Machine learning deployment for arms dynamics pattern recognition in Southeast Asia region

Zul Indra¹, Azhari Setiawan², Yessi Jusman³, Arisman Adnan⁴
¹Department of Informatics Engineering, Abdurrab University, Indonesia
²Department of International Relations, Abdurrab University, Indonesia
³Department of Electrical Engineering, Muhammadiyah Yogyakarta University, Indonesia
⁴Department of Mathematics, Riau University, Indonesia

ABSTRACT
Finding the most significant determinant variable of arms dynamic is highly required due to strategic policies formulations and power mapping for academics and policy makers. Machine learning is still new or under-discussed among the study of politics and international relations. Existing literature have much focus on using advanced quantitative methods by applying various types of regression analysis. This study analyzed the arms dynamic in Southeast Asia countries along with its some strategic partners such as United States, China, Russia, South Korea, and Japan by using ‘Decision Tree’ of machine learning algorithm. This study conducted a machine learning analysis on 55 variable items which is classified into 8 classes of variables videlicet defense budget, arms trade exports, arms trade imports, political posture, economic posture, security posture and defense priority, national capability, and direct contact. The results suggest three findings: (1) state who perceives maritime as strategic drivers and forces will seek more power for its maritime defense posture which is translated to defense budget, (2) big size countries tend to be an arms exporter country, and (3) state’s energy dependence often leads to a higher volume of arms transfers between countries.

Keywords: Arms dynamics, Decision tree, Machine learning, Pattern recognition, Preprocessing, Southeast Asia

1. INTRODUCTION
Machine learning is considered as one of the most essential subjects in scientific studies nowadays which the name firstly was coined in 1959 [1], [2]. It is a field of study for computer science and information technology which is focused on algorithms and statistical models. Machine learning algorithms are used in a wide variety of applications like email filtering and computer vision. It is closely related to computational statistics, which focuses on making predictions using computers [3]. Machine learning as the subset of ‘artificial intelligence’ is used to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead [4]. The patterns and inferences are utilized for various predictive analytics which is essential for social sciences, especially political science and international relations. There are many probabilities, type of interactions, and dynamics in the studies to formulate decision making. However, regarding the political analysis, there have been numerous and various researches that used advanced statistical methods but still rare in using machine learning as the main part of the research. Based
on international faculty survey by teaching, research & international policy (TRIP) on 2359 professors and individual who are affiliated with political science unit, professional or policy school and research institute worldwide showed the methods which is most employed by academics from various institutions in the study of international relations are: qualitative analysis (64.60%), policy analysis (16.30%), quantitative analysis (6.80%), other (5.20%), pure theory (3.30%), legal or ethical analysis (2.50%), formal modelling (0.80%), and experimental (0.50%) [5]. The quantitative analysis was the 3rd most employed in the study of international relations. Therefore, it can be concluded that the study for implementation of machine learning algorithm is still under-discussed in the study of international relations.

This argument is supported by facts obtained from the conducted literature studies related to research on quantitative politics and international relationship in the last 10 years [6]-[15]. During the last 10 years, most of the studies tried to find an inferential insight such as probability, causation, and effects by using certain models of regression. All of those previous researches have not applied pattern recognition for the analysis yet. Hence, this study is aimed to conduct research on the dynamics of the arms trade in Southeast Asia using machine learning algorithms in order to tackle the issue of this research gap. This study tries to explore the patterns of arms dynamics in Southeast Asia region countries focusing on several prominent variables, namely (1) economic posture, (2) political posture, (3) contiguities, and (4) defense posture and security prioritization. The analysis generated a model/pattern of arms trade dynamics by applying ‘Decision Tree’ algorithm.

2. RESEARCH METHOD

As stated in introduction section, this study is aimed to analyze the arms trade dynamics in Southeast Asia region by using the algorithms of machine learning. However, this study discovered that there are three main challenges to achieve this goal are (i) how to identify and collect the appropriate data, (ii) how to prepare the collected data to be able processed by machine learning and (iii) what is the most fit machine learning algorithm to be employed to analyze the data.

