

Predictive Machine Learning Framework for Medical Relief Supply Decision-Making in Disaster Response Operations

Roman B. Villones, Rachel J. Vergara

Graduate School, La Consolacion University Philippines, Philippines

DOI: <https://doi.org/10.51583/IJLTEMAS.2025.1411000016>

Received: 10 November 2025; Accepted: 20 November 2025; Published: 02 December 2025

ABSTRACT

This study aims to strengthen decision-making in medical relief supply management through computational modeling and machine learning. Specifically, it seeks to develop a predictive model that supports effective medical supply allocation during disaster response operations. By experimenting with multiple algorithms and optimization techniques, the study endeavors to identify the most suitable machine learning approach for accurate and real-time prediction of supply reorder needs. The study adopts the SEMMA methodology as the framework for developing decision-making models that focus on optimizing medical supply distribution during natural calamities. Both ensemble and classical algorithms were implemented to compare predictive performance. Ensemble methods included Random Forest, Gradient Boosting, HistGradientBoosting, AdaBoost, XGBoost, LightGBM, and CatBoost, while classical algorithms comprised of Decision Tree, Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Model evaluation utilized the F1-score and ROC-AUC metrics, with hyperparameter tuning conducted via Randomized Search and 5-fold cross-validation to ensure robust generalization. Results revealed that CatBoost achieved the highest F1-score (0.9316), outperforming other models such as XGBoost (0.9282), Gradient Boosting (0.9279), and LightGBM (0.9277). These findings underscore the superiority of gradient boosting ensemble methods in capturing complex relationships and improving prediction accuracy. Traditional models, including AdaBoost, Decision Tree, Logistic Regression, KNN, and SVM are yielded with lower F1-scores ranging from 0.8051 to 0.8738 only and indicating its reduced predictive capability after optimization. The CatBoost algorithm demonstrated the most reliable and accurate predictive performance by highlighting its potential as a robust decision- support tool for optimizing medical supply allocation during disaster response operations.

Keywords — Predictive Modeling, Machine Learning, Decision Support Systems, Disaster Response, Medical Relief Supply Management.

INTRODUCTION

In recent years, the growing complexity of disaster environments has encouraged the adoption of data-driven and machine learning-based decision-support systems in humanitarian logistics (Rodríguez-Espíndola et al., 2023). The predictive modeling and artificial intelligence (AI) approaches are increasingly used to forecast medical supply demand and improve real-time decision-making during emergencies (Bastani et al., 2023). Furthermore, the integration of big data analytics and optimization algorithms has enabled more adaptive and responsive medical relief operations where critical resources can be dynamically deployed to high-impact areas (Bag et al., 2021).

Despite this advancement, the real-world disaster response operations remain constrained by data fragmentation and uncertain demand mechanisms. Inconsistencies in data collection and the absence of unified decision-support frameworks often result in inefficiencies such as overstocking and shortages of essential medicines and equipment (Ma, K. et al., 2022). Additionally, many existing models lack validation through real-world experiments by limiting their reliability and adoption by organizations.

This study aims to strengthen decision-making in medical relief supply management through computational modeling and machine learning. Specifically, it seeks to develop a predictive machine learning model to support

decision-making for medical supply allocation during disaster response operations. Experimenting with multiple machine learning algorithms to evaluate predictive performance and suitability in disaster contexts. And implementing optimization techniques and cross-validation methods to fine-tune model accuracy for real-time deployment.

The study contributes to the field of logistics and AI-driven decision systems in several ways. In methodological contribution, it introduces a predictive framework that integrates supervised machine learning, optimization, and cross-validation by addressing the gap between theoretical modeling and applied decision-making. For empirical contribution, experiments on disaster-response datasets, this study will benchmark multiple algorithms and identify which are most effective for high-uncertainty and time-sensitive environments. By practical contribution, the resulting decision-support model enhances situational awareness and enables data-informed allocation of medical supplies by helping relief organizations respond faster and more efficiently to disasters.

REVIEW OF RELATED LITERATURE

The SEMMA methodology was developed by the SAS Institute. It provides a systematic framework for developing and evaluating data mining and machine learning models. It focuses on transforming raw data into reliable insights through an iterative and structured process (Al Alawi et al., 2022). In emergency logistics research, constructing composite indices to quantify localized demand pressure between supply and need, in such composite metrics combine observed demand, casualty/impact estimates, and infrastructure accessibility to guide prioritization when raw demand signals are noisy (Ma, Z. et al., 2023).

