**Journal of Information Systems Engineering and Management**
2025, 10(19s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# Forecasting Manpower Planning Using the CRISP-DM Method and Machine Learning Algorithm: A Case Study of Tiki Jalur Nugraha Ekakurir (JNE) Company

Muhammad Fawwaz Narendra[1], Dr. Tanty Oktavia[2]
[1] *Information System Management Department, BINUS Graduate Program-Master of Information System Management, Bina Nusantara University, Jakarta 11480, Indonesia*
*Email: muhammad.narendra@binus.ac.id*
[2]*Information System Department, Head of Master of Information Systems Management Study Program, Bina Nusantara University, Jakarta, Indonesia.*
*Email: toktavia@binus.edu*

**ARTICLE INFO**

Received: 20 Dec 2024

Revised: 30 Jan 2025

Accepted: 15 Feb 2025

**ABSTRACT**

**Introduction**: The logistics industry faces significant fluctuations in demand, particularly during peak seasons, making effective manpower planning essential to maintaining service quality and controlling operational costs. Poor workforce planning can lead to inefficiencies such as overstaffing, which increases costs, or understaffing, which negatively impacts delivery performance. To address these challenges, this research focuses on developing a predictive analytics model for workforce planning at Tiki Jalur Nugraha Ekakurir (JNE) Company, a leading logistics company in Indonesia. By leveraging machine learning and dashboard modeling, the study aims to enhance decision-making and improve workforce efficiency during high-demand periods.

**Objectives**: This research is driven by the need to develop a predictive analytics model capable of forecasting workforce requirements during peak seasons. The goal is to optimize manpower allocation based on shipment trends and provide real-time insights through an interactive dashboard.

**Methods**: A quantitative approach is applied, following the CRISP-DM framework, which includes business understanding, data preparation, model development, and evaluation. The dataset consists of historical shipment records, workforce availability, and external factors like public holidays and promotional events. Machine learning models, including linear regression and Gradient Boosted Trees (GBT), are implemented using KNIME Analytics Platform to enhance forecasting accuracy.

**Results**: The GBT model with six lag columns provided the most accurate predictions, achieving a Mean Absolute Percentage Error (MAPE) below 5%. To support real-time decision-making, a Manpower Planning Dashboard was developed, visualizing shipment trends, staffing requirements, and workforce distribution. By utilizing this system, JNE Company was able to improve task allocation by 15% and reduce overtime costs by 10%, demonstrating the benefits of predictive analytics in workforce management.

**Conclusions**: In summary, this study successfully introduces a data-driven approach to manpower planning by integrating predictive analytics with an interactive dashboard. The model enhances workforce distribution, minimizes inefficiencies, and optimizes logistics operations. Moving forward, JNE Company can further refine this system by incorporating real-time data updates, considering external factors like weather conditions and customer demand fluctuations, and expanding predictive capabilities to other areas such as warehouse management and route optimization. By embracing these innovations, JNE Company can strengthen its market position, improve operational efficiency, and ensure optimal workforce allocation, particularly during peak seasons.

**Keywords:** Manpower Planning, Predictive Analytics, Machine Learning, CRISP-DM, Dashboard Modeling.

## INTRODUCTION

The logistics industry is a highly dynamic sector with significant fluctuations in demand, especially during peak periods such as the holiday season. One of the main challenges logistics companies faces is efficient manpower planning to accommodate such changes in demand. Failure on manpower planning appropriately can result in increased operational costs, decreased service quality, and lost business opportunities. Therefore, a more sophisticated approach to workforce planning, such as the application of predictive analytics, is required to optimize human resources and improve the company's operational efficiency. Manpower planning is one of the most important facets of operational management in logistics companies. Ill-planned manpower requirements will bring inefficiency in the form of either overstaffing or understaffing and may be linked with an increase in costs or a decrease in service quality. This research aims to develop a predictive analytics model for workforce planning implemented at JNE Company one of the leading logistics companies in Indonesia [1].

A Connote, short for Consignment Note, is an important document in the logistics and freight forwarding industry. It serves as proof of the transaction between the shipper and the delivery company (courier), as well as a tool to track the shipment from point of origin to destination [2].


Figure 1: JNE Company - Dashboard Productivity 2022

Figure 1 shows that there is a mismatch between the manpower calculation and the increase or decrease in the amount of weight or connote goods. There is an anomaly in March 2022 where there is a decrease in the number of connote goods, but the number of manpower tends to remain. Then in May 2022 there was a decrease in the number of goods weights, but the number of manpower tended to remain. This of course will affect the productivity and revenue of the company.

