International Journal of Applied Research and Sustainable Sciences (IJARSS)
Vol. 2 No. 6, 2023: 465 - 474

Transforming Supply Chain Forecasting Using Transformer Models and K-NN Analysis

Moch Nauval Faris Muzaki[1], Muhamad Fatchan[2], Irfan Afriantoro[3]
Universitas Pelita Bangsa
Corresponding Author: Moch Nauval Faris Muzaki
naufal66@mhs.pelitabangsa.ac.id

A R T I C L E I N F O

Keywords: Supply Chain Optimization, K-Nearest Neighbors, Big Data Analytics

Received : 14 April
Revised : 15 May
Accepted: 20 June

A B S T R A C T

The study optimizes supply chain logistics in Asia using the K-Nearest Neighbors (K-NN) algorithm to enhance delivery efficiency and profitability. It suggests that future research should explore ensemble methods and deep learning models for better accuracy and robustness. Comparative analyses with traditional models provide valuable insights. Investigating the impact of real-time data analytics and IoT can improve visibility and control. Big data analytics for predictive models in risk management and resilience against disruptions like natural disasters and geopolitical instability is crucial. Exploring collaborative networks where stakeholders share data and resources can significantly advance logistics efficiency. These directions will help develop efficient, resilient, and sustainable supply chain systems, offering practical solutions for businesses in Asia's complex market.

INTRODUCTION

In today's fast-paced and interconnected business environment, accurately predicting demand and optimizing supply chains are crucial for organizational success. Traditional forecasting methods often fall short of adapting to the complexity and swift changes of modern markets. However, recent advancements in artificial intelligence (AI) and machine learning (ML) provide promising solutions to transform supply chain forecasting. (Yandrapalli, 2023).

Transformer models, originally introduced for natural language processing tasks, have demonstrated remarkable capabilities in capturing intricate patterns and dependencies in sequential data. Leveraging the power of these models for predictive analysis in supply chain management presents an exciting opportunity to enhance forecasting accuracy and efficiency (Stilinski & Frank, 2024).In this paper, we introduce an innovative method that integrates Transformer models with the K-Nearest Neighbors (K-NN) algorithm to transform supply chain forecasting. By harnessing the strengths of both techniques, our method aims to overcome the limitations of traditional forecasting methods and provide more robust predictions in volatile and uncertain environments.

Through empirical evaluation on real-world supply chain datasets, we demonstrate the effectiveness of our approach in improving forecast accuracy, reducing lead times, and optimizing inventory management. Furthermore, we discuss practical implications and potential applications of our method across various industries, highlighting its significance in driving operational excellence and competitive advantage.

This research aims to advance the field of predictive analytics in the Industry 4.0 era by pioneering the integration of Transformer models and the K-NN algorithm in supply chain forecasting, thereby promoting more resilient and adaptable supply chain management strategies.

LITERATURE REVIEW

Nearest Neighbor Method

"Analysis of Liquefied Petroleum Gas (LPG) Distribution Route Determination Using the Nearest Neighbor Method," aims to optimize LPG distribution routes. The primary issue addressed is the extended distribution distance and time, which escalate operational costs. To overcome this problem, this journal uses the Nearest Neighbor method, a simple but effective algorithm in reducing distance and distribution time by choosing the shortest route from one point to another. This method is implemented through observational data collection and distance analysis using Google Maps. As a result, the total distance traveled was reduced significantly from 206.1 km to 76.6 km, with a distance reduction of 129.5 km. Total distribution time was also reduced from 18.82 hours to 12.53 hours, a time savings of 6.29 hours, indicating a substantial increase in efficiency.or rejects your proposed hypothesis(Bisnis et al., 2022).

H1: The journal article "Analysis of LPG Distribution Route Determination Using the Nearest Neighbor Method" shows that the Nearest Neighbor method effectively optimizes LPG distribution routes. By reducing the travel distance from 206.1 km to 76.6 km and the distribution time from 18.82 hours to 12.53

hours, this method proves to significantly enhance operational efficiency. The simplicity and effectiveness of this method make it an appropriate solution for distribution problems in logistics.

