



INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv



Predictive Algorithms Analysis to Improve Sustainable Mobility

Oscar Dario León-Granizo ^{a,*}, Miguel Botto-Tobar ^{b,c}

^a Department of Mathematics and Physical Science, University of Guayaquil, Ecuador

^b Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands

^c Research Group in Artificial Intelligence and Information Technology, University of Guayaquil, Ecuador

Corresponding author: *oscar.leong@ug.edu.ec

Abstract— In this work, a comparative analysis of 3 prediction algorithms (Linear Regression, Neural Networks, and KNN) was carried out, which allows for studying georeferential coordinates of moving objects. Through an exhaustive study, it will be possible to know the predictions of each algorithm, which will make a comparison of results. This will help implement an algorithm with greater accuracy and effectiveness in a system developed as a research project called Intelligent System for Sustainable Mobility of the University of Guayaquil (SIAMS-UG), using open source tools that allow working with Machine Learning. It will be possible to analyze the forecasts of the congestions that are formed in the surroundings of the University of Guayaquil. This problem generates inconveniences for both students and administrative staff that are part of this institution. The methodology used for the project's development was the waterfall methodology, as it is a linear and simple implementation model where each phase of the project was emphasized, allowing the management of the results and the successful completion of the project.

Keywords— Analysis; algorithms; mobility; open source; prediction.

Manuscript received 15 Oct. 2021; revised 17 Jan. 2022; accepted 2 Mar. 2022. Date of publication 31 Mar. 2022.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

In the routine traffic that people carry out, it is common to find mobility problems. This has consequently delayed the work activities of many people. In Guayaquil, there is a high degree of inconvenience with vehicular congestion, which causes environmental contamination damaging the city's sustainability. It is novel and interesting to be able to anticipate vehicular flow problems to offer optimal solutions.

The present work is based on a comparative analysis of three predictive modeling algorithms focused on sustainable mobility, applied in the vicinity of the University of Guayaquil, to find the algorithm with the most accurate certainty probabilities that will be implemented as a module. The Intelligent Analysis System of the University of Guayaquil will allow students and administrative staff of the university to have an option that allows them to predict at what times or situations there may be high vehicular flow through the city's main roads. The data was obtained through a special investigation in which empirical documents were considered that evidenced some related works in recent years and managed to reflect similar themes that will guide finding suitable references for this study.

The growth of the automobile fleet, whether private or public, is creating massive chaos in terms of mobility since, on several occasions, they manage to saturate the roads within the city [1]. This is one of the reasons it is intended to avoid the negative effects of the excessive number of cars that increase day by day, causing environmental and social problems due to the massive vehicular flow and traffic jams, for which a model called “sustainable mobility.”

Many prediction algorithms allow finding a precise probability. For this analysis, only three "Linear Regression, Neural Networks, and KNN" were selected with these algorithms. The pertinent studies were carried out that compared their result, obtaining the necessary data to reach a decision, and determining the answers to the problem that encompasses this project [2].

The information collection was carried out through a thorough and detailed exploration where empirical documents were considered that show some works of recent years related to the subject. These reflect similar themes that served as a guide to obtaining this project's relevant reference.

The article "Prediction of traffic conditions through machine learning", performs a procedure in which they had to determine the possible states of traffic through machine learning, for which they used tools such as generic devices

and simulation of a laser sensor. However, they also used three prediction algorithms, Bayer-Navier Classifier, Decision Tree and Neural Networks, which allowed them to reach an analysis with a qualitative methodology using data extraction in frames, which is more complex for getting the necessary prediction [3].

The article "Analysis of vehicular congestion to optimize the transport system through artificial neural networks". It is based on an investigation with which they sought to reduce vehicular traffic by building an artificial intelligence model. They focused on performing this procedure with the neural network algorithm. The objective was to know the prediction to optimize this process. The researchers used ten input variables that served them for the training that would lead them to reach the best value of Least-squares. These assessments were made in two sample groups, in this case, the first training (training) and the second validation (test) of the data extracted from the interactions [4].

