



Predicting Transport Mode in the Rabat Region: A Machine Learning Approach

Khalid Qbouche^{1*}, Khadija Rhoulami²

¹ Laboratory of Research in Computer Science and Telecommunications (LRIT), Faculty of Sciences, Mohammed V University, Rabat 1014, Morocco

² Department of Communication Sciences and Techniques (DESTEC), FLSHR Mohammed V University 1040, Morocco

Corresponding Author Email: khalid_qbouche@um5.ac.ma

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290210>

ABSTRACT

Received: 29 September 2023

Revised: 1 December 2023

Accepted: 31 December 2023

Available online: 25 April 2024

Keywords:

machine learning, k-nearest neighbors, multi-layer perceptron neural net-RBF, Bayesian belief network, support vector machine, multi-agent system, daily mobility

The continuous increase in population density and the increase in average travel of people by different modes of transportation, whether public or private, play a crucial part in how the city's urban area develops. The latter contribute significantly to the emergence of many problems, including road congestion, loss of time, and pollution in urban areas, noise, and other issues. Machine Learning has extensively beneficial aspects of proposing models in this context. Therefore, we proposed four algorithms of the machine learning techniques that have been implemented to analyze and classify the displacement of an individual's database to support urban decisions, K-Nearest Neighbors (KNN), Artificial Neural Network-multilayer perceptron Neural (net-RBF), Bayesian Belief Network (BBN) and Support Vector Machine (SVM). We will compare their learning metrics using train/test and cross-validation. The obtained results show that net-RBF offers the best accuracy (92.77%), SVM classifier (89.87%), BBN classifier (87.33%), and KNN classifier (86.53%).

1. INTRODUCTION

The large population agglomeration that comprises a region of Rabat presents excellent challenges related to the development of an infrastructure capable of accommodating the traffic within this agglomeration and thus achieving flexibility and ease of movement, which requires appropriate planning to accomplish this challenge. The Rabat area is an essential urban pole with more than four million residents and needs, primarily, the coordination of efforts for a shared vision that contributes to the strengthening of infrastructure in the topic of urban transport and the development of forms of mobility, thus achieving a more flexible flow and more significant result. In addition, the center of Rabat remains the destination of most of the trips made in the aggregation. This means that thought must be given to ensure the fluidity of trips and target densely populated urban areas. In the context of daily mobility, which entails researching how people and goods move around a city daily. This includes the movement of vehicles, bicycles, and pedestrians and the infrastructure and systems that support these movements. One approach to studying and improving daily mobility is multi-agent systems (MAS). MAS are a type of system in which multiple agents interacted with each other to reach same goal. In addition, in the context of everyday mobility, MAS agents could represent individual vehicles, traffic lights and pedestrians. These agents would communicate and make decisions to avoid crashes, improve traffic flow, and lessen congestion. The effectiveness and safety of transportation in a city may be increased by employing MASs for daily movement. In this topic, we present

a model to support urban decisions in the region of Rabat based on a machine learning algorithm and multi-agent system model. Decentralized control, hybrid control, game-theoretic approach, machine learning-based approach, and formal approaches are a few of the ways that may be used to achieve it. Machine learning has altered our perspective on daily mobility. Many aspects of our everyday mobility have been improved thanks to machine learning algorithms and technological improvements. Machine intelligence has emerged as a crucial tool for enhancing our everyday journey, from streamlining traffic to anticipating travel patterns. Additionally, the development of intelligent transportation systems has benefited from machine learning. By learning from user behavior, these systems can provide personalized recommendations for modes of transportation, such as suggesting the most efficient combination of public transit and ride-sharing services. Machine learning algorithms can also analyze data to create models supporting public transport exploitation, ensuring better service and minimizing delays. Traffic management is one of the critical ways that machine learning is used in daily mobility. Machine learning algorithms can accurately predict traffic congestion and suggest alternative routes in real time by analyzing vast amounts of data from traffic sensors, GPS systems, and historical patterns. This saves commuters valuable time, reduces fuel consumption, and lowers carbon emissions, contributing to a more sustainable transportation system. Decentralized control distributes decision-making power among the agents; centralized control relies on a central system or controller; hybrid management combines elements of both; the game-

theoretic approach models agents as players in a game; a machine learning-based approach uses algorithms to learn and make predictions, and formal methods use mathematical techniques to verify system properties. The system's unique requirements and other factors like technological capability and moral ramifications determine the approach to use. In this study, we crafted a model within his research using four machine-learning algorithms to forecast individuals' transportation modes from the survey data supplied by HCP [1]. The results produced by the model can provide valuable insights to both the general public and decision-makers, aiding in formulating response strategies and resilience plans aimed at mitigating the effects of challenges on daily mobility. The primary contributions of this article are organized as follows: Section 2 introduces related work on solving Daily Mobility problems with multi-agent systems and machine learning models. Section 3 describes our proposed four algorithms that have been implemented (KNN, SVM, net-RBF, and BNN) in which we have used labeled data to implement the model developed and to be able to predict mode transport. Section 4 showcases the results achieved; comparing them with those obtained using the traditional multinomial logit model. The conclusion is then reached, followed by exploration of potential future extensions.

