12-4-2021

Air Temperature Prediction Using Different Datamining Approaches In Sulaymaniyah City In Iraq

Yusra Mohammed Mohammed Salih
Kurdistan Technical Institute, Sulaymaniyah

Sherko Murad
Kurdistan Technical Institute, Sulaymaniyah

Brzu Tahir
Kurdistan Technical Institute, Sulaymaniyah

Air Temperature Prediction Using Different Datamining Approaches in Sulaymaniyah City in Iraq

Yusra Mohammed Mohammed Salih[1] *, Sherko Murad[2], Brzu Tahir[3]

1 Kurdistan Technical Institute, Sulaymaniyah, Kurdistan Region, Iraq

ABSTRACT

Climate change has a historical impact at universal and local levels over the past era. Climate change is one of the greatest challenge issues in the globe for meteorological research. Air temperature estimation, in particular, has been measured as a significant feature in weather impression studies on industrial sectors, environmental, ecological, and agricultural. Accurately predicting air temperature guides to measure lifestyle, perform a key character for the government, industries, and public in development activities. In this paper, we investigate the use of various data mining approaches such as Support Vector Machine (SVM), Decision tree (DT), and Naïve Bayes for air temperature prediction within Sulaymaniyah City in Kurdistan, IRAQ. The metrological data is collected from the local Weather Forecast Department in the city within the range 2013 to 2018 inclusive. A dataset for the metrological data was developed and used to train the data mining algorithms. The proposed data mining algorithms were tested on the dataset to predict the air temperature and the performance of these algorithms were compared using standard performance metrics. Support vector machine has accomplished promising performance among using algorithms.

1. Introduction

Climate prediction is one of the fascinating challenges which are addressed by the meteorology departments all over the globe [1]. Weather parameters consist of pressure, temperature, precipitation, and humidity. Predicting these parameter values are major challenges for atmospheric research. Air Temperature denotes the value degree of exact hotness of air in a specific place and time and in addition to the time-based and longitudinal progress of the component in various climatical regions [2]. Air temperature prediction has been a serious climatic feature which impacts on the economy, health, energy consumption and food security [3]. Air temperature values, besides the other atmospheric constraints, are measured nearby the earth surface by qualified air temperature values are measured near the surface of the earth by trained observers and automatic weather stations.

Data Mining is one of effective processes used to remove relevant data and information from huge volume of structured and unstructured data [4]. Data mining can be considered as a general case in every data discovery models [1]. Various models are obtainable within the literature based on classification and prediction in different fields such as stock markets, banking sections, healthcare, and weather forecasting as well [5].

The main classification methods used, by researchers, for weather forecasting are Support Vector Machine (SVM), Naïve Bayes, Artificial Neural Network (ANN) and others [4, 5]. Data in different places as India, Korea, South Africa, Malaysia, and Kurdistan are used for predicting climate prediction[5]. For this paper, predicting air temperature is focused in Kurdistan region of Iraq, especially Sulaymaniyah city. The dataset collected from the local climate perdition, Weather forecasting, department in the city and tested for this study.

In this work different supervised learning algorithms such as Support Vector Machine (SVM), Decision Tree, and Naïve Bayes were tested to achieve a comparable investigation regarding their capability to predict the air temperature in Sulaymaniyah city, Kurdistan region, Iraq. Meteorological data are collected from the Weather Forecasting Department within the city. The collected data has been examined, leaked out, and normalized. The most suitable features have been selected to train the data mining approaches for predicting the air temperature.

The rest of the paper is well-organized as follows. Section II, review literature studies of data mining and weather forecasting based on different datasets in different locations. Section III, the area of study and date description are explained well. In section IV, Experimental setup of the research has been explained such

* Corresponding author
E-mail address: yusra.mohammad@kti.edu.krd (Instructor).
Peer-reviewed under the responsibility of the University of Garmian.

Passer Journal, Vol. 3 [2021], Iss. 2, Art. 1
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

as data collection, data preprocessing and data selection, and the algorithms which has been studied for the paper has been explained as well. In section V, the Result and Analysis Discussions are clarified. Finally, in section VI the paper is concluded and future work has been suggested.

2. Literature Review

Many researchers have tried to use data mining techniques in areas related to weather forecasting. various parameters of meteorological conditions like wind speed, humidity, rainfall, air temperature has been predicted using several data mining approaches. For weather forecasting several linear and non-linear datamining techniques are used by researchers.

Researchers in [6] investigates the study to predict rainfall, maximum temperature, wind speed and evaporation using data mining approaches. They used Decision Tree and Artificial Neural Network (ANN) algorithms to test the collected meteorological data. The result shows how the given parameters have persuaded the weather observed in a given period during the study.

In [5], researchers used Support Vector Machine (SVM), Naïve Bayes (NB), and Artificial Neural Network (ANN) supervised learning approaches to achieve a comparable investigation for rainfall forecasting in Sulaymaniyah city in Kurdistan region, Iraq. Regarding the ability for rainfall prediction and concerning the results, the SVM gave the best performance to predict the rainfall.

In [7] the author defines the abilities of several algorithms to predict various climatic phenomena such as wind, humidity, temperature, rain. These parameters concluded that the main techniques such as regression algorithms, artificial neural networks, grouping and decision trees are adequate to forecast climatic conditions. This illustrate that k-means grouping and decision trees are the most appropriate data mining approaches for weather forecasting.

