IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-59

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁂ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11573

Survey Study on Rain Prediction System

[1] Rushali Jakkan, [2] Vaishnavi Chavan, [3] Nikita Chavan, [4]Prof. Sunita Vani

[1,2,3,4]Department of Information Technology,

G.H. Raisoni Institute of Engineering and Technology, Pune, India

Abstract: Due to its complexity and durability, precipitation prediction has recently gained the highest research relevance. Among other applications, such as flood forecasting and pollutant concentration monitoring. Existing model uses a complex statistical model that is often too expensive for both calculations and budgets. It does not apply to downstream applications. Therefore, an approach using machine learning algorithms. It is being studied in combination with time series data as an alternative to overcome these shortcomings. To this end, this study presents a comparative analysis based on a simplified precipitation estimation model. Efficient traditional machine learning algorithms and deep learning architectures for this Downstream application. This paper presents a time-series method called Neuralprophet for predictions of Maharashtra's ten most popular cities. This method provides an estimate of rainfall using different atmospheric parameters like average temperature and cloud cover to predict the rainfall. The main advantage of this model is that this model estimates the rainfall based on the previous correlation between the different atmospheric parameters. Thus, an estimated value of what the rainfall could be at a given period and place can be found easily. Introducing Neural Prophet, the successor to Facebook Prophet, which sets the industry standard for a explainable, scalable, and easy-to-use prediction framework. NeuralProphet is a hybrid prediction framework based on PyTorch and trained using standard deep learning techniques, so developers can easily extend the framework. Local context is introduced in autoregressive and covariate modules that can be configured as classical linear regression or neural networks. It includes traditional statistical and neural network models for time series modeling used for forecasting and anomaly detection. This model produces high-quality predictions of time series data showing multiple seasonality's with linear or non-linear growth. Use this model to predict future temperatures in Maharashtra's most popular cities using historical temperature data from the same location.

INTRODUCTION

Precipitation is one of the most influential meteorological parameters in many aspects of our daily lives. The socio-economic impact of rainfall is significant, from damage to infrastructure during floods to disruption of transportation networks. Floods and similar extreme events are the result of climate change and are expected to occur more frequently and will have catastrophic consequences in the coming years. Interestingly, recent studies reveal that weather conditions can potentially increase air pollution in winter and summer, another important topic discussed alongside recent climate change. It has become. It is important to repeat that increased air pollution leads to health problems such as asthma and similar problems. Therefore, as a mitigation approach, many research have investigated and proposed precipitation prediction strategies in case of all contingencies. However, with the intention to enhance human mobility and enhance agricultural and business development, those methods want to offer green and well timed forecasts.

Another applicable component that contributed to the expanded use of ANN as an technique to predicting precipitation calls for very little know-how of the capacity to cope with the non-linearity of precipitation records and the relationships among the variables beneathneath consideration.

Therefore, the use of artificial neural networks (ANN) as a precipitation forecast model has attracted the attention of researchers. Another applicable component that contributed to the expanded use of ANN as an technique to predicting precipitation calls for very little know-how of the capacity to cope with the non-linearity of precipitation records and the relationships among the variables beneath Neath consideration. That is. However, predicting precipitation is even more difficult because it is often hampered by spatial and temporal variations in precipitation in the region. With that in mind, ANN variants, known as recurrent neural networks (RNNs), are best suited to address these types of challenges. While the task of predicting precipitation can be addressed with mathematical models, the use of different types of ANNs has emerged as an alternative that allows the development of lowcomplexity predictive approaches. However, like other machine learning applications, there is no such thing as a "free lunch". Therefore, the use and performance of predictive models depends on various design decisions, such as: B. Available data, predictive task objectives / approaches, and model implementation.

DOI: 10.17148/IJARCCE.2022.11573

Therefore, this study aims to investigate the suitability of Neuralprophet for the task of daily precipitation prediction compared to modern machine learning approaches. AutoML performs intelligent searches of various machine learning algorithms in a reasonable amount of time. The result is a more powerful regression model and its hyperparameters for predicting precipitation for a particular dataset. Used to get the value. Neural Prophet, a
time series model built on top of
ARNet and Facebook Prophet. This is an improved version of Facebook Prophet. Use the PyTorch framework as the backend. It's beginner friendly and you can start with Quick Install with a pip. This is a combination of traditional neural networks and statistical modeling to model the time series used to predict and detect anomalies. This model produces high quality predictions of seasonal time series data with linear or non-linear growth rates.

