2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Sep 3-4, 2021

2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 978-1-6654-1703-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICRITO51393.2021.9596117

A Comprehensive Study on Weather Forecasting using Machine Learning

Deepti Mishra
GLBajaj Institute of Technology & Management,
Greater Noida, India
itsdeepti.s@gmail.com

Pratibha Joshi
GLBajaj Institute of Technology & Management,
Greater Noida, India
joshipratibha1916@gmail.com2

***Abstract*- Weather forecasting which is a key player in everyday life is a remarkable advantage of science and technology. Weather prediction can support people as protecting the assets and lives of them. The persons involved in outdoor occupations can be benefited by the weather prediction as they needed to know the weather previously. In the paper, the concept of supervised learning is used, which is one of the learning techniques in machine learning. The paper presents the algorithm, which applies the concept of linear regression and artificial neural network, to predict the variation of temperatures such as high and low values as features in a linear combination. The main features focus on maximum or minimum temperature, mean values of humidity, and atmospheric pressures considering the previous two days. The paper focuses on prediction and analysis of weather forecasting which further entirely applies the concept of machine learning.**

Keywords- Prediction, Weather Forecasting, Linear Regression, Machine Learning.

I. INTRODUCTION

Weather forecasting is an essential element in the lives of everyone. It is one of the key applications of science and technology for estimating weather for a specified place within the period. The process of prediction of weather is tried informally from the last centuries, for example, for centuries people have endeavored to estimate the climate. In 650 BC, the Babylonians expected the climate from cloud designs furthermore as pseudoscience. In concerning 350 preceding Christ, logician outlined climate designs in Meteorological. Later, rationalist incorporated a book on expectation, alluded to as the Book of Signs. Chinese climate forecast mental article broadens at least as such a great amount back as three hundred preceding Christ that was conjointly round a similar time old Indian stargazer created climate forecast ways. In confirmation times, Christ himself watched unraveling and understanding local climate designs, by saying, "When evening comes, you state, 'It is an environmental condition, for the sky is red', and inside the morning, 'Today it'll be stormy, for the sky is red and cloudy.' you perceive the best approach to translate the vibes of the sky, any way you can't decipher the indications of the days."

Old expectation ways normally depended on discovered examples of occasions, conjointly named example acknowledgment. for instance, it would be discovered that if the nightfall was eminently red, the consequent day typically brought barometrical conditions [2]. This aptitude gathered over the ages to supply climate mental article. In any case, not all of those expectations demonstrate solid, and heaps of them have since been found to not ascend to thorough connected science testing.

But now science has provided tools and technologies for accurate outcomes. Machine learning plays a crucial role as a key player in weather prediction. It is gripping the opportunities very tightly while designing weather forecasting models and systems in the market. It can offer both comparative and analysis studies to compute between historical and current data of weather predictions. Weather forecasting models can be enhanced in an account of errors, inaccuracies by utilizing machine learning techniques. Machine learning includes three types of learning i.e., supervised learning, unsupervised learning, and reinforcement learning. Supervised learning techniques apply one of the statistical techniques named as a regression for prediction of weather. Regression provides statistical calculations and relations among different variables by enhancing prediction.

Prediction of weather is prepared by applying computation on aggregated quantitative information concerning the state of the atmosphere at a given place and analyzing climatology to project the amendment in the forecasting.

Comparing and contrasting the way of predictions in weather forecasting, in ancient times it is based on changes in air pressure, current climate, and atmosphere state or inclemency. Currently it is predicted using models feed into the computers based on past predictions. Manual intervention is needed to select the best model for making prediction effectively on new data. The idea includes various machine learning techniques and pattern recognition algorithms. Now a days, machine learning techniques applied with mathematical equations and error correction techniques are the standard ways to predict the weather. Hence, forecasts diminish correct as a result of the excellence between the current time and conjointly the time that the forecast is being created (the vary of the forecast) can increase the use of ensembles and model agreement facilitate slim the error and opt for the foremost likely outcome.

Inaccuracies in weather prediction have different effects on different folks. To some folks, these inaccuracies might mean obtaining wet on the walk to the automobile or having to require off an additional layer of garments, however to others, these inaccuracies will cause unnecessary evacuations, home injury, and presumably death. higher weather prediction would afford better information of once evacuation is required, once

folks must cover their homes, and once and wherever folks are most visiting want facilitate. Inaccurate weather prediction systems should be improved once such a lot is at stake. However, to know what must be improved, it is vital to know however the weather is presently foretold.

