Economic Classification and Prediction of Southeast Asian Countries During the COVID-19 Pandemic
(Analysis with Orange Data Mining)

Ni Nyoman Reni Suasih[1(✉)], Ida Ayu Nyoman Saskara[1], Putu Yudy Wijaya[2], and I. Gede Putu Kawiana[2]

[1] Faculty of Economics and Business, Udayana University, Denpasar, Indonesia
renisuasih@unud.ac.id

[2] Faculty of Economics, Business, and Tourism, Hindu Indonesia University, Denpasar, Indonesia

Abstract. World Bank data shows that during 2020 the economies of countries in the world (including Southeast Asian countries) contracted, and were in a recession. However, there are still countries whose economies are still growing during the COVID-19 pandemic. Economists are still debating the predicate of a recession, because even though it has contracted for two quarters, it doesn't necessarily mean a country is considered a recession. Therefore, this study analyzes the economic classification of Southeast Asian countries during the COVID-19 pandemic (growth or recession), as well as the governance of their economic attributes (export components, agriculture, industry, and manufacturing). This study uses analytical techniques with orange data mining. The results show that based on the Naïve-Bayes algorithm, there are three countries whose economies are classified as experiencing a recession, while the economies of seven other countries are classified as growing. The attributes with the highest rank scores are agriculture, exports, and COVID-19 cases.

Keywords: Economic growth · COVID-19 pandemic · Export · Agriculture · Industry · Manufacture

1 Introduction

The COVID-19 (Corona Virus Diseases-2019) pandemic, which began in Wuhan (China) at the end of 2019, has spread massively to almost all parts of the world [1]. Governments in various countries have imposed social restrictions and community activities (including economic activities) to prevent the further spread of COVID-19 [2]. Transactions of goods and services also experienced a slowdown due to government policies in various countries to carry out various restrictions. As a result, the pace of world GDP contracted for the first time in a decade. As noted by the World Bank that in 2020, the average pace of GDP fell drastically until it contracted by −3.40 (Fig. 1).

Discussions about global economic conditions lead to the word "recession". Until now, there is no official definition of recession, either in the scientific literature or in
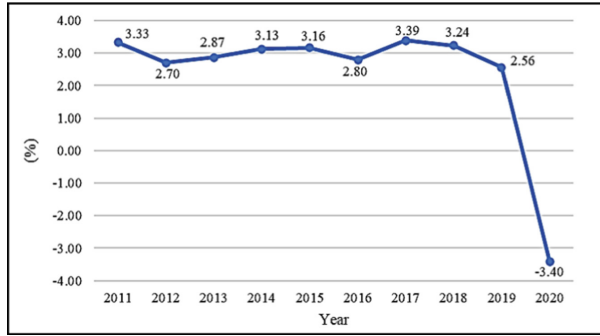
Fig. 1. World GDP growth (Annual %) (data.worldbank.org (processed)).

the practice of international organizations or statistical offices [3]. Indeed, most experts and analysts use the practical definition of a recession, namely the decline in real Gross Domestic Product (GDP) for two consecutive quarters. Although this definition is practical in understanding a recession, the focus on GDP alone is still considered narrow, and it is considered better to also consider a wider range of economic measures (such as employment, income, sales, and industrial production) in determining whether a country is in recession [4].

Exports are also one of the components of the macro-economy which is suspected to play an important role in the economy of a country. The previous research results from [5] with Granger causality effect state that exports have an impact on GDP growth. Especially during the COVID-19 pandemic where export restrictions (especially food) gave a shock not only to exporting countries, but also to importing countries [6].

Some economists consider the agricultural sector to be an important part of economic growth [7]. An example is the research result of [8] that agricultural transformation is closely related to economic growth in Indonesia. During the COVID-19 pandemic, the agricultural sector also experienced a decline, but it was still lower than the impact experienced by the non-agricultural sector [9], so it can be said that the agricultural sector is important in supporting the country's economic conditions during the pandemic.

Industrial development is also considered as a driver of economic growth. Like the research results of [10] which found that there was a significant relationship between industrial development and Senegal's economic growth, where industrialization would be very helpful in encouraging economic growth. However, during the pandemic, labor-intensive industries were greatly affected as a result of morbidity and containment policies [6].