LITERATURE REVIEW

Thirumalai, Chandrasegar, and colleagues discuss the quantity of rainfall received in previous years based on crop seasons and forecast rainfall for future years.
Rahi, Khabif, and Zaid are the three crop seasons. For early prediction, the linear regression approach is used. In this case, the variables rahi and Khabif were used as variables, and if one was supplied, the other could be predicted using linear regression. The standard deviation and mean were also determined in order to forecast crop seasons in the future. Farmers will be able to use this technology to determine which crops to harvest based on crop seasons. A. Geetha and G. M. Nasira have developed a model that anticipates weather situations such as rainfall, fog, thunderstorms, and cyclones, allowing individuals to take precautionary precautions. The decision trees were modelled using data mining techniques and a data mining programme called Rapid miner.
Trivandrum data set including parameters such as day, temperature, dew point, pressure, and so on. A decision tree algorithm is used to partition the dataset into training and testing sets. The precision is determined, and the actual and projected values are compared. The accuracy is 80.67, and it can be improved by using soft computing techniques such as fuzzy logic and evolutionary algorithms to get a higher value. The authors, Parmar, Aakash, Kinjal Mistree, and Mithila Sompura, explore the various methods for predicting rainfall for weather forecasting and their limitations.
Various neural networks algorithms used for prediction are discussed in detail, along with their steps. This categorises various approaches and algorithms used for rainfall prediction by various researchers in today`s era.
Finally, the paper's conclusion is presented.
Background research on some machine learning models, such as the ARIMA Model, Artificial neural networks, and types such as BackPropagation Neural Network Cascade Forward Back Propagation Network Layer Recurrent Network, SelfOrganizing Map, and Support Vector Machine, has been completed. Collected, surveyed, and table presents categorization of various rainfall prediction approaches. Dash, Yajnaseni, Saroji K. Mishra, and Bijaya K. Panigrahi have used artificial neural networks (ANN), extreme learning machines (ELM), and K-nearest neighbors to predict summer monsoons and post-monsoon precipitation. We used artificial intelligence technology such as KNN).
The dataset used is the time series data from the Indian Institute of Technology in Kerala from 1871 to 2016 (IITM).
The data is then preprocessed and normalized before being split into training and test groups. We used the data from 2010 as a training set and the data from 2011 to 2016 as a test set.
The above algorithms were used and their performance was calculated using MAE, RMSE, and MASE. Compared to others, the ELM algorithm provided more accurate results.
Many machine learning algorithms are used to predict precipitation, according to Singh, Gurpreet, and Deepak Kumar, with two methods, random forest and gradient boosting, and many machine learning methods such as AdaBoost and K-nearest neighbours. in conjunction with A combined approach is used. (KNN), Vector Machine (SVM), and Neural Network (NN) are all supported algorithms. These are for precipitation data from North Carolina from 2007 to 2017, and the performance is calculated using various indicators such as Fscore, Precision, Fit, and Recall. Finally, eight hybrid models were announced, with the best of them, the gradient boosting AdaBoost, producing excellent results. Prerika Sanghvi, Kar Kaveri, Neelima Thakur, and Kar Kaveri used a fuzzy logic approach to predict precipitation based on temperature data collected at various locations. The researchers employed the fuzzy model. Due to the inaccuracy of predictions caused by other climatic factors, we examined the advantages of the fuzzy system over other methods, taking into account other influencing factors such as humidity. Sardesh Pande, Kaushik D, and Vijaya R. Thool used artificial neural networks, backpropagation (BPNN), radial basis functions (RBFNN), and generalised regression (GRNN) to

DOI: 10.17148/IJARCCE.2022.11573

analyse precipitation data from India`s Nanded region. Maharashtra was taken into account, and the data was normalised between 0 and 1 Algorithms were used, and their performance was calculated and compared. BPNN and RBFNN outperformed GRNN. These methods focus on nonlinear machine learning methods such as gradient boosting decision tree models and deep neural networks for short-term precipitation predictions. Use AUC-calculated classifiers to evaluate performance. Correlates with F1 score, accuracy and accuracy, and RMSE regression index. The DNN algorithm outperformed the EC data. Moon, SeungHyun, and colleagues [9] "built" the early warning system. EWS`. A signal will be sent if the alert level is reached three hours ago. This was accomplished using machine learning techniques. Data from South Korea were collected between 2007 and 2012, performance was measured against multiple criteria, and confounding matrices were created.

PROPOSED SYSTEM

In this project, we are mainly focused on predicting the exact climate so that we can determine if the rain going to happen or not. For this goal, we are going to follow the following step to achieve our goal.

