International Journal of Humanities Social Science and Management (IJHSSM)
Volume 3, Issue 2, Mar.-Apr. 2023, pp: 1109-1115 www.ijhssm.org

A Study on Weather Forecasting For Effective Business Using Machine Learning

MR. V. S. ARUN[1], MS. S. R. RAMYA[2]
STUDENT[1], ASSISTANT PROFESSOR[2]
DEPARTMENT OF MANAGEMENT STUDIES,
KARPAGAM COLLEGE OF ENGINEERING, COIMBATORE – 641 032

Date of Submission: 09-04-2023 Date of Acceptance: 23-04-2023

ABSTRACT

The activities of many primary sectors depend on the weather for production, e.g., farming. The climate is changing at a drastic rate nowadays, which makes the old weather prediction methods less effective and more hectic. To overcome these difficulties, the improved and reliable weather prediction methods are required. These predictions affect a nation's economy and the lives of people. To develop a weather forecasting software system that main motivation of this work. The data analytics and machine learning algorithms, such as random forest regression, are used to predict weather conditions. Random forest algorithm is a supervised machine learning algorithm that is used widely in classification and regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the random forest algorithm is that it can handle the data set containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. In this paper, a low-cost and portable solution for weather prediction is built. The temperature is predicted by random forest regressor and the climate is classified by decision tree classifier.

Keywords: Weather forecasting, random forest algorithm, machine learning, weather prediction, exploratory data analysis, applications of machine learning.

I. INTRODUCTION

This research presents the rain prediction with the use of real-time data of temperature, humidity, and pressure using various sensors. Machine learning provides a capability to the systems to learn and improve from experience without being programmed by a user. With the introduction of machine learning concept, there is a great ease for data analysis and prediction. Machine learning does not require to understand the physical processes that control the atmosphere but uses the past data to predict future data. Therefore, this process may be used as a weather forecasting method. The random forest regression, a machine learning algorithm, uses an ensemble learning method in which two or more machine learning models are combined to form a single machine learning model.

1.1 Weather Forecasting Process

A weather forecast is made up of three steps: observation and analysis, extrapolation to determine the state of the atmosphere in the future, and estimation of specific variables.

II. INDUSTRY PROFILE - ANALYTICS INDUSTRY

Business analytics is a powerful tool in today's marketplace that can be used to make decisions and craft business strategies. Across industries, organizations generate vast amounts of data which, in turn, has heightened the need for professionals who are data literate and know how to interpret and analyse that information. Analytics is the practice of using data to help companies make more informed decisions. Analysts provide support in a wide range of areas, from product development to accounting.

Typical duties of analysts entail: Budgeting and forecasting, strategic planning, monitoring business efforts, analysing variance, gathering data for problem-solving, reporting data to stakeholders, creating business solutions and defining business requirements.

III. COMPANY PROFILE - NoviTech R&D Pvt Ltd

Incepted in 2017, NoviTech R&D Pvt Ltd is a Hardware/Software Development, Technical

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 3, Issue 2, Mar.-Apr. 2023, pp: 1109-1115 www.ijhssm.org

training, and Information technology services to the Companies & Educational Institutions. They are focused, Product/Project approach is derived from the cumulative experience of their team members. As a result of their combined practical business experience with strong technology delivery capabilities, they can comprehend the client's business requirements, propose IT and creative strategies in line with the vision of the business and execute the business plan by building or implementing high quality scalable technical solutions using latest cutting-edge technologies.

IV. RESEARCH PROBLEM

The activities of many primary sectors depend on the weather for production, for example, farming. The climate is changing at a drastic rate nowadays, which makes the old weather prediction methods less effective and more hectic. To overcome these difficulties, the improved and reliable weather prediction methods are required. These predictions affect a nation's economy and the lives of people. To develop a weather forecasting software system is the main motivation of this research. The data analytics and machine learning algorithms, such as random forest regression, are used to predict weather conditions. In this project, a low-cost and portable solution for weather prediction has been implemented which helps in carrying out a competent business.