Several methodologies are used for temperature and humidity investigation using data mining algorithms. From the correlated previous works, Support Vector Machine, Artificial Neural Network, Decision Tree, Naïve Bayes, and K-Nearest Neighbor with clustering have been commonly used for air temperature prediction. The main purpose of supervised machine learning techniques is selecting an appropriate technique with the right features to have better performance.

Researchers in [8] used K-nearest neighbor with clustering technique method to predict Temperature and Humidity in a particular time interval. The authors describe the method and clustering techniques to discover the hidden patterns inside a massive weather-related dataset. So, the saved information from the result can be used for weather prediction. High prediction acquired for temperature and humidity.

In [9], a nonlinear method as Artificial Neural Network (ANN) network's compared with regression methods for daily temperature prediction. The dataset used for this work is collected in Mumbai. The performance of each method for temperature prediction has been analyzed. As a result, ANN gives a significant better performance for temperature prediction. In [10], data collected from different cities in Belgium are merged and Multi-View Least Squares Support Vector Machines (MV LS-SVM) regression model is used to predict minimum and maximum temperature. The objective of LS-SVM is to advance the achievement of the model by taking information into account from all views as there are great number of considerations in climate forecasting.

More recently, many approaches have been studied for air temperature prediction in many locations as India, Turkey, Nigeria, Dubai and others [2, 11, 1, 12]. The most widely implemented datamining techniques for temperature prediction are SVM, ANN, K-nearest Neighbors, Naïve Bayes, and others [3, 4].

3. Study Area and Date Description

Sulaymaniyah**,** as shown in the Figure1, is a mountainous city in Kurdistan region, Iraq. Sulaymaniyah is 831 m above sea level and located at 35.53° N 45.45° E. Sulaymaniyah is surrounded by Azmar mountain from Northern-East (1700m), Goizha (1525m), and Qaiwan mountain, and surrounded by Baranan mountain from South (1373m), and from East Shahrazur plain started with 45km length and 15km width [13].

Figure 1: Geographical area of Sulaymaniyah City

Weather of Sulaymaniyah is similar to the Middle-East weather. Temperature in summer reaches approximately 45 ^{0}C and in Winter it reaches 3.5 ^{0}C or under 0 ^{0}C in some places. The average high temperature is 7.9 ^{0}C, in January, to 38,9 ^{0}C, in August [14]. The average low temperature is -0.2 ^{0}C, in January, to 24.1 ^{0}C, in August. Table 1, shows the Temperature in Sulaymaniyah during the year.

Mohammed Salih et al.: Air Temperature Prediction Using Different Datamining Approaches
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

Table 1: Temperature values for each month in the city

	January	February	March	April	May	June	July	August	September	October	November	December
Avg. temperature (˚C)	3.8	5.6	10.4	15.2	21.4	27.6	31.3	31.4	27.2	21.1	13.1	6.7
Min. temperature (˚C)	-0.2	1	5.3	9.7	14.5	20.3	23.9	24.1	19.7	14.1	7.6	2.3
Max. temperature (˚C)	7.8	10.3	15.5	20.8	28.4	35	38.8	38.8	34.7	28.2	18.7	11.1
Avg. temperature (˚C)	38.8	42.1	50.7	59.4	70.5	81.7	88.3	88.5	81.0	70.0	55.6	44.1
Min. temperature (˚C)	31.6	33.8	41.5	49.5	58.1	68.5	75.0	75.4	67.5	57.4	45.7	36.1
Max. temperature (˚C)	46.0	50.5	59.9	69.4	83.1	95.0	101.8	101.8	94.5	82.8	65.7	52.0
Precipitation/Rainfall (mm)	129	146	132	100	41	0	0	0	0	13	74	110

4. Experimental Setup

Primary goal of this work is to deal with air temperature detection in Sulaymaniyah city. The dataset used for this work were gathered from the Weather Forecast Department in the city within the period between 2013 and 2018. The recorded data are average temperature, humidity, steam, station pressure, speed and direction of the wind, cloudiness, and rainfall. Based on the selected features, a suitable dataset is created and were tested using data mining techniques to predict the predictor output variable air temperature in the city. The framework of proposed model for air temperature prediction system is illustrated in Figure 2. Classification and performance analysis of the study performed using Weka. Weka is one of the commonly used data mining software [5].
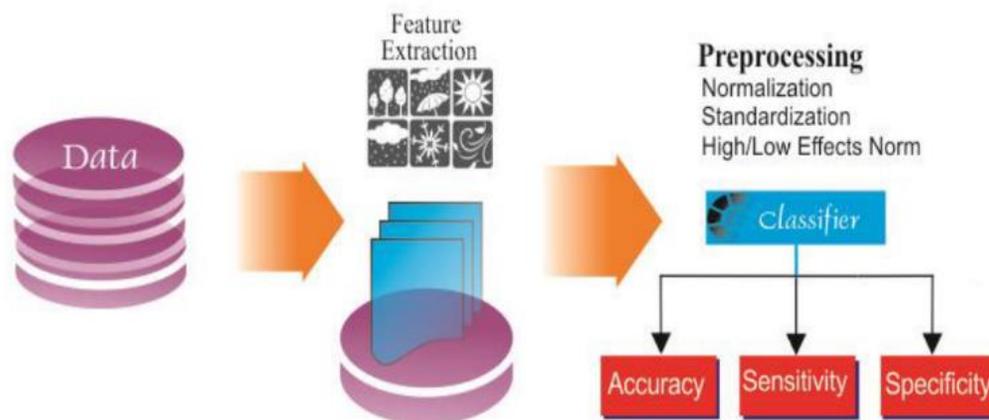