The precipitation forecasting model is employed. The first step is to convert data into the appropriate format for conducting experiments, which is followed by extensive data analysis and observation of variations in rainfall patterns. We predict rainfall by splitting the dataset into training and testing sets, then applying various machine learning approaches (MLR, SVR, etc.) and statistical techniques to compare and draw conclusions about the various approaches used. We try to minimise the error using a variety of approaches. Because the dataset is large, feature reduction is used to improve accuracy while also reducing computation time and storage. Principal Component Analysis is a technique for extracting necessary variables from a large set of variables. It extracts lowdimensional sets with the goal of capturing as much information as possible. When there are few variables, visualisation becomes more important.

It is accomplished by utilising a covariance matrix and obtaining Eigen values from it.

We reduced the attributes in our dataset using PCA by only considering rainfall data from three consecutive months and annual data from each subdivision. Techniques used include: Multiple Linear Regression Analysis: By fitting an equation to determined data, multiple regressions attempt to model the relationship between two or more variables and a response. Clearly, it is nothing more than an extension of simple regression toward the mean. The general form of a multivariable linear regression model is: y=+1x1+ 2x2+...+ kxk+ where y is the dependent variable and x1, x2,... xk are independent variables, and are coefficients. Multiple regression will model more complicated relationships that result from numerous options along with they should be used in cases where one explicit variable isn't obvious enough to map the link between the independent and also the variable quantity.

Data acquired from Worldweatheronline site using API with request module in python

data cleansing, deleting redundant columns, checking for missing values, feature selection, deciding which features are most significant

The neural prophet accepts only two columns as input and then builds a Neural prophet model for individual parameters (Date and one value column).

Applying the neural prophet model to specific characteristics such as temperature, cloud cover, precipitation, and humidity, and then combining the results

We use SVM Classifier to categorise the type of weather (cloudy, sunny, clear sky, etc.) based on these features, and then present the results.

Data Acquiring:

Step 1: Collect Data

Given the trouble you need to solve, you'll have to research and reap information that you'll use to feed your machine. The best amount of records you get is very critical seeing that it's going to immediately affect how nicely or badly your version will work. You might also additionally have the records in a present database otherwise you ought to create it from scratch. If it's miles a small assignment you may create a spreadsheet so that it will later be without problems

Impact Factor 7.39 ⁒ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11573

exported as a CSV file. It is likewise the not unusual place to apply the net scraping approach to routinely accumulate records from numerous assets together with APIs.
This global website provides official meteorological observations, weather forecasts, and climatological information for specific cities provided by National Meteorological & Hydrological Services (NMHS) around the world. NMHS conducts official meteorological observations in each country. Links to official weather service websites and tourist offices/organizations are also provided whenever available. In this version, a weather icon is displayed next to the written forecast for easy visual confirmation.

As of April 2022, WWIS provides official weather information for 3443 cities, 3299 cities are available with forecasts from 139 members, and 2185 cities are climatology from 170 members. It is available for information. We welcome suggestions for enriching the content of this website From this site, we fetch data for our system using the request module of python.

Data preprocessing:
Then we preprocessed the data, cleans data by removing unnecessary columns from the fetched data, and then check for missing values in our dataset. After applying all the above criteria we selected features and then we prioritized which features are important To achieve our goal.

building a Neural prophet model for individual parameters:

NeuralProphet is a python library for modeling time-collection records primarily based totally on neural networks. It's constructed on the pinnacle of PyTorch and is closely stimulated with the aid of using Facebook Prophet and AR-Net libraries. NeuralProphet vs. Prophet From the library name, you could ask what's the primary distinction between Facebook's Prophet library and NeuralProphet. According to NeuralProphet's documentation, the brought capabilities are[:

- Using PyTorch's Gradient Descent optimization engine making the modeling procedure a lot quicker than Prophet
- Using AR-Net for modeling time-collection autocorrelation (aka serial correlation)
- Custom losses and metrics
- Feed-ahead neural networks with customizable non-linear layers.

After implementing the neural profit model, classify the type of weather. The following are the different types of weather that we categorise: The sky could be cloudy, sunny, or clear, for example

Train our system :
You'll want to train the datasets to run quickly and notice an incremental improvement in the prediction rate. Remember to randomly initialise the weights of your version - the weights are the values that multiply or have an effect on the relationships between the inputs and outputs - so that you can be routinely adjusted via the chosen set of rules the extra you teach them.
Finally, we have completed our output.

