Available online at
http://uppp.uitm.edu.my/online-issues.html

e-ISSN: 2689-4934

ESTEEM Academic Journal Vol 20, March 2024, 12–23

ESTEEM
Academic Journal
(EAJ)

Flood damage cost prediction using random forest

Ainul Najwa Azahari[1], Norlina Mohd Sabri[2*]

[1]*College of Computing, Informatics and Mathematics, UiTM Cawangan Terengganu, Kampus Kuala Terengganu, Malaysia*
[2]*Research and Industrial Linkages Unit, UiTM Cawangan Terengganu, Kampus Kuala Terengganu, Malaysia*

ARTICLE INFO

Article history:
Received 20 January 2024
Revised 8 March 2024
Accepted 19 March 2024
Online first
Published 25 March 2024

Keywords:
Prediction
Machine Learning
Flood Damage
Damage Cost
Random Forest

DOI:
10.24191/esteem.v20iMarch.614.g472

ABSTRACT

Floods are one of nature's deadliest catastrophes, causing permanent and catastrophic damage on the socioeconomic system, agriculture and human life. The problems arise when floods could cause a lot of economic damage such as damage to buildings, agriculture and others. Flood damage estimation is a subject of study that has not received much attention. The objective of this research is to explore the Random Forest algorithm in the flood damage cost prediction. The damages specified by the Malaysia's Department of Irrigation (JPS) are structures such as culverts, MTB bridges, riverbank ruins, concrete main channels, farm roads, hydrological stations, agricultural and water drainage, JPS pump houses and tyres in Terengganu. Terengganu is one of the states in Malaysia which has to endure floods during the monsoon season by the end of the year. The methods employed in this research include data collection, data pre-processing, backend engine coding and user interface design. This project was implemented using the Python programming language. The data were collected from the annual flood report provided by the JPS Negeri Terengganu. The research used the rainfall and streamflow data from the year 2012 to 2022 as attributes to forecast the cost of the JPS structures damages in Terengganu. The prediction results showed that the best model achieved the accuracy of 91.47% with a Mean Squared Logarithmic Error (MSLE) of 0.48 and Coefficient of Determination (R^2) of 0.92. In the performance evaluation, the model with 80:20 training and testing data ratio produced the best result in predicting the flood damage cost. The potential enhancements to this research involve extending the scope to encompass all Malaysian states, incorporating diverse flooded structures and adding more input variables for a more improved and more reliable flood prediction system.

[1*] Corresponding author. *E-mail address*: norli097@uitm.edu.my
https://doi.org/10.24191/esteem.v20iMarch.614.g72

©Authors, 2024

1. INTRODUCTION

Prediction is a statistical technique that uses Machine Learning and data mining to anticipate and forecast possible occurrences using the help of past and current data. Machine Learning can continuously learn from the data to create and enhance the prediction model, which can be trained repeatedly to further improve the outcomes or for new scenarios [1]. Machine Learning (ML) is a beneficial technology for predictive analytics systems because it helps to speed up the processing and analysis of data. Prediction in data mining requires an algorithm that can simplify and speed up the process. Nowadays, predictions are widely used in various areas such as entertainment, medical, education and other areas. The application of data mining in various disciplines, such as the study and prediction of damage from natural catastrophes, has been or may continue to be very beneficial. Numerous studies have been conducted utilising data mining to extract important information from the patterns in huge flood data sets and to determine its use to obtain crucial information that may then be used for decision-making at high levels of governance [2]. Prediction is important in accurately predicting the flood damage because flood damage estimation is an essential element in the assessment of flood risk. The development of flood damage models constitutes a crucial step in reducing flood risk disasters given the many uses in mitigation and emergency planning, economic loss appraisal and the cost-benefit analysis for flood protection systems [3]. In order to prevent damage loss and encourage the construction of new homes outside of flood catastrophe zones, estimation of economic loss is crucial.

