TWF-Weather Forecasting

Manish Soni
Computer Science & Engineering
Raj Kumar Goel Institute of Technology

Pranjal Srivastava
Computer Science & Engineering
Raj Kumar Goel Institute of Technology

Ram Sharma
Computer Science & Engineering Assistant Professor
Raj Kumar Goel Institute of Technology

***Abstract*— In the past few decades, weather forecasting has grown in importance as a subject of study. The majority of the time, the researcher tried to create a linear relationship between the target data and the related input meteorological data. However, when nonlinearity in the makeup of meteorological data was discovered, attention switched to nonlinear forecasting of the weather data. Despite the fact that there is a wealth of literature on nonlinear statistics for weather forecasting, most of it calls for the nonlinear model to be defined before any estimation is carried out. However, because meteorological data is nonlinear and has a very erratic trend, Artificial Neural Network (ANN) has proven to be a more effective method for highlighting the structural relationship between the various components.**

Keywords—Linear Regression, regression, style, styling, insert

I INTRODUCTION

The use of science and technology to forecast the atmosphere's condition at a specific location is known as weather forecasting. Weather predictions are created by gathering quantitative information about the atmosphere's current condition and utilizing scientific knowledge of atmospheric processes to predict how the atmosphere will change. Forecasts become less accurate as the gap between the current time and the time for which the forecast is being made widens due to the chaotic nature of the atmosphere, the enormous computational power needed to solve the equations that describe the atmosphere, measurement error associated with the initial conditions, and an incomplete understanding of atmospheric processes.

Weather forecasts can be used for a variety of purposes. Weather warnings are crucial forecasts because they serve to safeguard property and human life. Forecasts based on precipitation and temperature are crucial.

Forecasts can be used to arrange activities around these events, as well as to prepare ahead and survive them since outdoor activities are severely restricted by heavy rain, snow, and wind chill. We have suggested a weather forecasting model employing artificial neural networks in order to very effectively anticipate the weather and to help solve all such issues.

The benefit that ANN has over other weather forecasting methods is that it uses a variety of algorithms to reduce error and provide us with a predicted value that is very close to the

real value. Such a network is simulated using more recent data to determine the direction of the weather trend.

Weather forecasts are made by collecting as much data as possible about the current state of the atmosphere (particularly the temperature, humidity, and wind) and using an understanding of atmospheric processes (through meteorology) to determine how the atmosphere evolves in the future. However, the chaotic nature of the atmosphere and incomplete understanding of the processes mean that forecasts become less accurate as the range of the forecast increases.

Traditional observations made at the surface of atmospheric pressure, temperature, wind speed, wind direction, humidity, and precipitation are collected routinely from trained observers, automatic weather stations or buoys

During the data assimilation process, information gained from the observations is used in conjunction with a numerical model's most recent forecast for the time that observations were made to produce the meteorological analysis. Numerical weather prediction models are computer simulations of the atmosphere.

They take the analysis as the starting point and evolve the state of the atmosphere forward in time using understanding of physics and fluid dynamics. The complicated equations which govern how the state of supercomputers to solve them.

Linear regression is a statistical technique that is widely used to predict continuous variables. It is a linear approach to modeling the relationship between a dependent variable and one or more independent variables. In this study, we will use linear regression to predict medical expenses.

II RELATED WORK

We have different weather parametric data. The available data covers four main domains related to weather: Humidity, Temperature, Wind Speed and Rainfall. Then further we will forecast the future trends of weather on the basis of past meteorological data. using Naive Bayes Algorithm are some of the methods used in the Weather Prediction System. Data can be viewed against four different types of scenarios/conditions. The scenarios are as follows:

1. Single place single point in time weather prediction
2. Multiple places single points in time weather prediction
3. Single place multi points in time weather prediction
4. Multi places multi points in time weather prediction

2.1 Weather Data Forecasting

2.1.1 Short Term Forecasting

We have used Weather API for performing short term display of temporary details about weather of entered location. It shows the daily weather for any entered region, depicting the maximum and minimum temperature of day & night, information like humid, wind and weather symbol for pictorial representation. The 2 or 3 days for the sites within the areas of a number of purposes of short-range prediction today is to supply numerous users with data on the anticipated weather over forthcoming million sq. kilometers to take necessary precautions beforehand and thus to cut back the harm of adverse weather.

