IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 8, No. 4, December 2019, pp. 360~366
ISSN: 2252-8938, DOI: 10.11591/ijai.v8.i4.pp360-366

Predicting fatalities among shark attacks: comparison of classifiers

Lim Mei Shi, Aida Mustapha, Yana Mazwin Mohmad Hassim
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia,
Parit Raja, 86400 Batu Pahar, Johor, Malaysia

Article Info

Article history:

Received Aug 2, 2019
Revised Oct 10, 2019
Accepted Oct 24, 2019

Keywords:

Bayes point machines
Data mining
Machine learning
Prediction
Support vector machines

ABSTRACT

This paper presents the comparisons of different classifiers on predicting Shark attack fatalities. In this study, we are comparing two classifiers which are Support vector machines (SVMs) and Bayes Point Machines (BPMs) on Shark attacks dataset. The comparison of the classifiers were based on the accuracy, recall, precision and F1-score as the performance measurement. The results obtained from this study showed that BPMs predicted the fatality of shack attack victim experiment with higher accuracy and precision than the SVMs because BPMs have "average" identifier which can minimize the probabilistic error measure. From this experiment, it is concluded that BPMs are more suitable in predicting fatality of shark attack victim as BPMs is an improvement of SVMs.

Corresponding Author:

Yana Mazwin Mohmad Hassim,
Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia,
Parit Raja, 86400 Batu Pahat, Johor, Malaysia.
Email: yana@uthm.edu.my

1. INTRODUCTION

Shark attack can be categorized into two types which are provoked and unprovoked. According to the International Shark Attack File (ISAF), 88 cases were confirmed to be unprovoked shark attacks on human out of the investigated 155 incidents of alleged shark-human interaction occurring worldwide in 2017 [1]. This result is higher than the most recent years (2012-2016), with average of 83 incidents annually [2]. Growing number of shark attacks have caused human to fear of shark [3]. Shark attack could be fatal and non-fatal. Nonetheless, only 5 among 98 unprovoked attacks were fatal worldwide in 2015 [4], which is around 5%. Human has bad impression because they tend to think that shark attack as their nature [5] eventhough in reality dogs or bees kill more people every year than sharks [6]. United States is the leading country that has the most shark attack, with 60 percent of the globe's 88 unprovoked shark attacks in 2017 [7]. However, the United States did not have any shark attacks that resulted in fatality. Australia, on the other hand, had 7% fatality rate in shark attacks, which means 1 out of 14 incidents reported in Australia has resulted in a fatality in 2017 [8]. The number of human-shark interactions is directly correlated with time spent by humans in the sea [9]. The higher the number of human-shark interaction, the higher the risk of being attacked by a shark. However, only certain species of shark attack are more likely to lead to fatality.

While research on predicting specifically shark attack is limited such as in [10-12], the literature has shown various data mining approaches used in analyzing fatalities of other animal attacks such as leopard [13], elephant [14], and snake [15]. The main motivation to predict fatalities among shark attacks is atributed to the increasing number of shark attack worldwide over the past five years [16]. However, some studies found that most of the shark attacks are not fatal. In unprovoked cases, shark attacks only when they are confused a person with natural prey due to its poor vision. Thus, the shark will approach, bite and swim away after verifying that the victim is not part of their diet. In fact, sharks have no particular liking for human flesh as it contains a lower level of fat than they need.

Besides, the false impression of human on shark must be corrected. Shark does not simply attack people as what movies portrayed [17]. Shark attacks are only triggered when human did certain actions or doings which may confuse or threaten the shark. The activities conducted by the victim will affect the chances of being attacked by a shark. Following recent trends, surfers and those participating in board sports accounted for most incidents as this group spends the most time in the surf zone, an area commonly frequented by sharks and may unintentionally attract sharks by splashing or paddling. Swimmers and waders accounted for 22% of incidents, 9% for snorkelers or free divers, 2% for scuba divers, 3% for body-surfers and 5% for those participating in other shallow water activities. The low awareness of human on shark attack is the next motivation for this research. Human is not aware of the effect of wearing bright clothes which will capture shark’s attention as the reflected light from the clothes can be confused with the brightness of the fish’s scales. Additionally, human who conducts water sports activities during shark feeding time is exposed to higher chances of being attacked as sharks are more sensitive early in the morning or late at night. Thus, this shark attack dataset will also determine if the time of conducting sports activities will affect the fatality of the victim of shark attack.

