Enhancing E-commerce Supply Chain and Shipping Efficiency with Machine Learning and Deep Learning Models

MSc Research Project
Data Analytics

Sathvika Bandarapu
Student ID: x23249927

School of Computing
National College of Ireland

Supervisor: Bharat Agarwal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sathvika Bandarapu
Student ID:	x23249927
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Bharat Agarwal
Submission Due Date:	12/12/2024
Project Title:	Enhancing E-commerce Supply Chain and Shipping Efficiency with Machine Learning and Deep Learning Models
Word Count:	6993
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	12th December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing E-commerce Supply Chain and Shipping Efficiency with Machine Learning and Deep Learning Models

Sathvika Bandarapu
x23249927

Abstract

E-commerce businesses are increasingly prioritizing with the efficient supply chain and shipping processes to meet rapidly growing customer demands while also mitigating the operational costs. Despite advancements in machine learning (ML) and deep learning (DL) techniques, existing works predominantly focuses on predicting mainly late delivery risk rather than estimating scheduled and actual shipping durations which is very critical aspect for enhancing supply chain resilience. To address the gap, this study aims to predict both scheduled and actual shipping durations using the DataCo SMART SUPPLY CHAIN dataset. This research aims to make predictions as a multi-output regression problem while also employing extensive preprocessing, feature selection, and hyperparameter optimization to mitigate the prediction error. Results shown that tree-based ML models outperform DL models in capturing complex, nonlinear relationships within tabular data, with XGBoost achieving the highest accuracy. These findings highlight the potential of advanced predictive analytics to optimize logistics, minimize delays, and improve decision-making in e-commerce supply chain operations.

1 Introduction

In the present dynamic e-commerce world, it is becoming increasingly important to have proper SCM (supply chain management) to drive business success. With the enhancement of consumer expectations of shipping speed and reliability, there is pressure building up on e-commerce enterprises to incorporate shipping optimization, and supply chain enhancement into their business models. When it comes to shipping the customer is dissatisfied and more importantly the overall profitability and competitiveness of a business. Considering such issues, more sophisticated approaches, including machine learning (ML) and deep learning (DL) approaches, are being implemented to enhance the reliability of shipping predictions and improve the supply chain for higher performance.

Several studies have examined the ability of machine learning in the supply chain of e-commerce. Previous studies show how demand forecasting, inventory control, and tracking and monitoring of shipments can be done using ML to allow companies to react faster to market fluctuations. Furthermore, the combination of ML and DL has been found to enhance the predictive models greatly in terms of effectiveness, particularly by providing a solution to dealing with multiple variables and complicated data. However, further work is still required to determine how the models can be improved concerning shipping time prediction including the identification of outliers, tuning of hyperparameters, and identification of the features that influence the predictions most.

1.1 Motivation

Shipping time estimation is a crucial factor in determining customer satisfaction and minimizing cost. The problem with traditional approaches to supply chain management is that they do not account for variables in real-time. The integration of machine learning models offers a significant enhancement in this respect, making it possible for firms to process large quantities of data, identify anomalies, and pinpoint the features that help make better predictions about shipping. The forecast can also avoid overstocking, cut on delays, and overall improve the supply chain, especially in the e-commerce business where speed of delivery is paramount.

1.2 Research Question

This study addresses the following research question:

How can machine learning and deep learning models be used to accurately predict shipping durations and optimize supply chain operations in the e-commerce industry?

The research question emerges from the increasing need for more accurate forecasts of shipping schedules and the operational performance of the supply chain. To address this research question, the DataCo SMART SUPPLY CHAIN dataset is employed, which contains detailed information regarding different supply chain activities such as shipping. Using the ML algorithms like Decision Trees, Random Forests, Linear Regression, XG Boost, and DL algorithms such as RNN, LSTM, and GRU (Gated Recurrent Units), this study will create predictive models to predict the *Actual Days for Shipping* and *Scheduled Days for Shipment*. These models will be later preprocessed, features will be selected and the models will be analyzed to identify important factors that affect shipping time the most.

1.3 Objectives

The core objectives of this study are as follows:

1. To analyze enhanced machine learning and deep learning techniques for accurate estimation of shipping duration, and improving the efficiency of the supply chain systems through overall effectiveness of supply chain management by minimizing delays and optimizing logistics processes.

2. To conduct a thorough analysis aimed at identifying the key attributes and features within the dataset that have the most significant impact on driving more efficient and accurate predictions of shipping times, which will, in turn, support better decision-making in supply chain operations.

3. To implement hyperparameter optimization across the machine learning and deep learning models to guarantee their best performance by fine-tuning them; the models will be able to give their best predictions for the given data.

In achieving these objectives, the study aims to provide a approach to how E-Commerce firms can enhance their shipping activities by integrating existing machine learning models to achieve optimal supply chain solutions.

The e-commerce environment is constantly changing and companies need to pay attention to their supply chain and delivery methods to stay competitive. Applying machine learning models leads to better accuracy of shipping prediction, less time on delivery, and increased customer satisfaction. This research will discuss how using predictive analytics and analyzing data with sophisticated techniques will help solve several issues that e-commerce SCM faces, alongside suggesting possible solutions to these problems.

2 Literature Review

2.1 Supply Chain Management (SCM) in E-Commerce

The management of the supply chain is critical to gaining a competitive advantage in the e-commerce industry because logistics is highly critical in satisfying the customers' expectations of timely delivery. A good SCM reduces cost, optimizes inventory, and demands forecasting, all of which are crucial in the dynamic e-commerce environment that requires agility to respond to changes in demand patterns (Wisetsri and Senarat; 2022). Organizational integration of technology in e-commerce firms can enhance the monitoring of supply chain inventory and provide real-time tracking of the products in a way that helps the firm have a better and more efficient supply chain system that directly affects customers' satisfaction and loyalty (Jana; 2021; Chin et al.; 2021). This complexity is further compounded in SCM in e-commerce by the need to develop more advanced models for demand forecasting and resources. Innovative data analysis and intelligent supply chain management help firms manage stocks and deliveries more efficiently to address the market's fluidity (Kalkha et al.; 2023). Still, predictive SCM models also assist e-commerce businesses to control their supply chain networks for disruptions, and lower costs while improving the network's general reliability (Sharma et al.; 2022).

