Weather Prediction Application Using Artificial Neural Network And Spark

Varshil Narola
Department of Computer Science
The University Of Texas at Dallas
Richardson, TX, 75080, USA
VXN220007

Gaurav Sharma
Department of Computer Science
The University Of Texas at Dallas
Richardson, TX, 75080, USA
GXS230001

Wasey Mulla
Department of Computer Science
The University Of Texas at Dallas
Richardson, TX, 75080, USA
WAM180001

Dhairya Desai
Department of Computer Engineering
The University Of Texas at Dallas
Richardson, TX, 75080, USA
DPD140130

***Abstract*— One of the most important and complex processes in the modern world is forecasting. Forecasting weather involves predicting how the atmospheric conditions will change over time, which is a natural process. Because of the harsh weather forecasts and methods, it has become a fundamental topic of research and study for scientists. The challenge of predicting the weather has always been difficult. Researchers and scientists have made many attempts to develop a linear relationship between the input and targeted weather data. The goal now is to anticipate non-linear weather data since weather is inherently dynamic and non-linear. Even though weather prediction is automated and mostly based on statistics, the results are not always reliable when using traditional methods. Artificial Neural Networks (ANN) are the most effective method for handling such issues because of their complicated and non-linear processes. Because of its increased accuracy, dependability, and efficiency, ANN makes weather forecasting simpler. The capabilities of artificial neural networks (ANNs) extend beyond historical data analysis to provide future forecasts, making them very suitable for weather forecasting. Neural Network is a highly intricate network with a malleable and adaptable structure. It is self-learning and uses the training data it already has to create a new intelligent pattern that can be used to predict the weather. A review of many Neural Network strategies for weather prediction is given, and it is found that the trained Neural Network can predict and categorize the weather variables with little error by only adding more hidden layers. This research uses a back propagation network (BPN) and predictive analysis method to train the network to forecast future weather. These surveys examine and illustrate the technological milestones that different researchers in this field have attained.**

Keywords—Artificial Intelligence, Artificial Neural Network (ANN), weather forecasting, training data, hidden neurons, Back propagation Network.

I. INTRODUCTION AND BACKGROUND WORK

Since many highly technologically sophisticated industries, like agriculture, crisis management, transport, communication, pollution dispersal, etc., rely on weather forecasting, it is regarded as a major feature. Given that weather is unpredictable and intense and that many factors must be taken into account, weather forecasting is evidently a difficult and intricate procedure. It is primarily caused by two factors: first, human activity, which helps people make more modern decisions on a daily basis, and second, the conveniences brought about by various technological advancements that are closely related to this field of study, such as advancements in computation and measurement systems.

All things considered, it is seen to be an important topic to focus on globally, as weather forecasts may have a profound effect on many different nations. As a result, forecasting the weather uses the requirements of research scientists to benefit the average person, greatly contributing to daily aspects. This demonstrates unequivocally why forecasting is impossible to anticipate using simpler methods and necessitates the use of sophisticated computing systems that can quickly analyze statistical and nonlinear data to provide patterns and rules that can be reviewed and trained using historical data. This kind of prediction is better suited for ANN, which will produce more accurate and effective results. The accuracy level will be higher than in the previous forecast, even though the error may or may not completely condense. Since the back propagation network (BPN) technique is by far one of the most often utilized training algorithms for weather prediction in the neural network area, a thorough review of the Neural Network idea is done in this study. In essence, neural networks are data-driven machine learning algorithms that may be used to predict statistical as well as binary outcomes. Together, they provide a hybrid system of smoothing techniques and regressions. Similar to smoothing algorithms, it leverages the available data to identify patterns. It has resemblance to regressions, which were built on the basis of cross-sectional data to establish a link between input and output variables.