V. OBJECTIVES OF THE RESEARCH

5.1 PRIMARY OBJECTIVE

The main objective of this work is to develop a low-cost, reliable, and efficient weather forecasting application using the machine learning concept in Python for effective business

5.2 SECONDARY OBJECTIVES

- ✓ To overcome difficulties in the activities of many primary sectors that depend on the weather for production, e.g., farming
- ✓ To improve nation's economy and the lives of people by giving weather predictions
- ✓ To develop a weather forecasting software system that is the main motivation of this study
- ✓ To understand the factors that contribute to climate change
- ✓ To improve decision making process that aids to build an effective business

VI. REVIEW OF LITERATURE

Talha Siddique et al. (2022) [15]A study was conducted on "A survey of uncertainty quantification in machine learning for space weather prediction". With the availability of data and computational technologies in the modern world, machine learning (ML) has emerged as a preferred methodology for data analysis and prediction. While ML holds great promise, the results from such models are not fully unreliable due to the challenges introduced by uncertainty. However, if the uncertainty in the data and the model parameters are not considered, such optimal solutions have a high risk of failure in actual world deployment.

Nagadevi et al. (2020) [10]A study was conducted on "Weather Forecasting using Machine Learning Algorithm". The weather is disordered in nature and hence has always been difficult for meteorologists in forecasting the weather accurately. Weather is measured from days to months where it includes the following components such as: Precipitation, Temperature, Sunshine, Wind, Lightning, Direction of Wind, Humidity, Cloud Cover, Speed of Wind, Snow, Pressure etc. This project aims to obtain the highest accuracy to predict weather using four different algorithms – K-Nearest Neighbour, Decision Tree, Random Forest Algorithm, Linear Regression.

Sue Ellen Haupt et al. (2018) [5]A study was conducted on "Machine learning for applied weather prediction". The National Centre for Atmospheric Research (NCAR) has a long history of applying machine learning to weather forecasting challenges. The Dynamic Integrated foreCasting (DICast®) System was one of the first automated weather forecasting engines. It is now in use in quite a few companies with many applications. Some applications being accomplished at NCAR that include DICast and other artificial intelligence technologies include renewable energy, surface transportation, and wildland fire forecasting.

VII. RESEARCH METHODOLOGY

The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables, as in the case of regression, and categorical variables, as in the case

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 3, Issue 2, Mar.-Apr. 2023, pp: 1109-1115 www.ijhssm.org

of classification. It performs better for classification and regression tasks.

7.1 Existing System

In previous, some authors proposed an apparatus to conceive an acknowledgment of continuous day to day weather information by k-closest neighbour and little measure of observed information like target year. And some of them proposed with the machine learning algorithm which is Support Vector Regression.

7.1.1 Disadvantages of Existing System

Sensitive to the scale of the data and irrelevant features. The major drawback of KNN is becoming significantly slows as the size of that data in use grows. SVM does not execute very well when the data set has more sound i.e., target classes are overlapping

