Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2020, 7(12):63-80

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Analysis and forecast of road traffic by the method of Data Mining and Data Warehouse: case of the northwest area of Lomé town in Togo

Mawugno Koffi Kodjo, Akim Adekunlé Salami, Eyouléki Tcheyi Gnadi Palanga, Ayité Sénah Akoda Ajavon

Research Team in Engineering sciences (ERSI), Department of Electrical Engineering, National High School of Engineers (ENSI), University of Lomé, 01 BP 1515 Lomé 01, TOGO

Abstract In the last decades, the big cities in Africa have been growing rapidly in population and surface area. This urban growth is at the root of the increase in different traffics including road traffic which is limited by the saturation of road networks at rush hours, noise pollution and pollution due to high CO_2 production, which is the main cause of the greenhouse effect, difficulties in parking. It is necessary to correct these limits as much as possible in order to improve road exploitation. In this article, our study is aimed at establishing forecasts of the road network parameters for the northwest part of Lomé town in Togo (vehicle flow on a section of road, occupation rate of a section of road and travel time on a section of road) using the techniques of Data Mining and Data Warehouse. It involves the collection and search for data relating to this zone to define the optimal run between two points in the studied zone.

Keywords Data Warehouse, Data Mining, road traffic, decision-making analysis, forecast analysis, virtuous circle of data mining

1. Introduction

Today, urban expressways in large metropolitan areas are experiencing significant and growing congestion. Congestion is a phenomenon that occurs when the number of vehicles on roads is greater than the capacity of the infrastructure. If demand exceeds the track capacity, then vehicles will be slowed down towards forming a cap causing network saturation. This represents many hours lost by users but also causes additional pollution of various kinds (chemical and noise) as well as road accidents, traffic jams, parking problems, etc., and therefore has both an economic and an environmental cost [4, 5, 6, 7, 25, 28].

It therefore becomes imperative to try to understand this phenomenon and to find traffic observation tools to improve the operation of road networks, without widening the existing roads and on so doing remedy the congestion problem. It is in this perspective and with the rapid development of information and communications technology in recent decades that network managers are turning to new innovative solutions such as computerized traffic data processing and dynamic traffic simulation [13, 24]. These tools put in place for effective traffic management strategies are tools for decision support but also for evaluating the traffic management actions [13, 25, 28].

The road network in the Northwest part of the city of Lomé in Togohas a large volume of complex data [8] and requires effective systems to view and analyze trends for its sustainable management. The current state of knowledge on the modeling of agent displacement in a space shows that generally, researchers use multi-agent systems to study road traffic. These systems are combined with the use of cellular automata [21, 23, 28, 29, 30].

Although it is a very common phenomenon, traffic congestion brings together very different realities, which must be taken into account. If simple models give satisfactory results, their limits must not be forgotten, since the behavior of motorists can be very complex (change of row, overtaking). Cellular automata appear to be a promising pathway for modeling such phenomena, as they accomplish the junction between microscopic and macroscopic viewpoints. Their plasticity makes it possible to test many different models in order to adapt the models to reality. However, the behavioral aspect of urban traffic has recently led to statistical methods being grafted onto flow models, probability models, and cybernetic systems. In this article we base our analysis and forecast on the method of Data Mining and Data Warehouse.

For a quick and easy analysis of a network, the data must be summarized and grouped together in a single system: the best approach to reach this solution is the construction of a Data Warehouse and implementation an OLAP application (on Line Analytical Processing) because on one hand, Data Warehouse can organize and conduct historical data for decision-making and on the other hand OLAP provides data displays as tables and statistical graphs of all kinds, which facilitates analysis.

To achieve our objective, this report adopts an approach consisting of three main sections. In this article our study is aimed at establishing forecasts of road network parameters for the North-West part of Lomé, capital of Togo (speed of circulation, rate of flow, rate of road occupation). To this end, we adopt an approach consisting of four main sections. In the first section, we define the concepts of data warehouse and data mining. In the second, we determine the model of road traffic satisfying the objectives of the study and adapt it to the process of data collection, stocking and searching [29, 30]. The third section gives an account of the selection of decision-making tools to be used and their application. Finally, the fourth and the last section permits to validate the used mathematical road traffic model.

2. Data Warehouse and Data Mining Methods

The concept of data warehouse was formalized for the first time in 1990 by INMON W. H.. According to INMON, "*The Data Warehouse is a collection of data subject oriented, integrated, nonvolatile and historized organized for the support of the decision support process*"[14]. Indeed, in a data warehouse it is necessary to maintain data history. Thus, for the same request made two months apart, specifying the reference date of the record will give the same result. With a structure defined by four classes of data organized in a historical axis and a synthetic axis as illustrated in Figure 1, the Data Warehouse aims to integrate different operational databases, allowing access to information historized and provide analysis on that data for decision making.

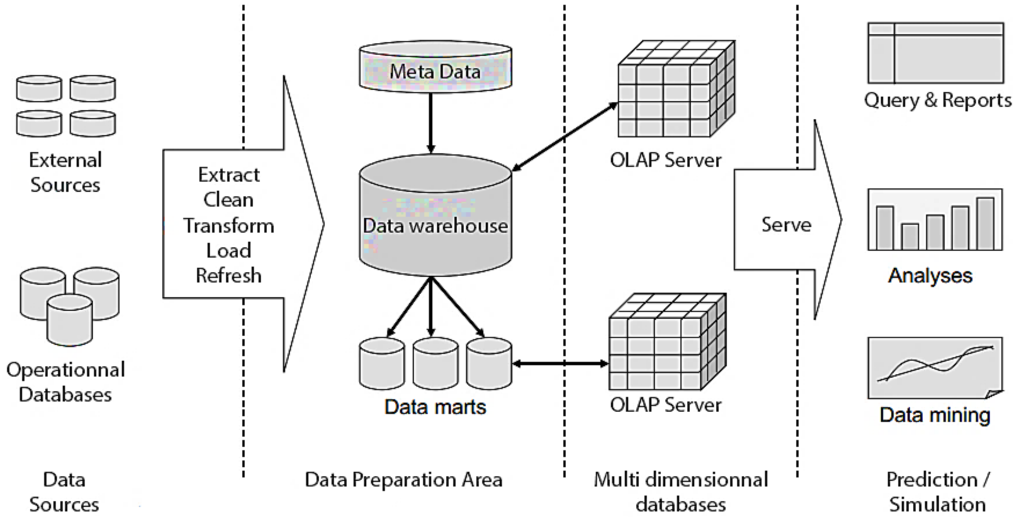

Figure 1: Architecture of Data Warehouse

2.1. Concept of Data Warehouse

According to PASQUIER N. [22], citing INMON W. H. [14],a data warehouse is "a collection of data oriented subject, integrated, non-volatile, historical, and organized to support the process of helping the decision" [19,26, 29].

Its construction requires a modeling phase. Dimensional modeling Data Warehouse highlights the concepts of dimensions and facts. A dimension corresponds to the axis of the analysis; it is described by one or more hierarchies. A fact models the subject of analysis; it consists of corresponding measures to the information of the analyzed activity. For example in the context of this paper, the traffic situation can be analyzed according to dimensions: Day, Year, Hour, Junction and Section.

But the Data Warehouse alone is sterile for the establishment of a decision-making tool unless and it is organized. Indeed, after the implementation of the data warehouse, you must find tools and techniques that can transform stored data into information for decision making. It is in this sense that Data Mining and tools such as OLAP are used for extracting and exploiting these data for analysis and decision making, while facilitating data navigation. Figure 1 illustrates the global architecture of the Data Warehouse system [8,19,22,29].

2.2. Concept of Data Mining

The term Data Mining means literally "drilling for data". As in the field of drilling its objective is to extract a resource which in our case is knowledge. That permits to reveal knowledge by means of the techniques of stocking and exploring data. Data Mining is a "complex of techniques permitting to transform data into knowledge or "informational action by using statistics and artificial intelligence to identify and exploit the useful information dozing in the database".

The principal tools of Data Mining, OLAP (On Line Analytical Process) are based on the multidimensional data base intended to exploit rapidly the dimensions of population data. Main part of OLAP solutions is based on the same principle: to restructure and to stock in the multidimensional frame the data issued from flatfiles or relational data. This multidimensional frame is also known under the name "hypercube". In order to formalize the OLAP frame at the end of 1993 E. F. CODD defines 12 rules which must be respected by the entire managing system [3,7,8,9]. The OLAP architecture servers decline in different versions: ROLAP (*Relational OLAP*), MOLAP (*Multidimensional OLAP),* and HOLAP *(Hybrid OLAP).*

Data mining is not a miracle medicine; anyway a multitude of problems of intellectual order could be regrouped in their formalization into one of the following tasks: Classification, estimation, forecast, grouping by similarity, segmentation (or clustering), description and optimization. The description being at the base of the whole task is realized by the technique called "analysis of consumer basket". To solve many problems, it is common for each potential, to add an evaluation function. The objective of optimization is to maximize or minimize that function.

Generally, the technique of data mining is not relevant to realizing a simple exploration. It is better to put it into a more global context called "the virtuous circle" presented in Figure 2 [8,9,10].

