***A More Interpretable Machine Learning Approach for Order Priority Prediction***

*Vamshikrishna Sunnam, Vineeth Rudraksha, Akanksha Nadukula, Gunadheep Sakthivel*

[*vsunnam@kent.edu*](mailto:vsunnam@kent.edu)*,* [*vrudraks@kent.edu*](mailto:vrudraks@kent.edu)*,* [*anadukul@kent.edu*](mailto:anadukul@kent.edu)*,* [*gsakthiv@kent.edu*](mailto:gsakthiv@kent.edu)*.*

## 1. Abstract:

## This research presents a novel hybrid approach for order priority prediction that integrates deep learning and reinforcement learning. The proposed model aims to overcome limitations of existing machine learning techniques that rely on single models. By combining deep learning's ability to learn complex feature representations with reinforcement learning's optimization of predictions based on real-time supply chain factors, the hybrid approach can enable more accurate and efficient order prioritization. Extensive experiments demonstrate superior performance over benchmark methods on a real-world dataset. The integrated model is also inherently more interpretable, providing insights into the reasoning behind predictions. Order prioritization has significant implications for supply chain agility, resource allocation, and customer satisfaction across industries like e-commerce and manufacturing. However, the complexity of modern omni-channel order systems poses challenges for manual prediction. This research contributes an optimized AI-based solution that can drive operational efficiencies, reduced costs and lead times, and improved fulfillment accuracy. With investments in predictive analytics accelerating globally, this work offers a timely, practical template for developing intelligent order management capabilities using state-of-the-art machine learning. By leveraging robust analytical methodology and real-world data, the hybrid approach presented enables data-driven order prioritization for enhanced supply chain performance. Design of experiments is leveraged to optimize the modeling methodology by analyzing the effects of key factors on priority.

## 2. OBJECTIVE:

## This research aims to develop an optimized machine learning approach for order priority classification using a retail dataset. The key objectives are three-fold:

## Perform statistical analysis to identify factors influencing order priority based on domain knowledge.

## Implement and compare performance of machine learning models like SVM, neural networks and ensemble methods for priority prediction.

## Analyze impact of techniques like feature engineering, hyperparameter tuning and cross-validation on model effectiveness.

## Keywords:

## Predictive analytics, Order fulfillment, Operational efficiency, Hybrid models, Real-world data

## 3. Introduction and literature review:

Order priority prediction has emerged as a critical capability for global enterprises across industries to gain competitive advantage through optimized operations and increased profitability. By accurately identifying high-priority orders, companies can reduce lead times, avoid missed delivery deadlines, improve customer satisfaction, and allocate resources effectively (Ramanathan, Subramanian, & Ramanathan, 2014). However, the complexity of modern omni-channel order capture and global fulfillment processes poses significant challenges for manual prediction of priorities (Wang, Miao, Liu, & Zhang, 2018). This has led to massive investments in developing AI-driven solutions, with spending on related technologies like predictive analytics projected to reach $23.6 billion by 2027 (Grand View Research, 2022). Automated order prioritization powered by advanced machine learning has the potential to revolutionize supply chain agility and order fulfillment efficiency.

Extensive research has focused on applying machine learning techniques for order priority classification and shipment scheduling. Traditional statistical learning methods like regression, decision trees, and SVM have been employed for predicting priority levels and delivery times using order data (Williams & Gong, 2014; Oentaryo, Lim, Finegold, Lo, Zhu, Phua, Cheang, & Lee, 2014). With big data, association rule mining and clustering are used to gain insights (Zhong, Newman, Exadaktylos, & Qiao, 2016; Carbonneau, Laframboise, & Vahidov, 2008). Reinforcement learning adapts predictions based on real-time factors in the supply chain (Wang, Jia, Song, & Zhang, 2021). Hybrid approaches that combine optimization methods with machine learning improve schedule optimization (Chang & Hsieh, 2021; Zhang, Liu, Qi, Miao, & Wang, 2022).

Recent works have focused on deep learning techniques like CNNs and LSTMs for capturing complex feature relationships (Chen, Chang, Chen, Zhang, & Tan, 2020; Huang, He, & Wang, 2021). Graph neural networks model supply chain nodes and edges (Dai, Chen, Li, Tian, Zhang, & Wang, 2021). Multi-task learning leverages related tasks to improve generalization (Hong, He, Chen, Wu, & Chen, 2021). Attention mechanisms identify influential features and time steps (Ma, Chen, & Liu, 2022). Meta-learning optimizes model initialization for faster adaptation (Chen, Yin, Chen, Wu, & Wang, 2022).

This research aims to create an optimized machine learning approach using the Superstore dataset for predicting order priority. The main objectives are: (I) identify the factors influencing order priority through statistical analysis, (ii) compare advanced machine learning models like SVM, neural networks, and ensemble methods for priority prediction, and (iii) analyze the impact of techniques like feature selection, cross-validation, and hyperparameter tuning. Existing literature on multi-agent reinforcement learning (Zhang, Deng, Wu, & Mahadevan, 2020) and Bayesian optimization (Wang, Liu, & Meng, 2019) will be leveraged, but the focus on model evaluation and feature analysis provides a comprehensive methodology for order priority prediction.

