

RAINFALL PREDICTION USING MACHINE LEARNING

Kanasani Renuka Devi¹, Alapati Padma Sree²,

Narra Bala Mary Sowmya³, Katuri Santosh Dutt⁴, Mrs. Sk. Shammi⁵

^{1,2,3,4} Student, Department of Computer Science and Engineering, Tirumala Engineering College

⁵ Professor, Department of Computer Science and Engineering, Tirumala Engineering College

Abstract - India is an agricultural country and its economy is largely based upon crop productivity and rainfall. For analyzing the crop productivity, rainfall prediction is required and necessary to all farmers. Rainfall Prediction is the application of science and technology to predict the state of the atmosphere. It is important to exactly determine the rainfall for effective use of water resources, crop productivity and pre planning of water structures. Using different data mining techniques it can predict rainfall. Data mining techniques are used to estimate the rainfall numerically. This paper focuses some of the popular data mining algorithms for rainfall prediction. Random Forest, K-Nearest Neighbor algorithm, Logistic regression, Decision Tree are some of the algorithms have been used. From that comparison, it can analyze which method gives better accuracy for rainfall prediction.

Key Words: Rain fall, Agricultural, Economy, Farmers, Water resources.

1. INTRODUCTION

1.1. OBJECTIVE OF THE PROJECT:

The goal is to develop a machine learning model for Rainfall Prediction to potentially replace the outdated supervised machine learning classification models by predicting results in the form of best accuracy by comparing supervised algorithm.

1.1.1. Necessity:

This prediction helps in predicting the rainfall and it helps in Over coming the crop productivity and to predict the state of atmosphere in agricultural countries. These models are very easy to use. It can work accurately and very smoothly in a different scenario. It reduces the effort workload and increases efficiency in work. In aspects of time value, it is worthy.

1.1.2. Software development method:

In many software applications program different methods and cases are followed such as, Waterfall model, Iterative model, Spiral model, V- model and Big Bang model. we used waterfall model in this application. we tried to use test case and cases of ware approaches.

1.1.3 Layout of the document:

This documentation starts with form a introduction. After introduction analysis and design of the project are described. In analysis and design of the project have many parts such as project proposal, mission, goal, target audience, environment. Use cases and test cases are explained below respectively. Finally, this documentation finished with result and Conclusion part.

2. LITERATURE REVIEW

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discuss published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes are search proposal and may be just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis.

A summary is a recap of important information about the source, but a synthesis is a re-organization, reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent event of them.

3. SYSTEM ANALYSIS

The project proposal is the term of documents. A project can describe the project proposal. It is the set of all plans of a project. Like, how the software works, what are the steps to complete the entire projects, and what are the software requirements and analysis for this project. In our project, we are doing all the steps and also risk and reward and other project dependencies in the project proposal.

To compare several machine learning models like logistic

regression, random forest, knn and decision tree. Plotting confusion matrix for each model after cleaning the data set so that we can easily find the best model among them. After finding best model we will draw ROC curve and classification report for that best fit model to predict train fall which is very essential for farmers.

The goal is to develop a machine learning model for predicting the rainfall.

The scope of this paper is to implement and investigate how different supervised binary classification methods impact default prediction. The model evaluation techniques used in this project are limited to precision, sensitivity, F1-score.

The overview of the project is to provide a best machine learning algorithm to the user. Therefore, the user can directly know whether the rainfall is occur or not through his best model.

4. EXISTING SYSTEM

Agriculture is the strength of our Indian economy. Farmer only depend upon monsoon to be their cultivation. The good crop productivity needs good soil, fertilizer and also good climate. Weather forecasting is the very important requirement of the each farmer. Due to the sudden changes in climate/weather, The people are suffer economically and physically. Weather prediction is one of the challenging problems in current state. The main motivation of this paper to predict the weather using various data mining techniques. Such as classification, clustering, decision tree and also neural networks. Weather related information is also called the meteorological data. In this paper the most commonly used weather parameters are rainfall, wind speed, temperature and cold.

4.1 PREPARING THE DATASET

This data set contains 145460 records of features extracted from kaggle, which is having Rain Tomorrow as a target column containing 2 values.

5. PROPOSED SYSTEM

5.1 Exploratory Data Analysis of Rainfall Prediction:

Multiple datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy.

5.1.2. Data Cleaning

In this section of the report will load in the data, check for cleanliness, and then trim and clean given data set for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

5.1.3 Data Collection

The data set collected for predicting given data

is split into Training set and Test set. Generally, we split the dataset into Training set and Test set. The Data Model which was created using machine learning algorithms are applied on the Training set and based on the test result.

5.1.4 Building the classification model

For predicting the rainfall, ML algorithm prediction model is effective because of the following reasons: It provides better results in classification problem. It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables. It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.