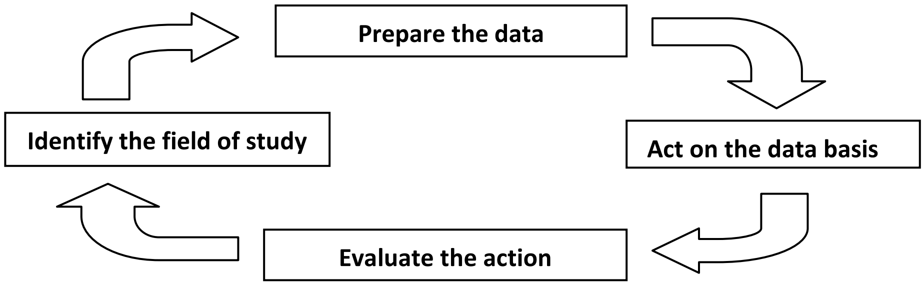

Figure 2: The virtuous circle of Data Mining

3. Model of Road Traffic Satisfying the Objectives of the Study

To model road traffic it is possible to adopt at least two points of view. The first approach, called that of Lagrange, is clearly more precise but much more difficult to apply because it requires a large number of data (position and speed of all vehicles at each moment). This approach is also called "microscopic modeling".

Quite opposite to the previous one, the second approach called Euler is less precise, but easier to realize by the computer. It stores at each moment only the density of vehicles on the small segments of the road. Besides, once the evolution of the number of vehicles on each small road segment is known, it becomes possible to reconstruct to a good approximation the each small road segment beginning from their initial position. This is "macroscopic modeling."

3.1. Microscopic Modeling: Law of Chase

The microscopic modeling of road circulation was part of the primary contributions to traffic science, thanks to the initial progress achieved by the models of GAZIS & al. [11,12]. The management of traffic flow in microscopic models is principally realized through the law of Chase describing the movement of vehicles as a function of vehicles which are around it on its way, and by another model aiming to show the changes of ways or overtaking.

The laws of Chase contained in the microscopic models are based on the following principle: the drivers react in their driving actions at stimuli generated by one or many surrounding vehicles. This reaction is revealed in a certain time gap τ, corresponding to the time of driver reaction and according to a certain sensibility to the stimulation.

$$\mathrm{Reaction}(t+\tau)=\mathrm{Sensibility}\times\mathrm{Stimulus}(t) \quad (1)$$

The first models conceived by General Motors engineers were based on the hypothesis that the acceleration of each driver is a linear function of the speed difference relative to the preceding vehicle, from where, comes the appellations car-following and a linear model associated with that model [2,11, 12]:

$$\ddot{x}_{i+1}(t+\tau)=\alpha\left(\dot{x}_i(t)-\dot{x}_{i+1}(t)\right) \quad (2)$$

where "α" is the sensibility of a driver to the speed differential, x is mono-directional space as a function of time t.

This first model was not satisfactory particularly because it did not take into consideration the space between two cars and their speed. The works of GAZIS D. C., HERMAN R. and ROTHEBURYR. W. [12] permitted to fill in these gaps by proposing the general expression (3) for the models of car-following:

$$\ddot{x}_{i+1}(t+\tau)=\alpha\left(\dot{x}_i(t)-\dot{x}_{i+1}(t)\right)\frac{\dot{x}_{i+1}^{m}(t+\tau)}{\left(x_i(t)-x_{i+1}(t)\right)^{l}} \quad (3)$$

The parameters *l* and *m* permit to stipulate the forms of independence in relation to the space between and the speed of cars. Other models have been proposed to precise certain interactions, but the basis of these models, described by the general formulas above, remain the same.

3.2. Macroscopic Modeling

Macroscopic models consider traffic as a continuous and homogenous flaw of vehicles. Built by analogy with the dynamics of liquids, they were introduced by the basic articles of LIGHTHILL, WHITHAM and RICHARD in 1954 used in [4,5,6, 27].

Traffic is described by the concentration χ (cars in one unit of length) of cars moving at speed V and generating a rate of flowφ(in cars in one unit of time). All these variables are defined as continuous functions of time t and of a supposed mono-directional space x. It is clear, that in the case of road traffic, the hypothesis of continuity is rather rough, empty spaces in the circulation being a very frequent fact in relation to the number of car particles.

The rate of flow is defined classically as a result of concentration and of the speed of traffic flow [1,5,6,7,23]:

$$\chi(x,t)=\varphi(x,t)V(x,t) \quad (4)$$

where V(x, t) is the average speed of cars constituting the flow.

The fundamental law of flow is the conservation of cars, which thus corresponds to the equation of conservation (5).

$$\frac{\partial\chi(x,t)}{\partial t}+\frac{\partial\varphi(x,t)}{\partial x}=0 \quad (5)$$

In fact, the cars are assimilated with material points (according to the logic of gas dynamics, where the molecules of gas are assimilated with material points). This model is classically called model LWR according to the names of the authors of the basic articles (LIGHTHILL, WHITHAM and RICHARDS) [3]. The disadvantages of macroscopic modeling are noted in the situations at the start at green lights (or more generally, the restart after stopping) [4, 5, 15].

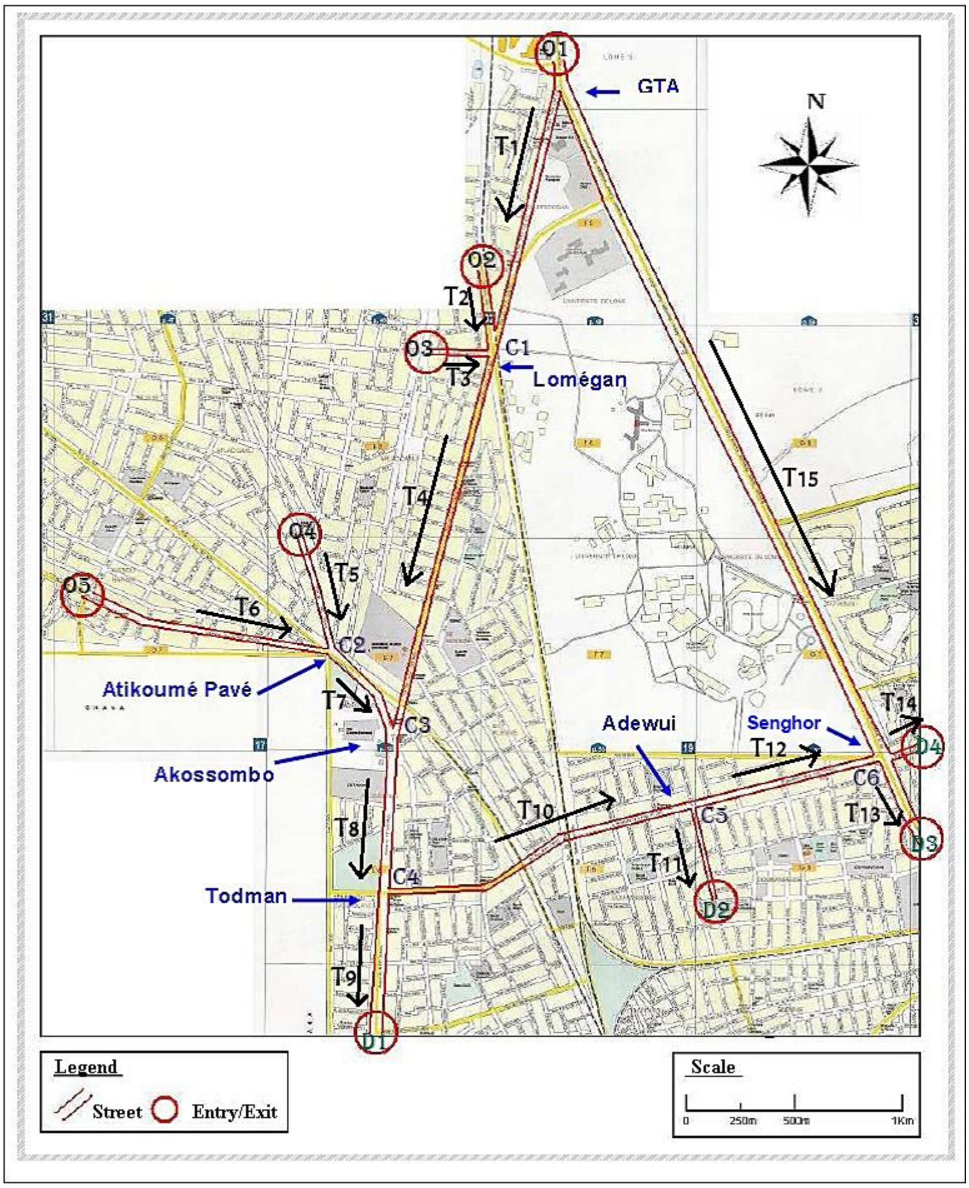

Figure 3: Road network of the study area (Lome Northwest)

3.3. Hybrid Modeling

The "Macro Particle Simulation Model" model calculates the traffic flow in the hybrid form: traffic is represented at the same time in the form of flow (concentration being calculated by space-time discretization) and in the form of packets of vehicles. The importance of this form of modeling is that it permits an individual representation of the decisions of users in the choice of their itinerary in the network (in the affectation model) [5,12]. The works of BOURREL [4], applied to hybridization by coupling between models. The idea of coupling of models is to apply the more adapted model to the representation of different zones of the system: microscopic model in the zone where the phenomenon is poorly described by the macroscopic model, the macroscopic model elsewhere; and all intended to limit the number of calculations [4,5,13,18].

The road network in the study area thus defined is composed of 6 crossroads as shown in Figure 3 [16]: Lomegan, Atikoume Pave, Akossombo, Todman, Adewui and Senghor. Each crossroad is fed by several sections.

