REPORT

Rainfall Prediction USING MACHINE LEARNING

1. INTRODUCTION

1.1 OVERVIEW:

Rainfall prediction remains a serious concern and has attracted the attentionof governments, industries, risk management entities, as well as the scientificcommunity. Rainfall is a climatic factor that affects many human activities likeagricultural production, construction, power generation, forestry and tourism,among others . To this extent, rainfall prediction is essential since this vari-able is the one with the highest correlation with adverse natural events suchas landslides, flooding, mass movements and avalanches. These incidents haveaffected society for years . Therefore, having an appropriate approach for rain-fall prediction makes it possible to take preventive and mitigation measures forthese natural phenomena .To solve this uncertainty, we used various machine learning techniques andmodels to make accurate and timely predictions. These paper aims to provide endto end machine learning life cycle right from Data preprocessing to implementingmodels to evaluating them. Data Preprocessing steps include imputing missingvalues, feature transformation, encoding categorical features, feature scaling andfeature selection. We implemented models such as Logistic Regression, DecisionTree, K Nearest Neighbour, Rule-based and Ensembles. For evaluation purpose,used Accuracy, Precision, Recall, F-Score and Area Under Curve as evaluation metrics.

1.2. PURPOSE:

Rainfall prediction is a critical task because many people rely on it, particularly in the agricultural sector. Rainfall forecasting is

difficult due to the ever-changing nature of weather conditions. We propose a Long Short-Term Memory (LSTM)-based prediction model capable of forecasting daily rainfall.

2. LITERATURE SURVEY

2.1. EXISTING PROBLEM:

Rainfall forecasting is very important because heavy and irregular rainfall can have many impacts like the destruction of crops and farms,damage of property so a better forecasting model is required for an early warning that can reduce the risks to life and property and also helps to manage the agricultural farms in a better way. Heavy rainfall is a cause for natural disasters like floods and

drought that square measure encountered by individuals across the world each year. Many models are developed to evaluate the rainfall and for predicting the likeliness of rain.

These models are based on both supervised and unsupervised machine learning algorithms. Taking into consideration of overall rainfall will not help us to know if it rains in specific conditions. Accuracy is the major concern in machine learning.

We are going to understand the data and then train the model accordingly to predict whether if it rains under given conditions or not.

2.1.1PROJECT FLOW:

1. User interacts with the UI to enter the input.
2. Entered input is analysed by the model which is integrated.
3. Once model analyses the input the prediction is showcased on the UI.

