International Journal on Advanced Science Engineering Information Technology

Vol.7 (2017) No. 3
ISSN: 2088-5334

The Application of Apriori Algorithm in Predicting Flood Areas

Nur Ashikin Harun[#1], Mokhairi Makhtar[#2], Azwa Abd Aziz[#3], Zahrahtul Amani Zakaria[#4], Fadzli Syed Abdullah[#5], Julaily Aida Jusoh[#6]

[#]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Tembila, Besut, Terengganu, Malaysia*
E-mail: [1]shikin_harun91@yahoo.com, [2]mokhairi@unisza.edu.my, [3]azwaaziz@unisza.edu.my, [4]zahrahtulamani@unisza.edu.my, [5]fadzlihasan@unisza.edu.my, [6] julaily@unisza.edu.my

***Abstract*— The changing of physical characteristics of the hydrological system have caused a lot of natural phenomenon, which leads to flooding as one of the major problems that cause economic damages and affect people's life. Therefore, the need for a systematic and comprehensive approach to flood area prediction is needed. This research proposed a flood area prediction model with the application of Apriori algorithm towards hydrological data sets. Department of Irrigation and Drainage Malaysia supply the data sets and flood report from year 2009 to 2015 (November until January) which consist of 7 district. The research begins with the data selection, pre-process the data, and data transformation, then the cleaned data will be tested with the Apriori algorithm. The rules will be evaluating using support, confidence and lift value to rank either it is best rules or not. The results show that each district generates best and crucial rules which consist the association of the villages and water level. Thus, hopefully the result can be use in flood management and can give early an early warning to the villagers at flood risk area.**

***Keywords*— data mining; Apriori algorithm; association rule; flood disaster; knowledge discovery in database**

I. INTRODUCTION

Floods have been befalling throughout the Earth history and are estimated to endlessly occur as the water cycle continues to run. According to the United Nations Office for Disaster Risk Reduction (UNISDR), flooding is the biggest natural disaster around the world for the year 1980 until 2011 with mark 3455 events all around the world. Besides, flood is the most devastating disaster and cause a lot of damages and trauma towards sufferers [1], [2]. The factors that cause flood at a certain area are geographical condition, metrological condition, planning problem, hydrological condition and environmental status due to human activities [3]. Due to the damages and loss, the effective management of flood risk is a spark issue all around the globe to have better management and prevention [4], [5].

In Malaysia, there are always flood occurrence event, especially during monsoon season. In Malaysia, there are distinct dry and rainy seasons with rainfall annually 3000 m, and an average of humidity is 80%. There is a total of 189 river basins in Malaysia including Sabah and Sarawak, which the main channels are flowing to the South China Sea and 85 of them are disposed to erratic flood [6]. For this research, Terengganu which is one of the low-lying area [7] were chosen as the research area. Terengganu experienced Northeast monsoon season, which exposed to heavy rainfall every early November and ends in March. Due to this phenomenon, floods always occur which sometimes unpredictable.

Therefore, this research applies data mining toward hydrological data to obtain new knowledge. Data mining is a step that applies data analysis and algorithms to produce a particular enumeration of patterns or model over the data [8]. Data mining has various methods and techniques that could be applied in many fields of research as a problem solver and also important to decision makers. Association rules are one of the major techniques of data mining. In [9] discovered the association rules when they need to do research on sales pattern on a large database. For instance, association rules are mining large datasets to find frequent item sets by considering the minimum support and minimum confidence [10]. Apriori Algorithm is the algorithm to mine frequent item sets that satisfied support and confidence by generating candidates in the process of joining and pruning [11]. This research uses Apriori Algorithm compare to other algorithms because it reduces the number of scans in the database to extract frequent item sets. Therefore, it maximizes the computational workload [12]. The Apriori algorithm had been applied in many fields such as medical [13], [14], education [15], weather forecasting [16] and disaster management [17], [18].

