

Machine Learning

Comparison of Random Forest and XGBOOST Methods on Weather in North Sumatera

Royhan Umri Sibuea ^{1*}, Halim Maulana ²

¹ Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatra, Indonesia

² Department of Information System, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatra, Indonesia

ARTICLE INFORMATION

Received: June 19, 2025
Revised: June 27, 2025
Available Online: June 30, 2025

KEYWORDS

Machine Learning
Weather Prediction
Random Forest
XGBoost

CORRESPONDENCE (*)

Phone: +62 852-6167-5895
E-mail: royhansibuea@gmail.com

A B S T R A C T

Accurate weather forecasting is crucial for various sectors, including agriculture, transportation, and disaster management. The weather data used includes variables such as humidity, temperature, and wind speed collected from weather stations across North Sumatra. The Random Forest method is an ensemble algorithm based on decision trees known for its ability to handle overfitting and provide accurate results. On the other hand, XGBoost is a boosting technique that improves model performance through iterative learning, correcting errors made by previous models. Research results show that both methods have their respective advantages in terms of accuracy and prediction speed. The Random Forest method yields a Root Mean Squared Error (RMSE) of 0.753732 and a Coefficient of Determination (R^2) of 0.736315. In contrast, XGBoost shows a slightly lower RMSE of 0.737818 and a higher R^2 of 0.747332. It is concluded that XGBoost performs slightly better in minimizing prediction errors (RMSE) and improving model fit to the data (R^2) compared to Random Forest.

INTRODUCTION

Weather is a combination of several components such as temperature pressure, the amount of water volume in clouds, wind speed and wind direction, air pressure, and air humidity. One of the things that people all over the world really need is weather forecasting so as not to interfere with their planned activities [1,2,3].

Weather prediction is the prediction of atmospheric conditions of a place using science and technology [4,5,6]. Machine learning methods can be used in weather prediction to study complex patterns and relationships between various weather variables, such as temperature, humidity, time, wind speed and region [7,8,9].

Researchers often use machine learning for classification or prediction cases. Machine learning is the application of computers and mathematical algorithms to use learning from data to make predictions in the future. Machine learning is also used to analyze and predict weather conditions accurately [10,11]. Machine learning is divided into two approaches, namely supervised learning and unsupervised learning. Both methods aim to produce an understanding of the data generated by programmed artificial intelligence. One approach to the Supervised learning method is classification, which is a technique for grouping data based on previously trained data [12,13,14]. Supervised learning algorithms consist of random forest, XGBoost, logistic regression, support vector machine, decision tree, adaboost, naive bayes, and K-nearest neighbor [15,16,17]. This study discusses the comparison between the random forest method and XGBoost.

Based on the description of previous research, this study will discuss weather classification and predict the weather in 2024 in North Sumatra using random forest and XGBoost. What is different from previous research is that the study only focuses on two methods, namely random forest and XGBoost, while previous research used several methods. Then this study adds the results of weather predictions in North Sumatra in 2024.

METHOD

Random Forest

Random forest is a collection of several (decision trees) or decision trees used to make predictions by breaking data into several categories based on certain characteristics and using certain value comparisons to make decisions [18,19]. The random forest algorithm has several parameters that can be set including the number of decision trees to be created, the criteria for measuring the distance between each node and the maximum limit of the depth of the decision tree [20,21].

XGBoost

XGBoost builds a regression tree structure using a more regular model, which allows for better performance and the ability to reduce model complexity to avoid overfitting. The final XGBoost prediction result is the sum of the prediction results from each regression tree. Data with imbalanced classes, algorithms based on decision trees are less effective [22,23].

