Profiling and Classification of Users Through a Customer Feedback-based Machine Learning Model

Jihane LARIOUI, Abdeltif EL BYED

Laboratory of Computer Science and Systems, HASSAN 2 University Faculty of Science.
Ain Chock Casablanca, Morocco

Abstract—**The systems aimed at predicting user preferences and providing recommendations are now commonly used in many systems such as online shops, social websites, and tourist guide websites. These systems typically rely on collecting user data and learning from it in order to improve their performance. In the context of urban mobility, user Profiling and Classification represent a crucial step in the continuous enhancement of services provided by our multi-agent system for multimodal transportation. In this paper, our goal is to implement and compare some machine learning (ML) algorithms. We will address the technical aspect of this implementation, demonstrating this model leverages customer feedback to develop a thorough understanding of individual preferences and travel behaviors. Through this approach, we can categorize users into distinct groups, enabling a finer personalization of route recommendations and transportation preferences. The ML model analyzes customer feedback, identifies recurring patterns, and continuously adjusts user profiles based on their evolution. This innovative approach aims to optimize the user experience by offering more precise and tailored recommendations, while fostering dynamic adaptation of the system to the changing needs of urban users.**

Keywords—Machine learning; urban mobility; multimodal transportation; multi-agent systems

I. INTRODUCTION

The expansion of urban mobility, driven by population growth and rapid urbanization, has introduced new complex challenges in transportation management. In densely populated urban areas, transportation systems must cope with increasing demand, often resulting in issues such as traffic congestion, longer travel times, and heightened pressure on existing infrastructure.

The data challenge in this context stems from the need to collect, process, and interpret massive and diverse amounts of information related to urban mobility. Data comes from various sources, including traffic sensors, navigation apps, public transportation, connected vehicles, and other Internet of Things (IoT) devices. Managing this variety of often unstructured data poses challenges in terms of collection, real-time processing, and analysis to derive actionable insights [7].

In response to this challenge, the application of machine learning (ML) techniques is becoming increasingly crucial. ML enables the processing of complex data and the discovery of hidden patterns, thereby providing insights to improve planning, operational efficiency, and user satisfaction in the field of urban mobility. In summary, effective urban mobility

management requires an innovative and technological approach to overcome the challenges related to the quantity and diversity of data generated by urban travel.

The machine learning algorithms we intend to implement play a crucial role within our multi-agent system [16, 17, 18], acting in complement to significantly improve overall performance and decision-making. These algorithms seamlessly integrate into our approach, leveraging the rich data collected on user history, transportation choices, and preferences. By enhancing the decision-making capabilities of our system, machine learning algorithms enable increased customization of the services offered [8]. By analyzing emerging patterns from this data, the algorithms can anticipate user needs, dynamically adjust route recommendations, and optimize interactions within the multi-agent system. Thus, this synergy between agent intelligence and the adaptive capabilities of machine learning algorithms aims to provide a more responsive, efficient, and perfectly aligned urban mobility experience tailored to each user's individual preferences.

The establishment of a user data collection procedure represents a crucial step in the evolution of our multi-agent system for multimodal information. This initiative aims to establish closer interaction with users by gathering information about their preferences, behaviors, and experiences during multimodal travel.

The first objective of our work is to exploit this data to enhance the system's capabilities through learning process. By analyzing users' choices and habits over time, the system will be able to adjust its recommendations and responses based on individual preferences. This increased personalization will not only improve user satisfaction but also optimize the overall system performance by making it more responsive and adaptive.

The application of machine learning (ML) in the field of urban mobility is of major interest due to several factors. Firstly, the variety of available datasets covers a wide range of aspects related to urban travel, such as location data, user preferences, traffic conditions, and more. ML offers the opportunity to extract complex patterns from this diverse data, enabling a thorough understanding of travel patterns [9].

Furthermore, the use of ML can contribute to smarter and more proactive decision-making in urban mobility management. Machine learning models can analyze emerging trends in real-time, anticipate congestion, and provide

personalized recommendations to users. This paves the way for more efficient optimization of transportation systems, thereby reducing travel times, greenhouse gas emissions, and traffic congestion.

User data collection will also refine the preference criteria used in the decision-making process, ensuring a more precise match between generated recommendations and specific traveler expectations. Therefore, this step represents a strategic advancement toward an intelligent multimodal transportation system capable of dynamically adjusting to evolving user needs and preferences, ultimately optimizing the urban mobility experience.

In summary, the variety of available datasets in the context of urban mobility presents a unique opportunity to harness the potential of machine learning to significantly improve urban transportation management, promoting smoother, sustainable, and adaptive mobility.

In this article, our goal is to implement the integration and implementation part of machine learning (ML) algorithms. We will address the technical aspect of this implementation, demonstrating how the ML algorithm is applied to process and enhance urban mobility data. This practical application will be a crucial step in realizing our theoretical approach in a functional system that is adaptive to the challenges of contemporary urban mobility.

The next section will be dedicated to the state of the art, followed by a presentation of the methodology for designing our multi-agent system for multimodal transportation, detailing the ML part and its relevance in our system while highlighting the essential contribution of the ML agent to our architecture. Then, we will explore the implementation part of these algorithms in our system. The results from these implementations will be examined in detail in the next section, demonstrating the concrete impact of integrating machine learning algorithms on the evolution of our architecture.

II. LITERATURE REVIEW

In recent academic literature, there has been significant interest in the realm of urban mobility, particularly regarding the integration of machine learning algorithms into intelligent transportation systems (ITS). The aim is to enhance the efficiency, effectiveness, and overall quality of transportation services within urban environments [2]. This interdisciplinary field brings together expertise from transportation engineering, computer science, data analytics, and urban planning to address the complex challenges inherent in urban transportation. Researchers have delved into various aspects of urban mobility, leveraging machine learning techniques to tackle diverse problems. One key area of focus involves predicting transportation demand patterns, such as the need for taxis, ride-sharing services, or public transit, with the goal of optimizing resource allocation and service provision. By harnessing historical travel data, demographic information, and other relevant factors, predictive models can anticipate fluctuations in demand and facilitate proactive decision-making by transportation authorities and service providers.

Another critical application of machine learning in intelligent transportation systems is route optimization and trip planning. By analyzing real-time traffic data, weather conditions, and user preferences, algorithms can generate optimal routes for vehicles, minimizing travel times, fuel consumption, and environmental impact. This not only benefits individual commuters but also contributes to overall traffic management and congestion mitigation efforts in urban areas.

