Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Real-Time Rainfall Prediction in Kathmandu, Kapan Area using Sensor Data with Machine Learning and Linear Regression

Biplov Paneru[1], Bishwash Paneru[2]

[1]Department of Electronics & Communication Engineering, Pokhara University, Kathmandu, Nepal

[2]Department of Applied Science & Engineering, Tribhuvan University Institute of Engineering, Lalitpur, Nepal

E-mail: [1]biplovp019402@nec.edu.np, [2]rampaneru420@gmail.com

Orcid: 0000-0003-2003-0648[1]

Abstract

The rainfall in the Kapan area is one of the most challenging aspects due to flooding in many parts of the area. The rainfall in the area of Milanchowk, Kapan, Nepal is one of the most challenging aspects that all the researchers, engineers, and geologists are trying to figure out. The research work is simply forecasting the weather and predicting it using machine learning techniques. The data from nearly half a month is collected in real time using sensor data through raspberry pi computer, and then the data is cleaned and processed for linear regression model-based precipitation. The work is carried out with a case study of the area, with appropriate steps taken to monitor in real-time and process the data accordingly.

Keywords: Rainfall, Linear Regression, humidity, Raspberry pi, Mean absolute error

1. Introduction

Natural disasters are troublesome for the country facing an economic crisis. The adoption of technologies for the purpose of prevention is one of the significant measures to be adapted. The Kapan area, which has one of the most challenging threats due to flood-related issues, has caused a headache to all the people residing in the area nearby the river large pipes. A dependent variable and one or more independent variables are modeled using the supervised

DOI: https://doi.org/10.36548/jscp.2023.3.004
Received:20.08.2023, received in revised form: 19.09.2023, accepted: 01.10.2023, published: 12.10.2023

learning algorithm of linear regression. The amount of rainfall in this situation is the dependent variable, and the characteristics such as temperature, humidity, wind speed, etc. [6-10]. that are used to predict it are the independent variables. The regression-based modeling allows the weather data to be processed easily in real time with the help of sensors and a powerful microcontroller, the Raspberry Pi. This makes a lot of data to be easily stored in a CSV (Comma Separated Values) document and then utilize it through its systemThe research is carried out after a case study of the overall context and the objectives of the research work are:

i. Identification of rainfall-related problems and system installation feasibility with a case study
ii. Data collection based on an appropriate location for processing
iii. Applying a machine learning model for the proper analysis of weather parameters and identifying solutions
iv. Applying pre-caution measures with the help of collected sensors data

2. Literature Review

The regional meteorological office in Bahir Dar City, Ethiopia provided the dataset that was used to compare the performance of three machine learning algorithms (Multivariate Linear Regression, Random Forest, and Extreme Gradient Boost). The Mean Absolute Error and Root Mean Squared Error techniques were used to assess the performance of the machine learning model. Extreme Gradient Boosting outperform the other machine learning methods, according to the study's results [1]. This research presents a new supervised learning machine learning model for rainfall prediction. Rainfall [16-18] is a major issue since it has an impact on every aspect of the world that people rely on. Today, it is challenging to anticipate rainfall in a reliable and consistent manner. In this study, the support vector machine (SVM) classifier and logistic regression produce the best results and the most accurate rainfall forecast [2]. This study suggests a novel approach based on linear regression analysis for predicting monthly precipitation. To forecast when it will rain, use quantitative information about the state of the atmosphere. Some machine learning systems are capable of recognizing complex information. an input-to-output mapping produced from a small number of samples. It can be difficult to predict precipitation with complete certainty since the process of precipitation can be affected by how quickly the atmosphere can change [3]. In order to predict the future rainfall situation, it is required to consider the variation in previous conditions. The method put forth uses linear regressions using a range of variables, such as temperature, humidity, and wind. This prediction should prove to be much more accurate since the recommended model tends to estimate rainfall based on the historical data for a certain geographic area. When compared to more well-known techniques for predicting rainfall, the model's performance is more accurate [4]. A variety of datasets are used to construct the framework of the training phase. The feature selection algorithm employs the greedy forward selection technique. The proposed system makes use of fuzzy logic. The fuzzy logic algorithm is a practical tool for mining boolean association rules from common item-sets. Using MATLAB instructions, the rainfall occurrence dataset can be converted into a comprehensible form after being collected using global forecasts [5].

3. Case study

This makes a lot of data easy to store in a .CSV (Comma Separated Values) document and then utilize it through its system. The research is carried out after a case study of the overall context and the objectives of the research. The Kapan area is one of the most prone areas to flooding, and many residing people suffer from the flood every year. One of the biggest markets in the area is the victim of economic as well as physical effects of flooding. The study of the area is made by the authors, who then carry out an approximation for rainfall prediction[11] in a feasible area. The sensors and the hardware unit are implemented on the roof of one of the local residents' homes.

