e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:02/ February -2022 Impact Factor- 6.752 www.irjmets.com

A SURVEY ON WEATHER FORECASTING AND THEIR TECHNIQUES

Dharsan G M*1

*1Student, Department Of Data Science And Business Systems, SRM Institute Of Science And Technology, Chennai, Tamil Nadu, India.

ABSTRACT

Weather forecasting is the process of technology used to predict the atmospheric conditions of a location. Many scientists have also attempted to predict the weather formally and informally for many years. Weather forecasts are done by collecting data of a specified location with different attributes of climate. This paper discusses about the various works done by different researchers on mission learning and deep learning algorithms and compares their performance using the best approach to optimize prediction and precision.

Keywords: Weather Forecast, Algorithms, Machine Learning, Deep Learning.

I. INTRODUCTION

Weather forecasting is the prediction atmospheric conditions such as temperature, humidity, dew point, rainfall, and wind speed on a specified location. Instruments like barometers, thermometer and radars are used in collecting data for prediction such as current weather conditions, weather patterns, tracking motion of air and clouds and previous weather patterns. Weather forecast systems are among the advanced equation system that computer need to solve. The prediction of weather forecasting system helps airport or naval if there is a sudden change in climate, helps farmers in the increase of their crop yield production, even mining people require climatic conditions to monitor the Earth's crust continuously. Floods and droughts can be prevented using accurate weather prediction. Data recorded by a Weather Station which are used for analysis. The combination of linear and nonlinear model is one of the foremost liked and was wide used in forecasting.

Machine Learning

Machine learning is a supervised technique used in data science to predict the dataset. A model is created which used to produce outputs a target value based on individual weights and value for each training variable. In each record, corresponding weights to each variable gives correlation between target and training variable. The training data must be sufficient to determine the best possible weights of all variables. When these weights are learned as accurate as possible, the model can predict the correct output or the target value given on the test data record. Utilizing simple machine learning techniques allows us to predict complex weather models. It has high possibilities in the realm of weather forecasting. Such models can be offered to public as web services easily. Some of the proposed machine learning techniques are Simple Linear Regression(SLR), Multiple Linear Regression(MLR), Random Forest Regression(RFR), Support Vector Machine(SVM).

Deep Learning

The rapid growth of technologies like Internet of Things, Cloud Computing, Wireless Sensor Network has helped Weather forecast enter the Big Data. Deep Learning and visualization techniques have made climate prediction more effective and accurate. Deep Learning techniques are separated as layers of neural networks to identify and extract patterns from the dataset. In order to predict in a very effective way several forecasting models using deep learning have been proposed.

The weather forecasting systems can be classified into three categories based on statistical models, Artificial intelligence models and hybrid models.

The major contributions of this paper are presented below:

1. A comprehensive review of various weather forecasting models.
2. A precise classification of weather forecasting models.
3. Comparison between various machine learning and deep learning models
4. Potential research directions in this area.

II. FORECASTING FRAMEWORK

The framework of a weather prediction model is as shown in Fig 1. The framework of weather prediction includes data acquisition, data preprocessing, model selection, and training, evaluation and visualization of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:02/ February -2022 Impact Factor- 6.752 www.irjmets.com

results. paper size. This data consists of both useful and useless information, and they are in the form of unstructured data. Once the data is gathered, the first step is to preprocess the data to remove missing values and to clean the data. Data preprocessing try to improve the quality of the input data. At this stage data is prepared for training. The inconsistencies, noise, missing values are removed by this process. Real world data contains many missing values. It can be caused due to values that are noted and due to data corruption. As each value is validated missing values can cause error to the model, so these missing values are replaced with mean values in most cases.

The data integration process merges data from different sources into a data store. Data reduction is used to reduce the size of the data by eliminating unwanted features or clustering. In the data transformation step, the data are normalized using standardization technique to convert suitable for processing. Data transformation techniques is also useful in reducing the training time of the models. Preprocessing is not mutually exclusive, hence they may function together. Once the pre-processing is done a suitable model is selected, trained and tested using the datasets. For effective forecasting of weather information appropriate algorithms are to be used.

