Feature Exploration for Prediction of Potential Tuna Fishing Zones

Devi Fitrianah, Nursidik Heru Praptono, Achmad Nizar Hidayanto, and Aniati Murni Arymurthy

***Abstract*—Prediction for potential fishing zone is one of the important activities concerning for the tuna fishing exploration, conservation and management. Accurate prediction will give more efficient in fishing activities. One of the way to predict is the classification techniques. Currently, as the state of the art, most of the methods utilize the chlorophyll and SST features. However, there are still other parameters that can be utilized. In this paper, the other parameters are then observed: ocean currents and salinity feature. First the results shows that, taking a part of ocean currents together with the chlorophyll and SST feature combination gives the improvement on the prediction. On other hand, this ocean currents feature is then substituted with the salinity, and the result shows that the combination between salinity, chlorophyll, and SST also increases the result. Finally, the ocean current and salinity parameters are combined together with chlorophyll and SST parameters and the result was surprising. It is found that the last feature combination which includes Chlorophyll, SST, Ocean current and salinity gives the highest result in classification (in Naïve Bayes reaches 69.03%, Decision Tree reaches 82.32% and SVM reaches 68.30% of accuracy) compared to the "baseline" feature combination including only Chlorophyll and SST (in Naïve Bayes reaches 57.44%, Decision Tree reaches 58.91% and SVM reaches 56.74% of accuracy). Therefore it is suggested that the proposed feature can be harnessed for the better prediction of potential fishing zone.**

***Index Terms*—Feature exploration, potential tuna fishing zones, classification, chlorophyll, sea surface temperature (SST), ocean currents, salinity.**

I. Introduction

Tuna fish, is one of the important fishery commodities in the world. There are many potential values added in a tuna fish, and therefore it is good to consume the tuna fish to fulfill the nutrition need of human life. Industries are now trying to explore the tuna fish regarding to the industrialization of tuna fish, for example either the canned tuna fish or fresh tuna fish [1].

However there is still a problem in tuna fish catching, regarding to the area that a fisherman should visit. The way to determine which trip that the fisherman should choose, however needs the good prediction of potential fishing zone [2], [3]. The prediction itself can be performed in many ways. One of the prediction methods that can be utilized is the classification techniques. Some research in prediction of potential fishing zone [4] have been done utilizing the physical and biochemical marine aspects either derived from remote sensing, or given by primary data from ground truth [5], [6]. However, most of them are still using the two common oceanographic parameters which are chlorophyll and SST feature although they utilize some specific technique or tools.

In this research the experiment employs the classification process. The ground truth data provided from PT Perikanan Nusantara (PTPN) enables us to validate this work. The goal of this research is the finding of the explored important feature combinations which give better result in predicting the potential tuna fishing zone. Instead from PTPN, it is also collected: some needed data from National Oceanic and Atmospheric Administration (NOAA) [7], [8]. Then experiment utilizes three different supervised learning methods, including Naïve Bayes, (which is probabilistic based), Decision Tree (which is information gain based), and the last, support vector machine-or SVM- which is a kernel based classifier. Finally the result evaluation and analysis is then described.

II. Literature Review

A. Related Works Regarding to Potential Fishing Zone

There are some works related to tuna fishing zone. Mansor *et al.* in 2001 [4] explored the satellite fish forecasting technique considering sea surface temperature (SST) and chlorophyll intensity of the sea. This work utilized satellite images, and geographical information system (GIS) as a Topical Fish Forecasting System (TroFFS) to do the prediction. The results shows that these two parameter (SST and chlorophyll) can be determined as important parameter besides of including upwelling, boundary, nutrient, and phytoplankton and many other, since the relationship amongst demersal features is difficult to establish.

Later, in 2005 Solanki *et al*. in their work [5] analyzed the remote-sensing-based methodology to predict the potential fishing zone. Still the same as Mansor did, this research also utilized chlorophyll and SST, but the scope of the object has been specified for some kind of species (Ribbon Fish, Cat Fish, and some others). In this case, chlorophyll is derived from ocean color monitor (OCM) while SST is derived from advanced very high resolution radiometer (AVHRR). Again, this result endorses the phenomenon that there is a high relationship between habitat-foods, and biochemical-physical parameters.

