International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 5 ǁ 2022 ǁ PP. 36-40

Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data

TARUN SARAWGI[1] TUSHAR SHARMA[2] YASH MISHRA[3]
Galgotias University , Gautam Budh nagar , Greater Noida
Under the supervision of Miss Archana(Ass.

Date of Submission: 28-04-2022 Date of acceptance: 09-05-2022

ABSTRACT

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Here this system will predict weather based on parameters such as temperature, humidity and wind. This system is a web application with effective graphical user interface. To predict the future's weather condition, the variation in the conditions in past years must be utilized. The probability that it will match within the span of adjacent fortnight of previous year is very high. We have proposed the use of K-medoids and Naive Bayes algorithm for weather forecasting system with parameters such as temp., humidity, and wind. It will forecast weather based on previous record therefore this prediction will prove reliable. This system can be used in Air Traffic, Marine, Agriculture, Forestry, Military, and Navy etc. Weather forecasting is the attempt by meteorologists to predict the weather conditions at some future time and the weather conditions that may be expected. The climatic condition parameters are based on the temperature, wind, humidity, rainfall and size of data set. Here, the parameters temperature and Humidity only are considered for experimental analysis.

General Terms

Data Mining, Classification, Prediction.

Keywords

Chi square, Classification, Naïve Bayes, Prediction, Weather Forecasting

1. INTRODUCTION

Weather forecasting has been a standout amongst the most experimentally and technologically troublesome issues over the world in the most recent century [1]. Environmental change has been looking for a great deal of consideration since a long time because of the sudden changes that happen. There are several limitations in better execution of weather forecasting thus it ends up hard predicting weather here and now with effectiveness [4][18]. Weather forecasting assumes a significant role in meteorology [6] [9].To makes an exact prediction is one of the significant troubles standing up to meteorologist wherever all through the world. Weather warnings are vital in light of the fact that they are utilized to ensure life and property. Forecasts dependent on Temperature, Outlook, Humidity and Wind are important to farming, and along these lines to traders inside product markets. Temperature forecasts are utilized by utility companies to assess request over coming days. Since outdoor activities are seriously reduced by substantial rain, snow and wind chill, estimates can be utilized to design activity around these occasions, and to prepare and survive them. Without precise weather forecasts individuals may end up in hazardous circumstances as they were unprepared for and end up harmed or worse-[1].

The difficulties of weather forecasting, among others, are learning weather representation utilizing an enormous volume of weather dataset. For this purpose, analysis of different data mining procedure is performed. Data mining techniques enables users to analyze data from a wide range of dimensions or angles, classify it, and condense the connections recognized. Some fundamental terms related to Data Mining are: Classification, Learning and Prediction. Classification is a data mining (machine learning) method used to predict aggregate participation for information cases. For instance, classification can be utilized to predict whether the weather on a specific day will be "sunny", "rainy" or "cloudy" [1].Learning refers to training and mapping contribution to yield information. It tends to be performed in two different ways: Supervised and Unsupervised learning. A supervised learning algorithm analyzes the training data and produces a derived capacity utilizing Classifier [17]. In machine learning, unsupervised learning alludes to the issue of trying to hidden structure in unlabeled information [1]. Since the precedents given to the learner are unlabeled, there is no mistake or reward signal to assess a potential solution. This recognizes unsupervised learning from supervised learning. Prediction identifies with modeling and the logical relationship of the model sooner or later. Finding patterns and data may prompt sensible predictions [17].

This paper exhibits a classifier approach utilizing Naive Bayes and Chi Square strategy for weather forecasting. In this system, state of weather is classified in some attribute like as Outlook, Temperature, Humidity, and Wind. Using those attribute the system will predict the class label as Weather Forecasting (Good/Bad).In the system two basic functions namely classification (training) and prediction (testing) will be performed.

The contents of this paper are illustrated as takes after: section two gives review of some related works, section three briefs on Methodology, section four incorporates methodology with design architecture, section five shows the experimental result and analyses it and last segment is committed to the conclusion.

2. RELATED WORK

In the most recent decade, numerous significant efforts to solve weather forecasting issue utilizing statistical modeling including machine learning systems have been reported with successful results [8][11][12][13]. Different Methods has been utilized in Weather Prediction System, for example, neural network-based algorithm utilizing Back Propagation Neural Network (BPN) and Hopfield Network [5], Recurrence Neural Network (RNN), Conditional Restricted Boltzmann Machine (CRBM), and Convolutional Network (CN) models [8], Artificial Neural Network and Decision tree Algorithms [6], predictive analysis in Apache Hadoop Framework utilizing Naive Bayes Algorithm [4].

