ISSN No:-2456-2165

K-Means Clustering and Artificial Neural Network in Weather Forecasting

Shubham Choramle; Jaydeep Lokhande; Vidwesh Choudhary; Pruthviraj Mohalkar; Dr. Vitthal Gutte
Department of Computer Engineering and Technology, Dr. Vishwanath Karad MIT WPU University
Address: MIT World Peace University, Sr No 124, Ex Serviceman Colony, Paud Road, Kothrud, Pune-411038
1032210941@mitwpu.edu.in TO jaydeeps.lokhande@gmail.com
1032210967@mitwpu.edu.in TO vidweshchoudhary@gmail.com
1032210657@mitwpu.edu.in TO shubham.choramle1330@gmail.com
1032210772@mitwpu.edu.in TO pruthvirajmoholkar3@gmail.com
vitthalgutte2014@mitwpu.edu.in TO vitthal.gutte@mitwpu.edu.in

Abstract- The research paper's abstract provides a brief introduction to the subject of forecasting the weather, discusses the applications and limitations of K-Means clustering and ANNs (artificial neural networks) in weather forecasting, and identifies a comparison of these two approaches as the primary goal of the study. It finishes by highlighting the significance of integrating different methodologies for precise weather forecasting and briefly referencing relevant work.

Keywords:- *"Weather Forecasting," "K-Means Clustering," "Artificial Neural Networks," "Ml," "Comparative Analysis," And "Meteorological Data."*

I. INTRODUCTION

On a mix of meteorological measurements, historical weather information, and meteorologists' expertise, classical weather forecasting is predicated. This process includes gathering and evaluating data from several meteorology sources of information, run computerized weather models, and employing meteorological expertise to assess the models' outputs.

On the other hand, forecasting of weather powered by machine learning makes use of advanced statistical and computational techniques. To forecast the weather, it entails utilizing machine learning algorithms that have been taught from historical weather data. Similar to traditional forecasting methods, machine learning-based systems incorporate meteorological data from many sources such as weather stations, satellites, and radar. They could, however, also make use of peculiar data sources, such as social networking sites or Internet of Things (IoT) devices.

[1] The use of predictive machine learning techniques for weather forecasting is covered in this paper. In particular, the research investigates how to make use of both linear and functional regression approaches to forecast the highest and lowest temperature for seven consecutive days, given meteorological information for the previous two days. The study's dataset includes meteorological information for Stanford, California, spanning the years 2011 through 2015. The article comes to the conclusion that even while professional weather predicting services beat the models, they may do so over longer time frames. Discussion of the study's shortcomings and recommendations for further investigation follow.

In our Paper two Algorithm are being studied 1. K-mean Clustering and 2. Artificial Neural Network (ANN) in weather prediction:

- K-means clustering is a well-liked machine learning approach that is employed for this purpose. K-means clustering refers to the combining of data points that are alike without the usage of predetermined labels. In conclusion, K-means operates in the way outlined below:
  - Initialization of the centroids of the 'K' cluster, either randomly or by a predetermined method.
  - Based on distance, frequently using a Euclidean distance metric, each information point gets assigned to the nearest centroid.
  - By averaging all the data points assigned to each cluster, centroids are recalculated.
  - Repeat steps 2 and 3 as necessary to achieve the appropriate number of iterations or until the centroids hardly move at all.
  - The basic objective of this technique is to generate 'K' different clusters of information points by maximising variance between clusters and minimising variance within clusters. K-means has several uses in picture segmentation, data analysis, and other circumstances where accumulating pertinent data is important.

- An Artificial Neural Network (ANN), a type of computational model, is inspired by how the human brain works. In a word, an ANN is built up of layered networks of linked nodes, or "neurons," that are coupled to one another. These layers commonly include an input layer, one or more layers that are hidden, and a layer for output.

An ANN's core components are as follows:

- Neurons: These act as computing components that process inputs, carry out a weighted summation, and then send the output of the result through an activation function.
- Weights: A weight is assigned to each connection between neurons, indicating the strength of the relationship. Through exercise, these weights are changed.

ISSN No:-2456-2165

- A network can model complex interactions within data thanks to the non-linearity introduced by an activation function.

An ANN's main goal is to learn from data by minimizing prediction errors in classification or regression tasks by adjusting the weights during training. ANNs are used in a variety of artificial intelligence and machine learning fields, such as image identification, natural language processing, autonomous driving, and many more.