To analyze the results, it was decided to divide the data as follows: 64% of the data was used in the training process, and 36% of the data for validations and tests, thus completing 100% of the information. Once the correct processes were carried out, they had an answer to their problem, where the training gave them a quadratic value $R=0.99976$, in the validation $R=0.99974$, and the test $R=0.99977$, in such a way that they were able to obtain a model of the Neural Network with $R=0.99976$, this being the result of its prediction [5].

Nowadays, it is inevitable to ignore the traffic congestion suffered by some areas of Guayaquil, including the university citadel, since it is one of the areas affected by the amount of urban transport or people who circulate daily in its surroundings, causing severe environmental and social problems.

II. MATERIALS AND METHOD

A. Non-experimental Research

For this project, the type of non-experimental research was used because the analysis to be carried out does not intend to modify its variables. Rather, only the variables will be measured to know their probability without manipulating any of them, and it is non-experimental research of cut cross-sectional because the data studied will be done at the same time without having to make comparisons with data from any previous stage [6].

B. Statistical Population

To characterize the population to be researched in this research, the database "Mobility" was used, which provides significant features such as vehicular trajectories, which are critical in representing the issue to be handled generically. The number of records in the database, which we will use as the study population for this research, is shown below:

TABLE I
POPULATION OF VEHICLE

Population	N°
Vehicle trajectories	201436
Total	201436

C. Statistic Sample

Since the executions of each algorithm meant a high processing cost which prevented obtaining the results efficiently, it was decided to work with three data packages, each one organized with a sample of 3000 random records that allowed obtaining the results. This guaranteed to obtain the results accurately [7].

TABLE II
RESEARCH SAMPLE

Population	N°	Research sample	N°
Vehicle trajectories	201436	Established sample	3000
Total	201436	Total	3000

D. Pearson Correlation Analysis

Through Pearson's correlation, it is verified how related the variables of the analyzes proposed for the research area, since according to [8] "It serves as a basis for the fulfillment of two main objectives, comparing groups and studying relationships". In this way, the connection they present will be more fully understood.

For the determination coefficient, our Pearson correlation result must be squared, and, in this way, the level of weakness in the linear relationship of X and Y will be known [9]. Once we have the Pearson coefficients, we proceed to obtain the hypothesis tests, which are defined as:

If $p < 0.05$, then H_0 is rejected, and H_1 is accepted

If $p > 0.05$, then H_0 is accepted and H_1 is rejected

$H_0 p = 0$ means that there is a correlation between the variables

$H_1 \neq 0$ means that there is no correlation between the variables

To reach the acceptance and rejection of hypotheses, the following formula must be applied:

1) Test Statistic:

$$t = \sqrt{\frac{n-2}{1-r^2}} \quad (1)$$

2) Critical Value

n = sample

$gl (n-2)$ = Degree of freedom

$\alpha = 0.05$ (significance level)

$t (/2, n-2)$ = Critical value

For the determination coefficient, our Pearson correlation result must be squared and, in this way, the level of weakness in the linear relationship of X and Y will be known [10]. Next, the analysis of the variables with which the study is carried out is established. For the first analysis, a Dataset containing 2 significant variables, an X (temperature) and a Y (duration), both with 3000 records, is considered.

3) Hypothesis Testing

- H_0 : The temperature variable is related to the duration variable.
- H_1 : The temperature variable is related to the duration variable.

Once having this result, the hypothesis test is achieved by obtaining p is greater than $\alpha (0.3648 > 0.05)$, given this result

we proceed to accept H_0 and reject H_1 . Therefore, it means that if there is a relationship between the variable temperature and variable duration. For the second analysis, a Dataset containing two significant variables, an X (hour) and a Y (duration), both with 3000 records, is considered.

E. Variables

1) *Independent variable*: Measure the indicators of the probability of certainty and margin of error of the algorithms' predictions.

2) *Dependent variables*: Analysis of predictive modeling algorithms (linear regression, Neural Networks, KNN) to measure situations that can cause traffic congestion.

F. Research Background

1) *Traffic congestion*: Kan [11] defines congestion as "the action and effect of congesting or becoming congested", while "congesting" means "obstructing or hindering the passage, circulation or movement of something", in this case, road obstructions due to traffic jams due to high vehicle growth.