This study is divided into three stages namely data collection, data pre-processing and machine learning deployment to tackle these challenges. Based on the literature review that has been carried out, it was found that there are several categories of data that have to be obtained during data collection stage. The necessary data are then elaborated to find patterns of defense budget, the number of imports and exports of arms for countries in the Southeast Asia region by using machine learning algorithms. The machine learning algorithm chosen in this study is a decision tree algorithm since it can generate a model of decision rules that are easy to understand [16], [17]. In addition, this algorithm is very useful in the data exploration process [18], [19]. It is really useful algorithm to find the hidden relationships between input variables and a target variable. Hence, decision tree is an appropriate algorithm to get model from a set of data collection since it combines data exploration and modeling. To accommodate the use of this decision tree algorithm, this study proposes a novel algorithm for data pre-processing to enable the collected data can be processed by this chosen algorithm. The overall architecture of research method to conduct this study is illustrated Figure 1.
2.1. Data collection stage

As in general studies related to machine learning which is used for data analytics, the early stage for this study is data collection. As previously mentioned, the first challenge to conduct this study is how to collect the appropriate data for arms dynamics in Southeast Asia which is then analyzed by using machine learning algorithm. Based on the conducted literature study, it is revealed that there are two fundamental data regarding the arms dynamics, namely total exports and total imports of arms. The total exports and imports of arms trade along with the defense budget data will be treated as the output variables in this study. Furthermore, this study has identified 5 groups of input variables that greatly influence the arms dynamics namely political posture, economic posture, security posture, national capabilities, and direct contiguity. Data collection took place by observing data from 1960 to 2018 significant variables for all countries in the Southeast Asia region added several countries that have important role in ASEAN, namely Russia, Japan, China, and South Korea.

2.2. Data preprocessing stage

The next stage in this study is data pre-processing. As previously mentioned, the second challenge in this research is how to prepare the data that has been collected during the data collection stage to be able be processed with machine learning algorithms. The use of decision tree algorithm raises problems in terms of data suitability since the collected dataset is numerical data. In other hand, the data required by the decision tree algorithm to conduct data classification is categorical data. To tackle this issue, this study proposes a novel algorithm for data preprocessing so that numerical data can be analyzed using a decision tree. The main idea of this novel algorithm is to convert numerical data into categorical data based on averaging the data [20]. The proposed algorithm will calculate the average data for each variable then convert each data for that variable into several categories. The flowchart of proposed preprocessing for input and output variables can be seen in Figure 2 and 3.

![Flowchart of proposed preprocessing](image)

Figure 2. Preprocessing for output variable

2.3. Machine learning deployment stage

The final stage of this study is analyzing data using machine learning. Based on the conducted study literature, the most appropriate algorithm to achieve the goal for this study is the decision tree algorithm. Decision tree algorithm is really useful algorithm to find the hidden relationships between input variables and an output variable. In addition, this algorithm can generate a model of decision rules that are easy to understand. The main concept of a decision tree is to convert data into a decision tree and decision rules. This decision tree consists of several types of nodes, namely root nodes, internal nodes and leaf nodes. This node is a representation of the data variables in the dataset.
3. RESULTS AND DISCUSSION

3.1. Data collection and preprocessing

The success of the data collection stage is determined by the quality of the data collected to support the results of data analysis at the machine learning deployment stage. The results of preprocessing are determined from the success of the proposed preprocessing algorithm to convert numerical data in the dataset into categorical data. Throughout the data collection stages, this study succeeded in identifying the variables needed to analyze arms trade patterns. There are 30 input variables that have been identified and retrieved as the influential variables that affect the patterns of arms dynamics. The details of these variables can be seen in Table 1.

