Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 417~428
e-ISSN: 2548-6861

Association Rule Mining for Truck Body Damage Pattern Analysis Using Apriori and CRISP-DM

Livanty Efatania Dendy[1], Rinabi Tanamal [2]
Information System for Business, Universitas Ciputra Surabaya
lefatania@student.ciputra.ac.id [1], r.tanamal@ciputra.ac.id [2]

Article Info

Article history:

Received 2025-12-30
Revised 2026-01-22
Accepted 2026-01-30

Keyword:

Association Rule Mining,
Apriori Algorithm,
CRISP-DM,
Truck Body Damage,
Preventive Maintenance.

ABSTRACT

This study investigates damage patterns in truck body components by applying the Apriori association rule mining algorithm within the CRISP-DM framework. The analysis is based on 281 historical repair records from CV Lestari's fleet throughout 2024. The dataset encompasses 14 attributes, including vehicle types, route categories, body materials, and load conditions. To ensure the robustness of the generated rules, parameter tuning was conducted using a grid search approach, resulting in minimum support and confidence thresholds of 15% and 60%, respectively. A total of 50 association rules were derived, with several rules demonstrating high confidence values and lift values above 1.1, indicating meaningful and non-random correlations. Notably, structural frame damage is strongly associated with mountainous routes and heavy loads, while door and hinge damage tends to occur in aluminum box-bodied trucks operating under medium loads. These patterns were aligned with practical insights from field technicians and further contextualized through technical recommendations, such as reinforcing high-stress points and adjusting inspection schedules for high-risk configurations. The findings support the formulation of predictive maintenance strategies, enabling companies to transition from reactive repairs to proactive, data-driven decision-making. By integrating rule-based insights into maintenance planning, the study contributes to reducing unexpected failures, optimizing inspection frequency, and enhancing overall fleet reliability.

I. INTRODUCTION

Indonesia's economic growth in recent decades has consistently been supported by several key sectors, one of which is the transportation and logistics sector. This sector not only functions as a connector between regions but also serves as an important indicator in assessing the efficiency and resilience of the national distribution system. Based on official data from Statistics Indonesia (Badan Pusat Statistik) for the year 2025, the Transportation and Warehousing sector recorded a contribution of 6.09 percent to the national Gross Domestic Product in the first quarter, which then increased to 6.21 percent in the second quarter and returned to 6.10 percent in the third quarter. Among all sub sectors, land transportation contributed the largest share, reaching 2.92 percent in the first quarter, 2.96 percent in the second quarter, and 2.93 percent in the third quarter. This illustrates the dominance of trucks in supporting Indonesia's logistics system [1]. The significant growth in the logistics sector has been driven by various factors including an increase in export and import volumes as well as the rapid rise of electronic commerce and regional distribution activities [2]. Along with the growing contribution of the transportation sector to GDP, the number of motor vehicles, especially trucks that play a role in national logistics distribution, has also shown an upward trend. According to data from BPS, the development of the number of trucks in Indonesia during the period from 2020 to 2024 can be seen in the following table [3].

This growth indicates an increase in the intensity of land freight vehicle usage, particularly trucks, in supporting national goods distribution activities. Freight transportation in Indonesia is generally dominated by the use of trucks as the primary mode of transport due to their high flexibility in reaching various routes, especially medium to short-distance routes that connect production centers, ports, and markets.

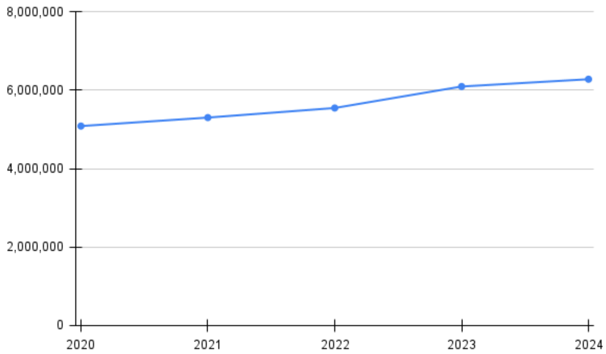

Figure 1. Growth in the Number of Motor Vehicles (Trucks), 2020 to 2024

Trucks and semi-trailers play a dominant role in Indonesia's freight transport system, both in terms of cargo volume and distribution reach [4]. This high reliance on land transportation directly demands long-term reliability of logistics vehicles, not only in terms of engine performance but also in the construction of the vehicle body and its structural strength. Therefore, the role of the vehicle body manufacturing industry is crucial in enhancing the efficiency, safety, and resilience of the national logistics distribution system.

In this context, various body manufacturing companies at both national and regional levels play a strategic role in the provision and maintenance of reliable commercial vehicle bodies. One such company is CV Lestari, a body manufacturer located in Mojotengah Subdistrict, Gresik Regency, East Java. The company operates in the field of production and assembly of commercial vehicle bodies, with a primary focus on manufacturing truck bodies for logistics distribution, agricultural industry, and construction material transportation. Through a made-to-order approach, CV Lestari designs and manufactures various types of bodies such as open beds, water tanks, dump trucks, and specialized cargo vehicles tailored to the needs of local and regional customers. In addition to manufacturing new truck bodies, CV Lestari also provides repair and maintenance services for commercial vehicles, particularly for the body and structural parts of the vehicle. As the demand for repair services continues to increase, the company has begun to encounter numerous cases of recurring damage with varying characteristics and causes. Most trucks brought into the workshop experience serious structural issues primarily caused by poor road conditions, such as uneven surfaces, large potholes, and road deformations. These irregularities in road surfaces expose trucks to repeated dynamic loads, especially when traveling at high speeds or with frequent usage [5]. Such loads generate vertical impact forces that are transmitted to the main structural components of the vehicle, including the frame, chassis, and inter-component joints, eventually leading to the formation of microcracks and permanent deformations that worsen over time [6]. Structural damage such as cracks in the frame, chassis deformations, or worn joints disrupt the vehicle's load distribution and increase the risk of mechanical failure during road operation, posing a danger to drivers and other road users. Therefore, early detection of minor cracks is essential, as micro-damage can escalate into major failures if left unaddressed [7]. Moreover, this condition accelerates material fatigue, increases the likelihood of sudden breakdowns, and significantly shortens the vehicle's service life [8].

