Journal of Engineering Science and Technology
Vol. 19, No. 1 (2024) 163 - 175
© School of Engineering, Taylor's University

APPLICATION OF DATA MINING FOR PREDICTING HORTICULTURAL COMMODITIES PRICE

DIAN DHARMAYANTI[1,*], ARSY OPRAZA AKMA[1],
EDDY SOERYANTO SOEGOTO[2], LIA WARLINA[3]

[1]Department of Informatics Engineering, Universitas Komputer Indonesia, Indonesia
[2]Department of Management, Universitas Komputer Indonesia, Indonesia
[3]Department of Urban and Regional Planning, Universitas Komputer Indonesia, Indonesia
*Corresponding Author: dian.dharmayanti@email.unikom.ac.id

Abstract

In Garut Regency, farmers receive their commodities at a selling price that collectors determined. This issue causes farmers to lose money because their income does not correspond with the market price for their goods. This loss has an impact on reducing farmer productivity. Therefore, this research aims to provide recommendations to farmers regarding commodity selling prices for particular periods based on forecast results. This research implements data mining so commodities selling prices can be predicted. Data mining is a way to find patterns or knowledge from past data. Multiple linear regression methods or algorithms are used to find patterns or knowledge with data mining. Multiple linear regression can be used to predict commodities selling prices based on rainfall factors and total production. The application developed can produce recommendations for farmers for commodity selling prices for specific periods based on prediction results, thereby reducing losses for farmers.

Keywords: Data mining, Horticultural commodities, Multiple linear regression, Price prediction.

1. Introduction

Indonesia is an agrarian country, with most of its population working in the agricultural sector. The agricultural sector serves as the primary source of income for a significant portion of the Indonesian population. With vast agricultural land available, the surrounding population can utilize this land for farming. In other words, a substantial number of people in the vicinity depend on the agricultural sector for their livelihood [1].

Garut Regency is one of the West Java regencies with great potential in agriculture. Garut Regency has vast agricultural land, covering an area of 242,388 hectares, which can be utilized for cultivating various agricultural commodities. Most of the Garut Regency's population relies on agricultural land to meet daily needs. With the support of the local government, especially the Department of Agriculture, Garut Regency produces agricultural commodities that can supply the needs of the local community, particularly the residents of Garut Regency and the surrounding regencies or cities [2]. Based on data from the Central Statistics Agency (BPS) in 2021 [3], Garut Regency is one of the top five regencies producing shallots, red, and bird's eye chili commodities.

The factors influencing commodity market prices include production quantity, market demand, seeds and fertilizers, rainfall, and land area [4]. The market prices of agricultural commodities affect the increase in farmers' production yield and enhance their income. High market prices assist farmers in achieving more significant profits, thereby boosting their production. On the other hand, low market prices lead to a decrease in farmers' production. High market prices and price stability are essential for farmers to attain profit and income stability [5]. Unstable market prices result in farmers' losses and provide intermediaries with opportunities to manipulate commodity price information at the farmer level.

By implementing data mining, it is possible to generate predictions for commodity sale prices [6]. Data mining is discovering interesting patterns and knowledge from large datasets [7]. Data mining can be used as a decision support tool, forecasting future trends to make practical steps in problem-solving [8]. Furthermore, correlation analysis is also used to find the relationship or association between two or more variables. Correlation analysis measures the relationship or association between two or more variables [9]. In addition, linear regression is a method used to predict the value of the dependent variable (response) based on the independent variables (predictors) [10]. Based on projected outcomes, this study attempts to give farmers recommendations regarding commodity prices for specific periods. In order to forecast commodity prices, this study uses data mining.

2. Related Works

Data mining is a technique for making estimates or judgments based on historical data in various fields, including business, education, and agriculture. One definition of data mining is the science of taking relevant information out of databases. Data mining identifies patterns and subtle relationships in data and infers principles that enable future prediction by combining machine learning, statistical analysis, modelling approaches, and database technologies. There are several application domains for decision trees and clustering algorithms, and each approach has pros and cons [11]. Through analysis of data originating from various educational technologies, a growing field called Educational Data Mining (EDM) is building upon and contributing to a wide range of fields. Researchers in EDM are combining data from intelligent tutoring systems, massively open online courses, educational games and simulations, discussion forums, and incentives to address topics related to cognition, metacognition, affect, language, and social discourse. The information includes thorough action and timing logs of student interactions with various user interfaces, including chat dialogs, games, simulations, enhanced problem-solving settings, and graded essays, questions, and essays [12].