Model selection and training is essential step in any forecasting system. The knowledge about types of forecasting models will help researchers to select a suitable model matching the application domain is selected and trained using the datasets. After completion of training, the model performance is evaluated using statistical quantatives error indicator such as MAE, MAPE, RMSE and R^2. Atlast the results are visualised using suitable plots. Plots like scatter plot, line plots can be used to visualize the results. These plots are used to analyse the difference between actual and predicted values. The difference between the predicted values and actual values can be more precisely viewed in semilog plots.

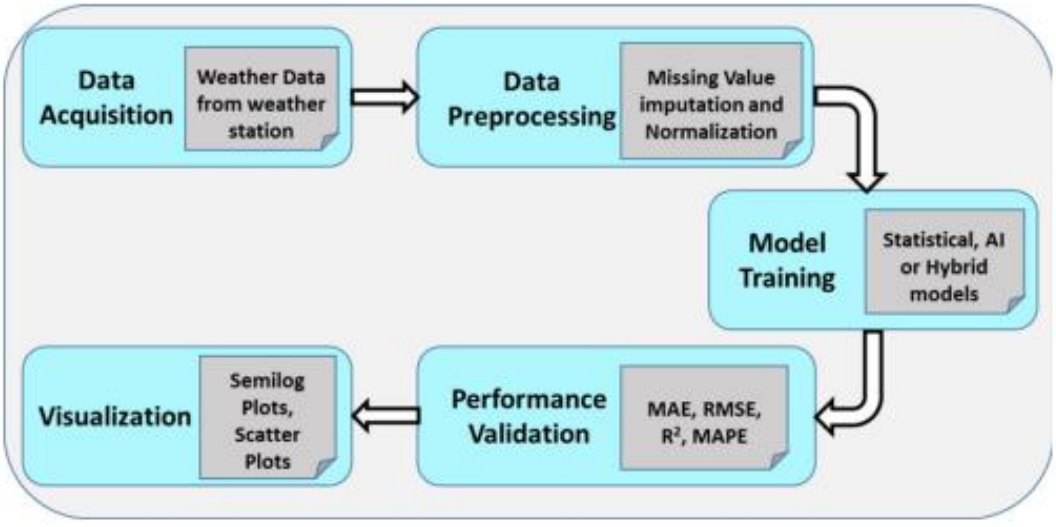

Fig. 1. The layout of the weather forecasting model.

Classification of Models

Based on multiple factors the forecasting systems are classified into various categories. The factor comprises the number of times each step predicted, the number of variables involved, the methodology used and the parameter to be predicted. The classification of forecasting model is shown in Fig. 2. The model can be univariate model or multivariate depending upon the number of variables in forecasting. If the temperature is dependent on other environmental factors, then such type of forecasting can be termed as multivariate. A univariate model depends only on one variable, whereas multivariate predicts the model based on several factors. Apart from this, forecasting model be single-step or multistep, depending on number of time to be predicted. Weather forecasting approaches comes in four different types based on time horizon they are, very short-term, short-term, medium-term and long term. Short- range and very short-range are more accurate than medium or long ranges.

Furthermore, based on methodology employed forecasting systems can be two basic types namely, deterministic and probabilistic. Deterministic methods provide accurate values of weather forecasts for a specific location. This survey focuses on deterministic forecasting models can be further classified as statistical

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:02/ February -2022 Impact Factor- 6.752 www.irjmets.com

models, hybrid models and artificial intelligence models. Depending on the parameter predicted, forecasting can be classified as temperature prediction, wind speed direction, dew point prediction, rainfall prediction models etc. as shown in Fig. 2. With the influence of weather in social and economic activities, bad events can be prevented by predicting weather conditions accurately. In past years several methods have been suggested to deal with the goal. The next section clearly explains about different categories weather forecasting models.

Classification based on methodology

Based on methodology used for forecasting, deterministic model can be classified into three types they are, statistical models, Artificial Intelligence and hybrid models.

