

# **SEASONAL CLIMATE PREDICTION USING MACHINE LEARNING TECHNIQUES FOR THE HOLY CITY OF MAKKAH**

Mohammed Arif (Corresponding author)

Computer Science Department, Umm Al Qura University

Makkah, KSA, E-mail: [Mohammad-arif@hotmail.com](mailto:Mohammad-arif@hotmail.com)

Mohammed A Al Ghamdi

Computer Science Department, Umm Al Qura University

Makkah, KSA, E-mail: [maeghamdi@uqu.edu.sa](mailto:maeghamdi@uqu.edu.sa)



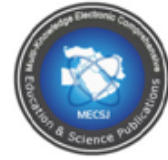
## **Abstract**

The Weather Conditions are affecting many areas of life and may at times be obstructed to accomplish a certain activity because of sudden and rapid change. Hence, the importance of predictability. So, as to avoid side effects that may result in significant damage. In this paper, we used data mining techniques to predict the specific temperature in the Mecca region which has a special for Muslims as it is one of the most important religious rites and in the Hajj season many Muslims come to Mecca. It is expected in 2030 that about 30 million Muslims come for Hajj, it is therefore important to predict the weather and temperature to avoid risks and minimize damage and take the necessary steps to prepare for the care of visitors. We also compared three algorithms: Long Short Term Memory (LSTM) with accuracy 91%, Multiple Liner Regression (MLP) with accuracy 95% and Decision Tree Algorithm with accuracy 96%

**Keywords:** Weather Prediction, Data Mining, Data Analysis, Algorithms, LSTM

## **1. Introduction**

The prediction of weather has been a hard task since years owing to different reasons such like the climate drastically unpredictable behavior. Even though the technology most recent advancement, it has been very hard, however the predicting able to make the weather prediction accuracy is the suspicious reason. Though in the present time, that range continues such like an object of research in the case that scientists and mathematicians are putting efforts to generate an algorithm either a model so it will precisely portend the weather.



The most motivations of the proposed work are to: i) applying the different machine learning algorithms over given data, ii) predict the weather condition of the holy city where millions of people visiting each year to Perform Hajj, iii) helping government authorities in order to develop the services being provided to the city's visitors based on the weather conditions to avoid expected accidents. Enormous developments have been existed in the sensors that are accountable in order to record the data of environment and scratch a noise that existing in it; right together with that modern models which have been suggested that contain varied properties relative to obtaining the prediction which is accurate. Presently, a type of the largest extensively applied techniques for weather predicting is information of mining which perform the weather. The mining of data presents method for analyzing of data in a statistical way and extracting or deriving like rules which able to be utilized in predicting works. The mining of data is considered to be techniques of exploration that has been extensively used in main domains where the future prediction is highly in demand such like the prediction of stocks prices within a time-bound period hereafter. Recently, scientists have recognized which the mining of data able to be utilized such like a tool of weather prediction either. The essential entity needs to mine is the meteorological data which collected more than a specified time period. The data perhaps become a piece of information that supplies the climatic status details like humidity, moisture, etc. The mining of data presents technique diversity to expect the consequences like, Regression, Classification, Clustering, etc. Data Classification is considered to be the method of data uniting to varied classes even groups build upon the connections between them.



The mining of data term indicates to techniques that are utilized in extracting the information that are needed to the data provided group which may become beneficial in objective statistical either to make predicts through learn manners in correlation and data among parameters varied. The accuracy of prediction exceedingly relies upon knowledge of prevailing weather condition over an extensive area. there are many techniques which help us to achieve our goals from data that we have, and these techniques are using by researchers with many subjects include genetics, economy, marketing and so on.If we used data mining techniques correctly, we can improve the business work by using predictive analysis because the techniques of data mining getting an efficient and increase the confidence of the prediction. Also, data mining techniques include the approaches which depend on multi-learning for tasks for classifying and mining big data and Machine Learning techniques which is one of data mining tools that help to build algorithms to learn and predict .this paper will include three algorithms of machine learning. Data mining include six type of classes to do tasks: Detection Outlier of data set to determine unusual data ,Dependency Modeling which identify the variable relationships Clustering here we search of grouping data ,Classifications of the data set for predict a new statues, Regression that try to model data and finally Summarization which enable to visualization data and images .all these classes can help to increase the revenue, accelerate to making the decisions, reduce risk and costs and improve work. consider data mining is basic owing to Artificial intelligence, machine learning and statistic consider data mining is basic in statistic, Artificial intelligence, and data mining of machine learning have different processes and it classifies for two types: first one, data preparing or preprocessing.