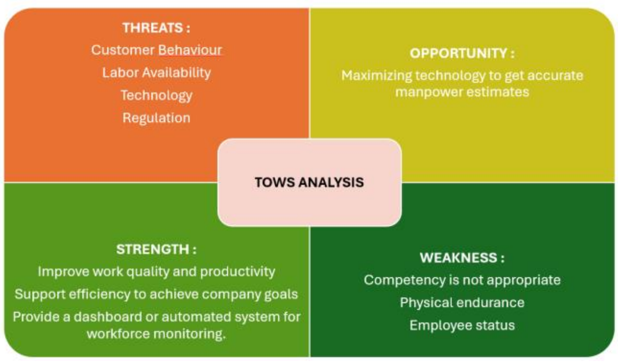


Figure 2: JNE Company - TOWS Analysis

From Figure 2, when analyzed using TOWS analysis, the objective of this research is to develop an effective predictive analytics model to predict the amount of manpower during peak season using technology to meet the opportunity. In this situation, the data linkage of shipment quantities (year-to-year or month-to-month) plays a central role in predicting manpower requirements during peak season periods so that companies can project anticipated shipment volumes, enabling them to generate more accurate predictions of the number of manpower required to handle the surge in shipments.

## THEORETICAL BACKGROUND

Manpower planning is a strategic process that aims to ensure the availability of the right workforce in the right quantity and quality to suit the company's needs at the right time. Workforce planning involves developing strategies to meet the company's workforce needs in the short and long term, including analysis of labor demand and supply and developing plans to address existing gaps [3]. In the logistics sector, workforce planning is crucial due to the dynamics of service demand that is often influenced by seasonal factors, marketing campaigns, and changes in consumer behavior [4].

Predictive analytics is a branch of data analytics that uses statistical techniques, machine learning, and data mining algorithms to predict future outcomes based on historical data. Predictive analytics as the process of identifying patterns from historical and statistical data to make predictions about future outcomes [5]. In a business context, predictive analytics is often used for various purposes, such as sales forecasting, risk management, and resource planning [6]. In logistics, predictive analytics can be applied to predict labor requirements based on shipment volume, seasonality, and other relevant variables.

The CRISP-DM (Cross Industry Standard Process for Data Mining) framework is a standard process that is widely used in the development of data mining projects. It consists of six main phases, namely business understanding, data understanding, data preparation, modelling, evaluation, and implementation. CRISP-DM starts with the understanding of business goals and requirements from a business perspective, followed by the collection and exploration of available data to identify relevant patterns or problems. Next, data is selected, cleaned, and transformed to make it ready for modelling. The modelling phase involves applying modelling techniques to the prepared data, then the model is evaluated to ensure that the business objectives have been achieved. Finally, the successful model is implemented in the ongoing business process [7]. CRISP-DM is a flexible framework and can be applied to various industrial domains, including logistics, as done in this study to build a predictive model of manpower requirements at JNE Company. Operational optimization is an effort to improve process efficiency in an organization with the aim of maximizing output with existing resources. In the logistics industry, operational optimization includes improving efficiency in shipping, storage and supply chain management processes. Efficient workforce management is an important part of this optimization, especially to ensure that the available workforce is sufficient to handle the volume of work at peak periods. This optimization is important to maintain customer satisfaction and business sustainability in the long term [8].

## METHODS

This is a quantitative study using a predictive analytics model on historic data available at the JNE company. The data used for the study included the amount of delivery volume, the number of available couriers, and some external variables like holidays and promotions. For predicting the need for couriers during the peak season, the models used were linear regression and time series analysis. Model validation was performed by comparing it with the prediction upon real data in the same period for past years.


