

Journal of Humanities and Social Sciences Research

www.horizon-JHSSR.com



RESEARCH ARTICLE Peer-reviewed | Open Access

A Web based Weather & Geoinformatics System using Multiple Linear Regression: Data Analytics Approach

Edie Boy M. Dela Cruz*

Department of Mathematics and Natural Sciences, College of Arts Sciences – Caloocan Campus, University of the East, Philippines

ARTICLE INFO

Article history RECEIVED: 12-Mar-25 REVISED: 07-Jun-25 ACCEPTED: 14-Jun-25 PUBLISHED: 15-Jul-25

*Corresponding Author Edie Boy M. Dela Cruz E-mail: <u>edieboy.delacruz@ue.edu.ph</u>

Citation: Edie Boy M. Dela Cruz (2025). A Web based Weather & Geoinformatics System using Multiple Linear Regression: Data Analytics Approach. Horizon J. Hum. Soc. Sci. Res. 7 (S), 69–86. <u>https://doi.org/10.37534/</u> <u>bp.jhssr.2025.v7.nS.id1295.p69-86</u>



ABSTRACT

Introduction: Accurate and timely weather forecasting is crucial for disaster preparedness, especially in flood-prone urban areas like Valenzuela City, Philippines. This study aims to develop a localized web-based weather and geoinformatics system that predicts rainfall probabilities using a data analytics approach. Methods: The system employs multiple linear regression (MLR) algorithms to analyze various meteorological parameters, including wind velocity, atmospheric pressure, temperature, and humidity. Historical weather data were mined to identify patterns and correlations influencing rainfall events. A centralized database was designed to manage the collection, storage, and accessibility of weather-related and geospatial data for key stakeholders. Results: Initial implementation of the system demonstrates the potential to produce reliable and location-specific rainfall forecasts. It also provides essential real-time information such as projected floodwater levels and the availability of nearby evacuation centers, thereby enhancing situational awareness and community responsiveness. Discussion and Conclusion: The integration of MLR-based predictive modelling within a web-enabled geoinformatics platform offers a costeffective and scalable solution for urban disaster risk reduction. The system is expected to support local government units and community members in making informed decisions during weather-related emergencies. Future work will focus on refining prediction accuracy through machine learning enhancements and expanding coverage to other high-risk urban areas.

Keywords: weather forecasting, geoinformatics, machine learning, multiple linear regression, data mining, disaster risk management.

1. INTRODUCTION

Weather prediction has emerged as one of the most challenging domains over the past generation, significantly influenced by advancements in data science. Modern forecasting leverages various technological tools that enhance meteorological analysis beyond traditional methods, which primarily rely on surface observations and aerial research (VamsiKrishna, 2015). Recent innovations, particularly in machine learning and data mining, have transformed approaches to analyzing historical data, facilitating better planning and forecasting across multiple fields, including E-commerce and market exchanges (Yadav & Khatri, 2019). As climate research increasingly relies on data mining algorithms to analyze evolving trends—such as temperature extremes, wind intensity, and precipitation—these methods extract valuable insights from extensive historical data, enhancing our understanding of climate variables over the last decade (Ekereke & Akpojaro, 2019).

In the Philippine context, the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) traditionally oversaw meteorological

Published by BP Services, eISSN.2682-9096 | Copyright © the author(s). This is an open access article distributed under the terms of CC-BY license (<u>https://creativecommons.org/licenses/by/4.0/</u>) DOI: <u>https://doi.org/10.37534/bp.jhssr.2025.v7.nS.id1295.p69-86</u>



forecasting. Meteorological forecasting involves the statistical analysis of potential atmospheric conditions. Climatic conditions refer to the state of the atmosphere at a specific time, expressed in units of key meteorological variables. The accuracy of meteorological predictions varies geographically due to regional disparities in weather patterns. In the Philippines, cloud cover, precipitation, and atmospheric conditions are subject to significant variability and are of particular interest to forecast users. To generate accurate forecasts, meteorologists must possess a comprehensive understanding of atmospheric conditions across a wide area. The precision of their predictions is contingent upon their awareness of regional climate patterns. Forecasting decisions are informed by a variety of predictive methods. The weather chart serves as a foundational tool for meteorologists."

The study aimed to develop a web-based weather forecasting using machine learning. It sought to answer the following questions: (1) How can a web-based system that will mine historical data, provide visualization, and suggest possible actions for the LGU of Valenzuela be developed?, (2) What parameters should the system consider to apply and provide real time results in terms of: (2.a) connectivity, (2.b) hardware, (2.c) (liaison), (2.d) personnel and (2.e) access control, (3) How should the system monitor and assess the weather and flood condition of Valenzuela city and provide possible course of action in the following areas: (3.a) floodwater monitoring, (3.b) high tide, (3.c) evacuation centers and (3.d) available resources. The study focused on the development of a system for weather forecasting and Geoinformatics in the City of Valenzuela. It used multiple linear regression for predicting future forecasts that used available parameter such as a) wind velocity, b)

atmospheric pressure, c) temperature, and d) humidity. The proposed system evaluated and provided weather forecast analysis beneficial to the city government in deciding for cancelation and suspension of classes in each barangay and an alert system for flooded areas.

Figure 1 illustrates the conceptual framework, that employs multiple linear regression for meteorological forecasting using relevant weather parameters. The primary objective of this study is to develop a localized, web-based geoinformatics system capable of predicting rainfall probabilities over the next three days, serving as an alternative weather source for Valenzuela City. Additionally, the system will facilitate communication between the municipal administration and various risk and disaster management officials.

The web-based system will benefit the community by providing timely flood data, information on pumping station levels, details about passable and non-passable areas, tidal data, and the availability of evacuation centers in each barangay. Furthermore, the municipal government will benefit from a centralized database management system that allows all relevant personnel—including barangay officials, officers, and staff—to contribute to the system's maintenance through their own accounts.

2. MATERIALS AND METHODS

2.1 Research Design

Calderon (2006) defined descriptive research as a purposive process of gathering, analyzing, classifying, and tabulating data about prevailing conditions, practices, processes, trends, and cause-effect relationships. This process allows for the adequate and accurate interpretation of such data, with or without minimal aid from statistical methods. Additionally, this approach



Figure 1. Conceptual Framework.

ascertains the prevailing conditions of facts in a group under study, providing either qualitative or quantitative descriptions—or both—of the general characteristics of the group.

This study employed descriptive research. The webbased system provided a dashboard to display information that assists the citizens of Valenzuela City. The prediction of percentage rainfall through data science aids the government in planning and making timely decisions. This platform also facilitates interactive communication among concerned officials and personnel.

