INTRODUCTION

1.1 Overview

Rainfall has been a major concern these days. Weather conditions have been changing for time being. Rainfall forecasting is important otherwise, it may lead to many disasters. Irregular heavy rainfall may lead to the destruction of crops, heavy floods that can cause harm to human life. It is important to exactly determine the rainfall for effective use of water resources, crop productivity, and pre-planning of water structures.

This comparative study is conducted concentrating on the following aspects: modeling inputs, Visualising the data, modeling methods, and pre-processing techniques. The results provide a comparison of various evaluation metrics of these machine learning techniques and their reliability to predict rainfall by analyzing the weather data.

We will be using classification algorithms such as Decision tree, Random Forest, KNN, and xgboost. We will train and test the data with these algorithms. From this the best model is selected and saved in pkl format. Once the model is saved, we integrate it with flask application and deploy the model in IBM.

Purpose:

The main purpose of this project is to weather its gone rain tomorrow or not (predicting rainfall)

LITERATURE SURVEY

2.1 Existing problem

The accurate and precise rainfall prediction is still lacking which could assist in diverse fields like agriculture, water reservation and flood prediction. The issue is to formulate the calculations for the rainfall prediction that would be based on the previous findings and similarities and will give the output predictions that are reliable and appropriate. The imprecise and inaccurate predictions are not only a waste of time but also the loss of resources and lead to inefficient management of crises like poor agriculture, poor water reserves and poor management of floods. Therefore, the need is not only to formulate the rainfall predicting system but also

a system that is more accurate and precise compared to the existing rainfall predictors.

2.2 Proposed solution

Spiral model of programming by creating a V0 and test it for feedback from the test sample. It will retrieve the possible alterations to create the next Version of the algorithm. The test will be by stimulating the neural network to retrieve results by archiving.

THEORITICAL ANALYSIS

To develop this project, we need to install the following software/packages:

Step 1: Anaconda Navigator

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning-related applications. It can be installed on Windows, Linux, and macOS. Conda is an open-source, cross-platform, package management system. Anaconda comes with great tools like Jupiter, Jupiter Notebook, Console, Spyder, Glue viz, Orange, RStudio, Visual Studio Code.

For this project, we will be using a Jupiter notebook and Spyder

Step 2: Python packages

To build Machine learning models you must require the following packages

Sklearn: Scikit-learn is a library in Python that provides many unsupervised and supervised learning algorithms.

NumPy: NumPy is a Python package that stands for 'Numerical Python. It is the core library for scientific computing, which contains a powerful n-dimensional array of object

Pandas: pandas are a fast, powerful, flexible, and easy-to-use open-source data analysis and manipulation tool, built on top of the Python programming language.

Matplotlib: It provides an object-oriented API for embedding plots into applications using general-purpose GUI (Graphical User Interface) toolkits

Step 3: Flask - Web framework used for building Web applications.

If you are using anaconda navigator, follow the below steps to download the required packages:

Open anaconda prompt as administrator

Type “pip install numpy ” and click enter

Type “pip install pandas” and click enter

Type “pip install scikit-learn” and click enter.

Type” pip install matplotlib” and click enter.

Type” pip install spicy" and click enter.

Type” pip install pickle-mixing” and click enter.

Type” pip install seaborn” and click enter.

Type “pip install Flask” and click enter.

Project Objectives

By the end of this project:

You will be able to understand the problem to classify if it is a regression or a classification kind of problem.

You will be able to know how to pre-process/clean the data using different data preprocessing techniques.

You will be able to analyze or get insights into data through visualization.

Applying different algorithms according to the dataset and based on visualization.

You will be able to know how to find the accuracy of the model.

You will be able to know how to build a web application using the Flask framework.

Project Flow

User interacts with the UI (User Interface) to enter the input values Entered input values are analyzed by the model which is integrated

Once the model analyses the input the prediction is showcased on the UI

To accomplish this, we must complete all the activities and tasks listed below

Data Collection.

Collect the dataset or create the dataset

Data Pre-processing.

Import the Libraries.

Importing the dataset.

Checking for Null Values.

Data Visualization.

Taking care of Missing Data.

Feature Scaling.

Splitting Data into Train and Test.

