Research Article

Rainfall Prediction with Machine Learning Algorithms

K Vivek #1, V Akhil #2, K Ganesh #3, K Nithish reddy #4, P Vamsi #5, S Kumar Swamy #6

#1 Asst. Professor, #2,3,4,5,6 Students

Department of Computer Science Engineering

QIS College of Engineering and Technology

Abstract

Predicting when and how much rain will fall is a difficult and unpredictable process that has far-reaching consequences for human civilization. Predictions that are both timely and accurate may be used to proactively reduce casualties and property damage. This research provides a series of experiments that employ popular machine learning methods to construct models that predict whether or not it will rain the next day in major Australian cities based on meteorological data for that day. Modelling inputs, modelling approaches, and pre-processing procedures are the focal points of this comparative analysis. The findings compare and contrast the performance of different machine learning methods in making accurate weather predictions using a variety of assessment measures.

Introduction

Farming is crucial to India's economy. Rainfall is crucial to agricultural success. It's also useful for conserving water. If farmers have access to historical rainfall data, they can better manage their crops and contribute to national economic development. Precipitation forecasting helps to avoid floods, which may save lives and valuables. Predicting rainfall is difficult for meteorologists due to the variability in both the time and intensity of precipitation. Researchers from many different domains, including meteorological data mining, environmental machine learning, functional hydrology, and numerical forecasting, face a significant hurdle when attempting to develop an effective rainfall prediction model: predicting. How to deduce the previous forecasts and apply the future predictions is a recurring challenge in these situations. The main process in rainfall is often made up of many smaller processes. Precipitation forecasts based only on the worldwide system have shown to be unreliable in the past. Out of all the advantages and services offered by the meteorological department, climate forecasting stands out as the most useful for nations all over the world. The task is very difficult since it requires precise calculations and all signals are implied without guarantee. Hydrological scientists have made precise precipitation forecasting a top priority since timely warning of severe weather may reduce the number of casualties and property damage caused by natural disasters. In order to better predict future precipitation, scientists have lately shown increased interest in the modular model theory and the integration of many models. In India, you may choose from a wide variety of methods for predicting rain. Rainfall predictions may be made using one of two main approaches in India. Many different kinds of computations are employed for weather forecasting, but the most common ones include regression, ANN, decision tree algorithm, fuzzy logic, and group process of data processing. Following information norms and relationships to acquire intangible and perhaps costly knowledge is the primary objective. The field is vast, but artificial NN has great promise.

Governments, businesses, risk management organisations, and scientists alike continue to be interested in improving our ability to forecast rainfall. Rainfall is a climatic component that influences a wide range of human endeavours [1], including farming, building, electricity generation, forestry, and tourism. Therefore, accurate rainfall forecasting is crucial, since it is the most strongly correlated variable with such unfavourable natural phenomena as landslides, floods, mass movements, and avalanches. For years, society has felt the effects of these occurrences [2]. Preventative and mitigating actions may be taken against natural disasters if an accurate method of rainfall prediction is available.

We employed a number of machine learning methods and models to resolve this ambiguity and provide reliable forecasts. The purpose of this work is to give a complete machine learning life cycle, beginning with data preparation and ending with model implementation and evaluation. Imputing missing values, transforming features, encoding category features, scaling features, and selecting features are all preprocessing operations for data. Logistic Regression, Decision Tree, K-Nearest Neighbour, Rule-based, and Ensembles are just few of the models we used. As a means of judgement.

The dataset considered in this article is comprised of daily weather reports from a wide variety of weather stations around Australia. The intended value of the raintomorrow variable is "did it rain the following day?" Clearly, we need a yes or no. Definitions are derived from

Literature Survey

The fundamental objective of this research is to analyse the various techniques proposed by the authors and create a real-time rainfall forecast system that fixes the problems with the currently available systems. The Udupi district in the Indian state of Karnataka is included in the system [1] that forecasts precipitation. Backpropagation in a cascade of FFNNs is employed. When compared to BPNN, the network's accuracy improves. Long-term precipitation forecasts using this method may not be reliable.

