Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12943
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Sales Conversion Optimization Analysis Using the Random Forest Method

Th. Dwiati Wismarini [1], Hari Murti [2)], Kristiawan Nugroho[3)*]
[1)2)3)]Universitas Stikubank, Indonesia
[1)]thwismarini@edu.unisbank.ac.id, [2)]harimurti@edu.unisbank.ac.id, [3)]kristiawan@edu.unisbank.ac.id

Submitted :Aug 28, 2023 | **Accepted** : Sep 19, 2023 | **Published** : Oct 3, 2023

Abstract: Companies are competing to be winners by improving their services and hoping that their product sales can increase in various ways, including by using optimization theory. However, the lack of data analysis is a problem that is often encountered in optimizing sales conversions. Various machine learning-based methods have also been used to help analyze sales conversion optimization. This research uses the Random Forest method which is one of the more robust machine learning methods compared to other methods, namely Adaptive Booster (AdaBoost) and K-Nearest Neighbor (KNN) in analyzing sales conversion optimization. The results showed that the Random Forest method had the best performance in classifying data, by using the 10 cross validation technique the results were obtained with a Mean Squared Error (MSE) value of 0.928 and a Root Mean Square Error (RMSE) of 0.963, better than the Adaptive Booster method. and K-Nearest Neighbor which has lower performance. Sales conversion optimization processing using Random Forest is proven to have the best performance as evidenced by the small Mean Squared Error and Root Mean Square Error which means it has an accurate level of performance compared to other methods.

Keywords: Sales Conversion, Optimization, Machine Learning,Methods, Random Forest

INTRODUCTION

The development of the world of business and information has made the world of commerce more advanced. Business people can communicate with each other and carry out buying and selling activities without being controlled by distance and time. Information technology is a bridge between the worlds so that any business transaction can be carried out more quickly, thus providing greater benefits to business people. Sales activities are one of the areas that are currently growing. After the Covid-19 pandemic ended, the buying and selling sector seemed to be stretching and running again. Various kinds of transactions are carried out in various ways either by meeting in person or using website media in selling their products and services to the public.

Sales via the internet, which is more popular, known as E-Commerce, is one of the prima donnas of today's trade. E-Commerce makes it easier for business people to communicate with each other and make transactions via the Internet(Riswandi, 2019). In addition, E-Commerce will also provide a wider range of trade so that it will reach more and more customers. Based on the increasing number of advantages in using E-Commerce, it is not surprising that currently the trend of trading using the internet is increasingly advanced and popular with business people in an effort to market their products.

Sales conversion (SC) is a term that is familiar to the world of business and trade. SC is a term that refers to the comparison between the number of prospects or potential customers who ultimately make certain purchases or actions desired, compared to the total number of existing prospects. Various forms of frameworks have been produced from research on sales conversions such as research conducted by Zimmermann (Zimmermann & Auinger, 2023) who created a framework for conversion rate optimization which produces a new framework that retailers can use to increase competitiveness in e-commerce.

One of the problems faced in sales conversion optimization is the lack of data analysis so that it cannot achieve the goals that have been set. This problem has also been tried to be overcome in various ways, including by optimizing the website so that it is hoped that the conversion rate will increase. The use of Artificial Intelligence and machine learning has also been used in several studies regarding conversion optimization such as in research conducted by(Risto Miikkulainen, Myles Brundage, Jonathan Epstein, Tyler Foster, Babak Hodjat, Neil Iscoe, Jingbo Jiang, Diego Legrand, Sam Nazari, Xin Qiu, Michael Scharff, Cory Schoolland, Robert Severn, 2020) who analyzed the conversion rate on landing pages.

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12943
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Machine learning is an approach that is often used in various studies. This research uses a method in Machine Learning that is often used in research, namely Random Forest. The Random Forest algorithm has advantages in terms of accuracy produced, efficiency in data storage(Supriyadi et al., 2020), and can solve non-linear problems. In several studies that have been carried out the Random Forest method has also succeeded in achieving quite good performance as in research conducted by (Normah et al., 2022) in detecting Covid-19 through CT-Scan images by incorporating Haralick and Color Histogram features into the process so as to produce an accuracy of 96.9%. Apart from that, in another study, (Mu'Alim & Hiday, 2022) used Random Forest for student majors at State Aliyah Madrasah with 210 training data and 90 test data resulting in an accuracy of 94.38%.