Figure 2: Framework of Proposed Air Temperature Prediction Model

4. 1 Data Collection, Preprocessing, and Data Arrangement

There are several government agencies and some educational institutions that provide access to weather data. The metrological data is collected from the local Weather Forecast Department in the city that gathers weather information over the geographical locations in Sulaymaniyah province. In this Study, we have used weather of Sulaymaniyah city. The daily weather data has been gathered within the range 2013 to 2018 inclusive.

The collected dataset consists of 2,157 instances; The data should be standardized totally to prepare the appropriate data for the proposed methods. This stage contains data preprocessing, and data selection. In data preprocessing stage, the suitable data are selected form the collected dataset and formatted to have 10 columns, 9 as input values and last column as target value. The dataset is preprocessed by examine and identifying the missing,

insufficient and duplicated data and removing them from the dataset. The prepared dataset was transformed into an appropriate format for the data mining approaches.

After the collection and cleaning dataset stage, data selection state has been started. The collected data is classified into nine input features (Wind speed, Wind direction, Humidity, Precipitation, Vapor, Sun, Cloudiness, Sea pressure, Station pressure). The strength of air temperature is represented by these nine input attributes to predict the output attribute air temperature. The measurement for air temperature prediction is divided in three different output values "cold", "mild" and "hot". If the amount of air temperature is greater than or equal to 20, the attribute is identified as "hot". The sample feature for "cold" temperature is created as if the value of air temperature is less than or equal to 10. The generated temperature value between (20 < Temperature > 10) identified as "mild" temperature.

Passer Journal, Vol. 3 [2021], Iss. 2, Art. 1
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

The measurements for air temperature result are shown in Figure 3:
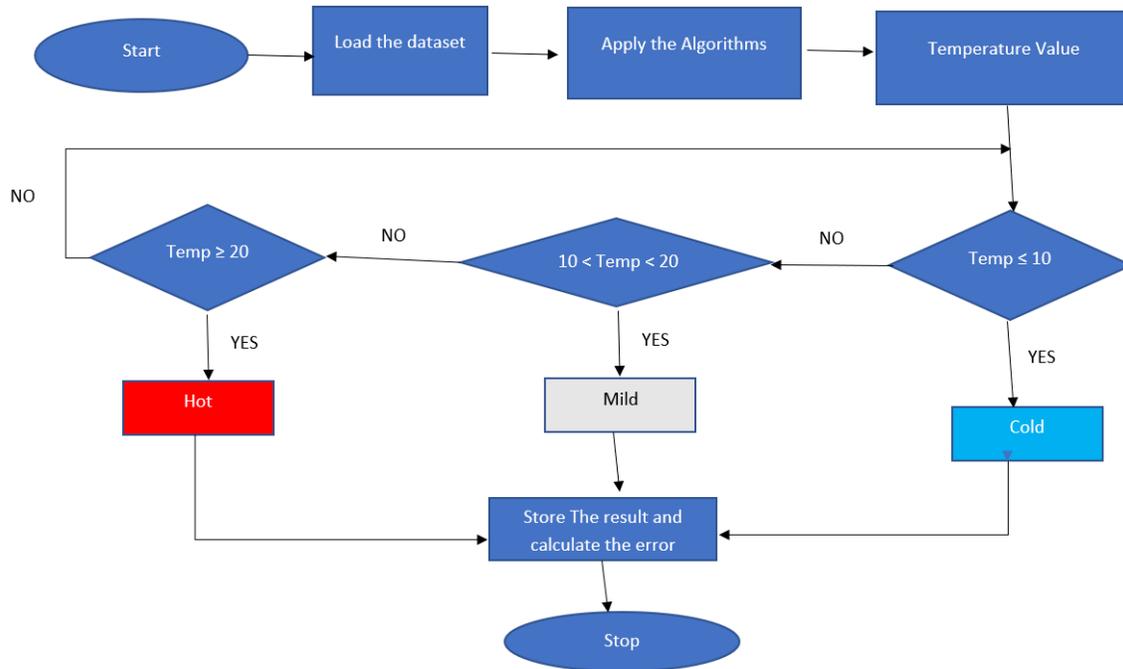

Figure 3: the training process and Descriptive Terms for Temperature value to identify the Cold, Mild, and Hot values

4. 2 Classification and Prediction

Classification is the procedure of selecting a model that defines and differentiates data classes or perceptions to determine the model for being able to use to predict the class of items whose class label is unknown [8]. Objective of this study, is to be able to predict the weather air temperature parameter value with higher accuracy, and prove the forecasting ability of different data mining techniques in the same context. Decision Tree, Support Vector Machine (SVM) with cross validation and Naïve Bayes algorithms are used to evaluate and analyze meteorological data sets. Lastly, each algorithms outcome is examined and the best one has been chosen to predict the air temperature as "Cold", "Mild" and "Hot".