The second is the mining of data. Actually, the cleaning step, integration, selection, and transformation represents the data preparation and integration of the last three steps consider as data mining. The methodology has been dependent on our paper on the data mining steps and we applied as the following steps:

- Data Selection: we choose the most important data for our paper aim, we focused on time as criteria and data type so we prefer numeric data it helps rapid analysis. And the data we choice: Year, month, day wind speed, wind direction, air temperature. Pressure and humidity to put them in data set
- Data Collection: the data we used in the paperwork was obtained from the General Authority of Meteorology and Environmental Protection in Saudi Arabia, the dataset included data about temperature, wind speed, wind direction, pressure, and humidity.
- Data Cleaning: cleaning data set from bad data like missing and duplicated data, to make it suitable with data mining .
- Data Transformation: here we transform data which has been chosen, for making it in a suitable form for data mining, we removed noise and make attribute clustering on data set.
- Data Mining: in this stage, we made some analysis on data set by using three different algorithms like long Short Term Memory algorithm, Multiple Linear Regression algorithm , and Decision Tree Algorithms and get the result of accuracy ratio

What's left in that paper is held to be followed like: In sect. 2 we discuss related work. Section 3, shows our work methods. Section 4 demonstrates the Experimental Design. Section 5 finished with results and discussion. Section 6, future work and Conclusion.



## 2. Related Work

Forecasting and Temperature prediction were studied through many different algorithms and data extraction techniques in order to predict the weather behavior in future. In [1][2] show that the LSTM model will be improved in many cases when using the spatial information. This paper [3] mention that precipitation is a spatiotemporal sequence forecasting problem, the researches proposed the convolutional LSTM (ConvLSTM). Experiments show that ConvLSTM better than FC-LSTM and ROVER algorithm in captures spatiotemporal correlations. A case of the research works contains paper [4] searchers proposed artificial Neural Net-works and Decision Tree algorithms to predict weather around. A relationship in between the attributes of weather is found by ANN to build model, while the tree of C5 Decision discovers the data trend to build a tree of classifier. In [5], comparison between Random Forests and Decision Trees. These papers [6][7] show a comparison between decision trees and k-means clustering. The searchers found that it is a good performance in data mining. With increasing the size of the data set, the accuracy is increasing too, but it is decreasing after a specific limit. In [6]for Wind speed future values prediction, Minimum Temperature, Radiation, Evaporation, Maximum Temperature, Rainfall, and so Recurrent network architectures were used with the TLFN neural network to build the ANN model, searchers used the TDNN memory component and they found it better than when using the Gamma memory component. In this paper [8], the Growth Algorithm of FP was utilized to create trees of decision for weather classifying parameters. Searchers apply FP Growth algorithm with the evaluation of MAE, MSE and SD.



The algorithm shows correct monthly rainfall prediction and it is more accurate than existing model neural network. In [9],[10] [11] comparison of in-depth performance has been among C4.5 and algorithm of Random Forest, contains discussion more than the algorithm suitability while applying to the diverse data set.

### **3. Methods**

#### *3.1 Long Short Term Memory*

In general, there are many algorithms to extract useful information from the dataset. Such clustering of data, Regression, Classification or genetic algorithm. This paper suggests using three different algorithms as a working model for weather forecasting. The first is networks of LSTM that are very suitable to processing, making predictions, and classifying build upon data of time series ever after, it able to be unknown duration lags among significant actions in a series of time. The algorithm of LSTMs were improved to transact with problems of vanishing gradient and exploding so it able to be faced while training RNNz that is traditional. the second is C4.5 DT which can be used for classification problems. It can deal with both continuous and discrete attributes. Finally, Multi Linear Regression is able enough to define the relative impact of a kind or more variables of predictor to the standard value and ability to identify outliers or anomalies. LSTM is a type of Recurrent Neural Networks which able to learn dependencies with long-standing. They were submitted by Schmidhuber and Hoch Reiter. Recurrent Neural Network have much success when applying it on problems like a translation and comment on images (image captioning) so, the Memory of Long Short Term importance comes to improve Network of Recurrent Neural work.



