UITMCNS
SEREMBAN

MATHEMATICS

IN APPLIED RESEARCH

RECENT **MATHEMATICAL PROJECTS** AT UITM, *SEREMBAN CAMPUS*

Vol. 004 | PREPARED BY: KPPIM, UiTM N. SEMBILAN. | DATE: 20 APR 2023

PERFORMANCE EVALUATION OF FOOD AND BEVERAGES INDUSTRY IN MALAYSIA USING GRA MODELS

FACTORS AFFECTING THE DIAGNOSIS OF ISCHEMIC HEART DISEASE

OPTIMAL VITAMINS INTAKE TO MAINTAIN A HEALTHY DIET USING WEIGHTED GOAL PROGRAMMING

SELECTION OF INSTITUTE FOR PUBLIC HIGHER EDUCATION (IPTA) AMONG FIRST-YEAR STUDENTS USING FUZZY AHP

PREDICTION OF RAINFALL IN KUALA LUMPUR

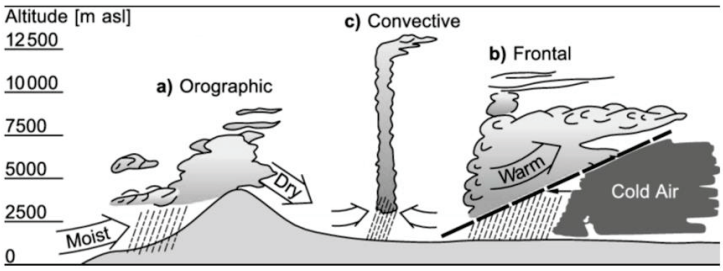
Hurun Ain Mohd Roslan[1], Siti Syakirin Sazali[2] and Haslinda Ab Malek[3]
[1,2,3] Mathematical Sciences Studies, College of Computing, Informatic and Media
Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus,
70300 Seremban, Negeri Sembilan, Malaysia
Corresponding author: haslinda8311@uitm.edu.my

Keywords: Binary logistic regression, Rainfall, Humidity

1. Introduction

Rainfall is one of the most important elements of weather patterns. It is an essential atmospheric phenomenon and influential meteorological parameter in many aspects of daily life (Elvis Selase, 2015). The frequency and amount of rainfall have long been acknowledged as having an impact on the environment, economy, society, and even biodiversity. It is a significant factor and one of the most challenging problems on global and local weather atmospheric circulation. Generally, rainfall events can begin or end at any time of the day and may last a few hours or several days.

The main types of rainfall are orographic, frontal or cyclonic and convectional. Convectional rainfall, frontal rainfall occurs due to what happens when a cold front meets a warm front. Meanwhile, orographic rainfall occurs on the upwind sides of the mountain ranges (Team, 2022). Figure 1 illustrates the information of orographic, frontal and convectional rainfall.



(Source: Wehren et al., 2010)
Figure 1: Information of orographic, frontal and convectional rainfall.

Currently, rainfall prediction is the science and innovation used to predict the state of the climate for a specific area, and the results of rainfall prediction should be accurate. It is mainly concerned with the rainfall prediction conditions for a given environment, as the climatic factors of rain affect many human activities like agricultural production, construction, power generation, and forestry (Hernandez et al., 2016). Thus, predicting rainfall is crucial since it has the strongest association with natural disasters like landslides and flooding. In addition, it can serve as an early warning since knowing the weather forecast far in advance helps manage agricultural farms more effectively and lowers hazards to life and property.

Due to the inability of the current drainage system to handle excessive water flow, Kuala Lumpur suffered torrential rains that could unexpectedly result in several flash floods in just two hours (Ayamany, 2022), The monsoon season typically occurs at the end of the year. However, in March 2022, Kuala Lumpur experienced unnaturally heavy rain. Usually, Kuala Lumpur will experience drought instead. Besides, the ministry's secretary-general, Datuk Seri Zaini Ujang stated that Kuala Lumpur receives 2500mm of precipitation annually, or around 208mm each month. Vigdor (2021) explains that flash floods are so named because sudden downpours can follow heavy rains. Hence, flash floods can happen so fast that they cause damage to live, vehicles, public assets, and infrastructure. Therefore, this study aims to determine the significant factors that contribute to rainfall in Kuala Lumpur.