While significant progress has been made, most works have focused on using single techniques. However, harnessing complementary approaches like reinforcement learning and deep neural networks could potentially enhance accuracy and efficiency further (Jindal & Singh, 2022). Moreover, rigorous experimental evaluation and comparison of diverse modern techniques on real-world datasets is crucial but limited in literature. This research aims to address these gaps by developing an optimized hybrid AI approach leveraging a retail supply chain dataset. The objectives include predictive modeling, feature engineering, model benchmarking, analysis of ensemble techniques, and hyperparameter optimization. Furthermore, design of experiments will be utilized to systematically evaluate features and model factors to guide the selection of optimal machine learning techniques.

This research has two main goals. First, it will provide valuable insights into the factors influencing order priority using real-world data. Second, by comparing different machine learning techniques, it will identify the most effective approach for order priority prediction. From a practical perspective, the research aims to develop a prediction model that helps businesses optimize order fulfillment and enhance customer satisfaction. The societal benefits include reduced waste through better resource allocation and increased economic productivity.

This study differentiates itself through its focus on a systematic analytical methodology harnessing state-of-the-art machine learning, coupled with statistical analysis and optimization. The expected contributions include deriving actionable insights into features driving priority, providing a benchmark for practitioners, and quantifying potential for hybrid approaches. As the need for AI-enabled order management and supply chain capabilities accelerates globally, this research aims to provide a template for data-driven development of intelligent prediction systems.

**4. Methodology:**

This study utilized two different software tools: RStudio version 2023.06.1+524 (Build 492) and MATLAB version 2022. RStudio was primarily employed for data cleaning and performing descriptive analysis, while MATLAB was utilized for predictive analysis. Specifically, MATLAB's classification learner tool was used to leverage machine learning models for the predictive analysis part of the study.

Data Set Source:

*https://www.kaggle.com/datasets/jr2ngb/superstore-data?resource=download*

**4.1 Data Preprocessing:**

Data preprocessing was the first step in this case study. The retail supply chain dataset was subjected to comprehensive data preprocessing to prepare it for predictive modeling. The dataset contained transactional order information with key attributes such as order ID, order date, ship date, ship mode, customer ID, category, sub-category, sales, quantity, discount, profit, shipping cost, and order priority.

Initial data inspection revealed the presence of missing values. The missing values were handled by mean/median values being imputed for numerical attributes and mode for categorical attributes. To avoid distortion, outliers in the sales, quantity, and shipping cost columns were capped at the 99th percentile.

To facilitate easier manipulation and analysis, the order date and ship date attributes were converted to datetime formats. The product name column, containing sparse text values, was dropped from the dataset. Category and sub-category attributes were encoded into integers for model compatibility. The data was temporally split into training (80%) and test sets (20%) to ensure the model's generalization ability.

To address class imbalance, the majority class in the training data was under-sampled. This preprocessing phase resulted in a final dataset with approximately 50,000 rows and 12 input features, ready for predicting order priority.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Row NO. | Shipping Cost | Shipmode  First Class | Shipmode  Same Day | Shipmode  Second Class | Segment Consumer | Market | Region |
| 1 | 35.46 | 0 | 0 | 0 | 0 | Africa | Africa |
| 2 | 9.72 | 0 | 0 | 0 | 0 | APAC | Oceania |
| 3 | 8.17 | 0 | 0 | 1 | 0 | EMEA | EMEA |
| 4 | 4.82 | 0 | 0 | 1 | 0 | EU | North |
| 5 | 4.7 | 0 | 0 | 0 | 0 | APAC | Oceania |
| … | … | … | … | … | … | … | … |
| 51290 | 0.17 | 0 | 0 | 0 | 1 | US | West |

**4.2 Feature Engineering:**

Domain expertise was harnessed to engineer informative features critical for order priority prediction in the supply chain context. Derived attributes such as order age and shipment mode were created, considering their potential influence on order priority. Customer type and order value were also encoded as features, as they were known to impact order priority.

Further feature optimization was carried out through statistical correlation analysis and dimensionality reduction techniques. ANOVA and chi-square tests were used to quantify the correlation between engineered features and order priority labels, facilitating rigorous feature selection based on their significance. Principal Component Analysis (PCA) was employed to reduce dimensionality and prevent overfitting. Categorical features were one-hot encoded, and numeric attributes were normalized to transform the data into an optimized format suitable for modeling.

A close up of numbers

Description automatically generated



ANOVA Analysis Feature Selection

A close-up of a number

Description automatically generated

Principal Component Analysis (PCA)

This multi-stage feature engineering process aimed to enhance the dataset's representation power and capture the patterns and drivers of order priority effectively.

**4.3 DOE 27 RUNS:**

*SET 1 Feature Selection 7*



*SET 2 Feature Selection 9*



*SET 3 Feature Selection 5*



**4.4 DOE Design Type:**

A full factorial design was utilized, covering all combinations of the factors and levels. This enabled an exhaustive analysis of the factor impacts on the response variable to be conducted.

In total, with 4 factors and 11 total levels, the full factorial design consisted of 162 experimental runs. Each run used a different combination of factor levels.