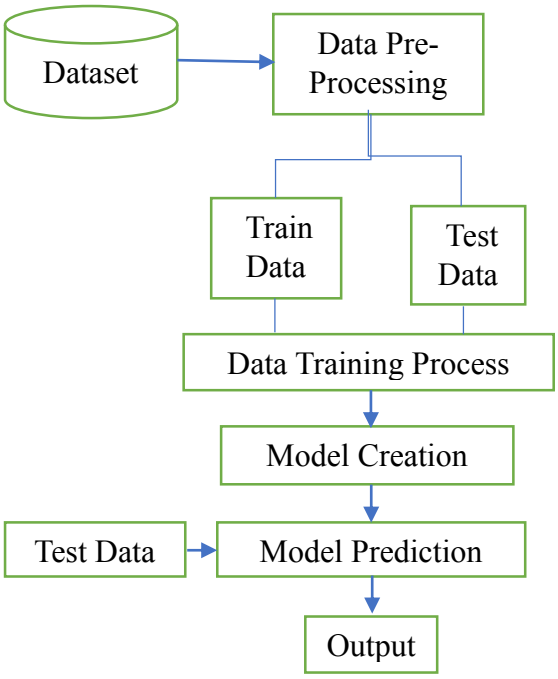

Fig. 1. Module Block Diagram

7.2.2 Module Description

i. Data Collection - Secondary Data: The data set used for this study is secondary data available at Kaggle. It is Delhi's weather data for the past 20 years which contains the details of temperature, humidity, pressure, and rain for each hour of a day.

ii. Data Pre-Processing: Pre-processing refers to the transformations applied to our data before providing the data to the algorithm. Data Pre-

7.2 Proposed System

In this project, the proposed system is based on Random Forest Regressor and Exploratory Data Analysis. We are using python tkinter to predict the climate based on temperature. In the exploratory data analysis part, we are visualizing the charts like plot chart, pie chart, distplot chart for heat, distplot chart for wet, distplot chart for rain and finally we are predicting the climate.

7.2.1 Advantages of Proposed System

Random Forest performs well comparing to other machine learning algorithms. It provides the highest accuracy and reduces overfitting. EDA helps to improve understanding of variables by extracting averages, mean, minimum, and maximum values, etc

processing technique is used to convert the raw data into an understandable data set. In other words, whenever the information is gathered from various sources it is collected in raw format that is not possible for the analysis.

iii. Training data and Test data: For choosing a model we split our dataset into train and test. Here data are split into 3:1 ratio that means training data having 70 percent and testing data having 30 percent. In this split process preforming based on train_test_split model. After splitting we get x_train, x_test and y_train, y_test

iv. Model Creation: Contextualise machine learning in your organisation. Explore the data and choose the type of algorithm. Prepare and clean the dataset. Split the prepared dataset and perform cross validation. Perform machine learning optimisation. Deploy the model.

v. Model Prediction: Predictive modelling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data. It works by analysing current and historical data and projecting what it learns on a model generated to forecast likely outcomes.

VIII. DATA ANALYSIS AND INTERPRETATION

8.1 Weather Prediction Process Using Random Forest Algorithm

Once we enter the weather forecasting data in the data fields, the weather prediction process will be automated and the result of temperature will be displayed on the screen when we click the "Temperature checker" button as shown in fig. 2. After checking the temperature, by clicking on the "climate classifier" button we need to enter the

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 3, Issue 2, Mar.-Apr. 2023, pp: 1109-1115 www.ijhssm.org

displayed temperature value in the "Temperature" text box displayed in climate classification window and then click "Climate checker" button as shown in fig. 3. Now a pop-up window named "Climate" will appear which shows the information of the climate (Sunny, Rainy or Cold) as shown in fig. 4.