| Type                          | Category                          | Data Source       | Indicator                                                                 |
|-------------------------------|-----------------------------------|-------------------|---------------------------------------------------------------------------|
| Input Variable                |                                   |                   | Political System, Regime Durability, State Fragility Index, Effectiveness  |
|                               |                                   |                   | Score, Legitimacy Score, Security Effectiveness, Security Legitimacy,      |
|                               |                                   |                   | Political Effectiveness, Political Legitimacy, Economic Effectiveness,     |
|                               |                                   |                   | Economic Legitimacy, Social Effectiveness and Social Legitimacy            |
| Political Posture [21], [22]  | Polity IV Project                 |                   | Political Effectiveness, Political Legitimacy, Economic Effectiveness,     |
| Economic Posture [23], [24]   | Worldbank                         |                   | Economic Legitimacy, Social Effectiveness and Social Legitimacy            |
| Security Prioritization       | SIPRI                             |                   | Military Expenditure (%GDP), Military Expenditure (Constant$) and Military |
| and Defense Posture [23], [25]|                                   |                   | Personnel                                                                 |
| [26]                          |                                   |                   | GDP (Gross Domestic Product), GDP per Capita, GDP Growth and Inflation    |
| National Capabilities [27]    | Correlates of War Project         |                   | Military Expenditure (%GDP), Military Expenditure (Constant$) and Military |
| Direct Contiguity [28]        | Correlates of War Project         |                   | Personnel                                                                 |
| Defense Budget                | SIPRI                             |                   | Defense Budget                                                            |
| Total of Arms Export          | SIPRI                             |                   | Air defense system, Number of Arms Export for Aircraft, Armored vehicles, |
| Total of Arms Import          | SIPRI                             |                   | Artillery, Missiles, Naval weapons, Sensors, Ships and Others             |

3.2. Machine learning deployment

As explained in the method section, the algorithm chosen for data analysis is the decision tree algorithm. This algorithm produces an arms trading pattern in the form of decision rules by discovering the root variables that greatly influence the pattern of the decision rule for each output variable. However, this study has three output variables, namely defense budget, total arms export and total arms imports. Based on the results of data conversion at the preprocessing stage, the values of these three data have been set to be two output values, namely high and low. The first step to conduct data analysis by using a decision tree...
algorithm is by calculating the total entropy for all data in the dataset. Once the total entropy has been obtained, the next process is to determine the node of the decision tree. The explanation of the decision tree algorithm analysis results for each output variable.

### 3.2.1. Defense budget

The decision tree for pattern of the defense budget variable is presented in Figure 4. The pattern of arms dynamics for defense budget output variables can be determined based on the data presented in Figure 4. The decision rules for Figure 4 are described as shown in.

![Decision tree for defense budget goal](image)

**IF** (Contsea = None AND GDP Percapita = High Positive)
**OR** (Contsea = Low Positive AND Sfi = Low Positive) **OR** (Contsea = Low Positive AND Sfi = High Positive)
**OR** (Contsea = Low Positive AND Durability = Middle) **OR** (Contsea = Low Positive AND Durability = Low Positive)
**THEN** Defense Budget = High
**ELSE** Defense Budget = Low

### 3.2.2. Total of arms export

The decision rules for Figure 5. The decision tree for pattern of the arms export variable is presented in Figure 6.

![Decision tree for arms export goal](image)
3.2.3. Total of arms import

The decision tree for pattern of the arms export variable is presented in Figure 6.

The decision rules for Figure 6 are described as shown in.

\[
\text{IF (Pec1000Coalton = Middle AND Durability = High Positive) OR (Pec1000Coalton = Low Positive AND Contland = Low Positive) OR (Pec1000Coalton = Low Positive AND Contland = High Positive) OR (Pec1000Coalton = Low Positive AND MilexpGDP = Low Positive) OR (Pec1000Coalton = Low Positive AND MilexpGDP = High Positive) THEN Arm import = High ELSE Arm import = Low}
\]

4. RESULTS AND DISCUSSION

Based on the hypothesis in the introduction section, there are three objectives in the study namely: (1) defense budget, (2) arms trade export, and (3) arms trade import. Each goal has different decision tree of machine learning outcome. The following is a detailed discussion of each objective.

4.1. Defense budget

Based on the result, defense budget is most significantly determined by state’s direct sea contiguity (contsea). It suggests that the more state has direct sea contiguity; the state will spend more for its defense budget. Another study also suggests similar preposition that contiguity has positive effect or positively associated with military spending [29]. Regarding the direct sea contiguity, it has three derivative variables which are: (1) GDP per Capita which occur for group of state with no direct sea contiguity, (2) state fragility...
index which occur for group of state with low number of direct sea contiguity, and (3) state’s durability which occur for group of state with high number of direct sea. According to the result, study on defense budget related to sea contiguity is important for the study on maritime power and/or maritime security. A state who perceives maritime as strategic drivers and forces will seek more power for its maritime defense posture. A study by United States department of The Navy also made classification about states by its maritime power [30].