Most of the Southeast Asian countries are still in the middle-income stage, so that the manufacturing sector is still the driving force of economic growth. It was also explained that the manufacturing sector has three important characteristics, namely attracting services, accelerating the rate of technological accumulation, and increasing the use of human capital and economic institutions [11].

The purpose of this study is to analyze the economic classification and predictions of Southeast Asian countries during the COVID-19 pandemic. In addition, this study also presents a ranking of the attributes of economic components that are assumed to be

related to the economy of Southeast Asian countries during the COVID-19 pandemic, namely exports, agriculture, industry, and manufacturing and of course the COVID-19 cases.

2   Literature Review

2.1   Economic Growth and Recession

Economic growth is an increase in the production of all people in a country from year to year [12]. Economic growth is often proxied by GDP, and is defined as an increase in output from the economic capacity to produce goods and services needed to increase the welfare of the people of a country [13]. In general, there are three approaches to measuring economic growth, namely: nominal measurement of growth, real output growth rate, and growth measured in per capita values [10].

The National Bureau of Economic Research (in Mazurek, 2012) defines a recession as a significant decline in economic activity that is evenly distributed in a country, lasting more than a few months, can be monitored on real GDP growth, real per capita income, employment, industrial production, also wholesale-retail sales. This understanding is considered to have weaknesses because it is only qualitative and does not present quantitative limitations [14].

Business cycle analysts generally use the concept of recession to refer to a phase of a weak economy, whose duration, depth, and spread exceed the usual limits [3]. Three key dimensions of recession or what is known as the "three Ds", namely duration, depth, and diffusion [15]. So, it can be said that in a recession there is a substantial decline in economic activity and has a broad impact and some of its impacts continue.

2.2   Data Mining

Data mining is a concept used to find hidden added value in databases. Data mining is a semi-automatic process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify potential and useful knowledge information stored in large databases [16]. Data mining is also a stage in knowledge discovery in databases which consists of data cleaning, data integration, data selection, data transformation, mining process, pattern evaluation, and knowledge presentation processes [17]. In simpler terms, a data mining process framework in three stages, namely data collection, data transformation, and data analysis [18, 19] as shown in Fig. 2.

Figure 2 shows the flow of information in the data mining process as an iterative process. The results of the evaluation of knowledge generated by data mining can lead to the need for more complete knowledge, improvement of data sets (datasets) or changes to the system [19].
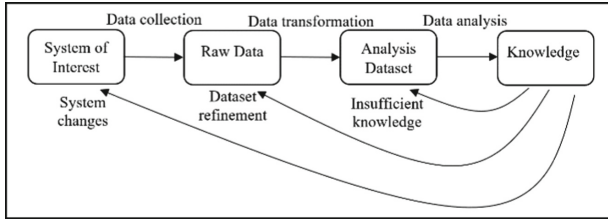
Fig. 2. Data mining process framework [18].

3 Methods

Based on the research objectives, a study can be categorized as exploratory, descriptive, or explanatory research [20]. Exploratory research is generally an early stage research that aims to explore, while descriptive research aims to describe, and explanatory research generally aims to test and can be a continuation of exploratory and descriptive research [21, 22]. Based on these criteria, this research can be classified as exploratory research.

The variables to be analyzed are as follows:

1) Economic condition, namely the country's economic condition, which in this study is presented in the category of growth or recession.
2) The severity of COVID-19, which in this study was proxied by the number of positive cases of COVID-19 in Southeast Asian countries.
3) Export, which in this study is proxied by the annual growth of exports of Southeast Asian countries (%).
4) Agriculture, which in this study is proxied by the annual growth of agriculture of Southeast Asian countries (%).
5) Industry, which in this study is proxied by the annual growth of industry of Southeast Asian countries (%).
6) Manufacturing, which in this study is proxied by the annual growth of manufacture of Southeast Asian countries (%).

The data in this study is secondary data from the world bank (data.worldbank.org) and tradingeconomics.com during the COVID-19 pandemic (2020 period). Furthermore, the data was analyzed by orange data mining using the Naïve Bayes algorithm which is suitable for classification and prediction of data according to the research objectives.