Analysis Design

The aim is to produce a more accurate comparison accuracy of the random forest and XGBoost methods on weather that suits user needs, therefore, researchers use a flowchart diagram to describe the comparison flow of these methods to provide an overview of the comparison process from one method to another using the python programming language to make it easier to understand and easy to understand. Below is a comparison flow of the random forest and XGBoost methods to determine which method is more accurate on weather in the form of a flowchart:

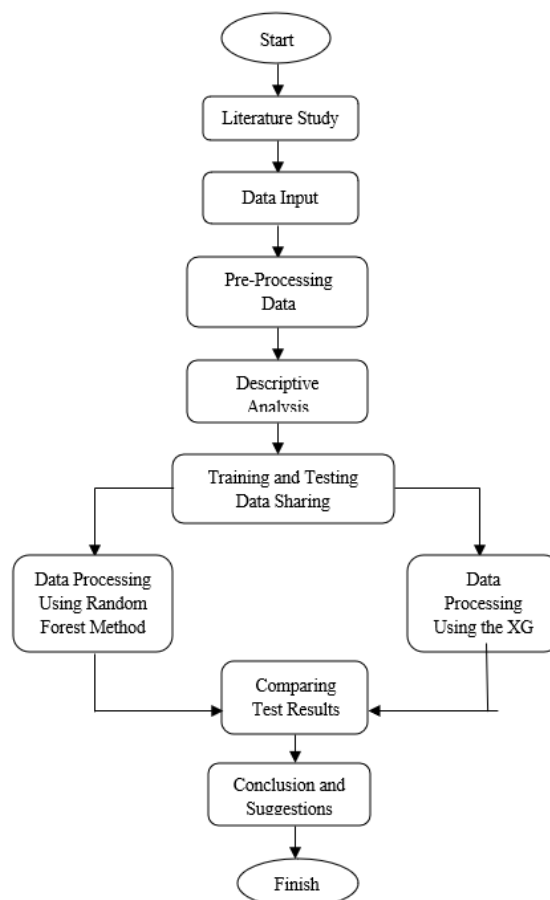


Figure 1. Research Flow

The explanation of the analysis design in the flowchart above is as follows:

1. The first process carried out by the Literature Study is to find references about the research problem. References that can be obtained from previous research journals, books related to research problems and information obtained from articles on the internet
2. The second process carried out is entering data into the program that will be used to carry out the data analysis process.
3. The third process carried out is Data Pre-Processing, the aim is to divide numerical data and categorical data.
4. The fourth process carried out is Descriptive Analysis, the aim is to see how the data will be used.
5. The fifth process carried out is dividing training data and testing data with 80% training data and 20% testing data from the total data
6. The sixth process carried out is data processing using the random forest and xgboost methods in order to prove a more accurate method for weather using variable data taken from BMKG previously, using the help of the programming language python.
7. The seventh process carried out is to compare the results of the two Random Forest and XGBoost methods
8. Then interpret the best method obtained.
9. The next process determines conclusions and suggestions. Conclusions will be taken from the whole of this study and then provide suggestions for further research.

