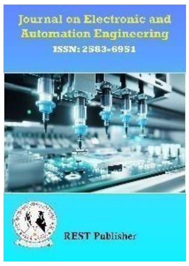
Journal on Electronic and Automation Engineering

Vol: 2(4), December 2023

REST Publisher; ISSN: ISSN: 2583-6951 (Online)

Website: https://restpublisher.com/journals/jeae/

DOI: https://doi.org/10.46632/jeae/2/4/2



Improving Weather Forecasting Accuracy Using Machine Learning

***Senthilkumar Meyyappan, A. Bharath Naik, A. Uma Sai, Ch. Keerthi**
Nalla Malla Reddy Engineering College, Hyderabad, India.
*Corresponding author Email: kathir_senthil@yahoo.co.in

Abstract: *Weather forecasting has several applications in our daily lives, ranging from agriculture to event planning. Previous weather forecasting models relied on a complex combination of mathematical instruments, which was insufficient to achieve a higher categorization rate. We offer fresh revolutionary approaches for estimating monthly rainfall using machine learning algorithms in this study. Weather forecasts are created by gathering quantitative information about the current state of the atmosphere. Machine learning algorithms may learn complicated mappings from inputs to outputs using only samples and with little effort. The dynamic nature of the atmosphere makes accurate weather prediction challenging. The fluctuation in weather conditions in previous years must be used to anticipate future weather conditions. It is extremely likely that it will match within the next two weeks of the preceding year. We proposed using linear regressions with the Random forest algorithm to forecast weather using characteristics such as temperature, humidity and wind. It will forecast weather based on prior records thus, this prediction will be accurate.*

1. INTRODUCTION

Weather forecasting is primarily concerned with predicting weather conditions in the future. Weather predictions provide essential information about the weather in the future. Weather forecasting methodologies range from relatively simple observation of the sky to highly complicated computerized mathematical models. Weather forecasting is critical for a variety of purposes. Climate monitoring, drought detection, severe weather prediction, agricultural and production, energy industry planning, aviation industry planning, communication, pollution dispersal, and so on are some of them. Weather Underground's free tier API web service will be utilized to acquire the data for this series. Since 2015, I've been interacting with the API via the requests library to retrieve weather data for the city of Lincoln, Nebraska. Once acquired, the data must be processed and aggregated into a format suitable for data analysis before being cleaned.
The planet is a very complicated area that is suffering from climate change. It is critical to predict precise weather without error, to ensure security and mobility, as well as a safe daily operation.
Weather forecasting is built by collecting massive amounts of data, which makes machine learning an indispensable tool, by employing some back testing methods and algorithms to generate an accurate prediction of the weather.
We will evaluate the methods through a series of experiments that will emphasize their performance and worth. The addition of ML to weather forecasting provides a significant advantage for the prediction make it more accurate.
Weather forecasting is a very effective application of science for the benefit of society. It can be used for Aerospatiale, agriculture, sport, musical events, maritime, renewable energy, aviation, and forestry, where it is critical to know what the weather will be. Finding the highest and minimum temperature is one of the linear regression methods that yields an accurate answer.

- Collecting the data.
- Exploring and preparing data
- Data preparation- creating random training and test data.
- Training model on data.
- Evaluating model performance.
- Improving model performance.

Weather simply refers to the condition of air on the earth at a given place and time. It is a continuous, data-intensive, multidimensional, dynamic and chaotic process. These properties make weather forecasting is a formidable challenge. Forecasting is the process of estimation in unknown situations from the historical data. Weather forecasting is one of the most scientifically and technologically challenging problems around the world in the last century. To make an accurate prediction is indeed, one of the major challenges that meteorologists are facing all over the world. Since ancient times, weather prediction has been one of the most interesting and fascinating domains. Scientists have tried to forecast meteorological characteristics using a number of methods, some of these methods being more accurate than others.
Meteorology knowledge is the foundation of scientific weather forecasting, which revolves around predicting the state of the atmosphere at a certain location. Human weather forecasting is an example of needing to make decisions in the face of uncertainty.
Weather predictions are frequently generated by collecting quantitative data about the current condition of the atmosphere and projecting how the atmosphere will evolve in the future using scientific understanding of atmospheric processes. The importance of growing information about the cognitive process in weather forecasting has been recognized in recent years.
Forecasting the weather becomes a task for human practitioners in which the details might be uniquely personal, even though most human forecasters use methodologies based on meteorology science in common to deal with the profession's obstacles. Weather forecasting is the prediction of how the current state of the atmosphere will change in the future. Ground observations, observations from ships, observations from airplanes, radio noises, doppler radar, and satellites are used to determine current weather conditions. This data is transported to meteorological centers, where it is collected, analyzed, and displayed in a variety of charts, maps, and graphs. Many thousands of observations are transferred onto surface and upper-air maps by modern high-speed computers.

