Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 1, May 2026
https://doi.org/10.47709/brilliance.v6i1.7744

Implementation of C5.0 Algorithm in Cement Stock and Purchase Management at PT. Maktal

Rezky Salman Maulida[1*], Azizah Zakiah[2]
[1,2]Universitas Widyatama, Indonesia
[1]rezky.salman14@gmail.com, [2]azizah.zakiah@widyatama.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 07-01-2026
Accepted: 17-01-2026
Published: 26-01-2026

Keywords:
C5.0 Algorithm; Classification; Predictioon; Random Forest Algoorithm; Stock Management

ABSTRACT

Stock management is a crucial activity in the supply chain of any company, including PT. Maktal, which operates in cement distribution. The current system, relying on experience and manual methods, potentially causes demand and supply mismatches, leading to excessive inventory costs (overstock) or product shortages (stockout). Implementation of Machine Learning offers a solution to enhance stock planning accuracy. This study aims to develop and compare machine learning models, specifically Decision Tree (C5.0) and Random Forest, in predicting cement stock need categories (Low, Medium, High) based on historical transaction data. The dataset includes historical cement sales and ordering transactions from 2020 to 2024. Numerical stock quantity was converted into categorical variables (Low, Medium, High) through a discretization process. Both algorithms were evaluated using accuracy, precision, recall, and F1-score metrics. Comparative results indicate that the Random Forest algorithm provides superior performance with an accuracy level of 79.91%, significantly higher than the Decision Tree algorithm. Feature importance analysis identified that the Purpose (customer type) and Month variables are the most influential predictors of stock categories. The Random Forest model proved effective and reliable as a data-driven decision support system to optimize stock planning and cement purchasing at PT. Maktal, reducing risks associated with demand uncertainty.

INTRODUCTION

Stock management is an essential and challenging activity within the framework of Supply Chain Management (SCM) that significantly impacts a company's financial and operational performance (Krajewski et al., 2020). Efficient inventory control ensures that goods are available when needed while minimizing the high costs associated with both *stockouts* (lost sales and customer dissatisfaction) and *overstocking* (increased storage and obsolescence expenses). For companies operating with large-volume, fast-moving items, such as PT. Maktal in cement distribution, the precision of stock planning is paramount.

PT. Maktal, handling various types of cement to meet diverse customer needs, currently faces challenges in optimizing its stock levels. The existing stock planning process heavily relies on manual methods and the experience of personnel, which often leads to planning inaccuracies. These inaccuracies manifest as temporary product shortages or, conversely, excessive stock accumulation, directly affecting the company's profitability and capital efficiency. Therefore, a modern, reliable, and data-driven approach is urgently required to transform stock planning from a reactive process into a proactive predictive strategy.

In recent years, the implementation of Machine Learning (ML) techniques has emerged as a powerful solution for improving prediction accuracy in various fields, including demand forecasting and inventory management. Among the popular classification methods, Decision Tree algorithms, particularly C5.0, are valued for their simplicity, speed, and interpretability (Lonang et al., 2023). However, a single Decision Tree can be prone to high variance and overfitting when dealing with complex, real-world data. This limitation is effectively addressed by Random Forest, an *ensemble* method that aggregates predictions from multiple Decision Trees to achieve higher stability and robustness. Comparative analysis of these two related algorithms provides valuable insight into the most suitable model for a specific business case.

Based on the issues and technological opportunities presented, this study aims to develop and compare the performance of the Decision Tree (C5.0) and Random Forest algorithms to solve the problem of stock management at PT. Maktal. Specifically, the study seeks to classify future cement stock needs into categorical levels (Low, Medium, or High) based on historical transaction data from 2020 to 2024. Furthermore, this research will identify the most influential input features (such as Customer Type/Purpose and Month) that drive demand patterns.

The primary contribution of this paper is the recommendation of an accurate, data-driven classification model with Random Forest showing superior preliminary results (79.91% accuracy) that can be directly integrated into PT. Maktal's decision support system. This is expected to provide decision-makers with a reliable forecast, enabling

E-ISSN : 2807-9035
Volume 6, **Number** 1, May 2026
https://doi.org/10.47709/brilliance.v6i1.7744

optimized cement procurement and contributing significantly to the reduction of operational costs associated with inefficient inventory levels.