This study aims to process sales coverage optimization data using Random Forest so that the best performance results of machine learning models can be obtained. Good sales conversion data will help businesses improve the effectiveness and efficiency of the sales process by increasing the percentage of customers who take action to buy the products or services offered. This research is written in several chapters which include an introduction which contains the background of the problem, related research and solutions to the problems faced then continued with the research method chapter which contains the methods used in the research along with the explanation. The next chapter is the results of the research which contains the results of a series of studies that have been carried out as well as the conclusion chapter which contains the conclusions from the research that has been carried out.

.

LITERATURE REVIEW

Sales Conversion Optimization

Sales Conversion Optimization, also known as Conversion Rate Optimization (CRO), is a systematic process of increasing conversion rates within a website, app or other digital channel. Conversions in this context can refer to various actions a business owner wants, such as purchasing a product, signing up for a newsletter, filling out a contact form, or even watching a video all the way through. The main goal of CRO is to optimize the ratio between the number of visitors who convert compared to the total number of visitors to that site or channel. CRO is very important in digital marketing efforts because it can generate a significant increase in business results without having to drastically increase the number of visitors. By understanding visitor behavior and making changes based on data evidence, companies can improve resource use efficiency and better achieve business objectives. Various studies on CRO have been carried out (Tomescu, 2020) which examine CRO in E-commerce that uses Machine Learning to measure the level of satisfaction with website usage. In addition (Deligiannis et al., 2020) also researched CRO which is related to predicting the right time to send messages to potential buyers. Then in another study (Peng & Chen, 2022) even used a Deep Learning approach to analyze CRO in the catering industry. The various studies that have been conducted on CRO have continued to develop until now to produce several recent analyzes using various approaches to analyze and predict various cases in research.

Machine Learning

The development of information technology has encouraged the progress of various research, including in the field of Machine Learning (ML). Machine learning is a branch of artificial intelligence that refers to the ability of computer systems to learn patterns from data (data-driven) and make decisions or predictions without being explicitly programmed. According to (Batta, 2018) ML is the scientific study of algorithms and statistical models that computer systems use to carry out certain tasks without being explicitly programmed. Meanwhile, according to (Roihan et al., 2020) machine learning is a branch of artificial intelligence that is more widely used to solve various problems. The development of Machine Learning has had a far-reaching impact, and this technology continues to develop and be applied in various industries to increase efficiency, productivity and analytical capabilities.

Random Forest (RF)

RF is a robust method in machine learning. RF is a machine learning algorithm used in classification, regression and other tasks. It falls into the ensemble learning category, which means this algorithm utilizes multiple machine learning models (usually decision trees) to produce more accurate and stable predictions. Random Forest is an evolution of the Decision Tree method with several Decision Trees; each Decision Tree has been trained using individual samples and each attribute is distributed to trees selected from a random subset of attributes(Supriyadi et al., 2020). RF has the advantage of producing models that have high accuracy. This is because this model utilizes many randomly constructed decision trees and combines the results of each tree to produce a final prediction.

METHOD

This research is a type of quantitative research that uses a scientific research approach that collects, analyzes, and interprets data in the form of numbers or numerical data. This approach aims to measure relationships, identify patterns, and test hypotheses objectively using statistical methods and other quantitative analyses. An explanation of the stages of the research carried out can be seen in Figure 1 as follows :

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12943
e-ISSN : 2541-2019
p-ISSN : 2541-044X

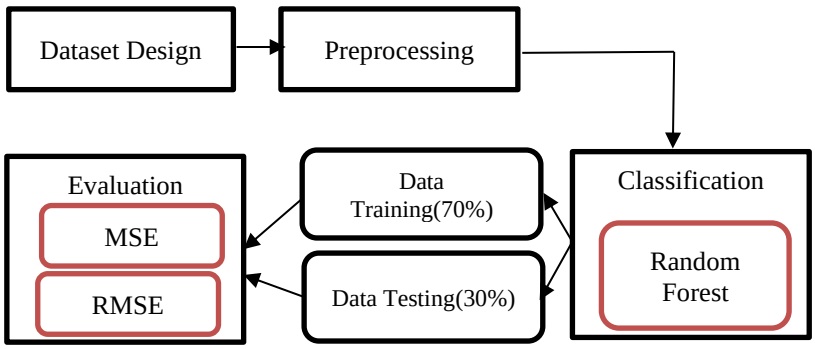

Fig. 1 Research Stages

Figure 1 shows several stages carried out in this study, namely :

Dataset Design

Dataset design refers to the process of planning, collecting, organizing, and preparing data to be used for specific purposes, such as training and evaluating machine learning models or data analysis. The steps involved in designing a dataset are very important to ensure that the data is of high quality, representative and in accordance with the objectives to be achieved. This study uses the sales conversion dataset obtained from https://www.kaggle.com/datasets/loveall/clicks-conversion-tracking

