

A CONCEPTUAL FRAMEWORK FOR PREDICTING FLOOD AREA IN TERENGGANU DURING MONSOON SEASON USING ASSOCIATION RULES

AZWA ABDUL AZIZ, NUR ASHIKIN HARUN, MOKHAIRI MAKHTAR, FADZLI SYED ABDULLAH, JULAILY AIDA JUSOH, ZAHRAHTUL AMANI ZAKARIA

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Terengganu, Malaysia

E-mail: ¹azwaaziz@unisza.edu.my, shikin_harun91@yahoo.com, mokhairi@unisza.edu.my, fadzlihasan@unisza.edu.my, julaily@unisza.edu.my, zahrahtulamani@unisza.edu.my

ABSTRACT

Over the last decade, flooding has been one of catastrophic disaster which causes economic damage, lost life and environmental deprivations. The need for flood prediction is rising since the decision maker lacks intelligent tools to predict flood areas. The data mining and geospatial visualization will test the temporal data which include river flow and rainfall data to find patterns and new information which then will use to predict areas that expose to flood. Since Terengganu situated at east peninsular, every year in October to March will be having heavy rainfall and increasing of sea levels. To this extent, we propose a framework to support flood prediction which the outcome will intention to better manage floods through prevention, protection and emergency response.

Keywords: *Data Mining, Association Rule, Geographical Information System(GIS), Flood Area Prediction, Hydrology*

1. INTRODUCTION

Data Mining is a knowledge discovery which extracting large data into useful information by analysing data from different perspective and summarizing it. The increasing of computing power, massive data collection and storage and along with data mining algorithms will help to find correlations or patterns. Nowadays, we have infinite data but lacking in useful information. IBM stated that every day we create 2.5 quintillion bytes of data which 90% of the data in the world has been created in the last two years. So we need data mining to excerpt into new knowledge. On the other hand, due to the advance of the technology, the interest in Data Mining (DM) which part of Knowledge Discovery in Database upswing swiftly and help to solve many problems in business science and also disaster management [1]. The raw data contain pattern and trends and other valuable information which DM will help to extract high-level information for the decision-maker. The goals of discovery are prediction and description of data. The prediction will predict the future behaviour of some entities and Description will present the pattern to a human-understandable form [2].

Data mining in predicting the natural disaster is very useful and being used by many scientists in this day and age [1, 3, 4, 12, 13]. As a matter of fact, we are at the whim of nature. The human need is to consume a large volume of disaster situational information. It is very time consuming when a quick data needed, but still we need to collect information by reading and assimilation situational information where there are high probability of exposure to redundant and repetitive information. Aforementioned, the information should synthesize from heterogeneous sources, be tailored to specific contexts or task at hand, and be summarized for effective delivery, and be immediately useful for making decision [3]. In a recent publication of the International Journal of Emergency Management, experts such as Adam Zagorecki of the Centre for Simulation and Analytics at Cranfield University UK point to the increasing abilities of data mining in the wake of natural disasters. The unstructured and structured sources will undergo the data mining process to gain new information. This source will lead a creation of a model so that those involved can more effectively deploy strategies of mitigating damage.

The changing of physical characteristics of the hydrological system caused a lot of natural phenomenon, whereas flood is one of the major problems which cause economic damages and affect people's life. Calamitous flooding is a major security concern in Malaysia. Flood generally develops over a period of days, when too much rainwater does not fit in the rivers and spreads over the land. Coastal areas are also at risk from sea flooding which happen recently in Kelantan because when the storms and big waves bring seawater into the land then flood occurred.

This study focuses on east of Malaysia where Terengganu becomes the study area. Terengganu is one of East Coast Malaysia state that faces flood problems during monsoon seasons. Terengganu experienced heavy rain when the northeast monsoon blows between November and March. Some areas suffered flooding during this period of time and it is inadvisable to visit any of the offshore islands as sea can be very rough. This natural phenomenon is insurmountable, but can be avoided if there are hasty alert to the community.