1. Statistical models

Statistical model are linear models that can be used for short term weather forecasting. Using past weather data statistical forecasting allows weather prediction. Linear regression is a statistical model, which is used to find linear relationship between dependent and independent variables. The aim of the linear regression is to find the best fit line. Best fit line is a line which covers the most number of points and error are as small as possible. A gradient descent algorithm is used to minimise the error as much as possible. Multiple linear regression is not as same simple linear regression as multiple linear regression has more than one independent arguments. ARMA, ARIMA, Multiple linear regression and VAR are the most commonly used statistical models. ARIMA is statistical model that use time series data to predict future observations. Cadenas et al. [2] VAR model is multivariate form of univariate auto regression model. With the help of univariate ARIMA and multivariate NARX model a wind speed forecast was predicted.

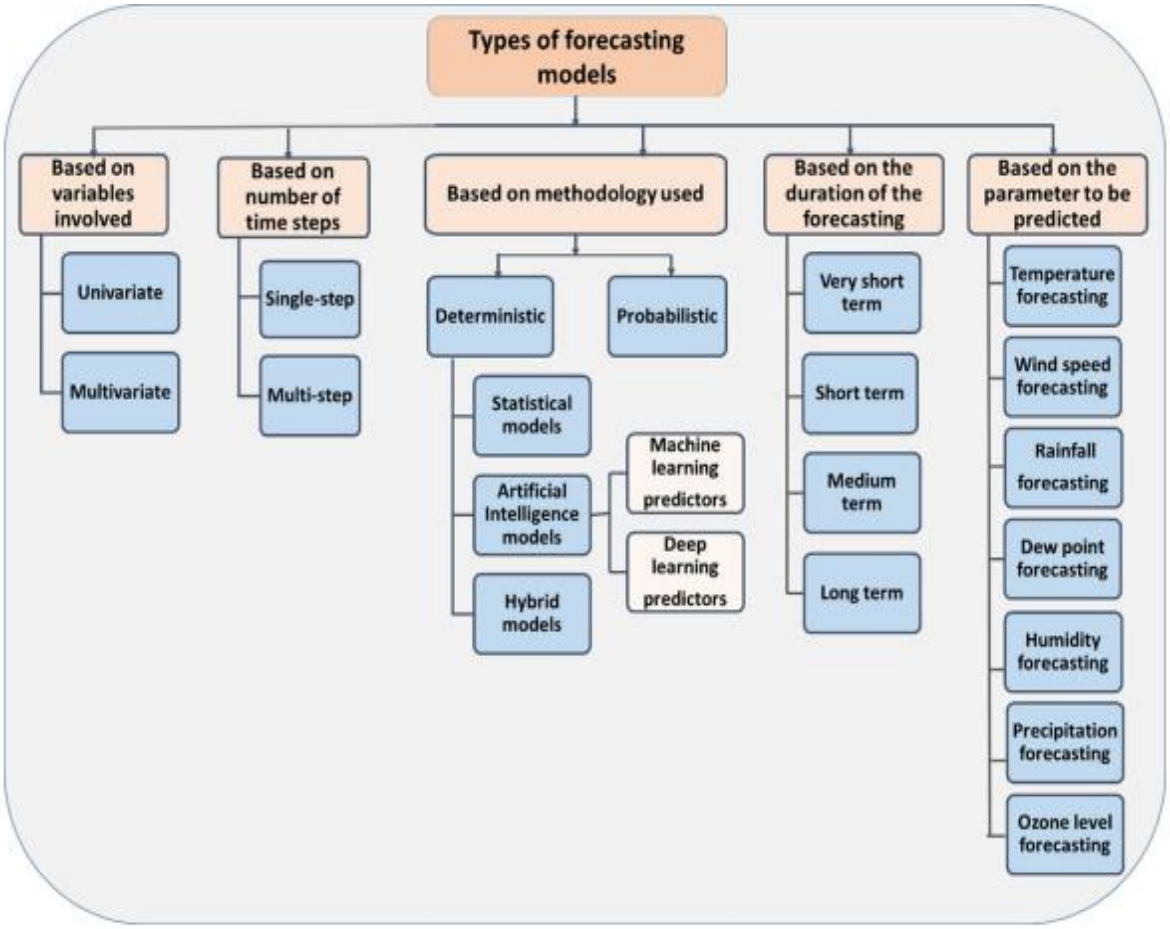

Fig 2: Classification of models

2. Artificial Intelligence models

The improvement of artificial intelligence has inspired the growth self-learning models. These model are more effective compared to statistical models. AI models deals with nonlinear data effectively and produce better

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:02/ February -2022 Impact Factor- 6.752 www.irjmets.com

forecasting performance. These models are further divided into machine learning predictors and deep learning predictors.