2.2 Machine Learning-Based Supply Chain Management

ML (machine learning) has enhanced SCM in the e-commerce industry by enhancing critical processes like demand forecasting, inventory management, and risk assessment. Decision trees, random forests, and support vector machines are some of the traditional ML algorithms that assist e-commerce companies to analyse big data and plan the stock accurately to avoid stock-outs and overstocking (Jana; 2021; Tuli and Gupta; 2024). Furthermore, SCM is advanced by integrated ML models since they use several algorithms for better performance in specific scenarios. For instance, the cost complexity pruning with a decision tree enhances the prediction of supply chain risks such as delayed deliveries when using a hybrid model (Tuan; 2022). More recent algorithms such as CatBoost improve supply chain processing because it is efficient in managing categorical data to give more accurate delivery predictions and customer satisfaction (Sayyad and Varshney; 2024).

ML is also applied to cross-border SCM and International e-commerce which requires a high level of accuracy and flexibility. A multi-objective optimization model of demand forecasting based on ANFIS for cross-border e-commerce accurately predicts high volatility in demand while minimizing forecasting errors, which enables e-commerce firms to regulate unpredictable global demand and mitigate delays. Moreover, it is used for real-time risk analysis and management of risk situations to prevent disruptions to SCM, by identifying early signs of possible disruptions before they occur (Aljohani; 2023; Mittal and Panchal; 2023). Integrating ML models not only streamlines logistics but also fosters resilience in supply chains by leveraging real-time data and predictive analytics for a more agile, robust operation.

2.3 Deep Learning-Based Supply Chain Management

DL (Deep Learning) has proved to be very significant in the development of SCM since it offers complex models that can help in the analysis of large complex data in e-commerce. The LSTM networks for example have been used to improve efficiency in logistics by addressing task scheduling and resource management which is important in demand fluctuations and logistics issues in SCM (Issaoui et al.; 2021). These DL models help to identify temporal patterns in SCM data and enable accurate forecasts of demand and scheduling, which in turn, minimizes operational expenses and optimizes logistical outcomes (Yu et al.; 2024). Furthermore, LSTM models provide accuracy in high-responsive environments, which places DL as a suitable tool for SCM applications.

Comparisons made with traditional ML models suggest that DL methodologies including CNNs (Convolutional Neural Networks) offer higher performance than conventional techniques in some SCM applications because they can identify complex data features, including cyclical demand patterns (Yu et al.; 2024). Using a combination of CNN and LSTM as analyzed by Abosuliman and Almagrabi (Abosuliman and Almagrabi; 2021), it has been found that in the field of logistics management, prediction accuracy is high since CNN captures spatial features and LSTM deals with temporal features. This approach improves the accuracy of demand and delivery time estimation and serves as a foundation for logistics and inventory in e-commerce SCM.

2.4 Predictive Modeling and Risk Mitigation in Supply Chain Management

Risk management and overall SCM benefit from predictive modeling because it helps to anticipate and prevent risks, especially in environments like e-commerce. Real-time insights about potential risks are provided to the companies by ML-driven predictive models and this helps the companies to take precautionary measures against delays or supply chain disruptions (Aljohani; 2023). These models improve SCM flexibility and help companies to promptly respond to disturbances and minimize disruptions' impact, thus sustaining operations.

Apart from risk identification, predictive models also assist organizations in cost control. For example, deep CNN models have demonstrated good performance in detecting high-order and non-linear relationships in SCM data, and enable companies to have better risk evaluations and cost optimization (Mittal and Panchal; 2023). This proactive approach not only helps to build supply chain management flexibility but also helps in building a strong SCM that can withstand pressure and maintain the operations function.

2.5 Gaps in Current Literature

Although there is a great enhancement in the development of ML and DL in supply chain management (SCM), there are still research gaps, especially for the forecasting of e-commerce shipping time. The literature review shows that despite a vast body of research on demand forecasting and general logistics optimization, very few studies address the problem of shipping time prediction—the key concern in e-commerce SCM for customer satisfaction. Recent works mainly aim at general SCM goals without considering specific features such as hyperparameter tuning for accurate shipping time prediction that can help to reduce delay. Studies in these domains would extend the required level of predictive precision for optimizing e-commerce operations. Furthermore, LSTM and CNNs have been applied in SCM; however, very little research concentrates on employing these approaches to forecast both 'Actual Days for Shipping' and 'Scheduled Days for Shipment. Existing works mainly address risk prediction or demand forecasting rather than on developing end-to-end SCM solutions that address these specific shipping metrics. These gaps could be filled to develop effective integrated predictive models that can minimize the delivery time and the shipping schedule which is in line with the goal of the study, to improve the e-operations and decisions making in the SCM of e-commerce.