The Solanski's work was also then validated and developed by Rahul *et al*. in 2011 [6] by using those two source of data (OCM and AVHRR) to derive the physical and

Manuscript received August 19, 2014; revised October 25, 2014. This work was supported in part by the Universitas Indonesia Research Grant 2014 (BOPTN).

The authors are with the Faculty of Computer Science, Universitas Indonesia, Depok 16424, Indonesia. Devi Fitrianah is also with the Faculty of Computer Science, Universitas Mercu Buana, Jakarta 11650, Indonesia (e-mail: fitrianah.devi@gmail.com, heru.pra@cs.ui.ac.id, nizar@cs.ui.ac.id, aniati@cs.ui.ac.id).

DOI: 10.7763/IJIEE.2015.V5.543

biochemical factors. Their concept employs the satellite image developed with mesoscale eddy-simulating models. However, this work focuses on long-term prediction, and not the on the oceanographic (physical or biochemical) features.

A rough cluster prediction has also been introduced by Jagannathan *et al.* in 2012 [2]. The physical aspects including the depth and the distance are included in this research. The case study also takes in India. They assume that there is strong relationship between these two physical parameters. Even though this work does not need any satellite data, there is a difficulty to deal with the differences characteristics between the specific areas.

In 2013, Ravindran *et al.* [3] utilize the works as Jagannathan's did (distance-depth aspects), to see whether there is any impact on fishing patterns and lifestyle changes. Their works included fuzzy c-means clustering, and the data is clustered into two: summer and winter. In their conclusion, the productivity of marine fisheries may be affected by ocean current conditions. However, it was just based on the analysis generation of the distance and depth features, not the actual ocean current data.

B. *Classification Techniques*

Classification is a technique in data mining to group the data instances into their proper classes. There are two activities available included which are training and testing. Since classification needs training phase, then this activities can be categorized as the supervised learning. Classification is used to build the models from the given example (or historical data) which have been gathered.

There are various domain problems that classification can take a part ranging from information retrieval [9], geoscience and remote sensing [10], web technology [11] and so on.

Since the training and testing process need the important variable (this is called as "feature"), it is possible that exploration can be equipped with the classification techniques to see which feature or features give(s) the maximum or optimum result. There are some famous-basic classification algorithms available, including probabilistic based [11] (e.g. Naïve Bayes Classifier, Hidden Markov Models), information gain based (decision tree) [12], [13], [14], kernel based (SVM) [15], [16] and many others.

III. Methodology

Overall the methodology is shown as in Fig. 1 below. There are some steps included, which are data gathering, data integration, classification and the last are evaluation and analysis. The detailed step for classification and evaluation process is given in the Fig. 4 later.

A. *Data Gathering and Integration*

The experiments include some different sources of available data. First the chlorophyll and SST data (including fish catching data) are obtained from PTPN, Indonesia. This data contains the information of activity (including the total of fish catched) of tuna fishing during January 2000 until December 2004 in 15 S to 8 S and 105 E to120 E of latitude and longitude respectively. Fig. 2 below represents the study area that is used in this experiment. This data is then used as the ground truth.

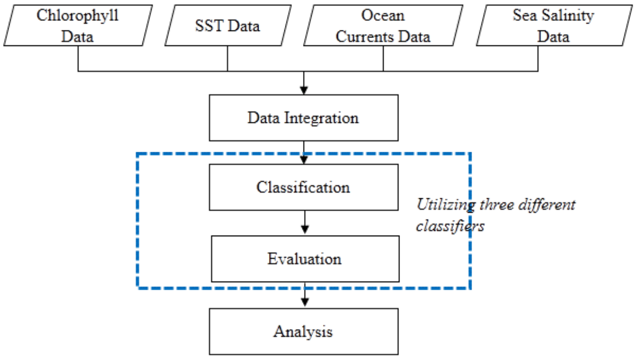


Fig. 1. Methodology for feature exploration.

