



Review of ML based Techniques for Weather Forecasting System using Data Mining

Ms. Lata kumawat¹ and Kalyani Tiwari²
M.Tech Scholar, CSE, Sage University, Indore¹
Assistant Professor, CSE, Sage University, Indore²
kumawat.lata029@gmail.com¹, ktiwari.official@gmail.com²

Abstract: *One of the most challenging task in real time environment weather prediction. Understanding the various factor that influence weather is as much as important knowing the specific methodology on how to perform the weather forecasting prediction .The weather forecasting is the use of science and technology to predict the condition of the weather for a given area. It is one of the most difficult issues the world. The main way we can forecast the weather include looking at current weather conditions tracing the motion of air and cloud in the sky finding previous weather pattern that resemble current ones ,examining changes in air pressure. This project aims to estimate the weather by utilizing predictive analysis. it is identify to the relationship utmost necessary among different attributes used in the dataset. These datasets can be obtained from various accessible databases and information repositories. We needed for the reason analysis for the data mining procedure before apply. This paper introduces a classifier approach for prediction of weather condition. This paper are shows how Naive Bayes and Chi square algorithm can be utilized for classification purpose. This system is a web application with effective graphical User Interface. The use will login to the system through his ID and password. User will enter some information such as current outlook, temperature, humidity and wind condition. This system will take this parameter and predict weather after analyzing the input information with the information in database. Will be performed two basic functions to be specific classification (training) And prediction (testing). The outcomes demonstrated that these data mining procedures can be sufficient for weather forecasting.*

General Terms: Prediction, Data mining, Classification

Keywords: Chi square, Naïve Bayes, Prediction, Weather Forecasting.

1. Introduction

Enable users to analyze data from a wide range of dimensions or angles, classify it, and condense the connections recognized of weather. Some fundamental terms related to Data Mining are:- Classification, Learning and Prediction[5]. Classification is data mining (machine learning) process used to predict Aggregate participation for information cases. For example, you can use the classification to predict that the weather will be like on a particular day. It will be “sunny”, “rainy” or “cloudy” [3].

Two different way are used for training and mapping:- Supervised and Unsupervised learning. Supervised: A supervised learning algorithm analyzes the training data and produces a derived capacity utilizing Classifier [18]. Unsupervised: In machine learning, unsupervised learning alludes to the issue of trying to hidden structure in unlabeled information [3]. 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 eventually. Finding patterns and data may prompt sensible predictions [18]. Weather forecasting is a complex and challenging science that depends on the efficient interplay of weather observation, data analysis by meteorologists and computers, and rapid communication systems. Meteorologists have acquired very respectful and appropriate skill for small scale weather forecasting. Which use Naive Bayes and Chi Square strategy for weather forecasting, which would show the classifier approach. In this system, the weather conditions are classified into the following characteristics:-

Outlook, Temperature, Humidity, and Wind. Using the characteristics, the system will predict the class level as the forecast weather (the weather will be good or bad). two basic functions will be done in the system classification (training) and prediction (testing), The contents of this paper are illustrated as takes: section first is introduction, 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 and Literature Survey

Weather forecasting has been one of the most challenging difficult around the world because of both its practical value in popular scope for scientific study, meteorology Weather is a continuous, dynamic multidimensional chaotic process, and data intensive and this properties make weather forecasting as emulating challenge. The chaotic features associated with atmospheric phenomena also attracted the attention of scientist. It is a complex phenomenon, due to which the assessment of climate, change, nature so that the nature and causes remain even day. Weather forecast is used for weather forecast like natural disaster ,earthquake, sudden, climate etc. inside weather forecasting we capture the dynamic behavior of weather. Such as minimum temperature or maximum temperature. In this we describe time series of parameters and prepare statistical models for weather forecasting. It is one of the most imperious and demanding operational responsibilities that must be carried out by many meteorological services all over the globes. The real time series data has several model which predict the feature weather data 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 [9][12][13][14]. Different Methods has been utilized in Weather Prediction System, for example, neural network-

based algorithm utilizing BPN (Back Propagation Neural Network) and Hopfield Network [4], RNN (Recurrence Neural Network), CRBM (Conditional Restricted Boltzmann Machine), and CN (Convolutional Network models) [9], Artificial Neural Network and Decision tree Algorithms [5], predictive analysis in Apache Hadoop Framework utilizing Naive Bayes Algorithm [2]. using these methods practically accurate weather forecast can be made .weather forecasting is widely used in business travel, behavioral, science, social and climate reengineering and many other field.