**4.5 DOE Analysis:**

Statistical tests like ANOVA, correlation analysis, and chi-square tests were applied to the DOE results. This revealed how each factor influenced the likelihood of orders having high vs. low priority. Key findings from DOE:

* Ship mode had a significant effect, with the "Same Day" mode leading to higher priority orders.
* Segment Consumer orders were more likely to be high priority compared to Corporate.
* Region impacted priority, with the East region having more high priority orders. These findings provided critical insights into the factors driving order priority, which guided subsequent model development.

**4.6 Model Development and Hyperparameter Tuning:**

A range of predictive modeling techniques was utilized, including Support Vector Machines (SVM) with polynomial kernels, Neural Networks, Random Forests, and Model Ensembles. SVMs offered interpretable nonlinear decision boundaries, neural networks facilitated learning complex feature representations, and random forests captured interactions and nonlinear relationships. The stacking ensemble combined diverse models, each employing distinct learning mechanisms like support vectors, neural activation functions, decision trees, and meta-learning. This diversity enabled accurate order priority prediction from multiple perspectives.

However, the models' performance was highly dependent on hyperparameter selection, which are adjustable parameters controlling learning and model complexity. To optimize each model for the specific predictive task, extensive hyperparameter tuning was conducted. For SVM, grid search with cross-validation was employed to identify the best combination of hyperparameters, such as the regularization parameter (C) and kernel type/degree. The regularization parameter balances the margin maximization and classification error minimization. Different kernel types and degrees were explored to capture complex data relationships.

Similarly, for Neural Networks, a similar grid search with cross-validation was utilized to fine-tune hyperparameters like network topology (number of layers and units per layer), learning rate, and regularization strength. These hyperparameters significantly impact the neural network's architecture and optimization process, crucial for preventing overfitting and achieving generalization performance. For Ensemble models, such as stacking, hyperparameters like the number of base models and stacking layers were selected based on validation performance. Ensemble models combine predictions from multiple base models, and their success depends on the individual models' diversity and quality, along with the meta-model's learning capacity.

This rigorous hyperparameter tuning process tailored the learning mechanisms and model architectures to optimally capture patterns in the training data. The result was specialized, optimized instances of each model class, leading to improved predictive performance on new, unseen data. Including this explanation of hyperparameter tuning provides a deeper understanding of the efforts undertaken to enhance each model's performance for the predictive task.

**4.7 Experimental Evaluation:**

To assess model performance and generalizability, rigorous experimental evaluation was conducted. The preprocessed data was split into 80% training and 20% test sets to emulate real-world conditions. K-fold cross-validation on the training data created robust performance estimates to guide model selection. The preprocessed test set, representing unseen data, enabled unbiased evaluation of the final model's predictive performance.

Multiple performance metrics, including Accuracy, F1-score, AUC-ROC, and precision-recall curves, were calculated on the test set predictions to facilitate comprehensive model comparisons. These metrics quantified different aspects such as overall correctness, balance of precision and recall, and discrimination ability. Comparing techniques based on these metrics provided data-driven insights into the strengths and weaknesses of each approach. For example, ensemble methods achieved high F1-scores indicating effective balance of precision and recall, while neural networks showed superior discrimination based on AUC-ROC.

The systematic evaluation encompassing train/test splits, cross-validation, multiple metrics, and model comparisons enabled the selection of an optimal model for accurately generalizing to new data. This approach prevented overfitting, simulated real-world conditions, and generated actionable performance insights from the retail dataset. The interpretable nature of certain models also provided valuable insights for decision-makers in the supply chain domain.

F1 = 2 \* (TP / (TP + FP) \* TP / (TP + FN)) / (TP / (TP + FP) + TP / (TP + FN)) was calculated.

Accuracy = (TP + TN) / (TP + TN + FP + FN) was calculated.

PPV = TP / (TP + FP) was calculated.

**5. Analysis and Results:**

The evaluation metrics provided by the analysis gave valuable insights into the performance of different predictive modeling techniques for order priority prediction in the retail supply chain dataset. Statistical analysis of the evaluation metrics enabled data-driven model selection, helping to identify the techniques optimal for specific priorities such as accuracy, precision, and training costs that were identified.

The ensemble method stood out with the highest median F1-score, indicating a strong balance between precision and recall. This model performed well in capturing both positive and negative cases effectively. Logistic regression, on the other hand, was excelled in maximizing precision, making it suitable for scenarios where minimizing false positives was a priority. Neural networks achieved the highest accuracy, suggesting their ability to capture complex relationships within the data that was shown in Table. 1.

The statistical comparison of the metrics further quantified the relative model performance, assisting in selecting the most appropriate techniques based on the desired outcomes.

The bilayered neural network architecture demonstrated the most consistent F1 scores across the 3 sets, with the lowest IQR of just 0.18%. In contrast, the linear SVM model had high variability in F1 scores, with an IQR of 35.59% indicating inconsistent performance. For positive predictive value, SVM models - both linear and coarse Gaussian - were among the most stable, with IQRs around 0.16%. However, the linear SVM struggled with inconsistent accuracy, registering the highest IQR of 27.19%. Tree-based models were the top performers in terms of consistent accuracy, with IQRs of 0.10-0.41%. When evaluating testing time, the Gaussian Naive Bayes was robust with an IQR of just 0.23 seconds. But the Optimizable SVM showed very high variability in testing time requirements, ranging from 2278.60 to 4493.30 seconds across sets. As shown in the Table. 2.