Figure 1. Flooded Area in Kapan

4. Proposed Work

The proposed work involves using a Raspberry Pi as a brain for system development, which involves sensors-based data capture, processing the data as a dataset, importing it, and finally applying a linear regression model.

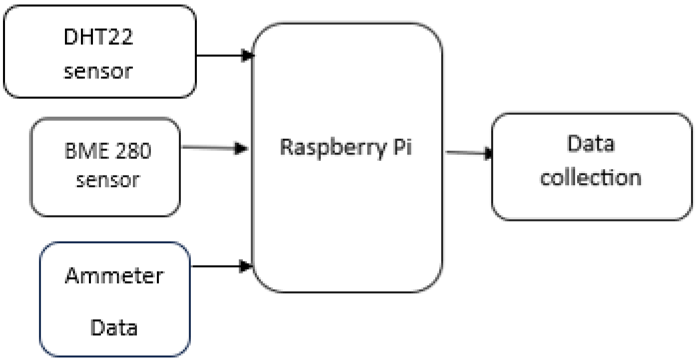

Figure 2. Block Diagram of Raspberry Pi System

4.1 Hardware Section

The Raspberry Pi is a powerful microcontroller as well as a single-board computer capable of handling large amounts of data, and its processing speed is also higher than that of ordinary microcontrollers. Sensors like DHT-22, pressure sensors, barometers, rain gauges, etc. have been utilized to extract data on the weather conditions. These data are very useful for the purpose of predicting rainfall and then forecasting weather too.

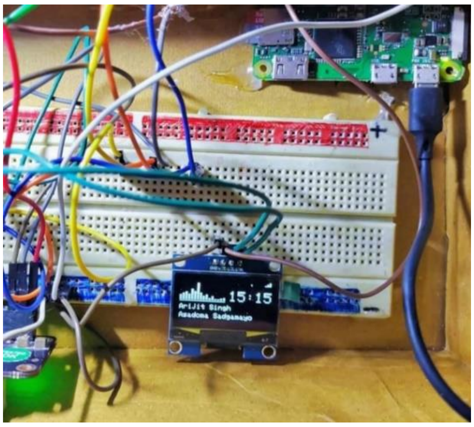

Figure 3. Hardware Setup of System

Temperature parameters, humidity and sea level pressures, wind levels, and visibility levels are the corresponding parameters included for the purpose of rain prediction. The sensors employed are tabulated in Table 1.

Table 1. Sensors and Functions

Sensor	**Function**
DHT22 sensor	Humidity prediction
BMP280 high precision sensor	Pressure, temperature, altitude
IR distance sensors	Light visibility detection
Digital anemometer	Wind speed & Direction

The data produced by sensors is viewed in an OLED display illustrated in figure.4. This is a preliminary process evaluation of the Raspberry Pi on an OLED display for tracking precise values whenever the researchers are free to track the information and assess research progress.

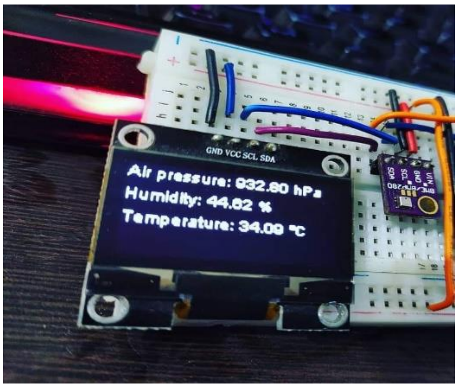

Figure 4. Sensor Data Monitoring on OLED

The OLED display is used to monitor the data for the BME 280 sensor and provides a range of values in real time. On the circuit board, additional sensors are similarly attached to give precise monitoring of environmental data.

Apart from these sensors, an OLED display has been employed in the system for the purpose of viewing the data. The data's collected are stored in the Raspberry Pi's storage in .CSV file format rapidly from sensors and then collected for future prediction and modeling.

i. Raspberry Pi as a Central Processing Unit

The Raspberry Pi serves as the central processing unit of the system, capable of handling large volumes of data and providing higher processing speeds compared to standard microcontrollers.

ii. Sensor Integration

Various sensors have been integrated into the system to gather weather-related data:

- DHT22 Sensor: This sensor is utilized for humidity prediction.
- BMP280 High Precision Sensor: It provides data on pressure, temperature, and altitude.
- IR Distance Sensors: These sensors are responsible for detecting light visibility levels.
- Digital Anemometer: It measures wind speed and direction.

