Weather Forecasting System using Machine Learning

R.B.Saroo Raj [1], Ankush Rai [2], Ronit Dharmik [3], Siddharth Bhattacharjee [4]
[1] Assistant Professor (Sr.G), Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.
[2, 3, 4] Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

Abstract – Weather prediction is an ongoing testing issue seen by the world in the most recent decade. The prediction is hard because of the adjustment in weather. Different weather prediction models have been outlined utilizing machine learning and Artificial Intelligence over the recent years. The genuine issue of precisely anticipating or forecasting the weather still remains. Numerical weather prediction is taking the current numerical information on weather conditions and applying machine learning algorithms on it to forecast the weather. This paper is the use of machine learning algorithms, direct relapse demonstrate from insights, and two enhancement techniques, Normal condition strategy, and Gradient plunge technique to anticipate the weather based on a couple of parameters. Two enhancement techniques have been utilized to look at the execution of the algorithms. The got results exhibited that the typical condition technique forecasts the weather with high exactness, though the inclination plummets strategy forecasts the weather with next to no accuracy.

Index Terms — Weather prediction, machine learning, efficiency, linear-regression model, normal equation method, gradient descent method, dataset, temperature, humidity, error.

1. INTRODUCTION

Weather forecasting is considered as the most difficult issue seen by the world in the most recent decade. This by implication affected the viable prediction of the weather information.

Machine learning is the capacity of the PC to learn without being expressly modified. It enables machines to discover concealed examples and bits of knowledge. In directed learning, we construct a model in light of named preparing information. The model is then utilized for mapping new precedents. Thus, in view of the watched weather examples from the previous, a model can be constructed and used to anticipate the weather. Numerical Weather Prediction (NWP) is a generally utilized prediction display. Generally, it takes the present weather conditions and procedures it to manufacture a model for anticipating the weather.

A few information mining techniques have been utilized in differentiated applications, for example, foreseeing precipitation, weather, tempests and surge. Weather forecasting falls under prescient mining which centers around the information examination, details the database, and forecasts the highlights of mysterious information.

To have a compelling prediction; it is expected to recognize the connection between's the traits of weather, which by implication have a job in the weather changes. Consequently, in this article, a model is proposed for compelling weather prediction by considering different traits together with their connections together with information mining techniques.

This examination work centers around understanding the weather prediction inconsistencies and in-effectiveness in view of straight relapse algorithms through typical condition show. The real commitment of this exploration work is to detail a proficient weather prediction demonstrate in light of the direct relapse algorithms. Consequently, keeping in mind the end goal to defeat the restrictions of weather prediction, this work utilizes ordinary condition model's speculation and contrasts it and the angle plunge model to give a superior thought of productivity of the models.

2. RELATED WORK

In this segment, an investigation is completed with the current weather prediction techniques accessible in the writing in view of the enhanced ideas of different analysts. Another weather prediction approaches in light of linear regression strategy is exhibited for which different weather characteristics have been utilized. The trial result is computed for various parameters, for each approach.

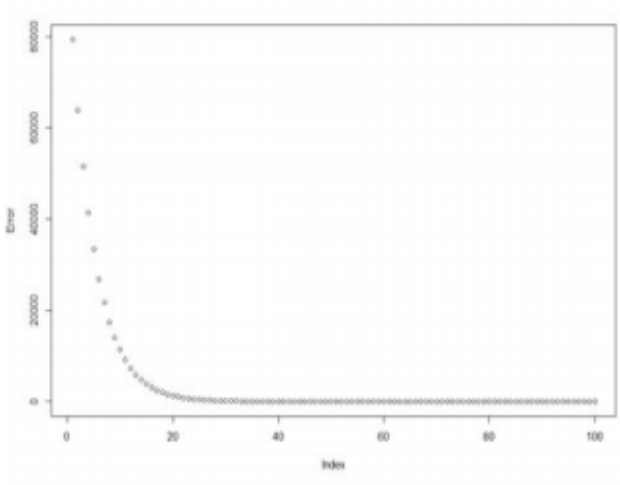

Fig. 1. Example of regression line

Linear regression is the most fundamental and every now and again utilized the prescient model for investigation. Regression gauges are for the most part used to portray the information between at least one autonomous and ward factors. Linear regression finds the best-fit through the focuses, graphically. The best-fit line through the focuses is known as the regression line.