2.1 BPN and Hopfield Network

In this work Back Propagation Neural (BPN) Network is utilized for initial modeling. The outcomes acquired by BPN model are sustained to a Hopfield Network. In BPN, the info and yield layer comprises of 3 neurons where as the hidden layer has 5 neurons and Hopfield Network display work with the assistance of training data set [5]. The system must perform Temperature or Wind Speed or Humidity flow with the end goal to establish equilibrium. This procedure will proceed iteratively and in every iteration bias and weight esteems should be updated until it converges.

2.2 RNN, CRBM and CN models

The goal of this work is to investigate the capability of profound learning technique for weather forecasting. The investigations, on deep networks [14], on energy-based models [15] have progressed toward becoming establishments for the emerging deep learning as deep architecture generative models in the most recent decade. Three climate estimating models will be investigated in this examination which are in particular: (I) Recurrence Neural Network (RNN), (ii) Conditional Restricted Boltzmann Machine (CRBM), and (iii) Convolutional Network (CN) [8]. Every one of these models will be prepared and tried utilizing the predetermined weather dataset. Parameter learning algorithm for each model, for instance: gradient descent for CRBM and CN, is executed to gain testing error below the predetermined threshold value and compared with the prominent time series forecasting models for example, Recurrent NN.

2.3 ANN and Decision Tree

Artificial Neural Networks (ANN) and Decision Trees (DT) were utilized to analyze meteorological data, accumulated with the end goal to develop classification rules for the Application of Data Mining Techniques in Weather Prediction. There are three fundamental components of a neuron model, which are, (i) an arrangement of synapses, interfacing links, every one of which is considered by a weight/strength of its own (ii) an adder, for summing the info signals, weighted by particular neuron's neural connections (iii) an activation function, for restricting the amplitude of neuron's yield [6]. The MLP network is prepared through the back-propagation learning algorithm. The Prediction is performed through Decision tree.

Table 1: Attribute of Meteorological Dataset

No.	Attributes	Class
1.	Outlook	Sunny, Overcast, Rainy
2.	Temperature	High, Mild, Cool
3.	Humidity	High, Normal
4.	Windy	True, False

2.4 Naive Bayes algorithm utilizing Hadoop

The project aims to forecast the chances of rainfall by utilizing predictive analysis in Hadoop. Predictive analysis models catch connections among numerous elements in a data set to evaluate chance with a specific arrangement of conditions to allocate a score or a weight. Here, Apache Hadoop Framework and Map Reduce Framework are utilized to decrease the data and Naïve Bayes Algorithm is utilized in classification and prediction [4]. Naïve Bayes Algorithm is classification technique based on Bayes Theorem. Naïve Bayes is anything but difficult to assemble and especially valuable for expansive datasets. It is exceptionally utilized in different looks into which contains substantial datasets, for example, Disease prediction [1]. Hadoop is open source programming and it is accustomed to storing large data set in a distributed computing environment, Hadoop makes it conceivable to run applications on system with several hardware nodes. The Hadoop Distributed File System (HDFS) is like the Google File System (GFS) and it utilizes large cluster of data and it gives appropriated distributed file system, fault- tolerant way.

3. METHODOLOGY

In this paper, the system predicts the future weather conditions based on current weather data. The data mining techniques namely Chi square test and Naïve Base statisticsare applied on the dataset to extract the useful information from the dataset. The System Methodology shows in fig. 1:

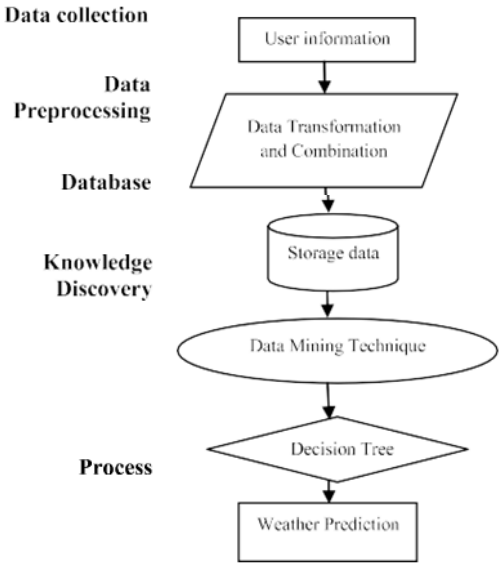

Fig.1: Methodology of Weather Forecast System

3.1 Data Collection and Preprocessing

The initial stage in data mining process is data collection and preprocessing. The crucial stage is data preprocessing, because only valid data will yield accurate output. The data is used in this project collected from users. Though the data set contained many attributes, data preprocessing step considered only the relevant information, ignoring the rest. Then data transformation performed, into a format, which is suitable for Data mining. Four Attributes are used to identify the Weather Forecasting .They are Shown in a table below:

It is to find out the **Class Level** of Weather Forecast where, **Class Levels** are **Good or Bad**.