The sales conversion dataset consists of 11 attributes as shown in table 2 as follows:

Table 2. Dataset Sample

No	Name	Type	Remarks
1	ad_id	Numeric	Feature
2	xyz_campaign_id	Numeric	Feature
3	fb_campaign_id	Numeric	Feature
4	Age	Categorical	Feature
5	Gender	Categorical	Feature
6	Interest	Numeric	Feature
7	Impressions	Numeric	Feature
8	Clicks	Numeric	Feature
9	Spent	Numeric	Feature
10	Total conversion	Numeric	Feature
11	**Approved conversion**	**Numeric**	**Target**

Preprocessing

In machine learning research, the preprocessing stage is an important step to take. Data preprocessing is the initial stage of data processing where inconsistent or noisy data will be processed (Alghifari & Juardi, 2021) so that the best data structure is obtained and is useful for cleaning data (Chuzaimah Zulkifli, 2018) so as to produce quality data.

Feature Extraction

Feature extraction is the process of converting raw or unstructured data into a more structured and informative representation, which can be used for further analysis or processing by computer algorithms. In text processing, extraction features are important in changing text formats from unstructured to structured(Faisal & Nugrahadi, 2020).

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12943
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Classification

Classification in data mining is the process of categorizing or grouping data into predetermined classes or labels based on certain characteristics or features of the data. The goal is to build a model that can predict classes or labels from previously unseen data. In this study, the Random Forest method was used which is an approach that can improve accuracy in generating attributes for each node which is carried out randomly (Suci Amaliah et al., 2022) so that it is often used in research. As a comparison, there are 2 other methods tested in this study, namely :

1. **Adaptive Boosting**

   The Adaptive Boosting Method (abbreviated as AdaBoost) is a technique in machine learning that also falls into the category of ensemble learning. The main goal of AdaBoost is to improve the performance of machine learning models by combining weak models (often called "weak" because they perform slightly above random chance) into a strong model. The strength of the AdaBoost approach is in optimization, which can be used together with the naïve Bayes algorithm as an estimator algorithm to produce accuracy that can increase results with a smaller margin of error (Byna & Basit, 2020).

2. **K-Nearest Neighbor(KNN)**

   The Adaptive Boosting Method (abbreviated as AdaBoost) is a technique in machine learning that also falls into the category of ensemble learning. The main goal of AdaBoost is to improve the performance of machine learning models by combining weak models (often called "weak" because they perform slightly above random chance) into a strong model.

Random Sampling

In machine learning, random sampling refers to the technique of randomly selecting samples from the dataset used for model training or testing. This technique is often used to ensure that the dataset used represents the variation that exists in the larger population data. The dataset is divided into a training set and a test set with a composition of 70:30, 80:20 and 90:10.

Evaluation

In machine learning refers to the process of measuring the performance of a model that has been trained on data that has never been seen before. In this study used 2 evaluation approaches namely :

1. **Mean Squared Error (MSE)**

   It is a metric used in statistics and data science to measure the difference between observed values (actual or predicted data) and values predicted by a model or estimate. According to Maricar (Maricar, 2019) MSE is a calculation used to calculate the average rank error, the calculation formula is as follows :

$$\text{MSE} = (Actual - Forecast)^2 \,/\, n - 1 \qquad (1)$$

   MSE is the result of reducing the actual and forecast values which are then squared.

2. **Root Mean Squared Error (RMSE)**

   RMSE is a metric widely used in statistics and machine learning to measure the accuracy of a predictive model or the error between the predicted value and the actual observed value. According to (Sanjaya & Heksaputra, 2020) RMSE is an alternative method for evaluating forecasting techniques used to measure the accuracy of the forecast results of a model.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i}^{n}(\bar{y}i - yi)}^{2} \qquad (2)$$

   Where $\bar{y}i$= the value of the forecasting results

   yi=actual value

   n=amount of data

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12943
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

Research on sales conversion optimization using a machine learning approach goes through several research stages which can be explained as follows :

Dataset Design

The dataset used is sales conversion optimization data with a data structure as shown in table 2 as follows :

Table 2. Dataset Sample

App_conv	Ad_id	Xyz_cm_id	Fb_cm_id	Age	Gender	Interest	Imppresion	Clicks	Spent	Total_conversion
1	708746	916	103916	30-34	M	15	7350	1	1.43	2
0	708749	916	103917	30-34	M	16	17861	2	1.82	2

