INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

ISSN
INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

Monsoon Prediction Using Linear Regression and Neural Network Model Using Scikit Libraries

Gopi S[1], Deepak B[2], Sethu B[3]

Associate Professor, Department of IT, Panimalar Engineering College, Chennai, India[1]

B. Tech Student, Department of IT, Panimalar Engineering College, Chennai, India[2,3]

ABSTRACT: Rainfall is the amount of precipitation that causes the water vapours in the atmosphere to condense and pour down as rain to the Earth. This is extremely useful and satisfies multiple needs of human, flora and fauna. Many types of water needs are mostly only quenched by rainfall as almost 91% water present in the Earth as water bodies are alkaline whereas rainfall is a freshwater source. So, it becomes quite important to predict the rainfall in a particular area so that necessary facilities can be made ready to preserve and utilise the fresh resource and satisfy our needs without facing much scarcity. This paper discusses the prediction of rainfall in various regions of India based on the dataset obtained over the course of 1901 - 2015. Linear regression algorithm is the main technique used here in terms of machine learning and artificial intelligence. This method is responsible for the prediction of rainfall over a region over a period of time.

KEYWORDS: Monsoon,Matplotlib, deep learning

I. INTRODUCTION

Water is at the core of sustainable development and is critical for socio-economic development, healthy ecosystems and for human survival itself. It is vital for reducing the global burden of disease and improving the health, welfare and productivity of populations. Water is also at the heart of adaptation to climate change, serving as the crucial link between the climate system, human society and the environment. Without proper water governance, there is likely to be increased competition for water between sectors and an escalation of water crises of various kinds, triggering emergencies in a range of water-dependent sectors. The physical world of water is closely bound up with the socio-political world, with water often a key factor in managing risks such as famine, epidemics, inequalities and political instability. Water is complex because it is linked to almost everything in the world. But complexity should not hinder understanding: Water is a precondition for human existence and for the sustainability of the planet. According to WHO i.e. World Health Organisation, every year the demand of water is quite increasing at least by 5% and hence it creates a lot of demand with lesser supply. Thus, there must be harvesting happening in order to obtain the maximum judicial utilisation of resources and hence satisfy our needs. And also, according to science, rainwater is a great source of freshwater and does not cause any harm to consume and is highly portable with some conditions. Statistics in India say that 66,000 litres of rainwater is collected every year and harvested. This data is based on the assumption that only 60% of the water is harvested and the rest being not harvested. So, it is quite important to predict the rainfall and also aim at harvesting the rest 40% so that there is enough water to battle scarcity, provided that the rest percentage is not very less in number. For this prediction, we need some data and a model to perform the process and hence reduce wastage of water thereby helping the nation in battling the water problems in near future or also in the long term. This can be achieved using AI and ML i.e Artificial Intelligence and Machine Learning where the algorithms can be used effectively to predict the rainfall precisely using the dataset obtained over the years. Let us discuss the data science part in the subsequent paras.

II. LITERATURE SURVEY

TITLE:A Data-Driven Approach for Accurate Rainfall Prediction by Shilpa Manandhar; Soumyabrata Dev; Yee Hui Lee; Yu Song Meng; Stefan Winkler.

In recent years, there has been growing interest in using precipitable water vapor (PWV) derived from global positioning system (GPS) signal delays to predict rainfall. However, the occurrence of rainfall is dependent on a myriad of atmospheric parameters. This paper proposes a systematic approach to analyze various parameters that affect precipitation in the atmosphere. Different ground-based weather features such as Temperature, Relative Humidity, Dew Point, Solar Radiation, PWV along with Seasonal and Diurnal variables are identified, and a detailed feature correlation study is presented. While all features play a significant role in rainfall classification, only a few of them, such as PWV, Solar Radiation, Seasonal, and Diurnal features, stand out for rainfall prediction. Based on these findings, an optimum set of features are used in a data-driven machine learning algorithm for rainfall prediction. The experimental evaluation using a 4-year (2012-2015) database shows a true detection rate of 80.4%, a false alarm rate of 20.3%, and an overall accuracy of 79.6%. Compared to the existing literature, our method significantly reduces the false alarm rates.