Every truck undergoing repairs at CV Lestari is recorded through a Repair Order Form, an administrative document that contains detailed information about the vehicle unit, body type, damage level, materials used, and repair actions taken. The data in the SPP is strategically valuable, as it reflects patterns and tendencies of damage across various vehicle types and operating conditions. However, to date, this data has only been used administratively and has not yet been analyzed to identify relationships between the underlying causes of damage. Therefore, it is necessary to transform the data into a structured digital format in order to enable its use as a foundation for data analysis [9].

To optimize the utilization of such data, a data driven analytical approach is needed one that can extract hidden patterns from the repair dataset. One such approach is data mining, a process that leverages statistics, mathematics, artificial intelligence, and machine learning to extract and identify useful information and knowledge from large datasets [10]. Within data mining, various methods are available to meet analytical needs; one of the most widely used is association rule mining, which is designed to uncover correlations or significant patterns among items within numerous transactions [11]. A commonly used algorithm in this method is the Apriori algorithm, known for its simplicity, ease of implementation, and effectiveness in determining frequent itemsets [12]. Using the Apriori algorithm, it is possible to generate association rules that represent relationships between items, such as: "if item A and B appear, then item C is also likely to appear," based on support and confidence values [13]. This method is particularly well-suited for analyzing large-scale transaction data, such as the SPP (Repair Order Form) data of CV Lestari, to identify the most frequent patterns of damage and uncover the relationships among their causes [14]. The application of association rule methods based on the Apriori algorithm has been widely adopted in various previous studies.

For example, in study [12] the Apriori algorithm was used to analyze the relationships among factors contributing to traffic accidents in Thailand. The results revealed that a combination of variables such as male gender, speeding, motorcycle use, straight road conditions, dry surfaces, and clear weather significantly increased the risk of fatal accidents. Another study [15] applied the Apriori algorithm to analyze product association patterns in retail sales data collected through a Point of Sale (PoS) system. The objective of the study was to identify product relationships that could be utilized to improve sales through relevant product recommendations for customers. The findings indicated that the association pattern with the highest confidence value of 0.61 and a minimum support of 0.003 occurred in March, which coincided with the highest monthly sales, amounting to IDR 295,509,934. These findings suggest that the application of the Apriori algorithm in PoS systems can positively impact retail business revenue.

In addition, study [16] confirmed the effectiveness of the Apriori algorithm in identifying consumer purchasing patterns by analyzing 4,371 transactions using the CRISP-DM approach, resulting in nine significant association rules that were utilized to optimize product arrangement and promotional strategies. A similar approach was applied in the automotive sector [17] to determine spare part recommendation packages based on customer purchasing patterns, where oil filters and air filters achieved confidence values of 68% and 63%, respectively, contributing to improved inventory management accuracy. Furthermore, the application of the Apriori algorithm to truck crash characteristics revealed associations among vehicle attributes, road conditions, and accident severity, demonstrating the capability of association rule mining to identify complex damage patterns in heavy vehicles [18]. In line with this, another study emphasized predictive maintenance for underground trucks and loaders in the mining industry, showing that historical maintenance data analysis supports improved asset availability and more proactive maintenance decision-making [19].

Understanding damage patterns is a crucial foundation for improving repair quality and operational management at CV Lestari. By applying association rule analysis based on the Apriori algorithm, the company can identify the most frequently occurring types of damage, discover relationships among causal factors, and detect recurring patterns that were previously unnoticed. This information allows a shift in the maintenance approach from reactive to more proactive and well planned, including in scheduling inspections, preventing small cracks from developing into major damage, and managing raw material inventory more efficiently.

This approach improves not only the structural reliability of vehicle bodies but also supports operational efficiency by reducing unexpected repair costs and optimizing the service process. Building on this, the study contributes a novel application of the Apriori algorithm in the domain of commercial truck body maintenance. Unlike prior works that typically focus on retail transaction data or POS systems, this research analyzes real-world repair order data collected from CV Lestari, a body manufacturing and repair company. It incorporates previously underexplored operational variables such as route type (mountainous, coastal, urban), load category, and body material composition. These context-specific variables enhance the practical relevance of association rule mining in supporting predictive maintenance strategies.

II. METHODOLOGY

This study employs the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as a framework, which has been widely applied in various business and research contexts due to its flexibility, adaptability across different domains, and well-structured standardized process with clearly defined phases [20]. This methodology consists of six phases as introduced in the process model developed by Wirth and Hipp [21]and the official guidelines by Chapman et al. [22]. The research phases applied in this study are illustrated in Figure 2.

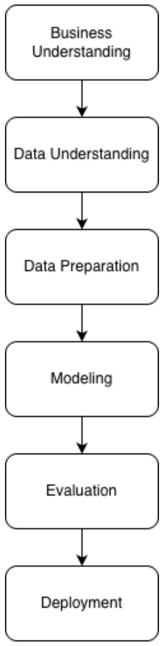

Figure 2. CRISP-DM Diagram

A. Business Understanding

This phase aims to develop a comprehensive understanding of the research objectives, including the analytical needs and the expected outcomes of the data mining process [20]. At this stage, the core problem is identified, the type of analysis to be used is determined, and the business objectives are translated into specific and measurable data mining tasks. In addition, this phase involves assessing available resources, constraints, and the scope of the data to ensure that the work remains relevant to practical needs and that the project aligns with the organization's strategies and requirements.

B. Data Understanding

This phase involves the process of data collection, exploration, structural review, and data quality assessment to understand the initial characteristics of the dataset. The main activities in this stage include examining data completeness, identifying value distributions, recognizing early patterns, and detecting potential issues such as missing values, duplicates, or other irregularities. The initial data exploration conducted during the Data Understanding phase helps identify anomalies and missing values, and these findings will guide the subsequent data preprocessing steps [23].

C. Data Preparation

In this phase, the data obtained during the Data Understanding stage is further processed to ensure it is ready for use in the modeling phase. This process includes data cleaning to remove incomplete or inconsistent values, followed by data transformation to adjust its structure according to the requirements of the methods to be used in the modeling phase [16].

D. Modeling

In this phase, the modeling process is conducted by applying the Apriori algorithm as a data mining technique to generate frequent itemsets based on a defined minimum support value. The minimum support serves as a threshold to determine which item combinations are considered to occur frequently enough in the dataset. Each itemset is then calculated for its support value using the formula presented below.

$$Support(A) = \frac{\text{Number of transactions containing } A}{\text{Total number of transactions}} \times 100$$

In contrast, the support for a pair of items is obtained using the equation presented below.

$$Support(A, B) = \frac{\text{Number of transactions containing } A \text{ and } B}{\text{Total number of transactions}} \times 100$$

Itemsets that meet the minimum support threshold are retained as frequent itemsets, while those that do not meet the threshold are eliminated. This process continues iteratively until no further item combinations satisfy the support criteria.