II. OBECTIVE

To do a comparative analysis of K-Means Algorithm and ANN in Weather Forecasting.

III. RELATED WORK

There are other Machine Learning algorithm aswell which are also used for weather prediction:

- Regression Models: Linear and polynomial regression are effective methods for identifying relationships between meteorological variables. By examining historical data together with variables like humidity, air pressure, and time of day, they help predict temperature, for example.
- Time Series Analysis: Time series analysis methods, such as ARIMA (AutoRegressive Integrated Moving Average), are used in the field of weather forecasting to predict and foresee trends, seasonality, and irregular variations in weather-related data.
- Random ForestRandom Forest is an ensemble learning approach that excels at handling regression as well as classification issues when used for weather forecasting. It may be used to forecast various meteorological conditions, such as rainfall or the possibility of storms.
- Support Vector Machines (SVM): Based on historical data and input features, SVMs are used for classification tasks like predicting the likelihood of various weather events (like rain or snow).
- Decision Trees: Decision trees can be used to categorize weather events and make decisions related to them, such as determining if inclement weather at an airport would cause aircraft delays.
- Ensemble Methods: To increase forecasting accuracy, multiple machine learning models are combined using techniques like AdaBoost and Gradient Boosting.
- Models for numerical weather prediction (NWP) The complicated mathematical and physical models used in NWP models, which are not ordinary machine learning techniques, recreate atmospheric phenomena. They play a significant role in weather forecasting.
- Deep Learning for Image Analysis: Convolutional Neural Networks are used for image analysis in the field of weather forecasting, particularly the recognition of cloud cover in satellite photos.

IV. METHODS

- It is uncommon to use K-Means clustering directly to improve weather forecasting accuracy. Instead, it is frequently used as a part of a larger data analysis or preprocessing pipeline, which could indirectly enhance weather forecasting models. The following elements are included in K-Means clustering's application to weather forecasting:

- Data Preprocessing: K-Means clustering is used in data preprocessing to find patterns or clusters in historical meteorological data. The data can be efficiently reduced in dimension or translated into categorical variables that indicate different weather regimes or conditions by grouping together comparable weather patterns. These data clusters can then act as features for more traditional weather forecasting models.
- Pattern Identification: A useful approach for locating repeating weather patterns or anomalies is K-Means clustering. It can be used, for instance, to find trends in historical data on temperature, humidity, or air pressure. Meteorologists' decision-making is improved and their understanding of the current meteorological conditions is deepened by their ability to recognize these patterns.
- Compression of data: K-Means clustering can sometimes be used to efficiently compress large amounts of weather data into a more manageable size. This is very useful for efficient data transmission and storage, especially in remote areas or with limited resources.
- Feature Engineering: The use of clustering outcomes as features in more intricate machine learning models, such as random forests, decision trees, or neural networks, enhances weather forecasting. When analyzing the data using the raw data, it might be challenging to spot hidden links or trends in the data. These qualities may help you do just that.
- Data Quality Control: K-Means clustering can be used to identify irregularities or anomalies in meteorological data, assisting in the detection of measurement errors or inconsistencies in the dataset. Data quality assurance. This procedure is important for improving the precision of the data used as input for prediction models.

While numerical weather prediction (NWP) mathematical models, data assimilation techniques, and the addition of different sources of data, like data from satellites, weather observatories, and radar, are the main ways to improve forecast accuracy, K-Means clustering can nonetheless indirectly improve weather forecasting.

[2] The simulations presented in the research are built using data from 2009 and 2010. The simulation is carried out to determine the precision of the strategy suggested in the study. The experiment is carried out using Java and the Weka software on a PC with a 2.26 GHz Core i3 processor and 4GB of memory running Windows 7 Home Basic. Any method's accuracy can be evaluated by contrasting its present value with its real value. The simulation outcomes demonstrate the excellent accuracy rate of the suggested incremental K-means clustering method for weather forecasting. By contrasting the actual weather with the forecasted weather, the accuracy of the

strategy is evaluated. As a result of the findings, it can be concluded that the methodology for forecasting weather is highly accurate. The suggested approach can be strengthened even more by utilizing more sophisticated methodologies and algorithms, according to the paper. The creation of more sophisticated weather forecasting algorithms and methods, which can be utilized to increase the suggested methodology's accuracy, is part of the planned work's future scope.