LITERATURE REVIEW

Stock Management and Classification

Effective Stock Management is defined as the process of controlling, planning, and organizing inventory within the supply chain to meet demand while minimizing costs (Krajewski et al., 2020). In a fluctuating market, the key challenge lies in accurately determining future stock needs (Monczka et al., 2020). Modern stock planning increasingly uses classification techniques in data mining to categorize demand into predefined level such as Low, Medium, or High rather than predicting exact numerical values (Lonang et al., 2023). This approach is often more robust for decision-making support as it provides actionable categorical insights for procurement teams (Wijaya et al., 2021). The present study employs this classification approach to categorize the volume of cement stock required by PT. Maktal (Zai, 2022).

Decision Tree Algorithm (C5.0)

Decision Tree is one of the most widely used supervised learning techniques for classification tasks. It operates by recursively splitting the dataset based on the best feature, resulting in a tree structure where internal nodes represent tests on attributes, branches represent the outcomes of the test, and leaf nodes represent the class labels (Lonang et al., 2023). The **C5.0** algorithm is an advancement of its predecessors, ID3 and C4.5, specifically designed to handle large datasets efficiently (Amalda et al., 2022). It uses the concept of Information Gain and Gain Ratio to select the splitting attribute that provides the greatest purity for the subsequent nodes. The main advantage of C5.0 is its fast-processing speed and ability to generate highly interpretable rules (Utomo et al., 2020) (Purwanti et al., 2023). However, a single Decision Tree can be susceptible to overfitting, meaning it performs exceptionally well on training data but poorly on unseen data (Normawati & Prayogi, 2021).

Random Forest Algorithm

Random Forest (RF) is an *ensemble* learning method developed by Leo Breiman that operates by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. RF addresses the high variance problem inherent in single Decision Trees through two primary mechanisms: Bagging (Bootstrap Aggregating) and Feature Randomness. Bagging involves training each tree on a different bootstrap sample of the data, and Feature Randomness involves using a random subset of features for splitting at each node. By aggregating the results of hundreds of independent trees, Random Forest significantly reduces overfitting and improves the generalization ability and robustness of the model, making it highly effective for complex classification tasks (Patel & Shah, 2021), including stock prediction (Saraswati et al., 2022).

Based on the review, while Decision Tree (C5.0) offers interpretability, Random Forest often provides superior predictive performance and robustness in real-world supply chain classification problems (Rachman et al., 2024). The current study differentiates itself by specifically comparing these two algorithms to classify cement stock needs at PT. Maktal and identifying the key transactional features unique to this context, thereby filling the identified gap.

METHOD

Research Framework

The systematic process of this research, from data collection to model evaluation, is illustrated in the following flow diagram (See Figure 1). This structured approach ensures the reproducibility of the study and clarity in the methodological execution.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 1, May 2026
https://doi.org/10.47709/brilliance.v6i1.7744

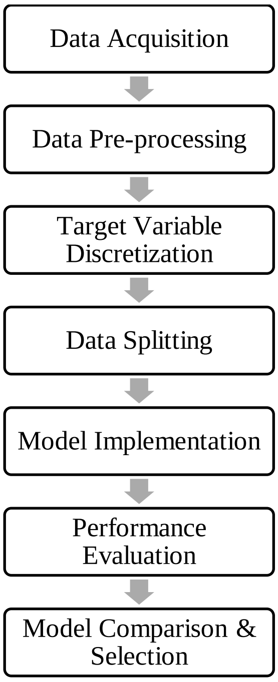

Fig. 1 Research Flow Diagram

Data Source

The data utilized in this study is secondary data obtained from the internal transaction records of PT. Maktal regarding cement sales and purchasing orders. The dataset covers a period from January 2020 to December 2024, encompassing various transactional attributes. The initial dataset comprises features such as Date of Transaction, Customer Type (Purpose), Type of Cement, and Quantity (in unit loads). The historical data ensures the models are trained on real-world patterns and seasonality inherent to the cement distribution business.

Data Pre-processing

Before model training, the raw transactional data underwent several pre-processing steps:

1. **Data Cleaning:** Identifying and handling missing values, inconsistencies, and outliers within the transactional records to ensure data quality.
2. **Feature Engineering:** Extracting relevant temporal features. The Date of Transaction was transformed into categorical features, notably Month and Year, to capture seasonal and annual trends, which are crucial factors in predicting demand patterns.
3. **Feature Encoding:** Non-numeric categorical features, such as Customer Type (Purpose) and Type of Cement, were converted into a machine-readable format using One-Hot Encoding to prevent the models from assuming ordinal relationships.

