Bulletin of Electrical Engineering and Informatics
Vol. 11, No. 5, October 2022, pp. 2764~2772
ISSN: 2302-9285, DOI: 10.11591/eei.v11i5.4082

Prediction measuring local coffee production and marketing relationships coffee with big data analysis support

Anita Sindar Ros Maryana Sinaga[1], Ricky Eka Putra[2], Abba Suganda Girsang[3]
[1]Department of Information Technology, STMIK Pelita Nusantara, Medan, Indonesia
[2]Department of Informatics, Faculty of Engineering, Surabaya State University, Surabaya, Indonesia
[3]Department of Computer Science, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received May 28, 2022
Revised Jun 23, 2022
Accepted Jul 13, 2022

Keywords:

Big data analysis
Coffee production
Machine learning
Modeling
Prediction

ABSTRACT

Following the increasing enthusiasm of the coffee market in Indonesia, a machine learning model is developed to study the relationship between coffee producers, consumers, production, and the market. Machine learning work flow is constructed in various stages; explore, develop, and validate the models. In this research, the building model predicts the production and market of late departure coffee based on labeled and unlabeled variables. The best predictions from the trained type of model algorithms of machine learning like tree accuracy of 85.7%, support vector machine (SVM) accuracy of 82.9%, and k-nearest neighbors, the accuracy of 82.9%, which produce three categories, namely, high production of 2 provinces, medium production of 21 provinces, and low production of 11 provinces. The accuracy classification is supported by the AUC value obtained for a high class, a medium class, and a low class. In addition, local coffee marketing modeling used in logistic regression was found with an accuracy of 88.9%, aiming to classify coffee interests between Arabica coffee and Robusta coffee. We found that the AUC value logistic regression for arabica coffee is about 0.94, while for Robusta is 0.92. The analysis of the classification modeling results shows a high level of accuracy of 93.0%.

Corresponding Author:

Anita Sindar Ros Maryana Sinaga
Department of Information Technology, STMIK Pelita Nusantara
Iskandar Muda Street, No. 1, Medan, 20154 Indonesia
Email: haito_ita@yahoo.com

1. INTRODUCTION

Director of processing and marketing of plantation products, Ministry of Agriculture of Indonesia, stated that Indonesia's coffee production in 2020 is 753,491 tons from an area of 1.2 million hectares of land or reaching up to 806 kg/ha. Meanwhile, in 2019, the production is about 752,511 tons from an area of 1,245,358 ha or reaching up to 803 kg/ha. Data from the International Coffee Organization (ICO) shows that Indonesian coffee consumption increases yearly [1]. The Indonesian government supports the coffee business and avoids low price pressures for coffee farmers [2], [3]. Indonesia's coffee production comprises 72% Robusta, 27% Arabica, and 1% Liberica. The five main coffee production centers in the province of Indonesia are South Sumatra, Lampung, North Sumatra, Aceh, and East Java. In most cases, the tendency of people to find out the quality of coffee is specified based on regional origin. Coffee marketing has its characteristics, and each coffee beans have its distinctive aroma according to its original's region. The government states that there are at least 16 types of Indonesian coffee in the international market. Seven of them are Sumatran Gayo Arabica coffee, Bali Kintamani Arabica coffee, Toraja Arabica coffee, Ijen Raung's Java Arabica coffee, Riau's

Rangsang Meranti Liberika coffee, Temanggung Robusta coffee. Coffee's marketing is needed to be planned and should be well prepared, including the choice of price, promotion, distribution, and services that satisfy the needs, as well as potential buyers.

In the business world, big data is used to collect customer information from every transaction that potentially occurs. Therefore, it can be used as a tool to prepare a market business. The information from big data will be used to identify consumer needs and determine the effective marketing strategies implemented in business through three critical components, namely volume, speed, and variety. Big data consists of a lot of information, and extensive data analysis is carried out on testing or processing large data sets and patterns of data correlation, which is valuable information to run a business. A supervised learning algorithm model is essential. In this problem, the machines work like learning media and depend on working with tasks to form an algorithm and the ambiguity between the classification and clustering algorithms. Many algorithms can be used to classify states or objects. Generally, if the number is higher than the limit range, it is considered a correct classification, but if it is lower, it is classified as incorrect. In cases to predict people with diabetes, machine learning algorithm three is used to form a decision model, comparing prediction results from naive Bayes, linear regression, and decision tree [4]. The previous article also informed the impact of light delays on customer satisfaction using the Naïve Bayesian method, which shows the worst score obtained from score, accuracy, and error (RMSE) [5]. In this present research, the classification of coffees using machine learning techniques is developed. In this case, the coffee is classified into four items, and the category results are used in the quantitative variables of the Bayes model to produce the minimum validation error with a rate of 48.56%. Technological innovations such as big data could be implemented to help companies process rapidly growing data from complex data with low-cost expenses. Big data can also provide information development for decision-making to innovate and improve the capabilities based on knowledge or insights [6]-[9].