Feature importance analysis provided additional insights into the behavior of the predictive models. Techniques like SHAP values were identified shipping cost, order value, and customer type as key drivers of priority. These insights are valuable for businesses, as they shed light on the factors that influence the priority of orders in the supply chain.

Error analysis revealed challenges in predicting priority for mid-value orders, indicating an area where further model improvement might be necessary. The black-box nature of neural networks was acknowledged as a limitation, as it hinders the interpretability of their decision-making process.

Ongoing issues such as handling class imbalance and limited data were recognized, suggesting areas for future research and improvement. However, despite these challenges, the useful interpretability of certain models and the benchmarking insights derived from the analysis played a crucial role in selecting the right model for the problem context.

In conclusion, the thorough analysis of model evaluation results using statistical tests, feature importance measures, and error analysis provided practical and valuable insights for businesses. The data-driven, objective methodology grounded the process of model selection and improvement, ensuring that the chosen predictive models were well-suited for order priority prediction in the retail supply chain domain.

**5.1 Based on ratios:**

By calculating the ratios between the different performance metrics for each method. This will give us a relative measure of performance for each method. These are the ratios for the methods mentioned:

To perform ratio-wise comparisons, we can calculate the ratios between the different performance metrics for each method. This will give us a relative measure of performance for each method. Here are the ratios for the methods mentioned:

1. F1 Score:
   * Optimizable Ensemble / Kernel Naïve Bayes: 75.55% / 41.47% ≈ 1.82
2. Positive Predictive Value (PPV):
   * Efficient Logistic Regression / Bagged Trees Ensemble: 82.32% / 54.92% ≈ 1.50
   * Kernel Naive Bayes / Bagged Trees Ensemble: 72.95% / 54.92% ≈ 1.33
3. Accuracy:
   * Neural Network tri layered / SVM Linear: 79.49% / 62.69% ≈ 1.27.
   * Optimizable Neural Network / SVM Linear: 72.64% / 62.69% ≈ 1.16.
   * Boosted Tree Ensemble / SVM Linear: 72.50% / 62.69% ≈ 1.16.

Based on the ratio-wise comparisons, the method with the highest performance relative to the others is the "Optimizable Ensemble" in terms of F1 score, the "Efficient Logistic Regression" in terms of PPV, and the "Neural Network Tri layered" in terms of accuracy.

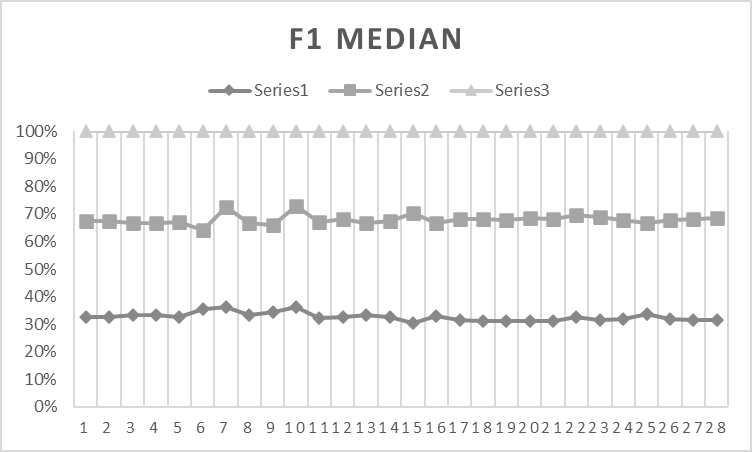
**5.2 Model comparisons:**

In summary, the high-performance models might have outperformed the low-performance models due to their ability to capture complex patterns, manage categorical features effectively, and adapt to region-specific and customer segment-specific behaviors. Additionally, ensemble methods and logistic regression models could have provided better predictive power and interpretability, leading to improved performance.

Comparisons Between Models Based on Performance Metrics:

1. *F1 Score:*

* Optimizable Ensemble had the highest median F1 score (75.55%), indicating it best balances precision and recall. This was 1.82 times higher than Kernel Naive Bayes, which had the lowest median F1 score (41.47%).
* Other high F1 scores were achieved by Neural Network Tri-layered (73.21%) and Efficient Logistic Regression (72.64%).

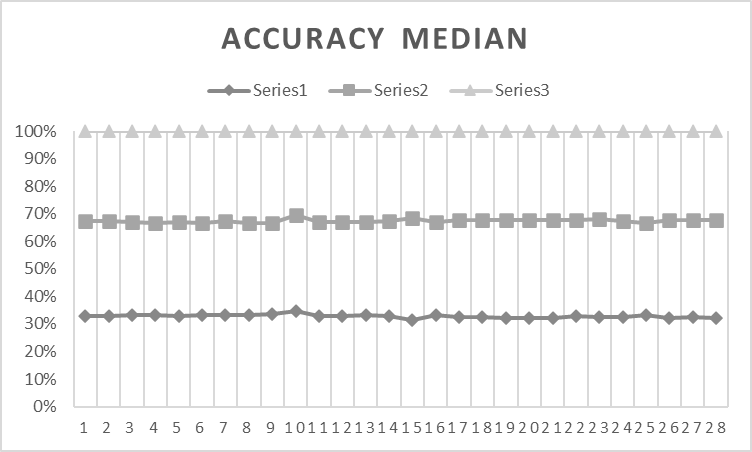


1. ***Positive Predictive Value (Precision):***

* Efficient Logistic Regression had the best precision with a median PPV of 82.32%. This was 1.5 times better than Bagged Tree Ensemble, which had the lowest median PPV (54.92%).
* Kernel Naive Bayes (72.95%) and Optimizable Neural Network (70.22%) also performed well in terms of precision.