E. Evaluation

The Evaluation phase aims to assess the quality of the association rules generated during the Modeling phase. In this stage, each rule is analyzed using three key metrics: support, confidence, and lift, to ensure that the rule not only appears frequently in the transaction data but also reflects a significant and meaningful relationship relevant to business decision-making. These metrics are commonly used in association rule mining as they provide a comprehensive view of the frequency, strength, and interestingness of relationships among items within a transaction dataset [15][24].

Among these, confidence measures the degree of certainty in the relationship between the antecedent and the consequent within a rule. A high confidence value indicates that when the antecedent occurs, the consequent is also likely to occur, making the rule strong and relevant for decision-making purposes. The calculation of confidence is shown in the equation below.

$$Confidence\ P(B|A) = \frac{\text{Number of transactions containing } A \text{ and } B}{\text{Number of transactions containing } A} \times 100$$

Lift is used to evaluate whether the relationship between two items is truly significant or merely coincidental. A lift value greater than 1 indicates that the occurrence of the antecedent increases the likelihood of the consequent occurring, meaning that the two items are meaningfully associated. The formula for calculating lift is presented in the equation below.

$$Lift(A \Rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A) \times \text{Support}(B)}$$

F. Deployment

In the Deployment phase, the results obtained from the data mining process are translated into outputs that can be utilized by stakeholders. This phase not only focuses on presenting the results in a technical format but also emphasizes how the findings can be applied in real operational contexts. The insights generated are typically presented in the form of reports, visualizations, pattern summaries, or recommendations that support decision-making processes. The goal of this stage is to ensure that the discovered patterns or rules provide added value and practical relevance.

Before full-scale implementation, it is important to validate the results through discussions or reviews with relevant stakeholders to ensure that the recommendations align with operational needs and conditions. In some cases, limited trials such as pilot testing or controlled experiments may be conducted to evaluate the effectiveness of the deployment before proceeding to full implementation.

III. RESULT AND DISCUSSION

In this study, the data were obtained from CV Lestari in the form of SPP (Repair Order Form) documents. The data were then converted into Microsoft Excel format to enable computational processing. Subsequently, the dataset was imported into the Jupyter Notebook environment and processed using the Python 3 programming language. The Apriori algorithm was applied to extract association patterns from the dataset. The processing results were then exported into CSV format for further analysis and interpretation in accordance with the stages of the CRISP-DM framework.

A. Business Understanding

At this stage, the study focuses on understanding the business needs and challenges faced by CV Lestari as a coachbuilding company engaged in the assembly, repair, and maintenance of truck bodies. The company has experienced an increase in the number of vehicles requiring repair; however, it does not yet have a system capable of identifying frequently occurring damage patterns. As a result, damage analysis is currently conducted manually, which is time-consuming and may lead to inaccuracies in decision-making.

This study aims to assist the company in gaining insights into the relationships among different types of damage through the application of the Apriori algorithm. By identifying combinations of damage that frequently occur together, the company can leverage the analysis results to improve repair process efficiency, estimate material requirements more accurately, design more appropriate service packages, and minimize the risk of recurring damage. Furthermore, the findings of this analysis can serve as a foundation for developing data-driven maintenance strategies and informing managerial decision making regarding vehicle repair prioritization and resource planning.

B. Data Understanding

The data used in this study were obtained from internal documents in the form of Repair Order Forms (SPP) belonging to CV Lestari, a coachbuilding company that recorded all truck body repair activities throughout 2024. These documents contain 281 service transaction records, representing actual damage conditions encountered in various truck units received by the workshop. The information documented includes vehicle specifications (such as body type and material), operational conditions (including route type and load category), and the types of body damage identified. Given the potential for incomplete or inconsistent records, additional data verification was conducted through direct interviews with technicians and workshop personnel who were directly involved in the repair processes. This step was taken to ensure the completeness and validity of the information, particularly for attributes not explicitly recorded in the forms.

As a result of the data collection and verification process, a final dataset consisting of 14 attributes was compiled, encompassing both technical and operational aspects of the vehicles. To provide a more comprehensive understanding of the dataset structure used in this study, the column names and attribute descriptions are presented in Table 1 below.

TABLE I
DATASET

No	Column Name	Column Description
1	Date	The date when the truck unit entered the workshop for coachbuilding repair or inspection.
2	Vehicle	The brand or model of the truck undergoing repair services.
3	Body Type	The type of truck body used (e.g., steel flatbed, aluminum box, drop-side body).
4	Material	The primary material used for the truck body (e.g., steel, aluminum, checker plate, or a combination).
5	Load Category	The typical load classification of the vehicle (light, medium, or heavy).
6	Operational Route	The characteristics of the truck's operational route (coastal, urban, or mountainous).
7	Corrosion / Material Damage	Damage conditions caused by corrosion, rust, oxidation, or material degradation of the truck body, including surface peeling, material thinning, perforation, and deterioration due to environmental exposure.
8	Floor Structure Damage	Physical damage to the truck body floor, including cracks, perforations, deformation, surface wear, loosened panels, or weakened floor joints caused by load stress or prolonged use.
9	Structural Frame / Support Damage	Damage to the main coachbuilding frame or supporting structures, including cracks, fractures, bending, misalignment, or structural weakening.
10	Welding / Joint Damage	Damage to welded areas or joints between truck body panels, including cracked welds, joint separation, weakened connections, or weld failures.
11	Door / Hinge Damage	Problems affecting truck body doors and hinges, including jamming, corrosion, misalignment, loosened or broken hinges, and improper door closure.
12	Paint / Finishing Damage	Damage to the exterior appearance of the truck body, including paint peeling, scratches, fading, or surface deterioration.
13	Electrical / Lighting Damage	Damage affecting the truck body electrical system, including lamps, wiring, connectors, or related electrical components.
14	Interior / Lining Damage	Damage affecting the interior body components, including inner panels, linings, or internal coverings.

To provide a clearer overview of the attribute distributions within the dataset, the following pie charts illustrate the proportions body types, load categories, and operational routes recorded in the truck repair data.