COMPARATIVE ANALYSIS
A forecast is a calculation or estimation of future occurrences, particularly financial or weather trends. Forecasting had been highly useful as a foundation for developing any action or policy prior to face any events till this year. In the tropics, for example, where several countries only have two seasons per year (dry and rainy), many countries, particularly those that rely heavily on agricultural commodities, will need to forecast rainfall in order to determine the best time to begin planting their crops and maximise their harvest. Another example is how a corporation can utilise forecasting to estimate raw material price movements and devise the best plan for maximising profit. The performance of the time recording strategies considered (ARMA mode, ANN and ANN techniques) is first analyzed and compared in terms of their ability to predict the spatially averaged precipitation of validation set storms. The output precipitation forecasts can be passed through a conceptual precipitation-spill conversion version in sequence, and the resulting flow forecasts can be analyzed and compared within the next session. The performance of various predictive strategies is examined using a trial and error

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⅔ Vol. 11, Issue 5, May 2022
DOI: 10.17148/IJARCCE.2022.11573

approach. Analyzing the results of precipitation predictions is very difficult and a number of quantitative measurements (root mean square error, mean absolute error, persistence factor, coefficient of performance, correlation coefficient) to integrate the effectiveness of overall performance. , Concordance index) was taken into account. In contrast, of the lead instance. The various strategies provided excellent results for exceptional predictive overall performance metrics and exceptional read instances, so it was not possible to clearly assess the overall performance type. After various precipitation prediction studies (French et al., 1992; Kuligowski and Barros, 1998), Correlation Coefficient (CC) was selected as the most effective advisor to assess the overall performance of precipitation prediction. it was done. The correlation coefficient is calculated by dividing the covariance of the predictions and the observations by the covariance created from their respective general deviations. The range is 21.0 to 1.0, and the higher the value, the higher the degree of matching.

The split-pattern and adaptive applications are described.

1.1. ARMA fashions: trial-and-error methods Low-order ARMA fashions were used in the split-pattern calibration, each in basic terms auto-regressive and with a moving-common component, provided that the sum of the autoregressive and moving-common orders, p 1 q; is equal to or less than 6. All of the examined ARMA models delivered nearly identical outcomes (imply CC 0–293). The examined ARMA fashions in the adaptive calibration had auto-regressive orders of one and a pair of moving-common orders less than or equal to three. For all of the ARMA fashions, the fashion of the acquired performances is characterised by way of tremendously top overall performance for a lead-time of one h, observed by way of a fall apart in correspondence with longer time horizons. The typical first-rate outcomes are provided by using ARMA models with simple configurations, indicating ARMA(1, zero) as the best appearing version (imply CC 0.281).

1.2. ANN:
Strategic Trial and Error Split Starting at 2-24, the sample application tested an architecture with multiple input node NIs. For each dimension of the input layer, hidden node (NH) diversity frequently increases from 2 to 8 nodes. The validation set's basic overall performance degradation, which indicates overfitting, is generally seen in small hidden layer dimensions with good results, such as 2-6 NH. The overall basic performance of the ANN architecture with regard to overall lead time. It is excellent because it expands the number of input nodes (NI) and increases the gain with the addition of 15-18 nodes (see Figure 2a). The networks tested as part of the adaptive calibration were very sparse due to the limited number of training samples (each NI and NH start at 2-5) (this is 100 times earlier). At ARMA, observations determined an honest assessment). The adaptive approach) can easily create complex networks. The maximum green network complexity of adaptively calibrated neural networks appears to correspond to 3 NI and NH input and hidden nodes.

1.3. Nearest neighbors:
the trial-and-error procedure A trial-and-error procedure was used for several nearest neighbours, K, ranging from 5 to 100, and a feature vector dimension, d, ranging from 2 to 12. The improvement in performance with increasing number of nearest neighbours is less noticeable for more than 20 neighbours, and increasing K beyond 70 results in no marginal improvement in overall performance (see Fig. 2b). For each given number of neighbour vectors K, small values of the feature dimension d (from 2 to 4) produced the best results.

CONCLUSION

Precipitation is one of nature's most important phenomena, affecting not only humans but all other organisms. Our study sought to develop a Neural Prophet and SVM Classifier prediction system capable of accurately and efficiently forecasting annual precipitation with minimal error. Finally, we analysed and predicted annual precipitation using a Nueral Prophet algorithm. Precipitation is one of nature's most important phenomena, affecting not only humans but all other organisms.