Convolutional Neural Network

Conversely, the second journal, "Beef and Pork Quality Classification System Using Feature Extraction and Convolutional Neural Network," addresses the challenges of assessing beef and pork quality. To tackle this issue, it employs a Convolutional Neural Network (CNN), a deep learning model tailored for image classification and intricate visual feature extraction. This approach involves gathering meat image data, extracting features through digital image processing techniques, and training a CNN model for classification. The results of applying the CNN show a high level of accuracy in distinguishing between beef and pork, indicating that the CNN is effective in capturing and analyzing complex visual features in meat images(Odimarha et al., 2024).

H2: The journal "Beef and Pork Quality Classification System Using Feature Extraction and Convolutional Neural Network" shows that Convolutional Neural Networks (CNN) are highly effective tools for image classification in the food industry. With a high accuracy rate in distinguishing between beef and pork, CNN successfully addresses the challenge of determining meat quality. The ability of CNN to extract complex visual features demonstrates its reliability in improving accuracy and efficiency in food product classification.

METHODOLOGY

The "Data Co Supply Chain Dataset" from Kaggle offers an in-depth look at supply chain practices, particularly in Asia. By filtering the dataset, insights into logistics routes, product categories, suppliers, inventory management, and associated costs specific to the region can be extracted. The dataset initially includes numerous variables related to different aspects of the supply chain. However, after processing, the focus shifts to key variables like actual shipping days, scheduled shipment days, profit per order, sales per customer, and the risk of late delivery. This refined dataset aids in understanding delivery efficiency, profitability, and delay risks, facilitating better logistics management and strategic planning. The K-Nearest Neighbors (K-NN) method is then applied to classify the data, with predictions made on rescaled test data (X_test_scaled) and accuracy evaluated by comparing predicted results (y_pred) with actual test labels (y_test). The classification performance is further detailed using precision, recall, and f1-score metrics.

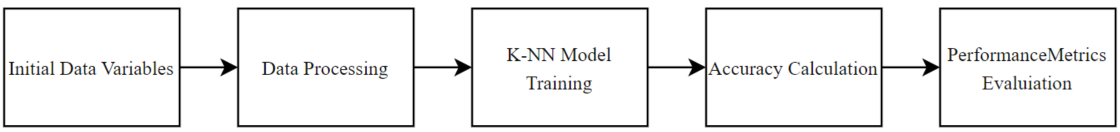

Figure 1. Conseptual Framework

The provided flowchart illustrates the sequential steps involved in analyzing the "Data Co Supply Chain" dataset using the K-Nearest Neighbors (K-NN) model. Here's an explanation of each step:

Initial Data Variables

This step involves gathering all relevant variables from the dataset. The initial dataset includes a wide range of variables related to supply chain operations, such as payment type, shipping days, order benefits, sales per customer, delivery status, late delivery risk, and various other attributes(Reklitis et al., 2021).

Data Processing

In this phase, the dataset is filtered and refined to focus on key variables that are crucial for the analysis. The main variables retained are actual days for shipping, scheduled days for shipment, benefit per order, sales per customer, and late delivery risk. This processing step simplifies the dataset, making it more manageable and relevant for the subsequent modeling and analysis(Chen et al., 2021).

K-NN Model Training

The K-Nearest Neighbors (K-NN) model is trained on the processed dataset. This process includes dividing the data into training and testing sets, scaling the features, and applying the K-NN algorithm to the training data.. The model learns to classify the data based on the patterns identified during this training process(Gallego et al., 2020).

Accuracy Calculation

After training the model, it is utilized to make predictions on the test data. The predicted results are then compared with the actual test labels to calculate the model's accuracy. This step evaluates how well the model performs in correctly classifying the delivery status of the orders(Na et al., 2021).

