

PREDICTION OF FLIGHT DELAYS USING FEATURE ENGINEERING, CATBOOST, AND BAYESIAN OPTIMIZATION TO IMPROVE MODEL PERFORMANCE

Ilham Maulana¹, Siti Ernawati², Risa Wati³

¹Ilmu Komputer / Fakultas Teknologi Informasi
Universitas Nusa Mandiri
k4ilham@gmail.com

²Sistem Informasi / Fakultas Teknologi Informasi
Universitas Nusa Mandiri
siti.ste@nusamandiri.ac.id

³Sistem Informasi / Fakultas Teknik dan Informatika
Universitas Bina Sarana Informatika
risawati.rwx@bsi.ac.id

Abstract

Flight delays have become a major issue in the aviation industry, impacting operational efficiency and customer satisfaction. This study proposes a CatBoostClassifier-based approach combined with Feature Engineering, Bayesian Optimization, and Random Over Sampling techniques to improve the accuracy of flight delay predictions. Based on model evaluation results, the use of Feature Engineering and Bayesian Optimization enhances performance compared to the baseline CatBoost model. The CatBoost+FE+Bayes combination achieves an accuracy of 83.32%, higher than the unmodified CatBoost model, which only reaches 82.95%. However, applying the Random Over Sampling technique in the CatBoost+FE+Bayes+ROS combination decreases model performance, reducing accuracy to 81.44%. Regarding other metrics, the CatBoost+FE+Bayes model demonstrates the highest F1-score of 0.62, indicating a balance between precision and recall. Additionally, the Area Under Curve (AUC) analysis reveals that CatBoost+FE+Bayes has the highest AUC value of 0.7793, followed by CatBoost+FE at 0.7768, and the unmodified CatBoost model at 0.7643. Meanwhile, the application of ROS leads to a decrease in AUC value to 0.6787. These findings suggest that utilizing Feature Engineering and Bayesian Optimization significantly enhances flight delay predictions. However, resampling techniques such as ROS do not always positively impact the tested model and can even degrade classification performance. The objective of this research is to develop a more accurate flight delay prediction model through the application of appropriate optimization techniques. The resulting model is expected to improve prediction quality and benefit the aviation industry by optimizing operational efficiency and minimizing the negative impact of delays on passengers.

Keywords: Flight delays; CatBoost; Feature Engineering; Bayesian Optimization; Random Over Sampling

Abstrak

Keterlambatan penerbangan menjadi permasalahan utama dalam industri penerbangan, yang berdampak pada efisiensi operasional dan kepuasan pelanggan. Penelitian ini mengusulkan pendekatan berbasis CatBoostClassifier dengan kombinasi teknik Feature Engineering, Bayesian Optimization, dan Random Over Sampling untuk meningkatkan akurasi prediksi keterlambatan penerbangan. Berdasarkan hasil evaluasi model, penggunaan Feature Engineering dan Bayesian Optimization memberikan peningkatan kinerja dibandingkan model CatBoost murni. Model dengan kombinasi CatBoost+FE+Bayes mencapai akurasi 83,32%, lebih tinggi dibandingkan CatBoost tanpa modifikasi yang hanya mencapai 82,95%. Namun, penerapan teknik Random Over Sampling dalam kombinasi CatBoost+FE+Bayes+ROS justru menurunkan performa model, dengan akurasi turun menjadi 81,44%. Dari segi metrik lainnya, model CatBoost+FE+Bayes menunjukkan F1-score tertinggi sebesar 0,62, yang menunjukkan keseimbangan antara precision dan recall.

Selain itu, analisis Area Under Curve (AUC) menunjukkan bahwa CatBoost+FE+Bayes memiliki nilai AUC tertinggi sebesar 0.7793, diikuti oleh CatBoost+FE dengan 0.7768, dan CatBoost tanpa modifikasi dengan 0.7643. Sementara itu, penerapan ROS menyebabkan penurunan nilai AUC menjadi 0.6787. Hasil penelitian ini menunjukkan bahwa pemanfaatan Feature Engineering dan Bayesian Optimization dapat meningkatkan prediksi keterlambatan penerbangan secara signifikan. Namun, teknik resampling seperti ROS tidak selalu memberikan dampak positif pada model yang diuji dan justru dapat menurunkan performa klasifikasi. Tujuan dari penelitian yaitu mengembangkan model prediksi keterlambatan penerbangan yang lebih akurat melalui penerapan teknik optimasi yang sesuai. Model yang dihasilkan diharapkan mampu meningkatkan kualitas prediksi serta memberikan manfaat bagi industri penerbangan dalam mengoptimalkan efisiensi operasional dan meminimalkan dampak negatif keterlambatan terhadap penumpang.