Terengganu has 70% lowlands and 30% are always exposed to flood. It is 3000m from sea levels and the flood usually to occur on monsoon season starting from October to March where the northeast monsoon brings heavy rainfall. Since 1990, Terengganu record high rainfall and river water level during the season. Moreover, in 2014 flood caused more than \$78.2 million cover on private houses, infrastructures, facilities and department buildings [19]. In December 2014, the most unscrupulous flood happens at the Kelantan state where 150,000 people from 36,000 families need to move to 309 safe locations [26]. The research carried out by studying sample data on river flow density from Terengganu main rivers. The flow density and flood areas during monsoon season will be under investigate to discover the association between two events. The goal of the research is to predict high possibilities of an area having flood disaster, so that any casualties of adverse impact can be avoided.

This study focuses only river floods. Rainfall over an extended period and an extended area can cause major rivers to overflow their banks. The water can cover enormous areas. Downstream areas may be affected, even when they didn't receive much rain themselves. This the normal type of flood happen during monsoon season due to the

high intensity of a rainfall. Flash floods are also common things in Malaysia. However, flash flood is not covered in this research as different elements that contribute to the flood types.

2. RELATED WORKS

In hydrology, hydrologist has gained attention in data mining because the hydrology field has very data-intensive domain and believe can solve selected hydrology task [4]. In fact, data mining will develop an efficient algorithm for discovering patterns in data and useful rules. Previous work [5] detects a flood pattern using a sliding window technique and temporal database, historical data and experimental results showed that by employing a regression technique, mathematical flood prediction were formulated. Neural Network also increasingly used in hydrological research where autocorrelation of time series and precipitation field past measure of flows and rainfalls [6, 7].

The forecast obtains by weighting predictions and neural predictor identified through a fuzzy approach of basin state. Besides, realizing predicting the flood event is extremely important whereby decision support also can be applied [8, 9]. A Web GIS (Geographical Information System) based decision support system has been developed and able to dynamically display observed and predicted flood extent for decision makers and general public. It is the combination of data integration, floodplain delineation and online map interfaces to show up the output [8].

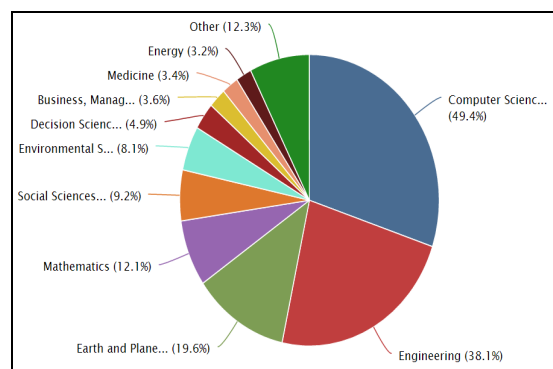


Figure 1: Data Mining In Disaster By Subject Area

Based on research that composed in comprehensive database [27], fig. 1 show percentage of the past ten years for flood predictions in Computer Science where the scientist tries to find the solution through computer aided.

Hence, the research to find a solution for the problems increases in demand, especially in the physical science field and in fact the result obtain more logical and accurate.

Fig. 2 showed the country/territory which conducts an experiment to find correlation between flood and mining data. China, which is the top country always face terrifying flood which affect people's life and economical state [9].

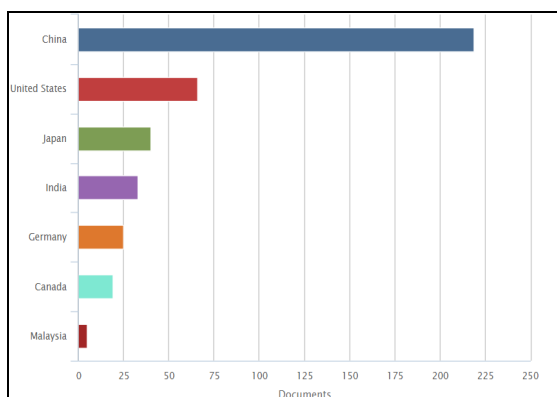


Figure 2: Data Mining In Disaster By Country

In Southeast Asia, various countries have been affected by floods with Malaysia topped the list. Due to the bad weather and dangerous level of water level, it then forced almost 237,037 people displaced to a safe place. Malaysia has been recorded 21 fatalities until December 2014 [19]. In 2014, Terengganu faced 3 waves of flood during monsoon and caused 68,184 evacuate by the authorities [18]. Since the 2014/2015 five peoples are dead in Terengganu due to flood. Above and beyond, infrastructure failure and damage are \$3 million include roads, buildings, parks, etc. Our focus on the study is concern flood happening because of high levels of water at river banks. The study areas encrustation flooding for a long time and has been a significant problem. Fig. 3 shows the study areas of this research.