Data mining research for price prediction was conducted on the Bucharest Stock Exchange (BSE). A multistep forward technique was employed to forecast. Three Romanian stocks traded on the BSE were subjected to neural networks intended for stock exchange rate predictions. A multistep forward technique was applied to forecast transient price changes. Eventually, an intelligent multi-agent system models that employs data mining and data stream processing techniques to assist users in making stock purchases or sales decisions can incorporate the study's findings [13].

Information technology applications in agriculture can enhance decision-making and increase farmer productivity. Data mining is essential in decision-making on many issues in the agriculture sector. Studies conducted in India demonstrated the function of data mining as it relates to the agricultural industry. Various data mining techniques are applied to the input data to find the optimal output-yielding method [14]. Agricultural data mining and statistical algorithms are improved using custom approaches and time series representation breakthroughs in precision and intelligent agriculture data mining and statistics. In order to build a precision and intelligent farm data mining and statistical analysis model that is based on big data analysis and design experiments. The study's findings demonstrate the particular impact of the developed model [15].

In agriculture, data mining is used to forecast yield levels of production, agricultural hindrances, including disease and pest outbreaks, weather patterns, and market pricing for crop products [16-20]. The issue of forecasting crop prices has grown in importance and should only be resolved with the data that is now available. This difficulty can be resolved with data mining approaches by identifying appropriate data models that support high-price forecast accuracy and generality [21].

Memory limits, processor power limitations, hard disk space issues, and other issues are common issues with traditional data mining algorithms. Distributed Data Mining, or DDM, on a distributed computer environment has emerged as a practical substitute in numerous applications to address the issues above. DDM techniques include distributed clustering, distributed frequent itemset mining, distributed frequent sequence mining, distributed frequent graph mining, and privacy-preserving DDM [22].

3. Method

The research methodology to be employed in this study is Cross Industry Standard for Data Mining (CRISP-DM) [23]. CRISP-DM makes data mining a commonly used solution for businesses or research units [24]. CRISP-DM has six stages in its lifecycle [25, 26]. The methodology for the research was adapted to produce the method depicted in Fig. 1.

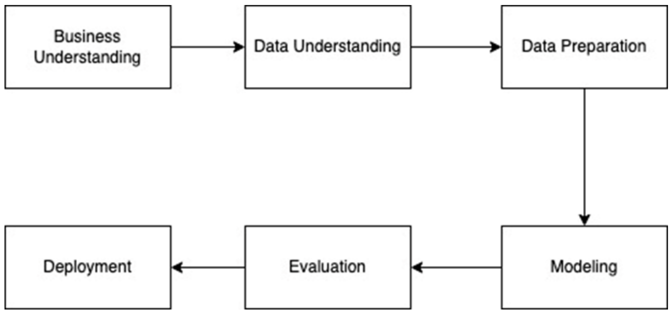

Fig. 1. Research methodology.

The following are the phases of work for the research that was conducted:

i) **Business understanding**
This stage involves understanding the business and business objectives, which are to reduce farmers' losses in selling harvested commodities and to predict the selling prices of commodities for the upcoming harvest period.

ii) **Data understanding**
In the second stage, data is collected by collaborating with the Garut Regency Department of Agriculture, the Meteorology, Climatology, Geophysics Agency, and the Ministry of Trade. Subsequently, an understanding of the collected data is established.

iii) **Data preparation**
In this stage, preparations are made to create the dataset from the production, weather, and market price data collected in the previous stage. Data is then cleaned, data selection for use is made, and unused attributes are removed.

iv) **Modelling**
In this stage, a prediction model for the selling prices of agricultural commodities in the Garut Regency is developed.

v) **Evaluation**
In this stage, testing and evaluation of the model created using multiple linear regression algorithms are conducted. Additionally, accuracy testing of the agricultural commodity price prediction model is carried out.

vi) **Deployment**
In the last stage, the model is implemented into an Android application that end-users can use. Data mining results can be represented as commodity selling price predictions.