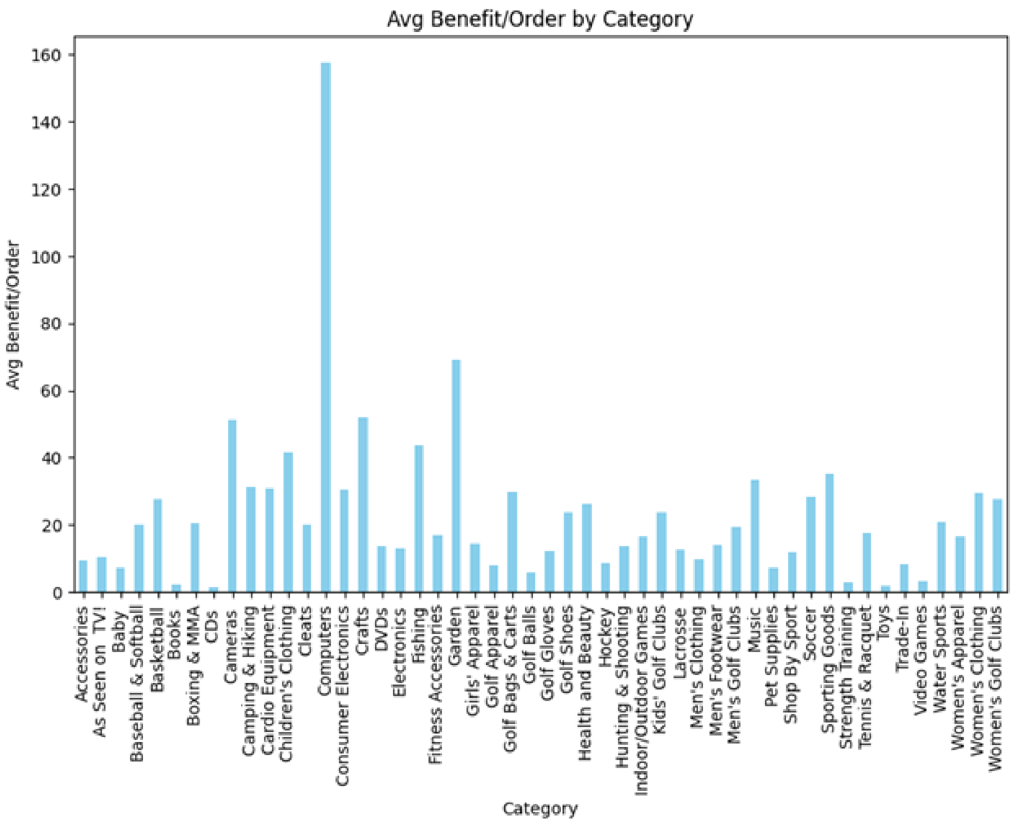
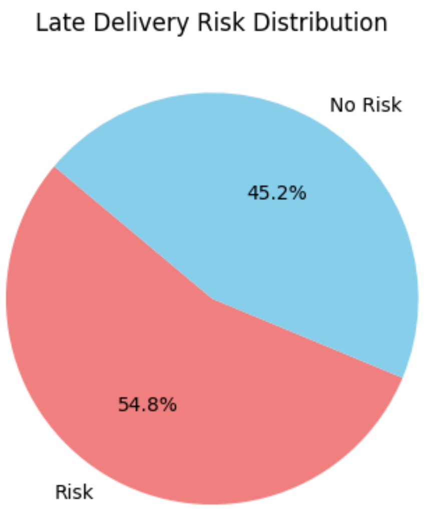


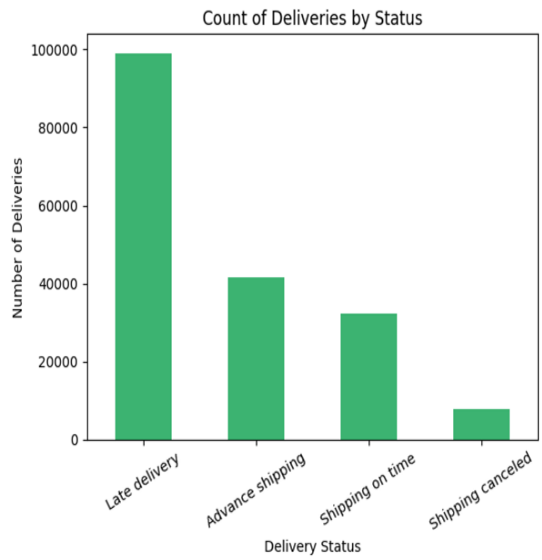
Figure 1: Average benifit recieved on various categories in dataset

3 Data Preprocessing

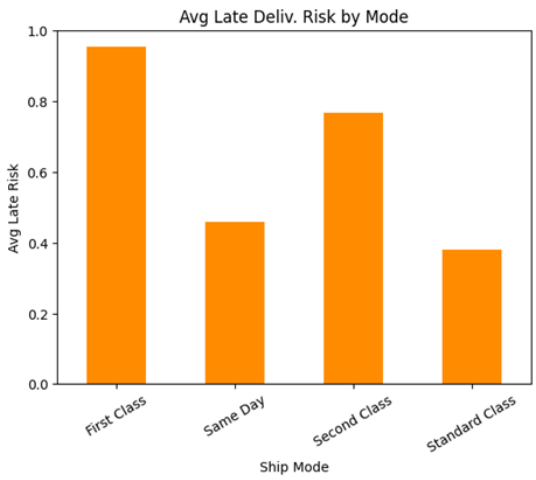
The preprocessing step involved preparing the supply chain management dataset for modeling. There are 1,80,000 rows and 48 columns in the dataset – a base large enough to feed



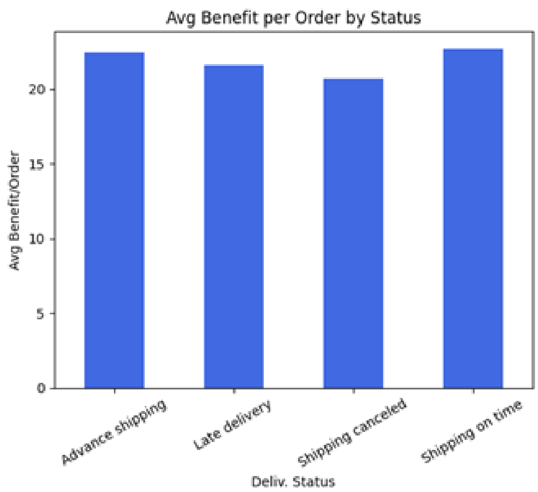
(a) Distribution of late delivery in the dataset



(b) Count of delivery by status



(c) Average late delivery risk by mode



(d) Average benefi per order by status

Figure 2: Various visualizations on supply chain dataset

strong and efficient machine learning models. The previous steps included data cleaning where missing values were treated, feature scaling & selection, feature engineering for categorical values, and defining target variables.

3.1 Dataset Overview and Missing Value Handling

The first exploration of the dataset provided a list of columns with missing values are checked. The *Customer Lname* and the *Customer Zipcode* contained 8, and 3 missing values respectively For the other two columns *Order Zipcode* and *Product Descriptions* there were a significantly large number of null values 155679 & 180519 respectively. Features like Order Zipcode and Product Description were excluded because they contained a high proportion of missing values. All records that were empty in the *Customer Lname* and *Customer Zipcode* were deleted to maintain the integrity of the remaining dataset.