4.1.1. Weather Parameters for Rain Prediction

The following parameters are considered for rain prediction:

i. Temperature

ii. Humidity

iii. Sea Level Pressure

iv. Wind Levels (speed and direction)

v. Visibility Levels

4.2 Hardware Setup

The hardware components, including Raspberry Pi and the various sensors, are arranged according to the specified configuration to facilitate data collection and processing.

i. OLED Display for Real-Time Monitoring

An OLED display unit has been incorporated into the system for the purpose of real-time data monitoring. This display provides a visual representation of the gathered data.

ii. Data Storage and Format:

The data acquired from the sensors are rapidly stored in the Raspberry Pi's storage in CSV (Comma Separated Values) file format. This format allows for efficient data organization and retrieval.

iii. Data Utilization for Prediction and Modeling:

The collected data is crucial for future prediction and modeling of weather conditions, particularly rainfall. It serves as the foundational dataset for developing predictive algorithms and forecasting models.

iv. Continuous Data Collection and Updating:

The system operates continuously, ensuring that data is collected and updated in real-time. This allows for accurate and up-to-date information for rain prediction.

4.2.1 Role of BMP280 Sensor

The BMP280 high precision sensor plays a key role in providing accurate data on pressure, temperature, and altitude. This information is essential for understanding atmospheric conditions, which contribute to rain prediction.

4.2.2 Role of IR Distance Sensors

The IR distance sensors are responsible for detecting light visibility levels. This data is a critical component in understanding overall weather conditions, as visibility is a crucial factor in assessing safety and environmental conditions.

4.2.3 Role of Digital Anemometer

The digital anemometer measures both wind speed and direction. This information is vital for comprehending wind patterns, which are influential in determining potential rain events.

Overall System Flexibility:

The Raspberry Pi-based system demonstrates flexibility in handling various sensors, making it adaptable to different weather monitoring scenarios beyond rainfall prediction.

4.3 ML Modelling & Data Cleaning

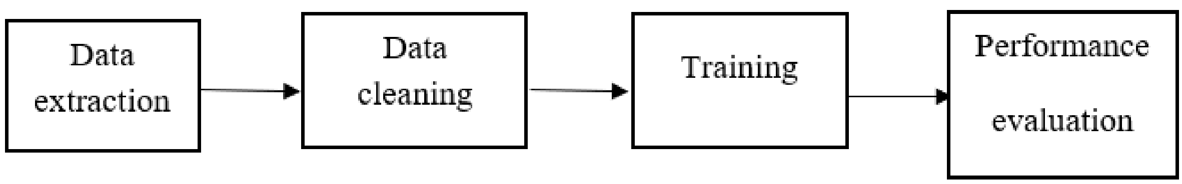

Figure 5. Linear Regression Procedure

4.4. Linear Regression

Linear regression [19] is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. It is a fundamental technique in statistics and machine learning for understanding and predicting relationships between variables.

Linear regression is widely used in various fields for tasks such as prediction, forecasting, and understanding the relationships between variables. It's important to note that linear regression assumes a linear relationship between the variables, so it may not be appropriate for all types of data. If the relationship is more complex, other regression techniques or machine learning models may be more suitable.

Initially the hypothesis function that maps the input variables to the output variables is built before performing the linear regression [17]. The hypothesis function in this instance is a linear equation of the following form:

$$Y = B0 + B1\ x\ B2\ x\ B3 ... + Bn\ x\ N$$ **........................Eqn 1.**

where b0, b1, b2,..., bn are the coefficients that are learned during training, x1, x2,..., xn are the input variables, and y is the expected amount of rainfall.

To train the model, the coefficient values that reduce the discrepancy between the training set's actual values and the predicted values are identified. This is accomplished by utilizing gradient descent or another optimization approach to reduce the mean squared error (MSE). The data is collected, preprocessed and trained. , Before defining a hypothesis function, train the model, and finally evaluate its performance.

Figure 6. Simple Linear Regression

4.5 Data Cleaning

Data cleaning is a crucial step in preparing data for any statistical or machine learning model, including linear regression. It involves identifying and correcting errors, inconsistencies, and outliers in the dataset to ensure that the model can learn from accurate and reliable information. Here are some important aspects of data cleaning in the context of linear regression:

1. Handling of Missing Values:
    - identifying and computing the missing values in the dataset.
    - deciding whether missing data handling strategy is preferable (such as removing rows or columns or imputation).
    - Imputation techniques include mean, median, and mode imputation in addition to more sophisticated techniques like regression imputation.
2. Outlier Detection and Treatment:
    - Outliers are located in the dataset. Outliers could have a major impact on the linear regression model.
    - selecting an outlier management method, which can involve data transformation or extreme data removal.
3. Dealing with Duplicates:
    - examining the dataset for duplicate records, then eliminating them.
4. Data Transformation:
    - Transforming variables if necessary. For example, the logarithms, square roots are taken, or other transformations are performed to make the data more linearly separable.
5. Categorical Variable Encoding:
    - The categorical variables in the dataset were encoded (e.g., using one-hot encoding) to make them compatible with linear regression.
6. Scaling and Standardization:

- The variables are inevitably scaled or normalized based on the regression algorithm. This guarantees that each variable makes an equal contribution to the model.