Furthermore, machine learning algorithms play a crucial role in detecting and predicting traffic congestion. By analyzing sensor data from traffic cameras, GPS-enabled vehicles, and infrastructure sensors, these algorithms can identify congested areas in real-time and forecast potential bottlenecks [3]. This information enables traffic management authorities to implement timely interventions, such as adjusting signal timings, rerouting traffic, or deploying additional resources to alleviate congestion and improve traffic flow.

Overall, the integration of machine learning into intelligent transportation systems holds immense promise for revolutionizing urban mobility. Through innovative research and practical implementation, researchers and practitioners aim to create safer, more efficient, and sustainable transportation networks that cater to the evolving needs of urban populations.

In research [1], the author investigates the prediction of passenger flow within urban areas, focusing on the city of Oslo, Norway. Traditional macro models for traffic flow simulation face limitations in capturing the complexity of real traffic patterns. In contrast, machine learning (ML) models offer promising alternatives. The study compares the effectiveness of a traditional Spatial Interaction model (SIM) with a selective ML model for traffic flow prediction. Results reveal that while the SIM is interpretable and requires fewer parameters, it struggles to accurately represent real flow dynamics compared to the ML model. Statistical analyses support these findings, highlighting the potential of ML models in discerning passenger movement trends and simulating traffic scenarios. The research provides a decision support system for urban planners and policymakers to forecast traffic flow accurately, contributing to the ongoing discussion on the role of machine learning in transportation modeling.

In the other hand, in study [4] the author focuses on human mobility as an interdisciplinary field encompassing physics and computer science. While various models and prediction methods have been proposed for understanding and forecasting human mobility, the emergence of multi-source heterogeneous data from handheld terminals, GPS, and social media has opened new avenues for exploring urban mobility patterns in detail. Such studies are crucial for applications spanning urban planning, epidemic control, location-based services, and intelligent transportation management. This survey focuses on human-centric perspectives within a data-driven context, examining mobility patterns at individual, collective, and hybrid levels. It also reviews prediction methods across four aspects and discusses recent developments while addressing open issues to guide future research. This comprehensive review serves as a valuable resource for newcomers seeking an overview of human mobility and provides insights for researchers aiming to develop unified mobility models.

The author in [5] in her thesis, delves into predicting user mobility using deep learning algorithms, with the aim of

enhancing service quality for users and reducing paging costs for telecom carriers. Through a comprehensive literature review, RNN, LSTM, and variants of LSTM are identified as suitable deep learning algorithms for the task. Subsequently, an experiment is conducted to evaluate the performance of these algorithms, both as a global model and individual models. The results reveal that individual models demonstrate superior performance in predicting user mobility compared to the global model. Hence, it is concluded that individual models represent the preferred technique for this purpose, offering valuable insights for optimizing mobility prediction strategies.

The author in study [6] on his paper tried to investigate the integration of AI, machine learning, and data analytics in smart transportation planning to enhance urban mobility sustainability. It tackles two core questions: how these technologies can optimize urban transportation systems and the potential benefits they offer. The methodology involves collecting transportation data, applying statistical analysis, and developing a simulation model calibrated with real-world data to evaluate various scenarios. Performance metrics like travel time and congestion levels are used to assess strategy effectiveness. The findings demonstrate improved transportation efficiency and sustainability in New York City but acknowledge limitations such as data availability and modeling assumptions. Overall, this research contributes to evidence-based decision-making in civil engineering, providing insights for stakeholders and urban planners striving for sustainable urban mobility.

Based on this research and several others, we observe the significant contribution that machine learning provides to urban mobility.

III. Presentation of the ML Methodology Design Approach

A. Contribution of Machine Learning to Urban Mobility

Machine Learning, as an essential component of artificial intelligence, offers promising perspectives for transforming the way we perceive and manage mobility in dynamic urban environments. This enlightened introduction by machine learning brings a new and adaptive dimension to multimodal information systems dedicated to mobility, providing substantial benefits for both users and urban mobility managers.

This section will explore the multiple contributions of ML, highlighting how these advanced techniques can be deployed to improve service personalization, optimize routes, predict demand, manage traffic, provide intelligent recommendations, and foster continuous evolution through feedback collection. The goal is to demonstrate how the integration of ML transforms urban mobility systems into intelligent entities capable of dynamically adapting to user needs and the complex challenges of urban life.

The contribution of Machine Learning (ML) in the field of urban mobility is significant and offers substantial benefits to improve efficiency, service personalization, and overall user experience quality. Indeed, Machine Learning can contribute to enhancing urban mobility management on several levels.

1) Personalization of services: ML algorithms enable understanding users' individual preferences by analyzing their travel habits, previous choices, and other relevant data. This personalization offers the opportunity to provide tailor-made route recommendations, considering each user's specific preferences.

2) Route optimization: ML techniques, such as decision trees, can be used to analyze historical travel data, real-time traffic conditions, and other relevant variables. By using this information, the system can recommend optimized routes that minimize travel time, costs, or other specific criteria.

3) Transport demand prediction: ML models can be employed to predict fluctuations in transport demand based on various factors, such as time of day, days of the week, and special events. These predictions enable better resource management and more efficient transport service planning.

4) Feedback collection and continuous improvement: ML algorithms can be applied to analyze user feedback, assess customer satisfaction, and identify potential areas for improvement in the mobility system. This allows for continuous adaptation to changing user needs.

In our context, we pay particular attention to the aspects of Feedback Collection and Continuous Improvement. This approach promotes continuous adaptation to changing user needs, thus constantly optimizing the services offered. By leveraging the collected data, these algorithms contribute to a thorough understanding of individual preferences, facilitating targeted adjustments to improve the overall urban mobility experience.

The integration of ML into multimodal information systems thus offers an intelligent approach to solving complex issues related to urban mobility, contributing to a smoother, personalized, and efficient user experience.

B. Proposed Approach

The design methodology we have adopted for the Machine Learning (ML) part of our system revolves around a systematic and iterative approach. We have implemented a process that integrates robust design principles while considering the specific features and particular requirements of urban mobility data.

The first step of our methodology involves clearly defining the objectives of ML modeling in our system. We identify specific aspects of urban mobility that we wish to address, such as predicting optimal routes, traffic management, or improving the user experience. This phase guides the whole design process.

In the second step, we proceed with data collection and preparation. This crucial step involves selecting relevant data sources, cleaning the data to remove inconsistencies and outliers, and preparing the data to make it usable by ML models. Data quality plays a central role in the overall system performance, hence the interest in the semantic layer which will be used to unify the data and solve the interoperability problem [13].

The third step refers to the model selection. Depending on the defined objectives and the characteristics of the data, we choose the most suitable ML algorithms. This selection may vary depending on specific requirements, such as route classification, travel time prediction, or other aspects related to urban mobility.