3.4. Objectives of the Study

This study involves the development of a support expert decision tool applied to traffic management in the northwest part of the city of Lomé, which allows to:

- receive real -time traffic predictions for the study area;
- apply different management actions in real -time traffic based on the results of dynamic simulations;
- estimate the average travel time of vehicles on a section of road therefore determine the cost of travel;
- detect a green wave to minimize downtime and thereby reduce journey time.

This in order to provide the following services:

- increase road safety (e.g. by simulating forced green waves);
- inform users about traffic conditions (fluid or congested) on a road section and redirect them to another section where traffic is fluid in the case of a congested traffic;
- resolve recurring or occasional traffic problems (congestion, incidents, special events, etc.) in a coordinated manner with the various stakeholders.

The system comes with some accuracy of traffic forecasts (based on data provided by the usual traffic measurement systems) that enable managers to anticipate network status at a future time horizon [17, 20, 27].

Such a tool will also help users to consider the possible scenarios, to adopt a number of actions on the road network.

So thanks to the system, operators will have access to all the necessary information useful for decision-making on the basis of the simulation of several predefined scenarios.

3.5. Simulation Model Adopted Traffic

The road network in the northwest part of the city of Lomé constituting the study area is a large network. To model a network of this size on a global level and with only a few calibration parameters, the macroscopic flow model of road traffic was adopted because this model allows to model traffic as a flow.

The model adopted must be able to:

- provide good control of network complexity to enable a better understanding of its internal mechanisms and parameters that determine its evolution;
- provide sufficient results to evaluate and compare various traffic management strategies;
- provide consistent solutions to improve traffic flow.

3.6. Conceptual Model of the Data Warehouse

To provide decision makers with an effective means of analysis of an important phenomenon such as congestion, the static, dynamic and temporal data of the road network should be integrated into a data warehouse that can support the proposed decision-making system. The data warehouse is modeled by a snowflake scheme as shown in Figure 4.

In figure 4, The Data Warehouse pattern consists of a fact table (the table Fact_Traffic) that contains useful information for the description of traffic such as the measurement information (number of vehicles, rates, occupancy, time course) and textual information on traffic conditions, and a set of dimension tables (tables Crossroads, Section, Time, Registration and DayOfYear) connected to the fact table by foreign keys. The primary keys of dimension tables are used as foreign keys in the fact table.

The fact table is the table that will provide all information relating to the creation of the analysis and forecasting model.

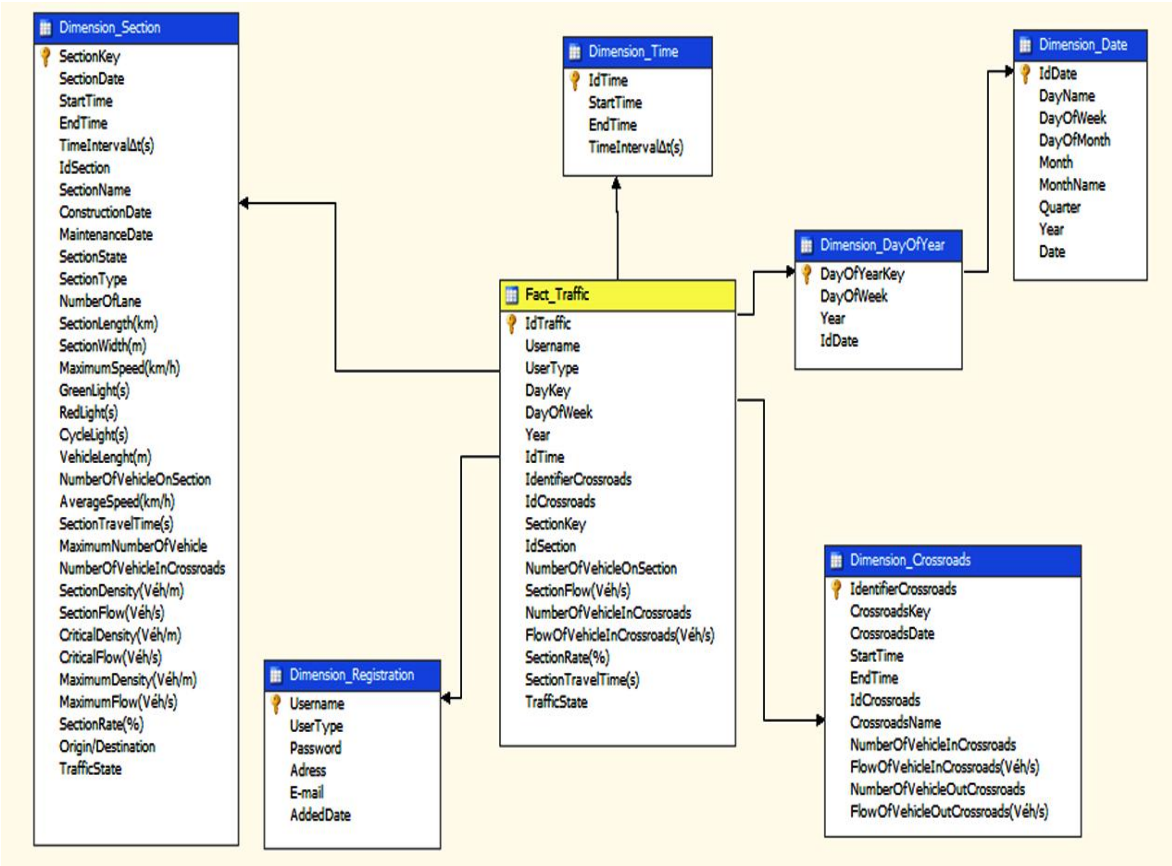

Figure 4: Data Warehouse flake in the road network in the city of Lomé (Northwest Zone)

3.7. Development of Decision-Making Tool

This section contains descriptions of various tools used and the various processes that led to the establishment of the decision-making tool.

The decision-making tool in place is structured around two main players: the administrator and the decision maker (user). Figure 5 shows the overall architecture of the developed decision support tool based on the data warehouse in place.

3.7.1. Presentation of the tools

The development of decision-making tools dedicated to the treatment and prediction of traffic variables was performed using the Microsoft SQL Server 2017 environment, which integrates database management system relational database and the system decision via its various components:

- SQL Server Management Studio 2017 (SSMS) dedicated to the creation and administration of relational databases and Data Warehouse;
- SQL Server Integration Services (SSIS) dedicated to the decision and with tools such as:
  - SQL Server Integration Services (SSIS) which helped feed the data warehouse with data from the source of production;
  - SQL Server Analysis Services (SSAS) which allowed to explore and predict the traffic variables using the data stored in the Data Warehouse;
  - SQL Server Reporting Services (SSRS) which helped generate reports to interpret the obtained results of the prediction.

3.7.2. Feeding the Data Warehouse

Once the data warehouse structure defined, the data must be integrated. The tool used for data integration is the SQL Server Integration Services 2017. This integration goes through three processes:

- initially , the raw data from different Excel files constituting the source of production will be pretreated ; pretreatment creates a hub for each Excel spreadsheet that contains a single table that includes all data from the counting of the crossroads based on the counting parameters;
- in a second stage, these pre-processed data will be extracted, processed and stored in the relational traffic database;
- finally, the data will be retrieved from the relational database to power first dimension tables and then the traffic fact table.

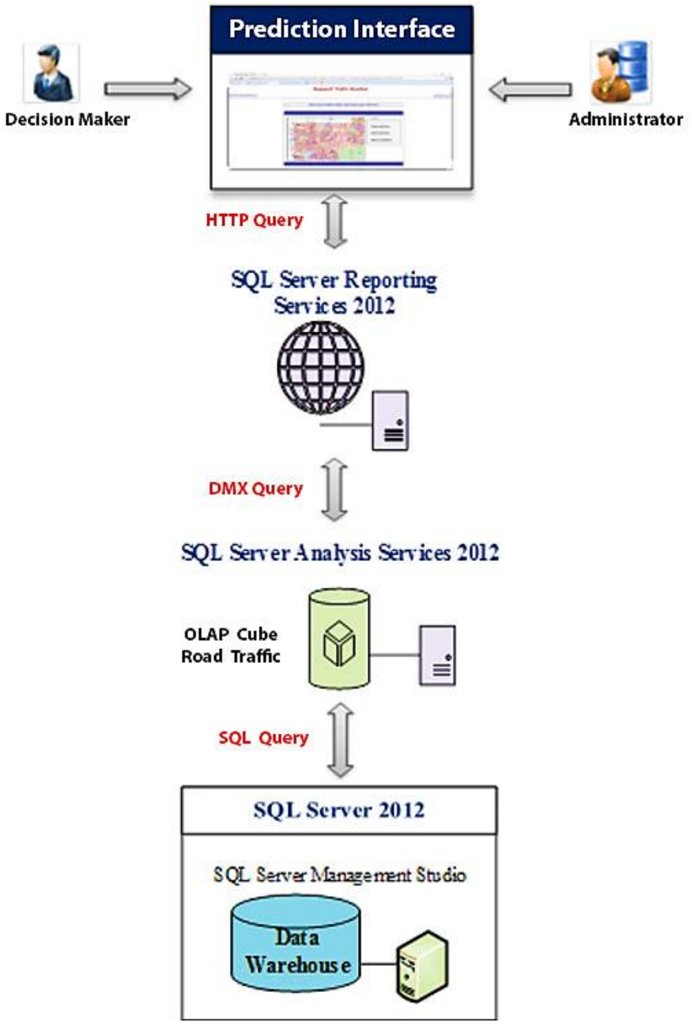

Figure 5: Architecture of the decision-making tool for road traffic of the study area