Target Variable Discretization

Since the objective is to classify the category of stock needs rather than predicting the exact quantity, the numerical target variable, Quantity, was transformed into a nominal categorical variable called Stock Needs Category.

Model Implementation and Training

The prepared dataset was randomly split into 80% for training data and 20% for testing data. Two classification models were implemented and compared:

1. **Decision Tree (C5.0)**: The C5.0 algorithm was trained using the training dataset, building a decision tree by calculating the information gain for each attribute to determine the optimal split points.
2. **Random Forest (RF)**: The Random Forest model utilized the bagging technique to construct multiple C5.0-style trees on diverse subsets of the data, combining their output to produce a more robust and generalized prediction.

Model Evaluation

The performance of both models was assessed using the 20% testing data. The evaluation relied on the Confusion Matrix and standard classification metrics, which are crucial for assessing model quality in classification tasks:

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 1, May 2026
https://doi.org/10.47709/brilliance.v6i1.7744

1. **Accuracy (A)**: The ratio of correctly predicted instances to the total number of instances.
2. **Precision (P)**: The measure of the proportion of positive identifications that were actually correct.
3. **Recall (R)**: The measure of the proportion of actual positives that were correctly identified.
4. **F1-Score (F1)**: The harmonic mean of Precision and Recall, providing a balanced measure of the model's performance across all classes.

RESULT

Exploratory Data Analysis

Initial exploration of the transactional data from 2020 to 2024 revealed significant characteristics concerning cement demand at PT. Maktal.

1. Sales Trend: The analysis of the monthly sales trend shows a clear seasonal pattern. Cement sales volume consistently increases significantly in the middle of the year (around April–October) and decreases at the beginning and end of the year, correlating with the typical construction project cycle in Indonesia (See Figure 2).
2. Product Distribution: 'Semen PCC' is the most frequently transacted product type, dominating the sales volume compared to 'Semen OPC' and 'Mortar Instan' (See Figure 3).
3. Customer Distribution: The highest transaction frequency is recorded from 'Toko Material Bangunan' (Building Material Stores), followed by 'Kontraktor Perumahan' (Housing Contractors), indicating that the retail and small-to-medium project segments are the primary drivers of demand.

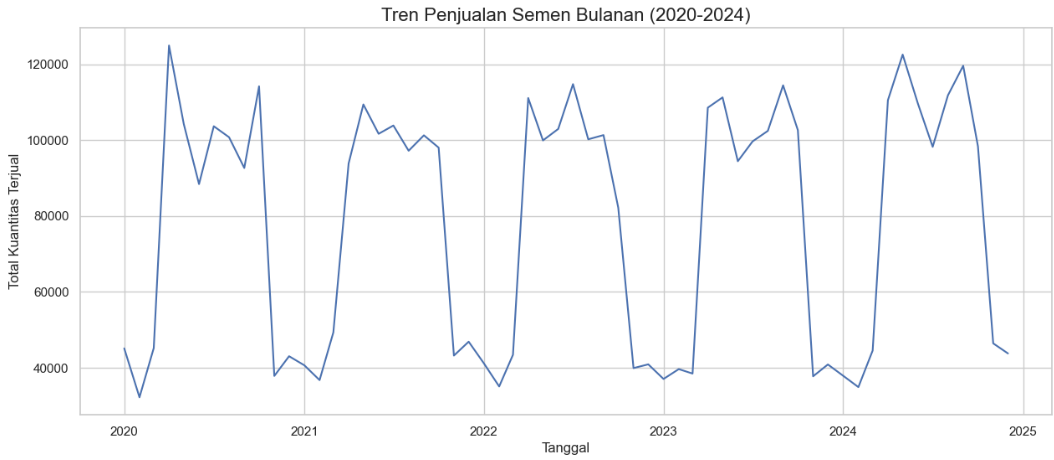

Fig. 2 Monthly Sales Trend (2020-2024)