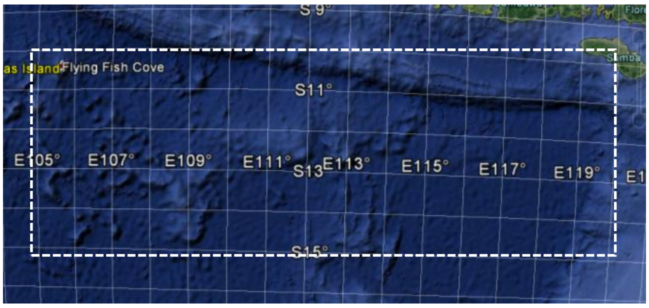


Fig. 2. Case study area (source: Google Earth, https://www.google.com/earth/).

A part from that, there is also use of the ocean current data. This data is obtained from Ocean Surface Current Analyses-National Oceanic and Atmospheric Administration (OSCAR-NOAA) [7]. It contains two different surface current wind direction, which are based on meridional (*u*) and zonal (*v*) wind direction. For simply, the representation of u and *v* direction can be described as in Fig. 3 below. The data is in netCDF format, containing some information parameters which are depth, time, latitude, longitude, *u* and *v* current, where depth value is stated by 15 m.

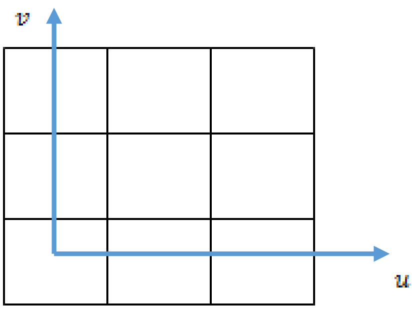


Fig. 3. Representation of direction for ocean current.

The other gathered data source is the salinity data. This data is obtained from WOA 2013, the National Oceanographic Data Center-(NODC, NOAA) [8].

The next step is the integration data. The analysis and integration of some different source of data into one are then performed. The process includes the way to match the data by comparing each attributes. Finally, the result is the data containing some parameters including time, spatial information (latitude-longitude), chlorophyll, SST, *u*, *v*, salinity, and the last: label (for potential or non-potential). The term of potential and non-potential fishing zone has been defined and clarified by the experts.

B. *Classification, Equipping 10-Folds Cross Validation*

The one-time-event classification process is described in

Fig. 4. In this step, there is a comparison of three different classifiers which are: Naïve Bayes (probabilistic based), Decision Tree (information gain based), and SVM Classifier (kernel based) to see the combination impact.

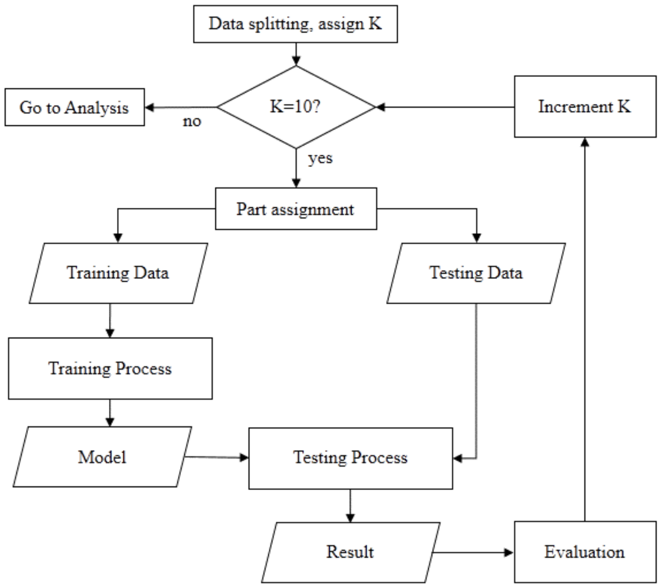


Fig. 4. Classification scenario.

