficient output. Studies conducted by (Hussain et al., 2017; Alex et al., 2019) compared ANN with other techniques like Multinomial logit and multiple linear regressions. ANN has the drawback of overfitting the training data, due to which the accuracy of prediction may get reduced.

The uncertainty of human behaviour can be easily captured by fuzzy logic models. It was successfully applied in four-stage planning models, network analysis, accident prevention and signal control (Rahman and Ratrout, 2009). These models also require a large amount of data and familiarity with the nature of the system for developing fuzzy set rules. Neural networks and fuzzy systems can be combined to have all their advantages and cure individual illnesses. Literature shows that ANN and fuzzy logic outperform traditional logit and probit models, and at the same time, neuro-fuzzy architecture outperforms and overcomes the limitations of ANN and fuzzy logic (Minal et al., 2018). All neuro-fuzzy architecture uses the gradient descent technique for learning its internal parameters. For a faster convergence of these parameters, it would be interesting to explore efficient algorithms of machine learning other than the backpropagation algorithm. This leads to the search for a simple machine learning algorithm which can be used to solve classification and regression problems and ends in KNN.

KNN is a supervised learning algorithm in the basic form and is easy to implement. It can perform complex classification tasks. It is also known as a lazy learning algorithm because it does not have a specialized training phase. Another feature of KNN is that it is a non-parametric learning algorithm i.e., it doesn't assume anything about the underlying data. It is an instance-based learning method which stores all available data points and classifies new data points based on similarity measures. It is very easy to implement because the major two parameters used in KNN are the K value and distance function (e.g., Euclidean or Manhattan etc.). KNN algorithm has been applied in many fields such as data mining, image processing, handwriting recognition, pattern recognition, and ECG disease classification for better performance (Gupta and Mittal, 2020). Studies conducted by (Chao Li. et al., 2012) and (Upadhyay et al., 2020) showed that KNN can be used for signature verification. Studies by (Saini et al., 2013) and (Jamaluddin et al., 2019) applied KNN in medical fields like urinary tract infection diagnosis classification of QRS-complex delineation and lymph node metastasis etc. KNN was applied in chemistry for the solubility of hydrocarbon (Lashkenaria and KhazaiePoulb, 2015). The study by (Preethi and Priyadarsini, 2017) reviewed the use of the KNN algorithm in medical and social networks. It was concluded that KNN is performing better than ANN in a study for classifying the spectrogram images in brain balancing (Mustafa et al., 2012).

KNN algorithm is usually used when the real-world data does not obey the theoretical assumptions and is used in other techniques like linear separability, uniform distribution of data, etc. This algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. The major benefits of the KNN algorithm are that it is versatile; no assumption is used and can be used for classification and regression problems. There are limited studies in which KNN is used in the field of transportation. KNN was applied for travel time and traffic prediction and found useful (Xu, 2018). The performance of the KNN algorithm in mode choice prediction has not been studied yet. This study intends to check the suitability of the KNN algorithm in mode choice modelling and its comparison with commonly and traditionally used MNL models.

2. Study Area and data collection

The study area chosen was Thiruvananthapuram, the capital of the state of Kerala in India. The growth of Thiruvananthapuram in recent years has induced enormous transportation demand. The city is in the southwestern part of India and located in latitudes 8°17' to 8°54' in the north and longitudes 76°41' to 77°17' in the east. As per the report of the 2011 census, Thiruvananthapuram district has a population of 3,307,284 with 55.75 % urban population, in which the working population is 276,641. The population density is 1506 persons per square kilometre (Thiruvananthapuram Municipal Corporation). The projected population for the year 2020 was found to be 3,369,767. Thiruvananthapuram district has 8.28 lakh households, of which 47.10% are in rural areas and 52.9% are in urban areas. The city is well connected to all over India as well as the world with roads, rails, waterways and airways. Public transport is well used to connect internally as well as externally. Kerala State Road Transport Corporation buses play a major role in the public transport system. Private buses on selected routes, auto-rickshaws and cabs are also a part of it. Commonly used private modes are cars and two-wheeler.

A random sampling method was adopted in the study for data collection. The scope of the study is limited to the mode choice behaviour of workers only, hence 1899 workers were interviewed with a predefined questionnaire. The questionnaire was divided into three sections: household characteristics, individual characteristics and trip characteristics. Household characteristics included household size and vehicle ownership. Individual characteristics included age, gender, employment, income and marital status. Trip characteristics included mode, purpose, travel time, cost and distance. t-tests for age groups showed that the sample population is a true representation of the target population.

3. Methodology

KNN and MNL models are developed in the study after coding the collected data. Both models are validated and the prediction accuracies are compared. KNN and MNL models are discussed in the following subsections.

