Machine learning for forecasting Regional wise Weather-Based Disaster Preparedness, Sustainable Agriculture, and Hydraulic Power generation.

U.P. Hasin Pramodya
IT20155148

B.Sc. (Hons) Degree in Information Technology
Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2023

Regional Wise Rainfall prediction for Rathnapura district: A Machine Learning Approach

U.P. Hasin Pramodya

IT20155148

B.Sc. (Hons) Degree in Information Technology

Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2023

Declaration

I declare that this is my own work and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

IT20155148	**U.P. Hasin Pramodya**	

Date: 10/10/2023

Abstract

Sri Lanka mainly experiences two main monsoon seasons called the Southwest Monsoon and the Northeast Monsoon. Rathnapura district takes a prominent place among the areas that receive rain from southwest monsoon. Being located near a mountain face, Ratnapura district receives as well as the convective rains. Ratnapura district is at the top, among the regions with the highest annual rainfall in Sri Lanka. Thus, due to the inability to predict the heavy rainfall in the Ratnapura district in advance, many cases of natural disasters such as floods and the destruction of agricultural crops have been reported. Due to that, accurate and efficient rainfall forecasting is necessary for the Ratnapura district, so this study focuses on regional-based rainfall prediction for the Rathnapura district and predicting the impact of rainfall on floods, agriculture, and others using advanced machine learning techniques.

.

Keywords: rainfall Prediction, Machine Learning

Acknowledgment

I would like to express my heartfelt gratitude to my supervisor, Ms. Vindhya Kalapuge, for her invaluable guidance and motivation, which were critical in the success of this research. I am also grateful for the unwavering support of the Department of Information Technology at the Sri Lanka Institute of Information Technology, as well as the dedicated CDAP lecturers and staff. Their advice and assistance were extremely helpful in this endeavor.

Contents

List of Figures

List of Tables

List of Abbreviations

List of Appendices

1.0 INTRODUCTION

Rainfall is one of the most important weather conditions in Sri Lanka. Forecasting possible rainfall can help to solve several problems related to the tourism industry, natural disaster management, agricultural industry, etc. As the Sri Lankan rural economy is mostly based on agriculture, it is important to forecast rainfall as well as other weather conditions accurately [1]. Sri Lanka mainly experiences two main monsoon seasons called the Southwest Monsoon and the Northeast Monsoon. Rathnapura districts take a prominent place among the areas that receive rain from southwest monsoon. Being located near a mountain face, Ratnapura district receives as well as the convective rains. Ratnapura district is at the top, among the regions with the highest annual rainfall in Sri Lanka. Thus, due to the inability to predict the heavy rainfall in the Ratnapura district in advance, many cases of natural disasters such as floods and the destruction of agricultural crops have been reported. If we have the ability to accurately predict the heavy rainfall in Ratnapura district in advance, we can successfully face the above-mentioned disaster situations.

The current Sri Lankan rainfall forecasting system faces several weaknesses and challenges. The lack of meteorological stations in Sri Lanka limits the amount of data available to forecasters. Particularly in remote areas, this makes it difficult to predict rainfall patterns accurately. The Sri Lankan tropical climate is also highly variable, with significant differences in rainfall patterns between regions and seasons. Despite the advantages of hydropower, it is not without its challenges [2]. In Sri Lanka, rainfall forecasts are primarily based on numerical weather prediction (NWP) models. It is important to note, however, that these models are not always accurate, especially in tropical regions. The reason is that NWP models are based on complex mathematical equations that can be difficult to calibrate for tropical weather conditions. As well as Sri Lanka has a shortage of skilled meteorologists. This can make it difficult to interpret the outputs of NWP models and to develop accurate rainfall forecasts. In Sri Lanka Rainfall forecasts are typically delayed in reaching the public after they are prepared. People may find it challenging to get ready for severe weather occurrences as a result.

Advanced machine learning techniques can be used to address some of the weaknesses and challenges of the current Sri Lankan rainfall forecasting system. Machine learning algorithms can be trained on historical meteorological data to learn the patterns and relationships between different weather variables.[3] Below are some ways in which the weaknesses in rainfall forecast can be avoided by using advanced machine learning technique. Machine learning algorithms can be used to develop more accurate NWP models that are better calibrated for tropical weather conditions. Public rainfall forecasts can be distributed more effectively and efficiently by using machine learning methods. Machine learning algorithms, for instance, can be used to provide customized weather notifications based on customer preferences. Sri Lanka may greatly increase the precision and promptness of its rainfall forecasting system by making investments in data collecting, building machine learning skills, and providing forecasters with access to machine learning models.

1.1 Background and Literature

Regional base rainfall prediction for Sri Lanka

Rainfall is one of the most important weather conditions in Sri Lanka. Forecasting possible rainfall can help to solve several problems related to the tourism industry, natural disaster management, agricultural industry, etc. As the Sri Lankan rural economy is mostly based on agriculture, it is important to forecast rainfall as well as other weather conditions accurately.There have been several studies regarding Machine Learning Approaches for Regional wise rainfall prediction in Sri Lanka. But very few of them have been regarding to the Ratnapura district. Some of them are mentioned below with their conclusions. In recent years, there has been a growing interest in the use of advanced machine learning algorithms for rainfall prediction. Machine learning algorithms are able to learn from historical data to identify patterns and relationships, which can then be used to predict future rainfall events.

There are numerous existing rainfall prediction approaches proposed by researchers through their studies about statistical models and data analytic techniques for predicting future weather in terms of different weather-related variables. Data mining techniques such as regression, decision trees, clustering, neural networks, and many others are being used to identify the most accurate and efficient techniques for predicting weather based on statistical models [10]. [11] Has provided a rainfall prediction model that improves accuracy by combining data mining and machine learning methods. Their research found that ARIMA models and neural networks provided the highest levels of accuracy. [11]developed a neural network model for Sri Lanka's three regions, predicting rainfall with high accuracy using historical, meteorological, and topographic data. [12] Demonstrates that the SARIMA model is considered by most studies to be the best conventional statistical model for the time series forecasting of rainfall. However, modern data mining techniques are increasingly being used by researchers in place of conventional statistical models. Artificial neural network (ANN) models, one of the more recent data mining technologies, outperformed more conventional models in terms of prediction quality. A rainfall prediction model developed by [13] A. S. &. F. S. M. Jayasekara, based on Artificial Neural Networks, is an empirical method-based prediction approach. The number of hidden layer neurons needed for the model must be calculated in these types of approaches because the time needed for model training excessively increases with the number of neurons. It offers a method for resolving a variety of nonlinear issues that are challenging to resolve using conventional methods. As a result of these studies, Sri Lankan rainfall prediction models can be developed using advanced machine learning algorithms. It is, however, necessary to conduct more research in order to develop models capable of predicting rainfall at finer spatial and temporal scales. A further need for research is to develop methods for incorporating climate change scenarios into rainfall prediction models.