II. PROBLEM DEFINITION

Accurate weather forecasting is crucial for various sectors like agriculture, disaster management, travel, and communication. However, traditional forecasting methods struggle with the inherent complexity of weather. Weather is a chaotic and dynamic system influenced by numerous factors, making it non-linear and difficult to model with linear relationships. Existing statistical methods often provide uncertain and inaccurate results. Their inability to handle non-linear weather data limits their effectiveness. To address these limitations, we need a more robust and adaptable approach to weather forecasting. This paper proposes Artificial Neural Networks (ANNs) as a promising solution. ANNs are powerful machine learning algorithms inspired by the structure and function of the human brain. Unlike traditional methods, ANNs can learn complex, non-linear patterns from historical weather data. This allows them to identify subtle relationships between various weather parameters and predict future weather conditions with greater accuracy. ANNs can learn from vast amounts of historical weather data, identifying patterns and trends that may not be evident with traditional methods. They can capture the complex, non-linear nature of weather, leading to more accurate forecasts. ANNs are self-learning systems that can continuously improve their accuracy as they are exposed to new data. This research aims to explore the potential of ANNs in weather forecasting. We will investigate the effectiveness of Backpropagation Networks (BPNs), a popular training algorithm for ANNs, in predicting future weather conditions. By leveraging the power of ANNs, we hope to significantly improve the accuracy and reliability of weather forecasts, ultimately benefiting various sectors and individuals who rely on weather information.

III. THEORETICAL AND CONCEPTUAL STUDY OF THE TECHNIQUE

This section delves into the underlying principles and mechanisms of the Artificial Neural Network (ANN) used for weather prediction. An ANN is a productive model for processing data that is made up of software application components. For data processing, biological neurons are like brains. Despite being mechanized and mathematical using conventional techniques, weather forecasting produces outcomes that are not always reliable. Neural correlations are the most widely utilized predictors in machine learning overall.

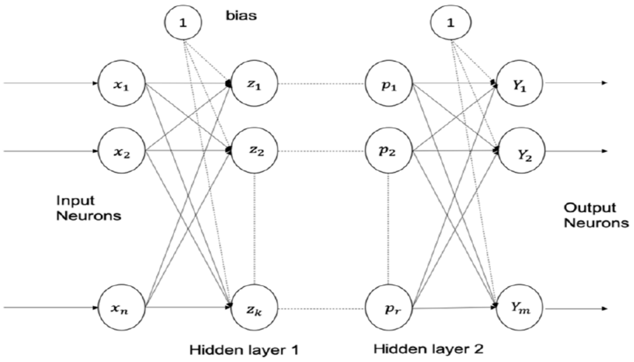

Fig. 1. Multi-Layer ANN Structure

An input layer contains all input variables (input neurons), an output layer contains all output variables (neurons), and lastly, a hidden layer is the layer between the input and output that contains hidden neurons. Typically, a neural network (NN) is represented as a network of connected nodes of a network called neurons, where each neuron is thought of as a variable. This layer is typically thought of as the neural network's hidden source. More layers and hidden nodes enable the network to record more intricate interactions. As a result, these ANN traits are more suited for identifying the climate issue. Typically, work is completed concurrently and in line with standard architecture. It appears that every muscle has an internal state of its own. This physiological state—which is dependent on the information the neuron receives—is referred to as an activation neuron.

A. *ANN in Weather Forecasting*

In predicting, accuracy and efficiency are crucial. The natural process of projecting future changes in the climate is called climate forecasting. Almost all of the inputs into a climate forecasting model are unique, and various techniques call for various sets of data that need to be properly maintained. Artificial intelligence (AI) approaches are often connected with nonlinear data, whereas statistical methods are typically linked with linear data. Neural networks, back propagation algorithms, genetic algorithms, and neuro-fuzzy logic are examples of AI-based learning algorithms. Accurate weather forecasting is possible with ANN. Since there are several characteristics in daily weather data that describe temperature, wind direction and speed, humidity, cloud range and density, and rainfall amount, etc. To prepare for humidity, rain, weather, or heat for the following day, these non-linear characteristics must be combined. For these kinds of systems, a complicated model that can learn from the training data it is given and generate the model automatically in order to get the desired outcomes is needed. Globally, weather forecasting may benefit nations in a pleasant way. In the day-to-day realm, weather forecasting greatly assists research scientists in meeting the demands of the general public. This is why weather forecasting requires sophisticated computer systems that can quickly assess statistical and nonlinear data to develop a model and a set of rules for observation and training from past occurrences. Forecasting cannot be done with basic tools. This kind of prediction is a good fit for ANN, which will produce more effective and accurate outcomes. The accuracy rate will rise in comparison to the prior forecast once more, even though the mistake could not be as little or as large. For this reason, one of the frequently used training techniques in neural networks for weather forecasting is the back-propagation network (BPN) algorithm. The field for input parameters and data selection is the initial stage in creating an ANN-based weather forecasting model.