Model training and evaluation represent an iterative phase, where we adjust the model parameters based on observed performance. This involves using techniques such as cross-validation to ensure model generalization to new data. The design methodology for the Machine Learning (ML) part of our system is not limited only to prediction and optimization but also extends to user experience personalization through user profiling and feedback collection [10].

The implementation of Machine Learning (ML) in our urban mobility system relies on the use of several algorithms, each tailored to specific aspects of personalization, prediction, and optimization. The Machine Learning algorithms presented in this section are classified into three main categories as shown in the Fig. 1 below: supervised learning, unsupervised learning, and reinforcement learning. This classification organizes these approaches based on their methodologies and respective objectives, thus providing a structured foundation for understanding how these algorithms contribute to the design and improvement of our multimodal information system based on agents.
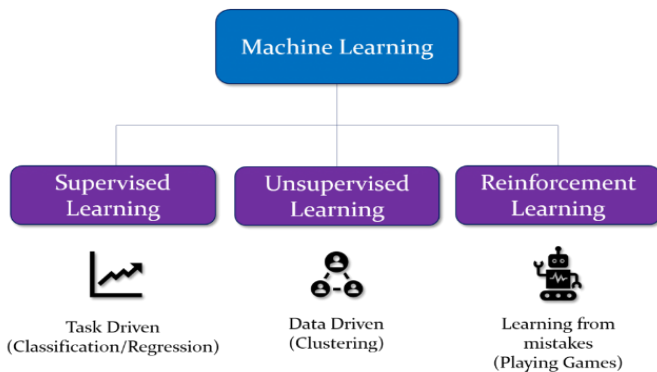

Fig. 1. Machine learning algorithms.

Among the Supervised Learning algorithms, we find:

- Decision Trees: Decision trees are used to model users' choices based on various characteristics. For example, a decision tree could determine what type of transportation a user prefers based on factors such as duration, cost, and safety. These models can be used, in our context, to understand user decision patterns and guide route recommendations accordingly.

Among the Unsupervised Learning algorithms, we find:

- K-Means: The K-Means algorithm is applied to group users into homogeneous clusters, thus identifying similar travel profiles. For instance, this algorithm could help us segment users into clusters to personalize route recommendations based on shared preferences within each group.

Among the Reinforcement Learning algorithms, we find:

- Reinforcement Learning: Reinforcement learning is used to improve the system over time by adjusting recommendations based on user feedback. In our case, considering the rewards and penalties of past route choices, the system adapts to provide more personalized suggestions.

This diversity of approaches allows our system to adapt to the various nuances and complexities of user preferences regarding urban mobility. Thus, in our approach to implementation, we will opt for the k-means algorithm for user classification and segmentation. The Decision Trees algorithm will be used for data learning, thus guiding route recommendations based on user preferences. Concurrently, the reinforcement learning algorithm will be implemented to adapt the system, offering personalized suggestions to each user.

User feedback collection is integrated into the process proactively. Mechanisms are put in place to solicit user feedback on their travel experiences. This feedback is then analyzed using ML models, allowing us to understand changing preferences, anticipate user needs, and continuously improve the service. This entire methodology of profiling and feedback collection using ML aims to create a personalized user experience while providing valuable data for decision-making in our multi-agent system dedicated to urban mobility. The process of profiling and feedback collection using machine learning (ML) is essential for creating a personalized user experience within our multi-agent system dedicated to urban mobility. This process combines the use of user profiling techniques and feedback data collection to optimize interaction between the system and users, while providing valuable insights for decision-making.

- User Profiling: The system collects data on user behavior, such as route preferences, preferred modes of transport, preferred schedules, and other relevant data.

- Feedback Collection: Interactive feedback mechanisms are integrated into the system, allowing users to express their preferences, provide feedback on routes, and rate their overall experience.

- Personalization of User Experience: Using user profiles and feedback information, the system adapts its route recommendations, thus offering a more personalized user experience in line with individual preferences.

- Enhanced Decision-Making: Data from profiling and feedback collection provide crucial information for decision-making in the multi-agent system. This information can be used to adjust recommendation policies, optimize routes, and overall improve service quality.

By integrating profiling and feedback collection through machine learning, our system aims to create a more intuitive and personalized interaction with users, while fueling a virtuous cycle of continuous improvement based on user feedback.

Finally, the implementation of the model in our multi-agent system is carried out in a way that ensures smooth integration with other components, especially the semantic layer and the agents responsible for decision-making. Our ML design methodology is centered on rigor, adaptability, and continuous optimization to ensure high performance in solving urban mobility-related problems.

C. Evolution of the Multi-agent System Architecture for Multimodal Transport

Our previous work [14, 15] has already presented the overall architecture of our multi-agent system for multimodal transport. However, this proposed architecture does not include the integration of the ML component, hence the interest in setting up an evolution of this architecture to include all the components of our system.

A major evolution of our architecture will be the subject of this section and will consist of integrating a dedicated Machine Learning agent, responsible for providing personalized recommendations to each user. This ML agent, now an integral part of our multi-agent system, will play a central role in refining travel needs. By leveraging extensive data on user history, this new agent will be able to analyze trends and individual preferences. Its ability to identify complex behavioral patterns will allow it to suggest alternative routes, thus refining the initial proposal based on each user's specific preferences. This fusion of agent intelligence and the expertise of the new ML agent aims to offer an even more precise, responsive, and tailored urban mobility experience to the evolving needs of each user.

By integrating a dedicated Machine Learning agent, our goal is to go beyond mere feedback collection by establishing a dynamic process of analysis and adaptation. This ML agent, by examining user feedback, will be able to assess customer satisfaction more thoroughly and discern subtle trends in their preferences.

The introduction of our new ML agent marks a significant advancement in our architecture, promising a more sophisticated interaction with the multi-agent system. By leveraging emerging patterns from user history, this ML agent will be able to suggest alternative and tailored routes, thus refining decision-making based on each user's specific preferences. The quality and quantity of the collected data will play a crucial role in the ML agent's ability to develop insightful and relevant travel suggestions. Thus, the establishment of a robust data collection procedure becomes a strategic element, allowing our ML agent to fully leverage available information to offer highly personalized and relevant travel suggestions.