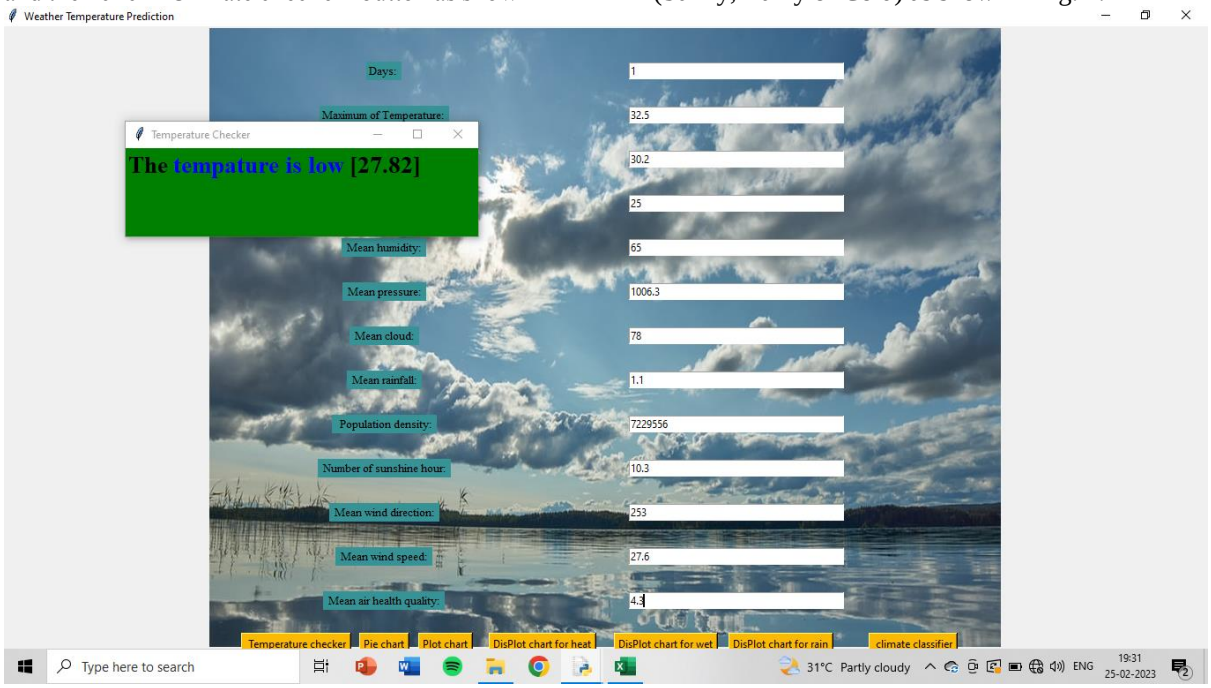

Fig. 2. Temperature Checker

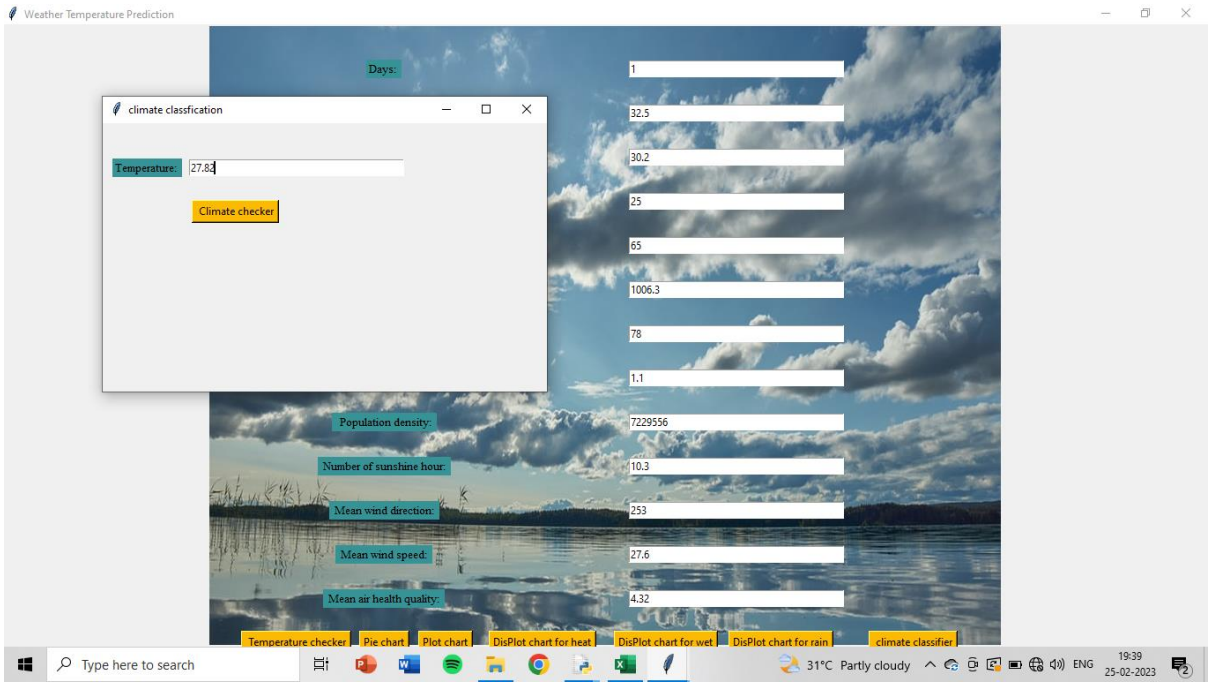

Fig. 3. Climate Checker / Classifier

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 3, Issue 2, Mar.-Apr. 2023, pp: 1109-1115 www.ijhssm.org

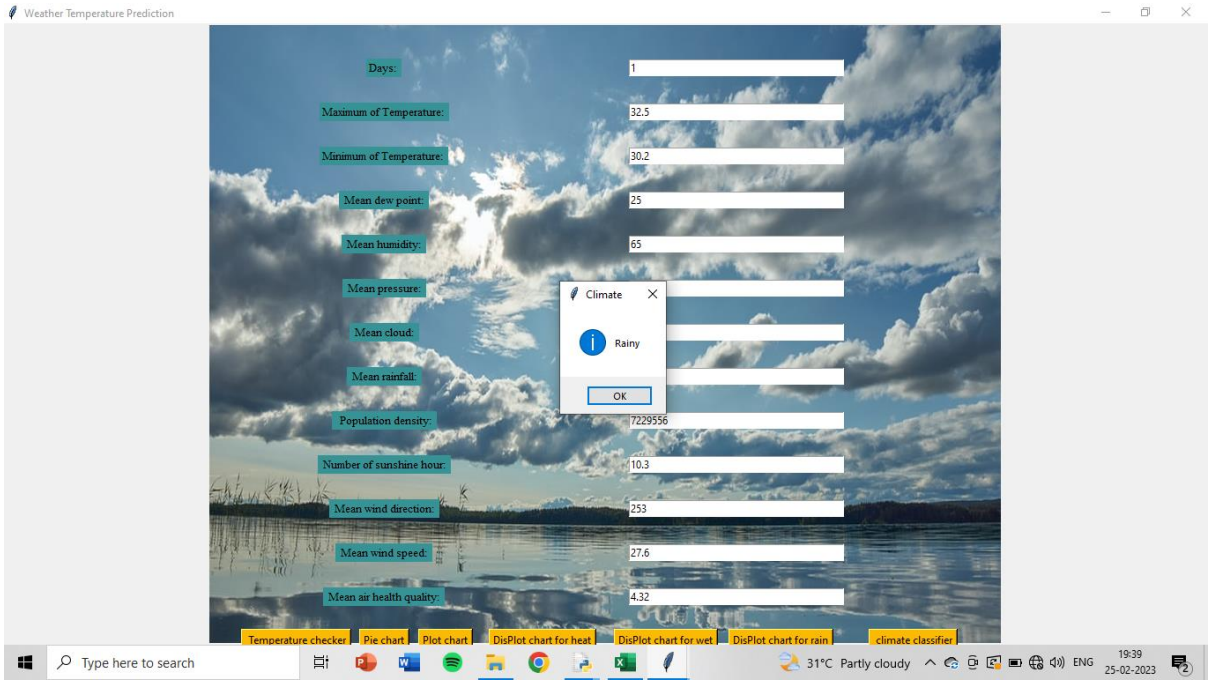

Fig. 4. Climate Result Pop-up Window

Interpretation

For day 1, we got the temperature as 27.82 °C which could give a "rainy" climate. Similarly, we can predict the climate for any day with the given weather parameters. R squared for train data is 0.987. R squared for test data is 0.906. Mean Absolute Error is 0.288. Mean Squared Error is 0.155. Root Mean Squared Error is 0.394. R^2 score tells us how well our model is fitted to the data by comparing it to the average line of the dependent variable. If the score is closer to 1, then it indicates that our model performs well versus if the score is farther from 1, then it indicates that our model does not perform so well. We achieved accuracy scores of approximately 99% and 91% for train data and test data respectively.