Malaysia has been frequently affected by floods, a serious natural calamity. According to JPS, areas with the flood risk covers 10.1% which is 33298 km^2 of the total area of Malaysia, involving an estimated of RM1.15 billion a year of annual flood loss. Floods can occur due to various factors such as heavy rain, high sea tide, obstruction of water flow in the drainage system and also the problem of shallow rivers. Floods have caused a lot of economic damage such as damage to buildings, agriculture and others. Flood damage estimation is a subject of study that has not received much attention, particularly in developing nations such as Malaysia. The majority of the studies on this issue have concentrated on calculating the probability of flooding while the studies on damage prediction are less common. Even though there has been a lot of improvement in forecasting this damage, more research is still required [1]. Based on the problem of flood damage, this research has proposed the prediction of flood damage cost using the machine learning algorithm.

Previously, other researchers have adopted various algorithms in predicting the flood damages. Among the algorithms that have been implemented were Decision Tree, Random Forest, Convolutional Neural Networks (CNN), Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Gaussian Naive Bayes, Support Vector Machine (SVM) and K Nearest Neighbour (KNN) [4], [1], [5], [3], [2], [6]. In this research, the Random Forest algorithm has been chosen due to its performance. Both classification and regression problems can be solved using Random Forest, often known as random decision forests. Random Forest is appropriate when training on data similar to a decision tree, and the output is obtained as a mean forecast [2]. The bagging technique used by Random Forest employs numerous decision trees, each of which is trained on a distinct data sample via replacement sampling. The key benefit of utilising Random Forest is that it looks for the best feature among a random subset of features when splitting a node, as opposed to the most important feature. Because of this feature, it produces good results in data mining [2]. Therefore, this research has proposed to implement the Random Forest algorithm in predicting the flood damage cost.

The objective of this research is to explore the Random Forest algorithm performance in the flood damage cost prediction. The scope of the study covers the damage cost of the structures of Jabatan Pengairan dan Saliran (JPS) due to flooding in Terengganu. Terengganu is located on the East Coast of Peninsular Malaysia with an area of 12,955 km^2. Terengganu has been chosen in this research as this state

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

has to endure flooding every year during the monsoon season. The damaged JPS properties are structures such as culverts, MTB bridges, riverbank ruins and concrete main channels. JPS has to calculate the structure damages every year and plan for the next budget. The flood damage prediction system could help JPS to prepare for the worst case scenario of the flood in terms of the costs. This paper is organised into five main sections which begin with the Introduction, Literature Review, Methodology, Results and Finding and finally the Conclusion and Recommendation section.

2. LITERATURE REVIEW

This section provides brief reviews on the similar research of flood damage prediction using machine learning algorithms and on the Random Forest algorithm approach.

2.1 Predictions of flood damage

Flood damage prediction is often solved by using machine learning. For broad spatial domains and intermediate resolution, ML provides an effective and computationally economical alternative to modelling flood damage risk [4]. The algorithm automatically formulates the rules (logic) of the data, which is one of the key benefits of ML. Additionally, ML can continuously learn from the data to create and enhance the prediction model, which can be trained repeatedly to further improve the outcomes or for new scenarios [1]. Many algorithms in ML are used for prediction and each algorithm gets different performance in each research.

There is much research that has been done in the prediction of flood damage. The purpose of this section is to find all of the similar works related to flood damage prediction and the algorithms used by the researchers. The first similar research was predicting flood damage probability across the conterminous United States (US). The majority of natural catastrophe damages in the US are caused by floods. The goal was to employ Random Forest to analyse the geographical distribution and underlying factors that influence the likelihood of flooding due to heavy rainfall and overflowing water bodies throughout the contiguous US. The model had an average area under the curve accuracy of 0.75 when classifying damage or no damage [4].

Another research explored the simplified automatic prediction of the level of damage to similar buildings affected by river flood in a specific area. Many structural components were affected by flooding brought in by overflowing rivers. With just three environmental parameters (minimum distance from the river, unevenness, and potential use of decision trees), the goal of this study was to forecast at least three degrees of affectation in structures. In general, the model's accuracy hovered around 80% [1].

A separate research examined the coupling of the machine learning and weather forecast to predict the farmland flood disaster for Yangtze River basin. Predicting rainfall and node water level, as well as calculating catastrophic damages were the goals of this study. Long Short-Term Memory (LSTM), CNN, Random Forest, and Multiple Perception (MLP) algorithms were employed in this study. The two models with the highest R^2 values were Random Forest (R^2 ranged from 0.7180 to 0.9803) and MLP (R^2 ranged from 0.5717 to 0.9965) [5].