3.2 Database

The Transformed dataset is store in database that is collected from user. So, there is no prviously stored store is in use. After real time data collection, Data mining techniques applied to predict weather condition.

3.3 Data Mining Technique

In this work, data classification is performed using two data mining technique: Chi square test and Naive Bays Statistics. The data which have to be classified is called training dataset, is fixed. By using this data with testing data, Weather Forecast will be possible. The algorithm of chi square and naïve bays finds relationships between the values of the predictors and the values of the target. The model learns from the training set and that knowledge is used as test data to predict in the scoringse

3.3.1 Chi Square Algorithm

Chi Square Algorithm is a predictive technique used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories. The Equation is as follows:

$\chi_c^2 = \Sigma(O_i - E_i)^2 / E_i$

where, The subscript "c" is the degrees of freedom, "O" is observed value and "E" is expected value. A chi square (X^2) statistic is used to investigate whether distributions of categorical variables differ from one another. In our project we use chi square statistic to determine the best attribute of weather forecast.

3.3.2 Naïve Bayes Algorithm

Naïve Bayes Algorithm is a classification technique based on Bayes Theorem. Naïve Bayes is easy to build and very much useful for large datasets. By using the Naïve Bayes equation we can find the future probability [12].The Equation is as follows:

$$P(c|x) = \frac{P(x|c) * P(c)}{P(x)}$$

Where, (c|x) is future probability of class(c, target), P(c) is the prior probability of the class, P(x|c) is the likelihood which is the probability of predictor given class, P(x) is the prior probability of predictor.

The condition of predicting weather of our project is as follows:

Class:

C1: Weather Forecasting = 'Good',

C2: Weather Forecasting = 'Bad'.

To find the class, C_i that maximizes P (X| Ci)*P (Ci)compute:

P(Weather = Good | x) ∝ P(Weather=Good) · [P(O = s | Weather = Good) · P(T = c | Weather = Good) · P(H = h | Weather = Good) · P(W = t | Weather = Good)]

P(Weather =Bad | x) ∝ P(Weather=Bad) · [P(O = s | Weather = Bad) · P(T = c | Weather = Bad) · P(H = h | Weather=Bad) · P(W = t | Weather =Bad]

IF P(Weather= Good| X) < P(Weather = Bad | X), so classify X as Weather= Bad

Otherwise, Classify X as Weather=Good.

Thus using the above probability prediction of the future chances of weather good or bad will be easy.

3.3.3 Decision Tree

The decision Tree generated from training data is helpful in making prediction. Construction of the decision tree is done by selecting the best possible attribute that will be able to split set of samples in most effective manner. The decision tree for this proposed system is figured below in fig. 2:

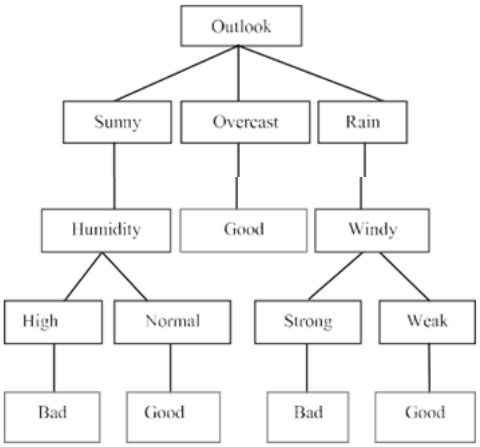

Fig. 2: Decision Tree

4. DESIGN AND ANALYSIS

This topic incorporates the methodology within the design of the system. This system analyzes and measures weather data. The architecture is given in fig.3, clarifies the working model of the project. The Architecture characterizes the behavior, structure and perspectives of our system.