In this present research, the cases of coffee production and marketing are studied using machine learning technology with big data analysis [10]. The proposed technology model establishes a periodic monitoring system for farmers' coffee production, guarantees quality and price, and expands coffee shops' interest [11]. In this case, the role of algorithms using big data is to study the trends that will occur in the future and apply the results of innovations to current business events [12]. Computers in a network series connection are connected and generated to continuously learn new knowledge or insights that would improve analytical skills and could be applied to solve the proposed research [13]. This, the proposed technology model, is to be applied to the data for which there is no direct information. This algorithm does not get a training data set, so it requires learning from existing data [14]. In this algorithm, no supervisor helps determine the performance and whether the resulting output is right or wrong, given some input samples without labels [15]. The unsupervised algorithm uses clustering and association methods [16], while the processing of unlabeled type data uses probability statistics [17]. The possibilities in an event are processed statistically by taking into account events that may occur, range 0, and events that do not happen, range 1 [18]. Statistics are closely related to machine learning [19] and are useful for finding patterns that have the smallest error values so that the results found are accurate. Furthermore, systems that use data mining processing that results in market segmentation into several clusters [20], analysis of grouping data with similarities to identify various groups of fields variable data of multiple items. Therefore, members of one cluster have more in common than members of other clusters [21]. The relationship between coffee production and coffee marketing is analyzed through machine learning modeling supported by BDA aimed at informing coffee marketing strategies according to the interests of coffee consumers. The availability of online data sourced from websites that were processed by intelligent machines improves big data work to map, group, and find trend solutions or market predictions as a result of big data analysis [22], [23].

2. METHOD

There are three problems raised in this study, namely firstly predicting coffee production in Indonesia, focusing on Arabica coffee, then classifying the interests of coffee consumers based on regional distinctive aroma characteristics. In the process stage, various data are accommodated in the database. The data is cleaned and divided into a set of user transactions representing the activity of each different user during a visit to the site. A supervised learning algorithm starts with processing data sets grouping them into training data and testing data from coffee production in coffee-producing Indonesian provinces.

2.1. Research frame

In this present case, the proposed technology model is used to establish a monitoring system in the form of periodic evaluation of coffee production by the farmers, guarantee the quality, and price, and expand the interest of coffee shops based on data grouping, in Figure 1.

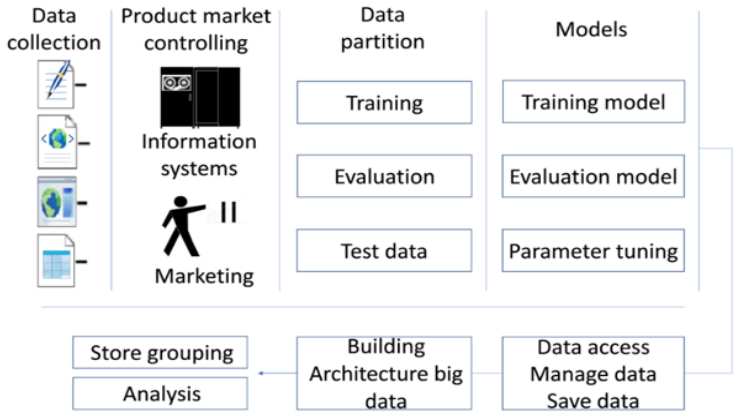

Figure 1. Research frame

2.1.1. Data collection

Data processing starts with collecting the original raw data, which significantly impacts the output results. Actual data in elemental form is assembled from definite and accurate sources so that the results are valid and can be used. In the preparation stage, processing the original raw data is cleaning the data by separating and filtering the raw data to eliminate unnecessary and invalid data. The actual data in the basic form will be seen whether there is incorrect data, similar, miscalculated, or data that does not exist and information is left behind. Data that is not needed is immediately discarded. The result of cleaning the data is converted into a suitable form so that it can be analyzed through further processing to ensure that the best quality data is entered into the processing unit after cleaning. Processing stage data produces output data ready to be processed as input for machine learning and big data analysis modeling.