3.1 KNN model

KNN is a supervised and versatile machine learning algorithm used for predicting regression and classification problems in various fields. It uses 'feature similarity' to predict the values for new data points i.e., the new data point is classified as a value based on how closely it matches the points given in the training set. The various steps used in the KNN algorithm are as follows.

Step 1- The first step of KNN is the classification of training as well as test data. Figure 1 gives the representation of two types of sample data (yellow square and blue diamond represent male and female commuters, axis represents the distance travelled and time taken by commuters) chosen for demonstrating the KNN algorithm.

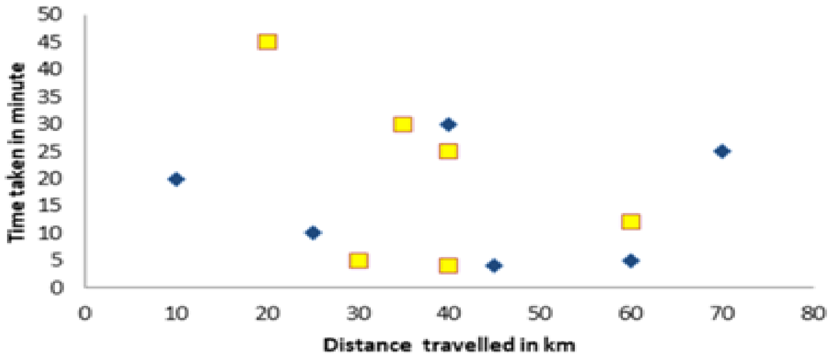


Figure 1: Sample data set.

Step 2 - Next step is the selection of the value of the nearest data points 'K', an integer, assuming it is 3 in this example. A new data point (commuter travelled 35 km in 25 min) represented as a red triangle in Fig. 2 is added to the data set.

Step 3 - Distance from test data is calculated with help of the Euclidean, Manhattan method for the classification of the new data point. The distance can be found using (1) and (2). The most commonly used method to calculate distance is Euclidean.

The $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ -Euclidean distance (1)

$\sum_{i=1}^{k}|x_i - y_i|$-Manhattan distance (2)

Step 4 - The distance is sorted in ascending order.

Step 5 - Next step is the choice of the top K rows from the sorted array which is near the new data point.

Step 6 - Assigning a class to the test point based on the most frequent class. The nearest points of the new data point (commuter travelled 35 km in 25 min) for K=3 is shown in Figure 2. Thereby, the new data point is classified as yellow square i.e., male commuter.

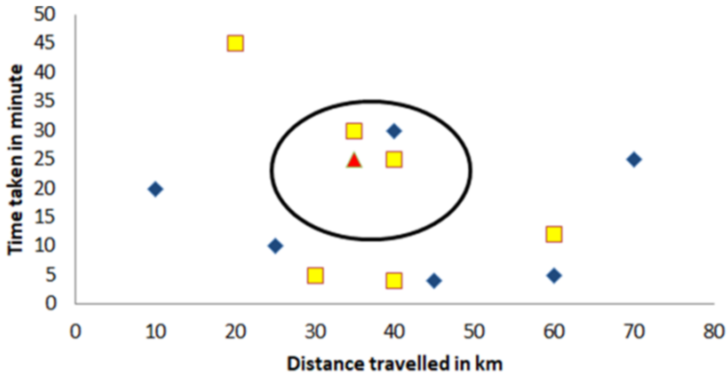


Figure 2: Classification of new data point

Step 7 - Probability can be found out by differentiable cost functions based on stochastic ("soft") neighbour assignments in the transformed space. Each point 'i' selects other points 'j' as its neighbour with a probability p_{ij} based on the softmax of the distance d_{ij}, shown in (3).

$$p_{ij} = \frac{e^{-d_{ij}}}{\sum_{k \neq i} e^{-d_{ik}}} \,, \qquad p_{ii} = 0 \qquad (3)$$

Where d_{ij} is the Euclidean distance given in (1).

The output of the KNN model obtained from SPSS consists of a predictor space diagram to show the, K selection error log used to select the K, predictor importance, quadrant map and peers chart. The predictor space diagram shows every point on the data set and the K selection error log helps to select the least error 'K' value. Predictor importance is used to analyse the importance of each variable in prediction. If a new data point is added, it is linked to its nearest data point and it will be shown in the predictor space diagram and quadrant map. The nearest neighbour of each feature is shown in the peers chart.