2. RELATED WORKS

Daily mobility is defined here as all the trips made each day, whatever the means of transport used (car, walk, scooter, etc.) and the reason for the destination (work, leisure, shopping, return home). Additionally, it has been observed that certain researchers have employed multi-agent systems platforms and mathematical models to address issues related to this phenomenon. We mathematically discovered a Bayesian Belief Network model, as well as a model multi-agent system, called the Gama Platform, for simulating person displacement in a studied area. In another work stream, as stated in Fan et al. [2] they have developed a mathematical contagion model to estimate the flow of floodwater through road networks and to offer insights into the impact of flooding on urban mobility networks. However, it was demonstrated as stated in Qbouche and Rhoulami [3] they used a multi-agent system Gama platform with quantitative data to draw this displacement for people in the Rabat region. As stated in Qbouche and Rhoulami [4], the Markov Chain model is been implemented by language Gaml on the Multi-agent System Gama Platform to manage individual behavior displacements. Furthermore, they used to simulate the propagation of COVID-19 in the Rabat region by including parameters such as the value of the distance between people with multi-agent systems as stated in Qbouche and Rhoulami [5]. That is for a multi-agent system simulating human scenarios. The Bayesian belief network mathematical model was employed to construct a model for simulating land-use changes resulting from human decisions as stated in Qbouche and Rhoulami [6]. As stated by Zhao et al. [7] exploits graph attention networks to facilitate agent communication. This concerns simulation models that have helped to understand the daily mobility of individuals and remain among the methods that have helped city decision-makers make decisions. However, artificial intelligence-based approaches have limitations. Numerous data mining techniques have been successfully used to address this issue. As stated in study by Wang et al. [8], it been used multi-agent

reinforcement learning (MARL) to control state of traffic signals in the urbain zone. Wang et al. [9] they introduced a novel strategy based Multi-agent reinforcement Learning (ATSC) and a CGB-MAQL algorithm to manage various intersection scenarios to mitigate congestion and promote environmental conservation. In addition, as stated in study by Qbouche and Rhoulami [10] machine learning models to predict the workplace using quantitative data by the j48 algorithm and simulation of this displacement to this workplace predicted using Gama platform the multi-agent system. As stated in study by Abouelela et al. [11] have developed models to forecast the daily shared e-scooter fleet, specifically predicting the daily number of trips per vehicle using machine learning algorithms. Among these models, the Gradient Boosting Machine (LightGBM) emerged as the top-performing model across various evaluation metrics. As indicated in the study by Xie et al. [12], it is been demonstrated that the Artificial Neural Network (ANN) exhibits greater reliability in predicting travel modes compared with Decision tree (DT) and multinomial logit (MNL) models. Furthermore, research conducted by Zhang and Xie [13] they are been demonstrated that Support Vector Machine (SVM) and ANN models are more robust and surpass the performance of the MNL model. Li et al. [14] they implemented the Support Vector Machine (SVM) model to predict motor vehicle crashes, is more robust and faster than neural networks. Furthermore, Wang et al. [15] applied the machine learning algorithm SVM to study mobility prediction and improve mobile communication; they obtained a high accuracy of 92.2%.

3. PROPOSED METHODOLODY

3.1 Dataset and attributes: Daily mobility in Rabat region and mode choices for displacement

In this research, we employed individual-level data obtained from the HCP (High Commission for Planning). This census, initiated in 2014, offers a representative sample of the Rabat region's population, covering various demographic and socioeconomic attributes including 'Level of study, Age, the aggregate level of study, large group profession, professional status, activity sector, workplace, and mode of transport'. The dataset comprises 100,000 instances specific to the Rabat region. Table 1 outlines the individual dataset, consisting of 8 attributes grouped into 7 class features representing distinct modes of transportation: Transport Personnel, Private Car, Bicycle/Motorcycle, Bus, Taxi, on foot, and Train. The second phase involves data preprocessing, a critical step aimed at preparing and changing data to format identical to train a machine learning model. Given the diverse sources of training and testing datasets and the presence of missing values, this step includes the application of normalization and standardization techniques to ensure that numerical data is consistently and comparably scaled. The third stage involves identifying the nature of the target variable. In our research, the target variable is categorical, specifically used for classifying the transport mode. The fourth phase entails to divide the dataset to train and test sets. This division is performed to ensure that both training and testing datasets contain properly regulated analysis values, mitigating the risk of overfitting or under fitting. The final stage involves utilizing the training subset of the dataset to train machine learning

models. This process enables the model to learn and generalize from the provided data. The parameters generated during this phase are subsequently employed in the final machine learning model, which is fine-tuned by optimizing the settings of each

method to achieve the highest accuracy when making predictions during testing. The complete process is shown in the following Figure 1.