2.1.2. Product market controlling

There are four products among analysis, information, economics, control, efficiency, and service. Of the many management variables, the best combination can be adapted to the product environment, price, promotion, and distribution.

2.1.3. Data partition

To analyze raw data using machine learning techniques, the data is then tidied up, looking for features that can represent the data. After the processing stage, it uses existing techniques to find the data. The processing stage is the most time-consuming step in machine learning experiments, while training a model may take much less time.

2.1.4. Model

Develop a machine learning model that requires space to learn. Selecting certain values as hyper parameters (different parameters in training) makes the model retested with feedback received when using data validation, and in essence, it is also a way of learning. Modeling is done through the build, train, and evaluation process. Machine learning modeling is done through the build algorithm. The decision tree method represents a function with input in the form of attributes that have a certain value and produces a single result in the form of a class. K-nearest neighbor is an algorithm with the principle that each dataset generally has the closest distance to other data sets with the same properties.

2.1.5. Manage data

An essential process in data analysis because by doing exploration, users will be able to save more time in the data analysis process and can find out errors in the data such as missing values, outliers, duplication, encoding, data noise, and incomplete data. In big data management, the main focus is on storing and processing data efficiently and securely. An analysis is intended to find out new insights or insights. This process uses analysis, machine learning, and artificial intelligence visualization to build a model.

2.1.6. Building architecture big data

Big data is capable of batch processing in real-time from the support of large amounts of data. In the process of storing large amounts of data from the data source, then the management layer receives the data and

converts the best quality data into a format that can be understood as a data source for big data analysis and processing by data analytic tools and stored according to the data format.

2.1.7. Analysis

The problem in this study is first to predict coffee production in Indonesia and then classify the interests of coffee consumers based on regional distinctive aroma characteristics. The relationship between coffee production and coffee marketing is analyzed through machine learning modeling supported by BDA aimed at informing coffee marketing strategies according to the interests of coffee consumers.

2.2. Machine learning algorithm

Technology can work alone to analyze data without being reprogrammed based on learning from a collection of statistics, mathematics, and data mining. The supervised algorithm model depends on training and testing on the task of forming the algorithm. The classification applies regression logistics, decision trees, and others. algorithms for numerical prediction/regression use linear regression, decision trees, and others. The data sets were derived from Web APIs and web services collected in data spreadsheets. The data collected is partitioned into three parts; training data, validation of the learning outcomes model, and test data, namely the data used for prediction [24]. Training data produces features used as indicators and selected according to modeling purposes. Validation data represents the model's performance on data that has not been studied previously. Accordingly, the results can be used for the tuning process. To improve the model's performance, tuning parameters are used [25].

2.3. Big data analysis

Building big data for analysis starts with preparing data sources, aggregators, stores, and apps. The data aggregator functioned to collect and distribute data [26]. The data store stores processing results on document-based NoSQL data [27]. Furthermore, the data is visualized on apps directly related to the user. Extensive data analysis can understand data by uncovering trends and patterns [28]. Machine learning accelerates complex and diverse data processing processes by applying decision-making algorithms. Machine learning can classify, categorize or categorize incoming data, identify patterns that occur, and turn data into information that contains insights according to business interests.

Machine learning forecasts incoming data and acts like an intelligent system that processes past data through past experiences to predict future business. Research is currently being developed to classify the relationship between coffee production and marketing so that business people have a picture of the performance created. This discussion is supported by previous research, such as predicting the future price of a coffee with comparative analysis based on its performance using linear regression, XGB, and LSTM techniques [29]. Related literature comparisons are in Table 1.

Table 1. Related literature comparison

Problem	Solution	Method
Diagnosis and treatment of type diabetes	Predict diabetes and its complications with a literature review	The review used 18 algorithm types from machine learning and deep learning.
Investors and other business people have a desire to know the future price of commodities to make informed business decisions	Predict the future price of a coffee with comparative analysis based on their performance	Using linear regression, XGB, and LSTM technique to build an android application by applying a prediction algorithm.
Agricultural product smart marketing problem and innovation in marketing smart agricultural products	Build an application for the public that will inform the marketing of agricultural products with a smart system	Technological system innovation in agricultural product marketing.

3. RESULTS AND DISCUSSION

In this present work, the data is obtained from the web structures in the form of logs, clickstreams, cookies, and queries. The problem from the coffee market is taken from a collection of types of coffee that are most in-demand by coffee lovers, which can be seen from the variables of coffee interest grouping. The initial is stage processing data sets from input data, then responding to produce output data, and making models that show work with logical calculations for the response to the new data. Guided learning is used with Regression classification techniques aimed at forming predictive modeling of coffee production with coffee marketing. Each domain defines each problem differently. Descriptive and plot statistical methods aim to analyze exploratory data, distribute probabilities according to the data, generate random numbers and then simulate them, and test hypotheses.