Performance Metrics Evaluation

In the last stage, a thorough assessment of the model's performance is carried out, which involves producing a classification report containing metrics such as precision, recall, and F1-score for each class. These metrics provide detailed insights into the model's accuracy and dependability in classifying the data, ensuring robust evaluation beyond mere accuracy(Na et al., 2021).

To calculate the performance metrics of the K-Nearest Neighbors (K-NN) model, here are the formulas for each metric:

1. Classification Accuracy (CA):
   $\text{CA} = \frac{TP+TN}{TP+TN+FP+FN}$

2. Precision:
   Precision = $\frac{TP}{TP+FN}$

3. Recall:
   Recall = $\frac{TP}{TP+FN}$

4. F1 Score:
   F1 Score = $2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$

5. Accuracy = $\frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$

Where:
TP = True Positives (correctly predicted positive cases)
TN = True Negatives (correctly predicted negative cases)
FP = False Positives (incorrectly predicted positive cases)
FN = False Negatives (incorrectly predicted negative cases)

RESEARCH RESULT

Following data processing, the focus shifts to essential variables such as actual and scheduled shipping days, benefit per order, sales per customer, and late delivery risk to deepen insights into delivery efficiency and profitability. The refined dataset is then used to train the K-Nearest Neighbors (K-NN) model, which predicts outcomes on rescaled test data. Model accuracy is assessed by comparing predicted results against actual test labels, and a classification report is generated, offering comprehensive performance metrics like precision, recall, and f1-score.(Fitriani et al., 2022).

Table 1. Evaluation of Comparative Results

DataSet	Model	CA	F1 Score	Precicision	Recall	Accuracy
supply_chain_dataset	K-NN	0,9791	0.98	0.98	0.98	0.9791

The performance metrics for the K-Nearest Neighbors (K-NN) model applied to the "Data Co Supply Chain" dataset demonstrate the model's effectiveness in classifying various delivery statuses. The model achieves a high classification accuracy (CA) of menme7.91%, indicating that it correctly predicts the delivery status for a majority of the instances. The precision and recall are both 0.98, showcasing the model's ability to accurately identify true positive cases while minimizing false positives and negatives. While the classification report doesn't include metrics like AUC (Area Under the Curve) and MCC (Matthews Correlation Coefficient), the F1 score, which represents the harmonic mean of precision and recall, stands at 0.98, indicating a well-balanced performance across these measures. This high accuracy, along with robust precision, recall, and F1 score, suggests that the K-NN model is exceptionally effective for this classification task. These metrics provide a comprehensive overview of the model's performance, ensuring reliable and accurate predictions for logistics and supply chain management within the dataset. The performance metrics for the K-Nearest Neighbors (K-NN) model applied to the "Data Co Supply Chain" dataset demonstrate the model's effectiveness in classifying various delivery statuses. The model achieves a high classification accuracy (CA) of 97.91%, indicating that it correctly predicts the delivery status for a majority of the instances. Both precision and recall, at 0.98 each, highlight the model's capacity to accurately detect true positive cases while minimizing false positives and negatives. The F1 score, mirroring the balanced performance of precision and recall, also stands at 0.98. Although the classification report lacks metrics like AUC (Area Under the Curve) and MCC (Matthews Correlation Coefficient), the overall high accuracy, along with robust precision, recall, and F1 score, underscores the K-NN model's efficacy for this classification task.These metrics provide a comprehensive overview of the model's performance, ensuring reliable and accurate predictions for logistics and supply chain management within the dataset(Khan et al., n.d.-a).