Naïve Bayes is a fast and simple probabilistic based on Bayes' theorem with the assumption that the features can stand alone. Naïve Bayes can only be used to classify data [23]. Naïve Bayes is an algorithm that is used to find data that has a value with the highest probability. So that a group of data can be sorted or grouped according to similarities and also appropriate. Naïve Bayes is also a method in Data Mining which is included in the ten most popular classification methods among other algorithms [24].

Regarding the measurement of performance models, here are some performance metrics that are commonly and often used [25] and can be obtained through the Naïve Bayes algorithm using the Orange application.

1) **Accuracy**. Accuracy can be illustrated how accurate the model is to classify correctly. Thus, prediction accuracy is the ratio of the number of true positives and true negatives to the overall data. In other words, accuracy is the degree of closeness of the predicted value to the actual value.
2) **Precision**. Precision can be described as the level of accuracy of the requested data with the prediction results provided by the model. Then it can be interpreted that the precision is the ratio of the positive correct prediction divided by the overall positive predicted result. In other words, of all positive classes that have been correctly predicted, how many data are truly positive.
3) **Recall**. Recall or sensitivity can be described as the success of the model in obtaining information. Therefore, it can be said that the recall is the ratio of true positive predictions divided by the total number of true positive data.
4) **Specificity**. Specificity can be described by the accuracy of predicting the negative divided by the total negative data.
5) **F1 score**. The F1 score can be said as a weighted comparison of the average precision and recall. F1 scores can be said to be the best if there is some kind of balance between precision and recall in the system.

Through Naïve Bayes analysis, a confusion matrix (also called classification matrix) can also be presented, namely the matrix which is used to assess the prediction accuracy [26], with a general structure as shown in Table 1.

Correlation (r) is an effect size, whose value is in the range of -1 (negative correlation) to 1 (positive correlation). The closer value to -1 or 1 means that the stronger correlation. Verbally, the absolute value of the correlation can be categorized as Table 2.

Table 1. General structure of the confusion matrix [26]

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

Table 2. General structure of the confusion matrix [27]

No	Value of Correlation	Interpretation
1	0.00 to 0.29	Little if any correlation
2	0.30 to 0.49	Low correlation
3	0.50 to 0.69	Moderate correlation
4	0.70 to 0.89	High correlation
5	0.90 to 1.00	Very high correlation

To answer the objectives related to attribute ranking, data mining analysis with Orange through the rank widget was carried out. There are several algorithms used in rank analysis, which are as follows [28].

1) Information Gain, is an algorithm that principally reduces the variance of attributes so that it shows closeness to the classification results of the object being tested.
2) Information Gain Ratio, is a refinement of the information gain algorithm, which means solving the weaknesses of the previous algorithm, where bias problems often arise.
3) Gini Impurity Index, is an algorithm that focuses on discriminants, where the attribute that makes the most discrimination will get a higher value and is considered better.
4) Chi-Square Distribution, is a non-parametric technique used to determine whether the observed frequency distribution is different from the expected theoretical frequency.
5) Relief, is an algorithm that does not depend on heuristic methods, and is based on a noise-resistant feature selection algorithm.
6) Fast Correlation-Based Filter (FCBF), is a correlation-based algorithm (Correlation-Based Filter/CBF) and its fast.

4 Results and Discussion

4.1 Overview of Southeast Asian Countries Economics

The economic growth of Southeast Asian countries during the COVID-19 pandemic (in 2020) was indeed mostly contracted, but there were several countries whose economies continued to grow positively. Figure 1 presents a comparison of the economic growth rates of Southeast Asian countries.