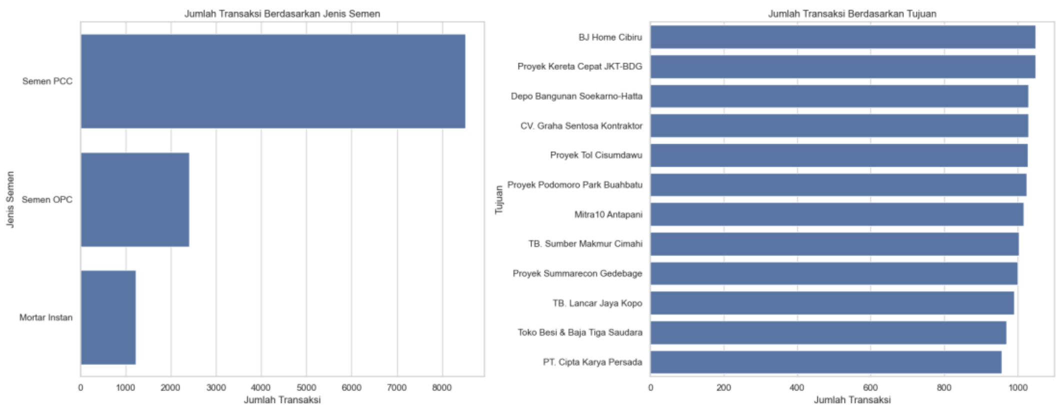

Fig. 3 Distribution of Transactions by Cement Type and Purpose

Model Performance Evaluation

Following data processing, two classification models, Decision Tree and Random Forest, were trained and evaluated on the testing dataset to predict the Stock Needs Category (Low, Medium, High).

Decision Tree (C5.0) Performance

The initial single Decision Tree model, based on the C5.0 logic, achieved an overall accuracy of 75.83%. While providing a foundational benchmark, its performance suggested room for improvement, particularly in differentiating the 'Medium' and 'High' categories, as seen in the confusion matrix analysis (See Figure 4).

E-ISSN : 2807-9035
Volume 6, **Number** 1, May 2026
https://doi.org/10.47709/brilliance.v6i1.7744

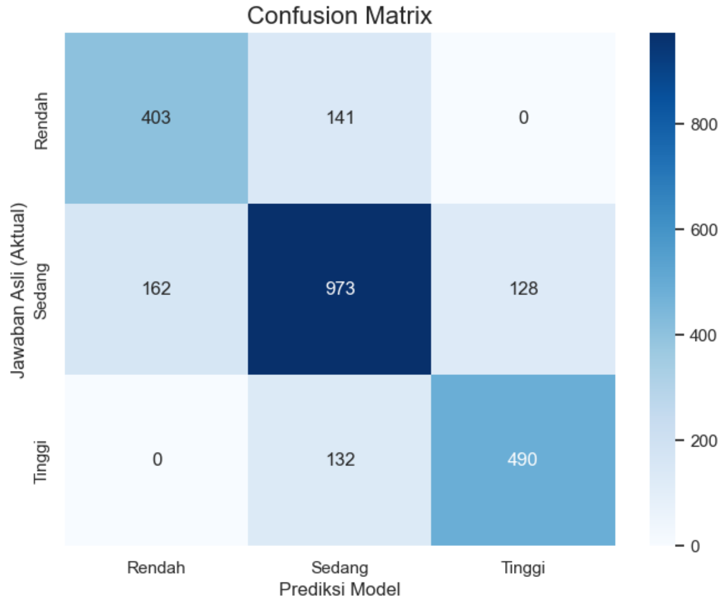

Fig. 4 Confusion Matrix Decision Tree

Random Forest Performance

The Random Forest model, an ensemble classifier, demonstrated superior performance compared to the single Decision Tree. It achieved a higher overall accuracy of 79.91% and displayed balanced performance across all classification metrics (Table 1).

Table 1. Classification Report (Random Forest)

	Precision	**Recall**	**F1-score**	**Support**
Low	0.81	0.69	0.75	544
Medium	0.77	0.86	0.81	1263
High	0.84	0.75	0.79	622
Accuracy			0.80	2429
Macro avg	0.81	0.77	0.79	2429
Weighted avg	0.80	0.80	0.79	2429

The confusion matrix for Random Forest (See Figure 4) visually confirms the balanced performance, with a high number of correctly classified instances along the diagonal, particularly for the dominant 'Medium' class.