Table 1. Rabat region individual's dataset attributes

Attributes	Data
Age	20-24, 25-30,30-35..., 60-65.
Large group of professional	workers in the domain agriculture, Craftsmen, skilled tradesmen (except agricultural workers), waters and forests, fishing domain, Non-farm Labourers, Material Handlers and Small Trade Workers
The aggregate level of study	Preschool, Superior, High school, college Primary, Qualifying secondary, Preschool, or No level of education
Professional Group	workers in agriculture, Workers in the fishing industry, Craftsmen, skilled tradesmen (except agricultural workers), Members of legislative bodies- local elected officials public service line managers and executives, workers in the forest domain, Inactive, Technicians and intermediate professions, the Director General, Professionals, Lawmakers - public service line managers and executives, Employees
Professional status	Caregiver/Apprentice, Private sector employee, Independent
Activity Sector	Agriculture- forestry and fishing, Extractive and manufacturing industries, Construction.
Workplace	Not determined, Non-fixed location, in the territory of the douar which is resident, Other provinces, in the municipality of residence, in the provincial territory where he resides, At home, places not determined.
Mode Transport	Transport Personnel, Private Car, Bicycle/Motorcycle, Bus, Taxi, On foot, Train.

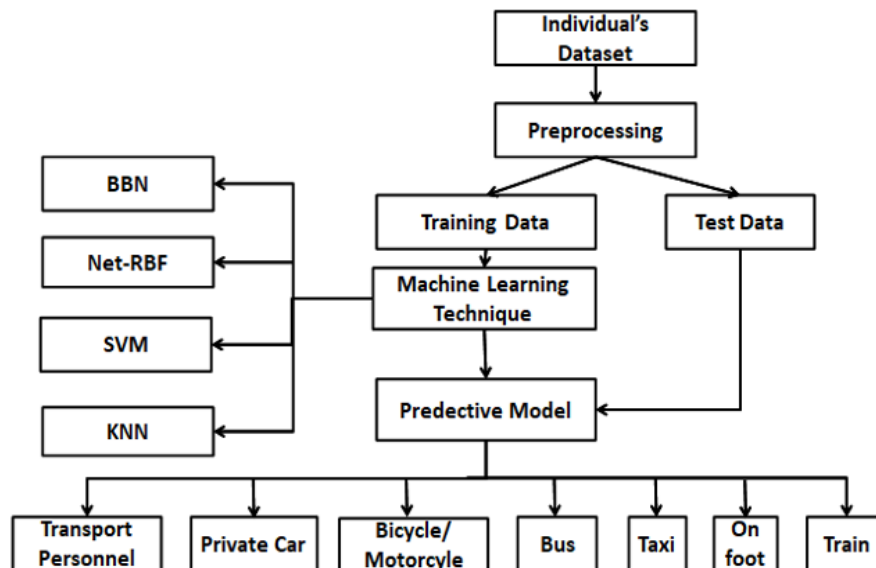


Figure 1. Proposed process of model mode travel predicted

3.2 Machine learning algorithm

Below is an explanation of this comparative research's various machine-learning methods.

3.2.1 k-nearest neighbors: KNN

The KNN or k-NN, it is a straightforward, nonparametric, supervised learning classifier, it is used to address classification and regression problems. One of the most straightforward and intuitive algorithms in machine learning is considered.. It is a good yet inefficient algorithm that waits until it receives the training sets before starting to train as stated by Zhang et al. [16].

The fundamental concept of K-Nearest Neighbors (KNN) is to identify the K closest data points in the training set to a specified test data point. In succession, either the majority class or the average of the labels from these K nearest neighbors is used to classify or predict the target to the test data.

Here are the steps for the KNN algorithm:

- Choose the value of k : k is the number of nearest neighbors that will be used to classify or predict the target value of a given test data point.
- Calculate the distance: It is between the test data point and all data points in the training set. While the Euclidean distance is the most frequently employed metric, various other distance metrics can also be considered. With X and Y be the respective coordinate points $((X_1, \dots, X_n), (Y_1, \dots, Y_n))$ in a relevant space:

$$D(x, y) = \sum_{i=1}^n (x_i - y_i)^2 \quad (1)$$

- Find the k nearest neighbors: Choose K data points from the train dataset that exhibit the minimal distances on the test dataset, as determined by the selected distance metric.
- Classify or predict the target value: if the task is

classification, assign the majority class label among the k nearest neighbors to the test data point. If the task involves regression, which calculates the average of the estimated values of the nearest neighbor K -factor values.

- Returns the predicted values: return the predicted target value for the test data point.