3.2 MNL model

MNL model is widely used due to its simplicity and ability to obtain the relationship with predictor variables. MNL is based on the theory that each option has a net utility for individual 'j' and the choice available by commuter 'q'. The net utility has two components, as in equation (4), where v_{jq} is the measurable part and ε_{jq} is a random part. v_{jq} is a function of measured attributes x; it is obtained as a linear combination of x, given in equation (5).

$$U_{jq} = V_{jq} + \varepsilon_{jq} \qquad (4)$$

$$V_{jq} = \Sigma\theta_{kj}X_{kjq} \qquad (5)$$

Where θ is the coefficient to be estimated.

The probability of alternative 'i' chosen by commuter 'q' can be formulated as (6).

$$P_n = \frac{e^{V_n}}{1+\sum_{m=M} e^{V_m}} \qquad (6)$$

Where, the probability that the commuter selects the mode n is given by Pn. Vn is the utility of chosen mode n. The utility of any mode is given by Vm and m in the equation is the set of all available modes.

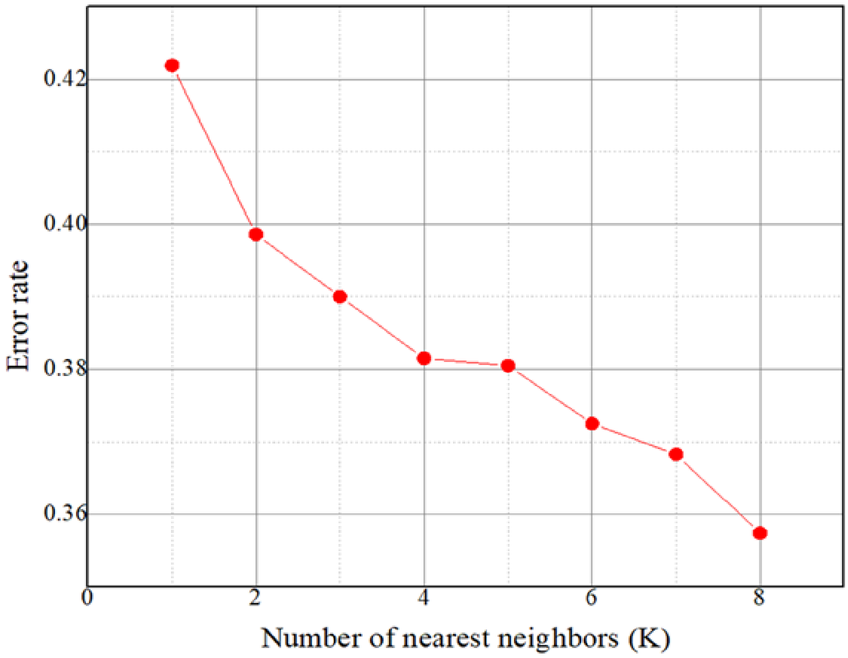
While developing the MNL model, one category of mode is chosen as the reference category to which all other modes are compared. The influence of variables on the mode choice of the workers can be understood by developing MNL.

4. Model development

It is intended to know the various attributes and characteristics that decide and influence the mode choice of workers. Both KNN and MNL models were developed using SPSS software. 80% of the sample was used for model development. Socioeconomic characteristics like gender, age, employment, vehicle ownership, household size, marital status, two-wheeler availability, car availability and trip-related characteristics like cost, time and distance of travel are considered in the study. The correlation matrix revealed that there is a high correlation between vehicle ownership and car availability. It also shows that travel time and cost are highly correlated with distance. Hence car availability, travel time and cost are not considered for model development. Household income is not considered in the study due to the reluctance of people to reveal their true income.

4.1 KNN model

71.5% of the active data set is used to train the nearest neighbour model. The remaining 28.5 % of the active data set was used as a holdout sample. The K value is selected by applying feature selection from K selection error log graph, which is a record of errors that are encountered by running algorithm. Normally K value, i.e., the optimum number of nearest neighbours, with the lowest error rate is selected by running the algorithm with different values of K. Small K values are sensitive to noise and higher values are computationally expensive (Chao Li. et al., 2012). Hence from Figure 3, the optimum K value is obtained as 8, which has the lowest possible error rate of 0.36. The importance of the independent variables used for modelling was obtained from the relative value of the variable importance chart (Figure 4). Variable importance explains the importance of each variable in making predictions and it does not have any relation to model accuracy. The predictor importance values obtained are 0.13 for vehicle ownership, distance, employment, gender and household size. The value obtained is 0.12 for age group, two-wheeler availability and marital status. This implies that the variables considered in the model are important and need not be eliminated.