From the PTA agent, we retrieve the initial parameters set by the user, namely the choice of modes of transportation, preferences in terms of time, cost, safety, and number of connections, as well as navigation data, including details of the routes taken. These data are then stored in a MongoDB collection. The ML agent then processes this data and performs its learning model in order to provide new route recommendations and classify each user according to their profile. This process is illustrated in Fig. 2.
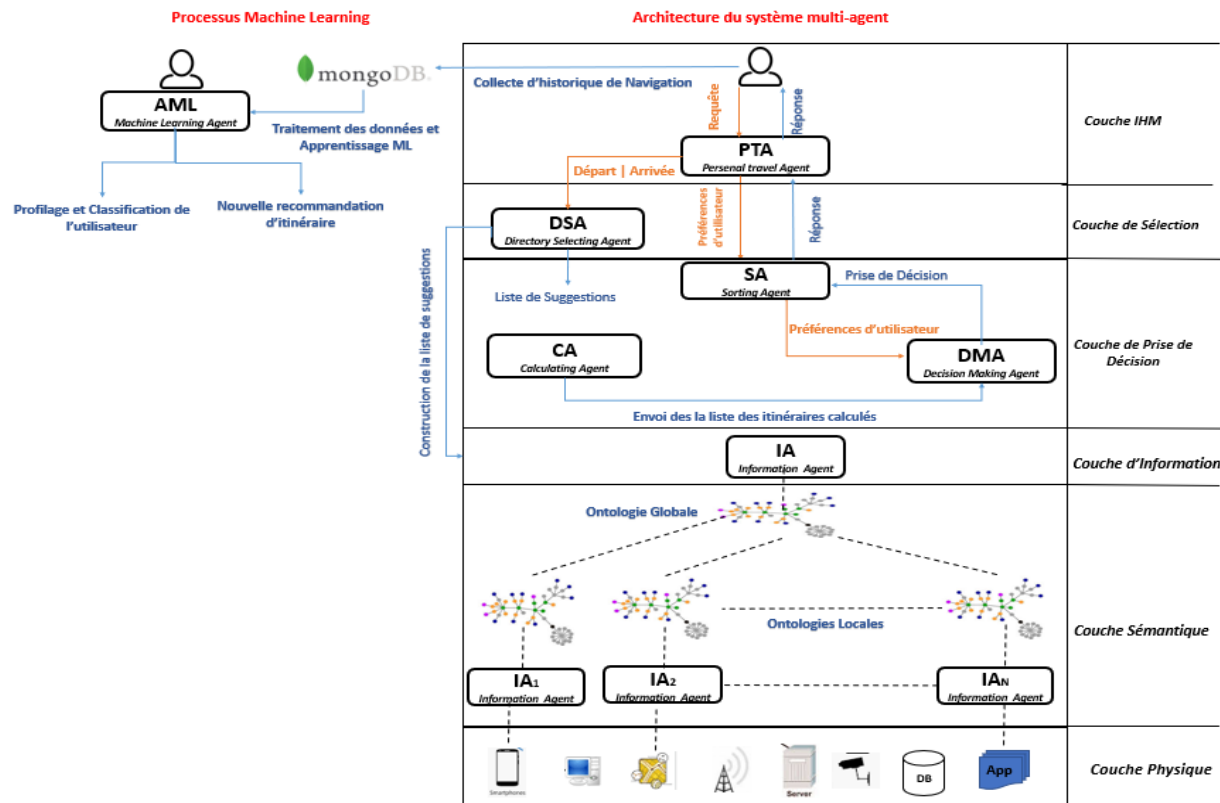

Fig. 2. New Multi-agent System Architecture including ML agent for urban mobility.

IV. ML Algorithms Implementation in the Multimodal Information System

In this section dedicated to the implementation of our solution in the multimodal information system, we will specifically focus on the integration of Machine Learning (ML) techniques to enhance our approach. We will delve into detail on how these algorithms are applied to improve service personalization, user profiling, feedback collection, and ultimately, how they contribute to informed decision-making in the field of urban mobility. The goal is to demonstrate how ML methods can be valuable assets in optimizing the user experience and providing smarter route recommendations in our multimodal information system.

We reiterate that our methodological approach relies on the integration of three Machine Learning (ML) algorithms within our multimodal information system. We will begin our implementation with the reinforcement learning algorithm, focused on analyzing user history to enhance route recommendations [11]. Indeed, the reinforcement learning algorithm is used to train an agent to make decisions in an environment to optimize performance. In parallel, utilizing the Decision Tree algorithm will allow us to gain deep insights into user behavior in the context of urban mobility. Lastly, we will implement the K-means algorithm on a representative sample of 500 users to segment the population into distinct clusters, thus fostering a more personalized approach in our route recommendations. This progressive approach aims to leverage the advantages of each algorithm to optimize the user experience in our system.

A. Data Collection Process

The comprehensive approach used in the data collection process aims to understand user behavior by analyzing their browsing history. User preferences in terms of time, cost, security, and connections are also considered. This approach enables the creation of detailed profiles for each user, thereby contributing to service personalization.

Data collection also includes an analysis of the routes chosen by users, including any modifications made along the way. These diverse data are consolidated to create a solid foundation for the development of an intelligent and adaptive urban mobility system.

The collected data, stored in MongoDB collections, are organized flexibly to allow for precise retrieval. Once captured, these data are used as a crucial resource for machine learning algorithms. The goal is to provide a personalized, responsive, and efficient urban mobility experience by optimizing recommendations and dynamically adjusting the system according to evolving user preferences.

Information regarding each user's past route choices has been collected and recorded in a structured manner. This includes details such as the routes taken, transport preferences, starting and ending points, schedules, etc. These data are stored in MongoDB collections that allow retrieval of the following information for each user: below is an example of a MongoDB collection structure for storing user navigation history and preferences. This collection is created in JSON format, encompassing all user journey information.

In the example illustrated in Fig. 3, we have:

- Each user is identified by a "userID."

- Navigation history is stored as an array "history," where each journey is represented by a document with details such as the departure location, arrival location, time, mode of transport, cost, duration, number of connections, security, and frequented zone.

- User preferences are stored in a "preferences" document, assigning weights to each criterion (time, cost, connections, security).

- Frequented zones are listed in a "frequentZones" array.