2.1. Machine learning predictors

For handling nonlinear datasets machine learning and deep learning are better models. SVM, ANN, ELM and Random Forests are some of the popular machine learning predictors. These predictors can also be used weather forecasting. This section provides the various learning techniques of machine learning. They include ML models, ANN based models and SVM based models.

2.1.1. ANN based models

Neural Networks are supervised and predictive. They are one of the most popular techniques for weather prediction because they can obtain relationships of past weather trends and future weather conditions. It functions as a nonlinear regression models with fast information processing capability and develop mapping of input and output variables. An Artificial Neural Network comprises a network of neurons known as nodes, which are connected to each other. The architecture of ANN of single hidden layer is shown in fig. 3. The neural network comprises of input layer, hidden layer and output layer. The node in the input layer transforms the given input and transfers it to the next node. Each node evaluates the value it obtains and then transforms the results according to the activation function. The values transformed flow from input layer through the hidden layers until the output layers are reached. In this section, various prediction models are developed using ANN.

Liu et al. [7] presented a temperature forecasting model based on neural network and studied the importance of intrusion in neural networks. Kashiwao et al. [3] observed that ANN predicts more accurate results to predict the probability of rainfall.

2.1.2. SVM based models

SVM is a supervised machine learning algorithm that can be used for both classification and regression problems. SVM is defined by hyper plane that has minimum distance to the training data of samples. An optimal hyper plane has the maximum margin of training data. Using the kernel trick SVM can find nonlinear solutions and they also work well with high dimensional data. Kernels are mathematical functions that converts inputs to the desired format. The various functions solved by SVM algorithms are linear, nonlinear, polynomial, and sigmoid.

2.1.3. Deep Learning Predictors

Deep learning is section of machine learning algorithm, where it uses neural approach to gain intelligence and from huge datasets, and it make use of supervised or unsupervised learning in deep architectures to learn hierarchical representations. Deep learning has been successfully implemented in many areas such as speech recognition, time series and genomics. Deep learning techniques are more reliable for time series prediction because of their ability to learn temporal dependence present in time series data. Big companies like Microsoft, IBM, Google and Facebook are doing active research in this

III. RELATED WORKS

Senlin zhu et al. [7] Improved support vector regression is being used to estimate the temperature of water. The water temperature of huge high-altitude reservoirs in western China is analysed and evaluated using a model for predicting reservoir water temperature that incorporates solar radiation. The support vector regression (SVR) model is built using the measured water temperature data in the reservoir over many years, and the genetic algorithm (GA) is utilized to refine the parameters, resulting in a better support vector machine (M-GASVR). SVR, ANN, GA-SVR, and M-GASVR models are all evaluated using root-mean-square error, mean absolute error, mean absolute percentage error, and Nash–Sutcliffe efficiency coefficient. The findings demonstrate that the ANN model is effective.

Sindhu P. Menon et al. [5]. The goal of the study is to demonstrate the impact of an Urban Heat Island (UHI) by using temperature as an independent variable and pollution and population as dependent factors. Time Series analysis is used to determine the trend in temperature, population, and pollution, while Multiple Linear Regression is used to predict temperature. By comparing anticipated and observed values for the years 2013-2016, the accuracy of the predicted values can be seen.

Shen Rong et al. [10] The purpose of this article is to look at how temperature affects the selling of iced products. Data on previous year's anticipated temperature and iced product sales are collected, followed by

data compilation and purification. Data mining theory is used to create a mathematical regression analysis model based on cleansed data. The process of studying the relationship between the independent and dependent variables is known as regression analysis. The linear regression model is utilized in this research to analyze and anticipate the sale of iced products based on temperature variations, which will give a foundation for the company.

TanviPatil et al. [8] The system's goal is to forecast the weather for a specific time period. Different types of variables are used to characterize the weather at any given time. Only significant attributes are used in the weather forecast procedure out of all of these attributes. The traits you choose are heavily influenced by the location you've chosen. The existing weather condition attributes are used to fit a model, and future fluctuations in the attributes are examined using machine learning techniques and extrapolating the data. The linear and logistic regression models are employed in this system.