Fig. 1 is a case of the best-fit line. Here, the line can be straight or bent relying upon the information. The best-fit line can likewise be a quadratic or polynomial which gives us a better solution to our inquiries. Two of the streamlining algorithms utilized in this examination are Normal Equation Method and Gradient Descent Equation. Autonomous Component Analysis (ICA) decides the free part closeness inside the spatial transient information. Here, neural systems together with nonlinear sanctioned connection evaluation are proposed to determine the relationship between Humidity, Pressure, Temperature, wind speed and other weather parameters.

3. PORPOSED MODELLING

In a creating nation where the real populace is subject to horticulture, weather conditions play a vital and essential job in the monetary development of the general country. Along these lines, weather prediction ought to be more exact and precise.

Fig. 2. System block diagram

Two algorithms are used to compare the efficiency when applied on the original dataset. Theory for our proposed model can be shown by the Equation 1:

AverageTemp = a + b (AverageDew) + c (AverageHum) + d (AveragePress) + e (AverageVisi) + f (AverageWind)

Here, parameters a, b, c, d, e, and f are known as the learning parameters. The enhancement algorithm takes in these parameters with the assistance of the preparation information sustained to it. Three kinds of weather parameters are anticipated: Temperature, Humidity, and Dew-point. Temperature is the proportion of hotness or chilliness, by and large, estimated utilizing thermometer. Units of temperature most much of the time utilized are Celsius and Fahrenheit. Mugginess is the amount of water vapor present in the environment. It is a relative amount. Dew point is the temperature of the air (which fluctuates as indicated by weight and mugginess) underneath which water beads start to consolidate and dew is framed. Table I demonstrates test information from the preparation set, a tremendous dataset is bolstered to the model. Parameters are prepared and they relate to the model. At last, it very well may be said that the machine has been prepared.

TABLE I
SAMPLE TRAINING DATA

Date	Avg Temp	Avg Dew	Avg Hum	Avg Pres	Avg Visi	Avg Wind
1/1/2014	22	21	87	1016	3	0
2/1/2014	21.5	21	88	1016	3	0
3/1/2014	20.5	19	83	1017	3	3
4/1/2014	17.5	17	84	1017	3	10
5/1/2014	20.5	19	84	1015	3	6
6/1/2014	18	17	84	1014	3	0
7/1/2014	20.5	19	88	1014	3	0
8/1/2014	21	19	84	1016	3	0
9/1/2014	20	19	86	1016	3	0
1/1/2014	22	21	87	1016	3	0

Finally, it can be said that the machine has been trained.The linear regression hypothesis is given in Equation 2:

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

Least-square equation is used in the algorithm of linear-regression method. The least-squares cost equation appears in Equation 3:

$$J(\theta_{0\ldots n}) = \frac{1}{2n} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

The advancement algorithms utilized in the examination are typical condition and inclination plunge. Typical condition is a systematic answer for the linear regression issue with a minimum square cost work. The vector type of typical condition shown in Equation 4:

$$\theta = (X^T X)^{-1} X^T y$$

Gradient Descent formula is given in Equation 5:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Python is used to write the code.

4. RESULTS AND DISCUSSIONS

Conspicuously, three parameters: Temperature, Humidity, and Dew-point are computed through the theory of the linear-

regression show. The mistake is likewise computed for every one of the algorithms. Parametric and add up to mistake is ascertained. The aftereffects of the ordinary condition algorithm indicate almost no deviation from the genuine yield. It tends to be derived from the charts and tables in the accompanying areas that the blunder delivered by inclination good technique is extremely noteworthy contrasted with the mistake created by typical condition strategy. In the accompanying figures, blue shading demonstrates the genuine yield, green shading is the prediction by ordinary condition and red shading demonstrates the prediction by inclination drop technique. The outcomes got for the substance temperature are appeared in Table III and Fig. 4.