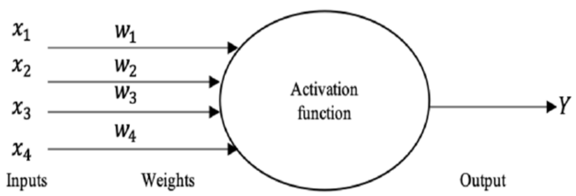

Fig. 2. Artificial Neuron Model

The basic idea is to use memory cells that have gating mechanisms installed so that information may be selectively remembered or lost over long durations.

Forward Pass-

- Input gates regulate the influx of fresh data into the memory cell in the forward pass.
- Forget Gates: These gates decide which data should be left behind from the prior cell state.
- Output Gates: Depending on the cell state at any given time, these gates control the information that is output.

Backward Pass-

- The backward pass technique involves computing the error gradient for each gate and adjusting the weights to reduce the error by applying the back-propagation through time (BPTT) algorithm.

The pseudocode for the given ANN is as follows:

```
for each epoch from 1 to num_epochs do
    for each training example (X_i, Y_i) in the dataset do
        // Forward Pass (Compute predictions)
        A_hidden, A_output = forward_pass(X_i, W_hidden, b_hidden, W_output, b_output)
        // Compute loss (e.g., cross-entropy loss)
        loss = compute_loss(Y_i, A_output)
        // Backward Pass (Compute gradients and update weights)
        gradients = backward_pass(X_i, Y_i, A_hidden, A_output, W_output)
        // Update weights and biases using gradients and learning rate
        W_hidden = W_hidden - alpha * gradients["dW_hidden"]
        b_hidden = b_hidden - alpha * gradients["db_hidden"]
        W_output = W_output - alpha * gradients["dW_output"]
        b_output = b_output - alpha * gradients["db_output"]
    end for
end for
```

B. *PROPOSED MODEL*

In this study, temperature data from the previous 20 years' temperature occurrences in June was gathered from the Kaggle data source. The data was then examined to uncover tendencies that would lead to the best possible forecasts in the future, ones that are almost exact matches to actual temperature events. The experimental study's analytical techniques involve data collection, analysis, transformation, modeling, and model performance analysis. They are based on the example depicted in FIG. 3.

1) Data Collection: Samples are collected during this phase, which also contains temperature, pressure, rain, and wind sensors. To ensure we have enough information for the event, the data are gathered again after a brief interval. In order to create a decent model, the suggested strategy calls for gathering a large number of data samples, which are then submitted for pre-processing.

2) Data Exploration: By seeing how the data are created, the researcher is able to "feel" the data through data analysis. Important insights are provided in this section.

3) Data Transformation: To get the data ready for the computation, it applies the suggested algorithm. It usually entails choosing the model's expensive components for which it is meant to be utilized. In the process, we created a matrix structure out of our data so that predictive models could use it with ease.

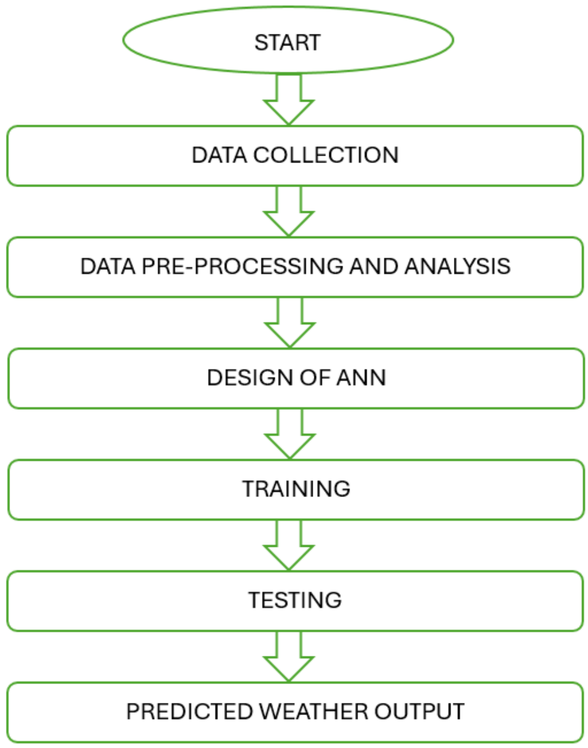

Fig. 3. Weather prediction process using ANN

4) Data Modeling: Currently, the gathered data will be simulated using the suggested method. Test data is provided to the trained model so that it may make predictions after training. These are accurate projections that do not overfit. dividing up the complete data set into training and analysis batches before utilizing the data.