This research aims to identify the association between water level and flood area during the monsoon season. Besides, a model was developed in this study by implementing association rule mining to predict the flood area in Terengganu. In flood forecasting, past research focuses more on the application of remote sensing [19]. Thus, this research applies Apriori Algorithm to predict flood area based on attributes and instances. Despite the vast number of studies available in the literature, the current study that uses data mining approaches can contribute to flood area prediction using hydrological data with cost effectiveness, reliable results and help in flood management in research area [20].

II. MATERIAL AND METHOD

This section explains the study area, the process of the research and also the data used and the algorithm that had been applied.

A. Study Area

The research focuses on Terengganu that consists of 7 districts. The districts are Marang, Dungun, Setiu, Kemaman, Besut, Kuala Terengganu and Hulu Terengganu. Terengganu is located in the North East Malaysia with the latitude of 3^{o} 53'U-5^{o} 50'U and longitude of 102^{o} 23'T-103^{o} 30'T. Fig. 1 shows the study area. Each district will generate best rules which show the association of village during flood happens. Thus, using the rules will we see the correlation of the rules with the water level at the main river and selected stations.

B. Research Workflow

In order to accomplish the research objectives, a process flow is designed with the research approach.

According to Fig. 2, the research starts with the collecting flood data from Department of Irrigation and Drainage Malaysia, Department Irrigation and Drainage Terengganu and Terengganu Flood Portal. The data is selected to create the desired datasets. Then, this dataset will be integrated to develop a dataset to be mined using the algorithm. The research finds that the output of phase one is the flood datasets, which is the collation of some (or all) data from the aforementioned data resources. In fact, in phase two, the association rule algorithm will be run using the flood datasets to generate association rule of flood area. Finally, the rules will be used at phase three to create a model of flood area prediction.

Moreover, this study follows the Knowledge Discovery in Database process flow as to apply the data mining in the research. The Knowledge Discovery Database has five phases which are the selection of data, data pre-processing, data transformation, data mining and lastly the interpretation of the data into desire reports. The phase could be loop or iterative to produce the output to be used in the next phase. Fig. 3 illustrates the phases of the KDD model.