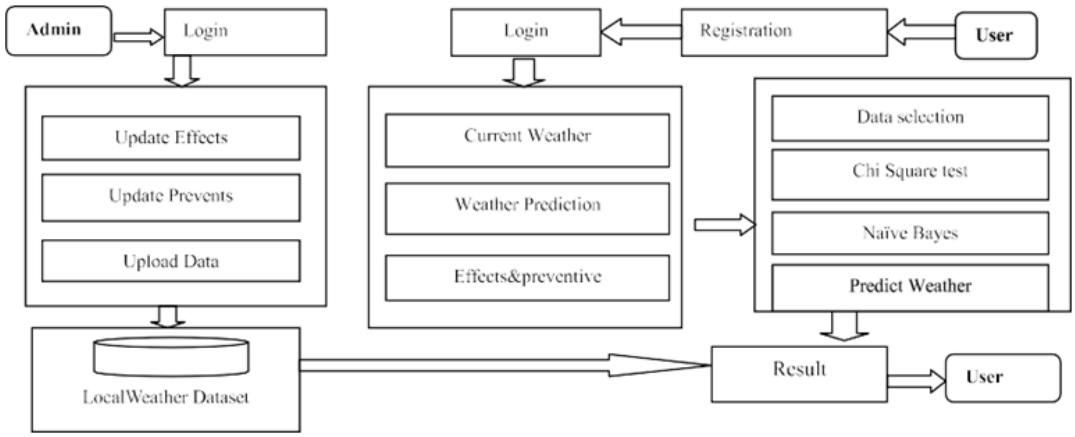

Fig. 3: System Architecture

Here, the job admin is to upload the data, for example, impact of the Outlook, Temperature, and Humidity, Windy and preventive measure and upload the dataset to system. Then again at a customer side client need to register to the application. After the login client gets the present weather condition. To predict the weather conditions in the proposed system, the data mining algorithms has been utilized.

In order to predict the next weather condition or upcoming weather condition the system required to take input of the weather conditions, based on the client input generate the next possible outcome of weather condition. To partition the information and to locate the weather condition Chi square test and Naive bayes are used here. After that final prediction of weather condition (Good or bad) will be performed.

5. EXPERIMENTAL RESULT AND ANALYSIS

The analysis and prediction of weather forecast are implemented using Java Language and using by tool of Eclipse and all data are stored by MySql server.

Chi square test summarizes the difference between our data and our independence hypothesis.

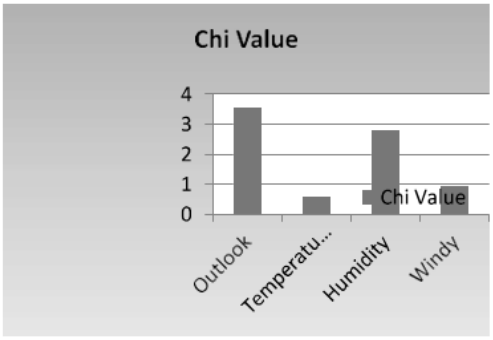

Fig4: Testing result by Chi square test

Chi square test is used to assess the observed value are significantly different from the expected value based on input attributes of the Training set .The model train the probability that a chi-square statistic having 2 degrees of more or less than significant level. Figure 4 depicts that the value of all attribute is more than significant level .Then system classifies dataset using Naïve bayes procedure. Naïve bayes calculating

method produces higher accurate outcomes when contrasted with traditional weather forecasting model.

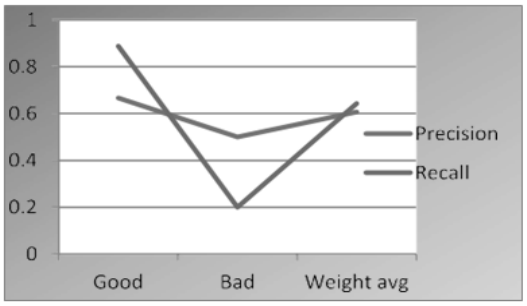

Fig. 6: Accuracy by class of training dataset.

6. CONCLUSION

This paper works with mix of Naïve Bayes and Chi Square algorithm to predict weather condition. The constant information i.e. time-series data is assembled and analysis is performed on this dataset utilizing an interface named Weather Prediction System, developed utilizing Java using Eclipse tools. This framework arranges the given information into various classifications and furthermore predicts the risk of the weather prediction of obscure example is given as an input. The system can be filled in as training tool for Meteorology Students. This methodology can decide the non-

an estimate for the class probability from the training set. Figure 5 depicts the weather forecast class bad is greater than the weather forecast class good.

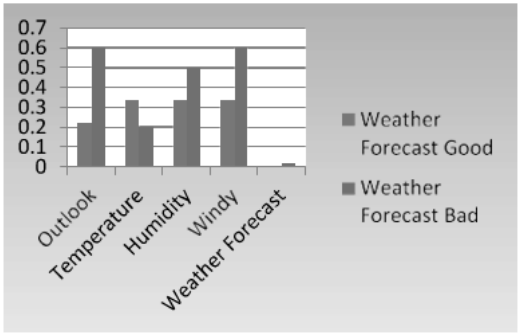

Fig 5: Testing result by Naïve Bayes Model.