8.2 Scatter Plots for the Weather Dataset – Correlation Analysis

(i) Wind Speed vs Mean Temperature

Null Hypothesis (H_0): There is no association between wind speed and mean temperature

Alternative Hypothesis (H_1): There is an association between wind speed and mean temperature

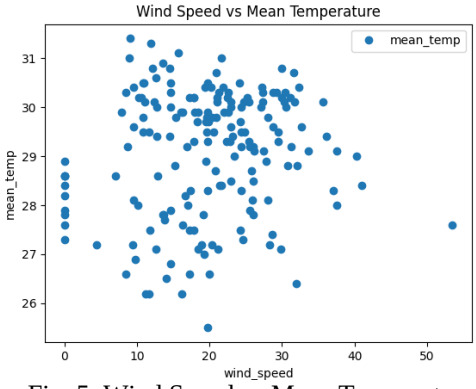

Fig. 5. Wind Speed vs Mean Temperature

Interpretation

The scatter plot is spread out and so there is no association between wind speed and mean temperature. From the scatter plot shown in fig. 5, we can say that there is no correlation between wind speed and mean temperature.

(ii) Bright Sunshine vs Mean Temperature

Null Hypothesis (H_0): There is no association between bright sunshine and mean temperature

Alternative Hypothesis (H_1): There is an association between bright sunshine and mean temperature

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 3, Issue 2, Mar.-Apr. 2023, pp: 1109-1115 www.ijhssm.org

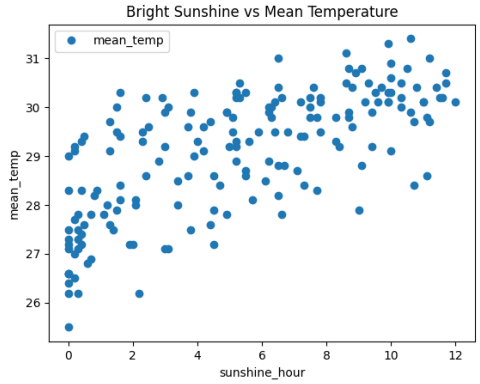

Fig. 6. Bright Sunshine vs Mean Temperature

Interpretation

The scatter plot is moderately spreading out in a linear upward direction and there is an association between bright sunshine and mean temperature. From the scatter plot shown in fig. 6, we can say that there is a moderate positive linear correlation between bright sunshine and mean temperature.

(iii) Mean Cloud vs Mean Temperature

Null Hypothesis (H_0): There is no association between mean cloud and mean temperature

Alternative Hypothesis (H_1): There is an association between mean cloud and mean temperature

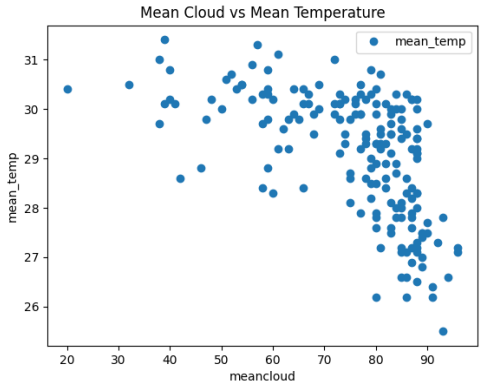

Fig. 7. Mean Cloud vs Mean Temperature

Interpretation

The scatter plot is strongly spread out in a U-shaped downward direction and so there is an association between mean cloud and mean temperature. From the scatter plot shown in fig. 7, we can say that there is a strong negative non-linear correlation between mean cloud and mean temperature.

