Weather Prediction System Using KNN Classification Algorithm

Yousif Elfatih Yousif

ABSTRACT

Data Mining is a technology that facilitates extracting relevant and which have factors in common from the set of data. It is the process of analysis data from different perspectives and discovering problems, patterns, and correlations in data sets that are useful for predicting outcomes that help you make a correct decision. Weather Prediction is a field of meteorology that is created by collecting dynamic data related to the current state of the weather such as temperature, humidity, rainfall, wind. In this paper, we designed a system using a classification method by k-Nearest Neighbors algorithm for predict whether through previous data to determine the expected temperature and humidity the prediction results were compared with real results, the comparison was good and acceptable.

Keywords: Correlations, classification, data Mining, discovering problems, k-Nearest neighbors algorithm, patterns.

Y. Elfatih Yousif
Department of Computer Engineering, Faculty of Engineering, Alzaiem Alazhri University, Khartoum, Sudan.
(e-mail: yousifsiddiq@gmail.com)

I. INTRODUCTION

Data mining is the process of analyzing data in different ways from a different perspective collected and summarized to give useful information for use or is the process of using a variety of Data analysis tools to discover patterns and relationships between different data that can be used to make the predictions correct. It is a mixture of statistics, mathematics and computer software [1]. Data mining had perfect attendance in the computer field and in society in general in last years, due to the wide availability of big amounts of data and the great need for turning such data into useful information and knowledge. The knowledge can be used for different applications such as sales analysis, medical fields, production control and science exploration [2]. Data mining can be used in all fields, for example, in the field of finance and business it can predict profit or loss and can also be used in Credit cards and the expected risks in this area and who is the expected customer to be good and who is bad. also, in the health field and predicting possible genetic diseases in the future, such as Diabetes, pressure, and heart disease. It can also be used in the security and military field, and it can also be used in the weather, as this paper presents [3]. The variety in the use of weather forecasts is innumerable, such as weather warnings; It protects and reduces the material loss of life and property. Weather forecasts that focus on knowing the temperature and the probability of precipitation since daily activities may change due to heavy rain, snow, wind or hail, weather forecasts can be used to plan the dates and locations of different activities during those weather conditions [4]. Data mining is one of the fastest-growing fields in computer science, and its popularity and spread came from the increasing need for tools that help analyze and comprehend large amounts of data. There are two kinds of

knowledge under this amount Data Huge: First Line-On Analytical Processing OLAP (Processing Analytical) and Data Mining, DM (Mining) and both types are analytical tools based on data warehouses Fig. 1. gives an overview of the conceptual structure of a data warehouse, data mining and OLAP [3].

All data collected by organizations will not serve any purpose without knowledge discovery. Data mining helps organizations visualize hidden patterns and trends in data sets that may not have been visible before. Whatever insights are revealed will lead to clear decisions [2].
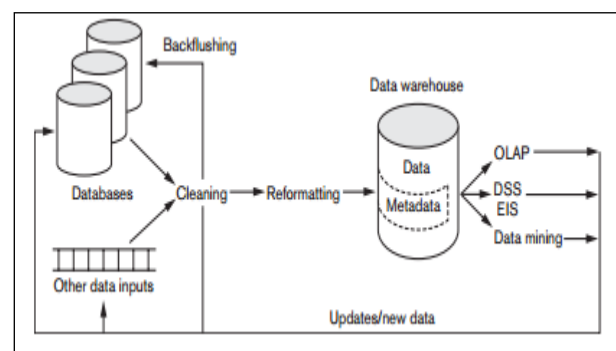

Fig. 1. Structure of a Data warehouse, Data mining and OLAP.

II. METHODOLOGY

This paper presented weather predicting model by using one of the most important data mining techniques the model was designed by classification using k-Nearest Neighbors algorithm, (KNN) is a simple algorithm, also is supervised machine learning algorithm that is widely used to solve classification problems. KNN algorithm assumes that similar things exist in close proximity. In other words, similar things

are near to each other, to design and implement the proposed system was used Visual Studio Environment.