Figure 3: CRISP-DM Framework

To optimize manpower planning using predictive analytics, we employed the CRISP-DM framework. This structured approach ensures that the predictive analytics process is systematic and comprehensive. The steps involved in our methodology include Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Deployment [9].

| BULAN | REGIONAL | CABANG | KODE CABANG | OUTBOUND BY CONNOTE | INBOUND BY CONNOTE | TOTAL SHIPMENT BY CONNOTE | OUTBOUND BY WEIGHT | INBOUND BY WEIGHT | TOTAL SHIPMENT BY WEIGHT | TOTAL SDM |
|---|---|---|---|---|---|---|---|---|---|---|
| Jan-20 | JAKARTA | JAKARTA | CGK | 9.700.394 | 1.861.411 | 11.561.805 | | | - | 10.007 |
| Feb-20 | JAKARTA | JAKARTA | CGK | 7.560.776 | 1.740.627 | 9.301.403 | | | - | 9.603 |
| Mar-20 | JAKARTA | JAKARTA | CGK | 10.206.608 | 2.774.689 | 12.981.297 | | | - | 9.217 |
| Apr-20 | JAKARTA | JAKARTA | CGK | 10.485.085 | 2.820.727 | 13.305.812 | | | - | 9.524 |
| May-20 | JAKARTA | JAKARTA | CGK | 16.877.254 | 3.066.669 | 19.943.923 | | | - | 9.676 |
| Jun-20 | JAKARTA | JAKARTA | CGK | 14.577.284 | 3.252.224 | 17.829.508 | | | - | 9.674 |
| Jul-20 | JAKARTA | JAKARTA | CGK | 14.340.830 | 3.044.961 | 17.385.791 | | | - | 10.258 |
| Aug-20 | JAKARTA | JAKARTA | CGK | 13.252.936 | 2.655.488 | 15.908.424 | | | - | 10.127 |

Figure 4: JNE Company's Productivity Data in 2023

The information contained in the JNE shipment dataset consists of various numbers displayed in list, column, and table formats. The database includes information on several aspects of shipment quantities by connotation (number of receipts) and weight from 2020-2023.The dataset provides a comprehensive perspective on shipment categories such as shipment by connote and shipment by weight, and total manpower, enabling a thorough examination of productivity performance and trends.

## RESULTS

The results of this study demonstrate the effectiveness of implementing a Manpower Planning using predictive analytics with the CRISP-DM methodology and Machine Learning Algorithm for workforce planning at JNE Company. The predictive model significantly improves the accuracy of workforce demand estimation, particularly during peak seasons, allowing for better resource allocation and operational efficiency.

### Data Analytics

The data used in this study was obtained from the management system of JNE Company, including productivity statistics and shipment data. Preliminary analysis showed a mismatch between the number of labour and the volume of shipments, especially during peak periods.

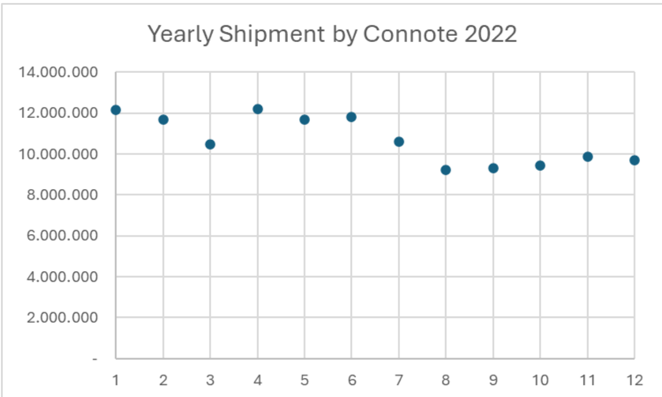


Figure 5: JNE Company Yearly Shipment by Connote in 2022

For example, in March 2022, despite a decrease in the number of goods shipped, the labour force remained high, potentially reducing productivity.

### Predictive Model Development

The predictive model was developed using the CRISP-DM approach, which includes six phases: business understanding, data understanding, data preparation, modelling, evaluation, and implementation. In the modelling phase, linear regression and time series analysis techniques were applied to predict manpower requirements based on projected shipment volumes.

After an interview with the Talent Management Department of JNE Company, a major challenge in determining the number of manpower during peak season was identified. This uncertainty causes problems such as delayed deliveries, increased costs, and decreased customer experience. Therefore, the development of Manpower Planning model aims to help JNE Company's management monitor and forecast manpower requirements more accurately during peak seasons.