Weather predictions provide essential information about the weather in the future. Weather forecasting employs a variety of methodologies, ranging from simple observation of the sky to extremely complicated computerized mathematical models. Weather forecasts can be made for one day, one week, or several months in advance. However, the accuracy of weather forecasts decreases dramatically after a week. Because of its chaotic and unpredictable nature, weather forecasting remains a challenging business. It is still a procedure that is neither entirely scientific nor entirely artistic. It is well recognized that people with little or no formal training can achieve significant forecasting ability.

Farmers, for example, are typically fairly competent of producing their own short- term forecasts of meteorological conditions that directly affect their livelihood, and the same may be said of pilots, anglers, mountain climbers, and so on. Weather events, which are frequently of a complicated character, have a direct impact on such people's safety and/or economic stability. Accurate weather forecast models are critical in third-world countries because agriculture is entirely dependent on the weather.

It is so critical to discover any patterns for weather parameters to depart from their periodicity, which would affect the country's economy. This dread has been heightened by the threat posed by global warming and the greenhouse effect.
Extreme weather events are becoming increasingly costly to society, inflicting infrastructure damage, injury, and death. Weather forecasting, as practiced by professionally educated meteorologists, is a highly developed ability that is based on scientific principles and methods and employs cutting-edge technical instruments. The significant gain in prediction accuracy that has occurred since 1950 is a direct result of technology advancements, basic and applied research, and weather forecasters' use of new information and procedures. High-speed computers, meteorological satellites, and weather radars are examples of tools that have significantly improved weather forecasting. Several other variables have greatly contributed to this increase in forecasting accuracy.

One example is the development of statistical tools for broadening and improving the accuracy of model predictions. Another benefit of meteorological satellites is greater observational capability. The ongoing improvement of the initial conditions prepared for the forecast models is a third major cause for the increase in accuracy. Statistical approaches anticipate a broader range of meteorological elements than models alone, and can tailor geographically less exact model projections to specific places. Satellites can now give near-constant viewing and remote sensing of the atmosphere on a worldwide scale. The improved initial circumstances are the product of more observations and better utilization of the observations in computational approaches.

Existing System: Because of the chaotic nature of the atmosphere, existing systems require significant processing capacity to solve the equations that describe the atmospheric conditions. This is due to an insufficient understanding of atmospheric processes, which means that forecasts become less accurate as the time difference between the present and the time for which the forecast is made rises. Weather is a continuous, data-intensive, multidimensional, dynamic, and chaotic process, and these characteristics make whether prediction difficult.

Disadvantages:

- Not well in prediction accuracy.
- It will not be supported with dynamic weather changes.

Proposed System: The proposed model will employ machine learning methods to forecast high and low temperatures based on a combination of all features. Because this technique cannot be utilised with classification data, it does not use weather classification data for each day. As a result, in the beginning of our project, just eight characteristics are chosen for use: maximum temperature, minimum temperature, mean humidity, and mean atmospheric pressure. A decision tree is the second algorithm to be employed. It searches for previous weather patterns that are similar to current weather patterns, and then predicts future weather conditions using the data from the historical weather patterns.

Advantages:

- Better efficiency and accuracy.
- A neural network is a powerful data modelling tool that is able to capture and represent complex input/output relationships.
- Able to scope with large weather dataset.
- Accurate weather prediction.