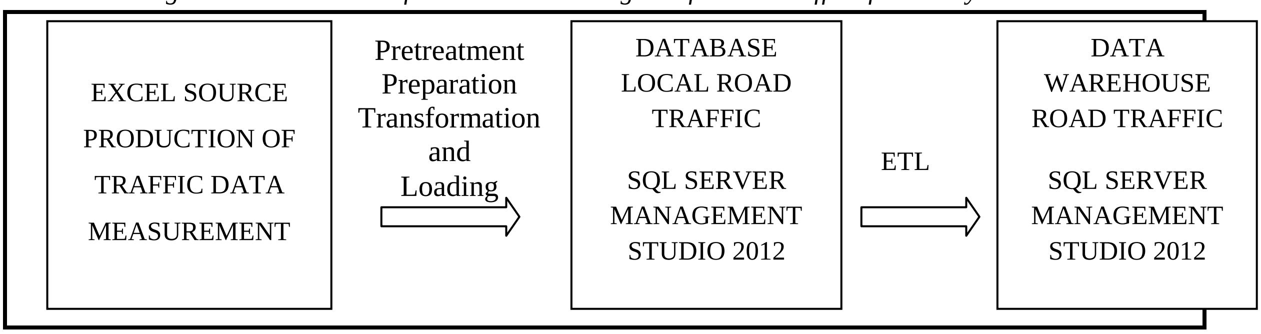

Figure 6: Architecture of the feeding process of the Data Warehouse road traffic

Figure 6 illustrates the steps of the feeding process of the Data Warehouse. Data is stored on the SQL Server 2017 server and access is made possible through SQL Server Management Studio 2017.

3.8. Creation of Analysis and Prediction of Traffic Attribute Models

The tool used to create analysis models is the Microsoft SQL Server Analysis Services. This tool has several algorithms for the analysis and prediction of attributes in continuous or discrete form.

The analysis models generated for the analysis and prediction of traffic attributes are based on two algorithms: Microsoft Decision Trees (MDT) for assessing continuous attributes (number of vehicles, rates, occupancy rates, travel time) and discrete (State traffic) and the Microsoft Naive Bayes (MNB) algorithm used for modeling of discrete attributes (State traffic).

These algorithms have been chosen on the basis of various tests and simulations based on thorough studies of their function.

The variables used to create the MDT model are summarized in Table 1 and the variables used to create the MNB model in Table 2.

Table 1: Variables used to create the "MTD"model

Variable Name	Data Type	Content Type	Use
Traffic Identifier	Long	Key	Key
Key Day and Year	Text	Discrete	Input
Name day	Text	Discrete	Input
Year	Long	Discrete	Input
Identifying time	Text	Discrete	Input
Key crossroads	Long	Discrete	Input
Identifying the crossroads	Text	Discrete	Input
Key section	Long	Continuous	Input
Identifier section	Text	Discrete	Input
Number of vehicles on the section	Long	Continuous	Predict
Flow of vehicles on the section	Double	Continuous	Predict
Number of vehicles entering the crossroads	Long	Continuous	Predict
Flow of vehicles entering the crossroads	Double	Continuous	Predict
Rate of the section	Double	Continuous	Predict
Travel time	Double	Continuous	PredictOnly
Traffic State	Text	Discrete	PredictOnly

Table 2: Variables used to create the "MNB" model

Variable Name	Data Type	Content Type	Use
Traffic Identifier	Long	Key	Key
Key Day and Year	Text	Discrete	Input
Name day	Text	Discrete	Input
Year	Long	Discrete	Input
Identifying time	Text	Discrete	Input
Key crossroads	Long	Discrete	Input
Identifying the crossroads	Text	Discrete	Input
Key section	Long	Continuous	Input
Identifier section	Text	Discrete	Input
State traffic	Text	Discrete	PredictOnly

Once the models are created, we must proceed to their exploration with the aim of generating DMX Prediction Queries (Data Mining Extension) used in the development of reports for data visualization.

After the exploration phase, we must deploy the templates on the server to the SQL Server 2017, with the aim of making it available for use by users and updated by the administrator.

3.9. Creating Data Visualization Reports

The data integrated into the Data Warehouse and the analysis models created and deployed, create reports to visualizing analytical data.

The creation of reports is done using the Microsoft SQL Server Reporting Services environment (SSIS). The first step is to connect to the data warehouse to access measurement data, then to integrate datasets relevant to each report to provide display data in tables and graphs. The previously generated DMX queries will be used to query and manipulate data stored in the data warehouse, with the aim of generating values of the prediction of traffic variables for each model.

3.10. Development of the Simulation Software

To ensure the proper development of the simulation software (PrédiSTraL), it was better to proceed first with its implementation in UML, in order to represent the software architecture. UML has defined the different actors that interact with the simulation interface, and the key software features including a general diagram of the various use cases shown in Figure 7.

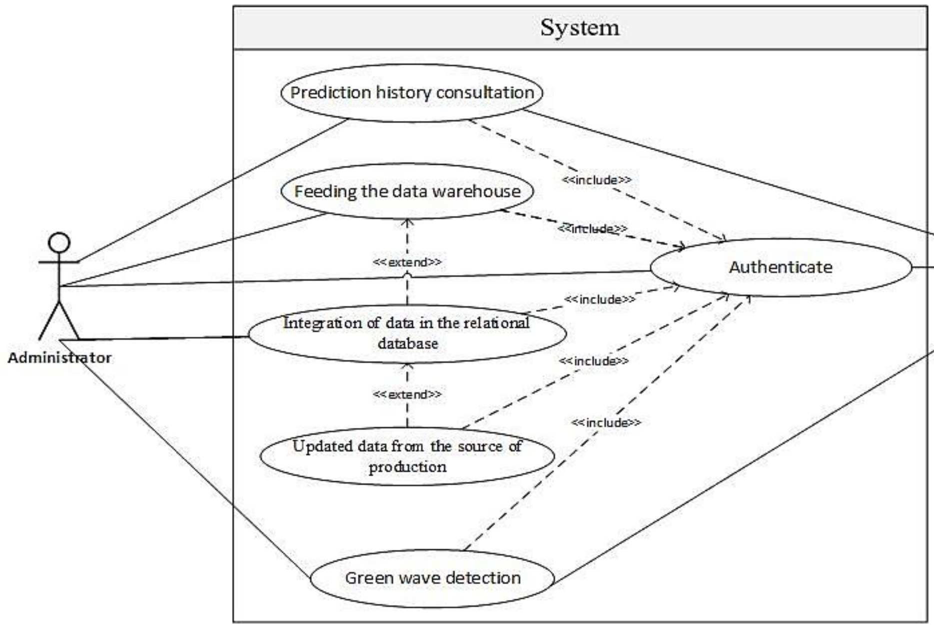

Figure 7: Case Diagram general use of PrédiSTraL

Figures 8 to 12 illustrate the different sequence diagrams of *PrédiSTraL*.

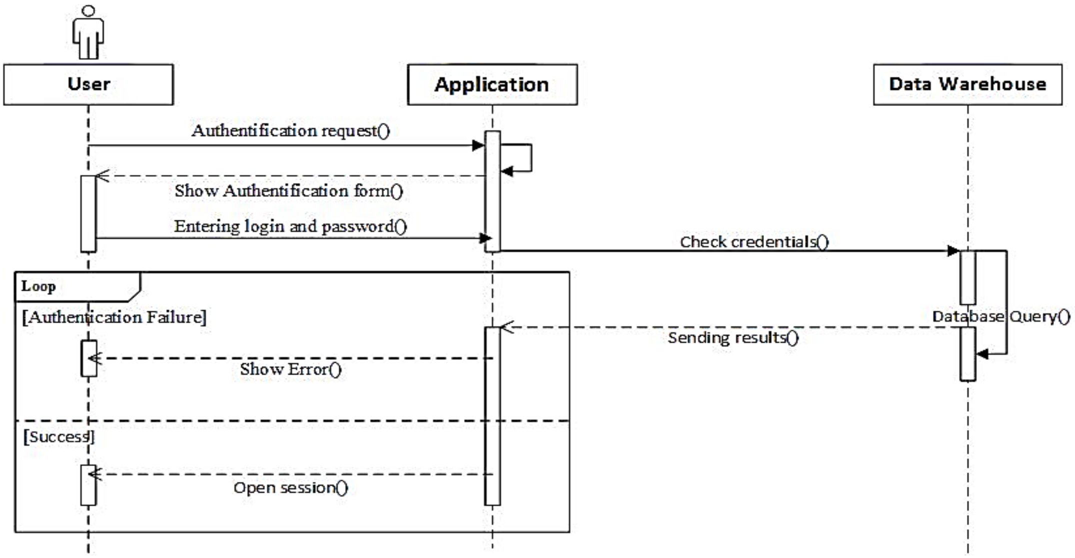

Figure 8: Sequence diagram "Authentication"