RESULTS AND DISCUSSION

Weather Chart in North Sumatra

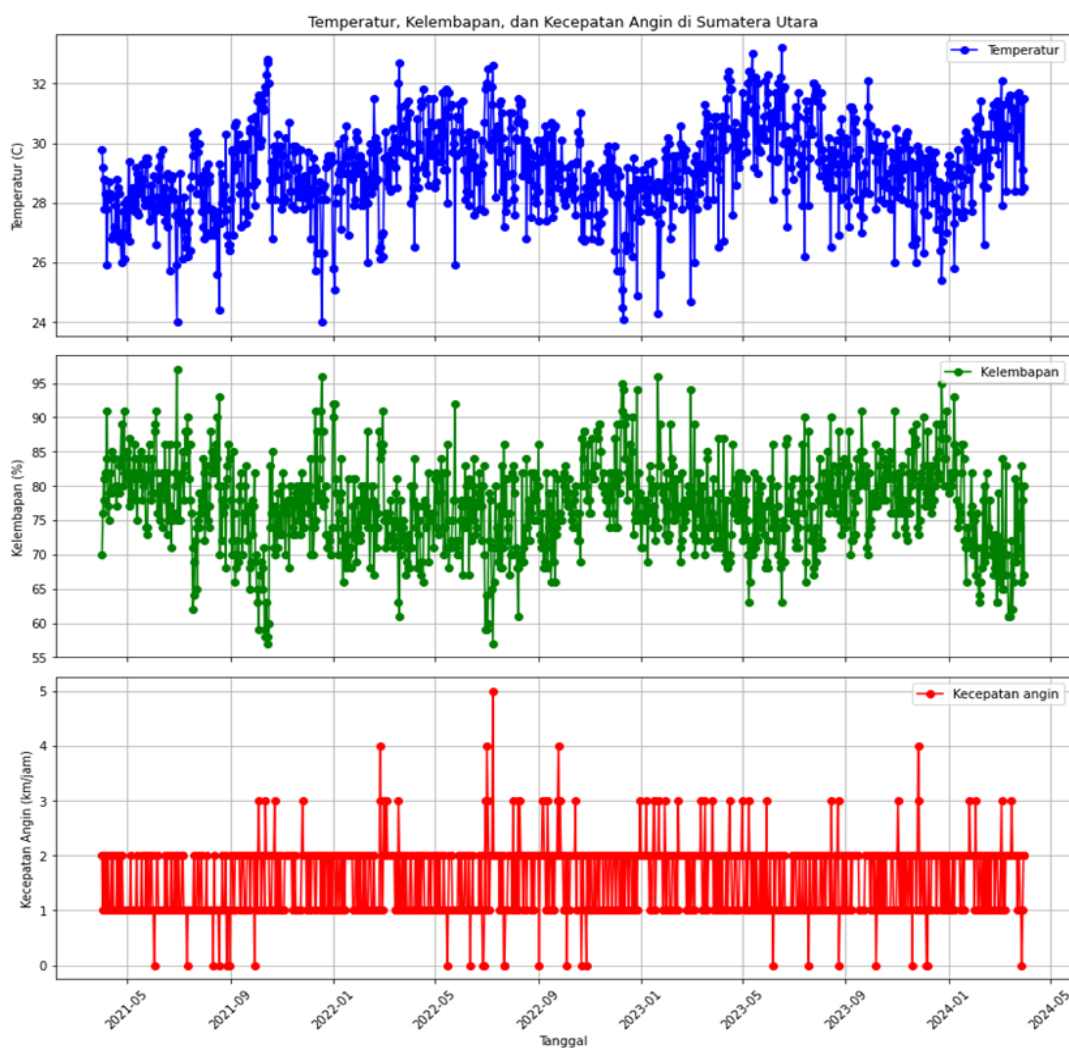


Figure 2. Weather Chart in North Sumatra

Figure 2. shows a graph of weather data in North Sumatra consisting of three parts, namely temperature, humidity, and wind speed measured from April 1, 2021 to March 31, 2024. The following is a detailed explanation of each graph:

1. Temperature (oC)

This graph shows the daily temperature changes in North Sumatra. Temperatures fluctuate between 24 oC and 33 oC with a general trend showing an increase in temperature during the observation period. The early months have lower temperatures and tend to increase until the middle of the year before decreasing slightly towards the end of the period.

2. Humidity (%)

This graph shows the daily humidity percentage in North Sumatra. Humidity ranges from 57% to 97% with most data being around 70% to 90%. Humidity is relatively stable throughout the observation period, although there are some significant fluctuations on certain days.

3. Wind speed (km/h)

This graph shows the daily wind speed in North Sumatra. Wind speeds are mostly 1 and 2 km/h, but there are some days with no wind and some even increase sharply to 4 or 5 km/h. This shows that the weather conditions are generally calm, but there are some days with significant strong winds.

Pre-Processing Data

Pre-processing data is done by dividing the data, namely training and testing with a ratio of 80:20 from the many data. So that 876 training data and 220 testing data are obtained.

Split Data

Split data is used to divide the dataset into subsets of training data and test data. In this test, `test_size = 0.2` means 20% of the total data will be used as test data, and the rest (80%) will be used as training data and `random_state = 42` is used to set the random seed as seen in Figure 3.

```
x_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
panjang_x_train = len(X_train)
print(f'Panjang X_train: {panjang_x_train}')
panjang_x_test = len(X_test)
print(f'Panjang X_test: {panjang_x_test}')
```

Figure 3. Training and Testing Data Sharing

Random Forest and XGBoost Modeling

Before modeling, you must determine the length of `x_train` and `x_text` from the entire data. Where the length of `x_train` = 876 and `x_text` 220. After the number of training and testing is determined, the next step is to build the Random Forest and XGBoost models as seen in Figures 4. and 5.

```
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Figure 4. Random Forest Model

```
model = xgb.XGBRegressor(objective='reg:squarederror', n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Figure 5. XGBoost Model

Data Analysis with Random Forest Method

Random forest is a machine learning technique that can be used for classification and regression tasks. Based on the initial research objective of predicting future weather, random forest regression was carried out where the predictor variables (X) were humidity and wind speed and the response variable (Y) was temperature.