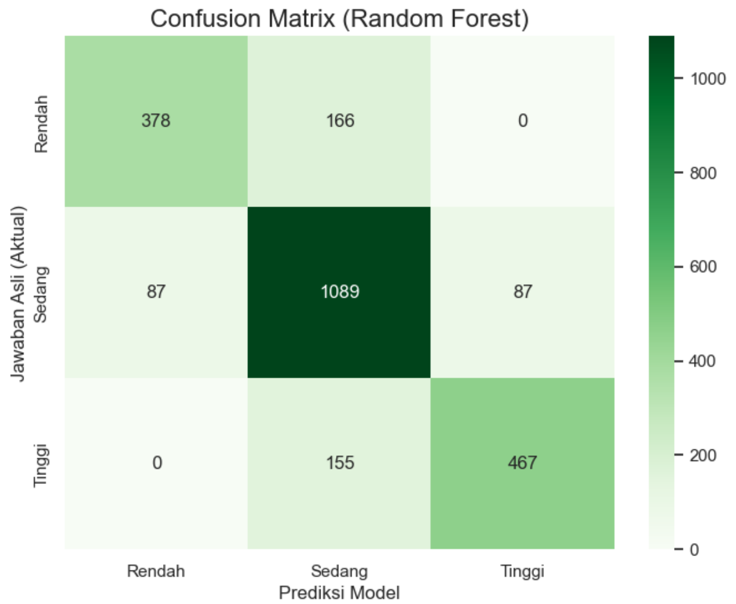

Fig. 5 Confusion Matrix Random Forest

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 1, May 2026
https://doi.org/10.47709/brilliance.v6i1.7744

DISCUSSION

Model Comparison

The Random Forest model was selected as the final model due to its consistently higher performance metrics, notably an overall accuracy of 79.91% compared to the 75.83% accuracy of the Decision Tree model. This finding is consistent with literature which suggests that ensemble methods effectively reduce the variance and bias inherent in single Decision Trees, resulting in better generalization on unseen data. [8888] In the context of PT. Maktal's stock management, the 4.08% improvement in accuracy provided by Random Forest translates directly into more reliable stock procurement decisions and reduced risk of costly prediction errors.

Furthermore, the evaluation of the confusion matrices is crucial. The results show that the model's prediction errors are non-fatal; for instance, instances that were *actually* 'High' were not mistakenly predicted as 'Low' and vice-versa (See Figure 5 & Figure 6). This crucial aspect ensures that the risk of critical stockouts which could severely impact customer projects—is minimized.

Feature Importance Analysis

Analysis of the Random Forest model revealed the key factors influencing the cement stock prediction (See Figure 6).

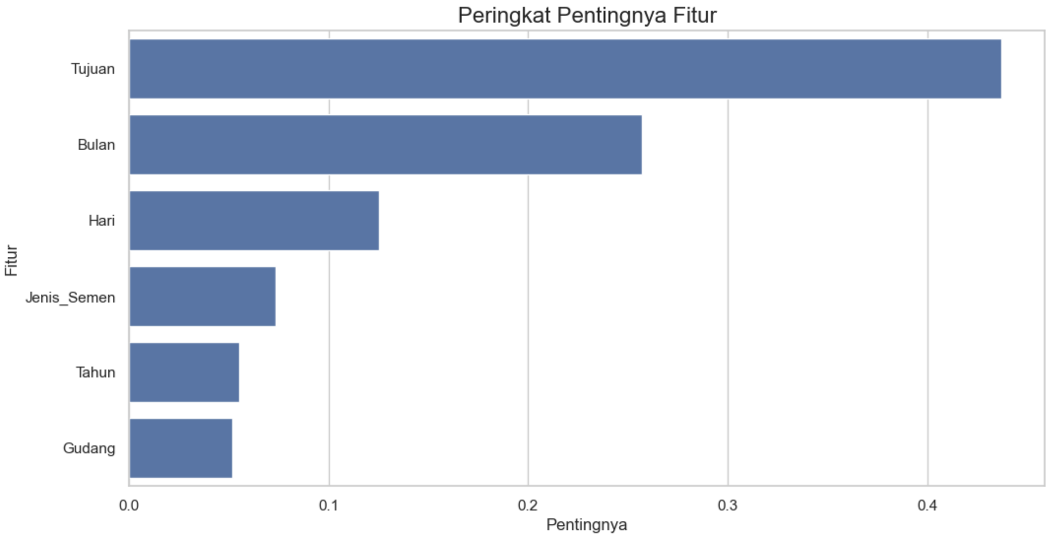

Fig. 6 Feature Importance Ranking (Random Forest)

The results indicate that Purpose (Customer Type) is the most dominant factor, accounting for approximately 42% of the prediction importance, followed closely by Month (around 27%).