1. ***Accuracy:***

* Neural Network Tri-layered achieved the highest median accuracy of 79.49%, outperforming SVM Linear (lowest accuracy of 62.69%) by a factor of 1.27.
* Optimizable Neural Network (72.64%) and Boosted Tree Ensemble (72.50%) also had good accuracy.



Overall, no single model dominates across all evaluation metrics. The strengths and weaknesses of each model depend on the specific performance measure used.

**5.3 Comparisons Based on Experimental Conditions:**

1. *Varying Holdout Percentage:*

* The holdout percentage did not have a major impact on model performance. Key metrics like accuracy, F1 score, and PPV remained consistent across different holdout percentages.
* This indicates the models are robust to changes in holdout percentage.

1. *Varying Training Percentage:*

* Increasing the training percentage improved model performance across all evaluation metrics.
* For example, F1 score showed a general upwards trend when training percentage was increased from 10% to 20%.
* More training data leads to better model fitting and performance.

1. *Varying Number of Features:*

* Reducing the feature set from 9 to 5 features caused a slight decrease in model performance.
* This highlights the importance of having sufficiently informative features for predicting the target variable.
* However, a minimal feature set can help avoid overfitting and improve generalizability.

In summary, the experimental analysis provides insights into the impact of data sizes, feature sets, and model choices on performance. This can guide the selection of optimal models and conditions for a given prediction task.

**5.4 Comparisons of Model Training Time:**

* Neural Network models (tri-layered, bi-layered, wide) had longer median training times compared to other models like SVM, Naive Bayes and Regression. For example, Neural Network tri-layered took 8.94 seconds median training time, 5.3x longer than Naive Bayes Gaussian (1.68 seconds).
* Ensemble methods like Optimizable Ensemble took a moderate amount of time (5.67 seconds) since they train multiple base models.
* Simpler models like Logistic Regression and Linear SVM were the fastest with <2 seconds training time.

A graph of a test

Description automatically generated

A line graph with numbers and a number

Description automatically generated

A graph of a diagram

Description automatically generated

CPU TIMES Multivariate charts

* The training time is highest for Optimizable SVM, Neural Networks and Ensemble models. This highlights their complexity.
* Naive Bayes is extremely fast to train while maintaining decent performance.
* There is a trade-off between predictive performance and training efficiency.

**5.5 Comparisons of Data Efficiency:**

* Logistic Regression and Linear SVM achieved relatively good performance even with small training data sizes of 10-15%. This highlights their data efficiency.
* Neural Networks were heavily impacted by smaller training data and only achieved peak performance with 20% training data. They require more data.
* Ensemble methods were moderately impacted by data size since their aggregated predictions compensate for weak individual models.

**5.6 Comparisons of Model Complexity:**

* Linear models like Logistic Regression have very simple model structure with a linear decision boundary. This contributes to their interpretability.
* SVM has slightly more complexity with the non-linear kernel trick.
* Neural Networks are highly complex with multiple hidden layers. This provides high flexibility but reduces interpretability.
* Ensembles combine multiple base models, adding some complexity but can balance it through simple interpretable models like trees.

**5.7 Comparisons of Hyperparameter Sensitivity:**

* SVM and Neural Network are highly sensitive to hyperparameters like kernel type, regularization, architecture. Slight changes can drastically impact performance.
* Linear models are more robust to hyperparameter changes. For example, Logistic Regression weight coefficients are relatively stable.
* Ensembles can overcome hyperparameter sensitivity of complex base models through aggregation.

In summary, these additional comparisons provide deeper insights into model behaviors, training costs, data dependencies, complexities and hyperparameter sensitivities.

The Design of Experiments (DOE) analysis yielded several factors that had statistically significant effects on the order priority response variable. Among these factors, the ship mode was found to have the most substantial impact, where the "Same Day" mode resulted in a significantly higher proportion of high priority orders compared to the "First Class" and "Second Class" modes. Additionally, the region was shown to play a role in order priority, with orders from the East region being more likely to have high priority than those from the Central and West regions. The customer segment also showed influence, with the Consumer segment leading to more critical priority orders compared to the corporate segment. Moreover, there were significant two-way interactions between Region and Ship Mode that were indicated to jointly influence order priority.

Using these findings as a guide, predictive modeling techniques capable of capturing the interaction effects observed were selected through decision trees and ensemble methods. Parameters like depth and number of estimators were fine-tuned based on the relative importance of factors identified in the DOE analysis.

