Civil and Environmental Engineering for Resilient, Smart and Sustainable Solutions

Edited by
Tahar Ayadat

MRF

Civil and Environmental Engineering for Resilient, Smart and Sustainable Solutions

1st International Conference on Civil and Environmental Engineering for Resilient, Smart and Sustainable Solutions, Civil Engineering Department - College of Engineering, Prince Mohammad Bin Fahd University
AL-Khobar - Saudi Arabia, 3-5 November 2024

Editor
Tahar Ayadat[1]

[1] Civil Engineering Department - College of Engineering, Prince Mohammad Bin Fahd University, AL-Khobar - Saudi Arabia

Peer review statement

All papers published in this volume of "Materials Research Proceedings" have been peer reviewed. The process of peer review was initiated and overseen by the above proceedings editors. All reviews were conducted by expert referees in accordance to Materials Research Forum LLC high standards.

Published under License by **Materials Research Forum LLC**
Millersville, PA 17551, USA

Published as part of the proceedings series
Materials Research Proceedings
Volume 48 (2025)

ISSN 2474-3941 (Print)
ISSN 2474-395X (Online)

ISBN 978-1-64490-341-4 (eBook)

Distributed worldwide by

Materials Research Forum LLC
105 Springdale Lane
Millersville, PA 17551
USA
https://mrforum.com

Manufactured in the United State of America

Table of Contents

Preface

On behalf of the conference committee, I would like to thank all the contributors and participants in the 1st International Conference on Civil and Environmental Engineering for Resilient, Smart and Sustainable Solutions (CEES2024) which had been held at Prince Mohammad Bin Fahd University (PMU) Khobar, Kingdom of Saudi Arabia on 3-5 November, 2024. I was honored to serve as Chairman of this important event. Firstly, I would like to thank the organizing committee for their unrelenting efforts to organize this congress. In the world community, we were all aware of how important research has been in contributing to the body of knowledge and the development of prospects. The conference aims to exchange scientific information and knowledge in the development of recent and future infrastructures that are resilient, smart, and sustainable. The conference provides an excellent environment for government policy makers, practicing professional engineers, researchers, university professors, students, and general public to extend their interests and expertise in addressing and solving the infrastructure issues faced by societies. The success of CEES2024 reflects a collective commitment to advancing civil and environmental engineering practices. As participants departed Al-Khobar, they carried with them not just new knowledge but also a renewed sense of purpose in shaping a sustainable future. Finally, I thank the keynote speakers, presenters and authors for their contributions.

Dr. Faisal Yousif Al Anezi
Conference Chair

Committees

Development model of small orifice flow using simple linear and multilinear regression

Dina P.A. HIDAYAT[1,a *], Sih ANDAJANI[1,b], Muhammad A.G. IMANULLAH[1,c]

[1]Universitas Trisakti, Kyai Tapa Grogol Jakarta, Indonesia

[a]dina.hidayat@trisakti.ac.id, [b]andajani@trisakti.ac.id, [c]adzkarghifary@gmail.com

Keywords: Laboratory, Linear Regression, Machine Learning, Model, Orifice Flow

Abstract. The concept of "small orifice flow" describes the movement of a fluid through a tiny hole or orifice. This is a typical occurrence in engineering and fluid mechanics. In machine learning, linear regression is a fundamental and extensively used technique for modelling the connection. In this research, an analysis of the influence of the variables H_1, H_2 and Q was carried out on the output value of the discharge coefficient (Cd). Model evaluation is done by calculating calibration parameters like the coefficient of determination. R square (R^2), also known as the coefficient of determination, describes the extent to which dependent data can be described by independent data. There are a total of 47 datasets that will be included in the development of linear and multilinear models. In this study, 90% (42) of the training data and 10% (5) of the total test data were used. Based on simple linear model, Q parameter having a stronger linear relationship with the Cd parameter compared to the H_1 and H_2 parameters with a strong multiple R value of 0,754 and moderate R_2 correlation value of 0.569. There is a significant influence between the variables Q and Cd based on significant F value. The multilinear regression method was chosen because of its statistical data handling capabilities and ease of creation, implementation, and operation. the multilinear regression model with H_1, H_2 and Q values as independent variables can explain the Cd value as the dependent variable very well with the higher value of multiple R and R^2, and small error value.