The Memory of Long Short Term consider a very specific kind of Recurrent Neural Networks that support achieve all wonderful results of Recurrent Neural Networks. LSTM has inner cells that carry information without change can remember data for long term time. The recurrent neural network controls over cell state completely by special structures named gates which is can add, edit and remove information.[12]. The Recurrent neural network contains a sigmoid layer that produces numbers between zero and one so, it helps the network to control over information when flowing ( $S_{(t)} = 1/1 + et$ ), the sigmoid layer determines how much to be permit by channels. Zero's value meaning nothing at all to flow whereas the value of a one is meaning leave all flow and so with that system we able to predict through the model which makes the value of a far distant important in the past in a modern predicting. So, this makes LSTM a useful model to the prediction system. Mathematically, by LSTM we can estimate the probability of conditional  $p(y_1, \dots, y_N | x_1, \dots, x_N)$  where  $(x_1, \dots, x_N)$  is a sequence of input and  $(y_1, \dots, y_N)$  is considered as its sequence of corresponding output together with the similar length. The computes of LSTM that probability of conditional firstly, by obtaining the fixed-dimensional representation  $v$  of the input sequence  $(x_1, \dots, x_N)$  given by the last hidden state of the LSTM, then after computing the probability of  $(y_1, \dots, y_N)$  together with formulation of a standard LSTM-LM whose initial hidden state is set to the representation  $v$  of  $(x_1, \dots, x_N)$ :

$$P(Y_1 \dots Y_n, X_1 \dots X_n) = \Pi(Y_t | v, Y_1 \dots Y_{t-1}) \quad (1)$$



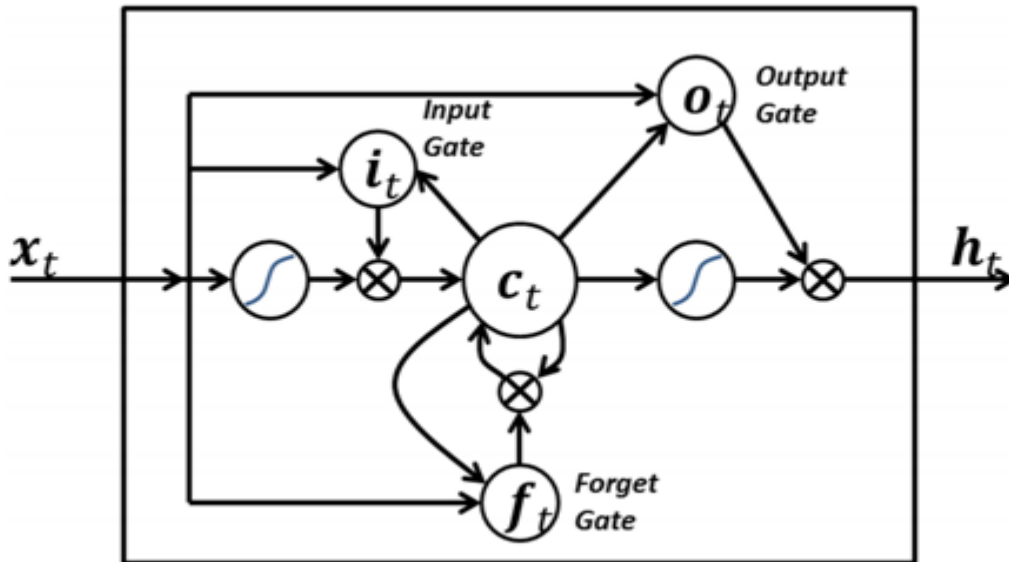


Fig. 1. Gate of simple LSTM together with only output, input, and gates of forget.

### 3.2 C4.5 Decision tree

Algorithms of decision tree are one of much-supervised learning algorithms in machine learning, it uses in solving problems of classification and regression, the decision tree is using the tree for representing problem-solving, the leaf represents a class or outcomes(continuous values, categorical), the internal node represent an attribute(features) and the branch represents the decision rule. We use it to represent any logic matter. The idea of how decision tree work is to create a tree for all data with processing each leaf or minimize error. The most important challenge in the decision tree is how to determine the attribute for each level,



There two way to selection attribute: information gain and gini index. DT is an algorithm for classification which can generate a decision tree algorithm to a specified dataset. So, it depends upon the definition of entropy. We can construct the algorithm of decision tree via select the preferable attribute which use for splitting the dataset in an effective way. The given attribute which has the highest entropy is chosen to split the dataset in a specific node. There are many attributes for each node to helping make decision [11][10] [13]. The C4.5 [14] algorithm is an extension of Quinlan earlier ID3 algorithm. The attributes are selected by entropy order that because of the higher entropy cause uncertainty in results. To make a prediction, each pattern is pushing below the tree. That it put the labels in a last node. This procedure is repeated for all trees then taken the average from all trees to represent the prediction. To base an algorithm of decision tree, we have to compute two kinds of entropy by utilizing frequency table as following: a) Entropy utilizing table of frequency of a one attribute:

$$Es = \sum_{i=1}^n - Pi \log 2Pi \quad (2)$$

b) Entropy utilizing table of frequency of two attributes:

$$E(T,X) = \sum_c^n P(c)E(c) \quad (3)$$

C4.5 advantages are over than other systems of Decision Tree algorithm: The algorithm naturally uses Process of Single Pass Pruning for Mitigating overlapping. It able to perform with both of Continuous and Discrete. Data of C4.5 able to deal with incomplete data issue perfectly.

