INTEGRATING PREDICTIVE ANALYTICS AND MACHINE LEARNING FOR WEATHER FORECASTING

Dhirendra Kumar Gupta

Department of Statistics and Operations Research
College of Science, Qassim University
Qassim -51911, Saudi Arabia
E-mail :d.gupita@qu.edu.sa

KEYWORDS
Prediction, Weather Forecasting, Accuracy, Climate, Preprocessing

ABSTRACT
To make accurate predictions about future weather and environmental conditions, predictive analytics makes use of cutting-edge data analysis tools like statistical modeling and machine learning. Predictive models are able to offer precise forecasts of important environmental variables including air quality, humidity, precipitation, and temperature by evaluating massive information collected from sensors, satellites, and weather stations. This study provides a comprehensive examination of findings from weather forecasting utilizing scatter plots, outputs from Ordinary Least Squares (OLS) models, computations of errors, and evaluations of accuracy, with a special emphasis on decision tree models. This methodological framework greatly aids in the progress of machine learning techniques by guaranteeing the creation of trustworthy models that can accurately anticipate future outcomes. The results show that machine learning approaches in weather forecasting have made great strides, leading to more accurate predictions.

I. INTRODUCTION

Predictive analytics for weather and environmental monitoring is a rapidly evolving discipline that uses complex statistical models, machine learning methods, and large volumes of data to predict weather conditions and monitor changes in the environment. The capacity to accurately anticipate weather patterns and environmental phenomena is more important than ever due to the fact that climate change and environmental degradation continue to provide major problems to human civilizations and natural ecosystems. Scientists and meteorologists may use predictive analytics to anticipate a variety of atmospheric variables, including temperature, precipitation, humidity, wind speed, and pressure. These variables are important for understanding both short-term weather occurrences and long-term climate trends. This method, which is based on data, makes it possible to create models that not only provide more precise weather forecasts but also provide more detailed information on changes in the environment, including air and water quality, deforestation, pollution, and even the migratory patterns of different species.

The first step in the process of predictive analytics is to gather a vast amount of data from a variety of sources. These sources include ground-based weather stations, satellites, weather balloons, remote sensing technologies, and environmental monitoring networks. These data sources provide a complete picture of the atmosphere and the environment by recording real-time information that represents the intricacies of weather systems and biological processes. The following phase is data processing, which entails cleaning, standardizing, and organizing raw data so that it can be analyzed. After that, this data is analyzed using a variety of modeling approaches which are able to find patterns and correlations in the data. After then, these patterns may be utilized to forecast meteorological events that will happen in the future or to evaluate environmental concerns, giving important information about what might happen in certain situations.

The capacity of predictive analytics to offer early warnings is one of the most important benefits of using it in weather and environmental monitoring. These early warnings may help save lives, safeguard property, and reduce the damage caused by natural catastrophes. When weather forecasts are accurate, governments, corporations, and emergency response teams can plan for severe weather occurrences in advance. For instance, anticipating the route of a hurricane or the arrival of a severe storm offers authorities the time they need to issue evacuation orders, block roadways, and distribute emergency supplies. In a similar way, predictive models assist local communities in implementing preventative methods, such as building flood barriers or establishing firebreaks, by anticipating the risk of floods or wildfires. This capacity to foresee is also very important in agriculture, where farmers may expect meteorological extremes like droughts, cold, or excessive rainfall. This helps them make better choices about when to sow, irrigate, and harvest. In addition, predictive analytics may enhance the management of water resources by projecting the availability of freshwater supplies, monitoring water quality, and anticipating flood risks in river basins.

Predictive analytics is an important tool for understanding and reacting to the changes brought about by climate change, which is continuing to affect weather patterns and environmental conditions throughout the world. Climate models are a kind of predictive analytics that provide information on the long-term patterns that will influence the climate of the Earth in the future. These models estimate future climatic scenarios by taking into consideration elements such as greenhouse gas emissions, solar radiation, ocean currents, and atmospheric composition. Policymakers may evaluate the possible effects of various climate policy choices and choose the most effective strategy for reducing climate change by conducting multiple simulations based on different amounts of emissions. Predictive analytics also helps with climate adaptation efforts by predicting the potential impacts of climate change on particular places. For example, it may anticipate the increasing frequency of droughts in dry regions or the growing danger of floods in coastal towns. This knowledge is crucial for building infrastructure, controlling agricultural output, and safeguarding vulnerable people from the most severe effects of climate change.