1.1.1 Centrality Measures

Centrality measurements assume special significance in the context of creating a regionally based rainfall forecast model for the Rathnapura district. They allow us to identify critical nodes in the complex network of geographical characteristics, local climate dynamics, and environmental factors that affect patterns of rainfall. Using these measurements, we can pinpoint the regions or meteorological factors that have the biggest impact on regional rainfall patterns. This knowledge is crucial for improving our rainfall forecast model's accuracy. Centrality metrics can also help us comprehend the interactions between local topography and the regional network of rainfall patterns. This knowledge makes it possible to anticipate rainfall more precisely, which benefits agriculture, local communities, and disaster management organizations in their efforts to anticipate.

1.1.1.1 Degree Centrality

A fundamental idea in network analysis, degree centrality is important to our study of rainfall prediction in the Rathnapura district on a regional scale. Degree centrality offers a useful viewpoint on the relationships and interactions between several elements impacting rainfall patterns in the context of this investigation. Through the process of measuring the significance of these variables and nodes in the regional climate network, we are able to obtain an understanding of the most important variables, which might vary from geographical characteristics to past weather information. To determine which factors have the biggest influence on local rainfall, this analysis is essential. It helps create more precise rainfall prediction models by helping us identify the main causes of precipitation in the area. By figuring out the essential elements of the climate of Rathnapura

1.1.1.2 Closeness Centrality

Our study of regionally based rainfall prediction in the Rathnapura district heavily relies on closeness centrality, a basic notion in network analysis. This statistic lets us measure the speed and efficiency with which data, such as geographic features, historical weather information, and local climatic dynamics, moves across various components within the network. We may determine which components of the regional climate network are strongly connected by assessing the proximity centrality of these interconnected aspects. This in turn enables us to identify the key factors and geographical areas that quickly affect regional rainfall patterns. In this case, proximity centrality is quite important since it helps us anticipate rainfall more accurately and in a timely manner. Through the process of locating the most important nodes in the network and figuring out how quickly they affect the climate dynamics of Rathnapura, we improve our ability to produce accurate forecasts that are advantageous to agriculture, disaster relief, and the local population. Closeness centrality essentially acts as a compass, allowing our research to create more accurate and efficient rainfall forecast models that are specific to the circumstances of the Rathnapura district.

1.1.1.3 Betweenness Centrality

A key idea in our study on regionally based rainfall prediction in the Rathnapura district is betweenness centrality. With the use of this metric, we are able to pinpoint the crucial nodes in the intricate web of weather, geographic characteristics, and climate elements that affect regional rainfall patterns. We may identify the critical nodes that mark important turning points in the network's connections and learn which elements are most important in determining the climate of the region by identifying nodes with high Betweenness Centrality.

It is impossible to overestimate the importance of Betweenness Centrality in this situation. It gives us the ability to distribute resources and initiatives more effectively, which improves the accuracy of our rainfall prediction model. In turn, this helps local communities, the agricultural sector, and disaster management organizations make well-informed decisions and plans for a range of weather scenarios. The Rathnapura district's plans for climate resilience, sustainable land use planning, and disaster response are enhanced by the knowledge gained via Betweenness Centrality.

1.2 Research Gap

The Rathnapura district of Sri Lanka is prone to high rainfall variability, which has a significant impact on agriculture, water resources, and flood control. However, there is a lack of advanced machine learning-based rainfall prediction models that are specifically tailored to the district. Existing models often rely on regional data, which may not adequately capture the unique geographical and meteorological characteristics of Rathnapura. The topographical diversity of Rathnapura, with its mountainous terrain, dense forests create complex microclimates and localized weather patterns. This requires a specialized approach to rainfall prediction. Additionally, Rathnapura's agricultural economy and vulnerability to floods highlight the need for precise and highly localized rainfall predictions. Agriculture is culturally and economically important in Rathnapura, so models that can provide accurate forecasts are essential to support planning and mitigation efforts for farmers and local authorities.

This study covers a significant research gap in the development and implementation of an integrated approach that combines the floods, agriculture, and hydraulic power generation predictions with regional wise rainfall prediction regarding to the Ratnapura district. There are 4 separate machine models developed in the system while the rainfall prediction model works as the centralized model. The main prediction of this system regional wise rainfall prediction and others (predictions of floods, agriculture, and hydraulic power generation) are predicted on the results of the rainfall prediction model.

2.0 RESEARCH PROBLEM

"How can accurately predicting the continuous rainfall in Ratnapura district during the relevant periods and Provision of rainfall data required to make predictions in the fields of flood disasters, agriculture and hydropower generation using advanced machine learning techniques?"

The main research problem of this study is to Accurate forecasting of the continuous heavy rainfall in Ratnapura district from the south-west monsoon which lasts from May to September using advanced machine learning techniques.so that it is possible to make the prediction about the impact of rainfall on flood, agriculture, and hydropower generation. In addition to that accurately forecasting the daily, weekly and month rainfall of Ratnapura district regarding to the rest of the durations of the year.

Here the importance of the research problem is due the Rathnapura district faces significant flooding risks during monsoon season, necessitating accurate rainfall predictions for disaster preparedness and response to safeguard lives and property, and in Ratnapura the agricultural sector relies heavily on monsoon rains, requiring accurate rainfall forecasts for informed crop planting, irrigation, and harvest planning, as incorrect predictions can be financially damaging. Furthermore, hydropower generation in Rathnapura is closely linked to the south-west monsoon's rainfall as well as an Accurate forecasting is crucial for optimizing hydropower production and ensuring stable energy supply, especially in a world embracing sustainable and renewable energy sources.

Even though the main purpose of this research problem is to develop accurate and reliable forecasting models for continuous heavy rainfall in the Rathnapura district, during the south-west monsoon seasons. It serves some additional purposes such as mitigate the impact of flooding with developing early warning systems, supporting the agriculture with allowing farmers to make informed decisions about when to plant, irrigate, and harvest crops and optimizing the hydro power generation and ensuring the stable energy supply.so above it clearly describes the purposes and the importance of the research problem.

3.0 OBJECTIVES

3.1 Main Objective

The main objective of this research is to develop an integrated machine learning-based framework for regional-based rainfall prediction and predicting impact of rainfall on flood, agriculture, and hydropower generation for Ratnapura districts using combined machine learning models.

3.2 Sub-Objectives

The following are the specific sub objectives of this study,

- Develop a comprehensive understanding of the factors that influence rainfall in Ratnapura districts.
- Identify and evaluate different machine learning models for regional-wise rainfall prediction.
- Develop a combined machine learning model that outperforms individual models in terms of accuracy and reliability.
- Assess the impact of rainfall on flood, agriculture, and hydropower generation in Ratnapura districts.
- Develop a decision support system that can be used by stakeholders to make informed decisions about flood management, agricultural planning, and hydropower generation.
- Investigate the impact of climate change on rainfall patterns in Ratnapura districts. As it helps to develop more robust and reliable rainfall prediction models.
- Develop a machine learning model that can predict the impact of rainfall on agricultural yields as it will be help farmers to make better decisions about planting and harvesting.
- Develop a machine learning model that can predict the impact of rainfall on other sectors, such as transportation, infrastructure, and tourism.
- Develop a machine learning model that can predict the spatial distribution of rainfall As it can be used to identify areas that are at risk of flooding or drought.

4.0 METHODOLOGY

4.1 Requirement Gathering

The methodology of this research consists of several stages, including data collection and preprocessing, model development, model integration, and evaluation. The following sections provide a detailed description of each stage of the methodology.