This finding provides significant managerial insight:

1. Purpose: The type of customer (e.g., large construction projects vs. small building material stores) dictates the volume and frequency of orders. Knowing the customer type is the primary input for predicting the necessary stock category.
2. Month: The high importance of the 'Month' variable confirms the strong influence of seasonality, where stock needs predictably peak during the middle of the year, aligning with Indonesia's non-rainy season construction activity cycle (See Figure 2).

The results confirm that decisions on cement procurement and inventory planning at PT. Maktal should prioritize these two critical dimensions to achieve the highest predictive accuracy and operational efficiency.

CONCLUSION

Based on the results and discussion presented, it is concluded that a robust machine learning model was successfully developed to optimize cement stock management and procurement at PT. Maktal. The comparative analysis confirms that the Random Forest algorithm is the superior model for classifying cement stock needs into the categories of Low, Medium, and High, achieving a high predictive accuracy of 79.91%. This performance is significantly higher than that achieved by the individual Decision Tree (C5.0) model.

Furthermore, the analysis successfully identified the most influential factors driving demand, namely the Purpose (Customer Type) and the Month of Transaction. The model evaluation ensures a reliable performance where critical prediction errors (predicting 'Low' when the actual need was 'High') are successfully minimized. In essence, this study

E-ISSN : 2807-9035
Volume 6, Number 1, May 2026
https://doi.org/10.47709/brilliance.v6i1.7744

contributes a verified and effective predictive tool that moves PT. Maktal's inventory strategy from reactive estimation to proactive, data-driven planning, enabling better decision-making for cement procurement.

REFERENCES

Amalda, R. N., Millah, N., & Fitria, I. (2022). Implementasi Algoritma C5.0 Dalam Menganalisa Kelayakan Penerima Keringanan UKT Mahasiswa ITK. *Teorema: Teori Dan Riset Matematika*, *7*(1), 101. https://doi.org/10.25157/teorema.v7i1.6692

Apriyadi, A., Lubis, M. R., & Damanik, B. E. (2022). Penerapan Algoritma C5.0 Dalam Menentukan Tingkat Pemahaman Mahasiswa Terhadap Pembelajaran Daring. *Komputa: Jurnal Ilmiah Komputer Dan Informatika*, *11*(1), 11–20. https://doi.org/10.34010/komputa.v11i1.7386

Krajewski, L. J., Ritzman, L. P., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2020). *Purchasing and Supply Chain Management*. Cengage Learning.

Lonang, S., Yudhana, A., & Biddinika, M. K. (2023). Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting. *Jurnal Media Informatika Budidarma*, *7*(4), 2109. https://doi.org/10.30865/mib.v7i4.6553

Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2020). *Purchasing and Supply Chain Management*. Cengage Learning.

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, *5*(2), 697–711.

Patel, S., & Shah, R. (2021). Boosting Techniques in Decision Trees: Enhancing Prediction Accuracy in Business Applications. *International Journal of Data Analytics*, *9*(4), 112–120.

Purwanti, E., Nor, R. U. N. U., & Soelistyono, S. (2023). Web Design for Stroke Early Detection Using Decision Tree C5.0. *Komputasi: Jurnal Ilmiah Ilmu Komputer Dan Matematika*, *20*(2), 135–147. https://doi.org/10.33751/komputasi.v20i2.8265

Rachman, A. N., Ramdani, C. M. S., & Insani, M. H. (2024). Implementation Of C5.0 Classification And Support Vector Machine Algorithm With Correlation-Based Feature Selection In Application Liver Disease. *International Journal of Applied Information Systems and Informatics (JAISI)*, *2*(1). https://doi.org/10.37058/jaisi.v2i1.10848

Saraswati, A., Aini, Q., & Utami, L. M. (2022). Perbandingan Algoritma Random Forest dan C4.5 untuk Prediksi Permintaan Barang. *Jurnal Ilmiah Informatika*, *8*(2), 112–120.

Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, *4*(1), 13–21. https://doi.org/10.30813/jbase.v4i1.2729

Utomo, D. P., Sirait, P., & Yunis, R. (2020). Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0. *Jurnal Media Informatika Budidarma*, *4*(4), 994–1006. https://doi.org/10.30865/mib.v4i4.2355

Wijaya, H. M., Deswantoro, G., & Hidayat, R. (2021). Analisis Perencanaan Supply Chain Management (Scm) Pada Pt. Kylo Kopi Indonesia. *Jurnal Ekonomi Manajemen Sistem Informasi*, *2*(6), 795–806. https://doi.org/10.31933/jemsi.v2i6.653

Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Portal Data*, *2*(3), 1–12.