Terengganu is one of the countries in Malaysia which experienced devastating flood every year. The Department of Irrigation and Drainage report that approximately 29,800 km² or 9% of the total Malaysia area were estimated exposed to flood. It is affecting 4.82 million people from the total population of the country because Malaysia located at low land and has 189 river basins.

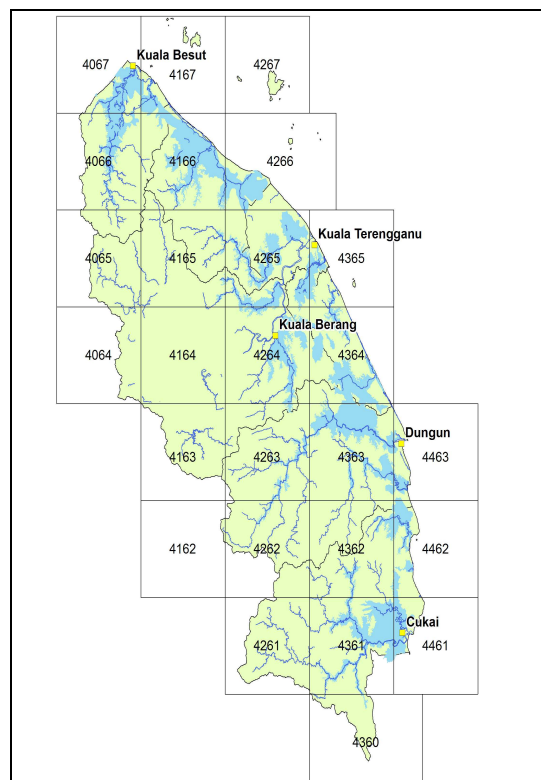


Figure 3: Area Of Study

3. METHODOLOGIES

The fundamental process of data mining process including data collection, data cleaning / transformation, data integration, data analysis will be conducted to build a geo-simulation model to support flood prediction.

3.1 Data Collection

Several government bodies have been identified as a main resources for a flood management (before, during, after) in Malaysia. There are:

- i. Department of Irrigation and Drainage (DoID)
(<http://www.water.gov.my/>)
- ii. Malaysian Meteorological Department (MetMalaysia)
(<http://www.met.gov.my/>)
- iii. National Security Council (NSC)
(<https://www.mkn.gov.my/>)
- iv. Department of Social Welfare (DoSW)
(<http://www.jkm.gov.my/>)

DoID is responsible for the rehabilitation of irrigation works in early year. In 1970/71, severe floods occurred in many parts of West Malaysia and the situation was so serious that a national disaster had to be declared on January 5th. 1971. Subsequent to this occurrence, flood mitigation and hydrology was made an additional responsibility of the DoID from 1972 onwards. Now, flood management is one of the main task to do by monitoring rainfall distribution and river density.

MetMalaysia is assign to fulfil the needs of meteorological, climatological and geophysical services for national security. Its main tasks are included monitoring rainfall distribution and weather forecasting. This body will produce early warning for any type disaster base on their scientific data collection and expertise. Meanwhile, NSC is responsible as the secretariat and main committees at the Federal and State level on issues involving national safety, public safety and crisis and disaster management. They managed the rescue operation at the time of the events. As a member of the National Disaster Management and Relief Committee, DoSW is responsible for the work of aid delivery and recovery of disaster victims. It has four (4) main roles and responsibilities:

- i. Preparing and maintaining evacuation centers
- ii. Distributing donations of food, clothing and other necessities.
- iii. Manage victims recovery
- iv. Provide guidance, advise and counseling services

The research is focus collecting spatial data between Novembers to January (monsoon season) each year. Some of the sources needs for this study are:

- river water levels data
- historical rainfall
- historical river height
- flood areas

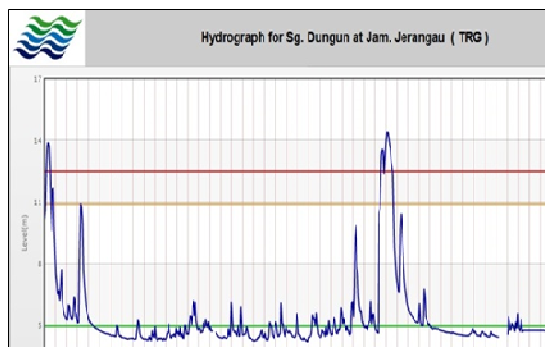


Figure 4: River Water level

Fig. 4 shows an example of data needs to be analysis from Terengganu State Government (TSG) Portal.