2.1.2 Long Term Forecasting

For long term forecasting we have used linear regression and decision tree regression. Linear regression is a mathematical technique that is used for finding the straight line that best-fits the values of a function that is linear for the axis. It is plotted on scatter graph as different data points of the database. When 'best fit' line of the relation is searched, it is used for the base for estimating and predicting the future values of the function. We do so by extending it without disturbing the slope of axis. It always uses 2 variables to analyse the data. One variable is taken as independent variable and the other is dependent variable. Independent variable is also taken as explanatory variable and affects the values of dependent variable. To use the linear regression, it is required to look for the relationship between different parameters. In long term forecasting we have predicted the weather trend of next few years on the basis of previous 30 years of data. For weather trends we have used the twenty four values for every year minimum value for each parameter in every month and maximum value for each parameter in every month. Our parameters are the as it as discussed before, for example: humidity, wind, rainfall and temperature. Each parameter has 3 values within 24 hours at 12pm at 3pm and at 12am. For regression we must have an independent and dependent variable and these parameters should have some relationship between them. Relationship among different - different parameters that is used in forecasting is as follows: Humidity depends upon rainfall. Temperature depends on average humidity. Rainfall depends on average temperature. Wind speed depends on average temperature. To use linear regression equation, first step is to determine if there is a relationship between the two variables.

2.2 Hopfield Network and BPN

Back Propagation Neural (BPN) Network is used in this work for initial modeling. The BPN model's results are carried over to a Hopfield network. The information and output layer in BPN has 3 neurons, whereas the hidden layer has 5. Hopfield Network display functions are supported by a training data set [5]. In order to achieve equilibrium, the system must perform temperature, wind speed, or humidity flow. The process will be repeated until it converges, and at each iteration, bias and weight values should be changed.

2.3 RNN, CRBM, and CN models

Investigating the capabilities of the deep learning technique for weather forecasting is the aim of this effort. In the most recent decade, the research on deep networks [14] and energy-based models [15] has advanced towards becoming foundations for the development of deep learning as deep architecture generative models. In this analysis, three climate estimating models—in particular, Recurrence Neural Network (RNN), Conditional Restricted Boltzmann Machine (CRBM), and Convolutional Network (CN)—will be examined [8]. The planned weather information will be used to prepare and test each of these models. Each model's parameter learning procedure, such as gradient descent for CRBM and CN, is run to reduce testing errors below the predetermined threshold value.

2.4 Decision Trees and ANN

To create classification criteria for the Application of Data Mining Techniques in Weather Prediction, meteorological data was gathered and analyzed using Artificial Neural Networks (ANN) and Decision Trees (DT). A neuron model is made up of three basic parts: (i) an arrangement of synapses, and interfacing links, each of which is given a weight or strength of its own; (ii) an adder, for summing the information signals, weighted by a specific neuron's neural connections. (iii) an activation function for limiting the neuron's yield amplitude [6]. Through the back-propagation learning process, the MLP network is set up. Decision trees are used to perform the prediction.

2.5 Using the Naive Bayes technique with Hadoop

The project makes use of Hadoop's predictive analysis to forecast the likelihood of rainfall. Predictive analysis models find connections between various components in a data set to assess chance under a particular set of variables to provide a score or a weight. Here, the data is reduced using the Apache Hadoop and Map Reduce frameworks, and the Nave Bayes algorithm is used for classification and prediction [4]. the naive Bayes algorithm

III PROPOSED WORK

In this study, the system makes weather predictions based on live weather information fetched from weather API. The dataset is subjected to the data mining techniques Chi-square test and Naive Base statistics in order to extract the dataset's important information.

This section describes the data provided by the OpenWeather API The parameters for which observations have been provided are temperature, wind speed, rainfall, and humidity. Met Office India observes these parameters at different times on a daily basis. the activity diagram of the web-based system we have developed for weather prediction. The process starts with the user having a choice of selecting a single station or multiple stations. The next step includes the selection of a single date or a range of dates. Now the parameter selection for which the user wants to see the predictions, such as temperature, wind, rainfall, or humidity. He will be shown the prediction based on the input he has given. For storing the previously stored data SQL server has been used. We have performed arithmetic operations on the data of the database for use in specific situations like calculating the average temperature. Predictions are shown using images, animations, graphs, etc. Data will be retrieved from the database and the prediction will be generated according to that data. The user will be allowed to see the details of that location. If the data is not present for situation then the user will be shown that the record is not found. Graphs will be generated for long-term forecasting. In this application, Open Weather Map API is used to show the location of the weather station(s).

Data Gathering and Preparation

The gathering and preparation of data is the first step in the data mining process. Data preparation is an essential step because only accurate data will result in reliable output. In this initiative, the information gathered from users is utilized. The data set has numerous attributes, however, during the data preprocessing step, only the pertinent information was taken into account. The data was then transformed into a format appropriate for data

mining. Weather Forecasting can be distinguished by four characteristics.