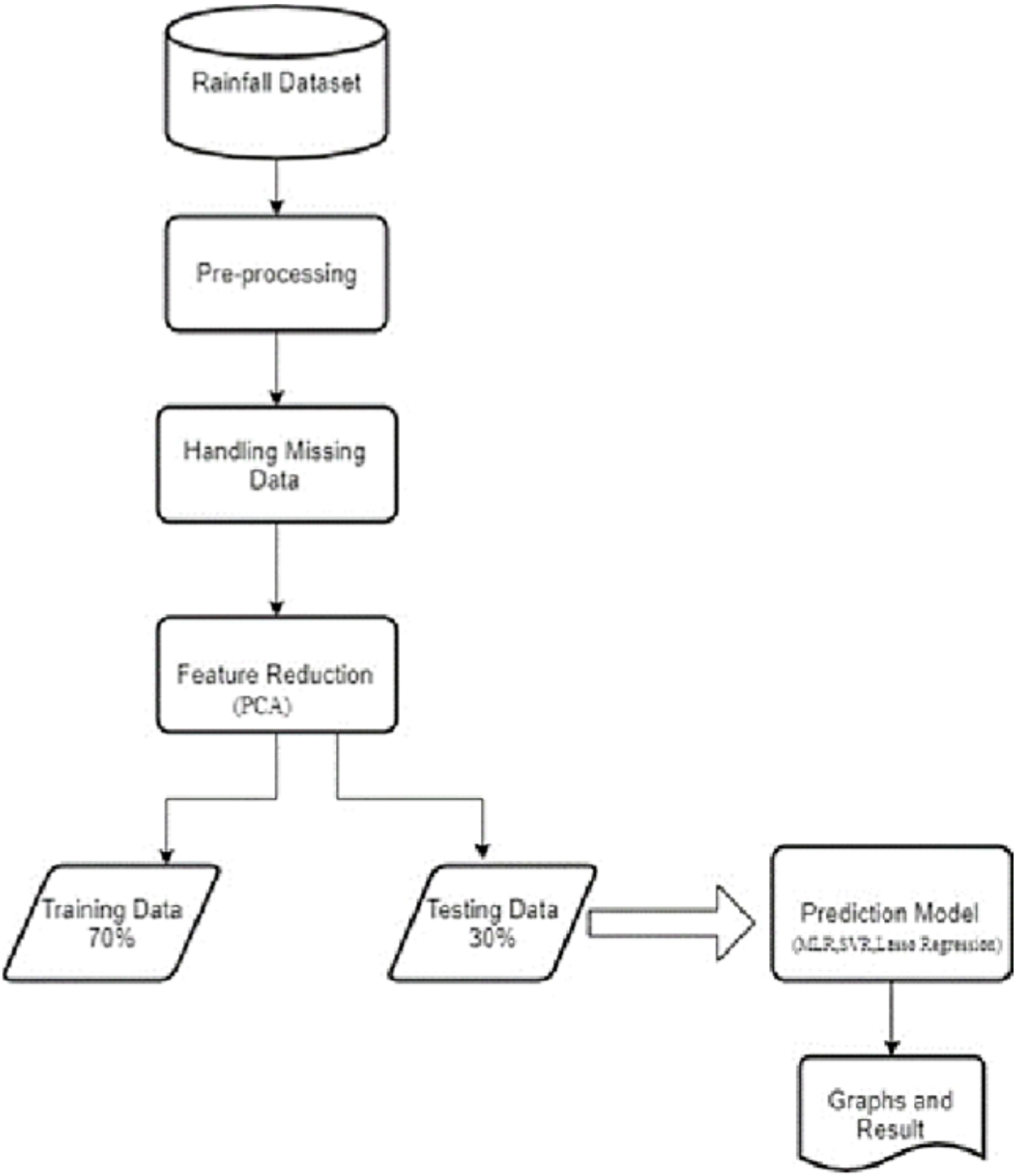
2.2 PROPOSED SOLUTION:

Building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scripting.

The implementation of the project is divided into seven sections. In the first section, we are going to import the required libraries and then study them.Next, we are going to prepare the dataset with required attributes, then transformations on data are performed,and then data analysis can be made usingcorrelation, followed by splitting of a dataset into train and test sets, finally,model training is done to know the best model that fit(s) our data for predicting rainfall.