7. Checking for Multicollinearity:

   - Multicollinearity, which happens when independent variables have a strong correlation, is addressed by detection and execution. The interpretation of coefficients may be affected by this.

8. Feature Selection:

   - selecting the features that will be included in the model. The performance of the model can be increased by choosing pertinent features.

9. Data Exploration and Visualization:

   - visual analysis of the data to spot trends, patterns, and potential problems. Finding anomalies can be made easier using visualization.

10. Data Sanity Checks:

    - Verifying the data for logical inconsistencies. Ensure, for instance, that the ranges and connections of the variables are appropriate.

11. Validation and Cross-Validation:

    - putting a validation set aside to gauge the model's effectiveness. Depending on the situation, several methods are used, such as k-fold cross-validation, to evaluate the generalization and stability of the model.

12. Documentation:

    - Following that, the cleaning procedures were meticulously documented. This improved the dataset's repeatability and comprehension.

4.6 Modeling and Training

After being cleaned, the data may be fed into the linear regression model. A relationship between a dependent variable and numerous independent explanatory variables can be formed linearly using linear regression. This is accomplished by drawing a line that most accurately, or with the fewest errors, matches the scatter plot. By changing the independent values in the line equation, this provides value predictions, or how much the dataset will be trained using Scikit-Learn's linear regression model. Once the model has been trained, adding own dataset to the various columns, like temperature, dew point, pressure, and others, to forecast the weather using these characteristics.

The amount of precipitation for the input is [[1.37729345]] inches the graph showing the trend in precipitation: Kapan Area Weather Precipitation versus Selected Attributes Graph:

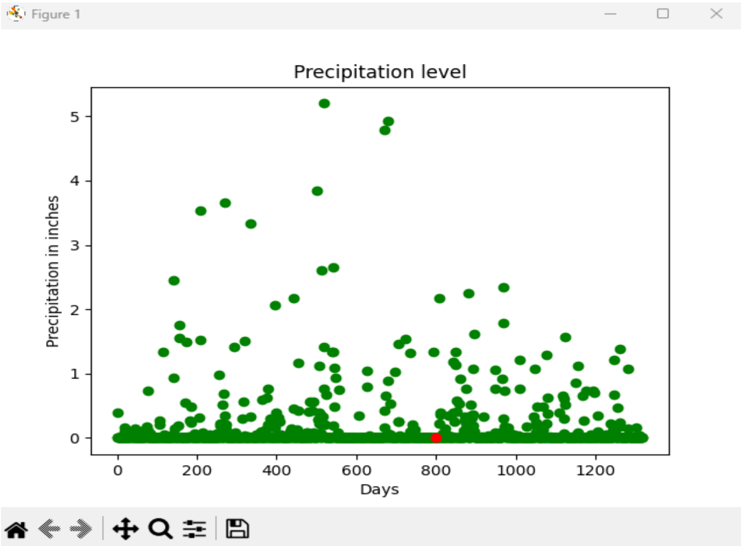

Figure 7. Precipitation Level

This graph visualizes the relationship between the independent variables (e.g., temperature, dew point, pressure) for a day (in red) and the dependent variable (precipitation). It helps in understanding how changes in these variables affect precipitation.

Overall, this process allows the authors to use historical data to build a model that can make predictions about future weather conditions based on specific characteristics (temperature, dew point, pressure, etc.). The specific prediction of 1.38 inches of precipitation suggests that, given the input data, this is the anticipated amount of precipitation.

