Information Model For Refining The Transportation Jamming In Sultanate Of Oman

Afra Alzidi¹, Boumedyen Shannaq ²

¹,²College of Business at University of Buraimi, Oman, Al-Buraimi, Sultanate of Oman.

Abstract – This work intends to build a model to predict the time hour/day/date of an accident on specific city and street location, which causes traffic congestion. The proposed model could alert the driver of an accident occurrence in specific location and time. Therefore, this Forecasting Information Model Scheme will minimize the traffic congestion as most of the drivers will change their path route. A large data set has been collected from the Traffic Database for the past 10 years. Data Mining Methodology has been customized to build the forecasting model. 4011 Time series instances have been used to train the forecasting model. The MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) have been used for Evaluation on training data. The MAE scored ~5.1, and RMSE scored ~6.8. The obtained results are promising and could be useful for improving road safety and traffic congestion strategy.

Keywords – Traffic Congestion, Traffic Accidents, Data Mining, Forecasting, Information Systems, Time series.

1. Introduction
In Traffic accidents are the main cause of death, and resulting in complex traffic congestion. They occur for many reasons, among which there are both technological and human factors. The accident can occur due to the fault of a tired driver, due to condition of the road surface or malfunction of the vehicle system.

The problem of traffic jams on the roads of large cities and central routes is more relevant than ever.[1] Every year there are more and more cars on the roads in Muscat city.[2] At the same time, there are many complaints about the roads itself; poor quality, inconvenient road interchanges, the absence of elevated/underground passages (many traffic lights) and bypass roads for cargo trucks.[3] The problem of traffic jams requires a solution - the sooner, the better both for an individual who is wasting time (well, except that a passenger can read a book or sleep in public transport), and for the country’s economy as a whole[4]. Road transport is the most unsafe of all available to drivers. According to all data, it is road accidents that put in first place in the number of dead and injured and resulting in complex traffic congestion. According to these parameters, cars significantly overtake rail, air and water transport. Traffic accidents are the main cause of death.[5] They occur for many reasons, among which there are both technological and human factors. The accident can occur due to the fault of a tired driver, due to conditions of the road surface or malfunction of the car system. However, third-
party factors such as the day of the week, weather conditions and the quality of the asphalt pavement often affect the risk of getting into an accident.

Identification of factors that significantly affect the risk of a road accident when solving the problem of improving road safety should be considered as a priority. This will allow making decisions that really can eliminate the external causes of accidents, and resulting in minimum traffic jam. This research paper is devoted to the analysis and identification of such causes, to predict the probability of an accident that could occur and cause a traffic jam in specific location, time, day and date. The results and conclusions made in this work can be useful for improving road safety and minimizing the traffic congestion. The proposed information systems prototype scenario, problem and solution are presented in Figure 1.

![Figure 1. Traffic Congestion scenario problem and solution](image)

To conduct an analysis in order to identify factors significantly affecting the size of an accident, a large amount of information is needed. In fact, the task at hand is very dimensional and requires a large number of observations to draw serious conclusions. As a source of information, a database of victims of accidents over several years was used. In total, about 47000 records are available; using random filtering, the number of source records was reduced to 37000. The practice of reducing the sample size is often used when working with large-scale sources - if you have a high-quality random number generator, this will not affect the significance of the result, but will save time.

Forecasting [6] is a prediction, which implies a state or description of possible or desirable aspects, states, solutions, problems of the future. Forecast is the result of the forecasting process, expressed in verbal, mathematical, graphic or other form of judgment about the possible state of the object and its environment in the future period of time. Forecasting is the process of building a prediction of the future based on historical data, current data (the current situation) and based on trend analysis. Risk and uncertainty are central factors in forecasting, and therefore, in accordance with best practices, the degree of uncertainty with respect to forecasts should be indicated. The correct approach to the
estimation of the forecasting method involves several stages. Five important steps should be highlighted as described in Figure 2.