Other than that, the decreasing number of white shark problem has motivated us to choose this dataset [18]. The negative image of the white shark and the fear it projected on humans often resulted in unwarranted killing of the species [19]. These actions are made worse by the proximity of white shark feeding and breeding areas to coastal human populations of the world’s sharks and rays [20]. Thus, the fatality of shark attacks victims is determined to understand better if shark is really a lethal animal. This paper presents the comparisons of different classifiers on predicting Shark attack fatalities. . In this study, we are comparing two classifiers which are Support vector machines (SVMs) and Bayes Point Machines (BPMs) based on four standard performance measurement : accuracy, recall, precision and F1-score. We aimed to seek a better classifer that can be use to predict the shark attack fatalities that can help to avoid this unwanted incident in the future.

2. RESEARCH METHOD

This study is carried out based on SEMMA methodology which consist of five phases; sample, explore, modify, model and assess [21]. Figure 1 shows the process flow of each phases in SEMMA method. Sample phase will select the data and determine the source of the dataset whereas exploring the input data is phase 2. Shark attack dataset is the input needed to do the analysis on fatality of shark attack victim. 10 out of 20 attributes will be used as the input of this analysis. The following modify phase include the process of preparing, repairing and transforming the data. This dataset has to be preprocessed before applying to algorithm. Otherwise, the result of analysis will be affected and not accurate. On the other hand, model phase will undergo the process of applying the algorithm or techniques to create the model that possibly provide the desired outcome. Classification algorithm is used in this analysis as the response class contains binary data which is either fatal or non-fatal. Classification is well known supervised learning task that has been previously used in other domains such as in music [22], heart disease [23], and traffics [24]. The final assess phase will evaluate the performance of each algorithm by using some standard metrics.

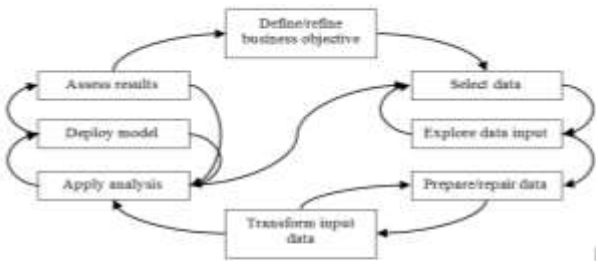


Figure 1. SEMMA methodology [21]

2.1. Dataset

Shark attack data is used to conduct this experiment. In this experiment, 6,095 instances, 22 attributes included 1 class attribute are involved. The 22 attributes are case number 1, date, year, type, country, area, location, activity, name, sex, age, injury, fatal, time, species, investigator or source, pdf, href formula, href, case number 2, case number 3, original order. The class attribute namely fatal. However, only 10 attributes which are type, country, area, location, activity, sex, age, fatal, time and species will be used to test the fatality of victims in the cases of shark attack. Others attributes such as case number 1, date, year, name, investigator or source, pdf, href formula, href, case number 2, case number 3 and original number will not be tested in the experiment. The data dictionary for this dataset is listed in Table 1.