\

Figure 3: K selection error log

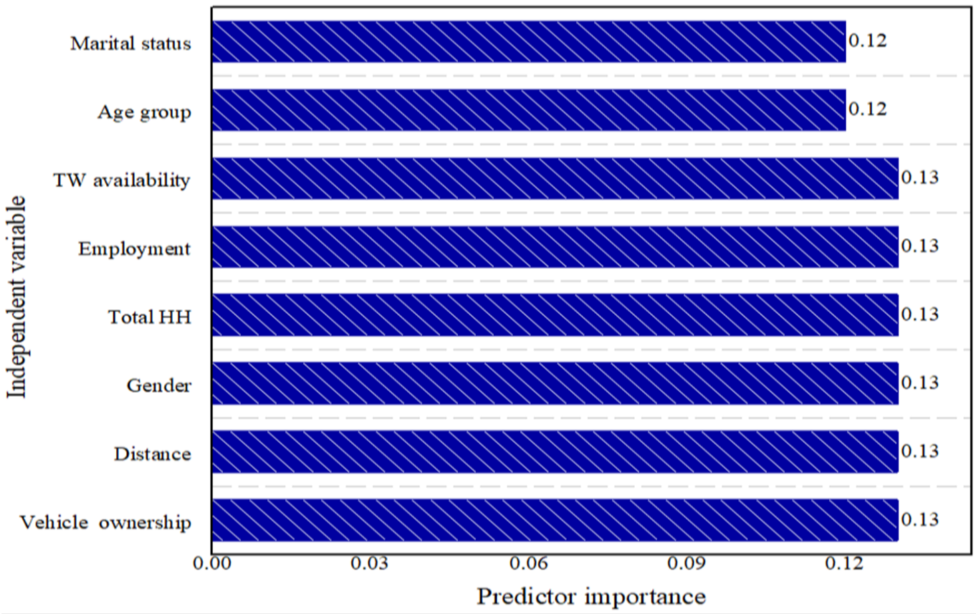


Figure 4: Predictor importance

Figure 5 shows predictor space, which is the interaction graph of feature space. If there are more than three features used for modelling, it results in a subspace. Features selected in the model are shown in axes. The location of points shows the values of features for all cases in both the training and holdout partitions. Training and holdout partitions are represented as circles and triangles respectively. The category value of the target variable is indicated by the various colours. Point (A) is Selected in the feature space diagram for the explanation of KNN algorithm, called a focal case, denoted by lines as shown in Figure 6. The focal case is usually linked to their K nearest neighbours. The nearest neighbours and distances from point ‘A’ in Table 1. The focal case and their K nearest neighbours on a scatter plot for diffrent features on axis are called a quadrant map (Figure 7). Features of K nearest neighbours and focal case are obtained from peers chart as shown in Figure 8.