Model Building

Import the model building Libraries

Initializing the model

Training and testing the model

Evaluation of Model

Save the Model

Application Building

Create an HTML file

Build a Python Code

Data Collection

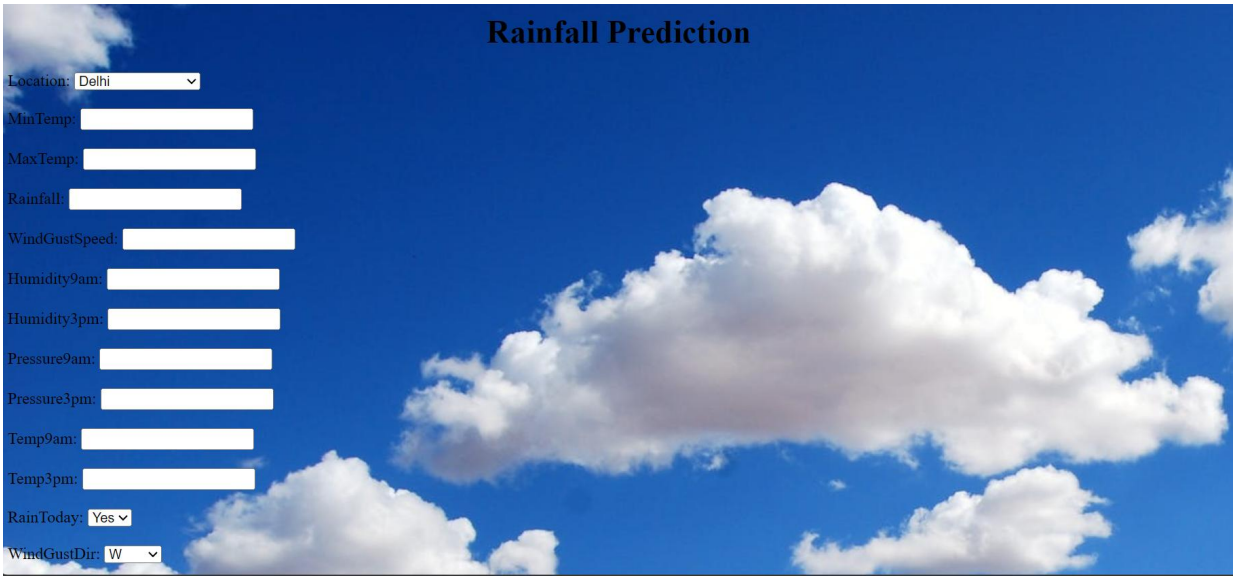
ML (Machine Learning) depends heavily on data, without data, it is impossible for an “AI” to learn. It is the most crucial aspect that makes algorithm training possible. In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions.

Data Pre-Processing

Data Pre-processing includes the following main tasks

- Import the Libraries.
- Importing the dataset.
- Checking for Null Values.
- Data Visualization.
- Feature Scaling.
- Splitting Data into Train and Test.

RESULT


Rainfall Prediction
Location: Delhi
MinTemp:
MaxTemp:
Rainfall:
WindGustSpeed:
Humidity9am:
Humidity3pm:
Pressure9am:
Pressure3pm:
Temp9am:
Temp3pm:
RainToday: Yes
WindGustDir: W


No chances of rain today

ADVANTAGES & DISADVANTAGES

Rainfall forecasting is important because heavy and irregular rainfall can have many impacts like destruction of crops and farms, damage of property so a better forecasting model is essential for an early warning that can minimize risks to life and property and managing the agricultural farms in better way.

APPLICATIONS

Precipitation is useful information for assessing vital water resources, agriculture, ecosystems, and hydrology. Data-driven model predictions using deep learning algorithms are promising for these purposes.

The ANN is a form of machine learning technique that has been widely used in rainfall prediction given its ability to identify complex non-linear relationships between input and output variables without the need to understand the nature of the physical processes.

CONCLUSION

Rainfall Prediction is the application area of data science and machine learning to predict the state of the atmosphere. It is important to predict the rainfall intensity for effective use of water resources and crop production to reduce mortality due to flood and any disease caused by rain. This paper analyzed various machine learning algorithms for rainfall prediction. Three machine learning algorithms such as MLR, FR, and XGBoost were presented and tested using the data collected from the meteorological station at Bahir Dar City, Ethiopia.

The relevant environmental features for rainfall prediction were selected using the Pearson correlation coefficient. The selected features were used as the input variables for the machine learning model used in this paper. A comparison of results among the three algorithms (MLR, RF, and XGBoost) was made and the results showed that the XGBoost was a better-suited machine learning algorithm for daily rainfall amount prediction using selected environmental features. The accuracy of the rainfall amount prediction may increase if the sensor data is incorporated for the study. But the sensor data was not considered in this study.

DISCUSSION

The environmental features used in this study taken from the meteorological station collected by measuring devices are analyzed their relevance on the impact of rainfall and selected the relevant features based on experiment result of Pearson correlation values as shown in Table for the daily rainfall prediction.