The decision tree model produces predictor variables and target variables. A decision tree, or grouping of the tree and a regression tree, estimates the description of responses to several data to be classified. To group the number of responses received, it is necessary to consider the results of the decisions in the tree starting from the beginning of the root node to the leaf node. The node of each leaf represents each response it received. Each decision tree obtained shows two nodes which consist of indicating a decision node and a leaf node. The resulting decision node can make any decision with many branches, while the leaf node shows a decision to be taken and contains no additional components.

This regression and classification tree are included in the supervised machine learning family, namely modeling, whose data is referred to as input variables, and there are variables called outputs. This regression tree and classification tree are included in the supervised machine learning family, namely modeling whose data is referred to as input variables and there are variables called outputs. The regression tree processes the output variable (target) with numeric type while the classification tree processes the output variable (target) with the categorical type (2 categories or more).

Classification learner was used, explaining the training data with the description of predictor X=IdProvince predictor Y=Average. Decision trees are interpreted, can be used quickly for calculations and predictions, and using less storage media results in poor prediction results. SVM class es coffee producer data as well as coffee marketing variables. To obtain the best line boundary called a hyperplane, it aims to group the predicted data results in the form of dots from one class from another. the best line boundary for SVM means the hyperplane with the highest margin between the two classes, in Figure 2.

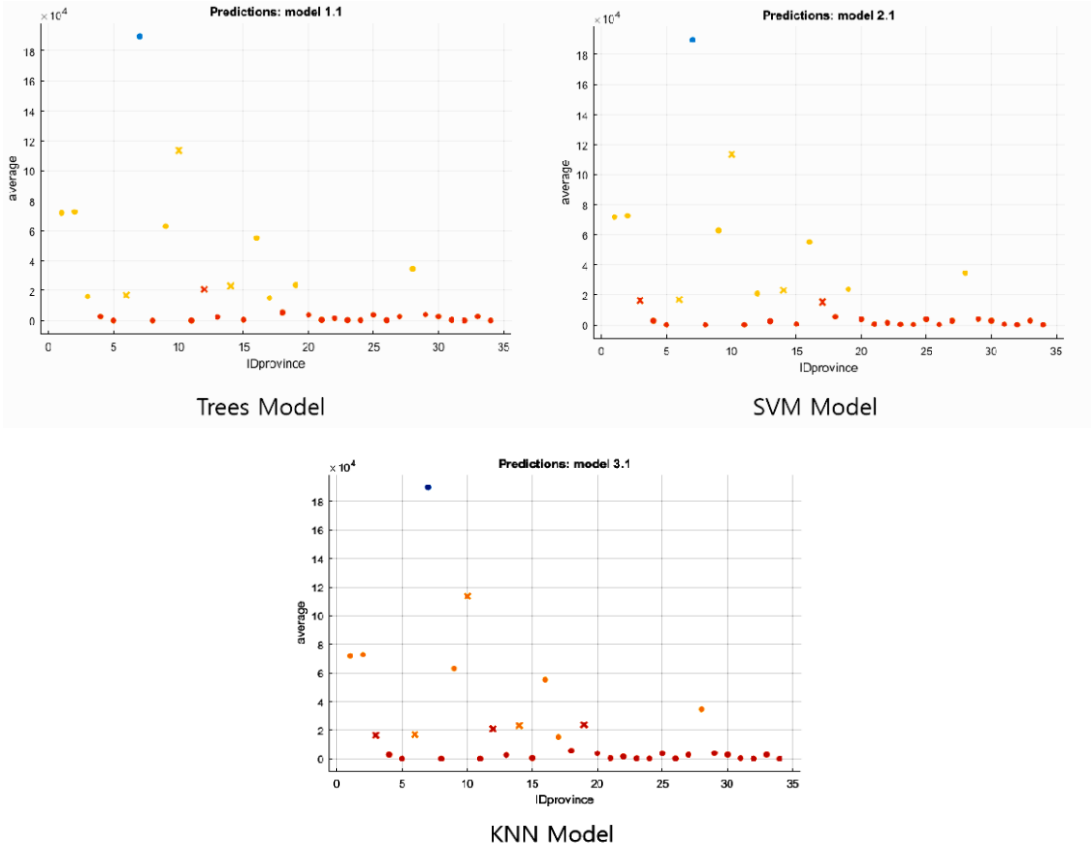

Figure 2. Model prediction

After classifying using the method is to calculate the performance, as for the performance equation using the confusion matrix. True positive indicate that explain predicts positive, and it is actual true negative suggests that explain expects negative. It is proper false-positive show kind 1 is an error that explains prediction of positive and false, false-negative show kind 2 is an error that explains predicted negative and it was wrong. From the results of the coffee production train data set for the last five years, the decision trees modeling is the most accurate at 85.7% compared to other algorithms, the and performance measurement results for classification problems, in Figure 3.

