Jurnal Mandiri IT
Vol. 14 No. 2, October (2025), pp. 188-194

ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Optimizing printer usage through data analytics for enhanced institutional efficiency

Fauwas Abdul Kadir[1], Sumanto[2]
[1,2]Informatics Study Program, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Indonesia

ARTICLE INFO

Article history:

Received Aug 3, 2025
Revised Aug 15, 2025
Accepted Sep 4, 2025

Keywords:

Data Analytic;
Environmental Impact;
Operational Efficiency;
Printer Usage;
Resource Management.

ABSTRACT

The advancement of information technology had simplified various workplace processes, including document processing and printing. In an institution, the use of printers played a crucial role in daily operations. However, without proper management, printer usage often became inefficient, leading to increased operational costs and unnecessary waste of resources. Therefore, an analytical system was needed to monitor and optimize printer usage. Such a system provided valuable insights by analyzing data generated from printing activities. This data analysis revealed patterns in work habits and allowed institutions to make informed decisions. As a result, institutions were able to improve operational efficiency, reduce costs, and minimize environmental impact. Paper and ink waste were significantly reduced by implementing data-driven policies. Overall, the integration of data analytics into printer management contributed to sustainable practices and better resource allocation in institutional environments.

Corresponding Author:

Fauwas Abdul Kadir,
Informatics Study Program,
Faculty of Engineering and Informatics,
Universitas Bina Sarana Informatika,
Jl. Kramat Raya No.98, Jakarta, 10450, Indonesia
Email: fauwazabdul@gmail.com

1. INTRODUCTION

The rapid development of information technology has significantly transformed the way organizations handle operational tasks, particularly in document processing and printing. Printers remain essential tools in various institutions, including offices, schools, and government agencies, where daily operations often involve high-volume printing activities. However, despite their importance, printer usage is frequently overlooked in terms of efficiency and sustainability.

Printers are seen as a critical point in green IT policies because they are resource-intensive (use paper, ink, electricity), generate significant waste, and have the potential for significant savings through relatively simple management policies compared to other devices. Employee printing behavior (e.g., repeated printing, unnecessary printing, or single-sided printing) significantly impacts resource consumption. In addition, analytics-driven policies enable institutions to implement sustainable printing practices as part of their Green IT strategies, ensuring that outcomes such as reduced paper demand or lower electricity consumption are not only cost-saving but also measurable contributions toward institutional sustainability goals. This means that printers are not just a technological tool, but also a cultural shift in supporting sustainability.

Unmonitored printer activities can result in excessive resource consumption, including paper, ink, and electricity, ultimately leading to increased operational costs and environmental waste. Studies such as those by Yin, J., & Li, C. (2022) emphasized that inefficient printing practices contribute to organizational inefficiencies and carbon footprint. Furthermore, research by Green IT advocates has highlighted the growing need to implement environmentally conscious policies through the optimization of device usage.

The integration of data analytics into printing systems offers a promising solution. By collecting and analyzing print-related data such as page count, color vs. monochrome usage, paper size, and print cancellations, institutions can identify patterns, detect anomalies, and implement informed decisions to improve performance.

Recent advancements in machine learning (ML) have made it possible to process such data at scale and extract valuable insights. ML techniques such as Random Forest, Support Vector Machines, and neural networks have been widely adopted in various domains to optimize resource usage and support automated decision-making (Hall & Rasheed, 2025; Mahajan et al., 2024). In sectors like telecommunications and education, ML has shown significant improvements in efficiency through anomaly detection and usage prediction (Thillaigovindhan et al., 2024; Khan et al., 2024). These successes highlight the potential of ML in addressing inefficiencies in overlooked processes such as printer usage.

Moreover, the adoption of explainable machine learning models ensures transparency and trust in their recommendations, which is especially critical in institutional environments (Pichler & Hartig, 2022). Integrating such approaches into printer usage monitoring allows institutions to make data-driven policies that not only cut operational costs but also support environmental sustainability goals.

This study aimed to analyze institutional printer usage through data-driven methods, providing insights into user behavior and operational efficiency. The outcomes were intended to support strategic policy-making to reduce unnecessary printing, cut costs, and promote sustainable practices.