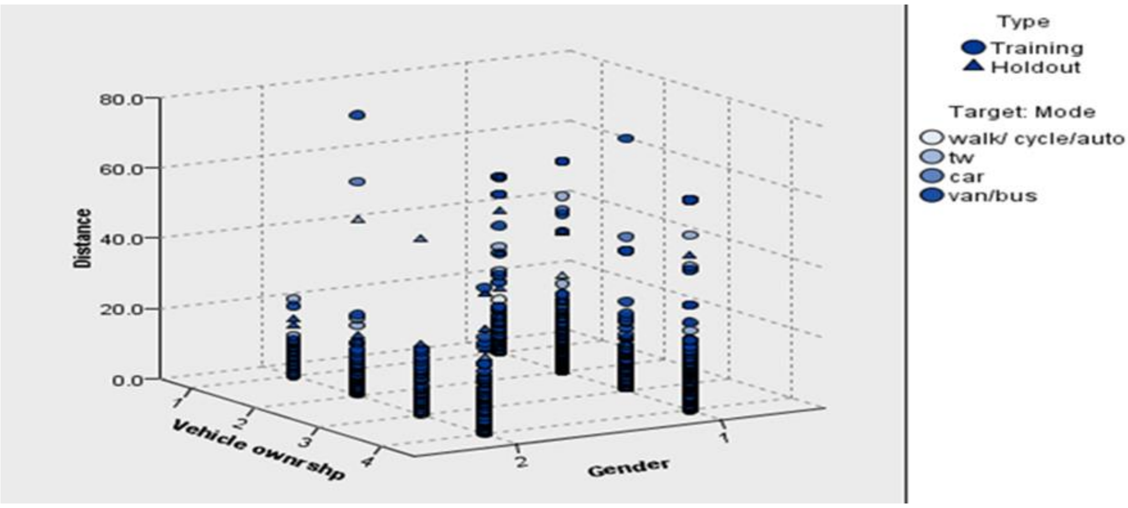


Figure 5: Predictor space

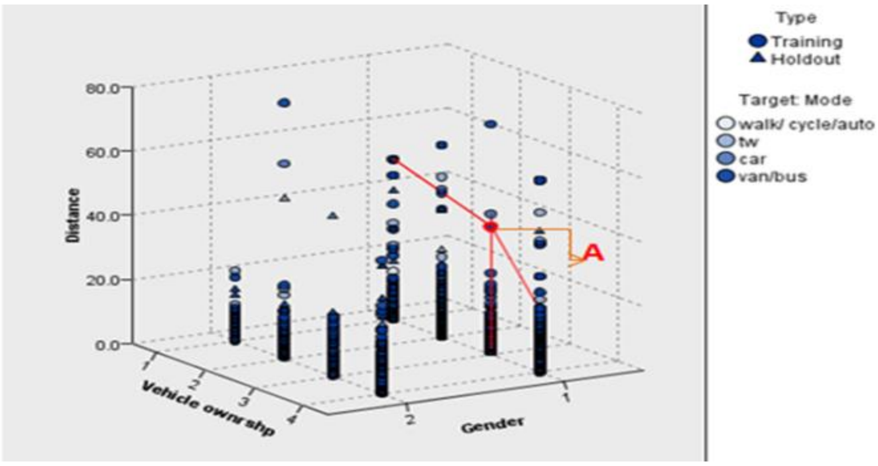


Figure 6: Predictor space with sample focal case

Table 1: K Nearest neighbours and distance

Focal Record	Nearest Neighbours								Nearest distances							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
281	719	1259	1604	24	493	492	1653	796	0.573	0.601	0.698	0.717	0.718	0.718	0.736	0.740