The following is a comparison of some predicted results and actual data:

Table 1. Actual Data and Predictions of Random Forest Method

No	Actual	Prediction	Difference
1.	28.6	28.278718	-0,321282
2.	31.0	30.923164	-0,076836
3.	27.4	27.273671	-0,126329
4.	28.7	29.616175	0,916175
5.	31.0	30.465685	-0,534315
...
216.	30.0	29.445380	-0,55462
217.	29.3	29.296954	-0,003046
218.	31.0	31.401435	0,401435
219.	27.9	28.471565	0,571565
220.	30.2	30.888383	0,688383

Table 1. shows that the predicted value is quite close to the actual value indicating that the model has quite good performance. The difference between the actual and predicted values is used to measure the accuracy of the prediction model used. The model evaluation is carried out by determining the Root Mean Squared Error (RMSE) and Coefficient of Determination (R2) values.

Prediction Data Results Graph with Actual

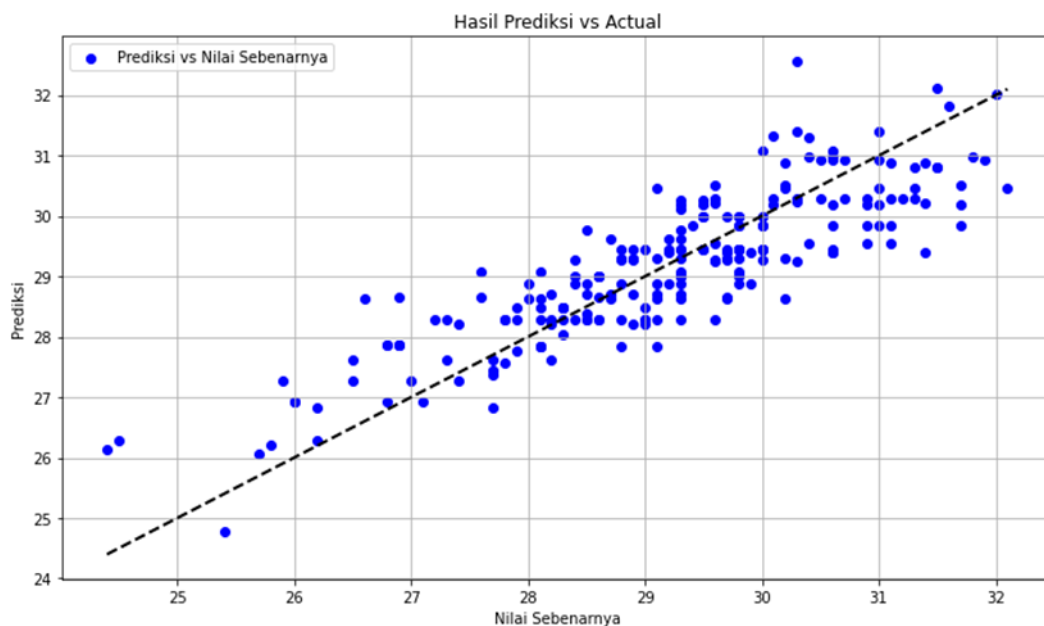


Figure 6. Prediction Results with Actual Random Forest Method

Figure 6. shows that the graph has a relationship between the predicted and actual values of a prediction model. The dashed black line represents the $y=x$ line which indicates the perfection of the prediction if all the points are exactly on it. Most of the blue points are scattered around this line, indicating that the prediction model is quite accurate although not perfect.

The denser distribution of points around the middle value on the x-axis (around 28-30) indicates that most of the actual and predicted values fall within that range. There is a clear trend that the model predictions increase as the actual values increase indicating that the model is able to capture the trend of the data well despite some variations. Overall, the model is quite good at predicting the actual values, but there is still room for improvement to reduce the deviation and improve the prediction accuracy.

Data Analysis with XGBoost Method

XGBoost (Extreme Gradient Boosting) is one of the most popular and powerful machine learning methods, especially for regression and classification problems. This method is part of the gradient boosting family that combines several weak learners to form a strong learner.