Next, a comparative study evaluated Expert-based versus data-driven flood damage models for data-scarce regions. The modelling of flood damage in numerous areas has been hampered by the lack of actual data. The goal was to compare the effectiveness of the two techniques using defined damage ratings. The conclusion of this research was that utilising the Random Forest technique, high class damage may be predicted with a probability of up to roughly 90% [3].

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

Lastly, a study focused on leveraging machine learning for predicting the flash flood damage in the Southeast of the US. Due to their sudden onset, flash floods were expensive natural hazards. The goal of this research was to use the Random Forest algorithm in order to estimate the property damage caused by flash floods, which is essential for proactive disaster management. Random Forest is 81% accurate in classifying detrimental occurrences [6].

Table 1 shows the summarisation of the similar research on flood damage prediction. Most of the research has shown that the algorithms could solve the different flood damage predictions with good accuracy. Previous studies have consistently shown that Random Forest outperforms other algorithms in flood damage prediction tasks. This research specifically opts for Random Forest to predict flood damage due to its successful track record in solving diverse problems. This research aims to thoroughly investigate the performance of Random Forest and anticipates favourable outcomes in the domain of flood damage prediction.

Table 1. Summary of the Similar Research on Flood Damage Prediction

No.	Technique/ Algorithm	Objective	Problem	Result	Reference
1	Random Forest	To analyse the spatial distribution and underlying drivers of flood damage probability (FDP) caused by excessive rainfall and overflowing water bodies across the conterminous United States	Floods are the leading cause of natural disaster damages in the United State	The model classified damage or no damage with an average area under the curve accuracy of 0.75	[4]
2.	Decision Tree	To predict at least three degrees of affectation in buildings, considering only three environmental factors (minimum distance from the river, unevenness and possible water communication)	Flooding due to overflowing rivers affects the construction elements of many buildings	Around 90% can be achieved in the "Recall" and "Precision" of "HighLevel-Affection" class, and an “Accuracy” around 80% in general	[1]
3.	Random Forest, Multiple Perception (MLP), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM)	Predict two key driving factors of waterlogging, rainfall and node water level and also estimate disaster losses	Floods have caused huge economic losses	The random forest and Multiple perception model (R^2 ranged from 0.7180 to 0.9803 and 0.5717 to 0.9965) performed best	[5]
4.	Random Forest	To evaluate the performance of expert-based and data-driven flood damage models based on developed damage grades	The scarcity of empirical data has limited flood damage modelling in several regions	High class damage is predicted with a probability of up to almost 90%	[3]
5.	Random Forest	Predicting property damage of flash floods is imperative for proactive disaster management	Flash floods are costly natural hazards, primarily due to their rapid onset	The accuracy of Random Forest in classifying damaging events is 81%	[6]

2.2 Random forest

Random Forest algorithm is a supervised machine learning technique that is built using Decision Tree algorithms. The "Forest" of this method is a collection of decision trees [7]. As a result of its versatility and simplicity, it is also one of the most often used algorithms. Random Forest is appropriate when training on data similar to a decision tree, and the output is provided in the form of a mean prediction [2].

The Random Decision Forest also known as Random Forest is used for classification, regression and

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

other tasks that are carried out by building several decision trees [8]. It makes use of ensemble learning, a method for solving complicated issues by combining several classifiers. In conventional Machine Learning, Random Forest is a well-liked and often applied Machine Learning model [9]. Random Forest is one of the greatest machine learning algorithms, which is extensively utilised because of its unique properties and outstanding classification performance. For instance, Random Forest models outperform other predictive models in terms of tolerance to noise in the data and resistance to overfitting [10].

A Random Forest employs the bagging (Bootstrap Aggregation) approach, which makes use of several decision trees, each of which is trained on a separate sample of data and is then used to aggregate the results. Therefore, the ultimate result is established by integrating several decision trees rather than relying on individual decision trees [2]. As a standard resampling technique, Random Forest validates the model and calculates population statistics using bootstrap techniques [10]. Random Forest is regarded as one of the top ensemble classifiers for high-dimensional data. Each tree in a Random Forest is based on the values of a randomly sampled vector, which has the same distribution across all the trees in the forest [11].

3. METHODOLOGY

The methodology section provides more detailed explanations on the experimental data, system architecture and the performance evaluation in the research.

