**American Journal of Engineering Research (AJER)**
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-11, Issue-03, pp-71-78
www.ajer.org
Research Paper Open Access

# Comparison of ANN and Multinomial Logit for Prediction of Mode Choice between Online Transportations and Private Vehicles

Puguh Budi Prakoso
*Department of Civil Engineering, Faculty of Engineering Universitas Lambung Mangkurat, Banjarmasin, Indonesia*

Yuslena Sari
*Department of Information Technology, Faculty of Engineering Universitas Lambung Mangkurat Banjarmasin, Indonesia*

***Abstract—****The presence of online transportation modes is a manifestation of the rapid development of information technology, especially smartphone and apps technology. Because of the critical issues of safety, service quality assurance and effectiveness of load factor to its capacity, online transportations in Indonesia cannot be classified and regulated as public transportation. The main function of online transportation is actually as a short distance feeder to public transportation. Unfortunately, because the public transportation is not adequate, so the online transportation functions slowly become informal public transportation mode. This in the future can cause problems in a city, especially when the increasing number of online transportations will rise the vehicle traffic together with private vehicles. To understand and prevent this problem becoming bigger in the near future, the mode choice between private vehicles and online transportations requires to be studied first. The purpose of this research is to obtain a reliable model of modal choice between private vehicles and online transportation by comparing the multinomial logit (MNL) probability method with the Artificial Neural Network (ANN), to study the accuracy of the performance of the two models for the mode selection between private vehicles and online transportation, and to evaluate the socio-economic factors underlying the mode choice between private vehicles and online transportation in the city areas of Banjarmasin and Banjarbaru.*
***Keyword:*** *mode choice, multinomial logit probability method, ANN, online transportation*

---

Date of Submission: 02-03-2022 Date of acceptance: 17-03-2022

---

## I. INTRODUCTION

Transportation planning is part of the decision making process of transportation policy to provide the best solution [1]. Transportation planning can also be described as a series of activity processes for selecting and deciding alternative choices of transportation facilities to achieve predetermined goals, by utilizing available resources efficiently[2].

The presence of online transportation is a manifestation of the rapid development of information technology, especially smartphone and apps technology. The online mode of transportation originated from the idea of making it easier for people to travel short distances with certainty of route, time and travel costs. Basically, the main function of online transportation is as a feeder to public transportation.

According to the function and definition, public transportation is a mass transportation. Unfortunately, because public transportation is inadequate in Indonesia, so the online transportation mode has switched its functions as the main public transportation mode. What is more, the online transportation has a lower capacity than mass transportation. This can cause problems in a city in the future, especially when the growing number of online transportations will increase the vehicle traffic together with private vehicles.

To prevent this from happening without control and to have long-term planning, a study is needed to analyze the modal choice between private vehicles and online transportations. In addition to return the online transportation function as a feeder transport, a study is needed on the effect of distance and travel costs on the selection of both modes.

The classical approach to investigate mode selection is by using Multinomial Logit (MNL) method. Nevertheless, due to recent developments in computer algorithm, other computation methods can also be implemented in mode choice prediction, such as Artificial Neural Network (ANN).

This paper will discuss the feasibility and capacity of ANN's ability to model discrete mode choice behavior and compare the performance of ANN predictions with MNL in term of their accuracy. The case taken in this study is from the results of a study regarding the choice of modes between online transportation and private vehicles in Banjarbaru and Banjarmasin, Indonesia.

## II. LITERATURE REVIEW

### *A. Online Transportation*

Online transportation is a type of transportation mode that uses the application as a link between the user and the driver. Online transportation makes it easy for customers because the driver will pick up the customer at a designated place without having to walk in the heat or the rain. Customers are also directly delivered to their destination. However, the use of online transportation is very dependent on the internet network, so if the network has problems the user cannot order public transportation services online [3]–[5].

### *B. Logit Multinomial (MNL)*

Multinomial logit, a logit model that is used to select more than two modes, is also called a multiple logit model. Can be formulated with the following equation:

$$\Pr(i) = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)} \quad (1)$$

Where:

i = Mode (i = 1,..., j)

Pr(i) = Alternative selection probabilities of i

$V_i$ = Deterministic Components of Alternative Utilities i

The multinomial logit model is one of the simplest and most frequently used models for mode choice [6], [7]. This model can be obtained by assuming that the random residues in Equation (1) are distributed with the Gumbel residues that are independently and identically distributed (IID). Logit model is a statistical tool applied for analyzing a situation related to a selection of options. It can be applied for the selection of transportation modes [8].