2.2 Software Development Process

During the software development phase, the Agile Programming Approach was utilized. First, the engagement of procedures and methods enabled the researcher to determine the reliability of the tool through monitoring and evaluation. Next, the researcher prioritized functionality over extensive documentary evidence. The validity and quality of the web development process enhanced the system's consistency.

Additionally, the researcher fostered constructive client cooperation during contract negotiations, which resulted in mutual advantages for all parties and ensured seamless execution while minimizing potential problems. Ultimately, time and understanding of the design increased, while the ability to implement effective improvements was reduced, leading to lower expenses. In contrast, agile designs routinely accommodated transitions. Finally, Agile approaches to organizing and prioritizing aided project teams in effectively adapting to challenges.

Figure 2 shows the different components agile modeling method approach for software development. The core of this technique ensured that appropriate modeling was better in providing instructions for the application of these critical success factors and ensured that the modeling algorithm was efficient. This approach aimed at improving the consistency of applications. Additionally, this strategy combines iterative and gradual process models, with an emphasis on process flexibility and usability. In recent years, agile methods have acquired development techniques and tractions to speed up project development. This interest in the issue demonstrates that effectively implementing agile approaches offers several advantages. The research was still in its early phases, and the analysis was in its early stages. Thus, the goal of this research was to investigate and acquire insights into current agile practices and methods, as well as to understand the benefits and downsides of agile processes, and to finally answer a variety of concerns about their application.

2.3 Development and Phasing of the System **Phase 1: Brain storming / Planning Stage**

The researcher identified various offices that would be involved in and benefit from this study. Additionally, different variables, parameters, and policies, including relevant protocols, were considered. For the initial system, the researcher incorporated weather attributes, geoinformatics, and other pertinent details about the city, as well as relevant contact numbers, directories, and evacuation centers.

Figure 3 illustrates the operation of the system. The collected data from previous forecasts in Valenzuela City served as the basis for new predictions. As an alternative weather source for Valenzuela, the Geoinformatics system offers weather forecasting capabilities that display the predicted percentage of rainfall for the next three days. This system also facilitates communication between the local government and various risk and disaster management personnel. Additionally, the local government implemented a centralized database management system that enables all authorities—including barangay officials and employees—to contribute to the system's updates.



Figure 2. Agile Modeling Methodology used for Software Development (Buric, 2013).



Figure 3. Operational Diagram.



Figure 4. Standard Rainfall Warning System for Philippines.

2.4 Philippine Public Storm Warning Signals

As an archipelagic region, the Philippines is particularly vulnerable to variable-intensity tropical cyclones, which occur at an annual rate of 15 to 20. The National Weather Service of the Philippines, now known as the State Climatological Agency, has established an official warning system for tropical cyclones to mitigate the damaging effects of these catastrophic events. A color-coded system using red, orange, and yellow shades corresponds to three stages of precipitation intensity expected during floods.

In Figure 4, a yellow rainfall warning is issued when the expected rainfall rate ranges from 7.5 mm to 15 mm within one hour and is likely to continue. In light of this alert, people are advised to stay informed about weather conditions and to be aware that flooding may occur in vulnerable areas. If the rainfall within one hour increases to between 15 mm and 30 mm, an orange rainfall alert will be issued for the affected areas. This orange warning indicates a significant risk of heavy rainfall for the population.

When measured precipitation ranges from 30 mm to 65 mm within one hour and is sustained for three hours, a red rainfall warning is issued. Once a red warning is announced by PAGASA, residents should prepare to take action, as large floods are anticipated, and individuals should be ready to seek safety. The weather disruptions within the Philippine Area of Responsibility (PAR) allow for dynamic adjustments to public storm signals based on the severity and predicted path of storms. The assigned warning percentages are influenced by storm intensity and movement, with variations in strength and location affecting the Public Storm Warning Signals for specific regions.

Weather charts, including surface and upper-air maps, illustrate key atmospheric conditions such as air pressure, wind, humidity, and temperature. Upper-air charts are generated twice daily at twelve-hour intervals to ensure timely data collection. Valenzuela City, situated at an average elevation of 38 meters above sea level and bordered by several municipalities, has its geographic data updated annually by the Valenzuela City Disaster Risk Reduction and Management (VCDRRM) office to monitor changes in infrastructure and surface levels. The system employs multiple linear regression (MLR) to predict rainfall probabilities by analyzing various explanatory factors, enhancing forecasting accuracy compared to single-variable regression. Advanced computational physics software is also used to address geographical and environmental research challenges, generating visual representations of critical data, including weather updates and alert notifications for the city.

Phase 2: Design (System and Software Design)

The researcher used various software tools in performing different algorithms and designs. One of which is Python, a high-level language that is suited for the system design. The system is a web-based design and deployed in alapaap.com as the main webhost. The researcher considered the current website of Valenzuela including the theme, content, and features.

2.5 Software Development Components

The following are the software components of the system: Python, Data Base Management, IP Host, and Server. Python is designed to be simple and easy to understand. Its formatting is clean and straightforward, often favoring English terms over punctuation marks found in other languages. Unlike many other languages, Python does not use curly brackets to delimit code blocks, and while semicolons can be used to end statements, they are rarely employed. In comparison to languages such as C or Pascal, Python features fewer syntactic exceptions and special cases (Bogdanchikov et al., 2013).

Meanwhile, a database management system (DBMS) is a piece of software that allows users to access data stored in databases. The goal of a DBMS is to make defining, storing, and retrieving data in a database as simple and straightforward as possible. The DBMS communicates with the application programs so that the database's contents may be accessed by a variety of applications and users. Furthermore, the DBMS maintains centralized database control, protects data from fraudulent or unauthorized users, and assures data privacy (Gunjal, 2003). Web hosting is a service that allows people and businesses to publish a website or web page on the Internet. A web host, also known as a web hosting service provider, is a company that offers technology and services required to access a website or webpage on the Internet. Websites are kept on servers, which are dedicated computers that host websites. What all Internet visitors need to do is put their website URL or domain into the browser to visit their site. Their machine will then establish a connection with their server, then their websites will be transmitted to them via the browse interface. The system used alapaap web host.

Phase 3. Development (Machine Learning) Phase

Linear Regression Algorithm is a statistical modelling technique that makes use of independent variables to predict the value of the dependent variables. Regression models often vary on the effect of the quality of predictor variables as well as the form of interaction among variables of the study.

Figure 5 shows that there is only one independent variable and it can be seen that there is a positive correlation between independent and dependent variables (x and y respectively). The red line is the regression line, while the dots correspond to the actual data. On the basis of the given data sets, linear regression will attempt to plot a line that better models the points. The line can be modeled on the equations seen below in (1).