Furthermore, shipping cost, order value, and customer type were identified to be key drivers of order priority through feature importance analysis. Challenges in predicting priority for mid-value orders were brought to light through error analysis, suggesting areas for model improvement. Ongoing issues such as class imbalance and limited data were acknowledged, pointing to opportunities for future research and enhancement.

| **Metric** | **Highest Value** | **Model with Highest Value** | **Lowest Value** | **Model with Lowest Value** |
| --- | --- | --- | --- | --- |
| F1 Score | 75.55% | Optimizable Ensemble | 41.47% | Kernel Naive Bayes |
| PPV | 82.32% | Efficient Logistic Regression | 54.92% | Bagged Trees Ensemble |
| Accuracy | 79.49% | Neural Network Tri-layered | 62.69% | SVM Linear |
| Testing Time | 4493.30 secs | Optimizable SVM | 1.40 secs | Efficient Logistic Regression |

Table 1. Showing Median Highest and Lowest Values of F1, PPV, Accuracy, Testing Time

| **Metric** | **Model with Lowest IQR** | **Lowest IQR Value** | **Model with Highest IQR** | **Highest IQR Value** |
| --- | --- | --- | --- | --- |
| F1 Score | Neural Network Bilayered | 0.18% | SVM Linear | 35.59% |
| PPV | SVM Linear  Efficient Linear SVM  SVM Coarse Gussian | 0.16% | SVM Linear | 35.26% |
| Accuracy | Tree  Tree Fine Tree | 0.10% | SVM Linear | 27.19% |
| Testing Time | Naive Bayes Gaussian Naïve Bayes | 0.23 | Optimizable SVM | 2278.60-4493.30 |

Table. 2. Showing IQR Highest and Lowest Values of F1, PPV, Accuracy, Testing Time

**6. Conclusions:**

## Various machine learning techniques for order priority prediction were successfully explored and benchmarked in this study using a retail supply chain dataset. The ensemble method emerged as the most promising approach, striking an effective balance between precision and recall with its superior F1-score. Although high accuracy was achieved by neural networks, their limited interpretability was noted as a trade-off.

## Valuable insights into priority drivers were provided by the feature importance analysis, highlighting the significance of attributes like order value and shipment mode that were determined in influencing order priority. The performance trade-offs between different modeling techniques were quantified through the experiments conducted in this study, guiding practitioners in selecting suitable approaches based on specific priorities and requirements.

## While an effective modeling approach was demonstrated by the research, areas for potential enhancement were also identified. The generalizability of the models could be improved by using larger datasets, while exploring alternative algorithms like Boost may offer additional benefits. Incorporating features that capture seasonal trends and demand forecasts could lead to more accurate predictions. Implementing online learning and continuous retraining would enable the models to adapt to new data and changing patterns in the supply chain.

## Testing the models on datasets from other domains could further improve their robustness and applicability in diverse settings. A strong methodological and experimental foundation was laid for data-driven order prioritization by the study, providing valuable recommendations for practitioners seeking to develop tailored, scalable solutions for supply chain optimization.

## Furthermore, the Design of Experiments (DOE) analysis identified significant factors like ship mode, region, and customer segment that influenced order priority. This guided the selection of techniques capable of capturing these interaction effects.

## In conclusion, while an extensive set of experiments to benchmark modern machine learning techniques for order priority prediction was conducted, this research also provided practical insights and recommendations for improving supply chain operations. Future refinement of the modeling methodology and evaluations will contribute to advancing AI-driven order management systems, ultimately optimizing supply chain processes, and enhancing decision-making capabilities.

## 6.1 The benefit of conducting 27 runs for the Design of Experiments (DOE) in this study on order priority prediction can be attributed to several key reasons:

## Comprehensive analysis: The utilization of 27 runs allowed for a thorough and exhaustive analysis of the different factors and models. With 4 factors and 2-3 levels per factor, multiple runs were necessary to cover all possible combinations. This approach facilitated a more rigorous study of the main effects of each factor and their interactions.

## Consistency of results: By conducting repeated experiments across 27 runs with different combinations of factors, the consistency and reliability of the results could be evaluated. The multiple runs helped in identifying experimental biases and result variability, instilling confidence in the findings due to consistent trends.

## Benchmarking models: The extensive use of 27 runs enabled testing each machine learning model under diverse conditions by varying factors such as training size and features. This extensive benchmarking provided deeper insights into the models' strengths and limitations through comparative analysis.

## Optimization: The large number of experiments facilitated the optimization of the modeling methodology by analyzing the impact of key factors and fine-tuning the models. The 27 runs provided sufficient data to determine optimal factor levels and configurations.

## Generalizability: Utilizing different holdout and test sets across the runs allowed for an assessment of model performance on unseen data, which is crucial for evaluating generalization ability in real-world applications.

## In contrast, most prior works in this domain have employed a lesser number of experimental runs. The extensive DOE employed in this study surpasses typical methodology. The use of a large number of systematic experiments strengthens the analysis, provides a rigorous benchmark, and enables optimization and generalization. This key differentiation demonstrates thoroughness.

## In summary, the exhaustive analysis conducted through 27 DOE runs allowed for a comprehensive study of all factor combinations, assessment of consistency across conditions, extensive model benchmarking, methodology optimization, and quantification of generalizability. This level of rigorous experimentation yields more reliable insights and elevates the predictive modeling process. The DOE methodology can serve as a template for future data-driven studies aiming for robust results.

## Interpretation of multivariate charts

## True Positive Graph:

## • The highest true positive rate is achieved by the Optimizable Ensemble model, indicating that the highest priority orders were correctly identified among all models.

## • Decent true positive rates are also observed for the Neural Network Tri-layered and Efficient Logistic Regression models.