3.2 Feature Selection and Reduction

The goal was to determine the best combination of features that could predict shipping durations, some unimportant columns were omitted. The excluded features included: *Benefit per order, Sales per customer, Delivery Status, Customer Fname, Customer Lname, Order Item Discount, Order Item Discount Rate, Order Item Profit Ratio, Order Profit Per Order, Order Region, Shipping date (DateOrders), and Order date (DateOrders).* Some of these features were omitted because they were not relevant to the target variables, or because they could not provide sufficient variation to be useful in the prediction models. This step helped to eliminate extra noise and made computations in the course of modeling less and more efficient.

3.3 Categorical Data Encoding

Categorical variables were converted into numerical form by using label encoding. This technique incorporated unique integers into the categories of the features in each one of them while making it suitable for machine learning models to accommodate categorical relationships.

3.4 Defining Target Variables

The study concentrates on forecasting two important target variables: *Days for shipping (real)* and *Days for shipment (scheduled)*. These variables are the most relevant ones in determining and improving shipping time. Due to the fact that the task involves the prediction of both targets at the same time, the model is cast as a multi-output regression problem, where the goal is to predict multiple dependent variables and for which there exists models that are designed to address this type of problem.

3.5 Final Dataset Preparation

After that, these preprocessing steps were performed and the dataset was ready for analysis. The cleaned data is accurate, consistent, complete, and formatted for the purpose of building sound and reliable predictive models.

4 Models and Methodology

The *Days for shipping (real)* and *Days for shipment (scheduled)* are continuous numeric variables, the multi-output regression approach of predicting them enters new challenges like handling features that are of different types, modeling non-linear relationships between the features, and achieving high accuracy of the predicted values for use in decision making of the supply chain. To overcome these challenges, both the machine learning and deep learning models were used. Each model was chosen because of its effectiveness in handling the various facets of the problem.

4.1 Decision Tree

The decision tree model was chosen to create the base model because of its simplicity and interpretability. When the regression problem is multi-output, then decision trees

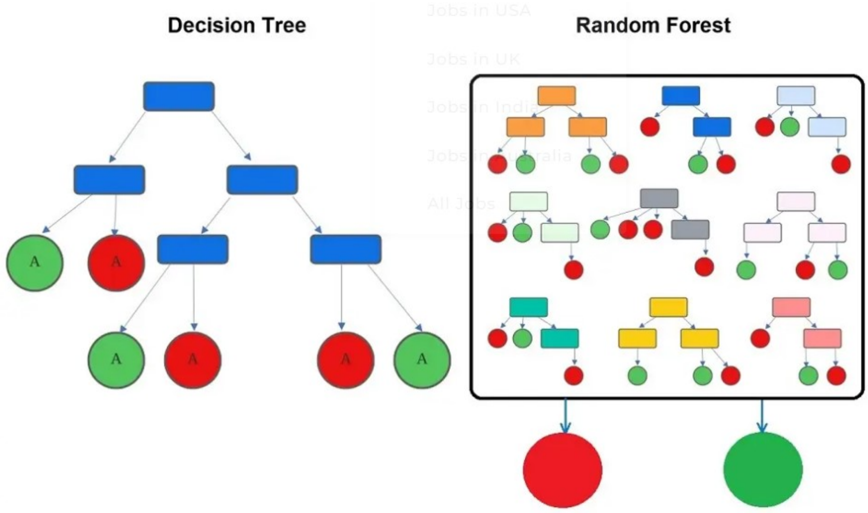


Figure 3: Basic architecture of Decision Tree and Random Forest (Breiman; 2001)

offer an explicit understanding of the nature of the impact that the features have on the output values. This interpretability is crucial for knowing which features, like the distance of the shipping location or volume of orders, most greatly contribute to shipping times. However, the decision tree is prone to overfitting and can only work independently in complex cases when faced with large datasets with many features.

4.2 Random Forest

Random Forest was chosen as an algorithm because it is an ensemble method that includes the decision trees to increase predictive power and avoid the formation of overfitting. The given model is good for datasets with numerical and categorical features as it does not require a separate data transformation. Random Forest has a high ability to capture feature interactions and it also works well in noisy data which fits the given problem. In multi-output regression tasks, it is used to combine different trees where it gives steady and better outputs for the target values.

4.3 Ridge Regression

Ridge regression was incorporated as a reference method for linear models regarding this multiple output regression problem. Even though the problem is intricate with non-linear interactions, Ridge regression provides an understanding of how much a linear model can capture the dependencies between the features and the target variables. The model uses L2 regularization which corrects for overfitting and is very effective when there is an issue of multicollinearity. This model is known for its computation efficiency and provides a comparison of model efficiency against complex models.

4.4 Gradient Boosting

Gradient Boosting was chosen because it is a technique that tries to minimize the errors of the previous models by making new predictions. This makes Gradient Boosting to be especially useful in large datasets where a little boost can greatly improve the prediction power. When it comes to the task of estimating shipping duration, Gradient Boosting

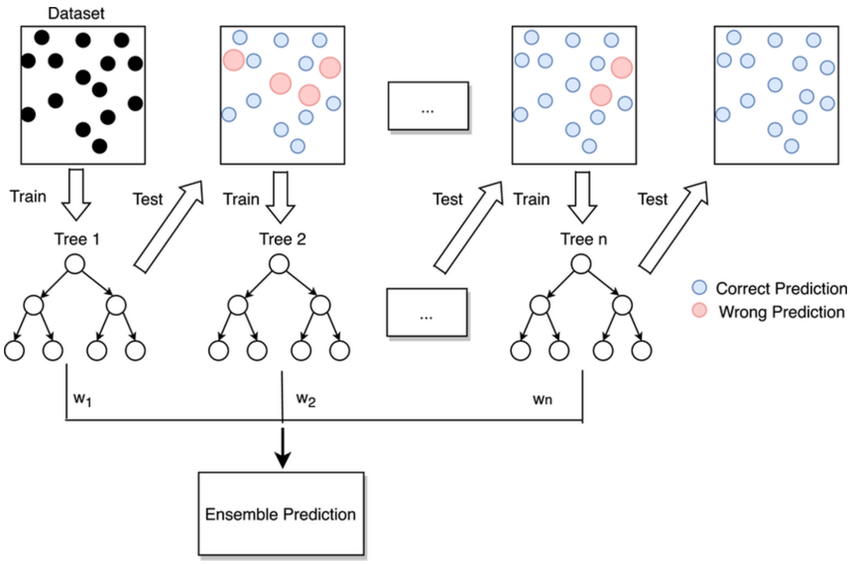


Figure 4: Basic flow of Gradient Boosting in making predictions using ensembling approach (Bentéjac et al.; 2021)

can incline non-linear interactions between features that are present. Furthermore, the model explains feature importance, which can be important when explaining shipping times.