Fig. 1 Research area

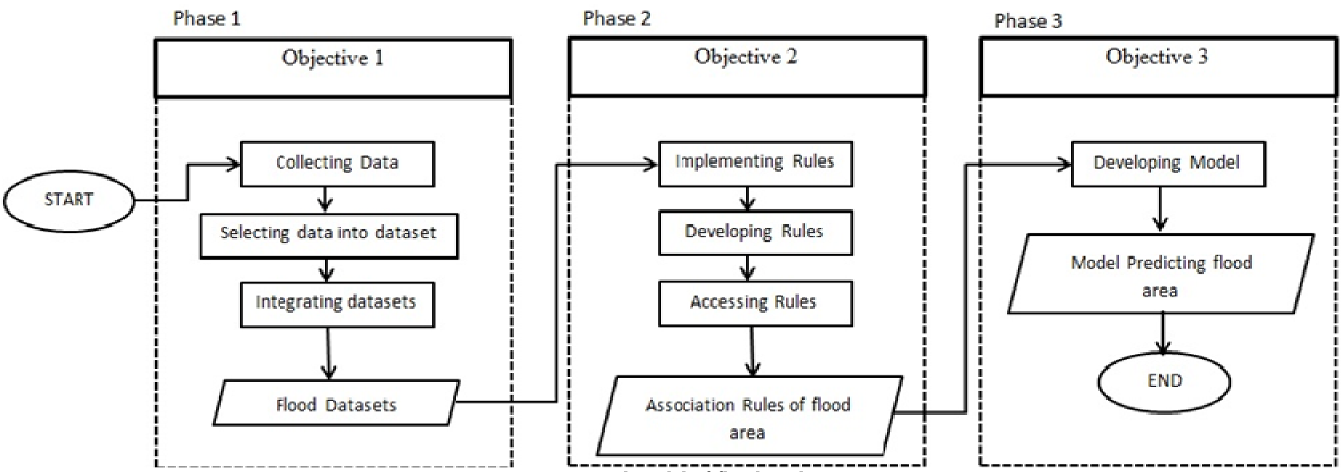

Fig. 2 Proposed model of flood prediction

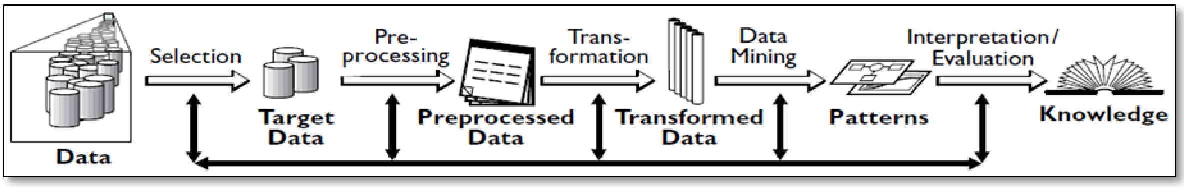

Fig. 3 KDD process

This research used secondary data which come from reports, newspaper and also data storage. The raw data for this research mainly come from the Malaysian Irrigation and Drainage Department. In order to extract the data, the selection of data set and subset requires an understanding of the domain. In this research, the main emphasis is on finding a correlation between river flow and flood area and developing a model to predict flood area by implementing association rule. At first, the data selection [22] phase is the initial phase to create a target data. At pre-processing phase, the target data will be cleaned from the missing values, outliers, inconsistent and much more. Then, at data transformation, the data will be transforming into the format of destination data. While this research uses WEKA as the aided tools, the data are in ".csv" format to make sure the WEKA can process and load it. In data mining phase, the association rules are applied to the data sets. Basically, the association rules apply if/then statements to discover the relationship between unrelated data in the data repository. The basic formula of association rules is

$$A \Rightarrow B \tag{1}$$

A will be the 'if statement' and B will be the 'then statement'. In association rules, A is the antecedent and B is the consequent. Support and Confidence are calculated to analyze which one is the best rules.

$$\text{Support} = P(A \cap B) = \frac{\text{number of transactions containing both } A \text{ and } B}{\text{total number of transactions}} \tag{2}$$

The support for a particular association rule A => B is the proportion of transactions in D that contain both A and B. The rule indicates how frequently the items in the rule occurred together.

$$\text{Confidence} = P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\text{number of transactions containing both } A \text{ and } B}{\text{number of transactions containing } A} \tag{3}$$

The confidence of the association rule A => B is a measure of the accuracy of the rule. It is determined by the percentage of the transaction containing A that also contains B. It is the conditional probability of the consequent given the antecedent.

On the other hand, to find the best rules, we will consider rules that have high support or high confidence and usually both. Rules that categorized as strong rules or best rules are rules which surpass certain minimum support (called minsup) and minimum confidence (called minconf) that we specified earlier.

For this research, we used Apriori algorithm to test on the datasets collected. Association rules have lots of other algorithms, but the Apriori is the most suitable algorithm for this research because of the particular data set and also the rules extraction solve the research problems. The advantage of this algorithm is it will shrink the search space in term of "if an item set Z is not frequent then for any item A, Z U A will not be frequent".

In obtaining the rules, there are two major steps applied [21] which are:

- Find all sets of items that have support value greater than the minimum support. These item sets are called large item sets. All others are called small item sets.
- Use the large item sets to generate the desired rules. It begins with finding all non-empty subsets of every large item set L. For every such subset A generate a rule of the form A => (L-A) if the ratio of the support (A) is at least minconf. All subsets of L must be considered to generate rules with multiple consequences.