Introduction

Small orifice flow refers to the flow of a fluid through a small orifice or hole. This is a common phenomenon in fluid mechanics and engineering, often encountered in applications such as:

- Flow Measurement: Orifices are frequently used as flow meters to measure the rate of fluid flow. By knowing the pressure difference across the orifice and the geometry of the orifice itself, the flow rate can be calculated using various equations (such as the Bernoulli equation or specific empirical formulas).
- Control Valves: Orifices can act as control valves in hydraulic and pneumatic systems. By adjusting the size of the orifice or the pressure conditions, the flow rate of the fluid can be controlled to achieve desired operational conditions.
- Cavitation: In some cases, small orifices can lead to cavitation—a phenomenon where rapid changes in pressure cause the formation of vapor bubbles within the fluid. This can potentially damage equipment or reduce the efficiency of the system.
- Noise Generation: Flow through small orifices can generate noise due to turbulence and pressure fluctuations. In engineering design, minimizing such noise is often a consideration.

The behavior of fluid flow through a small orifice is influenced by factors such as the size and shape of the orifice, the properties of the fluid (density, viscosity), and the pressure conditions upstream and downstream of the orifice. Mathematical models and empirical relationships help in

Materials Research Proceedings 48 (2025) 746-754 https://doi.org/10.21741/9781644903414-81

predicting and understanding the characteristics of small orifice flow in different applications. The flow through the hole in the vessel has its own characteristics including discharge calculation, discharge coefficient, speed, flow, and vessel emptying time. In field applications, the flow through this hole is an analogy of leaks/seepage in water reservoirs, whether ponds, reservoirs or other forms of storage.

Machine learning models are effective models used for phenomena/cases that have a lot of empirical data. In machine learning, linear regression is a fundamental and widely used technique for modeling the relationship between a dependent variable (often denoted as y and one or more independent variables (denoted as x). The goal of linear regression is to find the best-fitting linear relationship that predicts the dependent variable based on the independent variables. The aim of this research is to develop a machine learning model for flow through small holes using Simple Linear Regression (SLR) and Multiple Linear Regression (MLR) methods. The data for this research comes from flow experiments through small holes in the laboratory (Figure 1).

Linear regression is versatile and widely used across various fields, including:

- Predictive Modeling: Predicting outcomes based on historical data, such as predicting sales based on advertising expenditure.
- Economic Analysis: Modeling relationships between economic variables.
- Engineering: Predicting system behavior based on input parameters.
- Social Sciences: Analyzing relationships between variables in sociological studies.

The application of linear regression and multi-linear regression methods in the field of water resources has been developed in several studies, including: coordination analysis between marsh vegetation and hydrology change [1], water quality prediction [2], soil erodibility estimation [3] and non-revenue water prediction [4].

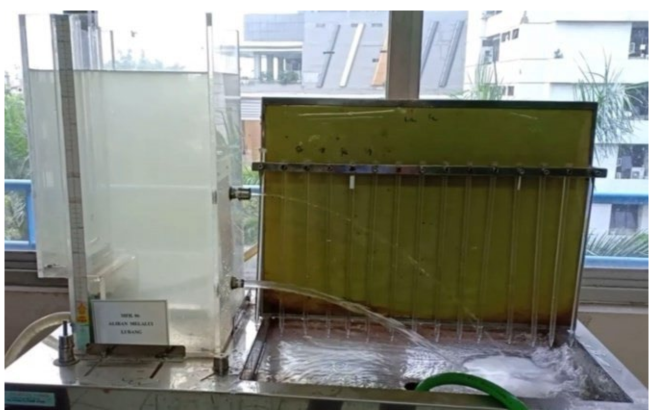

Figure 1. Small orifice flow experimental

Research Methodology

Machine learning (ML) is a machine that was developed to be able to learn by itself without direction from the user. Machine Learning is machine learning aimed at understanding and building a method that utilizes data to improve performance and is developed to be able to obtain various information independently[5]. The implementation of machine learning has been developed for several fields, including: traffic, economics, engineering, health and etc[6][7]. The machine learning model with a simple linear method uses each factor and analyzes its relationship with the output variable, namely the flow rate through the hole. Apart from that, multiple linear models are also used to analyze the influence of all factors together on output variables. The model built will also be analyzed for accuracy using calibration parameters and using several training data and test data scenarios. Linear regression models are useful for the following reasons [8]:

- Descriptive - determining the strength of the relationship between the outcome (dependent variable) and predictor variables.
- Adjustment - It includes the effects of covariates or confounders.