The use of predictive analytics for weather and environmental monitoring has several advantages, but it also has a number of limitations. The complexity and variety of environmental systems is one of the main challenges. There are many linked components that affect the weather and climate, and even little changes in one part of the system may have a big impact on the rest of the system. Because of this complexity, it is challenging to develop models that are dependable and precise in all places and under all circumstances. Furthermore, the quality and availability of data might be a limiting issue. Despite the fact that there is now a much larger quantity of data accessible for weather forecasting and environmental monitoring, there are still areas that are not covered, particularly in distant or impoverished locations. In addition, the quality, resolution, and temporal coverage of data from various sources may differ, which may lead to uncertainty in prediction models. In order to address these difficulties, it is necessary to continue investing in technology for collecting data and to create more advanced algorithms and approaches to manage the enormous complexity of environmental systems.

II. REVIEW OF LITERATURE

Mehendale, Pushkar. (2022). Climate change is an issue that has never been seen before, and it requires new and creative answers. Artificial Intelligence (AI) has become a powerful technology that can be used to combat climate change. Artificial intelligence is essential for monitoring the environment since it allows for the collecting and analysis of data in real time and the prediction of environmental changes. It helps to improve sustainable practices by making the best use of water, managing renewable energy sources, and lowering greenhouse gas emissions. AI can predict climate modeling and impact assessment, which helps in the development of proactive adaptation strategies. Furthermore, to ensure that AI technologies are utilized fairly and responsibly in the fight against climate change, ethical considerations such as data privacy and algorithmic bias must be considered.

Fathi, Marziye et al., (2021) Predicting the weather is an important and essential part of people's daily life. It evaluates the current atmospheric condition in light of the changes that are occurring. The term "big data analytics" refers to the process of mining massive datasets for previously unseen insights and patterns that might improve previous results. Today, big data is attracting attention from all walks of life, including the weather bureau. This means that weather forecasts will be more accurate and that big data analytics will help meteorologists make more exact predictions. This objective and its best solutions have prompted the proposal of a number of big data methods and technologies for the management and assessment of massive volumes of meteorological data collected from a variety of sources. When used to weather forecasting, big data analytics has the potential to solve some of the issues plaguing more conventional approaches to data management. This study lays forth a plan for conducting a comprehensive literature assessment of articles published between 2014 and 2020 about big data analytical approaches used in weather forecasting. Methods based on techniques, methods based on technology, and hybrid approaches are the three broad categories into which the current reviewed works fall. Along with comparing the aforementioned categories, this research also looks at how well they scale, how long it takes to execute, and other Quality of Service metrics.

Huang, Zi-Qi et al., (2020) We rely on reliable weather reports to plan our everyday activities. The suggested weather management system is a two-stage process that would track and forecast weather data. This system integrates data processing, bus mobility, sensors, and deep learning technologies to offer weather monitoring in real-time on buses and stations, as well as to accomplish weather forecasts using predictive models. Expanding the measurement coverage and providing additional sensory data, this study takes advantage of moving buses and local information processing. To achieve this, it makes use of sensor data collected from buses. I use the test environment's temperature, humidity, and air pressure data given by the weather sensors to train and validate the multilayer perceptron and long short-term memory models. Using the trained learning model, weather time series predictions are made in Phase II. Predicted weather data is compared with actual measurements obtained at the Taichung observation station by the EPA and the CWB to assess the accuracy of the predictions. The system's performance may be examined in this way. The research shows that the proposed system reliably monitors the weather and, using trained models, produces accurate one-day weather predictions.

Biswas, Munmun et al., (2018) To predict the weather in a given area, meteorologists use a variety of scientific and technical tools. There are few challenges on a global scale as difficult as this one. Making weather predictions using predictive analysis is the focus of this study. Therefore, it's important to study different data mining procedures thoroughly before using them. Using the Naive Bayes and Chi-square algorithms, this research shows how to construct a classifier method for weather prediction. A online application with an intuitive GUI, this system is easy to use. By entering their login credentials, users will be able to access the system. Current conditions, including weather, humidity, and wind speed and direction, will be input by the user. In order to generate a weather forecast using this parameter, the system will combine the provided data with database records. Therefore, the two primary tasks of classification (during training) and prediction (during testing) will be executed. The outcomes demonstrated that these data mining techniques are capable of producing sufficient weather forecasts.