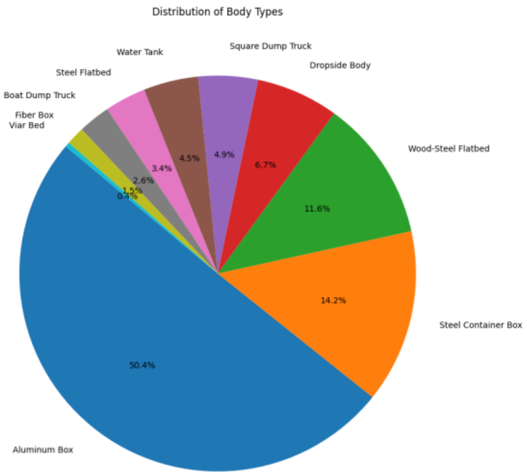

Figure 3. Distribution of Body Type

This chart shows that the most common operational route is the coastal route, representing approximately 41.0% of all truck usage. This is closely followed by the mountainous route at 37.7%, while the urban route is the least common with 21.3%. This distribution suggests that a significant number of trucks operate in geographically challenging environments, particularly in mountainous and coastal areas, which may contribute to the observed damage patterns.

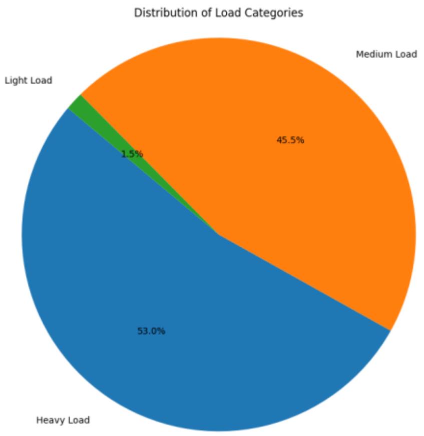

Figure 4. Distribution of Load Categories

The load category distribution indicates that heavy loads dominate the dataset, accounting for 53.0% of all entries. Medium loads follow with 45.5%, while light loads are rarely recorded only 1.5%. The prevalence of heavy and medium loads highlights the high operational stress placed on truck components, which is a critical factor in damage analysis.

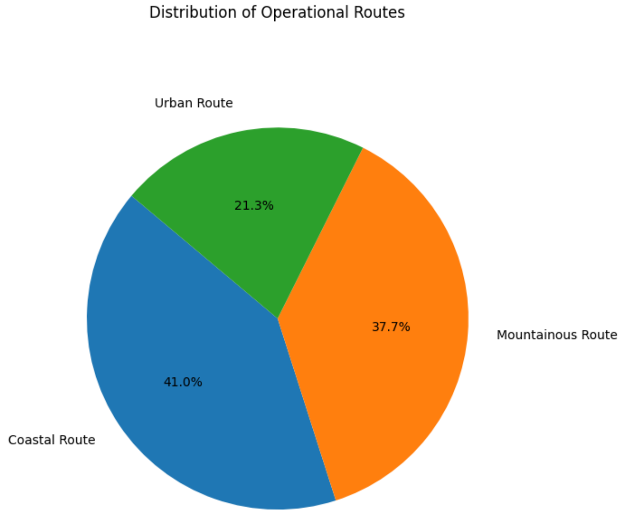

Figure 5. Distribution of Operational Routes

In terms of body type, aluminum box trucks are the most common, making up over 50% of the data. Other notable categories include steel container box 14.2% and wood-steel flatbed 11.6%. The dominance of aluminum based bodies may reflect fleet preferences for lighter materials, although these may also present unique vulnerabilities under certain operational conditions, such as medium-load door damage.

C. Data Preparation

The collected data are still in raw form and therefore require several selection and preprocessing steps to produce the required dataset. The processes carried out at this stage include:

1) *Data Selection*: At this stage, relevant attributes were selected for the data mining process. The date column was excluded because it does not contribute to the discovery of damage association patterns, as the research focuses on relationships among body type, load category, operational route, and types of damage. Therefore, this column was removed from the dataset. Meanwhile, all damage-related columns were retained, as they represent the primary features analyzed in this study.

2) *Data Cleaning*: In this process, data cleaning was performed to address incomplete and inconsistent records. Based on the data inspection results, six rows contained missing values in the Load Category column and seven rows contained missing values in the Operational Route column. These missing values resulted from limitations in the documentation of the *SPP* (Repair Order Form). Follow-up verification was conducted through direct interviews with workshop personnel responsible for handling the respective vehicles; however, in several cases, the information could not be confirmed due to memory limitations regarding the truck

conditions at that time. Records that could not be verified were therefore treated as missing values and subsequently removed from the dataset to prevent any adverse impact on the quality of the analysis.

3) *Data Transformation*: During the data transformation stage, all categorical attributes were converted into binary format to match the transaction item structure required by the Apriori algorithm. This process was performed using the one-hot encoding technique, in which each category of attributes such as Vehicle, Body Type, Material, Load Category, and Operational Route was transformed into a binary column with a value of 1 if the category appeared in a transaction and 0 otherwise. In addition, damage attributes that were originally labeled as Yes and No were also converted into binary representations. The transformation process resulted in a binary tabular dataset consisting of 46 binary attributes. All attributes were represented as integer values to ensure compatibility with frequent itemset generation and association rule mining. Representative examples of the transformed attributes, along with their data types, are presented in Table 2.

TABLE II
REPRESENTATIVE ATTRIBUTES RESULTING FROM ONE-HOT ENCODING

Attribute Category	**Column Name**	**Data Type**
Vehicle	Vehicle_Fuso	int64
Body Type	Body_Type_Dropside	int64
Load Category	Load_Category_Heavy	int64
Operational Route	Route_Mountainous	int64
Damage	Corrosion_or_Material_ Damage_Yes	int64

D. Modeling

At this stage, the modeling process began with determining the minimum support and minimum confidence values as the main parameters of the Apriori algorithm. These parameters were not selected arbitrarily; instead, a grid search approach was employed. This approach systematically evaluates multiple combinations of commonly used support and confidence values in association rule mining to identify threshold values that generate the most relevant and meaningful association rules. In the grid search process, support values ranging from 0.10 to 0.30 and confidence values ranging from 0.40 to 0.80 were examined. Each parameter combination was evaluated in a structured manner to ensure that the selected values were based on objective evaluation rather than subjective estimation [16]. The final parameters used in the modeling stage were chosen based on their ability to produce the most informative itemset relationship patterns within the dataset.