2. RESEARCH METHOD

This study employed a quantitative research methodology that utilized machine learning techniques to analyze institutional printer usage data. The research process was structured into six distinct stages. Each stage was designed to systematically process, analyze, and interpret the data to derive meaningful insights.

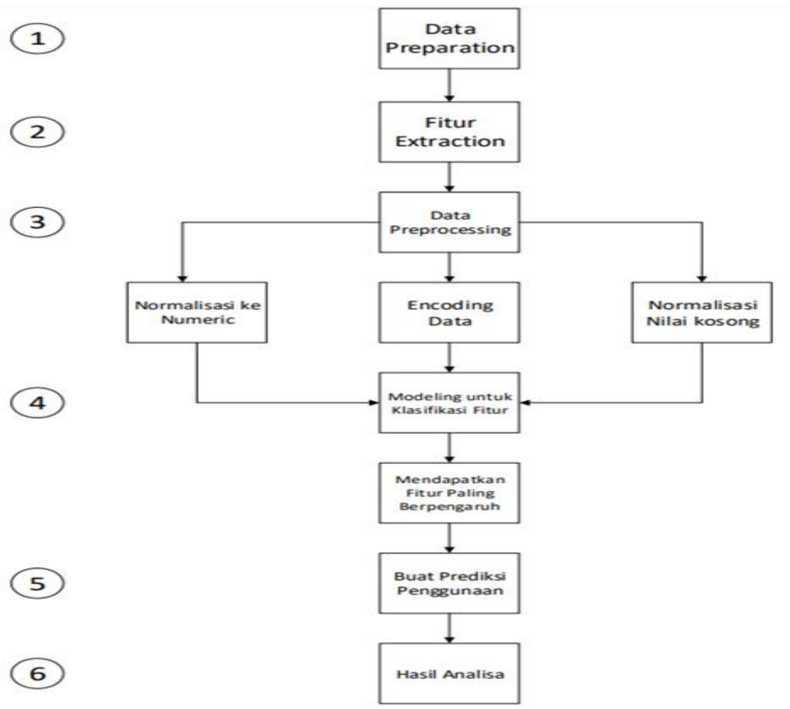

Figure 1. Research method data analysis flow diagram

1. Identify the data to determine the purpose of the data processing to be performed. Raw data taken from the printer server log is imported into CSV format. Several Python libraries are used to perform data processing and visualization, such as Numpy, Pandas, Seaborn, Matplotlib, and Sklearn.

2. Feature extraction was performed to identify and select the most relevant features for model implementation. This step helps in reducing noise and improving the model's accuracy. A correlation matrix was utilized to analyze the relationship between each feature. By examining the correlations, redundant or less impactful features could be excluded from the final model.