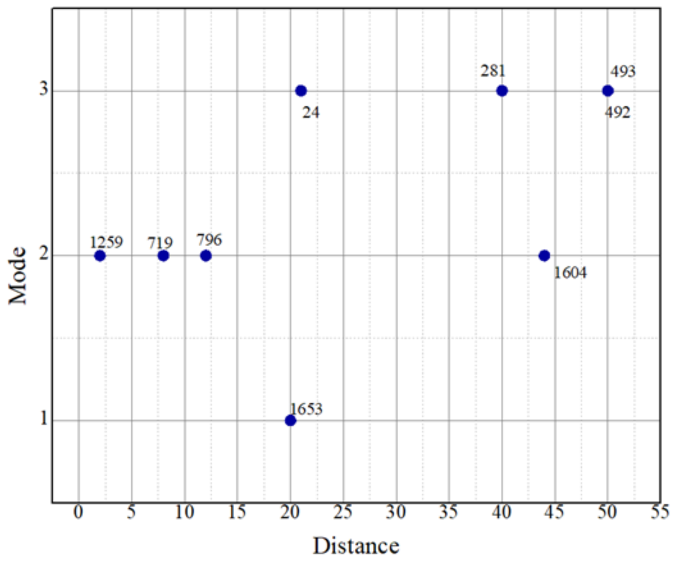


Figure 7: Quadrant map

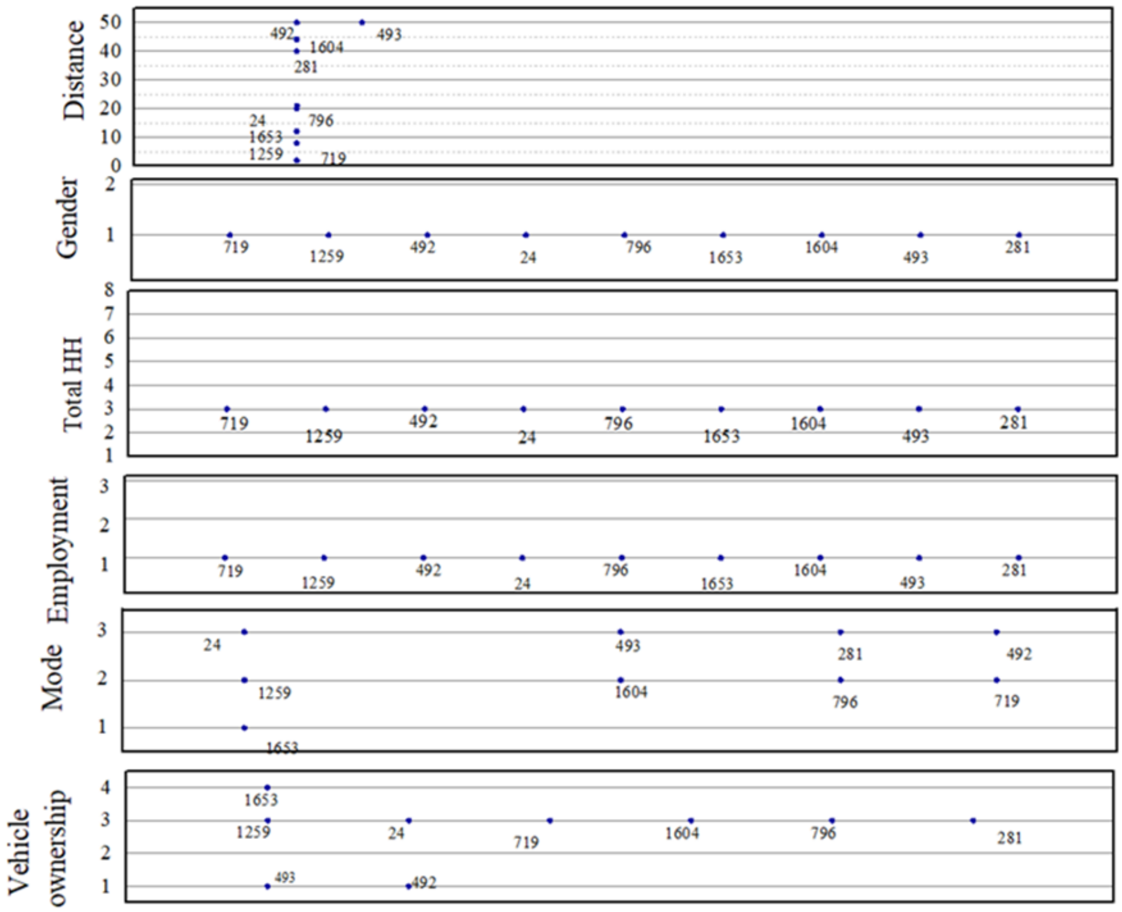


Figure 8: Peers chart

4.1 MNL model

MNL model was developed using independent variables viz; household size, two-wheeler availability, vehicle ownership, gender, age group, marital status, employment, distance and the dependent variable is the mode chosen by the commuters. The base mode chosen for MNL is walk/cycle/auto. The maximum likelihood estimation method was used for finding values for the coefficient, so as to maximize the likelihood and the model predicts the same choice made by the observed individual, to yield highly accurate estimates. Model fitting information of the final model is given in Table 2, which shows that the -2 log-likelihood value of the final model is less than that of the intercept-only model and the p-value of the chi-square test is 0.000. Both of them indicate that the developed model is good. This implies that there is a significant relationship between the chosen independent variables and the dependent variable.

Table 3 shows the parameter estimates of the developed model. Variables which are significant up to a 5% level of significance are included in the final model. Total household size, two-wheeler availability, gender, age, marital status, employment and distance are found significant for two-wheeler mode. Wald statistics showed that among these significant variables, two-wheeler availability is more significant and household size is less significant towards two-wheeler utility. As household size increases, the chance of using two-wheeler decreases. The utility of a two-wheeler is more for youngsters and males. Vehicle ownership, distance, employment, marital status and two-wheeler availability are obtained as significant for car utility and the order of significance decreases from vehicle ownership to two-wheeler availability. Two-wheeler availability, gender, employment and distance are found significant for van/bus. The significance level of variables increases from two-wheeler availability to distance. Females are more likely to use a van/bus than walk/cycle/auto. Males use two-wheeler or cars more than walk/auto/cycle. The model reveals that youngsters use two-wheeler more than walk/cycle/auto, and walk/cycle/auto is mainly used by aged people. The model also proves that as distance increases commuters prefer van/bus, car and two-wheeler than walk/cycle/auto. There is more chance for commuters who own both cars and two-wheeler to use car than other modes.

Table 2: Modal fitting information

Model	Model fitting criteria	Likelihood ratio tests		
	-2 log-likelihood	Chi-square	df	Significance
Intercept only	3998.123	1292.419	24	0.00
Final	2705.704			

Table 3: Parameter estimates of MNL model

Mode		B	Wald	Sig.
Two-wheeler	Intercept	11.649	78.334	0.000
	Household size	-.243	5.467	0.019
	Two-wheeler availability	-3.041	105.610	0.000
	Vehicle ownership	.107	.795	0.372
	Gender	-1.184	16.726	0.000
	Age group	-.456	7.319	0.007
	Marital status	-1.361	18.804	0.000
	Employment	-1.295	36.496	0.000
	Distance	.390	65.158	0.000
	Intercept	-.900	.382	0.536
	Household size	-.127	1.293	0.255
	Two-wheeler availability	1.079	11.839	0.001
	Vehicle ownership	1.455	88.243	0.000