III. EXPERIMENTAL SETUP

Data preprocessing, normalization, training, and assessment are the four stages that make up the suggested model's architecture. In order to maximize the accuracy of the model, the setup makes sure that the data is clean, properly prepared, and processed effectively. The process's several stages are illustrated by the flow diagram.

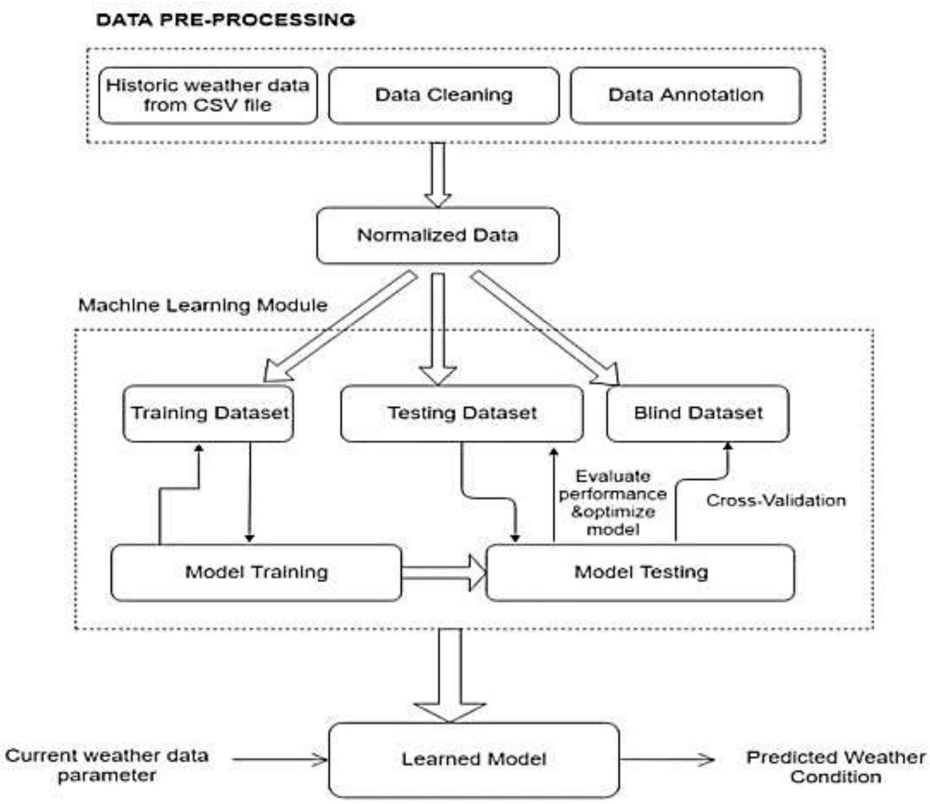

Figure 1: Architecture of proposed model

Data Preprocessing

An increase in the amount of preprocessing done on the data set will lead to more precise results. Simply said, it's the procedure by which we filter out irrelevant or useless data that is too loud to be meaningful. You won't receive accurate results until you eliminate all fields with null values or empty strings. As a result, developing the model is a crucial procedure.

Normalization

Machine learning module is another name for it. Once again, we blind the dataset for cross validation, after which we train and test the dataset to build a new model.

Learn Model

We finalize the process by learning from the model and making a prediction about the outcome. We have assessed the learning model correctly, which is vital. The training approach yields the artifact model here.

IV. RESULTS AND DISCUSSION

Scatter plots

Figure 2 shows the relationship between the actual and anticipated values from the linear regression model. The scatter plot might help you see trends or unusual results that don't match up with your predictions.