Block Diagram:

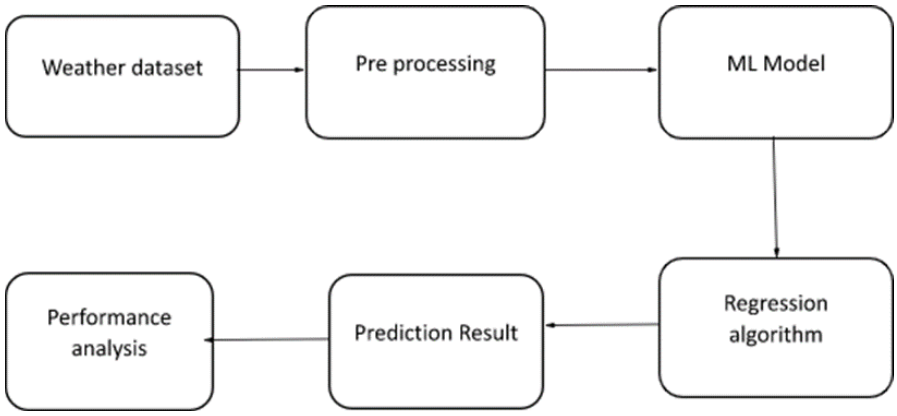


FIGURE 1. Block Diagram

Literature Survey: The study by Lu et al. (2012) focuses on efficiently processing k Nearest Neighbor joins (kNN joins) using the MapReduce framework, which is well-suited for data-intensive applications. The kNN join operation is commonly used in data mining to find the k nearest neighbors from one dataset for each object in another dataset, but it can be computationally expensive, especially with large volumes of data. To address this, the authors propose a mapping technique that uses pruning rules for distance filtering, which helps reduce both shuffling and computational costs. They also introduce two approximate techniques to minimize the number of replicas, further lowering the cost of shuffling. Through extensive testing on their cluster, the authors demonstrate that their approaches are efficient, robust, and scalable, making them valuable for processing kNN joins in large-scale environments.

2. PROJECT OVERVIEW

UML Diagrams: The Unified Modelling Language (UML) is used to specify, visualize, modify, construct and document the artifacts of an object-oriented software intensive system under development. UML offers a standard way to visualize a system's architectural blueprints, including elements such as:
Actors, business processes, (logical) components, activities, programming language statements, database schemas, and Reusable software components.
UML combines best techniques from data modelling (entity relationship diagrams), business modelling (work flows), object modelling, and component modelling. It can be used with all processes, throughout the software development life cycle, and across different implementation technologies. UML has synthesized the notations of the Boch method, the Object-modelling technique (OMT) and Object-oriented software engineering (OOSE) by fusing them into a single, common and widely usable modelling language. UML aims to be a standard modelling language which can model concurrent and distributed systems.

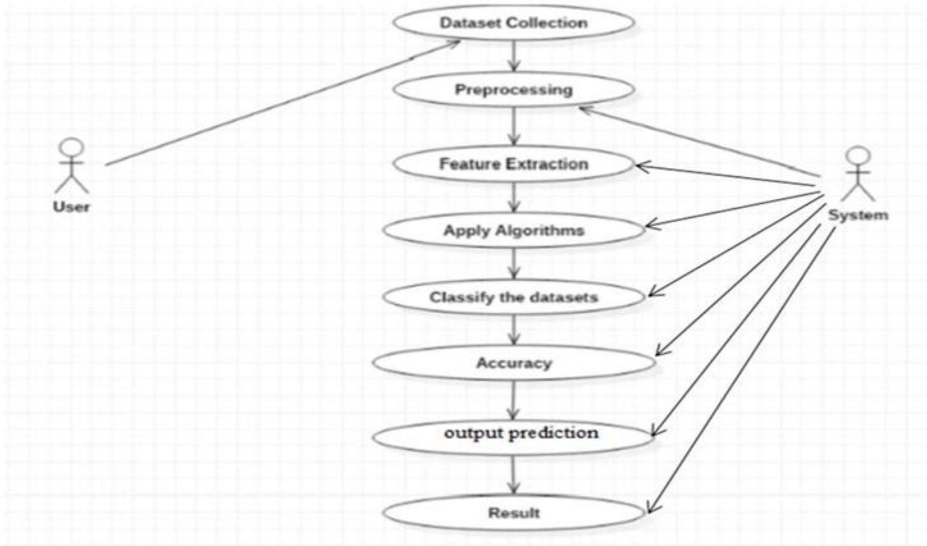


FIGURE 2. UML Diagram

Class Diagram: The class diagram is the main building block of object-oriented modelling. It is used for general conceptual modelling of the systematic of the application, and for detailed modelling translating the models into programming code. Class diagrams can also be used for data modelling. [1] The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed. In the diagram, classes are represented with boxes that contain three compartments: The top compartment contains the name of the class. It is printed in bold and cantered, and the first letter is capitalized. The middle compartment contains the attributes of the class. They are left-aligned and the first letter is lowercase. The bottom compartment contains the operations the class can execute. They are also left-aligned and the first letter is lowercase.