In the classification process, there are four different feature combinations to compare:

- f1, consists of Chlorophyll and SST (for further, it is called as the "baseline" feature): as the work of [4]-[6].
- f2, consists of Chlorophyll, SST and Ocean Current
- f3, consists of Chlorophyll, SST and Salinity
- f4, consists of Chlorophyll, SST, Salinity and Ocean Current.

C. Evaluation

In order to evaluate the works, experiment utilizes the accuracy and kappa value. Given the confusion matrix as on Table I below.

TABLE I: CONFUSION MATRIX

		Detected	
		Potential	*non-Potential*
Actual	*Potential*	*a*	*b*
	non-Potential	*c*	*d*

The accuracy (Acc) value is obtained from the Equation 1,

$$Acc = \frac{a+d}{N} \times 100\% \qquad (1)$$

while the kappa (κ) value is calculated by Equation 2 below.

$$k = \frac{\frac{(a+b)}{N} - \left(\left(\frac{(a+b)}{N} \times \frac{(a+c)}{N} \right) + \left(\frac{(b+d)}{N} \times \frac{(c+d)}{N} \right) \right)}{1 - \left(\left(\frac{(a+b)}{N} \times \frac{(a+c)}{N} \right) + \left(\frac{(b+d)}{N} \times \frac{(c+d)}{N} \right) \right)} \qquad (2)$$

Here, *a* is true positive, *d* is true negative, *c* is false positive and *b* is false negative. The value of *N* is the sum of *a*, *b*, *c*, and *d*. This experiment, is performed in 1000 different times and then result is analyzed.

IV. RESULT AND DISCUSSION

A. First Combination f1: Chlorophyll and SST Features

First, experiment starts with the combination of the chlorophyll and SST feature only as the work in [4]-[6], as the basic benchmark of the whole experiments. Therefore by considering this "baseline" feature combination it can be figured out, in which position that the proposed work is. From the experiment result which is shown in Table II, it can clearly be seen that the accuracy values obtained from the baseline feature f1 are 57.44% in Naïve Bayes classifier, 58.91% in Decision Tree, and 56.74% in SVM with the kappa values below 0.2 for overall.

B. Second Combination f2: Chlorophyll, SST and Ocean Currents Features

In the next experiment, the ocean currents information is then added into the baseline feature combination. This feature includes both *u* and *v* current directions as discussed in part IVA. The results as in Table II show that this combination give better result compared to the combination utilizing Chlorophyll and SST only. By adding the ocean current, there is an improvement of accuracy values for about 8% in Naïve Bayes, 6% in Decision Tree, and 9% in SVM with 0.3054, 0.2831 and 0.3253 for kappa value, respectively.

C. Third Combination f3: Chlorophyll, SST and Salinity

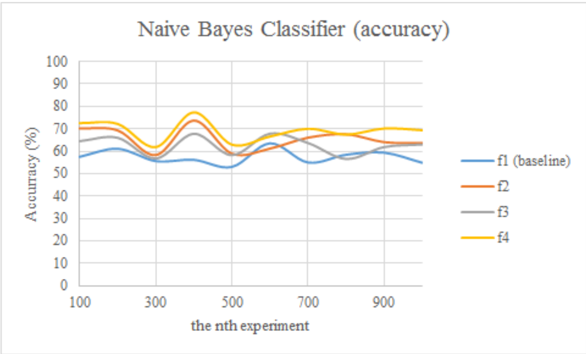
In the third experiment, the ocean current feature is excluded, and is substituted with the salinity, therefore it provides the f3 which are the combination of Chlorophyll, SST and Salinity. This f3 scenario is then getting compared with f1 feature combination scenario. As shown in Table 2, there is also improvement performance compared to the chlorophyll-SST feature (f1). In addition, compared to the chlorophyll-SST-Ocean (f2) current feature combination, the increasing values of f3 are still less than f2 for Naïve Bayes and SVM classifier, although in Decision Tree the increasing value is outperfomed.