3.2.2 Support vector machine: SVM

It is an algorithm, widely utilized in tasks such as: a classification, regression and outlier detection. It is a supervised learning algorithm, aiming to determine the optimal hyper plane that maximizes the margin between the two classes within the data set, as outlined by Abdar and Makarenikov [17]. By positioning this hyperplane effectively in an n -dimensional space, SVM seeks to achieve optimal generalization capabilities, as highlighted by Osman [18].

As a result, the equation below serves as the definition of the ideal choice function:

$$y = \text{sign}(f(x)) = \text{sing}(\sum_{i=1}^n w_i \cdot x_i + b) \quad (2)$$

With: $(x_1, y_1) \dots \dots \dots (x_n, y_n) \in IR$.

Here are the steps for the SVM algorithm:

- Data preparation: convert the data into a feature matrix X and target vector y , where X contains the features of each data point, and y contains the class labels
- Select the kernel: It determines the shape of the decision boundaries and the model's capacity to capture intricate patterns within the data. The choice of kernel depends on the nature of the data and the problem and hand.
- Training the SVM: it involves the process of teaching the algorithm to obtain the optimal hyper plane that best divides data samples belonging to different classes.
- Testing SVM: Once SVM has been trained, it can estimate the class labels of fresh data points. This is done on applying the trained SVM model to the feature matrix X of the new data points.
- Evaluation: Cross-validation or a validation set should be used to assess the SVM's performance. Accuracy, precision, recall, and F1 score are typical classification problem assessment criteria.

3.2.3 Artificial neural network multilayer perceptron: net-RBF

As elucidated by Broomhead and Lowe [19], an artificial neural network (ANN) with a radial basis function (RBF) employs radial basis functions as activation functions within its hidden layer. This ANN-RBF model operates under the assumption that the output for input patterns resembling a given input x will likely correspond closely to the expected value for x . Primarily utilized for pattern recognition and function approximation tasks, his architecture consists of three layers: an input layer, a hidden layer featuring radial basis functions, and an output layer. It uses its hidden layer to transform input data into a higher-dimensional space. It is connected to each neuron in the hidden layer, typically a Gaussian function mathematical centered at a specific point in the input space. In the Figure 2, we present RBF Neural Network architecture.

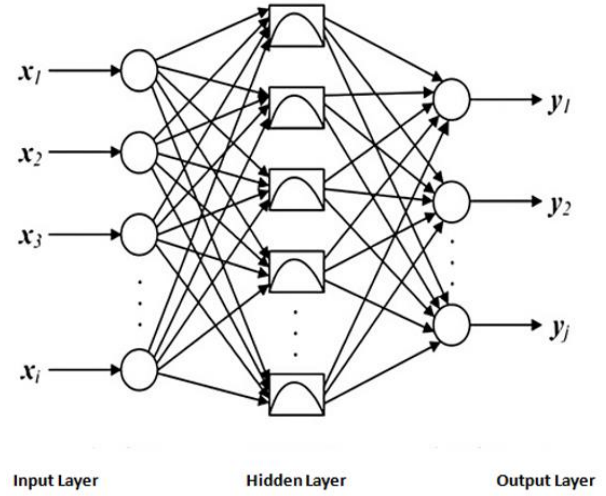


Figure 2. Multilayer perceptron architecture in artificial neural networks.

The value of the output is determined based on the input using the following equation:

$$y = \sum_{i=1}^h w_i \cdot \phi(\|x - c_i\|) \quad (3)$$

With h is the number of hidden neurons, w_i , and c_i the weight associated and the center of the i -th RBF. Furthermore, ϕ is a radial basis function (RBF). During the training process, the objective is to optimize the two sets of values, namely w and c .

3.2.4 Bayesian belief network: BBN

It is a collection of random variables with their probabilistic interconnections. It is typically represented as a tuple (G, P) , additionally G denotes a directed acyclic graph (DAG) accompanied by a node corresponding to each variable in the model, as described by Broomhead and Lowe [19]. Meanwhile, P represents a multivariate probability distribution defined for each variable. In a Bayesian Belief Network, a set of nodes, each corresponding to variables within the system under investigation, embodies the key factors influencing the phenomenon being studied. Each node encompasses a set of mutually exclusive states. The network comprises directed links that encode conditional dependencies among these nodes, delineating the relationships between variables. Additionally, each node incorporates a set of probabilities that specify the likelihood of it being in a particular state based on the states of its parent nodes. To illustrate, consider a scenario with three nodes. Can be representing the relationship among these variables by this following expression mathematic:

$$P(B/A) = \frac{P(B,A)}{P(A)} \quad (4)$$

In the equation provided, P represents a multivariate probability distribution.