4.5 XGBoost

XGBoost is an advanced implementation of Gradient Boosting and was selected due to its computational speed and scalability. It uses several optimization features like tree pruning and regularization which enhances the model's efficiency and accuracy. XGBoost is especially convenient to work with missing values that are typical in datasets with incomplete information. Custom objective functions in the case of multi-output regression make XGBoost provide accurate and stable predictions for *Days for shipping* and *Days for shipment*. The simplicity and efficiency of the model also suggest that it is well-suited to the large and high-dimensional dataset.

4.6 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks were applied to utilize temporal dependencies in the data. The nature of the presented dataset does not involve strictly consecutive time-series data but LSTMs are aimed to detect long-term temporal dependencies and latent temporal relationships. Consequently, LSTM networks may identify hidden patterns between shipping durations and other factors like order processing durations or shipping calendars. Hence, the LSTM model's characteristic of memorizing past information while excluding unrelated data enables it to capture important details that may be important in predicting both real and scheduled shipping days.

4.7 Gated Recurrent Units (GRU)

It is similar to LSTMs but is even more computationally, known as Gated Recurrent Units (GRU). GRUs are the function used to account for both temporal dependencies

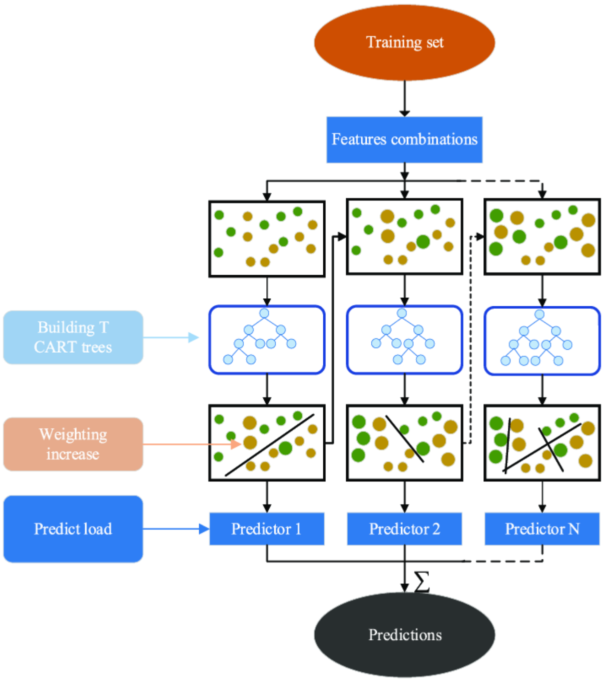


Figure 5: Basic flow of XGBoost in making predictions (Chen and Guestrin; 2016)

and faster training with the help of the model. Like LSTMs, the GRUs need to find dependencies of data within a sequence, but it is more effective for large sets due to the fewer parameters involved. The GRUs are more beneficial in cases where training is a bit time-consuming and it is suitable for identifying sequential patterns in shipment scheduling and actual shipping time.

4.8 Reasoning for Model Selection

The above models were chosen because each can help to solve different parts of the multi-output regression problem. Both Random Forest and XGBoost were used as they have fewer hyperparameters to tune compared to neural networks are well suited for higher dimensionality data, and can capture non-linear data relationships. These models are most suitable to be used in shipping duration predictions for mixed-type data.

On the other hand, deep learning models like LSTM and GRU were used in the model to determine if there are temporal patterns in the data that can be extracted that are hidden through feature engineering. LSTM and GRU networks are meant to find long-term temporal dependencies in sequences of data which could be useful when estimating shipping time.

Overall, the above models provide a rather holistic strategy to the problem, with each model providing its advantages in terms of model interpretability, predictive accuracy, computational complexity, and the capacity to discover nonlinear structures in the data.

4.9 Evaluation Metrics

To analyze the models performance, several evaluation metrics were assigned like

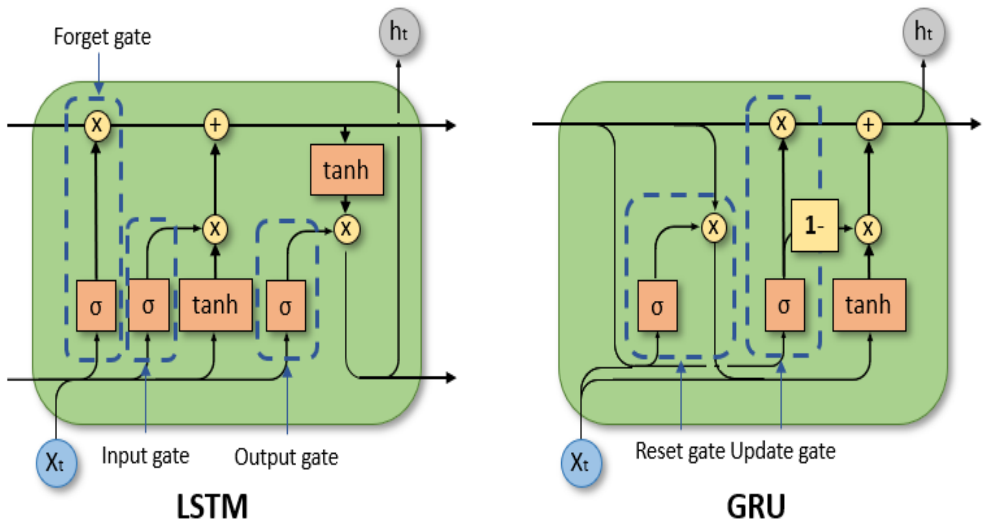


Figure 6: Architecture of LSTM and GRU models(Zargar; 2021)

4.9.1 Mean Squared Error (MSE)

MSE estimates the mean squared deviation between the predicted values and the actual one. It is especially appropriate when large errors or outliers must be suppressed due to the squaring of differences which leads to a higher penalty for greater deviation. When it comes to estimating shipping times, MSE gives a clear pointer of how accurate the projected shipping durations are to the actual shipping times; the smaller the MSE, the better performing the model. As it was expected, the Random Forest and XGBoost models were found to have the lowest MSE which implies its potential to optimize the deviations of predicted results from actual results.