Kata Kunci: Keterlambatan Penerbangan; CatBoost; Feature Engineering; Bayesian Optimization; Random Over Sampling

INTRODUCTION

The aviation industry plays a crucial role in global mobility, enabling fast and efficient travel for millions of people every day. However, flight delays pose a significant challenge, affecting airline operational efficiency, leading to financial losses, and even impacting passenger satisfaction (Hatipoğlu & Tosun, 2024)(Ardhana et al., 2022)(Jha, Jha, Pandey, & Babiceanu, 2019) as well as airport management. Departure time, aircraft type, and weather conditions can be processed to provide better insights into the factors contributing to delay (Herdian, Kamila, & Agung Musa Budidarma, 2024). Therefore, building an accurate flight delay prediction model is crucial for airlines to take swift action and improve service quality.

In this study, a machine learning-based approach is used to predict flight delays by optimizing model performance through Feature Engineering, CatBoost, and Bayesian Optimization. Feature Engineering is employed to extract and transform relevant features, enabling the model to learn more effectively (Verdonck, Baesens, Óskarsdóttir, & vanden Broucke, 2024). It can also transform a set of problematic features into strong factors for the targeted property, enabling faster computational processes and improved accuracy (Hu et al., 2021). CatBoost, as one of the gradient boosting algorithms that excels in handling categorical data, is chosen for its ability to improve prediction accuracy (Alfarhood et al., 2024). In machine learning, many automated solutions have been developed for hyperparameter optimization, with Bayesian Optimization being one of the most popular methods (BO) (Cho et al., 2020). Bayesian Optimization is used to determine the optimal and faster combination of hyperparameters to

enhance the overall performance of the model. (Victoria & Maragatham, 2021).

Research on flight delay prediction has advanced rapidly between 2020 and 2024, focusing on the application of feature engineering techniques, the CatBoost algorithm, and Bayesian optimization to improve model accuracy. Several studies, including the research conducted by Ningthoukhongjam et al., highlight the importance of feature engineering in identifying significant variables that influence flight delays. This study developed a hybrid model combining random forest, decision tree, and gradient boosting, demonstrating improved accuracy in predicting delay (Ningthoukhongjam, Alam, Kumar, & G, 2024). Additionally, research applying ensemble gradient boosting techniques for flight delay prediction found that the CatBoost algorithm effectively handles the complexity of flight data, achieving an accuracy score of 0.68 (Khan, Akbar, & Zahed, 2022). Research on hyperparameter optimization has explored the use of Bayesian optimization for tuning machine learning model hyperparameters. It found that the proposed model significantly enhances performance by identifying the optimal hyperparameter combination more efficiently than traditional methods (Victoria & Maragatham, 2021).

This study aims to develop a more accurate flight delay prediction model by leveraging appropriate optimization techniques. The resulting model is expected to provide better predictions and be beneficial for the aviation industry in improving operational efficiency and reducing the negative impact of delays on passengers (Qalbi & Jayadi, 2020).

RESEARCH METHODS

The research was conducted using an experimental method, involving several stages from

dataset preparation to evaluation. Figure 1 illustrates the stages carried out during the research process.

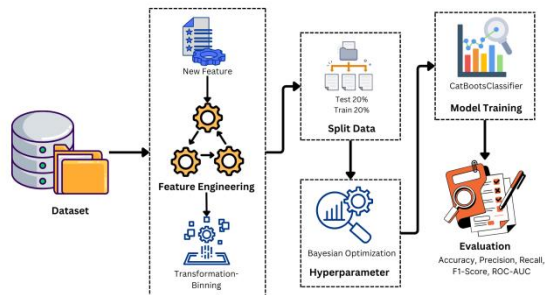


Figure 1. Research Stages

3.1. Dataset

This study uses the Flight Delays Fall 2018 dataset, obtained from Kaggle <https://www.kaggle.com/competitions/flight-delays-fall-2018/data>. The dataset consists of 100,000 data. Eight features and one target variable are used in this experiment. Each feature and the target variable are explained in Table 1.