4. Results and Discussion

The results of each stage's development are shown in the following section.

4.1. Business understanding

The business understanding phase aims to comprehend the problem, identify objectives, and use data mining tools and techniques. The problem understanding conducted involves assisting farmers in determining the selling prices of agricultural commodities based on price prediction results. The predicted selling prices can be used to determine the prices in the next harvest period.

Shallot commodities have a 3-month planting period. Bird's eye chili commodities have a three-month planting period and can be harvested during a 6-month harvest period. Red chili commodities have a planting period of 70-75 days.

The objective of data mining in this research is to generate price recommendations for commodities based on prediction results. The Head of the Farmers' Association and individual farmers can use these price recommendations. The tools and techniques utilized in this research include Python, Kotlin, Visual Studio Code, Android Studio, and the regression method.

4.2. Data understanding

The data understanding phase consists of two stages: (1) data collection and (2) data description used in this research.

4.2.1. Data collection

Rainfall data was obtained from the Meteorology, Climatology, and Geophysics Agency downloaded from the West Java Central Statistics Agency website. Commodity price data was obtained from the National Strategic Food Price Information Center website. Commodity production data was acquired from the Garut Regency Department of Agriculture.

4.2.2. Data description

The obtained rainfall data represents the precipitation in the West Java region for 2021, categorized by month and measured in millimetres (mm). The commodity price data collected pertains to agricultural commodity prices in West Java for 2021, categorized by month and denominated in Indonesian Rupiah (IDR). The production data acquired from the Garut Regency Department of Agriculture for 2021 is categorized by month and measured in tons.

4.3. Data preparation

Data preparation, or data preprocessing, is a stage that aims to adjust the dataset to make it usable according to modelling needs. It consists of the following steps: (1) Data integration, (2) Attribute selection, (3) Data cleaning, (4) Data outlier refinement, and (5) Data transformation.

4.3.1. Data integration

The data to be used for mining is stored in different .xlsx files; therefore, data integration is necessary. Table 1 shows the result of data integration.

4.3.2. Attribute selection

The attributes used for the data mining process were adjusted according to the requirements, resulting in three datasets corresponding to each commodity. Unused attributes were removed. The attributes selected for rainfall data include the '2021' attribute to match the data period of the year 2021. For price data, the selected attribute is 'Province: West Java'. For production data, the used attributes match the commodity to be predicted. Table 2 as an example illustrates the attribute selection for shallot commodities.

Table 1. Results of data integration.

Month	Rainfall (mm)			Commodity's Price (IDR)			Production (Ton)		
	2020	2021	2022	Shallot	Red Chilli	Bird's Eye Chili	Shallot	Red Chilli	Bird's Eye Chili
January	207.6	304.3	106.6	17 150	33 150	27 950	3 459	10 224	5 068
February	336.6	486.8	150.3	16 050	30 800	60 450	4 607	10 428	6 568
March	292.5	233.0	113.2	18 350	31 850	57 500	2 848	14 111	5 534
April	271.4	505.1	316.6	22 150	38 000	53 250	2 998	13 996	5 809
May	292.3	510.3	228.5	21 350	26 550	37 450	4 328	8 432	3 084
June	30.3	311.1	463.7	16 200	15 950	20 850	5 078	13 554	4 443
July	63.7	115.6	358.1	18 150	12 400	20 550	2 768	12 058	5 826
August	41.6	399.5	384.9	20 900	20 400	29 350	2 587	7 190	2 156
September	87.7	317.3	353.7	16 350	13 750	10 400	2 313	4 420	1 419
October	327.3	566.5	492.3	13 750	14 150	11 650	1 775	3 159	1 201
November	207.3	183.6	321.0	15 100	20 750	15 900	1 586	3 329	1 080
December	262.1	279.1	224.1	8 750	28 650	16 250	894	2 565	664

Table 2. Results of attribute selection.