TITLE: Accuracy Enhancement Using Machine Learning and Forecasting Techniques by Urmay Shah; Sanjay Garg; Neha Sisodiya; Nitant Dube; Shashikant Sharma.

The paper is focused to provide the insights of climate to the clients from various businesses, e.g. agriculturists, researchers etc., to comprehend the significance of changes in climate and atmosphere parameters like precipitation, temperature, humidity etc. Precipitation estimate is one of the critical investigations in the field of meteorological research. In order to predict precipitation, an endeavor is made to a couple of factual procedures and machine learning techniques to forecast and estimate meteorological parameters. For experimentation purpose daily observations were considered. The accuracy assessment of forecasting model experimentation is done using validation of results with ground truth. The experimentation demonstrates that for forecasting meteorological parameters ARIMA and Neural Network works best, and best classification accuracy in comparison to other machine learning algorithms for forecasting precipitation for next season was given by the Random Forest model.

TITLE:Portable Autonomous Rain Prediction Model Using Machine Learning Algorithm by Varad Abhyankar; Aksh Gurnain Singh; Poulami Paul; Aryan Mehta; S. Vidhya

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. While climate refers to an average over vast ranges of time .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.

III. PROPOSED SYSTEM

In the proposed solution, we can provide a more precise prediction with quick response by performing model pre-processing. This can be used in normal systems and low specification also without loss of quality in performance in terms of accuracy. The prediction is not biased with recent climatic conditions as outliers are removed here and only important points are considered by data visualisation.

Proposed Architecturre
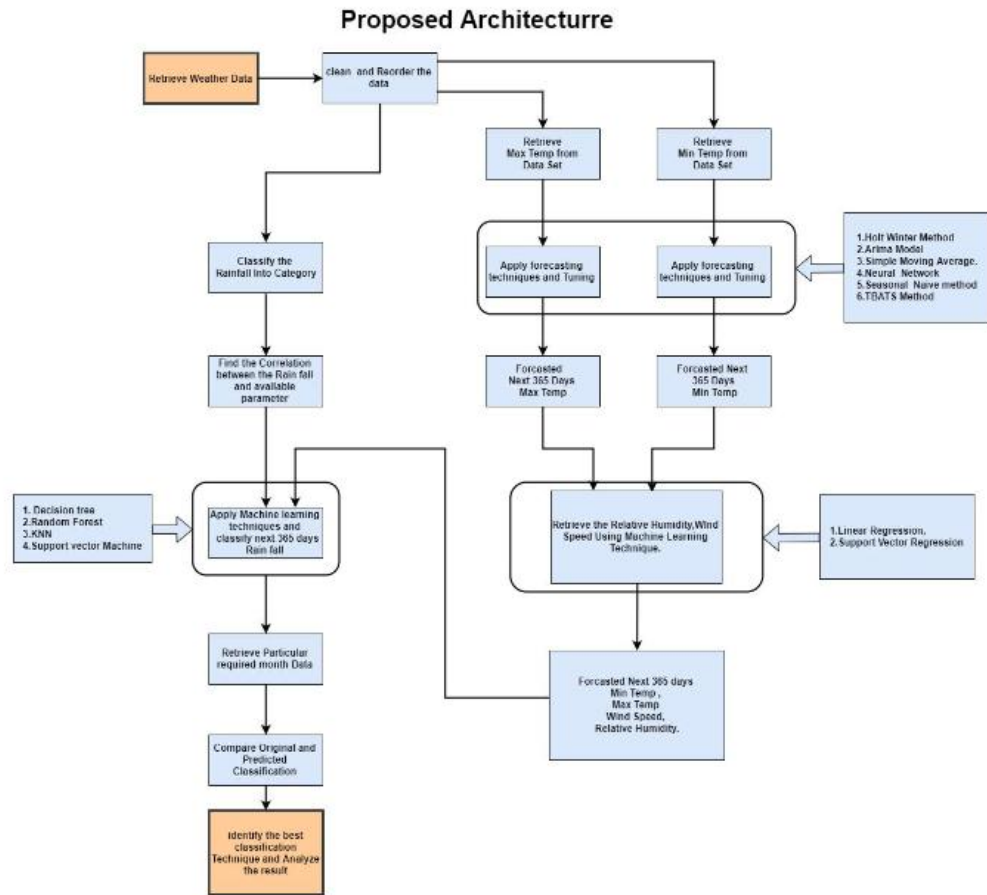
IV. IMPLEMENTATION