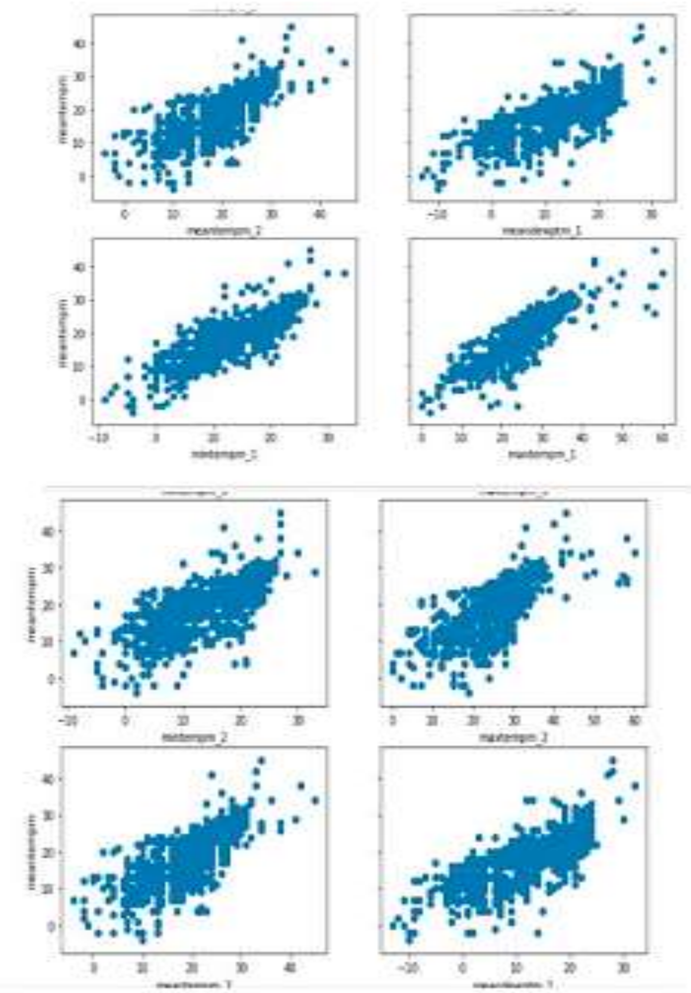

Figure 2: Scatter plots for linear regression model

SEEJPH Volume XXVI, S1, 2025, ISSN: 2197-5248; Posted:05-01-2025

OLS model Results

By using the Ordinary Least Squares (OLS) model, we may determine if the relationship between the two variables is linear. The OLS model's output sheds light on the linear regression model's data fitting accuracy.

Out[12]:

OLS Regression Results

Dep. Variable:	meantempm	R-squared:	0.982
Model:	OLS	Adj. R-squared:	0.982
Method:	Least Squares	F-statistic:	5979
Date:	Fri, 06 Jul 2018	Prob (F-statistic):	0.00
Time:	20:08:07	Log-Likelihood:	-2443.9
No. Observations:	987	AIC:	4906
Df Residuals:	978	BIC:	4950
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
mindewptm_3	0.0821	0.022	3.665	0.000	0.038	0.126
maxdewptm_2	-0.1758	0.027	-6.533	0.000	-0.229	-0.123
mindewptm_2	-0.1459	0.029	-5.081	0.000	-0.202	-0.090
maxtempm_3	0.1630	0.021	7.908	0.000	0.123	0.203
meandewptm_1	-0.1103	0.052	-2.118	0.034	-0.212	-0.008
mindewptm_1	0.2859	0.044	6.537	0.000	0.200	0.372
mintempm_1	0.7105	0.135	5.260	0.000	0.445	0.976
maxtempm_1	0.8414	0.126	6.673	0.000	0.594	1.089
meantempm_1	-0.7120	0.254	-2.806	0.005	-1.210	-0.214

Figure 3: OLS model results

Calculated Errors

The errors in the model's predicted values are shown in Figure 4. In order to enhance the model's performance, it is necessary to decrease these mistakes, which provide insight into the accuracy of the model's predictions.

```
The Explained Variance: 0.85
The Mean Absolute Error: 2.10 degrees celcius
The Median Absolute Error: 1.30 degrees celcius
```

Figure 4: Errors for Estimates

Accuracy Measurement of Decision Tree

Decision tree models rely on feature selection, as seen in Figure 5. In order to simplify the model without losing any important information, the relevant characteristics are chosen.

Out[38]:

	maxhumidity_3
0	84
1	92
2	92
3	80
4	80

Figure 5: Select features

Figure 6 shows the trained decision tree model that was applied to the dataset once the relevant characteristics were identified. Using data-driven decision criteria, this tree can forecast future events.