2) *Programming Language used in the project*:

- Python. It is an easily understood programming language for the introduction of study in this area of learning. Python is a simple, fast, and light language and is ideal for learning, experimenting, practicing, and working with machine learning, neural networks, and deep learning [12]. If we collect the advantages of the Python language, we could define it as a language that has everything for learning. Subasi, as mentioned [13]: Python is a multiparadigm language in which imperative, functional and object-oriented aspects coexist natively. These paradigms are very well decoupled, which allows entry to the language to be made progressively starting, for example, with an imperative style and later including functional and object-oriented elements.
- Anaconda Python. Anaconda Python is a data analysis distribution that includes many libraries and packages. Highlighting some of its features:
 - a) Enables you to manage and implement Python packages, dependencies, and an ecosystem for the rise of Data Science.
 - b) It includes programming environments such as Spyder, RStudio, and Jupyter and data analysis tools such as Numba.
 - c) Allows users to have access to more advanced learning resources.
- Spyder. As a development environment, Spyder is an open-source environment made in Python. Designed for subject matter experts such as analysts and data scientists [14].
- Libraries for data analysis. The implementation of the algorithms for the analysis will depend a lot on the types of libraries that are used by the features and functions they provide. Libraries involved:
 - a) Matplotlib: Python standard library for graphing statistical data results [15].

- b) NumPy: Besides providing mathematical functions, it contains a universal data structure that facilitates data analysis and exchanges involved in algorithms [15].
- c) Pandas: It reads a large amount of data, the data structure developed in this library leads us to manipulate the data in one and two dimensions [16].
- d) Scikit-learn: Designed for machine learning. The executions of algorithms to perform supervised and unsupervised learning through problems such as regression, and classifications can be developed by Scikit-learn [17].

G. Least-squares

To be able to measure more accurately and estimate certain parameters away from the differences in the results concerning a real value, the least-squares method aims to minimize the sum of possible errors squared. With the following equation:

$$Q = \sum_{i=1}^n (Y_i - \beta_0 + \beta_1 * X_i)^2 \quad (2)$$

It can be determined that this method seeks to take the line of best fit through the graph with the data related to the study.

H. Methodological Design of the Research

To meet the first objective, the analysis of the database was carried out, the steps followed for the study of the data are shown below:

1) *Backup copy of the "Mobility" database*: As a first step, the backup of the "Mobility" Database was carried out, which is part of a larger research project, originally this database is hosted in the services of AWS (Amazon Web Service) and has information of the trajectories of the objects in the movement of certain days and hours in which the data collection was carried out [18].

2) *Restore data in PostgreSQL*: We proceeded to execute the restoration of the data contained in the "Mobility" database directly in the PostgreSQL administration platform (PgAdmin4) that will later serve us to develop the dataset according to the analyzes determined for this investigation.

3) *Analysis of the restored database*: Once the restoration was carried out, each part of the data was verified in detail, both its tables, its fields, and its relationships to proceed to choose which of them would be specifically worked within the study of the predictions of the flow vehicle and thus improve sustainable mobility.

4) *Data training*: Once the previous steps have been developed, the data is cleaned, eliminating possible fields with 0 or null records, avoiding any bias of unwanted values. Therefore, 3 Datasets are created that contain the necessary information for the study of the three analyzes that will be considered to carry out the investigation [19]. After the creation of the datasets, the execution and training of the three prediction algorithms (Neural Networks, Linear Regression, and KNN) began, thus managing to measure for each model the probabilities and margin of error, which will be compared to later determine which is the best.

5) *Dataset 1*: This dataset called settemperature created from the relationship of the `inf_tray`, `inf_tray_det`, and `gen_climas` tables, stores a randomly produced list of 3000 records for each field (temperature, duration). The classification field is generated from these fields, which will be very useful later. This dataset is very important to perform the data training in the three algorithms from which the predictions of the analyses established in this work will be obtained.