REFERENCES

1. S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, “An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives,” Expert Systems with Applications, vol. 85, pp. 169–181, 2017.View at: Publisher Site | Google Scholar
2. J.-H. Seo, Y. H. Lee, and Y.-H. Kim, “Feature selection for very short-term heavy rainfall prediction using evolutionary computation,” Advances in Meteorology, vol. 2014, 15 pages, 2014.View at: Publisher Site | Google Scholar

DOI: 10.17148/IJARCCE.2022.11573

3. L. Wu, Y. Fan, X. He, and G. Yu, "Change trend of extreme precipitation events in the Wuhan city (China) from 1951 to 2012," in International Conference on Energy, Environment and Materials Engineering, Shenzhen, Guangdong, China, February 2014.View at: Google Scholar

4. M. Li, C. Chen, and J. Jiang, "Features of flood risk map drawing for downtown areas of Wuhan city and its application," China Water Resources, vol. 5, pp. 49–51, 2017.View at: Google Scholar

5. H. Messer and O. Sendik, "A new approach to precipitation monitoring: a critical survey of existing technologies and challenges," IEEE Signal Processing Magazine, vol. 32, no. 3, pp. 110–122, 2015.View at: Publisher Site | Google Scholar

6. P. Petkov and T. Slavov, "Stochastic modeling of mems inertial sensors," Cybernetics and Information Technologies, vol. 10, no. 2, pp. 31–40, 2010.View at: Google Scholar

7. T. Alexandrov, S. Bianconcini, E. B. Dagum, P. Maass, and T. S. McElroy, "A review of some modern approaches to the problem of trend extraction," Econometric Reviews, vol. 31, no. 6, pp. 593–624, 2012.View at: Publisher Site | Google Scholar

8. R. B. Cleveland, W. S. Cleveland, and I. Terpenning, "STL: a seasonal-trend decomposition procedure based on loess," Journal of Official Statistics, vol. 6, no. 1, p. 3, 1990.View at: Google Scholar

9. L. Zhu, Y. Wang, and Q. Fan, "MODWT-ARMA model for time series prediction," Applied Mathematical Modelling, vol. 38, no. 5-6, pp. 1859–1865, 2014.View at: Publisher Site | Google Scholar

10. P.-S. Yu, T.-C. Yang, S.-Y. Chen, C.-M. Kuo, and H.-W. Tseng, "Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting," Journal of Hydrology, vol. 552, pp. 92–104, 2017.View at: Publisher Site | Google Scholar

11. T. Guo, Z. Xu, X. Yao, H. Chen, K. Aberer, and K. Funaya, "Robust online time series prediction with recurrent neural networks," in 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 816–825, Montreal, QC, Canada, October 2016.View at: Publisher Site | Google Scholar

12. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.View at: Publisher Site | Google Scholar

13. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," in 9th International Conference on Artificial Neural Networks: ICANN '99, Edinburgh, UK, September 1999.View at: Publisher Site | Google Scholar

14. Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," Tech. Rep., 2015, http://arxiv.org/abs/1511.03677.View at Google Scholar

15. X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," Data Mining and Knowledge Discovery, vol. 26, no. 2, pp. 275–309, 2013.View at: Publisher Site | Google Scholar

16. H. Izakian, W. Pedrycz, and I. Jamal, "Fuzzy clustering of time series data using dynamic time warping distance," Engineering Applications of Artificial Intelligence, vol. 39, pp. 235–244, 2015.View at: Publisher Site | Google Scholar

17. C. Chen, D. Gu, and Y. Lu, "Change of climatic seasons between the urban area and suburb in Wuhan from 1960 to 2009," Journal of Meteorology and Environment, vol. 31, no. 1, pp. 96–100, 2015.View at: Google Scholar

18. N. Tax, I. Vranich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," in Advanced Information Systems Engineering. CAiSE 2017, E. Dubois and K. Pohl, Eds., Lecture Notes in Computer Science, pp. 477–492, Springer, Cham, 2017.View at: Publisher Site | Google Scholar

19. S. Han, J. Kang, H. Mao, et al., "ESE: efficient speech recognition engine with sparse LSTM on FPGA," in FPGA '17 Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, pp. 75–84, Monterey, CA, USA, February 2017.View at: Publisher Site | Google Scholar

20. M. Shahrban, J. P. Walker, Q. J. Wang, A. Seed, and P. Steinle, "An evaluation of numerical weather prediction based rainfall forecasts," Hydrological Sciences Journal, vol. 61, no. 15, pp. 2704–2717, 2016.View at: Publisher Site | Google Scholar

21. K. Bandara, C. Bergmeier, and S. Smyl, "Forecasting across time-series databases using long short-term memory networks on groups of similar series," Tech. Rep., 2017, http://arxiv.org/abs/1710.03222.View at Google Scholar