4.9.2 Mean Absolute Error (MAE)

MAE on the other hand does the average calculation of the absolute difference between the predicted values and the actual values. It is a less noisy measure relative to the Mean Absolute Deviation because it does not overstate larger errors, and provides a clear sign of the average size of the prediction errors. MAE is particularly useful when the actual measurement of the prediction error in the same unit as the target variable (here shipping days) is required. In this analysis, two models were found to be best suited for predicting the shipping durations, namely Random Forest which was able to predict the actual values with an average error of 0.244.

4.9.3 R^2 Score

The closer the R^2 score to 1, the better the model reflects the variability of the target variable, while a lower score close to 0 points to the fact that the model does not predict the data trends. As for this multiple-output regression problem, R^2 is more useful than for a single-output case, as it generalizes over all the models at once, as to how well each fits the variability in the actual and expected days of shipment. Comparing all the models, the highest value of R^2 was achieved by XGBoost equal to 0.912, which is a sign that the used model can fit the existing nonlinear dependencies correctly.

4.9.4 Model Comparison

The analysis of these metrics can ensure an overview of the performance of each model. MSE shows large errors are being averaged in the model, MAE gives a direct measure of accuracy and R^2 shows how well the model explains variations. According to these parameters, it can be concluded that the ensemble methods, namely Random Forest and XGBoost, are better than other models in terms of model accuracy and error mitigation. Conversely, the LSTM and GRU had slightly higher errors of MSE and MAE, but they provide a fairly good approximation of the underlying patterning in the data.

Model	**MSE**	**MAE**	**R^2 Score**
Machine Learning Models			
Decision Tree	0.538	0.299	0.795
Random Forest	0.254	**0.244**	0.903
Ridge Regression	1.109	0.774	0.556
Gradient Boosting	0.788	0.488	0.699
XGBoost	**0.232**	0.249	**0.912**
Deep Learning Models			
LSTM	0.798	0.513	0.696
GRU	0.798	0.502	0.696

Table 1: Average model performance evaluation on various models where results

5 Results and Discussion

The results from the various Machine Learning and Deep Learning models are presented in this section.

5.1 Decision Tree

The Decision Tree model was moderately good with an MSE value of 0.538, MAE value of 0.299, and R^2 score of 0.795. This model is easy to understand and computes very fast but it did not perform as well as other methods such as the ensemble methods. A major weakness of Decision Trees is that it easily overfits especially when working with large numbers of features. In this case, the model probably did not generalize well, and therefore it exhibits a higher error than that of an ensemble technique such as Random Forest or XGBoost, the latter of which averages the result of many trees.

5.2 Random Forest

The Random Forest model did better with an MSE value of 0.254, an MAE value of 0.244, and an R^2 score of 0.903. The high accuracy of this model is because of its ensemble learning that embeds the idea of using many decision trees to make a final prediction to minimize the variance. As a result, Random Forest is capable of capturing intricate patterns and can handle the problem with different types of data. A high R^2 value in the model reveals that the proposed model can capture the major characteristics of the data and predict the shipping durations effectively.

5.3 Ridge Regression

Among all the machine learning models used in this study, Ridge Regression which is a linear model, produced the lowest results as indicated by MSE of 1.109 MAE of 0.774, and R^2 of 0.556. This model failed to perform well enough because it could not handle the non-linear relationship present in the datasets. Shipping times are dependent on several interacting and potentially non-linear factors that are difficult to capture with linear models. The high and moderate error values and low value of R^2 also indicate that the Ridge Regression model is not applicable for this multi-output regression problem since the interactions between the target variables and other features and non-linearity are high.

5.4 Gradient Boosting

Gradient Boosting model, the model achieved an MSE of 0.788, an MAE of 0.488, and an R^2 score of 0.699. It was better than Ridge Regression, but not as powerful as the Random Forest and XGBoost models. The concept behind Gradient Boosting is the ability to build trees in stages, each subsequent tree is trained to minimize the mistakes of the previous tree. Despite being a strong model, this approach tends to be more sensitive to overfitting, the tuning parameters are not optimal –which might explain its poor performance on this task. Nevertheless, it can work with complex data relationships, but other ensemble methods are more effective.

5.5 XGBoost

The best performance was achieved by XGBoost, the optimized gradient boosting, where the MSE value is 0.232, the MAE value is 0.249, and the R^2 score is 0.912. This model was consequently able to attain the least error measures, which signifies high predictive ability. XGBoost outperforms the other algorithms, particularly for large datasets with intricate features that this data set possesses. The overfitting is handled by regularization and weak learners are improved by boosting, the two factors that make it more effective than other models. The high R^2 score can be interpreted as XGBoost provided a good fit for predicting shipping duration variance and was the best-performing machine-learning model for this problem.

5.6 LSTM (Long Short-Term Memory)

The deep learning approach using the LSTM model gave an MSE of 0.798, an MAE of 0.513, and an R^2 score of 0.696. However, the implementation of the LSTM model which is a powerful model capable of learning complex relationships between variables failed to provide improved results compared to the best machine learning models. A possible explanation for this could be that the model failed to identify the dependencies between the features in the dataset and the dependency was not temporal. Moreover, most deep learning models demand more preprocessing and hyperparameter optimization than shallow learning models; which might have affected LSTM's generalization of this data.