A. Support Vector Machine (SVM)

Among the ML-based approaches, SVM algorithm has been measured as one of the most powerful and accurate methods. It is developed with a kernel-based procedure and has been processed in regression, forecasting, and classification purposes. The primary goal of SVM is to produce an optimal hyper-plane w.x + b = 0 in a new area, where weight vector is referred as w, x refers to the attribute values, and b is the bias. The hyperplane is shaped by plotting the input data x, into a high-dimensional feature space through a nonlinear plotting procedure.

B. Decision Tree

Decision Tree (DT) is data mining approach used for various real-world purposes as an energetic key to classification problem [15]. DT is a tree-like graph which made up of root, internal and leaf nodes. The DT algorithm uses breadth-first approach or depth-first greedy approach to divide the dataset of records until all the data items belong to a target item value being identified. DT

algorithm, Compared to other data-mining techniques, is commonly used in climate forecasting since it is powerful method to data scales or distributions [16].

C. Naïve Bayes

Naive Bayes algorithm is the simple supervised machine classifier. It is mainly based on the Bayesian theorem and is used when the dimensionality of the inputs is high, which suppose that the functions are statistically sovereign. Since it assumes the attribute list $X_1 \ldots X_n$ is all of tentatively sovereign of each other given Y. So, the assumption value clarify the illustration of P(X|Y) and the estimating problem from the highlight training data [17].

4. 3 Performance Evaluation

The performance and efficiency of the proposed methods were examined using the evaluation parameters. Root Mean Square Error (RMSE), Sensitivity (Recall), Precision (Accuracy), and F-measure, with the ROC area under the curve. The confusion matrix is used to test the evaluation and performance of the given algorithms [18]. Confusion matrices denote amounts from predicted and real values. The number of negative examples classified accurately is denoted as True Negative "TN" output. Similarly, the number of positive examples classified accurately is represented as True Positive "TP" output. The number of actual negative examples classified as positive is denoted as the term False Positive "FP" and the number of actual positive examples classified as negative is denoted as False Negative value "FN" [19]. The assessment metrices formula are represented below:

$$Sensitivity\ (Recall) = \frac{TP}{TP+FN} = TP\ Rate$$

Mohammed Salih et al.: Air Temperature Prediction Using Different Datamining Approaches
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

$$Specificity = \frac{TN}{TN+FP} = 1 - FP\ Rate$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+PN}$$

$$Precision = \frac{TP}{TP+FP}$$

Standard deviation, Root Mean Square Error (RMSE) [3], calculates the variance between the investigated data with the observed accurate data. This measure is more sensitive to big calculation errors:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y_i}|^2}$$

F-Measure, average of recall and precision, is used to select best algorithm evaluating the performance, because sometimes the precision and recall giving issues in algorithm selection for performance evaluation [5].

$$F-Measure = \frac{Precision*Recall*2}{Precision+Recall}$$

The ROC, Receiver Operating Characteristic, area under curve, shortened as AUC. It measures the complete evaluation of a dual classifier as a single scalar value. The range of AUC value is within [0.5–1.0], where the performance of a random classifier

denoted as lowest rate and the highest rate denotes an excellent classifier (e.g., with a classification error rate equal to zero) [20].

5. Results and Analysis Explanation:

2,157 instances overall with 10 features are composed to create a dataset suitable for the designs. The dataset is split into 9 attributes as input values and 1 attribute as an output which is the air temperature prediction. The dataset is divided into three sets of air temperature prediction, Cold, Mild, and Hot. Weka data mining tool is used to test the standardized data set with various methods and the evaluation performance of each method is examined.

5. 1 Tested Methods

5. 1. 1 Performance Evaluation using Decision Tree

Decision tree with J48 algorithm and 10-fold cross validation is used to test the dataset. The properly identified instances are 1929 and imperfectly identified instances are 228 together with the root mean square error (RMSE) 0.2507. The method shows acceptable achievement with an accuracy of 89.4298%. The ROC area under the curve for cold, Mild, and Hot is showed in Figure 4. False Positive Rate expressed by X-axis and True Positive Rate expressed by Y-axis.
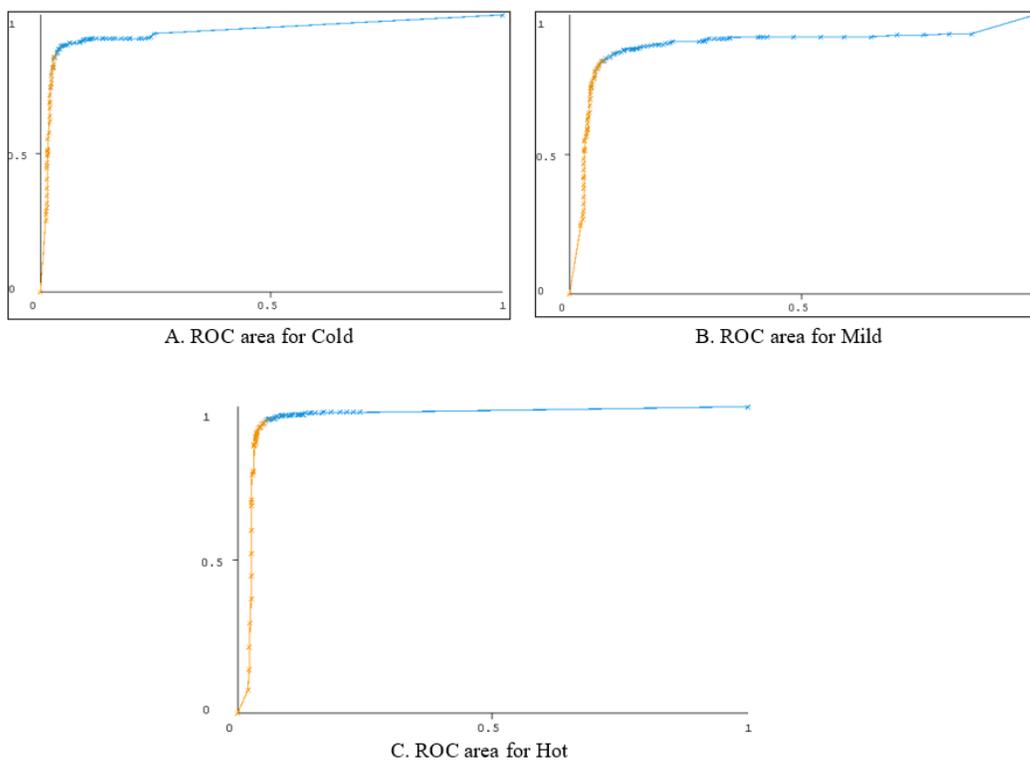