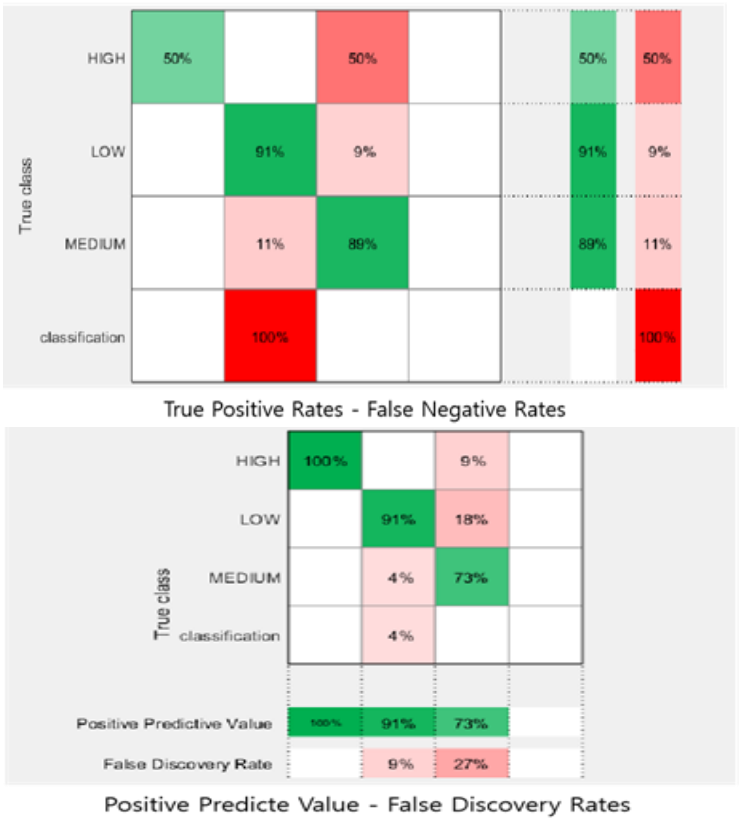

Figure 3. Positive negative rates decision trees

Represents the maximum limit of the plate parallel to the dividing line that does not indicate the interior data point. Classification between neighbors with the closest distance usually produces a level of prediction accuracy that shows the best results within the low-dimensional limit but describes the high-dimensional state: coffee production modeling, Algorithm performance measurement is carried out to determine which algorithm is the best optimal.

F1-score will be used to measure the performance of each algorithm. After knowing the accuracy of each algorithm, then the performance test is carried out. The model on the data set uses K-fold Cross-Validation, which is in charge of testing system classification performance with the number of folds=5. In testing using K-fold cross-validation using training data, the data used for model validation uses test data. The data set is divided into several K-partitions randomly. To further ensure the accuracy of the data set then tested using the AUC on the ROC, in Figure 4.

AUC is a square-shaped area whose value is always between 0 and 1. AUV value for positive high-class production=0.97, AUV value for positive medium class production=0.92, AUV value for positive low-class production=0.91. Random performance produces an AUC value of 0.5 because it shows a diagonal line. The resulting points are shown on the curve at the area (0,0) and the area (1,1). If the resulting AUC<0.5, then the statistical model being evaluated has a meager success of the accuracy value shown in the resulting prediction model is inferior if used. AUC value for positive class Robusta=0.94 while AUC value for positive class Arabica=0.96. Validation of classi cation modeling in Table 2.

Accuracy measurements were carried out to review the level of accuracy of the grouping results based on existing data and classification purposes, the relationship between production and marketing for Arabica Coffee and Robusta Coffee showed that each class had high accuracy. Big data analysis using machine learning algorithm modeling will ensure fast decision making. Big data descriptive analysis can classify coffee production by province as well as classify coffee specialties from regional coffee types. Big data prescriptive analysis helps make better decisions and determines what will happen using optimization and mathematical simulations. The application of machine learning, as well as analysis of big data in the coffee business, has recently continued to develop, especially regarding the classification of coffee beans aimed at predicting the quality of coffee beans, the roasting process, and the quality of coffee produced by coffee production [30]-[33].