The structure [2] The meteorological variables of maximum and lowest temperature, relative humidity, wind speed, and wind direction were employed in the ANN model developed by G. Geetha and R. Selvaraj to estimate monthly rainfall over the Chennai area. Based on their findings, they made weekly rainfall forecasts for parts of Chennai. ANN-based prediction is more accurate than the standard multiple linear regression method. This algorithm has a forward pass and a backward pass that both function. The data is sent to the forward layer, where it is processed and sent on over the network to the subsequent layers. After evaluating the output of the preceding layer, the final result is generated at the backword layer. In [3], authors suggested a deep mining KNN-based rainfall forecast system. Unlabeled data may be categorised with the use of a single K value, which is then used to calculate the total number of closest neighbours. By grouping datasets with similar properties into the same kind of cluster, we may use KNN to establish their classification. Training for classification or regression is unnecessary with this technique. If the wrong value of K is used, this system may not provide satisfactory results.

Existing System

Accurately forecasting rainfall is crucial because extreme precipitation may trigger a wide variety of catastrophes. Predictions should be reliable in order to encourage individuals to take preventative action. There is both short-term and long-term precipitation forecasting. The most reliable predictions are those made for the near future. The primary difficulty is developing a model for forecasting rain over an extended period of time. Predictions of extreme rainfall, which have important implications for human life and the economy, might be inaccurate.

Disadvantages

Predicting future precipitation only requires looking back at past years' worth of data. Classification and regression are only two of the numerous methods that may be used, and errors between the actual and predicted values, as well as accuracy, can be calculated. Since various methods provide varying degrees of precision, it is crucial to choose the appropriate approach and describe it appropriately.