6) *Dataset 2*: The second dataset also contains 3000 records for each field (distance, duration) taken at random, this is designated by the name `settime`. To create this dataset, only the `inf_trayectoria_det` table was needed, which contained the necessary fields for the study (distance, duration) that were very useful to determine the third field identified as classification, after this the instructions will be executed in every algorithm.

7) *Dataset 3*: Finally, in the last dataset named `sethora`, 3000 records were also stored that were saved randomly, two fields were entered (time, duration), the first column is a derivation of the date field located in the `inf_trayectoria_det` table, the second column (duration) also belongs to the same table, thanks to this it was possible to create the third column called classification after having the package established, the training for each predictive model is performed.

The same amount was taken for training and test data for all analyses.

TABLE III
DATA DISPOSITION

Availability for training and testing	Data
Training data	(70%) 2100
Test data	(30%) 9000
Total	(100%) 3000

I. Explanation of Each Proposed Analysis

1) *Analysis 1*: What is the probability and margin of error that, depending on the temperature, the duration of the trajectories are high or low so that it can be predicted if there is traffic congestion? For this study, the temperature was considered because it refers to the climate. It was taken into account that one of the reasons why there is congestion is due to changes in the environment since there may be the possibility of rain and the roads become somewhat dangerous for them, the vehicles will take longer to complete their trajectories. , based on this analysis, it is decided to predict that, depending on the temperature, the possible durations that moving objects will use to complete their journey will be known [20].

2) *Analysis 2*: What is the probability and margin of error that the durations of the trajectories are high or low depending on the distance traveled from the point of origin to the point of arrival? It originated because it is very feasible to predict how long it will take depending on the distance it will travel from the starting point to the destination where it is intended to arrive since, if the probability of the prediction is high, it means that within the distance to travel more vehicles are making the trajectories. Therefore, a large vehicular influx can be estimated. Otherwise, if the duration is shorter, the places where these distances will be traveled will not have

much vehicular flow. Thus, it will be possible to determine if there is congestion.

3) *Analysis 3*: What is the probability and margin of error that, depending on the day's hours, the trajectories' duration is high or low to predict if there is vehicular congestion? The following query arose because one of the problems of a large amount of vehicular movement occurs according to the hours in which they travel. This is due to the needs of people when they have to travel by some means of transport to their jobs. , educational institutions or to manage any activity they wish to carry out, regularly there have always been specific times when mobility is complicated, so it is necessary to predict exactly at what times the journey times will be a little higher or lower to determine if there will be congestion on the roads surrounding the University of Guayaquil.

J. Analyses Executed on the Algorithms

1) Linear Regression:

The simple linear regression algorithm was used to carry out the proposed analyses since there are two unique variables for each study presented, one dependent variable and one independent. The analysis of set temperature resulted in the probability that the predictions obtained are accurate was 0.007008 with a margin of error of 0.992991.

```
LinearRegression()
[0.76783194]
y=0.767831935494548*x+-17.06525385329434
Coefficient : [0.76783194]
Temperature [26.1] [26.1] [24.4] [26.1] [26.1]
Prediction [2.97515966 2.97515966 1.66984537 2.97515966 2.97515966]
Probability of certainty (r^2): 0.007008106433558314
Margin of error : 0.9929918935664417
```

Fig. 1 Linear regression results

Within the responses obtained, the first five data of the predictions achieved through the linear regression model are displayed, the same information that is used for comparison with the real data. As can be seen in fig. 2, having unadjusted data, forecasts about durations that are not as effective can be verified with the established temperature and time data.