3.1 Experimental data

This research was carried out in Terengganu, Malaysia. The data were collected from the JPS Terengganu hydrology section through online application and annual flood reports of Terengganu. The state is divided into 8 districts, which are Besut, Setiu, Hulu Terengganu, Kuala Nerus, Kuala Terengganu, Marang, Dungun and Kemaman. There are 8 river basins in Terengganu namely Besut River, Ibai River, Dungun River, Keluang River, Kemaman River, Marang River, Paka River and Setiu River.

The requested data were state district, rainfall in millimetres (mm) units, river streamflow in cubic metres per second (m^3/s) units and damage cost of JPS's structures in Malaysian ringgit (RM). Data was taken every day from January 2012 until July 2022, which was for 10 years duration. The rainfall over the four months (November, December, January and February) and along with the streamflow represent the attributes for the prediction system. The prediction is expected to be able to predict the structural damage cost in Malaysian ringgit. A total of 403 data points were gathered from the annual flood report to assess the costs incurred due to damages. The types of possible damaged JPS structures are culverts, MTB bridges, riverbank ruins, concrete main channels, farm roads, hydrological stations, agricultural and water drainage, JPS pump houses, tires or boundaries and others. Table 2 shows the attributes for the prediction of flood damage cost and the predicted output.

Table 2. Attributes of the prediction and the predicted output

Prediction Attributes	Rainfall in November in millimetres (mm)
	Rainfall in December in millimetres (mm)
	Rainfall in January in millimetres (mm)
	Rainfall in February in millimetres (mm)
	River Streamflow in cubic metres per second (m^3/s)
Prediction Output	Damage Cost in Malaysian ringgit (RM)

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

3.2 System architecture

Fig. 1 illustrates the system architecture for the project. This system starts with data preparation. Data is collected from the JPS Terengganu. After collecting data, data is pre-processed by replacing missing values, removing noisy data and saving the cleaned data into a csv file. The processed data will go to the next section to be trained and tested. Data is trained and tested using percentage splits of 90:10, 80:20 and 70:30. Random Forest regression will be applied to the data. Then, calculation of the Mean Squared Logarithmic Error (MSLE), Coefficient of determination (R^2) and Accuracy for performance evaluation will be conducted. After testing and evaluation, the model of Random Forest Prediction will be created. On the other hand, the user will input the data of rainfall and river streamflow in the user interface and connect with the Random Forest Prediction model. After the prediction process, the model will give the result of the predicted flood damage cost to the user.

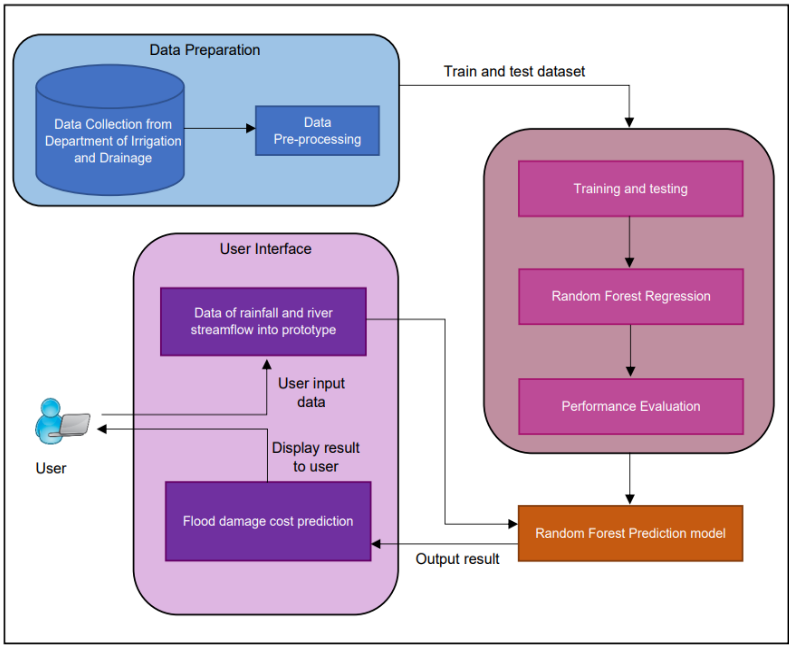

Fig. 1. System architecture for flood damage cost prediction.

