Journal of Emerging Information System and Business Intelligence
ISSN: 2774-3993
Journal homepage: https://ejournal.unesa.ac.id/index.php/JEISBI/

Prediction of Goods Damage in Land Transportation Services (Trucking) Using Naïve Bayes

Moh. Fatihul Farras Dzulfaqqor[1], I Kadek Dwi Nuryana[2]
[1]*Universitas Negeri Surabaya, Surabaya, Indonesia*
mohfatihul.20102@mhs.unesa.ac.id, *dwinuryana@unesa.ac.id*

ABSTRACT

Land transportation plays a crucial role in supporting economic dynamics, yet cargo damage remains a significant challenge that affects the supply chain, efficiency, customer satisfaction, and financial stability. This research aims to analyze the factors influencing cargo damage and predict its likelihood using the Naïve Bayes classification algorithm. A case study was conducted at PT. Tuntas Smart Solusi, a logistics company in Gresik, Indonesia. The Knowledge Discovery in Databases (KDD) framework was employed to process historical shipping data from 2021 to 2023, incorporating variables such as cargo type, shipping route, weather conditions, and load capacity. The results indicate that adverse weather conditions, excessive load weight, and rough routes significantly contribute to cargo damage rates. The Naïve Bayes classifier demonstrated high predictive accuracy, validated using k-fold cross-validation, proving its effectiveness in logistics risk assessment. The findings offer strategic recommendations for logistics companies to minimize damage risks, including optimized packaging strategies, route selection improvements, and predictive monitoring systems. By integrating machine learning-based predictive analytics, logistics firms can enhance operational efficiency, reduce financial losses, and improve overall service quality.

Keyword: Land transportation, Cargo damage prediction, Naïve Bayes, Logistics analytics, Risk assessment, Predictive modelling.

Article Info:	**Corresponding Author**
Article history:	Moh. Fatihul Farras Dzulfaqqor
Received February 24, 2025	Universitas Negeri Surabaya, Surabaya, Indonesia
Revised October 01, 2025	mohfatihul.20102@mhs.unesa.ac.id
Accepted October 13, 2025	

1. INTRODUCTION

The logistics sector, particularly land transportation, is vital for economic sustainability. Efficient transportation ensures the smooth distribution of goods across various regions, supporting trade and commerce. However, cargo damage during transit remains a significant challenge, affecting service quality, financial stability, and customer trust. Studies have shown that damaged goods can lead to financial losses, delayed shipments, and reputational harm for logistics companies.

PT. Tuntas Smart Solusi, a logistics company in Gresik, Indonesia, frequently encounters cargo damage issues. Traditional damage handling methods, such as manual inspection and post-damage compensation, are reactive rather than proactive. These conventional approaches often lead to increased operational costs and customer dissatisfaction. Given the growing demand for reliable and efficient logistics services, companies must adopt advanced techniques to enhance their risk management strategies.

Machine learning (ML) techniques have gained prominence in predictive analytics, allowing businesses to forecast potential risks and take preventive measures. One such technique, the Naïve Bayes classifier, is widely recognized for its efficiency in probabilistic classification tasks. By analyzing historical data, including factors such as weather conditions, route types, and load capacities, the Naïve Bayes model can predict the likelihood of cargo damage, enabling logistics firms to implement proactive strategies.

This study explores the impact of various factors on cargo damage and evaluates the performance of the Naïve Bayes algorithm in predicting such occurrences. The research aims to provide valuable insights for logistics companies to enhance their operational efficiency, reduce financial losses, and improve customer satisfaction through data-driven decision-making. Furthermore, the study contributes to the broader field of predictive logistics by demonstrating the applicability of machine learning models in real-world transportation scenarios.