3. THEORITICAL ANALYSIS:

3.1. BLOCK DIAGRAM:



3.2. HARDWARE AND SOFTWARE REQUIREMENTS:

Jupiter Notebook

Pandas

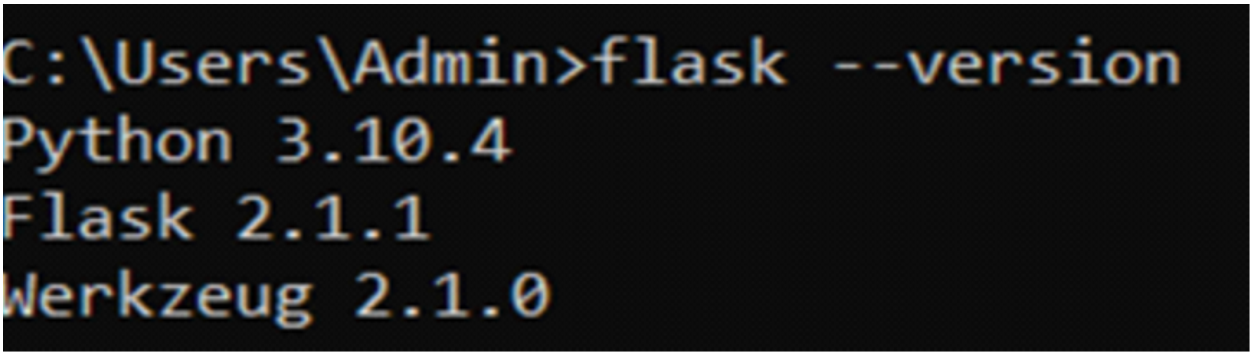
Numpy

Matplotlib

Linear regression

Logistic Regression

Flask Software



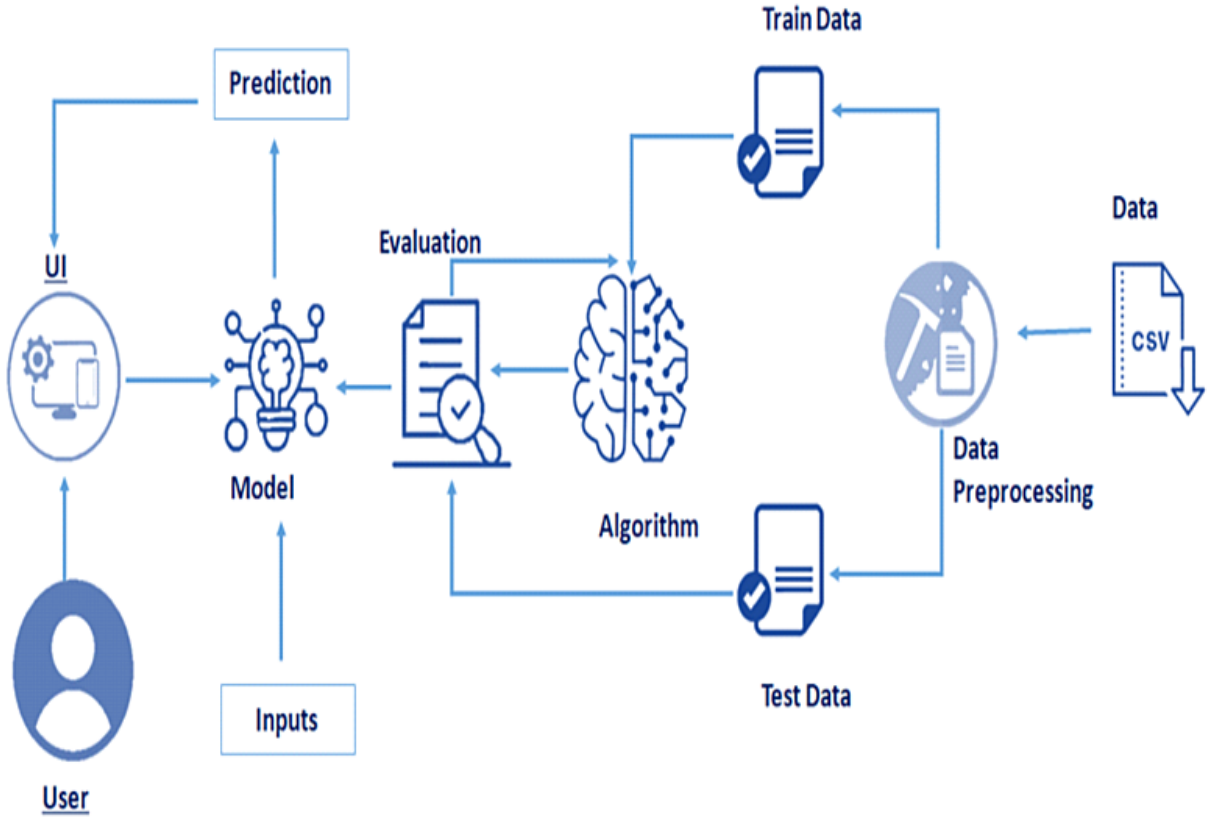
4. <u>EXPERIMENTAL INVESTIGATIONS:</u>

In this paper, the data set under consideration contains daily weather observations from numerous weather stations. The target variable is Rain-Tomorrow which means: Did it rain the next day? Yes or No. The data set consists of 23 features and 142k instances. Below are the fea-tures.