FIGURE 3. Class Diagram

Sequence Diagram: Sequence Diagrams Represent the objects participating the interaction horizontally and time vertically. A Use Case is a kind of behavioral classifier that represents a declaration of an offered behavior. Each use case specifies some behavior, possibly including variants that the subject can perform in collaboration with one or more actors. Use cases define the offered behavior of the subject without reference to its internal structure. These behaviors, involving interactions between the actor and the subject, may result in changes to the state of the subject and communications with its environment. A use case can include possible variations of its basic behavior, including exceptional behavior and error handling.

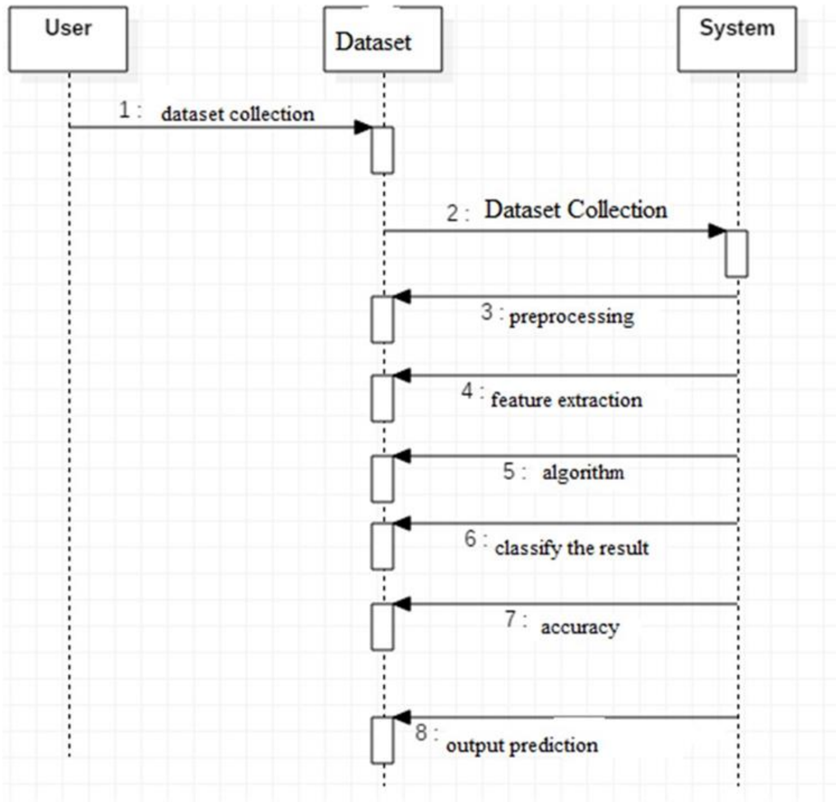


FIGURE 4. Sequence Diagram

Activity Diagrams: Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

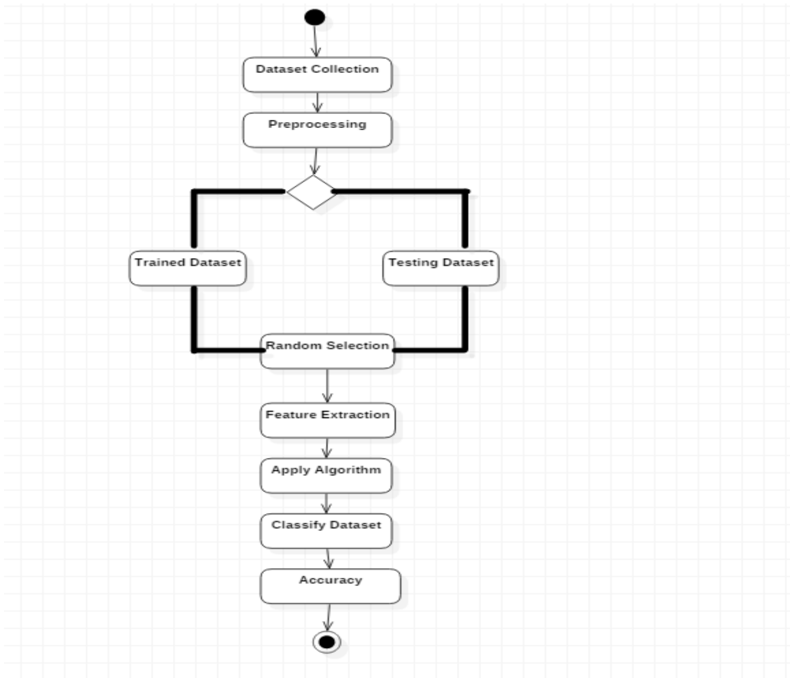


FIGURE 5. Activity Diagram

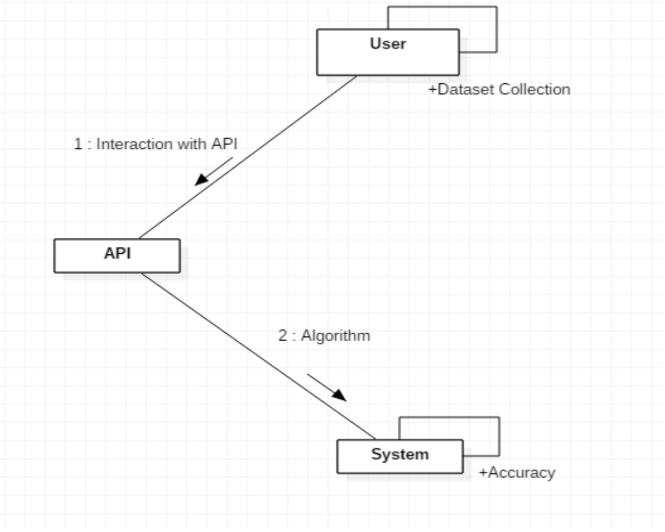


FIGURE 6. Collaborations Diagram