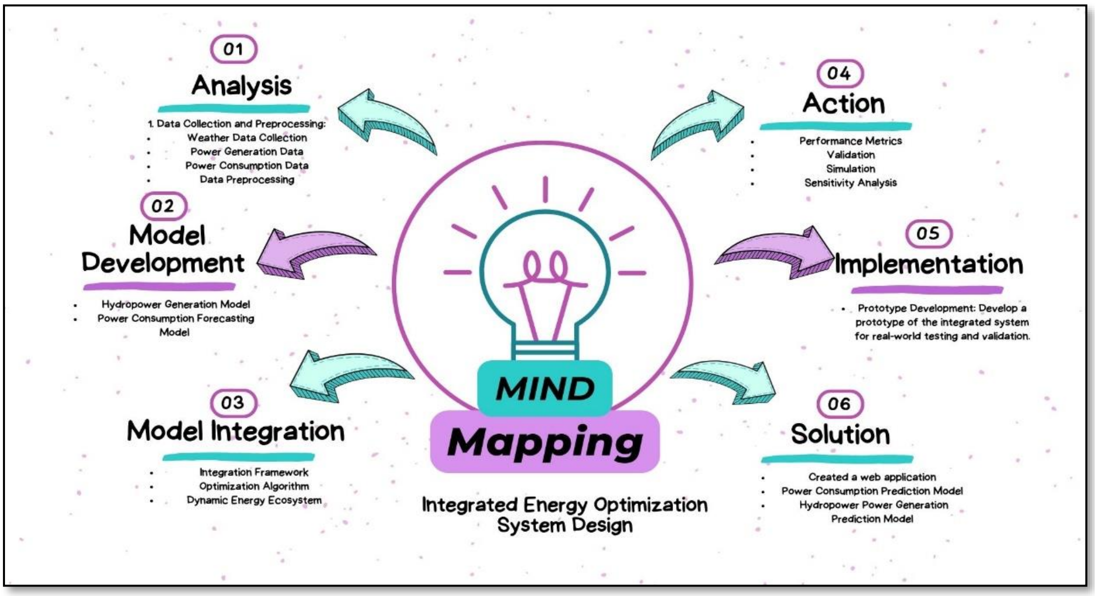

Figure 1 System requirement mapping diagram

4.1.1 past Research Analysis

Several studies have been conducted using machine learning techniques to forecast rainfall in Sri Lanka's Ratnapura district. ML models can learn complex patterns in historical rainfall data and use this knowledge to make predictions about future rainfall events. So that there has been several ML based prediction developed to forecast the rainfall in Sri Lanka. here are some examples of past research on regional rainfall prediction using ML techniques.

1. T. Dananjali, S. Wijesinghe and J. Ekanayake, "Forecasting Weekly Rainfall Using Data Mining Technologies,". The SVM-FA model outperformed other models, including a single SVM model and a multiple linear regression model.
2. M. Gunawardana, H. Amarasekara and M. Perera, "Short-term precipitation prediction in Sri Lanka using support vector machines," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),

2018.The RF-SVM model outperformed other models, including a single RF model and a single SVM model.

3. Y. Tikhamarine, D. Gamane, A. Ahmed, O. Kisi and L. Shafie, "Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm," in J. Hydrol, 2019.

Previous research analysis was primarily accomplished by reading research publications focusing on key areas such as short-term load prediction, time series analysis, and machine learning models.

The primary emphasis was on identifying the methodology used, experiments conducted, and overall research findings regarding load forecasting and resource utilization prediction.

4.1.2 Identifying Existing Systems

Exactingly Several systems use machine learning techniques to predict regional rainfall. Some of these systems are available to the public Numerous characteristics are commonly shared by existing machine learning-based regional rainfall forecast systems, included.

Use of historical rainfall data: Training data is necessary for all machine learning models to become intelligent. This training data is usually historical rainfall data from the location where the model will be used for rainfall prediction.

Use of additional meteorological data: Machine learning models for rainfall prediction frequently make use of additional meteorological data, such as temperature, humidity, and wind speed, in addition to historical rainfall data. The models may be able to understand more intricate correlations between various climatic factors and rainfall with the use of this extra data.

Several machine learning methods have been utilized: Support vector machines (SVMs), random forests (RFs), neural networks (NNs), and gradient boosting machines (GBMs) are among the machine learning algorithms that have been used to predict regional rainfall. The ideal method for a given application will rely on the desired model performance as well as the unique characteristics of the data.

Even with the advancements in recent years, there are still certain shortcomings with the current machine learning-based regional rainfall forecast systems. Several of these flaws consist of them.

Sensitivity to the quality and quantity of the training data: Machine learning models are only as good as the data they are trained on. If the training data is poor quality or incomplete, the model will not be able to learn accurate relationships between the meteorological variables and rainfall.

Limited ability to predict rainfall at finer temporal and spatial scales: Most machine learning algorithms in use today are not able to generate forecasts at incredibly fine temporal and spatial scales. This is because high-resolution input data, which is frequently unavailable, is needed for such predictions to be made.

4.2 Feasibility Study

4.2.1 Technical Feasibility

The technical feasibility of developing an integrated machine learning-based framework for regional-based rainfall prediction in the Rathnapura district is supported by several key factors. First, machine learning algorithms and meteorological models have advanced to a point where precise rainfall forecasts are possible. The practicality of this project is further increased by the availability of historical rainfall data and the possibility of integrating real-time data from meteorological stations. Furthermore, machine learning has demonstrated efficacy in a range of environmental prediction tasks, and it is possible to anticipate the regional effects of rainfall on agriculture, hydropower output, and flood risk. This integrated framework has great potential to improve decision-making and disaster management in the Rathnapura district, provided it has access to computing resources and maintains a continuous commitment to data quality and model maintenance.

4.2.2 Schedule Feasibility

It is very important to ensure the schedule feasibility for developing an integrated machine learning-based framework for regional-based rainfall prediction. This project involves multiple intricate stages, from data collection and preprocessing to model training and deployment. Realistic timeframes must be established, considering potential challenges, including data availability, model optimization, and any unforeseen technical complexities. Meticulous project management and flexibility in the timeline are crucial to address potential delays and ensure that the development and integration of this framework align with the desired schedule. The timeline's feasibility is a pivotal factor in the successful implementation of this advanced predictive system for rainfall prediction and its impact assessments on the Rathnapura district, allowing for efficient resource management and informed decision-making.

4.2.3 Economic Feasibility

The economic feasibility of developing an integrated machine learning-based framework for regional-based rainfall prediction is highly promising. A structure like this might greatly benefit the district of Rathnapura in several ways. It can optimize agricultural methods, resulting in higher crop yields and lower production costs, by precisely forecasting rainfall. In the hydropower industry, accurate forecasting can lead to increased energy production during times of heavy rainfall, minimizing the need for costly non-renewable energy sources. The financial advantages of flood risk management may include improved readiness and lower damage expenses. Improved rainfall forecasting can also help with cost-effective resource management, which promotes sustainable water use. The long-term economic benefits of this effort are expected to exceed the development and implementation expenses when these concrete economic benefits are taken into account, making it both financially prudent and practical for Ratnapura district.