To conclude, a defense budget can be explained by identifying state’s spatial condition-in this context is direct sea contiguity—which is positively associated with GDP per Capita, State’s Fragility index, and state’s durability. GDP per Capita will significantly determine defense budget among state with no direct sea contiguity or a landlocked country. State’s fragility will significantly determine defense budget among state with few direct sea contiguities. State’s durability will significantly determine defense budget among state with many direct sea contiguities.

4.2. Arms export

Based on the result of machine learning decision tree model outcome, size of the state which is translated to its total population has a significant relation with its arms trade export. Big size population country tends to export more in the arms trade. This analysis result was corresponded with classical theory on economic and manufactured production that there is a general positive relationship between population size and density with export performance of the country [31], [32].

Regarding the decision tree model of arms export, total of population variable has one derive variable that is direct land contiguity. It suggests that if the population of the state is big and it has much direct contiguity by land, the state tend to get higher amount of arms trade export and vice versa. Existing literatures also have explained how land contiguity was associated with not only international conflicts but also military expenditures [33], [34]. Our analysis result suggests new insight that state’s number of direct land contiguity was positively associated with the amount of arms trade export. To conclude, arms export can be explained by identifying state’s total of population which is positively associated with state’s number of direct land contiguity.

4.3. Arms import

In term of arms trade import, our machine learning decision tree model outcome suggests primary energy consumption significantly determine the amount of state’s arms trade import. As the root variable, primary energy consumption has three derivative variables which are: (1) state durability which occur for group of state with low primary energy consumption, (2) direct land contiguity occur for group of state with middle primary energy consumption, and (3) military expenditure by percent of GDP occur for group of state with high primary energy consumption. A study by Vincenzo Bove on the relationship between energy consumption and arms trade stated that energy dependence leads to a higher volume of arms transfers between countries [35]. The study found that a country with a recent discovery of new oil fields will increase its import of weapons from oil-dependent economies by 56% [35]. In deeper discussion, oil might play an even larger role in influencing economic and political decisions [35]. To conclude, arms import can be explained by identifying state’s consumption on primary energy which is determined by State’s Durability, number of direct land contiguity, and amount of military expenditure by percent of GDP.

5. CONCLUSION

This research presents robust evidence that ‘decision tree’ of machine learning algorithm can classify and predict defense budget, arms trade export, and arms trade import in Southeast Asia countries. Each variable has its own root variables or its most significant factor/variable to determine those dependent variables. Defense budget was determined by state’s direct sea contiguity. Arms export was determined by state’s total population. Arms import was determined by primary energy consumption. To sum up, arms dynamics in the Southeast Asia countries-and some strategic countries partners as well-was highly associated with its direct sea contiguity, total population, and primary energy consumption. In spite of that, this research presents three theoretical implications. First, a state who perceives maritime as strategic drivers and forces will seek more power for its maritime defense posture which is translated to defense budget. That is why sea contiguity according to our analysis determine defense budget significantly. Second, big size countries tend to be an arms exporter country. Yet, it is not just about the size of the country, because according to the research analysis, it must be a country with many direct land contiguities. Third, relationship between energy consumption and arms trade stated that energy dependence leads to a higher volume of arms transfers between countries. Among the states with high primary energy consumption, a high level of military expenditure by percent of GDP will lead to a high level of arms trade import and vice versa. Interestingly, the combination of these three findings can be tools/supports for
academics and decision/policy makers to perceive and calculate the power projections and regional political landscape in order to formulate certain strategic policies and foreign policy. However, further research needs to be done in order to deepen and enrich this research topic, such as by applying random forest and unsupervised learning algorithms. The random forest algorithm is a decision tree development algorithm hence its implementation is expected to cover the limitation of the decision tree algorithm used in this study, which is that it can only classify two classes. In terms of implementing unsupervised learning algorithms, cluster of countries can be formed based on their weapon trading patterns by using an unsupervised learning algorithm. These country clusters can be an important information to conduct analysis for the polarization of these countries in their political and defense policies.