A. ROC area for Cold

B. ROC area for Mild

C. ROC area for Hot

Figure 4: ROC area for the target output (Cold, Mild, Hot) for Decision Tree algorithm

5. 1. 2 Performance Evaluation using Support Vector Machine (SVM)

For dataset classification, a poly-kernel Support Vector Machine is used. Every individual data is charted as a point in n-dimensional space, where n is the number of entities, with the value of a specific coordinate. Assessment performance for the dataset is examined using 10 folds cross-validation test option in

SVM. Properly Identified instances were 1964 and imperfectly identified instances are 193 with the Root Mean Square Error (RMSE) 0.372. The classifier shows superior performance evaluation in comparison to other optimization algorithms with an accuracy of 91.0524%. The ROC area under the curve for cold, Mild, and Hot is showed in Figure 5. False Positive Rate expressed by X-axis and True Positive Rate expressed by Y-axis.
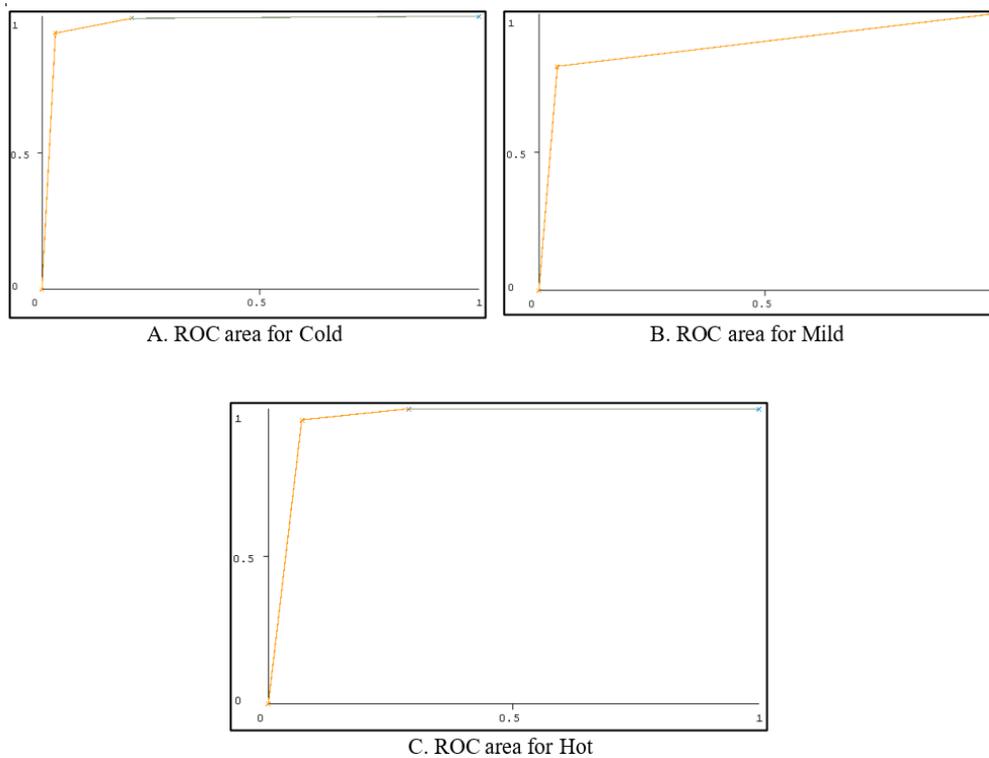
Passer Journal, Vol. 3 [2021], Iss. 2, Art. 1
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

A. ROC area for Cold

B. ROC area for Mild

C. ROC area for Hot

Figure 5: ROC area for the target output (Cold, Mild, Hot) for SVM algorithm

5. 1. 3 Performance Evaluation using Naïve Bayes

The Naïve Bayes algorithm, with Gaussian distribution and 10-fold cross-validation test choice, is selected to assess performance on the dataset. 1776 instances are properly identified, and imperfectly identified instances are 381 together with the root mean square error (RMSE) 0.299. The classifier offers a satisfactory performance evaluation of 82.3366%. The ROC area under the curve for cold, Mild, and Hot is showed in Figure 6. False Positive Rate expressed by X-axis and True Positive Rate expressed by Y-axis.
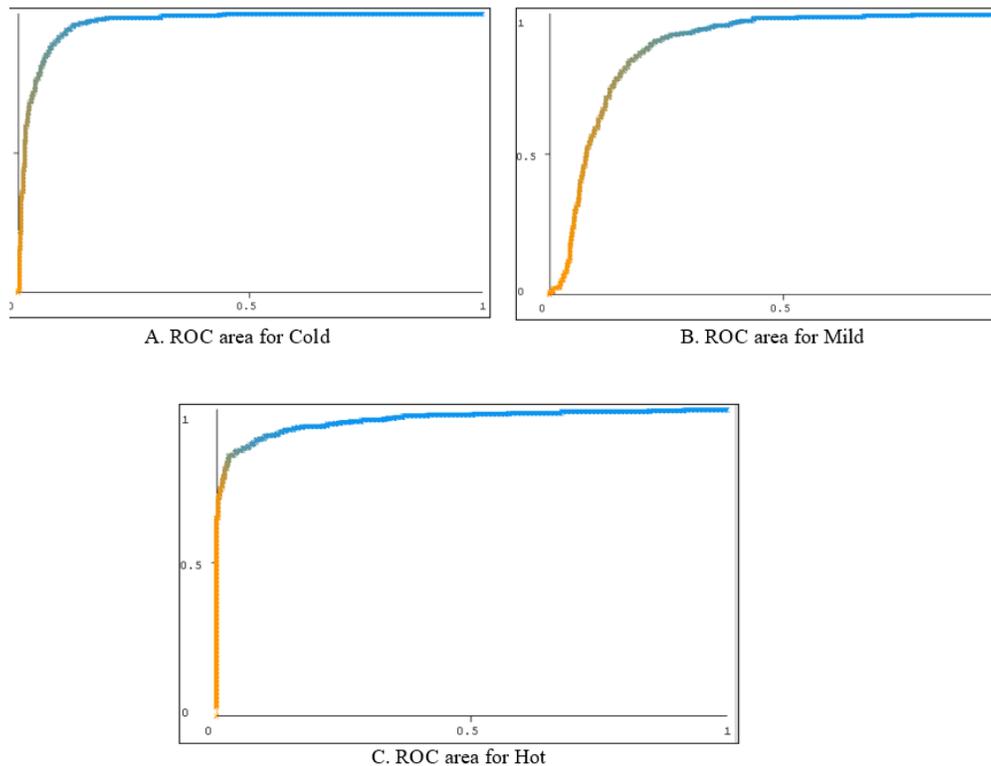

A. ROC area for Cold

B. ROC area for Mild

C. ROC area for Hot

Figure 6: ROC area for the target output (Cold, Mild, Hot) for Naïve Bayes algorithm

Mohammed Salih et al.: Air Temperature Prediction Using Different Datamining Approaches
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

According to the results, the hottest month of the year in the city of Sulaymaniyah is July in which the highest temperature in the month reaches 40 degrees Celsius and the coldest month is February, where the temperature drops below freezing, and it is the time of the snowfall. The predicted output of air temperature in each month during the year is exposed in the Figure 7.
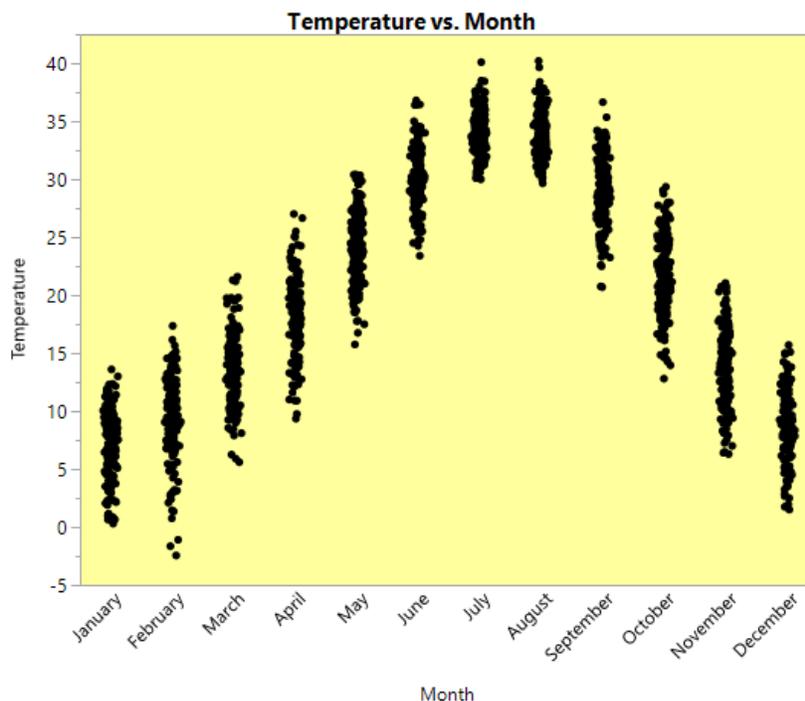

Figure 7: Air temperature in each month of the year

The results of the prediction profiler, expressed in Figure 8, show that as the predictor variables of humidity, precipitation, station pressure and sea pressure increase, the air temperature decreases. This indicates the relation between the target value of air temperature and these predictor variables are negative. In contrast, as vapor increases, air temperature also increases, and shows the positive relationship between these two variables. On the other hand, by increasing the other predictor variables, the air temperature remains at the same level, namely, there is no vital connection between these predictor variables and the target variables of air temperature.
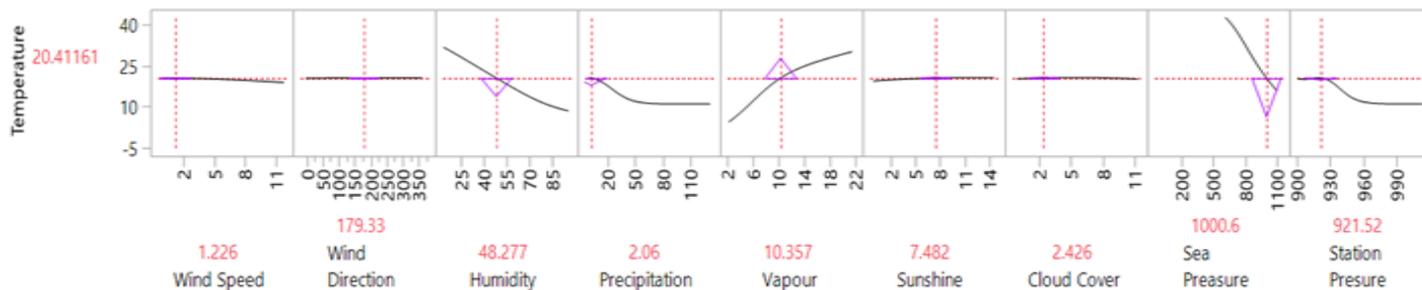

Figure 8: Prediction Profiler for Air Temperature

5. 2 Comparison on the algorithms:

The performance of the suggested datamining algorithms for the dataset are analyzed as shown in Table 2. The outcome expressly shows that SVM, with 10-fold cross-validation test option, is the best algorithm for predicting air temperature, which it provides the 91.5624%. The Accuracy result for air temperature of each algorithm is plainly denoted in Figure 9.

Table 2: Comparable Examination for The Suggested Algorithms

Algorithm	Accuracy	Precision	Recall	F- measure	ROC Area
SVM	91.0524%	0.871	0.938	0.903	0.973
Decision tree (J48)	89.4298%	0.852	0.846	0.849	0.936
Naïve Bayes	82.3366%	0.813	0.726	0.767	0.965

Passer Journal, Vol. 3 [2021], Iss. 2, Art. 1
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

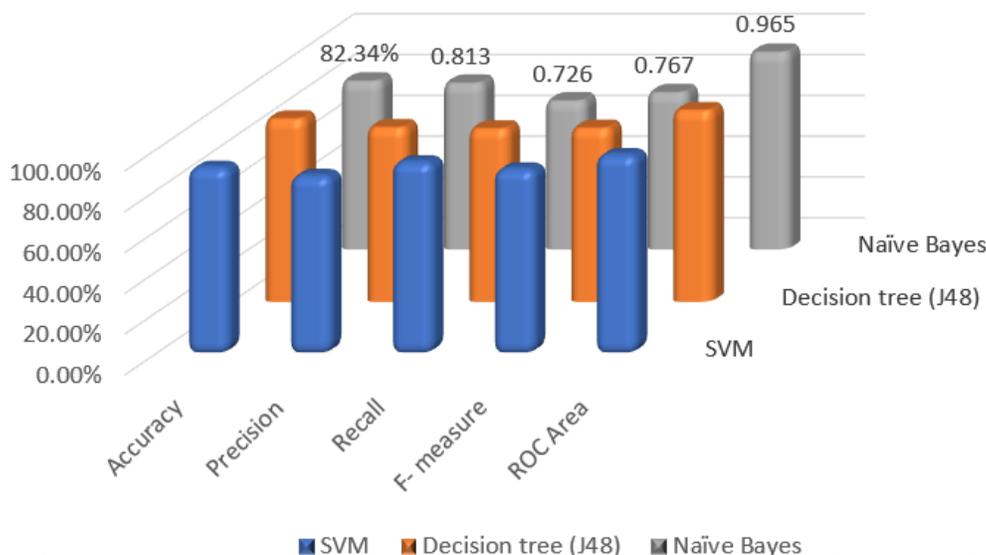## Air Temperature Evaluation Result

Figure 9: Comparison between three different data mining techniques for air temperature prediction

6. Conclusion and Future Work

The objective of this work is to predict the air temperature in the city of Sulaymaniyah using various data mining approaches such as Support Vector Machine (SVM), Decision tree (DT), and Naïve Bayes. The data composed from Weather Forecast Department in Sulaymaniyah city and used for this work. The historical data covers the date from 2013 to 2018. The analysis performance of the used data mining approaches is achieved using three precision measures: Accuracy, recall and F-measure. According to the results, the predictive variables of humidity, precipitation and sea pressure are significant and have a negative influence on air temperature. In other words, as each unit of these climate predictor variables increases, the climate air temperature decreases. In contrast, air temperature increases as steam increases one more unit in the climate. In other words, steam has a positive effect on air temperature. In addition, this study found that the air temperature exceeds 40 degrees Celsius in the month of July, and becomes the hottest month of the year in the Kurdistan region of Iraq, while the month of February records the lowest degree below zero air temperature. Based on the results, the used classification techniques worked sensible, but SVM is the best data mining approach for predicting air temperature, which provides 91.57% in return. Future work suggestion is that improvement in data mining algorithms can be obtained for air temperature prediction with better performances. As well as data for the year 2019 to current date can be added to test the dataset and getting better performance.