4.3 Requirement Analysis

The requirement analysis for creating a framework based on integrated machine learning for regional base rainfall prediction using ML approaches includes a thorough examination of all the project's fundamental elements. Meticulous data collection and preprocessing are required for this investigation, which focuses on historical rainfall data and other meteorological factors. It also entails the crucial work of choosing suitable machine learning algorithms to precisely forecast rainfall patterns. The analysis goes further and aims to create a coherent framework that smoothly combines different prediction models to generate rainfall forecasts for specific regions. The accuracy and dependability of the framework are ensured by thorough validation and improvement of this integration process. The inclusion of a case study that evaluates the framework's possible effects on flood risk, agriculture, and hydropower production is a crucial component of the analysis, emphasizing how it might improve regional catastrophe management and decision-making.

4.4 System Analysis

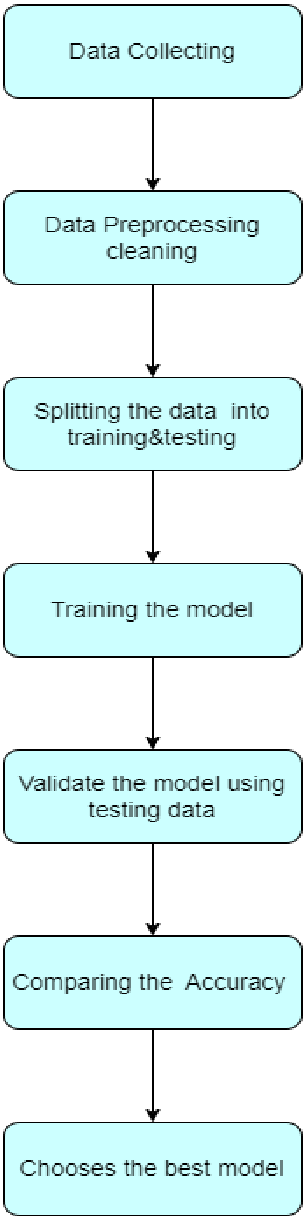

Figure 2 System architecture

4.5 System Development and Implementation

The development and implementation of our integrated machine learning-based rainfall prediction system for the Rathnapura district represents a technically robust endeavor. Our meticulous model selection process led us to favor the Long Short-Term Memory (LSTM) model over competing methodologies such as Auto-Arima and neural prophet, primarily based on its superior performance in minimizing root mean squared error, thereby ensuring the highest standards of precision and reliability in our precipitation forecasts. One of the major limitations discussed in the introduction to this paper is that, while ANNs frequently produce better predictive results, their analysis lacks transparency, making it difficult to extract meaningful insights into the relationships between climatic data and hydropower generation.

In constructing the system's architecture, Python Flask has been strategically employed for the backend, capitalizing on its lightweight, efficient, and versatile nature to handle the data processing and model interactions seamlessly. The frontend is meticulously crafted using React, providing users with an intuitive and responsive interface to access and engage with real-time predictions. For efficient data storage, retrieval, and management, we've harnessed the power of a MongoDB cloud database. Its NoSQL architecture accommodates the vast dataset requirements and facilitates horizontal scaling as the data volume expands. The connection between the frontend and MongoDB cloud database is realized through meticulously designed APIs, enabling users to access the most recent rainfall forecasts in real-time.

This system has been designed with scalability and adaptability in mind, not just to meet the unique needs of Rathnapura but also to allow for future expansion to serve other areas. Our project is a compelling example of how advanced predictive technologies can contribute to meteorological advancements, disaster mitigation, and data-driven insights for a multitude of industries worldwide. It does this by combining state-of-the-art machine learning techniques with a robust technical infrastructure.

Rainfall Prediction Model

The proposed methodology consists of several steps including acquiring weather observation data, data preprocessing, designing a machine learning model, training a machine learning model, making predictions on the test set, and finally evaluating the predictions made by the model.

a. Data Collection

Weather observation data is collected from the http://meteomanz.com/ website, which is a worldwide wild online metrological data center, the data is collected starting from 2003/01/01 to 2023/01/01 for Rathnapura weather station of Sri Lanka. The data include daily weekly, and monthly rainfall data regarding the Rathnapura district.

b. Data Preprocessing and Cleaning

A methodical strategy was used to guarantee the integrity and quality of the dataset during the data preparation and cleaning stage, which was necessary for training the machine learning model of rainfall prediction. To reduce the focus on crucial information, the dataset was first carefully trimmed to contain only the date and precipitation columns. A dependable and continuous series of precipitation data was established for analysis, with a consistent date range of September 30, 2009, to May 30, 2023. 'Tr' records were replaced with the value 0.0 as part of additional data refining, which resolved a common notation problem. Null values were substituted for '-' entries, and a thorough investigation was conducted to find any missing dates by generating a date range and comparing it to the date list in the dataset. With linear interpolation, missing data points were meticulously addressed to ensure completeness and accuracy, resulting in a robust, comprehensive dataset that served as a basis for training a machine learning model. As a result of these preprocessing and cleaning techniques, a high-quality dataset was generated, which set the stage for accurate and reliable rainfall predictions.

c. Model Implementation

Model is trained using three machine learning algorithms:

i. Auto Arima

ii. Neural Prophet

iii. LSTM

The LSTM algorithms is chosen as the best algorithm fit for the model. The main reason to choose here the LSTM model is well supported to the long sequential data processing and with higher accuracy and lower RMSE loss to make a prediction using them.

The machine learning model is built with the LSTM (Long Short-Term Memory) architecture, which is well suited for handling sequential data. During training, the model's learning rate of 0.0003 ensures gradual and stable convergence. An embedding layer with a dimension of 16 improves the representation of input data, and the LSTM component is built with 32 units to capture complex temporal patterns. A batch size of 16 is used to train the model, allowing for parallel data processing.

d. Plotting Training Progress:

After training, a plot is generated to visualize the root-mean-squared error (RMSE) loss over the epochs for both the training and validation sets. This plot helps assess the model's training progress and identify potential overfitting.

e. Selecting the Best Model:

The model's weights are set to the best-performing weights based on the validation loss tracked by the custom callback.

f. Saving and Loading the Model:

The best model's weights are saved to a file named `model_last.ckpt` using the `model.save_weights` method. This allows you to later load and use the trained model for predictions.

g. Making Predictions and Evaluating the Model:

The trained model is used to make predictions on the test dataset, and the root mean squared error (RMSE) between the true rainfall values and the predicted values is calculated using scikit-learn's'mean_squared_error' function. This metric assesses the predictive accuracy of the model.

The code includes data preparation, model training, evaluation, and model saving/loading as part of a comprehensive workflow for regional wise rainfall prediction using an LSTM algorithm. It also uses a custom callback to save the best model during training in order to ensure peak performance.