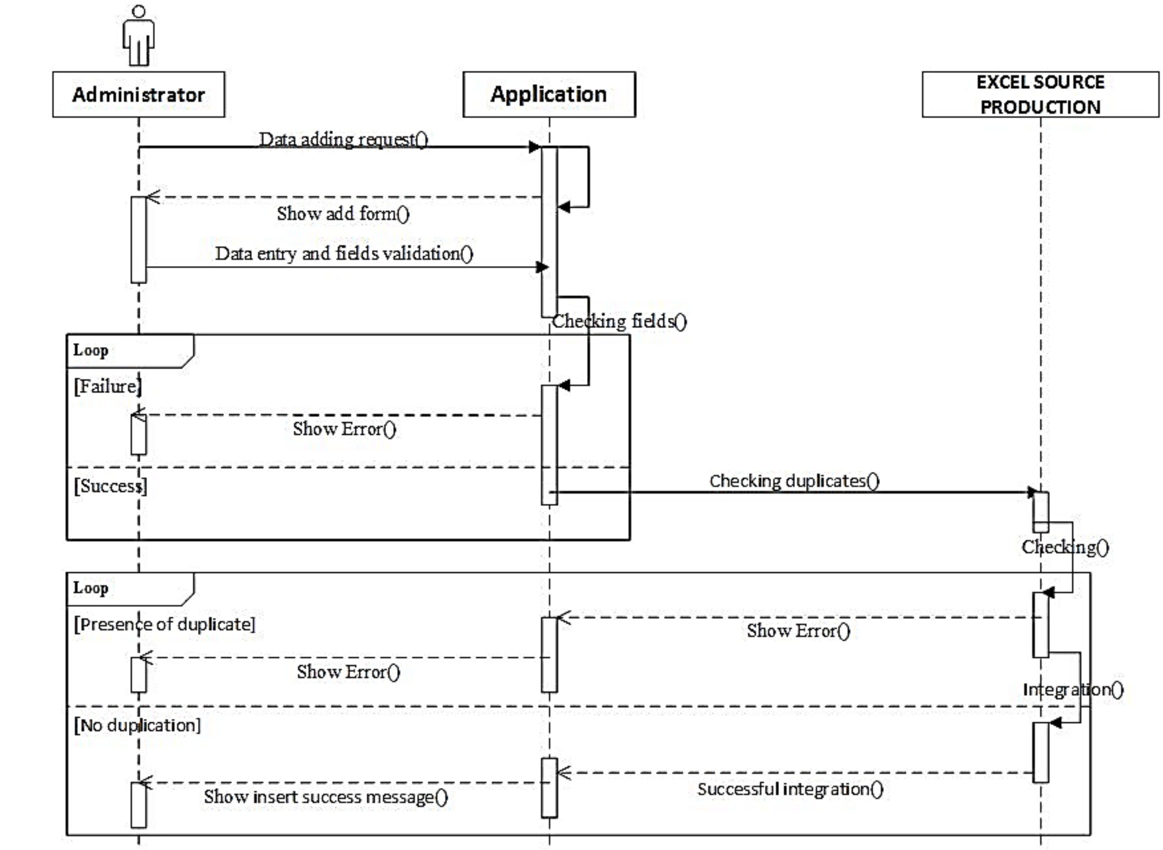

Figure 9: Sequence Diagram "Inserting new data"

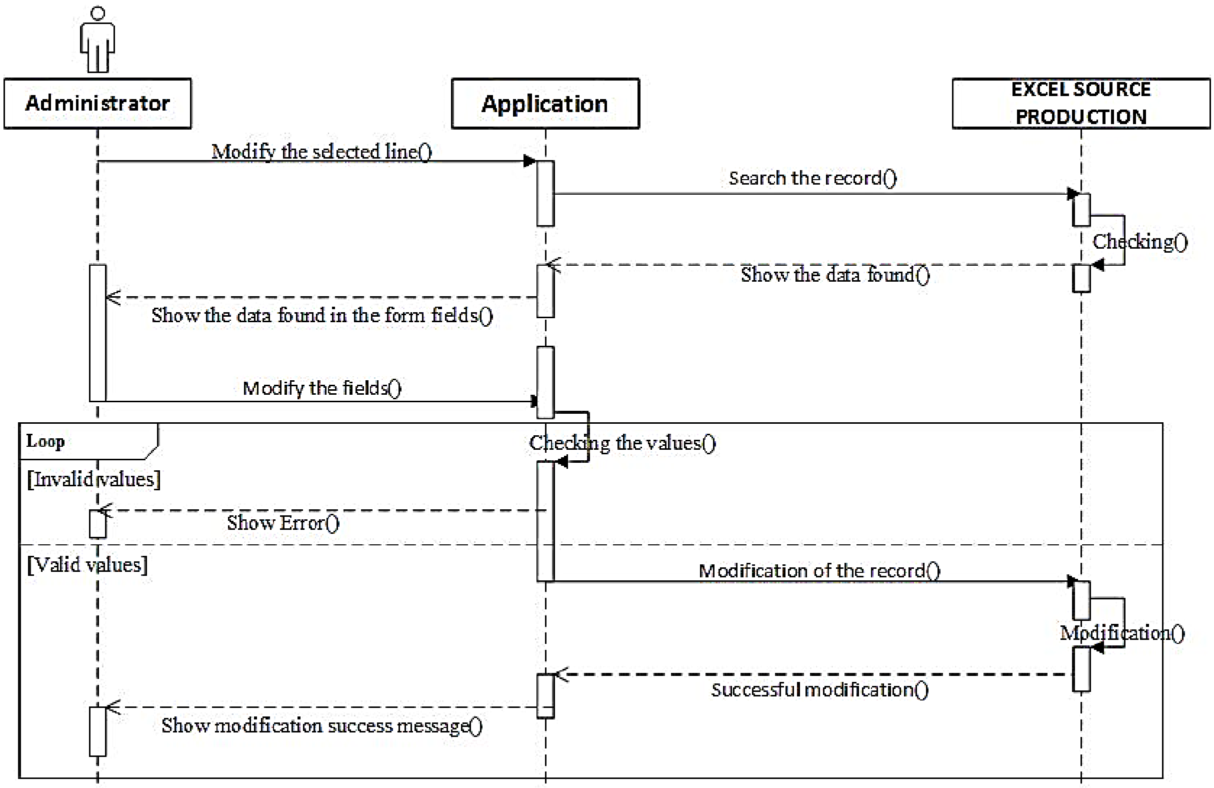

Figure 10: Diagram of sequence "Data modification"

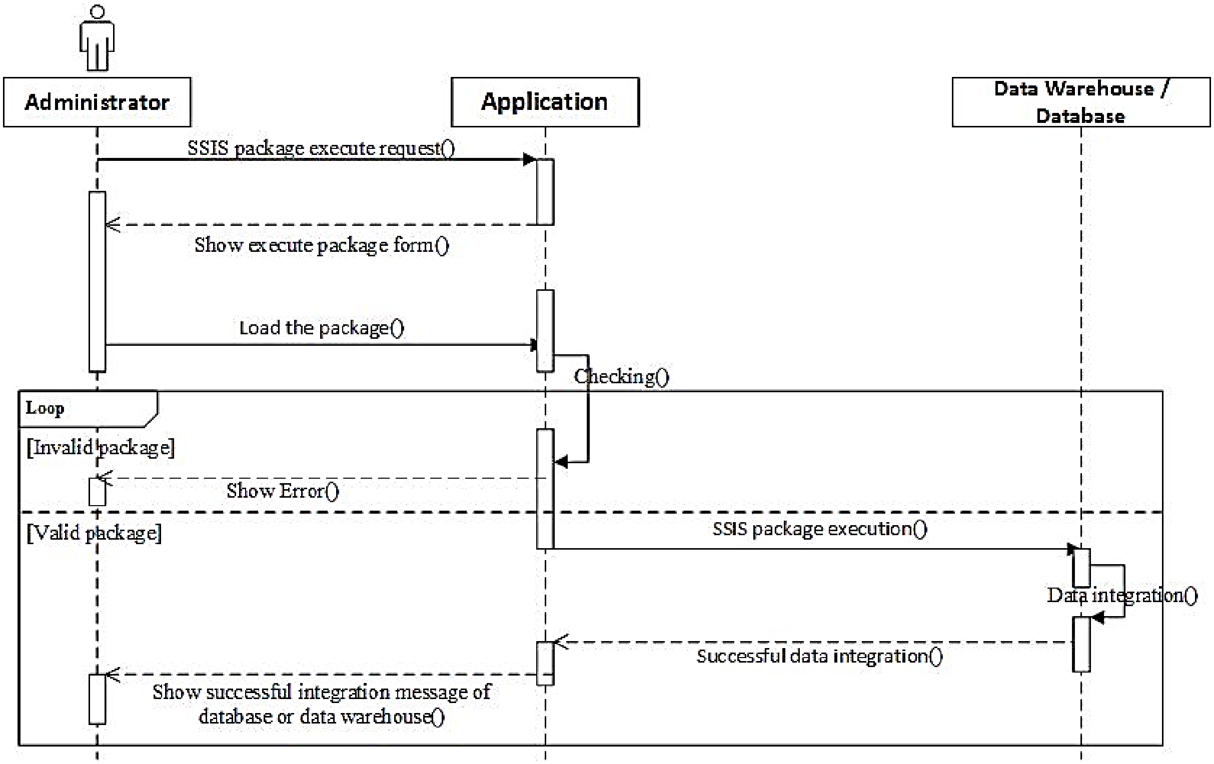

Figure 11: Diagram of sequence "Data Integration"