Month	Rainfall (mm)	Price (IDR)	Production (Ton)
January	304.3	17 150	3 459
February	486.8	16 050	4 607
March	233.0	18 350	2 848
April	505.1	22 150	2 998
May	510.3	21 350	4 328
June	311.1	16 200	5 078
July	115.6	18 150	2 768
August	399.5	20 900	2 587
September	317.3	16 350	2 313
October	566.5	13 750	1 775
November	183.6	15 100	1 586
December	279.1	8 750	894

4.3.3. Data cleaning

The data used in this research does not require data cleaning since it does not contain null values or duplicates. The data is ready to be used for the modelling stage.

4.3.4. Data outlier refinement

Outlier data refinement is performed on rainfall, price, and production attributes in the shallot, red chili, and bird's eye chili datasets. Outliers can be identified using the Interquartile Range (IQR) method. Outliers were not removed because the identified outlier values do not result from data collection or input errors; therefore, the data are retained. However, these outliers were smoothed using the binning method. Binning itself has four methods: Equal width binning, Equal frequency binning, binning by clustering, and binning by predictive value [27-29]. The binning method consists of three stages: sorting the data in ascending order, dividing the data into several bins, and smoothing by bin means. Table 3 is the results of the data binning performed on the shallot dataset.

Table 3. Results of outlier smoothing.

Month	Rainfall (mm)	Price (IDR)	Production (Ton)
January	304.3	16.962	3.459
February	486.8	13.412	4.607

March	233.0	20.687	2.848
April	505.1	20.687	2.998
May	510.3	20.687	4.328
June	311.1	16.962	5.078
July	115.6	16.962	2.768
Agustus	399.5	20.687	2.587
September	317.3	16.962	2.313
October	566.5	13.412	1.775
November	183.6	13.412	1.586
December	279.1	13.412	894

4.3.5. Data transformation

Data transformation is scaling data into a smaller range, typically between 0 and 1. The transformation technique employed here is min-max normalization. The data to be normalized includes rainfall and production data in the shallot, red chili, and bird's eye chili datasets.

4.4. Modelling

The modelling stage involves selecting a suitable model for the data and mining objective. Multiple Linear Regression is the modelling method used to generate price predictions for commodities, which provide price recommendations to farmers. However, before modelling, a correlation analysis is conducted among variables.

The method used for correlation analysis is Pearson Correlation, employed to understand the relationships between the observed variables, namely rainfall and production, with the selling price variable as the target variable. The reason for using the Pearson Correlation method is that it requires the data to have a normal distribution and linear relationships [30-32]. Table 4 presents the results of the correlation analysis for shallot commodities.

Table 4. Correlation analysis results.

No.	X	Y	Correlation Coefficient	Relationship Level
1	Rainfall	Price	0.238794079	Weak
2	Production	Price	0.49280976	Currently
3	Rainfall	Production	0.26539885	Weak
4	Production	Price	0.505381	Currently

The method used for predicting selling prices in this research is multiple linear regression. The correlation analysis results show that the variables production, rainfall, and production correlate highly. Therefore, multiple linear regression uses these variables for the modelling stage. Table 5 shows the results of the modelling for shallot commodities.

Table 5. Data transformation.

Month	Production (Ton)	Rainfall and Production	Prediction (IDR)
January	0.613	0.257	17.524

February	0.887	0.731	18.214
March	0.467	0.122	17.063
April	0.503	0.434	16.952
May	0.821	0.718	17.964
June	1.000	0,434	18.891
July	0.448	0.000	17.086
August	0.405	0.255	16.713
September	0.339	0.152	16.540
October	0.211	0.211	15.992
November	0.165	0.025	15.964
December	0.000	0.000	15.339

4.5. Evaluation

After the modelling, the created model is evaluated to measure its accuracy. Linear regression or predictions can be tested using several methods, such as Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) [15]. The evaluation used in this research involves utilizing Root Mean Square Error (RMSE), presented in Table 6.

Table 6. Evaluation of prediction results.