- The user's activity time is recorded with start and end times to have an idea of the busiest hours.

```
]{
    "_id": ObjectId("5fdcfb1adbe3507f00f7aa0a"),
    "userID": "user123",
    "historique": [
        {
            "trajetID": "trajet001",
            "depart": "Bd Elqods-Panoramique",
            "arrivee": "Bd Mekka",
            "heure": ISODate("2022-01-01T08:30:00Z"),
            "modeTransport": "Tramway",
            "cout": 10,
            "duree": 15,
            "correspondances": 0,
            "securite": "Moyenne",
            "zoneFrequente": "Ain Chock"
        },
        {
            "trajetID": "trajet002",
            "depart": "Bd La Corniche",
            "arrivee": "Bd Zerktouni",
            "heure": ISODate("2022-01-05T12:15:00Z"),
            "modeTransport": "BUS",
            "cout": 5,
            "duree": 30,
            "correspondances": 1,
            "securite": "Elevée",
            "zoneFrequente": "Maarif"
        },
        // ... autres trajets
    ],
    "preferences": {
        "temps": 0.6,   // Poids associé à la préférence de temps
        "cout": 0.4,    // Poids associé à la préférence de coût
        "correspondances": 0.3,
        "securite": 0.7
    },
    "zonesFrequentes": ["Ain Chock", "Maarif", "Hay Hassani"],
    "heureActivite": {
        "debut": "08:00",
        "fin": "18:00"
    }
-}
```

Fig. 3. Example of user MongoDB data collection.

This organization allows for efficient retrieval of user-specific information. These rich data will serve as raw material for the future machine learning algorithms that we plan to implement.

B. Application of the Reinforcement Learning Algorithm

The reinforcement learning algorithm constitutes a central pillar of our approach, focusing particularly on the in-depth analysis of user history to refine route recommendations. By utilizing the data stored in MongoDB collections, the algorithm explores users' past choices, examining previously taken routes, preferred modes of transportation, as well as adjustments made based on individual preferences [12].

This analysis process enables the reinforcement learning algorithm to discern significant behavioral patterns. By

learning from past experiences, the algorithm can assign rewards or penalties to certain actions or routes, thus contributing to the refinement of future recommendations. For example, if a user prefers shorter routes with fewer transfers, the algorithm will adjust its recommendations accordingly, favoring routes that meet these specific criteria.

A thorough analysis of historical data is conducted to identify behavioral patterns, recurring preferences, and individual user trends. This step allows for insights into past choices. To accomplish this modeling, it is important to define an agent, states, actions, the environment, as well as the reward for each action maintained, as illustrated in Fig. 4.
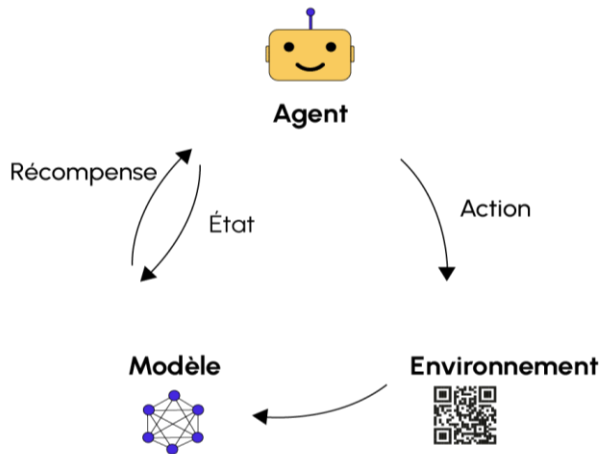

Fig. 4.   Reinforcement algorithm modeling.

Based on the identified behavior models, the reinforcement algorithm has been developed. It utilizes historical data to dynamically adjust the weights of criteria in route recommendations, giving more importance to aspects preferred by each user. The development of the reinforcement algorithm involves the practical implementation of agent logic and interactions with its environment. In our urban mobility context, the reinforcement algorithm aims to recommend personalized routes based on the user's past actions and the evolution of the environment. The approach aims to leverage these histories to dynamically adjust route recommendations, paying particular attention to recurring choices and emerging user preferences. In essence, this approach is based on the idea that users' past decisions can provide valuable insights into their future preferences. Thus, by understanding and modeling these behaviors, the reinforcement algorithm contributes to further personalizing route suggestions, thereby enhancing the overall user experience of the multimodal system.

C. Application of the Decision Tree Algorithm

In this section, we will delve into the application of the Decision Tree algorithm within our multimodal transportation information system dedicated to urban mobility. The main objective is to better understand the decision-making patterns of users when they choose their routes by identifying the key factors that influence these choices.

The Decision Tree algorithm is a supervised learning method that is well-suited for analyzing complex decisions. By successively dividing the data into subsets based on specific features, this algorithm allows for easy visualization and interpretation of decision-making processes. The Decision Tree algorithm, or decision tree, is a supervised learning method used in the field of artificial intelligence. Its main objective is to model decision-making by creating a tree-like structure based on specific criteria. This decision tree allows for a clear and hierarchical visualization of the different possible decision paths based on the data characteristics. The main concepts associated with the Decision Tree algorithm are illustrated in Fig. 5.
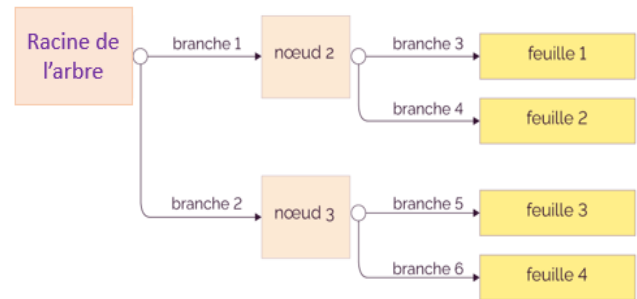

Fig. 5.   Representation of the modeling of the decision tree algorithm.

Root Node: The initial division of data based on the criterion that maximizes the separation of classes.

Internal Nodes: Decision points in the tree that determine the next feature to evaluate.

Leaves: Terminal nodes of the tree representing classes or predicted values.

Branches: Paths between nodes, indicating how the data is segmented.

Splitting Criteria: Features guiding the division of data at each node.

Building the decision tree is a key step in applying the Decision Tree algorithm to model user behavior in urban mobility. Here's how this step was carried out:

Selection of Root Node: The algorithm begins by selecting the feature, at the root node, that divides the dataset into the most homogeneous subsets in terms of user behavior. The chosen feature at this stage is the one that provides the most information for decision-making. In our case, we set the feature as the choice of transportation method (bus, tram, etc.), with criteria such as availability, frequency, cost, security, travel duration, and number of transfers for each mode of transport.

Data Set Division: Once the feature is selected, the dataset is divided into subsets based on the possible values of this feature. Each subset is associated with a branch from the root node. In our case, we separated the data into subgroups for each mode of transport based on the following splitting criteria: Frequency of use, availability, security, pricing, travel duration, and number of transfers.

Process Repetition: The division process is repeated for each child node, selecting the most informative feature each time to split the current subset. This process continues until a stopping condition is met, such as a maximum number of levels in the tree or sufficient purity of the subsets. In our case,

the features for Child Nodes are the specific line choices for each mode of transport, and for criteria, we have average travel time, serviced stations, security, cost, and number of transfers.