Additionally, we set up the rank of rules that will be generated according to lift metric. Lift use to measure the quality and interestingness of the rule. In this research lift is the third measure to evaluate the quality of the rule after the support and confidence. Lift is defined as

$$\text{Lift } (A \Rightarrow B) = \frac{support(A \cup B)}{(Support\ (Antecedent) \times Support\ (Consequence))} \tag{4}$$

If some rule had a lift of 1, it showed that the probability of the relationship between antecedent and consequent are independent of each other. When two events independent, hence no rule can be haggard involving the events. Besides, if the lift is bigger than 1 value, the two occurrences are dependent to one another. Likewise, the rules are potentially useful for predicting consequences in future data sets.

Lastly, in data interpretation phase, the patterns or rules or information generated from the phases before are then will be interpreted. The results are stored in .arff format. Then, based on the best rules obtained, a model is developed to visualize the extracted patterns.

III. Results and Discussion

Table 1 shows the data used for the experiments. The data loaded in CSV format and should normalize into nominal data. The data used are the village name (attributes) and flood status (instances) either it is Yes or No. The association rules are generated which we can predict the association with one village to another. Therefore, with the association rules and water level repositories, we make an analysis to investigate the association between the water level and flood area. The results for each district are given in Table 2 to 8.

TABLE I
DATA USED FOR EXPERIMENTING

District	No of Attributes	No of Instances
Setiu	36	26
Marang	27	9
Kemaman	38	34
Besut	50	40
Dungun	47	12
Hulu Terengganu	49	38
Kuala Terengganu	44	29

The results (Table 2 to Table 8) indicate that the best rules generated for each district and the association with the water level. Table 2 showed the analysis of Marang results. The association of the village denoted that, for example, "KG. TEMALA = Y ==> KG. KUBU = Y". If Kg. Temala having a flood, then Kg. Kubu will have a flood with 100% confidence and 40% support values. The dependence of the rules is 2.25 which they depend on each other. Whereas during the flood happen, the water level at the nearest river is in danger state. We can conclude that, if the water level tends to be a danger, the villager at Kg. Temala should be extra cautious because there are risks in facing flood. Kg. Kubu should know that after if Kg. Temala in risk, they also face the same situation.

TABLE II
ANALYSIS OF MARANG RESULT

Association Rules	Support	Confidence	Lift	River	Station	Water Level
KG. TEMALA = Y ==> KG. KUBU = Y	0.4	1	2.25	Sg. Marang	Jambatan Pengkalan Berangan	13.95 (Danger)
KG. TEMALA = Y ==> KG. ALOR WAN SYED = Y KG. KUBU = Y	0.4	1	2.25	Sg. Marang	Jambatan Pengkalan Berangan	14.12 (Danger)
KG. ALOR WAN SYED = Y KG. KUBU = Y ==> KG. TEMALA = Y	0.4	1	2.25	Sg. Marang	Jambatan Pengkalan Berangan	14.14 (Danger)
KG. ALOR WAN SYED = Y ==> KG. TEMALA = Y	0.4	0.67	1.5	Sg. Marang	Jambatan Pengkalan Berangan	3.08 (Danger)
KG. ALOR WAN SYED = Y ==> KG. KUBU = Y	0.4	0.67	1.5	Sg. Marang	Jambatan Pengkalan Berangan	3.08 (Danger)
KG. TEMALA = Y KG. KUBU = Y ==> KG. ALOR WAN SYED = Y	0.4	1	1.5	Sg. Marang	Jambatan Pengkalan Berangan	3.10 (Danger)
KG. PASIR PUTIH = Y 7 ==> KG. JENANG = Y	0.6	1	1.29	Sg. Marang	Jambatan Pengkalan Berangan	2.1 (Alert)

