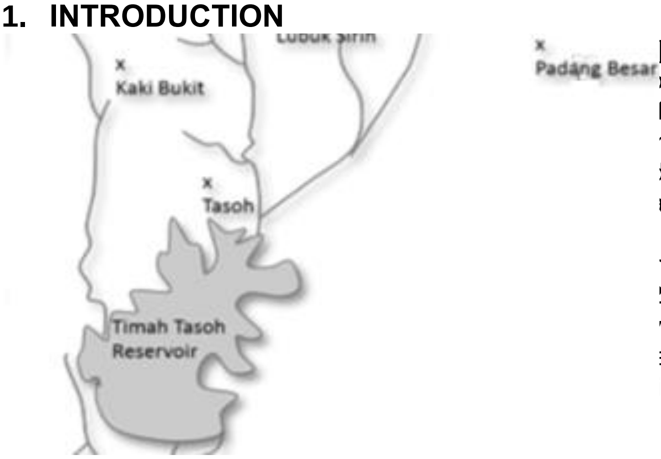
Ant Colony Optimization For Predicting Flood Based On River Water Level

Nurfaezah Mohamad Zahir, Rizauddin Saian

Abstract: Predicting flood is crucial in the South East Asia region as flood will affect life and property of the people in the region. The main objective of this study is to develop a classification model for predicting flood based on river water level. The study is conducted in Perlis, Malaysia. Perlis is a small state which is situated in the northern part of Malaysia. For the purpose of this study, data from six rainfall distribution stations in Perlis starting from year 2000 until 2014 is used. There are two classes that is used to classify the class of the river water level which are danger and normal. This study used a variant of Ant Colony Optimization algorithm called Ant-Miner to develop the classification model. The finding shows that Ant-Miner produced a better predictive model with better predictive accuracy as compared to J48.

Index Terms: Ant colony optimization, Ant-Miner, classification, flood, Perlis, reservoir, river.

1. INTRODUCTION



], as
od is
ffect
river
d. In
ised
the
n for
2010
, six
el of
reh,
in

Fig. *1*. The water of these six areas will flow into Timah Tasoh Reservoir. The reservoir will hold the water until it is no longer capable to do so. At this point, the reservoir will have to release the water, and the water will flow into the adjacent area, such as Sungai Arau. Hence, Sungai Arau area will be flooded. Flood could become a serious matter once peoples start losing their property and the life of beloved ones. One inch of flood can cause costly damage to property [3]. Hence, prediction of flood is important so that people can prepare themselves and avoid any losses that might occur if they receive early information regarding flood in their area.

2 LITERATURE REVIEW

Classification is defined as the process of building the model to describe and distinguish the data class or concept. Classification is used to identify the data whose class is unknown into an appropriate data class. Common classification methods are decision trees, neural network, genetic algorithm, and naive Bayes classification. Classification works in two steps which are learning and classification. Learning step is where classification algorithm builds the classifier by analysing the training set made by data tuples. The data will be trained and by using classification algorithm, classification rules are generated. In classification step, the data that has been tested will be used to estimate the accuracy of classification rules. Once the accuracy is considered acceptable, rules can be applied to new data tuple [4].

***Fig. 1**. Timah Tasoh reservoir.*

Rule induction is a process to train data to generate new information in the form of classification. This method is one of the basic tools for the data mining where the classification type is used of the desired output. The result or decision obtained from dataset is known as a rule that is very easy to understand and it also will give a clear and direct explanation to user. Classification rule is generally expressed with IF-THEN rule which is IF {term1 AND term2 AND …} THEN {prediction}.The rule consequent is the result of classification [5].The left-hand side of the rule is known as rule antecedent or condition consists of terms. Each term is a triple element which are

• *Nurfaezah Mohamad Zahir*
Faculty of Computer and Mathematical Sciences,Universiti Teknologi MARA, Malaysia. E-mail: znurfaezah@yahoo.com
• *Rizauddin Saian*
Faculty of Computer and Mathematical Sciences,Universiti Teknologi MARA, Malaysia. E-mail: rizauddin@uitm.edu.my