The following is a comparison of predictions and actual data.

Table 2. Actual Data and Predictions of XGBoost Method

No	Actual	Prediction	Difference
1.	28.6	28.240177	-0,359823
2.	31.0	30.840273	-0,159727
3.	27.4	27.326345	-0,073655
4.	28.7	29.536985	0,836985
5.	31.0	30.482248	-0,517752
...
216.	30.0	29.336044	-0,663956
217.	29.3	29.149357	-0,150643
218.	31.0	31.102798	0,102798
219.	27.9	28.497131	0,597131
220.	30.2	31.048227	0,848227

Table 2. overall shows that this prediction model works quite well with most of them being close to the actual values. These values can be used to evaluate the performance of the model using metrics such as Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2).

Prediction Data Results Graph with Actual

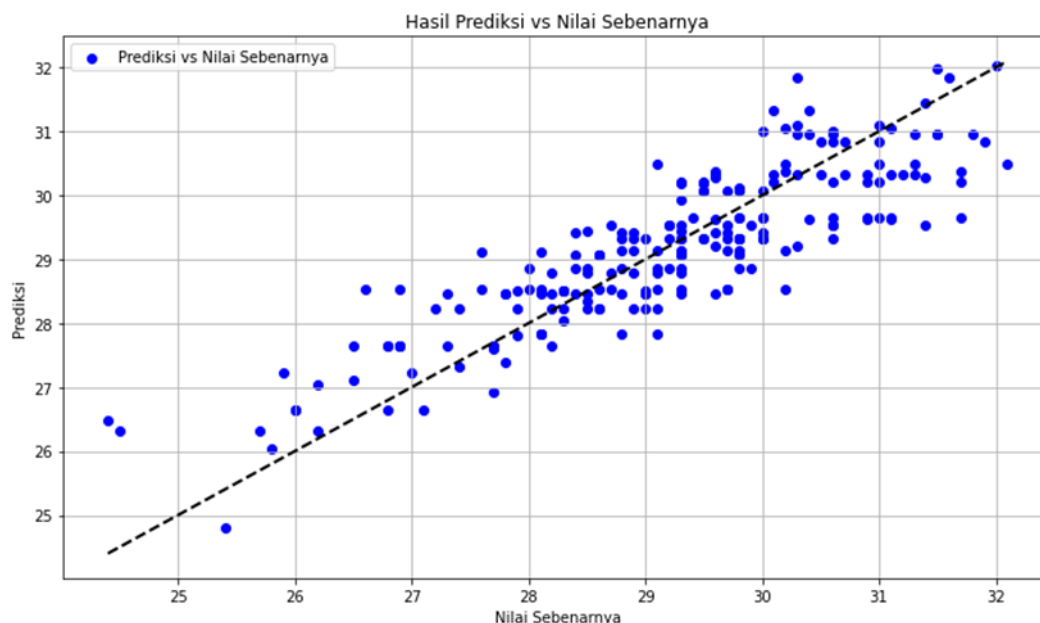


Figure 7. Prediction Results with Actual XGBoost Method

Figure 7. shows a comparison plot between predicted and actual values. Each blue dot on the graph represents a pair of predicted and actual values. The black dotted line is the reference line where the predicted values are equal to the actual values. Most of the data points are spread around the reference line indicating that the prediction model has a good trend although there is some scatter around the line. Overall, the resulting pattern shows a strong positive correlation between predicted and actual values.

CONCLUSION

The conclusion of this study is: The random forest method produces a Root Mean Squared Error (RMSE) value of 0.753732 and a Coefficient of Determination (R2) of 0.736315. On the other hand, XGBoost shows a slightly lower RMSE value of 0.737818 and a higher R2 reaching 0.747332. The smaller (closer to 0) the RMSE value, the more accurate the prediction results are, while the larger (closer to $+\infty$) the worse the prediction results are, and if the R2 value is closer to $-\infty$, the worse the prediction results are, while the R2 value is closer to 1, the more accurate the prediction results are. It is concluded that XGBoost has slightly better performance in terms of minimizing prediction errors (RMSE) and improving model fit to data (R2) compared to random forest. Although the difference is not significant, these results indicate that XGBoost is more suitable for obtaining more accurate predictions and in accordance with actual data. Seeing the results of the predicted values with the XGBoost method is closer to the actual value than the random forest method. So XGBoost has better performance in capturing complex data patterns and variations. Below is a graph of the results of the predicted data with the actual XGBoost method.