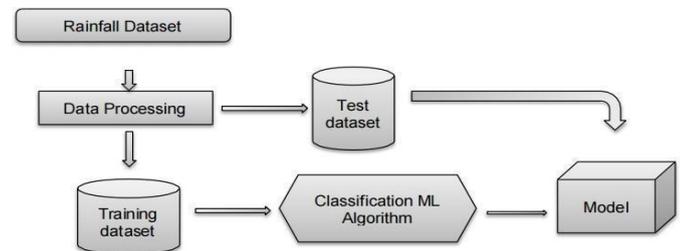
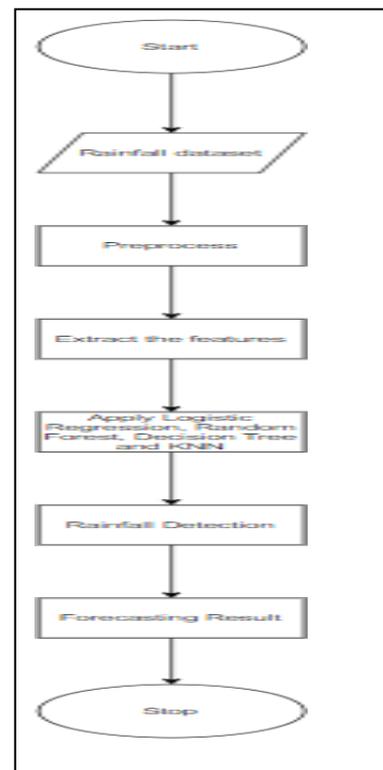


Fig 5 : Architecture of Proposed model

5.1.5 Advantages

- Performance and accuracy of the algorithms can be calculated and compared.
- Numerical Weather Prediction
- Statistical Weather Prediction

5.1.6 FLOW CHART



6. METHODS AND ALGORITHMS USED

SYSTEM STUDY:

To develop his model we use new modern technologies which are Machine Learning using Python for predicting rainfall.

System requirement specifications:

Hardware requirements:

- Processor :Intel
- RAM :2GB
- Hard Disk :80GB

Software requirements:

- OS :Windows
- Framework: Flask
- Technology: Machine Learning using Python
- Web Browser: Chrome, Microsoft Edge
- Codeeditor : Visual Studio Code, Google Colab,

SOFTWARE ENVIRONMENT:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English key words frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is Interpreted – Python is processed at run time by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive –You can actually it at a Python prompt and interact with the interpreter directly to write your programs.

Python is Object-Oriented –Python supports Object-Oriented style or technique of programming that encapsulates code with in objects.

Python is a Beginner's Language – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

7. SYSTEM TESTING

TESTING PLAN

Software testing is the process of evaluation a software item to detect differences between given input and expected output. Also to assess the feature of a software item. Testing assesses the quality of the product. Software testing is a process that should be done during the development process. In other words software testing is a verification and validation process.

Verification :

Verification is the process to make sure the product satisfies the conditions imposed at the start of the development phase. In other words, to make sure the product behaves the way we want it to.

Validation :

Validation is the process to make sure the product satisfies the specified requirements at the end of the development phase. In other words, to make sure the product is built as per customer requirements.

BASICS OF SOFTWARE TESTING :

There are two basics of software testing: black box testing and white box testing.

Black box testing :

Black box testing is a testing technique that ignores the internal mechanism of the system and focuses on the output generated against any input and execution of the system. It is also called functional testing.

White box testing :

White box testing is a testing technique that takes into account the internal mechanism of a system. It is also called structural testing and glass box testing. Black box testing is often used for validation and white box testing is often used for verification.

1. TYPES OF TESTING :

There are many types of testing like :

- Unit Testing
- Integration Testing
- Functional Testing
- System Testing
- Stress Testing
- Performance Testing
- Usability Testing
- Acceptance Testing
- Regression Testing
- Beta Testing

Unit Testing :

Unit testing is the testing of an individual unit or group of related units. It falls under the class of white box testing. It is often done by the programmer to test that the unit he/she has implemented is producing expected output against given input.

Integration Testing :

Integration testing is testing in which a group of components are combined to produce output. Also, the interaction between software and hardware is tested in integration testing if software and hardware components have any relation. It may fall under both white box testing and black box testing.

Functional Testing :

Functional testing is the testing to ensure that the specified functionality required in the system requirements works. It falls under the class of black box testing.

System Testing :

System testing is the testing to ensure that by putting the software in different environments (e.g., Operating Systems) it still works. System testing is done with full system implementation and environment. It falls under the class of black box testing.

Stress Testing :

Stress testing is the testing to evaluate how system behaves under unfavorable conditions. Testing is conducted at beyond limits of the specifications. It falls under the class of black box testing.

Performance Testing :

Performance testing is the testing to assess the speed and effectiveness of the system and to make sure it is generating results within a specified time as in performance requirements. It falls under the class of black box testing.