The accuracy of the suggested approach for forecasting the weather using incremental K-mean clustering was calculated to be around 83.3%. By comparing the true worth with the present value of the new approach, the accuracy was calculated. The simulation results demonstrate the high accuracy rate of the suggested methodology, which makes it a trustworthy tool for weather forecasting.

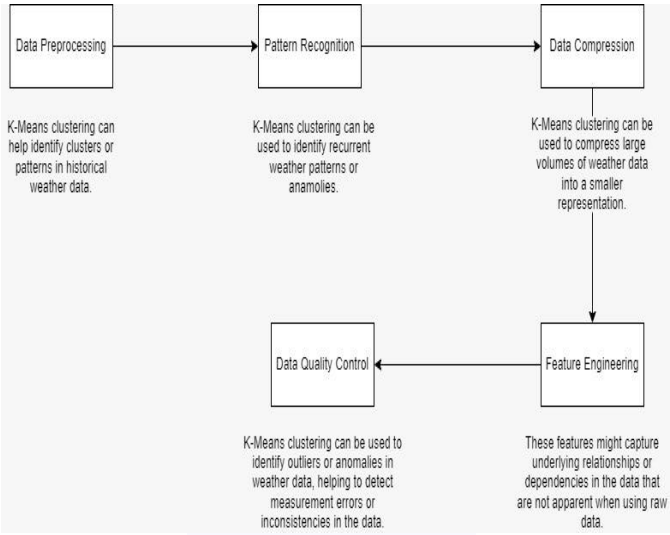

Fig 1 K-Means Flow Chart

- Artificial Neural Networks (ANNs) have been applied to weather prediction tasks Although they are only one component of complex weather forecasting models, artificial neural networks (ANNs) find uses in weather prediction tasks. The following is an outline of how ANNs are used in weather prediction:
- Data Input: The historical weather data used to train ANNs includes elements like temperature, humidity, wind speed, atmospheric pressure, and more. These informational components act as the neural network's input features.
- Output: Depending on the precise task at hand, the ANN's output in the context of weather prediction may vary. It could entail predicting future values for variables like temperature, precipitation, wind patterns, or even categorizing weather as "rainy" or "sunny."
- Training: The Artificial Neural Network (ANN) is trained using historical data as well as observed weather conditions. The main goal is to optimize the weights and biases of the network so that it can produce accurate predictions based on the input data.
- Model Complexity: Weather forecasting is fundamentally complex since it takes into account a wide range of variables, including atmospheric dynamics, geographic characteristics, and other elements. To properly handle this complexity, ANNs can be incorporated into a thorough ensemble of models or act as a component inside a broader weather forecasting system.
- Integration: By combining several models and data inputs, a comprehensive weather forecasting system can be created using the forecasts generated by an artificial neural network (ANN). The use of ANNs as one tool among many in a varied toolkit is a common ensemble method in modern weather forecasting.
- Real-time Data: Current information from weather stations, satellites, and other sources is essential for accurate weather forecasting. ANNs are able to continuously examine new data and update their predictions in light of the most recent information.

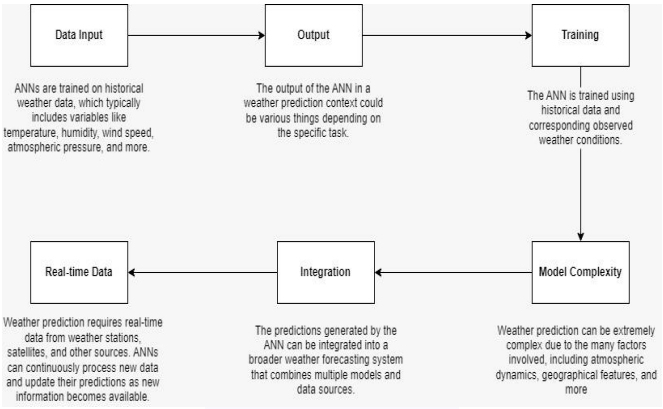

Fig 2 ANN Flow Chart

[4] The use of a back propagation neural network system to forecast rainfall in India is discussed in this paper. According to the authors, 52% of all jobs in India are in the agricultural sector, making it the country's most common occupation. However, there are insufficient irrigation infrastructure, since only 52.6% of the area was irrigated in 2009–2010. As a result, farmers continue to rely largely on rainfall, especially during the monsoon season. Therefore, it is essential for agriculture planning and management to estimate rainfall accurately.