REFERENCES

- [1] Al'afi, A. M., Widiart, W., Kurniasari, D., & Usman, M. (2020). Peramalan Data Time Series Seasonal Menggunakan Metode Analisis Spektral. *Jurnal Siger Matematika*, 1(1), 10–15. <https://doi.org/10.23960/jsm.v1i1.2484>
- [2] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>
- [3] Dwiyantri, Z. A., & Prianto, C. (2023). Prediksi Cuaca Kota Jakarta Menggunakan Metode Random Forest. *Jurnal Tekno Insentif*, 17(2), 127–137. <https://doi.org/10.36787/jti.v17i2.1136>

- [4] Farhanuddin, Sarah Ennola Karina Sihombing, & Yahfizham. (2024). Komparasi Multiple Linear Regression dan Random Forest Regression Dalam Memprediksi Anggaran Biaya Manajemen Proyek Sistem Informasi. *Journal of Computers and Digital Business*, 3(2), 86–97. <https://doi.org/10.56427/jcbd.v3i2.408>
- [5] Gusliana, I. (2021). Bab II Landasan Teori. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- [6] Sari, I.P., Al-Khowarizmi, A.K., & Batubara, I.H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, Vol. 2 No. 1, page 139-144
- [7] Sari, I.P., Batubara, I.H., & Al-Khowarizmi, A.K. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. *International Journal of Economic, Technology and Social Sciences (Injects)*, Vol. 2 No. 1, page 331- 338
- [8] Sari, I.P., Fahroza, M.F., Mufit, M.I., & Qathrunad, I.F. (2021). Implementation of Dijkstra's Algorithm to Determine the Shortest Route in a City. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, Vol. 2 No. 1, page 134-138
- [9] Sari, I.P., Al-Khowarizmi, A.K., Ramadhani, F., & Sulaiman, O.K. (2023). Implementation of the Selection Sort Algorithm to Sort Data in PHP Programming Language. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, Vol. 4 No. 1, page 377-381
- [10] Manurung, A.A., Nasution, M.D., & Sari, I.P. (2023). Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Disease Classification. *2023 11th International Conference on Cyber and IT Service Management (CITSM)*
- [11] Ramadhani, F., Satria, A., & Sari, I.P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer 2 (2)*, 58-62
- [12] Sari, I.P., Batubara, I.H., Ramadhani, F., & Wardani, S. (2022). Perancangan Sistem Antrian pada Wahana Hiburan dengan Metode First In First Out (FIFO). *Sudo Jurnal Teknik Informatika 1 (3)*, 116-123
- [13] Ramadhani, F., Satria, A., & Sari, I.P. (2022). Aplikasi internet berbasis website sebagai E-Commerce penjualan komponen sport car. *Blend Sains Jurnal Teknik 1 (2)*, 69-75
- [14] Sari, I.P., Al-Khowarizmi, A., & Ramadhani, F. (2021). User Interface Prototype Using User Centered System Design Method in Motorvice Information System. *2021 International Conference on Computer Science and Engineering (IC2SE) 1*, 1-6
- [15] Ramadhani, F., Al-Khowarizmi, A.K., & Sari, I.P. (2021). Improving the Performance of Naïve Bayes Algorithm by Reducing the Attributes of Dataset Using Gain Ratio and Adaboost. *2021 International Conference on Computer Science and Engineering (IC2SE) 1*, 1-5
- [16] Sitompul, D.N., Rahmatika, A., & Sari, I.P. (2023). Application of The Sales and Purchase Program Using The Rapid Application Development Model. *Al'adzkiya International of Computer Science and Information Technology (AIOCSIT) Journal*, Vol. 4 No. 1, page 6-16
- [17] Sari, I.P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. *Hello World Jurnal Ilmu Komputer 2 (3)*, 146-157
- [18] Batubara, I.H., Sari, I.P., Siregar, E.F.S., & Lubis, B.S. (2021). Meningkatkan Kemampuan Penalaran Matematika Melalui Metode Penemuan Terpandu Berbantuan Software Autograph. *Seminar Nasional Teknologi Edukasi Sosial dan Humaniora 1 (1)*, 699-705
- [19] Apdilah, D., & Sari, I.P. (2021). Optimization Of The Fuzzy C-Means Cluster Center For Credit Data Grouping Using Genetic Algorithms. *Al'adzkiya International of Computer Science and Information Technology (AIOCSIT) Journal*, Vol. 2 No. 2, page 156-163
- [20] Sulaiman, O.K., & Sari, I.P. (2021). Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule. *Al'adzkiya International of Computer Science and Information Technology (AIOCSIT) Journal*, Vol. 2 No. 2, page 128-135
- [21] Sari, I.P., Hariani, P.P., Al-Khowarizmi, A.K., Ramadhani, F., Sulaiman, O.K., Satria, A., & Manurung, A.A. (2024). CLUSTERING HIV/AIDS DISEASE USING K-MEANS CLUSTERING ALGORITHM. *Proceeding International Seminar on Islamic Studies*. Vol. 5, No. 1 (2024), 1668-1676
- [22] Sari, I.P., Ramadhani, F., Satria, A., & Sulaiman, O.K. (2023). Leukocoria Identification: A 5-Fold Cross Validation CNN and Adaboost Hybrid Approach. *2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. 486-491
- [23] Sari, I.P., Al-Khowarizmi, A.K., Sulaiman, O.K., & Apdilah, D. (2023). Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, Vol. 4 No. 2, page 402-412