Based on the grid search results, the optimal parameters were identified as a minimum support of 15% and a minimum confidence of 60%, which resulted in a total of 50 association rules. A minimum support threshold of 15% was selected to ensure that only frequently occurring patterns in the data were considered during the analysis. This threshold effectively filters out infrequent item combinations, thereby producing rules that are more representative of actual conditions in the field. Meanwhile, the minimum confidence value of 60% was established to ensure a sufficient level of reliability in the relationships between antecedents and consequents within the generated rules. This threshold indicates that the probability of the consequent occurring given the antecedent is strong enough to support operational decision-making, such as identifying dominant damage patterns or formulating more accurate repair recommendations. The combination of these parameters ensures that the resulting association rules are both significant and meaningful, retaining only frequently occurring damage patterns with strong relationships for further analysis.

Subsequently, the Apriori algorithm was applied to identify association patterns within the coachbuilding damage data using the optimal parameters obtained from the grid search. The algorithm generated frequent itemsets and association rules that represent dominant damage patterns across all vehicle units. In this stage, only itemsets with a value of “Yes” were retained to ensure that the resulting frequent itemsets accurately reflect the presence of damage conditions. This process enables the identification of the most frequently occurring types of damage, which are prioritized for further analysis. Table 3 presents the frequent 1-itemsets with the highest support values, representing the most dominant damage types observed in the dataset.

TABLE III
1 ITEMSETS

No.	**Damage Type**	**Support**
1	Structural Frame / Support Damage	0.839552
2	Door / Hinge Damage	0.481343
3	Floor Structure Damage	0.365672
4	Paint / Finishing Damage	0.332090
5	Welding / Joint Damage	0.305970
6	Corrosion / Material Damage	0.235075
7	Electrical / Lighting Damage	0.197761

Based on the results of the frequent 1-itemset analysis presented in Table 3, several damage categories exhibit relatively high occurrence rates in the service transaction data. Structural Frame/Support Damage emerges as the most dominant type of damage, with a support value of 83.9%, indicating that the majority of truck units experience damage to the main frame or supporting structure. Furthermore, Door/Hinge Damage and Floor Structure Damage also show substantial occurrence frequencies, with support values of 48.1% and 36.6% respectively.

Meanwhile, other damage types such as Paint/Finishing, Welding/Joint, Corrosion/Material, and Electrical/Lighting Components demonstrate lower support values, indicating that these types of damage occur less frequently compared to structural damage. These findings suggest that issues related to the vehicle frame and main structural components constitute the dominant problems and should therefore be prioritized in maintenance and repair activities at CV Lestari.

After obtaining an overview of the most frequently occurring damage types through the frequent 1-itemset analysis, the analysis was extended to identify frequent 2-itemsets. At this stage, the Apriori algorithm identifies combinations of two items that co-occur within service transactions with a minimum support threshold of 0.15 (15%). Each combination consists of one vehicle-related factor, such as body type, material, load category, or operational route, and one damage type with a Yes value.

The frequent 2-itemset analysis provides an initial insight into the relationships between vehicle characteristics and the types of damage that occur. Through these co-occurrence patterns, associations between operational conditions and specific body damage types can be identified prior to the formation of association rules. Table 4 presents the list of frequent 2-itemsets with the highest support values, representing the most dominant damage patterns observed in the analyzed vehicle units.

TABLE IV
2 ITEMSET

No	Support	Factor	Associated Damage
1	0.507	Aluminum Box Body Type	Structural Frame/Support Damage
2	0.447	Aluminum Box Body Type	Door/Hinge Damage
3	0.444	Steel + Aluminum Material	Door/Hinge Damage
4	0.395	Aluminum Box Body Type	Structural Frame/Support Damage
5	0.391	Medium Load Category	Door/Hinge Damage
6	0.391	Steel Material	Structural Frame/Support Damage
7	0.376	Door/Hinge Damage	Structural Frame/Support Damage
8	0.376	Mountainous Route	Structural Frame/Support Damage
9	0.358	Coastal Route	Structural Frame/Support Damage
10	0.350	Mitsubishi Vehicle Type	Structural Frame/Support Damage

Based on the results of the frequent 2-itemset analysis, the most frequently co-occurring item combinations were identified from the vehicle body damage data throughout 2024. The most dominant pattern indicates a strong association between Structural Frame/Support Damage and various vehicle-related factors. This damage type exhibits the highest co-occurrence rate with vehicle factors, with a support value of 50.7%, making it the most frequent 2-itemset pattern observed in the analyzed units at CV Lestari. Conversely, damage patterns with lower support values represent less frequently occurring combinations, indicating that these factors or damage types are comparatively rarer in the observed dataset.

In addition, Door/Hinge Damage, Paint/Finishing Damage, and Welding/Joint Damage are also frequently found to co-occur with several vehicle characteristics, including body type, load category, and body material. For instance, the Aluminum Box body type is commonly associated with Door/Hinge Damage, while medium and heavy load categories tend to co-occur with structural and finishing-related damage, with co-occurrence rates exceeding 35%. These patterns suggest that the physical characteristics of vehicles play a significant role in influencing the types of damage that occur.

Overall, the frequent 2-itemset results indicate that structural damage, particularly damage to the frame and supporting components, represents the most recurrent type of damage. This pattern is closely related to high operational workloads and intensive vehicle usage, especially for units operating on specific routes or carrying heavy loads. These findings provide valuable insights for the company in identifying the primary factors contributing to vehicle damage and can serve as a basis for implementing preventive maintenance strategies, periodic inspections, and more appropriate body material selection.

E. Evaluation

During the evaluation stage, the quality of the generated association rules was assessed using three main metrics: support, confidence, and lift. The support and confidence thresholds used in this phase were 15 percent and 60 percent, respectively, and were determined using a grid search strategy, as detailed in the Modeling section. These metrics were employed to evaluate the strength of the relationships between causal factors (antecedents) and damage types (consequents). To assess the overall quality of the generated association rules, an analysis of the confidence value distribution was conducted for all rules that passed the filtering process.