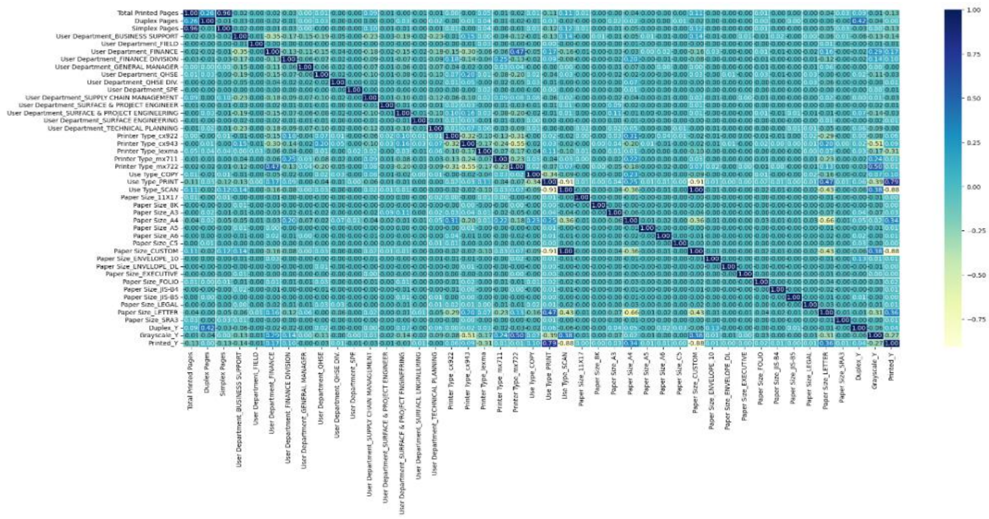

Figure 2. Correlation matrix for each variable

3. Data preprocessing involved cleaning and transforming raw print logs into a structured dataset suitable for machine learning modeling. This included handling missing values, encoding categorical variables, and normalization. Outlier detection and feature selection were applied to enhance model performance and reduce noise (Khan et al., 2024; Hall & Rasheed, 2025). Data processing was performed using three methods. First, data was normalized into numeric form. Second, encoding was performed using the one-hot encoder method. Third, data cleaning and normalization were performed to handle missing values and duplication. Encoding and normalization were used to transform categorical data into numeric form. From this numeric data, the following statistical data was obtained.

Table 1. Statistical calculations of numeric data

Variable	Total Printed	Duplex	Simplex
count	124878	124878	124878
mean	4.21	0.19	4.03
std	13.59	3.77	13.10
min	1	0	0
25%	1	0	1
50%	1	0	1
75%	3	0	3
max	941	270	941

4. Once the data is converted to numerical values, it can be implemented into a machine learning model for classification. Here, we used a grid search technique from five classification models, resulting in the Random Forest Classifier model, which selected the most influential features, achieving the highest accuracy of 97% on the data. The Random Forest Classifier was chosen as the primary algorithm due to its robustness against overfitting and its capability to handle high-dimensional data (Pichler & Hartig, 2022). The dataset was split into training (80%) and testing (20%) subsets using stratified sampling to maintain class distribution. The model was trained to classify printing patterns based on usage behavior (e.g., efficient vs. wasteful) and to detect anomalies that may indicate unusual or excessive printing. Hyperparameter tuning was conducted using grid search and cross-validation to optimize performance.

Table 2. Gridsearch technique to find the best classification model.

Model	Accuracy
Linear SVC	73.82%
KNeighbors	97.38%

Model	Accuracy
Random Forrest	97.40%
Decision Tree	94.79%
SGD Classification	73.27%

5. From the classification results, the five most influential features were identified, as shown in the figure. These five features were used to implement a machine learning forecasting model to predict future print runs. The results were visualized using charts and reports to identify departments with inefficient usage patterns. Recommendations for policy improvement, such as print quotas or duplex defaults, were proposed based on findings. This approach follows the data-driven governance and green IT principles discussed by Yin, J., & Li, C. (2022).

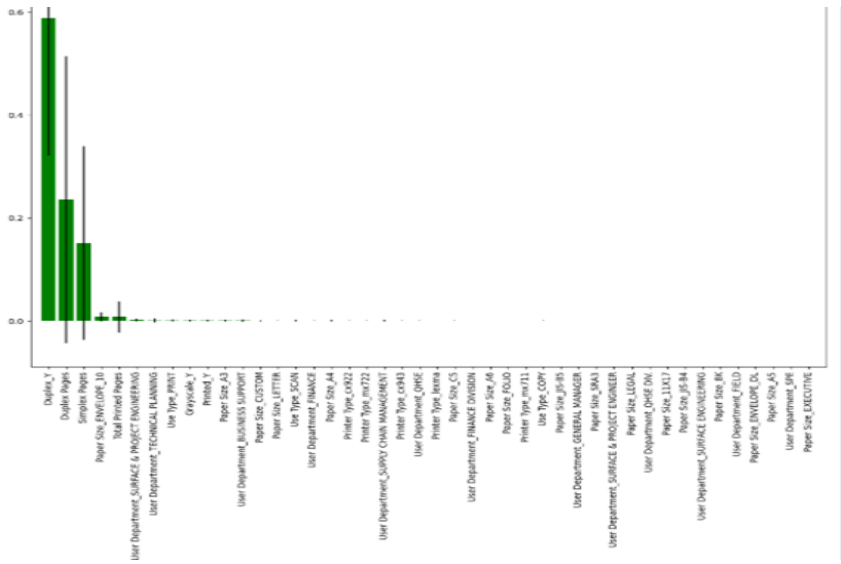

Figure 3. Feature important classification results

6. All the data obtained during the research were collected systematically. This step ensured that relevant information from each phase was gathered accurately. The collected data were then used to draw conclusions based on the analysis conducted.