5.7 GRU (Gated Recurrent Unit)

The next deep learning technique similar to LSTM was the GRU model, which had an MSE of 0.798, MAE of 0.502, and R^2 of 0.696. As a result, the computational complexity of GRU models is less than LSTM because of their less complex architecture, however, in this particular case both GRU and LSTM performed equally well. It is necessary to notice that the performance of the GRU model is also limited due to the same reasons as in the case of the LSTM model, related to the problem of addressing relationships within the data. This is just like LSTM, where it could have been an issue with the non-sequential setting of the problem, which prevents GRU from outperforming other models.

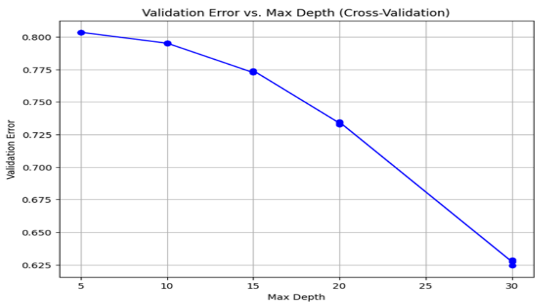
5.8 Discussion on Machine Learning Models Outperformed Deep Learning Models

From the results obtained, it is clear that the machine-learning models of Random Forest and XGBoost were superior to the deep-learning models of LSTM and GRU. This outcome can be attributed to several factors:

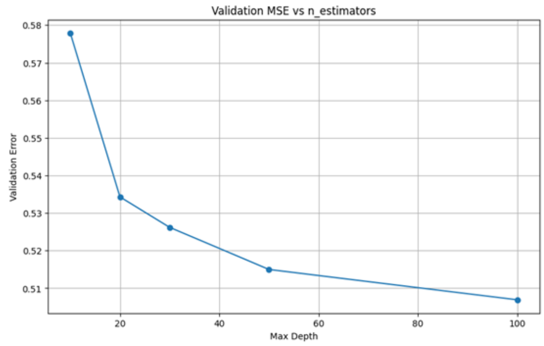
- **Tabular Data Suitability:** In machine learning, Tree-based models are Random Forest and XGBoost which are appropriate for tabular data This type of data has numerical, categorical, and structured features. Such models can process such data efficiently without the need for extensive transformation, or the use of large datasets that have been labeled. However, LSTM and GRU models, which are deep learning models, are more suitable for sequential or unstructured data and with less accuracy for tabular data where the interaction between features and the relationships between them are handled efficiently by tree-based models.
- **Dataset Complexity:** Although deep learning models are very good at capturing complex patterns in big data, the patterns in this data set were better aligned with machine learning models such as Random Forest and XG Boost. These models are better suited for multi-dimensional, and non-linear data, without excessive data preparation and parameter tuning.
- **Model Interpretability and Efficiency:** Random Forest and XGboost are other machine learning models, which are more understandable and computationally efficient. The fact that they are ensemble methods, they are better equipped to handle problems of overfitting and must have played a role in their enhanced performance over deep learning models that generally require more data and computational resources for optimal performance.
- **Generalization:** LSTM (Long short-term memory) and GRU (gated recurrent units) based deep learning models generally require big data and fine-tuning to provide generalization capabilities. It is possible that these models were overfitting or could not generalize as well because of the data set used here and, therefore, had lower predictive accuracy than the machine learning models that could perform well with the available data.

However, deep learning models when trained with large amounts of training data can yield highly accurate results but in this particular task, the machine learning models

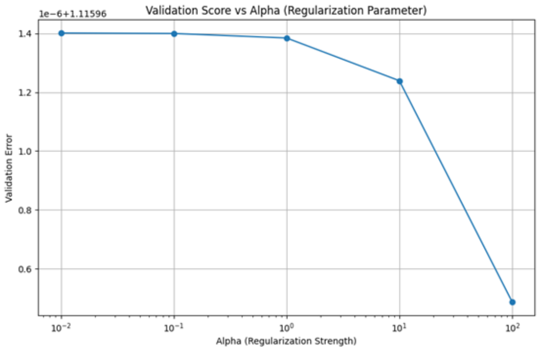
Random Forest and XG Boost were found to give better results because of their strong generalization capacity, the capacity to handle non-linearity and mitigates overfitting. These studies indicate that traditional methods of machine learning such as Random Forest and XGBoost are even more effective in the e-commerce segment for predicting shipping durations than deep learning approaches.



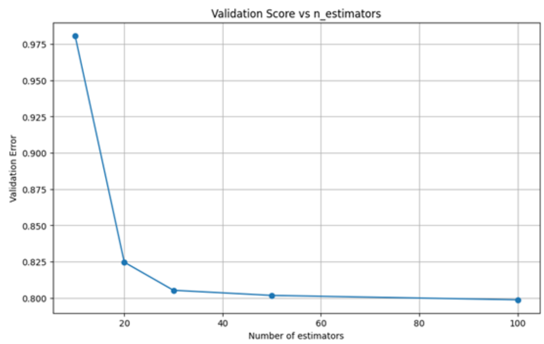
(a) Validation error vs max depth on decision tree



(b) Validation error vs number of estimators in random forest



(c) Validation error vs regularisation constant in ridge regression



(d) Validation error vs number of estimators in gradient boosting

Figure 7: Hyper-parameter tuning of various models on supply chain dataset using cross validation

5.9 Discussion on Variability in Real vs. Scheduled Shipping Predictions

Real shipping forecasts are always much more diversified because shipping times are very volatile and depend on many factors. While expected shipping durations are usually standard and directly related to the type of service, the shipping type or the destination, real shipping times are more sensitive to traffic jams, bad weather, slow processing in warehouses, and other logistic issues. This variability is a problem for predictive models because it complicates the range of factors that contribute to real shipping durations. This complexity is well captured by the dataset itself in which nearly half of the samples are classified as late deliveries. From table 2 can observe that real of shipping is more prone to error compared expected days of shipping and this underlines the fact that the real-world shipping process is irregular and non-linear. These challenges lead to higher prediction errors for real shipping durations as opposed to scheduled durations as was seen in the analysis.