Class Assignment: Each leaf of the tree represents a class or final decision. The examples in each leaf share similar characteristics and are grouped based on these similarities. These classes can be interpreted as user behavior segments. In our case, the defined classes depend on different user segments based on transportation preferences. Examples of Classes: "Users preferring tram for safety and speed", "Users opting for express bus due to lower cost".

Validation and Adjustment: Once the tree is built, it is typically validated using separate data to assess its generalization ability. If necessary, adjustments can be made to optimize the model's performance. To validate the model in our case, we used a separate dataset to evaluate the predictive performance of the tree. Following this evaluation, we adjusted the model by adapting some parameters, such as the depth of the tree, to optimize generalization.

The diagram in Fig. 6 below provides a simplified example of a decision tree schema for modeling user behavior in the context of urban mobility. We consider criteria such as security, cost, travel time, and number of transfers. Note that this is an abstract representation, and the criterion values may vary depending on the data.
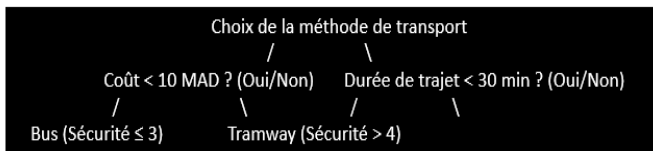

Fig. 6. Representation of the decision tree algorithm.

In this representation, if the trip cost is less than 10 MAD and the travel duration is less than 30 minutes, the user may choose the bus if security is moderate (security rating $\leq$ 3) or the tram if security is high (security rating > 4). In this context, we identify two types of users: "*Users preferring the tram for safety and speed*" and "*Users opting for the bus due to lower cost.*"

In conclusion of the Decision Tree algorithm application, we have successfully modeled user preferences in urban mobility. By understanding the criteria influencing their choices, such as safety, speed, and cost, we are better equipped to personalize route recommendations and enhance the experience of each user.

D. Application of the K-Means Algorithm

The K-means, also known as the centroid-based clustering algorithm, is a method of unsupervised machine learning. Its main objective is to divide a dataset into several clusters, where each cluster is represented by a central point called a centroid. The algorithm assigns each data point to the cluster whose centroid is closest to it, minimizing the sum of squared distances between the data points and their respective centroids.

In the context of urban mobility, applying K-means to our user data will allow us to group travelers with similar

behaviors. These behaviors may include specific preferences such as preferred mode of transportation, frequent travel times, visited geographical areas, etc. By segmenting users meaningfully, we can better understand the different profiles within our system.

The K-means algorithm is a clustering algorithm that aims to partition a dataset into K clusters, where each cluster is characterized by its center of gravity, called a centroid. The number of clusters K is a parameter that the user must specify before applying the algorithm. The principle of the K-means algorithm can be summarized in several steps as illustrated in Fig. 7.
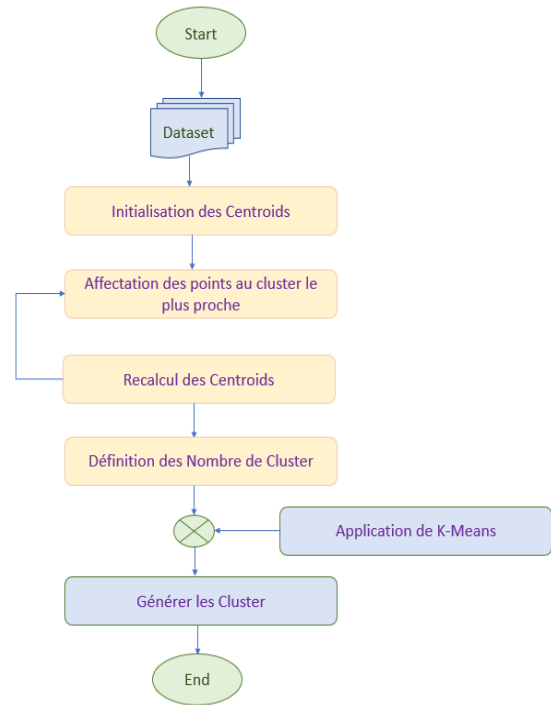

Fig. 7. K-Means algorithm steps.

In the context of our analysis, we utilized the elbow method to determine the optimal number of clusters (k) in the context of multimodal transportation and urban mobility. The results indicate that the elbow of the inertia cost graph was identified around k=3. This suggests that three clusters appear to be the optimal number to segment user data based on their preferences, behaviors, or characteristics related to urban mobility. Each cluster could represent a distinct group of users sharing similarities in their choices of transportation modes, preferred routes, or other relevant criteria. Thus, with k=3, we are able to better understand the diversity of user behaviors in the context of multimodal transportation, which can then guide the customization of services and recommendations to more accurately meet the specific needs of each identified group. Therefore, the implementation of the K-means algorithm in our context resulted in three different clusters. In the following chapter, we will examine the results of this experimentation in detail, including the types of clusters and the characteristics identified for each profile type.

V. EXPERIMENTATION AND RESULTS

This section is dedicated to analyzing the results obtained from the application of key algorithms in our system, while providing an in-depth insight into trends, behaviors, and user preferences regarding urban mobility. These results not only strategically group users but also shed light on how this information can be leveraged to improve system efficiency and respond more personally to the needs of each user.

A. Result of Applying the Reinforcement Learning Algorithm

The reinforcement learning algorithm, based on each user's navigation history, provides the system with deep insights into the preferences and travel behavior of each individual. This thorough understanding of the user serves as a solid foundation for subsequent steps, including the application of k-means and the decision tree algorithm. Using a sample of 500 users, the reinforcement learning algorithm was able to categorize the types of navigation histories as follows:

1) Fast profile: This type of user prioritizes the fastest routes and prefers express modes of transportation and direct routes. From these results, we observed that this type is less sensitive to costs and transfers.

2) Economic profile: This type of user prioritizes the most economical routes and favors routes with reduced costs, even if they involve longer travel times. They may also opt for cheaper modes of transportation.

3) Balanced profile: This type of user seeks a balance between cost, duration, and comfort. They accept moderate travel times for balanced routes and also have moderate sensitivity to costs and travel times.

4) Cautious profile: This type of user places great importance on safety by avoiding risky areas, even if it means taking detours. They may choose safer modes of transportation, even if they are slightly more expensive.

These profiled pieces of information will then be used in the classification process with k-means and the decision tree, allowing for a more tailored customization of route recommendations for each user.