- [24] Sari., I.P, Batubara., I.H, Al-Khowarizmi., A, & PP Hariani. (2022). Perancangan Sistem Informasi Pengelolaan Arsip Digital Berbasis Web untuk Mengatur Sistem Kearsipan di SMK Tri Karya. *Wahana Jurnal Pengabdian kepada Masyarakat 1 (1)*, 18-24
- [25] Habibi., F, Qathrunada., I.F, & Anggraini., T. (2023). "Design and Build a Tourism Website Using Shopify Framework". *Hanif Journal of Information Systems*. Vol. 1 No. 1, 2023.
- [26] Sari., I.P, A Syahputra, N Zaky, RU Sibuea, & Z Zakhir. (2022). Perancangan sistem aplikasi penjualan dan layanan jasa laundry sepatu berbasis website. *Blend sains jurnal teknik 1 (1)*, 31-37
- [27] Sari., I.P, A Azzahrah, FQ Isnaini, L Nurkumala, & A Thamita. (2022). Perancangan sistem absensi pegawai kantoran secara online pada website berbasis HTML dan CSS. *Blend sains jurnal teknik 1 (1)*, 8-15
- [28] Septiana., D. (2024). Forecasting Rice Prices with Holt-Winter Exponential Smoothing Model. *Hanif Journal of Information Systems*. Vol. 1 No. 2, 2024.
- [29] Sari.,I.P, & Ramadhani., F. (2021). Pengaruh Teknologi Informasi Terhadap Kewirausahaan Pada Aplikasi Perancangan Jual Beli Jamu Berbasis WEB. *Prosiding Seminar Nasional Kewirausahaan 2 (1)*, 874-878.
- [30] Satria., A, Ramadhani., F, & Sari, I.P. (2023). Rancang Bangun Sistem Informasi Penerimaan Peserta Didik Baru (PPDB) Sekolah Menengah Kejuruan Telkom 2 Medan Menggunakan Codeigniter. *Wahana Jurnal Pengabdian kepada Masyarakat 2 (1)*, 23-31
- [31] Sari., I.P, A Jannah, AM Meuraxa, A Syahfitri, & R Omar. (2022). Perancangan Sistem Informasi Penginputan Database Mahasiswa Berbasis Web. *Hello World Jurnal Ilmu Komputer 1 (2)*, 106-110.
- [32] Mahardika., F, & Abdillah., M.L. (2024). Design of Unified Modeling Language Information System for Motorcycle Unit Selling and Buying Transactions using the Waterfall Method. *Hanif Journal of Information Systems*. Vol. 1 No. 2, 2024.
- [33] Sari., I.P, & Batubara., I.H. (2021). Perancangan Sistem Informasi Laporan Keuangan Pada Apotek Menggunakan Algoritma K-NN. *Seminar Nasional Teknologi Edukasi dan Humaniora (SiNTESa) 1 (2021 - ke 1*
- [34] Sari., I.P, & Batubara., I.H. (2021). User Interface Information System for Using Account Services (Joint Account) WEB-Based. *International Journal of Economic, Technology and Social Sciences (Injects)*, 462-469
- [35] Sari., I.P, Al-Khowarizmi., A, & Batubara., I.H. (2021). Implementasi Aplikasi Mobile Learning Sistem Manajemen Soal dan Ujian Berbasis Web Pada Platform Android. *IHSAN: JURNAL PENGABDIAN MASYARAKAT 3 (2)*, 178-183