Recent benchmarking literature categorizes machine learning models for classification tasks, for example, a comparative study of gradient-boosting algorithms under hyperparameter tuning on multiple datasets then noting a trade-off between accuracy and computational cost (Florek & Zagdański, 2023). Well-tuned boosting implementations typically achieve top performance on complex tabular tasks, while ensemble strategies provide robustness across noisy and imbalanced settings (Ayodele, 2023). At the same time, applied evaluations continue to include classical models as interpretable baselines to quantify practical gains from more complex learners (El Guabassi et al., 2021).

Recent analyses, threshold-sensitive and threshold-independent metrics together like F1-score for balancing precision/recall in imbalanced and ROC-AUC for overall discrimination across thresholds (Li, 2024). In addition, empirical work highlights AUC's relative stability across prevalence scenarios including the prediction accuracy for determining the model's performance in healthcare system (Owusu-Adjei et al., 2023). For hyperparameter tuning, modern applied studies increasingly use randomized search strategies as efficient in reproducible alternatives to exhaustive grids (A Ilemobayo et al., 2024). In addition, hyperparameter optimization documented that randomized sampling as a practical approach for large model families and limited compute budgets (Ali et al., 2023).

METHODOLOGY OF THE STUDY

In this study, SEMMA serves as the methodological foundation for building decision-making models aimed at optimizing medical supply allocation during natural calamity relief operations.

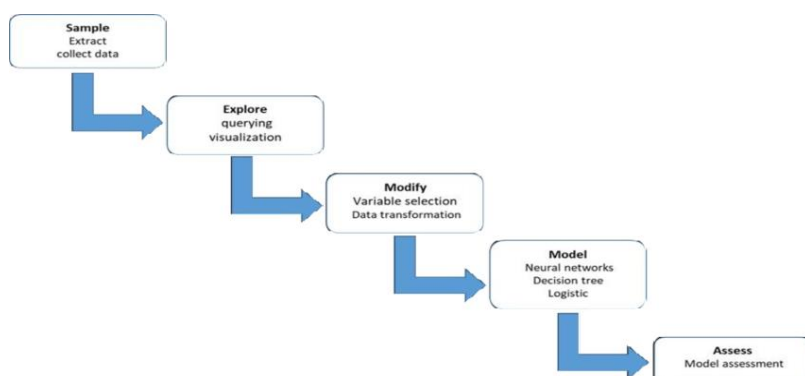


Fig. 1. SEMMA Methodology according to Al Alawi et al. (2022)

Sample

The dataset used in this study consists of 76,000 records and 10 attributes, representing critical variables that influence medical supply allocation during natural calamity relief operations. The dataset includes features such as Item ID and District, which identify the supply items and their geographic distribution. Operational variables such as Inventory Level, Relief Aid, and Inventory Re-Ordered capture logistical and replenishment patterns. Contextual variables like Calamities provide disaster-related information affecting supply demand. Derived indicators such as Simple Ratio Method, Weighted Demand Index, Normalized Score, and Overall Demand Score were computed to quantify supply prioritization and regional demand intensity.

Explore

This phase focuses on identifying patterns, correlations, and data quality issues that may influence model accuracy.