For data mining, a form of supervised machine learning method is used: the back propagation neural network model. In this instance, moisture, dew point, and atmospheric pressure data are used to train the model in order to forecast rainfall. Using 250 patterns for training and 120 testing patterns, the authors divided their data in half for training and the other third for testing. They claim to have achieved testing accuracy of 94.28% and training accuracy of 99.79%.

[3] The usage of artificial neural networks, or (ANN) for weather forecasting is discussed in the paper. It emphasizes that owing to the irregular nature of weather, conventional techniques of forecasting the weather have limits and that ANN can offer more precise forecasts by learning from prior data and modifying its parameters. The PDF also includes examples of research on temperature, a thunderstorm, rain, and weather in particular places. These works are all examples of research on predicting weather using ANN. The PDF comes to the conclusion that ANN is an effective tool for weather prediction and may be used in place of conventional meteorological methods.

ISSN No:-2456-2165

[5] They have presented a temperature forecast model that makes use of advanced wireless data gathering techniques. This model mixes artificial neural networks with the strength of statistical software. An artificial neural network integrates smoothly with wireless technologies and statistical software to evaluate data, acquire insights from them, and create accurate temperature projections for the future.

[6] By applying the Levenberg Marquardt Back-Propagating Feed Forward Neural Network, it was suggested to take a novel method to classifying and predicting weather. The Levenberg BP is by far the most effective backpropagation algorithm available. The Weather Classification & Prediction system built around BPNN fundamentally functions as a predictive toolbox, designed to capture data such as humidity, temperature, pressure, and the direction of the wind - these variables function as input neurons to the BP neural network. For the purpose of training the neural network, both historical and present atmospheric data are gathered.

[7] They developed a neural network-based algorithm to predict temperature and used the neural network software package, which provides a number of training or learning options. The Back Propagating neural network technique (BPN) was one of these methods used. Utilizing real-time data, the researchers assessed this idea and contrasted the results with projections made by the meteorological department. The outcomes show that the model has potential for making accurate temperature predictions.

[8] It has been noted that major modes of the complex climate factors, in particular, also affect Australian rainfall. To improve our comprehension and forecasting abilities, there have been few attempts to thoroughly evaluate the combined impact of these indexes on rainfall. Given the complexity of rain as an atmospheric phenomena, traditional linear approaches might not be able to adequately capture its complex features. Using Artificial Neural Networks, or ANN, as a technique, this study aims to identify a non-linear link between rain in Victoria and the lag indices which affect the area. The results show that when using the lag indices for springtime rainfall predictions, ANN modeling produces much better correlations than conventional linear techniques. Incorporating these indicators into an ANN model results in a striking improvement in model correlation, with numbers for all three case study stations in Victoria, Australia, situated in Horsham, Melbourne, Australia, and Orbost, respectively, reaching 99%, 98%, and 43%.

[9] This Paper is a survey article on rainfall prediction using artificial neural networks. It discusses the challenges and techniques involved in predicting rainfall accurately. The article covers various neural network architectures such as MLP, BPN, RBFN, SOM, and SVM, and compares them with other forecasting techniques such as statistical and numerical methods. The survey concludes that the forecasting techniques that use MLP, BPN, RBFN, SOM, and SVM are suitable to predict rainfall than other forecasting techniques. The article also provides extensive references in support of the different developments of ANN research.

[10] In a city catchment in west Sydney, Australia, artificial neural networks, or ANNs, are used to forecast rainfall, as discussed in the paper file. Based on previously known geographical and temporal rainfall patterns, the research analyzes the effectiveness of three different types of artificial neural networks (ANN in forecasting rainfall at numerous sites simultaneously over the next 15 minutes. The fundamentals of artificial neural networks (ANN and their construction are also covered in the study, along with the crucial concerns of determining the complexity and lag order of the networks. The authors point out that while none of the networks accurately predicted the peak rainfall rate, three different types of networks gave fair predictions of rain one time step in advance. According to the article, better predictions should be predicted when more data are gathered for model training and more control factors are found to be added to the network inputs. The research comes to the conclusion that the network with smaller lags may have performed marginally better than those with higher lags because there exists an optimum complexity of the network for the issue under consideration, given the data at hand.

