International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Manuscript Rainfall Prediction using Multilayer Perceptron (MLP)

Manoj Chhetri[1]

[1]College Of Science and Technology, Royal University of Bhutan

---

Abstract - In this study, we undertook an analysis focused on forecasting hourly rainfall. This analysis framed the problem as binary classification, with rainfall events classified into two categories: "rainy" (the positive class) and "non-rainy" (the negative class). We leveraged independent climatic variables from the current hour to predict rainfall conditions for the following hour. Our data source was the CST weather station, providing records of 8 hourly weather parameters. To make predictions, we harnessed a commonly used machine learning model, the Multilayer Perceptron (MLP), resulting in an accuracy rate of 79%.

***Key Words*:** rainfall, MLP, time series

1.INTRODUCTION

Rainfall prediction constitutes a vital realm of research, given its substantial implications for agriculture, the management of water resources, and disaster preparedness. Decision trees, a widely adopted machine learning approach, are particularly favored for rainfall prediction because of their proficiency in handling intricate data patterns and delivering interpretable outcomes.

Phuntsholing is characterized by a subtropical climate featuring well-defined seasons. During the summer months, spanning from June to August, the weather turns hot and humid, with temperatures spanning the range of 25 to 35 degrees Celsius. This period is marked by frequent monsoon rains, contributing to the region's lush greenery. The monsoon season ushers in copious amounts of precipitation, sometimes leading to flash floods and landslides in the surrounding areas.

Traditional rainfall prediction methods relied heavily on statistical models and numerical weather simulations. These methods often struggled to capture the intricate and nonlinear patterns in rainfall data. MLP, a type of artificial neural network, has emerged as an alternative approach that holds promise for improving the accuracy of rainfall predictions.

In this scholarly article, we offer an in-depth examination of rainfall prediction through a case study employing decision trees within the CST region, specifically Phuntsholing. Our primary objective is to assess the performance and applicability of decision tree models in forecasting rainfall patterns within the CST area, relying on historical weather data.

2. Body of Paper

2.1. Data Collection

The dataset was collected at the College of Science and Technology, Rinchending, Bhutan, using the WatchDog 2900ET Weather Station. Measurements were taken at 10-minute intervals, precisely at the geographical coordinates of 26.89 North Latitude and 89.39 East Longitude. Subsequently, the collected data was transformed into hourly averages and meticulously reviewed for any potential errors, with the Center for Renewable and Sustainable Energy Development at the College of Science and Technology overseeing the correction process.

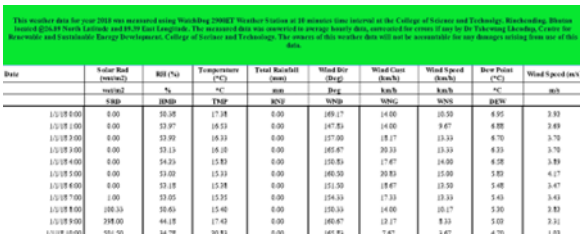

Fig -1: Dataset

2.2. Data Preprocessing

The collected dataset was preprocessed using the following data preprocessing pipeline:

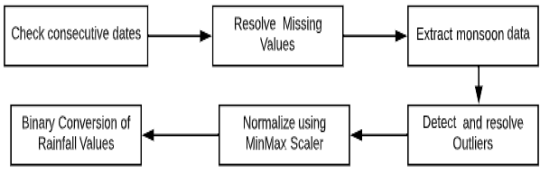

Fig -2: Data Preprocessing Pipeling

The dataset underwent a standard preprocessing pipeline. It was initially received as four distinct sheets, each organized by year. Our initial step involved a comprehensive review to ensure the presence of continuous dates and hours within each annual dataset.

We performed a thorough check for missing values and subsequently removed any rows containing such gaps. The dataset encompassed records for every month throughout the year. However, given the typical occurrence of rainfall during the monsoon season, including all records introduced a substantial class imbalance issue, as rainfall is infrequent during other periods. To address this, our study exclusively

focused on data from the monsoon season, which spans from July to September.

In order to identify and manage potential outliers, we employed a basic box plot analysis. Any outliers identified during this process were substituted with their respective mean values.

2.3. Data Preprocessing and Conversion

Machine learning usually involves converting the dataset into training and testing sets. In our research we perform an 80-20 split whereby 80% of the dataset were used for training and the remaining 20% were used for testing.

