Weather Prediction Using Machine Learning

Dr Ka Jayabalaji [1] , Sudhakar. S [2] , Kaviyarasu. B [3] , Ankush Dev. K [4]

[1] *ASSOCIATE PROFESSOR, DEPARTMENT OF SOFTWARE SYSTEMS, SRI KRISHNA ARTS AND SCIENCE COLLEGE, COIMBATORE, INDIA,* jayabalajika@skasc.ac.in

[2] *STUDEN, II Bsc AI Ml, SRI KRISHNA ARTS AND SCIENCE COLLEGE, COIMBATORE, INDIA,* sudhakars22bai052@skasc.ac.in

[3] *STUDENT, II Bsc AI Ml, SRI KRISHNA ARTS AND SCIENCE COLLEGE, COIMBATORE, INDIA,* kaviyarasub22bai022@skasc.ac.in

[4] *STUDENT, II Bsc AI Ml, SRI KRISHNA ARTS AND SCIENCE COLLEGE, COIMBATORE, INDIA,* Ankushdevk22bai003@skasc.ac.in

A B S T R A C T

Customarily, environment evaluation has been performed dependably by regarding the climate as a fluid. The future condition of the climate is recorded by figuring out thermodynamics and the mathematical place of the fluid components. Nevertheless, this traditional arrangement of differential conditions as observed by physical models is at times unstable under oscillating effects and uncertainties when estimating the underlying states of air. This demonstrates an inadequate comprehension of natural varieties, so it limits environment conjectures to 10-day time spans since environment projections are basically inconsistent. However, AI is modestly generous for most barometric weakening impacts contrasted with customary methods. One more ideal place of AI is that it doesn't rely upon the actual laws of natural cycles.

Keywords: inadequate, precipitation, divergent

1. INTRODUCTION

Weather prediction is the task of predicting the atmosphere at a future time and a given area. This has been finished through actual conditions in the good 'ol days in which the environment is viewed as liquid. The present status of the climate is investigated, and the future state is anticipated by settling those conditions mathematically, however we can't decide extremely exact climate for more than10 days and this can be improved with the help of science and technology. Machine learning can be used to process immediate comparisons between historical weather forecasts and observations. With the utilization of AI, weather conditions models can all the more likely record for expectation mistakes, like misjudged precipitation, and produce more exact forecasts. Temperature forecast is critical in an enormous number of utilizations, including environment related examinations, energy, farming, clinical, or and so forth. There are various sorts of AI computations, which are Straight Relapse, Polynomial Relapse, Arbitrary Backwoods Relapse, Counterfeit Brain Organization, and Repetitive Brain Organization. These models are arranged ward on the genuine data gave of any area. Contribution to these models is given, for example, if anticipating temperature, least temperature, mean air weight, greatest temperature, mean dampness, and order for 2 days Considering this Base Temperature and Greatest Temperature of 7 days will be achieved

2. Machine Learning

Machine learning is relatively robust to disturbances and does not require other physical variables for prediction. Therefore, machine learning is a much better choice in the development of weather forecasting. Before the development of technology, weather forecasting was a tough nut to crack. Weather forecasters relied on satellites, data modeling atmospheric conditions with less accuracy. The accuracy and predictability of weather forecasting and analysis has greatly increased over the last 40 years with the use of the Internet of Things. With advances in data science and artificial intelligence, scientists now predict weather with great accuracy and predictability Purpose (briefly)

This project aims to predict temperature using different algorithms such as linear regression, random forest regression and decision tree regression. The initial value should be numerically based on several additional factors such as maximum temperature, minimum temperature, cloud cover, humidity and sunshine hours per day, precipitation, pressure and wind speed.

3. Background

In the current situation, the Indian Observatory makes traditional weather forecasts. There are four general methods of weather forecasting. The first method is the climatological method, which looks at weather statistics collected over several years and calculates averages. Another method is the analog method of finding a day in the past with similar weather to the current forecast. The third method is the Persistence and Trends method, which has no

skills predicts the weather because it is based on past trends. The fourth method is numerical weather forecasting, which makes weather forecasts based on some atmospheric conditions such as temperatures, wind speed, high and low pressure systems, rain, snowfall and other conditions. Thus, these traditional methods have many limitations. It not only predicts the maximum temperature for the current month, but it also predicts without using machine learning algorithms. Therefore my project is to increase the accuracy and predict the weather at least one month into the future by applying machine learning techniques.

4. USE OF ALGORITHMS:

There are different regression methods for temperature forecasting and different functional regressions that use datasets for calculation and research. 80% of the data volume is used in To Train calculations and 20% of the data volume is designated as the test set. For example, if we need to predict the temperature of Kanpur, India using these machine learning calculations, we use 8 years of data to do the calculations and 2 years of data as the test data. Unlike weather forecasting, which essentially uses machine learning algorithms dependent on physics and differentiation iterations, artificial intelligence is additionally used for temperature forecasting: which includes models such as linear regression, decision tree regression, random forest regression. Finally, machine learning has greatly changed the world view of weather estimation with high accuracy and predictability. In addition, greater progress will be made in the next few years in using these advances to accurately predict the climate to avoid typhoons, tornadoes, and thunderstorms.

