

PREDICTION OF DENGUE CASES BASED ON THE METEOROLOGICAL FACTORS BY USING MACHINE LEARNING

Nur Safura Fatimah Mohd Salleh and Muhammad Nabil Fikri Jamaluddin
College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, Perlis Branch
nursafurafatihahb@gmail.com and nabilfikri@uitm.edu.my

ABSTRACT - Dengue that is affected by meteorological factors, including the rainfall and temperature had caused many deaths. Nurulhusna AH even added that “Malaysia is endemic for dengue”. To effectively manage and prevent dengue outbreaks, accurate prediction systems are essential. However, currently Malaysia still lack of system that makes predictions about dengue. Thus, the objective of this project is to develop a web-based system that can predict the number of dengue cases, along with the week and state that have the highest number of dengue cases. System will be developed by using python language, to apply machine learning techniques namely Multiple Linear Regression (MLR), Random Forest Regression, Gradient Boosting Regression (GBR), and Support Vector Regression (SVR). Based on the project, Support Vector Regression (SVR) algorithm was identified as the most suitable model to make prediction on the number of dengue cases. Besides, the week with highest number of dengue cases vary for each state. It also discovered that Selangor is the state with the highest number of dengue cases. In the future, the researcher can perform feature selection to identify significant factors impacting dengue transmission through exploratory data analysis and machine learning-based techniques, other than using a more accurate dataset.

Keywords: Prediction, dengue cases, meteorological factors, machine learning.

1. INTRODUCTION

Dengue cases will peak between 2023 and 2025 (Arumugam, 2023). Thus, we need to have an accurate prediction systems by machine learning that bring many benefits (Raval et al., 2015). Therefore, this project’s objective is to develop a machine learning system for predicting dengue cases, identifying the state and week with the highest number of cases in Malaysia, selecting the most accurate prediction model and evaluating the predicted value with MAPE. It include model developed using Multiple Linear Regression (MLR), Random Forest Regression, Gradient Boosting Regression (GBR), and Support Vector Regression (SVR) technique. Based on Mean Absolute Percentage Error (MAPE), the models were compared. One is chosen to make predictions on the number of dengue cases across 14 states. On top of that, through the implementation of the dengue prediction system using Python programming language, several key findings were observed. Firstly, it was identified that the week with the highest number of dengue cases varied across different states. Secondly, it was determined that the state with the highest number of dengue cases was Selangor. Thirdly, the Support Vector Regression (SVR) algorithm was identified as the most suitable model for making predictions across all states in Malaysia. Last but not least, for future work, future researcher can perform feature selection to identify significant factors impacting dengue transmission through exploratory data analysis and machine learning-based techniques, other than using a more accurate dataset.

2. METHODOLOGY

This project use waterfall model. It’s first phase is literature study, where few materials were studied to increase understanding towards the project. Second phase is data preparation and training. It included data gathering, data merging, data cleaning and data training. The trained data model in this phase were used to developed a prediction system in third phase called development. Web based system is prepared through coding activity, by using python flask. After preparing HTML form, the trained model will be embedded to it. After making prediction, MAPE will be used to find the accuracy of the trained model.

3. RESULTS AND DISCUSSION

3.1 Result for analysis and evaluation

In the end, the week with the highest number of dengue cases varied across different states, indicating the importance of localized predictions. It means that the prediction models and analyses are tailored for each state. Other than that, it was found that the state with the highest number of dengue cases was Selangor. Thus, it is emphasizing the need for targeted interventions in that region by the responsible people. The prediction value produced from the trained data model were evaluated by using the evaluation metric which is Mean Absolute Percentage Error (MAPE). From this evaluation, all the results gained is less than 50% which is considered as an indication that the forecast is reasonable.

4. NOVELTY OF RESEARCH / PRODUCT

The system facilitates mosquito control and health campaigns in specific areas which aiding in combating dengue. It also serves as an early warning system, providing crucial advance information on dengue trends for health officials to devise effective policies, interventions, and monitoring operations. Healthcare workers benefit from improved dengue management and can develop tailored vaccination campaigns. Furthermore, the system is good for practitioner such as it supports research and development of dengue prediction methodologies, offering valuable insights for future infectious disease forecasting studies using diverse machine learning algorithms.

5. CONCLUSION

All of the project's objective are achieved. The week with highest number of dengue cases differ for each state, and Selangor suffer for being the state with highest number dengue cases. Support Vector Regression (SVR) algorithm was identified as the most suitable model to make prediction on the number of dengue cases. For future work, researcher can perform feature selection to identify significant factors impacting dengue transmission through exploratory data analysis and machine learning-based techniques, other than using a more accurate dataset.

REFERENCES

- Arumugam, T. (2023, January 1). "Dengue jab offers hope, but rigorous tests vital." NST Online. <https://www.nst.com.my/news/nation/2023/01/866085/dengue-jab-offers-hope-rigorous-tests-vital>
- IMR-SEAMEO 2021 Webinar: Dengue in Malaysia: An Endemic Amidst Covid-19 Pandemic—YouTube. (2021, June 25). https://www.youtube.com/watch?v=wqjg3jU7M_c
- Raval, D., Bhatt, D., Kumhar, M. K., Parikh, V., & Vyas, D. (2015). *Medical Diagnosis System Using Machine Learning*. 7(1), 6.