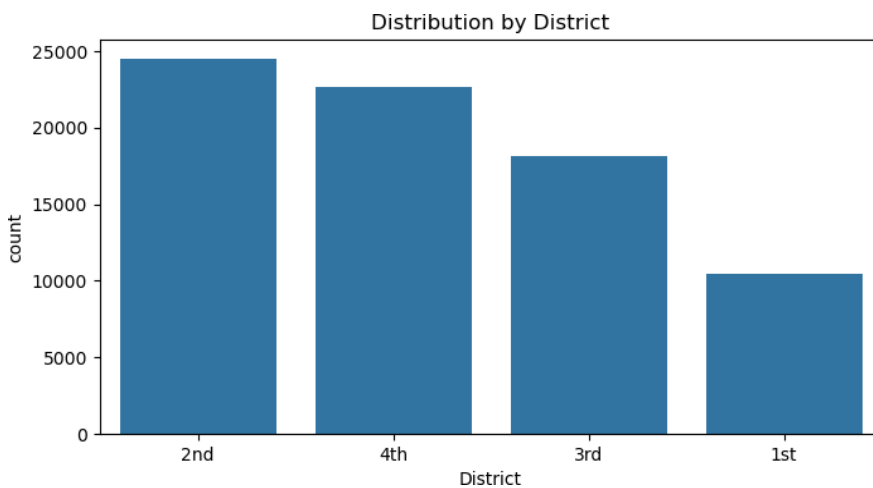


Fig. 2. Distribution of Medical Relief across Districts during the Calamities

Figure 2 presents the distribution of data entries by district in the NCR, Philippines. The 2nd District recorded the highest count, followed by the 4th and 3rd Districts, while the 1st District showed the lowest frequency. This indicates a significant variation in data concentration across districts, suggesting that the 2nd and 4th Districts may have higher levels of calamities encountered compared to the others.

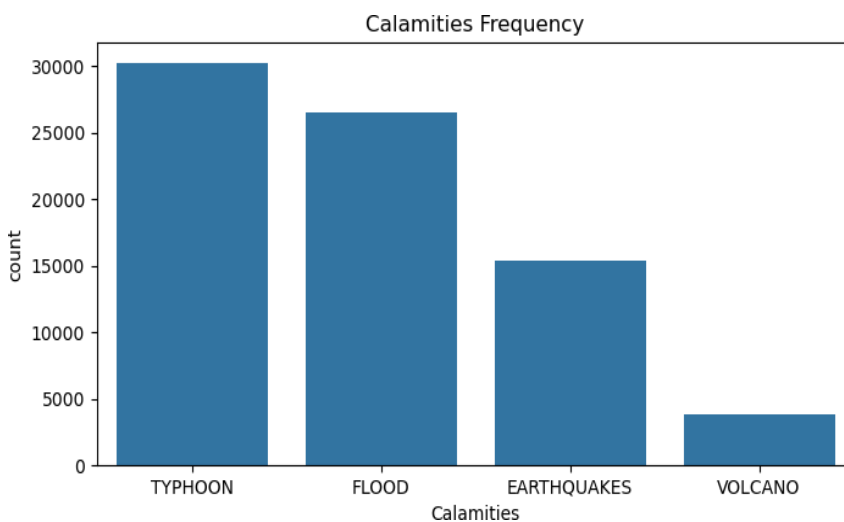


Fig. 3. Distribution of Medical Relief by Calamities

Figure 3 presents the frequency distribution of various calamities. Among the four categories, Typhoons recorded the highest count, followed by Floods and Earthquakes, while Volcano incidents showed the lowest occurrence. This indicates that typhoons are the most frequently experienced calamity in the NCR, Philippines, suggesting a strong climatic influence in the observed area whereas volcanic activities are relatively rare.

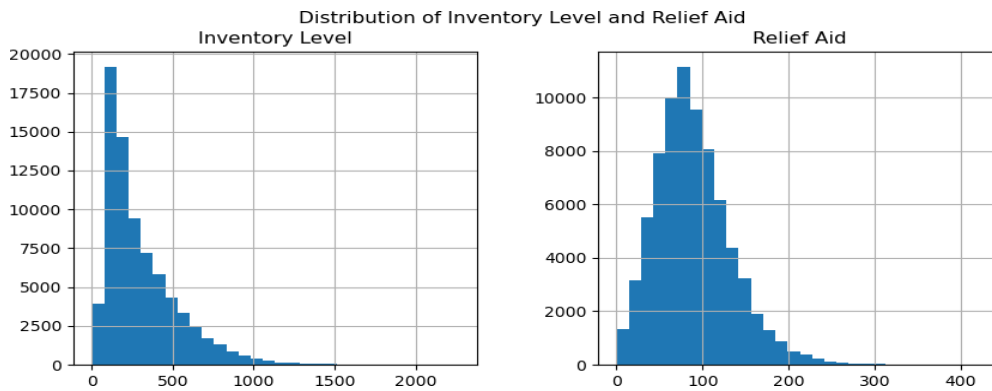


Fig. 4. Distribution of Inventory Level and Relief Aid

Figure 4 illustrates the distribution patterns of Inventory Level and Relief Aid. The histogram for inventory levels (left) shows a right-skewed distribution, indicating that most inventory counts are relatively low with fewer instances of high inventory accumulation. Similarly, the relief aid distribution (right) also demonstrates a right-skewed pattern, where most of the relief aid quantities are concentrated at lower values. This suggests that both inventory and relief aid are generally maintained at moderate levels with occasional peaks possibly during times of high demand or major calamities.