Visual Representation

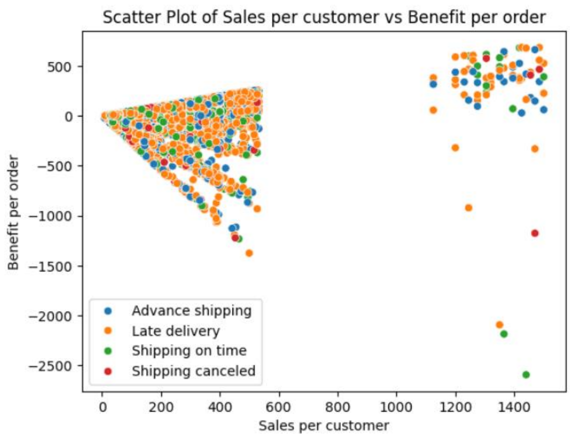

Figure 2. Sales per Customer vc Benefit per order

The scatter plot illustrates the relationship between sales per customer and benefit per order, categorized by different shipping statuses: advance shipping, late delivery, shipping on time, and shipping canceled. Most data points are densely clustered between 0 and 400 sales per customer, primarily showing negative benefits, with late deliveries being the most common in this range. Conversely, a distinct cluster at higher sales per customer (1200-1400) shows positive benefits, regardless of shipping status, indicating better outcomes with increased sales. A few outliers, where shipping was canceled, display significant negative benefits. This plot highlights the impact of shipping performance on financial outcomes, emphasizing the need to address late deliveries and cancellations to enhance profitability and customer satisfaction(Sena & Ariyachandra, 2023).

DISCUSSION

The examination of the "Data Co Supply Chain" dataset, with a focus on logistics in Asia, offers significant insights into the effectiveness and profitability of supply chain activities. Initial data processing consisted of filtering essential variables like actual and scheduled shipping days, benefit per order, sales per customer, and late delivery risk, which were vital for comprehending delivery performance.

Using the K-Nearest Neighbors (K-NN) model, the data was trained and tested to evaluate its predictive capabilities. The model achieved a high accuracy of 97.91%, demonstrating its effectiveness in classifying delivery statuses. The classification report indicated precision, recall, and F1-scores of 0.98 across most categories, signifying the model's reliability in predicting correct delivery outcomes and identifying true positives with minimal errors.

Further, the scatter plot analysis between sales per customer and benefit per order, categorized by shipping statuses, revealed critical trends. A dense cluster of data points at lower sales per customer levels showed predominantly negative benefits, particularly with late deliveries. This suggests that delays in delivery are significantly impacting profitability at lower sales volumes. Conversely, higher sales per customer were associated with positive benefits, regardless of shipping status, indicating that increased sales correlate with better financial outcomes.

The presence of outliers where shipping was canceled highlights areas of substantial loss, emphasizing the importance of mitigating order cancellations to improve overall supply chain performance. These findings underscore the necessity for businesses to focus on reducing late deliveries and cancellations to enhance customer satisfaction and profitability.

Overall, this comprehensive analysis underscores the critical relationship between efficient delivery logistics and financial performance in supply chain management. By leveraging the K-NN model and focusing on key performance metrics, businesses can gain deeper insights into operational efficiencies and make informed strategic decisions to optimize their supply chain operations. This study highlights the potential for using machine learning models in logistics to drive improvements in delivery performance and profitability, particularly in the dynamic and complex market conditions prevalent in Asia(Khan et al., n.d.-b).

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

In conclusion, this study analyzed the "Data Co Supply Chain" dataset to evaluate supply chain logistics within Asia, employing the K-Nearest Neighbors (K-NN) model to classify delivery statuses and identify key performance metrics. The K-NN model demonstrated high accuracy, with precision, recall, and F1-scores of 0.98, effectively predicting delivery outcomes. The scatter plot analysis revealed that late deliveries significantly impact profitability, particularly at lower sales per customer levels where the benefit per order is negative. Conversely, higher sales per customer were consistently associated with positive benefits, indicating that increasing sales volumes lead to better financial outcomes. Additionally, substantial losses were observed in cases of canceled shipping, emphasizing the need to prevent order cancellations through more reliable and consistent shipping processes. These findings provide strategic insights for businesses to optimize logistics and enhance profitability in Asia's supply chain sector. By leveraging machine learning models like K-NN, companies can identify operational inefficiencies and make data-driven decisions to improve their supply chain operations. This study underscores the critical relationship between efficient delivery logistics and financial performance, highlighting the potential for strategic enhancements in supply chain management(Helo & Hao, 2022).