In general, the stages of this research is shown in the Figure 2, consist of:

- Data preprocessing

The fundamental and first stage in transforming unprocessed data into information that is usable is data preparation or data preprocessing. Raw data can generally be noisy, redundant, or incomplete. All of those issues can be fixed and used to create machine learning models[9]. Data preprocessing consists of data preparation, data cleaning (outlier and duplicate data), data training and testing.

- Develop linear regression model

A linear regression model when the dependent variable (Y) is approximated by the independent variable (x) and the intercept value (i), has an equation of the form such as:

$$Y = ax + i$$

- Develop multi linear regression model

While multi-linear equations approach a dependent variable with several independent variables (x_1, x_2…), such as:

$$Y = ax_1 + bx_2 + \cdots + i$$

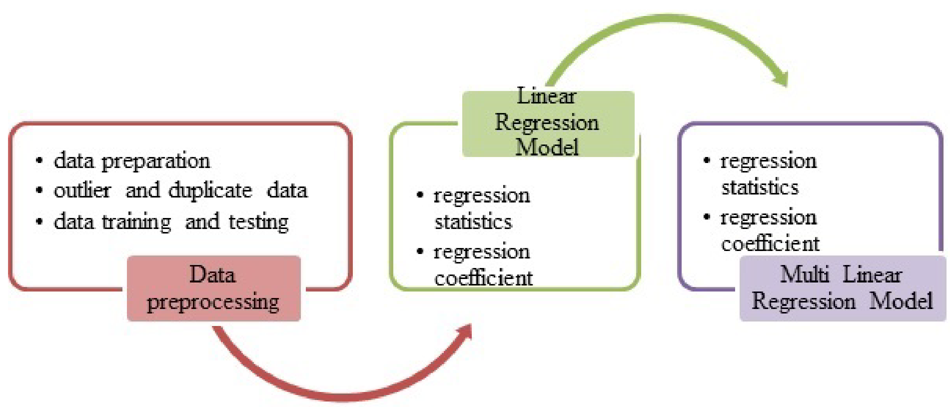

Figure 2. Research step

In this research, an analysis of the influence of the variables H_1, H_2 and Q was carried out on the output value of the discharge coefficient (Cd). H_1 is the initial depth of water, H_2 is the final depth of measurement and Q is water discharge. Model evaluation is carried out by calculating calibration parameters such as the coefficient of determination. R square (R^2) is also called the coefficient of determination which explains how far dependent data can be explained by independent data. The coefficient of determination (R^2) is used to assess the goodness of fit in the linear regression model[10]. R square has a value between 0 – 1 with the better condition that the closer to one with R^2 interpretation as shown as Table 1.

Table 1. R^2 interpretation

Interpretation	R^2 value [11]	R^2 value [12]
Very strong influence	≥0.7	≥0.82
Strong influence	0.4-0.69	0.49-0.81
Moderate influence	0.3-0.39	0.17-0.48
Weak influence	0.2-0.29	0.05-0.16

Linear machine implementations can use various software used to process data, such as Microsoft Excel, SPSS, Python, R and others. In this research, data processing software was used by Google Collaboratory with the Python programming language. All processing was tested using Google Collaboratory by utilizing libraries that support linear regression such as Pandas, Matplotlib, and Seaborn. Functions and benefits of the libraries used include the Pandas library used for data analysis that will be used.

Result and Discussion

Data preprocessing

Data set preparation was carried out by collecting experimental data, especially small orifice flow. Apart from that, data set preparation was carried out by changing the file format to CSV. Data preprocessing consists of the process of checking duplicate data for the same time, missing data/no data, outlier data that is different from the trend of the majority of data and data visualization. Based on the total practical data collected, there are a maximum of 6 data sets which will be eliminated in the initial data processing because they are outside the trend of other data (outliers) (Figure 3). There are a total of 47 datasets that will be included in the development of linear and multilinear models.