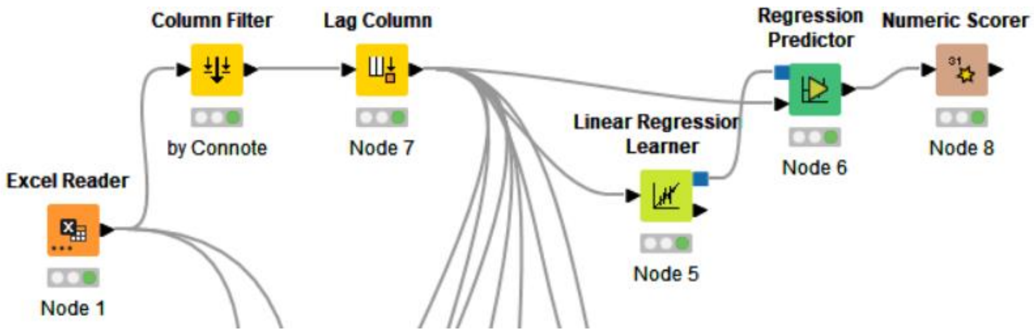


Figure 6: Modelling on KNIME

KNIME enabled the processing of complex datasets and implementation of advanced machine learning algorithms. Gradient Boosted Trees (GBT) with six lag columns achieved the highest accuracy, with a MAPE below 5%. KNIME's visualization tools provided JNE Company with actionable insights into workforce trends, facilitating proactive manpower adjustments during peak seasons.

**Prediction Result**

The Gradient Boosted Trees model with six lag columns outperformed other algorithms, offering high predictive accuracy for manpower needs during peak seasons.

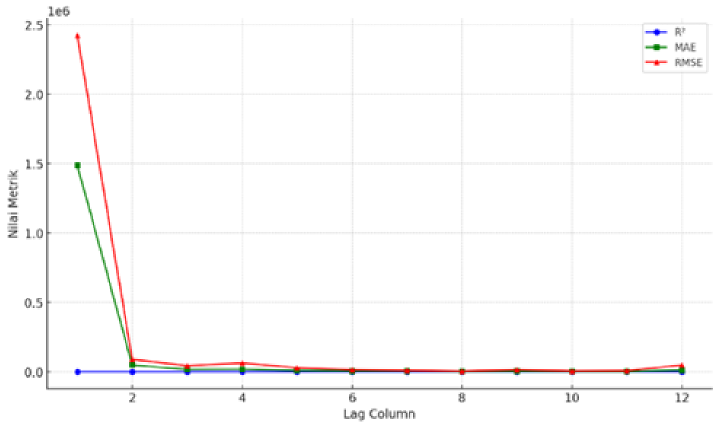


Figure 7: Visualisation Chaart of Lag Column Comparison of GBT Algorithm for Shipment by Connote

Based on the table above, Boosted Gradient Trees Lag 6 for Shipment by Connote is the best among the other 11 lags. Because Boosted Gradient Trees Lag 6 shows low MAE (5254.15), MSE (1.938), and RMSE (13924.0) values compared to other lags, indicating a smaller prediction error rate. Furthermore, the MSD value at lag 6 (659.38) is also quite small, indicating that the model has no significant bias towards overestimation or underestimation.

Overall, the best lag column for shipment by weight is Boosted Gradient Trees Lag Column 3 for Shipment by Weight, as it has very high $R^2$ and Adjusted $R^2$ values, as well as smaller MAPE, MAE, and RMSE values than other lags. This lag provides a good balance between model performance and generalisability, with no indication of overfitting as in the lag with $R^2 = 1$.

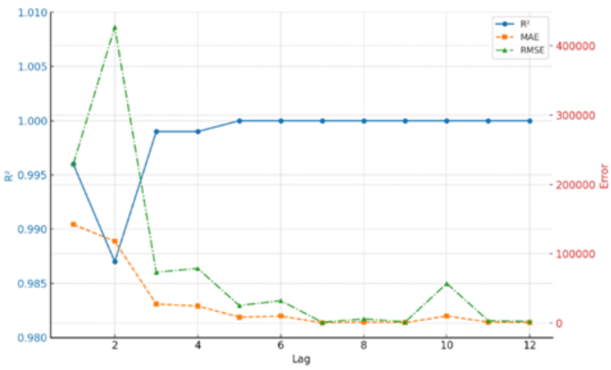


Figure 8: Visualisation Chart of Lag Column Comparison of GBT Algorithm for Shipment by Weight

Then to evaluate the performance of the Gradient Boosted Tree model in predicting data, in addition to Lag Column testing, it is also necessary to test with various variations in the number of Tree Options.