The objective of the paper is to apply the concept of prediction in weather forecasting. Prediction is a kind of supervised learning in machine learning techniques, which applies a trained dataset. The rest of the paper includes the sections of the literature of survey, methodology applied, proposed new approach and discussion of experimental results.

II. Literature Review

Weather forecasting has been one of the foremost challenging problems around the world due to its sensibility towards data related to the scope for scientific analysis and prediction. Weather forecasting may be a continuous, dynamic, multi-dimensional chaotic method additionally data-intensive further these characteristics create foretelling an interesting challenge. It's one of the foremost domineering and hard operational responsibilities that has got to be administrated by various earth science services everywhere the world. In the current scenario, there are many statistical and predictive systems for weather forecasting are accessible [1].

Various organizations in the Asian nation and abroad have done demonstrating mistreatment supported statistic knowledge manipulation. the assorted methodologies viz. datum decomposition models, Exponential smoothing models, ARIMA models which are based on autoregression further apply previous time series values of weather. ARIMA models have various kinds such as seasonal ARIMA models, vector ARIMA models with versatile statistic values, ARMAX models one of the types of ARIMA having variables unlike ARIMA, that has been used for prognostication functions. Numerous training models have applied for analysis and pattern identification by gathering the data of rain in different locations of the world. It is an entirely changed statistical way to investigate weather forecasting values by applying different functions. To provide outcomes accurately within a limited time may be a major challenge for the research project [2]. Regression may be an applied math experimental technique and it should be widely employed in various businesses, the behavioral sciences, social and climate rephrasing, and plenty of alternative areas [3].

The weather has destroyed cities, injury has value trillions of greenbacks, and tragic events have killed thousands of individuals, globally. Thankfully, these terrible aspects of weather don't affect most of the people, however, the weather still plays a job within the lives of each person on Earth. attempting to know the weather is not new. many alternative cultures had, and still have, gods and shrines dedicated to the terribly topic. Today, as humanity turns to the longer term and pushes boundaries in science and technology, weather prediction can be far better [3]. Typically, skilled weather prediction is formed mistreatment mathematical simulations. Numerical models are accustomed to simulating what is presumably visiting happen supported the known state of the atmosphere. Forecasters use an associate array of models then, supported the outputs of those simulations, manufacture their predictions.

The persistence methodology of prediction assumes that weather nowadays goes to be the identical as weather tomorrow. the tactic uses generalized data regarding however the weather for a region is meant to be, then supported what the weather was that day, it predicts the following day to be similar [4]. The meteorology methodology is extremely almost like the persistence method. However, this methodology takes a lot of comprehensive observe the past once predicting the longer term. Markov chains are one methodology of prediction. These chains use a recent pattern of behaviour and continuous knowledge to form predictions. the pc would produce a matrix, which might be used for analysis and predictions once there is a stable event. ANNs perform by finding out a particular data set and therefore the outcomes from such a knowledge set, and then, from the finding out, produce an associate algorithmic program to use for extrapolation [5]. These networks perform by finding out than learning from the number of errors they need. The error is employed to tweak the various elements of the associate algorithmic program. Once the algorithmic program goes through many alternative coaching iterations, it develops some way to predict knowledge from a given data set Weather prediction is not a simple task [6]. There is an associate array of various factors, methods, and knowledge that must be understood before correct weather prediction may be achieved.

Supervised learning plays a crucial role in designing a weather prediction model. The Bayesian classifier is one of the methodologies of supervised learning that can be applied for weather predictions. Attributes such as humidity, temperature, and wind are used as main constraints in predicting weather [7].

Decision tree as name suggests is a tree like structure distributing the data into different classes and groups. A newly entered data in the model can be verified and judged regarding its adjustment in any class. It is comprised of nodes, branches, and leaf. Node on the top is considered as root node, other nodes in the tree structured are known as internal nodes and classes are denoted in leaf nodes. The tree is the outcome of calculation of information gain and entropy of the attributes in the dataset which is used for processing. The methods from top node that is root node to class node that is leaf node depicts classification rules [8].

A combination of the number of decision trees for better prediction results. In explicit, trees that are full become horrendously profound will in general be told incredibly sporadic examples: they overfit their instructing sets, for example, have the low inclination, anyway horribly high fluctuation. Irregular backwoods are some methods for averaging various profound call trees, prepared on totally various parts of the indistinguishable training set, to diminish the change.