External Entity: An outside system that sends or receives data, communicating with the system being diagrammed. They are the sources and destinations of information entering or leaving the system. They might be an outside organization or person, a computer system or a business system. They are also known as terminators, sources and sinks or actors. They are typically drawn on the edges of the diagram.
Process: Any process that changes the data, producing an output. It might perform computations, or sort data based on logic, or direct the data flow based on business rules. A short label is used to describe the process, such as "Submit payment."

Data Store: Files or repositories that hold information for later use, such as a database table or a membership form. Each data store receives a simple label, such as "Orders."

Dataflow: The route that data takes between the external entities, processes and data stores. It portrays the interface between the other components and is shown with arrows, typically labelled with a short data name, like "Billing details."

Dataflow Diagram:

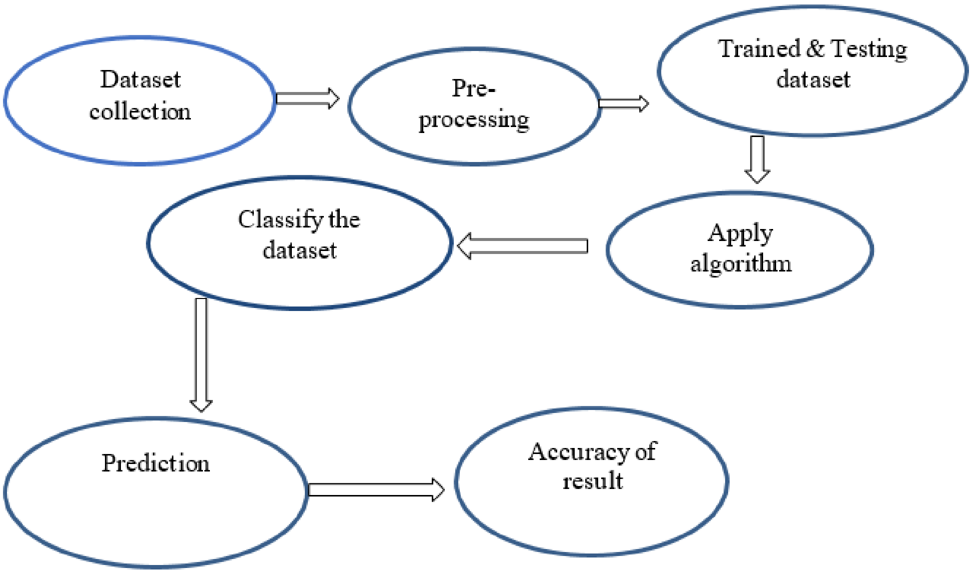


FIGURE 7. Dataflow Diagram

3. RESULTS

Import Libraries:

Here we use different libraries of python for importing the data and analyzing the data.

```
#!pip install tensorflow
import tensorflow as tf
tf.compat.v1.enable_eager_execution()
```

```
# Import pandas to load the data.
import pandas as pd
df = pd.read_csv("istanbul.csv")

# Display the first ten rows of the data frame.
df.head(10)
```

	date_time	maxtempC	mintempC	totalSnow_cm	sunHour	uvIndex	uvIndex.1	moon_illumination	moonrise	moons
0	01-01-2009 00:00	5	2	0.0	8.7	1	0	31	11:09 AM	10: F
1	01-01-2009 01:00	5	2	0.0	8.7	1	0	31	11:09 AM	10: F
2	01-01-2009 02:00	5	2	0.0	8.7	1	0	31	11:09 AM	10: F

FIGURE 9. Importing the Data Using Different Python Libraries

```
# Format the date_time column and make it the index.
df.date_time = pd.to_datetime(df['date_time']).dt.strftime('%m/%d/%Y,  %H:%M')
df.index = df['date_time']

# Extract the number of rows and columns by using the shape of the data.
numRows,numColumns = df.shape
# Extract the time interval.
last_date, first_date = df.iloc[0].date_time, df.iloc[-1].date_time
# Check the availability of the data.
na_cols = df.columns[df.isna().any()].tolist()

# Print the information.
print(f"There are {numRows} rows and {numColumns} columns in the initial dataset.")
print(f"The data represents the time frame between the dates '{last_date}' and '{first_date}'.")
if not na_cols:
  print("There are no NA rows.")
else:
  print(f"Columns in the dataset which include NA rows: {na_cols}.")
```

```
There are 87672 rows and 26 columns in the initial dataset.
The data represents the time frame between the dates '01/01/2009,  00:00' and '01/01/2019,  23:00'.
There are no NA rows.
```

FIGURE 10. Analyzing the Imported Data

```
# Visualize a plot of loss on the training and validations sets over training epochs.

plt.figure()

plt.plot(range(len(hist.history['loss'])), hist.history['loss'], 'b', label='Training loss')
plt.plot(range(len(hist.history['val_loss'])), hist.history['val_loss'], 'r', label='Validation loss')
plt.title('Training History')
plt.xlabel("Epochs")
plt.ylabel("Mean Absolute Error")
plt.legend()

plt.show()
```

FIGURE 11. Checking for Loss of Data or Any Error Occured

```
In [17]: y_pred= classifier.predict(xtest)

In [18]: from sklearn.metrics import accuracy_score
         accuracy_score(ytest, y_pred)

Out[18]: 0.2959470762679644
```

Random forest

```
In [19]: from sklearn.ensemble import RandomForestClassifier
         classifier1= RandomForestClassifier(n_estimators= 10, criterion="entropy")
         classifier1.fit(xtrain, ytrain)

Out[19]: ▾ RandomForestClassifier
         RandomForestClassifier(criterion='entropy', n_estimators=10)

In [20]: y_pred1= classifier1.predict(xtest)

In [21]: from sklearn.metrics import accuracy_score
         accuracy_score(ytest, y_pred1)

Out[21]: 0.23252984563911488

In [22]: import pickle
         filename = 'weather.sav'
         pickle.dump(classifier1, open(filename, 'wb'))
         # Loading the saved model
         loaded_model = pickle.load(open('weather.sav', 'rb'))

In [ ]:
```

FIGURE 12. Accuracy Result of Weather Prediction Using Random Forest Algorithm

FIGURE 13. Accuracy Result of Weather Prediction Using k-NN Algorithm

4. ADVANTAGES AND DISADVANTAGES

Advantages:

- Forecasting the weather is important because it can predict how the climate will change in the future.
- We can calculate the likelihood that snow and hail will reach the surface using latitude.
- We can determine the thermal energy that the sun is exposing a region to.
- The ability to plan military operations based on anticipated weather conditions benefits military personnel.
- To increase the likelihood of winning the war, the military can plan their Fig.hts while taking into account the predicted weather.