Table 1. Features and Target Used in the Experiment

Type	Name	Description
Fitur	Month	Month of Departure
Fitur	DayofMonth	Day of Departure within a Month
Fitur	DayOfWeek	Day of Departure within a Week
Fitur	DepTime	Time of Departure
Fitur	UniqueCarrier	Unique Airline Code
Fitur	Origin	Departure Airport
Fitur	Dest	Destination Airport
Fitur	Distance	Distance Between Origin and Destination Airports
Target	dep_delayed_15min	Delay Over 15 Minutes (whether the flight was delayed by more than 15 minutes)

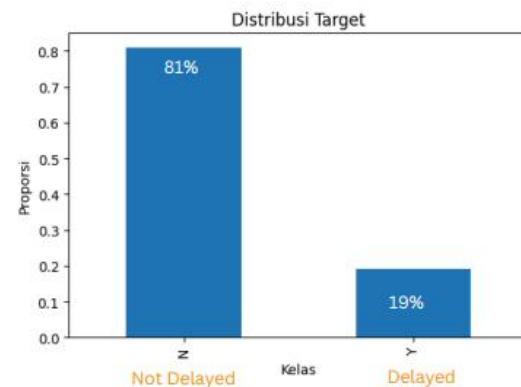


Figure 2. Data Distribution graph Based on Target

Figure 2 shows the data distribution graph based on the target variable used in the experiment. The graph illustrates the distribution of the target into two classes: Not Delayed and Delayed. It is observed that 81% of the data falls into the Not Delayed category, while only 19% belongs to the Delayed category. This indicates that the dataset is imbalanced, with a significantly larger number of samples in the Not Delayed category compared to the Delayed category. Such an imbalance can impact the performance of machine learning models. Therefore, a technique to address data imbalance, namely Random Over Sampling (ROS), was chosen. Random Over Sampling (ROS) is a data balancing technique used to handle class imbalance issues in datasets (Fitriani, Yasin, & Tarno, 2021). This technique works by randomly duplicating samples from the minority class to balance the number of samples with the majority class (Qalbi & Jayadi, 2020).

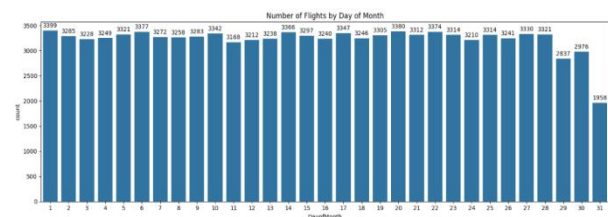


Figure 3. Graph of the Number of Flights in One Month

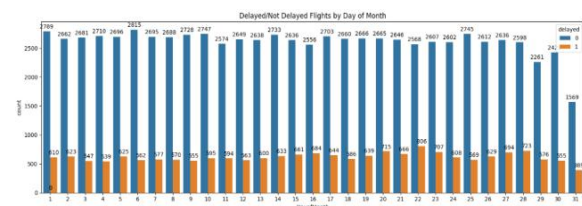


Figure 4. Graph of the Number of Delayed (Delayed = 1) and Non-Delayed (Non-Delayed = 0) Flights in One Month

Figure 3 shows the number of flights in one month, while Figure 4 presents the graph of delayed flights (delayed = 1, orange) and non-delayed flights (non-delayed = 0, blue) based on the day of the month. The peak number of flights occurred on the 6th day, with the highest total flights (2,815), followed by relatively stable numbers on other days. The last days of the month (30 and 31) show a decline in total flights, but the delay ratio remains relatively high. Certain days, such as the 17th and 30th, exhibit an increase in delayed flights, possibly indicating external factors influencing delays.

3.2. Feature Engineering

Feature engineering is the process of creating, selecting, or modifying features (variables) in a dataset to improve the performance of a machine learning model (Rajendran & Karthi, 2022). Feature Engineering is performed to improve the model's accuracy in predicting flight delays. It is a crucial step in developing predictive models, where raw data is processed and transformed into a more meaningful and suitable format for analysis (Herdian et al., 2024) (Yun, Yoon, & Won, 2021). A standard model can be enhanced into a highly useful and accurate predictive and decision-making tool with well-designed features (Ahmmad, Labassi, Alsuraiheed, Mahmood, & Khan, 2024).