Table 1 : Overview of data sources

Source	Period	Attributes	Data Type	Station information
DoID	2000-2015	-Station name, -River level, -Rainfall data - Flood Area	Hourly Data	All station around Terengganu.
MetMalaysia	Past-2015	-Weather forecast, -Temperature,	Hourly Data	States of Malaysia
National Security Council (NSC)	Past-2015	-Flood report, -Flood stage	Monthly Data	State of Malaysia
Terengganu Flood's Report Portal (Run by Terengganu State Government)	2010-2015	-River level, -Rainfall Data -Traffic movement -Flood victims data	Hourly Data	All station around Terengganu
DoSW		1. Number of relief center 2. Number of victims		

Table 1 shows summarized of data availability for each agency. However, from our pre-requirement analysis, we found data are scattered amongst those agencies. Therefore, it becomes one of the main challenges to integrate and ensure the consistency data for each data provider.

However, TSG had produced a portal that contain several crucial information on flood

disaster (<http://etindakan.terengganu.gov.my/>). Fig. 5 showing the river water level in the portal (data from 2001 onward) and Figure 6 shows the rainfall distribution in Terengganu areas. Both data are crucial to the research because this will be the main parameter to perform mining algorithms.

data will be used to predict the area around the flood coordinate. GIS is a tool to help create maps, integrate information, visualize scenarios, present powerful ideas and develop effective solutions by using latitude (3° 53'U - 5° 50'U) and longitude (102° 23'T – 103° 30'T) of the study area. Those data will be transforms and extracted to structured format (DBMS).

StationID (Photo)	Station Name (Graph)	District (State) (Trend text)	Last Update Time	Daily Rainfall				Rainfall since Midnight 13/12/2015		
				07/12/2015	08/12/2015	09/12/2015	10/12/2015		11/12/2015	12/12/2015
5424001	Kg.Keruak	Besut	13/12/2015-09:00	17	14	15	12	12	3	4
5524001	Kg.La	Besut	13/12/2015-09:00	7	9	6	0	0	0	4
5625003	Angga	Besut	13/12/2015-09:00	14	24	2	0	0	15	3
5724003	Jam.Jerteh	Besut	13/12/2015-09:00	12	37	0	0	0	0	2
5626001	Institut Pertanian Tak Dor	Besut	13/12/2015-09:00	12	22	0	1	0	0	0
5626011	Sek.Keb. Kg.Jabi	Besut	13/12/2015-09:00	8	21	1	0	0	0	3
5426003	Kg.Besut	Setiu	13/12/2015-09:00	33	26	5	1	0	2	7
5527024	Jambatan Permaisuri	Setiu	13/12/2015-09:00	22	26	8	1	0	24	4

Figure 5: River Water Level Data

StationID (Photo)	Station Name (Cross-section)	District	River Basin (Trend)	Last Update Time	River Level (Graph)	Normal Level	Alert Level	Warning Level	Danger Level
5424401	Sa_Besut di Jambatan Keruak	Besut	Sa_Besut	13/12/2015-09:00	30.28	28.50	33.00	34.00	35.00
5524401	Sa_Besut di Kg. La	Besut	Sa_Besut	13/12/2015-09:00	17.30	16.50	19.80	20.80	21.80
5724411	Sa_Besut di Jam.Jerteh	Besut	Sa_Besut	13/12/2015-09:00	0.58	2.30	7.30	8.10	8.90
5426401	Sa_Setiu di Kampung Besut	Setiu	Sa_Setiu	13/12/2015-09:00	15.02	13.30	16.40	17.40	18.40
5128438	Jambatan Permaisuri	Setiu	Sa_Setiu	13/12/2015-09:00	5.76	5.50	7.60	8.20	8.80
5428401	Sa Chalok di Jambatan Chalok	Setiu	Sa_Setiu	13/12/2015-09:00	5.62	5.00	7.00	7.70	8.40

Figure 6: Rainfall Data

Our aim is to integrate all the multiple-source data sets into a single intelligible and consistent data repository to conduct a comprehensive analysis.