The recorded dataset is a timeseries dataset and all the records are ordered chronologically, with a timestamp associated with each observation. Machine learning models like decision tree requires the dataset to consist of a set of independent variables(y) and a dependent variable(x). Using the variables in set 'y' the decision trees calculate 'x'. Since we want to predict the status of rainfall in the next hour using the parameters which are currently available to us, we convert the timeseries to a machine learning problem by shifting the binary rainfall records by a timestamp of one. So, the climatic parameters available becomes independent variables in set 'y' and the status of rainfall becomes the dependent variable 'x'.

2.4. MLP Experiments

Multilayer Perceptron (MLP) is a fundamental type of artificial neural network used in machine learning and deep learning. It's designed to mimic the structure and function of the human brain, with interconnected layers of nodes, or neurons, that process and transform data. An MLP typically consists of three layers: an input layer, one or more hidden layers, and an output layer. Each neuron in one layer is connected to every neuron in the subsequent layer, and these connections have associated weights that determine the strength of the connections. During training, the model learns by adjusting these weights to minimize the difference between its predictions and the actual target values.

MLPs are highly versatile and capable of solving a wide range of problems, including regression, classification, and pattern recognition. They excel at capturing complex patterns in data, making them suitable for tasks like image and speech recognition, natural language processing, and even time series forecasting. However, training an MLP can be computationally intensive, and it requires a sufficient amount of labeled data to perform effectively.

The researchers in this study experimented with various setups or configurations of the Multilayer Perceptron (MLP) model, which is a type of neural network commonly used in machine learning. They wanted to find the best architecture that would yield accurate predictions for their specific task, which, in this case, was likely rainfall prediction.

After trying different combinations of hidden layers and the number of neurons in each layer, they settled on a configuration with three hidden layers. These hidden layers contained 32, 64, and 128 neurons, respectively. This configuration, with its gradually increasing number of neurons in the hidden layers, appeared to provide the best performance and accuracy for their particular dataset and prediction task. Essentially, they found that this 3-layer architecture with 32-64-128 configuration struck the right balance in capturing the intricate patterns in their data, leading to more accurate and reliable predictions.

2.4. Model Evaluation

Accuracy serves as a foundational assessment metric utilized in the realms of machine learning and statistics to gauge the performance of a classification model. Its role lies in quantifying the model's proficiency in correctly assigning class labels or categories to data points concerning the total volume of data points available for evaluation.

In the computation of accuracy, the customary practice entails dividing the count of accurate predictions, encompassing both true positives and true negatives, by the overall count of predictions made, which includes true positives, true negatives, false positives, and false negatives. This computation yields a numerical value between 0 and 1. An accuracy score of 1 signifies perfect alignment between the model's predictions and the actual outcomes, while a score of 0 denotes complete incongruity, indicating that the model's predictions are entirely erroneous.

Accuracy emerges as an uncomplicated metric, offering clarity and ease of interpretation. Consequently, it stands as one of the most frequently employed metrics in an array of classification tasks.

Throughout various configurations of the Multilayer Perceptron (MLP) model, the research or experimentation process revealed that the highest level of accuracy achieved was 82%. This means that, among the different settings and architectural choices tested for the MLP, the particular configuration associated with this 82% accuracy score emerged as the most effective in accurately predicting or classifying the desired outcomes. In other words, this specific MLP setup exhibited the best performance in terms of aligning its predictions with the actual data, resulting in an accuracy rate of 82%. This finding underscores the significance of this particular configuration in achieving optimal results for the given task or dataset.

3. CONCLUSIONS

We carried out numerous experiments involving various Multilayer Perceptron (MLP) configurations. The most notable result was an 82% accuracy achieved with an architecture comprising three hidden layers with a neuron configuration of 32-64-128. It is worth noting that further enhancements in accuracy can be attained through more comprehensive dataset preprocessing. In the future, we plan to explore different deep learning models and compare their performance against the results obtained using MLP.

ACKNOWLEDGEMENT

The dataset was generously provided by Mr. Gom Dorji from the electrical department, for which the authors extend their gratitude. Additionally, the authors express their appreciation for the valuable assistance and guidance received from the research officer at the College of Science and Technology

REFERENCES

1. Adamowski, J., & Chan, H. F. (2011). Comparison of two data-driven techniques in modeling and forecasting the standard precipitation index. Journal of Hydrology, 397(3-4), 329-339.
2. Zhu, Y., Wu, Z., & Huang, N. E. (2017). Predicting rainfall for a vulnerable watershed by using a deep learning model. Journal of Hydrology, 548, 604-611.
3. Li, X., Wang, S., & Xu, J. (2017). Rainfall prediction with long short-term memory neural networks. Mathematical Problems in Engineering, 2017.
4. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin Heidelberg New York (1996).
5. Gao, X., Shen, H., & Wang, W. (2019). A hybrid deep learning model for rainfall-runoff modeling. Water, 11(6), 1267.