{attribute, operator, value}. It refers to a conjunction of attributes that had been tested, for example, attribute {Padang Besar = No rain}. The right-hand side is known as the predicted decision. For example, in IF {Padang Besar = No rain} THEN {Water Level = Normal}, the water level represents a decision. Parpinelli proposed Ant-Miner algorithm in 2001

$$H\left(W|A_i = V_{ij}\right) = -\sum_{w=1}^{k}\left(P\left(w|A_i = V_{ij}\right)\cdot \log_2 P\left(w|A_i = V_{ij}\right)\right) \quad (1)$$

[6]–[8], a Sequential Covering algorithm which is the based on Ant Colony Optimization algorithm, that extract rules directly from data. Sequential Covering algorithm [9] discovers rules in greedy fashion based on a certain evaluation measure. In other words, this algorithm selects terms using some heuristic. Ant-Miner uses a heuristic measure as evaluation measure to fill in the antecedent part of the rule, by selecting the best term to be included into the partial rule. The heuristic measure (2) is the normalization of entropy measures between terms (1). The algorithm selects one best rule from a set of discovered rules, based on a quality measure using some fitness function.

where

- W is the class attribute.
- a is the total number of attributes.
- x_i is set to one if the attribute A_i was not yet used by the current ant, zero otherwise.
- k is the number of classes.

$$\eta_{ij} = \frac{log_2\left(k\right) - H\left(W|A_i = V_{ij}\right)}{\sum_{i}^{a} x_i \cdot \sum_{j}^{b_i} \log_2(k) - H\left(W|A_i = V_{ij}\right)} \quad (2)$$

Ant-Miner differs from other Sequential Covering algorithms implementation because this algorithm also depends on a value called the pheromone, which contributes to the

$$P_{ij}\left(t\right) = \frac{\tau_{ij}\cdot \eta_{ij}}{\sum_{i}^{a}\sum_{j}^{b_i}\tau_{ij}\cdot \eta_{ij}}, \forall i \in I \quad (3)$$

behaviour of exploration of the algorithm. Hence, Ant-Miner uses a probability (3) that is proportional to the product of heuristic value and pheromone level for that term, to add terms to a rule. Dorigo in his book, called this transition rule, random proportional transition rule [10].

where

- τ_{ij} is the amount of pheromone at time t.
- I is the set of attributes that are not yet used by the ant.

Ant-Miner initialized equally the pheromone level, τ_{ij} using (4).
where

- a is the total number of attributes
- b_i is the number of values in the domain of attribute i.

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t)Q, \forall i, j \quad (5)$$

Ant-Miner updated the pheromone level after each ant colony has selected the best rule from a set of rules constructed by many ants in a colony using (5).

where Q is the quality of a rule.

For the next ant colony, terms in the rule antecedent that have been selected by the previous ant colony will have a higher level of pheromone and will probably are more favoured than other terms.Ant-Miner selects the best rule after each ant colony has created a set of rules, based on rule's quality. This algorithm measures rule's quality using a fitness function that depends on the product of sensitivity and specificity, which were adapted from the field of information retrieval (IR). The quality of rules, Q is calculated using (6).

where

- TP is the number of cases covered by the rule and having the same class as that being predicted by the rule.

$$Q = \frac{TP}{TP+FN} \times \frac{TN}{FP+TN} \quad (6)$$

- FP is the number of cases covered by the rule and having a different class from that being predicted by the rule.
- FN is the number of cases that are not covered by the rule while the class predicted by the rule.
- TN is the number of cases that are not covered by the rule and having a different class predicted by the rule.

In conclusion, by applying this method which is Ant-Miner technique that can help and give people a direct and clear explanation about their problems. Ant-Miner could develop a better classification model to predict the water level in Sungai Arau since it is able to uncover series of rules for classification task.