TABLE III
ANALYSIS OF BESUT RESULT

Association Rules	Support	Confidence	Lift	River	Station	Water Level
KG. PASIR AKAR = Y ==> KG. TENANG = Y KG. PADANG BUAL = Y	1.7	1	2.35	Sg. Besut	Kg. La	21.74 (Warning)
KG. PASIR AKAR = Y 17 ==> KG. TENANG = Y KG. KAYU KELAT = Y 17	1.7	1	2.35	Sg. Besut	Jambatan Jerteh	8.33 (Warning)
KG. PASIR AKAR = Y ==> KG. TENANG = Y KG. BUKIT MALI = Y KG. PADANG BUAL = Y	1.7	1	2.35	Sg. Besut	Kg. La	21.34 (Warning)
KG. TENANG = Y KG. PADANG BUAL = Y ==> KG. BUKIT MALI = Y KG. PASIR AKAR = Y	1.7	1	2.35	Sg. Besut	Kg. La	20.35 (Alert)
KG. BUKIT MALI = Y KG. PASIR AKAR = Y ==> KG. TENANG = Y KG. KAYU KELAT = Y	1.7	1	2.35	Sg. Besut	Jambatan Jerteh	9.61 (Danger)
KG. BUKIT MALI = Y ==> KG. KAYU KELAT = Y	1.7	1	2.35	Sg. Besut	Jambatan Jerteh	8.12 (Warning)
KG. TENANG = Y KG. BUKIT MALI = Y ==> KG. PADANG BUAL = Y	1.90	0.89	2.11	Sg. Besut	Kg. La	19.14 (Normal)

TABLE IV
ANALYSIS OF KEMAMAN RESULT

Association Rules	Support	Confidence	Lift	River	Station	Water Level
KG. DADONG = Y ==> KG. TELADAS = Y	0.3	1	1.79	Sungai Tebak	Jambatan Tebak	19.879 (Danger)
KG. AIR PUTIH = Y ==> KG. TEBAK = Y	0.25	1	4.25	Sungai Tebak	Jambatan Tebak	19.283 (Danger)
KG. BATU 14 = Y KG. TELADAS = Y ==> KG. AIR PUTIH = Y	0.25	1	4.25	Sungai Tebak	Jambatan Tebak	18.475 (Warning)
KG. BATU 14 = Y KG. TELADAS = Y ==> KG. TEBAK = Y	0.25	1	4.25	Sungai Tebak	Jambatan Tebak	17.522 (Alert)
KG. BATU 14 = Y KG. BUKIT MENTOK = Y ==> KG. AIR PUTIH = Y	0.25	1	4.25	Sungai Tebak	Jambatan Tebak	16.671 (Normal)
KG. BATU 14 = Y KG. BUKIT MENTOK = Y ==> KG. TEBAK = Y	0.25	1	4.25	Sungai Tebak	Jambatan Tebak	18.247 (Warning)
KG. TEBAK = Y ==> KG. BATU 14 = Y KG. BUKIT MENTOK = Y	0.25	1	4.25	Sungai Tebak	Jambatan Tebak	16.516 (Normal)

TABLE V
ANALYSIS OF SETIU RESULT

Association Rules	Support	Confidence	Lift	River	Station	Water Level
CHALOK KEDAI = Y ==> KG. LANGKAP = Y	0.3	1	2.89	Sg. Setiu	Jambatan Permaisuri	9.84 (Danger)
CHALOK KEDAI = Y ==> KG. BESUT = Y KG. LANGKAP = Y	0.3	1	2.89	Sg. Setiu	Jambatan Permaisuri	9.16 (Danger)
MERBAU MENYUSUP = Y ==> PENGKALAN MERBAU = Y	0.25	1	4.33	Sg. Setiu	Jambatan Permaisuri	8.48 (Warning)
MERBAU MENYUSUP = Y 6 ==> KG. BESUT = Y PENGKALAN MERBAU = Y 6	0.25	1	4.33	Sg. Setiu	Jambatan Permaisuri	8.29 (Warning)
MERBAU MENYUSUP = Y 6 ==> CHALOK KEDAI = Y PENGKALAN MERBAU = Y 6	0.25	1	4.33	Sg. Setiu	Jambatan Permaisuri	7.86 (Alert)