Simple Linear Regression is another name for it. It uses a straight line to illustrate the connection between two variables. By determining the slope and intercept that define the line and minimizing regression errors, linear regression aims to construct a line that comes closest to the data. A multiple linear regression occurs when two or more explanatory factors have a linear relationship with the dependent variable. See equations 1 and 2 below.

$$y = bo + b1 * x1 \tag{1}$$

Nonlinear regression is used by statisticians because many data connections do not follow a straight line. Both track a certain reaction from a collection of factors graphically, which is comparable. Nonlinear models, on the other hand, are more difficult to understand than linear models since the function is generated using a sequence of assumptions that may be based on trial and error.

Multiple Linear Regression is defined as when It is unusual for a dependent variable to be explained by only one factor. Multiple regression entails utilizing more



Horizon J. Hum. & Soc. Sci. Res. 7 (S): 69 – 86 (2025)

than one independent variable to explain a dependent variable. There are two types of multiple regressions: linear and nonlinear. Multiple regressions are predicated on the premise that the dependent and independent variables have a linear relationship. It also presupposes that the independent variables have no significant connection. As previously stated, there are a variety of benefits to employing regression analysis. Businesses and economists can utilize these models to assist them in making practical decisions.

$$y = bo + b1 * x1 + b2 * x2 + b3 * x3 + \dots \dots + bn * xn$$
(2)

Phase 4: Quality Assurance and Deployment

Criteria Evaluation Using ISO/IEC 25010. Criteria evaluation is used in evaluating the execution of the algorithms and formulas. Analytics (Descriptive, Prescriptive, and Predictive) & Descriptive Analytics, which includes data collection and data analysis to offer overview of the history and response to: "What occurred?" The system will advise what to do in certain situations and the community. On the other hand, Predictive Analytics, using mathematical simulations and modeling skills to analyze the potential response to: "What will happen?" Finally, Prescriptive Analytics, using design and simulation algorithms, offers guidance on future results and answers: "What should we do?"

3 RESULTS

The main objective of this research was to provide a localized web-based Geoinformatics system with weather forecasting features that deliver rainfall percentages for the next three days, serving as an alternative source of weather updates for Valenzuela City. This system will also facilitate interaction between the city government and various responsible officers assigned to Risk and Disaster Management. The web-based system will greatly benefit the community by providing timely updates on flood data, the levels of pumping stations in preparation for Disaster Risk Management, monitoring of passable and impassable areas, high tide and low tide information, and current updates on the availability of evacuation centers in each barangay. Additionally, the city government will be equipped with a centralized database management system that allows all barangay officials, officers, and personnel to contribute to system updates through their designated accounts.

How can a web-based system that will utilize historical data, provide visualization, and suggest possible actions for the LGU of Valenzuela be developed?

Valenzuela City Government has two weather station sites. One is located in Barangay Ugong and the other is located in Barangay Pulo. These two stations provide weather data to the city government as one of their sources in forecast update. Ugong Station monitors areas near it such as Marulas, Gen T. De Leon, Parada, Mapulang Lupa, Bagbaguin, Paso De Blas, Canumay West and East, Lawang Bato, Punturin, Bignay, Maysan, and Lingunan while Pulo Station monitors most of the flooded areas such as Wawang Pulo, Coloong, Tagalag, Malanday, Pasolo, Malinta, Katuhatan, Rincon, Daandanan, Pasolo, Isla, Bisig, Mabolo, and other areas. These two stations were acquired and purchased by the city government in 2017.

Machine Learning using Multiple Linear Regression

Multiple Linear Regression is a widely used regression method that models the linear relationship between a single dependent variable and a set of independent variables. It is one of the most important regression algorithms, specifically designed to analyze the linear relationship between a single continuous dependent variable and multiple independent variables. In many linear regressions, the target variable (Y) is expressed as a linear combination of multiple predictor variables used in the system for forecasting). Since the multiple linear regression, the equation is an extension of simple linear regression, the equation can be expressed as follows:



Figure 6. Location of Weather Station (Polo and Ugong Station).



Figure 7. Multiple Linear Regression Illustration.

```
X = dataset[[ 'Wind Speed', 'Wind Direction', 'Atmospheric Pressue',
    'Relative Humiditiy', 'Air Temperature', 'Dew Point', 'Precipitation']] # our features
    y = dataset['Rain'] # what we want to predict
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.0000001)
    lm = LinearRegression()
    lm.fit(X_train, y_train)
    # print(lm.intercept_) # the intercept
    # print(lm.coef_) # the coeficient
```

Figure 8. Source Code Importing Data Set.

Table 1. Sample Data Set from (Polo and Ugong Stations).

	Independent Variable					Dependent Variable	
Hourly	Wind Spd. & Wind Dir.	Atm.Pressu	Rel. Hum.	Air.Temp.	Dew Point	Precip.Int	Rain Percetage
Monitoring	2.723729 & 43.78088	1015.04	82.82833	27.27046	24.17673	0	0

Source: Author, 2024



Figure 9. Actual and Normalized Wind Velocity of Polo Station.



Figure 10. Actual and Normalized Wind Velocity in Ugong Station.

Steps in Data Preparation - The first step in data preparation is composed of two subprocesses namely Importing Libraries and Data Set. The source code in Figure 8 shows the system imported libraries and define data sets.

Data Frames hold information in the form of rectangular grids that can be easily examined. Each column in the grid is a vector representing data for a single variable, while each row contains values for an individual instance. This means that the values in a Data Frame's rows do not need to be of the same data type; they can be numeric, character, logical, and so on. Data Frames are two-dimensional labeled data structures that can accommodate diverse types of columns in Python.

Weather Data / Parameters

The two satellites provide weather data such as wind, atmospheric pressure, humidity, air temperature, and dew point. They also give the percentage of the chance of rain on the said date. The collective data include the hourly monitoring for both stations. Weather data are stored and monitored in the office of Valenzuela in the Disaster Risk Management Building. This is also where the reading and forecasted data from the satellite are transmitted to the Weather Monitoring Board. Weather data collected from weather stations have an uneven spatial distribution and density. A single weather station's data is solely indicative of that station's location. Therefore, conversion is required to create weather data for locations between stations. One approach is interpolation, which involves creating new data points within the range of a discrete set of known data points. In figure 3.4 x variables serve as independent variables and y represents the dependent variable which is the rain.

Figure shows the analysis between the Actual and Normalized Wind Velocity. A linear regression analysis is an improved version of ratio-based normalization technique. Normalizing fits the original numbers into a certain range; standardization fits them into a distribution with a mean of 0 and a standard deviation of 1.