Table 2. Statistical parameter

index	H1	H2	Q	Cd
count	53.0	53.0	53.0	53.0
mean	48.283018867924525	47.283018867924525	24.083475471698115	0.6628773584905662
std	1.2766747445936684	1.2766747445936684	2.3088139097735256	0.06315472039580936
min	46.0	45.0	17.39	0.4781
25%	47.0	46.0	23.206	0.6398
50%	48.0	47.0	23.73	0.6609
75%	49.0	48.0	24.852	0.6811
max	50.0	49.0	31.8867	0.891

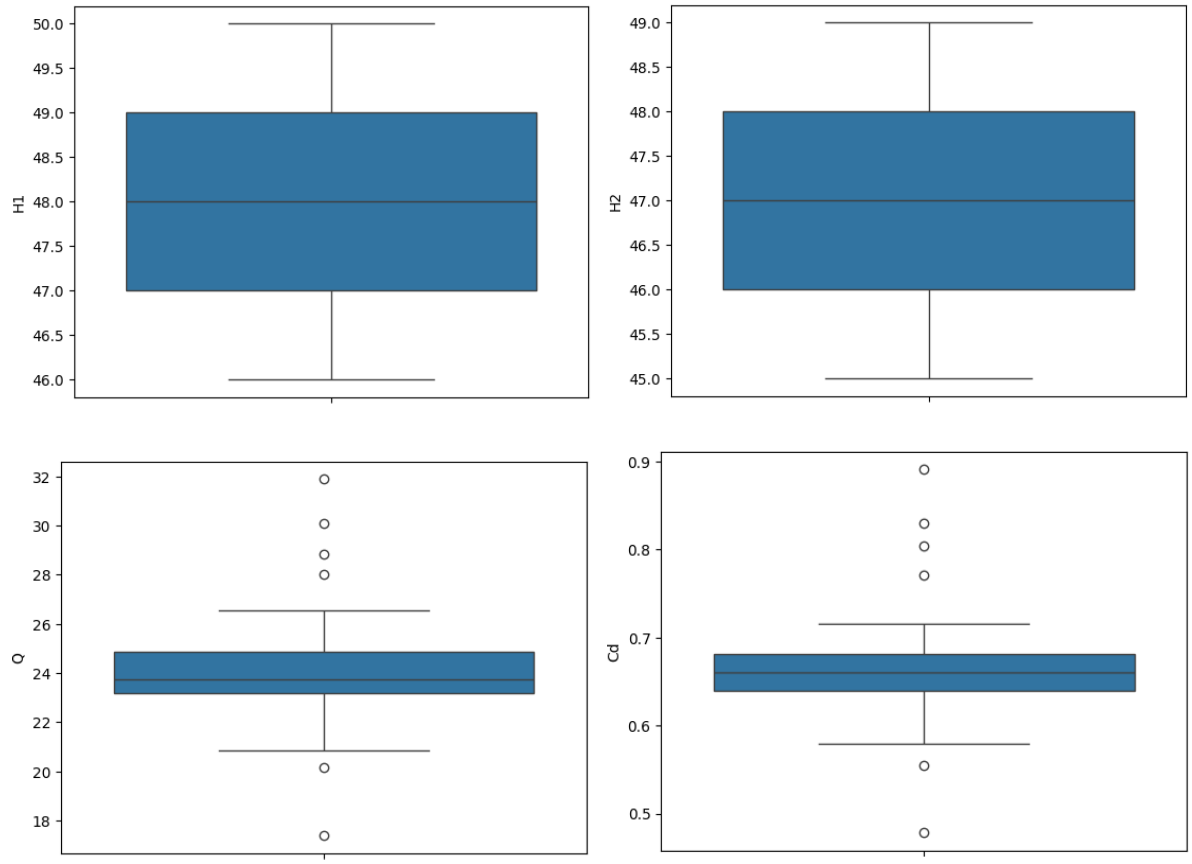

Figure 3. Boxplot distribution of parameter model

Dataset and training

The processing process in machine learning requires data that becomes training data, learning or training and data that is used for learning as training data (training set). The machine learning process will divide the processed dataset into 2 parts, namely data that becomes training data and test data[13][14]. The amount of training data when added to the testing data will be the total dataset. The number of percentages in the training data will later be determined based on repetitions of the cross-validation process[15]. In this study, 90% (42) of the training data and 10% (5) of the total test data were used. The value of training and testing data is shown in Figure 4 and Table 3.