REFERENCES
[1] J. McCarthy and E. Feigenbaum, “In Memoriam Arthur Samuel: Pioneer in Machine Learning,” AI Magazine, vol. 11, no. 3, p. 11, 1990, doi: 10.1609/aimag.v11i3.840.
[2] A. L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers,” in IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210.
[3] J. H. Friedman, “Data Mining and Statistics: What’s the Connection?,” Computing science and statistics, vol. 29, no. 1, pp. 3-9, 1998.
[4] R. M. Neal, “Pattern Recognition and Machine Learning,” New York: Springer, vol. 49, no. 3, 2007.
[5] D. Maliniak, S. Peterson, R. Powers, and M. J. Tierney, “TRIP 2014 World Survey Report,” Williamsburg, 2014.
[6] P. E. Shea and P. Poast, “War and Default,” Journal of Conflict Resolution, vol. 62, no. 9, pp. 1876-1904, 2018, doi: 10.1177/0022002717707239.
[7] M. Digiuoppi and P. Poast, “Arms versus Democratic Allies,” British Journal of Political Science, vol. 48, no. 4, pp. 981-1003, 2018.
[8] C. Kaoutzianis, P. Poast, and J. Urpelainen, “Not letting ‘bad apples’ spoil the bunch: Democratization and strict international organization accession rules,” The Review of International Organizations, vol. 11, no. 4, pp. 399-418, 2016, doi: 10.1007/s11558-015-9237-5.
[9] P. Poast, M. J. Bommarito, and D. M. Katz, “The Electronic World Treaty Index: Collecting the Population of International Agreements in the 20th Century,” SSRN Electronic Journal, pp. 1-29, 2015, doi: 10.2139/ssrn.2652760.
[10] D. Wiens, P. Poast, and W. R. Clark, “The Political Resource Curse: An Empirical Re-evaluation,” Political Research Quarterly, vol. 67, no. 4, pp. 783-794, 2014.
[11] P. Poast, “Central Banks at War,” International Organization, vol. 69, no. 1, pp. 63-95, 2014, doi: 10.1017/S0020818314000265.
[12] P. Poast and J. Urpelainen, “Fit and Feasible: Why Democratizing States Form, not Join, International Organizations,” International Studies Quarterly, vol. 57, no. 4, pp. 831-841, 2013, doi: 10.1111/isqu.12031.
[13] W. R. Clark, S. N. Golder, and P. Poast, “Monetary institutions and the political survival of democratic leaders,” International Studies Quarterly, vol. 57, no. 3, pp. 556-567, 2013, doi: 10.1111/isqu.12013.
[14] J. Erasy, “Causal Inference with Observational Data Introduction to Matching,” in Analytics, Policy, and Governance, New Haven: Yale University Press, 2015, pp. 1-37, doi: 10.11177/1536867X0800700403.
[15] W. D. Berry et al., “Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?,” American Journal of Political Science, vol. 54, no. 1, pp. 248-266, 2010, doi: 10.1111/j.1540-5970.2009.00429.x.
[16] M. Somvanshi, P. Chavan, S. Tambade and S. V. Shinde, “A review of machine learning techniques using decision tree and support vector machine,” 2016 International Conference on Computing Communication Control and Automation (ICCCUBEA), 2016, pp. 1-7, doi: 10.1109/ ICCUBEA.2016.7860040.
[17] S. Gavankar and S. Sawarkar, "Decision Tree: Review of Techniques for Missing Values at Training, Testing and Compatibility," 2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS), 2015, pp. 122-126, doi: 10.1109/ AIMS.2015.29.
[18] G. G. Moisen, “Classification and Regression Trees,” in Jorgensen, Sven Erik; Fath, Brian D. (Editor-in-Chief). Encyclopedia of Ecology, CRC press, vol. 1, pp. 582-588, 2008.
[19] W. Loh, “Classification and regression trees,” Wiley interdisciplinary reviews: data mining and knowledge discovery, vol. 1, no. 1, pp. 14-23, 2011, doi: 10.1002/widm.8.
[20] Z. Indra, A. Setiawan, A. B. Saputra, and Y. Jusman, “Arms Dynamics Classification System for South East Asia Region by Using K-NN Algorithm,” Journal of Physics: Conference Series, vol. 1811, no. 1, p. 12095, 2021.