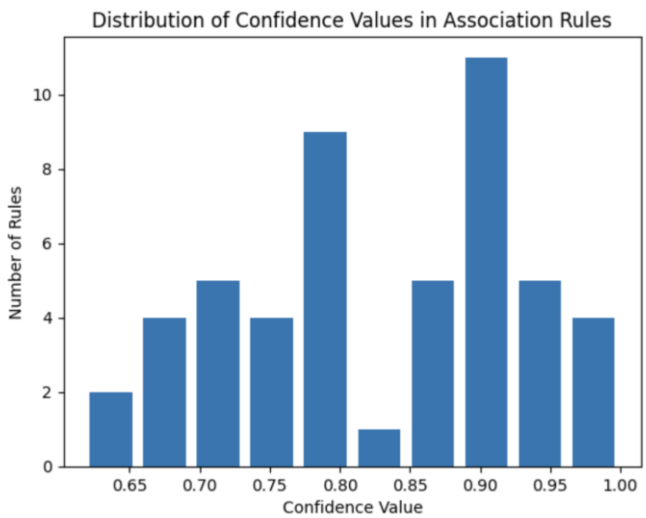

Figure 6. Distribution of Confidence Values in Association Rules

Based on Figure 6, most association rules exhibit confidence values above 0.65, with several rules approaching 1.00. This distribution indicates that the majority of the generated association rules have a high level of reliability, suggesting that the relationships between vehicle-related factors and damage types are consistent and trustworthy. In addition, the absence of rules with low confidence values indicates that the selected confidence threshold has effectively filtered out less meaningful associations.

After evaluating the overall quality of the association rules through the confidence distribution, the analysis was further extended to examine individual association rules in more detail based on their support, confidence, and lift values.

TABLE V
ASSOCIATION RULES

Antecedent	Consequent	Support	Confidence	Lift
Mountainous Route	Structural Frame/Support Damage	0.377	1.000	1.191
Mountainous Route, Heavy Load	Structural Frame/Support Damage	0.243	1.000	1.191
Aluminum Box Body Type, Mountainous Route	Structural Frame/Support Damage	0.172	1.000	1.191
Mountainous Route, Steel + Aluminum Material	Structural Frame/Support Damage	0.168	1.000	1.191
Heavy Load	Structural Frame/Support Damage	0.507	0.958	1.141
Hino Vehicle Type	Structural Frame/Support Damage	0.205	0.948	1.130
Hino Vehicle Type, Heavy Load	Structural Frame/Support Damage	0.179	0.941	1.121
Steel Material, Heavy Load	Structural Frame/Support Damage	0.295	0.940	1.120
Coastal Route, Heavy Load	Structural Frame/Support Damage	0.243	0.929	1.106
Aluminum Box Body Type, Medium Load	Door/Hinge Damage	0.381	0.911	1.892

Based on the generated association rules, Mountainous Route emerges as the most dominant risk factor strongly associated with Structural Frame/Support Damage. This is evidenced by the top four rules, all of which exhibit a confidence value of 1.000, indicating that all trucks operating on mountainous routes experienced structural frame damage.

However, it is important to note that perfect confidence values may be influenced by a relatively small sample size in certain categories. For instance, the number of trucks operating on mountainous routes is limited, which may inflate confidence scores. Thus, such results should be interpreted with caution and further validated using broader datasets.

Additionally, the lift metric provides further insight, with consistent values around 1.191, reinforcing that the associations are not due to chance. Coupled with support values between 0.16 and 0.37, this confirms that the patterns not only appear frequently but also represent meaningful relationships. These high-lift rules are particularly valuable for informing preventive maintenance strategies. In practical terms, high lift values indicate that certain damage types are significantly more likely to occur under specific operational conditions. Therefore, rules with high lift can be used to establish maintenance inspection priorities, such as focusing preventive checks on vehicles frequently operating on mountainous routes or carrying heavy loads. This approach allows companies to allocate maintenance resources more efficiently and reduce the likelihood of unexpected failures.

In addition to route-related factors, the Heavy Load Category also shows a strong association with structural damage. Rules with Heavy Load as the antecedent exhibit a confidence value of 0.9577, meaning that more than 95% of heavily loaded trucks experienced structural frame damage. When heavy load conditions are combined with specific vehicle types, such as Hino, the confidence value remains high (0.94), reinforcing the notion that excessive loading is closely associated with accelerated deterioration of structural components. This pattern is further supported by lift values greater than 1.1, indicating an increased likelihood of damage compared to normal operating conditions.

Conversely, Door/Hinge Damage exhibits a different pattern. The combination of Aluminum Box Body Type and Medium Load Category produces an association rule with a confidence value of 0.9107 and a lift value of 1.892, which is among the highest lift values observed across all rules. This finding suggests that lighter aluminum body materials tend to be more susceptible to door and hinge damage when operating under medium load conditions. With a support value of 0.38, this pattern also occurs relatively frequently, making it operationally relevant. In addition to the numerical analysis, a visual representation is provided in Figure 7 to enhance understanding of how operational factor combinations relate to specific damage outcomes.

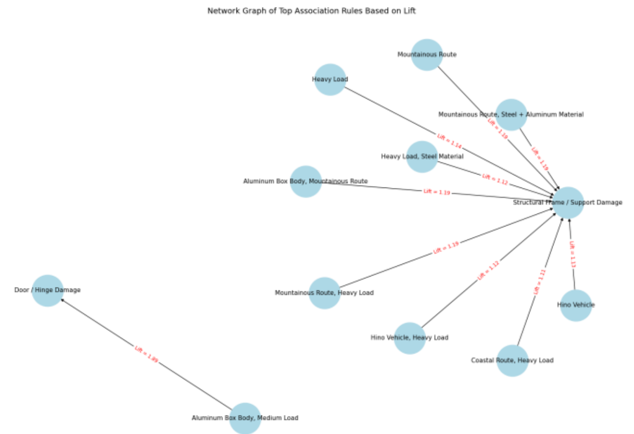

Figure 7. Visualization of Top Association Rules Based on Lift

Figure 7 presents a visual representation of the relationships between combinations of operational factors and the most frequently occurring types of truck body damage. In this visualization, each node represents an antecedent or consequent, and the directed edges illustrate association rules, with the strength of each relationship indicated by its corresponding lift value.

Visually, structural frame damage emerges as the central node, receiving connections from various combinations of factors such as route types, load categories, and vehicle characteristics. In contrast, door/hinge damage is linked to only one specific combination, yet still exhibits a notably strong association. This visual layout not only highlights the dominant patterns but also underscores the complexity of factor interactions influencing damage outcomes. The graph reinforces the importance of designing context aware maintenance strategies grounded in historical data insights. This finding aligns with previous research by Hong et al. [18], who applied association rule algorithms to analyze accident patterns in heavy duty vehicles. Their study demonstrated that specific combinations of road types and vehicle configurations significantly influence accident severity, thereby supporting the relevance of rule based approaches in systematically identifying operational damage patterns.