Table 1. Data dictionary

Number	Attributes	Description
1	case_number	Date when the incident was reported
2	Date	Date when the incident was happened
3	Year	Year when the incident was happened
4	Type	Type of shark attack
5	country	Country where shark attack took place
6	Area	Area where shark attack took place
7	location	Location where shark attack took place
8	activity	Activity conducted by the victim
9	Name	Name of victim
10	Sex	Sex of victim
11	Age	Age of victim
12	Injury	Types of injury happened to the victim
13	fatal_y_n	Fatality of victim
14	Time	Time of incident
15	species	Species of shark
16	investigator_or_source	Investigator involved in the case
17	Pdf	Case name
18	href_formula	Reference hyperlink
19	Href	Reference hyperlink 2
20	case_number_2	Related case number 2
21	case_number_3	Related case number 3
22	original_order	Order of file

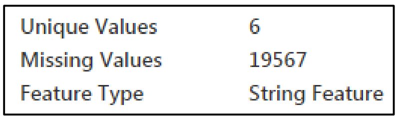
The shark attack dataset was sourced from a website called data.world. Data.world is a website that contains numerous different types of raw datasets, published by numerous contributors from different countries. It is a platform for people to collaborate, contribute and solve problems relating to datasets. New data ranging from finance to health to sports and politics can be discovered from various sources through this website. User would only need to filter the searches according to their preferences. This dataset is collected from Global Shark Attack File [25]. The aim of this website is to provide current and historical data on shark or human interactions for those who seek accurate and meaningful information and verifiable references. On the other hand, this workspace is contributed by Shruti Jayaram Prabhu on 22 June 2017. This dataset is publicly shared and contains shark attack reportings from over a century.

2.2. Pre-Processing

Before building the classification model using the dataset, the dataset was first pre-processed to cater issues such as missing and continuous values. Data to be analyzed by data mining techniques can be incomplete, noisy and inconsistent. The purpose of data cleaning is to clean the data to be analyzed [26]. There are many different cleaning modes available for user to select such as removing the entire column or removing the entire row. In shark attack raw dataset, many missing values were found and these missing values would be removed in order to provide a better and accurate result during mining process. Figure 2 shows one of the attribute with missing value in shark attack dataset.

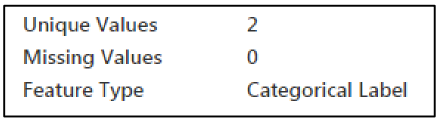
In [27], SMOTE module was used to treat the imbalanced dataset. This is because this shark attack dataset is bias to non-fatal of shark attack victim. Thus, this imbalanced dataset needs to be corrected to produce accurate result during analysis. Data transformation is used to transform or consolidate the data into forms appropriate for mining [28]. The type, activity, sex, species and age attributes in shark attack dataset would be transformed to categorical feature type. This is because the values of data in those attributes can be sorted into groups or categories. Categorical data must be cast categories so that the computer can treat them correctly when using classification algorithm. Figure 3 shows one of the attribute that is transformed from string feature type to categorical feature type in order to use the classification learning.

Next, data reduction is used to reduce the number of attributes which are not relevant to the analysis without compromising the integrity of the original data and yet producing the quality knowledge [29]. This is to reduce the complexity of data, making the analysis process become quicker and increase its efficiency. Only 10 attributes out of 22 attributes would be used in this analysis. The injury, case number, date, year, country, area, location, name, investigator or source pdf, href formula, href, case number (2), case number (3) and original order are excluded in this analysis. This is because these attributes are not significant and not useful in this analysis. Type, country, area, location, activity, sex, age, time, species and fatality of victims attributes are useful in this analysis and thus they are included in this analysis. Figure 4 shows the attribute used in analysis.



Unique Values	6
Missing Values	19567
Feature Type	String Feature

Figure 2. Missing values of fatal attribute



Unique Values	2
Missing Values	0
Feature Type	Categorical Label

Figure 3. Categorical feature type of fatal attribute

project shark attack

rows	columns
1914	10

Figure 4. Attribute used in analysis

Different algorithm would require different specific content types in order to function correctly. In data discretization, values are put into buckets so that there are a limited number of possible states [30]. Data in the columns can be discretized to enable the use of the algorithms to produce mining model. Age attribute in shark attack dataset would be categorized and converted to higher conceptual level through the level of hierarchies. The values for age attribute would be divided into several categories with fixed size of interval. With this being done, the dependency between the class and the interval are increased and provide a more accurate result.