As Fig. 3 shows, there are only three countries whose economies continue to grow, namely Brunei Darussalam, Laos, and Vietnam. Meanwhile, seven other countries experienced a contraction in their economic growth.
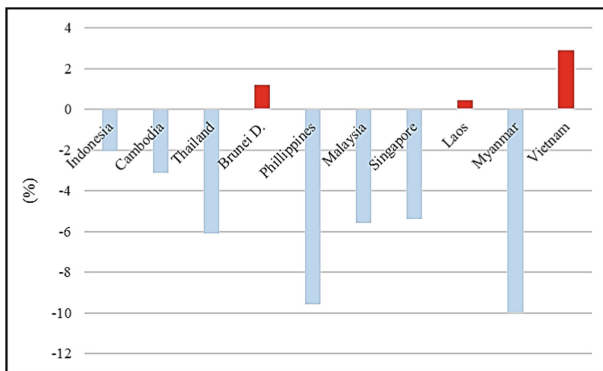

Fig. 3. Comparison of Southeast Asian Countries GDP growth rate (in 2020) (data.worldbank.org (processed)).
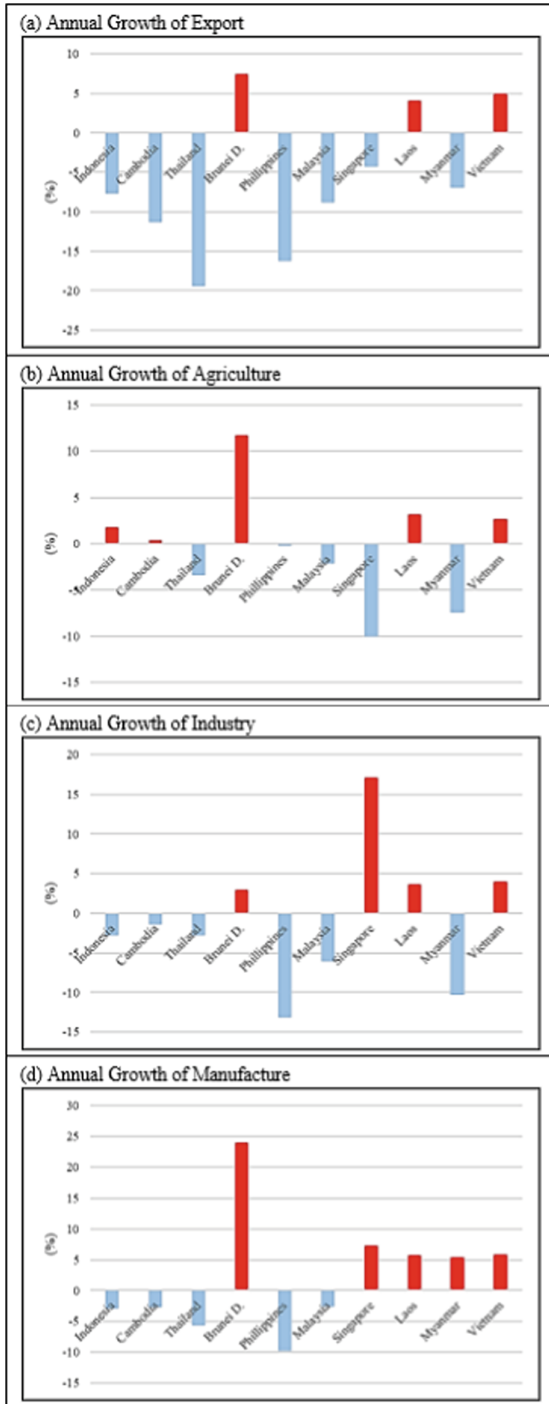
Fig. 4. Comparison of the growth rate of selected economic attributes (Year 2020) (data.worldbank.org and tradingeconomics.com (processed)).
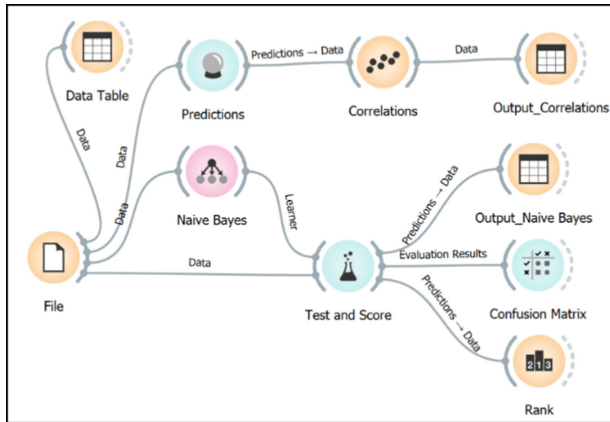
Fig. 5. Research workflow (Widgets).