It is a powerful tool for modeling complex systems, such as those involving human behavior, because it allows for efficient reasoning and inference. In studying individual behavior, a Bayesian belief network can represent the various factors that influence a person's actions. For example, a network could include variables such as a person's mood, the time of day, the location they are in, and any external stimuli they are exposed to. The network would also include the probabilistic

relationships between these variables, such as the likelihood that a person will be in a good mood if they have had enough sleep the previous night. It is possible to make predictions about a person's behavior based on the values of these variables by this model. For example, if a person is in a crowded public place and is exposed to loud noises, the network could predict that they will likely become agitated. This prediction could be used to inform decisions about how to manage the situation best, such as by providing a quieter area for the person to retreat to. In addition to making predictions, a Bayesian belief network can identify the most important factors influencing a person's behavior. Analyzing the network makes it possible to determine which variables have the strongest effects on the outcome of interest. This information can benefit the creation of interventions that target the most influential factors to achieve the desired behavior change. Overall, a Bayesian belief network helps study individual behavior because it allows for modeling complex, probabilistic relationships between multiple variables. It can be used to predict behavior and identify the key factors that influence it, which can inform the development of effective interventions.

In machine learning and artificial intelligence, Bayesian Belief Networks (BBNs) are frequently employed to model uncertain and intricate domains. This approach has demonstrated success in various domains, including medical diagnostics, decision support systems, semantic search, and bioinformatics, as evidenced by studies conducted by Andersen et al. [20-22].

In the realm of machine learning, BBNs find utility in diverse tasks such as classification, anomaly detection, missing data imputation, and feature selection, as highlighted in the research conducted by Chen et al. [23]. These networks can be trained from data by amalgamating expert knowledge with data-driven learning techniques.

4. EXPERIMENTS AND RESULTS

4.1 Training techniques

Cross-validation is a method employed to analyze the performance of models created in machine learning algorithms by dividing the available data into multiple subsets. It is then train it and test it on different combinations of these subsets. The primary goal of cross-validation is to offer an unbiased predict of the model's performance, utilizing all available data for training and testing without the risk of overfitting to a single data split. This method helps to provide a more reliable evaluation of the model's generalization capability and robustness.

4.2 Learning metrics

Learning metrics, also referred to as evaluation metrics, have a critical in measuring the performance of a machine learning model created. These metrics are instrumental in evaluating the accuracy and efficacy of the model's predictions, enabling stakeholders to assess the quality of the model's output and identify areas for improvement. By leveraging learning metrics, practitioners can gauge how well the model is performing and make informed decisions regarding model refinement and optimization.

4.2.1 Accuracy

Accuracy is a widely used evaluation metric in machine learning, quantifying the percentage of correctly classified instances by a model. True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) have been used to calculate accuracy using the following formula:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

4.2.2 Confusion matrix

This refers to the evaluation of a classification model's performance when applied to a set of test data with known actual values. Another term for it is a confusion matrix or an error matrix. A confusion matrix displays the value of:

- True Positive (TP);
- True Negative (TN);
- False Positive (FP);
- False Negative (FN);

Does a classification model make predictions. These values are computed in comparison the predicted class labels obtained from the model created with the actual class labels present in the reserved test data. As shown in Table 2.

Table 2. Confusion matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

4.2.3 Precision

Precision is the percentage of accurately predicted positive instances for the chosen mode of transportation. It determines accuracy using the following formula:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

4.2.4 Recall

Also known as sensitivity, it represents the accuracy in estimating the true positive mode of transportation:

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

4.2.5 F-measure

It is known as the F1 score, is a widely employed metric for assessing the performance of a classification model. It can be computed using the subsequent formula:

$$measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (8)$$

5. RESULTS

In this section, we present our experimental findings. Employing the renowned Weka tool, we conducted supervised learning utilizing four distinct machine learning algorithms: KNN, BBN, SVM, and net-RBF. These algorithms were implemented using the popular Scikit-learn Python library. To evaluate the performance of the machine learning model, we utilized both Train/Test split and Cross-validation techniques.

This approach aids in gauging the model's effectiveness on unseen data and mitigating overfitting, wherein the model merely memorizes the training data instead of identifying

underlying patterns.

Table 3. Assessing metrics utilizing a partition of train/test data

Classifier	KNN	BBN	SVM	Net-RBF
Accuracy	88.09%	87.59%	90.29%	92.94%
Precision	89%	88.3%	90%	89%
Recall	90%	87.6%	91%	93%
ROC	89%	89.9%	96.9%	98.9%
F1-score	89%	87.7%	90%	91%

The dataset was split randomly, with 80% of the individuals' dataset destined to train the model, while the remaining 20% to test. The training set utilized for model training, the validation set for tuning hyper parameters and model selection,

and the test set for final evaluation. The initial results obtained from the Train/Test Split are presented in Table 3.

As a secondary outcome, as depicted in Table 4, we employed 10-fold cross-validation, allocating 80% to train and 20% to test.