### 3.2 Regression of Multiple Linear

The regression of multiple linear (MLR) objective [15] consider as an analysis of the relationship among several dependent variables and independent variables. The model represents such the relation in between the dependent variable  $Y_i$  and the  $p$  vector of repressors  $X_i$  is linear. The followed performs an equation of MLR: when a consider an intercept,  $B$  consider a slope. In forecasting objectives, the equation of linear regression shall suit the forecasting model in an observed data set of  $Y$  and  $X$  values.

$$Y = a + B_1X_1 + \dots + B_nX_n \quad (4)$$

The work idea of Multiple Linear Regression based on ordinary least squares, we make the model with good fit such that minimize the difference between actual and predict. Multiple Linear Regression seek to achieve efficiently estimators and unbiased value and consistent,  $R^2$  is a way to determine how good is model by compute the sum of squares. Coefficient of determination :

$$R = SSR/SST \quad (5)$$

Where: SSR: Sum of Squarer Regression and SST: sum of squarer Total

## 4. Experimental Classification Results and Analysis

### 4.1 Data Set

Data set was taken from General Authority for Meteorology and Environ- mental Protection, from Saudi Arabia, in period (2008, 2018) for all days for each month. Data set includes 3791 records and 22 columns contains :STATION NUM- BER,STATION NAME, YEAR, MONTH, DAY, AIR TEMPERATURE MAX DB,



AIR TEMPERATURE MEAN DB , AIR TEMPERATURE MIN DB, WIND PREVAILING DIRECTION, WIND MEAN SPEED, WIND MAX DIRECTION, WIND MAX SPEED, SKY COV OKTES MEAN , RELATIVE HUMIDITY MIN, RELATIVE HUMIDITY MAX, PRESSURE MIN SEA LEVEL, PRESSURE MIN STATION LEVEL, PRESSURE MAX STATION LEVEL,PRESSURE MAX SEA LEVEL, PRESSURE MEAN SEA LEVEL, and PRESSURE MEAN STATION LEVEL. All data is about Mecca region weather.

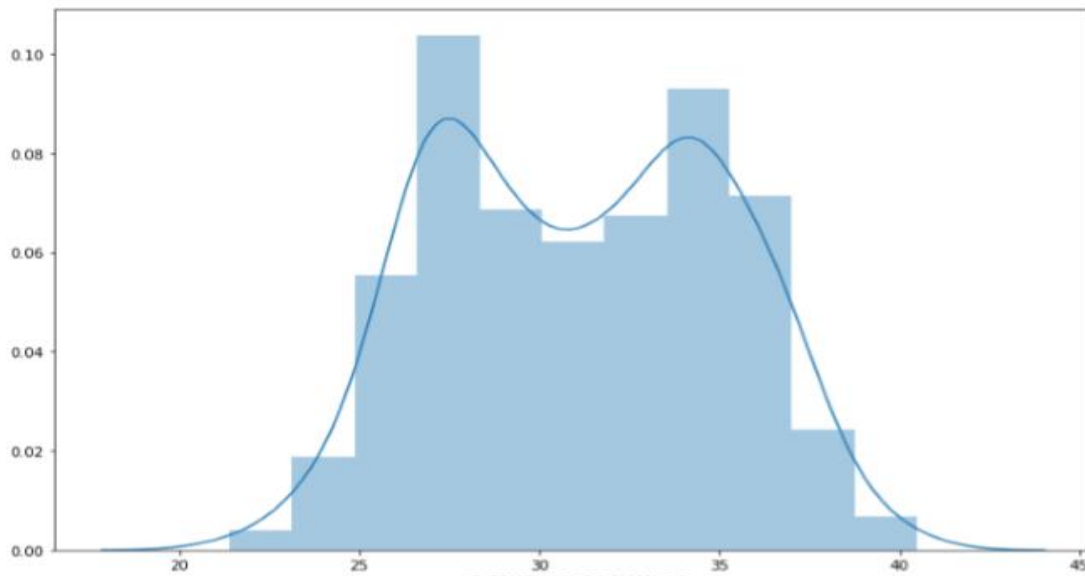
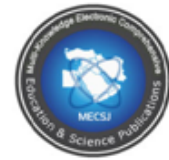