3.3 Performance evaluation

The performance metrics used in this research for the evaluation of the Random Forest performance are the Mean Squared Logarithmic Error (MSLE), Coefficient of Determination (R^2) and the accuracy evaluation.

Mean squared logarithmic error (MSLE)

Mean Squared Logarithmic Error is determined by averaging the squared discrepancies between the actual and anticipated values after a log transformation. MSLE has superior numerical characteristics than percentage-based errors. During model training, it strikes a balance between data points with radically different orders of magnitude. The error profile becomes flatter as a result of the logarithm, which also lessens the impact of the bigger numbers. The MSLE formula is as in Eq. (1):

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

$$\text{MSLE} = \frac{1}{n}\sum_{i=1}^{n}(log(yi+1) - log(\hat{y}i+1))^2 \quad (1)$$

where, the predicted value is y = {y1, ̂y2, ..., ̂yn}, the actual value was y = {y1, y2,...,yn}[12].

Coefficient of determination (R^2)

The coefficient of determination (R^2) is a number between 0 and 1 that measures how well a statistical model predicts an outcome. R^2 is a statistical metric in a regression model that establishes the percentage of variation in the dependent variable that can be explained by the independent variable. R^2 or "goodness of fit," measures how well the prediction model matches the actual data and also serves as a measure of prediction quality. The R^2 formula is as in Eq. (2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - yi)^2} \quad (2)$$

where, the predicted value is y = {y1, ̂y2, ..., ̂yn}, the actual value was y = {y1, y2,...,yn} [12].

Accuracy

The accuracy is calculated as the percentage departure of the projected target from the actual target with some allowable error [13]. Accuracy refers to how well the predicted values match the actual values of the target field in the uncertainty caused by statistical fluctuations and noise in the input data values. The accuracy formula is in Eq. (3):

$$\text{Accuracy} = \left(1 - \frac{(1-R^2)\ (n-1)}{n-k-1}\right) \times 100 \quad (3)$$

where n is the total of data, k is the number of features and R^2 is from the result of coefficient of determination.

4. RESULTS AND FINDING

The performance of the Random Forest prediction model has been evaluated using the three assessment methods. This section presents the recorded and tabulated evaluation results based on the three percentage ratio splits (90:10, 80:20 and 70:30).

4.1 Mean squared logarithmic error (MSLE)

The assessment outcomes achieved by applying the MSLE approach is shown in Table 3. The 90:10 split has the largest error of 0.68 among the percentage splits. The errors produced by the 80:20 and 70:30 splits are both 0.48. The value shows the measure of the ratio between the actual and predicted values. To summarise the results that have been shown in Table 3, the chart of MSLE of the prediction is shown in Fig. 2. The chart shows that the MSLE has increased from the split 70:30 until split 90:10. A lower MSLE value indicates better model performance. In this evaluation, the value 0.48 is lower and better than the value of 0.68 as achieved by the 90:10 split.

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

Table 3. Mean squared logarithmic error results

Percentage Split	MSLE
70 : 30	0.48
80 : 20	0.48
90 : 10	0.68

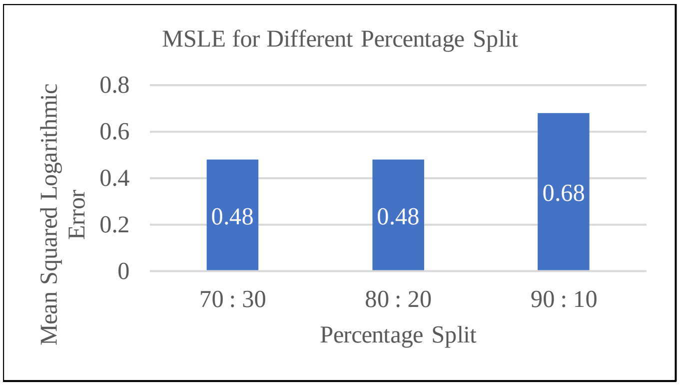

Fig. 2. Chart of MSLE for different percentage splits.