B. Result of Applying the Decision Tree Algorithm

The results interpreted in this section are based on the data collected from 500 users in an urban mobility system, including information on estimated travel time, associated cost, and route choices. To explore the results, we will first follow the modeling process of the decision tree algorithm according to the following steps:

1) Data collection: We have a dataset with examples of users. We remind that data collection is done through a JSON file from the MongoDB collection. MongoDB is used as a database to store user information, including their navigation history, preferences, activity hours, frequented zones, etc. The data stored in this MongoDB database is then used for training the decision tree model and other machine learning algorithms. In the example below, we present only (see Table I) a sample proposal made by the system to represent the results:

TABLE I.   SAMPLE DATA FOR DECISION TREE APPLICATION

Travel Duration	Cost	Safety on a scale of 1 to 5	Route Choice
20Min	5	1	A
30Min	8	2	A
25min	6	1	B
35min	10	3	B

2) Creation of the decision tree: Root of the Tree: Division based on the estimated travel duration.

- Internal Node 1: Division of trips under 30 minutes.

- Internal Node 2: Division of trips with a perceived safety of 4 or more.

  o Leaf 1: Prediction "A" for the route.

  o Leaf 2: Prediction "B" for the route.

- Internal Node 3: Division of trips with a perceived safety of less than 4.

  o Leaf 3: Prediction "B" for the route.

- Internal Node 4: Division of trips of 30 minutes or more.

  o Leaf 4: Prediction "B" for the route.

  o Leaf 5: Prediction "A" for the route.

3) Detailed interpretation: We added a division based on the number of transfers at each internal node. The second internal node divides trips with a perceived safety of 4 or more based on the number of transfers. The third internal node divides trips with a perceived safety of less than 4 based on the number of transfers. According to the results obtained on the types of decisions made by users "Users preferring the tramway for safety and speed," "Users opting for the express bus due to lower cost," here is the interpretation:

Users preferring the tramway for safety and speed: We identified a group of users through the application of the Decision Tree algorithm who attach great importance to safety and speed when choosing their mode of transportation.

Users opting for the express bus due to lower cost: We identified a segment of users through the application of the Decision Tree algorithm for whom cost is a determining factor in the choice of mode of transportation. The model's decisions indicate that these users are directed towards the bus, which is probably associated with lower costs compared to other modes of transportation.

In summary, these interpretations highlight the specific preferences of certain user groups regarding criteria such as safety, speed, and cost. This information is crucial for customizing route recommendations and improving the user experience in the urban mobility system.

4) Model evaluation: Evaluating a decision tree model typically involves using a separate dataset called a test set.

This test set consists of data that the model has not seen during its training. Here's how it could be done:

Test Set: A separate dataset from the training set, usually around 20 to 30% of the total data, is reserved for evaluation.

Prediction on the Test Set: The decision tree model is used to predict route choices on this test set.

Comparison with True Values: The model's predictions are compared to the actual route choices of the test set.

Evaluation Metrics: Different metrics can be used, such as accuracy (the percentage of correct predictions), confusion matrix, recall, precision, etc.

This allows the model trained on a sample to make predictions on new cases, thus contributing to customizing route recommendations based on user preferences and characteristics. This approach would take into account the number of transfers in addition to travel duration, cost, and perceived safety to understand user route choices in an urban mobility context.

C. Result of Applying the K-Means Algorithm

The application of the K-Means algorithm in the context of urban mobility has generated significant results that help to better understand user behaviors and preferences regarding transportation. The clusters obtained represent distinct groupings of users based on their criteria for route selection, such as travel time, cost, number of transfers, safety, etc. These results play a crucial role in personalizing route recommendations, thereby contributing to an optimized user experience tailored to each profile.

Using a sample of 500 users with various characteristics associated with their travels and navigation history, we applied the K-Means algorithm to group these users into clusters based on these characteristics. On the same dataset presented previously, we present a proposal made by the system to represent the results with the following features:

TABLE II. SAMPLE DATA FOR K-MEANS APPLICATION

Users	Cost	Safety on a scale of 1 to 5	Number of transfers
User 1	5	1	1
User 2	8	2	0
⋮	⋮	⋮	⋮
User 500	10	3	2

These data will be used for applying the k-means algorithm to group these users into three clusters as shown in Table II. We start by initializing the centroids by randomly selecting three users from the database to represent the initial centroids. Each user is described by attributes such as travel time, cost, number of transfers, security, etc. According to the elbow method applied, we obtained k=3 clusters as the optimal number of clusters for our context. The results obtained with k=3 provide a balanced representation of the different user behaviors. Then we apply our K-means algorithm, the implementation results in three distinct clusters. These clusters help understand users' travel preferences and can be used to personalize route recommendations. These clusters also classify users into three different profiles: Economic, Standard, and Premium.

Cluster 1 Economic Profile: the chosen criterion among cluster users is the lowest possible cost. These users are very cost-sensitive and are willing to sacrifice travel time, security, or the number of transfers to minimize their expenses. They prefer economical options even if it means compromising on other aspects. Generally, these users prefer to use the bus as a means of transportation.

Cluster 2 Standard Profile: These users choose criteria such as acceptable security, minimal number of transfers, flexibility in terms of time and cost. These users are moderate and flexible, seeking a balance between cost, security, and travel time. They are willing to accept acceptable security and a minimal number of transfers, while maintaining some flexibility in terms of time and cost.

Cluster 3 Premium Profile: the predominant criteria in this cluster are: minimal travel time, high security. These users are willing to pay a premium price for an optimal travel experience. They seek routes with minimal travel time and maximum security. Cost is a secondary consideration as long as the other criteria are met.

Each cluster represents a specific segment of travelers with distinct preferences, which allows for adapting route recommendations and services to meet their specific needs. These differentiated profiles will serve as a solid foundation for further personalizing route recommendations, thereby contributing to a more tailored and efficient user experience. The combination of these results reinforces the relevance of our multi-agent approach, where each agent can dynamically adapt to the specific characteristics of each user cluster, thus improving the overall management of urban mobility. The next step will be to further explore these profiles to refine recommendations and offer an even more personalized user experience.

D. Discussion and Results Evaluation

The algorithms implemented in our system were chosen to address specific objectives, thus contributing complementarily to the overall decision-making process in urban mobility.

The primary objective of the Reinforcement Algorithm is to learn from users' past behaviors to recommend personalized routes. The Decision Tree Algorithm was implemented to understand the factors influencing users' route choices. The aim is to gain clear insights into the specific motivations guiding users' choices and to better understand and analyze the decisions made by users. The purpose behind implementing the K-Means Algorithm is the classification and segmentation of users based on their general profiles. This approach aims to simplify personalization by grouping users with similar preferences, thereby facilitating the recommendation of routes tailored to each segment.