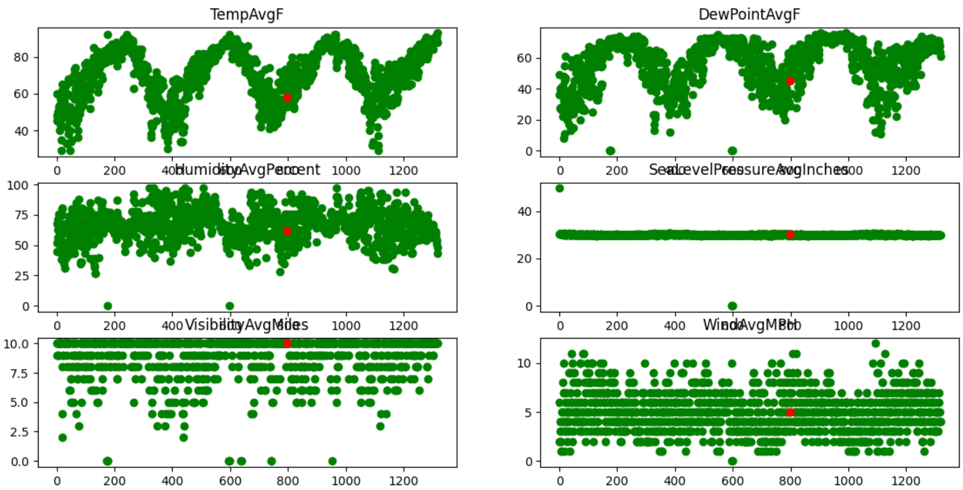

Figure 8. Graphs for Obtained Data

In this context from fig.2., a coefficient of approximately 1.377 suggests that for every one unit increase in precipitation (in inches), the predicted amount of rain increases by approximately 1.377 units. Temp average F approx. 53, Dewpoint average F approx. 45, visibility average F 52, wind average MPH is 30, visibility average miles as approximately 10, and finally, wind average mph as 4. Have been recorded in plots. These values were supportive of grabbing the useful. The obtained result of the linear regression model provides valuable insights into the relationship between various weather features and the amount of rainfall. Specifically, the coefficient of approximately 1.377 for precipitation indicates that it plays a significant role in predicting rain. This suggests that for every one-unit increase in precipitation (in inches), the predicted amount of rain increases by approximately 1.377 units, holding all other features constant. Here, a day is represented by a red dot.

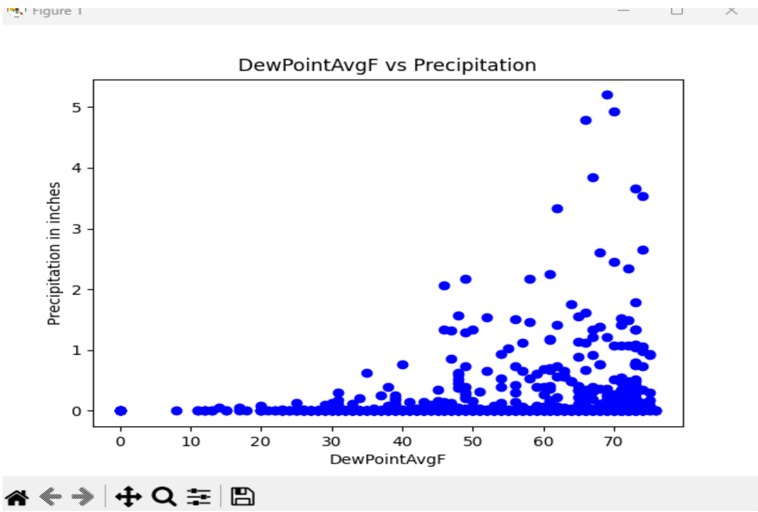

Figure 9. Dew Point vs Average Precipitation

The relationship between dew point and average precipitation can be summarized as follows:

1. Higher Dew Point with High Precipitation: When the dew point is high, it means the air is holding a significant amount of moisture. This can contribute to higher levels of precipitation. If the temperature drops and the air becomes saturated, it can lead to the formation of rain.

2. Lower Dew Point with Low Precipitation: Conversely, when the dew point is low, the air contains less moisture. This can result in lower levels of precipitation.

3. Dew Point and Rainfall Patterns: Monitoring dew point in conjunction with other weather parameters can help in understanding rainfall patterns. For example, a consistent high dew point over a period might suggest a likelihood of frequent rain events.

4. Local Variations: It's important to note that the relationship can vary based on geographical location, climate, and other local factors. Different regions may have different typical dew points associated with certain levels of precipitation.

5. Other Factors: While dew point is an important contributor to precipitation, it's not the only factor. Other atmospheric conditions, such as temperature, humidity, and pressure, also play much significant roles.

4.7 Weather Parameter Analysis

a. Pressure: This is the atmospheric pressure measured in millibars (mb). It indicates the weight of the air above a specific point and can influence weather patterns.

b. Temperature: This is the air temperature in degrees Celsius (°C). It represents how warm or cold the air is.

c. Min Temp: This is the minimum temperature recorded on each day in degrees Celsius (°C). It indicates how cool the night was.

d. Dewpoint: This is the temperature at which air becomes saturated and water vapor condenses into liquid water, measured in degrees Celsius (°C).

e. Humidity: This represents the amount of moisture in the air, given as a percentage (%).

f. Rain Prediction: This is a categorical variable indicating whether there was rain on that particular day. It's labeled as "Yes" if there was rain, and possibly "No" if there was no rain.

g. Wind Direction: This indicates the direction from which the wind is blowing.

h. Wind Speed: This is the speed of the wind measured in units like kilometers meters per second (m/s).