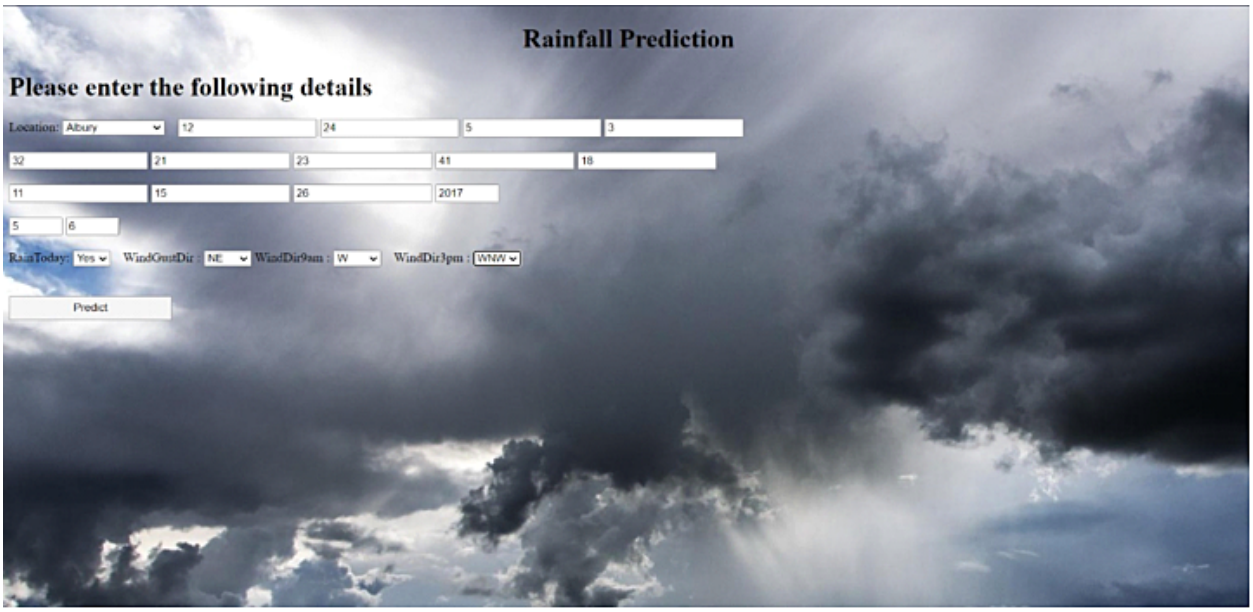
Table 1. Data set Description

Feature	Description
Date	The date of observation
Location	The common name of the location of the weather station
MinTemp	The minimum temperature in degrees celsius
MaxTemp	The maximum temperature in degrees celsius
Rainfall	The amount of rainfall recorded for the day in mm
Evaporation	The so-called Class A pan evaporation (mm) in the 24 hours to 9am
Sunshine	The number of hours of bright sunshine in the day.
WindGustDir	The direction of the strongest wind gust in the 24 hours to midnight
WindGustSpeed	The speed (km/h) of the strongest wind gust in the 24 hours to midnight
WindDir9am	Direction of the wind at 9am
WindDir3pm	Direction of the wind at 3pm
WindSpeed9am	Wind speed (km/hr) averaged over 10 minutes prior to 9am
WindSpeed3pm	Wind speed (km/hr) averaged over 10 minutes prior to 3pm
Humidity9am	Humidity (percent) at 9am
Humidity3pm	Humidity (percent) at 3pm
Pressure9am	Atmospheric pressure (hpa) reduced to mean sea level at 9am
Pressure3pm	Atmospheric pressure (hpa) reduced to mean sea level at 3pm
Cloud9am	Fraction of sky obscured by cloud at 9am.
Cloud3pm	Fraction of sky obscured by cloud at 3pm.
Temp9am	Temperature (degrees C) at 9am
Temp3pm	Temperature (degrees C) at 3pm
RainToday	1 if precipitation exceeds 1mm, otherwise 0
RISK_MM	The amount of next day rain in mm.
RainTomorrow	The target variable. Did it rain tomorrow?