In the same context, each algorithm has made a unique contribution to our understanding of user preferences and behaviors in the complex context of urban mobility.

- Reinforcement Algorithm: By identifying four distinct clusters, this algorithm offers detailed granularity to understand the various nuances in users' preferences. The "Prudent," "Economic," "Fast," and "Balanced" clusters enable route recommendations to be adapted according to these specific profiles.

- Decision Tree Algorithm: With two well-defined clusters, this algorithm highlights specific characteristics influencing route choices. The categories "Users preferring the tramway for safety and speed" and "Users opting for the bus due to lower cost" offer clear insights into users' motivations.

- K-Means Algorithm: By classifying users into three clusters – "Economic," "Standard," and "Premium" - K-Means offers segmentation based on criteria such as cost, speed, and safety. This approach simplifies personalization by grouping users with similar preferences.

Indeed, without the Reinforcement Algorithm, we would not benefit from a robust user navigation history. It provides an essential foundation by learning from past behaviors and creating relevant clusters, thus enabling fine personalization.

Likewise, in the absence of the Decision Tree Algorithm, our ability to understand how decisions were made by users would be limited. This algorithm offers valuable insights by identifying clusters based on specific patterns, thereby contributing to the transparency of the decision-making process.

Lastly, the integration and implementation of the K-Means Algorithm have been highly successful in classifying users into well-defined profiles. K-Means not only simplifies user segmentation by grouping those with similar preferences but also facilitates the recommendation of routes tailored to each profile.

This variety of algorithms has been carefully integrated to leverage their respective strengths, thus creating a robust system capable of adapting to individual user needs while optimizing overall urban mobility management.

By combining these approaches, our system aspires to offer a comprehensive solution that integrates fine-grained behavior-based personalization, understanding of influential factors, and global classification of user preferences. The ultimate goal is to provide route recommendations that are both effective for individual users and beneficial for overall urban mobility management.

VI. Conclusion

The introduction of a machine learning (ML) approach represents a significant evolution. This work has constituted a deep dive into the implementation and results of machine learning algorithms, in which we have addressed the outcomes and evaluation of these different ML algorithms integrated into our system. Each of the three algorithms, namely Reinforcement, Decision Tree, and K-Means, has been meticulously applied to address specific objectives within the system.

This combination of approaches has enriched the system's ability to provide smarter and more personalized urban mobility solutions. By integrating these algorithms complementarily, we have succeeded in understanding the complexity of user preferences and dynamically adapting to their needs. As part of the system's continuous improvement, this approach enables adaptation to changing user needs and ensures efficient urban mobility management.

In conclusion, this technical achievement harmoniously blends the domains of semantics, decision-making, and machine learning to give rise to an innovative agent-based information system dedicated to the constant optimization of urban mobility.

References

[1] O. Parishwad, S. Jiang, K. Gao. Investigating machine learning for simulating urban transport patterns: a comparison with traditional macro-models Multimodal Transp., 2 (3) (2023), Article 100085.

[2] Q. Ma, S. Li, H. Zhang, Y. Yuan, L. Yang. Robust optimal predictive control for real-time bus regulation strategy with passenger demand uncertainties in urban rapid transit.

[3] M. Zhong, J.D. Hunt, J.E. Abraham, W. Wang, Y. Zhang, R. Wang. Advances in integrated land use transport modeling.Advances in Transport Policy and Planning, Vol. 9, Elsevier (2022), pp. 201-230.

[4] J.Wang, X.Kong , F. Xia, Lijun Sun, Urban Human Mobility: Data-Driven Modeling and Prediction, ACM SIGKDD explorations newsletter, 2019.

[5] HR Pamuluri, Predicting user mobility using deep learning methods,Master Thesis, 2020.

[6] A. Makanadar and S. Shahane, "Urban Mobility: Leveraging AI, Machine Learning, and Data Analytics for Smart Transportation Planning- A case study on New York City," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10307632.

[7] M. Smerkol, M. Sulajkovska, E. Dovgan, M. Gams, Traffic simulation for mobility policy analysis, 2021.

[8] E. Dovgan, M. Smerkol, M. Sulajkovska, M. Gams, Supporting decision-making in the urban mobility policy making, 2021.

[9] M. Smerkol, M. Shulajkovska, E. Dovgan, M. Gams, Visualizations for mobility policy design, 2021.].

[10] I. Kaczmarek, A. Iwaniak, A. Swietlicka, ´ M. Piwowarczyk, A. Nadolny, A machine learning approach for integration of spatial development plans based on natural language processing, Sustain. Cities Soc. 76 (2022) 103479.

[11] M. Sulajkovska, M. Smerkol, E. Dovgan, M. Gams, Machine learning-based approach for estimating the quality of mobility policies, 2021.

[12] B. Charbuty, A. Abdulazeez, Classification based on decision tree algorithm for machine learning, Journal of Applied Science and Technology Trends 2 (2021) 20–28.

[13] LARIOUI J., & EL BYED A., Towards a Semantic Layer Design for an Advanced Intelligent Multimodal Transportation System. International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9(2): 2471–2478. https://doi.org/10.30534/ijatcse/2020/236922020.

[14] LARIOUI J., & EL BYED A., Multi-Agent System Architecture Oriented Prometheus Methodology Design for Multi modal Transportation. International Journal of Emerging Trends in Engineering Research, 2020, 8(5): 2118-2125. https://doi.org/10.30534/ijeter/2020/105852020.

[15] LARIOUI J., & EL BYED A., Multi Agent System based on MCDM Approach for Multi Modal Transportation Problem Resolution. Journal of Hunan University (Natural Sciences）, 2021, Vol. 48. No. 7 : 1-10.

[16] LARIOUI J., & EL BYED A., A Multi-Agent Information System Architecture For Multimodal Transportation. In Embedded Systems and Artificial Intelligence Proceedings 2019. Fez, Springer, 2019: 795-803. https://doi.org/10.1007/978-981-15-0947-6_75.

[17] LARIOUI J., & EL BYED A., An Advanced Intelligent Support System for Multi-modal Transportation Network Based on Multi-Agent Architecture. Advanced Intelligent Systems for Applied Computing Sciences, 2020, 4: 98-106. http://dx.doi.org/10.1007/978-3-030-36674-2_10.

[18] LARIOUI J., & EL BYED A., An Agent-based architecture for Multi-modal Transportation Using Prometheus Methodology Design. In Innovations in Smart Cities Applications Volume 4. Virtual Safranbolu, Türkiye, Springer, 2021: 325-343. https://link.springer.com/chapter/10.1007/978-3-030-66840-2_25.