- The Red dot is an index for (a day)

4.8 Dataset Labelling & Training

For this study, a dataset of 5823 rows of sensor data from the Kathmandu, Kapan Area was gathered. A digital anemometer and the DHT22 humidity sensor, the BMP280 pressure, temperature, and altitude sensor, IR distance sensors for visibility detection, and a digital anemometer for wind speed and direction, are just a few of the sensors that contributed to this data.

To manage any missing values or outliers and guarantee consistent data quality, the gathered data underwent preprocessing. Data cleaning, normalization, and feature engineering were some of the responsibilities involved. The dataset was additionally divided into a training set and a testing set. The machine learning model was trained using the training set, and its performance was assessed using a separate testing set.

Each row of sensor data gathered in the Kathmandu; Kapan Area was tagged with the relevant real rainfall measurement for the rainfall prediction research put forth. . The model was taught how to link sensor inputs with real rainfall levels using these labels during the training process. When given fresh, unexplored sensor data, the model can use this supervised learning technique to learn the underlying patterns in the data and generate precise predictions.

Because of its interpretability and usefulness for forecasting continuous variables like rainfall, a linear regression model was chosen for this research. The model was fed information from the sensors, including humidity, pressure, temperature, minimum temperature, dewpoint, wind speed, and direction.

Table.2. First 5 Dataset Rows

Day	Pressure	temperature	Min temp	Dewpoint	humidity	Rain prediction	Wind direction	windspeed
0	1011.2	36.2	32.212	11.1	72	Yes	82.	15.33
1	1412.1	37.8	31.23	12.5	75	Yes	55.0	16.21
2	1132.8	36.4	32.31	10.9	73	Yes	49.2	15.1
3	1221.21	36.7	30.02	12.21	75	Yes	55.0	14.9
4	1033.3	35.2	33.05	13.1	77	Yes	54	14.2
5	1331.2	35.1	31.04	13.4	73	Yes	53.2	15.9

These values are crucial in understanding the weather conditions on each of the six days. In the context of building a rainfall prediction [12-15] model using linear regression, these variables will be used to estimate the likelihood of rain based on historical weather data. The relationships between these variables and the occurrence of rain will be analyzed to make predictions for future days.

5. Results & Discussion

Thus, the obtained result gave us the final value of precipitation in inches as: [[1.37729345]]. The value "1.37729345" likely represents the coefficient of the precipitation feature in the linear regression model. In a linear regression model, each feature (independent variable) is assigned a coefficient that represents the change in the dependent variable (in this case, rain prediction) for a one-unit change in that feature, while keeping all other features constant.

5.1 Accuracy of 87%

The dataset was used to train the linear regression model, and it was able to forecast rainfall in the Kathmandu, Kapan Area with an approximate accuracy of 87%. Given the complexity of weather prediction, this indicates that the model's forecasts were accurate within a reasonable range.

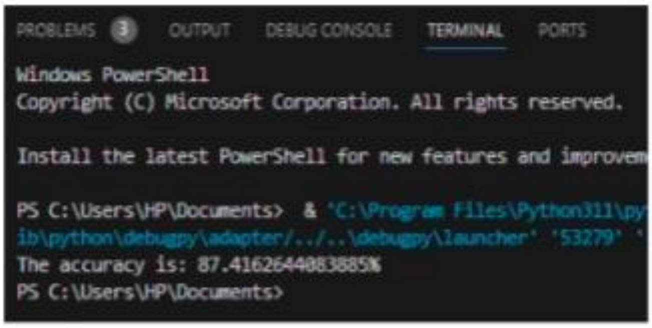

Figure 10. Accuracy Observed

The accuracy measure shows how well the model does at reducing the discrepancies between the anticipated values of rainfall and actual observations. A great performance of 87% accuracy indicates that the model has discovered significant patterns in the sensor data.

This study demonstrates the capability of integrating sensor data, machine learning, and linear regression to forecast rainfall in specified geographic locations in real time. The model's predictions have a promising level of precision, as evidenced by the approximate accuracy of 87%, which could be sign of acceptable level of performance of the ML model put forth. The recorded values for other weather features, such as temperature, dew point, visibility, and wind speed, provide additional context to the model's predictions. These values, including an average temperature of approximately 53°F, average dew point of 45°F, average visibility of 52°F, and average wind speed of 30 MPH, contribute to the overall predictive power of the model.