In addition, an overview of several components of the economy that are used as attributes, namely the growth rate of exports, agriculture, industry, and manufacturing is presented in Fig. 4(a)-(d).

Figure 4 also shows that several countries are still able to maintain a positive growth rate on each attribute, especially Brunei Darussalam, Laos, and Vietnam.

4.2 Data Mining Analysis with Orange

Orange is a data mining tool that is very useful for visual programming and explorative data analysis. Orange is also known as Python, and Orange has multiple components known as widgets [29]. The Orange widgets or workflows in this study are as presented in Fig. 5.

4.3 Naïve Bayes Analysis (Southeast Asian Countries Economic Classification)

The confusion matrix can be used to determine the location of the misclassification of the Naïve Bayes model. Table 3 describes the confusion matrix obtained from the evaluation results of the Naïve Bayes model.

Table 3 shows that there are seven countries whose economies are predicted to grow, and three countries whose economies are predicted to be in recession. However, of the seven countries whose economies are predicted to grow, there are four countries that have miscalculated. The country is supposed to be in recession, but the model predicts growth.

The results of the Naïve Bayes algorithm analysis on this model show the output of several components which can be described as Table 4. As the output in Table 4 shows, it is clear that all measurements of the model's performance show the lowest value of 0.6, which means the model is good.

The Naïve Bayes output also presents recommendations for the economic classification of Southeast Asian countries (Table 5).

Table 3. Confusion matrix - Naive Bayes

	Predicted Negative	Predicted Positive	Sum
Actual Negative	3	0	3
Actual Positive	4	3	7
Sum	7	3	10

Table 4. Evaluation of Naïve Bayes model results

Performance Metrics	Results	Evaluation
AUC	1.000	AUC is used to represent probabilities. The range of AUC values is 0–1, so an AUC value of 1 means that the model is excellence.
Accuracy (CA)	0.600	CA functions for the accuracy of the selected dataset. The accuracy of this model is 60%, which means that the level of closeness of the predicted value to the actual value is 60%.
Skor F1	0.600	F1 is the comparison between recall and precision, where the result is 60%.
Precision	0.829	The precision value is 82.9%, which means that the country's economy is actually growing from what is predicted to grow by 0.829.
Recall	0.600	The 60% recall shows the ratio of predictions for the economy to grow compared to the overall economic data that actually grew.

Based on the analysis using the Naïve Bayes algorithm, Table 5 presents the classification of countries whose economy is in growth or recession during the COVID-19 pandemic with an annual growth approach to the components of exports, agriculture, industry, and manufacturing as well as targets for the classification of economic growth. The result is that countries with a contracted economy are not necessarily classified as recession, and vice versa, there are three countries classified as experiencing a recession (Indonesia, Cambodia, and the Philippines), and seven countries classified as having their economies growing.

4.4 Correlation Between Economic Components and Their Ranking

Furthermore, to complete the data mining analysis, predictions of correlations between the components of the Southeast Asian countries economy have been determined previously. The output of the analysis is presented in Table 6.

Table 5. Classification based on Naïve Bayes output

No	Country	Classification of Economic (GDP) Growth (Year 2020)	Naïve-Bayes Classification
1	Indonesia	Negative	Recession
2	Cambodia	Negative	Recession
3	Philippines	Negative	Recession
4	Singapore	Negative	Growth
5	Vietnam	Positive	Growth
6	Thailand	Negative	Growth
7	Brunei Darussalam	Positive	Growth
8	Malaysia	Negative	Growth
9	Laos	Positive	Growth
10	Myanmar	Negative	Growth

Source: data.worldbank.org (processed)

Table 6. Pearson correlation between attributes

No	Correlation of Attribute	Pearson Correlation	Correlation Interpretation
1	Export – Manufacture	+0.844	High correlation
2	Agriculture – Export	+0.531	Moderate correlation
3	Export – Industry	+0.522	Moderate correlation
4	Industry – Manufacture	+0.516	Moderate correlation
5	COVID -19 Cases – Manufacture	−0.450	Low correlation
6	Agriculture – Manufacture	+0.449	Low correlation
7	COVID-19 Cases – Industry	−0.408	Low correlation
8	COVID-19 Cases – Export	−0.332	Low correlation
9	Agriculture – Industry	−0.046	Little if any correlation
10	Agriculture – COVID-19 Cases	−0.002	Little if any correlation

Based on Table 6, it can be seen that the correlation with the highest value is the correlation between export growth and manufacturing, which is categorized as positive and has a high correlation. Then followed by the correlation between agricultural and export growth as well as export and industrial growth which is positive and moderate correlation.