Feature Engineering techniques that can be applied in the experiment include creating new features (new feature) and transformation-binning.

- 1) New Feature: Creating new features based on the combination of existing features to better represent the original data. Table 2 presents the new features used in the experiment.

Table 2. New Features Resulting from the Combination of Existing Features

New Feature	Description
Route	Combination of Origin and Destination Airports
UniqueCarrier_Origin	Combination of Airline Code with Origin Airport
UniqueCarrier_Dest	Combination of Airline Code with Destination Airport
is_weekend	Whether the Flight is on a Weekend
Hour & minute	Departure Time
hour_sq & hour_sq2	Departure Hour Feature (Squared Hour)

- 2) Transformasi-Binning: Converting numerical data into categorical data by dividing numerical values into several intervals or groups (bins). This method helps the model handle data more effectively. The table presents the transformation-binning features used in the experiment.

Table 3. New Features Resulting from Transformation

Transformasi - Binning Feature	Description
Season	Summer, Autumn, Winter, Spring diambil dari Month
Departure Time (DepTime_bin)	Morning, Afternoon, Evening diambil dari DepTime
Distance (Dist_bin)	Vshort, Short, Mid, Midlong, Long diambil dari Distance

3.3. Split Data

This experiment splits the data into two parts: 80% for training and 20% for testing. This division ensures that the model learns patterns from the majority of the available data while the remaining portion is used to evaluate the model's performance on unseen data. By allocating 80% for training data, the model gains sufficient information to recognize patterns and generalize effectively, while the 20% test data serves to assess how well the model classifies new data with a good level of accuracy.

3.4. Hyperparameter dengan Bayesian Optimization

To improve model performance, hyperparameter tuning is performed using Bayesian Optimization (Wu et al., 2019). The hyperparameter search process utilizes the scikit-learn library, which plays a crucial role in exploring various hyperparameter combinations (Ernawati & Wati, 2024). The tool used for automatic hyperparameter optimization using the Bayesian Optimization method is BayesSearchCV. This tool serves as an alternative to GridSearchCV or RandomizedSearchCV for finding the best hyperparameter combinations and can compare test results in machine learning model (Ernawati, Wati, & Nuris, 2022). BayesSearchCV uses probability to determine more promising parameter searches by leveraging information from previous iterations. With a probabilistic approach, BayesSearchCV finds optimal results faster compared to more exploratory searches like GridSearchCV.

Bayesian Optimization applies Bayes' Theorem to select the best values in an optimization process. The equation for Bayes' Theorem is: (Victoria & Maragatham, 2021)

$$P(Z | Y) = \frac{P(Y | Z)P(Z)}{P(Y)} \quad (1)$$

Where $P(Z|Y)$ is the posterior probability. $P(Y|Z)$ is the likelihood, which represents the probability of obtaining YYY given that ZZZ is known. $P(Z)$ is the prior probability, which is the initial probability before observing the data YYY. $P(Y)$ is the marginal probability.

3.5. Model Training-CatBoost

CatBoostClassifier is a gradient boosting-based machine learning algorithm (combining multiple simple models) specifically designed to efficiently handle categorical data (Fitriani et al., 2021) (Sadaf, 2023). This algorithm is part of CatBoost, a machine learning library developed by Yandex. The advantage of CatBoostClassifier lies in its ability to handle categorical features and imbalanced datasets efficiently (Fitriani et al., 2021) (et al., 2023), Making it an effective choice for various classification task.

3.6. Evaluasi Model

The trained model is evaluated using Accuracy, Precision, Recall, and F1-Score metrics to measure its effectiveness in predicting flight delays. Additionally, ROC-AUC is used to assess the model's ability to distinguish between on-time and delayed flights.

RESULTS AND DISCUSSION

In the experiment phase, a series of tests were conducted to evaluate the model's performance in predicting flight delays. The dataset used in this study was divided into two parts: 80% as training data and 20% as testing data. This split was intended to ensure that the model could learn from the majority of available data, while the test data was used to assess the model's performance on previously unseen data.

The model training process utilized the CatBoost algorithm, which was chosen for its ability to handle categorical data directly and its superior performance compared to other boosting algorithms. Additionally, Feature Engineering was applied to extract and process influential features related to flight delays, such as departure schedules, weather conditions, and air traffic congestion. To improve prediction accuracy,

Bayesian Optimization was employed for hyperparameter selection, allowing the model to achieve optimal performance without relying on brute-force parameter search.