2. METHODS

This section explains the steps taken to complete this research. Several stages were carried out, as shown in:

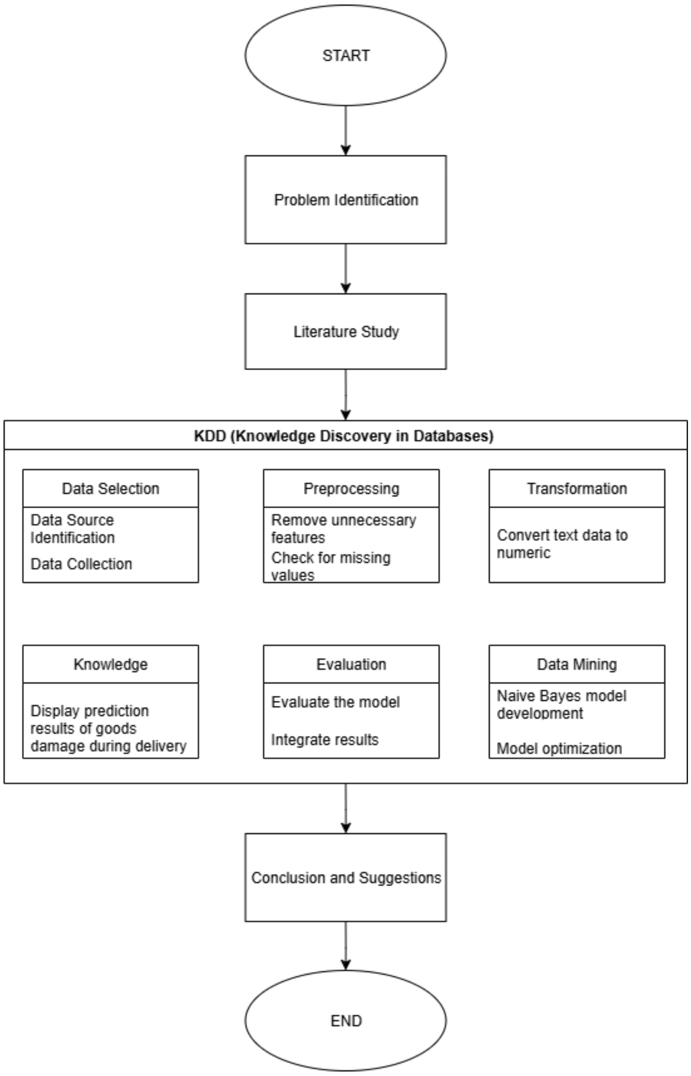

Figure 1. Research Flow

2.1 Problem Identification

The initial stage of this research is problem identification. At PT. Tuntas Smart Solusi, the issue involves handling damaged goods during land transportation services. This challenge arises due to uncertainties related to goods damage during shipping, which can affect customer trust in land transport service companies. This stage thoroughly details the framework of problems faced by PT. Tuntas Smart Solusi, specifically in identifying whether the shipped products are damaged or intact, providing a clear picture of the issues that need to be addressed.

2.2 Literature Review

In the literature review stage, the researcher collects and analyzes various sources relevant to the research topic. This involves reviewing scientific journals, books, articles, and other documents discussing concepts, theories, and previous research findings related to the identified problem. The literature review aims to understand the research context, identify knowledge gaps, and gain insights into methodologies used in prior studies. Additionally, it helps in establishing a solid theoretical foundation and provides references for comparing research results later.

2.3 Data Selection

Data selection is a crucial step in developing a business model for handling damaged goods during land transportation (trucking) services. This step ensures that the data used in analysis and modeling is relevant, high-quality, and reflects real conditions. It is essential for the successful application of the Naive Bayes method in data analysis.

The objectives of data selection include:

- Selecting data directly related to the occurrence of goods damage during transportation.
- Eliminating irrelevant data or noise that may interfere with analysis.
- Providing high-quality data for analysis using the Naive Bayes method.

2.3.1 Data Source Identification

In the initial stage of data source identification, the researcher determines the research objectives: classifying damaged goods during shipping at PT. Tuntas Smart Solusi. The primary data source includes shipping records from 2021-2023 stored in Excel format by PT. Tuntas Smart Solusi. This data contains detailed information on product names, shipment quantities, and damaged goods.

The researcher ensures that the collected data is accurate and provides a comprehensive overview of customer preferences. Initial data validation, cleaning, and structuring processes are conducted to maintain data integrity.

2.3.2 Data Collection

After identifying data sources, the next step is data collection. This process is crucial as it forms the foundation for analysis and modeling. The data collection method is designed to match the characteristics of the required data, specifically data on goods damage during shipping at PT. Tuntas Smart Solusi.