3 METHODOLOGY

odel
four
odel
by

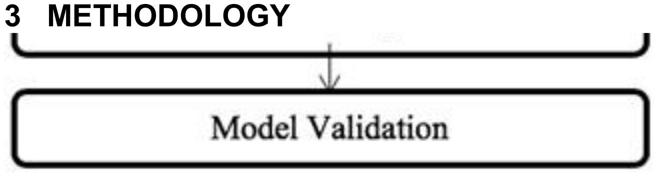


Fig. *2. Researchframework.*

Table 1 *Rainfall Data for Padang Besar Station Samples (2000-2014).*

Month	Day	2000	2001	...	2013	2014
Jan	1	0	0	...	0	0
	2	18	0	...	4.5	1
	...	...	...	...	...	...
Oct	1	0	9	...	4	0.5
	2	4	8.5	...	4	11
	3	13.5	1	...	0	24
	...	...	...	...	...	...
	31	1	55	...	16	0
Dec	1	0	7.5	...	15	0

. Data pre-processing will change the continuous attributes of the data to nominal. Model development will train the dataset using Ant-Miner. Finally, k-fold cross validation is used in model validation to determine the average predictive accuracy. The rainfall data was provided by Department of Irrigation and Drainage Malaysia, Perlis [11] from year 2000 until 2014. The rainfall distribution is based on six stations which are from Padang Besar, Tasoh, Kaki Bukit, Lubuk Sireh, Wang Kelian and Arau. **Error! Reference source not found.** shows the data from Padang Besar station from year 2000 until 2014. Data collected is in daily rainfall distribution. The dataset containing six attributes which are Padang Besar, Tasoh, Lubuk Sireh, Kaki Bukit, Wang Kelian and Arau to identify the class attribute whether the level of river water in Sungai Arau is "normal" or "danger". **Error! Reference source not found.** depicted the sample of rainfall data for Padang Besar Station from year 2000 to 2014. **Error! Reference source not found.** shows the categorization of rainfall intensity.

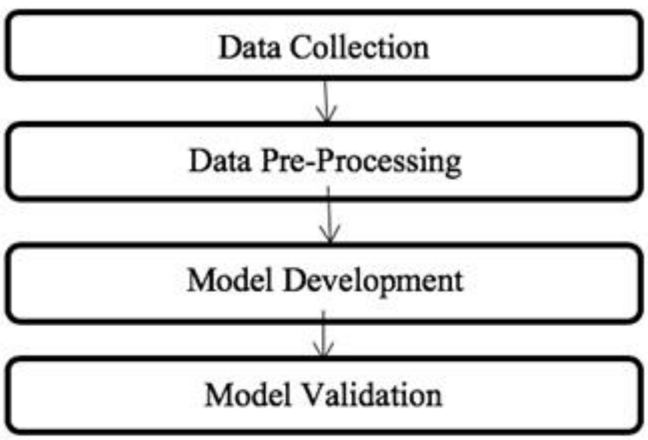


***Fig. 2**. Researchframework.*

***Table 1** Rainfall Data for Padang Besar Station Samples (2000-2014).*

Month	Day	2000	2001	...	2013	2014
Jan	1	0	0	...	0	0
	2	18	0	...	4.5	1
	...	...	...	...	...	...
Oct	1	0	9	...	4	0.5
	2	4	8.5	...	4	11
	3	13.5	1	...	0	24
	...	...	...	...	...	...
	31	1	55	...	16	0
Dec	1	0	7.5	...	15	0

***Table 2** Categorization of Rainfall Intensity.*

Rainfall (mm)	Category
0-1	No rain
1-10	light
11-30	moderate
31-60	Heavy
>60	Very heavy

(Source: Department of Irrigation and Drainage Malaysia. [11])

This study used Simple K-means for data clustering provided by Waikato Environment for Knowledge Analysis version 3.6 [12] as shown in

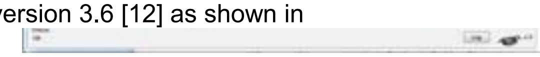


Fig. 3. The number of clusters is set to two: "normal" and "danger".