The analysis results indicated that the optimized model achieved higher accuracy than the baseline model without optimization. Based on evaluation metrics such as accuracy, precision, recall, and F1-score, the model with Bayesian optimization demonstrated significant improvements in classifying flight delays. Furthermore, an error analysis was conducted to identify prediction errors, revealing that weather conditions and departure schedules had a substantial impact on the model's prediction accuracy. Figure 5. The following presents the model's performance based on evaluation metrics.

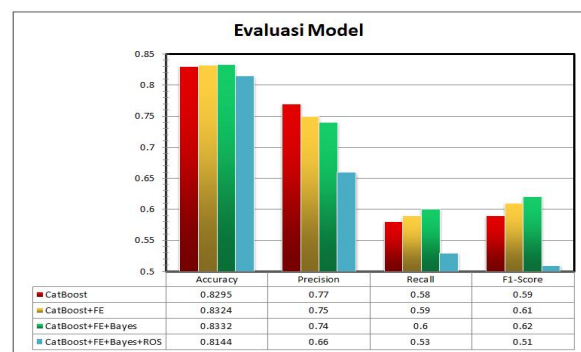


Figure 5. Model Performance Based on Evaluation Metrics

Based on the experimental results, it can be concluded that the combination of Feature Engineering (FE), Bayesian Optimization, and CatBoost produces the best-performing model for predicting flight delays. The following are some key findings from the evaluation metrics analysis:

1. Accuracy Improvement with Feature Engineering and Bayesian Optimization
 - a) The initial CatBoost model had an accuracy score of 0.8295
 - b) After applying Feature Engineering (FE)
 - c) With the addition of Bayesian Optimization, the accuracy increased to 0.8332, indicating that hyperparameter tuning successfully improved the model's performance.
2. Precision and Recall
 - a) Precision slightly decreased after Bayesian Optimization (from 0.77 in the initial CatBoost model to 0.74 in the CatBoost+FE+Bayes model), but recall improved from 0.58 to 0.6

b) The model with the FE + Bayesian Optimization combination achieved the highest F1-score (0.62) compared to other models

3. Impact of ROS (Random OverSampling) on Model Performance

The CatBoost+FE+Bayes+ROS model showed a decrease in precision to 0.66 and recall to 0.53, negatively affecting the F1-score (0.51). This indicates that applying ROS in this scenario did not provide a significant improvement and could potentially disrupt the model's balance

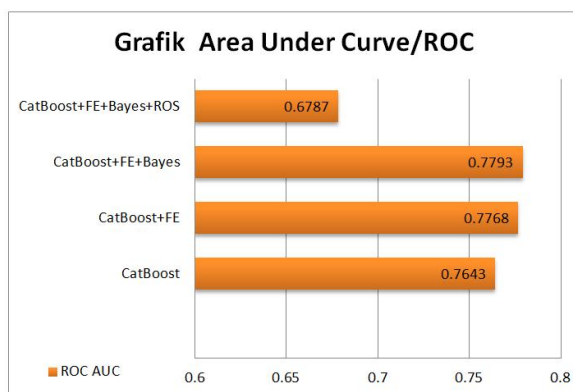


Figure 6. Graph Area Under Curve (AUC)/ ROC

Based on Figure 6, the Area Under Curve (AUC) / ROC graph displayed allows for several conclusions to be drawn:

1. Increase in AUC with Feature Engineering (FE) dan Bayesian Optimization (BO)

The unmodified CatBoost model has an AUC value of 0.7643. After applying Feature Engineering (FE), the AUC value increases to 0.7768, indicating that feature engineering improves model performance. When Bayesian Optimization is applied after Feature Engineering (CatBoost+FE+Bayes), the AUC value further increases to 0.7793, which is the highest value in the graph.

2. Impact of Using Random Over Sampling (ROS)

The CatBoost+FE+Bayes+ROS model experienced a significant decrease in AUC to 0.6787. This indicates that oversampling with ROS may not be suitable for this dataset and instead made the model less accurate.