Month	Actual (Y)	Prediction (Y')	Y-Y'	(Y-Y')^2
January	16962	17524	-562	316269
February	13412	18214	-4802	23060535
March	20687	17063	3624	13132262
April	20687	16952	3735	13949911
May	20687	17964	2723	7415730
June	16962	18891	-1929	3722518
July	16962	17086	-124	15409
August	20687	16713	3974	15792245
September	16962	16540	422	177794
October	13412	15992	-2580	6654764
November	13412	15964	-2552	6514883
December	13412	15339	-1927	3715060
Amount		94467379		

$$RMSE = \sqrt{\frac{94467379}{12}} \tag{1}$$

$$RMSE = 2806 \tag{2}$$

After calculating the RMSE, the prediction of shallot commodity prices has an error value of 2806. An RMSE value approaching zero can be interpreted as an accurate prediction result. Based on the RMSE calculation, this RMSE value serves as the range for the predicted values, which can be observed in Table 7.

Table 7. RMSE evaluation.

Month	Actual (IDR)	Minimum (IDR)	Maximum (IDR)	Prediction (IDR)

January	16.962	14.156	19.768	17.524
February	13.412	10.606	16.218	18.214
March	20.687	17.881	23.493	17.063
April	20.687	17.881	23.493	16.952
May	20.687	17.881	23.493	17.964
June	16.962	14.156	19.768	18.891
July	16.962	14.156	19.768	17.086
August	20.687	17.881	23.493	16.713
September	16.962	14.156	19.768	16.540
October	13.412	10.606	16.218	15.992
November	13.412	10.606	16.218	15.964
December	13.412	10.606	16.218	15.339

4.6. Deployment

The deployment carried out in this research involves developing a web-based Android application.

4.7. Implementation of result

The system implementation is the stages of software development that follows the system design phase previously created. This stage encompasses hardware, software, and user interface development.

4.7.1. Hardware implementation

The following is the hardware implementation environment presented in Table 8.

Table 8. Hardware implementation.

No.	Hardware	Specification
1	Processor	2 vCPU
2	Memory	1 GB
3	Storage	20 GB

4.7.2. Software implementation

Table 9 shows the software implementation environment.

Table 9. Software implementation.

No.	Software	Specification
1	Operation System	Ubuntu 22.04
2	Programming Language	Python and Kotlin
3	Framework	Flask
4	IDE dan Text Editor	Visual Studio Code and Android Studio
5	Version Controlling	GitHub

4.7.3. Interface implementation

The user interface implementations carried out on the website and Android application are presented in Figs. 2-5.

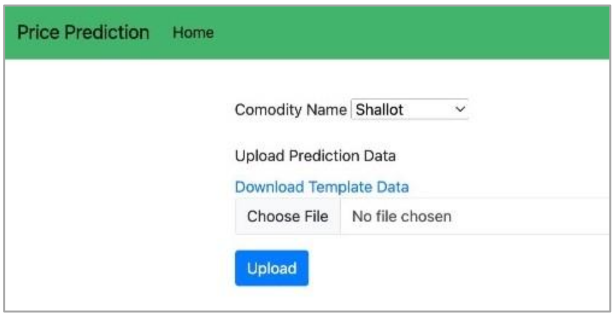

Fig. 2. Web data upload.

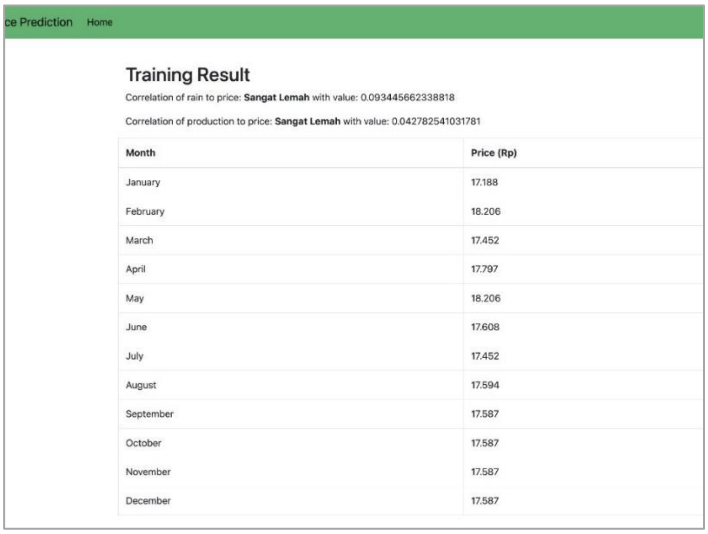

Fig. 3. Web data training results.