All of those results are not only important in the research on coffee business research but also in other scopes of research such as agriculture and food. The present result also could support the previous results and

provide more information on the relationship between coffee production and coffee marketing which is important in coffee marketing strategies and mapping of coffee bean variety which is important in the agriculture research field.

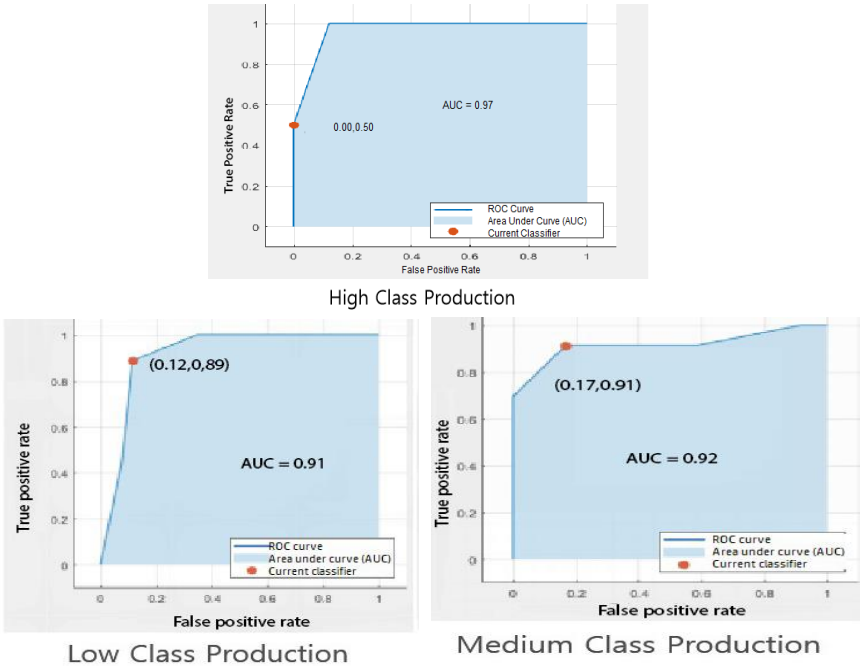

Figure 4. Class productions curve

Table 2. Validation type model

Type model	Class	AUC	ROC	Level of accuracy
Decision trees	High product	0.97	0.00-0.97	High
	Medium product	0.92	0.12-0.89	High
	Low product	0.91	0.17-0.91	High
Logistic regression	Arabica coffee	0.94	0.00-0.86	High
	Robusta coffee	0.92	0.14-1.00	High

4. CONCLUSION

The present research shows that the local coffee production by province can be predicted to be high, medium, or low, obtained from a big data analysis of coffee production over the last 5 years. The ROC curve for the classification of coffee production and local coffee marketing shows that there are three categories of coffee production and two types of coffee specialization which are stated to be true. Development of future research to be able to collect more data sets of types of coffee throughout Indonesia is essential, therefore, the relationship between coffee production and coffee marketing can be found precisely.

ACKNOWLEDGEMENTS

Thank you so much to Airlangga Editing Program 2022 for your kind help and valuable discussion.

REFERENCES

[1] I. Usman and P. V. Nanda, "Green business opportunity of coffee ground waste through reverse logistics," *Journal for Global Business Advancement*, vol. 10, no. 6, pp. 721-737, 2018, doi: 10.1504/JGBA.2017.091941.
[2] E. Candelo, C. Casalegno, C. Civera, and F. Mosca, "Turning Farmers into Business Partners through Value Co-Creation Projects. Insights from the Coffee Supply Chain," *Sustainability*, vol. 10, no. 4, p. 1018, Mar. 2018, doi: 10.3390/su10041018.
[3] N. Kartika and F. P. Setiawan, "Lean Production Implementation on Coffee Beans Business," *Journal of Theory and Applied Management*, vol. 13, no. 3, p. 294, Dec. 2020, doi: 10.20473/jmtt.v13i3.19577.
[4] L. F-Aparicio, J. Noguez, L. Montesinos, and J. A. G-García, "Machine learning and deep learning predictive models for type 2