III. PROCESS OF KNOWLEDGE DISCOVERY

Data mining aims to discover and extract useful patterns, and it is a useful technology that has become important with the rapid development and spread of the use of databases, and the use of data mining facilitates in all areas the ability to explore the most important information in the databases. Data mining techniques are used to build future predictions and explore behavior and trends, allowing for correct decisions to be estimated. It also answers many questions in a short time, especially questions that are difficult to answer using traditional statistical methods, data mining mean knowledge discovery from data, depicted in Fig. 2 and consists of an iterative sequence of the following steps:

1. Data cleaning: firstly must delete noise and inconsistent data.
2. Data integration: in this stage where combined between multiple data sources
3. Data selection: where data relevant to the analysis task are retrieved from the database
4. Data transformation: data are transformed into forms appropriate for mining by performing summary or aggregation operations.
5. Data mining: the main process to be applied in order to extract data patterns
6. Pattern evaluation: in this stage identify the truly interesting patterns representing knowledge based on some interestingness measures.
7. Knowledge presentation: using knowledge representation and visualization techniques to present the mined knowledge to the user [5].
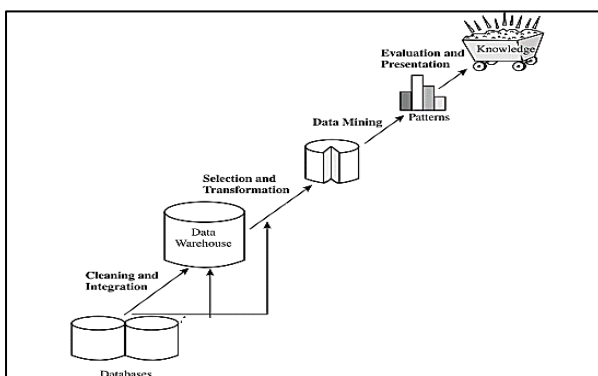

Fig. 2. Data mining as a step in the process of knowledge discovery.

IV. TYPES OF DATA MINING

When we deal with data mining, we find that there are two methods that can be used in data mining

1) Description Methods

This method works to get specific patterns that describe the data, this method contains the following types:

A. Clustering

Its objective is to identify trends within the data and this type tries to find groups of elements that exist together, and it divides the data into a group of categories based on similar properties [6].

B. Association Rule Discovery

Finding new dependency rules which will predict the occurrence of an item based on occurrences of other items, based on a set of records each of which contains some number of items from a given collection data [7].

C. Sequential Pattern Discovery

Sequencing techniques use time-based analysis to extract useful information and are similar to clustering in that they are used to determine which elements occur together and which elements occur first [6].

1) Prediction Methods

This method will take real variables to predict unknown or future values of other variables, this method contains the following types:

D. Classification

Classification is used to solve many problems by analyzing a set of data and placing it in the form of categories that can be used later to predict future data [8].

E. Regression

Predict a new value of a given continuous-valued variable based on the other values of variables, assuming a linear or non - linear model of dependency [7].

F. Deviation Detection

This method checks for abnormal deviations from normal behavior applications.
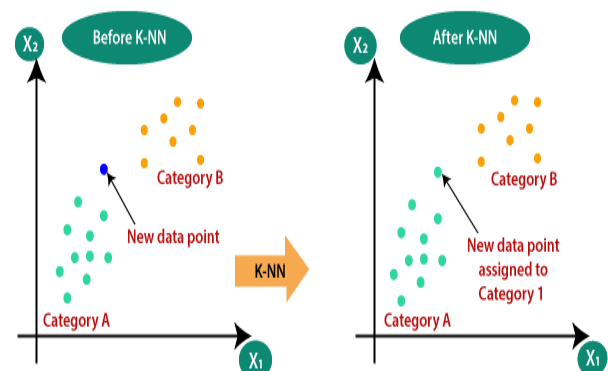
V. PROPOSED SYSTEM

Fig. 3.K-NN Algorithm.