2. Material and Method

2.1. *Study Area and Data*

This study used secondary data with six variables and 381 observations in Kuala Lumpur from September 1st to September 30th, 2021. Rainfall events is the response variable, whereas maximum temperature, minimum temperature, maximum humidity, minimum humidity, and wind speed are the predictor variables. Harun et al. (2020) stated that Kuala Lumpur is the largest metropolitan area in Malaysia has a population of 7.2 million and a land area of 2793 square kilometers. This represents approximately a quarter of the population of Malaysia. Kuala Lumpur has a tropical climate that is hot and humid throughout the year. There is minimal difference from one season to another. Kuala Lumpur experiences a lot of rain even though it is the driest month of the year. Besides, the monsoon season runs from March to April and September to November, with monthly rainfall varying depending on the season.

2.2. *Binary Logistic Regression*

Binary logistic regression is a relationship between a set of predictors and a binary response variable. It evaluates the association between one or more independent variables and a categorical target variable. A binary regression is used to analyses how changes in the predictor values relate to changes in the probability that an event will occur. In this research, binary logistic regression was used to determine the significant factors that influence rainfall, and a binary response has only two values, which are rain and no rain.

Binary logistic regression, by design, overcomes many of the restrictive assumptions of linear regression. The assumption of binary logistic regression includes the outcome must be discrete, or the dependent variable must be dichotomous in character, there should be no outliers in this data and there should be no high intercorrelations (multicollinearity). The dependent variable in logistic regression is a logistic transformation of the changes (also known as logit):

$$\log(\text{odd}) = \text{logit(P)} = \text{In}\left(\frac{p}{1-p}\right) = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + \ b_5X_5 \qquad (1)$$

Logit, also known as log odds, is a value that can range from 0 to infinity and tells how much more likely an observation is to be a member of the target group than the other group.

$$\text{odds} = \frac{\text{p}}{1-\text{p}} \qquad (2)$$

The odds ratio estimates the change in the odds of being a member of the target group when the predictor is increased by one unit. It is calculated using the predictor's regression coefficient as the exponent.

3. Result and Discussion

3.1. *Omnibus Test of Model coefficients*

The Omnibus Test was acquired in order to assess the effectiveness. If the p-value is below the significance level (α=0.05), the model performs effectively. According to Table 1, the chi square test's (p-value=0.001) was lower than 0.05 ($\chi_{0.05,5}$ =64.677). This result demonstrated how significantly the new model outperformed the baseline model.

Table 1. Omnibus Test of Model Coefficients

	Chi-Square	Degrees of Freedom	P-Value
Step	64.577	5	<0.001
Block	64.577	5	<0.001
Model	64.577	5	<0.001

3.2. *Hosmer and Lemeshow Test*

The Hosmer and Lemeshow test was used to assess how well the model fit the data. If the p-value exceeds the significance level, α=0.05, the model is thought to be well-fit. According to the result in Table 2, the p-value for the Hosmer and Lemeshow Test is 0.512. It was determined that the model fit very well since the p-value was greater than 0.05.

Table 2: Hosmer-Lemeshow Goodness of Fit Test

Chi-Square	Degrees of Freedom	P-Value
7.231	8	0.512

3.3. *Cox and Snell R^2 and Nagelkerke R^2*

By contrasting the log-likelihood of the model and the baseline model, the Cox and Snell R^2 and Nagelkerke $R^2$was determined. The values of the Cox and Snell R^2 and Nagelkerke $R^2$are 0.156 and 0.416 respectively, as shown in Table 3. This indicates that 41.6 percent of the variation in rainfall can be explained by factors related to maximum temperature, minimum temperature, maximum humidity, minimum humidity, and wind speed, while the remaining 58.1 percent may be explained by variables not included in the binary logistic regression model.