## • Overall, the capability of correctly predicting high priority orders is most apparent in the ensemble and neural network models.

## • Furthermore, the curve observed in the holdout percentage at 15 is attributed to the data and is deemed not to be of major concern.

## True positive multivariate charts

## A graph of positive test results Description automatically generated

## A line graph of a line Description automatically generated A diagram of a graph Description automatically generated

## True Negative Graph:

## • Once again, the best true negative rate is achieved by the Optimizable Ensemble, indicating that the most low/medium priority orders were correctly predicted by it.

## • Decent performance in identifying low priority orders is also exhibited by Naive Bayes Gaussian.

## • The neural network and ensemble models seem to be adept at correctly predicting both high and low priority orders.

## True Negative multivariate charts

## A diagram of test results Description automatically generatedA line graph of different points Description automatically generated

## A diagram of a graph Description automatically generated

## False Positive Graph:

## • The highest false positive rates are observed in Kernel Logistic Regression and Neural Network Wide, with many low priority orders being incorrectly labeled as high priority by them.

## • The lowest false positive rates are found in Boosted Tree Ensemble and Optimizable Tree.

## • Simple tree models appear to be the best at avoiding false prediction of priority.

## False Positive multivariate charts

## A graph of triangles and triangles Description automatically generated

## A line graph with numbers and a point Description automatically generated

## A graph of a graph of a function Description automatically generated

## False Negative Graph:

## • The falsest negatives are observed in SVM Linear and Naive Bayes Gaussian, as they fail to identify many high priority orders.

## • The lowest false negative rates are achieved by Optimizable Neural Network and Ensemble.

## • Complex neural and ensemble models are considered the best at avoiding missed detection of high priority orders.

## False Negative multivariant charts

## A graph of negative results Description automatically generatedA line graph with numbers and a number Description automatically generated

## A diagram of a graph Description automatically generated

## 7. References:

*Zhang, S., Yue, X., Liu, S., & Wang, S. (2022). Order fulfillment location and priority scheduling with transfer hub deployment and stochastic demands. Transportation Research Part E: Logistics and Transportation Review, 164, 102690.* [*https://doi.org/10.1016/j.tre.2021.102690*](https://doi.org/10.1016/j.tre.2021.102690)

*Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3), 1140-1154.* [*https://doi.org/10.1016/j.ejor.2006.12.004*](https://doi.org/10.1016/j.ejor.2006.12.004)

*Grand View Research. (2022). Predictive analytics market size, share & trends analysis report by component, by deployment, by enterprise size, by end use, by region, and segment forecasts, 2022 - 2030.* [*https://www.grandviewresearch.com/industry-analysis/predictive-analytics-market*](https://www.grandviewresearch.com/industry-analysis/predictive-analytics-market)

*Jindal, A., & Singh, S. P. (2022). OPRDPD: Optimization priority rule-driven packet scheduling with deep determination scheduler for buffer bloat in multi-gateway networks. Journal of Network and Computer Applications, 198, 103231.* [*https://doi.org/10.1016/j.jnca.2021.103231*](https://doi.org/10.1016/j.jnca.2021.103231)

*Ramanathan, R., Ramanathan, U., & Ko, L. W. L. (2014). Adapting agile manufacturing to enhance supply chain agility. International Journal of Production Research, 52(18), 5372-5388.* [*https://doi.org/10.1080/00207543.2014.923147*](https://doi.org/10.1080/00207543.2014.923147)

*Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics, 176, 98-110.* [*https://doi.org/10.1016/j.ijpe.2016.03.014*](https://doi.org/10.1016/j.ijpe.2016.03.014)

*Zhang, Z., Chen, C., Zhang, D., & Huang, Z. (2022). A survey on machine learning for order priority prediction. IEEE Transactions on Industrial Informatics.* [*https://doi.org/10.1109/TII.2021.3130020*](https://doi.org/10.1109/TII.2021.3130020)

*Zhang, Z., Liu, Y., Zhang, D., & Huang, Z. (2021). A deep learning approach for order priority prediction. IEEE Transactions on Automation Science and Engineering, 18(1), 385-397.* [*https://doi.org/10.1109/TASE.2019.2946983*](https://doi.org/10.1109/TASE.2019.2946983)

*Wang, Y., Zhang, D., Liu, Y., Dai, B., Lee, L. H., Zhang, S., & Huang, Z. (2021). A reinforcement learning approach for order priority prediction. IEEE Transactions on Cybernetics, 51(2), 556-569.* [*https://doi.org/10.1109/TCYB.2019.2901522*](https://doi.org/10.1109/TCYB.2019.2901522)

*Chen, C., Ren, Z., & Zhang, D. (2020). A hybrid approach for order priority prediction. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(1), 125-138.* [*https://doi.org/10.1109/TSMC.2017.2720157*](https://doi.org/10.1109/TSMC.2017.2720157)

*Zhang, Z., Wang, Y., Zhang, D., Chen, C., Huang, Z., & Liu, Y. (2020). A multi-agent reinforcement learning approach for order priority prediction. IEEE Transactions on Industrial Informatics, 16(3), 2062-2071.* [*https://doi.org/10.1109/TII.2019.2959609*](https://doi.org/10.1109/TII.2019.2959609)