### *C. Artificial Neural Networks (ANN)*

In contrast to MNL the ANN method consists of three parts, namely the input layer, hidden layer and output layer. Backpropagation learning algorithm is applied in the data processing method. This algorithm utilizes a learning pattern based on the error correction rule. The input vector (X) is entered through the input layer and forwarded to the hidden layer. And finally, the output (Y) is obtained as a response from the artificial neural network. In this process the weights between layers are initially determined randomly. The next process is carried out in the opposite direction (backward) to correct the existing weights using the error correction rule. This process will continue to repeat until the error in the output layer reaches the specified limit. The general form of the function of the artificial neural network in Equation 2:

$$Y = f(\sum x_n w_n) \quad (2)$$

where:

Y : Output

f : Transfer function

Wn : Correspondence weight

Xn : Input

For example, consider the Y neuron in Fig.1:

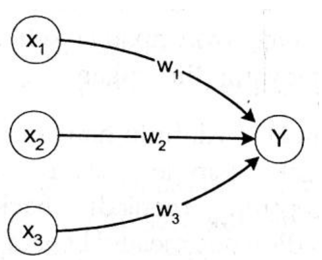


Fig. 1. A Simple Model of Artificial Neural Networks

## III. METHODOLOGY

The method proposed in the study can be seen in Figure 2:

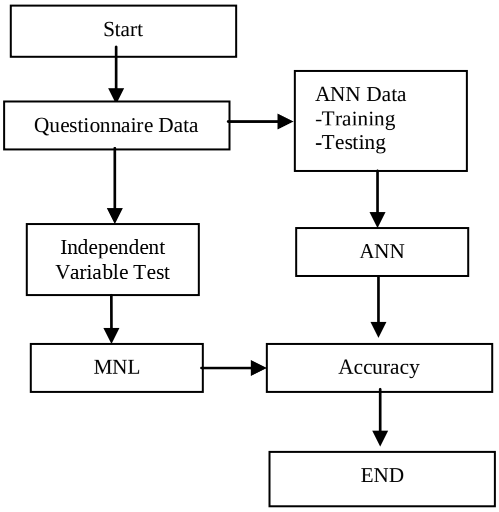


Fig. 2. Purposed Method

### *A. Data*

Primary data were taken using a questionnaire from 200 respondents. The questionnaires are divided into three categories of distances namely short, medium and long distances. From each of these distance categories there are nine variations of questions, so the number of respondents' answers is as many as 5400 answers.

Secondary data is operating costs of private vehicles and tariffs of online transportation. The field survey was conducted in Banjarbaru, Banjarmasin and surrounding areas with respondents being private vehicle users both cars and motorbikes and users of online transportation services with consideration in order to obtain an objective answer to mode selection. The data collection survey is conducted in March to April 2019.

Questions asked include the respondent's socio-economic background, characteristics of users and characteristics of travel. From the survey, there are initially 16 independent variables as they can be seen in Table I.

TABLE I. INITIAL INDEPENDENT VARIABLES

| Variable | Description |
|---|---|
| X1 | Distances |
| X2 | Travel Costs |
| X3 | Ages |
| X4 | Educations |
| X5 | Sex/Gender |
| X6 | Jobs |
| X7 | Numbers of working/studying days/week |
| X8 | Monthly incomes |
| X9 | Numbers of private vehicle ownership |
| X10 | Abilities of driving |
| X11 | Trip purposes |
| X12 | Monthly transportation costs |
| X13 | Origins |
| X14 | Destinations |
| X15 | The use of daily transportation modes |
| X16 | Reasons to use modes |

The choices of transportation modes are Private Motorcycle (PM), Online Motorcycle (OM), Private Car (PC) and Online Car (OC).

For ANN, the data will be divided into two, namely training and testing data, each 50% of the total data.

### *B. Correlation Test of Independent Variables*

Correlation test on independent variables, which is carried out to find out relevant factors is done by using the ANOVA test in Microsoft Excel. The 16 predetermined factors are then tested with the criteria of Significance-F value $\leq 0.05$, p-value $\leq 0.05$ and RSQUARE value which results in the highest correlation.

After several iterations and reductions, the independent variables that are less significant, 7 significant variables are obtained, namely distances, travel costs, ages, gender, and numbers of private vehicle ownership, monthly transportation costs, and the use of daily transportation modes.