TABLE III
ERROR IN TEMPERATURE

Parameters	Normal Equation	Gradient Descent
a	154.3793	0.4349
b	0.900699	1.872743
c	-0.15097	-0.92408
d	-0.13633	0.050531
e	0.089188	0.59752
f	-0.12941	-0.72156
Error	0.599866	41.56697

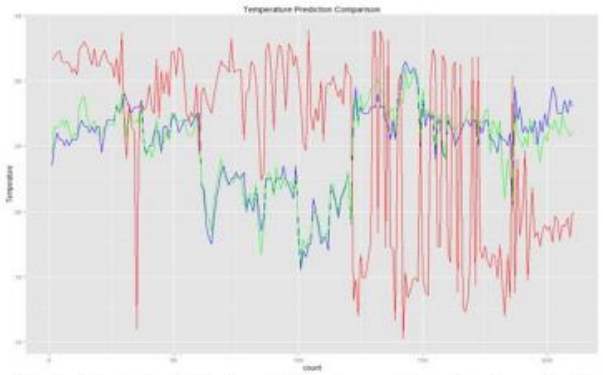

Fig. 4. Actual output (blue), prediction by normal equation (green) and prediction by gradient descent (red) for temperature

TABLE IV
ERROR IN HUMIDITY

Parameters	Normal Equation	Gradient Descent
a	-1022.97	0.180542
b	-2.3754	-0.38342
c	4.065956	-1.14582
d	1.059731	0.094575
e	-0.43401	1.354212
f	-0.63411	0.717851
Error	16.05951	134.9208

It tends to be reasoned from the above diagram that the distinction between blue lines (genuine yield) and green lines (results by typical condition strategy) are extremely insignificant, which connotes the exactness and viability of the ordinary condition technique. The outcomes got for the element dampness are appeared in Table IV and Fig. 5.

Table IV demonstrates the insignificant mistake utilizing ordinary condition and the extensive deviation from the real yield delivered by slope plummet algorithm, demonstrating the wastefulness of the angle drop show. Stickiness assumes an exceptionally huge job in anticipating precipitation and beads in the air. It is likewise the main factor in a significant number of our everyday exercises. It likewise influences the temperature. As the relative mugginess diminishes, the temperature increments, and the other way around.

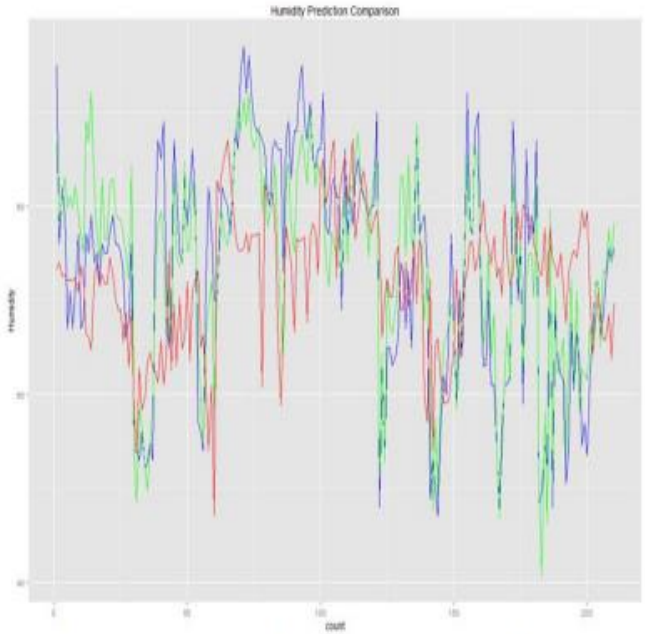

Fig. 5. Actual output, prediction by normal equation and prediction by gradient descent for humidity

It very well may be derived from Fig. 5 that ordinary condition predicts more precise outcomes when contrasted with inclination plunge strategy, even on account of dampness. The outcomes got for the element dew-point are appeared in Table V and Fig. 6.

