Date	**10.09.2022**
Team ID	**PNT2022TMID25913**
Project Name	**Exploratory Analysis of Rain Fall Data in India for Agriculture**
Maximum Marks	

PAPER 1

TITLE : Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting

AUTHORS: Barrera-Animas, A.Y., Oyedele, L.O., Bilal, M., Akinosho, T.D., Delgado, J.M.D. and Akanbi, L.A.
2022. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, p.100204.

ABSTRACT: Rainfall forecasting has gained utmost research relevance in recent times due to its complexities and persistent applications such as flood forecasting and monitoring of pollutant concentration levels, among others. Existing models use complex statistical models that are often too costly, both computationally and budgetary, or are not applied to downstream applications. Therefore, approaches that use Machine Learning algorithms in conjunction with time-series data are being explored as an alternative to overcome these drawbacks. To this end, this study presents a comparative analysis using simplified rainfall estimation models based on conventional Machine Learning algorithms and Deep Learning architectures that are efficient for these downstream applications. Models based on LSTM, Stacked-LSTM, Bidirectional-LSTM Networks, XGBoost, and an ensemble of Gradient Boosting Regressor, Linear Support Vector Regression, and an Extra-trees Regressor were compared in the task of forecasting hourly rainfall volumes using time-series data. Climate data from 2000 to 2020 from five major cities in the United Kingdom were used. The evaluation metrics of Loss, Root Mean Squared Error, Mean Absolute Error, and Root Mean Squared Logarithmic Error were used to evaluate the models' performance. Results show that a Bidirectional-LSTM Network can be used as a rainfall forecast model with comparable performance to Stacked-LSTM Networks. Among all the models tested,

the Stacked- LSTM Network with two hidden layers and the Bidirectional-LSTM Network performed best. This suggests that models based on LSTM-Networks with fewer hidden layers perform better for

EXISTING TECHNOLOGY:
This paper compares the prediction performance of rainfall forecasting models based on LSTM-Networks architectures and modern Machine Learning algorithms. In this paper, 2 models based on LSTM-Networks, 3 models based on Stacked-LSTM, and 1 Bidirectional-LSTM Networks model are compared with an XGBoost to do an Automated Machine Learning approach.

DISADVANTAGES: LSTM networks take very long to train and consume a lot of memory which in turn makes the process expensive. This makes prediction unaffordable for masses and people in the Agricultural sector to use.

PAPER 2

TITLE : Portable Autonomous Rain Prediction Model Using Machine Learning Algorithm

AUTHORS: Abhyankar, V., Singh, A.G., Paul, P., Mehta, A. and Vidhya, S.
2019, March. Portable Autonomous Rain Prediction Model Using Machine Learning Algorithm. In *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)* (pp. 1-4). IEEE.

ABSTRACT: In this paper we propose to create an Arduino based weather forecasting system using various sensors. The objective of this paper is to present to you a portable weather monitoring and rain forecasting system. Weather refers to the conditions of the atmosphere by measuring day to day temperature changes and describes various attributes of the troposphere [1]. While climate refers to an average over vast ranges of time [2]. With advancement of technology the act of automating weather monitoring is not only possible but also beneficial. This system will be of interest and use to weather enthusiasts and companies or individuals whose work depends on weather conditions. For example this system will be helpful in agriculture for monitoring the weather and giving warnings when the weather conditions become adverse to the planted crops. Some sports depend specifically on the weather and this system can be used for aiding sportsmen who participate in such sports.

EXISTING TECHNOLOGY:
Arduino based weather forecasting system using various sensors along with Machine Learning is used.DHT11 sensor module is used to collect data about temperature and humidity of the

surroundings. LDR is used along with a fixed value resistor in a voltage divider arrangement to get resistance of the LDR which leads us to the light intensity of the surroundings. Logistic regression algorithm is used to calculate rain probability based on current humidity and HLSI values. Logistic regression algorithm will fit the best fitting curve in the dataset and give the current probability of rainfall.

DISADVANTAGES: This paper, despite having a good hardware system , i.e sensors detecting rainfall based on parameters like Humidity and Temperature , lacks any GSM Module or Front end to give real time notifications and messages to users about the weather conditions or rainfall.