V. COMPARISON

Table no. 1: ANN vs K-Means

ARTIFICIAL NEURAL NETWORK	K-MEANS CLUSTERING
-ANNs are like weather forecasters who learn from past weather data to predict future conditions.	K-means is like a tool that helps group similar weather data together.
-They analyze various factors like temperature, humidity, and wind patterns to make predictions.	It doesn't predict the weather itself but can organize historical data into clusters based on similarities.
-ANNs are good at capturing complex patterns in weather data, providing detailed forecasts.	K-means can be used to identify regions with similar weather patterns but doesn't make future weather predictions.

VI. LIMITATIONS

- Its restricted consideration of the temporal dimension is a problem when using K-means clustering in the context of weather forecasting. K-means analyzes fixed parameters like temperature and humidity to group similar weather situations, but it does not sufficiently account for how these conditions change dynamically over time. Since weather is constantly changing, K-means is less useful for forecasting dynamic weather conditions since it struggles to understand the time-dependent intricacies of weather patterns and changes.

- Artificial neural networks' (ANNs') probable inability to effectively handle the complexity of weather systems is a limitation when used for weather forecasting. There are many interconnected components that make up weather, and ANNs might not be able to fully capture all of them. As a result, they might not provide completely accurate

forecasts, especially in cases where weather conditions are harsh or changing quickly.

VII. CONCLUSION

Artificial neural networks, or ANNs, have shown to be effective tools in the field of weather prediction due to their ability to directly predict future weather conditions using existing data. K-means clustering typically performs the task of analyzing and clustering data within meteorological datasets, as opposed to being able to provide weather predictions. Weather forecasting systems typically use a number of strategies, include physical models, statistical methods, and machine learning algorithms like ANNs, in order to provide thorough and reliable forecasts.

REFFERENCES

[1]. Mark Holmstrom, Dylan Liu, Christopher Vo, "Machine Learning Applied to Weather Forecasting", Stanford University, December 15, 2016.
[2]. Sanjay Chakraborty, N.K.Nagwani, Lopamudra Dey, "Weather Forecasting using Incremental K-means Clustering**",** June 18, 2014**.**
[3]. Abhishek Saxena, Neeta Verma, Dr K. C. Tripathi, "A Review Study of Weather Forecasting Using Artificial Neural Network Approach", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 2 Issue 11, November – 2013.
[4]. Enireddy.Vamsidhar, K.V.S.R.P.Varma P.Sankara Rao, Ravikanth satapati, "Prediction of Rainfall Using Back propagation Neural Network Model", International Journal on Computer Science and Engineering, Volume 02, No. 04,pp 1119-1121,2010.
[5]. P.P. Kadu, Prof. K.P. Wagh, Dr. P.N. Chatur, "A Review on Efficient Temperature Prediction System Using Back Propagation Neural Network", International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459), Volume 2, Issue 1, pp 52-55, January 2012
[6]. Arti R. Naik, Prof S. K. Pathan , "Weather Classification and Forecasting using Back Propagation Feed-Forward Neural Network", International Journal of Scientific and Research Publications, Volume 2,Issue 12,Dec 2012
[7]. Ch. Jyosthna Devi, B. Syam Prasad Reddy, K.Vagdhan Kumar, B. Musala Reddy, N.RajaNayak, "ANN Approach for Weather Prediction Using Back Propagation", International Journal of Engineering Trends and Technology, ISSN: 2231-5381, Volume 3, Issue 1, pp 19-23, 2012
[8]. F. Mekanik and M. A. Imteaz, "A Multivariate Artificial Neural Network Approach for Rainfall Forecasting: Case Study of Victoria", Australia, Proceedings of the World Congress on Engineering and Computer Science 2012, Vol I WCECS 2012, October 24- 26, 2012, San Francisco, USA
[9]. Deepak Ranjan Nayak , Amitav Mahapatra, Pranati Mishra, "A Survey on Rainfall Prediction using Artificial Neural Network", International Journal of Computer Applications (0975 – 8887),Volume 72, No.16, ,pp 32 - 40,June 2013.
[10]. Kin C. Luk, J. E. Ball and A. Sharma, "An Application of Artificial Neural Networks for Rainfall Forecasting", Mathematical and Computer Modeling, Volume 33, pp 683-693, 2001.