Recommendations

Based on the analysis and conclusions drawn from the "Data Co Supply Chain" dataset, the following recommendations are proposed to improve supply chain logistics and enhance profitability:

1. Utilize Advanced Predictive Analytics: Harness machine learning techniques like K-Nearest Neighbors (K-NN) to drive insights, to continuously monitor and predict delivery statuses. This can help in proactively identifying potential delays and taking corrective actions to ensure timely deliveries.
2. Focus on Reducing Late Deliveries: Develop strategies to minimize late deliveries, which significantly impact profitability. This could include optimizing routing algorithms, improving coordination with logistics partners, and enhancing real-time tracking and communication systems
3. Increase Sales Volume: Invest in marketing and sales strategies to drive higher sales volumes per customer, as higher sales are associated with positive benefits. Promotional campaigns, customer loyalty programs, and expanding market reach can contribute to achieving this goal.

ADVANCED RESEARCH

Building upon the findings and recommendations from this study, several avenues for advanced research can be pursued to further optimize supply chain logistics and enhance profitability within the Asian market. Future research could explore the integration of sophisticated machine learning algorithms, such as ensemble methods and deep learning models, to improve the accuracy and robustness of delivery status predictions. Comparative studies between these advanced techniques and traditional models like K-NN could provide insights Exploring the most optimal strategies for distinct supply chain contexts is crucial. Analyzing the influence of real-time data analytics and the Internet of Things (IoT) on supply chain efficiency presents another noteworthy avenue. Utilizing IoT devices for live tracking and advanced analytics can offer heightened visibility and management capabilities. Additionally, delving into the significance of big data analytics in supply chain enhancement is imperative. Research focusing on leveraging extensive data from diverse origins to augment predictive models and decision-making processes is essential. Additionally, research could delve into strategies for improving supply chain resilience and risk management, developing models to predict and mitigate risks such as natural disasters and geopolitical instability. Sustainable supply chain practices, customer-centric supply chain models, and the potential of blockchain technology to enhance transparency and security are also promising areas for advanced research. Finally, exploring the benefits of collaborative supply chain networks, where multiple stakeholders share data and resources, could lead to significant advancements in logistics efficiency.

By embarking on these progressive research avenues, academics and industry professionals can play a pivotal role in fostering the evolution of supply chain systems that are not only more efficient but also resilient and sustainable. These efforts will yield practical solutions tailored to the intricacies of the dynamic and swiftly changing Asian market, benefiting businesses across various sectors(Bharadiya et al., 2023).

ACKNOWLEDGMENT

We extend our heartfelt appreciation to our colleagues for their invaluable input and constructive feedback on this paper. Special thanks to our advisor at Universitas Pelita Bangsa for supporting this research, making it possible to complete this paper. We also appreciate the open-source community for providing the datasets and tools that facilitated our research.

REFERENCES

Bharadiya, J., Praful Bharadiya, J., & Praful Bharadiya, J. A. (2023).): 24-30 with Big Data Analytics. *American Journal of Artificial Intelligence*, *7*(1), 24–30. https://doi.org/10.11648/j.ajai.20230701.14

Bisnis, J., Sains, D., Meinar, R., Febriany, E., Wiludjeng, S., & Purwaningdyah, S. (2022). ANALISIS PENENTUAN RUTE DISTRIBUSI LIQUEFIED PETROLEUM GAS (LPG) TABUNG 3 KG MENGGUNAKAN METODE NEAREST NEIGHBOR PADA PT. RADE PUTRA UTAMA. *Jurnal Bisnis, Ekonomi, Dan Sains*, *2*(2), 288–294. https://doi.org/10.33197/BES.VOL2.ISS2.2022.1602