Table 7. Rank-scores by attribute selection algorithm

No	Attributes	Attribute Selection Algorithms					
		Inf. Gain	Gain Ratio	Gini	Chi-square	Relief	FCBF
1	Agriculture	**0.881**	**0.477**	**0.420**	**5.250**	**0.074**	**1.618**
2	Export	**0.881**	**0.477**	**0.420**	**5.250**	**0.185**	0.000
3	COVID-19 Cases	**0.606**	**0.307**	**0.287**	**4.298**	−0.042	**0.738**
4	Manufacture	0.406	0.206	0.187	3.048	**0.095**	0.000
5	Industry	0.406	0.206	0.187	3.048	0.010	0.000

[*]The numbers marked in bold are the highest three values in the sequence for each algorithm.

This study also aims to analyze the ranking of the components of the Southeast Asian economy based on several algorithms as previously described, namely the information gain algorithm, information gain ratio, Gini impurity index, chi-square distribution, relief, and Fast Correlation-Based Filter (FCBF). The following Table 7 presents the rank of the economic components as research attributes in each algorithm.

The output in Table 7 shows that the attributes of agriculture, exports, and COVID-19 cases received the highest scores by four algorithms (information gain, Gini ratio, Gini impurity index, and chi-square). While the relief algorithm places export, followed by manufacture and agriculture as the attribute with the highest score. Unlike other algorithms, FCBF places COVID-19 cases and agriculture as attributes with the highest values. Therefore, in general it can be said that the highest dominant attribute rankings are agriculture, export, and COVID-19 Cases.

5   Conclusion

Based on the results of data mining analysis with Orange (Nave-Bayes algorithm) it can be concluded that the classification of a country's economy is not always classified according to its economic growth rate. If it is associated with the attributes of economic components, it can be classified that as many as three Southeast Asian countries experienced a recession, while the economies of seven other countries could still grow.

Furthermore, the results of the analysis on several algorithms (information gain, information gain ratio, Gini impurity index, chi-square distribution, relief, and FCBF) show that the attribute components with the highest rank scores are agriculture, exports, and COVID-19 cases. Meanwhile, when viewed from the correlation between attributes, the absolute value of the correlation with the highest value in a row is the correlation between export and manufacture, agriculture and export, and export and industry. This shows that the export component has a significant correlation with almost all other attributes.

References

1. Suasih, N. N. R., Saskara, I. A. N., & Wijaya, P. Y. (2021). Community mobilities in Indonesia during the COVID-19 pandemic and the case of Bali. *Regional perspectives of COVID-19 in Indonesia: IRSA book series on regional development, no.19* (pp. 71–87).

2. Wijaya, P. Y., & Suasih, N. N. R. (2020). Peran local celebrity endorsement pada iklan di media sosial terhadap pembelian produk kuliner di provinsi bali pada masa stay home akibat pandemi covid-19. *Jurnal Bisnis Strategi*, *29*(2), 119–133. https://doi.org/10.14710/jbs.29.2.119-133

3. Abberger, K., & Nierhaus, W. (2008). How to define a recession. *CESifo Forum, 9*(4), 74–76.

4. Claessens, S., & Kose, M. A. (2009). Back to basics: What is a recession? *Finance & Development, 2009*, 52–53.

5. Stait, A. (2005). Are exports the engine of economic growth? An application of cointegration and causality analysis for Egypt, 1977–2003. *Economic Research Working Paper, no. 76*.