Table 4. Performances algorithms comparison

Classifier	KNN	BBN	SVM	Net-RBF
Accuracy	86.5%	87.33%	89.87%	92.77%
Precision	85%	88.10%	88%	89%
Recall	86%	87.30%	90%	93%
ROC	99.2%	98.5%	98%	98.4%
F1-score	85%	87.5%	88%	91%

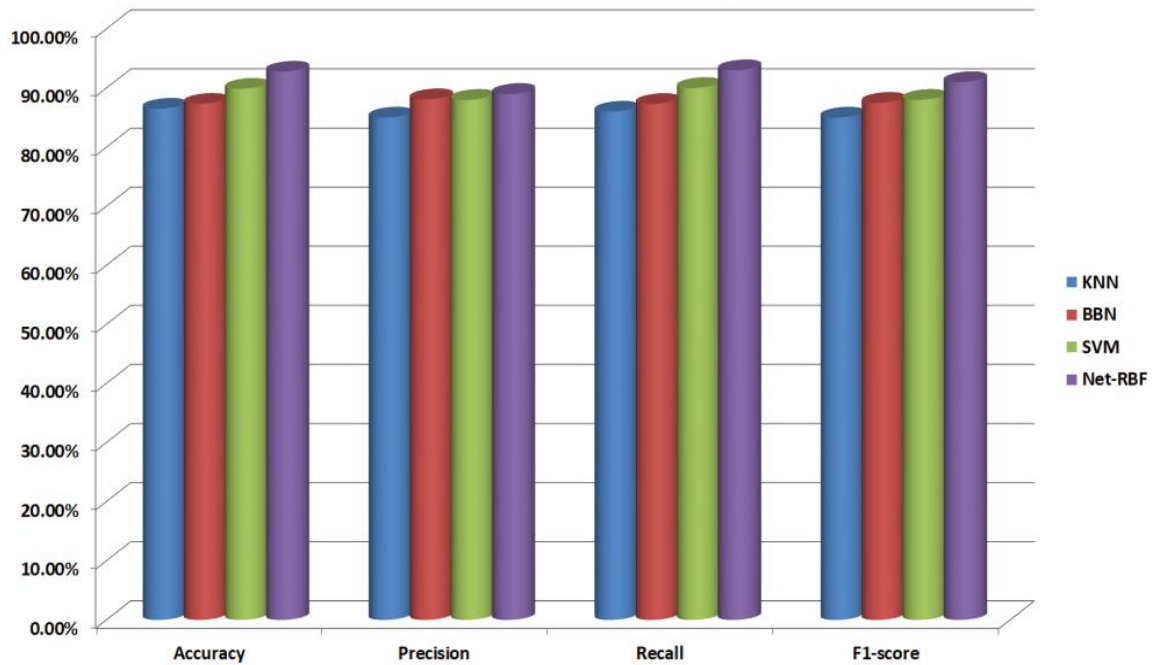


Figure 3. Comparative diagram of algorithms proposed concerning evaluation criteria: Accuracy, Precision, Recall, F1- score

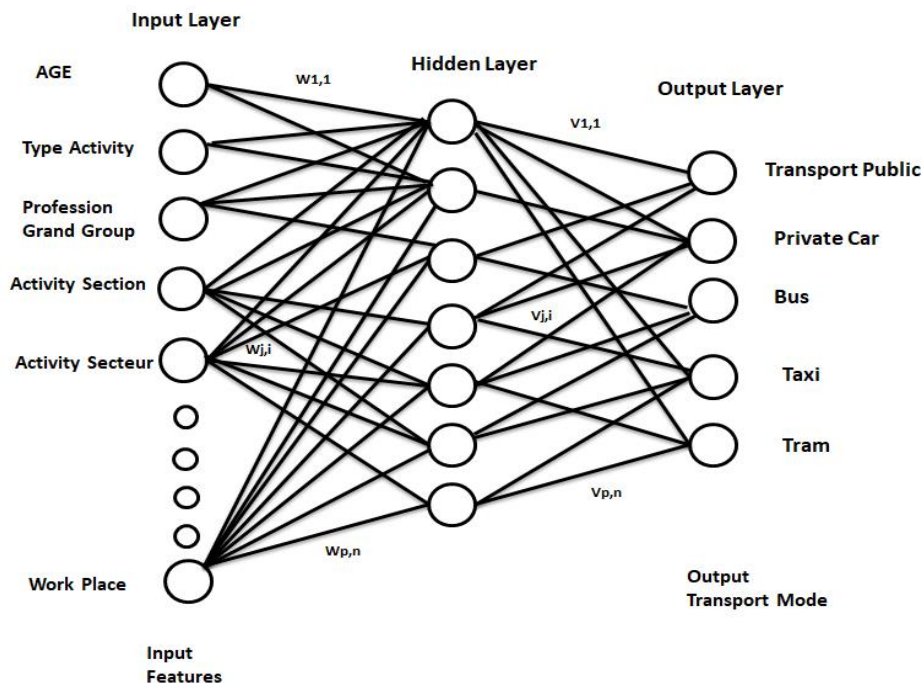


Figure 4. The net-RBF model's architecture was applied to predicting employers' transport modes in the Rabat region