Disadvantages:

- It is very difficult to predict the weather accurately; it is expensive to keep track of so many variables from so many different sources. The cost of the computers required to do the requisite millions of calculations is high. If the weather does not match the forecast, the weather forecasters are held responsible.
- Technology Dependence: Accurate forecasts cannot be achieved if technology fails or is not available, as is the case with weather forecasting.
- Limited Reach: It is challenging for residents of many rural or sparsely populated places to prepare for severe weather because weather forecasts are not readily available in these locations.
- Complicated Terminology: Some people may find it difficult to grasp weather predictions because to the complicated terminology used in forecasting.

5. CONCLUSION

Thus, employing multiple machine learning techniques, this system has enabled the development of a smart weather forecast system. The deep learning algorithms are developed in a way that will increase the system's overall accuracy and boost sales in the agricultural industry. In order to train machine learning models to make accurate forecasts, the weather patterns are examined under various conditions.

Future Scope: Machine learning algorithms using by proficient climate or weather determining directions or forecasting, in spite of the fact that the error in their execution. Results are intrinsically a high and accurate as it is steady for exceptions and forecasting, so one approach to enhance the straight relapse show is by accumulation of more information using random forest classifier and decision tree and Gaussian Navie Bayes. Showing that the decision of model was efficient and effective that its expectations can be enhanced by promote accumulation of information under the proposed scheme.

REFERENCES

[1] Abramson, Bruce, et al. ”Hailfinder: A Bayesian system for forecasting severe weather.”International Journal of Forecasting12.1 (1996): 57-71.
[2] Cofno, Antonio S., et al.” Bayesian networks for probabilistic weather prediction.”15th Eureopean Conference on Artificial