Data Collection Methods:

a. Interviews
    - Structured interviews with administrative staff to gather additional information about factors affecting goods damage during shipping.

ISSN: 2774-3993

b. Observations
   - Direct observation of business processes related to handling damaged goods during shipping.

c. Document Analysis
   - Reviewing company policies, business process management documents, and operational procedures.

2.5 Data Pre-processing

Data pre-processing is an essential step in the Knowledge Discovery in Databases (KDD) process to clean and prepare raw data for analysis and modelling. This stage involves handling missing values, removing outliers, and ensuring relevant features are included.

2.6 Data Transformation

Data transformation involves converting raw data into a more suitable format for analysis. In this research, Term Frequency-Inverse Document Frequency (TF-IDF) is used to convert text data into numerical features for machine learning algorithms.

2.7 Data Mining

Data mining is a critical step in knowledge discovery. This research employs the Naive Bayes algorithm to build a predictive model for damaged goods during transportation. The model development and optimization process ensures accurate and reliable results.

2.8 Model Evaluation

Model performance evaluation is conducted using cross-validation to measure accuracy in a multinomial distribution context. Cross-validation is a common technique in machine learning to objectively assess model performance and prevent overfitting or underfitting.

2.9 Research Approach

This research adopts a mixed-methods approach, combining both qualitative and quantitative methodologies. A case study approach is applied to analyze the handling of damaged goods at PT. Tuntas Smart Solusi. Additionally, the Knowledge Discovery in Databases (KDD) method and Naive Bayes algorithm are used to identify patterns and make predictions regarding damage handling during shipping.

2.10 Research Ethics

This research adheres to ethical principles, including:

- Informed Consent: Participants are informed about the research purpose and their rights before providing consent.
- Anonymity: No personal identifiers are recorded.
- Confidentiality: All collected data remains confidential and is used solely for research purposes.

2.11 Research Limitations

This research has several limitations, including sample size constraints and restricted access to some confidential company information. Additionally, the developed business process model may be projective and not fully represent actual conditions.

Through this research methodology, we aim to gain a deep understanding of damaged goods handling in land transportation services at PT. Tuntas Smart Solusi and provide recommendations for improving business process efficiency.

3. RESULTS AND DISCUSSION

Data Collection The objective of this data collection stage was to obtain a reliable dataset for further analysis to identify the causes and patterns of cargo damage during shipments by PT. Tuntas Smart Solusi. A total of 200 shipment records from PT. Tuntas Smart Solusi in Gresik, East Java, were collected for the period 2021–2023. This dataset was utilized to develop a predictive model aimed at reducing cargo damage rates and improving operational efficiency.

Preprocessing After data collection, preprocessing was conducted, involving two primary steps: removing irrelevant features and handling missing values. These steps were essential to ensure that the data used was relevant and free from issues that could affect the accuracy of the analysis.

1. Feature Removal
   - Irrelevant features such as date, item details, and delivery duration were excluded from the dataset.
   - The processed data was stored in an Excel DataFrame.
2. Missing Value Analysis
   - A missing value check was conducted to ensure data completeness.
   - The analysis showed that there were no missing values in any columns of the dataset.

```
Jumlah Barang yang Di Pick Up    0
Jumlah Barang yang Rusak         0
Ekspedisi                        0
Cuaca                            0
KelebihanMuatan                  0
```

Figure 1. Missing values checking results

Cargo Damage Analysis Prior to model implementation, an analysis was conducted to identify key factors contributing to cargo damage. Three primary variables were identified: expedition service, weather conditions, and load capacity. The damage ratio was calculated to provide insights into cargo damage risks.

1. Expedition Services
   - The highest damage ratio was recorded for TNC (2.37%), followed by Ngalimun (2.30%), Halim (2.27%), and ACMI (2.18%).