TABLE VI
ANALYSIS OF KUALA TERENGGANU RESULT

Association Rules	Support	Confidence	Lift	River	Station	Water Level
KG. TANJUNG DAMAI = Y ==> TAMAN PUSU TIGA = Y	0.3	1	2.23	Sg. Nerus	Kg. Bukit	12.841 (Danger)
KG. BANGGOL PERADONG = Y ==> KG. BARU TETAMBAH = Y	0.3	1	4.14	Sg. Nerus	Kg. Bukit	12.762 (Danger)
TAMAN PUSU TIGA = Y KG. JERAM = Y ==> KG. BANGGOL PERADONG = Y	0.3	1	4.14	Sg. Nerus	Kg. Bukit	14.299 (Danger)
TAMAN PUSU TIGA = Y KG. JERAM = Y ==> KG. BARU TETAMBAH = Y 7	0.3	1	4.14	Sg. Nerus	Kg. Bukit	13.174 (Danger)
KG. BANGGOL PERADONG = Y ==> TAMAN PUSU TIGA = Y KG. BARU TETAMBAH = Y	0.3	1	4.14	Sg. Nerus	Kg. Bukit	12.928 (Danger)
KG. BARU TETAMBAH = Y ==> TAMAN PUSU TIGA = Y KG. BANGGOL PERADONG = Y	0.3	1	4.14	Sg. Nerus	Kg. Bukit	12.735 (Danger)

TABLE VII
ANALYSIS OF HULU TERENGGANU RESULT

Association Rules	Support	Confidence	Lift	River	Station	Water Level
KG. KEPAH=Y ==> KG. MATANG=Y KG. MENERONG=Y	0.3	1	3.45	Sg. Terengganu	Kg. Tanggol	6.83 (Normal)
KG. KEPAH=Y KG. PENGKALANG AJAL=Y ==> KG. MATANG=Y KG. MENERONG=Y	0.3	1	3.45	Sg. Berang	Kg. Menerong	25.894 (Danger)
KG. PAYA BESAR=Y KG. MENERONG=Y ==> KG. MENJING=Y KG. BATU 23=Y	0.3	1	3.17	Sg. Berang	Kg. Menerong	25.309 (Danger)
KG. MENJING=Y KG. BATU 23=Y ==> KG. PAYA BESAR=Y KG. MENERONG=Y	0.3	1	3.17	Sg. Berang	Kg. Menerong	23.079 (Warning)
KG. PAYA BESAR=Y ==> KG. BATU 23=Y	0.4	0.84	1.88	Sg. Terengganu	Kg. Tanggol	7.44 (Normal)
KG. CHETING=Y ==> KG. MENERONG=Y	0.4	0.79	1.58	Sg. Berang	Kg. Menerong	25.498 (Danger)
KG. CHETING=Y ==> KG. PENGKALANG AJAL=Y	0.4	0.79	1.36	Sg. Berang	Kg. Menerong	25.547 (Danger)
KG. PENGKALANG AJAL=Y ==> KG. MENERONG=Y	0.4	0.82	1.64	Sg. Berang	Kg. Menerong	24.055 (Warning)

TABLE VIII
ANALYSIS OF DUNGUN RESULT

Association Rules	Support	Confidence	Lift	River	Station	Water Level
KUALA JENGAI = Y ==> JERANGAU = Y 5	0.45	1	1.5	Sg. Dungun	Jambatan Jerangau	11.92 (Warning)
KUALA JENGAI = Y ==> PASIR RAJA = Y JERANGAU=Y	0.45	1	1.5	Sg. Dungun	Jambatan Jerangau	13.37 (Danger)
JERANGAU = Y 8 ==> PASIR RAJA = Y 8	0.45	1	1.5	Sg. Dungun	Jambatan Jerangau	13.85 (Danger)
KUALA JENGAI = Y 5 ==> PASIR RAJA = Y 5	0.45	1	1.33	Sg. Dungun	Jambatan Jerangau	13.11 (Danger)