5. FLOW CHART:


Train Data
Prediction
Data
UI
Evaluation
CSV
Model
Algorithm
Data
Preprocessing
Test Data
Inputs
User

6. RESULT:



7. (I)<u>ADVANTAGES OF Rainfall Prediction:</u>

Rainfall forecasting is very 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 also managing the agricultural farms in better way.

(II) DISADVANTAGES OF GDP PREDICTION:

Problems concern availability, timeliness, and quality of observational data; time constraints on forecast preparation; the nature and reliability of communication systems available for forecast dissemination; and the makeup and requirements of the user community.

8. APPLICATIONS OF Rainfall Prediction:

Rainfall prediction is one of the challenging and uncertain tasks which has a significant impact on human society. Timely and accurate predictions can help to proactively reduce human and financial loss. This study presents a set of experiments which involve the use of prevalent machine learning techniques to build models to predict whether it is going to rain tomorrow or not based on weather data for that particular day in major cities.

9. CONCLUSION:

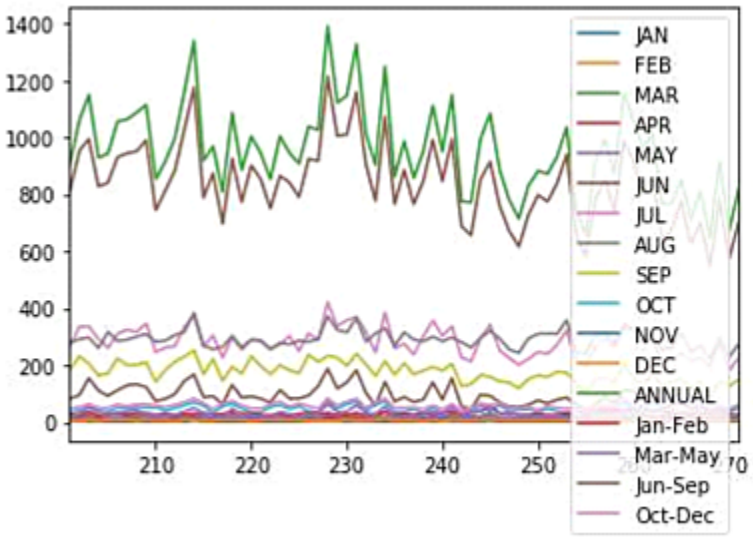
1. Thus the Rainfall is predicted using the given dataset.

10. FUTURE SCOPE:

e explored and applied several preprocessing steps and learnedthere impact on the overall performance of our classifiers. We also carried acomparative study of all the classifiers with different input data and observedhow the input data can affect the model predictions.We can conclude that weather is uncertain

and there is no suchcorrelation among rainfall and the respective region and time. We figured certainpatterns and relationships among data which helped in determining importantfeatures. Refer to the appendix section.As we have a huge amount of data, we

can apply Deep Learning models suchas Multilayer Perceptron, Convolutional Neural Network, and others. It wouldbe great to perform a comparative study between the Machine learning classifiersand Deep learning models.



11. BIBILOGRAPHY:

References:-

https://www.analyticsvidhya.com/blog/2021/06/pr

edictive-modelling-rain-prediction-in-australia-with-python/

https://thecleverprogrammer.com/2020/09/11/rainfall-prediction-with-machine-learning/

https://www.geeksforgeeks.org/ml-rainfall-prediction-using-linear-regression/