### *C. MNL*

Microsoft Office Excel 2013 with the Add-Ins Solver and The Real Statistics Resource Pack [19] and additional Visual Basic Application (VBA) will be used for the iteration process to obtain the utility function coefficients of the seven independent variables that have been tested previously.

The results of multinomial logit regression modeling (MNL) with seven independent variables and OM as a reference mode using Add-Ins Solver and The Real Statistics Resource Pack in MS Excel 2013 can be seen in the following Table II below:

TABLE II. MNL PARAMETERS OF SEVEN INDEPENDENT VARIABLES

| Regression Parameter | | | | | Model Parameter | |
|---|---|---|---|---|---|---|
| Coefficient | PM | PC | OC | OM | | |
| Intercept ($a_0$) | 4.320597 | -14.889 | -65.1141 | Ref | Σ likelihood | -5949.62 |
| $a_1$ | -9.2E-05 | 0.001854 | 0.002053 | Ref | R-Sq (L) | 0.13035489 |
| $a_2$ | -0.30539 | -0.12315 | -0.09164 | Ref | R-Sq (CS) | 0.2812932 |
| $a_3$ | 0.446913 | -0.03867 | -0.6992 | Ref | R-Sq (N) | 0.30553816 |
| $a_5$ | 0.429154 | 0.100393 | 0.121641 | Ref | Chi-Sq | 1783.62965 |
| $a_9$ | -0.14087 | -0.36198 | -0.21217 | Ref | p-value | 0.00 |
| $a_{12}$ | -0.43423 | -0.79377 | -0.09201 | Ref | | |
| $a_{15}$ | 4.320597 | -14.889 | -65.1141 | Ref | | |
| Model Accuracy | | | | | | |
| Every Mode | 81.37% | 33.56% | 18.07% | 36.69% | | |
| Total | 53.67% | | | | | |

It appears that the test results of the model using seven independent variables produce a MNL model that is less accurate. This can be seen from the RSQUARE value and model accuracy. So, by trial error, the number of independent variables is reduced again to get the best results. The final result shows that there are three significant independent variables, namely the distances, travel costs and vehicle ownership, which provide the best mode selection probability estimation model. The results of multinomial logit regression modeling (MNL) with three independent variables using Add-Ins Solver and The Real Statistics Resource Pack [19] can be seen in Table III.

TABLE III. MNL PARAMETERS OF THREE INDEPENDENT VARIABLES

| Regression Parameter | | | | | Model Parameter | |
|---|---|---|---|---|---|---|
| Coefficient | PM | PC | OC | OM | | |
| Intercept ($a_0$) | -3.93122 | 1.81662 | 7.951887 | Ref | Σ likelihood | -1813.161052 |
| $a_1$ | 9.990346 | -38.0933 | -50.3648 | Ref | R-Sq (L) | 0.734973661 |
| $a_2$ | -0.00087 | 0.002638 | 0.003116 | Ref | R-Sq (CS) | 0.84468881 |
| $a_9$ | 0.497639 | 0.435702 | -0.05787 | Ref | R-Sq (N) | 0.91749346 |
| Model Accuracy | | | | | Chi-Sq | 10056.5523 |
| Every Mode | 84.85% | 91.46% | 78.61% | 70.20% | p-value | 0.000 |
| Total | 81.13% | | | | | |

Thus, the mathematical equations of the utility functions as follows:

$U_{PM}$ = -3.93122 + 9.990346$x_1$ - 0.00087$x_2$ + 0.497639$x_9$ (3)

$U_{PC}$ = 1.81662 - 38.0933$x_1$ + 0.002638x2 + 0.435702$x_9$ (4)

$U_{OC}$ = 7.951887 - 50.3648$x_1$ + 0.003116$x_2$ - 0.05787$x_9$ (5)

Where:

$U_{PM}$, $U_{PC}$, $U_{OC}$ = utilities of private motorcycle (PM), private car (PC) and online car (OC)

$x_1$ = variable distance; (1) for short distances ≤ 5 km, (2) for medium distances from 5 to 10 km and (3) for long distances > 10 km

$x_2$ = variable travel costs (fares) in Rupiah

$x_9$ = variable ownership of private vehicles; (1) if one does not have any, (2) if one has 1-3 cars, (3) if one has 1-3 motorbikes, (4) if one has 1-3 motorbikes and cars, (5) if one has more than 3 car and (6) if one has more than 3 motorbikes.