Car	Gender	-.577	3.485	0.062
	Age group	-.046	.066	0.798
	Marital status	-1.663	19.624	0.000
	Employment	-1.229	28.902	0.000
	Distance	.412	71.353	0.000
Van/bus	Intercept	1.907	1.946	0.163
	Household size	-.077	.487	0.485
	Two-wheeler availability	-.643	4.873	0.027
	Vehicle ownership	.075	.342	0.559
	Gender	0.889	9.145	0.002
	Age group	-0.093	0.268	0.605
	Marital status	0.138	0.179	0.672
	Employment	-1.537	45.176	0.000
	Distance	0.446	84.796	0.000

5. Validation of the model

There are two phases for model validation. The first phase is measuring model statistics and the second is measuring the prediction success of the model. Model statistics have been verified in the previous section. Those models qualifying whole verification of model statistics are only considered for the prediction success table. Statistical checks and goodness of fit cannot be performed on a KNN model. The quality of the model is expressed in terms of prediction accuracy.

A test set of 474 was taken for validation purposes. In order to validate the KNN model, Euclidean distance was found for each variable and each mode by equation 2, and the probability for each mode was obtained by equation 3. MNL model was validated by finding the probabilities of each mode as a function of the systematic portion of the utility of all alternatives as per equation 5. Tables 4 and 5 describe the prediction accuracy of KNN and MNL models respectively. The prediction ability of the KNN algorithm is obtained as 73.84% and that of MNL is 59.07%.

Table 4: Prediction accuracy of KNN model

Mode	Correctly predicted	Wrongly predicted
Two-wheeler	195	45
Van/bus	72	23
Car	68	42
Walk/cycle/auto	15	14
Prediction accuracy (%)	73.84	

Table 5: Prediction accuracy of MNL model

Mode	Correctly predicted	Wrongly predicted
Two-wheeler	190	50
Van/bus	61	34
Car	21	89
Walk/cycle/auto	8	21
Prediction accuracy (%)	59.07	

6. Conclusion

The mode choice model is important for the efficient design of transportation systems, which helps the planners to determine what mode of transport will be used by the commuters and what modal share results. This study checked the feasibility of applying an emerging simple machine learning technique, KNN, to mode choice modelling of work trips and it is compared with the widely used MNL model.

From a wide range of explanatory variables such as socio-demographic attributes, at individual and household levels, mode and trip characteristics given as input for MNL and KNN models, the variables found relevant by both the models are the same. The Important and significant variables obtained for both models are vehicle ownership, distance, employment, gender, household size, two-wheeler availability and distance. Among these, two-wheeler availability, employment and distance were found significant in MNL for all the modes. The same variables were found most important in the KNN model also. This shows that KNN is able to capture the variable importance as same as that of the MNL model.

The KNN model is found superior with higher prediction accuracy of 73.84% than the MNL model which has an accuracy of 59.07%. The higher accuracy of the KNN model is due to a memory-based approach rather than an instance-based, hence it allows the algorithm to respond quickly to changes in the input during real-time use and it is non-parametric in nature. The developed KNN model can be used for predicting the mode shift towards public transport when policy implications are made. It can also be transferred and applied to other areas where similar socio-economic, mode and trip characteristics are existing. The model can be refined by collecting more data on sustainable modes like walk and cycle, as future scope.

References

Alex, A. P., Manju V. S. Isaac, K.P. (2019) “Modelling of Travel Behaviour of Students using Artificial Intelligence”, Archives of Transport, DOI: https://doi.org/10.5604/01.3001.0013.6159.

Almasri, E., Alraee, S. (2013) “Factors Affecting Mode Choice of Work trips in Developing Cities—Gaza as a Case Study”, Journal of Transportation Technologies, 3, 247-259, DOI: http://dx.doi.org/10.4236/jtts.2013.34026.

Ashalatha R., Manju V.S., Zacharia, A.B. (2013) “Mode Choice Behavior of Commuters in Thiruvananthapuram City”, Journal of Transportation Engineering, DOI:10.1061/(ASCE)TE.1943-5436.0000533.

Chao Li, Zhang s., Zhang H., Pang L., Lam K., Hui C., Zhang S. (2012) “Using the K-Nearest Neighbour Algorithm for the Classification of Lymph Node Metastasis in Gastric Cancer”. Hindawi Publishing Corporation Computational and mathematical methods in medicine, Volume 2012, Article ID 876545, DOI:10.1155/2012/87/65/45.

Chanda, R., Sen, S., Roy, S.K. (2016) “Mode Choice Modelling of Work Trips: a Case Study of Kolkata”, https://www.researchgate.net/publication/312046163.

De jong G., Daly A., Pieters M., Vellay C., Bradley M., Hofman F. (2003) “A Model for Time of Day and Mode Choice Using Error Components Logit”, Transportation Research Part E: Logistics and Transportation Review, Volume 39, Issue 3, May 2003, Pages 245-268.