IX. FINDINGS

From the secondary data, we found that the rain percentage is 21%, 47% and 32% of more, less and no rain respectively, the wet percentage is 56% and no wet percentage is 44%, and the heat percentage is 68% and no heat percentage is 32%. There is no correlation between wind direction and mean temperature. There is no correlation between population and mean temperature. There is a moderate negative non-linear correlation between rainfall and mean temperature. There is a strong negative linear correlation between mean humidity and mean temperature. There is no correlation between mean dew point and mean temperature. There is a low positive linear correlation between pressure and mean temperature. There is low positive linear correlation between air health quality and mean temperature.

X. SUGGESTIONS

Weather forecasting application using the machine learning concept can be built utilizing low-cost Internet of Things (IoT) devices, such as temperature and humidity sensors, in collecting weather dataset. The dataset can include more weather parameters such as flooding parameters like tide prediction, water level, wave height etc, that can increase the accuracy of the weather forecasting process. In addition to weather forecasting, the system could be built with disaster forecasting and management. Automatic temperature detection system could be installed that uses a GPS tracker for finding the location in which the temperature and climate could be analysed.

XI. CONCLUSION

The random forest classification provides deep insights into the data dependencies. The main result expected in this work is about whether rain will happen or not on a particular day according to the sensor data. In this work, out of total 7339 data sets, the data sets are divided into 75% as the training set and 25% as the test set, i.e., 5504 data for training and 1835 for testing. The splitting between the training set and test set is random. The size of the test set is 1835, out of which the number 1491 represents the count of correct predictions that rain will not happen and 122 is the number of correct predictions that rain will happen. Therefore, total 1613 correct predictions are made from 1835 test data.

REFERENCES

[1]. Navin Sharma, Pranshu Sharma, David Irwin & Prashant Shenoy. (2011). Predicting solar generation from weather forecasts using machine learning.

[2]. Afan Galih Salman, Bayu Kanigoro &Yaya Heryadi. (2015). Weather forecasting using deep learning techniques.
[3]. Mark Holmstrom, Dylan Liu & Christopher Vo. (2016). Machine Learning Applied to Weather Forecasting.
[4]. Sebastian Scher & Gabriele Messori. (2018). Predicting weather forecast uncertainty with machine learning.
[5]. Sue Ellen Haupt, Jim Cowie, Seth Linden, Tyler McCandless, Branko Kosovic & Stefano Alessandrini. (2018). Machine learning for applied weather prediction.
[6]. Risul Islam Rasel, Nasrin Sultana & Phayung Meesad. (2018). An Application of Data Mining and Machine Learning for Weather Forecasting.
[7]. Siddharth Singh, Mayank Kaushik, Ambuj Gupta & Anil Kumar Malviya. (2019). Weather Forecasting Using Machine Learning Techniques.
[8]. Bin Wang, Jie Lu, Zheng Yan, Huaishao Luo, Tianrui Li, Yu Zheng & Guangquan Zhang. (2019). Deep Uncertainty Quantification: A Machine Learning Approach for Weather Forecasting.
[9]. Nalluri Sravani, Ramasubbareddy Somula & Kannayaram. (2019). Weather Prediction Using Clustering Strategies in Machine Learning.
[10]. Nagadevi S., Ramesh V., & James H. (2020). Weather forecasting using machine learning algorithm.
[11]. A H M Jakaria, Md Mosharaf Hossain &Mohammad Ashiqur Rahman. (2020). Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee.
[12]. Xiaoli Ren, Xiaoyong Li, Kaijun Ren, Junqiang Song, Zichen Xu, Kefeng Deng & Xiang Wang. (2021). Deep Learning-Based Weather Prediction: A Survey.
[13]. Mihir Bhawsar, Vandan Tewari & Preeti Khare. (2021). A Survey of Weather Forecasting based on Machine Learning and Deep Learning Techniques.
[14]. Bogdan Bochenek & Zbigniew Ustrnul. (2022). Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives.
[15]. Talha Siddique, Md Shaad Mahmud, Amy M. Keesee, Chigomezyo M. Ngwira & Hyunju Connor. (2022). A Survey of Uncertainty Quantification in Machine Learning for Space Weather Prediction.