Table 3 specifies that when the water level at alert and warning, the village at flood risk area should be ready. The association of the village shows the prediction of the risky area if a village tends to flood. The water level in alert state at Sg. Besut makes this village “KG. TENANG = Y KG. PADANG BUAL = Y ==> KG. BUKIT MALI = Y KG. PASIR AKAR = Y” were having a flood with 100% confidence, 17% support, and 2.35 lift value. Table 4 illustrates the result of Kemaman District. The association rules of Kemaman have high dependency among rules which is 4.25. Therefore, the rules are strong and if a village has flooded, the probability of the associated village tend to flood is high. The confidence values mostly are 100%.

Setiu district’s result as in Table 5, If Jambatan Permaisuri in Danger level, the village nearby such as Kg. Chalok Kedai, Kg. Besut will easily have flooded. Moreover, at the alert level, this village “MERBAU MENYUSUP = Y 6 ==> CHALOK KEDAI = Y PENGKALAN MERBAU = Y 6” is already having flood due to the rise of river water and affected their area. As in Table 6, Kuala Terengganu will be having a flood when the water level is in danger level. Most of the rules have 30% support values and 100% confident, while lift value is 4.14. The main river is Sg. Nerus and the danger level will make the area has flooded.

Table 7 shows the result of Hulu Terengganu district. There are two stations in the experiment because the villages were situated near them tend to flood. The station has a different range of water level. Therefore, we cannot simply associate the rules with water level easily without considering the topographic factors. Table 8 shows the analysis [23] of Dungun district. Sungai Dungun is the main river which at danger level, it causes a flood to a nearby village. Moreover, best rules have 45% support, 100% confident and 1.5 lift value.

IV. Conclusion

The Apriori algorithm normally applies in business transactions, therefore, this research experiment the algorithm using hydrological data sets. The results by implementing the Apriori algorithm produced best rules and created the association of flood area. On the other hand, we can use the rules and create a model to help in flood management. Optimistically, this research can extend to a bigger case study and help in flood management which it is one of biggest catastrophe in Malaysia. Hopefully, the resulting model can help in flood management, especially by giving early warning to residents in flood potential areas in addition to saving lives and property.

Acknowledgment

The presented work has been funded by the Ministry of Higher Education Malaysia under the Research Acculturation Grant Scheme (RAGS) reference code RAGS/1/2014/ICT07/UniSZA/1. The authors would like to thank NSC, MetMalaysia and DoID for supplying the data of flood in Terengganu and to all those who participated in this research.

References

[1] Y. Jiao, S. Wang, Y. Zhou, and L. Wang, "Uncertainty analysis of flood disaster assessment using radar imagery," in *Proc. IEEE IGRSS'07*, 2007, p. 4729.

[2] H. Yahya, M. Z. M. Rodzi, and M. N. Ahmad, "Understanding the knowledge transfer process in the flood management domain," in *Proc. IEEE ICCIS'14*, 2014, p. 1.

[3] I. Priyadarshinee, K. Sahoo, and S. Mallick, "A model for flood prediction and prevention using wireless sensor network," in *Proc. ICETCC'16*, 2016, p. 22.

[4] K. Jenkins, S. Surminski, J. Hall, and F. Crick, "Assessing surface water flood risk and management strategies under future climate change: An agent-based model approach environment," in *Proc. GRICCE'16*, 2016, p. 1.

[5] M. S. Khalid and S. Shafiai, "Flood disaster management in Malaysia: An evaluation of the effectiveness flood delivery system," *International Journal of Social Science and Humanity*, vol. 5, pp. 398-402, Apr. 2015.