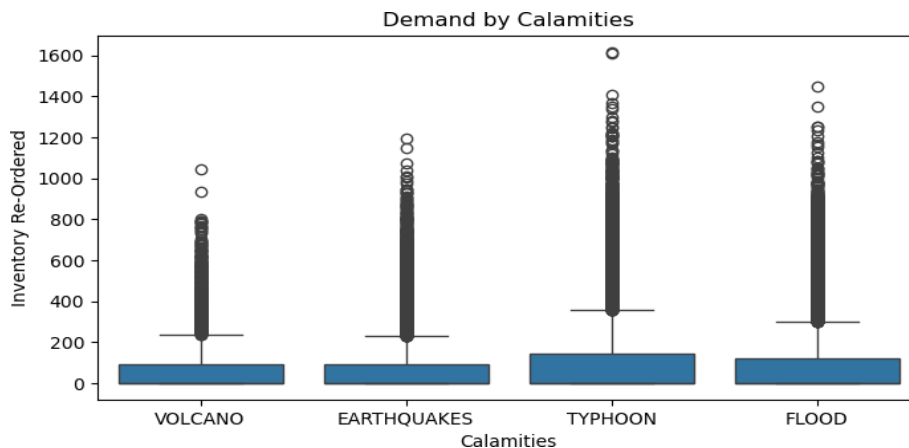


Fig. 5. Distribution of Inventory Re-Ordered by Type of Calamity

Figure 5 illustrates the distribution of inventory re-ordered across different types of calamities. The boxplot shows that all calamity categories exhibit a wide range of inventory re-orders with numerous outliers that indicates an instance of exceptionally high demand. Among the calamities, typhoons appear to have slightly higher median and upper-range values, suggesting that they tend to generate greater inventory demand compared to other disaster types. This trend implies that typhoon-related events may exert a more substantial impact on inventory replenishment activities and reflecting their frequency and severity in affected regions.

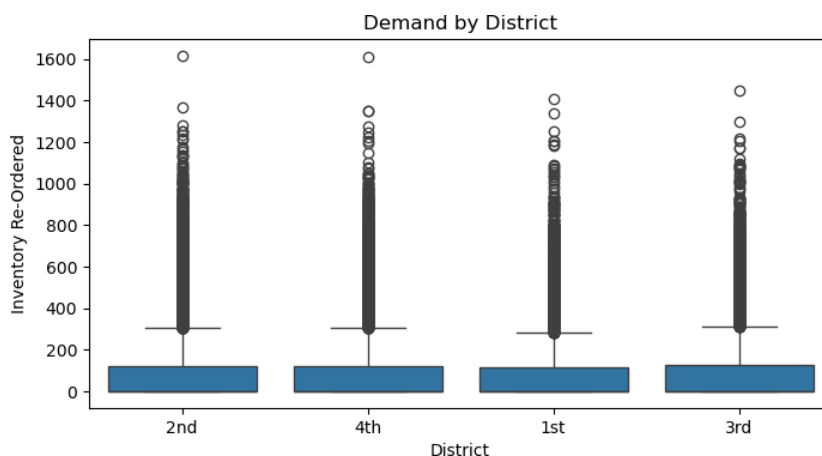


Fig. 6. Distribution of Inventory Re-Ordered by District

Figure 6 presents the distribution of inventory re-ordered across the four districts. The boxplot indicates that all districts exhibit a similar pattern of inventory demand with a concentration of lower re-order quantities and several outliers representing high-demand instances. Among the districts, the 2nd District displays a slightly higher median and broader range of re-orders, implying a relatively greater demand for inventory replenishment. This suggests that the 2nd District may experience more frequent or severe calamity-related disruptions compared to the other districts and necessitating increased inventory re-supply activities.