2.3. Algorithms

This paper adopted the classification technique for predicting shark attack fatalities. The experiments were carried out using the Azure ML tool [31] with 10-fold validation method to evaluate the SVM and BPM classifiers performance. Cross-validation model module was used in Azure ML to perform this validation process. Cross-validation parameter partitioned the data into 10 folds to estimate for classification model. 9 sets were used to train the classifier while the performance of classifier was assessed on the 1 left subset. This was then iterated ten times as subsets were included in training and test sets. The average performance is considered as the final performance of a classifier. Quality of data set can be determined by comparing the accuracy statistics for all the folds. Since the shark dataset used only have two classes; fatal or not fatal, two-class type of algorithm was be selected for the classification experiment. Support vector machines (SVMs) and Bayes Point Machines (BPMs) can be used in the classification of supervised learning dataset. BPMs are a type of linear classification algorithm which was introduced by Ralf Herbrixh, Theore Grapel, and Colin Campbell in 2001. BPMs are known as an “average” classifier that can efficiently approximate the theoretical optimal Bayesian average of several linear classifiers based on their ability to generalize [10]. This classifier is used to minimize the probabilistic error measure. The “average” classifier is known as Bayes point. BPMs algorithm used in Azure Machine Learning are based on Infert.Net and can perform better than the other Bayesian algorithm.

Number of iteration of test can be set in Azure ML. The higher number of iterations, the higher the accuracy of the result. BPMs are more robust and less prone to over-fitting of the data whereby the production of analysis are too similar to the data and causing it to fail to fit additional data or predict future observations reliably. It can also reduce the need to perform performance tunings and therefore time needed to run the experiment can be decreased. Expectation propagation is used in BPMs as the message-passing algorithm which passes the message to other nodes across the edges of model and thus produces a fast and accurate result [32]. SVM algorithm was introduced [33]. This algorithm will assign data to one class or the other by discovering hyperplanes which cleanly segregate data into classes [34]. New data points can be easily classified once ideal hyperplanes are discovered.

- Two-Class Bayes Point Machine. The formula is shown in (1), where A and B are events and P(B)≠0.

$$\mathrm{P}(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{1}$$

- Two-Class Support Vector Machine. The formula is shown in (2).

$$\frac{w}{||w||}.(\mathrm{x2} - x1) = width = \frac{2}{||w||} \tag{2}$$

2.4. Evaluation Metrics

The evaluation metrics used in the experiments are accuracy, precision, recall and F1 score. Accuracy performs best if false positives and false negatives have similar cost [35]. Precision and Recall will be used if the cost of false positives and false negatives are very different. True positive happens when predicted class and actual class are true whereas true negative happens when predicted class and actual class are both false. False positive happens when actual class is false but predicted class is true whereas false negative happens when predicted class is false but actual class is true. On the other hand, F1 score is used when more realistic measure of classifier's performance is required as arithmetic mean between a poor precision and a very high recall can be avoided [36].

- Accuracy. Accuracy is the ratio of summation of true positive and true negative to the total events. The formula for calculating accuracy is shown in (3).

$$\text{Accuracy} = \frac{\text{True Positive+True Negative}}{\text{Total events}} X\ 100\% \tag{3}$$

- Precision. Precision is the ratio of true positive to the total predicted positive observations. The formula for calculating precision is shown in (4).

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False Positive}} X\ 100\% \tag{4}$$

- Recall. Recall (Sensitivity) is the ratio of correctly predicted positive observations to the all observations in actual class. The formula for calculating recall is shown in (5).