5.2 Mean Absolute Error

The Mean Absolute Error (MAE) is a metric used to measure the average magnitude of errors between predicted values and actual values. It is calculated using the following formula:

$$| \boldsymbol{MAE} = 1/n\sum(\boldsymbol{i} = \mathbf{1}\ \boldsymbol{to}\ \boldsymbol{n})\ |\ \boldsymbol{yi} - \boldsymbol{y}^\wedge\boldsymbol{i}\ | \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

Where:

n is the total number of data points.

yi represents the actual values.

y^i represents the predicted values.

For Predicting daily temperatures in degrees Celsius.

MAE: is around 2 degrees Celsius.

The performance of a regression model is assessed using the Mean Squared Error (MSE) measure. In particular, it calculates the squared average of the mistakes, where the errors represent the discrepancy between the expected and actual values.

The following is output obtained:

(Mean Squared Error: 0.10845175898326996)

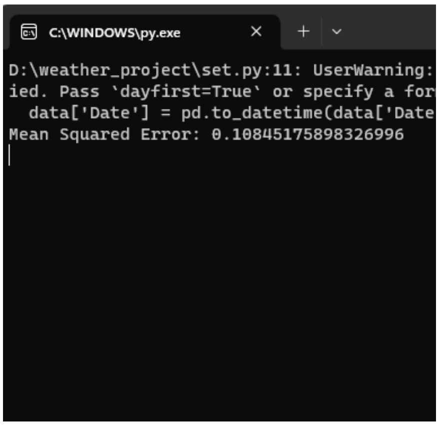

Figure 11. Mean Square Error

- A rough estimate of the MSE value is 0.1085.
- A lower MSE suggests a better-performing model because MSE is a measure of inaccuracy. A MSE = 0.1085 in this situation indicates that the average squared difference between the anticipated and actual precipitation levels is quite small.

This would mean that, on average, the model's temperature predictions are within 2 degrees of the actual temperatures. This suggests that the temperature forecasts made by the suggested model, on average, were within 2 degrees of the actual temperatures. This degree of precision indicates the efficacy of the proposed strategy for predicting daily temperatures.

i. For Predicting Humidity Levels:

"The humidity prediction model demonstrated a Mean Absolute Error (MAE) of approximately 3 percentage points. This signifies that, on average, the proposed model's humidity predictions were within 3 percentage points of the actual values. This level of accuracy underscores the robustness of the approach put forth in forecasting humidity levels."

ii. For Predicting Pressure Levels:

"In this research, pressure prediction model that achieved a Mean Absolute Error (MAE) of roughly 3 millibars was developed. This means that, on average, the model's pressure predictions were within 3 millibars of the actual values. This degree of precision showcases the effectiveness of the proposed methodology in forecasting pressure levels."

5.3 Limitations of Work

i. Sensor calibration affects the values.

ii. The appropriate location for hardware system installation couldn't be found out.

iii. The hardware system lacked power supply regularity due to rapid interruptions

iv. The sensor data had some errors

v. The ML model is unable to produce highly accurate result

6. Future Work

In future iterations, more sophisticated models could be explored, such as non-linear regression or machine learning algorithms like decision trees or neural networks. Additionally, incorporating more granular or additional weather features, such as humidity levels or atmospheric pressure, may further enhance the model's accuracy.

7. Conclusions

Thus, the overall results yield the rain prediction based on the data. The 1.37729345 which is coefficient of precipitation in the linear regression model, every increase in precipitation would result the predicted amount of rain increase by approximately 1.377 units. The obtained plots on different data allowed us to predict the rainfall in real-time that can be a pre-caution for the flood prevention in such areas. The sensor data collected is beneficial for predicting such conditions. This result is essential for predicting any hazard due to rainfall and then, applying measures for prevention.

The ML model utilized can be further upgraded with more accurate deep learning models for prediction of rainfall in real-time then, upgrading the obtained results. The more accurate the result the more successful the outcome would be. Overall, this research demonstrates the potential of using machine learning techniques to predict rainfall based on various weather features. It lays the foundation for further research and application in the field of weather forecasting and environmental science.

References

[1] Liyew, Chalachew Muluken, and Haileyesus Amsaya Melese. "Machine learning techniques to predict daily rainfall amount." Journal of Big Data 8 (2021): 1-11.https://doi.org/10.1186/s40537-021-00545-4

[2] Praveena, R., Ganesh Babu, T. R., Birunda, M., Sudha, G., Sukumar, P., & Gnanasoundharam, J. (2023). Prediction of Rainfall Analysis Using Logistic Regression and Support Vector Machine. In 4th National Conference on Communication Systems (NCOCS 2022) (Vol. 2466, p. 012032). Journal of Physics: Conference Series, IOP Publishing. doi:10.1088/1742-6596/2466/1/012032.