So, the probability models for each mode are:

$$P_{OM} = \frac{1}{1 + \exp(U_{PM} + U_{PC} + U_{OC})} \quad (6)$$

$$P_{PM} = \frac{\exp(U_{PM})}{1 + \exp(U_{PM} + U_{PC} + U_{OC})} = P_{OM} * \exp(U_{PM}) \quad (7)$$

$$P_{PC} = \frac{\exp(U_{PC})}{1+\exp(U_{PM}+U_{PC}+U_{OC})} = P_{OM} * \exp(U_{PC}) \quad (8)$$

$$P_{OC} = \frac{\exp(U_{OC})}{1+\exp(U_{PM}+U_{PC}+U_{OC})} = P_{OM} * \exp(U_{OC}) \quad (9)$$

Where:
$P_{OM}$, $P_{PM}$, $P_{PC}$, $P_{OC}$ = probability of selection of online motorcycle (OM), private motorcycle (PM), private car (PC), and online car (OC)

*D. ANN*

Calculation of prediction of mode selection with ANN method is done using datasets obtained from the questionnaire results. There are also three independent variables applied in ANN, which are the same with the ones used in MNL. With this ANN method, the data undergoes two processes, namely: the first is called the Training Set Process phase, where input data is processed together with their respective weights to produce outputs that are close to the expected output (target, T).

The second phase is called the Testing Process Set, where the input data above is entered back into the mathematical function obtained to obtain a new output. These processes will iterate (loop) so that the actual output is close to or equal to the target output (Y = T).

The chosen mode of the resulting prediction is assumed to be a variable Y, which is an independent variable. The factors that influence the emergence of predictions are; variable input X.

Calculations using the ANN method are performed using the MATLAB software. In this study a combination of the three parameters (independent variables) was tried on the predicted results. The indication of the significance of the influence of the three parameters, namely distance, travel costs and ownership of the vehicle being tested. The best model obtained for network architecture is 4-5-4, shown in the following figure:

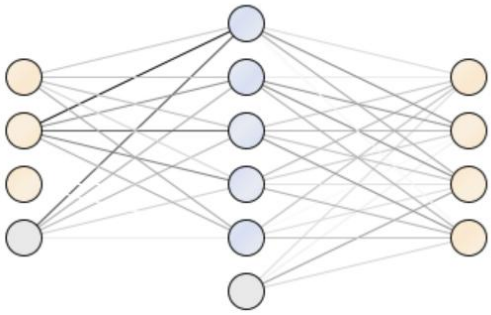

Fig. 3. The Best Architectural Model of ANN

*E. Assessment and Evaluation*

The performance of MNL and ANN approaches to predict the mode choice between online transportation and private vehicle modes will be evaluated through parameters of precision, recall, accuracy, and RSQUARE.

## IV. RESULT AND DISCUSSION

The mode choice of two private vehicle modes and two online transportation modes can be seen in this following Figure 4:

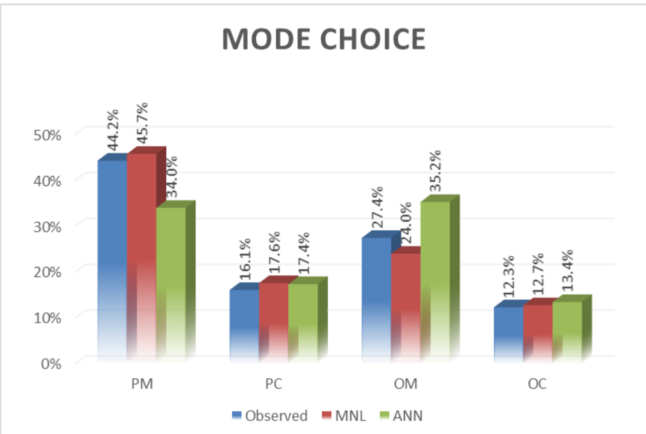


Fig. 4. Results of Mode Choice

To compare the performance between MNL and ANN, evaluation parameters of precision and recall are estimated from confusion matrix. The following is a confusion matrix table with mode prediction modeling calculated using the MNL method:

TABLE IV. CONFUSION MATRIX MODEL MNL

| | true 4 | true 3 | true 2 | true 1 | Class Precision |
|---|---|---|---|---|---|
| pred. 4 | 409 | 0 | 0 | 0 | 100.00% |
| pred. 3 | 0 | 405 | 0 | 85 | 82.65% |
| pred. 2 | 44 | 51 | 622 | 0 | 86.75% |
| pred. 1 | 70 | 214 | 0 | 800 | 73.80% |
| Class Recall | 100.00% | 81.98% | 100.00% | 100.00% | Accuracy 82,81% |

The following table is a confusion matrix modeling prediction mode which is calculated using the ANN method:

TABLE V. CONFUSION MATRIX MODEL ANN

| | true 4 | true 3 | true 2 | true 1 | Class Precision |
|---|---|---|---|---|---|
| pred. 4 | 141 | 0 | 0 | 0 | 100.00% |
| pred. 3 | 0 | 810 | 0 | 0 | 100.00% |
| pred. 2 | 0 | 0 | 245 | 1 | 99.59% |
| pred. 1 | 0 | 0 | 0 | 1503 | 100.00% |
| Class Recall | 100.00% | 100.00% | 100.00% | 99.93% | Accuracy: 99,96% |

It can be seen from Table IV and V that the accuracy analyzed from confusion matrix of ANN is superior to one of MNL, namely 82.81% against 99.96%.
Comparison of the accuracy and RSQUARE values according to output of MNL and ANN compared to the observed values is shown in the following Figure 5:

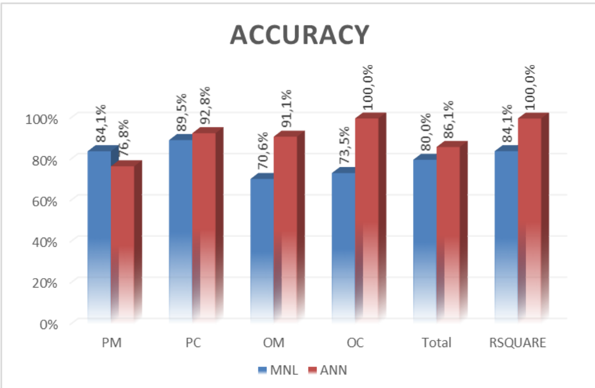


Fig. 5. Results of Mode Choice

From the results of validation based on output accuracy, it can be seen that the MNL model outperforms the ANN model, namely 84.1% and 76.8%. Whereas PC, OM and OC MNL were outperformed by ANN. On the MNL, PC has an accuracy of 89.5% and an accuracy of ANN 92.8%. In OM, ANN accuracy of 91.1% outperformed MNL accuracy of 70.6%. Whereas at OC, ANN accuracy reaches 100% and MNL accuracy is only 73.5%.
The average total accuracy of MNL was 80% and ANN accuracy was 86.1%. For the RSQUARE value, ANN is superior to MNL, namely the MNL RSQUARE value of 84.1% and ANN of 100%.
By considering ANN as more accurate model than MNL, the influence of the distance to the mode selection can be analysed using ANN, as it is depicted in Figure 6.

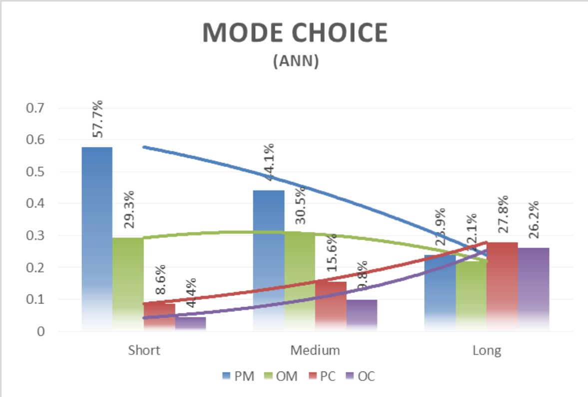


Fig. 6. Influence of Distance to Mode Choice Predicted with ANN Model

Some points, which can be figured out from Figure 6 are: (1) the probability of using private motorcycle (PM) will be reduced when the distance is increased, (2) online motorcycle (OM) gets the higher choice in the medium distance between 5 to 10 km, (3) the probabilities of selecting private car (PC) and online car (OC) are raised when the distance is increased.

From the results shown in the Figure 6, it can be concluded that online motorcycle (OM) is selected more by the users for short to medium distance trips. This also means that the function of OM is potential to be managed as a feeder for public transportation, when public mass transportation modes will be provided in the future. Meanwhile, the online car (OC) can still be an option for medium to long distance trips more than 10 km, competing the private car (PC).