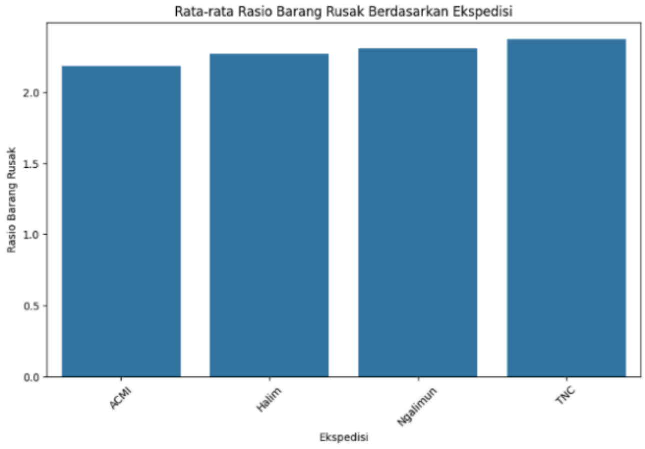

Figure 2. Graph of the average ratio of damaged goods based on expedition

2. Weather Conditions
   - The lowest damage ratio was recorded during clear weather (2.19%), while drizzling conditions showed a higher damage ratio (2.5%).
   - Rainy conditions had a slightly lower damage ratio (2.10%) than drizzling but were still higher than clear weather.

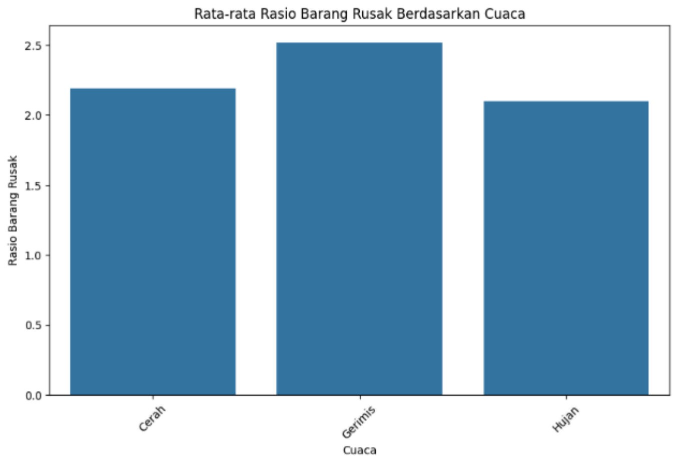

Figure 3. Graph of the average ratio of damaged goods based on weather

3. Load Capacity
   - Overloaded shipments recorded the highest damage ratio (2.35%), followed by underloaded shipments (2.32%) and balanced loads (2.09%).

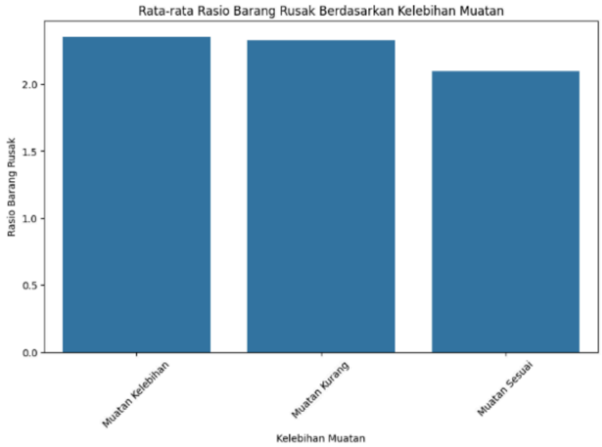

Figure 4. Graph of the average ratio of damaged goods based on weather

Model Development A Naïve Bayes model was implemented using historical shipment data. The dataset was transformed using Term Frequency-Inverse Document Frequency (TF-IDF) before training the model.

Model Optimization Using Cross-Validation To ensure robust model performance, k-fold cross-validation was applied with k=3, 5, and 10. The results were as follows:

- K-Fold 3: Training accuracy = 72.77%, Testing accuracy = 72.93%
- K-Fold 5: Training accuracy = 67.37%, Testing accuracy = 66.50%
- K-Fold 10: Training accuracy = 65.61%, Testing accuracy = 63.50%