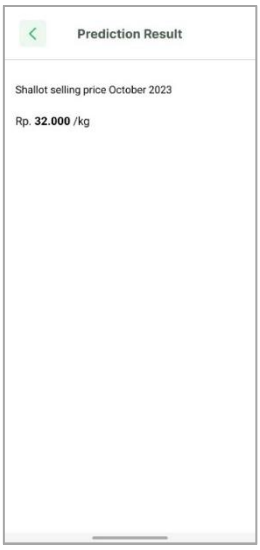

Fig. 4. Input data android. **Fig. 5. Android prediction result.**

Based on the analysis result, the prediction results only have four months outside the RMSE range: February, March, April, and August. Therefore, it can be concluded that the prediction results are reasonably good.

5. Conclusion

Based on the analysis and testing conducted in this research, it can be concluded that the commodity price prediction system can assist the Head of the Farmers' Association and individual farmers in predicting selling prices, which can be used as price recommendations for commodities. As for recommendations for the further development of this research, it is suggested to include additional attributes or factors such as fertilizer and seed types, expand the range of agricultural commodities to be predicted, and increase the amount of training data used for modelling to enhance prediction accuracy.

Acknowledgements

The authors thank Kementerian Pendidikan dan Kebudayaan and Riset dan Teknologi (Kemendikbud Ristek) Republik Indonesia for the research funding and the Office of Agriculture in the Garut Regency for the data and consultation.

References

1. Soeryanto, E.S.; Warlina, L.; Supatmi, S.; Rafdhi, A.A.; Jumansyah, R.; and Saputra, H. (2022). SITAMPAN: Mobile application for planting and harvesting of horticultural crops in Garut regency. *MATRIX-Jurnal Manajemen Teknologi dan Informatika*, 12(3), 123-136.
2. Purnomo, H.; Fitrah, F.R.; Maulana, R.; and Pratadina, M.M. (2021). Implementation of information system in Indonesian traditional beverage businesses. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 2(1), 15-24.
3. Luckyardi, S.; Soegoto, E.S.; Supatmi, S.; Warlina, L.; and Hassan, F. (2022). Marketing strategy for local superior commodities and regional economic contributions of Indonesia. *Journal of Eastern European and Central Asian Research*, 9(1), 1-9.
4. Lesmana, R.; and Rifaldi, M.I. (2023). A computational bibliometric analysis of e-groceries analysis using vosviewer. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 4(1), 75-88.
5. Bolaji, H.O.; and Ibrahim-Raji, T.B. (2023). ICT-based literacy evaluation in Nigeria educational sector: Case study in Kwara state. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 4(1), 1-10.
6. Pratama, A.F. (2021). Sustainable and digitized certification of palm oil production: Its impact on the environment in Indonesia. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 2(2), 23-30.
7. Wibowo, B.D.S. (2022). XBRL open information model for risk-based tax audit using machine learning. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 3(1), 21-46.