Fig.2 Temperature Analysis



```
data.info()
<'class 'pandas.core.frame.DataFrame>
Int64Index: 3791 entries, 0 to 207
:(Data columns (total 22 columns)
STATION_NUMBER          3791 non-null int64
STATION_NAME            3791 non-null object
YEAR                    3791 non-null int64
MONTH                   3791 non-null int64
DAY                     3791 non-null int64
AIR_TEMPERATURE_MEAN_DB 3791 non-null float64
AIR_TEMPERATURE_MAX_DB  3791 non-null float64
AIR_TEMPERATURE_MIN_DB  3791 non-null float64
WIND_PREVAILING_DIRECTION 3791 non-null object
WIND_MEAN_SPEED         3789 non-null float64
WIND_MAX_DIRECTION      3791 non-null int64
WIND_MAX_SPEED          3791 non-null int64
RELATIVE_HUMIDITY_MEAN  3791 non-null int64
SKY_COV_OKTES_MEAN     3791 non-null float64
RELATIVE_HUMIDITY_MAX   3791 non-null int64
RELATIVE_HUMIDITY_MIN   3791 non-null int64
PRESSURE_MIN_SEA_LEVEL  3791 non-null float64
PRESSURE_MIN_STATION_LEVEL 3791 non-null float64
PRESSURE_MAX_STATION_LEVEL 3791 non-null float64
PRESSURE_MAX_SEA_LEVEL  3791 non-null float64
PRESSURE_MEAN_STATION_LEVEL 3791 non-null float64
PRESSURE_MEAN_SEA_LEVEL 3791 non-null float64
(dtypes: float64(11), int64(9), object(2)
memory usage: 681.2+ KB
```

Fig 3. Data informations

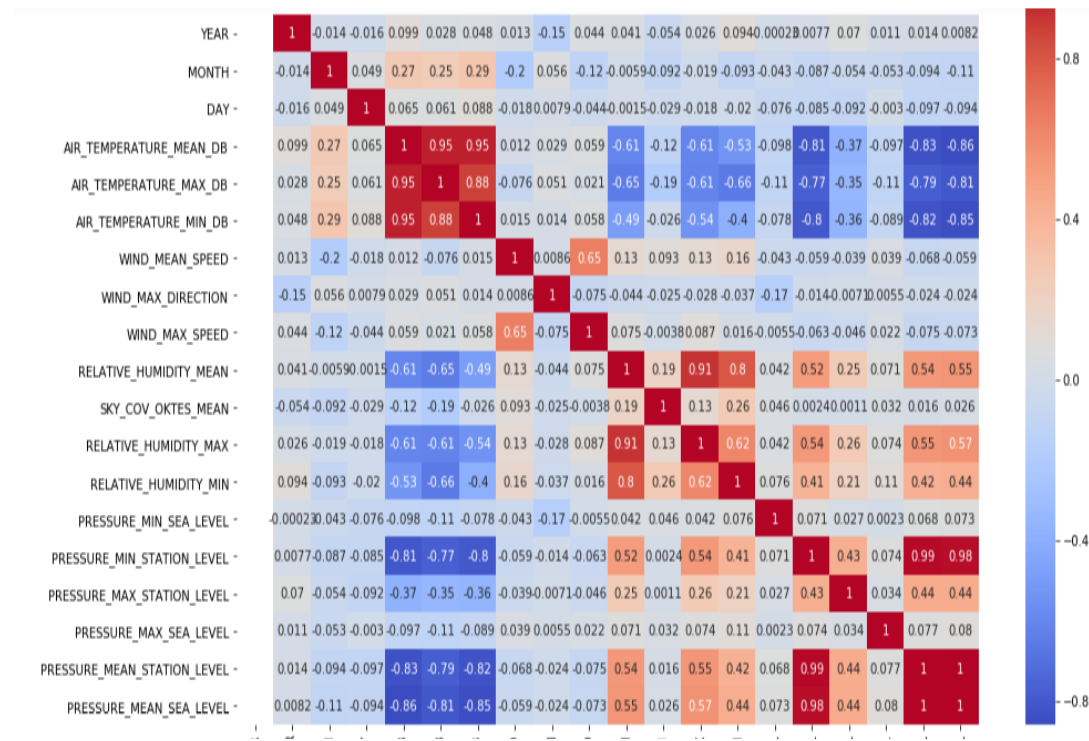


Fig 4. Heatmap



## 4.2 Experiment

In this experiment, we are using Python Libraries, TensorFlow and Keras as Tools and using SKlearn package as one of machine learning tool that contains many methods and functions to apply on the dataset. Firstly, preparing data sets and cleaning it, then split the data set for training dataset and testing dataset either. Finally, using that algorithms for the training process of the model.