Database

The user-collected database contains the transformed dataset. Therefore, no previously stored store is currently in use. Following the acquisition of real-time data, data mining techniques are used to forecast the weather. 3.3.1 Data Mining Method Chi square test and Naive Bays Statistics are two data mining techniques used in this paper to classify the data. Training dataset refers to the fixed data that must be classified. Forecasting the weather will be achievable by combining this data with testing data. Relationships between the values of the predictors and the values of the target are discovered using the chi square and naive bays algorithms. The model gains information from the training set, and then applies that understanding to test data to make predictions in the scoring.

Design And Examination

The technique is incorporated into the system design in this subject. This system measures and analyses weather information. The architecture, which is shown in Fig. 3, makes the project's functioning model clear. The Architecture describes how our system behaves, is structured, and has viewpoints.

IV Conclusion

In order to develop a trustworthy and precise weather prediction and forecasting model, this paper proposes and encourages the usage of the linear regression concept. This concept is a part of machine learning. The variables of temperature, humidity, and pressure can be combined in this extremely effective weather prediction model to generate precise weather forecasts. Additionally, this paradigm facilitates daily decision-making. Even better outcomes can be obtained when it is applied to larger, more pristine datasets. Because unprocessed data may potentially diminish the model's efficacy, pre-processing the datasets is helpful for making predictions.

Future work on this topic will involve adding more weather-related attributes to the forecasting process and collaborating with other classification algorithms to improve prediction accuracy.

V. RESULT

Some intriguing features have been added to the web-based Weather Report project, such as managing and handling exception errors directly by the system, which will be invisible to the user, to make it bug-free. The customer is given a variety of options so they can choose different weather channels based on their needs and interests. Its pattern recognition algorithm will be able to alert users about inclement weather.

REFERENCES

[1] Ch.Anwar ul Hassan,[1] awaid Iqbal, Saddam Hussain, Hussain AlSalman, Mogeeb A. A. Mosleh,[4]and Syed Sajid Ullah, "A Computational Intelligence Approach for Predicting Medical Insurance Cost" *Hindawi Mathematical Problems in Engineering Volume 2021, Article ID 1162553.*

[2] Sam Goundar, Suneet Prakash, Pranil Sadal, Akashdeep Bhardwaj, "Health Insurance Claim Prediction Using Artificial Neural Networks" *International Journal of System Dynamics Applications Volume 9 • Issue 3 • July-September 2020.*

[3] Rama Devi Burri, Ram Burri, Ramesh Reddy Bojja, Srinivasa Rao Buruga "Insurance Claim Analysis Using Machine Learning Algorithms" *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue- 6S4, April 2019.*

[4] Belisario Panay, Nelson Baloian , José A. Pino , Sergio Peñafiel , Horacio Sanson and Nicolas Bersano "Feature Selection for Health Care Costs Prediction Using Weighted Evidential Regression" *Sensors 2020, 20(16), 4392.*

[5] Viktor von Wyl "Proximity to death and health care expenditure increase revisited: A 15-year panel analysis of elderly persons." *Health Econ Rev* 9, 9 (2019).

[6] K. Bhatia, S. S. Gill, N. Kamboj, M. Kumar and R. K. Bhatia, "Health Insurance Cost Prediction using Machine Learning," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-5, doi: 10.1109/INCET54531.2022.9824201.

[7] Kaushik K, Bhardwaj A, Dwivedi AD, Singh R. Machine Learning-Based Regression Framework to Predict Health Insurance Premiums. Int J Environ Res Public Health. 2022 Jun 28;19(13):7898. doi: 10.3390/ijerph19137898. PMID: 35805557; PMCID: PMC9265373.

[8] E. M. van den Broek-Altenburg and A. J. Atherly, "Using Social Media to Identify Consumers' Sentiments towards Attributes of Health Insurance during Enrollment Season," *Applied Sciences*, vol. 9, no. 10, p. 2035, May 2019, doi: 10.3390/app9102035.

[9] Nidhi Bhardwaj , Rishabh Anand "Health Insurance Amount Prediction" *International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 9 Issue 05, May-2020.*

[10] *Aman Kharwal "Health Insurance Premium Prediction with Machine Learning"issue 26,10,20201,*

[11] *Eline M.van den Broek-Altenburg, Adam J.Atherly "Using Scoial Media to Identify Consumers' Sentiments towards Attributes of Health Insurance during Enrollment Season", Center for Health Services Research , The Larner College of Medicine, University of Vermont, Burlington, VT 05405, USA.*