4.2 Coefficient of determination (R^2)

The results of the evaluation utilising the R^2 approach are shown in Table 4 with various percentage splits. With an R^2 of 0.73, the 90:10 split has the lowest value. On the contrary, the 80:20 split has the greatest R^2 value, 0.92. The R^2 values for the 70:30 split is 0.91. This R^2 value shows how good Random Forest is in this prediction. To summarise the results that have been shown in Table 4, the chart of R^2 of the prediction is shown in Fig. 3. Predictive models should avoid having low R^2 values. For the testing phase, the split 80:20 has been chosen based on the highest R^2 result that has been achieved.

Table 4. Coefficient of determination (R^2) results

Percentage split	Coefficient of determination (R^2)
70 : 30	0.91
80 : 20	0.92
90 : 10	0.73

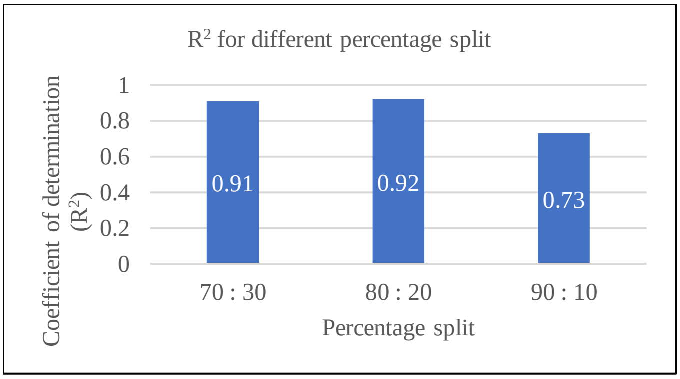

Fig. 3. Chart of Coefficient of Determination (R^2) for different percentage splits.

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

4.3 Accuracy

The assessment results achieved using the accuracy method are shown in Table 5, which also illustrates how well the model performs at different percentage splits. The 80:20 split has the highest accuracy of the percentage splits, reaching a good value of 91.47%. In contrast, the accuracy of the 90:10 split is the lowest, coming in at 69.14%. To summarise the results that have been shown in Table 5, the chart of accuracy of the prediction is shown in Fig. 4. A higher accuracy value indicates better model performance. For the testing phase, the split 80:20 has been chosen based on the best accuracy result of 91.47% that has been achieved. This high accuracy has shown that Random Forest could generate good and reliable performance in this prediction problem.

Table 5. Accuracy results

Percentage Split	Accuracy (%)
70 : 30	90.61
80 : 20	91.47
90 : 10	69.14

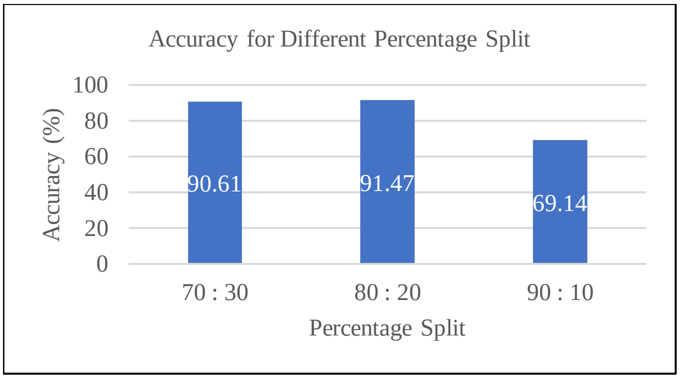

Fig. 4. Chart of accuracy for different percentage splits.

4.4 Comparison of results

Compared to previous studies, this research has produced the highest accuracy in predicting the flood damage cost which is 91.47%, from the percentage split of 80:20. In the research by [4], the distinction between damage and no damage was distinguished with an impressive average area under the curve accuracy score of 0.75. [3] produced a result in which the occurrences of high-class damage were predicted with a probability close to 90%. The research of [6] also has led to an 81% accuracy rate for classifying detrimental occurrences using the Random Forest algorithm. This shows that the accuracy result of the Random Forest prediction in this research is better and the algorithm is capable of predicting the flood damage cost reliably.

4.5 System interface

The user interface development of the research is based on the Streamlit Web App. Streamlit is a free and open-source platform which is used to create stunning machine learning and data science web apps in a quick and easy way. Streamlit is a Python-based library created especially for engineers working in machine learning. In this prototype, users are required to insert the rainfall data by month (November, December, January and February) in millimetres (mm) and river streamflow in cubic metres per second

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

(m^3/s) into the system. Then, the user has to click the predict button. The system will display the cost of flood damage in Malaysian ringgit (RM). The predicted cost would be displayed as shown in Fig. 5.