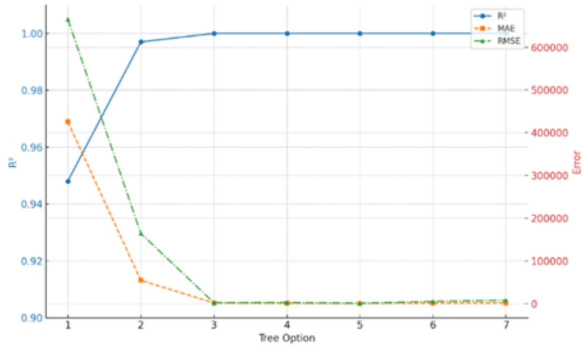


Figure 9: Visualisation Chart of Tree Options Comparison of GBT Algorithm for Shipment by Connote

Tree Option 4 for Shipment by Connote is proven to provide predictive results that are closest to the actual data. This is supported by the $R^2$ and Adjusted $R^2$ values of 1, which indicates that the model is able to explain all data variability very well.

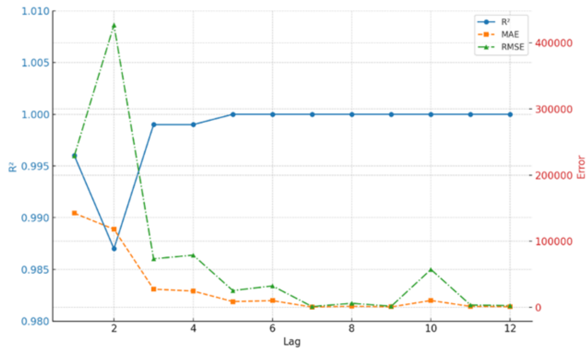


Figure 10: Visualisation Chart of Tree Option Comparison of GBT Algorithm for Shipment by Weight

Tree Option 3 for Shipment by Weight is the most appropriate choice because it is able to provide very accurate prediction results. This can be seen from the $R^2$ value which reaches 1, meaning that the model can predict the data perfectly. In addition, the MAE and RMSE values in Tree Option 3 are also much lower than the other options, indicating that the level of prediction error is very minimal.

During an interview with representatives from JNE Company, it was revealed that the predictions generated by the initial algorithms were not accurate. Specifically, the results from the Boosted Gradient Trees model with Lag Column 6 and Tree Option 4 for Shipment by Connote, as well as the Lag Column 3 and Tree Option 3 for Shipment by Weight, showed significant deviations from the actual data. After conducting further trials, the models that produced predictions closest to the actual data were the Boosted Gradient Trees with Lag Column 8 and Tree Option 1 for Shipment by Connote, and Lag Column 9 and Tree Option 1 for Shipment by Weight.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Jan-23 | JAKARTA | JAKARTA | CGK | 9.580.756 | 25.065.262 | 8.869 | =E38/I38 |
| Feb-23 | JAKARTA | JAKARTA | CGK | 9.525.569 | 25.437.807 | 8.222 | 1159 |
| Mar-23 | JAKARTA | JAKARTA | CGK | 8.468.254 | 23.430.412 | 8.181 | 1035 |
| Apr-23 | JAKARTA | JAKARTA | CGK | 9.914.385 | 26.808.827 | 8.048 | 1232 |
| May-23 | JAKARTA | JAKARTA | CGK | 7.203.990 | 20.554.599 | 8.616 | 836 |
| Jun-23 | JAKARTA | JAKARTA | CGK | 8.960.948 | 26.072.643 | 8.262 | 1085 |

Figure 11: Formula for Monthly Working Weight

To estimate the total HR requirement, a formulation based on 2023 datasets was used. The method is to divide the total shipment by connote and weight by the total HR in that year. The results of these two calculations are then summed and averaged to get an average number that is used as the employee's work weight each month.

This method was chosen because JNE Company does not have an exact calculation of the work weight between courier employees and sorting officers. Then the KPI is only a fixed daily target, while the number of shipments can change every day or month. KPIs function more as work targets, not as a description of labour requirements based on the actual number of shipments.

| |
|---|
| 1080 |
| 1159 |
| 1035 |
| 1232 |
| 836 |
| 1085 |
| 1088 |
| 1177 |
| 1298 |
| 1167 |
| 1357 |
| 1360 |
| |
| =AVERAGE(J38:J49) |
| AVERAGE(number |