This comes to the detriment of a minor low increment inside the inclination and a couple of loss of interpretability, anyway, normally enormously supports the presentation inside the last model.

Artificial Neural Network is also emerging as a significant technique for weather prediction models [9] [10]. The neural network technique offers more precise and accurate results in computation of weather forecasting. It is an effective technique for developing reliable model for prediction of weather. The outcome of one layer may the input to connected layer. Due to this mechanism, the neural network technique is raised as a robust model for weather forecasting [11].

III. METHODOLOGY

A. Regression

Regression is a machine learning algorithm that uses supervised learning in which applies already trained data. Regression is a statistical method used for prediction. Regression is one of the techniques of statistics using the concept of the dependent variable and independent variable. Even though the Regression technique includes different types such as linear regression, multiple regression, and logistic regression, but linear regression is applied in the methods technique part in the paper.

B. Linear Regression

It is a linear approach to predict the relationship between dependent variable X and independent variable Y. It is a good approach for prediction calculations. Here, in the Figure1 value of Y can be calculated from the linear combination of the X. It is called linear because here the dependent variable is linear to the independent variable.

$$Y = \alpha 1 + \alpha 2.X \quad (1)$$

Here, $\alpha 1$ and $\alpha 2$ are the variables representing the interception of the line and linear regression coefficient, respectively. This equation is used to make predictions about the data.

In straightforward linear regression process, the relationship between the variables is direct indicator used for parameter modelling which are further selected from the data [12]. Outliers can also be detected applying regression technique to the data. Outliers can impact the data by disturbing the accuracy and preciseness of the results.

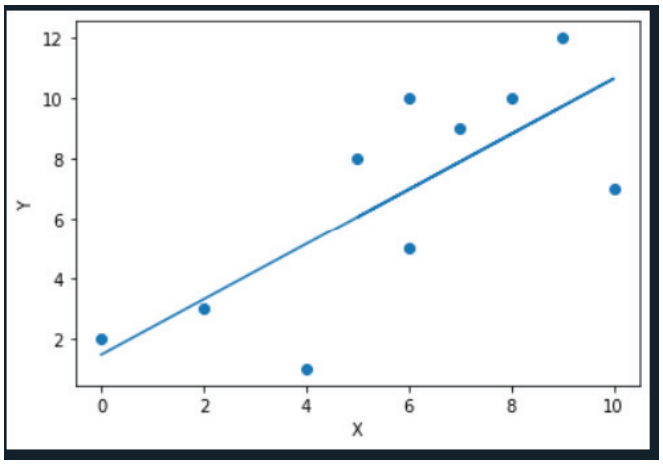

Fig. 1. Graph showing the independent variable X with the dependent variable Y.

As a wide range of multivariate investigation, basic relapse centers around the opportunity dispersion of the reaction given the estimations of the indicators, rather than on the possibility conveyance of those factors, that will be that the space of factual strategy.

Direct relapse was the essential style of multivariate examination to be contemplated carefully and to be utilized broadly in reasonable applications. this is regularly because models that depend directly on their obscure parameters are simpler to suit than models that are non-straightly connected with their parameters and since the connected math properties of the resulting estimators are simpler to work out [13].

To perform our analysis, we need to extract previous data from the internet using site Kaggle.

C. Artificial Neural Network

The concept of artificial neural network is also applied in calculation of results. Artificial neural network is the gathering of nodes for computation alike human brain structure. Each node is connected like synapse in human brain with other nodes in the network for transferring data. There is calculation of input and output at the nodes by some non-linear function with sum of its inputs. Synaptic weights are attached at each node which are further used in calculation of inputs and outputs. All the connections of nodes are bounded and designed in layers. Neural network is the connection of processing nodes which receive inputs and generate outcome. Processing nodes are interconnected to each other in the layered form. All the layers are connected with each other.

IV. IMPLEMENTATION & RESULTS

A. Data retrieval

Here, linear regression is applied to predict weather forecasting. The latest dataset is used in July for the top 10 cities of 10 countries, available on Kaggle1 [14].

The training set contains data starting from 01-07-2020 to 31-07-2020. In contrast the test set is the next day after 31-07-2020 Depending on the weather conditions and attributes we here predicting the weather for the top 10 cities of 10 countries during covid-19 lockdown.