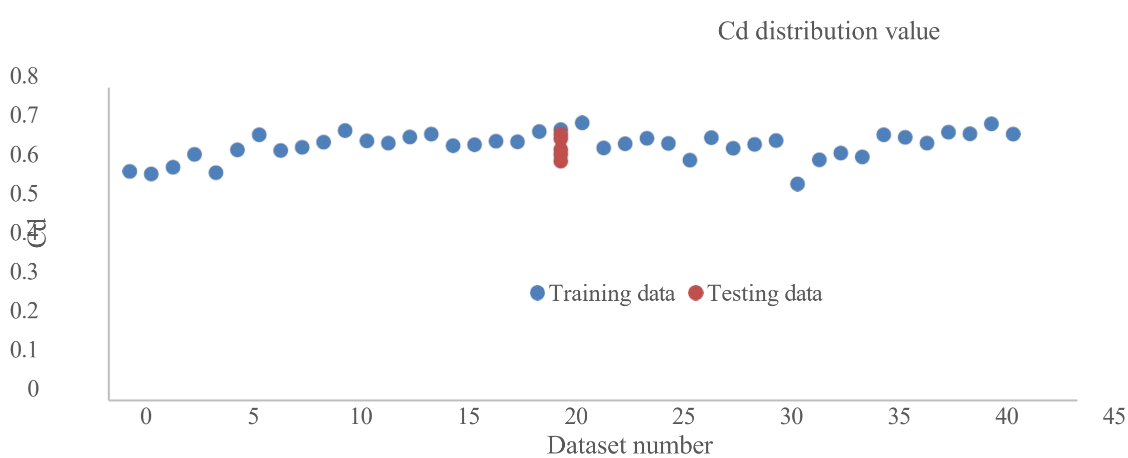

Figure 4. Training and testing distribution data

Table 3. Training and Testing Data

H_1	H_2	Q	Cd
47	46	23.667	0.6699
47	46	22.6757	0.6419
50	49	23.8	0.6294
49	48	25.4545	0.6805
50	49	23.5087	0.6121

Simple Linear Regression Model

Regression models state that the dependent variables are predicted by the independent variables [16]. A case model with a single independent variable is known as simple linear regression [17]. Simple regression modeling with a total of 42 data resulted in the Q parameter having a stronger linear relationship with the Cd parameter compared to the H_1 and H_2 parameters with a strong multiple R value of 0,754 and moderate R_2 correlation value of 0.569. Based on the significance value of F, there is a significant influence between the variables Q and Cd with a significant F value of 7.81608E^{-09}. Table 4 presents regression statistics result for simple linear regression. The simple linear regression equation for the relationship of each variable Q, H_1 and H_2 to the discharge coefficient Cd is shown in Table 5.

Table 4. Regression statistics result for simple linear regression

Regression Statistics	*H_1*	*H_2*	*Q*
Multiple R	0.311868	0.311868	0.754633358
R Square	0.097261	0.097261	0.569471505
Standard Error	0.03367	0.03367	0.023252342
Observations	42	42	42
Significance F	0.044368	0.044368	7.81608E-09

Table 5. Regression coefficient for simple linear regression

Regression Coefficients	*H_1*	*H_2*	Q
Intercept	1.076975	1.068174	0.186493115
x	-0.0088	-0.0088	0.019559729

Multi Linear Regression Model

Sometimes a single independent variable yields insufficient results. This indicates that insufficient information from a single independent variable is needed to forecast the appropriate amount of the dependent variable. We may overcome this issue by using more independent variables and developing a multi linear regression analysis to obtain a meaningful connection [18]. Multilinear regression is the regression model using one dependent variable and more than one independent variables [19]. The multilinear regression method was chosen because of its statistical data handling capabilities and ease of creation, implementation, and operation[20]. Based on multilinear regression analysis with 42 datasets, we obtained very good R_2 and multiple R_2 values close to 1, with very small significant error and F values. Therefore, the multilinear regression model with H_2 and Q values as independent variables can explain the Cd value as the dependent variable very well. Regression statistics result and equation coefficients for multiple linear regression is shown in the Table 6.