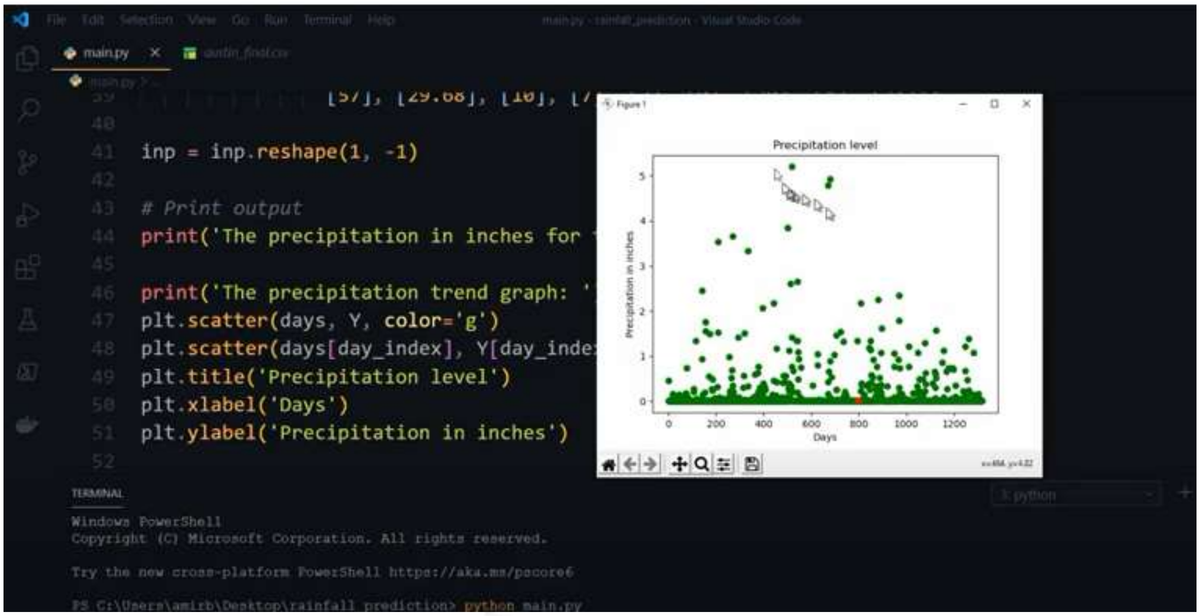
Proposed System

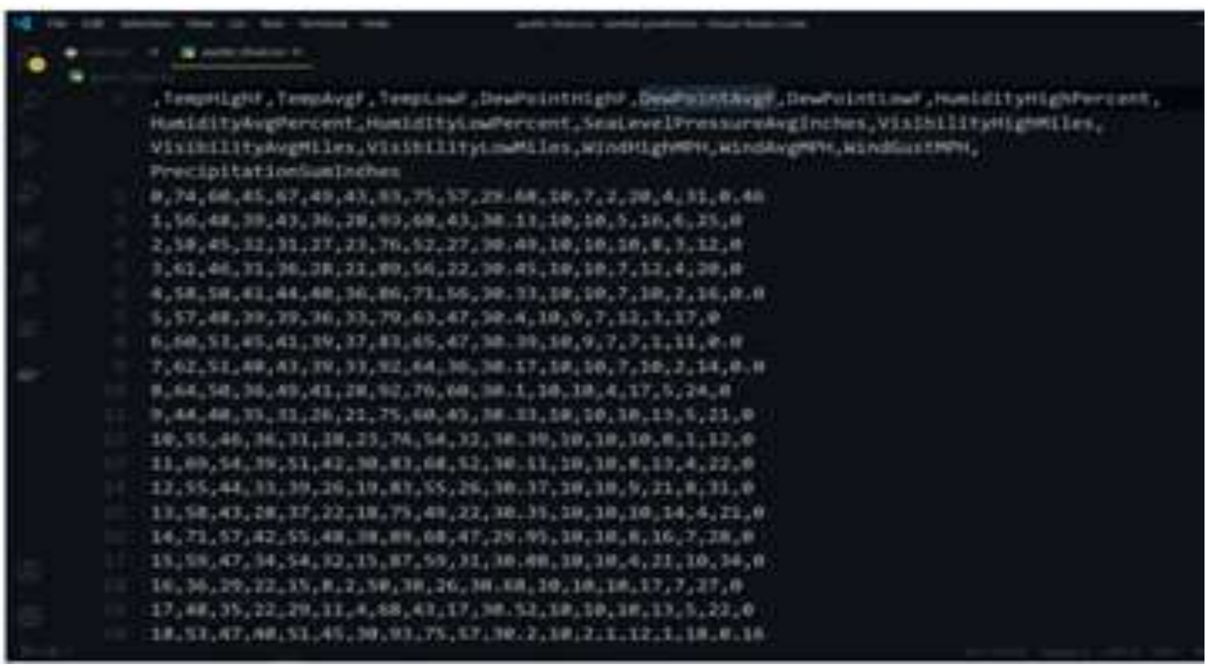
It's a contributing factor to the yearly occurrence of natural catastrophes like flooding and drought. Countries like India, whose economy relies heavily on agriculture, place a premium on accurate rainfall predictions. Due to the ever-changing characteristics of the atmosphere, conventional methods of calculating precipitation provide insufficient results. Regression might be used in machine learning for rain forecasting. The goal of this project is to give a streamlined overview of the methods and approaches used in the field of precipitation prediction and a comparative analysis of the available machine learning methods for those who aren't specialists in the field.