Preprocessing Dataset

The preprocessing process is an important step in dataset processing. In this research, the Orange application was used to carry out the preprocessing process with the display of the process results as shown in Figure 2 as follows :

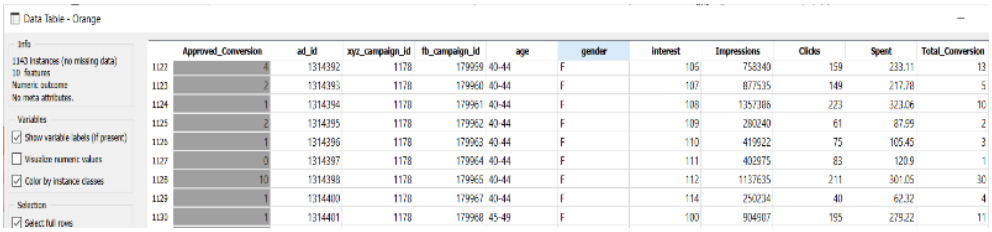

Fig. 2 Preprocessing Result

Classification

The Orange application is also used in carrying out the classification process using the Random Forest method. Orange is an open source data mining application and is widely used by researchers (Hozairi et al., 2021). The classification process uses the Random Forest method and is then compared with other methods, namely the KNN and Adaptive Booster (AdaBoost) methods. The results of the performance comparison of the three approaches above can be seen in table 3 as follows:

Table 3. Methods Performance

Methods	**MSE**	**RMSE**
K-Nearest Neighbor (KNN)	1.035	1.017
Adaptive Boosting(AdaBoost)	1.185	1.088
Random Forest	**0.966**	**0.938**

Table 3 shows the performance of the 3 methods which are compared with each other by measuring the MSE and RMSE. Both of these metrics measure how close the model's predictions are to the observed values, and their goal is to minimize the difference between the predictions and the true values.
Random Forest is clearly proven to be superior when compared to KNN and AdaBoost in predicting sales conversion optimization. The visualization results of the performance measurement of the tested method are shown in Figure 3 as follows :

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12943
e-ISSN : 2541-2019
p-ISSN : 2541-044X

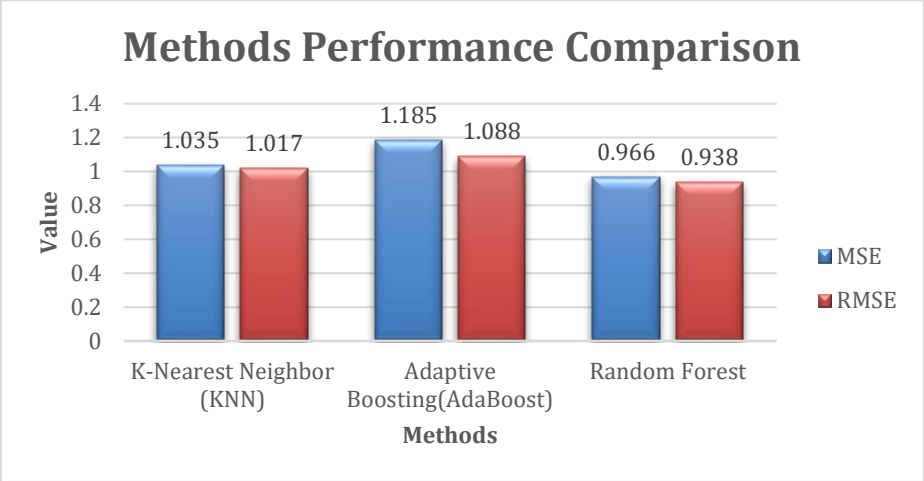

Fig. 3 Method Performance Graph

In Figure 3 it can be seen that the random forest method has the smallest MSE and RMSE values, namely MSE of 0.966 and RMSE of 0.938. It can be taken from the analysis that the random forest method has the best performance in predicting sales conversion optimization compared to the KNN and AdaBoost methods.

DISCUSSION

The data from the test results shown in table 3 shows that the Random Forest method has the smallest MSE and RMSE values, meaning the better the predictive model performance. This indicates that the model predictions are closer to the actual observed values. When compared with other similar studies, the results of measuring the performance of machine learning models can be seen in table 4 as follows :

Table 4. Comparison of similar studies

Methods	MSE
Gradient Boosting(Lee et al., 2021)	0.984
Decision Tree(Lee et al., 2021)	0.973
Proposed Method	**0.966**

It can be seen that the proposed method (Random Forest) has the best performance compared to the other 2 methods. Random Forest uses many decision trees that are built randomly. Each of these trees will have different feature selection policies, so they are likely to have different biases. However, when the prediction results from all these trees are combined, these biases tend to compensate for each other, resulting in more accurate predictions, in addition to Random Forest having a low tendency to overfitting. Overfitting occurs when the model is too complex and fits the noisy details in the training data. By using many trees, each of which tends to overfit in a different way, Random Forest can reduce the risk of overfitting, so that prediction results on test data are more consistent and have lower MSE and RMSE.