The mean for actual values of wind velocity in Polo station within seventy-two hours monitoring was 1.15516 while the standard deviation was 0.46579 respectively. Polo station had an actual value of 0.35834 and 1.437672 peak while the normalized value was -1.7107 and 1.7129. While in the Figure 3. 4 The mean for actual values of wind velocity in Ugong station within seventy-two hours of monitoring was 1.22657 while the standard deviation was 0.53287. For the actual value in Ugong Station, 2.338253 and 0.4807527 were the high and low peaks while for the normalized value, -1.3996 and 2.0869 for low and high peaks.

The MLR model fitted to the training equipment

After the system had prepared the dataset for training, the regression model was fitted to the training set. Linear regression model needed to fit the variables which were the weather parameters, namely the X train and Y train used as the container/storage for parameters

and rainfall forecasting. For Intercept and Coefficient, the simple linear regression model is essentially a linear equation of the form $y = c + b^*x$, where y is the dependent variable (outcome), x is the independent variable (predictor), b is the line's slope; also known as the regression coefficient, and c is the intercept.

Figure 12 shows the analysis of rainfall forecast by the actual wind velocity and the normalized wind velocity using multiple linear regression. The mean value of actual forecast was 0.0394 while the system forecast mean was 0.0375. Meanwhile, 0.0513 was mean value for normalized wind velocity. Using correlation (Pearson) see equation 5, the correlation coefficient was actual forecast and multiple linear regression. Non normalized wind velocity was 0.98622 while normalized with velocity was 0.97475 with a difference of 0.01147.

Figure 13 shows the analysis of rainfall forecast by the actual wind velocity and the normalized wind velocity using multiple linear regression. The mean value of the actual forecast was 0.0452 while the system forecast mean was 0.0561. Meanwhile, 0.0405 was the mean value for normalized wind velocity. Using correlation (Pearson) see equation 5, the correlation coefficient of the actual forecast and multiple linear regression using non normalized wind velocity was 0.98518 while normalized wind velocity was 0.98518 which had 0.036808 difference.

Predicting the test set's outcome. Checking the model's performance is the model's final step. The system will do so by predicting the outcome of the test set. A y pred vector will be created for prediction. The code is as follows:

The formula is Y = a + bX + e, where a is the intercept, b is the slope of the line, and e is the error term. The mean absolute error of a model with respect to a test set is the



Figure 11. Source Code of Fitting Regression from Dataset.



Horizon J. Hum. & Soc. Sci. Res. 7 (S): 69 - 86 (2025)



print(predictions)
mae = metrics.mean_absolute_error(y_test, predictions) #mae
print(mae)
mse = metrics.mean_squared_error(y_test, predictions)
print(mse)
rmse = np.sqrt(metrics.mean_squared_error(y_test, predictions))
print(rmse)

Figure 14. Source Code of fitting prediction from dataset.



Figure 15. System Interfaces with Rain Forecast.

average of the absolute values of the individual prediction errors across all instances in the test set. In contrast, the mean squared error is calculated by taking the distances from the points to the regression line (these distances represent the "errors") and squaring them. Squaring is necessary to eliminate any negative signs and to give more weight to larger differences. This process results in the mean squared error, which represents the average of a set of squared errors. Together, these two processes contribute to more accurate rainfall forecasts, providing a percentage of expected rainfall. In this context, the independent variable is the rainfall forecast, while the dependent variables are the weather.

This is the Home Page of the Web-Based Weather and Geoinformatics System for Valenzuela City, featuring sections on Home, Evacuation Areas, Geoinformatics, and the Admin Panel. In this section, the web-based system provides information on the actual temperature from both satellites, wind speed, cloud formation, wind direction, the times of sunset and sunrise, and the percentage of expected rainfall.

This is the Home Page of the Web-Based Weather and Geoinformatics System for Valenzuela City, featuring sections on Home, Evacuation Areas, Geoinformatics, and the Admin Panel. In this section, the web-based system provides information on the actual temperature from both satellites, wind speed, cloud formation, wind direction, the times of sunset and sunrise, and the percentage of expected rainfall.

Figure 17 displays a map featuring three colors that represent the level of alertness for each barangay. This map is manually monitored by the assigned administrator for the system, indicating various evacuation center locations along with complete details and information.



Figure 16. Flood Alert System and Status of each Barangay.



Figure 17. Flood Monitoring Updates and Status of each Barangay.

Finding comparable events, or K-nearest neighbors, in which to make forecasts is a simple yet successful forecasting method that has been documented in books dating back to the 11th century (Chen and Shah, 2018). Meanwhile, multiple linear regression is an extension of simple linear regression in which values on an outcome (Y) variable are predicted from two or more predictor (X) variables. There are three principal objectives of multiple linear regression (Frey, 2018). Two models were the most popular in terms of forecasting as shown in Table 3

Table 3 shows the correlation of forecasting rate of multiple linear regression and K-nearest neighbors'

regression with the actual prediction from Valenzuela City (Ugong and Polo Stations). The Pearson correlation of multiple linear regression based on its forecasted data versus the actual was 98.62% which was higher compared with the K-nearest neighbors' regression with the rate of 96.92 percentage. Using the T-test, the two models have. The difference between the target value and the prediction value also shows that the value of error is small and the prediction results using KNN Algorithm have an accuracy of 97.27% (Novitasar, 2019). In multiple linear regression, the R2 reflects the correlation coefficient between the observed and fitted values of the outcome