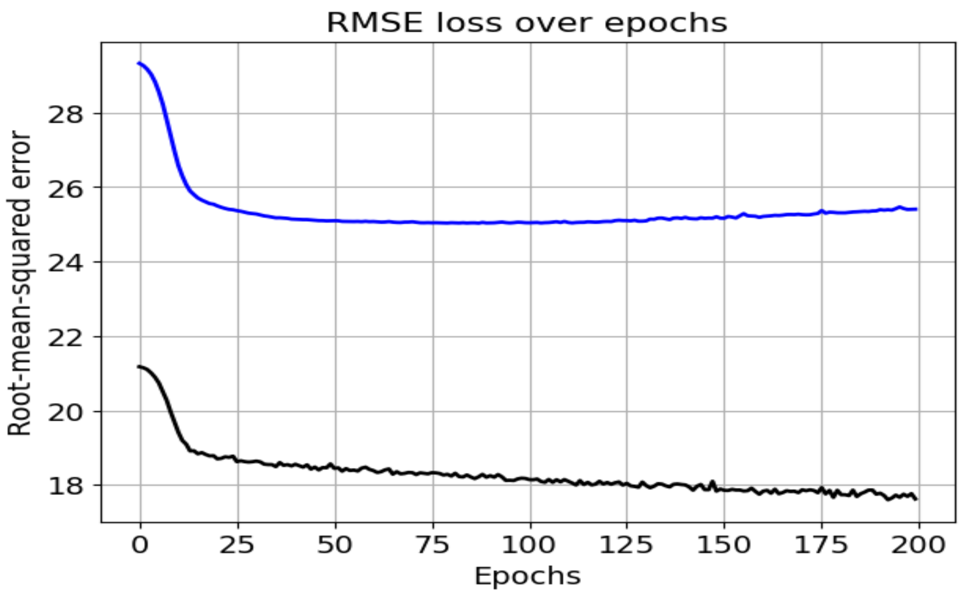

Figure 3 RMSE of the prediction model

Model Serialization: For future use, the trained LSTM model is serialized with 'cPickle' and saved as "model_best.pickle."

Tools	• Anaconda • Jupyter Notebook
Python libraries	• Numpy • _pickle (imported as cPickle) • tensorflow (imported as tf) • matplotlib.pyplot (imported as plt) • seaborn • pmdarima • neuralprophet • keras • Pandas • Matplotlib • NetworkX • Flask