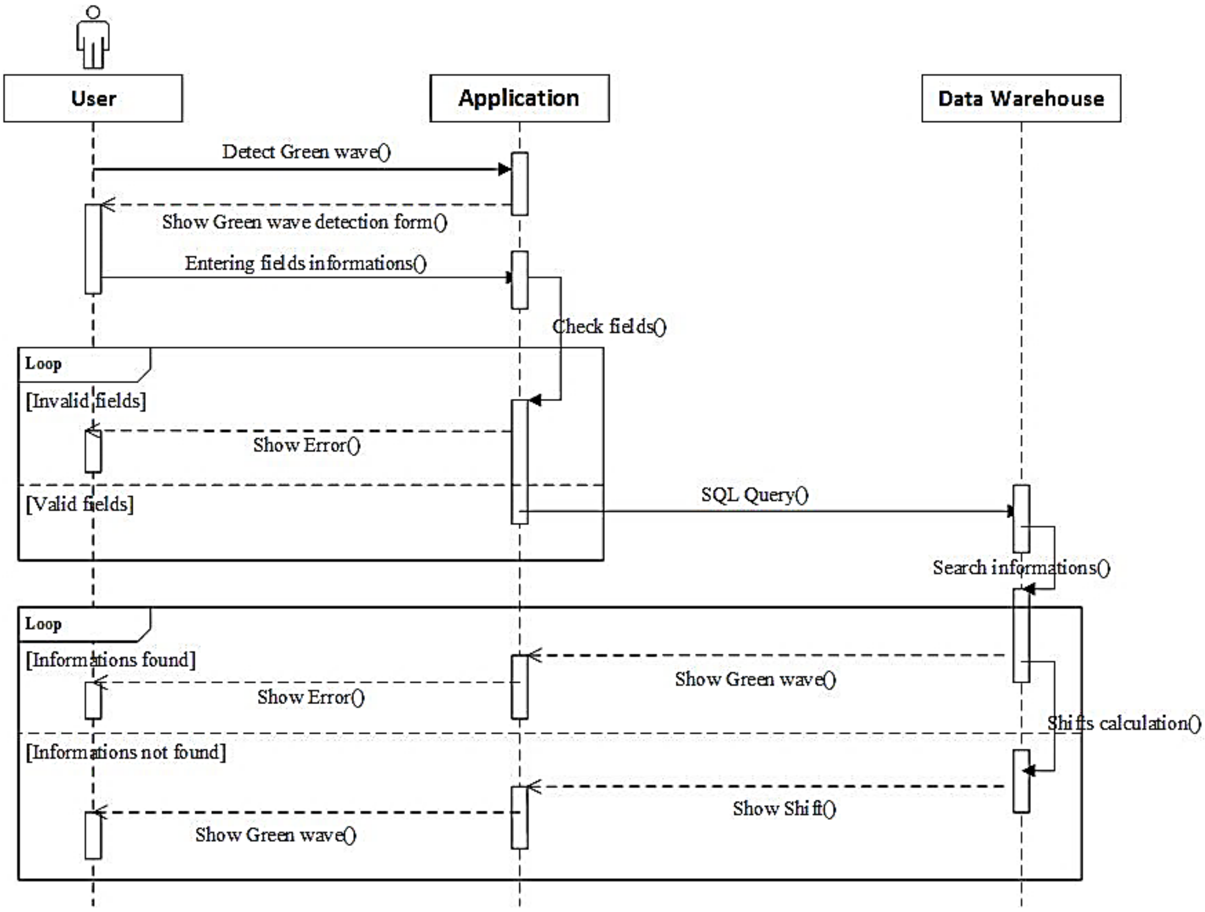

Figure 12: Sequence Diagram "green wave detection"

The tool used to obtain the different UML diagrams is Microsoft Visio 2017 diagram. The main human actors who use the system are: administrator and user. Lifeline sequence diagrams used are: the actors, the simulation software and the data warehouse or the Excel file generation source.

After the modeling phase, comes the actual development. The simulation software has been implemented in the Visual Studio 2017 environment Microsoft designer with programming languages C # that helped develop graphical interfaces, and SQL that allowed to connect to the data warehouse and to administer data.

4. Analysis and Interpretation Of Results

The functionality of the simulation and decision tool software are presented through a set of graphical interfaces. Two user categories are considered: the administrator and the decision maker.

Access to the system is ensured by a conventional authentication with username and password. This authentication interface provides access to the main interface of the simulation software. The host interface provides access to capabilities for policy makers and to the administrator.

4.1. Features for the Administrator

The administrator performs tasks related to the administration of the Excel files from the source of production data and the data warehouse on which is based the decision-making tool developed.

Through the administrator interface (Figure 13), the administrator can perform update operations on information stored in Excel file sources namely the insertion of a new row of data, changing line data, or search for specific information.

Once the data in Excel files sources updated, the administrator can perform the integration package of data in the relational database or power the data warehouse via ETL directly from the interface of the presented simulation software.

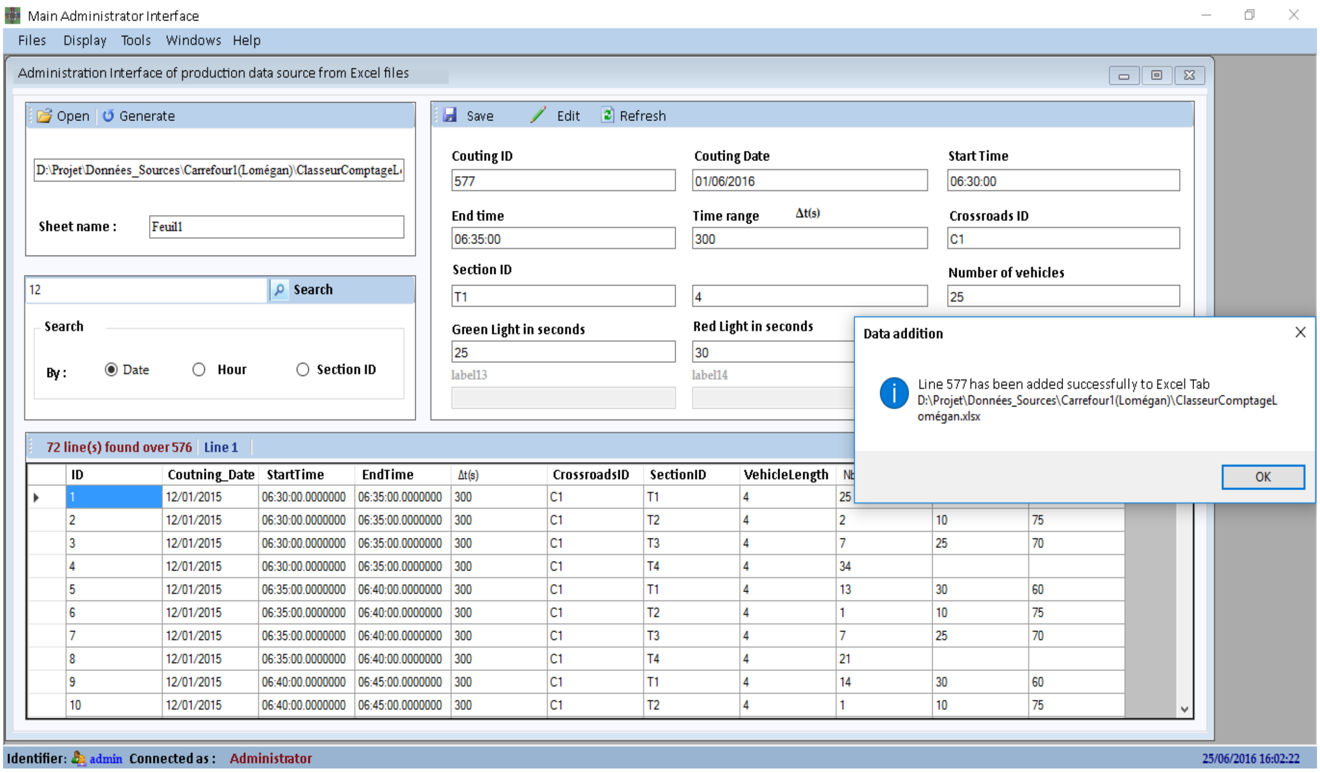

Figure 13: Data administration interface

4.2. Features for the User

The features for the user are: green wave detection and prediction of traffic attributes.

4.2.1. Green wave detection

Illustrated in Figure 14, the green wave detection is to synchronize crossroads belonging to the same road, by shifting the switch to the green different successive fires of these crossroads, so that a large number of vehicles can cross all these crossroads without stopping, thus reducing the number of stops.