Table 3: Model Summary

Cox and Snell R^2	0.156
Nagelkerke R^2	0.416

3.4. *Classification Table*

The classification table is used for assessing predictive accuracy. The results which are shown in Table 4, show that the model was completely correct in forecasting when it might not rain. On the other hand, specifications measure a model's ability to anticipate unfavourable events with accuracy; in this case, the model was able to predict rainfall in the absence of rain with a rate of 37.5%. Meanwhile, the accuracy of the model is measured by the percentage of correct predictions. Overall, it was confirmed that adding the variables to the model improved the model's accuracy, with 96.1% indicating an improvement overall. As a result, it shows that the model used in this study effectively predicts a successful outcome.

Table 4: Classification Table Including All Variable (Block 1)

Observed		Rainfall Events		Percentage Correct
		No Rain	Rain	
Rainfall Events	No Rain	357	0	100
	Rain	15	9	37.5
Overall Percentage				96.1

3.5. *Variables in The Equation*

Variables in the equation are used to determine how much the variable significantly contributes to the model. According to Table 5, the minimum humidity is one of the major characteristics that significantly leads to rainfall (p-value=0.005). As a result, it can be concluded that rain is 1.248 times more likely when humidity is at its lowest. However, as the p-value is bigger than the significant value, the factors such as maximum temperature, minimum temperature, maximum humidity, and wind speed did not significantly affect rainfall.

Table 5: Variables in the Equation

Factor	B	Standard Error	Wald	df	Sig.	Exp(B)
Maximum Temperature	0.056	0.177	0.099	1	0.753	1.057
Minimum Temperature	-0.148	0.197	0.560	1	0.454	0.863
Maximum Humidity	0.165	0.115	2.047	1	0.153	1.179
Minimum Humidity	0.221	0.078	7.951	1	0.005	1.248
Wind Speed	0.169	0.297	0.324	1	0.569	1.184
Constant	-27.56	23.292	1.4	1	0.237	0.000

4. Conclusion

Several locations in Malaysia, including Kuala Lumpur, suffered extremely strong rains in December 2021, causing flooding and landslides. According to binary logistic regression, the minimum humidity is the only variable that significantly influences rainfall in Kuala Lumpur. It is hoped that this study will serve climatologists by assisting in the early identification of natural disasters, providing information on rainfall probability, imparting knowledge to future scholars, and improving societal benefit. It is recommended that further research can be done to enhance the current method for creating a better weather prediction model, such as multinomial logistic regression, to raise the predictions' accuracy.

References

Ayamany, K. (2022). Putrajaya says unusual heavy rain caused kl flash floods, vows long-term mitigation efforts to face climate change. Retrieved from https:// www.malaymail.com/news/malaysia/2022/03/07/putrajaya-says-unusual -heavy-rain-caused-kl-flash-floods-vows-long-term-mit/2046018

Elvis Selase, A. (2015). Precipitation and rainfall types with their characteristic features.

Harun, Z., Reda, E., Abdulrazzaq, A., Abbas, A. A., Yusup, Y., & Zaki, S. A. (2020). Urban heat island in the modern tropical kuala lumpur: Comparative weight of the different parameters. Alexandria Engineering Journal, 59, 4475–4489. doi: 10.1016/ j.aej.2020.07.053

Hernandez, E., Sanchez-Anguix, V., Julian, V., Palanca, J., & Duque, N. (2016). ´ Rainfall prediction: A deep learning approach. In (p. 151-162). doi: 10.1007/ 978-3-319-32034-2 13

Team, C. (2022). Rainfall: Different types explained in layman's language. Retrieved from https://www.clearias.com/rainfall/

Vigdor, N. (2021). What is flash flooding, and how can you avoid it? Retrieved from https://www.nytimes.com/article/flash-flood-safety.html

Wehren, B., Weingartner, R., Schadler, B., Viviroli, D. (2010). General Characteristics of ¨ Alpine Waters. The Handbook of Environmental Chemistry, 17–58. https://doi.org/10.1007/978-3-540-88275-62