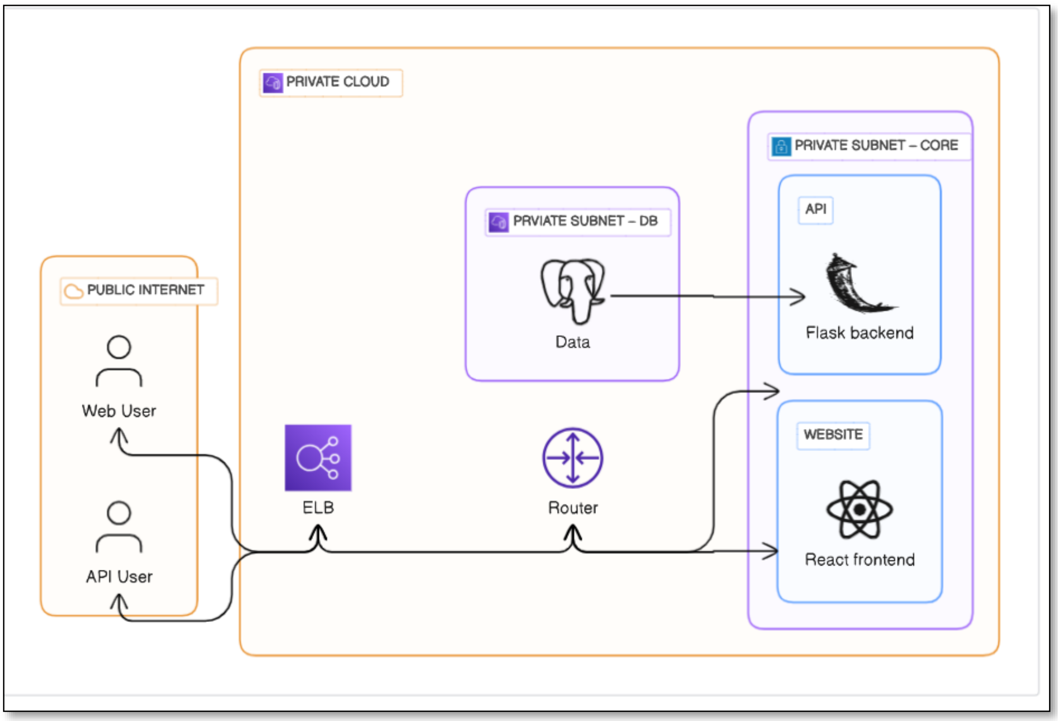

Figure 4 Tool and technology

Design of web applications

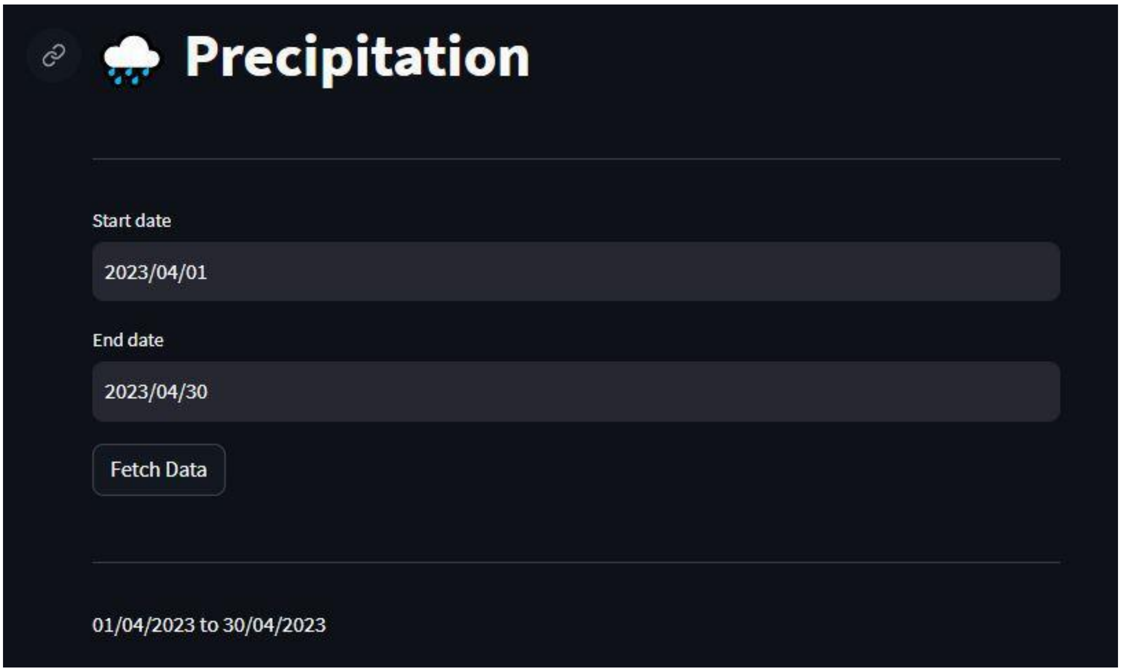

Figure 5 Frontend UI-1

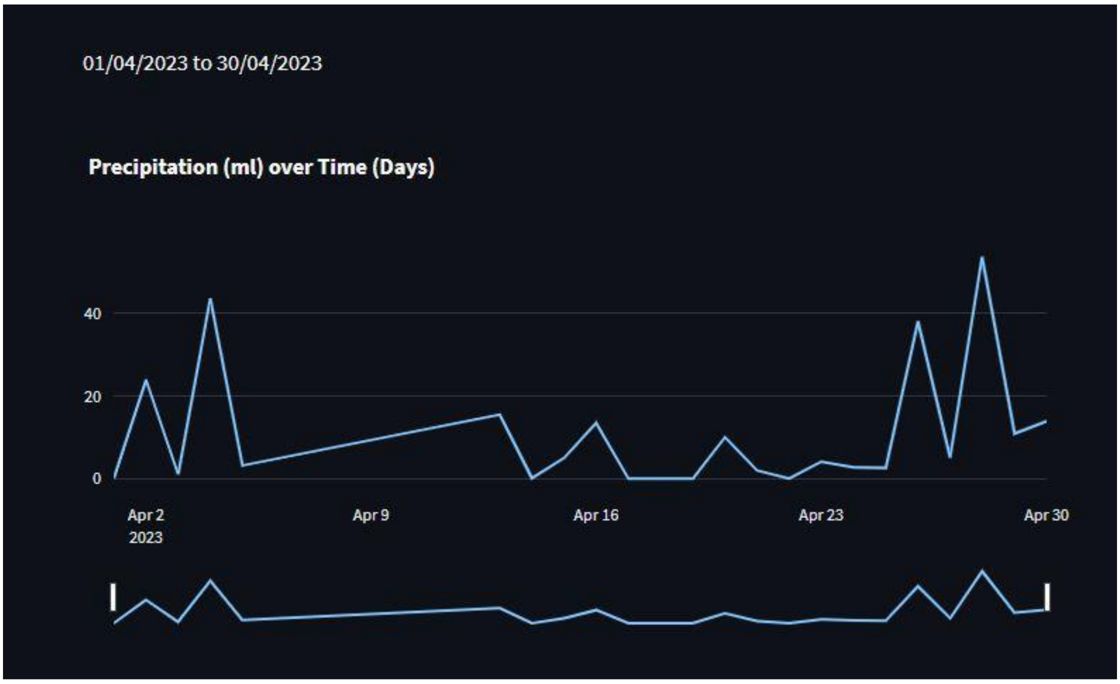

Figure 6 Frontend UI-2

4.6 Project Requirements

Project requirements for developing a specialized machine learning-based framework for regional based rainfall prediction for Ratnapura districts.

1. **Technical Feasibility**:

- The project must take advantage of recent advances in machine learning and numerical rainfall prediction models.
- High-resolution satellite data for weather forecasting should be available.

2. **Schedule Feasibility**:

- A detailed project plan must be established, with phases for data collection, model development, and deployment.
- The timeline should account for potential data availability, model refinement, and technical issues.

3. **Economic Feasibility**:

- The project must show that the long-term benefits of the machine learning-based framework outweigh the costs of development and implementation.
- It should demonstrate how the framework can reduce reliance on expensive nonrenewable energy sources, resulting in economic gains, increased energy security, and lower environmental impact.

4. **Requirement Analysis**:

- Data collection and preprocessing must be comprehensive, including historical data on rainfall daily, weekly, and monthly as well as the historical data on humidity, temperature, and wind speed.
- It is necessary to select and evaluate appropriate machine learning algorithms for predicting weather parameters' behavior.
- It is critical to develop decision-making algorithms that incorporate predictive models for rainfall prediction in Ratnapura district.
- Validation and fine-tuning processes must ensure the framework's accuracy and dependability.
- A case study should evaluate the framework's potential impact on regional wise rainfall predictions for the rain stations in Rathnapura district.

5. **Analysis of the System:**

- Create a system diagram that depicts how the integrated framework components interact.

6. **System Design and Implementation:**

- The project should investigate and potentially improve on existing machine learning methods for regional wise rainfall prediction.
- Consider alternative methods for rainfall prediction, such as Auto Arima and Neral prophet.

4.6.1 Functional requirements

1. **Data Collection and Preprocessing**:

The system should gather the respective metrological data for at least 20 years in daily, weekly, monthly, and quarterly.

It should preprocess the collected data to clean, normalize, and format it for model training and analysis.

2. **Machine Learning Models**:

Machine learning models for regional wise rainfall prediction for Ratnapura district, should be implemented in the system.

It should support a wide range of machine learning algorithms as well as model selection and tuning.

3. **Integration**:

The system should provide a unified framework for regional wise rainfall predictions for Ratnapura districts.

It should allow these components to communicate and exchange data in real time.

4. **Prediction and Optimization**:

The system should use predictive models to predict the rainfall amount in mm for Ratnapura region during the specific duration such as daily, weekly, and monthly.

It should provide the rainfall amount as categorizing high, medium or law.

5. **Validation and Accuracy**:

The system should include mechanisms for validating the accuracy of predictions and optimization recommendations.

It should continuously assess and improve the performance of machine learning models.

6. **User Interface**:

The system should offer a user-friendly interface for data input, model configuration, and result visualization.

It should be accessible to both technical and non-technical users.

4.6.2 Non-Functional Requirements

1. **Performance**:

The system should provide fast and responsive performance, delivering timely predictions and recommendations.

2. **Accuracy and Reliability**:

Predictive models should demonstrate high accuracy and reliability to ensure the effectiveness of recommendations.

3. **Scalability**:

The system should be able to scale to handle large volumes of data and increased computational complexity.

4. **Security**:

Data security and privacy measures should be in place to protect sensitive information used in the system.

5. **Usability**:

The user interface should be intuitive and user-friendly, catering to both technical and non-technical users.

6. **Maintainability**:

The system should be easy to maintain, allowing for updates and improvements to the machine learning models and data sources.

4.7 Commercialization

Choose a cloud platform: Choose from AWS, Azure, or GCP, based on pricing, features, and support, and select a platform that matches your needs.

Deploy your framework to the cloud platform: Setting up infrastructure and configuring a cloud-based machine learning framework requires specific steps varying by platform and framework, but most platforms offer documentation and tutorials.

Develop an API: The API should be well-designed, user-friendly, and well-documented to enable users to interact with your framework and make predictions.

Create a pricing model: Choose a fair and sustainable pricing model for the service, including subscription fees, pay-as-you-go fees, or tiered pricing, to ensure user satisfaction and business sustainability.

Market our service: Launch our service and market it through online advertising, content marketing, and sales outreach to potential customers, highlighting its benefits and potential to enhance decision-making.