diabetes: a systematic review," *Diabetol Metab Syndrome*, vol. 13, no. 148, pp. 1-22, Dec. 2021, doi: 10.1186/s13098-021-00767-9.
[5] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou and D. Zhao, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning," in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140-150, Jan. 2020, doi: 10.1109/TVT.2019.2954094.
[6] X-Y. Liu, "Agricultural products intelligent marketing technology innovation in big data era," *Procedia Computer Science*, vol. 183, pp. 648–654, 2021, doi: 10.1016/j.procs.2021.02.110.
[7] I. Nurkasanah, "Reinforcement Learning Approach for Efficient Inventory Policy in Multi-Echelon Supply Chain Under Various Assumptions and Constraints," *Journal of Information Systems Engineering and Business Intelligence*, vol. 7, no. 2, p. 138, Oct. 2021, doi: 10.20473/jisebi.7.2.138-148.
[8] T. N. Rochmah, M. N. Fakhruzzaman, and T. Yustiawan, "Hospital staff acceptance toward management information systems in Indonesia," *Health Policy and Technology*, vol. 9, no. 3, pp. 268–270, Sep. 2020, doi: 10.1016/j.hlpt.2020.07.004.
[9] S. Zulaikha, H. Mohamed, M. Kurniawati, S. Rusgianto, and S. A. Rusmita, "Customer predictive analytics using artificial intelligence," *Singapore Economic Review*, pp. 1–12, Aug. 2020, doi: 10.1142/S0217590820480021.
[10] V. Bhatnagar, G. Singh, G. Kumar, and R. Gupta, "Internet of things in smart agriculture: applications and open challenges," *International Journal of Students Research in Technology & Management*, vol. 8, no. 1, pp. 11–17, Jan. 2020, doi: 10.18510/ijsrtm.2020.812.
[11] C. Miles, "The combine will tell the truth: On precision agriculture and algorithmic rationality," *Big Data & Society*, vol. 6, no. 1, p. 205395171984944, May 2019, doi: 10.1177/2053951719849444.
[12] J. L. Belmonte, A. S-Robles, A.-J. M-Guerrero, and M. E. P-González, "Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science," *Symmetry*, vol. 12, no. 4, p. 495, Mar. 2020, doi: 10.3390/sym12040495.
[13] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *PLoS ONE*, vol. 12, no. 4, p. e0174944, Apr. 2017, doi: 10.1371/journal.pone.0174944.
[14] A. Y. Sun and B. R. Scanlon, "How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions," Environmental Ressearch Letters, vol. 14, no. 7, p. 073001, July 2019, doi: 10.1088/1748-9326/ab1b7d.
[15] C. Shang and F. You, "Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era," *Engineering*, vol. 5, no. 6, pp. 1010–1016, Dec. 2019, doi: 10.1016/j.eng.2019.01.019.
[16] N. Altini *et al*., "Segmentation and Identification of Vertebrae in CT Scans Using CNN, k-Means Clustering and k-NN," *Informatics*, vol. 8, no. 2, p. 40, June 2021, doi: 10.3390/informatics8020040.
[17] I. Q. Utami, M. N. Fakhruzzaman, I. Fahmiyah, A. N. Masduki, and I. A. Kamil, "Customized moodle-based learning management system for socially disadvantaged schools," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3325–3332, Dec. 2021, doi: 10.11591/eei.v10i6.3202.
[18] M. Smith and F. Alvarez, "A machine learning research template for binary classification problems and shapley values integration," *Software Impacts*, vol. 8, p. 100074, May 2021, doi: 10.1016/j.simpa.2021.100074.
[19] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, no. 3, p. e0194889, Mar. 2018, doi: 10.1371/journal.pone.0194889.
[20] C. Li, Y. Chen, and Y. Shang, "A review of industrial big data for decision making in intelligent manufacturing," *Engineering Science and Technology, an International Journal*, vol. 29, p. 101021, May 2022, doi: 10.1016/j.jestch.2021.06.001.
[21] H.-B. Wang and Y.-J. Gao, "Research on parallelization of Apriori algorithm in association rule mining," *Procedia Computer Science*, vol. 183, pp. 641–647, 2021, doi: 10.1016/j.procs.2021.02.109.
[22] M. Nacchia, F. Fruggiero, A. Lambiase, and K. Bruton, "A Systematic Mapping of the Advancing Use of Machine Learning Techniques for Predictive Maintenance in the Manufacturing Sector," *Applied Sciences*, vol. 11, no. 6, p. 2546, Mar. 2021, doi: 10.3390/app11062546.
[23] I. Q. Utami and F. Ramdani, "GEMAR: web-based GIS for emergency management and ambulance routing," *Informatics for Health and Social Care*, vol. 47, no. 2, pp. 1–9, Aug. 2022, doi: 10.1080/17538157.2021.1948856.
[24] S. Rawat, A. Rawat, D. Kumar, and A. S. Sabitha, "Application of machine learning and data visualization techniques for decision support in the insurance sector," International Journal of Information Management Data Insights, vol. 1, no. 2, p. 100012, Nov. 2021, doi: 10.1016/j.jjimei.2021.100012.
[25] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Comput Mater*, vol. 5, no. 83, Dec. 2019, doi: 10.1038/s41524-019-0221-0.
[26] C. Lehmann, L. G. Huber, T. Horisberger, G. Scheiba, A. C. Sima, and K. Stockinger, "Big Data architecture for intelligent maintenance: a focus on query processing and machine learning algorithms," *Journal of Big Data*, vol. 7, no. 61, Dec. 2020, doi: 10.1186/s40537-020-00340-7.
[27] C.-O. Truică, E.-S. Apostol, J. Darmont, and T. B. Pedersen, "The Forgotten Document-Oriented Database Management Systems: An Overview and Benchmark of Native XML DODBMSes in Comparison with JSON DODBMSes," *Big Data Research*, vol. 25, p. 100205, Jul. 2021, doi: 10.1016/j.bdr.2021.100205.
[28] M. Q. Shabbir and S. B. W. Gardezi, "Application of big data analytics and organizational performance: the mediating role of knowledge management practices," *Journal of Big Data*, vol. 7, no. 47, Dec. 2020, doi: 10.1186/s40537-020-00317-6.
[29] M. Supriya, and A. Deepa, "Machine learning approach on healthcare big data: a review," *Big Data and Information Analytics*, vol. 5, no. 1, pp. 58–75, 2020, doi: 10.3934/bdia.2020005.
[30] J. Y. Kim, "Coffee Beans Quality Prediction Using Machine Learning," *SSRN*, 1-7, 2022, doi: 10.2139/ssrn.4024785.
[31] Y.-C. Hung, F.-S. Lee, and C.-I. Lin, "Classification of coffee bean categories based upon analysis of fatty acid ingredients," *Journal of Food Processing and Preservation*, vol. 45, no. 9, p. e15703, doi: 10.1111/jfpp.15703.
[32] J. A. Suarez-Peña, H. F. Lobaton-García, J. I. Rodríguez-Molano, W. C. Rodriguez-Vazquez, "Machine Learning for Cup Coffee Quality Prediction from Green and Roasted Coffee Beans Features," In *book: Applied Computer Sciences in Engineering, WEA 2020, Communications in Computer and Information Science*, vol 1274, pp. 48-59, doi: 10.1007/978-3-030-61834-65.
[33] M. Okamura, M. Soga, Y. Yamada, K. Kobata, D. Kaneda, "Development and evaluation of roasting degree prediction model of coffee beans by machine learning," *Procedia Computer Science*, vol. 192, pp. 4602-4608, 2021, ISSN 1877-0509, doi: 10.1016/j.procs.2021.09.238.