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive+False Negative}} X\ 100\% \tag{5}$$

- F1 score. F1 score is the average of precision and recall, it reaches its best value at 1 and worst at 0. The formula for calculating F1 score is shown in (6).

$$\text{F1 score} = \frac{\text{2 x precision x recall}}{\text{precision+recall}} \tag{6}$$

3. RESULTS AND DISCUSSION

The purpose of the experiments is to compare the performance of Bayes Point Machine and Support Vector Machine algorithms in shark attack dataset for accuracy, precision, recall and F1 score. The results showed that Bayes Point Machine performs better in this dataset compared to Support Vector Machine. BPMs have higher accuracy than SVMs. This is because Bayes-optimal classifier will minimize average error when marginalizing over all possible boundaries and all possible samplings of the data by finding the boundary in a fixed space which is closest to this classifier. Besides, BPMs also has a higher precision than SVMs. This is because the sampling scheme used in BPMs is very simple and efficient, thus making it to be applicable to large data sets such as shark attack dataset. The summary are shown in Table 2.

Table 2. Experimental results

Algorithm	Accuracy	Precision	Recall	F-Measure
Two-Class Bayes Point Machine	0.952	0.899	0.983	0.939
Two-Class Support Vector Machine	0.816	0.754	0.762	0.758
Two-Class Logistic Regression	0.803	0.740	0.740	0.740
Two-Class Boosted Decision Tree	0.854	0.794	0.829	0.811

Other than that, BPMs have a higher recall value compared to SVMs. This is because BPMs propose a novel differentiable loss function called trigonometric loss function which will normalize the likeliness of desirable characteristic before setting up a Bayesian framework using standard Gaussian processes techniques. BPMs have higher accuracy than that of Logistic Regression. This is because BPM have intuitions that can specify the prior in the shark attack dataset. With intuitions, BPM can make prediction on the model through the posterior. BPM can simply work by identify few important independent variables compared to Logistic Regression which need to include all important independent variables. This enables BPM to be well operate when certain class of shark attack dataset is changed or edited as BPM can evaluate the variables from the most important variable by itself. Hence, Logistic Regression is outperformed by BPM. In addition, BPMs have higher accuracy than that of Decision Tree. This is because BPM can have bigger training set compared to Decision Tree. This ensure the low classification error rate as bigger training set consists of more number of classes. BPM can admit the training error in shark attack dataset to avoid existing of noisy data.

4. CONCLUSIONS

In this paper, the comparison of two classifiers' performance on fatality of shark attack victim was carried out. The dataset was run on two classifiers; Support vector machines (SVMs) and Bayes Point Machines (BPMs) and their performance were analysed. Based on the result, it is shown that BPMs was able to predict the result with higher accuracy and precision as compare to SVMs due to the ability of BPMs to minimize the average error when marginalizing over all possible boundaries and possible samplings of the data. From this work, we can conclude that between these two classifiers, the BPMs are more suitable in predicting the fatality of shark attack victim. A future work may be carried out to seek a better classifer that can be efficiently used to predict the fatality of shark attack victim in order to avoid such an unwanted incident in the future.

ACKNOWLEDGEMENTS

This research is supported by Universiti Tun Hussein Onn Malaysia.