[21] I. V. Polity, “Polity IV Project: Political Regime Characteristics and Transitions, 1800-2013” Vienna, 2012.
[22] M. G. Marshall and B. R. Cole, “Global Report on Conflict, Governance and State Fragility 2008,” Foreign Policy Bulletin, vol. 8, no. 1, 3-21, 2008.
[23] PwC, “Global Defense Perspectives: Mapping Prioritization and Posture in a Challenging World,” London, U.K., 2015.
[24] K. M. Kadera and G. L. Sorokin, “Measuring national power,” International Interactions, vol. 30, no. 3, pp. 211-230, 2004, doi: 10.1080/03050620409429097.
[25] D. J. Singer, “Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985,” International Interactions, vol. 14, no. 145-132, 1987, doi: 10.1080/03050628808434695.
[26] D. J. Singer, S. Bremer, and J. Stuecky, “Capability Distribution, Uncertainty, and Major Power War, 1820-1965,” in Peace, War, and Numbers, B. Russet, Ed. Beverly Hills: SAGE Publications, 1972.
[27] S. M. Walt, “Alliance Formation and the Balance of World Order,” in *The Perils of Anarchy: Contemporary Realism and International Security*, M. Brown, S. Lynn-Jones, and S. Miller, Eds. Cambridge: MIT Press, 1995.
[28] D. M. Stinnet, J. Tir, P. Schafer, P. F. Diehl, and C. Gochman, “The Correlates of War Project Direct Contiguity Data, Version 3,” *Conflict Management and Peace Science*, vol. 19, no. 2, pp. 58-66, 2002, doi: 10.1177/073889420201900203.
[29] T. Böhme et al., “Forecasting military expenditure,” *Research and Politics*, vol. 1, no. 1, p. 205316801453590, 2014, doi: 10.1177/2053168014535909.
[30] U. S. Marine Corps, U.S. Navy, and U.S. Coast Guard, “Naval Operations Concept 2010: Implementing the Maritime Strategy,” United States. Department of the Navy; United States. Marine Corps; United States. Coast Guard, Washington D.C., 2010.
[31] T. K. Morrison, “The Effects of Population Size and Population Density on the Manufactured Exports of Developing Countries,” *Southern Economic Journal*, vol. 43, no. 3, p. 1368, Jan. 1977, doi: 10.2307/1057796.
[32] E. Nuroglu, “The Impact of Population on Bilateral Trade Flows in the case of OIC,” in *2nd International Conference on Islamic Economic Integration*, 2010.
[33] A. Q. Flores, “Alliances as contiguity in spatial models of military expenditures,” *Conflict Management and Peace Science*, vol. 28, no. 4, pp. 402-418, Sep. 2011, doi: 10.1177/0738894211413064.
[34] P. D. Senese, “Territory, contiguity, and international conflict: Assessing a new joint explanation,” *American Journal of Political Science*, vol. 49, no. 4, pp. 769-779, 2005, doi: 10.2307/3647696.
[35] V. Bove, “How the arms trade is used to secure access to oil,” *The Conversation*, 2018.

**BIOGRAPHIES OF AUTHORS**

**Zul Indra** is a currently as Lecturer at Universitas Abdurrah, Indonesia. He obtained his Master in Computer and Information Sciences from Universiti Teknologi PETRONAS. Presently pursuing his Ph.D. in Computer Science. His research interest includes software development, data mining, and machine learning.

**Azhari Setiawan** is a currently as Lecturer at Universitas Abdurrah, Indonesia. He obtained his Master in International Relations from University of Indonesia. His research interest includes International Relations, ASEAN Security Community, strategic studies, and international security.

**Yessi Jusman** is a currently as Lecturer at Universitas Muhamadhiy Yogyakarta, Indonesia. She obtained her PhD from Universiti Malaya. Her research interest includes intelligent system, data processing, and machine learning.

**Arisman Adnan** is a currently as Lecturer at Universitas Riau, Indonesia. He obtained a PhD in Statistics, University of Newcastle UK. He is working on statistical modelling, ordinal data, and statistical computing.