6. Espitia, N. R., & Ruta, M. (2020). Covid-19 and food protectionism: the impact of the pandemic and export restrictrions on world food markets. *Policy Research Working Paper, no. 9253, World Bank Group: Macroeconomics.* Trade and Investiment Global Practise.

7. Jatuporn, C., Chien, L.-H., Sukprasert, P., & Thaipakdee, S. (2011). Does a long-run relationship exist between agriculture and economic growth in Thailand? *International Journal of Economics and Finance, 3*(3), 227–233.

8. Askar, M. W. (2015). Agricultural transformation, economic growth, and poverty in Indonesia. *Jurnal Ilmiah Administrasi Publik, 1*(1), 17–21.

9. Beckman, J., & Countryman, A. M. (2021). The importance of agriculture in the economy: Impacts from COVID-19. *American Journal of Agricultural Economics, 103*(5), 1595–1611. https://doi.org/10.1111/ajae.12212

10. Ndiaya, C., & Lv, K. (2018). Role of industrialization on economic growth: The experience of senegal (1960-2017). *American Journal of Industrial and Business Management, 8*, 2072–2085. https://doi.org/10.4236/ajibm.2018.810137

11. Su, D., & Yao, Y. (2016). Manufacturing as the key engine of economic growth for middle income economies. *ADBI working paper series, no. 573*, Tokyo: Asian Development Bank Institute.

12. Nuraini, I., & Hariyani, H. F. (2019). Quality economic growth as an indicator of economic development. *Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi dan Pembangunan, 20*(1), 80–86. https://doi.org/10.23917/jep.v20i1.7104

13. Balami, D. H. (2006). Macroeconomic theory and practice. *Salawe prints, Off Leventies, Wulari, Maiduguri.*

14. Mazurek, J. (2012). On some issues concerning definition of an economic recession. *MPRA paper, no. 43381.* https://mpra.ub.uni-muenchen.de/43381/.

15. Fiedler, J. L. (1990). Increasing reliance on user fees as a response to public health financing crises: A case study of El Salvador.

16. Wicaksono, A. E. (2016). Implementasi Data Mining Dalam Pengelompokan Data Peserta Didik di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma Berbasis k-Means (Studi Kasus SMAN 16 Bekasi). *Jurnal Teknologi Rekayasa, 21*(3), 206–216.

17. Marisa, F. (2013). Educational data mining (Konsep dan Penerapan). *Jurnal Teknologi Informasi, 4*(2), 90–97.

18. Nilakant, K. (2004). Application of data mining in constraint based intelligent tutoring system.

19. Ayub, M. (2007). Proses data mining dalam Sistem Pembelajaran Berbantuan Komputer. *Jurnal Sistem Informasi, 2*(1), 21–30.

20. Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students, 5th edn.* Pearson Education.
21. Mudjiyanto, B. (2018). Tipe Penelitian Eksploratif Komunikasi. *Jurnal Studi Komunikasi dan Media, 22*(1), 65–74.
22. Makri, C., & Neely, A. (2021). Grounded theory: A guide for exploratory studies in management research. *International Journal of Qualitative Methods, 20*, 1–14. https://doi.org/10.1177/16094069211013654
23. Oktanisa, & Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 5*(5), 567–576. https://doi.org/10.25126/jtiik.201743299
24. Zaeki, M. (2020). *Analisa topik pendidikan dalam Al-Quran dengan Pendekatan text mining.* Universitas Islam Negeri Ar-Raniry Banda Aceh.
25. Derisma. (2020). Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining. *Journal of Applied Informatics and Computing (JAIC), 4*(1), 84–88.
26. Kantardzic, M. (2020). *Data mining: concepts, models, methods, and algorithms.* Willey.
27. Asuero, A. G., Sayago, A., & Gonzalez, A. G. (2006). The correlation coefficient: An overview. *Critical Reviews in Analytical Chemistry, 36*, 41–59. https://doi.org/10.1080/10408340500526766
28. Coban, A., & Tarimer, I. (2018). Feature selection for gender classification in TUIK life satisfaction survey.
29. Kukasvadiya, M. S., & Divecha, N. H. (2017). Analysis of data using data mining tool orange. *International Journal of Engineering Development and Research, 5*(2), 1836–1840.