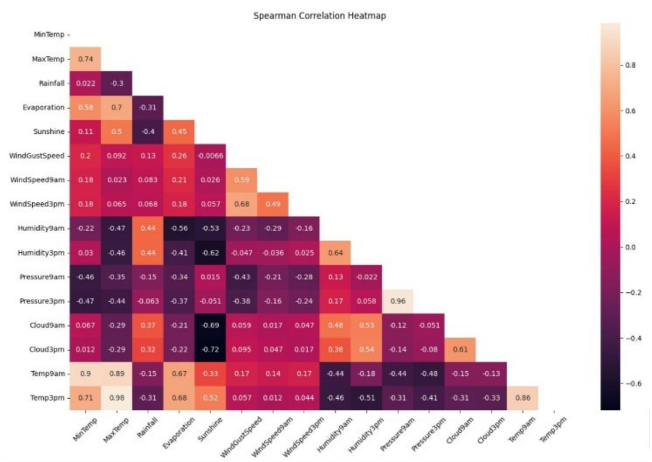

Fig. 4. Corelation Chart from Data Exploration Phase

5) Data Evaluation: This is the last phase of the model-based scoring, when the efficacy of the collaborative system in comparison to cutting-edge algorithms is evaluated.

The success of a predictive model relies on a substantial amount of data. This data is collected from sensors and goes through multiple stages. First, data is pre-processed to ensure its quality for the chosen algorithm. Then, the model is trained on a portion of the data. Finally, the model's performance is evaluated on unseen data to assess its generalizability. This process helps create accurate and reliable predictive models.

IV. RESULTS AND ANALYSIS

In this study, a backpropagation network with a predictive analytic technique employing Neural Network is used to provide an approximate weather forecasting output. The gradient descent process known as the "back propagation approach" is taught and simulated by varying the network's weights in order to minimize output error. Here, the accuracy between the expected and actual results is tested using the mean square error function. The outcome is more accurate the lower the error value. When weather is predicted using ANN, it shows that the temperature will climb progressively over the next few days and years. The same epoch is used for all validation, training, and best-fit results, during which the least gradient is reached using the input data from hidden layers, tests, training, and validation, together with the accompanying R values for each, and the output that results. As can be seen in Fig. 5, the Neural Network prediction result indicates a about significant accuracy in weather prediction. We can see that the error rate is also going down.

RMSE is a common metric used to measure the difference between predicted values by a model and the actual values. In weather forecasting, it typically refers to the difference between the predicted weather parameter (e.g., temperature, precipitation) and the actual observed value. Lower RMSE values indicate better model performance, signifying that the model's predictions are on average closer to the real values.

Accuracy is a broader metric that represents the overall correctness of the model's predictions. In weather forecasting, it might represent the percentage of times the model correctly predicted a specific weather event (e.g., sunny, rainy) or the value of a weather parameter within a certain acceptable range. Higher accuracy indicates better model performance. Looking at the table, ideally, we can see a consistent decrease in RMSE values and a corresponding increase in Accuracy values as the number of iterations (epochs) progresses. This would indicate that the ANN model is effectively learning from the training data and improving its prediction accuracy over time.

The table shows consistently low RMSE values and high Accuracy values across all iterations, it suggests that the ANN model is performing very well. It's effectively capturing the underlying patterns in the weather data and generating accurate predictions. A gradual decrease in RMSE and increase in Accuracy over iterations indicates that the model is learning and improving.

The model's performance seems to be significantly impacted by the feature selection. Based on the available measurements, it appears that there were no notable performance changes caused by differences in batch size and learning rate within a feature. The reasons for these outcomes may have been the following:

- Feature Importance: The model's robustness and reduced sensitivity to hyperparameter changes are indicated by the consistent performance inside a particular feature.
- Data Quality: Model performance may be impacted by the dataset's attributes and quality. The model is less responsive to changes in the hyperparameters since the dataset is not very well-structured and tidy.
- Changes to the hyperparameters are not resolving the possibility that the model is either overfitting or underfitting the data. Understanding model fit may be gained by keeping an eye on the performance throughout training and validation.