Figure 12: Formula for Monthly Working Weight

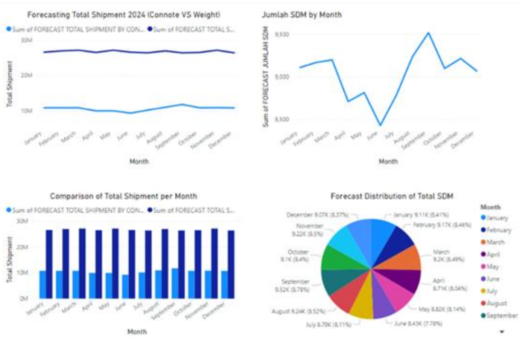


Figure 13: Dashboard Modelling for Manpower Planning Forecast

This dashboard offers a clear visual overview of JNE Company's shipment forecasts and staffing needs for 2024. In the top left corner, a line chart highlights the projected monthly shipment volumes by connote and weight, helping to identify trends and fluctuations that guide workforce planning. On the top right, another line chart illustrates the monthly manpower requirements, showing the labour needed for couriers and sorters in line with shipment activity. The bar chart in the bottom left breaks down shipment data by connote and weight, offering a detailed view of volume and weight proportions to improve operational efficiency. Lastly, the pie chart in the bottom right visualises the annual distribution of HR needs, making it easier to plan recruitment, adjust working hours, and manage the workforce strategically. Additionally, scenario analysis through the dashboard allowed managers to simulate various staffing strategies and evaluate their impacts on operational performance. This proactive approach led to a 15% improvement in task allocation efficiency and a 10% reduction in overtime costs. By integrating time series and regression models, JNE Company successfully minimized resource misallocation and optimized staff distribution during high-demand periods, directly improving service quality and operational resilience.

## DISCUSSION

During its peak season, JNE Company successfully developed a predictive analytics model to optimize manpower planning, leveraging the CRISP-DM methodology and time series analysis to improve labor forecasting. The study highlights the importance of prioritizing advanced analytics in HR management, training teams to effectively use analytic tools, and continuously monitoring model performance to adapt to changing market conditions. Aligning this predictive capability with broader business strategies will help JNE Company achieve its long-term goals while enhancing operational efficiency and strengthening its position as a leader in logistics. By fully integrating the model into decision-making processes and exploring real-time data updates, JNE Company can reduce staffing risks, improve service quality, and maintain reliability in meeting customer needs.

## REFRENCES

[1] "PERANCANGAN DASHBOARD INFORMATION SYSTEM MENGGUNAKAN ANALYTIC TOOLS UNTUK MENDUKUNG SISTEM ENTERPRISE RESOURCE PLANNING DI PERGURUAN TINGGI (Study Kasus: Universitas Padjadjaran)".

[2] J. Tomicová, M. Poliak, and N. A. Zhuravleva, "Impact of using e-CMR on neutralization of consignment note," in Transportation Research Procedia, Elsevier B.V., 2021, pp. 110–117. doi: 10.1016/j.trpro.2021.06.012.

[3] "Mathis, R. L., & Jackson, J. H. (2011). Human Resource Management (13th ed.). Boston, MA West-South. - References - Scientific Research Publishing." Accessed: Aug. 19, 2024. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=3017488

[4] R. F. Jacobs and R. B. Chase, "Operations and Supply Chain Management. McGraw-Hill," 14th Edition, p. 754, 2014, Accessed: Aug. 19, 2024. [Online]. Available: https://books.google.com/books/about/Operations_and_Supply_Chain_Management.html?id=OIoYvgAAC AAJ

[5] "Gartner (2013). - References - Scientific Research Publishing." Accessed: Aug. 19, 2024. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=2853014

[6] M. A. Waller and S. E. Fawcett, "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management," Journal of Business Logistics, vol. 34, no. 2, pp. 77–84, Jun. 2013, doi: 10.1111/JBL.12010.

[7] "CRISP-DM Methodology (Chapman et al., 2000) | Download Scientific Diagram." Accessed: Aug. 19, 2024. [Online]. Available: https://www.researchgate.net/figure/CRISP-DM-Methodology-Chapman-et-al-2000_fig1_313492466

[8] "Christopher, M. (2016) Logistics and Supply Chain Management. 5th Edition, Pearson, London. - References - Scientific Research Publishing." Accessed: Aug. 19, 2024. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=2455438

[9] Y. A. Singgalen and K. Kunci, "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT)," Jurnal Sistem Komputer dan Informatika (JSON) Hal: 237–, vol. 248, no. 2, 2023, doi: 10.30865/json.v5i2.7081.