Overall, the evaluation results based on support, confidence, and lift indicate that truck damage patterns do not occur randomly, but are strongly influenced by combinations of operational factors such as travel routes, load categories, body materials, and vehicle types. These findings can serve as a basis for formulating practical recommendations, including route optimization, load control policies, and targeted component inspections, tailored to the specific risk profiles identified in this study.

F. Deployment

The deployment stage focuses on leveraging the discovered association rule patterns as a basis for improving business processes, particularly in the management of truck body maintenance at CV Lestari. The results of this study indicate that several operational factors, such as mountainous routes, heavy load categories, and specific combinations of routes and body materials, are consistently associated with an increased risk of structural frame damage. These findings provide a strong foundation for the implementation of data-driven preventive maintenance strategies. Interviews with workshop personnel further confirm that the identified patterns align with technicians' practical experience in the field, thereby reinforcing the practical relevance of the analytical results. Consequently, the findings of this study can support operational decision-making and the formulation of more targeted maintenance policies.

In practical terms, the identified patterns can be utilized to prioritize preventive maintenance activities by assigning trucks that frequently operate on high-risk routes or carry heavy loads to early inspection schedules, thereby reducing the likelihood of severe structural damage. Furthermore, the relationships observed between body type or material and specific damage types enable the workshop to provide more accurate technical recommendations to customers, particularly in selecting body materials that are better suited to certain operational conditions. In addition, the digitalization of vehicle damage records and operational data into a structured database can facilitate systematic documentation, further analysis, and the development of risk monitoring dashboards for fleet management. To operationalize the discovered association rules into structured maintenance strategies, a direct mapping can be established between high-confidence rules and specific technical interventions. For example, the rule linking mountainous routes to structural frame damage (confidence = 1.000, lift = 1.191) supports the implementation of regular underframe inspections and structural reinforcements for trucks that frequently operate in such environments. Trucks routinely traversing mountainous areas such as Lembang or Pacet may be scheduled for bimonthly visual inspections and underframe thickness testing. Upon detection of material fatigue or corrosion, the workshop can apply reinforcement plates at critical stress points and conduct re-welding procedures to preserve structural integrity.

Similarly, the rule showing that aluminum box bodies combined with medium load conditions are associated with door and hinge damage (confidence = 0.91, lift = 1.89) can be translated into hinge alignment inspections and the reinforcement of door components on vehicles with this configuration. Other rules can also inform more specialized preventive actions. For instance, the strong association between coastal routes and heavy loads with structural damage (confidence = 0.929) may justify scheduled anti-corrosion coating applications and fatigue testing at key structural joints. Trucks with steel bodies carrying heavy loads which present a confidence level above 0.94 for structural damage should be prioritized for re-welding or the installation of gusset plates, especially in older units.

In the long term, these patterns can serve as the foundation for developing a predictive maintenance Standard Operating Procedure (SOP) tailored to CV Lestari's fleet. Each operational factor combination (route, load, and body type) can be assigned a custom inspection checklist based on its associated risk profile. This structured, data-driven approach can enhance fleet reliability, extend vehicle lifespan, and reduce both downtime and emergency repair costs.

Such an approach is in line with the findings of Patil et al. [19] who demonstrated that applying predictive maintenance analytics to heavy-duty mining vehicles significantly improves asset availability and reduces unscheduled breakdowns, highlighting the practical benefits of transitioning from reactive to proactive maintenance frameworks. The importance of implementing such rule-based preventive maintenance frameworks is further supported by studies showing that risk-based inspection models, when informed by pattern discovery techniques, can greatly enhance maintenance prioritization and operational safety [25].

Although the analytical results reveal strong damage patterns, as indicated by high support, confidence, and lift values, the findings of this study remain recommendatory in nature. Time constraints limited the scope of the research, preventing the inclusion of experimental validation methods such as A/B testing or direct field evaluations. Therefore, full-scale operational implementation is recommended to be conducted gradually, through pilot testing and data-driven evaluation, to ensure that the proposed strategies effectively reduce damage incidence and maintenance costs before being widely adopted.

IV. CONCLUSION

This study applied the Apriori association rule mining algorithm within the CRISP-DM framework to analyze historical truck body repair data at CV Lestari. The primary objective was to identify dominant damage patterns and to examine the relationships between vehicle operational factors and types of truck body damage. Through data transformation into a transactional format and systematic evaluation of association rules using support, confidence, and lift, the study successfully identified meaningful and operationally relevant patterns. The analysis revealed that structural frame and support damage is the most dominant type of damage observed in the service records. The highest-quality association rules indicate that mountainous routes are strongly associated with the occurrence of structural damage, as reflected by the highest confidence values. In addition to route-related factors, the heavy load category also shows a consistent association with an increased risk of structural damage, suggesting that operational load plays a critical role in accelerating the deterioration of structural components.

Conversely, the results demonstrate that damage patterns are not uniform across all components. Door and hinge damage, for instance, exhibits a stronger association with specific combinations of body type and medium load category, rather than with route conditions. This finding highlights that different types of damage are influenced by distinct operational and physical factors, indicating that a one-size-fits-all maintenance approach may not be effective.

Overall, the findings confirm that truck body damage patterns do not occur randomly but are shaped by the interaction of operational factors and vehicle characteristics. These insights provide a valuable basis for implementing data-driven preventive maintenance strategies, such as prioritizing inspections for high-risk vehicles and offering more targeted technical recommendations. However, as this study remains recommendatory in nature and does not include experimental validation, future research is encouraged to conduct pilot studies or field evaluations to assess the real-world effectiveness of the proposed strategies.

However, it is important to acknowledge that this study is based on a single case study involving CV Lestari and limited to the year 2024. As such, the generalizability of the findings remains limited. The association rules and insights presented here are context-specific and may not fully apply to other organizations, regions, or timeframes. Therefore, further studies involving more diverse datasets across multiple companies and longer periods are recommended to strengthen external validity and support broader implementation.

REFERENCES

[1] Badan Pusat Statistik, “Distribusi PDB Menurut Lapangan Usaha Seri 2010 Atas Dasar Harga Berlaku (Persen),” Badan Pusat Statistik. Accessed: Nov. 10, 2025. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/MTA2IzI=/-seri-2010--distribusi-pdb-menurut-lapangan-usaha-seri-2010-atas-dasar-harga-berlaku--persen-.html

[2] F. Syifa *et al.*, “Peran E-Commerce Dalam Transformasi Ekspor Impor Di Era Digital,” *Sinergi Jurnal Riset Ilmiah*, vol. 1, no. 12, pp. 1332–1340, Dec. 2024, doi: 10.62335.