2.1 BPN and Hopfield Network

A neural network is a group of connected I/O units where each connection has a weight associated with its computer programs .it helps you to build predictive models from large database. In this work, Back Propagation Neural Network is utilized for initial modeling. Back propagation is a short for backward propagation of errors. The outcomes acquired by BPN model are sustained to a Hopfield Network. In BPN, the info and yield layer comprises of 3 neurons whereas the hidden layer has 5 neurons and Hopfield Network display work with the assistance of training data seta [4]. A hope field network is just one layer of neurons relating to the size of the input and output. Hope field network also provide model for understanding human memory. 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

RNN a linear SVC and a five layer neural network is used to predict the weather using time series. The goal of this work is to investigate the capability of profound learning technique for weather forecasting. The investigations, on deep networks and the result of this model are analyzed and compare on the basis of root mean between the predicted and actual value [17], on energy-based models [16] 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) [9]. 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

Accurate weather forecasting plays a very important role in the industrial sector in today world and we used for forecasting natural disasters as well. The parameters used in this analyze, wind speed, pressure and point visibility and humidity.[7] Collect Meteorological data is analyzed to develop rules and end goals for classification of data mining techniques in artificial neural neural network and decision trees. Creating a model that makes prediction based on input parameters based on input parameters in which prediction is made through judgment tree secrecy and high level interpretation capability. however each node in decision tree is a neural network making low level decision . three fundamental components of a neuron model, like (i) synopsis is a system that interfaces each link(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 [5]. The MLP network is prepared through the back-propagation learning algorithm. It is a tree that is created by dividing the source dataset into subset. The Prediction is performed through Decision tree.

2.4 Naive Bayes algorithm utilizing Hadoop

Machine learning algorithms have the advantage of making use of the powerful hadoop distributed computing platform and map reduce programming model to process data in parallel. 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 Naive Bayes Algorithm is utilized in classification and prediction [2]. Naive Bayes Algorithms[4]

Classification technique based on Bayes Theorem. Naive bayes classifier is one of the supervised learning algorithm that can be programmed in from of map reduce. Naive 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 [3]. 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. [4]The HDFS (Hadoop Distributed File System) is like the GFS (Google File System) and it utilizes large cluster of data and it gives appropriated distributed file system, fault- tolerntway.

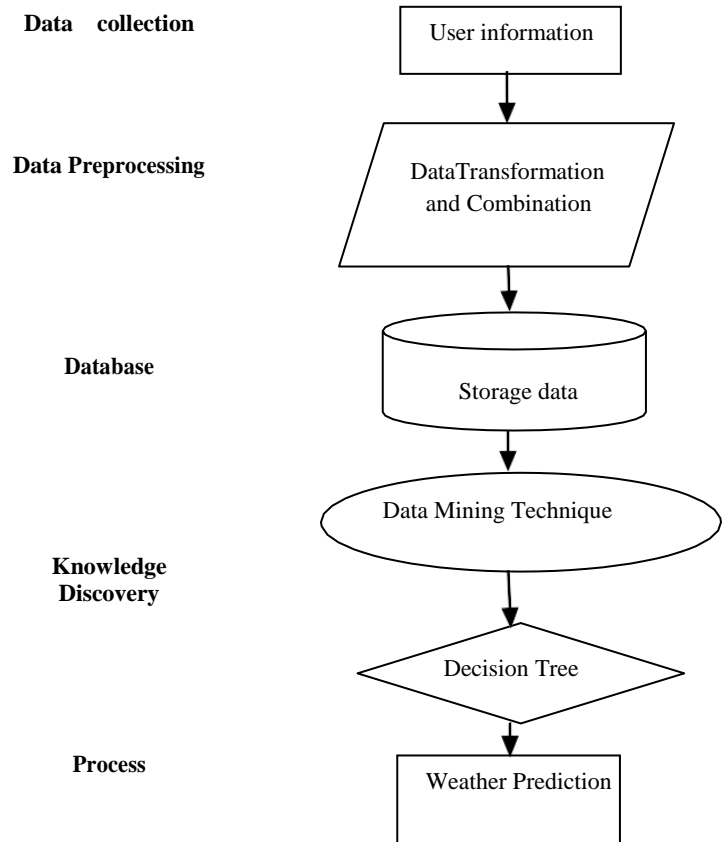


Fig.1

3. Research Object and Methodology

The objective are the research study are:

India is place where the weather is highly variable e.g. Kanyakumari, tamil nadu, etc. developing effective and predictive models for weather analysis, back propagation neural network, hopfield network, recurrence neural network, condition restricted boltman machine convolution network ,artificial neural network, naive bays algorithm, chi square algorithm etc. to develop an efficient reliable and effective weather forecasting system with the help of radius of basic function and will display the weather information in numerical form. The scope of this research is to evaluate the performance of other neural



network models with the chosen data set. In this paper, the system predicts the future weather conditions based on current weather data so which gives as the weather forecast. The data mining techniques namely Chi square test and Naïve Base statistics are applied on the dataset to extract the useful information from the dataset. The System Methodology shown in fig. 1.

3.1 Data collection preprocess

in the process of data mining data collection and preprocessing is done in the initial stage in which preprocessing data would be very important because only valid data will yield accurate output user collected data is used in project through the data is used in this project. Through the data set contained many attributes, but only relevant information was considered in the data preprocessing phase and ignoring the rest of the information. The data was than converted in to a format suitable for data mining to determine weather ,the weather was good or bad.

3.2 Database

The Transformed dataset is store in database that is collected from user. So, there is no previously stored is in use., Data mining techniques applied to forecast real time weather condition.

3.3. Data mining technique

Data mining is used to extract data from a very large data set and we can say that the data is filtered and classified so that we can study the data and filter the data. Data mining or knowledge discovery in database has been popularly recognize as an important research issue with board application. Data classification is performed using teo data mining technique: chi square test and naïve bays statics. The data that is classified for weather forecasting is called training data. Weather forecasting was done using the same with data. The algorithm of chi square and naïve bays find relationship between the values of the target. The model learns from the training set and that knowledge is used to test data to predict in the scoring.

There are four stage of data mining:-

1. Data source
2. Data gathering
3. Model
4. Deploying model

3.4 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:

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.

3.5 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 [13].Thus using the above probability prediction of the future chances of weather good or bad will be easy. various ststistical methods are used to process operational numerical weather prediction .due to the characteristics of attributes of continuous value data discretization are done during tha data preprocessing then the naïve bayesian are used to weather forecast

3.6 Decision Tree

A decision Tree generated from training data that would help in predicting. 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. decision tree learning is a method commonly used in data mining .the goal is to create model that predict the value of a target parameter based on several input parameter. A tree can be made to learn by splitting the source data set into subset based on an attributes value test [6].The decision tree for this proposed system is figured below in fig 2.

Table 1

No.	Attributes	Class
1.	Outlook	Sunny, overcast, rainy
2.	Temperature	High ,mild ,cool
3.	Humidity	High ,normal
4.	Windy	Strong, week