·					
	Polo Station	1	Ugong Station		
K-nearest neighbors regression	Multiple Linear Regression	Actual Forecast : Rainfall Precentage	K-nearest neighbors regression	Multiple Linear Regression	Actual Forecast : Rainfall Precentage
0.092	0.183	0.133	0.133	0.100	0.117
0.008	0.017	0.014	0.003	0.011	0.033
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.007	0.010	0.017
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.082	0.000	0.001	0.008	0.009	0.017
0.018	0.000	0.000	0.013	0.023	0.033
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.009	0.050	0.076	0.013	0.023	0.033
	K-nearest neighbors regression 0.092 0.008 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.082 0.018 0.018 0.000 0.000	Polo Station K-nearest neighbors regression Multiple Linear Regression 0.092 0.183 0.092 0.183 0.008 0.017 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.018 0.000 0.018 0.000 0.018 0.000 0.018 0.000 0.000 0.000	Polo Station K-nearest neighbors regression Multiple Linear Regression Actual Forecast : Rainfall Precentage 0.092 0.183 0.133 0.092 0.183 0.133 0.092 0.183 0.133 0.008 0.017 0.014 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.001 0.000 0.000 0.002 0.000 0.000 0.003 0.000 0.000 0.004 0.000 0.000 0.005 0.000 0.000 0.005 0.000 0.000	Polo Station Actual Forecast : Rainfall Precentage K-nearest neighbors regression Multiple Linear Regression Actual Forecast : Rainfall Precentage K-nearest neighbors regression 0.092 0.183 0.133 0.133 0.092 0.183 0.133 0.133 0.092 0.183 0.133 0.133 0.092 0.183 0.134 0.003 0.008 0.017 0.014 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.001 0.008 0.000 0.000 0.001 0.003 0.000 0.000 0.000 0.013 0.000 0.000 0.000	Polo Station Ugong Station K-nearest neighbors regression Multiple Linear Regression Actual Forecast : Rainfall Precentage K-nearest neighbors regression Multiple Linear Regression 0.092 0.183 0.133 0.133 0.100 0.092 0.183 0.133 0.133 0.100 0.008 0.017 0.014 0.003 0.011 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Table 2. Comparative Analysis of different Model.

Source: Author, 2024

Table 3. Comparison to different Model.

	Multiple Linear Regression	K-nearest neighbors regression
Mean	0.0375	0.0336
Variance	0.0092	0.0038
Pearson Correlation	98.62%	96.92%
	T Test	
t Stat	-0.4311	-0.4610
P(T<=t) one-tail	0.3348	0.3241
t Critical one-tail	1.6991	1.6991
P(T<=t) two-tail	0.6696	0.6482
t Critical two-tail	2.0452	2.0452

Source: Author, 2024

variable (y) in multiple linear regression. As a result, R will always be positive and will vary between zero and one. The amount of variation in the outcome variable y that can be predicted using the values of the x variables is represented by R2. The model explains a considerable percentage of the variation in the outcome variable if the R2 value is near 1. The R2 has the drawback of continually increasing when additional variables are added to the model, even if those variables are only tangentially related to the answer (James et al., 2014).

Provide real time results in terms of:

Connectivity - The main server of Valenzuela City is located at the Alert Division Building, utilizing PLDT and

Table 4. Average mbps of	of Service Provider	Usage
--------------------------	---------------------	-------

	SPEED		
Internet Provider in Philippines	PLDT (60mbps)	Globe Telecom (60mbps)	
Average Speed	57.99 mbps	57.48 mbps	

Source: Author, 2024

Globe Telecommunication as its service providers. The connectivity speed is 60 Mbps, and the server operates 24 hours a day, seven days a week. The webpage functions effectively under the same conditions as the internet providers. Additionally, citizens of Valenzuela can access the web-based platforms even in free data mode.

Table 4 presents the average performance of the internet connection speed for PLDT as the primary provider and Globe Telecom as the secondary provider. The City **Government** of Valenzuela required this setup for backup purposes in case one provider fails to perform. According to the head of the Communication Division, who manages their social media accounts, they have not experienced any connectivity failures thus far, thanks to the presence of network backup. The average performance recorded daily (as shown in Table 4) was 57.99 Mbps for PLDT and 57.48 Mbps for Globe Telecom, respectively

Hardwarea - The two satellites are powered by a 12-15 V source and operate 24 hours a day, seven days

a week. The average voltage used at Ugong Station was 12.91 volts, while Polo Station utilized 12.94 volts. Both stations are equipped with backup generators to be used in the event of service interruptions. Valenzuela City acquired the Ugong and Polo stations, which function continuously throughout the week. Additionally, the city has backup sources, such as generators, capable of sustaining operations for 2-3 days during electric power interruptions. The monitoring device is located at the Alert Division Office, which operates with a power requirement of 5V. Valenzuela City also employs CCTV cameras as remote monitoring platforms.

Liaison Officers - The web-based system is composed of Dashboard, Barangays, Streets, Evacuations, Pumping Stations, Flood Levels, and Users Account. For the "Dashboard," it displays all weather parameters and forecast including the geoformations of Valenzuela City. For "Barangay" section, the admin users can edit all information for each barangay including area, population, and officers. Meanwhile, in the "Users" section, that

Table 5.	Average	Voltage	Source	of	Two	Satellites
----------	---------	---------	--------	----	-----	------------

Voltage Source in	VOLTAGE SOURCE			
Satelite	Ugong Station	Polo Station		
Average Voltage	12.91 V	12.94 V		

Source: Author, 2024



Figure 18. Creation of Unified Users and Access.

admin can add and delete users that are involved within communication. The system creates generic ways for the creation of accounts.

Barangay Officials – Committee on Disaster Risk Management Office (See Figure 4.7 for reference) Each Barangay office has a department head in the Disaster Risk Management office who is responsible for implementing rules and policies, local monitoring if needed, and supervision of personnel for initial action. They will give each an access for system to input an hourly observation if needed that will contribute to the data monitoring of the city. The committee head in Disaster Risk Management from Barangay Councilor will be given an access based on practice of the City Government for the Unified Response Team.

Personnel - In Figure 19 is part of the system where the Head Admin of the portal can monitor all accounts of personnel, offices, and divisions involved in monitoring and decision making for the suspension and response.

The **Flood Control Division** is responsible for monitoring the 35 pumping stations across Valenzuela City, especially during storm signals. This division operates 24 hours a day, seven days a week, ensuring the continuous monitoring of floods and water flow from various rivers in and around Valenzuela City. However, it is important to note that the division currently lacks a data storage mechanism for archiving previous flood observations.

The **Public Safety Division**, operating under the Public Order and Safety Management Office, develops strategies and programs to maintain peace and order within the city's jurisdiction. This division plays a key role in improving the safety of services in the city by providing updated safety protocols for each barangay, as necessary.

The **Traffic Management Office** manages and regulates any activities that could impact or obstruct vehicular traffic flow within the city. It is tasked with granting permissions for activities that may hinder traffic, ensuring smooth transportation across Valenzuela. As part of its responsibilities, the office also assists in flood

			a	Search	0	10 • 📰 • Add User
Date	First Name	Last Name	User Name	Department	Role	Delete
July 25, 2023	Edie Boy	Dela Cruz	edie@gmail.com	NA	Administrator	Delete
August 1, 2024	John Paul	Gonzales	jpgnza@gmail.com	NA	Staff	🔟 Delete
September 2, 2025	James	Rauliz	jrauzil@gmail.com	NA	Administrator	Delete

Figure 19. Users Account Section – Administrative Account.