Figure 2. Process of estimating forecasting model

➢ The systematic study of the nature of the investigated object or process to select an adequate forecasting method;
➢ The allocation of two groups among the available data to develop forecasts and to verify the results;
➢ The refinement of the source data in order to detect errors;
➢ Development of forecasts and assessment of the reliability of the results;
➢ The interpretation of the results obtained and the implementation, if necessary, of refinement and addition of forecasts.

2. Literature Review

The current studies in traffic accidents have been increased in the literature. However, the development of reliable methodology for predicting and preventing traffic accidents to reduce accident rates on highways, need more efforts and development. [7] Proposed to analyze the state of road safety, methods and measures for the prevention of road traffic accidents in the sultanate of Oman. The predictive analytics of road accidents in Oman using machine learning approach.

The main goal in conducting various data analyzes is to search for patterns to predict what may happen in the future. For the traffic industry [8] proposed neural network principal component analysis in the forecasting the road traffic accident. Forecasting traffic accidents using a neural network, evaluating the effectiveness of its use Parameters, the ratio between theoretical, model values. The results of neural network modeling of the possibility of getting into a traffic accident. Researchers and experts conduct various tests to understand traffic mechanisms. [9] Proposed forecasting model for the freeway traffic accidents base on Markov model. To make rational decisions to improve road safety, it seems advisable to investigate the dependence of the accident rate on the state of roads in the region as a whole and in the areas where traffic accidents are concentrated. The solution of the tasks set is impossible without a statistical analysis of the level of development and technical condition of the road network and the operation of vehicles in the administrative regions of the region. Given the existing level of financing for the repair and maintenance of the road network, even a slight increase in cargo traffic in the administrative regions of the region will lead to an increase in the total number of road accidents. When rebuilding the areas where traffic accidents are concentrated, it is first
necessary to pay attention to improving the connection qualities of the pavement and broadening the roadway. In the case of limited visibility and the presence of curves of small radius, it is necessary to install road signs and perform horizontal marking, taking into account the specifics of the road section.

How do we predict driver behaviors by analyzing some related variables, Can we predict the likelihood that an experience driver will successfully overcome the expected accidents. Can we find a connection between the specific behavior of one particular accident? [10] Introduced a model of traffic accident prediction based on convolutional neural network. More and more attention is paid to process optimization, mainly in the form of lower production costs. Cost reduction can be achieved by upgrading equipment, but this approach entails a lot of costs for design, purchase, reconstruction, etc., and is also accompanied by lost profits during the idle time of the reconstructed object. But it is also possible to use a mathematical approach to search for inefficiencies in the technological process. [11] argued that only IoT based framework for Vehicle Over-speed detection could solve the traffic accidents problem.

The concept and classification of traffic accidents. Remote computer simulation of an accident. Appointment and principle of operation of the AI Monitor complex. Functional characteristics of the alarm sensor. Finite element method and computer calculation programs. [12] The communication technology-based solution for reducing road traffic congestion in smart cities was proposed. The research work analyzes the phenomenon of “smart city” (“Smart City”) through the prism of an innovative technological base that currently exists in the world. The authors investigate the theoretical foundations of the concept of “smart city”, its multi-factor model. Particular attention is paid to technological innovations such as the Internet of Things, Big Data, unmanned vehicles, robotics; the article analyzes the deep-seated changes and possible risks arising from the implementation of the above technologies, gives recommendations on the anticipation of possible negative consequences. The study allows to assess the significant potential of the concept of “smart cities”. [13] Design and implementation of an eye blinking detector system for automobile accident prevention.

Relevance and goals. The article is devoted to issues of increasing transport accessibility of large cities. Regulated intersections in large cities are sources of increased costs, and at such intersections during peak hours there are queues that are accompanied by time losses, as well as frequent emergency situations. In order to eliminate the above problems, it is proposed to introduce intelligent transport systems into the process of controlling the movement of vehicles in large settlements, the use of which will reduce the level of congestion of roads and increase their traffic capacity, optimize the use of road transport and increase the accessibility of the transport complex as a whole. [14] ‘Variable Speed Limit (VSL) Based Model for Advanced Traffic Management through VANET’.