It is obvious from Table V and Fig. 6 that the angle drop technique is unsatisfactory for foreseeing dew-point, though results created by the typical condition strategy are exceptionally dependable. The diagrams above demonstrates the effectiveness of the ordinary technique. With bigger dataset, better and more exactness can be accomplished.

TABLE V
ERROR IN DEW-POINT

Parameters	Normal Equation	Gradient Descent
a	164.8517	-0.62643
b	0.469616	-1.18322
c	0.134735	0.48247
d	-0.16361	0.026811
e	0.077588	0.369279
f	0.077965	-2.42383
Error	0.740194	156.8855

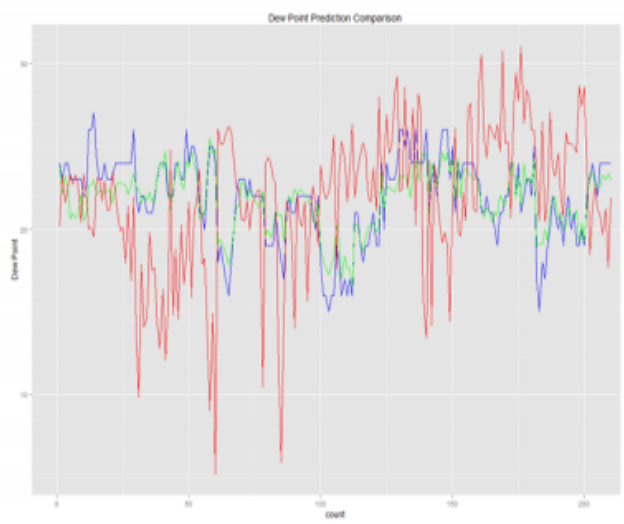

Fig. 6. Actual output, prediction by normal equation and prediction by gradient descent for dew-point

5. CONCLUSION

This exploration recommends and proposes a productive and exact weather prediction and forecasting model utilizing linear regression ideas and typical condition show. Every one of these ideas is a piece of machine learning. The typical condition is an extremely productive weather prediction model and utilizing the substance's temperature, dampness and dew-point, it tends to be utilized to make solid weather predictions. This model likewise encourages basic leadership in everyday life. It can yield better outcomes when connected to cleaner and bigger datasets. Pre-handling of the datasets can be viable in the prediction as natural information can likewise influence the effectiveness of the model.

REFERENCES

[1] Radhika Y, Shashi M. Atmospheric temperature prediction using support vector machines. International Journal of Computer Theory and Engineering. 2009 Apr; 1(1):1793– 8201.
[2] Han J, Kamber M. Data mining: Concepts and techniques. Morgan and Kaufmann; 2000
[3] Smith BA, McClendon RW, Hoogenboom G. Improving air temperature prediction with artificial neural networks. International Journal of Computational Intelligence. 2007; 3(3):179–86.
[4] Hsieh WW, Nonlinear canonical correlation analysis of the tropical pacific climate variability using a neural network approach. Journal of Climate. 2001.
[5] Casas DM, Gonzalez AT, Rodrigue JEA, Pet JV. Using data mining for short-term rainfall forecasting. Notes in Computer Science. 2009; 5518:487–90.
[6] Smith BA, McClendon RW, Hoogenboom G. Improving air temperature prediction with artificial neural networks. International Journal of Computational Intelligence. 2007; 3(3):179–86.
[7] S. Sripada, E. Reiter, I. Davy, SumTime-Mousam: Configurable marine weather forecast generator, Expert (2003) 4–10.
[8] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1992.
[9] P. Gorniak, D. Roy, Grounded semantic composition for visual scenes, J. Artificial Intelligence Res. 21 (2004) 429–470.