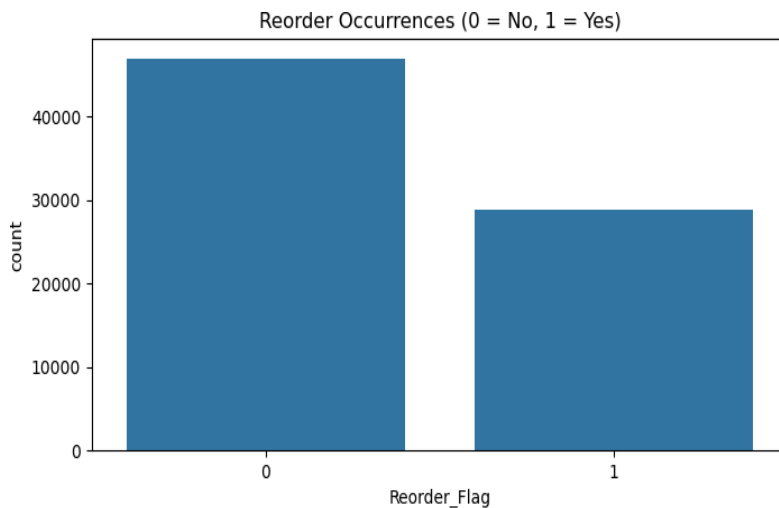


Fig. 7. Distribution of Reorder Occurrences

Figure 7 shows the frequency of reorder occurrences based on the Reorder_Flag variable, where 0 represents “No” and 1 represents “Yes.” The chart reveals that items marked with No reorder have a higher count compared to those that required reordering. However, a considerable number of items were flagged for reorder, signifying certain supplies reached critical levels and required replenishment to meet ongoing demand. This suggests that while inventory management is generally stable, there are instances where high consumption that creates a pressing need for restocking to ensure continuous supply availability during crucial periods.

Modify

In this phase, a total of 114 duplicate rows were removed and reducing the dataset from 76,000 to 75,886 unique records. To facilitate the application of classification algorithms, a binary target variable was created from the Inventory Re- Ordered attribute. This transformation converted the quantitative reorder values into a categorical indicator representing decision outcomes. A value of 1 denoted that an item was reordered, while 0 indicated no reorder action. To enhance the dataset’s predictive capability, the inventory sufficiency and shortage indicators were computed. These derived features assessed whether the current inventory levels could meet the expected demand. The Inventory_Shortage variable represented the difference between available stock and projected demand, where negative values indicated insufficient supply and a higher likelihood of reorder. Additionally, a binary flag Is_Low_Inventory, was introduced to simplify classification by marking instances of low stock conditions. To further enhance feature representation, a Demand Pressure Score was computed by combining multiple demand-related metrics into a single standardized indicator. The score was calculated as a weighted sum of the Simple Ratio Method (40%), Weighted Demand Index (30%), and Normalized Score (30%), resulting in a unified measure of demand intensity for each item.

For model training and evaluation, the Reorder_Flag variable was designated as the target variable, representing the binary classification outcome of whether an item required reordering. The dataset was then partitioned into training (80%) and testing (20%) subsets to ensure proper model validation and performance assessment. This split allowed the models to learn underlying patterns from the training data while maintaining an independent dataset for unbiased evaluation of predictive accuracy and generalization capability.

Before proceeding to model training and evaluation, a data preprocessing procedure was implemented to standardize and prepare the input variables. A ColumnTransformer was utilized to apply different transformations to numerical and categorical features. Specifically, numerical attributes were standardized using

the StandardScaler to normalize feature distributions and improve model convergence, while categorical variables were passed through without alteration using the passthrough method to retain their original encoding. This preprocessing step ensured that all input features were properly scaled and formatted, promoting consistency and stability across the various machine learning algorithms applied in this study.

Model

For the Modeling phase, multiple machine learning algorithms were employed to develop predictive models for medical supply reorder decisions. Both ensemble methods including Random Forest, Gradient Boosting, HistGradientBoosting, AdaBoost, XGBoost, LightGBM, and CatBoost are implemented. Classical algorithms such as Decision Tree, Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) were also applied. Each model was trained using the preprocessed training dataset and evaluated on the held-out test set to determine its suitability for supporting data-driven decision-making in disaster relief operations.