Showing 1 to 3 of 4 entries





Figure 21. Users Account Log In.

monitoring by recording flood levels on a per-street basis, issuing advisories, and updating the status of streets whether they are passable or not—based on current conditions.

The access protocol, as illustrated in Figure 20, outlines that the admin has the authority to add and delete user accounts within the system. The system is regulated by three division offices, with the Valenzuela City Disaster Risk Reduction and Management Office (VCDRRM) serving as the Head Admin, responsible for assigning accounts to the appropriate personnel or offices. The VCDRRM Office, specifically through the Valenzuela City Rescue Unit, plays a crucial role during calamities and emergencies by assisting residents, while the Planning and Research Office Division monitors storm and flood updates. As the main host of the system, the Rescue Office is central to operations. The Digital Communication Office manages the creation, development, and monitoring of the city's digital platforms, including the Valenzuela City webpage, social media channels, and applications, using these tools to disseminate and cascade information effectively. The Valenzuela Command Center facilitates access to the system for various offices that contribute to data encoding within the centralized database management system. Each barangay is granted one access point for contributing updates, supported by working CCTV cameras for real-time monitoring. The Flood Control Division, which has its own database management platform for monitoring pumping stations, rivers, and flood levels, shares its data directly to the centralized system, ensuring information is visible through the city's webpage. This collaborative approach provides essential information to various offices and the wider community, enhancing disaster preparedness and response.

The Valenzuela disaster risk officer updates the Geoinformatics, Barangay Information, Contact Persons, Contact Numbers, plotting of rescue places, and assessment of data gathered which serve as the main host of the system. Other supporting offices include Digital Communication offices, which is responsible in information dissemination through digital platforms, Traffic Management office for traffic updates during rainy weather, and Public Safety Division.

While certain places are more vulnerable to flooding than others, installing flood warning systems near any major canal or body of water gives important information that may help preserve property and save lives. In fact, the most successful flood warning systems go beyond installing gages and telemetry equipment; they also use skilled personnel and well-established protocols to offer the earliest possible notice regarding whether a flood is predicted, when it will occur, or how severe it will be. Individuals, towns, and organizations that want to set up and operate flood warning systems can use this handbook for guidance. Valenzuela City Administration using social Media Platforms like Facebook and Twitter in updating citizens. Offices use CCTV cameras in monitoring floods.

Figure 23 shows the different tides. The regular rise and fall of the ocean's waters are referred to as high and low tides. When the tide reaches its maximum level, it covers a large portion of the beach. When the water recedes to its lowest level and moves away from the beach, it is said to be at low tide.

The City of Valenzuela had compiled a list of 97 approved evacuation centers spread around the city's 33 barangays, the bulk of which are public schools, covered courts, and barangay halls, with a few private schools and churches thrown in for good measure. The list was compiled after the Department of Education (DepEd) Division of City Schools - Valenzuela issued Division Memorandum No. 170, Series of 2013, instructing district and school supervisors to prepare public schools in the city for use as evacuation centers.

4 DISCUSSION

The researcher utilized weather data from the Ugong and Polo stations, which are optimally located to serve the entire City of Valenzuela. By employing data science techniques, the system leverages historical data through algorithms embedded within the platform. Multiple

Flood Levels

Barangay	Street	Water Level	Passable
Arkong Bato	Arkong Street	2.0	Yes
Karuhatan	Something Street	5.0	Yes
Bagbaguin	Street in Karuhatan	11.0	No
Canumay West	Test Street	11.0	No

Figure 22. Floodwater Level from Typhoon.



Figure 23. Tides (High Tides and Low Tides Record Level).

Tides						
Day	1st Tide	2nd Tide	3rd Tide	4th Tide		
11 Sot	12:36am 🛦 0.82 m	.6:49am ▼ 0.27 m	1:16pm 🛦 0.79.m	6:41pm ¥ 0.44 m		
12 Sun	1:01am 🔺 0.95 m	8:09am ▼ 0.22 m	2:33pm 🛦 0.61 m	6:34pm 🔻 0.49 m		
13 Mon	1:34am 🔺 1.07 m	9:420m ¥ 0.17 m				
14 Tue	2:18am 🔺 1.17 m	11:30om ¥ 0.1 m				
15 Wed	3:16am 🛦 1.22 m	1:11pm ¥ 0.02 m				
16 Thu	4:28am 🔺 1.25 m	2:24pm ¥ -0.04 m				
17 Fri	5:50om 🛦 1,26 m	3:18pm ▼ -0.05 m				

Figure 24. Geoinformatics of Centers.

linear regression is used to predict the percentage chance of rainfall based on historical weather data, including humidity, wind speed, atmospheric pressure, air temperature, and dew point. The system also harnesses historical data from the city government's weather records using the same statistical tool. Furthermore, the system provides various visualizations using geo-maps for flood monitoring updates, securing geoinformation for barangays, monitoring pumping stations, and suggesting possible actions through collaborative inputs from different concerned departments, which assists the Valenzuela City Disaster Risk Reduction and Management (VCDRRM) office in providing effective responses and actions.

Parameters for Real-Time System Functionality:

Connectivity: The system is accessible 24 hours a day, seven days a week. The City Government of Valenzuela maintains a strong internet connection, using Globe as the primary service provider and PLDT as a backup, ensuring continuous operation of the web-based system. Each office and barangay have its own internet connection, allowing for updates even during inclement weather conditions. Hardware Parameters: Similar to connectivity, the two satellites operate continuously, 24/7, with backup power supply generators to ensure uninterrupted service. Valenzuela also utilizes CCTV cameras as an alternative monitoring method for floods. However, according to the VCDRRM Office and Traffic Advisory Office, there are times when these cameras may malfunction. The system serves as an additional means of data collection, allowing for human intervention and inputs through actual observations to provide accurate information.

Liaison and Personnel: Different officers and personnel play crucial roles in this platform. Each has an account in the system for inputting and updating data. The Traffic Management Office, using CCTV cameras, can update information regarding street or road conditions, including flood levels and whether routes are passable or impassable. During the rainy season, the Barangay Councilor assigned to the Disaster Risk Management Office updates the flood levels in at-risk areas. This collected data can inform future protocols and policies created by the Flood Control Division and the Public Safety Division Office.

Access Protocol: The VCDRRM office serves as the main administrator responsible for decision-making based on the inputs and data gathered from various offices. The department head controls user accounts for different concerned offices and their respective personnel. He also updates the Geoinformatics-Flood Monitoring sections by manually changing the status (Green, Orange, or Red) in conjunction with the corresponding alert protocols and advisories. The Digital Communications Unit can utilize this information, which may be shared through various social media platforms.