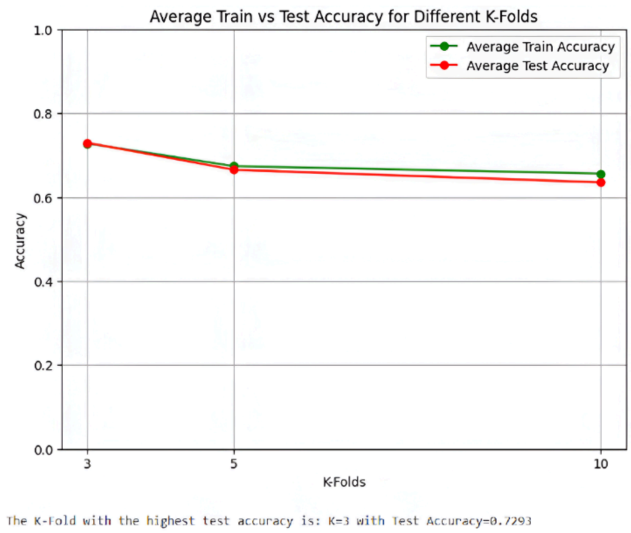

Figure 5. Graph of average model accuracy

Confusion Matrix Analysis

- K-Fold 3: The model correctly identified 146 damaged cargo cases, but 50 cases were misclassified.

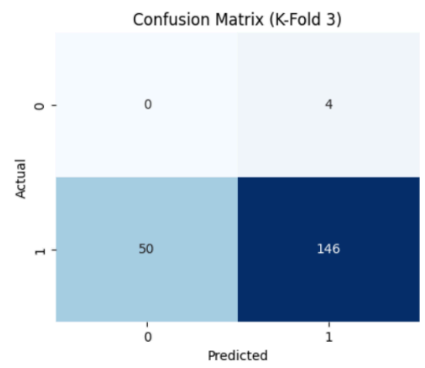

Figure 6. Confusion Matrix K-Fold 3

- K-Fold 5: The model detected 131 damaged cargo cases correctly, with 65 misclassified cases.

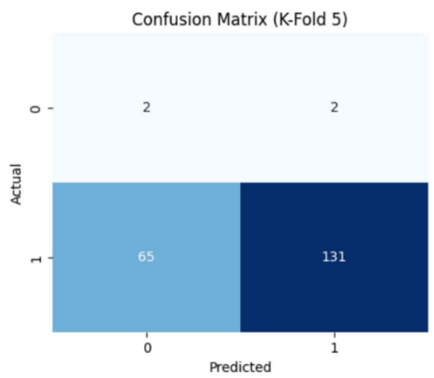

Figure 7. Confusion Matrix K-Fold 5

- K-Fold 10: The model had 125 correct predictions but misclassified 71 cases.

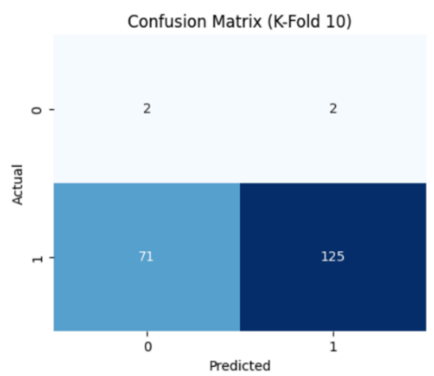

Figure 8. Confusion Matrix K-Fold 10

Evaluation of Gaussian and Multinomial Naïve Bayes Models

- Gaussian Naïve Bayes showed an overall stable performance, but its accuracy slightly declined as k increased.

ISSN: 2774-3993

- Multinomial Naïve Bayes achieved the highest accuracy at K=3 (91.5%) and declined slightly at K=5 and K=10.

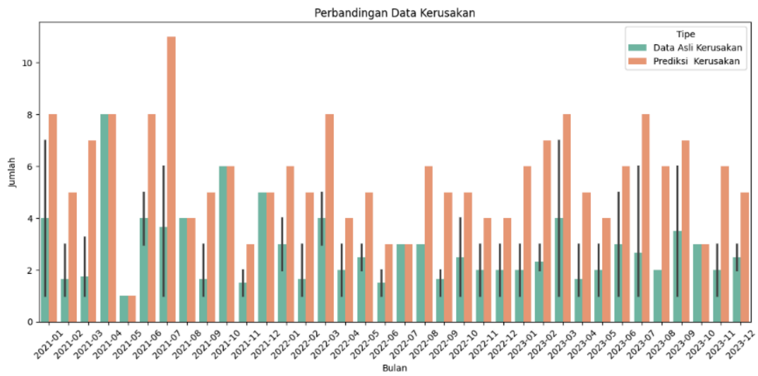

Figure 9. Comparison of Damage Data Prediction Results

Implementation of Predictive System A web-based system was developed for cargo damage prediction using Naïve Bayes classification.