Intelligence (ECAI). 2002.
[3] Kasturi, Kalyan, Rajani Dharanikota, Khaleelu Rehman, Senthilkumar Meyyappan, and Akhil Kommineni. "Low-Cost, IOT-Based Child Safety Monitoring Robot with User-Friendly Mobile App." In *Smart Sensors Measurement and Instrumentation: Select Proceedings of CISCON 2021*, pp. 503-511. Singapore: Springer Nature Singapore, 2023.
[4] Krasnopolsky, Vladimir M., and Michael S. FoxRabinovitz." Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction."Neural Networks19.2 (2006): 122-134.
[5] Lai, Loi Lei, et al." Intelligent weather forecast."Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on. Vol. 7. IEEE, 2004.
[6] Ng, Andrew. "CS229 Lecture Notes Supervised Learning" 2016.
[7] Radhika, Y., and M. Shashi." Atmospheric temperature prediction using support vector machines."International Journal of Computer Theory and Engineering1.1 (2009): 55.
[8] " Stanford, CA" in Weather Underground, The Weather Company, 2016. [Online]. Available: https://www.wunderground.com/us/ca/paloalto/zmw: 94305.1.99999. Accessed: Nov 20, 2016.
[9] Kumar, M. Senthil, and P. Sivakumar. "Enhancing the quality of technical education accreditation system using current communication system." *International Journal of Applied Engineering Research* 9, no. 23 (2014): 20421-20432.
[10] Gupta, Subham Kumar, Seema Rawat, and Praveen Kumar. "A novel based security architecture of cloud computing." In Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2014 3rd International Conference on, pp. 1-6. IEEE, 2014.
[11] Saini, Parag, Tanupriya Choudhury, Praveen Kumar, and Seema Rawat. "Proposal and implementation of a novel scheme for image and emotion recognition using Hadoop." In Smart Technologies for Smart Nation (SmartTechCon), 2017 International Conference On, pp. 1358-1363. IEEE, 2017.
[12] Weather.com,http://www.weather.com
[13] Kadambari, Sanchita, Seema Rawat, and Praveen Kumar. "A Comprehensive Study on Big Data and Its Future Opportunities." In Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 277-281. IEEE Computer Society, 2014.
[14] Meyyappan, S., and S. Chandramouleeswaran. "Impact of Mobility on the routine of enhanced-DSDV protocol in mobile ad-hoc networks." *Int J Appl Eng Res IJAER* 13 (2018): 11674-11679.
[15] Gupta, Subham Kumar, Seema Rawat, and Praveen Kumar. "A novel based security architecture of cloud computing." In Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2014 3rd International Conference on, pp. 1-6. IEEE, 2014.
[16] Sridhathan, C., M. Senthil Kumar, and K. Arutselvan. "Unlocking mobile devices using improved face recognition and eye blinking technique." *International Journal of Applied Engineering Research* 13, no. 24 (2018): 16907-16909.
[17] Wiki (2013). Applications and organizations using Hadoop. http://wiki.apache.org/hadoop/PoweredBy
[18] Gartner Research Cycle 2014, http://www.gartner .com
[19] Kumar, M. Senthil. "Energy Efficient Techniques for Transmission of Data in Wireless Sensor Networks." (2015).
[20] K. Morton, M. Balazinska and D. Grossman, "Paratimer: a progress indicator for MapReduce DAGs", In Proceedings of the 2010 international conference on Management of data, 2010, pp.507– 518.
[21] Vimalarani, C. I., and M. Senthilkumar. "Energy Efficient PCP protocol for k-coverage in Sensor networks." *Proc IEEE* (2010).
[22] Lu, Wei, et al. "Efficient processing of k nearest neighbor joins using MapReduce", Proceedings of the VLDB Endowment, Vol. 5, NO. 10, 2012, pp. 1016- 1027.
[23] J. Dean and S. Ghemawat, "Mapreduce: simplied data processing on large clusters", in OSDI 2004: Proceedings of 6th Symposium on Operating System Design and Implementation.
[24] Kumar, M. Senthil, and Ashish Chaturvedi. "Energy-Efficient Coverage and Prolongs for Network Lifetime of WSN using MCP." (2012).
[25] Sohn T., Lee J.H., Lee S.H. and Ryu, "Statistical prediction of heavy rain in South Korea" Advances in Atmospheric Sciences, Vol. 22, 2005.
[26] Kumar, M. Senthil, and Ashish Chaturvedi. "A novel enhanced coverage optimization algorithm for effectively solving energy optimization problem in WSN." *Research Journal of Applied Sciences, Engineering and Technology* 7, no. 4 (2014): 696-701.
[27] Kannan, M. Prabhakaran S. and Ramachandran, P. "Rainfall forecasting using data mining technique". International Journal of Engineering and Technology,Vol. 2, No. 6, pp. 397-401, 2010.
[28] Chattopadhay S. "Multiplayer feed forward artificial neural network model to predict the average summer monsoon rainfall in India". Acta Geophysica, Vol. 55, No. 3, pp. 369-382, 2007.
[29] Kumar, R. Dinesh, C. Sridhathan, and M. Senthil Kumar. "Performance Evaluation of Different Neural Network Classifiers for Sanskrit Character Recognition." *Business Intelligence for Enterprise Internet of Things* (2020): 185-194.
[30] Hayati M. and Mohebi Z. "Temperature forecasting based on neural network approach". World Applied Science Journal, Vol. 2, No. 6, pp. 613-620,2007.
[31] Kal N., Jim C. and Moula C. "The inventory policy using ESWSO measure for the ARIMA lead-time demand and discrete stochastic lead-time". Journal of Academy of Business and Economics, Vol. 10, No. 2, 2010.
[32] Badmus M.A. and Ariyo O.S." Forecasting cultivated areas and production of maize in Nigeria using ARIMA model". Asian Journal of Agricultural Sciences, Vol. 3, No. 3, pp. 171-176, 2011.
[33] Saima H., Jaafar J., Belhaouari S. and Jillani T.A. "ARIMA based Interval Type-2 Fuzzy Model for Forecasting". International Journal of Computer Applications, Vol. 28, No. 3, pp. 17-21,2011.
[34] Sridhathan, Senthilkumar, and M. Senthil Kumar. "Plant infection detection using image processing." *International Journal of Modern Engineering Research (IJMER)* 8, no. 7 (2018): 13-16.
[35] Sharma M.A. and Singh J.B. "Use of Probability Distribution in Rainfall Analysis". New York Science Journal, Vol. 3, No. 9, pp. 40-49,201.