To monitor and assess the weather and flood conditions in Valenzuela City and provide possible courses of action, the web application serves as an integrated system platform that collects updated inputs from assigned and authorized personnel or offices. For weather forecasting, the system predicts the percentage of rainfall from the current day up to the next three days using multiple linear regression, taking into account historical weather data collected from the two stations. The VCDRRM office uses this predicted data as one of their alternative sources for weather conditions, and the entire community, along with other officers, can also monitor the percentage chance of rainfall as it is displayed in the web application.

For Floodwater Monitoring, the web-based system includes two monitoring components on the dashboard: pumping stations and flood levels per street and barangay. Assigned engineers monitor all pumping stations and river conditions in Valenzuela, updating the status and conditions as needed by the VCDRRM. For High Tide Monitoring, the system provides information through data mining for daily and hourly updates regarding high tide and low tide conditions, including potential tide heights (source: Web Crawler).

In terms of Evacuation Centers, the Geoinformatics feature allows citizens to identify available spaces in evacuation centers, including contact numbers, status, and population capacity, along with exact locations. The administrator is responsible for adding and updating the status of these evacuation centers. Lastly, the Available Resources section displays contact numbers for various offices and rescue teams in Valenzuela, facilitating access to essential services.

To evaluate the system's performance, the ISO/ IEC 25010 software standard criteria were employed. Nine evaluators manually checked the system, navigated its features, and assessed its functionality. The ISO/IEC 25010 standard includes seven criteria and twenty-two sub criteria (Yorke & Vidovich, 2016). The web application received an overall rating of 4.85 on a five-point scale, where 5 represents the highest level of satisfaction and 1 the lowest. Usability was the system's strongest attribute, earning a score of 5, indicating that evaluators were satisfied with its understandability, learnability, operability, and attractiveness. Reliability scored 4.76, the lowest among the categories, with Recoverability being the weakest subcategory, averaging 4.64. Overall, the system met the ISO/IEC 9126 software standard criteria for software development.

5 CONCLUSION

The following conclusions are drawn from the findings presented above regarding the Web-based Weather and Geoinformatics System for Valenzuela City:

To effectively provide weather forecasts and geoinformatics for Valenzuela City, certain parameters must be considered. The system requires inputs and updates from the relevant offices under established liaison, personnel, and access protocols (including user administration). According to the Head of the VCDRRM, the system is operational and accessible to both administrators and citizens 24/7.

For historical weather data, the administrator should supply this information to the software developer for direct storage in the database. However, during the implementation period approved by high officials, it will become easier to encode weather data. Currently, the system relies on manual data entry. Other offices will also contribute to updating various variables, such as flood levels, weather data, and pumping station conditions, all of which are essential for generating accurate results. The system has successfully integrated these parameters, considering and presenting high tide and low tide status, available resources, and evacuation centers.

The web-based system delivers real-time results. As shown in Figure 4.3, using Pearson correlation, the system, which employs multiple linear regression, demonstrates an accuracy of 98.62% compared to actual forecasts. In contrast, the K-Nearest Neighbors regression model achieves an accuracy of 96.92%. Parameters such as connectivity and hardware play significant roles in this system. The City Government of Valenzuela utilizes PLDT and Globe Telecom as its main and backup providers, with average connectivity speeds of 57.99 Mbps and 57.48 Mbps, respectively (see Table 4). In terms of hardware, the two stations (Ugong and Polo) operate continuously, 24/7, with average power supplies of 12.91 V and 12.94 V, respectively (see Table 5). The system also provides accounts for different users under the liaison, personnel, and access protocol (see Figure 18).

The system facilitates flood monitoring not only for barangays but also for individual streets (see Figure 22). Through this platform, residents of Valenzuela will receive timely updates relevant to their areas, overseen by the VCDRRM. Additionally, the system provides updates for high tide and low tide schedules, which are important considerations for decision-making (see Figure 24). The platform also offers a new feature for evacuation centers (see Figure 23), enabling residents to identify the availability of all registered evacuation centers, including contact persons and numbers, availability status, capacity, and exact locations.

Using the ISO/IEC 25010 software standard to evaluate the system yielded scores of 4.84 for functionality, 4.76 for reliability, 5.00 for usability, 4.86 for efficiency, and 4.80 for maintainability and portability, resulting in an average performance score of 4.85. Therefore, the system is considered to be functioning effectively.

Acknowledgment

The author would like to express his gratitude to the editors and editorial staff of JHSSR for their assistance during publication period.

Funding

The author received no financial support for the research.

Declaration of Conflicting Interests

The author declares that he has no competing interests.

References

- Ahmed, H. A., & Mohamed, S. W. (2021). Rainfall prediction using a multiple linear regression model. 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 12(2), 312–326. <u>https://doi.org/10.1109/iccceee49695.2021.9429650</u>
- Aviles, J., & Esquivel, R. (2019). Mining social media data of Philippine higher education institutions using the naive Bayes classifier algorithm. SSRN Electronic Journal, 1(2), 20–40. <u>https://doi.org/10.2139/ssrn.3379025</u>
- Bagtasa, G. (2021). Analog forecasting of tropical cyclone rainfall in the Philippines. Weather and Climate Extremes, 32, 323– 328. <u>https://doi.org/10.1016/j.wace.2021.100323</u>
- Bowen, T. (2015). Social protection and disaster risk management in the Philippines: The case of Typhoon Yolanda (Haiyan). Policy Research Working Papers. <u>https:// doi.org/10.1596/1813-9450-7482</u>
- Bogdanchikov, A., Suliyev, R., & Zhaparov, M. (2013). Python to learn programming. International Conference on Science & Engineering in Mathematics, Chemistry and Physics, 423, 31–50. <u>https://doi.org/10.1007/423-012027</u>
- Dandan, H., Yajuan, Z., Junfeng, L., Chen, L., Mo, X., & Zhihai, S. (2017). Research on a centralized data-sharing model based on Master Data Management. MATEC Web of Conferences, 139, 00195. <u>https://doi.org/10.1051/</u> <u>matecconf/201713900195</u>
- Datta, A., Si, S., & Biswas, S. (2019). Complete statistical analysis for weather forecasting. Computational Intelligence in Pattern Recognition, 1, 751–763. <u>https://doi.org/10.1007/978-981-13-9042-5_65</u>
- Darjan, K., Dragisa, S., Miodrag, B., Mladjan, M., & Pavle, B. (2019). The evaluation of websites in the textile industry by applying the ISO/IEC 9126-4 standard and the EDAS method. Industria Textila, 69(06), 489–490. <u>https://doi.org/10.35530/it.069.06.1520</u>
- Ekereke, L., & Akpojaro, J. (2019). Mathematical tools for predicting the weather condition of coastal regions of Nigeria: A case study of Bayelsa State, Nigeria. Journal of Applied Sciences and Environmental Management, 23(5), 871. <u>https://doi.org/10.4314/jasem.v23i5.16</u>
- Grover, A., Kapoor, A., & Horvitz, E. (2015). A deep hybrid model for weather forecasting. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 21(15), 379–386. <u>https://doi.org/10.1145/2783258.2783275</u>
- Gupta, S., Indumathy, K., & Singhal, G. (2016). Weather prediction using normal equation method and linear regression techniques. International Journal of Computer Science and Information Technologies (IJCSIT), 7(3), 1490– 1493. <u>https://doi.org/10.31120/iicsit.b1490.097456</u>
- Janani, P., & Sabastian, P. (2014). Analysis on weather forecasting and techniques. International Journal of Advanced Research in Computer Engineering & Technology. <u>https://doi.org/10.4314/10.1016/c2013-0-13585-4</u>
- Jones, D. M., Watton, J., & Brown, K. J. (2005). Comparison of hot rolled steel mechanical property prediction models using linear multiple regression, non-linear multiple