- Multinomial Naïve Bayes System

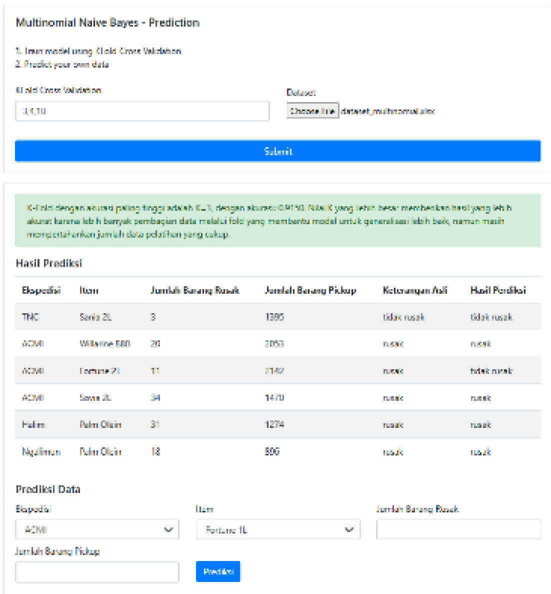

Figure 10. Multinomial Naïve Bayes System View

- Gaussian Naïve Bayes System

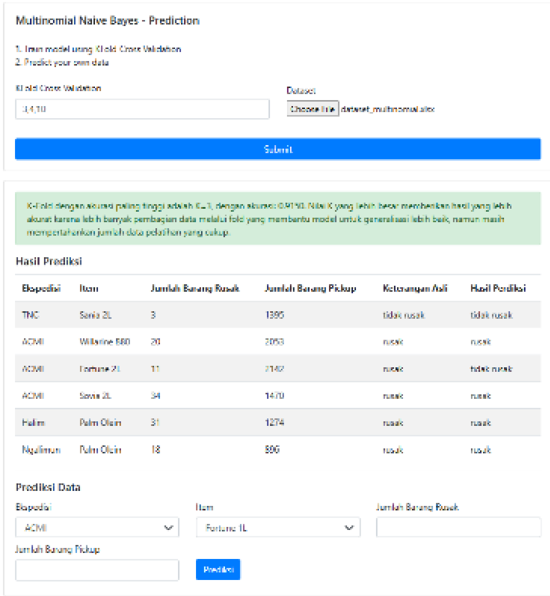

Figure 11. Gaussians Naïve Bayes System View

This system allows users to input shipment data, process predictions, and instantly display the probability of cargo damage. It serves as a practical tool for logistics decision-making and operational improvement.

CONCLUSION

This study has analyzed factors affecting product damage during shipping using trucking services and has applied the Naïve Bayes method for predictive modeling. The research identified key variables that contribute to product damage, including the type of goods, packaging methods, handling procedures, travel conditions, and transportation vehicles.

The implementation of the Naïve Bayes algorithm demonstrated its effectiveness in classifying and predicting damage probabilities with a reasonable level of accuracy. By utilizing historical shipment data, this model provides an early warning system for potential damage, allowing companies to take proactive measures. The cross-validation approach further confirmed the model's reliability, with varying performance levels across different K-Fold configurations.

Based on the findings, several improvements can be recommended, including enhanced packaging procedures, optimized shipping routes, staff training on proper handling, and advanced monitoring systems to detect potential damage early. Strengthening collaboration with vendors for damage claims and resolutions is also crucial to ensuring efficient logistics operations.

Note: Ensure that all relevant figures, such as missing value analysis, correlation matrices, and accuracy comparisons, are included in the article for better data visualization.

ACKNOWLEDGEMENTS

The author expresses sincere gratitude to PT. Tuntas Smart Solusi for providing valuable shipping data, which was crucial for this study. Their support in sharing operational insights greatly enhanced the research's depth and reliability.

Special thanks to Universitas Negeri Surabaya (UNESA), particularly the Faculty of Engineering, Department of Informatics Engineering, and Information Systems Program, for their academic guidance and research facilities. The support from faculty members and advisors was invaluable in refining the methodology and analysis.