The operating of K-NN classifier is as follows:

1. Value of K is initialized.
2. The distance between input sample and coaching samples is calculated:

$$dist((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \quad (2)$$

3. Sort the distances.
4. Take high K- nearest neighbors.
5. Apply easy majority.
6. Predict category labels with additional neighbors for input sample [9]. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data. the Fig 4 shown the steps of proposed system.

Weather prediction has often been a wonderful field of research because changes in climatic conditions greatly affect people, in this paper we use k-Nearest Neighbors algorithm for weather prediction .The main idea on which k-Nearest Neighbors works is that the info is connected in an exceedingly feature space. wherefore, to search out the space among knowledge points, all the points are considered in order. The distance is calculated using Euclidian distance based on the data type of data classes used. A single value of K is given, and it is used to find the total number of nearest neighbors that determine the class label for unknown sample. If the worth of K=1, then it's referred to as thenearest neighbor classification [9]. Suppose there are two classes, i.e., class A and class B, and that we have a brand new x1, thus this information can be that of those classes. to resolve this problem, we need a K-NN algorithm. With the assistance of K-NN, we will simply determine the class or category of a specific dataset. take into account Fig 3.
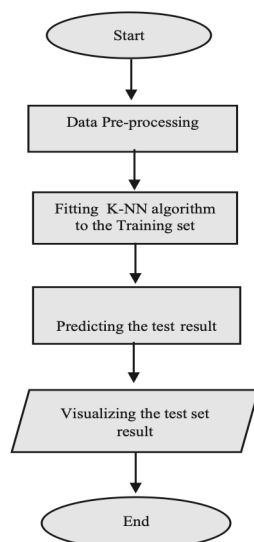

Fig. 4. Steps of proposed system.

VI. RESULT AND DECUSSATION

Fig. 5. Entry form.

This paper presented the work to predict day weather conditions in this system has been registered the previous seven days weather and weather conditions of past seven years for the same day with the same time chosen for all values, this data will be saved together in a database file before entered into the system, entering the specified values shown in Fig 5.

Fig. 6. Predicted Form.

The weather is predicted for today It will be predicted using a k-Nearest Neighbors algorithm to find the distance among data points, all the points are considered in order and predicted the weather (temperature and humidity) the expected values will be displayed in Fig 6.

This system was implemented by Visual Studio, Visual Studio code could be a source-code editor created by Microsoft for Windows, Linux, and macOS. options embrace support for debugging, syntax highlight, intelligent code completion, code refactoring, and embedded Git.

TABLE I: THE TEMPERATURE COMPARISON

NO	Predicted Temperature	Actual Temperature
1	23	24
2	23	23
3	24	23
4	24	25
5	26	26
6	25	24
7	24	24
8	25	25
9	26	24
10	23	23
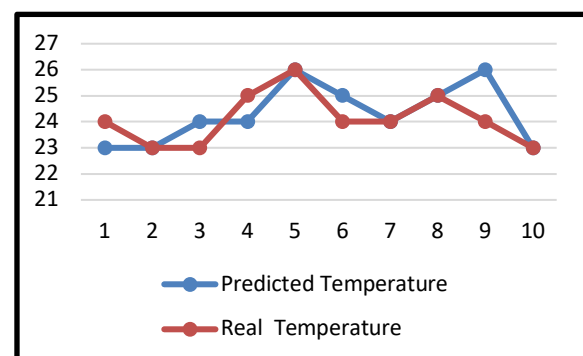
Fig. 7. Temperature comparison.

TABLE II: THE HUMIDITY COMPARISON

NO	Predicted Humidity	Actual Humidity
1	31%	31%
2	33%	34%
3	33%	32%
4	34%	34%
5	30%	31%
6	34%	34%
7	32%	30%
8	33%	33%
9	33%	35%
10	34%	34%

After the results appeared, a comparison was made between the actual temperature and the predicted temperature, shown in Table I and Fig. 7, and we notice a remarkable convergence of the results.