regression, and non-linear artificial neural networks. Ironmaking & Steelmaking, 32(5), 435–442. <u>https://doi.org/10.1179/174328105x48151</u>

- Prabakaran, S., Kumar, & Tarun, P. (2017). Rainfall prediction using modified linear regression. ARPN Journal of Engineering and Applied Sciences, 12(12), 3715–3718. https://doi.org/10.6608
- Raziullah, M., Abbas, R., & Leonora, L. (2019). Rainfall prediction using regression model. International Journal of Recent Technology and Engineering (IJRTE), 8(2S3), 543–546. <u>https://doi.org/10.35940/ijrte.b1098.0782s319</u>
- Reddy, N., Patil, N. S., & Laddimath, R. S. (2021). Climate change impact assessment on precipitation over the Ghataprabha sub-basin, India, under future climate change scenarios. International Journal of Recent Technology and Engineering, 8(5), 45–48. <u>https://doi.org/10.21203/rs.3.rs-890363/v1</u>
- Ribera, P., García-Herrera, R., & Gimeno, L. (2008). Historical deadly typhoons in the Philippines. Weather, 63(7), 194– 199. <u>https://doi.org/10.1002/wea.275</u>
- Richey, R. C., & Klein, J. D. (2005). Developmental research methods: Creating knowledge from instructional design and development practice. Journal of Computing in Higher Education, 16(2), 23–38. <u>https://doi.org/10.1007/ bf02961473</u>
- Seref, B., & Bostanci, E. (2018). Sentiment analysis using naive Bayes and complement naive Bayes classifier algorithms on Hadoop framework. 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). <u>https://doi.org/10.1109/ismsit.2018.8567243</u>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. Journal of Business Research, 70, 263–286. <u>https://doi.org/10.1016/j.jbusres.2016.08.001</u>

- Suyatno, J. A., Nhita, F., & Rohmawati, A. A. (2018). Rainfall forecasting in Bandung Regency using the C4.5 algorithm. 2018 6th International Conference on Information and Communication Technology (ICoICT), 324–328. <u>https://doi. org/10.1109/icoict.2018.8528725</u>
- Vaníček, J. (2012). Data gathering for science and research. Agricultural Economics (Zemědělská Ekonomika), 50(1), 29–34. <u>https://doi.org/10.17221/5163-agricecon</u>
- VamsiKrishna, G. (2015). An integrated approach for weather forecasting based on data mining and forecasting analysis. International Journal of Computer Applications, 120(11), 26–29. <u>https://doi.org/10.5120/21273-4093</u>
- Venkata Subba Reddy, P. (2021). Data mining and fuzzy data mining using MapReduce algorithms. In Data Mining
 Methods, Applications and Systems. <u>https://doi.org/10.5772/intechopen.92232</u>
- Wei, H. H., Subramanian, A. C., Karnauskas, K. B., DeMott, C. A., Mazloff, M. R., & Balmaseda, M. A. (2021). Tropical Pacific air-sea interaction processes and biases in CESM2 and their relation to El Niño development. Journal of Geophysical Research: Oceans, 126(6), 228–228. <u>https:// doi.org/10.1029/2020jc016967</u>
- Yadav, R., & Khatri, R. (2019). Weather forecasting using modified K-means intra and inter-clustering algorithm. International Journal of Recent Technology and Engineering, 8(2S3), 517– 521. <u>https://doi.org/10.35940/ijrte.b1092.0782s319</u>
- Yorke, J., & Vidovich, L. (2016). Theoretical frameworks for policy analysis. In Learning Standards and the Assessment of Quality in Higher Education: Contested Policy Trajectories (pp. 69–87). <u>https://doi.org/10.1007/978-3-319-32924-6_4</u>

Biographical Statement of Author(s)

Dr. Edie Boy M. Dela Cruz is a seasoned academic and professional specializing in information technology, engineering, and education. He earned his Doctorate in Information Technology from the University of the East, Manila (2022), and is currently completing a Doctor of Education with



a specialization in Educational Management at the National University, Manila. He also holds a Master's degree in Engineering Science from the University of the East, Manila (2016), and is pursuing a Master of Science in Computer Engineering with a concentration in Data Science and Engineering at the Polytechnic University of the Philippines. He obtained his Bachelor of Science in Computer Engineering from the University of the East, Caloocan (2013).

Dr. Dela Cruz currently serves as the Curriculum Development Coordinator at the University of the East. He has previously held leadership roles as Dean of the College of Engineering at both Colegio De San Lorenzo and AMA University. His scholarly work includes SCOPUS-indexed publications on IoT-based solid waste management systems and web applications employing fuzzy logic methodologies. He is also the author of *Lecture Guide and Worksheets – Statics of Rigid Bodies*.

A frequent keynote and resource speaker, Dr. Dela Cruz has delivered talks on virtual learning environments, data mining trends, developments in information technology, leadership, and research methodology. His continued commitment to academic excellence and technological innovation plays a vital role in shaping the next generation of engineers and IT professionals.

Dr. Edie Boy M. Dela Cruz

Department of Mathematics and Natural Sciences College of Arts and Science Caloocan Campus University of the East Philippines Email: edieboy.delacruz@ue.edu.ph edieboydelacruz63@gmail.com



BY Open Access: This article is published under the CC BY 4.0 (Creative Commons Attribution 4.0 International) license which enables re-users to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

CC BY includes the following elements:

• BY: credit must be given to the creator.

To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.