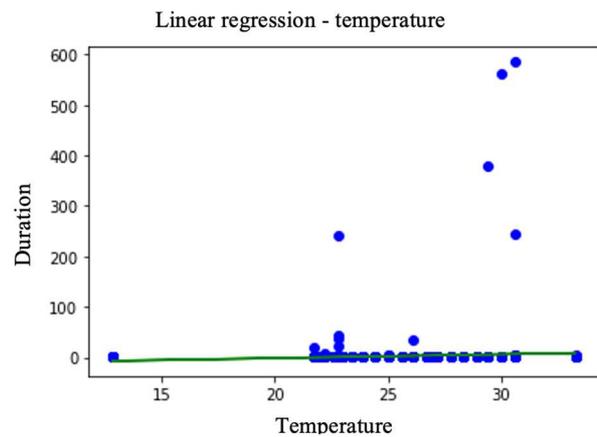


Fig. 2 Linear regression result – temperature

As can be seen in the graphs, the relationship of the data is very dispersed and does not consider the high points with greater intensity and only focuses on the nearby points, which is where the linear regression predictive model is created that

works with the function of root mean square error cost that focuses on minimizing the sum of squares (fewer errors).

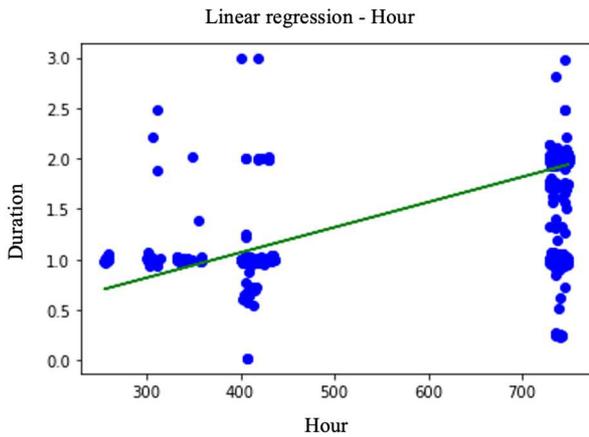


Fig. 3 Linear regression result - Hour

Observing the evident mismatch in the study data was taken as an alternative to verify it; through an adjustment, there can be a perfect combination between the parameters, in this case, the dependent and independent variable, and thus confirm that the probabilities of predictions given by the smallest number of squared errors can become more precise.

As can be seen, linear regression is a model that depends a lot on a close relationship of its variables X, Y, as verified in the analyzes carried out since the first 2 Datasets do not have variables so connected that the probabilities of certainty of the predictions. They gave very low, unlike the settiempo data set with an absolute relationship in its variables, which presented a higher probability of certainty than the previous ones.

To obtain better results, the above must be considered. For predictive analysis with this type of algorithm, it is advisable to consider the data that best fit this model. It is preferable to eliminate possible aberrant data because they greatly influence the alteration of the result [21].

2) Neural Networks

As can be seen, Linear regression is a model that depends a lot on a close relationship of its variables X, Y as verified in the analyses carried out since the first 2 Dataset not have variables. So, connected the probabilities of certainty of the predictions gave very low, unlike the set of data settiempo that has an absolute relationship in its variables which presented a higher probability level of certainty than the previous ones [22].

```

Redes neuronales
Hour [732] [738] [735] [737] [748]
Prediction [1.91705245 1.93206076 1.9245566 1.92955937 1.95707461]
Probability 0.6303159840925454
Margin of Error : 0.3696840159074546
    
```

Fig. 4 Neural networks result

In this new network model of 15 neurons by the 2 hidden layers, the result of the predictions had a slightly more favorable result. Within the analysis by temperature result with a probability of certainty of 0.7278 and margin of error of 0.2721 and in the analysis of time a probability of certainty of 0.6303 and margin of error of 0.3696.