Pinki sagar et al. [12] This work shows a prediction technique based on linear equations for forecasting future sequence trends for time series data sets, as well as humidity prediction based on temperature measured every 10 minutes. Using linear regression-based methods, this paper also examines error rates during prediction. This paper discusses two techniques, one for one-dimensional stream data and the other for two-dimensional stream data. Both algorithms are used to analyse mistakes in time series multivariate data sets.

Anusha et al. [1] The weather state is predicted in this article based on the supplied attributes. The majority of present systems rely on statistical methodologies such as Support Vector Machine (SVM), which are unable to provide accurate predictions since they do not account for quick changes in weather conditions. The suggested method employs the concept of multi-linear regression, which has the potential to outperform existing methods.

Ramesh et al. [9] The minimum and maximum temperatures were predicted. At the location of Chennai, India, numerical weather parameters are used to create seven-day minimum and maximum temperature forecast models using multiple linear regression. With the best R-Square and lowest MAE, RMSE in an independent test dataset, regression-based minimum temperature prediction models give superior accuracy than maximum temperature forecast models. The analysis also points us that for both minimum and maximum temperature, the forecast ability is good at shorter lead days and gradually declines as lead days increase.

IV. CONCLUSION

As a whole, the sole purpose of this study is to learn about the different techniques used in prediction of weather forecast. This study provides insight about weather forecast and different methods involved to predict them. It is summarized and concluded that hybrid techniques have optimized results compared to other techniques. Hence, choosing the optimized technique and applying them for yielding better performance is an important task and a challenging risk.

V. REFERENCES

[1] N. Anusha, M. Sai Chaitanya, Guru Jithendranath Reddy (2019). Weather Prediction using Multi Linear Regression algorithm. IOP Conf. Series: Materials Science and Engineering 590 012034.

[2] Cadenas, E. Rivera, W. Campos-Amezccua, R. Heard, "Wind speed prediction using a univariate ARIMA model and multivariate NARX model", Wnergies 9(2), 109.

[3] Kashiwao, K. Sarigul, M. Avci, M. Kandirmaz, T. Nakayama, K. Ando, S. Ikeda, K. Lee, M. Bahadori, A neural network based local rainfall prediction system using meteorological data on the internet: a case study using data from Japan meteorological agency.Appl. Soft Comput. 56, 317-330 ,

[4] Sahil Ratra (2020). Weather Prediction using Machine Learning Techniques.

[5] Sindhu P. Menon, Ramith Bharadwaj, Pooja Shetty, Prajwal Sanu, Sai Nagendra (2017). Prediction of Temperature using Linear Regression. International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT).

[6] Liu, X., Zhang, C., Liu, P., Yan, M., Wang, B., Zhang, J. and Higgs, R., 2018. Application of temperature prediction based on neural network in intrusion detection of IoT. Security and Communication Networks, (2018).

[7] Senlin Zhu, Emmanuel Karlo Nyarko, Marijana Hadzima-Nyarko (2018). Modelling daily water temperature from air temperature for the Missouri River. DOI 10.7717/peerj.4894.

[8] TanviPatil, Kamal shah (2021). Weather Forecasting Analysis using Linear and Logistic Regression Algorithm. International Research Journal of Engineering and Technology (IRJET)

[9] K. Ramesh, R. Anitha, Selva Gopal (2014). Linear Regression Based Lead Seven Day Maximum and Minimum Air Temperature Prediction in Chennai. Research Journal of Applied Sciences, Engineering and Technology 7(11): 2306-2310.

[10] Shen Rong, Zhang Bao wen (2018). The research of regression model in machine learning field. MATEC Web of Conferences 176, 01033.

[11] A. H. M. Jakaria, Md Mosharaf hossain, Mohammad Ashiqur Rahman (2020). Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee.

[12] Pinki Sagar, Prinima Gupta, Indu Kashyap (2019). Prediction of Humidity Depending on Temperature with Low Error Rate using regression model. International Journal of Engineering and Advanced Technology (IJEAT).