Out[44]:

	maxhumidity_3
920	79
651	96
972	93
513	93
464	63

Figure 6: Trained data

The decision tree model's accuracy is shown in Figure 7. It displays the decision tree's performance on the test data, which is usually evaluated using metrics like F1-score, recall, accuracy, or precision.

```
In [53]: accuracy_score(y_true = y_test, y_pred = predictions)

Out[53]: 0.9090909090909091
```

Figure 7: Accuracy measurement of decision tree

V. CONCLUSION

The study shows that machine learning algorithms have a lot of promise for improving weather forecasts with predictive analytics. We have constructed strong models that can make accurate predictions by following a methodical strategy that involves data preprocessing, normalization, training the models, and assessment. Decision tree accuracy assessments, error analysis, scatter plots, and Ordinary Least Squares (OLS) models were used to further confirm the efficacy of the suggested approach. According to the results, machine learning, especially when applied to decision tree algorithms, has the potential to greatly enhance the accuracy of weather forecasts, leading to developments in the field of meteorology. Additional variables, real-time forecasting applications, and model development might be the focus of future research.

REFERENCES: -

[1] P. Mehendale, "AI for Climate Change Mitigation: Predictive Analytics and Sustainable Solutions," *EAI Endorsed Trans. Cloud Syst.*, vol. 9, no. 5, pp. 140–144, 2022.

[2] M. Fathi, M. H. Kashani, S. M. Jameii, and E. Mahdipour, "Big Data Analytics in Weather Forecasting: A Systematic Review," *Arch. Comput. Methods Eng.*, vol. 29, no. 5, pp. 1–29, 2021.

[3] S. D. Latif, A. N. Ahmed, Y. F. Huang, and A. El-Shafie, "Evaluation of deep learning algorithm for inflow forecasting: A case study of Durian Tunggal Reservoir, Peninsular Malaysia," Natural Hazards, vol. 109, no. 1, pp. 351–369, 2021.

[4] Z.-Q. Huang, Y.-C. Chen, and C.-Y. Wen, "Real-Time Weather Monitoring and Prediction Using City Buses and Machine Learning," *Sensors (Basel, Switzerland)*, vol. 20, no. 18, p. 5173, 2020.

[5] V. Kumar and D. Kumar, "A systematic review on firefly algorithm: past, present, and future," *Arch. Comput. Methods Eng.*, vol. 28, no. 4, pp. 3269–3291, 2020.

[6] M. Biswas, T. Dhoom, and S. Barua, "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data," *Int. J. Comput. Appl.*, vol. 182, no. 34, pp. 20–24, 2018.

[7] Q. Xu et al., "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1283–1291, 2015.

[8] S. Kumar, M. K. Tiwari, C. Chatterjee, and A. Mishra, "Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis, and bootstrap method," Water Resources Management, vol. 29, no. 13, pp. 4863–4883, 2015.

[9] A. Tiwari and S. K. Verma, "Cloudburst predetermination system," IOSR Journal of Computer Engineering, vol. 17, no. 3, pp. 47–56, 2015.

[10] D. Chauhan and J. Thakar, "Boosting decision tree algorithm for weather prediction," *Journal of Advanced Database Management & Systems*, vol. 1, no. 3, pp. 21–30, 2014.

[11] R. Sarvizadeh, M. H. Kashani, F. S. Zakeri, and S. M. Jameii, "A novel bee colony approach to distributed systems scheduling," *Int. J. Comput. Appl.*, vol. 42, no. 10, pp. 1–6, 2012.

[12] D. Bergholt and P. Lujala, "Climate-related natural disasters, economic growth, and armed civil conflict," *J. Peace Res.*, vol. 49, no. 1, pp. 147–162, 2012.

[13] S. S. Badhiye, B. V. Wakode, and P. N. Chatur, “Analysis of temperature and humidity data for future value prediction,” *International Journal of Computer Science and Information Technologies*, vol. 3, pp. 3012–3014, 2012.

[14] A. Costello et al., "Managing the health effects of climate change. Lancet and University College London Institute for Global Health Commission," *Lancet*, vol. 373, no. 9676, pp. 1693–1733, 2009.

[15] P. Lynch, "The origins of computer weather prediction and climate modeling," *J. Comput. Phys.*, vol. 227, no. 7, pp. 3431–3444, 2008.