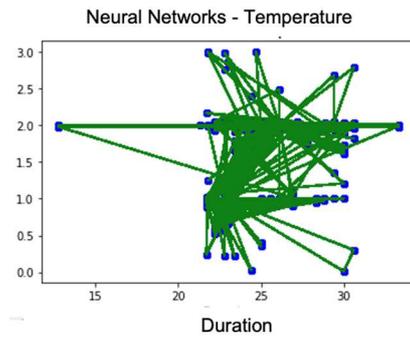


Fig. 5 Neural networks result - temperature

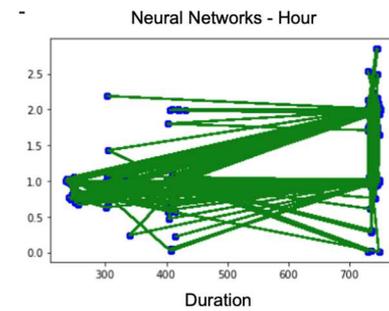


Fig. 6 Neural networks result - hour

The experiments carried out in the Neural Networks algorithm decided to show only the first five records of the dataset to help verify the results. When comparing them, it is visualized that the predictions are very close to reality, and the probability of certainty was very good, demonstrating that to have success in the forecasts, it is essential to work with a more adjusted data distribution since, in this way, it is possible to train efficiently [23]. Experiment with settiempo: In the study of settiempo analysis, it was presented answer that the probability that the predictions shown are successful is 0.981 with a margin of error of 0.01893.

K. K-Nearest-Neighbor (KNN)

To cover each of the analyses, with the KNN algorithm, the same procedures of the two algorithms analyzed above are determined, but in this case adding a characteristic of the algorithm. A classification field is required to carry it out that allows the algorithm to interpret and solve the problems raised in each study.

The analysis of the environment at setthora carried out in KNN showed as a response in the execution that the probability that the predictions obtained according to the analysis are true is 0.98865, corresponding to the average duration within the established interpretation classification. It maintains a similar probability thanks to the classification criteria from which the data was established.

```

clasificacion
1.0 402
2.0 1291
3.0 1288
dtype: int64
K-MN Classifier Hora
Prediction => [3. 3. 3. ... 3. 3. 3.]
Probability | => 0.9865951742627346
Margin of Error => 0.013404825737265424
    
```

Fig. 7 Classification result

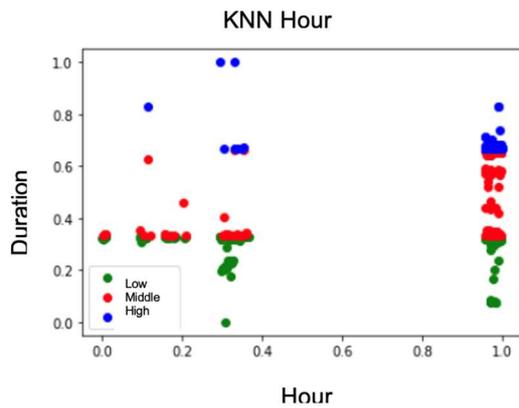


Fig. 8 Classification result - hour

In each of the results, good results are observed in terms of the probabilities of certainty and margin of error with the K-NN algorithm. Based on that, it should be noted that the predictions were 100% accurate because a well-distributed classification criterion was adopted. As this is an algorithm that works with classification where a categorical variable of integer values is required, it was necessary to proceed to create them for their execution and analysis.

Once all the results were collected after carrying out the pertinent analyses and experiments, it was found that the predictive algorithm with the greatest precision is the Neural Network model, unlike Linear Regression and KNN, which presented the best probabilities and the least margins and error. In total, nine executions were carried out, all of them contributed to the evaluation and determination of the optimal algorithm for this type of analysis.

TABLE IV
RESULTS OF THE INVESTIGATION

Analysis	Linear regression		Neural networks		KNN	
	Probabil-ity of certainty	Marg-in of error	Probabil-ity of certainty	Marg-in of error	Probabil-ity of certainty	Marg-in of error
settemperat						
ura	0,36993	0,630	0,7278	0,272	0,71066	0,289
Sethora	0,5806	0,419	0,6364	0,363	0,68	0,32
Settiempo	0,8651	0,134	0,981	0,018	0,94	0,06

III. RESULTS AND DISCUSSION

After the experimentation process with the analyzes executed in each algorithm, the comparison is generated using a general box, as shown in table 4. It was verified that the algorithm that presented the best results was that of Neural Networks, reaching more satisfactory probabilities for the analysis. With the analysis to measure the vehicular flow through a change in the environment, a value close to 0.73% probability of success was obtained. The margin of error of 0.27%, a probability of 0 is also reached 63% for success. As a result, the vehicular flow depends on a certain hour type, with a margin of error of 0.36%. Finishing the reading of the obtained results to analyze the time that a trajectory would take to get from one point to another. A probability of 0.98% was obtained with a margin of error of 0.02%, determining it in this way as the optimal algorithm since it managed to achieve the best probabilities of certainty and the lowest margins of error. For the development of this research, 3