Authors Contribution

This paper focuses on air temperature forecasting in the city of Sulaymaniyah. The dataset is gathered from the local forecasting department in the city. This study focuses on three different supervised data mining techniques to forecast the air temperature in the region, including SVM, Naive Bayes, and Decision Tree.

The result of the algorithms is compared and analyzed based on their ability to forecast the air temperature in the area of study.

References

1. G. V. Krishna, "A Review of Weather Forecasting Models-Based on Data Mining and Artificial Neural Networks," *Comput. Scince Electron. J.*, vol. 6, pp. 214–222, 2015, doi: 10.090592/IJCSC.2015.617.

2. A. You, M. A. Y. Be, and I. In, "Bayesian network model for temperature forecasting in Dubai Dubai," vol. 100006, no. October, 2018, doi: 10.1063/1.5064935.

3. J. Cifuentes, G. Marulanda, A. Bello, and J. Reneses, "Air temperature forecasting using machine learning techniques: A review," *Energies*, vol. 13, no. 6, pp. 1–28, 2020, doi: 10.3390/en13164215.

4. D. Chauhan and J. Thakur, "Data Mining Techniques for Weather Prediction: A Review," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 2, no. 8, pp. 2184–2189, 2014.

5. Y. Mohammed, "COMPARABLE INVESTIGATION FOR RAINFALL FORECASTING USING DIFFERENT DATA MINING APPROACHES IN SULAYMANIYAH CITY IN IRAQ," vol. 4, no. 1, pp. 11–18, 2020, doi: 10.18488/journal.72.2020.41.11.18.

6. F. Olaiya and A. B. Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies," *Int. J. Inf. Eng. Electron. Bus.*, vol. 4, no. 1, pp. 51–59, 2012, doi: 10.5815/ijieeb.2012.01.07.

7. C. Sahu, "A N I NTELLIGENT A PPLICATION O F F UZZY I D 3 T O," vol. 2, no. 1, pp. 17–22, 2013.

8. S. Badhiye, "Temperature and Humidity Data Analysis for Future Value Prediction using Clustering Technique : An Approach," no. February 2012, 2015.

9. A. Sharaff and S. R. Roy, "Comparative Analysis of Temperature Prediction Using Regression Methods and Back Propagation Neural Network," *2018 2nd Int. Conf. Trends Electron. Informatics*, no. Icoei, pp. 739–742, 2018.

Mohammed Salih et al.: Air Temperature Prediction Using Different Datamining Approaches
Mohammed salih et. al. Passer 3 (2021) 114-122

Passer

10. L. Houthuys, Z. Karevan, and J. A. K. Suykens, "Multi-View LS-SVM Regression for Black-Box Temperature Prediction in Weather Forecasting," pp. 1102–1108, 2017.

11. M. Yadav, S. Jain, and K. R. Seeja, *Prediction of air quality using time series data mining*, vol. 56. Springer Singapore, 2019.

12. M. Yesilbudak, S. Sagiroglu, and I. Colak, "A new approach to very short term wind speed prediction using k -nearest neighbor classification," *ENERGY Convers. Manag.*, vol. 69, pp. 77–86, 2013, doi: 10.1016/j.enconman.2013.01.033.

13. S. Governate, "Sulaymaniyah Governate," 2018. [Online]. Available: http://slemani.gov.krd/so/pageDetail.php?secID=3#.

14. P. Unit, "Sulaymaniyah Governorate Profile," no. December 2015.

15. Z. U. Khan and M. Hayat, "Hourly Based Climate Prediction Using Data Mining Techniquesby Comprising Entity Demean Algorithm," vol. 21, no. 8, pp. 1295–1300, 2014, doi: 10.5829/idosi.mejsr.2014.21.08.21413.

16. M. Muqeem and N. Javed, "A Critical Review of Data Mining Techniques in Weather Forecasting," vol. 5, no. 4, pp. 1091–1094, 2016, doi: 10.17148/IJARCCE.2016.54266.

17. S. D. Jadhav and H. P. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," *Int. J. Sci. Res.*, vol. 5, no. 1, pp. 1842–1845, 2016, doi: 10.21275/v5i1.nov153131.

18. Y. M. M. Salih, T. Cevik, and A. Kattan, "Detection of Motorway Disorders by Processing and Classification of Smartphone Signals Using Artificial Neural Networks," *Int. J. Nat. Sci. Res.*, vol. 4, no. 3, pp. 56–67, 2016, doi: 10.18488/journal.63/2016.4.3/63.3.56.67.

19. A. Kulkarni, D. Chong, and F. A. Batarseh, *Foundations of data imbalance and solutions for a data democracy*. Elsevier Inc., 2020.

20. A. Hanley, J. Mcneil, and D. Ph, "under a Receiver Characteristic," pp. 29–36.