## V. CONCLUSION

The research that has been carried out demonstrates the results that the ANN model achieves better prediction accuracy (86.1%) than the MNL model (80.0%). In particular in the confusion matrix, ANN accuracy is better than MNL. It can be concluded that the ANN method has a better level of precision than the MNL method.

Based on the results of the analysis of the two models, it is evident that the factors of distance, cost and number of vehicle ownership affect the choice of the transportation modes.

From the mode choice composition, it can be seen as well that online motorcycles, which are now very popular as transportation modes are desired by the users for short to medium distance trips. This can be an important factor of this mode's characteristic, which is potential to be managed as feeder for mass transportation mode when it is available in the near future.

For future research, sensitivity analysis can be exploited as an option to examine the impact of variables and to compare the results by the model for different variables: cost and distance. ANN is highly potential to be combined with other algorithms for classifying the modes, which can be useful for better and more accurate prediction of mode choice probability.

## REFERENCE

[1]. R. Corporation, "Transportation Planning, Climate Change, and Decisionmaking Under Uncertainty," *Transportation (Amst).*, pp. 1–26, 2008.
[2]. A. Abe, "Transport Operations Management," 2014.
[3]. D. O. T. F. T. Administration, "Shared Use Mobility , Transportation Technology and Intercity Transit Services," no. June, 2018.
[4]. S. Shaheen, E. Martin, A. Cohen, A. M. Abhinav, and A. B. Tsrc, "Mobile Apps and Transportation : A Review of Smartphone Apps and a Study of User Response to Multimodal Traveler Information," no. October, 2016.
[5]. T. Document, "Transportation Demand Management," no. April, 2009.
[6]. M. Srivastava and C. R. Sekhar, "Web Survey Data and Commuter Mode Choice Analysis Using Artificial Neural Network," *Int. J. Traffic Transp. Eng.*, vol. 8, no. 3, pp. 359–371, 2018.
[7]. H. Omrani, "Predicting travel mode of individuals by machine learning," *Transp. Res. Procedia*, vol. 10, no. July, pp. 840–849, 2015.
[8]. S. Indriany, A. Widyantoro, and I. W. W, "Analisis Pemilihan Moda Dengan Model Multinomial Logit Untuk Perjalanan Kerja Dari Kota Tangerang Selatan-Dki Jakarta," *Portal J. Tek. Sipil*, vol. 10, no. 1, pp. 24–32, 2019.
[9]. Z. K. Gurmu and W. D. Fan, "Artificial neural network travel time prediction model for buses using only GPS data," *J. Public Transp.*, vol. 17, no. 2, pp. 45–65, 2014.
[10]. B. Gültekin Çetiner, M. Sari, and O. Borat, "A neural network based traffic-flow prediction model," *Math. Comput. Appl.*, vol. 15, no. 2, pp. 269–278, 2010.
[11]. T. Pamuła, "Neural networks in transportation research-recent applications," *Transp. Probl.*, vol. 11, no. 2, pp. 27–36, 2016.
[12]. L. Detwiler and L. Detwiler, "Using Backpropagation Neural Networks for the Prediction of Residual Shear Strength of Cohesive Soils Using Backpropagation Neural Networks for the Prediction of Residual Shear Strength of Cohesive Soils Thesis Review Committee :," 2014.

[13]. J. V. Pérez-Rodríguez, S. Torra, and J. Andrada-Félix, "STAR and ANN models: Forecasting performance on the Spanish 'Ibex-35' stock index," *J. Empir. Financ.*, vol. 12, no. 3, pp. 490–509, 2005.

[14]. S. Pan and K. Duraisamy, "Data-driven discovery of closure models," *SIAM J. Appl. Dyn. Syst.*, vol. 17, no. 4, pp. 2381–2413, 2018.

[15]. B. Duan, L. Han, Z. Gou, Y. Yang, and S. Chen, "Clustering Mixed Data Based on Density Peaks and Stacked Denoising Autoencoders," *Symmetry (Basel).*, vol. 163, no. 11, pp. 1–23, 2019.

[16]. C. Saranya and G. Manikandan, "A Study on Normalization Techniques for Privacy Preserving Data Mining," *Int. J. Eng. Technol.*, vol. 5, no. 3, pp. 2701–2704, 2013.