



Rainfall Prediction Using Machine Learning

M.Ajay¹, M.Harshita², K.Sai Kiran Reddy³, K.Vara Prasad⁴

^{1,2,3,4}UG Students, Department of Electronics and Communication Engineering, Tirumala Engineering College, Narasaraopet, Guntur Dist. Andhra Pradesh

ABSTRACT

This project seeks a distinctive and efficient machine learning system for the prediction of rainfall. The main motive of the project is to predict the amount of rainfall in a particular place well in advance. We predict the amount of rainfall using the past data. The learning of data can be done by applying different algorithms. The rainfall parameters in this study are collected, trained and tested to achieve the sustainable results through different models. The monthly rainfall predictions obtained after training and testing are then compared with actual data to ensure the accuracy of the model. We apply various statistical and machine learning approaches in the prediction and make analysis over various approaches. By using various approaches, we try to minimize the error. The results of this project outline that the model is successful in predicting the monthly rainfall data with the particular parameters.

I INTRODUCTION

This Rainfall Prediction is an important task that provides crucial information on the conditions ahead. This climatic factor has impact on many human activities like power generations, forestry, tourism and agricultural production. Prediction of rain can also help us saving from adverse natural events such as avalanches, landslides and flooding. These incidents have high impact on society for years. Therefore, having an appropriate approach for rainfall prediction makes it possible to take pre-emptive safety measures for these natural phenomena.

II LITERATURE REVIEW

Machine learning is one of the applications of artificial intelligence (AI) that provides computers, the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to learn from themselves. The main aim is to allow computers to learn automatically without human intervention and also adjust actions accordingly.

III EXISTING METHOD

Rainfall prediction model based on several ANN based architectures have been proposed to predict rainfall. The learning algorithm tries to find out the optimal set this the training phase can be thought of as an optimization problem where an error function is usually minimized.



IV PROPOSED SYSTEM

To predict the rainfall in advance to take pre-emptive safety measures from different natural phenomena that results in human and financial loss. So by using data mining techniques and machine learning models we can predict the rainfall.

Advantages:

1. Give good accuracy for the considered dataset.
2. Can easily calculate mean absolute error for each model and can calculate mean and standard deviation for each dataset accurately.

V SOFTWARE REQUIRED

PYTHON: The most recent major version of Python is Python3. However, Python 2, although not being updated with anything other than security updates, is still quite popular. It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Net Beans or Eclipse, Anaconda which are particularly useful when managing larger collections of Python files. Python was designed for its readability. Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses. Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose. In the older days, people used to perform Machine Learning tasks manually by coding all the algorithms and mathematical and statistical formula. This made the process time-consuming, tedious and inefficient. But in the modern days, it has become very much easy and efficient compared to the olden days by various Python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reasons is its vast collection of libraries. Python libraries that are used in Machine Learning are:

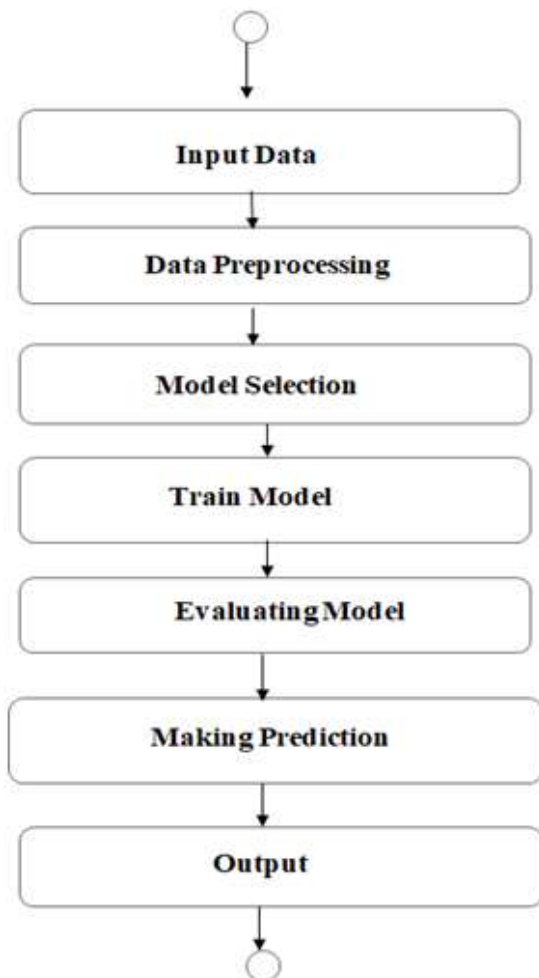
1. **NumPy** is a very popular Python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High-end libraries like TensorFlow use NumPy internally for manipulation of Tensors.
2. **SciPy** is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation.
3. **Scikit-learn** is one of the most popular Machine Learning libraries for classical Machine Learning algorithms. It is built on top of two basic Python libraries, NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit-learn can also be used for data-mining and data-

analysis, which makes it a great tool who is starting out with Machine Learning.

4. Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

5. Matplotlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy whena programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, histogram, error charts, bar chats, etc.

VI BLOCK DIAGRAM OF PROPOSED METHOD

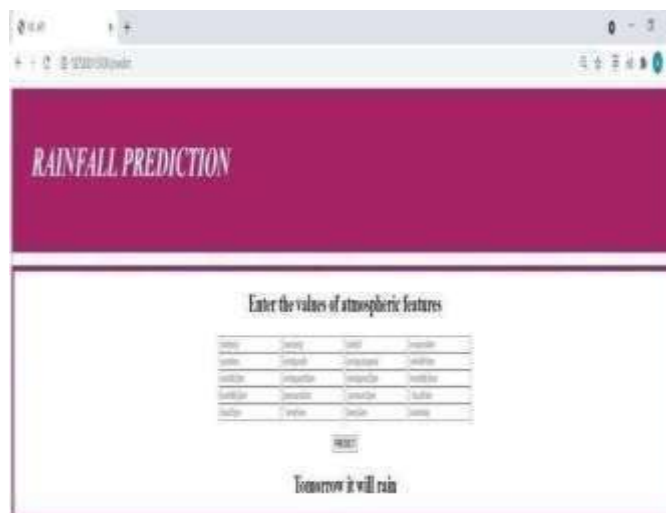


VII DATA PREPROCESSING

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. We have carried below preprocessing steps;

1. **Missing Values:** As per our EDA step, we learned that we have few instances with null values. Hence, this becomes one of the important step. To impute the missing values, we will group our instances based on the location and date and there by replace the null values by there respective mean values.
2. **Feature Expansion:** Date feature can be expanded to Day, Month and Year and then these newly created features can be further used for other preprocessing steps.
3. **Categorical Values:** Categorical feature is one that has two or more categories, but there is no intrinsic ordering to the categories. We have a few categorical features - WindGustDir, WindDir9am, WindDir3pm with 16 unique values. Now it gets complicated for machines to understand texts and process them, rather than numbers, since the models are based on mathematical equations and calculations.
4. **Activity Diagram:** Activity Diagrams describe how activities are coordinated to provide a service which can be at different levels of abstraction. Typically, an event needs to be achieved by some operations, particularly where the operation is intended to achieve a number of different things that require coordination, or how the events in a single use case relate to one another, in particular, use cases where activities may overlap and require coordination.

VIII RESULT





CONCLUSION

In this paper, we explored and applied several preprocessing steps and learned their impact on the overall performance of our classifiers. We also carried a comparative study of all the classifiers with different input data and observed how the input data can affect the model predictions. We can conclude that Australian weather is uncertain and there is no such correlation among rainfall and the respective region and time.

REFERENCES

1. Abhishek, K., Kumar, A., Ranjan, R., & Kumar, S. (2012). A Rainfall Prediction Model using Artificial Neural Network. *IEEE Control and System Graduate Research Colloquium*, 15.
2. M.Sireesha, S. N. TirumalaRao, SrikanthVemuru, Frequent Itemset Mining **Algorithms**: A Survey Journal of Theoretical and Applied Information Technology Vol - 96, No .3, Feb - 2018 ISSN - 1992-8645, Pages – 744 – 755.
3. M. Sireesha, Srikanth Vemuru and S. N. TirumalaRao, "Coalesce based binary table: an enhanced algorithm for mining frequent patterns", *International Journal of Engineering and Technology*, vol. 7, no. 1.5, pp. 51-55, 2018.
4. M.Sireesha, S. N. TirumalaRao, SrikanthVemuru, Optimized Feature Extraction and Hybrid Classification Model for Heart Disease and Breast Cancer Prediction *International Journal of Recent Technology and Engineering* Vol - 7, No 6, Mar - 2019 ISSN - 2277-3878, Pages–1754–1772.