JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6467

OPTIMIZATION CVRP WITH MACHINE LEARNING FOR IMPROVED CLASSIFICATION OF IMBALANCED DATA FOOD DISTRIBUTION

Muhammad Syahputra Novelan[1*]; Solly Aryza[2]

Magister of Information Technology[1]
Electrical Engineering [2]
Universitas Pembangunan Panca Budi, Medan, Indonesia[1,2]
www.unpab.ac.id[1,2]
putranovelan@dosen.pancabudi.ac.id[1*], sollyaryza@dosen.pancabudi.ac.id[2]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *The classification of imbalanced data in food delivery distribution is an important issue that needs to be considered to ensure fairness and efficiency in the food distribution system. This research answers these problems by improving the accuracy of the classification of delivery locations that have imbalanced demand data, so that high priority areas are not neglected. Generating more efficient and cost-effective distribution routes, taking into account vehicle capacity and delivery urgency. Reducing delivery time and potential food waste due to delays or non-optimal route allocation. This study addresses the problem of improving the accuracy of delivery location classification that has imbalanced demand data, so that high priority areas are not neglected. Generate more efficient and cost-effective distribution routes, taking into account vehicle capacity and delivery urgency. Reduce delivery time and potential food wastage due to delays or non-optimal route allocation. This study uses the research stages of data collecting, data preprocessing, and implementation of K-Means and K-NN methods. The results of CVRP testing with K-Means show the value of cluster 7 acc=80, precc=85, recall=84. cluster 9 acc=85, precc=90, recall=91. cluster 11 acc=88, precc=93, recall=94. While the results of CVRP testing with K-NN show the value of K 7 acc=89, precc=88, recall=85. value of K 9 acc=87, precc=90, recall=91. value of K 11 acc=95, precc=97, recall=94. The optimization results show that this approach not only improves operational efficiency but also increases the accuracy of food delivery, which will affect the availability of traditional markets.*

Keywords: *classification, food distribution, imbalanced data, machine learning, optimization*

Intisari— *Klasifikasi data yang tidak seimbang dalam distribusi pengiriman makanan merupakan masalah penting yang perlu diperhatikan untuk memastikan keadilan dan efisiensi dalam sistem distribusi makanan. Penelitian ini menjawab permasalahan tersebut dengan meningkatkan akurasi klasifikasi lokasi pengiriman yang memiliki data permintaan yang tidak seimbang, sehingga daerah yang memiliki prioritas tinggi tidak terabaikan. Menghasilkan rute distribusi yang lebih efisien dan hemat biaya, dengan mempertimbangkan kapasitas kendaraan dan urgensi pengiriman. Mengurangi waktu pengiriman dan potensi pemborosan bahan makanan akibat keterlambatan atau alokasi rute yang tidak optimal. Penelitian ini membahas masalah peningkatan akurasi klasifikasi lokasi pengiriman yang memiliki data permintaan yang tidak seimbang, sehingga area yang memiliki prioritas tinggi tidak terabaikan. Menghasilkan rute distribusi yang lebih efisien dan hemat biaya, dengan mempertimbangkan kapasitas kendaraan dan urgensi pengiriman. Mengurangi waktu pengiriman dan potensi pemborosan bahan makanan akibat keterlambatan atau alokasi rute yang tidak optimal. Penelitian ini menggunakan tahapan penelitian pengumpulan data, preprocessing data, dan implementasi metode K-Means dan K-NN. Hasil pengujian CVRP dengan K-Means menunjukkan nilai cluster 7 accuracy = 80, precision = 85, recall = 84. Cluster 9 acc=85, precc=90, recall=91. Cluster 11 acc=88, precc=93, recall=94. Sedangkan hasil pengujian CVRP dengan K-NN menunjukkan nilai K 7 accuracy=89, precision=88, recall=85. Nilai K 9 accuracy = 87, precision = 90, recall = 91. Nilai K 11 accuracy = 95, precision = 97, recall = 94. Hasil optimasi menunjukkan bahwa pendekatan ini tidak hanya meningkatkan efisiensi operasional tetapi* *juga meningkatkan akurasi pengiriman bahan makanan, yang akan berdampak pada ketersediaan pasar tradisional.*

Kata Kunci: *classification, food distribution, imbalanced data, machine learning, optimization*

INTRODUCTION

Imbalanced data in machine learning refers to a situation where the number of examples of each class in a dataset is unbalanced, which may result in a biased model being built. In many applications, such as medical diagnosis or fraud detection, minority classes are often more important despite being fewer in number. When models are trained on unbalanced data, they tend to overlook the minority classes, resulting in poor performance in detecting or classifying such data [1][2][3]. Imbalanced data refers to a situation where the distribution of classes in a dataset is unbalanced, where one class has a much larger number of examples than another. This phenomenon often occurs in various fields, including agriculture [4][5]. The problem of shipping food distribution from Berastagi to Medan involves several challenges that can affect the efficiency and effectiveness of the distribution process. First, geographical conditions and sometimes poor road infrastructure can hamper transportation. Winding and often damaged roads in mountainous areas can cause delays in the delivery process and increase the risk of damage to sensitive foodstuffs [6].

The classification of imbalanced data in food delivery distribution is an important issue that needs to be addressed to ensure fairness and efficiency in the food distribution system. In many cases, data pertaining to certain types of food, such as organic vegetables or local food products, is often less than that pertaining to common food products such as rice or wheat. This imbalance may cause classification models built to analyze and predict food delivery patterns to overlook minority classes. As a result, decisions taken in distribution planning can be biased, which can negatively impact food access and availability in different communities. Class imbalance is one of the influential characteristics on the performance of machine learning algorithms and this issue has received attention in academia and industry [7][8][9].

Therefore, it is important to apply appropriate techniques to handle imbalanced data, such as oversampling, undersampling, or specially designed algorithms, so that the model can accurately recognize and predict all food types, thus supporting more equitable and fair food distribution [8][10]. To improve accuracy and fairness in food distribution classification, it is important to apply machine learning techniques that are able to handle the problem of imbalanced data. Various methods, such as oversampling the minority class, undersampling the majority class, or the use of special algorithms designed for imbalanced situations, such as K-Means and K-Nearest Neighbors, have been used [10][11]. In this study, the authors implemented the capacity vehicle routing problem algorithm to limit the amount of food distribution transportation. Vehicle Routing Problem (VRP) is one of the important issues in transportation system. Vehicle Routing Problem (VRP) is a distribution system problem that aims to create an optimal vehicle path for a group of vehicles whose capacity is known, in order to meet customer demand with a known location and number of requests [12][1][13]. Research [14] presents comprehensive research results on the application of optimization and machine learning in last-mile logistics. The authors classify the literature into three main approaches: vehicle routing problem (VRP) optimization models, machine learning models, and hybrid approaches, and identify the main research gaps in each area. One way to overcome this difficulty is to develop hybrid approaches that combine machine learning as a predictive tool (e.g. in demand or service time) and metaheuristics as an optimization solution tool.

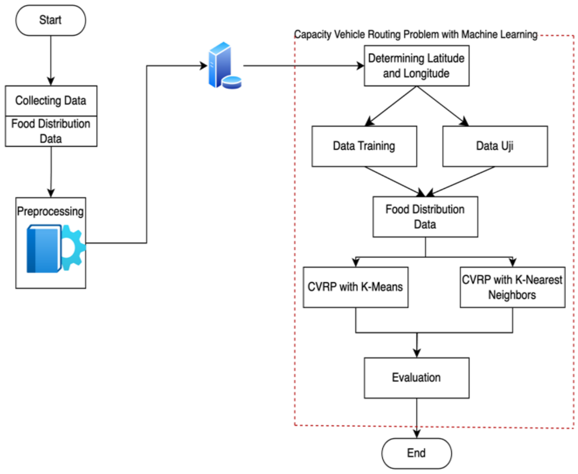
Research [15] presents research results on the application of optimization and machine learning techniques in last-mile logistics, with a focus on the classification of vehicle route optimization model approaches, machine learning, as well as mixed approaches. It is shown that the use of machine learning algorithms, especially for demand prediction and anomaly detection, can significantly improve the efficiency of logistics distribution. However, there are still unresolved issues, especially regarding the integration of AI-based predictive methods with complex optimization systems in real-time in the context of unbalanced data. Handling imbalanced data in food distribution is empirically proven to significantly improve the sensitivity and specificity of the model. This study aims to identify the factors determining the classification of food distribution destination locations and provide their impact on the efficiency of distribution routes from Berastagi to Medan in order to design an optimal distribution strategy based on vehicle capacity limitations, geographical conditions, and regional priorities.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6467

MATERIALS AND METHODS

This research method is based on Berastagi city food delivery data. This dataset will be used for Imbalanced Classification of Food Distribution Data. The following is the research method used in this research:



Source: (Research Results, 2025)
Figure 1. Research Stages

Based on Figure 1, the stages of this research are divided into several main steps that form a workflow in solving the Capacitated Vehicle Routing Problem (CVRP) with a machine learning approach. The stages include:

Data Collection

Systematic and structured data collection is critical to the success of this research. Accurate and relevant data enables the development of more effective CVRP optimization models, as well as improving the accuracy of product demand prediction in the context of food distribution. With the right approach to data collection and processing, this research is expected to contribute significantly to the efficiency and effectiveness of the food distribution system [16][17]. The data used in this study are data from food delivery companies in the Berastagi area which are sent to several locations in Medan City. Data collection methods were carried out through documentation and semi-structured interviews. Documentation was obtained from transaction records and delivery logs available in the company's internal system, including delivery history for a certain period (for example, data dated June 1, 2025 as shown in Table 1).

Semi-structured interviews were conducted with the company's operational staff, including couriers and field officers, to explore additional information regarding distribution patterns, geographic challenges, and route grouping practices that have been implemented manually. This approach was chosen to obtain quantitative data as well as a contextual understanding of the food distribution process in the region.

Table 1. Food Distribution Dataset

Delivery Time	Farmers	Recipient	Recipient Address
2025-06-01 08.11.15	SERTU SAFRIZAL SARAGIH	admin	SUMATERA UTARA, MEDAN, Medan Pajak Induk Tuntungan
2025-06-01 07.47.06	EMA MANURUNG	TTD RETUR	Sumatera Utara, Medan, Pajak Sambas
2025-06-01 14.45.32	ESMERAIDA SITEPU	Basaulian Aritonang	Sumatera Utara, Medan, ,,Pajak Sukaramai
2025-06-01 09.49.13	PT. PANGKLALAN BARU INDAH	Seller	Sumatera Utara, Medan. Pajak Simpang Limun
2025-06-01 09.04.27	Taty Nursery Tambunan	Salman Nasution ,031	SUMATERA UTARA, MEDAN, Pajak Ikan Lama
2025-06-01 08.02.17	RISMA SONDANG	CP/Kurir Pickup	SUMATERA UTARA, MEDAN, Pajak Kampung Lalang
2025-06-01 08.00.12	Ambar Murboreno	Joni	SUMATERA UTARA, MEDAN, Pajak Induk Pancing
2025-06-01 08.01.59	Bachtera jusup purba	CP/Kurir Pickup	Sumatera Utara, Medan, Pajak Titi Kuning

Source: (Research Results, 2025)

Based on Table 1, food distribution data from farmers to various market locations in Medan on June 1, 2025 is shown. The information presented includes delivery time, farmer name, recipient, and destination address. This data reflects variations in distribution destination locations such as Induk Tuntungan Tax, Sambas Tax, Sukaramai Tax, and others, which are important inputs in CVRP modeling.

Data Preprocessing

Data preprocessing is the initial process that will transform the data entered into the system into data in a proper format and ready for processing. Performing data cleaning, also known as data cleansing, is the process of identifying and correcting or removing errors, inconsistencies, and inaccuracies present in a data set. Various necessary processes such as merging, deforming, reducing, and discretizing are some examples of preprocessing. This stage includes several main processes: [18][19].

Data Cleaning: Includes removing and correcting missing values, duplications, input errors, and inconsistencies between attributes. This cleaning aims to reduce bias and improve model accuracy.

Data Transformation: This process includes scale normalization (e.g. on request volume and delivery distance), encoding categorical variables (if any), and converting location data into geographic coordinates (latitude-longitude) using API.

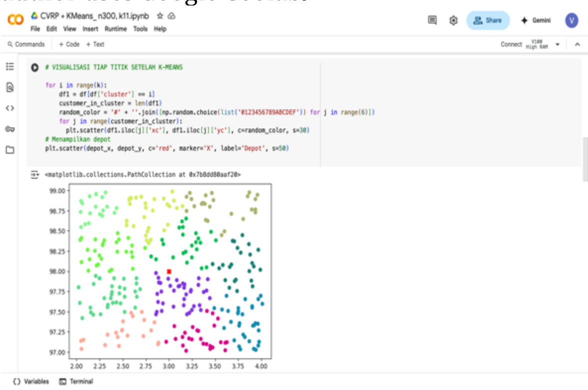
Reduction and Discretization: Irrelevant or redundant features are removed, while some continuous variables may be kept in discrete intervals to support the clustering process.

Implementasi CVRP dengan Machine Learning

This implementation is carried out in accordance with the selected optimization that will be applied to the preprocessing dataset for the classification of imbalanced food distribution data. This implementation technique uses the capacity vehicle routing problem classification with machine learning.

CVRP with K-Means

In this section the author performs CVRP optimization with K-Means, where at this stage the author uses Google Coolab.

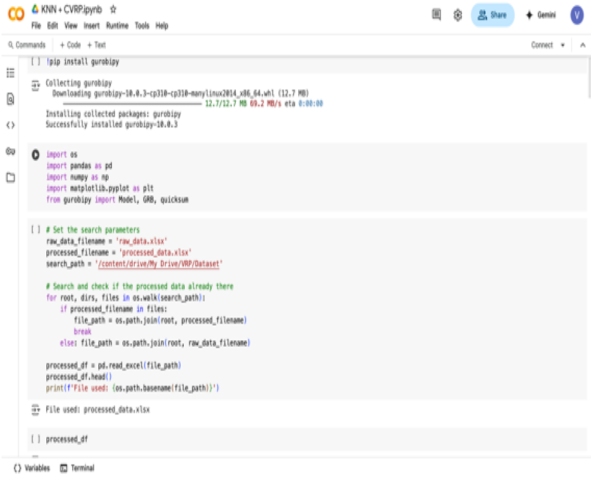


Source: (Research Results, 2025)
Figure 2. Implementation of CVRP with K-Means

Figure 2 shows the visualization results of the distribution destination location clusters formed using the K-Means algorithm. Each point on the graph represents one distribution destination address, which has previously been converted into geographic coordinates (longitude and latitude). Different colors on the graph indicate the results of the grouping (clusters) of the K-Means process based on geographic position similarities.

CVRP with K-Nearest Neighbors

At this stage the author uses Google Collab for the implementation of CVRP with K-NN.



Source: (Research Results, 2025)
Figure 3. CVRP Implementation with K-NN

Figure 3 shows the initial stages in data processing using Python in Google Colab, specifically for reading and managing Excel files (.xlsx) containing food distribution data.

Evaluation

Evaluation metrics are created in a similar manner to the process followed in binary and multiclass classification, specifically for your Label task. These metrics provide a quantitative assessment of the model's performance in handling multiple classes, providing insight into aspects such as precision, recall, F1 score, and accuracy for each class in binary and multiclass classification problems. This provides a solid basis for evaluating the effectiveness of the model in distinguishing between different classes in multiclass and binary classification databases [18].

$$\square\square\square = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Prec = \frac{TP}{TP+FP} \quad (2)$$

$$Rec = TP_{rate} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = \frac{2}{precision^{-1}+recall^{-1}} = 2\ x\ \frac{precision\ x\ recall}{precision+recall} \quad (4)$$

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6467

RESULTS AND DISCUSSION

In this section, we will discuss the implementation of food distribution with the capacity vehicle routing problem. first, the capacity vehicle routing problem algorithm, combined with the K-Means algorithm and second, combined with the K-NN algorithm. The study aims to determine the optimal vehicle route by minimizing the total distance traveled while meeting the capacity constraints of the amount of food transported.

Input: A dataset consisting of service points (e.g., customers or delivery locations), vehicle capacity, and a distance matrix between points.

Output: An optimized food distribution route with the allocation of service points to the appropriate vehicles, ensuring that each vehicle does not exceed the specified capacity. Confirm the capacity of the vehicle. Note that all vehicles have the same capacity.

$$\sum_{i=1}^{n} \sum_{j=2}^{n} qjxijk \leq Q. \quad \forall k \in \{1, \dots, p\} \qquad (5)$$

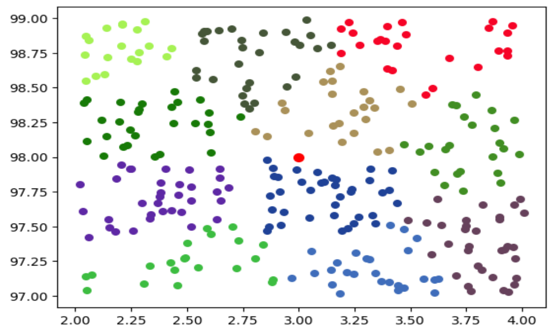
CVRP with K-Means

At this stage the author performs CVRP optimization with K-Means, where the first stage makes cluster points and calculates the distance of each point to be sent for food distribution as in the following table.

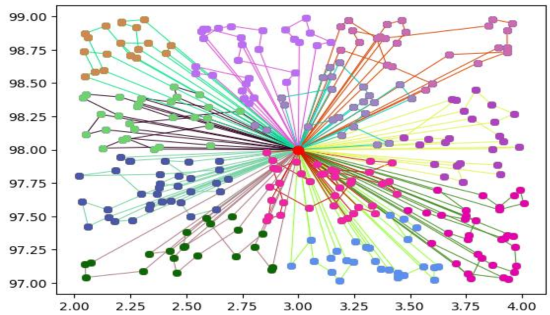
Table 2. Cluster Point Process Data

No	XC	YC	Cluster
0	3.107627	98.542614	10
1	3.440379	97.669627	6
2	3.215527	97.170581	0
3	3.099766	97.816338	6
4	2.857310	97.469824	6
...			
298	3.955839	98.949463	
299	3.931669		4
		97.355824	
300	3.823111	97.659416	4

Source: (Research Results, 2025)



Source: (Research Results, 2025)
Figure 4. Visualization of Each CVRP Point with K-Means

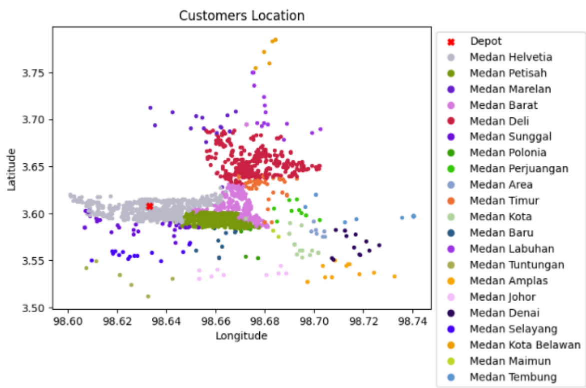


Source: (Research Results, 2025)
Figure 5. CVRP Optimization Results with K-Means

In figures 4 and Figure 5 Food recipient location refers to the place where the goods or consignment will be received or forwarded by the intended recipient. This could be a home address, business address, or any other location where the consignee has the presence or facilities to receive the goods. The location of the consignee is important in the delivery and logistics process, as it is the final point in the supply chain where the product or goods are handed over to the customer.

CVRP with K-Nearest Neighbors

At this stage, the optimization process is carried out by applying CVRP with K-NN. This stage is carried out using Google Collab based on the following stages: The implementation stages include: (1) pre-processing of unbalanced food distribution data; (2) application of K-NN algorithm for classification of demand points based on geographical proximity and demand volume; (3) establishment of distribution routes based on vehicle capacity using simple heuristics; and (4) evaluation of model performance using classification accuracy and route efficiency metrics (total distance and capacity fulfillment).

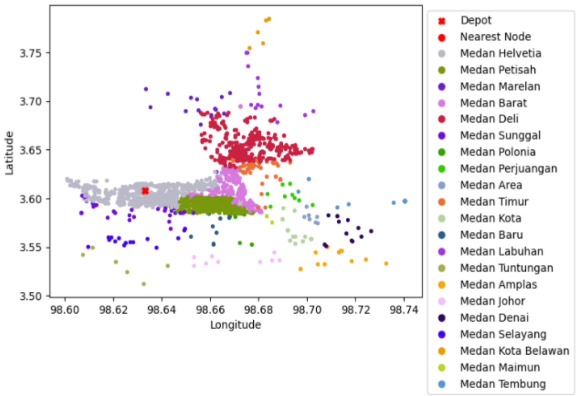


Source: (Research Results, 2025)
Figure 6. Data on Food Distribution Centers

Figure 6 Random customer center point data with neighbors and depots is a collection of

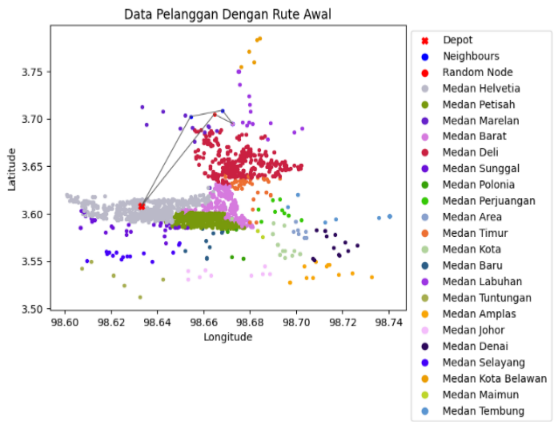
VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i4.6467

information that includes the geographical location of randomly scattered customer points in a certain area, as well as information about the neighbors of each customer point and the location of the depots or distribution centers that serve them.



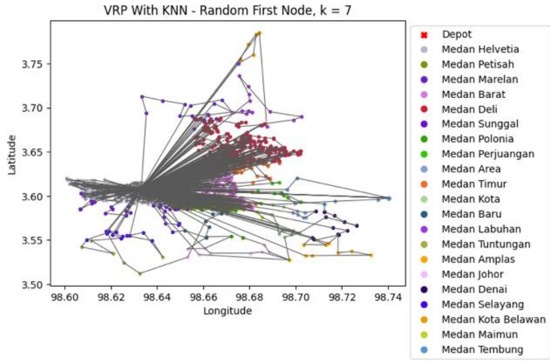
Source: (Research Results, 2025)
Figure 7. Random Center Point Data

Figure 7 Random customer center point data with neighbors and depots is a collection of information that includes the geographical location of randomly scattered customer points in a certain area, as well as information about the neighbors of each customer point and the location of the depots or distribution centers that serve them.

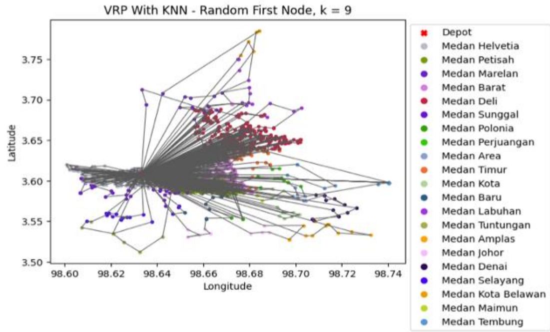


Source: (Research Results, 2025)
Figure 8. Initial Route Food Distribution Data

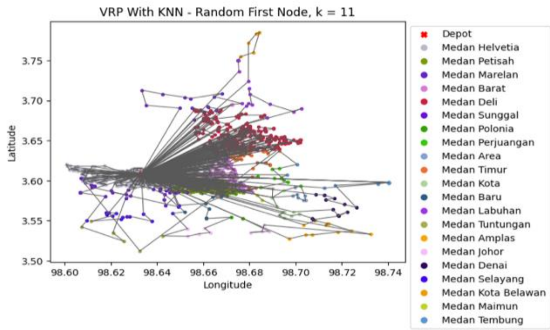
In figure 8 The location of the consignee refers to the place where the goods or shipment will be received or forwarded by the intended recipient. This could be a home address, business address, or any other location where the consignee has the presence or facilities to receive the goods. The location of the consignee is important in the delivery and logistics process, as it is the final point in the supply chain where the product or goods are handed over to the customer.



Source: (Research Results, 2025)
Figure 9. Result K=7 CVRP Optimization with KNN



Source: (Research Results, 2025)
Figure 10. Result K=9 CVRP Optimization with KNN



Source: (Research Results, 2025)
Figure 11. Result K=11 CVRP Optimization with KNN

From the test results, the implementation of the Capacity Vehicle Routing Problem (VRP) model with K-Nearest Neighbor (KNN) involves the process of mapping optimal delivery routes to visit a number of customers using the K-NN algorithm to determine efficient routes. The result of K=7, K=9, K=11 Capacity Vehicle Routing Problem Optimization with K-Nearest Neighbors is an optimization solution for delivery routes that satisfy vehicle capacity constraints and consider distance or travel time. In this case, K refers to the number of nearest customers considered by the KNN algorithm.

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6467

Table 3. Percentage Results of Vehicle Routing Problem Model Implementation with K-Nearest Neighbor

Data sets	Value K	Total Customers	Total Testing Time	Total Distance	Total Routes	Average Distance
Murni Cargo	7	2579	66.005 seconds	15851km	285	10.004 km / route
Murni Cargo	9	2579	43.584 seconds	17887 km	285	10.031 km / route
Murni Cargo	11	2579	66.361 seconds	2564194 km	190	10.782 km / route

Source: (Research Results, 2025)

Overall, the implementation of the CVRP model with KNN can bring great benefits to companies or organizations in organizing and managing food distribution deliveries more efficiently, reducing operational costs, and improving customer service.

Evaluation

Performance Evaluation of Capacity Vehicle Routing Problem Optimization Model with Machine Learning in Classification of Imbalanced Vehicle Routing Data can be done through several evaluation metrics commonly used in machine learning and optimization, as well as through performance analysis of the resulting solution. The following are some performance evaluation methods that can be used:

Table 4. CVRP Testing Results with K-Means

Datasets	Value Centroid	Performance				
		Total Customers	Total Testing Time	Accuracy	Precision	Recall
Murni Cargo	7	2579	58.624 seconds	80	85	84
Murni Cargo	9	2579	47.231 seconds	85	90	91
Murni Cargo	11	2579	57.365 seconds	88	93	94

Source: (Research Results, 2025)

Table 5. CVRP Testing Results with K-Nearest Neighbors

Datasets	Value K	Performance				
		Total Customers	Total Testing Time	Accuracy	Precision	Recall
Murni Cargo	7	2579	66.210 seconds	89	88	85
Murni Cargo	9	2579	40.652 seconds	87	90	91
Murni Cargo	11	2579	55.651 seconds	95	97	94

Source: (Research Results, 2025)

CVRP test results with K-Means show cluster value 7 acc=80, precc=85, recall=84. cluster value 9 acc=85, precc=90, recall=91. cluster value 11 acc=88, precc=93, recall=94. While the results of CVRP testing with K-NN show the value of K 7 acc=89, precc=88, recall=85. the value of K 9 acc=87, precc=90, recall=91. the value of K 11 acc=95, precc=97, recall=94.

The Capacity K-Nearest Neighbors implementation percentage result is a performance evaluation metric that gives an idea of how effective the algorithm is in solving routing problems considering vehicle capacity. By using the percentage result of CVRP implementation, we can evaluate the performance of the algorithm in solving the routing problem considering vehicle capacity, and also understand how effective the algorithm is in improving operational efficiency.

CONCLUSION

This research explores the application of CVRP with K-Means and K-NN using food distribution data from Berastagi to Medan expedition company. This research highlights the importance of applying Machine Learning techniques in CVRP Optimization to improve the classification of imbalanced data in food distribution. The optimization results show that this approach not only improves operational efficiency but also increases the accuracy of food delivery, which will have an impact on the availability of traditional markets. This research successfully demonstrates that the integration of CVRP algorithm with Machine Learning techniques, specifically K-Means and K-NN, is able to improve the performance of food distribution from Berastagi to Medan, both in terms of operational efficiency and accuracy of unbalanced demand classification.

The test results show that: K=9 value gives the fastest execution time in KNN test, while the highest accuracy is achieved at K=11 for K-NN test (acc=95, precc=97, recall=94). Clustering using K-Means with 11 clusters also showed the best classification performance. Overall, this approach marks a significant advancement in the effort to optimize data-driven logistics distribution, especially in the context of imbalanced data. Although efficiency improvements have not been fully demonstrated, CVRP optimization with Machine Learning in the classification of unbalanced food distribution data marks a step forward in the conduct of this research. Suggestions for future research require handling Unbalanced Data: Using techniques such as SMOTE (Synthetic Minority Oversampling Technique) to address data imbalances more systematically. Economic and Social Impact Evaluation: Assess the impact of this distribution efficiency on traditional markets, food prices, and customer satisfaction.

ACKNOWLEDGMENT

This research was made possible thanks to financial support from Universitas Pembangunan Panca Budi. The author would like to thank you for your support.

REFERENCE

[1] A. Arifuddin, A. Utamima, F. Mahananto, R. A. Vinarti, and N. Fernanda, "Optimizing the Capacitated Vehicle Routing Problem at PQR Company: A Genetic Algorithm and Grey Wolf Optimizer Approach," *Procedia Comput. Sci.*, vol. 234, pp. 420–427, 2024, doi: 10.1016/j.procs.2024.03.023.

[2] A. K. Hamoud *et al.*, "A prediction model based machine learning algorithms with feature selection approaches over imbalanced dataset," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, pp. 1105–1116, 2022, doi: 10.11591/ijeecs.v28.i2.pp1105-1116.

[3] S. S. Kaddi and M. M. Patil, "Ensemble learning based health care claim fraud detection in an imbalance data environment," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 3, pp. 1686–1694, 2023, doi: 10.11591/ijeecs.v32.i3.pp1686-1694.

[4] A. Ugarković and D. Oreški, "Supervised and Unsupervised Machine Learning Approaches on Class Imbalanced Data," *Proc. Int. Conf. Smart Syst. Technol. SST 2022*, pp. 159–162, 2022, doi: 10.1109/SST55530.2022.9954646.

[5] R. Asín-Achá, A. Espinoza, O. Goldschmidt, D. S. Hochbaum, and I. I. Huerta, "Selecting fast algorithms for the capacitated vehicle routing problem with machine learning techniques," *Networks*, pp. 1–22, 2024, doi: 10.1002/net.22244.

[6] I. Santoso, M. Purnomo, A. A. Sulianto, and A. Choirun, "Machine learning application for sustainable agri-food supply chain performance: A review," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 924, no. 1, 2021, doi: 10.1088/1755-1315/924/1/012059.

[7] Yakshit, G. Kaur, V. Kaur, Y. Sharma, and V. Bansal, "Analyzing various Machine Learning Algorithms with SMOTE and ADASYN for Image Classification having Imbalanced Data," *Proc. 2022 IEEE Int. Conf. Curr. Dev. Eng. Technol. CCET 2022*, pp. 1–7, 2022, doi: 10.1109/CCET56606.2022.10080783.

[8] Dennis, I. R. Budianto, R. K. Azaria, and A. A. S. Gunawan, "Machine Learning-based Approach on Dealing with Binary Classification Problem in Imbalanced Financial Data," *2021 Int. Semin. Mach. Learn. Optim. Data Sci. ISMODE 2021*, pp. 152–156, 2022, doi: 10.1109/ISMODE53584.2022.9742834.

[9] J. Fellers, J. Quevedo, M. Abdelatti, M. Steinhaus, and M. Sodhi, "Selecting between evolutionary and classical algorithms for the CVRP using machine learning: Optimization of vehicle routing problems," *GECCO 2021 Companion - Proc. 2021 Genet. Evol. Comput. Conf. Companion*, no. July 2021, pp. 127–128, 2021, doi: 10.1145/3449726.3459459.

[10] A. Senen, J. Jamani Jamian, H. Satya Dini, E. Supriyanto, and D. Anggaini, "Sectoral electricity micro-spatial load forecasting based on partitional clustering technique," *IAES Int. J. Artif. Intell.*, vol. 13, no. 3, pp. 3533–3544, 2024, doi: 10.11591/ijai.v13.i3.pp3533-3544.

[11] Z. Zainuddin and A. A. N. Risal, "Balanced clustering for student admission school zoning by parameter tuning of constrained k-means," *IAES Int. J. Artif. Intell.*, vol. 13, no. 2, pp. 2299–2311, 2024, doi: 10.11591/ijai.v13.i2.pp2301-2313.

[12] M. S. Novelan, S. Efendi, P. Sihombing, and H. Mawengkang, "Vehicle Routing Problem Optimization With Machine Learning in Imbalanced Classification Vehicle Route Data," *Eastern-European J. Enterp. Technol.*, vol. 5, no. 3(125), pp. 49–56, 2023, doi:

VOL. 10. NO. 4 MAY 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i4.6467

10.15587/1729-4061.2023.288280.

[13] A. Naguib, W. A. Yousef, I. Traoré, and M. Mamun, “On Statistical Learning of Branch and Bound for Vehicle Routing Optimization,” 2023.

[14] N. Giuffrida, J. Fajardo-Calderin, A. D. Masegosa, F. Werner, M. Steudter, and F. Pilla, “Optimization and Machine Learning Applied to Last-Mile Logistics: A Review,” *Sustain.*, vol. 14, no. 9, 2022, doi: 10.3390/su14095329.

[15] A. O. Adegbenjo and M. O. Ngadi, “Handling the Imbalanced Problem in Agri-Food Data Analysis,” *Foods*, vol. 13, no. 20, 2024, doi: 10.3390/foods13203300.

[16] G. Abisheva, N. Goranin, B. Razakhova, T. Aidynov, and D. Satybaldina, “Specifics of Data Collection and Data Processing during Formation of RailVista Dataset for Machine Learning- and Deep Learning-Based Applications,” *Sensors*, vol. 24, no. 16, 2024, doi: 10.3390/s24165239.

[17] Y. D. et al. Mai, Thanh; M, Shahbaz; Tong, “Pr ep rin t n ot pe er r iew Pr ep rin t n ot pe er ed,” *Fusion*, no. September, pp. 1–8, 2023, doi: 10.2139/ssrn.4949181.

[18] Soni, M. A. Remli, K. M. Daud, and J. Al Amien, “Improving imbalanced class intrusion detection in IoT with ensemble learning and ADASYN-MLP approach,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 2, pp. 1209–1217, 2024, doi: 10.11591/ijeecs.v36.i2.pp1209-1217.

[19] K. Chandra Sekhar and K. S. Kumar, “Data Preprocessing and Visualizations Using Machine Learning for Student Placement Prediction,” *Proc. Int. Conf. Technol. Adv. Comput. Sci. ICTACS 2022*, pp. 386–391, 2022, doi: 10.1109/ICTACS56270.2022.9988247.