REFERENCES

[1] "Yearly Worldwide Shark Attack Summary", retrieved from https://www.floridamuseum.ufl.edu/shark-attacks/yearly-worldwide-summary/, 2018.
[2] D. G. Caldicott, R. Mahajani, M. Kuhn, "The anatomy of a shark attack: A case report and review of the literature," *Injury*, vol. 32, no. 6, pp. 445-453, 2011.
[3] C. McCagh, J. Sneddon, D. Blache, "Killing sharks: The media's role in public and political response to fatal human–shark interactions," *Marine Policy*, vol. 62, pp. 271-278, 2015.
[4] E. Clua, B. Séret, "Unprovoked fatal shark attack in Lifou Island (Loyalty Islands, New Caledonia, South Pacific) by a great white shark, Carcharodon carcharias," *The American Journal of Forensic Medicine and Pathology*, vol. 31, no. 3, pp. 281-286, 2010.
[5] A. Kock, R. Johnson, "White shark abundance: Not a causative factor in numbers of shark bite incidents. Finding a balance: White shark conservation and recreational safety in the inshore waters of Cape Town," pp. 1-19, 2006.
[6] Z. Lucas, W. T. Stobo, "Shark-inflicted mortality on a population of harbour seals (Phoca vitulina) at Sable Island, Nova Scotia," *Journal of Zoology*, vol. 252, no. 3, pp. 405-414, 2010.
[7] J. Searing, "The Big Number: 53 shark attacks in U.S. waters," Retrieved from https://www.washingtonpost.com/national/health-science/the-big-number-53-shark-attacks-in-us-waters/2018/06/29/ad1f75d0-7aec-11e8-aeee-4d04c8ac6158_story.html?noredirect=on&utm_term=.38a03a3ef958, 2018.
[8] R. Crossley, C. M. Collins, S. G. Sutton,C. Huveneers, "Public perception and understanding of shark attack mitigation measures in Australia," *Human dimensions of wildlife*, vol. 19, no. 2, pp. 154-165, 2014.
[9] C. Neff, R. Hueter, "Science, policy, and the public discourse of shark "attack": A proposal for reclassifying human–shark interactions," *Journal of environmental studies and sciences*, vol. 3, no. 1, 65-73, 2013.
[10] R. Herbrich, T. Graepel, C. Campbell, "Bayes point machines," *Journal of Machine Learning Research*, vol. 1, pp. 245-279, 2001.
[11] J. Frost, "Overfitting Regression Models: Problems, Detection, and Avoidance," Retrieved December 9, 2018 from http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm, 2018.
[12] N. Olivo, "Using Shark Attacks to Understand Bayesian Networks," Retrieved December 9, 2019 from https://medium.com/@NatalieOlivo/shark-bites-920299b908b2, 2018.
[13] D. G. Nabi, S. R. Tak, K. A. Kangoo, M. A. Halwai, "Injuries from leopard attacks in Kashmir," *Injury*, vol. 40, no. 1, pp.90-92, 2009.
[14] S. K. Das, S. Chattopadhyay, "Human fatalities from wild elephant attacks: A study of fourteen cases," *Journal of Forensic and Legal Medicine*, vol. 18, pp. 154-157, 2011.
[15] M. K. Sadoon, "Snake bite envenomation in Riyadh province of Saudi Arabia over the period (2005–2010)," *Saudi Journal of Biological Sciences*, vol. 22, no. 2, pp. 198-203, 2015.
[16] C. Pepin-Neff, T. Wynter, "Shark Bites and Shark Conservation: An Analysis of Human Attitudes Following Shark Bite Incidents in Two Locations in Australia," *Conservation Letters*, vol. 11, no. 2, 2017.
[17] C. Neff, "The Jaws effect: how movie narratives are used to influence policy responses to shark bites in Western Australia," *Australian Journal of Political Science*, vol. 50, no. 1, pp. 114-127, 2015.
[18] N. K. Dulvy, S. L. Fowler, J. A. Musick, R. D. Cavanagh, P. M. Kyne, L. R. Harrison, C. M. Pollock, "Extinction risk and conservation of the world's sharks and rays," *eLife*, vol. 3, 2014.