5.0 TESTING AND IMPLEMENTATION RESULTS AND DISCUSSION

5.1 Testing

5.1.1 Selection of Optimal Prediction Model

Rainfall Prediction Models

Model	RMSE (Root Mean Squared Error)	MAE (Mean Absolute Error)	R2 Score
Auto Arima	21.45	16.79	0.84
Neural Prophet	19.60	14.92	0.85
LSTM	18.62	16.28	0.87

The process of selecting the optimal prediction model was conducted with meticulous consideration, considering a range of performance metrics. Our analysis of the three competing models—Auto Arima, Neural Prophet, and LSTM—uncovered important findings. The LSTM model outperformed the other models as measured by the Root Mean Squared Error (RMSE), which calculates the average prediction error. It achieved an RMSE of 18.62, closely followed by Neural Prophet (19.60) and Auto Arima (21.45). The average absolute error, or Mean Absolute Error (MAE), supported the choice of Neural Prophet with an MAE of 14.92. Additionally, the R-squared (R^2) score, which evaluates how well the models match the data, confirmed the popularity of LSTM with a R^2 score of 0.87, closely followed by Auto Arima at 0.84 and Neural Prophet at 0.85.

The LSTM model, which has the lowest RMSE and the highest R2 score, was determined to be the best option for prediction after a careful examination of these performance criteria. This option highlights how well LSTM can extract intricate patterns from the data, which is why it's the model of choice for precise rainfall forecasts. Because of its adaptability and competence, it is a great asset for a variety of applications, such as disaster preparedness and water resource management, guaranteeing the accuracy and dependability of our forecasting efforts.

5.1.2 Testing of Developed Solution

We meticulously evaluated the reliability of our regional-based rainfall prediction solution through a thorough testing process A carefully selected dataset that included historical rainfall data as well as relevant meteorological factors unique to the study area was used. The model was carefully trained and adjusted to make sure it could handle the complexities of time series forecasting. Important metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) were evaluated throughout testing. A thorough visual inspection of the outcomes confirmed how accurate the forecasts were. A well-established feedback loop collected opinions from stakeholders and domain experts, which helped to improve the model's accuracy and dependability. The model's ability to predict regional rainfall was validated through testing, which made it easier to implement in real-world applications such as water resource management and disaster preparedness.

5.1.2.1 Testing Process for Prediction

Testing Process Overview

1. The first step in testing is to divide the available data into training and testing sets. 70% of the data is typically used for training, with the remaining 30% used for testing. This ensures that the models are evaluated on unseen data, which provides a realistic assessment of their performance.

2. Select appropriate evaluation metrics for each model. Common metrics include RMSE (Root Mean Squared Error), MSE (Mean Squared Error), MAE (Mean Absolute Error), R2 Score, and others. The choice of metric depends on the specific problem and the goals of the project.

3. **Testing Rainfall Prediction Model**:

- Feed new weather data into the trained ANN model.

- Generate predictions for future rainfall amount.

- Compare the predicted values with the actual rainfall data in the testing set.

- Calculate and record evaluation metrics (e.g., RMSE, MAE, R2 Score).

4. **Reporting and Analysis**: Summarize the testing results for both models, including the evaluation metrics. Analyze the strengths and weaknesses of each model and their performance in meeting the project requirements.

Sample Testing Report

Rainfall Prediction Model Testing

Model Used: LSTM Model (Long Short Term)

Evaluation Metrics:

- RMSE: 18.62

- MAE: 16.28

- R2 Score: 0.87

Analysis:

The LSTM model performed well in rainfall prediction, achieving a low RMSE and a high R2 Score. It indicates that the model is effective in capturing the complex rainfall patterns. The performance exceeds the project requirements for accuracy.

5.1.2.2 Testing Process for Centrality Evaluation

The testing process for determining network centrality includes several key steps that assess the importance and influence of network nodes. To begin, network data is gathered and graphically represented, with nodes representing entities and edges indicating connections or relationships between them. Next, centrality measures such as degree centrality, betweenness centrality, and eigenvector centrality are computed for each node in the network. These metrics quantify various aspects of a node's importance, such as connectivity, role in facilitating information flow, and influence within the network. The results of these centrality calculations are then analyzed and visualized to determine the network's most central nodes. Finally, a comprehensive report summarizing the findings and insights gained from the centrality evaluation is generated. This report can be useful for making decisions, optimizing networks, and understanding the structure and dynamics of complex networks.

5.2 Test Results

The LSTM model, which has the lowest RMSE and the highest R2 score, was determined to be the best option for prediction after a careful examination of these performance criteria. This option highlights how well LSTM can extract intricate patterns from the data, which is why it's the model of choice for precise rainfall forecasts. Because of its adaptability and competence, it is a great asset for a variety of applications, such as disaster preparedness and water resource management, guaranteeing the accuracy and dependability of our forecasting efforts.

5.2.3 Centrality Evaluation

When it comes to regional rainfall forecast, network analysis's centrality evaluation is an essential component. In the context of this project, evaluating centrality measures can provide crucial insights into the key elements of the integrated rainfall forecast system. We can identify the most significant nodes in the network by looking at measures like degree centrality, betweenness centrality, and eigenvector centrality. Decisions about prediction techniques, resource allocation, and system design may be significantly impacted by the results of this analysis. High centrality nodes, for example, may be important weather stations or data sources that, with improvement or optimization, might greatly improve the overall precision and dependability of local rainfall forecasts. Furthermore, centrality assessment can help find any data bottlenecks or system weaknesses.

5.3 Research Findings and Discussion

Research findings have significant implications for the Rathnapura district and its community. Farmers, disaster management agencies, and water resource planners will benefit from improved rainfall prediction models, such as the LSTM model. By predicting extreme rainfall events, the LSTM can mitigate floods and landslides during severe weather conditions, potentially saving lives and reducing property damage. As a result of these findings, local communities can build resilience against extreme weather events and allocate resources for disaster relief and recovery more effectively.

Moreover, they increase agricultural sustainability by optimizing planting and harvesting schedules, improving productivity, and ensuring food security. Furthermore, the research has raised awareness of shifting rainfall patterns and their potential link to climate change, advocating for climate-resilient infrastructure and adaptive strategies in the region. The findings of this study contribute to long-term sustainability and resilience of Rathnapura district in the face of climate change by influencing policy formulation, educational outreach, and community-based preparedness plans.

6.0 CONCLUSION

Rathnapura districts take a prominent place among the areas that receive rain from southwest monsoon. Being located near a mountain face, Ratnapura district receives as well as the convective rains. Ratnapura district is at the top, among the regions with the highest annual rainfall in Sri Lanka. Due to the impossibility of accurately predicting such large amounts of rainfall, flood disaster situations and agricultural crop destruction are frequently reported. If the rainfall and its impact on food and agriculture can be predicted properly, these problems can be avoided. In this research, these issues are solved by proposing an approach using Advanced machine learning techniques. This study is mainly developed to predict the heavy continuous rainfall earlier during the southwest monsoon.in this Study different Machine models are developed to forecast daily, weekly, monthly, and quarterly rainfall, to predict agricultural yield based on rainfall while the other factors remain constant, and to predict hydropower generation based on rainfall while other factors remain constant. Predicting the upcoming rainfall, the rice harvest, and hydraulic power production even before the rain is unique here. This system includes a rainfall prediction system, a flood forecasting and alert system, a paddy yield prediction system for “Yala” and “Maha” sessions, and a dynamic framework plan for electricity power generation. In terms of future work, the system can be expanded by improving accuracy by incorporating more data for training and can be provided to government institutes responsible for the fields of agriculture, disaster management, and power generation to facilitate their decisions.