5.1 Accuracy

In this Project, Accuracy is the general rightness of the model and is determined as the whole of correct classifications separated by the aggregate number of classifications.

Accuracy = (TP + TN)/ (TP + TN + FP + FN).

Precision is a proportion of the accuracy gave that an explicit class has been predicted. It is characterized as

Precision = TP/ (TP + FP)

Where, TP and FP are the numbers of true positive and false positive predictions for the considered class.

Recall is a proportion of the capacity of a prediction model to choose examples of a specific class from a data set. It is likewise called sensitivity, and compares to the True positive rate.

Recall = Sensitivity = TP / (TP + FN)

Where, TP and FN are the numbers of true positive and false negative predictions for the considered class.

The System is computed and demonstrated utilizing the figure 6. In this figure the X axis contains the methods implemented and the Y axis demonstrates the percentage accuracy of the system. As indicated by the acquired execution the proposed

[6] Wang, ZhanJie & Mujib, A B M. (2017). "The Weather Forecast Using Data Mining Research Based on Cloud Computing". Journal of Physics, pp 1-6

[7] Sushmitha Kothapalli, S. G. Totad, "A Real-Time Weather Forecasting and Analysis", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017), pp 1567-1570

[8] A. G. Salman, B. Kanigoro and Y. Heryadi, "Weather forecasting using deep learning techniques," *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2015, pp. 281-285.

[9] Introduction to Data Mining and Knowledge Discovery, Third Edition, Two Crowds Corporation, http://www.twocrows.com/introdm.pdf, accessed on 12 April 2009.

[10] A. Gautam and P. Bedi, "MR-VSM: Map Reduce based vector Space Model for user profiling-an empirical study on News data," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, 2015, pp. 355-360.

[11] Chen, S.-M., and J.-R. Hwang. "Temperature prediction using fuzzy time series." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 30.2 (2000): 263-275.

linear relationship that exists between the historical data (temperature, wind speed, humidity, and so forth.,) provided to the system during the training phase and on that premise, make a prediction of what the weather would be in future. The Future work of this project is to incorporate more attribute of weather condition to predict and to work with other classification algorithm to become more accurate in prediction.

7. REFERENCES

[1] Dhanashree S. Medhekar, Mayur P. Bote, Shruti D. Deshmukh, "Heart Disease Prediction System using Naive Bayes", INTERNATIONAL JOURNAL OF ENHANCED RESEARCH IN SCIENCE TECHNOLOGY & ENGINEERING, VOL. 2, ISSUE 3, MARCH.-2013, pp 1-5

[2] Mehrnoosh Torabi, Sattar Hashemi, "A Data Mining Paradigm to Forecast Weather", The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012),IEEE, pp 579-584.

[3] Amruta A. Taksande, P. S. Mohod, "Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm", International Journal of Science and Research (IJSR), Volume 4 Issue 6, June 2015, pp 3048-3051

[4] Mr. Sunil Navadia, Mr. Jobin Thomas, Mr. Pintukumar Yadav, Ms. Shakila Shaikh, "Weather Prediction: A novel approach for measuring and analyzing weather data", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), (I-SMAC 2017), IEEE, pp 414-417

[5] Ghosh *et al.*, "Weather Data Mining using Artificial Neural Network," *2011 IEEE Recent Advances in Intelligent Computational Systems*, Trivandrum, 2011, pp. 192-195.

[12] Maqsood, I., M. R. Khan, and A. Abraham. "An ensemble of neural networks for weather forecasting." Neural Computing & Applications 13.2 (2004): pp 112-122.

[13] Kwong, K. M., Liu, J. N. K., Chan, P. W., and Lee, R. "Using LIDAR Doppler velocity data and chaotic oscillatory-based neural network for the forecast of meso-scale wind field." 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). (pp. 2012-2019).

[14] Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. "Greedy layerwise training of deep networks.," Advances in neural information processing systems, vol. 19 (153) 2007.

[15] Ranzato , M., Y., Boureau, 1.., Chopra, S., &LeCun, Y. "A unified energy-based framework for unsupervised learning." In Proc.Conference on AI and Statistics (AI-Stats), vol. 20, 2007.

[16] Linkon Chowdhury , Md.Sarwar Kamal & Sonia Farhana Nimmy, "Artificial System to Compare Energy Status in the Context of Europe Middle East", Global Journal of Computer Science and Technology Volume 12 Issue 8 Version 1.0 ,April 2012, pp25-30

[17] Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.

[18] A. Gautam and P. Bedi, "MR-VSM: Map Reduce based vector SpaceModel for user profiling-an empirical study on News data," *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, 2015, pp. 355-360.