Dataset loading is the first process involved here in given the proposed solution. Here the data is initially obtained from the Kaggle platform and then further used. The data is available as csv i.e. comma separated values and hence can be loaded. CSV is a simple file format used to store tabular data, such as a spreadsheet or database. Files in the CSV format can be imported to and exported from programs that store data in tables, such as Microsoft Excel or OpenOffice Calc. Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. The dataset taken has 4188 rows x 19 columns and has subdivision values of rainfall in all states of India during the period of 1901 - 2015.The data obtained might contain some empty spaces and hence can be leading to some unwanted deviations during the process. This must be changed and hence accuracy can improve greatly if this is done. This process involves a lot of packages as present in the software that helps in the printing and checking of the data and hence derive a conclusion from that. Let us explore some of the important packages used here to get clarity. Data pre-processing is an important step in the data mining process. As we know that our use case is based on the supervised learning part, we hence choose to use any of such models namely Linear Regression in our case. It is the most well-known and popular algorithm in machine learning and statistics. This model will assume a linear relationship between the input and the output variable. It is represented in the form of a linear equation which has a set of inputs and a predictive output. Then it will estimate the values of coefficients used in the representation. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: −100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), and missing values, etc. Analysing data that has not been carefully screened for such problems can produce misleading results. In programming, an algorithm is a process or set of rules to be followed in order to achieve a particular goal. An algorithm is characterized by its running time (run-time), whether in terms of space or time. As data scientists, we are interested in the most efficient algorithm so that we can optimize our workflow. Predictive analytics is the process of using data analytics to make predictions based on data. This process uses data along with

analysis, statistics, and machine learning techniques to create a predictive model for forecasting future events. For example, suppose the marketing manager needs to predict how much a given customer will spend during a sale at his company. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, python, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib.

V. CONCLUSION

Thus, we have hence built a model in data science using the old dataset and therefore we have obtained the best possible accurate result so as to obtain a clear prediction of rainfall in any of the sub divisions in India. The result can be used in terms of increasing the ways to foresee the rainfall based on past events, harvest water and hence save for the future. This can reduce water scarcity and also help in the betterment of the consumption of resources without wasting it. This also can help the government in a variety of ways in serving its people if implemented right.

REFERENCES

[1] Shilpa Manandhar, Soumyabrata Dev,Yee Hui Lee,Yu Song Meng,Stefan Winkler.,"A Data-Driven Approach for Accurate Rainfall Prediction",IEEE Transactions on Geoscience and Remote Sensing, Vol.57, Issue 11 Nov2019.

[2] Urmay Shah,Sanjay Garg, Neha Sisodiya, Nitant Dube,Shashikant Sharma.,"Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques",2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC),Issue27 June2019.

[3] Varad Abhyankar, Aksh Gurnain Singh,Poulami Paul,Aryan Mehta, S. Vidhya.,"Portable Autonomous Rain Prediction Model Using Machine Learning Algorithm" 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) ,Issue 14 Nov 2019.

[4] Ashneel Chand,Ravneil Nand.,"Rainfall prediction using Artificial Neural Network in the South Pacific region"2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE),Issue 07 August 2020.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor:
7.488

ISSN
INTERNATIONAL STANDARD SERIAL NUMBER
INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462   6381 907 438   ijircce@gmail.com

www.ijircce.com

Scan to save the contact details