[3] Badan Pusat Statistik, “Perkembangan Jumlah Kendaraan Bermotor Menurut Jenis (Unit), 2024,” Bada Pusat Statistik. Accessed: Dec. 17, 2025. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/NTcjMg==/perkembangan-jumlah-kendaraan-bermotor-menurut-jenis--unit-.html

[4] A. Budiharjo, P. Buana STJ, Pratikso, R. Mudiyono, and R. Y. Anindita, “The Analysis of Freight Transport in Indonesia: Trailer and Semi-Trailer,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 15, no. 3, pp. 930–939, Jun. 2025, doi: 10.18517/ijaseit.15.3.20319.

[5] O. Slavinska, A. Bubela, O. Razboinikov, O. Davydenko, O. Ivanushko, and I. Kozarchuk, “Assessment of the Dynamic Impact of a Truck on the Bridge Pavement Based on the Proposed Mathematical Model of Vehicle Movement,” *Nanotechnol. Percept.*, vol. 20, no. S1, Mar. 2024, doi: 10.62441/nano-ntp.v20iS1.20.

[6] M. Lin, C. Hu, S. M. Easa, and Z. Jiang, “A New Approach to Predict Dynamic Loads Considering Highway Alignment Using Data Mining Techniques,” *Applied Sciences*, vol. 12, no. 11, p. 5719, Jun. 2022, doi: 10.3390/app12115719.

[7] A. K. Ary, A. R. Prabowo, and F. Imaduddin, “Structural Assessment of Alternative Urban Vehicle Chassis Subjected to Loading and Internal Parameters Using Finite Element Analysis,” 2020.

[8] F. Lisowski and E. Lisowski, “Testing and Fatigue Life Assessment of Timber Truck Stanchions,” *Applied Sciences*, vol. 10, no. 17, p. 6134, Sep. 2020, doi: 10.3390/app10176134.

[9] R. Tanamal and J. Mourent, “Rancang Bangun Aplikasi Point Of Sales Berbasis Website Pada Toko Ramai Rambipuji,” *JATISI*

(Jurnal Teknik Informatika dan Sistem Informasi), vol. 12, no. 3, Sep. 2025, doi: 10.35957/jatisi.v12i3.11903.

[10] B. H. Situmorang, A. Isra, D. Paragya, and D. A. A. Adhieputra, "Apriori Algorithm Application for Consumer Purchase Patterns Analysis," *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, vol. 21, no. 1, pp. 15–20, Jan. 2024, doi: 10.33751/komputasi.v21i1.9260.

[11] S. Supriadi, G. Gunawan, RG. G. Alam, and D. Abdullah, "Implementation Of Apriori Data Mining Algorithm To Increase Sales Of Caringin Shop," *Jurnal Komputer, Informasi dan Teknologi*, vol. 4, no. 1, p. 14, Aug. 2024, doi: 10.53697/jkomitek.v4i1.1799.

[12] I. Mohamad, R. Kasemsri, V. Ratanavaraha, and S. Jomnonkwao, "Application of the Apriori Algorithm for Traffic Crash Analysis in Thailand," *Safety*, vol. 9, no. 3, p. 58, Aug. 2023, doi: 10.3390/safety9030058.

[13] D. Dwiputra, A. Mulyo Widodo, H. Akbar, and G. Firmansyah, "Evaluating the Performance of Association Rules in Apriori and FP-Growth Algorithms: Market Basket Analysis to Discover Rules of Item Combinations," *Journal of World Science*, vol. 2, no. 8, pp. 1229–1248, Aug. 2023, doi: 10.58344/jws.v2i8.403.

[14] S. Marselina, J. H. Jaman, and D. E. Kurniawan, "Sales Analysis Using Apriori Algorithm in Data Mining Application on Food and Beverage (F&B) Transactions," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 218–223, Nov. 2023, doi: 10.30871/jaic.v7i2.5026.

[15] I. A. Ashari, A. Wirasto, D. Nugroho Triwibowo, and P. Purwono, "Implementasi Market Basket Analysis dengan Algoritma Apriori untuk Analisis Pendapatan Usaha Retail," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 701–709, Jul. 2022, doi: 10.30812/matrik.v21i3.1439.

[16] A. Ivana and I. Maryati, "Implementation of Apriori Algorithm in Identifying Purchase Relationships at Bluder Cokro Pakuwon Mall," *Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 556–563, Apr. 2025, doi: 10.30871/jaic.v9i2.9154.

[17] Y. A. Alhillah, W. Priatna, and A. Fitriyani, "Implementation of Apriori Algorithm for Determining Spare Parts Product Recommendation Packages," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 212–217, Nov. 2023, doi: 10.30871/jaic.v7i2.5589.

[18] J. Hong, R. Tamakloe, and D. Park, "Discovering Insightful Rules among Truck Crash Characteristics using Apriori Algorithm," *J. Adv. Transp.*, vol. 2020, pp. 1–16, Jan. 2020, doi: 10.1155/2020/4323816.

[19] S. D. Patil, A. Mitra, K. Tuggali Katarikonda, and J.-D. Wansink, "Predictive asset availability optimization for underground trucks and loaders in the mining industry," *OPSEARCH*, vol. 58, no. 3, pp. 751–772, Sep. 2021, doi: 10.1007/s12597-020-00502-4.

[20] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[21] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," in *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, London, UK, 2000, pp. 29–39.

[22] P. Chapman *et al.*, "CRISP-DM 1.0 Step-by-step data mining guide," DaimlerChrysler, Munich, Germany, 1999. Accessed: Dec. 21, 2025. [Online]. Available: https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf

[23] S. Mohammed *et al.*, "The effects of data quality on machine learning performance on tabular data," *Inf. Syst.*, vol. 132, p. 102549, Jul. 2025, doi: 10.1016/j.is.2025.102549.

[24] J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, "A survey on the use of association rules mining techniques in textual social media," *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 1175–1200, Feb. 2023, doi: 10.1007/s10462-022-10196-3.

[25] E. Oropallo, P. Piscopo, P. Centobelli, R. Cerchione, E. Nuevo, and A. Rodríguez-Prieto, "A decision support system to assess the operational safety and economic benefits of risk-based inspection implementation strategies," *Saf. Sci.*, vol. 177, p. 106570, Sep. 2024, doi: 10.1016/j.ssci.2024.106570.