B. Results

It is known that data is available in many forms sometimes unstructured and containing invalid fields, missing values [15][16][17]. To make data usable, it should be cleaned the data so that it can be used. Here the algorithm contains the only integer, so it is to pop the values other than that.

Also, the attributes which are not needed in our prediction are dropped out and place the zero where the data entry is null. So here, it is needed to clean our data before using it.

After the data is cleaned, then that updated data can be applied to train the model. Here, Scikit for the linear regression model is used.

The purpose of the work is to predict the next day various weather conditions using linear regression. Here prediction of the weather is based on parameters such as temperature, Deviation point, humidity, wind speed, pressure, etc.

Here in Figure 2, the precipitation level is in blue, and the day predicted is in red.

Similarly, in Figure 3, blue shows the magnitude of the feature such as temperature, humidity, etc. while red shows the predicted value of the day based on parameter.

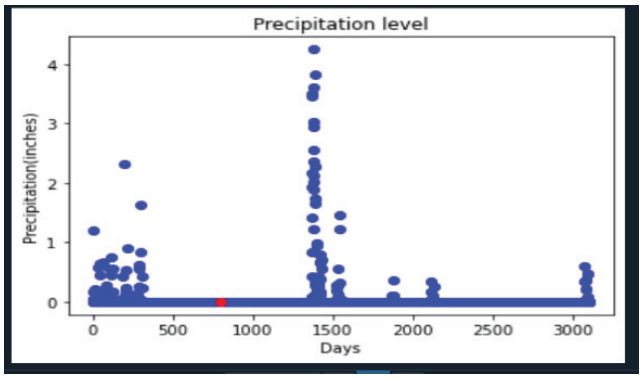

Fig. 2. Graph showing the predicted result of precipitation level

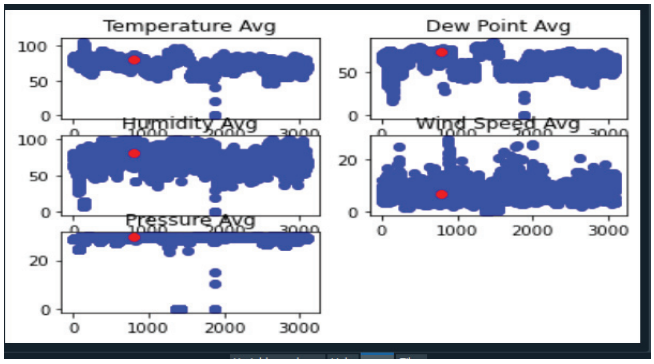

Fig. 3. Graph showing the data with different magnitudes

Here, it has shown that whenever the attributes are changing the prediction value is also changing.

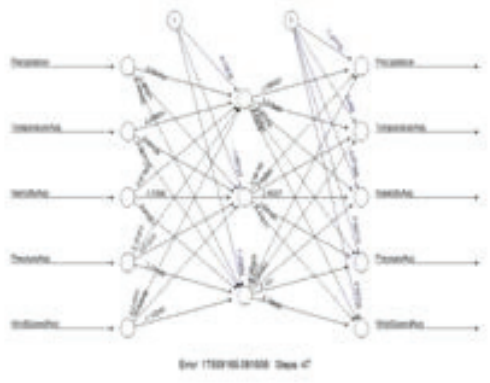

Fig. 4. Showing results for ANN

V. Proposed Approach

The proposed approach follows the machine learning technique that is supervised learning. The approach applies decision tree and probability calculation. Initial step includes the input of data which is 3000 X 5. In next step, the dataset is partitioned in 10 X 5 data, each time executing in a cycle following decision tree algorithm which is supervised learning approach. Probability is compiled on resulted outcome. Confusion matrix is generated on results for accuracy and specificity.

Input: Weather dataset

Output: Weather Prediction

Steps

1. Matrix 3000X5
2. Partition of data 10X5 matrix.
3. Applying decision tree at every partition cycle
4. Probability calculation on resulted outcome.
5. Calculation of confusion matrix.

VI. Discussions

The results of weather forecasting depend on current state of atmosphere including the parameters temperature, humidity and wind while considering the atmospheric pressure as main constraint. Here, it can be predicted that on the next day humidity, deviation point is high, temperature is average, and the wind speed is comparatively low. So, there is high chance of rainfall. But from Figure 4 also depicts that increase in temperature and wind affects the weather maximum. Both plays crucial role in depicting weather.