Usability Testing :

Usability testing is performed to the perspective of the client, to evaluate how the GUI is user-friendly? How easily can the client learn? After learning how to use, how proficiently can the client perform? How pleasing is it to use its design? This falls under the class

of black box testing.

Acceptance Testing :

Acceptance testing is often done by the customer to ensure that the delivered product meets the requirements and works as the customer expected. It falls under the class of black box testing.

Regression Testing :

Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing.

8. RESULTS, DISCUSSIONS AND PERFORMANCE ANALYSIS**PERFORMANCE ANALYSIS :**

Website performance optimization, the focal point of technologically superior website designs is the primary factor dictating the Rainfall occurred or not. After all, unimpressive website performance kills admission process when the torture of waiting for slow Web pages to load frustrates visitors in to seeking alternatives – impatience is a digital virtue! And also the ml algorithms used in our project will give the best accurate result to the user for Rainfall prediction.

We created the following six chapter in-depth speed optimization guide to show you how important it is to have a fast loading, snappy website! Countless research papers and benchmarks prove that optimizing your sites' speed is one of the most affordable and highest ROI providing investments!

Lightning-fast page load speed amplifies visitor engagement, retention, and boosts sales. Instantaneous website response leads to higher conversion rates, and every 1 second delay in page load decreases customer satisfaction by 16 percent, page views by 11 percent and conversion rates by 7 percent according to recent Aberdeen Group research.

Algorithm	Accuracy
Logistic Regression	0.8
Random Forest	1.0
K Nearest Neighbors	0.9
Decision Tree Classifier	1.0

DISCUSSIONS :

While discussions provide avenues for exploration and discovery, leading a discussion can be anxiety-producing: discussions are, by their nature, unpredictable, and require us as instructors to surrender a certain degree of control over the flow of information. Fortunately, careful planning can help us ensure that discussions are lively without being chaotic and exploratory without losing focus. When planning a discussion, it is helpful to consider not only cognitive, but also social/emotional, and physical factors that can either foster or inhibit the productive exchange of ideas.

9. CONCLUSION

This project represented the Machine Learning Approach for predicting the rainfall by using 4 ML algorithms like Logistic Regression, Random Forest Classifier, Decision Tree and KNN. Comparing the 4 algorithms and choosing the best approach for rainfall prediction. This project provides a study of different types of methodologies used to forecast and predict rainfall and Issues that could be found when applying different approaches to forecast in rainfall.

Because of nonlinear relationships in rainfall data sets and the ability to learn from the past, makes a superior solution to all approaches available. The future work of the project would be the improvement of architecture for light and other weather scenarios. Also, can develop a model for small changes in climate in future. An algorithm for testing daily basis data set instead of accumulated data set could be of paramount Importance for further research.

10. REFERENCES

- Xiong, Lihua, and Kieran M. OConnor. "An empirical method to improve the prediction limits of the GLUE methodology in rainfallrunoff modeling." *Journal of Hydrology* 349.1-2 (2008): 115-124.
- Schmitz, G. H., and J. Cullmann. "PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments." *Journal of hydrology* 360.1-4 (2008): 1-14.
- Riordan, Denis, and Bjarne K. Hansen. "A fuzzy casebased system for weather prediction." *Engineering Intelligent Systems for Electrical Engineering and Communications* 10.3 (2002): 139-146.
- Guhathakurta, P. "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network." *Current Science* 90.6 (2006): 773-779.
- Pilgrim, D. H., T. G. Chapman, and D. G. Doran. "Problems of rainfall-runoff modelling in arid and semiarid regions." *Hydrological Sciences Journal* 33.4 (1988): 379-400.
- Lee, Sunyoung, Sungzoon Cho, and Patrick M. Wong. "Rainfall prediction using artificial neural networks." *journal of geographic information and Decision Analysis* 2.2 (1998): 233- 242..
- French, Mark N., Witold F. Krajewski, and Robert R. Cuykendall. "Rainfall forecasting in space and time using a neural network." *Journal of hydrology* 137.1-4 (1992): 1-31.
- Charaniya, Nizar Ali, and Sanjay V. Dudul. "Committee of artificial neural networks for monthly rainfall prediction using wavelet transform." *Business, Engineering and Industrial Applications (ICBEIA), 2011 International Conference on. IEEE, 2011.*
- Noone, David, and Harvey Stern. "Verification of rainfall forecasts from the Australian Bureau of Meteorology's Global Assimilation and Prognosis(GASP) system." *Australian Meteorological Magazine* 44.4 (1995): 275-286.
- Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators." *Neural networks* 2.5 (1989): 359-366.
- Haykin, Simon. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- Rajeevan, M., Pulak Guhathakurta, and V. Thapliyal. "New models for long range forecasts of summer monsoon rainfall over North West and Peninsular India." *Meteorology and Atmospheric Physics* 73.3-4 (2000) 211-