relevant factors were considered that are essential to consider when knowing if there is vehicular congestion near the university zone in this area. In this case, it was decided to work with the temperatures, distances, and times of day in which this problem could occur. The time it would take for a trajectory to complete its journey was predicted for each of them. As a final point, the Neural Networks algorithm was implemented in the Intelligent Systems for Sustainable Mobility of the University of Guayaquil to help students and administrative staff make inquiries about forecasting whether there is abundant vehicular flow in the surroundings. From the educational center and thus be able to take other alternate routes to be able to move and arrive at an optimal time to fulfill their functions and in the same way contribute to sustainable mobility.

IV. CONCLUSION

Through the random processing of the 201,436 vehicle trajectory records for the implementation and analysis of the Linear Regression, Neural Networks, and KNN algorithms, the results of each experiment were obtained according to each exposed situation. This study compared the results of the probabilities of each analysis. The neural network algorithm (with a multilayer model: 2 hidden layers of 15 neurons each) is defined as the most accurate based on the records of the trajectories collected by the Intelligent System for Sustainable Mobility of the University of Guayaquil.

It is determined that the operation of a neural network through supervised machine learning can be adjusted thanks to the experience of data processing and the results obtained. The configuration of the model and the correlation of the data determine the algorithm's efficiency. Implementing the predictive analysis module in the Intelligent Analysis System for Sustainable Mobility of the University of Guayaquil has the three prediction analyses that influence the vehicular flow studied during this project. The probabilities for each predictive analysis and the georeferenced coordinates related to the results are obtained.

REFERENCES

- [1] J. Aráuz, "Supporting Mobility Planning in Small Cities and Communities with Low-power, Machine Learning Based Sensing," *Procedia Comput. Sci.*, vol. 185, pp. 109–118, Jan. 2021, doi: 10.1016/J.PROCS.2021.05.012.
- [2] Y. Xia, W. Pang, and X. Zhang, "Mining relationships between performance of link prediction algorithms and network structure," *Chaos, Solitons & Fractals*, vol. 153, p. 111485, Dec. 2021, doi: 10.1016/J.CHAOS.2021.111485.
- [3] M. Vázquez, R.; Fernández, J. M.; Simón, F.; Ing, and A. Olivetti, "Traffic Status Prediction using Machine Learning," 2018.
- [4] H. De La Cruz Lopez and V. Jurado Mamani, "Analysis of traffic congestion to optimize the transport system through artificial neural networks," 2019.
- [5] K. Nanath, S. Balasubramanian, V. Shukla, N. Islam, and S. Kaitheri, "Developing a mental health index using a machine learning approach: Assessing the impact of mobility and lockdown during the COVID-19 pandemic," *Technol. Forecast. Soc. Change*, vol. 178, p. 121560, May 2022, doi: 10.1016/J.TECHFORE.2022.121560.
- [6] D. Patiño Pérez, C. Munive Mora, L. Cevallos-Torres, and M. Botto-Tobar, "Predicción de la Efectividad de las Pruebas Rápidas Realizadas a Pacientes con COVID-19 mediante Regresión Lineal y Random Forest," *Ecuadorian Sci. J.*, vol. 5, no. 2, pp. 31–43, 2021, doi: 10.46480/esj.5.2.108.
- [7] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: A survey of dataset development and use in machine learning research," *Patterns*, vol. 2, no. 11, p. 100336, Nov.

- 2021, doi: 10.1016/J.PATTER.2021.100336.
- [8] D. Edelmann, T. F. Móri, and G. J. Székely, “On relationships between the Pearson and the distance correlation coefficients,” *Stat. Probab. Lett.*, vol. 169, p. 108960, Feb. 2021, doi: 10.1016/J.SPL.2020.108960.
- [9] C. P. Kuo and J. S. Fu, “Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions,” *Sci. Total Environ.*, vol. 758, p. 144151, Mar. 2021, doi: 10.1016/J.SCITOTENV.2020.144151.
- [10] D. Yacchirema and A. Chura, “SafeMobility: An IoT- based System for safer mobility using machine learning in the age of COVID-19,” *Procedia Comput. Sci.*, vol. 184, pp. 524–531, Jan. 2021, doi: 10.1016/J.PROCS.2021.03.066.
- [11] Z. Kan, M. P. Kwan, D. Liu, L. Tang, Y. Chen, and M. Fang, “Assessing individual activity-related exposures to traffic congestion using GPS trajectory data,” *J. Transp. Geogr.*, vol. 98, p. 103240, Jan. 2022, doi: 10.1016/J.JTRANGE.2021.103240.
- [12] N. Gray, S. Ferson, M. De Angelis, A. Gray, and F. Baumont de Oliveira, “Probability bounds analysis for Python,” *Softw. Impacts*, vol. 12, p. 100246, May 2022, doi: 10.1016/J.SIMPA.2022.100246.
- [13] A. Subasi, *Practical Machine Learning for Data Analysis Using Python*. Academic Press, 2020. doi: 10.1016/B978-0-12-821379-7.00001-1.
- [14] R. G. McClarren, *Getting Started in Python*. Academic Press, 2018. doi: 10.1016/B978-0-12-812253-2.00002-9.
- [15] H. Belyadi and A. Haghighat, *Introduction to machine learning and Python*. Gulf Professional Publishing, 2021. doi: 10.1016/B978-0-12-821929-4.00006-8.
- [16] H. Izadkhah, “An introduction of Python ecosystem for deep learning,” in *Deep Learning in Bioinformatics*, Academic Press, 2022, pp. 31–66. doi: 10.1016/B978-0-12-823822-6.00010-X.
- [17] O. León-Granizo and M. León-Granizo, “Desarrollo de un asistente virtual (chatbot) para mejorar el acceso a la información recurrente por los estudiantes de Instituciones de Educación Superior,” *Ecuadorian Sci. J.*, vol. 4, no. 2, pp. 111–116, 2020.
- [18] L. Cevallos-Torres, M. Botto Tobar, A. Díaz Cadena, and O. León-Granizo, “Decision making in inventory control by using artificial neural networks,” *Sustain. Eng. Innov.*, vol. 4, no. 1, pp. 66–75, 2022, doi: 10.37868/sei.v4i1.id150.
- [19] D. Mechergui and P. Jayakumar, “Efficient generation of accurate mobility maps using machine learning algorithms,” *J. Terramechanics*, vol. 88, pp. 53–63, Apr. 2020, doi: 10.1016/J.JTERRA.2019.12.002.
- [20] V. Papatathanasopoulou, C. Antoniou, and H. N. Koutsopoulos, “Data-Driven Traffic Simulation Models: Mobility Patterns Using Machine Learning Techniques,” *Mobil. Patterns, Big Data Transp. Anal. Tools Appl. Model.*, pp. 263–295, Jan. 2019, doi: 10.1016/B978-0-12-812970-8.00011-7.
- [21] G. Blanc-Pihuave, L. Cevallos-Torres, and J. Arteaga-Vera, “Modelo computacional de clasificación de aprendizaje de máquina supervisado, para el análisis de datos cardiovasculares y pronóstico médico,” *Ecuadorian Sci. J.*, vol. 4, no. 2, pp. 71–79, 2020, doi: 10.46480/esj.4.2.83.
- [22] L. Cevallos-Torres, M. Botto-Tobar, O. León-Granizo, and A. Cortez-Lara, “Application of a Heuristic-Diffuse Model to Support Decision-Making in Assessing the Post-seismic Structural Damage of a Building,” in *Applied Technologies*, 2020, pp. 411–423.
- [23] D. Patiño Perez, R. Silva Bustillos, M. Botto-Tobar, and C. Munive Mora, “Análisis de Imágenes de Rayos X por Medio de Redes Neuronales Artificiales,” *Ecuadorian Sci. J.*, vol. 5, no. 1, pp. 55–60, 2021, doi: 10.46480/esj.5.1.50.