Therefore, it can be stated as during COVID-19 period, weather seems to be suitable and good.

VII. Conclusion & Future Work

Weather forecasting has come a long time with the early civilizations and to be continued till now. People have continuously been inquisitive regarding climate and wished to and tried to find out numerous strategies to predict weather earlier. Early predictions were created mostly on personal expertise and assumptions while not based on experiments, testing, and analysis.

Weather prediction has many alternative issues. Even the only weather predictions are not good. The one-day forecast sometimes falls among 2 degrees of the temperature. Though this inaccuracy is not unhealthy, as predictions are created for more in time. Also, once predicting precipitation, accuracy maybe even worse. what is more, weather prediction, in areas wherever the climate lacks consistency, is off by even a lot of. as an example, in a very place like geographical area wherever temperatures have nice variance, the temperature predictions are a lot of inaccurate than an area just like the tropics.

Generating predictions on weather, data involves vast knowledge of statistical weather models and scientific information required for climatology coordinating with the observations received from the atmosphere. This study discusses meteorology evolved and developed from the early human civilizations to contemporary weather prediction.

Though previously it relates to rather out-dated strategies, now the scope is to apply new tools and technologies for weather prediction. Nowadays, Weather prediction became an emerging topic for researchers as it is associated with various

domains and further it requires both science and technology for processing, computations, and designing the forecasting model.

REFERENCES

[1] P. C. Reddy and A. S. Babu: Survey on weather prediction using big data analytics, in Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, (2017)

[2] M. Wiston and M. KM: Weather Forecasting: From the Early Weather Wizards to Modern-day Weather Predictions, Journal of Climatology & Weather Forecasting. (2018), vol 6,pp 1-9

[3] Janani.B and PriyankaSebastian: ANALYSIS ON THE WEATHER FORECASTING AND TECHNIQUES, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), (2014),

[4] Y. Quan and L. Yuchang: Research on weather forecast based on neural networks, IEEE. (2000), DOI: 10.1109/WCICA.2000.863402

[5] M. Wiston and M. KM: Weather Forecasting: From the Early Weather Wizards to Modern-day Weather Predictions, Journal of Climatology & Weather Forecasting, (2018).

[6] B. N. Goswami: The Challenge of Weather Prediction. (1997) pp 1-15

[7] M. Biswas, T. Dhoom, and S. Barua: Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data, International Journal of Computer Applications, vol. 182, no. 4, (2018) 20-24

[8] G. Krishna: An Integrated Approach for Weather Forecasting based on Data Mining and Forecasting Analysis, International Journal of Computer Applications.

[9] K. Abhishek, M.P.Singh, S. Ghosh, and A. Anand: Weather forecasting model using Artificial Neural Network, Procedia Technology, vol. 4, (2012) 311-318.

[10] Y. Quan and L. Yuchang: Research on weather forecast based on neural networks, in Proceedings of the 3rd World Congress on Intelligent Control and Automation (Cat. No.00EX393), Hefei, China, 2000.

[11] A. J. Litta, S. M. Idicula and U. C. Mohanty, "Artificial Neural Network Model in Prediction of Meteorological," International Journal of Atmospheric Sciences, pp. 1-15, 2013.

[12] Z. Wang and A. B. M. M. Mujib: The Weather Forecast Using Data Mining Research Based on Cloud Computing, Journal of Physics Conference Series. (2017),

[13] I. A. J. and W. T. Y.: Weather Forecasting Models, Methods and Applications, International Journal of Engineering Research & Technology (IJERT). (2013), vol 2, pp1945-1956

[14] Kaggels: https://www.kaggle.com/bhargavyagnik99/weather-100-top-10-cities-of-10-countries-july, [Online].

[15] V. K. Didal1, Brijbhooshan1, A. Todawat, and K. Choudhary: Weather Forecasting in India: A Review, Int.J.Curr.Microbiol.App.Sci, (2017) 577-590

[16] N. Krishnaveni , A.Padma: Weather forecast Prediction and analysis using Sprint algorithm:,Journal of Ambient Intelligence and humanized computing, Springer,(2020).

[17] P. Kapoor, S.S.Bedi: Weather Forecasting using Sliding Window, International Scholarly Research Notices,Hindawi, (2013)