The predictive performance of each machine learning model was evaluated using the F1-score and ROC-AUC metrics. The F1-score measures the balance between precision and recall, indicating how accurately the models identify regions that require urgent medical supply reorders while accounting for both misses and false alarms. The ROC-AUC metric evaluates the models' ability to rank regions according to urgency by providing insight into predictive discrimination independent of a specific classification threshold. To further enhance model performance, Randomized Search with 5-fold cross-validation was applied for hyperparameter optimization for systematically identifying the best parameter configurations for each algorithm. Together, these metrics and optimization techniques offer a comprehensive assessment of model effectiveness and guiding the selection of the most reliable algorithms for decision support in natural calamity relief operations.

Assess

Table I Performance Comparison of Machine Learning Algorithms Used

#	Algorithms	F1-score	ROC-AUC	Rank
1	Random Forest	0.999913	1.000000	1
8	CatBoost	0.999913	1.000000	1
9	Logistic Regression	0.999740	1.000000	3
6	XGBoost	0.999913	0.999979	4
3	Gradient Boosting	0.999827	0.999970	5
5	AdaBoost	0.999913	0.999968	6
2	Decision Tree	0.999913	0.999913	7
7	LightGBM	0.999827	0.999883	8
4	HistGradientBoosting	0.999827	0.999882	9
11	SVM	0.946383	0.994896	10
10	KNN	0.956899	0.991523	11

Table 1 presents the comparative performance of eleven machine learning algorithms in predicting medical supply reorders during natural calamity relief operations. Random Forest and CatBoost achieved the highest performance, both with an F1-score of 0.9999 and ROC-AUC of 1.000, ranking first. Logistic Regression followed closely in third place, with slightly lower F1-score but perfect ROC-AUC. Other ensemble algorithms such as XGBoost, Gradient Boosting, AdaBoost, Decision Tree, LightGBM, and HistGradientBoosting also demonstrated high predictive capability, with F1-scores and ROC-AUC values near 1.000. Meanwhile, SVM and KNN showed lower F1-scores of 0.9464 and 0.9569, despite with high ROC-AUC. Overall, the table highlights that ensemble-based models, offer the most accurate and reliable predictions for decision-making in medical supply allocation.

Table III Optimized Performance of Machine Learning Algorithms Using Randomized Search CV (5-Fold)

#	Algorithms	Best F1- score	Rank
8	CatBoost	0.931610	1
6	XGBoost	0.928203	2
3	Gradient Boosting	0.927939	3
7	LightGBM	0.927723	4
4	HistGradientBoosting	0.925507	5
1	Random Forest	0.919562	6
5	AdaBoost	0.873800	7
2	Decision Tree	0.868372	8
9	Logistic Regression	0.821085	9
10	KNN	0.814562	10
11	SVM	0.805065	11

Table 2 summarizes the best F1-scores achieved by eleven machine learning algorithms after hyperparameter optimization using Randomized Search CV with 5-fold cross-validation. CatBoost emerged as the top-performing algorithm with an F1- score of 0.9316, followed closely by XGBoost (0.9282), Gradient Boosting (0.9279), and LightGBM (0.9277), demonstrating the effectiveness of gradient boosting ensemble methods in modeling reorder decisions. HistGradientBoosting and Random Forest also performed strongly by achieving F1-scores above 0.92. Traditional and simpler models, including AdaBoost, Decision Tree, Logistic Regression, KNN, and SVM, exhibited lower F1-scores ranging from 0.8051 to 0.8738 that indicating comparatively reduced predictive capability after hyperparameter tuning. Overall, the results highlight that optimized ensemble-based algorithms can provide a superior decision-making performance in predicting medical supply reorders during natural calamity relief operations.

Table III Best Model and Its Hyperparameters

	Best Parameters of CatBoost	Values
1	learning_rate	0.05
2	l2_leaf_reg	3
3	iterations	1000
4	depth	6
5	verbose	0
6	random_state	42

Table 3 shows an after hyperparameter optimization using Randomized Search CV and CatBoost was identified as the top-performing algorithm. The optimal hyperparameters for this model included a learning rate of 0.05,

L2 regularization (l2_leaf_reg) of 3, 1,000 iterations, and a tree depth of 6, with verbose set to 0 to suppress training output and random_state set to 42 to ensure reproducibility. These parameters balance model complexity and generalization by enabling CatBoost to effectively capture non-linear relationships and interactions in the dataset while maintaining robust predictive performance in disaster relief operations.