## 5. Results and Discussion

The presented section supplies a comprehension of the evaluated parameters and results. Those parameters are showing the efficiency of the suggested technique. Memory of Long Short Term ( LSTM ) consider recurrent neural networks which can be used to learn patterns in sequence data. The data to be analyzed is a time series data with variables like mean air temperature, mean wind speed, year, month, etc. The mean air temperature is to be predicted using all other variables that can provide information about it. The data for a day is used to predict the mean temperature of the next day. The following are the variables that are being used to predict the daily temperature. The training data was reprocessed before make the LSTM model on it. (LSTM) achieved high accuracy for predict temperature which is 91%

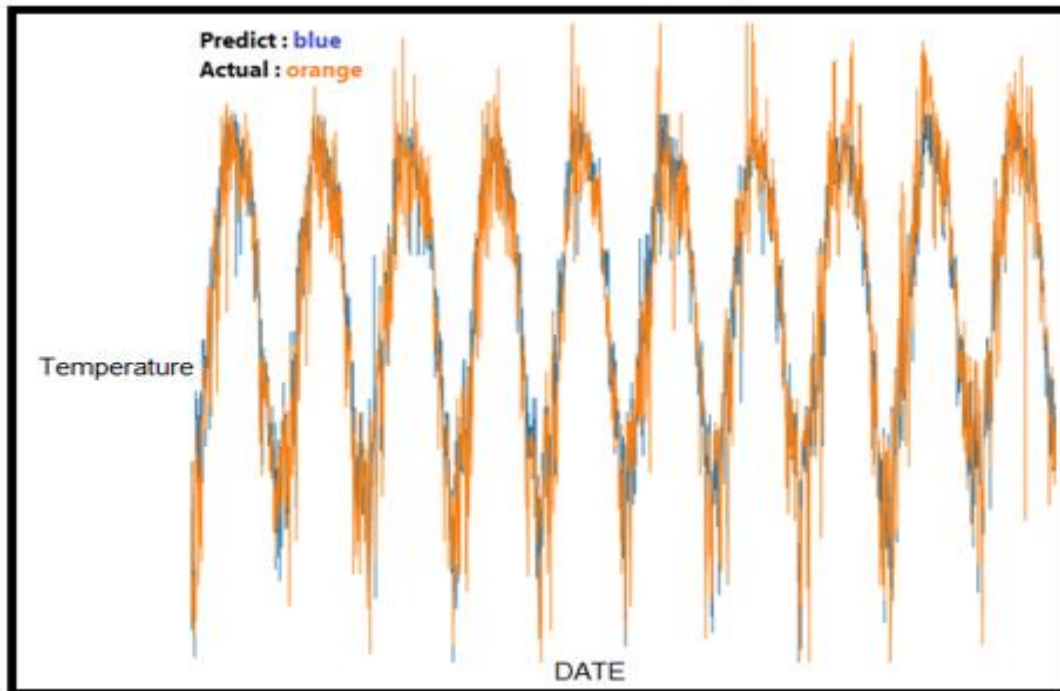


Fig.5 . shows the Prediction of Temperature based on LSTM. the orange color for actual data and the blue for predict

Decision Tree of C4.5: With algorithm of C4.5 Decision Tree, we also achieve high accuracy which is 96% with  $R = 0.9641$