[19] R. A. Martin, D. K. Rossmo, N. Hammerschlag, "Hunting patterns and geographic profiling of white shark predation," *Journal of Zoology*, vol. 279, no. 2, pp. 111-118, 2009.
[20] G. Bianucci, M. Bisconti, W. Landini, T. Storai, M. Zuffa, S. Giuliani, A. Mojetta, "Trophic interaction between white shark, carcharodon carcharias, and cetaceans: A comparison between Pliocene and recent data from central Mediterranean Sea," In *Proceedings of the 4th European Elasmobranch Association Meeting*, pp. 33-48, 2000.
[21] A. I. R. L. Azevedo, M. F. Santos, "KDD, SEMMA and CRISP-DM: A parallel overview," *IADS-DM*, 2008.
[22] D. Mohammed, K. A. Karawi, P. Duncan, F. L. Francis, "Overlapped Music Segmentation using a New Effective Feature and Random Forests," *AIES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no 2, 2019.
[23] H. Karim, S. R. Niakan, R. Safdari, "Comparison of Neural Network Training Algorithms for Classification of Heart Diseases," *AIES International Journal of Artificial Intelligence (IJ-AI)*, vol. 7, no 4, 2018.
[24] P. R. Iyer, S. R. Iyer, R. Ramesh, M. R. Anala, K.N. Subramanya, "Adaptive real time traffic prediction using deep neural networks," *AIES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 2, 2019.
[25] E. Ritter, M. Levine, "Use of forensic analysis to better understand shark attack behavior," *Journal of Forensic Odontostomatology*, vol. 22, no. 2, pp. 40-46, 2004.
[26] E. Rahm, H. H. Do, "Data cleaning: Problems and current approaches," *IEEE Data Eng. Bull.*, vol. 23. no. 4, pp. 3-13, 2020.
[27] N. Qazi, K. Raza, "Effect of feature selection, SMOTE and under-sampling on class imbalance classification," In *Proceedings of 2012 UKSim 14th International Conference on Computer Modelling and Simulation (UKSim)*, pp. 145-150, 2012.
[28] J. Han, J. Pei, M. Kamber, "*Data mining: concepts and techniques*", 2011.
[29] E. Namey, G. Guest, L. Thairu, L. Johnson, "Data reduction techniques for large qualitative data sets," *Handbook for team-based qualitative research*, vol. 2, no. 1, 137-161, 2008.
[30] R. Jin, Y. Breitbart, C. Muoh, "Data discretization unification," *Knowledge and Information Systems*, vol. 19, no. 1, 2009.
[31] M. Bihis, S. Roychowdhury, "A generalized flow for multi-class and binary classification tasks: An Azure ML approach," In *2015 IEEE International Conference on Big Data (Big Data)*, pp. 1728-1737, 2015.
[32] T. P. Minka, "Expectation propagation for approximate Bayesian inference," In *Proceedings of the Seventeenth conference on Uncertainty in Artificial Intelligence*, pp. 362-369, Morgan Kaufmann Publishers Inc., 2001.
[33] C. W. Hsu, C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.
[34] O. Benarchid and N. Raissouni, "Support Vector Machines for Object Based Building Extraction in Suburban Area using Very High Resolution Satellite Images, a Case Study: Tetuan, Morocco." *AIES International Journal of Artificial Intelligence (IJ-AI)*, vol. 2, no. 1, 2013.
[35] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861-874, 2006.
[36] G. Tsoumakas, I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining*, vol. 3, no. 3, pp. 1-13, 2007.

BIOGRAPHIES OF AUTHORS

Lee Jun Mei received the secondary education in Science Stream from Sekolah Menengah Jenis Kebangsaan (C) Chan Wa II, Seremban, Negeri Sembilan. Then, she continues her pre-university study at St. Paul Institution, Seremban, Malaysia. In September 5, 2016, She received B. degree in Information Technology from UTHM. She hopes that she can become an excellent system analyst in the future.

Aida Mustapha received the B.Sc. degree in Computer Science from Michigan Technological University and the M.IT degree in Computer Science from UKM, Malaysia in 1998 and 2004, respectively. She received her Ph.D. in Artificial Intelligence focusing on dialogue systems. She is currently an active researcher in the area of Computational Linguistics, Soft Computing, Data Mining, and Agent-based Systems.

Yana Mazwin Mohamad Hassim graduated with a PhD degree from Universiti Tun Hussein Onn Malaysia (UTHM) in 2016. Earlier, in 2006 she completed her Master's degree in Computer Science from Universiti of Malaya (UM). She received her Bachelor of Information Technology (Hons) degree majoring in Industrial Computing from Universiti Kebangsaan Malaysia (UKM) in 2001. Her research area includes neural networks, swarm intelligence, optimization and classification.