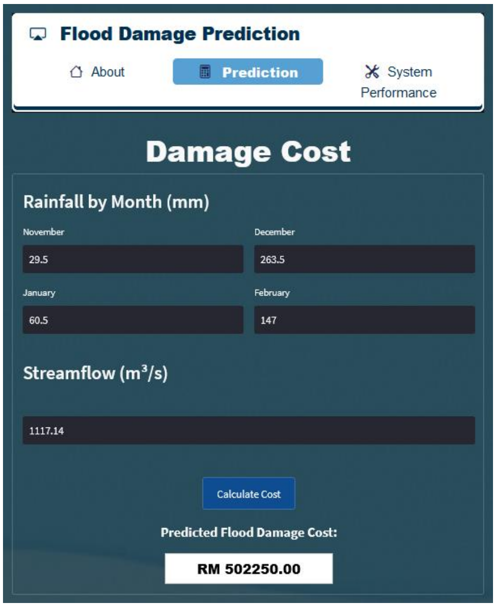

Fig. 5. Prediction Model User Interface.

4.6 Limitation

There are several limitations throughout the implementation of the research. Firstly, Terengganu is the only Malaysian state that was included in the scope of the research. The availability of data caused this restriction to occur. The detailed information on flood damage costs that is necessary for training the data and evaluating expected costs against actual data is not widely available from other Malaysian states. This data shortage means that without additional data sources, the predicting powers of the research might not be completely generalised to other states in Malaysia. The capacity of the model to precisely predict flood damage costs in various geographical and climatic conditions could be constrained by its dependence on data from a single state.

The second limitation of this research is it focuses exclusively on the damage to JPS's structures. Although estimating the costs of damage to JPS infrastructure is important, it does not account for all the structures damaged by floods, including households, farms and other public and private buildings. The difficulty in gathering and precisely predicting damage losses for other types of structures led to the choice to focus on JPS structures. Due to privacy and data access restrictions, obtaining thorough and trustworthy data for diverse buildings, particularly homes and private properties, can be much more difficult and time-consuming. Furthermore, predicting damage losses for these buildings calls for a larger, more varied dataset, which might not be easily accessible in all regions.

The third limitation of the research is related to the input variables, which are restricted to river streamflow and rainfall data by months (November, December, January, and February). Although these inputs are essential for predicting flood damage, the research might benefit from the inclusion of more pertinent variables, such as flood depth and other elements that directly affect damage. The level of damage

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

to buildings and infrastructure may be greatly impacted by the flood depth, which is a crucial measure of how severe the flooding was. Incorporating other relevant input variables, such as topography, soil type, land use, and building features, might also improve the capacity of the model for prediction.

5. CONCLUSION AND RECOMMENDATION

In conclusion, the primary aim of the research is to predict flood damage costs by implementing Random Forest algorithm. Based on the findings of the performance evaluation methods, Random Forest model delivers exceptional performance, achieving an accuracy of 91.47%, an R^2 score of 0.92, and an MSLE value of 0.48. This research makes significant contributions by accurately predicting the probable costs of flood damage to JPS's structures, thereby significantly advancing the field of flood damage assessment and research. This research achieves its objectives by implementing the Random Forest in predicting flood damage. The research successfully predicts possible flood damage costs by using the capability of the Random Forest model, and also exhibits high performance based on the outcomes of the performance evaluation. The limitations of this research are that this research only involved one state in Malaysia, namely Terengganu, focusing only on the damage to JPS's structures and using limited input variables. Therefore, the recommendations to improve this research are to include all of the states in Malaysia, broaden the scope to incorporate all kinds of flooded structures and add more input variables. The significance of the research is that the prediction could save the budget to repair JPS's structural damages by optimising the allocation of resources from the government. Furthermore, future flood damage costs can be reduced by making good decisions to improve damaged structures. In future work, hopefully this research can be further enhanced and given more focus on broader aspects of flood damage forecasting. The performance of the Random Forest algorithm would also be compared with other prediction algorithms such the Neural Network and Support Vector Machine.