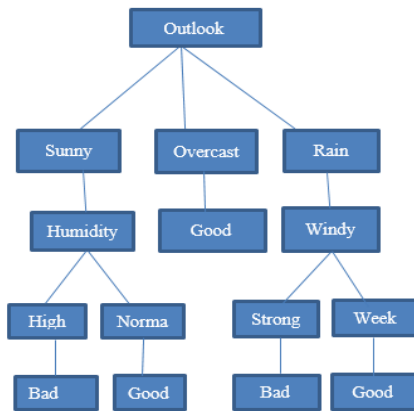


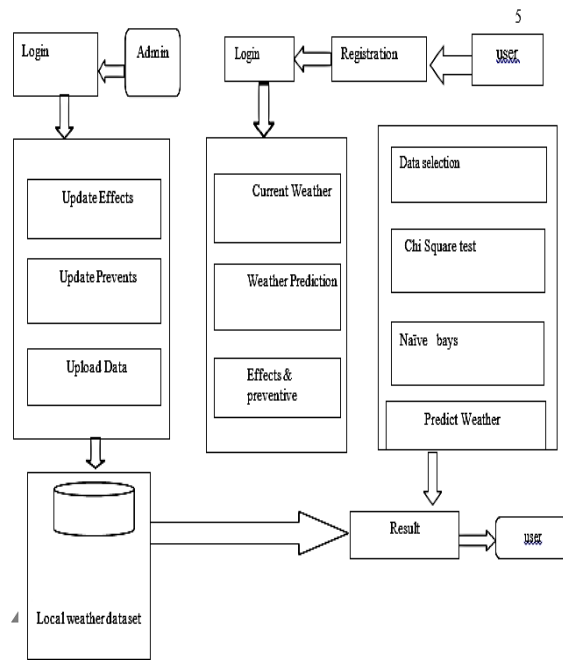
Fig.2 Decision Tree

3.7. Design and Analysis

Incorporates methodology into the design of the subject system. This system analyzes and measures weather data. The architecture is given in fig.3, clarifies the working model of the project and which explains the behavior and outlook of our system. In this we include preparing and disseminating information about weather conditions based on the collection and analysis of observations in relation to weather forecasting and forecasting, displaying information of weather related events in numerical form. The preparation of equations controlling the speed of weather and changes in atmospheric conditions by numerical solution of prediction is called long range or extended range climate forecasting. A data mining algorithm is used to estimate weather conditions. In this the administrator has to upload the data. Such as uploading the outlook temperature and humidity wind and the effect of preventive measure dataset in which client will login and the weather condition of the place will be received by entering the name of the place where the weather is to be known. The system generates outputs based on the client's input to predict upcoming weather conditions. Chi square test and naive bays are used to segment the information and find out the weather conditions. Then the weather is predicted to be good or bad.

4. Future Work and Conclusions

Precisely forecasting the weather will have a profound impact on our daily lives as well as food security and disaster management. In this we will use new technology like AI so that we will be able to predict the weather accurately. We will prepare such a model in which experts will be able to give information about natural disaster. We will be able to get weather information a few hours or one day in advance by the model. Scientists are surprised that unpredictable global and weather patterns are rapidly emerging. Therefore through this model, we will also be able to get the current weather which will help the farmers daily. Through this model we will be able to do business, travel and many other things in the right way. For example, if we have to go for a walk or do business then through this model we will get the information of the weather 2 or 4 days before that the weather is good or bad, based on that we will travel or do our work. Weather forecasting works with a mixture of chi square test and naive bays algorithm using an interface called weather prediction on the dataset, in which continuous information is analyzed. ANN technique is very good at mapping inputs and outputs. Through this method the historical data provided in the system determines the nonlinear relationship between temperature, wind speed, humidity etc. on the basis of this the weather is predicted as to what the weather will be like. Work with other classification algorithms to make weather forecasting more accurate. It deals with the combined use of different climates. This paper presents a direct overview of weather forecasting techniques along with time series data. The papers have been reviewed to emphasize the time scale of forecasting methods on the diversity of forecasting methods which is very useful as it gives a much better understanding and important role.



References

[1] 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, pp579-584.

[2] 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,AnalyticsandCloud),(I-SMAC2017),IEEE, pp 414-417

[3] 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, ISSUE3, MARCH.-2013, pp 1-5

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

[5] Wang, ZhanJie & Mujib, A B M. (2017). “The Weather Forecast Using Data Mining Research Based on Cloud Computing”. *Journal of Physics*, pp1-6

[6] pages downloaded from http://en.wikipedia.org/wiki/Decision_tree_learning

[7] 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, pp3048-3051

[8] 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

[9] 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.

[10] *Introduction to Data Mining and Knowledge Discovery*, Third Edition, Two Crowds Corporation, <http://www.twocrows.com/introdm.pdf>, accessed on 12 April2009.

[11] 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.

[12] 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.

[13] 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.

[14] 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).

[15] 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.

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

[17] 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

[18] Han, J., Kamber, M.: “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers,2006.

[19] 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.