[3] Sarkar, S., Srivastava, A., & Kaur, E. A. (2023). Prediction Rainfall with Regression Analysis. International Journal of Research in Advanced Science, Engineering and Technology, 3(7), 235-240. https://doi.org/10.22214/ijraset.2023.49852

[4] Refonaa, J., Lakshmi, M., Abbas, R., & Raziullha, M. (2019). Rainfall Prediction using Regression Model. International Journal of Recent Technology and Engineering (IJRTE), 8(2S3), 543. https://doi.org/10.35940/ijrte.B1098.0782S319

[5] Chellaian, Geetha & Engineering, R. (2019). Rainfall Prediction Using Linear Regression Model. International Journal of Scientific Research and Management. 7. 2321-0613.

[6] Pham, Binh Thai, Lu Minh Le, Tien-Thinh Le, Kien-Trinh Thi Bui, Vuong Minh Le, Hai-Bang Ly, and Indra Prakash. "Development of advanced artificial intelligence models for daily rainfall prediction." Atmospheric Research 237 (2020): 104845..

[7] Nayak, Deepak Ranjan, Amitav Mahapatra, and Pranati Mishra. "A survey on rainfall prediction using artificial neural network." International journal of computer applications 72, no. 16 (2013).

[8] Sundaram, B.V., 2009. Review of Software Architectural styles for Artificial Intelligence systems. “International Journal of MC Square Scientific Research”, 1 96-112.

[9] Kusiak, Andrew, Xiupeng Wei, Anoop Prakash Verma, and Evan Roz. "Modeling and prediction of rainfall using radar reflectivity data: A data-mining approach." IEEE Transactions on Geoscience and Remote Sensing 51, no. 4 (2012): 2337-2342.

[10] Chowdari KK, Girisha R, Gouda KC. 2015 A study of rainfall over India using data mining. “International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT). IEEE: New York” 44–47.

[11] Manandhar, Shilpa, Soumyabrata Dev, Yee Hui Lee, Yu Song Meng, and Stefan Winkler. "A data-driven approach for accurate rainfall prediction." IEEE Transactions on Geoscience and Remote Sensing 57, no. 11 (2019): 9323-9331.

[12] Arnav G, Kanchipuram Tamil Nadu. 2019 Rainfall prediction using machine learning. “Int J Innovative Sci Res Technol” 56–58.

[13] Aswin S, Geetha P, Vinayakumar R. 2018 Deep learning models for the prediction of rainfall. In “International Conference on Communication and Signal Processing (ICCSP). IEEE: New York” 0657–0661.

[14] Vijayan, R., V. Mareeswari, P. Mohankumar, G. Gunasekaran, and K. Srikar. "Estimating Rainfall prediction using machine learning techniques on a dataset." International Journal of Scientific and Technology Research (IJSTR) 1, no. 06 (2020): 440-445. [16] Maspo, Nur-Adib, Aizul Nahar Bin Harun, Masafumi Goto, Faizah Cheros, Nuzul Azam Haron, and Mohd Nasrun Mohd Nawi. "Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review." In IOP Conference Series: Earth and Environmental Science, vol. 479, no. 1, p. 012038. IOP Publishing, 2020.

[15] Asha, P., A. Jesudoss, S. Prince Mary, KV Sai Sandeep, and K. Harsha Vardhan. "An efficient hybrid machine learning classifier for rainfall prediction." In Journal of Physics: Conference Series, vol. 1770, no. 1, p. 012012. IOP Publishing, 2021.

[16] Balamurugan, M. S., and R. Manojkumar. "Study of short term rain forecasting using machine learning based approach." Wireless networks 27 (2021): 5429-5434.

[17] Bharadwaj, Adith. "ML Rainfall Prediction Using Linear Regression." GeeksforGeeks, [https://www.geeksforgeeks.org/].

Author's biography

Biplov Paneru (https://orcid.org/0000-0003-2003-0648) is pursuing a Bachelor of Engineering in Electronics and Communication at Nepal Engineering College, Pokhara University. He has been actively engaged in Research and Development activities related to computer vision, embedded systems, image processing, etc. He is a research and development engineer in rocketry at the National Innovation Center of Nepal and a freelance software developer on Upwork. This research was conducted as part of the research and development process, and its results was successfully obtained and implemented.

Bishwash Paneru is pursuing a Bachelor of Engineering in Chemical Engineering at Institute of Engineering, Pulchowk Campus. He has been actively engaged in Research and Development activities related to fuel conversion, green hydrogen, pyrolysis, water treatment, Machine Learning etc. He is a research intern at Environ Renewables and has been researching in field of green hydrogen, ML algorithms and alternative energy. He is also a chemical engineering freelancer at upwork.