6. ACKNOWLEDGEMENTS

Special recognition is directed towards Universiti Teknologi MARA Cawangan Terengganu for invaluable support towards the advancement of the research in the university. Special acknowledgement also goes to the Department of Irrigation and Drainage (JPS) Terengganu and to all individuals who were directly or indirectly involved, contributing to the completion of this research.

7. CONFLICT OF INTEREST STATEMENT

There is no conflict of interest related to this research. The research was conducted in an unbiased manner and no financial or personal affiliation could impact the results or interpretation presented in this research.

8. AUTHORS' CONTRIBUTIONS

Ainul Najwa Azahari: Conceptualisation, methodology, formal analysis, investigation and writing-original draft; **Norlina Mohd Sabri**: Supervision, review, editing and validation.

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024

9. REFERENCES

[1] D. M. García, J. R. G. Torga, M. D. Pinheiro and J. Moyano, "Simplified automatic prediction of the level of damage to similar buildings affected by river flood in a specific area," *Sustainable Cities and Society,* vol. 88, pp. 104251, 2022. Available: http://dx.doi.org/10.1016/j.scs.2022.104251

[2] Snehil and R. Goel, "Flood Damage Analysis Using Machine Learning Techniques," *Procedia Computer Science,* vol. 173, pp. 78 - 85, 2020. Available: http://dx.doi.org/10.1016/j.procs.2020.06.011

[3] M. B. Malgwi, M. Schlogl and M. Keiler, "Expert-based versus data-driven flood damage models: A comparative evaluation for data-scarce regions," *International Journal of Disaster Risk Reduction,* vol. 57, pp. 102148, 2021.

[4] E. L. Collins, G. M. Sanchez, A. Terando, C. C. Stillwell, H. Mitasova, A. Sebastian and R. K. Meentemeyer, "Predicting flood damage probability across the conterminous United States," *Environmental Research Letters,* vol. 17, no. 3, pp. 034006, 2022. Available: http://dx.doi.org/10.1016/j.ijdrr.2021.102148

[5] Z. Jiang, S. Yang, Z. Liu, Y. Xu, Y. Xiong, S. Qi, Q. Pang, J. Xu, F. Liu and T. Xu, "Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin," *Environmental Modelling & Software,* vol. 155, pp. 105436, 2022.

[6] A. Alipour, A. Ahmadalipour, P. Abbaszadeh and H. Moradkhani, "Leveraging machine learning for predicting flash flood damage in the Southeast US," *Environmental Research Letters,* vol. 15, no. 2, pp. 024011, 2020. Available: http://dx.doi.org/10.1088/1748-9326/ab6edd

[7] S. Dhanka and S. Maini, "Random Forest for Heart Disease Detection: A Classification Approach," *2021 IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES),* pp. 1 - 3, 2021.

[8] M. S. Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika and E. Aswini, "Credit Card Fraud Detection Using Random Forest Algorithm," *2019 3rd International Conference on Computing and Communications Technologies (ICCCT),* pp. 149 - 153, 2019.

[9] S. Fan and M. Fu, "Music Genre Recommendation Based on MLP & Random Forest," *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE),* pp. 331 - 334, 2022.

[10] A. Zermane, M. Z. M. Tohir, H. Zermane, M. R. Baharudin and H. M. Yusoff, "Predicting fatal fall from heights accidents using random forest classification machine learning model," *Safety Science,* vol. 159, 2022.

[11] A. T. Prihatno, H. Nurcahyanto and Y. M. Jang, "Predictive Maintenance of Relative Humidity Using Random Forest Method," *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC),* pp. 497 - 499, 2021.

[12] Y. Jiang, J. Huang, W. Luo, K. Chen, W. Yu, W. Zhang, C. Huang, J. Yang and Y. Huang, "Prediction for odor gas generation from domestic waste based on machine learning," *Waste Management,* vol. 156, pp. 264 - 271, 2022.

[13] V. K. Gupta, A. Gupta, D. Kumar and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," *Big Data Mining and Analytics,* vol. 4, no. 2, pp. 116 - 123, 2021. Available: http://dx.doi.org/10.26599/BDMA.2020.9020016

© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.24191/esteem.v20iMarch.614.g472

©Authors, 2024