3.2 Data Cleaning and Transformation

The collected data will be analysed and processed to get the most useful and relevant data. The raw data are in heterogonous formats (word, pdf, excel, web) and need to extract into manageable records. The data will be evaluated through data completeness whereas we will use statistical measure such as parameter mean or median or decision tree to complete the unvalued data. So, the expected values are predictable. Then, the data from different sources will integrate in a logical manner to eliminate data redundancy and detect any value conflicts.

When the secondary data become the sample data, the whole data will be rearranged according the year of monitoring and stations. GIS

3.3 Data Integration

Data integration is the process where multisource data will be combined and transform into an organized format. As the parameter of flood studies is scattered in the research, this process is critical to ensure the consistency and credibility of data, thus ensure the accuracy of mining results. Fig. 7 is showing the integration of data between multiple agencies to create a single consistent data source.

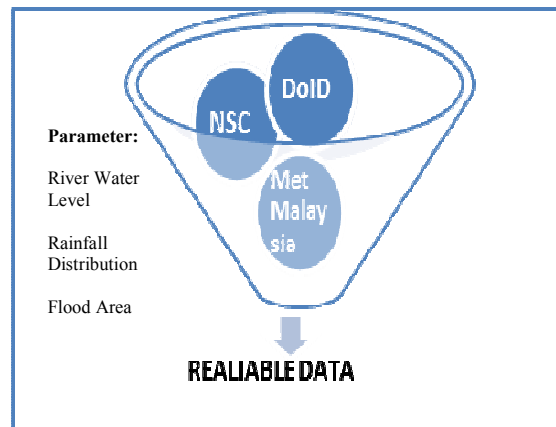


Figure 7: Integration Of Data

3.4 Data Analysis

Information regarding an accurate estimation of extreme events such as flood magnitudes and their frequency of occurrence is great importance in the planning, designing, and management of hydraulic structures such as dams, spillways, culverts, and storm water management systems. Recent catastrophic floods in Australia, Brazil, Pakistan, Thailand and United States call for reliable flood forecasts and long-lead times so that we can better prepare and respond to disastrous events [20].

Association Rules (ARs) are one of Data Mining (DM) techniques that recently used in several analyses of disaster events. It has become one of a major algorithm in data mining research in

order to find a pattern of an element that influences other elements. Early researches in ARs are focused on market analysis finding of an item that have correlation between one and others. It is important to determine a strategy to predict customers' needs. Bala [21] study a negative AR using 8,418 sales transactions for 45 grocery items was collected from various retail outlets. A relation between each item is figure out using negative ARs. ARs also have been used for predicting the factors of catastrophic events. Lee et, al. [22] tried to find the unknown characteristics of earthquakes using AR mining methods global earthquake data occurred since 1973. Dhanya & Kumar [23] are applying ARs to predict floods in India using climate inputs.

The general concept of ARs is followed:

Let $I = \{I_1, I_2, \dots, I_p\}$ be a set of p items and $T = \{t_1, t_2, \dots, t_n\}$ be a set of n transactions, with each t_i being a subset of I . An *association rule* is a rule of the form $X \rightarrow Y$, where X and Y are disjoint subsets of I having a support and a confidence above a minimum threshold [24].

Let us denote by $|X, Y|$ the number of transactions that contain both X and Y . The support of that rule is the proportion of transactions that contain both X and Y : $sup(X \rightarrow Y) = |X, Y| / n$. This is also called $P(X, Y)$, the probability that a transaction contains both X and Y . Note that the support is symmetric: $sup(X \rightarrow Y) = sup(Y \rightarrow X)$.

Let us denote by $|X|$ the number of transactions that contain X . The confidence of a rule $X \rightarrow Y$ is the proportion of transactions that contain Y among the transactions that contain X : $conf(X \rightarrow Y) = |X, Y| / |X|$. An equivalent definition is: $conf(X \rightarrow Y) = P(X, Y) / P(X)$, with $P(X) = |X| / n$.