Table 6. Regression statistics result and equation coefficients for multiple linear regression

Regression Statistics	
Multiple R	0.979753586
R Square	0.95991709
Standard Error	0.00718527
Observations	42
Significance F	5.10054E-30
Coefficients	
Intercept	0.930049102
H_1	0
H_2	-0.01889717
Q	0.025797796

Data Testing for Simple and Linear Regression Model

The multilinear regression equation with the coefficients in table 4 is then applied to the previously selected data test. The cd value of the data and the resulting model is not significantly different with a small absolute value of total error of 0.019. In more detail, the model error calculation and comparison of Cd values are shown in Table 7 and Figure 5.

Table 7. Error calculation for Cd value with multilinear regression

H_1	H_2	Q	Cd	Cd model	abs error
47	46	23.667	0.670	0.671	0.001
47	46	22.676	0.642	0.646	0.004
50	49	23.800	0.629	0.618	0.011
49	48	25.455	0.681	0.680	0.001
50	49	23.509	0.612	0.611	0.002
		total error			0.019

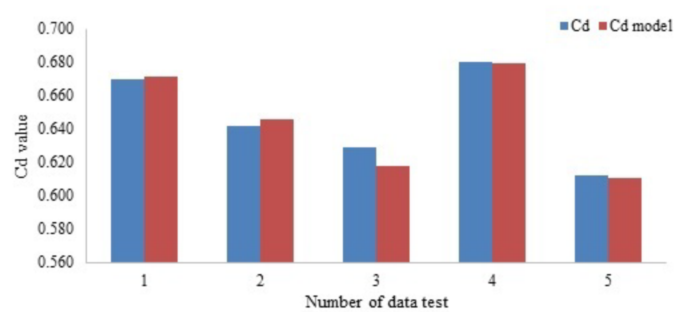

Figure 5. Comparison of Cd data and Cd multilinear regression model

Conclusion

Small orifice flow refers to the flow of a fluid through a small orifice or hole. Simple regression modeling with a total of 42 data resulted in the Q parameter having a stronger linear relationship with the Cd parameter compared to the H_1 and H_2 parameters with a strong multiple R value of 0,754 and moderate R_2 correlation value of 0.569. Based on the significance value of F, there is a significant influence between the variables Q and Cd with a significant F value of 7.81608E^{-09}.

The multilinear regression method was chosen because of its statistical data handling capabilities and ease of creation, implementation, and operation. Based on multilinear regression analysis with 42 datasets, we obtained very good R_2 and multiple R_2 values close to 1, with very small significant error and F values. Therefore, the multilinear regression model with H_2 and Q values as independent variables can explain the Cd value as the dependent variable very well.

References

[1] B. Fu *et al.*, "Spatio-temporal coupling coordination analysis between marsh vegetation and hydrology change from 1985 to 2019 using LandTrendr algorithm and Google Earth Engine," *Ecol. Indic.*, vol. 137, no. February, p. 108763, 2022. https://doi.org/10.1016/j.ecolind.2022.108763.

[2] A. C. P Fernandes, A. R Fonseca, F. A. L. Pacheco, and L. F. Sanches Fernandes, "Water quality predictions through linear regression - A brute force algorithm approach," *MethodsX*, vol. 10, no. January, p. 102153, 2023. https://doi.org/10.1016/j.mex.2023.102153.

[3] M. A. Rehman, N. Abd Rahman, A. N. H. Ibrahim, N. A. Kamal, and A. Ahmad, "Estimation of soil erodibility in Peninsular Malaysia: A case study using multiple linear regression and artificial neural networks," *Heliyon*, vol. 10, no. 7, p. e28854, 2024. https://doi.org/10.1016/j.heliyon.2024.e28854.

[4] M. R. Elkharbotly, M. Seddik, and A. Khalifa, "Toward Sustainable Water: Prediction of non-revenue water via Artificial Neural Network and Multiple Linear Regression modelling approach in Egypt," *Ain Shams Eng. J.*, vol. 13, no. 5, p. 101673, 2022. https://doi.org/10.1016/j.asej.2021.101673.