CONCLUSION

The application of the SEMMA methodology proved highly effective in guiding the development of machine learning models for medical supply decision-making during natural calamity relief operations. Its structured approach ensured proper data sampling, preprocessing, feature engineering, modeling, and performance assessment, resulting in robust and reliable predictive models. Among the algorithms tested, CatBoost emerged as the top-performing model by achieving the highest F1-score and demonstrating superior capability in accurately predicting reorder decisions.

For future studies, it is recommended to expand the experimentation to include additional machine learning algorithms to explore alternative predictive approaches that may further improve decision-making accuracy. Moreover, the use of hyperparameter optimization techniques, such as Randomized Search with cross-validation of 5-fold, should be applied systematically to enhance model performance and ensure generalizability across different disaster scenarios.

Ethical Approval

This study did not involve human participants or animal subjects. Therefore, ethical approval was not required for the conduct of this study.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this research. No financial, personal, or institutional relationships influenced the outcomes or interpretations presented in this study.

Data Availability

The data supporting the findings of this study are available to the community in the NCR, Philippines by upon request. No proprietary or confidential data were used in this research.

ACKNOWLEDGMENT

The researchers express their sincere gratitude to La Consolacion University Philippines and the Graduate School faculty for their guidance and support throughout this study. Appreciation is also extended to colleagues and peers who provided valuable insights, and to our families for their encouragement during the completion of this research.

REFERENCES

1. Rodríguez-Espíndola, O., Ahmadi, H., Gastélum-Chavira, D., Ahumada-Valenzuela, O., Chowdhury, S., Dey, P. K., & Albores, P. (2023). Humanitarian logistics optimization models: An investigation of decision-maker involvement and directions to promote implementation. *Socio-Economic Planning Sciences*, 89, 101669.
2. Bastani, P., Sadeghkhan, O., Bikine, P., Mehralian, G., Samadbeik, M., & Ravangard, R. (2023). Medication supply chain resilience during disasters: exploration of causes, strategies, and consequences applying Strauss and Corbin's approach to the grounded theory. *Journal of Pharmaceutical Policy and Practice*, 16(1), 99.
3. Bag, S., Pretorius, J. H. C., Gupta, S., & Dwivedi, Y. K. (2021). Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. *Technological Forecasting and Social Change*, 163, 120420.
4. Ma, K., Yan, H., Ye, Y., Zhou, D., & Ma, D. (2022). Critical Decision-Making Issues in Disaster Relief

- Supply Management: A Review. Computational intelligence and neuroscience, 2022(1), 1105839.
6. Al Alawi, S. J. S., Jamil, J. M., & Shaharane, I. N. M. (2022). Predicting student performance using data mining approach: A case study in Oman. Mathematical Statistician and Engineering Applications, 71(4), 1389-1398.
 7. Ma, Z., Zhang, J., & Gao, S. (2023). Research on emergency material demand based on urgency and satisfaction under public health emergencies. PLoS One, 18(3), e0282796.
 8. Florek, P., & Zagdański, A. (2023). Benchmarking state-of-the-art gradient boosting algorithms for classification. arXiv preprint arXiv:2305.17094.
 9. Ayodele, A. (2023). A comparative study of ensemble learning techniques for imbalanced classification problems. World Journal of Advanced Research and Reviews, 19(1), 1633-1643.
 10. El Guabassi, I., Bousalem, Z., Marah, R., & Qazdar, A. (2021). Comparative analysis of supervised machine learning algorithms to build a predictive model for evaluating students' performance.
 11. Li, J. (2024). Area under the ROC Curve has the most consistent evaluation for binary classification. PloS one, 19(12), e0316019.
 12. Owusu-Adjei, M., Ben Hayfron-Acquah, J., Frimpong, T., & Abdul- Salaam, G. (2023). Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. PLOS Digital Health, 2(11), e0000290
 13. A Ilemobayo, J., Durodola, O., Alade, O., J Awotunde, O., T Olanrewaju, A., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Ifeanyi, A., E Odezuligbo, I., & E Edu, O. (2024). Hyperparameter tuning in machine learning: A comprehensive review. Journal of Engineering Research and Reports, 26(6), 388-395.
 14. Ali, Y. A., Awwad, E. M., Al-Razgan, M., & Maarouf, A. (2023). Hyperparameter search for machine learning algorithms for optimizing the computational complexity. Processes, 11(2), 349