D. Better Combination, f4: Chlorophyll, SST, Salinity, and Ocean Current

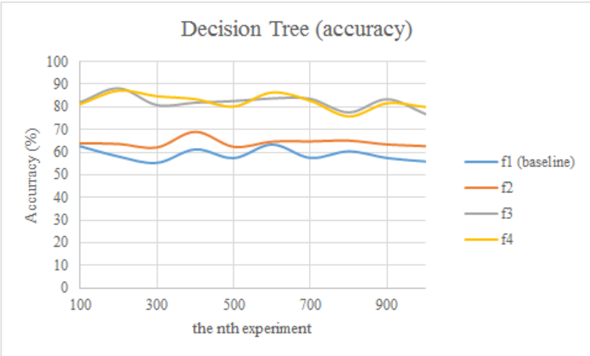
TABLE II: THE OVERALL EXPERIMENTAL RESULTS

Feature Combination	Naïve Bayes		Decision Tree		SVM	
	average Acc (%)	average *k*	average Acc (%)	average *k*	average Acc (%)	average *k*
f1: (Baseline) Chlorophyll, SST	57.44	0.1490	58.91	0.1780	56.74	0.1348
f2: Chlorophyll, SST, Ocean Current (u, v)	65.28	0.3054	64.16	0.2831	66.29	0.3253
f3: Chlorophyll, SST, Salinity	62.62	0.2522	82.07	0.6412	63.55	0.2712
f4: Chlorophyll, SST, Salinity, Ocean Current (u, v)	**69.03**	**0.3807**	**82.32**	**0.6462**	**68.30**	**0.3658**

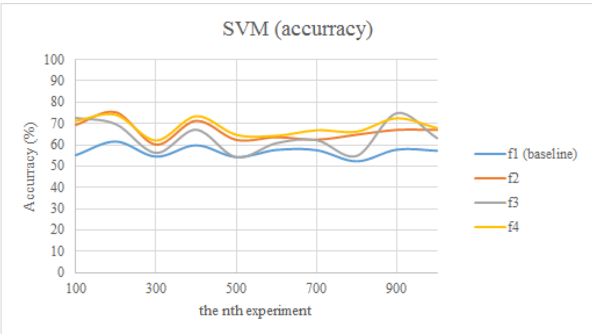
The last experiment shows the best result amongst the other combinations. It can be seen from Table II that this combination gives the highest accuracy values (69.03%-Na ïve Bayes, 82.32%-Dec Tree and 68.30%-SVM respectively). Also, the kappa values reach the outstanding achievement (0.3807-Na ïve-Bayes, 0.6462-Dec Tree and 0.3658-SVM respectively) compared to another combination.



(a)



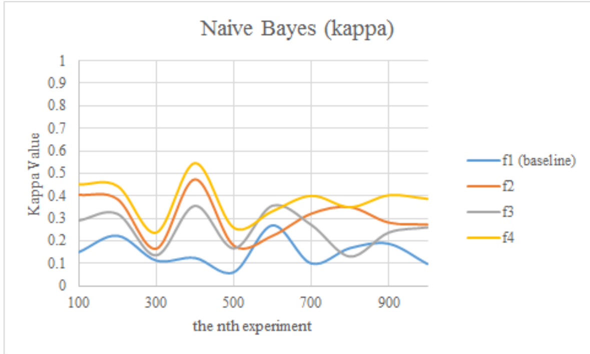
(b)



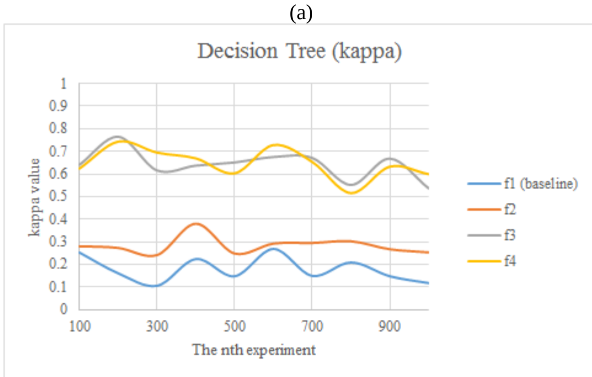
(c)

Fig. 5. Graphic of accuracy during the experiment (a) Na ïve Bayes (b) Decision Tree, and (c) SVM.