In general, this experiment proves that the combination of Feature Engineering, CatBoost, and Bayesian Optimization can enhance the performance of flight delay prediction. With these results, the developed model can be implemented in flight management systems to help airlines and

airports anticipate delays and improve operational efficiency.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study has examined the effectiveness of CatBoostClassifier in predicting flight delays by applying various performance enhancement techniques, including Feature Engineering (FE), Bayesian Optimization (Bayes), and Random Over Sampling (ROS). Based on the evaluation results, it can be concluded that the application of Feature Engineering (FE) and Bayesian Optimization significantly improves the performance of the CatBoost model. The CatBoost+FE+Bayes model achieved an accuracy of 83.32%, higher than the pure CatBoost model (82.95%) and the model with only Feature Engineering (83.24%). Additionally, this model also had the highest F1-score (0.62) compared to other models, demonstrating a better balance between precision and recall. The analysis of the Area Under Curve (AUC) shows that the combination of Feature Engineering and Bayesian Optimization yielded the best results. The CatBoost+FE+Bayes model had the highest AUC value (0.7793), followed by CatBoost+FE (0.7768) and unmodified CatBoost (0.7643). However, the application of Random Over Sampling (ROS) reduced the AUC value to 0.6787, indicating that this method was not effective in the context of the dataset used. The use of Random Over Sampling (ROS) did not provide the expected performance improvement. The CatBoost+FE+Bayes+ROS model experienced a decrease in accuracy (81.44%) compared to the model without ROS. Additionally, the F1-score dropped to 0.51, indicating that resampling the data disrupted the balance between precision and recall in this model.

For future research, the study can be expanded by exploring other data balancing techniques such as SMOTE or ADASYN and testing deep learning models to further improve the accuracy of flight delay predictions.

Suggestion

Based on the findings of this study, several recommendations can be considered for future research to improve the accuracy and effectiveness of flight delay predictions. First, further exploration of data balancing techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or ADASYN (Adaptive Synthetic Sampling) can be conducted to address data imbalance without degrading model performance, as observed with the application of Random Over

Sampling (ROS) in this study. Second, in addition to using CatBoostClassifier, it is recommended to compare the results with deep learning models, such as Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM), which can capture sequential patterns in flight data. Third, a more in-depth analysis of external factors that may influence flight delays, such as weather conditions, air traffic, and airline operational factors, should be conducted. Lastly, to enhance model interpretability, Explainable AI (XAI) techniques such as SHAP (SHapley Additive Explanations) can be applied to understand the contribution of each feature to the model's decisions. With further development, this research is expected to provide more accurate solutions that can be implemented in the aviation industry to significantly reduce the impact of flight delays.