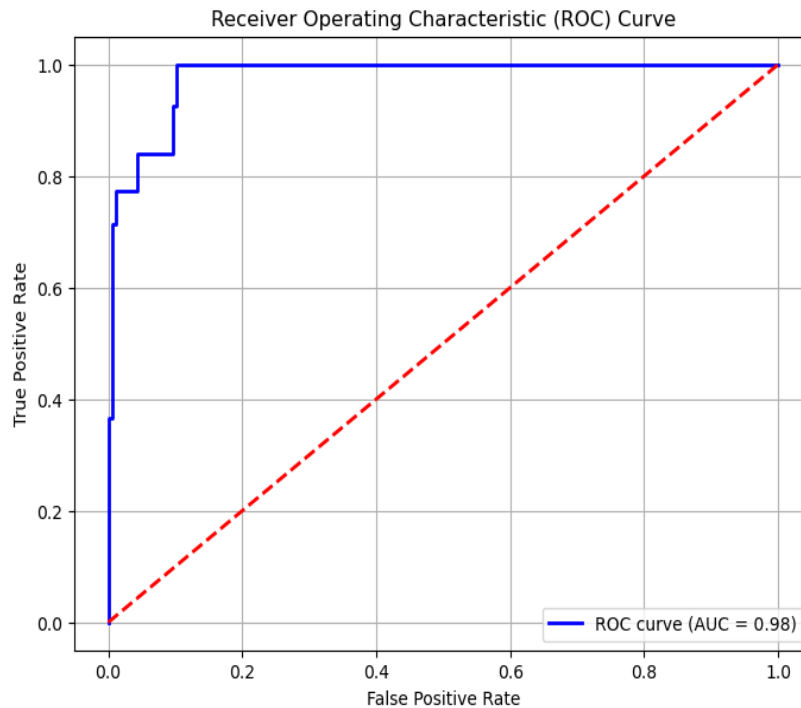


Figure 5. The ROC curve net-RBF mode

In Figure 3, we will show the comparative performance of these algorithms.

To better understand the model's efficiency, Figure 4 illustrates the architecture of the net-RBF, delineating its input, hidden, and output layers. Meanwhile, Figure 5 showcases the ROC curve for our classifiers, specifically highlighting the precision of each classifier when employing the net-RBF algorithm. By presenting the ROC curve, we offer a visual depiction that elucidates the performance disparities among different classifiers, aiding in a comprehensive understanding of their efficacy.

6. CONCLUSIONS

In this paper, we introduced a model to predict transport modes in the Rabat region using some machine learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Bayesian Belief Network (BBN), and Artificial Neural Network with multilayer perceptron architecture (netRBF). The model incorporates a comprehensive set of variables encompassing transport, workplace and residence locations, as well as social and demographic factors such as Level of study, Age, aggregate level of study, large group profession, professional status, activity sector, workplace, and transport mode. These variables are utilized alongside real-world data to study the mobility patterns of workers in the region Moroccan Rabat. Performance evaluation of the algorithms proposed is conducted by value of metrics like accuracy value, precision value, recall value, and F-score value. Results indicate that net-RBF exhibits the highest accuracy (92.77%), followed by SVM classifier (89.87%), BBN classifier (87.33%), and KNN classifier (86.53%). Future research endeavors will explore alternative methodologies with the aim of refining the predictive accuracy of transportation mode prediction through the adoption of innovative machine learning techniques.

ACKNOWLEDGMENT

This research is carried out as a part of the project Plateforme logicielle d'intégration de stratégies d'immunisation contre la pandémie COVID-19 funded by the grant of the Hassan II Academy of Sciences and Technology of Morocco.

REFERENCES

- [1] Institutional website of the Haut-Commissariat au Plan of the Kingdom of Morocco. <https://www.hcp.ma/>, accessed on Aug. 18, 2023.
- [2] Fan, C., Jiang, X., Mostafavi, A. (2020). A network percolation-based contagion model of flood propagation and recession in urban road networks. *Scientific Reports*, 10(1): 13481. <https://doi.org/10.1038/s41598-020-70524-x>
- [3] Qbouche, K., Rhoulami, K. (2021). Simulation daily mobility in Rabat Region. In *Proceedings of the 4th International Conference on Networking, Information Systems & Security*, pp. 1-4. <https://doi.org/10.1145/3454127.3454128>
- [4] Qbouche, K., Rhoulami, K. (2022). Towards for an agent-based model to simulate daily mobility in Rabat Region. In: Saidi, R., El Bhiri, B., Maleh, Y., Mosallam, A., Essaïdi, M. (eds) *Advanced Technologies for Humanity. ICATH 2021. Lecture Notes on Data Engineering and Communications Technologies*, 110: 3-9. https://doi.org/10.1007/978-3-030-94188-8_1
- [5] Qbouche, K., Rhoulami, K. (2022). Simulation and forecasting the Covid-19 pandemic in the Rabat Region. In: Motahhir, S., Bossoufi, B. (eds) *Digital Technologies and Applications. ICDTA 2022. Lecture Notes in Networks and Systems*, 454: 120-130. https://doi.org/10.1007/978-3-031-01942-5_12