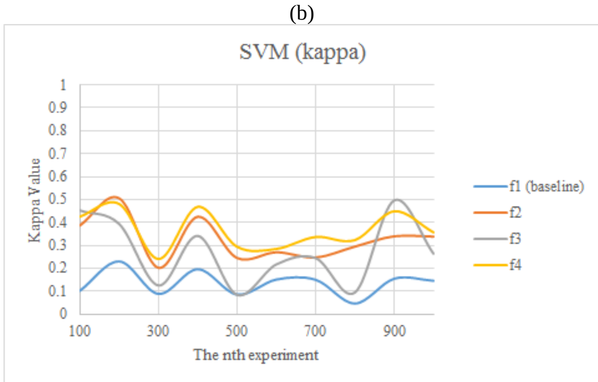
Fig. 5 above represents the accuracy values for the whole experiment. Either salinity or ocean currents feature can give better result in predicting a potential tuna fishing zone compared to the base line feature. If there is a use of the decision tree, it is enough to add the salinity only into f1, since the differences between f3 and f4 is so close. However, if Na ïve Bayes or SVM are utilized, it is better to add both ocean current and salinity parameter into f1. From the given result, it is also believed that f2 is not always better than f3 and vice versa. Those depend on the classifier used. The use of f2 is better in Na ïve Bayes and SVM but not in Decision Tree. Aligned with accuracy values, the kappa value for them during the experiment are shown in Fig. 6.



(a)



(b)



(c)

Fig. 6. Graphic of Kappa value during the experiment (a) Na ïve bayes (b) Decision tree, and (c) SVM.

V. Conclusion

The classification such as Na ïve Bayes, Decision Tree or SVM can be used either to predict the potential fishing zone or to explore the important feature for prediction. Instead of two common oceanography parameters (chlorophyll and SST), there are also two other important parameters which can improve the accuracy: ocean current and/or salinity parameter. Adding one or both ocean current and salinity feature can give the better prediction result.

References

[1] B. Mahyudin. (2012). Kebutuhan Teknologi Untuk Pengembangan Penangkapan Ikan. [Online]. Available: http://www.bbppi.info/index.php?pilih=news&mod=yes&aksi=lihat&id=164

[2] J. Sharath *et al*., "Potential fishing zone estimation by rough cluster predictions," in *Proc. The Fourth International Conference on Computational Intelligence, Modelling, and Simulation*, 2012, pp. 82-87.

[3] N. Rivandran, A. Samraj, and C. Kavitha, "Impact on fishing patterns and life cycle changes of Kanyakumari fisherman due to fading potential fishing zones," in *Proc. 2013 International Conference on Green High Performance Computing*, 2013, pp. 1-6.

[4] M. Shattri *et al.*, "Satellite fish forecasting in south china sea," presented at the 22nd Asian Conference on Remote Sensing, 2001.

[5] Solanki *et al.*, "Evaluation of remote-sensing-based potential fishing zones (PFZs) forecast methodology," *Continental Shelf Research*, vol. 25, pp. 2163-2173, 2005.

[6] P. R. C. Rahul *et al*, "Interlacing ocean model simulations and remotely sensed biophysical parameters to identify integrated potential fishing zones," *The IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 789-793, 2011.

[7] The US Government's Web Portal: Real Time Ocean Surface Current Data - National Oceanic and Atmospheric Administration (OSCAR-NOAA). [Online]. Available: http://www.oscar.noaa.gov/datadisplay/oscar_datadownload.php

[8] The US Government's Web Portal: National Oceanographic Data Center - National Oceanic and Atmospheric Administration (NODC-NOAA) for Salinity Data. [Online]. Available: http://www.nodc.noaa.gov/OC5/woa13/woa13data.html

[9] J. Shweta and B. Nigam, "Categorizing the document using multi class classification in data mining," in *Proc. International Conference on Computational Intelligence and Communication System*, 2011, pp. 251-255.