REFERENCES

- Ahmmad, J., Labassi, F., Alsuraiheed, T., Mahmood, T., & Khan, M. A. (2024). Classification of Feature Engineering Techniques for Machine Learning under the Environment of Lattice Ordered T-Bipolar Soft Rings. *IEEE Access*, 12. <https://doi.org/10.1109/ACCESS.2024.3406388>
- Alfarhood, M., Alotaibi, R., Abdulrahim, B., Einieh, A., Almousa, M., & Alkhanifer, A. (2024). Predicting Flight Delays with Machine Learning: A Case Study from Saudi Arabian Airlines. *International Journal of Aerospace Engineering*, 2024, 1–11. <https://doi.org/10.1155/2024/3385463>
- Ardhana, V. Y. P., Syam, M. Y., Ramadani, E. F., Sampetoding, E. A. M., Syahril, M., Manapa, E. S., & Mardzuki, R. (2022). Prediksi Flight Delay Berbasis Algoritma Neural Network. *Journal of Informatics, Electrical and Electronics Engineering*, 2(1), 26–30. <https://doi.org/10.47065/jieeee.v2i1.429>
- Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., & Rhee, W. (2020). Basic Enhancement Strategies When Using Bayesian Optimization for Hyperparameter Tuning of Deep Neural Networks. *IEEE Access-SPECIAL SECTION ON SCALABLE DEEP LEARNING FOR BIG DATA*, 8, 52588–52608. <https://doi.org/10.1109/ACCESS.2020.2981072>
- Darmawan, A. ... Nugraha, R. A. (2023). Implementasi Catboost Dengan Menggunakan Hyper-Parameter Tuning Bayesian Search Untuk Memprediksi Penyakit Diabetes. *Jurnal Komputasi*, 11(2), 148–156. <https://doi.org/10.23960/komputasi.v11i2.13746>
- Ernawati, S., & Wati, R. (2024). Evaluasi Performa Kernel SVM dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter dan VADER Lexicon. *Jurnal Buana Informatika*, 15(01), 40–49. <https://doi.org/10.24002/jbi.v15i1.7925>
- Ernawati, S., Wati, R., & Nuris, N. (2022). Support Vector Classification with Hyperparameters for Analysis of Public Sentiment on Data Security in Indonesia. *Jurnal Riset Informatika*, 5(1), 85–92. <https://doi.org/DOI:https://doi.org/10.34288/jri.v5i1.189>
- Fitriani, R. D., Yasin, H., & Tarno, T. (2021). Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *Jurnal Gaussian*, 10(1), 11–20. <https://doi.org/10.14710/j.gauss.v10i1.30243>
- Hatipoğlu, I., & Tosun, Ö. (2024). Predictive Modeling of Flight Delays at an Airport Using Machine Learning Methods. *Applied Sciences (Switzerland)*, 14(13), 1–19. <https://doi.org/10.3390/app14135472>
- Herdian, C., Kamila, A., & Agung Musa Budidarma, I. G. (2024). Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi. *Technologia : Jurnal Ilmiah*, 15(1), 93. <https://doi.org/10.31602/tji.v15i1.13457>
- Hu, M., Tan, Q., Knibbe, R., Wang, S., Li, X., Wu, T., ... Zhang, M. X. (2021). Prediction of Mechanical Properties of Wrought Aluminium Alloys Using Feature Engineering Assisted Machine Learning Approach. *Metallurgical and Materials Transactions A: Physical Metallurgy and Materials Science*, 52(7), 2873–2884. <https://doi.org/10.1007/s11661-021-06279-5>
- Jha, R. K., Jha, S. B., Pandey, V., & Babiceanu, R. F. (2019). Flight Delay Prediction using Hybrid Machine Learning Approach: A Case Study of Major Airlines in the United States. *Journal of Chemistry Social*, 44(5), 871–874. <https://doi.org/https://doi.org/10.48550/>
- Khan, R., Akbar, S., & Zahed, T. A. (2022). Flight Delay Prediction Based on Gradient Boosting Ensemble Techniques. *ICOSST 2022 - 16th International Conference on Open Source Systems and Technologies*.



- <https://doi.org/10.1109/ICOSST57195.2022.10016828>
- Ningthoukhongjam, J., Alam, M. S., Kumar, P., & G, M. (2024). Feature Engineering and Hybrid Machine Learning Approach for Flight Delay Prediction. *2024 International Conference on Data Science and Network Security (ICDSNS)*. <https://doi.org/10.1109/ICDSNS62112.2024.10690998>
- Qalbi, N., & Jayadi, A. (2020). Aspek Hukum Ganti Kerugian Keterlambatan Penerbangan (Flight Delay) Maskapai Penerbangan Komersial Indonesia. *Jurnal Media Iuris*, 2(3), 302–315. <https://doi.org/10.24252/aldev.v2i3.14642>
- Rajendran, R., & Karthi, A. (2022). Heart Disease Prediction Using Entropy Based Feature Engineering And Ensembling Of Machine Learning Classifiers. *Elsevier-Expert Systems With Applications*, 207(C). <https://doi.org/doi.org/10.1016/j.eswa.2022.117882>
- Sadaf, K. (2023). Phishing Website Detection using XGBoost and Catboost Classifiers. In IEEE (Ed.), *2023 International Conference on Smart Computing and Application (ICSCA)*. <https://doi.org/10.1109/ICSCA57840.2023.10087829>
- Verdonck, T., Baesens, B., Óskarsdóttir, M., & vanden Broucke, S. (2024). Special Issue On Feature Engineering Editorial. *Machine Learning*, 113(7), 3917–3928. <https://doi.org/10.1007/s10994-021-06042-2>
- Victoria, A. H., & Maragatham, G. (2021). Automatic Tuning Of Hyperparameters Using Bayesian Optimization. *Springer-Evolving Systems*, 12(1), 217–223. <https://doi.org/10.1007/s12530-020-09345-2>
- Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *Journal of Electronic Science and Technology*, 17(1), 26–40. <https://doi.org/10.11989/JEST.1674-862X.80904120>
- Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction Of Stock Price Direction Using A Hybrid Gaxgboost Algorithm With A Three-Stage Feature Engineering Process. *Expert Systems with Applications*, 186(July), 115716. <https://doi.org/10.1016/j.eswa.2021.115716>