PAPER 3

TITLE : Performance Comparison of Data Mining Techniques for Rain Prediction Models in Indonesia

AUTHORS: Muchamad Taufiq Anwar, Wiwien Hadikurniawati, Edy Winarno, Wahyu Widiyatmoko

ABSTRACT: Rain prediction is a crucial topic that continues to gain interest across the globe. Rain has a massive impact on various aspects of human life such as in agriculture, health, transportation, etc, and also some natural disasters. Various impacts of rain on human life prompts us to build a model to understand and predict rain to provide early warning for various use cases in various fields. Previous research on rain modeling using Data Mining (DM) techniques had suffered from low accuracy caused by the limited availability of the training data and their meteorological attributes. This research aims to address those issues by building the rain model using a richer and more abundant rain data in Indonesia. Four DM techniques are used and compared in this research i.e. the C4.5/J48, Random Forest (RF), Naïve Bayes (NB), and Multilayer Perceptron (MLP). The experimental results showed that the MLP and J48 algorithm can provide the best accuracy (up to 78,4%), which is better than previous research. Other key findings in this research include: (a) the selection of DM techniques has little effect on the model accuracy; (b) a larger training dataset generally improves model accuracy and a larger test dataset is necessary to get a representative realworld test accuracy, and (c) the two most influential attributes in rain modeling are the relative humidity and the minimum temperature, and we suggest to include cloud condensation nuclei in the next research to complete the model.

EXISTING TECHNOLOGY: This study compares four DM techniques used for the rain prediction model, i.e. J48, RF, NB, and MLP. Results showed that MLP and J48 algorithms can provide the best accuracy (up to 78,4%) compared to other algorithms, although the difference is small.

DISADVANTAGES: Disadvantages of MLP include too many parameters because it is fully connected. Parameter number = width x depth x height. Each node is connected to another in a very dense web — resulting in redundancy and inefficiency.

PAPER 4

TITLE : Annual and Non-Monsoon Rainfall Prediction Modeling Using SVR-MLP: An Empirical Study From Odisha

AUTHORS: XIAOBO ZHANG, SACHI NANDAN MOHANTY, AJAYA KUMAR PARIDA, SUBHENDU KUMAR PANI, BIN DONG, XIAOCHUN CHENG

ABSTRACT: Rainfall is a natural demolishing phenomenon. On the other side, it also serves as a major source of water when conserved through proper channels. For this issue, estimation of rainfall is of utmost importance. The present study employed rainfall forecasting in annual as well as non-moon sessions in Odisha (India). The total annual rainfall and relative humidity data were collected from 1991-2015 from the Department of Forest and Environment Govt. of Odisha. Support Vector Regression and Multilayer perceptron implemented for prediction of maximum rainfall in annual and non-monsoon sessions. Input parameters like average temperature in a month, wind velocity, humidity, and cloud cover was considered for predicting rainfall in a non-monsoon session. The performance of the results was measured with MSE (mean squared error), correlation coefficient, coefficient of efficiency and MAE (mean absolute error). The results of SVR were compared to those of MLP and simple regression techniques. MLP being a computationally intensive method, SVR could be used as an efficient alternative for runoff and sediment yield prediction under comparable accuracy in predictions.SVR-MLP may be used as a promising alternative forecasting tool for higher accuracy in forecasting and better generalization ability.

EXISTING TECHNOLOGY: The motivational factor in this study is to traditional optimization methods, which emphasize accurate and exact computation for the global optimum of a continuous function and avoid being trapped into one of the local optima, but may fall down on achieving the global optimum. In this study SVR-MLP was implemented for achieving global optima.

DISADVANTAGES: SVR algorithm is not suitable for large data sets. It does not perform very well when the data set has more noise i.e. target classes are overlapping. In cases where the number of features for each data point exceeds the number of training data samples, the SVR will underperform

PAPER 5

TITLE : Evaluation of Rain Attenuation Models in Satellite Links under Tropical and Equatorial Climates

AUTHORS: W. D. Dias, M. Carleti, S. S. L. Moreira, L. L. Mendes

ABSTRACT: Attenuation induced by rain is the most relevant propagation effect in satellite communications at frequencies above 10 GHz. Recent research suggests that classical and widely accepted prediction models, such as ITU-R P.618-11 and Crane, do not perform well in the intense rain regime found in tropical regions. Over the last decade, several models formulated specifically for equatorial and tropical climate were proposed. This paper summarizes some of the new rain attenuation models, describing its parameters and applications. The models were tested comparing predicted attenuation values to measurements available at the ITU propagation data bank (ITU DBSG3), taken from 10 locations over the Brazilian territory. The prediction error evaluation was carried out according to the ITU Recommendation P.311. The evaluation results show that some of those prediction methods can provide better accuracy than ITUR P.618-11 for climates found in Brazil. The use of these models satellite link budget planning can lead to a target reliability while avoiding the over dimensioned systems.