CONCLUSION

Research in the field of Sales Conversion Optimization or also known as Conversion Rate Optimization (CRO) is increasingly being carried out as a solution to weak data analysis which results in ineffective decision making. This study uses the Machine Learning approach using the Random Forest method to analyze sales conversion optimization data. Compared to other approaches such as KNN and AdaBoost, Random Forest has the best predictive performance so that it can be used as a suitable model for Conversion Rate Optimization. However, the Random Forest method needs to be tested and compared with other methods such as Neural Network which also has robust performance. Apart from that, the RF method also needs to be tested using a dataset with a larger quantity and number of features so that the best model can be obtained in the next research.

ACKNOWLEDGMENT

This research was successfully published due to support from internal research funds originating from the Directorate of Research, Community Service and Publication (DPPMP) of Universitas Stikubank.

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

REFERENCES

Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, *9*(02), 75–81. https://doi.org/10.33884/jif.v9i02.3755

Batta, M. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, *18*(8), 381–386. https://doi.org/10.21275/ART20203995

Chuzaimah Zulkifli, U. (2018). Pengembangan Modul PreprocessingTeks untuk Kasus Formalisasi dan Pengecekan Ejaan Bahasa Indonesia pada Aplikasi Web Mining Simple Solution (WMSS). *Jurnal Matematika Statistika Dan Komputasi*, *15*(2), 95. https://doi.org/10.20956/jmsk.v15i2.5718

Deligiannis, A., Argyriou, C., & Kourtesis, D. (2020). Predicting the optimal date and time to send personalized marketing messages to repeat buyers. *International Journal of Advanced Computer Science and Applications*, *11*(4), 90–99. https://doi.org/10.14569/IJACSA.2020.0110413

Faisal, M. R., & Nugrahadi, D. T. (2020). *Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah*. *8*(1), 62–69.

Lee, J., Jung, O., Lee, Y., Kim, O., & Park, C. (2021). A comparison and interpretation of machine learning algorithm for the prediction of online purchase conversion. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(5), 1472–1491. https://doi.org/10.3390/jtaer16050083

Mu'Alim, F., & Hiday, R. (2022). Implementasi Metode Random Forest Untuk Penjurusan Siswa Di Madrasah Aliyah Negeri Sintang. *Jupiter*, *14*(1), 116–125. https://www.neliti.com/publications/441871/implementasi-metode-random-forest-untuk-penjurusan-siswa-di-madrasah-aliyah-nege#cite

Normah, Rifai, B., Vambudi, S., & Maulana, R. (2022). Random Forest Classifier untuk Deteksi Penderita COVID-19 berbasis Citra CT Scan. *Jurnal Teknik Komputer AMIK BSI*, *8*(2), 174–180. https://doi.org/10.31294/jtk.v4i2

Peng, Z., & Chen, M. (2022). New Media Marketing Strategy Optimization in the Catering Industry Based on Deep Machine Learning Algorithms. *Journal of Mathematics*, *2022*. https://doi.org/10.1155/2022/5780549

Risto Miikkulainen, Myles Brundage, Jonathan Epstein, Tyler Foster, Babak Hodjat, Neil Iscoe, Jingbo Jiang, Diego Legrand, Sam Nazari, Xin Qiu, Michael Scharff, Cory Schoolland, Robert Severn, A. S. (2020). Ascend by Evolv: Artificial Intelligence-Based Massively Multivariate Conversion Rate Optimization. *AI Magazine*.

Riswandi. (2019). Transaksi On-Line (E-Commerce) : Peluang dan Tantangan Dalam Perspektif Ekonomi Islam. *Jurnal Econetica*, *13*(April), 15–38.

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *5*(1), 75–82. https://doi.org/10.31294/ijcit.v5i1.7951

Sanjaya, F. I., & Heksaputra, D. (2020). Prediksi Rerata Harga Beras Tingkat Grosir Indonesia dengan Long Short Term Memory. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *7*(2), 163–174. https://doi.org/10.35957/jatisi.v7i2.388

Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(3), 121–127. https://doi.org/10.35580/variansiunm31

Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis*, *13*(2), 67–75. https://doi.org/10.51903/e-bisnis.v13i2.247

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.