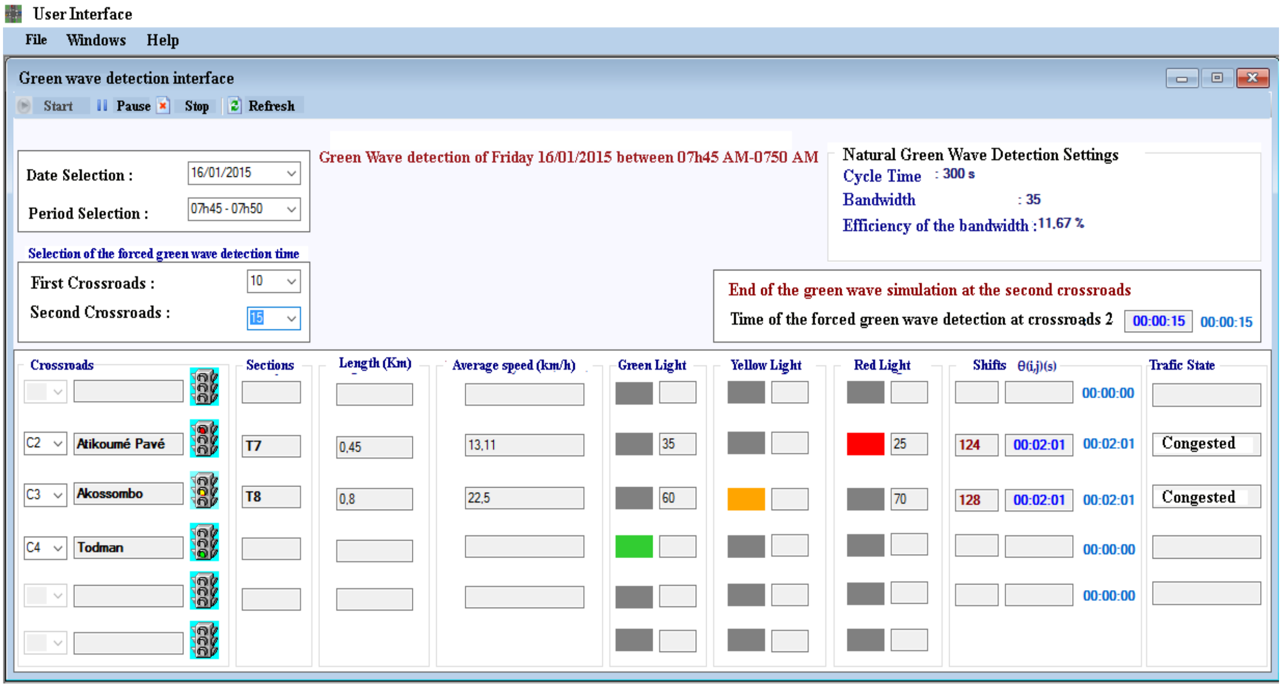

Figure 14: Illustration of a green wave detection process at C1, C2, C3 and C4 crossroads

In the example of Figure 14, the green wave detection is simulated of Atikoume Pave (C2), Akossombo (C3) and Todman (C4) crossroads. Initially, all the red lights of the three hubs are on, but the other lights (green and orange) remain off. Once the simulation starts, the green light from the first crossroad (Atikoumé Pave) lights up, allowing to cross this junction by vehicles initially waiting.

After the lapse of the time required for lighting the green light at the first crossroad, its orange light comes on causing the ignition of the green light from the second crossroad (Akossombo): this leads to a slow down of vehicles at the first crossroad and crossing at the second crossroad by vehicles initially waiting.

After the lapse of the time required for igniting the green light at the second crossroad, its orange light comes on causing the ignition of the first stoplight at the first crossroad ahead of the third crossroad (Todman): this leads to the stopping of vehicles at the first junction and a slowing down of vehicles at the second junction and crossing at the third and final crossroads by vehicles initially waiting.

4.2.2. Prediction of traffic attributes

The developed tool allows the decision maker to select the parameters on which it wishes to base prediction queries. Figure 14 shows the results of the prediction of a 2015 Monday at Lomegan crossroad for the Microsoft Decision Tree model.

For statistical validation of the prediction at various crossroads, we will use indicators to calculate the error rate. One of the most commonly used indicators is the Relative Root Mean Square Error (RRMSE), of which the formula is:

$$RRMSE = 100\sqrt{\frac{\sum_i (x_i - y_i)^2}{\sum_i {y_i}^2}} \qquad (6)$$

where x_i is the measurement of simulated traffic for the period of time i and y_i is the measurement of actual traffic for the period of time i.

Generally it is accepted that the data from a simulation are [5, 17].

- excellent if $RRMSE < 10$
- quite satisfactory if $RRMSE < 15$.

These values of the $RRMSE$ are considered as standard defined thresholds in a traffic simulation context [4,5, 17, 24, 25].

The values of the $RRMSE$ for predicting traffic crossroads (see Figure 3) by the MDT model are presented in Tables 3, 4, 5, 6, 7 and 8.

Table 3: Relative Root Mean Square Error values for Lomegan crossroads (C1)

Crossroads C1									
	Section T1				Section T2				
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				
Days	**AA**	**BB**	**CC**	**DD**	**AA**	**BB**	**CC**	**DD**	**TA**
Monday	0	0.01	0.69	0.26	0	0.07	4.40	0.70	1.86
Tuesday	0	0.01	0.72	0.31	0	0.37	5.35	0.97	2.53
Wednesday	0	0.01	21.95	0.30	0	0.34	148.18	0.83	2.02
Thursday	0	0.01	0.69	0.29	0	0.31	12.50	0.90	2.19
Friday	0	0.01	0.82	0.26	0	0.08	4.04	0.91	1.96
	Section T3				Section T4				
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				
Days	**AA**	**BB**	**CC**	**DD**	**AA**	**BB**	**CC**	**DD**	**TA**
Monday	0	0.04	1.00	0.29	0		0.28	0.38	4.65
Tuesday	0	0.06	1.29	0.39	0		0.37	0.32	4.72
Wednesday	0	0.05	3.91	0.34	0		19.46	0.41	4.03
Thursday	0	0.06	1.81	0.39	0		0.81	0.38	4.58
Friday	0	0.04	1.40	0.39	0		0.28	0.33	5.07

NOTE :
AA - RRMSE of the number of vehicles on the section
BB - RRMSE of the vehicles flow on the section
CC - RRMSE of the occupation rate on the section
DD - RRMSE of the travel time on the section
TA - RRMSE of the total number of vehicles entering the crossroads

Table 4: Relative Root Mean Square Error values for Atikoume Pave crossroads (C2)

	Crossroads C2													
	Section T5				Section T6				Section T7					
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)	
Days	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	of the occupation rate on the section	of the travel time on the section	of the total number of vehicles entering the crossroads	of the total flow of vehicles entering the crossroads
Monday	**0**	**0.03**	**1.03**	**0.92**	**0**	**0.03**	**1.22**	86.43	**0**		**0.28**	313.4	**0**	**0.85**
Tuesday	**0**	**0.03**	**1.67**	**0.86**	**0**	**0.03**	**0.86**	100.74	**0**		**0.42**	**1.63**	**0**	**0.74**
Wednesday	**0**	**0.02**	**1.12**	**0.7**	**0**	**0.02**	**0.83**	84.48	**0**		**0.52**	142.06	**0**	**0.78**
Thursday	**0**	**0.07**	**2.58**	**2.06**	**0**	**0.08**	**1.3**	87.24	**0**		**0.34**	89.08	**0**	**1.49**
Friday	**0**	**0.04**	**1.37**	**1.05**	**0**	**0.04**	**1.06**	71.19	**0**		**0.56**	85.3	**0**	**1.03**

Table 1: Relative Root Mean Square Error values for Akossombo crossroads (C3)

	Crossroads C3													
	Section T4				Section T7				Section T8					
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)	
Days	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the total number of vehicles Entering The crossroads	Of the total flow of vehicles entering the crossroads
Monday	**0**	**0.02**	**1.84**	**0.17**	**0**	**0.02**	**0.27**	313.4	**0**		**0.4**	229	**0**	**0.73**
Tuesday	**0**	**0.02**	**1.99**	**0.2**	**0**	**0.02**	**0.31**	**1.63**	**0**		**0.33**	67.27	**0**	**0.87**
Wednesday	**0**	**0.02**	**1.78**	**0.15**	**0**	**0.02**	**0.45**	142.06	**0**		**0.51**	85.5	**0**	**0.99**
Thursday	**0**	**0.09**	**5.12**	**0.18**	**0**	**0.04**	**0.7**	89.24	**0**		**0.37**	86.76	**0**	**1.18**
Friday	**0**	**0.02**	**1.87**	**0.19**	**0**	**0.02**	**0.33**	85.3	**0**		**0.28**	85.0	**0**	**1.2**

Table 6: Relative Root Mean Square Error values for Todman crossroads (C4)

	Crossroads C4													
	Section T10				Section T8				Section T9					
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)	
Days	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the total number of vehicles Entering The crossroads	Of the total flow of vehicles entering the crossroads
Monday	**2.35**		**2.27**	**0.32**	**0**	**0**	**0.39**	228.99	**1.08**		**0.44**	88.49	**0**	**0.01**
Tuesday	**2.37**		**1.75**	**0.55**	**0**	**0**	**0.31**	67.27	**1.05**		**0.3**	90.14	**0**	**0.01**
Wednesday	**3.25**		**2.06**	**0.35**	**0**	**0**	**0.51**	85.5	**2.22**		**0.67**	93.45	**0**	**0.01**
Thursday	**4.22**		**3.54**	**0.51**	**0**	**0**	**0.38**	86.76	**4.07**		**0.88**	63.13	**0**	**0.02**
Friday	**3.79**		**2**	**0.37**	**0**	**0**	**0.29**	85.07	**2.6**		**0.99**	173.39	**0**	**0.01**

Table 7: Relative Root Mean Square Error values for Adéwui crossroads (C5)

	Crossroads C5													
	Section T10				Section T11				Section T12					
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)	
Days	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the total number of vehicles Entering The crossroads	Of the total flow of vehicles entering the crossroads
Monday	**0**	**0**	**1.7**	**0.32**	**5.3**		**1.51**	**5.88**	**12.87**		**2.7**	**0.62**	**0**	**0.01**
Tuesday	**0**	**0**	**12.07**	**0.55**	**11.88**		23.37	**5**	15.07		15.97	**0.5**	**0**	**0.01**
Wednesday	**0**	**0**	**1.91**	**0.35**	**7.14**		**1.28**	**2.56**	17.29		**2.71**	**0.61**	**0**	**0.01**
Thursday	**0**	**0**	**13.3**	**0.51**	**6.46**		**7.17**	**3.68**	**9.79**		22.65	**0.56**	**0**	**0.01**
Friday	**0**	**0**	**0**	**0.37**	**2.9**		**2.49**	**3.08**	**7.2**		**1.59**	**0.71**	**0**	**0.01**

Table 8: Relative Root Mean Square Error values for Senghor crossroads (C6)

Crossroads C6									
	Section T12				Section T13				
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				
Days	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate on the section	Of the travel time on the section	Of the number of vehicles on the section	Of the vehicle flow on the section	Of the occupation rate of the section	Of the travel time on the section	Of the total number of vehicles entering the crossroads
Monday	**0**	**0.02**	**3.79**	**0.51**	**0.43**		**0.31**	**1.44**	**0.54**
Tuesday	**0**	**0.03**	**2.25**	**0.62**	**0.77**		**0.24**	**1.59**	**0.79**
Wednesday	**0**	**0.02**	**3.43**	**0.57**	**0.25**		**0.39**	**1.21**	**0.66**
Thursday	**0**	**0.01**	**3.11**	**0.63**	**0.67**		**0.26**	**1.91**	**1.06**
Friday	**0**	**0.01**	**2.78**	**0.7**	**0.41**		**0.42**	**1.73**	**0.56**
	Section T14				Section T15				
	Relative Root Mean Square Error (RRMSE)				Relative Root Mean Square Error (RRMSE)				
Days	AA	BB	CC	DD	AA	BB	CC	DD	Of the total flow of vehicles entering the crossroads
Monday	**7.78**		**15.24**	**1.75**	**0**	**0.03**	**0.22**	**0.54**	**0.74**
Tuesday	**9.35**		**4.16**	**1.51**	**0**	**0.03**	**0.36**	**0.79**	**0.99**
Wednesday	**4.05**		**9.56**	**1.4**	**0**	**0.02**	**0.35**	**0.66**	**1.37**
Thursday	25.16		**14.1**	**1.46**	**0**	**0.02**	**0.37**	**1.06**	**1.63**
Friday	**6.41**		**9.74**	**2**	**0**	**0.02**	**0.48**	**0.56**	**0.88**

NOTE :
AA - RMSE of the number of vehicles on the section
BB - RMSE of the vehicles flow on the section
CC - RMSE of the occupation rate on the section
DD - RMSE of the travel time on the section
TO - RMSE of flow of vehicles entering the croassroads

Analysis of Tables 3, 4, 5, 6, 7 and 8 allow to notice that the *RRMSE* values of different continuous predictions of traffic attributes presented in the *RRMSE* tables confirm the satisfactory results obtained for the 6 crossroads, with the exception of certain attributes in some sections of certain crossroads: values of *RRMSE*<10 are excellent and those < 15 are quite satisfactory. Indeed, some crossroads such as Todman, Atikoume Pave and Akossombo present the *RRMSE* values of travel time sections that are very high and therefore unsatisfactory.

What we can deduce from these values are twofold. First, Lomegan and Adewui Senghor crossroads show satisfactory results of the *RRMSE* for all attributes predicted while AtikoumePave crossroads, Akossombo and Todman show satisfactory results of the *RRMSE* for all attributes except the "travel times of the section" even of the results are more than encouraging. Second, the values of travel time of crossroad sections error rate at Atikoume Pave, Akossombo and Todman are high enough meaning that the sections concerned of the three simulated junctions have the characteristic of very high densities and therefore a congested traffic. Therefore, a slight variation in the simulation is sufficient to cause substantial errors and thus a relatively large *RRMSE*. So we can relativize the error rate attached to sections of these three crossroads and conclude that the property "travel time" of certain sections of the three crossroads is not as finely predicted as other attributes.

5. Conclusion

The results obtained and the proposed analysis suggest that the process of analysis and forecasting of traffic that is the subject of this document is effective to solve the problem of predicting different attributes of urban traffic as vehicle flow on a section of road, occupation rate on a section of road, travel time on a section of road. However, with encouraging results, it is quite possible to envisage a better setting for traffic variables. This implies that the developed system has not yet expired; implying that improvements are needed.

The implementation of the system required the use of many technologies and tools integrated into the GUI suite of SQL Server 2017 and historical data on traffic measures from the counting of vehicles at the six crossroads. These data are far from complete; indeed, to finalize the validation of the developed system, we need a lot more traffic data. This is the technical track and research that remain to be explored for improving the system:

- to provide the collection of video data from traffic sensors and infrared stations installed at crossroads;
- to identify areas prone to traffic accidents and to analyze road risk for the improvement of road safety;
- to incorporate climate factors such as weather and geographic factors (related to the topology of the road network).

References

[1]. ALAM I, AHMED MF, ALAM M, ULISSES J, Pattern mining from historical traffic big data,2017 IEEE Region 10 Symposium (TENSYMP), - ieeexplore.ieee.org, 2017

[2]. ALAM I, FARID DM, ROSSETI R J F, The prediction of traffic flow with regression analysis, Emerging Technologies in Data Mining and Information Security, Springer 2019

[3]. BENBERNOU S. *Entrepôt de données et OLAP.* EGC 2009 (Extraction et Gestion des Connaissances), Actes, Strasbourg du 27 au 30 janvier 2009.

[4]. BOURREL E. *Modélisation dynamique de l'écoulement du trafic routier : du macroscopique au microscopique.* Thèse de doctorat à l'Institut National des Sciences Appliquées de Lyon, Décembre 2003.

[5]. CHAMPION A. *Mécanisme de Coordination Multi-Agent fondé sur des jeux : Application à la simulation comportementale de trafic routier en situation de carrefour.* Thèse de doctorat en Automatique et Informatique des Systèmes Industriels et Humains présentée à l'Université de Valenciennes et du Hainaut-Cambrésis, Décembre 2003.

[6]. CHANG, G. L. *Macroparticle traffic simulation model to investigate peak-period commuterdecision dynamics.* Transportation Research Records, 1985.

[7]. CHANUT S. *Modélisation dynamique macroscopique de l'écoulement d'un trafic routier hétérogène poids lourd et véhicules légers.* Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 2005.

[8]. CNAML (Conservatoire National des Arts et des Métiers de Lille). *Ingénierie des Systèmesdécisionnels.* Data warehouse et data mining, juin 1998.

[9]. CODD Edgar F. *Does your DBMS run by the rules*? Computer World, 1985.

[10]. FRANCO J.-M., EDS-Institut-Prometeus. *Le Data warehouse, le Data Mining.* Eyrolles 1996.

[11]. GAZIS D. C. *The origins of traffic theory*. Operations Research, 2002, Vol. 50 № 1, pp. 60-77.

[12]. GAZIS D. C., HERMAN R., ROTHEBURY R.W. *Non-linear flow the leader models of trafficflow.* Operations Research, 1961, Vol.9, pp. 545-567.

[13]. GEROLIMINIS N., DAGANZO C. F. *Macroscopic modeling of traffic in cities.*86th Annual Meeting Transportation Research Board Washington, D.C. January 2007.

[14]. INMON William H. *Building the Data Warehouse.*2° Ed 1996 WILEY.

[15]. KAMARIANAKIS Y., SHEN W., real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO, Applied stochastic models in business and industry, 2012

[16]. KEOPS, CITAFRIC, *La Ville de Lomé.* Plan Guide Lomé, 2005.

[17]. LANA I, DEL SER J, VELEZ M, Road traffic forecasting: Recent advances and new challenges,IEEE Intelligent Transportation Systems Magazine (Volume: 10, Issue: 2 , summer 2018)

[18]. LECLERQ L. *Modélisation du trafic*. LICIT (ENTPE/INRETS), 2005.

[19]. LEFEBURE R., VENTURI G. *Le Data Mining*, Eyrolles 1998.
[20]. MINISTERE DU TRANSPORT DU TOGO. Le secteur du transport situation générale, http://www.un.org/esa/dsd/dsd_aofw_ni/ni_pdfs/NationalReports/togo/transport.pdf.
[21]. NGUYEN Q. T., BOUJU A., ESTRAILLER P. *Multi-agent architecture with space-time components for simulation of urban transportation systems*. Procedia Social and Behavioral Sciences, 54, 365-374 (2012).
[22]. PASQUIER N. *Fouille de données:* OLAP et Data Warehousing, Université de Nice Sophia-Antipolis, 2006 (10).
[23]. RANJINI K., KANTHIMATHI A., YASMINE Y. *Design of Adaptive Road Traffic Control System through Unified Modeling Language.*International Journal of Computer Applications (0975 – 8887), Volume 14 – No.7, February 2011.
[24]. REY T, KORDON A, WELLS C , *Applied data mining for forecasting using SAS*2012 - books.google.com
[25]. RUSKIN H. J., WANG R. *Modeling traffic flow at an urban signalized intersection*. In: Proceedings of International Conference on Computational Science, pp. 381–390 2002.
[26]. SANDOVAL V. *L'informatique décisionnelle.* Edition Hermès, 1997.
[27]. SAUVAGEOT J. *Modélisation du trafic routier*, IREM Université Paris7, 2002
[28]. TAMAS P.*Modeling nonlinear road traffic networks for junction control*. International Journal of Applied Mathematics and Computer Science,2012, Vol. 22, No. 3, pp723–732 DOI: 10.2478/v10006-012-0054-1.
[29]. TUFFERY S. *Data Mining et Statistique Décisionnelle.* Éditions Technip, 2005, 4^{e} édition 2012, préface de Gilbert Saporta.
[30]. ZAMITH M., LEAL R., KISCHINHEVSKY M., CLUA E., BRANDAO D., MONTENEGRO A. LIMA E. A. *Probabilistic Cellular Automata Model for Highway Traffic Simulation*. Procedia Computer Science 1 (2012) 337–345 (2012).