[10] Y. F. Zhong *et al*., "A hybrid object oriented conditional random field classification framework for high spatial resolution remote sensing imagery," *The IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7013-7073, 2014.

[11] A. Waleed *et al*., "Intelligent naïve bayes-based approaches for web proxy catching," *Knowledge Based Systems*, vol. 31, pp. 162-175, 2012.

[12] X. W. Li, "Application of decision tree classification method based on information entropy to web marketing," in *Proc. the Sixth International and Mechatronics Automation*, 2014, pp. 121-127.

[13] P. Narasimha *et al*., "A gini index based elegant decision tree classifier to predict precipitation," in *Proc. the 7th Asia Modelling Symposium*, pp. 46-54, 2013.

[14] P. N. Heru *et al*., "An automatic detection method for high density slums based on regularity pattern of housing using gabor filter and gini index," in *Proc. International Conference on Advanced Computer Science and Information System (ICACSIS)*, pp. 347-351, 2013.

[15] Ch. Xu, Y. Liu, and Z. Y. Li, "Algorithm research of face image gender classification based on 2-D gabor wavelet transform and SVM," in *Proc. International Symposium on Computer Science and Computation Technology*, 2008, pp. 312-315.

[16] Changchui *et al*., "An object based change detection approach using high-resolution remote sensing image and GIS data," in *Proc. International Conference on Image Analysis and Signal Processing*, 2010, pp. 565-569

Devi Fitrianah was born in Jakarta, Indonesia in 1978. She received her bachelor degree in computer science from Bina Nusantara University, Indonesia in 2000 and master degree in information technology from Universitas Indonesia in 2008. She is currently pursuing the Ph.D degree in computer science at Faculty of Computer Science Universitas Indonesia.

At present, she is a research assistant in the image processing, pattern recognition and GIS laboratory. Her research interests are image procesing, data mining, applied remote sensing and geographic information system.

Nursidik Heru Praptono was born in Banjarnegara, Central Java, Indonesia on May 8, 1988. He got his bachelor degree from the Dept. of Computer Science, Bogor Agricultural University (IPB), and the master degree, master of computer science from Faculty of Computer Science, Universitas Indonesia (FASILKOM-UI).

Heru Praptono is now with Faculty of Computer Science, Universitas Indonesia, as a junior lecturer, and a research assistant. The lab which he joins in is Laboratory of Pattern Recognition, Image Processing, and GIS (PRIPGIS). His research interests fall into image processing, machine learning, and remote sensing. In addition, he is also interested with software engineering and enterprise architecture applications.

Achmad Nizar Hidayanto was born in 1976. He received the B.S. degree in computer science from Universitas Indonesia, in 1999 and the M.S. degree in computer science from Universitas Indonesia, in 2002. He received his Ph.D. in computer science from Universitas Indonesia in 2008.

Dr. Hidayanto currently is appointed as a head of Information Systems/Information Technology Department, Faculty of Computer Science, Universitas Indonesia. He is the author of more than 40 articles written in journals and conferences. His research interests are related to information systems information technology, e-learning, information systems security, change management, distributed systems and information retrieval.

Aniati Murni Arymurthy is a professor at the Faculty of Computer Science, Universitas Indonesia. She was graduated from the Department of Electrical Engineering, Universitas Indonesia, Jakarta, Indonesia. She earned her master of science from the Department of Computer and Information Sciences, the Ohio State University (OSU), Columbus, Ohio, USA. She also holds Doktor from Universitas Indonesia and a sandwich program at the Laboratory for Pattern Recognition and Image Processing (PRIP Lab), Department of Computer Science, Michigan State University (MSU), East Lansing, Michigan, USA. She is the head of Laboratory for Pattern Recognition and Image Processing and GIS (PRIPGIS), Faculty of Computer Science, Universitas Indonesia. Her research interest includes the use of pattern recognition and image processing methods in several applications such as remote sensing, biomedical application, cultural artefak, agriculture and e-livestock.