3. RESULTS AND DISCUSSIONS

The total printer usage per division was calculated after completing the forecasting process. This analysis provided a clear overview of how each division utilized printing resources over time. By comparing the forecasted data with actual usage, patterns and discrepancies were identified. These insights were then used to evaluate printing efficiency and optimize resource allocation across divisions.

From a cross-divisional efficiency policy perspective, these findings provide a basis for management to direct more department-specific strategies, such as implementing a duplex printing policy, limiting the number of pages per user, or integrating a digital document management system to reduce the need for repeated printing. Furthermore, uneven distribution of the print load can trigger the need for redistribution of equipment or policies across divisions, for example, by shifting some documentation processes to a shared digital system to more evenly distribute the print load.

Table 3. Result prediction of printer usage in each division

Division	Estimated Print Usage
Business Support	39446
Finance	26106
Supply Chain Management	13283
Technical Planning	13102
Surface & Project Engineering	9454
QHSE	9224

Division	Estimated Print Usage
Finance Division	7707
General Manager	5530
QHSE Division	652
Surface & Project Enginer	208
Surface Engineering	156
Field	8
SPE	2

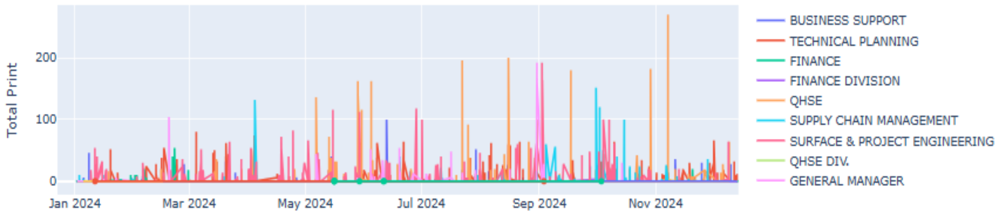

Figure 4. Graphical prediction of printer usage in 1 year

Recommendations for efficient printer usage included the implementation of a duplex printing policy to maximize paper efficiency. This approach aimed to significantly reduce paper consumption and minimize waste. The forecasting model developed from the processed data demonstrated a prediction accuracy of 80–81%. This level of accuracy indicates the model's reliability in supporting data-driven decision-making for printer usage optimization.

accuracy			1.00	24976
macro avg	0.80	0.81	0.80	24976
weighted avg	1.00	1.00	1.00	24976

Figure 5. The results of the accuracy measurement level of the model.

Second, document digitization and print quotas. By reducing unnecessary print volume, ink/toner costs are reduced. The resulting impact is reduced toner waste and printer electricity. Compare the average print run per employee before and after the policy, then convert this to potential ink and electricity reductions.

4. CONCLUSION

The conclusions drawn after analyzing the data obtained were as follows the "Business Support Department" had the highest print volume (39,446 pages). This reflects its high technical requirements. Print Efficiency majority of prints were grayscale (82.6%), reflecting efficient use of color ink. However, duplex printing usage was still low (6.2%) because most users printed on one side of the paper. Distribution of print job types is "reports" (34.4%) and "Presentations" (29.0%). other documents, such as invoices and emails, tended to require fewer prints. A predictive model accuracy of 80–81% can serve as an initial basis for developing printer usage policies, but it is insufficient to serve as a single, long-term reference. There is still a 19–20% chance of error, so policies need to be flexible and evaluated periodically.

The success of internal document digitization can be measured by reducing print volume. Print volume reduction is achieved by comparing the average number of pages printed before and after digitization. The reduction in print volume will be visible in printer log data and operational expenses for paper, ink, and device maintenance. If the average monthly print count before digitization was around 40,000 pages from Business Support and after standard implementation, it drops to 20,000 pages, a measurable 50 percent savings is achieved. Suggestions from practitioners that can be implemented for office policy include: a) Increasing the use of duplex printing by requiring two-sided printing for internal documents to reduce paper consumption and providing training to users on how to utilize the duplex printing feature; b) Limiting color printing for documents for use of color printing to external documents or important presentations with implementing an approval system for color printing; c) Using Monitoring print activity with implementing a regular reporting system to monitor print activity in each department and identifying users with high print activity for further evaluation; d) Implementing digitization of internal documents by reducing printing for documents such as reports or emails by adopting digital formats and using digital collaboration tools for internal presentations.