[5] N. HASYDNA and R. K. DINATA, "Machine Learning." Unimal Press, 2020, [Online]. Available: http://repository.unimal.ac.id/id/eprint/6707.

[6] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper," vol. 5, no. April, pp. 75–82, 2020.

[7] R. G. Wardhana, G. Wang, and F. Sibuea, "Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia," *J. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 40–45, 2023. https://doi.org/10.24076/joism.2023v5i1.1136.

[8] H. C. Kraemer and C. Blasey, "Linear Regression Analysis," *How Many Subj. Stat. Power Anal. Res.*, pp. 73–85, 2017. https://doi.org/10.4135/9781483398761.n6.

[9] G. K. Adari, M. Raja, and P. Vijaya, "Chapter 3 - Machine learning in genomics: identification and modeling of anticancer peptides," A. K. Tyagi and A. B. T.-D. S. for G. Abraham, Eds. Academic Press, 2023, pp. 25–68.

[10] C. L. Cheng, Shalabh, and G. Garg, "Coefficient of determination for multiple measurement error models," *J. Multivar. Anal.*, vol. 126, pp. 137–152, 2014. https://doi.org/10.1016/j.jmva.2014.01.006.

[11] R. Bomvet, D. Saura, R. Jr, and M. Andante, "Detection of cyanide in freshwater fishes relative to sex dimorphism using landmark-based geometric morphometrics in Agusan del Sur, Philippines," *Int. J. Biosci.*, vol. 12, no. 2, pp. 177–193, 2018. https://doi.org/10.12692/ijb/12.2.177-193.

[12] G. A. Nurunnisha, A. Rohmattulah, M. R. Maulansyah, and O. Sinaga, "Analysis of Consumer Acceptance Factors Against Fintech At Bandung SMEs," *Palarch's J. Archeol. Egypt / Egyptol.*, vol. 17, no. 5, pp. 841–856, 2020.

[13] A. Assidiq, P. Hendikawati, and N. Dwidayati, “Perbandingan Metode Weighted Fuzzy Time Series, Seasonal Arima, dan Holt-Winter’s Exponential Smoothng untuk Meramalkan Data Musiman,” *Indones. Gedung D7 Lt.1, Kampus Sekaran Gunungpati*, vol. 6, no. 2, pp. 129–142, 2017, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/ujm.

[14] Y. Astuti, I. R. Wulandari, A. R. Putra, and N. Kharomadhona, “Naïve Bayes untuk Prediksi Tingkat Pemahaman Kuliah Online Terhadap Mata Kuliah Algoritma Struktur Data,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, p. 28, 2022. https://doi.org/10.26418/jp.v8i1.48848.

[15] E. Karyadiputra and N. Hijriana, “Analisis Penerapan Algoritma Naive Bayes Untuk Klasifikasi Prioritas Pengembangan Jalan Di Provinsi Kalimantan Selatan,” *Technol. J. Ilm.*, vol. 9, no. 2, p. 105, 2018. https://doi.org/10.31602/tji.v9i2.1374.

[16] H. Roopa and T. Asha, “A Linear Model Based on Principal Component Analysis for Disease Prediction,” *IEEE Access*, vol. 7, pp. 105314–105318, 2019. https://doi.org/10.1109/ACCESS.2019.2931956.

[17] D. Maulud and A. M. Abdulazeez, “A Review on Linear Regression Comprehensive in Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 1, no. 2, pp. 140–147, 2020. https://doi.org/10.38094/jastt1457.

[18] A. Kayode Coker, “CHAPTER 1 - Numerical Computation,” in *Fortran Programs for Chemical Process Design, Analysis, and Simulation*, vol. 1, 1995, pp. 1–102.

[19] G. K. Uyanık and N. Güler, “A Study on Multiple Linear Regression Analysis,” *Procedia - Soc. Behav. Sci.*, vol. 106, pp. 234–240, 2013. https://doi.org/10.1016/j.sbspro.2013.12.027.

[20] O. Ø. Smedegård, T. Jonsson, B. Aas, J. Stene, L. Georges, and S. Carlucci, “The implementation of multiple linear regression for swimming pool facilities: Case study at Jøa, Norway,” *Energies*, vol. 14, no. 16, pp. 1–23, 2021. https://doi.org/10.3390/en14164825.