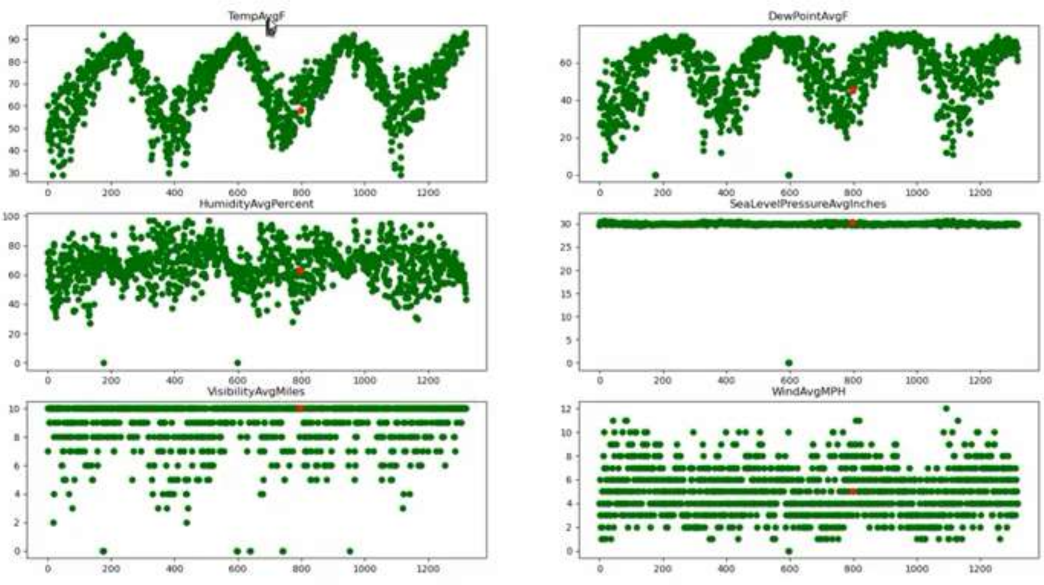
Advantages

One of its strengths is that it may be used to examine associations between a single dependent variable and a large number of potential mediators. Second, it lets scientists regulate confounding variables. Third, regression analyses how several elements interact with one another. Using the regression line as a foundation for estimates also helps in obtaining the measure of inaccuracy.

Result


inp = inp.reshape(1, -1)
# Print output
print('The precipitation in inches for
print('The precipitation trend graph: '
plt.scatter(days, Y, color='g')
plt.scatter(days[day_index], Y[day_inde
plt.title('Precipitation level')
plt.xlabel('Days')
plt.ylabel('Precipitation in inches')
Precipitation level
Days
Precipitation in inches
TERMINAL
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.
Try the new cross-platform PowerShell https://aka.ms/pscore6
PS C:\Users\amirb\Desktop\rainfall prediction> python main.py





Conclusion

In this research, we investigated and implemented several preprocessing procedures in order to discover how they affected the general efficiency of our classifiers. We also compared the performance of each classifier using data from a variety of sources to see the impact that data had on model predictions. It's safe to say that Australian weather is unpredictable, and that there is no regular pattern of rainfall that corresponds to a certain geographical area or specific time of day. We discovered associations and patterns in the data that allowed us to zero in on crucial details. Please see the supplemental materials. We can use Deep Learning models like Multilayer Perceptron and Convolutional Neural Network since we have so much data. A fantastic research project would compare Machine learning classifiers versus Deep learning models.

References

1. Climate Change and Human Health: Hazards and Responses, From the January 2003 issue of the World Health Report

2. Geomorphology, natural hazards, vulnerability, and disaster prevention in low-income countries. Alcntara-Ayala, I

3. Nicholls argues for better monitoring and prediction of weather and climate disasters. Dangers of the Environment.

4. Exploratory data analysis is the best approach to kick off a data science project, according to indatalabs (online).

5. Please refer to the Pandas documentation at get_dummies.html for further information.

6. Online documentation for Sckit-Learn may be found at featurehasher.html.

Sckitlearnonlinedocumentation stable/modules/generated/sklearn.preprocessing.minmaxscaler.html

7. Sckit-Learn Online Documentation feature_selection.selectkbest.html Sckit Learn Documentation.

8. Available at: https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python.

9. [Online] Raheel Shaikh, Feature Selection Techniques in Machine Learning with Python

For more information on Imbalanced-learn, see out its online documentation at Readthedocs.io/en/stable/introduction.html.

10. Ripple Down Rule Learner (RIDOR) Classifier for the IRIS Dataset, V. Veeralakshmi and D. Ramyachitra,

11. [Online] Indicators for Assessing Your Machine Learning Algorithm, Aditya Mishra Metrics to Evaluate Your Machine Learning Algorithm: