International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Predicting Rainfall using Machine Learning Techniques

#1M.RATNA KUMARI, #2 ALEKHYA KAVURI
#ASSISTANT PROFESSOR,#2 PG SCHOLAR
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS
QIS COLLEGE OF ENGENEERING AND TECHNOLOGY,ONGOLE
VENGAMUKKALAPALEM(V),ONGOLE,PRAKASAM DISTICT,ANDRA PRADESH

ABSTRACT

Rainfall prediction is one of the challenging and uncertain tasks which has a significant impact on human society. Timely and accurate predictions can help to proactively reduce human and financial loss. This study presents a set of experiments which involve the use of prevalent machine learning techniques to build models to predict whether it is going to rain tomorrow or not based on weather data for that particular day in major cities. This comparative study is conducted concentrating on three aspects: modeling inputs, modeling methods, and pre-processing techniques. The results provide a comparison of various evaluation metrics of these machine learning techniques and their reliability to predict the rainfall by analyzing the weather data.

INTRODUCTION

India’s welfare is agriculture. The achievement of agriculture is dependent on rainfall. It also helps with water resources. Rainfall information in the past helps farmers better manage their crops, leading to economic growth in the country. Prediction of precipitation is beneficial to prevent flooding that saves people's lives and property. Fluctuation in the timing ofprecipitation and its amount makes forecasting of rainfall a problem for

meteorological scientists. Forecasting is one of the utmost challenges for researchers from a variety of fields, such as weather data mining, environmental machine learning, functional hydrology, and numerical forecasting, to create a predictive model for accurate rainfall. In these problems, a common question is how to infer the past predictions and make use of future predictions. A variety of sub-processes are typically composed of the substantial process in rainfall. It is at times not promis ing to predict the precipitation correctly by on its global system. Climate forecasting stands out for all countries around the globe in all the benefits and services provided by the meteorological department. The job is very complicated because it needs specific numbers and all signals are intimated without any assurance. Accurate precipitation forecasting has been an important issue in hydrological science as early notice of stern weather can help avoid natural disaster injuries and damage if prompt and accurate forecasts are made. The theory of the modular model and the integrati2on of different models has recently gained more interest in rainfall forecasting

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

to address this challenge. A huge range of rainfall prediction methodologies is available in India. In India, there are two primary methods of forecasting rainfall. Regression, Artificial Neural Network (ANN), Decision Tree algorithm, Fuzzy logic and team process of data handling are the majority frequently used computational methods used for weather forecasting The basic goal is to follow information rules and relationships while gaining intangible and potentially expensive knowledge. Artificial NN is a promising part of this wide field

Rainfall prediction remains a serious concern and has attracted the attention of governments, industries, risk management entities, as well as the scientific community. Rainfall is a climatic factor that affects many human activities like agricultural production, construction, power generation, forestry and tourism, among others [1]. To this extent, rainfall prediction is essential since this variable is the one with the highest correlation with adverse natural events such as landslides, flooding, mass movements and avalanches. These incidents have affected society for years [2]. Therefore, having an appropriate approach for rainfall prediction makes it possible to take preventive and mitigation measures for these natural phenomena .

To solve this uncertainty, we used various machine learning techniques and models to make accurate and timely predictions. These paper aims to provide end to end machine learning life cycle right from Data preprocessing to implementing models to evaluating them. Data Preprocessing steps include imputing missing values, feature transformation, encoding categorical features, feature scaling and feature selection. We implemented models such as Logistic Regression, Decision Tree, K Nearest Neighbour, Rule-based and Ensembles. For evaluation purpose.

Case Study

In this paper, the data set under consideration contains daily weather observations from numerous Australian weather stations. The target variable is RainTomorrow which means: Did it rain the next day? Yes or No. The dataset is availableathttps://www.kaggle.com/jsphyg/ weather-dataset-rattle-package and definitions are adapted from http://www.bom.gov.au/climate/dwo/IDCJD W0000. shtml.

The data set consists of 23 features and 142k instances. Below are the features.

Feature	Description
Date	The date of observation
Location	The common name of the location of the weather station
MinTemp	The minimum temperature in degrees celsius
MaxTemp	The maximum temperature in degrees celsius
Rainfall	The amount of rainfall recorded for the day in mm
Evaporation	The so-called Class A pan evaporation (mm) in the 24 hours to 9am
Sunshine	The number of hours of bright sunshine in the day.
WindGustDir	The direction of the strongest wind gust in the 24 hours to midnight
WindGustSpeed	The speed (km/h) of the strongest wind gust in the 24 hours to midnight
WindDir9am	Direction of the wind at 9am
WindDir3pm	Direction of the wind at 3pm
WindSpeed9am	Wind speed (km/hr) averaged over 10 minutes prior to 9am
WindSpeed3pm	Wind speed (km/hr) averaged over 10 minutes prior to 3pm
Humidity9am	Humidity (percent) at 9am
Humidity3pm	Humidity (percent) at 3pm
Pressure9am	Atmospheric pressure (hpa) reduced to mean sea level at 9am
Pressure3pm	Atmospheric pressure (hpa) reduced to mean sea level at 3pm
Cloud9am	Fraction of sky obscured by cloud at 9am.
Cloud3pm	Fraction of sky obscured by cloud at 3pm.
Temp9am	Temperature (degrees C) at 9am
Temp3pm	Temperature (degrees C) at 3pm
RainToday	1 if precipitation exceeds 1mm, otherwise 0
RISK_MM	The amount of next day rain in mm.
RainTomorrow	The target variable. Did it rain tomorrow?

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

LITERATURE SURVEY

1.Climate Change and Human Health: Risks and Responses

The long-term good health of populations depends on the continued stability and functioning of the biosphere's ecological and physical systems, often referred to as life-support systems. We ignore this long-established historical truth at our peril: yet it is all too easy to overlook this dependency, particularly at a time when the human species is becoming increasingly urbanized and distanced from these natural systems. The world's climate system is an integral part of this complex of life-supporting processes, one of many large natural systems that are now coming under pressure from the increasing weight of human numbers and economic activities.

By inadvertently increasing the concentration of energy-trapping gases in the lower atmosphere, human actions have begun to amplify Earth's natural greenhouse effect. The primary challenge facing the world community is to achieve sufficient reduction in greenhouse gas emissions so as to avoid dangerous interference in the climate system. National governments, via the UN Framework Convention on Climate Change (UNFCC), are committed in principle to seeking this outcome. In practice, it is proving difficult to find a politically acceptable course of action—often because of apprehensions about possible short-term economic consequences.

This volume seeks to describe the context and process of global climate change, its actual or likely impacts on health, and how human societies should respond, via both adaptation strategies to lessen impacts and collective action to reduce greenhouse gas emissions. As shown later, much of the resultant risk to human populations and the ecosystems upon which they depend comes from the projected extremely rapid rate of change in climatic conditions. Indeed, the prospect of such change has stimulated a great deal of new scientific research over the past decade, much of which is elucidating the complex ecological disturbances that can impact on human well-being and health—as in the following example.

The US Global Change Research Program (Alaska Regional Assessment Group) recently documented how the various effects of climate change on aquatic ecosystems can interact and ripple through trophic levels in unpredictable ways. For example, warming in the Arctic region has reduced the amount of sea ice, impairing survival rates for walrus and seal pups that spend part of their life cycle on the ice. With fewer seal pups, sea otters have become the alternative food source for whales. Sea otters feed on sea urchins, and with fewer sea otters sea urchin populations are expanding and consuming more of the kelp that provides breeding grounds for fish. Fewer fish exacerbate the declines in walrus and seal populations. Overall, there is less food available for the Yupik Eskimos of the Arctic who rely on all of these species.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Global climate change is thus a significant addition to the spectrum of environmental health hazards faced by humankind. The global scale makes for unfamiliarity—although most of its health impacts comprise increases (or decreases) in familiar effects of climatic variation on human biology and health. Traditional environmental health concerns long have been focused on toxicological or microbiological risks to health from local environmental exposures. However, in the early years of the twenty-first century, as the burgeoning human impact on the environment continues to alter the planet's geological, biological and ecological systems, a range of larger-scale environmental hazards to human health has emerged. In addition to global climate change, these include: the health risks posed by stratospheric ozone depletion; loss of biodiversity; stresses on terrestrial and ocean food-producing systems; changes in hydrological systems and the supplies of freshwater; and the global dissemination of persistent organic pollutants.

Climate change and stratospheric ozone depletion are the best known of these various global environmental changes. Human societies, however, have had long experience of the vicissitudes of climate: climatic cycles have left great imprints and scars on the history of humankind. Civilisations such as those of ancient Egypt, Mesopotamia, the Mayans, the Vikings in Greenland and European populations during the four centuries of Little Ice Age, all have both benefited and suffered from nature's great climatic cycles. Historical analyses also reveal widespread disasters, social disruption and disease outbreaks in response to the more acute, inter-annual, quasi-periodic ENSO (El Niño Southern Oscillation) cycle (1). The depletion of soil fertility and freshwater supplies, and the mismanagement of water catchment basins via excessive deforestation, also have contributed to the decline of various regional populations over the millennia.

2. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries

The significance of the prevention of natural disasters is made evident by the commemoration of the *I*nternational *D*ecade for *N*atural *D*isaster *R*eduction (IDNDR). This paper focuses on the role of geomorphology in the prevention of natural disasters in developing countries, where their impact has devastating consequences. Concepts such as natural hazards, natural disasters and vulnerability have a broad range of definitions; however, the most significant elements are associated with the vulnerability concept. The latter is further explored and considered as a key factor in understanding the occurrence of natural disasters, and consequently, in developing and applying adequate strategies for prevention. Terms such as natural and human vulnerabilities are introduce and explained as target aspects to be taken into account in the reduction of vulnerability and for prevention and mitigation of natural disasters. The importance of the incorporation not only of geomorphological

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

research, but also of geomorphologists in risk assessment and management programs in the poorest countries is emphasized.

3. Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation

The last few years have seen enormous damage and lossof life from climate and weather phenomena. The mostdamaging events have included the severe 1997/98 ElNiño (with its near-global impacts), HurricaneMitch, and floods in China in mid-1998. What have welearnt regarding the causes, variability, andpredictability, of these phenomena? Can we predict theoccurrence of these extreme events, and therebymitigate their damage? This paper reviews what we havelearnt in the last decade or so regarding thepredictability of these climate and weather extremes.The view starts with the largest (El Niño) scales,and works towards the scale of individualthunderstorms. It focuses on the practical outcomes ofour improved knowledge with regard to decreasing theimpact of natural disasters, rather than describing indetail the scientific knowledge underlying theseoutcomes. The paper concludes with a discussion ofsome of the factors that still restrict our ability tomitigate the deleterious effects of atmospheric andclimatic hazards.

4. Exploratory Data Analysis: the Best way to Start a Data Science Project

Exploratory Data Analysis is a set of techniques that were developed by Tukey, John Wilder in 1970. The philosophy behind this approach was to examine the data before building a model. John Tukey encouraged statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. Today Data scientists and analysts spend most of their time in Data Wrangling and Exploratory Data Analysis also known as EDA. But what is this EDA and why it is so important? This article explains what is EDA and how to apply EDA techniques to a dataset.

SYSTEM ANALYSIS

EXISTING SYSTEM

Rainfall prediction is important as heavy rainfall can lead to many disasters. The prediction helps people to take preventive measures and moreover the prediction should be accurate. There are two types of prediction short term rainfall prediction and long term rainfall. Prediction mostly short term prediction can gives us the accurate result. The main challenge is to build a model for long term rainfall prediction. Heavy precipitation prediction could be a major drawback for earth science department because it is closely associated with the economy and lifetime of human.

Disadvantages

We can just do it by having the historical data analysis of rainfall and can predict the rainfall for future seasons. We can apply many techniques like classification, regression according to the requirements and also we can calculate the error between the actual and prediction and also the accuracy.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Different techniques produce different accuracies so it is important to choose the right algorithm and model it according to the requirements

PROPOSED SYSTEM

Accuracy of rainfall statement has nice importance for countries like India whose economy is basically dependent on agriculture. The dynamic nature of atmosphere, applied mathematics techniques fail to provide sensible accuracy for precipitation statement. The prediction of precipitation using machine learning techniques may use regression. Intention of this project is to offer non-experts easy access to the techniques, approaches utilized in the sector of precipitation prediction and provide a comparative study among the various machine learning techniques.

Advantages

1. It is a powerful technique for testing relationship between one dependent variable and many independent variables.

2. It allows researchers to control extraneous factors.

3. Regression asses the cumulative effect of multiple factors.

4. It also helps to attain the measure of error using the regression line as a base for estimations.

IMPLEMENTATION

1. System Overview

The rainfall prediction system is developed using machine learning algorithms to analyze historical weather data and forecast rainfall occurrence or quantity. The implementation consists of multiple stages including data acquisition, preprocessing, model training, and prediction.

2. Data Acquisition

The dataset is collected from reliable meteorological sources such as weather departments or open-source platforms. The dataset includes attributes such as:

- Temperature (minimum and maximum)
- Humidity
- Wind speed
- Atmospheric pressure
- Rainfall (target variable)

The collected data is stored in CSV format and imported into the system using Python libraries.

3. Data Preprocessing

Data preprocessing is an essential step to ensure the quality of input data. The following operations are performed:

- **Handling Missing Values:** Missing entries are replaced using mean or median values.
- **Data Cleaning:** Removal of duplicate and inconsistent records.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

- **Feature Scaling:** Standardization using normalization techniques to bring all features to a similar range.
- **Encoding:** Conversion of categorical data into numerical form using label encoding.

4. Feature Selection

Relevant features that influence rainfall are selected to improve model performance. Correlation analysis is used to identify highly contributing attributes. Irrelevant or redundant features are removed to reduce overfitting and computational complexity.

5. Dataset Splitting

The dataset is divided into:

- **Training Set (80%)** – used to train the model
- **Testing Set (20%)** – used to evaluate the model

This ensures that the model generalizes well on unseen data.

6. Model Implementation

Different machine learning algorithms are implemented and compared:

a) Linear Regression

Used for predicting continuous rainfall values based on input features.

b) Logistic Regression

Used for classification (Rain / No Rain).

c) Decision Tree

Creates a tree-like model based on feature conditions.

d) Random Forest

An ensemble method that combines multiple decision trees to improve accuracy and reduce overfitting.

e) Support Vector Machine (SVM)

Used for classification with high-dimensional data.

f) Artificial Neural Network (ANN)

Captures complex nonlinear relationships in weather data.

7. Model Training

Each model is trained using the training dataset. During training:

- Input features are fed into the algorithm
- The model learns patterns and relationships
- Hyperparameters are tuned to optimize performance

Methodology

the overall architecture include four major components: Data Exploration and Analysis,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Data Pre-processing, Model Implementation, and Model Evaluation, as shown in Fig.

Over All Architecture

Data Exploration and Analysis

Exploratory Data Analysis is valuable to machine learning problems since it allows to get closer to the certainty that the future results will be valid, correctly interpreted, and applicable to the desired business contexts [4]. Such level of certainty can be achieved only after raw data is validated and checked for anomalies, ensuring that the data set was collected without errors. EDA also helps to find insights that were not evident or worth investigating to business stakeholders and researchers

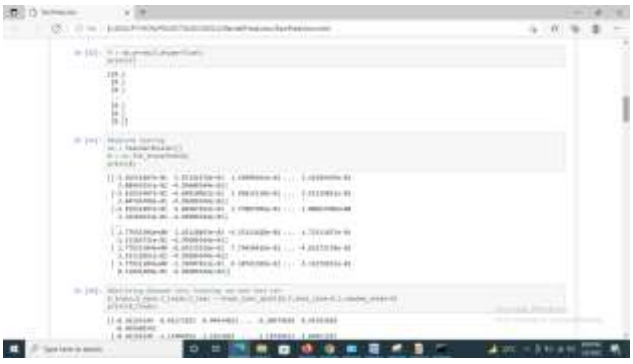
We performed EDA using two methods - Univariate Visualization which provides summary statistics for each field in the raw data set (figure 2) and Pair-wise Correlation Matrix which is performed to understand interactions between different fields in the data set.

RESULTS

Packeages

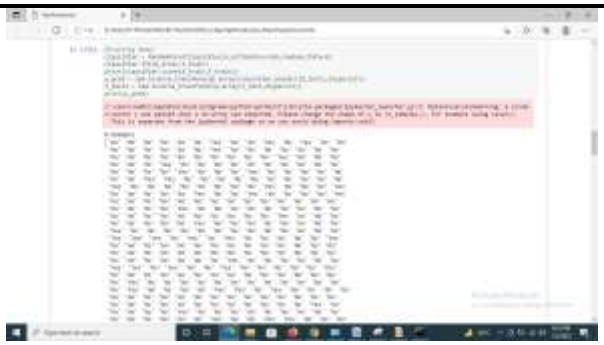
Analysis

Training



Algorithms

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



Random Forest

Bagging Classifer

Gradient Boosting

CONCLUSION

Rainfall prediction using machine learning techniques provides a reliable and efficient approach compared to traditional statistical methods. By leveraging historical weather data and advanced algorithms such as Random Forest and Neural Networks, the system can identify complex patterns and improve prediction accuracy.

The implementation demonstrates that ensemble and deep learning models outperform basic models in handling nonlinear relationships within meteorological data. Accurate rainfall prediction is crucial for agriculture, disaster management, and water resource planning, making this system highly beneficial for real-world applications.

However, the performance of the model depends heavily on the quality and quantity of data. Future enhancements may include integrating real-time data streams, satellite imagery, and deep learning models such as LSTM for time-series forecasting.

Overall, the project proves that machine learning is a powerful tool for environmental prediction and can significantly contribute to smarter and data-driven decision-making systems.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

REFERENCES

1. World Health Organization: Climate Change and Human Health: Risks and Responses. World Health Organization, January 2003

2. Alcntara-Ayala, I.: Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. Geomorphology 47(24), 107124 (2002)

3. Nicholls, N.: Atmospheric and climatic hazards:Improved monitoring and prediction for disaster mitigation. Natural Hazards 23(23), 137155 (2001)

4. [Online] InDataLabs, Exploratory Data Analysis: the Best way to Start a Data ScienceProject.Available:https://medium.com/@InDataLabs/ why-start-a-data-science-project-with-exploratory-data-analysis-f90c0efcbe49

5.[Online]PandasDocumentation. Available:https://pandas.pydata.org/ pandas-docs/stable/reference/api/pandas.get_dummies.html

6. [Online] Sckit-Learn Documentation Available:https://scikitlearn.orgstable/modules/generated/sklearn.feature\xtraction.FeatureHasher. html

7. [Online] Sckit-Learn Documentation Available:https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

8. [Online] Sckit Learn Documentation Available:https://scikitlearn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html

9. [Online] Raheel Shaikh, Feature Selection Techniques in Machine Learning with PythonAvailable:https://towardsdatascience.com/feature-selection-techniques-imachine-learning-with-python-f24e7da3f36e

10.[Online]ImbalancedlearnDocumentation Available:https://imbalancedlearn.readthedocs.io/en/stable/introduction.html

11. V. Veeralakshmi and D. Ramyachitra, Ripple Down Rule learner (RIDOR) Classifier for IRIS Dataset. Issues, vol 1, p. 79-85.

12. [Online] Aditya Mishra, Metrics to Evaluate your Machine Learning Algorithm Available: https://towardsdatascience.com/ metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38

AUTHOR PROFILE

Mrs. M Ratna kumari is an Assistant Professor In the Department of Master of Computer Application at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned M.Tech(CSE) in Chennai Bharath University, and her now pursuing PHD in

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

her research intrests include Machine Learning with AI programming language. She is committed to advancing research and forecasting innovating while mentoring students to excel in both academic & professional pursuits.

STUDENT PROFILE :

Alekhya Kavuri is a postgraduate student pursuing a MCA in the deparrtment of computer Applications at QIS College of Engineering & Technology, Ongole autonomous college in prakasam dist. She completed undergraduate degree in BSC (computer science) from ANU.with a keen intrest in research and practical learning, She is actively involved in academic projects and technical activities related to her field.