A traffic management model based on accounting for telecommunication load, which can be used in software-defined data transmission networks based on IP, is considered. [15] Study on the ‘Forecast Method of Road Accidents Based on Extension of Cubic Spline Curve’. Existing methods for conducting automotive technical expertise require the selection of some parameters based on the intuition and experience of the expert. Also, when determining the deceleration, the make of the vehicle and the degree of its load, road conditions are not taken into account. In the process of analysis it was found that the use of special software can significantly increase the efficiency of the work to solve the tasks, speeds up the calculation process, qualitatively reduces the likelihood of errors of an arithmetic nature and makes it possible to visualize the results of the study. [16] ‘Assessing Surface Transportation Network Component Criticality: A Multi-Layer Graph-Based Approach. The work is devoted to the analysis of the development possibilities of transport systems of large cities and urban
agglomerations. The main attention is paid to the tools for forming the optimal topology of the agglomeration transport network. The paper formulates the sequence of forming the graph of the transport network of a large city, functioning on the basis of the interaction of rail transport.[17]

Performance impact of relay cardinality on sender-driven messages dissemination in VANETs. The invention relates to wireless communication systems, in particular to a repeater intended for use in wireless communication systems, in which a wireless communication device is installed, which allows for interaction with base stations that transmit data using a repeater to enable control of the gain of the repeater and its power output.[18] A ‘LoRaWAN-based Camel Crossing Alert and Tracking System’ has been proposed. Some of the harm vehicle accidents affected when a group of camels crossing the roads. The author argued that this is a severe problem in Sultanate regions where there is a large population of camels. The author proposed an information system to send an alert when a camel or group of camels close to the road, therefore the drivers or the responsible authorities could take an action to avoid a harm accident vehicle and camel. The disadvantages the proposed system is the cost of the development and maintenance. [19] Conduct a survey in the form of ‘MANETs and VANETs clustering algorithms’. [20] proposed classification model as ‘Calibration and Pre-Processing Techniques for a Smartphone-Based Driving Events Identification and Classification System’. [21] works on Highway Driving Events Identification and Classification using Smartphone. Proposed to develop the theoretical requirements of a descriptive information model of the traffic emergency, which is characterized by parameters that affect the occurrence of accidents with justification for the choice of parameters that make up the model. In summary the task of using the forecasting model, define a set of variables for each of the parameters that affect the occurrence of a traffic accident and develop a mathematical apparatus to describe the selected parameters that affect the likelihood of a traffic accident. Justify the method of analysis of statistical data of road accidents, based on determining the range of spatial coordinates. To develop a mathematical model for determining the likelihood of road traffic accidents based on statistical information about road traffic accidents, in relation to the scene of an accident. Justify the use of fuzzy logic to implement the decision-making algorithm to reduce the number of road accidents and develop software to support decision making to improve road safety. Give an economic assessment of measures to improve road safety, carried out using the results of studies. More information about the Information modeling could be found in [22,23,24,25,26].

3. The proposed forecasting Methodology
A model is a simplified image of an object from real life, in which its most important characteristics are reflected, from the point of view of research. A method is a complex technique, an ordered set of simple techniques aimed at developing a forecast as a whole. A forecasting system is an ordered set of techniques, technical tools, designed to predict complex phenomena or processes. Figure 3 describes the proposed methodology for building the forecasting model.
Figure 3. The proposed forecasting Methodology

The way to achieve the goal is to proceed from the knowledge of the most general laws. Methodology - a certain combination of techniques (methods) for performing prognostic operations, obtaining and processing information about the future based on uniform forecast development methods. Forecasting methodology is a field of knowledge about forecasting methods and systems.

3.1. Prediction Method Categories

Qualitative forecasting methods are subjective, based on the opinions of consumers and experts. Qualitative methods are suitable when historical data are not available. These methods are used, as a rule, for medium-term and long-term solutions. Examples of good forecasting methods are market research. Quantitative forecasting models are used to predict future data as a function of historical data. They are suitable for use when historical numerical data are available and when data dynamics are expected to persist in the future. These methods are usually used for short and medium term forecasting. Examples of quantitative forecasting methods are: moving averages, exponential smoothing, and multiplicative seasonal indices.

3.2. BUILDING FORCASTING MODEL

In fact, people almost everyday encounters certain tasks related to the time series. Most often the question arises - what will happen with our indicators in the next day / week / month / etc. - how many cars will pass this rout, how many accidents could occur, how many actions users will perform, and so on. The forecasting problem can be approached in different ways, depending on what quality the forecast should be, for what period we want to build it, and, of course, how long it takes to select and adjust the model parameters to obtain it. After we understand the traffic accident and congestion problem, reviewing the previous studies, the following subsections describe the proposed data analysis process to build the forecasting model. At the stage of setting the problem, it is necessary to determine what the purpose of the analysis is. In particular, it is required to answer a number of questions, the main of which is what exactly needs to be determined as a result of the analysis. Also on this list:
Will it be necessary to make predictions based on a data mining model or just find meaningful patterns and relationships?
If a forecast is required, which dataset attribute should be predicted?
How are the columns connected? If there are multiple tables, how are they related?
How is the data distributed? Is the data seasonal? Does the data provide an accurate picture of the subject area?

3.2.1. Data Collection
The stage of data preparation includes identifying data sources for analysis, combining data and clearing it. The data used collected from different Traffic databases and on different servers. Moreover, some data were presented in the form of text files, spreadsheets, or in other formats. In the process of combining and converting data, the capabilities of SQL Server Integration Services have been used as demonstrated in Figure 4, Figure 5, and Figure 6. This allows this work to significantly automate the preparation process.

Figure 4. Sample of collected DB ERD

Figure 5. Sample of collected DB tables
The collected DB contains 56 variables and 30771 traffic recorders.

3.2.2. Data Cleaning

The data collected in this way usually needs additional processing, called cleaning. During the cleaning process, if necessary, removal of “outliers” (uncharacteristic and erroneous values), processing of missing parameter values, numerical conversion (for example, normalization), etc. can be performed.

B.1 Missing data:

Figures: Figure 7, Figure 8 and Figure 9 demonstrates Sample of performances of the data cleaning process on the collected DB with missing and noisy data collected in erroneous formats.
3.2.3. Data transformation
The next step is to study the data, which will allow us to understand how adequately prepared the set represents the studied subject area. Here, the search for the minimum and maximum values of the parameters, the analysis of the distribution of values and other statistical characteristics, the comparison of the results with ideas about the subject area can be carried out. This step is done in order to convert the data into appropriate forms suitable for the mining process.

3.2.4. Data Reduction
To bring data in the correct Time series format, SQL procedures have been performed. The process of this task demonstrated in Figure 10.

After performing the SQL procedures the data set improved and to match all the requirements of the forecasting model using the time series format.

Figure 11 shows the accident-number variable that configured as the class attribute for the proposed forecasting model, 2 minimum numbers, 55 maximum of accident occurs with specific parameters.
4. Results and Discussion

Traffic Information accumulated in a variety of enterprise databases, and finally transformed to be a time series, and arranged in chronological order and produced at successive points in time. The proposed time series analysis is carried out with the aim of:

- Determining the nature of the series
- Predicting the future values of the series.

(Time  station_code  acc_time)
Cross validation

- Break the sample into k blocks
- Each in turn acts as Test

Figure 14 demonstrates the proposed strategy for managing the training data and test data.

Figure 14. Process of data training and data testing

In the process of determining the structure and patterns of the time series, it is supposed to detect: noise and emissions, trend, seasonal component, cyclical component. Determining the nature of the time series in this work used as a kind of “intelligence” of the data. The analyst’s knowledge of the presence of a seasonal component is necessary, for example, to determine the number of sample records that should take part in the construction of the forecast. In order to focus on the traffic congestion of forecasting, this work considers time series only within the framework of solving the forecasting problem. Two fundamental differences in the time series from a simple sequence of observations: The members of the time series, unlike the elements of a random sample, are not statistically independent. The members of the time series are not equally distributed. Figure 15 demonstrates the profile of the Transformed Data Set.

Figure 15. Transformed Time series profile

The weka. classifiers. functions. Multilayer Perceptron’ classifier that uses back propagation to learn a multi-layer perceptron to classify instances, described in Figure 16.
Figure 16. Results from Multilayer Perceptron’ classifier

The ‘weka. classifiers. functions. Linear Regression’ class for prediction.

Experiment Details

Scheme:
Linear Regression -S 0 -R 1.0E-8 -num-decimal-places 4
Lagged and derived variable options:
-F ACCEDENT_numbers,acc_code_STREET, station_code,acc_time -L 1 -M 7 -G Date -day of week -weekend

Relation:     FINAL DATA FOR CAST WITH LOC.arff
Instances:    4400
Attributes:   5
  Date
  ACCEDENT_numbers
  acc_code_STREET
  station_code
  acc_time
Transformed training data:
  ACCEDENT_numbers
  acc_code_STREET
  station_code
  acc_time
  Day Of Week
  Weekend
  Date-remapped
  Lag_ACCEDENT_numbers-1
  Lag_ACCEDENT_numbers-2
  Lag_ACCEDENT_numbers-3
  Lag_ACCEDENT_numbers-4
  Lag_ACCEDENT_numbers-5

Figure 17, demonstrates the predictions for training data.
Figure 17. Predictions for training data

The Evaluation on training data used the Mean absolute error (MAE) and represented as mean percent error (MPE)

The MPE is the average percent forecast error. The main problem of this error is that in an unstable number series with large outliers, any slight fluctuation in the fact or forecast can significantly change the error indicator and, as a result, the accuracy of forecasting. In addition, the error is asymmetric: the same deviations in plus and minus affect the error indicator differently.

\[ MPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\Phi_i - \Pi_i}{\Phi_i} \]

The obtained result in this work achieves \( \sim 5.1 \) calculated as Mean Absolute Error (MAE) = \( \frac{\text{sum}(|\text{predicted} - \text{actual}|)}{N} \).

For each position, the forecast error is calculated (the fact is subtracted from the forecast) - Error

<table>
<thead>
<tr>
<th>insta</th>
<th>actual</th>
<th>predicted</th>
<th>conf</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>34</td>
<td>27.6445</td>
<td>11.057456</td>
<td>-6.9551</td>
</tr>
<tr>
<td>13</td>
<td>32</td>
<td>27.6734</td>
<td>11.06344236</td>
<td>5.4504</td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>20.4445</td>
<td>4.456155055</td>
<td>7.4445</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>20.3197</td>
<td>7.311140141</td>
<td>12.1057</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>24.1222</td>
<td>6.11544591</td>
<td>1.1231</td>
</tr>
<tr>
<td>17</td>
<td>14</td>
<td>23.5206</td>
<td>7.921451909</td>
<td>9.5206</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>13.7266</td>
<td>-2.10911229</td>
<td>-2.7094</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>26.1052</td>
<td>10.621456909</td>
<td>7.6392</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>21.9131</td>
<td>6.905453747</td>
<td>1.9131</td>
</tr>
<tr>
<td>21</td>
<td>25</td>
<td>16.9141</td>
<td>0.0334356502</td>
<td>-8.6596</td>
</tr>
<tr>
<td>22</td>
<td>35</td>
<td>17.5069</td>
<td>1.053136122</td>
<td>-17.4391</td>
</tr>
<tr>
<td>23</td>
<td>22</td>
<td>25.656</td>
<td>9.600144217</td>
<td>3.656</td>
</tr>
<tr>
<td>24</td>
<td>11</td>
<td>21.659</td>
<td>5.455147052</td>
<td>16.659</td>
</tr>
</tbody>
</table>

Figure 18, demonstrates the obtained results for the future predictions from end of training data.

<table>
<thead>
<tr>
<th>Time</th>
<th>ACCENT_numbers</th>
<th>acc_code</th>
<th>STEET</th>
<th>station_code</th>
<th>acc_time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-01-26</td>
<td>22.9455</td>
<td>2.0455</td>
<td>4.214</td>
<td>12.3506</td>
<td></td>
</tr>
<tr>
<td>2020-01-24</td>
<td>14.9594</td>
<td>2.403</td>
<td>3.9711</td>
<td>14.5703</td>
<td></td>
</tr>
<tr>
<td>2020-01-23</td>
<td>11.1036</td>
<td>3.9701</td>
<td>3.3276</td>
<td>19.0208</td>
<td></td>
</tr>
<tr>
<td>2020-01-22</td>
<td>22.1274</td>
<td>3.9993</td>
<td>2.3587</td>
<td>15.3039</td>
<td></td>
</tr>
<tr>
<td>2020-01-21</td>
<td>23.0669</td>
<td>3.2997</td>
<td>3.1392</td>
<td>13.0653</td>
<td></td>
</tr>
<tr>
<td>2020-01-20</td>
<td>22.7004</td>
<td>2.7009</td>
<td>3.2417</td>
<td>19.0226</td>
<td></td>
</tr>
<tr>
<td>2020-01-19</td>
<td>20.0947</td>
<td>2.7074</td>
<td>3.5912</td>
<td>14.4646</td>
<td></td>
</tr>
<tr>
<td>2020-01-18</td>
<td>20.2000</td>
<td>2.5052</td>
<td>2.9605</td>
<td>15.5615</td>
<td></td>
</tr>
<tr>
<td>2020-01-17</td>
<td>21.2405</td>
<td>3.2621</td>
<td>2.5894</td>
<td>13.4807</td>
<td></td>
</tr>
<tr>
<td>2020-01-16</td>
<td>23.2409</td>
<td>2.5619</td>
<td>3.6012</td>
<td>19.0674</td>
<td></td>
</tr>
<tr>
<td>2020-01-14</td>
<td>22.1292</td>
<td>2.9437</td>
<td>2.7734</td>
<td>14.2468</td>
<td></td>
</tr>
<tr>
<td>2020-01-13</td>
<td>22.2652</td>
<td>2.9449</td>
<td>2.5349</td>
<td>15.0596</td>
<td></td>
</tr>
<tr>
<td>2020-01-12</td>
<td>20.7111</td>
<td>3.1110</td>
<td>2.7142</td>
<td>14.0366</td>
<td></td>
</tr>
<tr>
<td>2020-01-10</td>
<td>22.6295</td>
<td>3.0900</td>
<td>2.7143</td>
<td>13.704</td>
<td></td>
</tr>
<tr>
<td>2020-01-09</td>
<td>22.7424</td>
<td>2.9426</td>
<td>2.837</td>
<td>13.9261</td>
<td></td>
</tr>
<tr>
<td>2020-01-08</td>
<td>20.6940</td>
<td>3.2581</td>
<td>3.0464</td>
<td>15.6853</td>
<td></td>
</tr>
<tr>
<td>2020-01-07</td>
<td>20.6165</td>
<td>3.1161</td>
<td>2.4941</td>
<td>14.14</td>
<td></td>
</tr>
<tr>
<td>2020-01-06</td>
<td>20.9575</td>
<td>3.0969</td>
<td>2.7972</td>
<td>15.0067</td>
<td></td>
</tr>
<tr>
<td>2020-01-05</td>
<td>23.8319</td>
<td>3.395</td>
<td>2.4525</td>
<td>14.1446</td>
<td></td>
</tr>
<tr>
<td>2020-01-04</td>
<td>23.2629</td>
<td>3.0646</td>
<td>2.7908</td>
<td>14.2664</td>
<td></td>
</tr>
<tr>
<td>2020-01-03</td>
<td>20.1741</td>
<td>3.1399</td>
<td>2.7955</td>
<td>13.7109</td>
<td></td>
</tr>
<tr>
<td>2020-01-02</td>
<td>23.3241</td>
<td>2.8238</td>
<td>2.7739</td>
<td>13.5202</td>
<td></td>
</tr>
<tr>
<td>2020-01-01</td>
<td>23.1632</td>
<td>3.1061</td>
<td>2.7753</td>
<td>13.7201</td>
<td></td>
</tr>
<tr>
<td>2020-01-00</td>
<td>24.0016</td>
<td>3.1067</td>
<td>2.6644</td>
<td>14.3702</td>
<td></td>
</tr>
</tbody>
</table>
Percent forecast error is calculated for each position (forecast error is divided by the actual indicator) - Percent Error. The arithmetic mean of all percent forecast errors is found (percent errors are summed and divided by the number)