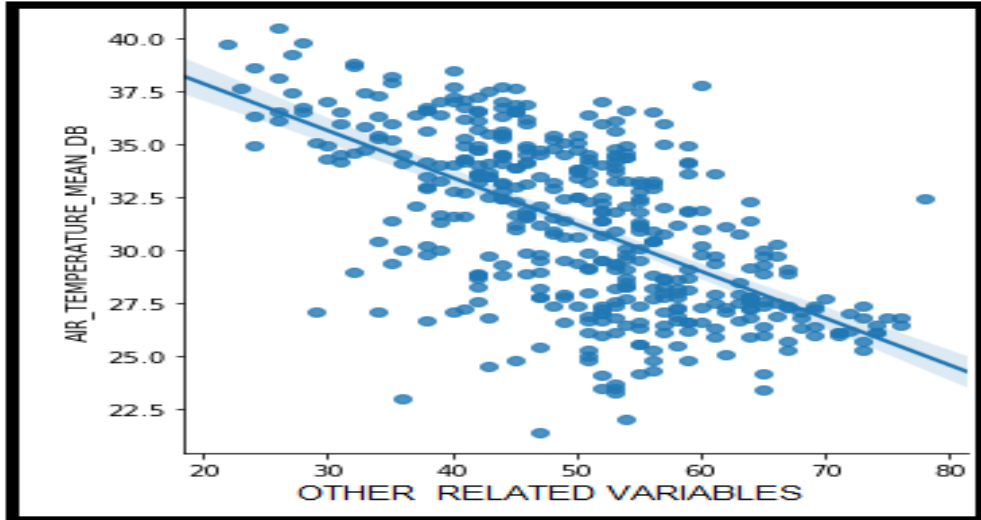


Fig.6 shows the Prediction of Temperature based on the C4.5 DT Algorithm. The points represent other related data and the line for prediction of temperature

Multiple Linear Regression: this algorithm achieve high accuracy which is 95% with R: 0.95551, it is a very excellent result. We used different variables such as wind speed, relative humidity to predict the temperature degree.

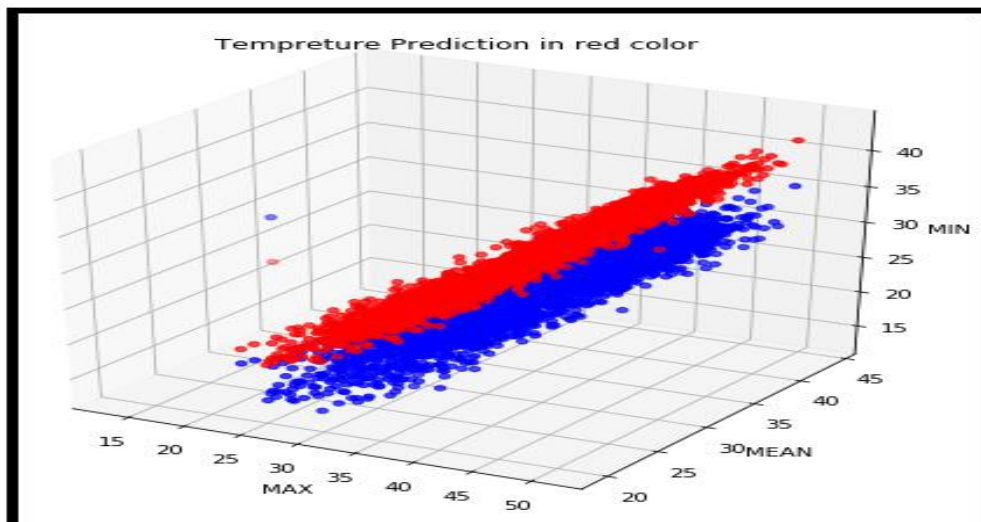
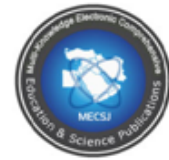


Fig.7 shows the Prediction of Temperature based on the MLR Algorithm

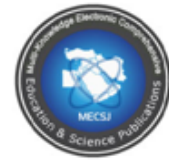




In accordance with the performance comparing, the algorithm of C4.5 confirmed that it's better than other algorithms supplying better impacts. The three algorithms are furthermore utilized studying the new data of concerning weather. The filter of resample was utilized in data pre-processing, moreover in the selection step of data. This has resulted in linear parameters consideration such as min temperature, max temperature, max humidity, mean temperature, min humidity, wind speed, mean humidity, rainfall and cloud cover. The later step for the implementation was to employ the algorithm of long short term memory for the dataset. The data that filtered was provided like an input in the algorithm, the data to be analyzed is a time-series data with variables like mean air temperature, mean wind speed, year, month, etc. The final step was to apply Multiple Linear Regression to the data set. By defining the model in analysis of multiple linear regression, significant consideration is considered as the model suit combining independent variables in a several linear regression, the model always increases the quantity of demonstrated variety for the variable dependent (typically presented as  $R^2$ ). Thus, applying a lot of independent variables except any theoretical justification perhaps caused model of over-fit. Algorithm of Decision tree was predicted like an output for the algorithm of C4.5, the decision tree will include nodes collection and each node is composed of attributes or attributes set as criteria for splitting the node in further classification together with accuracy 96%. By the next Table, we show the Accuracy value for each Algorithm. Results of Models are shown in Table 1