ACKNOWLEDGEMENTS

All praise and gratitude to God Almighty, for all His blessings and gifts have enabled the author to complete this journal to the best of his ability. Blessings and peace be upon the Prophet Muhammad, his companions, and their families, who have guided humanity from one age to another, filled with knowledge. This journal was created to fulfill the responsibility for the semester courses completed by the author, and as one of the graduation requirements for a bachelor's degree in the Informatics Department, Faculty of Engineering and Informatics, Bina Sarana Informatika University. During the course of this research and in developing this journal, the author has received guidance, direction, guidance, advice, and facilities that have been very helpful until the end of writing this journal. The author realizes that without the guidance and encouragement from all parties.

REFERENCES

Antosz, K., Kozłowski, E., Sęp, J., & Zabinski, T. (2022). The use of random forests to support the decision-making process for sustainable manufacturing. Journal of Physics: Conference Series, 2198(1). https://doi.org/10.1088/1742-6596/2198/1/012006

Ariatmanto, D., & Rifai, A. M. (2024). The Impact of Feature Extraction in Random Forest Classifier for Fake News Detection. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 8(6), 730–736. https://doi.org/10.29207/RESTI.V8I6.6017

Data-Driven decisions: The role of analytics in Print Management - ManagedPrint. (n.d.). Retrieved August 8, 2025, from https://managedprint.com/data-driven-decisions-the-role-of-analytics-in-print-management

Ganaie, M. A., Tanveer, M., Suganthan, P. N., & Snasel, V. (2022). Oblique and rotation double random forest. Neural Networks, 153, 496–517. https://doi.org/10.1016/j.neunet.2022.06.012

Hu, J., & Szymczak, S. (2022). A review on longitudinal data analysis with random forest in precision medicine. Briefings in Bioinformatics, 24(2). https://doi.org/10.1093/bib/bbad002

Ilham, I. (2024). Predicting Plant Growth Stages Using Random Forest Classifier: A Machine Learning Approach. Indonesian Journal of Data and Science, 5(2), 155–165. https://doi.org/10.56705/IJODAS.V5I2.167

Khan, H. A;, Jamil, S. ;, Piran, M. J. ;, Kwon, O.-J. ;, Lee, J.-W., Khan, H. A., Jamil, S., Piran, M. J., Kwon, O.-J., & Lee, J.-W. (2024). A Comprehensive Survey on the Investigation of Machine-Learning-Powered Augmented Reality Applications in Education. Technologies 2024, Vol. 12, Page 72, 12(5), 72. https://doi.org/10.3390/TECHNOLOGIES12050072

Kostidjan, O. D., Purwananto, Y., & Yuniarti, A. (2024). Skin Cancer Classification Using a Hybrid Pre-trained CNN with Random Forest Classifier. Jurnal RESTI, 8(4), 506–515. https://doi.org/10.29207/RESTI.V8I4.5857

Li, L. (n.d.). Causal Effect Random Forest Of Interaction Trees For Learning Causal Effect Random Forest Of Interaction Trees For Learning Individualized Treatment Regimes In Observational Studies: With Individualized Treatment Regimes In Observational Studies: With Applications To Education Study Data Applications To Education Study Data. Retrieved August 8, 2025, from https://scholarship.claremont.edu/cgu_etd

Mahajan, Y., Patil, R., Pattanaik, S., Firake, T. S., Kodulkar, R., Damre, S. S., & Uplaonkar, D. (2024). A Review of Techniques and Applications for Machine Learning and Deep Learning. International Journal of Intelligent Systems and Applications in Engineering, 12(16s), 182–187. https://www.ijisae.org/index.php/IJISAE/article/view/4804

Nugroho, M. W. (2025). Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi. Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi), 9(4), 1562–1571. https://doi.org/10.35870/JTIK.V9I4.4236

Ong, A. K. S., Chuenyindee, T., Prasetyo, Y. T., Nadlifatin, R., Persada, S. F., Gumasing, M. J. J., German, J. D., Robas, K. P. E., Young, M. N., & Sittiwatethanasiri, T. (2022). Utilization of Random Forest and Deep Learning Neural Network for Predicting Factors Affecting Perceived Usability of a COVID-19 Contact Tracing Mobile Application in Thailand “ThaiChana.” International Journal of Environmental Research and Public Health, 19(10), 6111. https://doi.org/10.3390/IJERPH19106111