EXISTING TECHNOLOGY:

The main objective of this article is to describe the formulation of the latest proposed rain attenuation models in the literature and compare the adherence of these models with the measurement data obtained in different locations in the Brazilian territory. In this way, it is possible to determine the model that best fits the realities of Brazil. The main application of these models is to allow the definition of the power margin necessary for the level of signal reception does not fall below a pre-set value for a percentage of time arbitrarily small.

DISADVANTAGES:

The absorption and dispersion of the electromagnetic wave are the main physical phenomena responsible for the attenuation caused by rain. However, due to the random nature of this meteorological phenomenon, in an Earth-space link, it is only possible to predict the long-term statistical behavior deadline.

PAPER 6

TITLE : Dynamical Short-Term Prediction of Rain Attenuation in the W Band : A Time-Series Model With Simpler Structure and Higher Accuracy

AUTHORS: Bin Xue, Ningning Tong, Xin Xu, and Xingyu He

ABSTRACT: Most existing rain attenuation prediction models are tested only at 55 GHz. Few W-band models have been tested, and these models present problems associated with high complexity and cannot predict with high accuracy in the short term. A dynamical short-term prediction method is proposed here, and it has a simpler structure and higher accuracy than previous rain attenuation models and can theoretically be implemented at any station and frequency. The method uses the relationship among time series to establish an autoregression integrated moving average (ARIMA) model, conducts a stationary test, transforms the nonstationary series into a stationary series, and estimates the parameters. First, we compared the ARIMA(1,1,6) models under the conditions of different polarizations, prediction intervals, and time-series lengths to identify the optimized ARIMA(1,1,6). Then, the optimized model was compared with the International Telecommunication Union Radiocommunication Sector (ITU-R) and Silva Mello (vertical polarization, 0.1 GHz, 50 time series) models. Finally, the forecasts and simulated series were compared. The results show that the prediction error among the three models does not exceed ,10-3 and the forecasts are consistent with the simulation results. Thus,

the model can be applied to forecast rain attenuation in the W band, and it has the advantages of a simple structure and high prediction accuracy.

EXISTING TECHNOLOGY:
This article proposed a dynamical rain attenuation short-term prediction model in the W band based on the nonstationary time-series ARIMA model.The results show that the polarization modes have little effect on the model parameters, which can subsequently be neglected.The proposed model showed good reliability and stability and can be applied to forecast rain attenuation in the W band.

DISADVANTAGES:
The Potential cons of using ARIMA models are Computationally expensive and Poorer performance for long term forecasts.Also cannot be used for seasonal time series. It is less explainable than exponential smoothing and Difficult to predict turning points.

PAPER 7

TITLE : Machine learning techniques to predict daily rainfall amount

AUTHORS: Liyew, C.M. and Melese, H.A.,
2021. Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, *8*(1), pp.1-11.

ABSTRACT: Predicting the amount of daily rainfall improves agricultural productivity and secures food and water supply to keep citizens healthy. To predict rainfall, several types of research have been conducted using data mining and machine learning techniques of different countries' environmental datasets. An erratic rainfall distribution in the country affects the agriculture on which the economy of the country depends on. Wise use of rainfall water should be planned and practiced in the country to minimize the problem of the drought and flood occurring in the country. The main objective of this study is to identify the relevant atmospheric features that cause rainfall and predict the intensity of daily rainfall using machine learning techniques. The Pearson correlation technique was used to select relevant environmental variables which were used as an input for the machine learning model. The dataset was collected from the local meteorological office at Bahir Dar City, Ethiopia to measure the performance of three machine learning techniques (Multivariate Linear Regression, Random Forest, and Extreme Gradient Boost). Root mean squared error and Mean absolute Error methods were used to

measure the performance of the machine learning model. The result of the study revealed that the Extreme Gradient Boosting machine learning algorithm performed better than others.

EXISTING TECHNOLOGY: In this paper, the rainfall was predicted using a machine learning technique. Three machine learning algorithms such as Multivariate Linear Regression (MLR), Random Forest (RF), and gradient descent XGBoost were analyzed which took input variables having moderately and strongly related environmental variables with rainfall. The better machine learning algorithm was identified and reported based on the performance measure using RMSE and MAE

DISADVANTAGES: XGBoost does not perform so well on sparse and unstructured data. A common thing often forgotten is that Gradient Boosting is very sensitive to outliers since every classifier is forced to fix the errors in the predecessor learners. The overall method is hardly scalable.