[6] S. G. Diya, M. B. Gasim, M. E. Toriman, and M. G. Abdullahi, "Flood in Malaysia historical reviews, causes, effects and mitigations approach," *International Journal of Interdisciplinary Research and Innovations*, vol. 2, pp. 59-65, 2014.

[7] M. B. Gasim, J. H. Adam, M. E. Toriman, S. Abd Rahim, and H. Juahir, "Coastal flood phenomenon in Terengganu, Malaysia: Special reference to Dungun," *Research Journal of Environmental Scienc*es, vol. 1, pp. 102-109, 2007.

[8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, pp. 37-54, Mar. 1996.

[9] R. Agrawal, T. Imieliński, and A. Swami, "Mining association in large databases," in *Proc. ACM SIGMOD ICMD'93*, 1993, p. 207.

[10] W. Novitasari, A. Hermawan, Z. Abdullah, R. W. Sembiring, and T. Herawan, "A method of discovering interesting association rules from student admission dataset," *International Journal of Software Engineering and Its Applications*, vol. 9, pp. 51-66, Aug. 2015.

[11] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. ICVLD'94*, 1994, p. 487.

[12] P. Tanna and Y. Ghodasara, "Using Apriori with WEKA for frequent pattern mining," *International Journal of Engineering Trends and Technology*, vol. 12, pp. 127-131, Jun. 2014.

[13] Q. Fan, C. J. Zhu, J. Y. Xiao, B. H. Wang, L. Yin, X.L. Xu, and F. Rong, "An application of Apriori algorithm in SEER breast cancer data," in *Proc. IEEE ICAICI'10*, 2010, p. 114.

[14] K. P. Merry, R. K. Singh, and S. S. Kumar, "Apriori-hybrid algorithm as a tool for colon cancer microarray data classification," *International Journal of Engineering Research and Development*, vol. 4, pp. 53-57, 2012.

[15] Z. Abdullah, T. Herawan, A. Noraziah, and M. M. Deris, "Mining least association rules of degree level programs selected by students," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, pp. 241-253, 2014.

[16] T. R. Baitharu and S. K. Pani, "A comparative study on associaton rule mining algorithms using weather dataset," *International Journal of Advanced Technology in Engineering and Science*, vol. 3, pp. 291-295, Aug. 2015.

[17] L. Hu, G. Zhuo, and Y. Qiu, "Application of apriori algorithm to the data mining of the wildfire," in *Proc. IEEE ICFSKD'09*, 2009, p. 426.

[18] P. Cortez and A. Morais, "A data mining approach to predict forest fires using meteorological data," in *Proc. NTAI-EPIA'07*, 2007, p. 512.

[19] G. Boni, L. Ferraris, L. Pulvirenti, G. Squicciarino, N. Pierdicca, L. Candela, A. R. Pisani, S. Zoffoli, R. Onori, C. Proietti, and P. Pagliara, "A prototype system for flood monitoring based on flood forecast combined with COSMO-SkyMed and Sentinel-1 data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, pp. 27941-2805, 2016.

[20] M. Cisty and J. Bezak, "The application of data mining methods for short time flows prediction in flood warning systems," in *Proc. RACMHE'13*, 2013, p. 92.

[21] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo, "Finding interesting rules from large sets of discovered association rules," in *Proc. ACM ICIKM'94*, 1994, p. 401.

[22] I. M. Yassin, A. Zabidi, M. S. A. M. Ali, N. M. Tahir, H. A. Hassan, H. Z. Abidin, and Z. I. Rizman, “Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 630-637, Oct. 2016.

[23] M. N. M. Nor, R. Jailani, N. M. Tahir, I. M. Yassin, Z. I. Rizman, and R. Hidayat, “EMG signals analysis of BF and RF muscles in autism spectrum disorder (ASD) during walking,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 793-798, Oct. 2016.