8. Pejić, A.; and Molcer, P.S. (2021). Predictive machine learning approach for complex problem solving process data mining. *Acta Polytechnica Hungarica*, 18(1), 45-63.
9. Gogtay, N.J.; and Thatte, U.M. (2017). Principles of correlation analysis. *Journal of the Association of Physicians of India*, 65(3), 78-81.
10. Darwis, M.; Hasibuan, L.H.; Firmansyah, M.; Ahady, N.; and Tiaharyadini, R. (2021). Implementation of k-means clustering algorithm in mapping the groups of graduated or dropped-out students in the management department of the national university. *JISA (Jurnal Informatika dan Sains)*, 4(1), 1-9.
11. Tamilselvi, R.; and Kalaiselvi, S. (2013). An overview of data mining techniques and applications. *International Journal of Science and Research (IJSR)*,2(2), 2319-7064.
12. Koedinger, K.R.; D'Mello, S.; McLaughlin, E.A.; Pardos, Z.A.; and Rosé, C.P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333-353.
13. Chang, T.H.; Wang, N.; and Chuang, W.B. (2021). Stock price prediction based on data mining combination model. *Journal of Global Information Management (JGIM)*, 30(7), 1-19.
14. Badapanda, K.; Mishra, D.P.; and Salkuti, S.R. (2022). Agriculture data visualization and analysis using data mining techniques: Application of unsupervised machine learning. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 20(1), 98-108.
15. Rao, Z.; and Yuan, J. (2021). Data mining and statistics issues of precision and intelligent agriculture based on big data analysis. *Acta Agriculturae Scandinavica, Section B—Soil and Plant Science*, 71(9), 870-883.
16. Lata, K.; and Chaudhari, B. (2019). Crop yield prediction using data mining techniques and machine learning models for decision support system. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 6(4), 391-396.
17. Maurya, R.; Yadav, S.; and Sharma, T. (2019). Estimation of major agricultural crop with effective yield prediction using data mining. *International Journal of Innovative Technology and Exploring Engineering*, 8(1), 170-174.
18. Sellam, V.; and Poovammal, E. (2016). Prediction of crop yield using regression analysis. *Indian Journal of Science and Technology*, 9(38), 1-5.
19. Hasibuan, L.H.; Munaji, A.A.; Deolika, A.; Chitayae, N.; and Setiawan, E. (2022). Student attendance information system using QR code (quick response) at ITSNU Kalimantan. *JURNAL SISFOTEK GLOBAL*, 12(2), 101-107.
20. Hammer, R.G.; Sentelhas, P.C.; and Mariano, J.C. (2020). Sugarcane yield prediction through data mining and crop simulation models. *Sugar Tech*, 22(2), 216-225.
21. Kaur, M.; Gulati, H.; and Kundra, H. (2014). Data mining in agriculture on crop price prediction: Techniques and applications. *International Journal of Computer Applications*, 99(12), 1-3.
22. Gan, W.; Lin, J.C.W.; Chao, H.C.; and Zhan, J. (2017). Data mining in distributed environment: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(6), 1-19.

23. Khumaidi, A. (2020). Data mining for predicting the amount of coffee production using CRISP-DM method. *Jurnal Techno Nusa Mandiri*, 17(1), 1-8.
24. Brzozowska, J.; Pizoń, J.; Baytikenova, G.; Arkadiusz, G.O.L.A.; Zakimova, A.; and Piotrowska, K. (2023). Data Engineering in CRISP-DM process production data-Case study. *Applied Computer Science*, 19(3), 83-95.
25. Kedong, L.Y.U.; and Li, K. (2022). Research on the safety fault diagnosis system of mine based on information fusion. *International Journal of Research and Applied Technology (INJURATECH)*, 2(2), 31-40.
26. Jaggia, S.; Kelly, A.; Lertwachara, K.; and Chen, L. (2020). Applying the CRISP-DM framework for teaching business analytics. *Decision Sciences Journal of Innovative Education*, 18(4), 612-634.
27. Soegoto, E.S.; Ananta, H.; Zaki, I.; and Ranau, M.I.N. Implementation of management information system using machine learning technology. *International Journal of Research and Applied Technology (INJURATECH)*, 2(2), 2022.
28. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, 49(4), 764-766.
29. Barbato, G.; Barini, E.M.; Genta, G.; and Levi, R. (2011). Features and performance of some outlier detection methods. *Journal of Applied Statistics*, 38(10), 2133-2149.
30. Khozani, Z.S.; Khosravi, K.; Pham, B.T.; Kløve, B.; Wan Mohtar, W.H.M.; and Yaseen, Z.M. (2019). Determination of compound channel apparent shear stress: Application of novel data mining models. *Journal of Hydroinformatics*, 21(5), 798-811.
31. Khosravi, K.; Daggupati, P.; Alami, M.T.; Awadh, S.M.; Ghareb, M.I.; Panahi, M.; and Yaseen, Z.M. (2019). Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. *Computers and Electronics in Agriculture*, 167(1), 105041.
32. Khosravi, K.; Mao, L.; Kisi, O.; Yaseen, Z.M.; and Shahid, S. (2018). Quantifying hourly suspended sediment load using data mining models: Case study of a glacierized Andean catchment in Chile. *Journal of Hydrology*, 567(1), 165-179.