De luca S. (2005) “Modelling Airport Choice Behaviour for Direct Flights, Connecting Flights and Different Travel Plan”, Journal of Transport Geography, 22,148-163,2012.

Gupta V. and Mittal M. (2020) "Efficient R-peak Detection in Electrocardiogram Signal based on Features Extracted using Hilbert Transform and Burg Method", Journal of the Institution of engineers (India): series B 101, 23-43 (2020).

Hussain H. D., Mohammed A. M., Salman A. D., Rahmat R.A. O. K. Borhan M.N. (2017) "Analysis of Transportation Mode Choice using a Comparison of Artificial Neural Network and Multinomial Logit Models", ARPN Journal of Engineering and Applied Sciences, vol. 12, no. 5, march 2017, http://www.arpnjournals.org/jeas/research_papers/rp_2017/jeas_0317_5785.pdf

Jamaluddin M.N.F., Malik S.N.A., Fauzi S.S.M., Razak T.R., Halim I.H.A., Mohammed A.H., Gining R.A.J.M. (2019) "An Application of Presumptive Diagnosis for Urinary Tract Infection via KNN Algorithm Approach", Proceedings from the 3rd international conference on the future of ASEAN (ICoFA) 2019-Volume 2, https://doi.org/10.1007/978-981- 15- 3434-8_32.

Karlaftis M. G. and Vlahogianni E. I. (2011) "Statistical Methods Versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights", Transportation Research Part C: Emerging Technologies 19(3):387-399, DOI: 10.1016/J.TRC.2010.10.004.

Krishnapriya, M. G., and Soosan George T. "Mode choice behaviour of students, integrating residential location characteristics: a study from Kochi City, India." European Transport: 2020.79 (2020).

Lashkenaria M. S. and KhazaiePoulb A. (2015) "Application of KNN and Semi-Empirical Models for Prediction of Polycyclic Aromatic Hydrocarbons Solubility in Supercritical Carbon Dioxide", polycyclic aromatic compounds ISSN: 1040-6638, DOI:10.1080/10406638.2015.1129976.

Minal S., Chalumuri R.S., Errampalli M. (2018) "Development of Neuro-Fuzzy-based Multimodal Mode Choice Model for Commuter in Delhi" IET Intelligent Transport Systems, ISSN 1751-956X, DOI: 10.1049/iet-its.2018.5112

Mustafa M., Taib M.N., Murat Z.H., Sulaiman N. (2012) "Comparison between KNN and ANN Classification in Brain Balancing Application via Spectrogram Image", Journal of computer science and computational mathematics, volume 2, issue 4, DOI:10.20967/jcscm.2012.04.004.

Preethi S. and Priyadarsini P.L.K (2017) "A Survey on the Applications of K-Nearest Neighbour Algorithm and its variants", Research journal of pharmaceutical, biological and chemical sciences. January-February 2017 ISSN: 0975-8585.

Pulugurta S., Arun A., Errampalli M. (2013) "Use of Artificial Intelligence for Mode Choice Analysis and Comparison with Traditional Multinomial Logit Model", Procedia–Social and behavioral Sciences, Volume 104. https://doi.org/10.106/j.sbspro2013.11.152.

Rahman S.M., Ratrout N.T. (2009) "Review of the Fuzzy Logic Based Approach in Traffic Signal Control: Prospects in Saudi Arabia", Journal of Transportation Systems Engineering and Information Technology 9 (5):58-70, DOI: 10.1016/s1570-6672(08)60080-x.

Ratrout, N.T., Gazder, U., Al-Madani, H.M.N. (2014) "A Review of Mode Choice Modelling Techniques for Intra-City and Border Transport", World Review of Intermodal Transportation Research, Vol. 5, No. 1, DOI: 10.1504/WRITR.2014.065055.

Saini I., Singh D., Khosla A. (2013) "QRS Detection using K- Nearest Neighbour Algorithm (KNN) and Evaluation on Standard ECG Data Basis", Journal of advanced research (2013) 4, 331-344, DOI: 10.1016/j.jare.2012.05.007.

Upadhyay A., Nadar S., Jadhav R. (2020) "Comparative Study of SVM & KNN for Signature Verification", Journal of Statistics & Management Systems ISSN 0972-0510 (Print), ISSN 2169-0014 (Online) Vol. 23 (2020), No. 2, pp. 191–198 DOI: 10.1080/09720510.2020.1724619.

Xu D., Wang Y., Peng P., Beilun S., Deng Z., Guo H. (2018) "Real-time Road Traffic State Prediction based on Kernel-KNN", Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2018.1491073.