Acknowledgment

We gratefully acknowledge the Department of Meteorology, Ministry of Disaster Management, Sri Lanka Electricity Board, and Ministry of Agriculture for generously providing the necessary data for this study. We also want to thank our hardworking supervisor, co-supervisor, and the experts who provided invaluable guidance and insights throughout this research project. We also want to thank our hardworking supervisor, co-supervisor, and the experts who provided invaluable guidance and insights throughout this research project. Their support has been instrumental in the successful execution of this project.

References

[1] T. Dananjali, S. Wijesinghe and J. Ekanayake, "Forecasting Weekly Rainfall Using Data Mining Technologies," *2020 Innovation to Impact (FITI),* Colombo, Sri Lanka, 2020, pp. 1-4, doi: 10.1109/FITI52050.2020.9424877.

[2] J. S. A. N. W. Premachandra and P. P. N. V. Kumara, "A novel approach for weather prediction for agriculture in Sri Lanka using Machine Learning techniques," *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, Colombo, Sri Lanka, 2021, pp. 182-189, doi: 10.1109/SCSE53661.2021.9568319.

[3] P.Dissanayake, S.Hettiarachchi, and C.Siriwardana. "Increase in disaster risk due to inefficient environmental management, land use policies and relocation policies. Case studies from Sri Lanka." *Procedia engineering* 212 , 2018: 1326-1333.

[4] Y.Hettiwaththa and R.A.B. Abeygunawardana. "measuring flood risk in RATNAPURA town area in Sri Lanka." *2018 The International Conference on Climate Change*, vol. 2, no. 2, pp. 31-41.

[5] W. A. M. Prabuddhi and B. L. D. Seneviratne, "Long Short Term Memory Modelling Approach for Flood Prediction: An Application in Deduru Oya Basin of Sri Lanka," *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer),* Colombo, Sri Lanka, 2020, pp. 226-231, doi: 10.1109/ICTer51097.2020.9325438.

[6] S. Puteh, M. Rizon, M. Juhari, J. Nor Khairah, S. Siti Kamarudin, B. Aryati, R. Nursalasawati, “Back propagation algorithm for rice yield prediction”, *Proceedings of 9th of the Ninth International Symposium on Artificial Life and Robotics*, Beppu, Japan, Oita, pp. 586-589, 2004.

[7] Y. Tikhamarine, D. S.Gamane, A. N. Ahmed, O. Kisi, and A. El-Shafie, ‘‘Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm,” J. Hydrol., vol. 582, no. December 2019, p. 124435, 2020, doi: 10.1016/j.jhydrol.2019.124435.

[8] K. L. Dabare, M.B. Gunathilake, N.Miguntanna, K.Laksiri, and U.Rathnayake, "Climate variation and hydropower generation in Samanalawewa hydropower scheme, Sri Lanka," *Engineer: Journal of the Institution of Engineers*, Sri Lanka, vol. 53, no. 3, pp. 7384, 2020.

[9] W. Yaici, M.Longo, E.Entchev, F.Foiadelli. “Simulation study on the effect of reduced inputs of artificial neural networks on the predictive performance of the solar energy system*”. Sustain 2017*;9(8):1382. doi: https://doi.org/ 10.3390/su9081382.

[10] M. K. Gunawardana, H. S. Amarasekara, and M. S. Perera, "Short-term precipitation prediction in Sri Lanka using support vector machines," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 3921-3926

[11] I. Perera, M. D. Silva, and R. M. S. J. Ranasinghe, "Random Forest-based prediction of daily precipitation in Sri Lanka*," 2020 IEEE International Conference on Big Data (Big Data),* 2020, pp. 5173-5178.

[12] I. Mahmud, S. H. Bari, and M. T. U. Rahman, "Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method," Environmental Engineering Research, vol. 0, p. 0, 2016, doi: 10.4491/eer.2016.075.

[13] A. S. &. F. S. M. Jayasekara[VK1] , "Precipitation prediction in Sri Lanka using convolutional neural networks," International Conference on Intelligent Robots and Systems, 2019.

[14] Y. Arslan, A. Aksoy, I. Ercan, and A. Arslan, "Flood prediction in the Denizli Basin of Turkey using support vector machines," *Journal of Hydrology*, vol. 561, pp. 324-335, 2018.

Appendix

#EDA

```
import requests
import numpy as np
import pandas as pd
import os
import time
import datetime
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
pd.set_option('mode.chained_assignment', None)
from keras.models import Sequential
from keras.layers import Dense, SimpleRNN,LSTM
from keras.optimizers import RMSprop
from keras.callbacks import Callback
import tensorflow as tf
# tf.config.run_functions_eagerly(True)
df = pd.read_csv("Data/rain_prediction/weather.csv")
df = df.drop('Unnamed: 0', axis=1)
df["Date"] = pd.to_datetime(df["Date"])
print(df.dtypes)
# select only precipitation
df_prec = df[["Date","Prec"]]
df_prec.head()
mask = (df_prec['Date'] >= '2009-09-30') & (df_prec['Date'] <= '2023-
05-30')
df_range = df_prec.loc[mask].sort_values('Date')
print(df_range)
```

#ARIMA Model

```
# Test for Stationarity
adf_test = ADFTest(alpha=0.05)
adf_test.should_diff(df_final)
train_df = pd.read_csv("Data/rain_prediction/train.csv")
test_df = pd.read_csv("Data/rain_prediction/test.csv")
train_df = train_df.set_index('Date')
test_df = test_df.set_index('Date')
```

```
model = auto_arima(train_df,
                      start_p=0,
                      start_q=0,
                      test='adf',       # use adftest to find optimal
'd'
                      max_p=4,
                      max_q=4,          # maximum p and q
                      m=1,              # frequency of series
                      d=None,           # let model determine 'd'
                      seasonal=False,   # No Seasonality
                      start_P=0,
                      D=0,
                      trace=True,
                      error_action='ignore',
                      suppress_warnings=True,
                      stepwise=True)
model_predictions = model.predict(1498)
pred = pd.DataFrame(model_predictions,index=test_df.index)
pred
plt.figure(figsize=(8,5))
plt.plot(train_df,label="train")
plt.plot(test_df,label="test")
plt.plot(pred,label="pred")
```

Neural Prophet

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from neuralprophet import NeuralProphet
import copy
train_df = pd.read_csv('Data/rain_prediction/train.csv')
test_df = pd.read_csv('Data/rain_prediction/test.csv')
train_df.rename(columns = {'Date':'ds', 'Prec':'y'}, inplace = True)
test_df.rename(columns = {'Date':'ds', 'Prec':'y'}, inplace = True)
train_df.head()
train_df['ds'] = pd.DatetimeIndex(train_df['ds'])
train_df.info()test_df['ds'] = pd.DatetimeIndex(test_df['ds'])

test_df.info ()train_df = train_df.append(test_df)

model = NeuralProphet(
    yearly_seasonality=True,
    daily_seasonality=False,
```

```
)
df_train, df_val = model.split_df(train_df, freq='D', valid_p = 0.2)
metrics = model.fit(df_train, freq='D',
validation_df=df_val,epochs=200,learning_rate=0.001)
```