The author is especially grateful to I Kadek Dwi Nuryana, S.T., M.Kom., for his expert guidance, constructive feedback, and mentorship. Appreciation is also extended to peer reviewers, fellow researchers, and colleagues for their insightful critiques and academic discussions that enriched this study.

Heartfelt thanks go to the author’s family and friends for their unwavering support and encouragement. Lastly, gratitude is extended to the broader academic and professional community in predictive analytics, logistics management, and AI, whose research laid the foundation for this study.

This research was made possible by the collective efforts of all contributors, and the author hopes it will benefit logistics efficiency, cargo safety, and predictive analytics in transportation.

REFERENCES

[1] Afandi, S., Hidayat, A., & Zulkifli, A. (2022). "Penerapan Metode Naive Bayes untuk Klasifikasi Sentimen pada Layanan Pengiriman." *Jurnal Informatika*, 8(2), 155-165.

[2] Adzy, L. B., & Pambudi, A. (2023). "Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi." *Jurnal Sistem Informasi*, 14(3), 210-222.

[3] Wicaksono, D. F., Basuki, R. S., & Setiawan, D. (2024). "Peningkatan Performa Model Machine Learning XGBoost Classifier melalui Teknik Oversampling dalam Prediksi Penyakit AIDS." *Jurnal Ilmu Komputer*, 15(2), 98-110.

[4] Sagala, G. J., Timur, Y., & Samuel, S. (2024). "Sentiment Analysis on ChatGPT App Reviews on Google PlayStore Using Random Forest Algorithm, Support Vector Machine, and Naïve Bayes." *Jurnal Kecerdasan Buatan*, 13(1), 112-125.

[5] Romli, I. M., & Kurniawan, E. (2022). "Klasifikasi Kerusakan Barang dengan Menggunakan Komparasi Algoritma C4.5 dan Naive Bayes." *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 11(4), 200-213.

[6] Putry, N. M., & Sari, B. N. (2022). "Komparasi Algoritma KNN dan Naïve Bayes untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus." *Jurnal Teknik Informatika*, 17(3), 144-157.

[7] Nashiruddin, A. Z., & Hidayat, R. (2022). "Perancangan Sistem Pakar dalam Identifikasi Kerusakan Mesin Cuci Berbasis Naïve Bayes." *Jurnal Ilmu Komputer dan Informasi*, 8(1), 76-89.

[8] Nurjanah, A., Aulia, N., & Rasywir, E. (2022). "Analisis Tingkat Kepuasan Konsumen terhadap Layanan Ekspedisi J&T Express dengan Metode Algoritma Naive Bayes." *Jurnal Manajemen dan Bisnis*, 19(2), 233-245.

[9] Gantini, T., & Wijaya, Y. K. (2021). "Rekayasa Proses Bisnis Layanan Pengiriman Paket." *Jurnal Teknik Industri*, 9(4), 121-135.

[10] Maulana, A., & Purnamasari, I. A. (2024). "Analisis Klasifikasi Indeks Kualitas Udara Kota di Indonesia Menggunakan Metode K-Nearest Neighbor dan Naïve Bayes." *Jurnal Pengolahan Data dan Informatika*, 12(2), 88-101.

[11] Fitika Andraini, T. H. K. (2019). Tanggung Jawab Atas Keterlambatan, Kerusakan Atau Kehilangan Pengiriman Barang Di Pt. Pandu Logistics. *Jurnal Ilmiah Dinamika Hukum*, *20*(1), 40–50. https://doi.org/10.35315/dh.v22i1.8260

[12] Hamzar, R. A., & Setyaningsih, F. A. (2023). Dampak Ukuran Dataset Pelatihan Terhadap Performa LSTM Network dalam Konteks Harga Saham. *Coding Jurnal Komputer Dan Aplikasi*, *11*(2), 238. https://doi.org/10.26418/coding.v11i2.62142

[13] Permata, A. R. (2023). Analisis Data Penelitian Kesehatan: Perbandingan Hasil antara SmartPLS, R dan IBM SPSS Health Research Data Analysis: Comparison of Results between SmartPLS, R and IBM SPSS. *Jurnal Sains Natural*, *1*(1), 17–22.