- [6] Qbouche, K., Rhoulami, K. (2022). Simulation daily mobility in rabat region using multi-agent systems models. *Journal of ICT Standardization*, 10(2): 293-304. <https://doi.org/10.13052/jicts2245-800x.10210>
- [7] Zhao, P., Yuan, Y., Guo, T. (2022). Extensible hierarchical multi-agent reinforcement-learning algorithm in traffic signal control. *Applied Sciences*, 12(24): 12783. <https://doi.org/10.3390/app122412783>
- [8] Wang, X., Ke, L., Qiao, Z., Chai, X. (2020). Large-scale traffic signal control using a novel multiagent reinforcement learning. *IEEE Transactions on Cybernetics*, 51(1): 174-187. <https://doi.org/10.1109/TCYB.2020.3015811>
- [9] Wang, T., Cao, J., Hussain, A. (2021). Adaptive traffic signal control for large-scale scenario with cooperative group-based multi-agent reinforcement learning. *Transportation Research Part C: Emerging Technologies*, 125: 103046. <https://doi.org/10.1016/j.trc.2021.103046>
- [10] Qbouche, K., Rhoulami, K. (2022). Simulation daily mobility using J48 algorithms of machine learning for predicting workplace. *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning*, 454: 120-130. https://doi.org/10.1007/978-3-031-01942-5_12
- [11] Aboulela, M., Lyu, C., Antoniou, C. (2023). Exploring the potentials of open-source big data and machine learning in shared mobility fleet utilization prediction. *Data Science for Transportation*, 5(2): 5. <https://doi.org/10.1007/s42421-023-00068-9>
- [12] Xie, C., Lu, J., Parkany, E. (2003). Work travel mode choice modeling with data mining: Decision trees and neural networks. *Transportation Research Record*, 1854(1): 50-61. <https://doi.org/10.3141/1854-06>
- [13] Zhang, Y., Xie, Y. (2008). Travel mode choice modeling with support vector machines. *Transportation Research Record*, 2076(1): 141-150. <https://doi.org/10.3141/2076-16>
- [14] Li, X., Lord, D., Zhang, Y., Xie, Y. (2008). Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention*, 40(4): 1611-1618. <https://doi.org/10.1016/j.aap.2008.04.010>
- [15] Wang, D., Zhou, Q., Partani, S., Qiu, A., Schotten, H.D. (2021). Mobility prediction based on machine learning algorithms. In *Mobile Communication - Technologies and Applications*, 25th ITG-Symposium, Osnabrueck, Germany, Osnabrueck, Germany, pp. 1-5.
- [16] Zhang, M.L., Zhou, Z.H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7): 2038-2048. <https://doi.org/10.1016/j.patcog.2006.12.019>
- [17] Abdar, M., Makarekiov, V. (2019). CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer. *Measurement*, 146: 557-570. <https://doi.org/10.1016/j.measurement.2019.05.022>
- [18] Osman, A.H. (2017). An enhanced breast cancer diagnosis scheme based on two-step-SVM technique. *International Journal of Advanced Computer Science and Applications*, 8(4): 158-165. <https://doi.org/10.14569/ijacsa.2017.080423>
- [19] Broomhead, D., Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2: 321-355.
- [20] Andersen, L.R., Krebs, J.H., Andersen, J.D. (1991). STENO: An expert system for medical diagnosis based on graphical models and model search. *Journal of Applied Statistics*, 18(1): 139-153. <https://doi.org/10.1080/02664769100000012>
- [21] Bacon, P.J., Cain, J.D., Howard, D.C. (2002). Belief network models of land manager decisions and land use change. *Journal of Environmental Management*, 65(1): 1-23. <https://doi.org/10.1006/jema.2001.0507>
- [22] Mihaljević, B., Bielza, C., Larrañaga, P. (2021). Bayesian networks for interpretable machine learning and optimization. *Neurocomputing*, 456: 648-665. <https://doi.org/10.1016/j.neucom.2021.01.138>
- [23] Chen, S., Webb, G. I., Liu, L., Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems*, 192: 105361. <https://doi.org/10.1016/j.knosys.2019.105361>