Table 1. Accuracy in each algorithms

algorithms	Performance of Algorithms	Accuracy
<b>DT</b>	The most	<b>96%</b>
<b>MLR</b>	Excellent	<b>95%</b>
<b>LSTM</b>	Excellent	<b>91%</b>

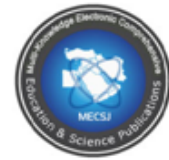


## 5. Conclusion

In this paper, we apply three different algorithms for weather prediction in region of Mecca in Saudi Arabia, we found the accuracy value for each one, these algorithms applied in data set which includes 22 parameters. Although the three algorithms influence was built to become relatively suitable such they fall in the recommended algorithms category in classification and problems of weather prediction. Later, the decision tree evidence to become better than the another two algorithms. As future work, we will apply more complicated algorithms like Dynamic Programming [16] which is a great way to optimize the solution by storing the result of the sub problem. This optimization decrease the exponential complexities to polynomial. Another way is Genetic Algorithms [17][18], is individual solution repeatedly chooses the parent randomly. This algorithm can use to deal with different types of problems like nonlinear problems.

## References

- [1] Karevan Z, Suykens J (2018) Transductive feature selection using clustering- based sample entropy for temperature prediction in weather forecasting. *Entropy* 20(4):264
- [2] Qing X, Niu Y (2018) Hourly day ahead solar irradiance prediction using weather forecasts by lstm. *Energy* 148:461-468
- [3] Xingjian S, Chen Z, Wang H, Yeung DY, Wong WK, Woo Wc(2015) Convolutional lstm network : A machine learning approach for precipitation nowcasting In. *Advances in neural information processing system*, pp 802-810
- [4] Mandale MA, Jadhawar B (2015) Weather forecast prediction: a data mining application. *Int J Eng Res Gen Sci* 3 (2)
- [5] Ali J, Khan R, Ahmed N, Maqsood I (2012) Random forests and decision trees. *International Journal of Computer Science Issues(IJCSI)* 9 (5):272



- [6] Chauhan D, Thakur J (2014) Data mining techniques for weather prediction: A review. *International Journal on Recent and Innovation Trends in Computing and Communication* 2(8):2184-2189
- [7] Kumar A, Sinha R, Bhattacharjee V, Vema DS, Singh S(2012) Modeling using k-means clustering algorithm, In: 2012 1<sup>st</sup> International Conference on Recent Advances in Information Technology (RAIT), IEEE, pp 554-558
- [8] Taksande AA, Mohod P (2015) Applications of data mining in weather forecasting using frequent pattern growth algorithm. *International Journal of Science and Research*
- [9] Joshi S, Pandey B, Joshi N (2015) Comparative analysis of naïve bayes and j48 classification algorithms. *International Journal of Advanced Research in Computer Science and Software Engineering* 5 (12):813
- [10]Patil TR, Sherekar S, et al.(2013) Performance analysis of naïve bayes and j48 classification algorithm for data classification. *International journal of computer science and applications* 6 (2):256-261
- [11]COE J (2012) Performance comparison of naïve bayes and j48 classification algorithms. *International Journal of Applied Engineering Research* 7(11):2012
- [12]Zaytar MA, El Amrani C (2016) Sequence to sequence weather forecasting with long short term memory recurrent neural networks. *International Journal of Computer Applications* 143(11):7-11
- [13]Heinermann J, Kramer O (2016) Machine Learning ensembles for wind power prediction. *Renewable Energy* 89:671-679
- [14]Ivan N (2017) Crime prediction using decision tree (j48) classification algorithm.
- [15]Adamowski J, Fung Chan H, Prasher SO, Ozga-Zienlinski B, Sliusarieva A (2012) comparison of multiple linear and nonlinear regression, autoregressive integrated



moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in montreal, canada. *Water Resources Research* 48(1)

- [16]Adeyanju I (2015) Generating weather forecast texts with case based reasoning. arXiv preprint arXiv: 150901023
- [17]Aybar-Ruiz A, Jimenez-Fernandez S, Cornejo-Bueno L, Casanova-Mateo C, Sanz-Justo J, Salvador-Gonzalez P, Salcedo-Sanz S (2016) A novel grouping genetic algorithm extreme learning machine approach for global solar radiation prediction from numerical weather models inputs, *Solar Energy* 132:129-142
- [18]Singh S, Bhambri P, Gil J (2011) Time series based temperature prediction using back propagation with genetic algorithm technique, *International Journal of Computer Science Issues (IJCSI)* 8 (5):2