5. METHODOLOGY

The data set used in this layout is collected from Kaggle which is "Historical weather data for Indian cities" from which we have selected the data for Kanpur city. The dataset was created keeping in mind the need for such historical weather data in the community. Data sets of India's eight largest cities by population. The dataset was accessed using the worldweatheronline.com API and the wwo_hist package. The datasets contain hourly weather data between 01-01-2009 and 01-01-2020. Each city has data for more than 10 years. This data can be used to visualize the change in data due to global warming or to predict the weather for the coming days, weeks, months, seasons, etc.The data was extracted using the world weatheronline.com API and we cannot guarantee the accuracy of the data. The main object of this dataset can be used to predict the weather for the next day or week using the huge amounts of data in the dataset. Additionally, this information can also be used for visualization to help understand the impact of global warming on various aspects of weather such as rainfall, humidity, temperature, etc. In this project, we focus on forecasting the temperature of Kanpur city using different machine learning algorithms and different regressions. Using different regressions on historical weather data of Kanpur city, we predict the temperature, for example we first use multiple linear regression, then decision tree regression and then random forest regression.

6. EXPERIMENTATION

The record was recently divided into a train series and a test series. Each piece of information is simply labeled. First, let's take the train master. We train our model using histograms and graphs. The feature extracted in this way is stored in a histogram. This process is done for each trainset data. Now we build our classifier model. We consider the classifiers which are linear regression, decision tree regression and random forest regression. We train our model using a histogram. The most important thing in this process is to set these parameters accordingly so that we get the most accurate results. After finishing the training, we will do tests. We now extract the features of each data variable in the test set using feature extraction techniques and then compare its values with the histogram values formed by the train set. The output for each test day is then predicted. Now to calculate the accuracy, we compare the predicted value to the specified value. Different metrics we use confusion matrix, R2 score etc.
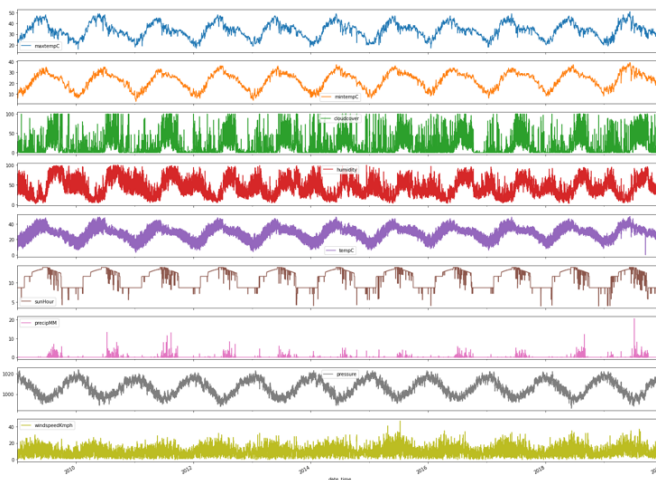
date_time	Actual	Prediction	diff
2013-07-10 08:00:00	34	33.209030	0.790970
2015-11-04 20:00:00	25	25.275755	-0.275755
2015-09-21 09:00:00	34	31.975338	2.024662
2017-02-16 11:00:00	28	20.496727	7.503273
2012-07-21 01:00:00	28	28.401085	-0.401085
...	...	...	...
2019-03-30 09:00:00	37	33.187428	3.812572
2015-11-12 12:00:00	32	28.483724	3.516276
2019-12-31 05:00:00	8	15.177361	-7.177361
2019-08-02 17:00:00	35	35.363251	-0.363251
2019-10-22 08:00:00	26	27.890691	-1.890691

19287 rows × 3 columns

7. RESULT AND DISCUSSION

The results of the implementation of the project are demonstrated below.

Multiple linear regression:

This regression model has a high mean absolute error, so it turned out to be the least accurate model. Below is a snapshot of the actual result of applying a multiple linear regression design.

Decision Tree Regression:

This regression model has a high mean absolute error, so it turned out to be the least accurate model. Below is a snapshot of the actual result of applying a multiple linear regression model.

8. CONCLUSION

All machine learning models: linear regression, divergent linear regression, decision tree regression, random forest regression outperformed expert climate analyzers, although their execution error decreased significantly in the following days, indicating that in the long run our models can outperform genius professionals. Linear regression was shown to be a low-slope, high-variance model, while polynomial regression was shown to be a high-slope, low-variance model. A linear regression model is a naturally high-variance model because it is unstable to outliers, so one way to improve a linear regression model is to collect more data. However, the practical regression had a high sensitivity, indicating that the decision of the model was poor and its predictions could not be improved by adding more information. This trend can be expected because the structural judgment evaluates the temperature as dependent on the climate of the previous two days, which may be too short to even consider capturing the slope in the climate necessary for a practical regression. If the number were based on the climate of the last four or five days, the bias of the practical regression model could probably be reduced. In any case, this would require significantly more computation time along with relearning the weight vector w, so it is left for future work. Speaking of random forest regression, it turns out to be the most accurate regression model. This is probably the most popular regression model used because it is very accurate and versatile. Below is an overview of the implementation of Random Forest in a project. Weather forecast has a large test that predicts accurate results, which are used in many permanent settings such as electrical offices, air terminals, tourist centers, etc. The problem with specifying this is parameter confusion. Each parameter has an alternative arrangement of quality ranges.

References

[1] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[2] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[3] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] K. Elissa, "Title of paper if known," unpublished.