Chen, D. Q., Preston, D. S., Swink, M., Chen, D. Q. ;, Preston, D. S. ;, & Jiang, J. (2021). How Big Data Analytics Affects Supply Chain Decision-Making: An Empirical Analysis. *Journal of the Association for Information Systems*, *22*(5), 1224–1244. https://doi.org/10.17705/1jais.00713

Fitriani, S. A., Astuti, Y., & Wulandari, I. R. (2022). Least Absolute Shrinkage and Selection Operator (LASSO) and k-Nearest Neighbors (k-NN) Algorithm Analysis Based on Feature Selection for Diamond Price Prediction. *2021 International Seminar on Machine Learning, Optimization, and*

Data Science, ISMODE 2021, 135–139. https://doi.org/10.1109/ISMODE53584.2022.9742936

Gallego, A. J., Calvo-Zaragoza, J., & Rico-Juan, J. R. (2020). Insights into Efficient k-Nearest Neighbor Classification with Convolutional Neural Codes. *IEEE Access*, *8*, 99312–99326. https://doi.org/10.1109/ACCESS.2020.2997387

Helo, P., & Hao, Y. (2022). Artificial intelligence in operations management and supply chain management: an exploratory case study. *Production Planning & Control*, *33*(16), 1573–1590. https://doi.org/10.1080/09537287.2021.1882690

Khan, A. M., Protasov, S., Adil, ·, & Khan, M. (n.d.-a). *Using Proximity Graph Cut for Fast and Robust Instance-based Classification in Large Datasets Proximity Graph Cut Application to Instance-based Classification*. https://doi.org/10.13140/RG.2.2.26103.14243

Khan, A. M., Protasov, S., Adil, ·, & Khan, M. (n.d.-b). *Using Proximity Graph Cut for Fast and Robust Instance-based Classification in Large Datasets Proximity Graph Cut Application to Instance-based Classification*. https://doi.org/10.13140/RG.2.2.26103.14243

Na, J., Wang, Z., Lv, S., & Xu, Z. (2021). An Extended K Nearest Neighbors-Based Classifier for Epilepsy Diagnosis. *IEEE Access*, *9*, 73910–73923. https://doi.org/10.1109/ACCESS.2021.3081767

Odimarha, A. C., Ayodeji, S. A., & Abaku, E. A. (2024). MACHINE LEARNING'S INFLUENCE ON SUPPLY CHAIN AND LOGISTICS OPTIMIZATION IN THE OIL AND GAS SECTOR: A COMPREHENSIVE ANALYSIS. *Computer Science & IT Research Journal*, *5*(3), 725–740. https://doi.org/10.51594/CSITRJ.V5I3.976

Reklitis, P., Sakas, D. P., Trivellas, P., & Tsoulfas, G. T. (2021). Performance Implications of Aligning Supply Chain Practices with Competitive Advantage: Empirical Evidence from the Agri-Food Sector. *Sustainability 2021, Vol. 13, Page 8734*, *13*(16), 8734. https://doi.org/10.3390/SU13168734

Sena, M. P., & Ariyachandra, T. (2023). An Examination of Tableau as a Supplement to Excel to Enhance Data Literacy Skills. *Information Systems Education Journal*, *21*(4), 15–22. https://isedj.org/;https://iscap.info

Stilinski, D., & Frank, L. (2024). *Harnessing Predictive Analytics and Generative AI for Proactive Supply Chain Management: A Comprehensive Overview*. https://www.researchgate.net/publication/379574199

Yandrapalli, V. (2023). Revolutionizing Supply Chains Using Power of Generative AI. *International Journal of Research Publication and Reviews Journal Homepage: Www.Ijrpr.Com*, *4*, 1556–1562. www.ijrpr.com