Following the success of ARs research accuracy results, we hope the research can find a pattern of flood areas in Terengganu based on river flow density. As a result, the output of the research is hoped to be used as an early flood alerts mechanism to mitigate the adverse impacts of monsoon variability and avoid casualties. Fig. 8 shows the conceptual framework for the research.

The purpose of designing this framework is to guide the research with the concept and methodology that being used. The framework has four layers which each layer crucial in carrying out the research. The first layer show the departments that supply data, which the data are not in a ready-used since some of them provide us the graphs,

chart which need to interpret first into reliable dataset. Then, the second layer indicates the dataset need for this research. Next layer is the process where this research took part. The phases start with the data collection and cleansing. The data which have the redundancy, incomplete, corrupt, duplicate will be scrubbed and alter to make sure the data is consistent, harmonize and help the research produce an accurate result in layer four which is flood prediction model.

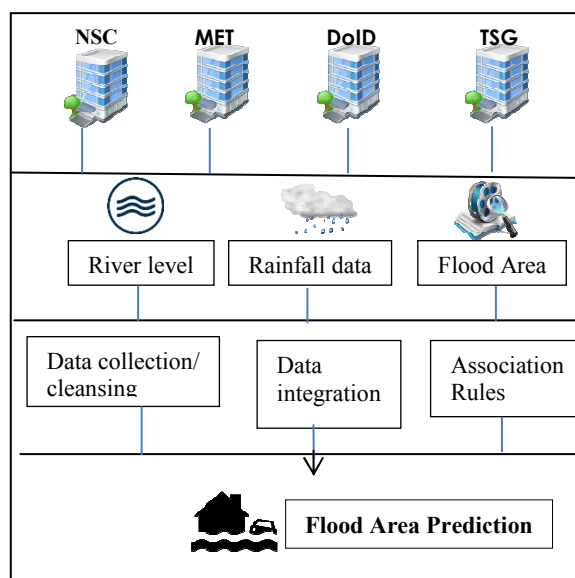


Figure 8: Conceptual Framework For Flood Prediction Model

4. EXPECTED RESULT

River water levels and rainfall measurements are two parameters that be selected for this study. There are 38 river/rain stations along Terengganu area its main tributaries ranging from year 2009 until 2010. Then it will integrate with common flood areas. GIS application will be used to find others potential in flood areas near to the Terengganu main rivers and area nearby stations.

It hopes that certain patterns of flood areas will be generated by using ARs. This is important expected find out as it becomes an input for predicting spot of flood occurrence. Hence, it develops an early warning mechanism to ensure the communities obtain fundamental information before disaster happen.



5. CONCLUSION AND FUTURE WORK

In this paper, we presented an approach to predict potential flooding area using spatial data and GIS. This study aims to inaugurate data mining paradigms with novel geological visualization with the help of intelligent computer involving software proxies. First, we integrate data to produce patterns and association between variables. This research provides the foundation for a revised data mining techniques that can result in enhancements in the preclusion, alleviation, response and salvage from flood events of Terengganu. The future works should consider integrating more variables to make a prediction better and more precisely.

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 to NSC, MetMalaysia and DoID for supplying the data of flood in Terengganu and to all those who participated in this research.

REFERENCES:

- [1] Paulo Cortez and Anibal Morais, "A Data Mining Approach to Predict Forest Fires using Meteorological Data", 2007.
- [2] Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth, "Knowledge Discovery and Data Mining : Towards a Unifying Framework", *Proceedings of KDD-9*, 1996, pp.82-88.
- [3] Li Zheng, Chao Sheng, Liang Tang, Tao Li, Steve Luis and Shu-Ching Chen, "Applying Data Mining Techniques to Address Disaster Information Management Challenges on Mobile Devices", *Proceedings of KDD-1*, August 21-24, 2011, pp. 283-291.
- [4] Milan Cisty and Juraj Bezak, "The Application of Data Mining Methods for Short Time Flows Prediction in Flood Warning Systems," *Recent Advances in Continuum Mechanics, Hydrology and Ecology*, 2013, pp. 92-97.
- [5] Ku Ruhana Ku, Mahamud, Norhayani Zakaria, Norliza Katuk, and Mohamad Shbier, "Flood Pattern Detection Using Sliding Window Technique," *Third Asia International Conference on Modelling & Simulation*, 2009, pp. 45-50.
- [6] Giorgio Corani and Giorgio Guariso, "Coupling Fuzzy Modelling and Neural Networks for River Flood Prediction," *IEEE Transaction on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 35, No 3, August 2005.
- [7] Fazlina A.R, Abd Manan S., Zainazlan M.Z. and Ramli Adnan, "Flood Water Level Modeling and Prediction Using NARX Neural Network: Case Study at Kelang River," *IEEE 10th International Colloquium on Signal Processing & its Application (CSPA2014)*, Mac 7-9, 2014, pp. 204-207.
- [8] Darka Mioc, Francois Anton and Brandford George Nickerson, "Decision Support for Flood Event Prediction and Monitoring," 2007.
- [9] Yan Li and Manchun Li, "Application and Research on Flood Risk Assessment Decision Support System in the Lower Yellow River," 2011.
- [10] Omar Al-Azzam, Deli Sarsar, Kirubel Seifu, Mehdi Mekni, "Flood prediction and Risk Assesment Using Advanced Geo-Visualization and Data Mining Technique : A case study in the Red-Lake Valley," *Proceedings of Applied Computational Science*, 2014.
- [11] Masond Bakhtyari Kia, Saled Piratesh, Biswajeet Pradhan, Ahmad Rodzi Mahmud, Wan Nor Azmin Sulaiman, Abbas Moradi, "An artificial neural network model for flood simulation using GIS : Johor River Basin, Malaysia," *Environ Earth Sci, Springer articles*, December 31, 2011.
- [12] Daniela Stonajova, Pance Panov, Andrej Kobler, Saso Dzeroski and Katerina Taskova, "Learning to Predict Forest Fires With Different Data Mining Techniques," 2006.
- [13] Ch. Lucas, St. Werder and H.-P. Bahr, "Information Mining for Disaster Management," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (3/W49), 2007, pp. 75-80.
- [14] C.T.Dhanya and D.Nagesh Kumar, "Fuzzy Association Rules for Prediction of Monsoon Rainfall," *4th Indian International Conference on Artificial Intelligence (IICAI-09)*, 2009, pp. 1299-1309.
- [15] Thomas Landssdall-Welfare, Seatviga Sudjkar, et. Al, "On the Coverage of Science in the Media : A Big Data Study on the Impact of the Fukushima Disaster," *IEEE International Conference on Big Data*, 2014, pp.60-66.
- [16] Qiang Yang, 2006.10 Challenging Problems in Data Mining Research.



- [17] Mary McGlohon, "Data Mining Disasters: a report,".
- [18] Department of Drainage and Irrigation, Portal Banjir Terengganu [Online]. From <http://jpsweb.terengganu.gov.my/>
- [19] National Security Council, Portal Bencana, [Online]. From <http://portalbencana.mkn.gov>
- [20] Wang, D., Ding, W., Yu, K., Wu, X., Chen, P., Small, D., L. & Islam, S., "Towards Long-lead Forecasting of Extreme Flood Events: A Data Mining Framework for Precipitation Cluster Precursors Identification," *Proceedings of KDD 13, ACM, 2013*.
- [21] Bala, P., K., A Technique for Mining Negative Association Rules, ACM, January, 2009.
- [22] Lee, J., A., Han., J & Chi, K., H. "Mining Quantitative Association Rule of Earthquake Data," *ICHIT'09*, August 27–29, 2009.
- [23] Danya, C., T. & Kumar, D., N., "Data mining for evolution of association rules for droughts and floods in India using climate inputs," *Journal Of Geophysical Research*, Vol. 114, D02102, 2009
- [24] Merceron, A. and Yacef, K., "Interestingness Measures for Association Rules in Educational Data," *1st International Conference on Educational Data Mining (EDM08)*, Montreal, Canada, 2008.
- [25] IBM website. Big Data and Information Management. [Online] From : <http://www-01.ibm.com/software/data/bigdata/>
- [26] Department of Drainage and Irrigation Kelantan, eBanjir Portal. [Online] From <http://did.kelantan.gov.my/>.
- [27] World Research Online database. [Online] From www.scopus.com