The Root Mean Square Error RMSE) also have been selected to evaluate the training data set.

RMSE - RMS prediction error. Approximately the same problem as in MPE and MAPE: since each deviation is squared, any small deviation can significantly affect the error rate. It is worth noting that there is also an MSE error, from which the RMSE is obtained by extracting the root. But since MSE gives the calculated units of measure squared, then using this error will be a little wrong.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\Phi_i - \Pi_i)^2}
\]

The obtained result in this work achieves ~ 6.8 calculated as Root Mean Square Error (RMSE): \(\sqrt{\text{sum}((\text{predicted} - \text{actual})^2) / N}\).

For each position, the square of deviations is calculated (the difference between the fact and the forecast squared) - Square Error

Then the arithmetic mean is calculated (the sum of the squared deviations divided by the number) - MSE - Mean Square Error

We extract the root from the result - RMSE

To convert to a percentage or to a “normalized” standard error, you need to:

1. Divide by the difference between the maximum and minimum values of indicators
2. Divide the difference between the third and first quartiles of indicator values
3. Divide by the arithmetic mean of the values of the indicators (the most common option). Figure 19, demonstrates the obtain results for the proposed forestry model.

<table>
<thead>
<tr>
<th>Date</th>
<th>Predicted</th>
<th>Actual</th>
<th>Predicted</th>
<th>Actual</th>
<th>Predicted</th>
<th>Actual</th>
<th>Predicted</th>
<th>Actual</th>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>401</td>
<td>401</td>
<td>401</td>
<td>401</td>
<td>401</td>
<td>401</td>
<td>401</td>
<td>401</td>
<td>401</td>
<td>401</td>
</tr>
<tr>
<td>Max absolute error</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
</tr>
<tr>
<td>Min absolute error</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
<td>0.00857</td>
</tr>
</tbody>
</table>

Figure 19. Results of evaluation on training data

MAE and RSME Evaluation on training data of a Total number of instances: 4401 for the ACCEDENT_numbers, acc_code_STREET, station_code, acc_time attributes.

Figure 20 describes the step ahead predictions for the ‘accident_numbers’ class attribute with a 95% confidence.
The obtain results demonstrated in Figure 21, are promising and matching the goals of this work.

5. Conclusions
The increased number of vehicles running on the Muscat roads resulting in increase in the total traffic congestion in the Sultanate, especially in the Governorate of Muscat. It is one of the important issues affecting the majority of the population, in a way that calls for serious thinking in finding solution and treatment. The proposed research work investigates and proposed to select an appropriate forecasting method. The allocation of two groups among the available data, used to develop the forecasting model and to verify the results. The refinement of the source data in order to detect errors, development of forecast model and assessment of the reliability of the results, the interpretation of the results obtained and the implementation, if necessary, of refinement and addition of forecasts. Data Mining Methodology has been customized to build the forecasting model, where 4011 Time series instances have been used to train the forecasting model. MAE and RMSE have been used for Evaluation on training data. The Mean Absolute Error (MAE) scored ~5.1, and Root Mean Square Error (RMSE) scored ~6.8. The obtained results are promising and could be useful for improving the road safety and traffic congestion strategy.

References


http://www.webology.org