BIOGRAPHIES OF AUTHORS

Anita Sindar Ros Maryana Sinaga Bachelor's degree in Informatics Engineering from the ST INTEN Bandung, Indonesia. I received a master's degree in informatics engineering from the Computer Science Study Program, Bina Nusantara University Jakarta, Indonesia. Currently a Lecturer Computer Science in the Department Information Technology STMIK Pelita Nusantara in Medan, Indonesia. Research interest in image processing, machine learning, and algorithms. She can be contacted at email: haito_ita@yahoo.com.

Ricky Eka Putra received Doctor from Informatics Engineering ITS, Surabaya. Currently, he is a lecturer at Department of Informatics, Surabaya State University, Surabaya, Indonesia. He can be contacted at email: rickyeka@unesa.ac.id.

Abba Suganda Girsang obtained Ph.D. degree in the Institute of Computer and Communication Engineering, Department of Electrical Engineering and National Cheng Kung University, Tainan, Taiwan, in 2014. He graduated bachelor from the Department of Electrical Engineering, Gadjah Mada University (UGM), Yogyakarta Indonesia, in 2000. He then continued his master's degree in the Department of Computer Science in the same university in 2006–2008. He was a staff consultant programmer in Bethesda Hospital, Yogyakarta, in 2001 and worked as a web developer in 2002–2003. He then joined the faculty of Department of Informatics Engineering in Janabadra University as a lecturer in 2003-2015. He also taught some subjects at some universities in 2006–2008. His research interests include swarm intelligence, business intelligence, machine learning and media social text mining. He can be contacted at email: agirsang@binus.edu.