- [36] Mudafri., H.A. (2024). Design of a Web-Based Coffeeshop Ordering Information System. *Hanif Journal of Information Systems*. Vol. 1 No. 2, 2024.
- [37] Sari., I.P, Hariani., P.P, Satria., A, & Manurung., A.A. (2023). Rancang Bangun Sistem Informasi Pengelolaan Arsip Materi Ajar Berbasis Web untuk Guru MAS Darul Falah. *Wahana Jurnal Pengabdian kepada Masyarakat 2 (2)*, 59-65
- [38] Ramadhani., F, & Sari., I.P. (2021). Pemanfaatan Aplikasi Online dalam Digitalisasi Pasar Tradisional di Medan. *Prosiding Seminar Nasional Kewirausahaan 2 (1)*, 806-811
- [39] Sari., I.P, Sulaiman., O.K, & Apdillah, D. (2024). Rancang Bangun Game Zombie Menggunakan Kodular Berbasis Android. *Jurnal Minfo Polgan 13 (1)*, 293-302
- [40] Ichsan., A, Siambaton., M.Z, & Nasution., K. (2023). "Android-Based Practical Work Student Registration Form Application System Design". *Hanif Journal of Information Systems*. Vol. 1 No. 1, 2023.
- [41] Sari., I.P, Sulaiman., O.K, Ramadhani., F, & Satria., A. (2023). Perancangan Sistem Manajemen Surat Berbasis Web Pada Kantor Camat Tano Tombangan Angkola. *INCODING: Journal of Informatics and Computer Science Engineering 3 (2)*, 61-76.
- [42] Jannah., A, Meuraxa., A.M, & Azzahrah., A. 2023. "Web Based E-Commerce System Design at EXO Shop Using The Waterfall Method". *Hanif Journal of Information Systems*. Vol. 1 No. 1, 2023.
- [43] Sari., I.P, Al-Khowarizmi., A, , Jannah., A, Meuraxa., A.M, & Tanjung., M.I. (2023). Web-Based Offline Game Suit Design: A Model Overview. *Journal of Computer Science, Information Technology and Telecommunication Engineering 4 (2)*, 389-394.
- [44] Guntur., S, Ichsan., A, & Sari., I.P. (2024). Designing a Web-Based Mail Management System at the Beringin Helvetia Sub-district Office. *Altafani: Jurnal Pengabdian Masyarakat 1 (1)*
- [45] Sari., I.P, Sulaiman., O.K, Al-Khowarizmi., A, & Azhari., M. (2023). Perancangan Sistem Informasi Pelayanan Masyarakat pada Kelurahan Sipagimbar dengan Metode Prototype Berbasis Web. *Blend Sains Jurnal Teknik 2 (2)*, 125-134.
- [46] Hutasuhut., B.K, Sari., I.P, & Al-Khowarizmi, A.K. (2023). Analysis the Effect of Digitalization and Technology on Web-Based Entrepreneurship. *Journal of Computer Science, Information Technology and Telecommunication Engineering*

- [47] Naibaho, R.F., & Sari, I.P. (2024). Implementasi Metode Gray Level Co-Occurrence Matrix Menganalisis Tekstur Kulit Wajah. *sudo Jurnal Teknik Informatika* 3 (4), 172-182
- [48] Erlangga, F., & Sari, I.P. (2024). Perancangan Sistem Untuk Merekomendasikan Produk Skincare Menggunakan Metode NLP. *Portal Riset dan Inovasi Sistem Perangkat Lunak* 2 (4), 1-11