Also, a comparison was made between the actual humidity and the predicted humidity, shown in Table II and Fig. 8, and we notice a remarkable convergence of the results. weather prediction approaches are challenged by complex weather phenomena with limited observations and past data, for this reason, we must always improve and renew the method of weather prediction.
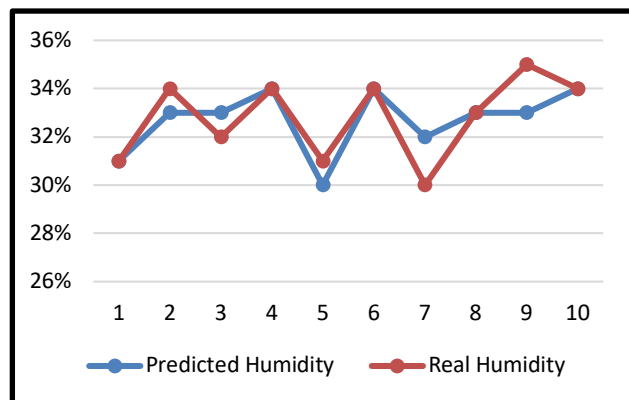

Fig. 8. Humidity comparison.

VII. CONCLUSION

Data mining is one of the highly efficient methods of analyzing and compiling data and reaching relationships that increase predictability and provide the best possible goals for knowledge exploration. this paper concludes that using Data mining techniques for weather prediction (temperature and humidity) gives good results and can be considered as an alternative to traditional metrological approaches.

A comparison is made in this paper, which shows that good results between predicted and real values the K-NN algorithm are one of best suited classifications technique for this application, one of the major concerns of Weather prediction is its accuracy and its timely output. For this reason, the main dependence in weather forecasting is on the basis of data mining techniques, such as what was presented in this paper using the KNN classification algorithm.

REFERENCES

[1] Introduction to Data Mining and Knowledge Discovery. Two Crowds Corporation, [Internet] 2009. Available from: http://www.twocrows.com/introdm.pdf
[2] Olaiya F, Barnabas A. Applications of Data Mining Techniques in weather prediction and climate change studies. *MECS.* 2012.
[3] Shu-Hsien L, Pei-Hui C, Pei-Yuan H. *Data mining techniques and applications – A decade review from 2000 to 2011.* 2012
[4] Folorunsho O. Application of Data Mining Techniques in Weather Prediction and Climate Change Studies. *I.J. Information Engineering and Electronic Business.* 2012; 1: 51-59.
[5] Dickey DA, N. Carolina State U., Raleigh NC. Introduction to Predictive Modeling with Examples. *Statistics and Data Analysis*. 2012
[6] Kapoor P, Singh Bedi S. Weather Forecasting Using Sliding Window Algorithm. *ISRN Signal Processing.* 2013.
[7] Jayavani K. Statistical classification in machine intellegent. *ISR Journals and Publications.* 2014;1.
[8] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. 2014.
[9] Gupta D, Ghose U. A Comparative Study of Classification Algorithms for Forecasting Rainfall. *IEEE.* 2015.
[10] Hayati M, Mohebi Z. Temperature Forecasting Based on Neural Network Approch. *World Applied Science Journal.* 2007; 2(6): 613-620.
[11] Dunne S, Ghosh B. Weather adaptive traffic prediction using neuro-wavelet models. *WCTR.* 2013.
[12] Wang J, Su X. An improved K-Means clustering algorithm. *IEEE.* 2014.
[13] Hand, Manilla, Smyth, Principles of Data Mining. *The MIT Press.* 2001
[14] Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. 2012.
[15] Witten IH, Frank E. Data Mining Practical Machine Learning Tools and Techniques. 2015.