Iteration	Training Accuracy	Test Accuracy	Training RMSE	Testing RMSE
01	85.49	84.07	0.38	0.40
02	85.99	84.39	0.37	0.40
03	86.50	84.59	0.37	0.39
04	86.78	84.68	0.36	0.39
05	87.33	84.68	0.36	0.39
06	87.37	84.90	0.36	0.39
07	87.61	84.58	0.35	0.39
08	87.73	84.58	0.35	0.39
09	87.94	84.42	0.35	0.39
10	88.15	84.46	0.35	0.39
11	88.26	85.75	0.34	0.39
12	88.43	84.50	0.34	0.40
13	88.58	84.70	0.34	0.40
14	88.72	84.27	0.34	0.40
15	88.79	84.19	0.34	0.40
16	88.79	84.20	0.33	0.40
17	88.96	84.21	0.33	0.40
18	88.96	83.90	0.33	0.40
19	89.26	83.86	0.33	0.40
20	89.13	83.93	0.33	0.40

Fig. 5. Logs

V. CONCLUSION AND FUTURE WORKS

This research reviews many artificial neural network weather forecasting approaches and uses the backpropagation methodology to test them further. We may deduce that the more accurate the anticipated outcome, the smaller the mean square error value. This article presents the suggested Artificial Neural Network model, which includes validation and best-fit data, the number of hidden layers employed, all input and output parameters, testing and verification of the statistical dataset, and the neurons' content, weight, bias, learning rate, and activation function. Although the approach shown here can produce decent results, different Neural Network techniques might be used to replace the primary component of the prediction system in the future to improve performance. Furthermore, while the primary goal of this experiment is to anticipate local weather, the study might also be used to other fields such as electrical load forecasting, stock market value prediction, travel and tourism, etc.

ACKNOWLEDGEMENT

All team members worked together to complete this report, and we sincerely thank everyone who helped make it a reality.

First and foremost, we would like to express our sincere gratitude to Dr. Anurag Nagar, our professor, for all of his help and advice as well as for giving us the resources and knowledge base that have helped to shape and improve the quality of the research. We express our gratitude to the University of Texas at Dallas for offering the facilities, equipment, and supportive atmosphere required to carry out this study. We are appreciative of the open-source community for providing the frameworks and technologies, such as Databricks, that made the code easier to implement.

At last, we would like to thank our family and friends for their moral support.

REFERENCES

[1] Riyazuddin, Y. Md, S. Mahaboob Basha, and K. Krishna Reddy.

"An Approach for Prediction of Weather System by Using Back propagation Neural Network."

[2] Pappula, Praveen, Ramesh Javvaji, and B. Rama. "Experimental Survey on Data Mining Techniques for Association rule mining." International Journal of Advanced Research in Computer Science and Software Engineering(2014).

[3] Gill, Er Jasmeen, Er Baljeet Singh, and Er Shaminder Singh.

[4] "Training back propagation neural networks with genetic algorithm for weather forecasting." In IEEE 8th International Symposium on intelligent Systems and Informatics, pp. 465-469. IEEE, 2010.

[5] Stack Overflow. (n.d.). Showing an error: This MinMaxScaler instance is not fitted yet. Call 'fit' with appropriate arguments before using this method. Retrieved from https://stackoverflow.com/questions/56058486

[6] Apache Spark Documentation. (n.d.). pyspark. ml.feature.VectorAssembler. Retrieved from h t t p s : //spark.apache.org/docs/3.1.3/api/python/reference/api/pyspark

[7] Apache Spark Documentation. (n.d.). pyspark.ml.feature.MinMaxScaler. Retrieved from https://spark.apache.org/docs/latest/api/python/reference/api/pyspark. ml.feature.MinMa

[8] S.S. De, University of Kolkata, Artificial Neural Network Based Prediction of Max. & Min. Temperature in the Summer-Monsoon month over India, Applied Physics Research, Vol.1, No.2, Nov- 2009