***Fig. 3.** WEKA settings for K-Means.*

4 RESULTS

This study trained the data using 10, 20, 30, 40, 50, 60 and 70 number of ants and the average predictive accuracy is depicted in **Error! Reference source not found.**. The followings are the values for the other parameters: minimum case per value is 5, maximum uncovered cases is 10, rule for convergence is 10, and number of iterations is 100. It is found that the highest average predictive accuracy is 96.15 %. The average predictive accuracy is better than J48 that is only 57.78 %. J48 is a class for generating a C4.5 [13] decision tree in WEKA [12]. As an addition, the number of rules and the number of terms is small, 11 and 10, respectively.

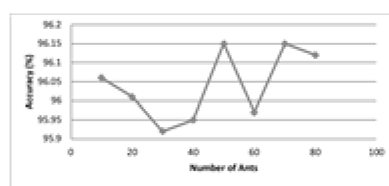


***Fig. 4.** Average predictive accuracy for various number of ants.*

5 CONCLUSIONS

The study has developed a classification model for river water level in Perlis. The findings show that the predictive accuracy of the classification prediction model is promising at the value of more than 90 percent. As an addition, the predictive accuracy of the classification model is better as compared to J48. The classification model established in this study can be used by the state and people of Perlis to prepare themselves in the event of flood. Hence, it can reduce the effect of flood disaster. This study recommends future research to obtain better percentage of accuracy in flood prediction. Since the class for rules constructed by Ant-Miner will only be determined after the creation of each rule, the selected terms are not focused relevance. As an addition, Ant-Miner might face the problem of stagnation, where optimized rule cannot be found, and thus make the program to run forever. So, it would be interesting to improve the terms selection strategy to reduce the stagnation problem.

REFERENCES

[1] S. Lavender et al., ‘Application of Earth Observation To A Ugandan Drought and Flood Mitigation Service’, in Proceedings of the 2017 Conference on Big Data from Space. BiDS’2017, 2017, pp. 469–472.
[2] M. Chini, R. Pelich, R. Hostache, P. Matgen, J. M. Delgado, and G. Sabatino, ‘Large Scale Flood Recurrence Map Using Sar Data’, in Proceedings of the 2017 Conference on Big Data from Space. BiDS’2017, 2017, pp. 67–70.
[3] A. Kaschube, ‘Drowning in Problems: Floods Threaten Properties and Owners’ Security’, J. Prop. Manag., vol. 71, no. 2, pp. 50–52, 2006.
[4] J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques. Elsevier, 2012.
[5] P. Jin, Y. Zhu, K. Hu, and S. Li, ‘Classification Rule Mining Based on Ant Colony Optimization Algorithm’, in Intelligent Control and Automation, Springer, 2006, pp. 654–663.
[6] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, ‘An Ant Colony Based System for Data Mining: Applications to Medical Data’, in Proceedings of the genetic and evolutionary computation conference (GECCO-2001), 2001, pp. 791–797.
[7] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, ‘An Ant Colony Algorithm for Classification Rule Discovery’, in Data Mining: A Heuristic Approach, 2002, pp. 191–208.
[8] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, ‘Data mining with An Ant Colony Optimization Algorithm’, IEEE Trans. Evol. Comput., vol. 6, no. 4, pp. 321–332, 2002.
[9] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Pearson International Edition, 2006.
[10] M. Dorigo and T. Stützle, Ant Colony Optimization. Scituate, MA, USA: Bradford Company, 2004.
[11] ‘Department of Irrigation and Drainage Ministry of Water, Land and Natural Resources’, 2014.
[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, ‘The WEKA Data Mining Software: An Update’, SIGKDD Explor Newsl, vol. 11, no. 1, pp. 10–18, Nov. 2009.
[13] R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers, 1993.