Ong, A. K. S., Prasetyo, Y. T., Yuduang, N., Nadlifatin, R., Persada, S. F., Robas, K. P. E., Chuenyindee, T., & Buaphiban, T. (2022). Utilization of Random Forest Classifier and Artificial Neural Network for Predicting Factors Influencing the Perceived Usability of COVID-19 Contact Tracing “MorChana” in Thailand. International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 7979, 19(13), 7979. https://doi.org/10.3390/IJERPH19137979

Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. Methods in Ecology and Evolution, 14(4), 994–1016. https://doi.org/10.1111/2041-210X.14061;CTYPE:STRING:JOURNAL
Ramadhan, R. F., & Ashari, W. M. (2024). Performance Comparison of Random Forest and Decision Tree Algorithms for Anomaly Detection in Networks. Journal of Applied Informatics and Computing, 8(2), 367–375. https://doi.org/10.30871/JAIC.V8I2.8492
Satria, B., Siswa, T. A. Y., & Pranoto, W. J. (2024). Optimasi Random Forest dengan Genetic Algorithm dan Recursive Feature Elimination pada High Dimensional Data Stunting Samarinda. JURNAL MEDIA INFORMATIKA BUDIDARMA, 8(3), 1778. https://doi.org/10.30865/MIB.V8I3.7883
Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. Stata Journal, 20(1), 3–29. https://doi.org/10.1177/1536867X20909688/ASSET/EC601327-319C-4C16-97CA-BC08F874F55A/ASSETS/IMAGES/10.1177_1536867X20909688-IMG9.PNG
Shahhosseini, M., & Hu, G. (2020). Improved Weighted Random Forest for Classification Problems. 42–56. https://doi.org/10.1007/978-3-030-66501-2_4
Singh, N., Sharma, R. L., & Yadav, K. (2023). Sustainable development by carbon emission reduction and its quantification: an overview of current methods and best practices. Asian Journal of Civil Engineering, 24(8), 3797–3822. https://doi.org/10.1007/S42107-023-00732-Z
Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Systems with Applications, 237. https://doi.org/10.1016/J.ESWA.2023.121549
Susetyoko, R., Purwantini, E., Iman, B. N., & Satriyanto, E. (2023). An Improved Accuracy of Multiclass Random Forest Classifier with Continuous Attribute Transformation Using Random Percentile Generation. International Journal on Advanced Science, Engineering and Information Technology, 13(3), 943–953. https://doi.org/10.18517/IJASEIT.13.3.18379
Thillaigovindhan, S. K., Roslee, M., Mitani, S. M. I., Osman, A. F., & Ali, F. Z. (2024). A Comprehensive Survey on Machine Learning Methods for Handover Optimization in 5G Networks. Electronics 2024, Vol. 13, Page 3223, 13(16), 3223. https://doi.org/10.3390/ELECTRONICS13163223
Trochoutsos, C., & Sofias, Y. (2022). THE IMPORTANCE OF DATA ANALYSIS IN THE MODERN ERA OF PRINT PRODUCTION. International Symposium on Graphic Engineering and Design, 853–859. https://doi.org/10.24867/GRID-2022-P94
Vergni, L., & Todisco, F. (2023). A Random Forest Machine Learning Approach for the Identification and Quantification of Erosive Events. Water 2023, Vol. 15, Page 2225, 15(12), 2225. https://doi.org/10.3390/W15122225
Yin, J., & Li, C. (2022). Data governance and green technological innovation performance: A curvilinear relationship. Journal of Cleaner Production, 379, 134441. https://doi.org/10.1016/J.JCLEPRO.2022.134441
Zheng, J., Shi, J., Lin, F., Hu, X., Pan, Q., Qi, T., Ren, Y., Guan, A., Zhang, Z., & Ling, W. (2023). Reducing manufacturing carbon emissions: Optimal low carbon production strategies respect to product structures and batches. Science of The Total Environment, 858, 159916. https://doi.org/10.1016/J.SCITOTENV.2022.159916