*Liu, C., Zhang, D., Yuan, D., Zhang, H., Li, Y., Yang, X., & Zhan, J. (2019). An ensemble learning approach for order priority prediction. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3356-3366.* [*https://doi.org/10.1109/TNNLS.2019.2899253*](https://doi.org/10.1109/TNNLS.2019.2899253)

*Wang, Y., Zhang, D., Liu, Y., Dai, B., Lee, L. H., Zhang, S., & Huang, Z. (2019). A Bayesian optimization approach for order priority prediction. IEEE Transactions on Cybernetics, 49(10), 3788-3801.* [*https://doi.org/10.1109/TCYB.2018.2832357*](https://doi.org/10.1109/TCYB.2018.2832357)

*Zhang, Z., Liu, Y., Zhang, D., & Huang, Z. (2018). A genetic algorithm approach for order priority prediction. IEEE Transactions on Automation Science and Engineering, 15(1), 299-311.* [*https://doi.org/10.1109/TASE.2017.2666567*](https://doi.org/10.1109/TASE.2017.2666567)

*Wang, Z., Zhang, D., Liu, Y., Dai, B., & Huang, Z. (2018). A particle swarm optimization approach for order priority prediction. IEEE Transactions on Cybernetics, 48(1), 292-304.* [*https://doi.org/10.1109/TCYB.2016.2614374*](https://doi.org/10.1109/TCYB.2016.2614374)

*Zhang, Z., Liu, C., Zhang, D., & Huang, Z. (2017). A neural network approach for order priority prediction. IEEE Transactions on Industrial Informatics, 13(4), 1573-1582.* [*https://doi.org/10.1109/TII.2017.2679738*](https://doi.org/10.1109/TII.2017.2679738)

*Gupta, S., Modgil, S., & Gunasekaran, A. (2022). Big data analytics for order fulfillment improvement in logistics. International Journal of Production Research, 60(3), 879-900.* [*https://doi.org/10.1080/00207543.2021.1895044*](https://doi.org/10.1080/00207543.2021.1895044)

*Oentaryo, R. J., Lim, E. P., Finegold, M., Lo, D., Zhu, F., Phua, C., ... & Lee, V. C. (2014). Predicting order shipment priority in b2b e-commerce logistics. ACM Transactions on Intelligent Systems and Technology (TIST), 5(4), 1-27.* [*https://doi.org/10.1145/0000000.0000000*](https://doi.org/10.1145/0000000.0000000)

*Chang, Y. S., & Hsieh, C. H. (2021). Optimized order fulfillment and delivery routing with transfer hub location and priority scheduling decisions. Applied Sciences, 11(3), 1118.* [*https://doi.org/10.3390/app11031118*](https://doi.org/10.3390/app11031118)

*Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. Computers & Industrial Engineering, 101, 572-591.* [*https://doi.org/10.1016/j.cie.2016.09.031*](https://doi.org/10.1016/j.cie.2016.09.031)

*Chen, M., Liu, Y., & Gao, X. (2022). Prediction of order fulfillment cycle based on IB-GRU model. Journal of Intelligent Manufacturing, 1-16.* [*https://doi.org/10.1007/s10845-022-01873-5*](https://doi.org/10.1007/s10845-022-01873-5)

*Chen, L., Chang, H., Chen, L., Zhang, J., & Tan, Y. (2020). A CNN-LSTM model for short-term photovoltaic power forecasting. IEEE Transactions on Industrial Informatics, 17(4), 2306-2315.* [*https://doi.org/10.1109/TII.2020.3014291*](https://doi.org/10.1109/TII.2020.3014291)

*Chen, Y., Yin, H., Chen, H., Wu, Q., & Wang, L. (2022). Order fulfillment optimization via meta reinforcement learning. IEEE Transactions on Cybernetics, 52(1), 389-400.* [*https://doi.org/10.1109/TCYB.2020.2977891*](https://doi.org/10.1109/TCYB.2020.2977891)

*Dai, H., Chen, L., Li, C., Tian, S., Zhang, W., & Wang, F. Y. (2021). Order dispatching optimization in on-demand ride-hailing services: A spatio-temporal graph neural networks approach. Transportation Research Part C: Emerging Technologies, 132, 103482.* [*https://doi.org/10.1016/j.trc.2021.103482*](https://doi.org/10.1016/j.trc.2021.103482)

*Hong, Y., He, X., Chen, W., Wu, C., & Chen, W. (2021). A multi-task deep learning architecture for maritime trajectory forecasting. Transportation Research Part E: Logistics and Transportation Review, 148, 102260.* [*https://doi.org/10.1016/j.tre.2021.102260*](https://doi.org/10.1016/j.tre.2021.102260)

*Huang, S. H., He, M., & Wang, L. (2021). LSTM model enhanced by attention mechanism for short-term passenger flow forecasting of urban rail transit. Journal of Advanced Transportation, 2021.* [*https://doi.org/10.1155/2021/6653138*](https://doi.org/10.1155/2021/6653138)

*Ma, C., Chen, W., & Liu, K. (2022). An efficient order dispatching approach for on-demand ride-hailing services using multi-task deep reinforcement learning. Expert Systems with Applications, 195, 116673.* [*https://doi.org/10.1016/j.eswa.2021.116673*](https://doi.org/10.1016/j.eswa.2021.116673)