Days for Shipping (Real)	Predicted (Real)	Days for Shipment (Scheduled)	Predicted (Scheduled)
6	5.16	4	3.99
2	2.55	4	3.99
5	4.35	4	3.99
4	3.68	4	3.99
5	4.01	4	3.99
4	4.38	4	3.99
2	3.73	4	3.99
3	2.79	2	2.00
6	5.86	2	2.00
5	4.66	4	3.99

Table 2: Comparison of Actual and Predicted Values for Shipping and Shipment Days on best performing XGBoost model. From results can observe that most of the error coming from days of shipping compared to scheduled as the shipping days are more prone to variability as can see in data almost 50% samples are late delivered. Similar trend appeared across all models.

On the other hand, the error of the scheduled shipping predictions is less for all the models because these values are less volatile. When a shipping schedule is established, for a specific destination or type of product, the delivery time is more or less fixed. This predictability means that machine learning models can generalize and predict scheduled durations well. However, real shipping values are not always reflected by the data because there are always factors such as short term disruption or delay due to operational reasons. These unpredictable elements bring noise and hence, increase the model error when estimating real shipping durations. This trend—where real shipping predictions have higher variance than the scheduled ones—was seen across all the models trained in the study, including traditional machine learning models such as Random Forest and XGBoost, as well as deep learning models such as LSTM and GRU.

6 Hyperparameter Tuning

The selection of hyperparameters is one of the critical tasks in the machine learning and deep learning model. In this study, hyperparameters for each model were tuned through hyperparameter grids to identify optimal parameters that would improve prediction accuracy. For the Decision Tree Regressor hyperparameters like `max_depth` and `min_samples_split` were tuned to control the depth of the tree and the minimum samples required to split an internal node of the tree respective. From the configuration analysis, the most optimal was found to be `max_depth = 30` and `min_samples_split = 5`. Likewise, for the Random Forest Regressor, the hyperparameter was the number of trees (`n_estimators`) of 100 found successful in tuning. The Ridge Regression was implemented using the `alpha` parameter for tuning, and the value of 100 was observed to be the best.

For the Gradient Boosting technique where an ensemble of weak learners is employed, the number of boosting stages, `n_estimators`, was varied and it was found that the model's best performance was obtained with 100 stages of boosting. This model was then wrapped in a MultiOutputRegressor to deal with multiple output targets. Similar to the model parameters tuning process of XGBoost, the number of boosting stages ranged between 10 and 500 with 500 trees as the best choice. The number of estimators was also fine-tuned on XGBoost and it led to a better R-squared score and lower errors than other models.

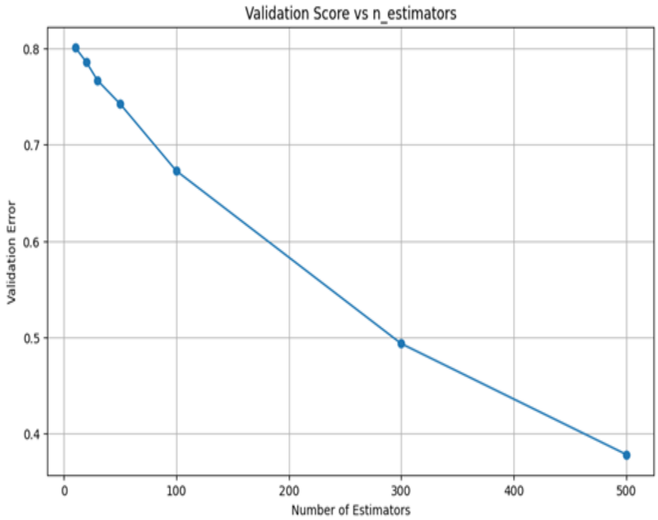
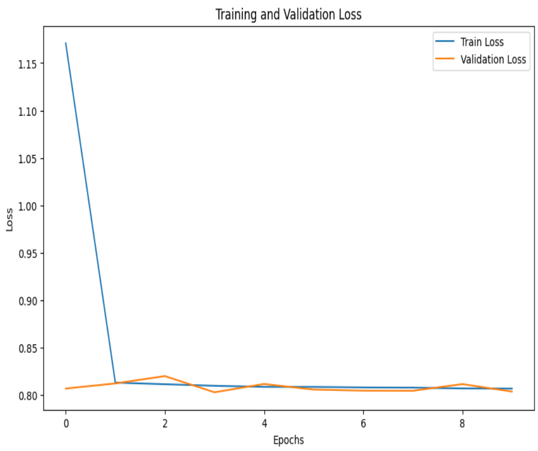
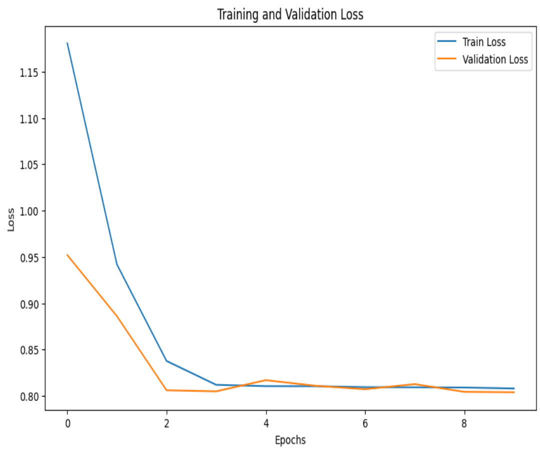


Figure 8: Validation error vs number of estimators in XGBoost model



(a) Training loss vs validation loss using LSTM model



(b) Training loss vs validation loss using GRU model

Figure 9: Loss plots of deep learning models

For other deep learning models such as LSTM and GRU, hyperparameter tuning was centered on whether the models had an efficient convergence during training. The models' performance was also evaluated through a validation dataset (10% of the training data) to check the validation loss to see whether the model was overfitting or underfitting. Unlike the tuning of parameters in deep learning including learning rate and batch size, the goal was to achieve good convergence of the model by looking at the validation loss. The validation data was useful on whether the model was being improved during training or if more tweaking was needed. While models like LSTM and GRU can handle such complex relations, they need more work on the tuning and validation checks as compared to the traditional machine learning models, which could be the reason why machine learning models were performing better.

7 Conclusion and Future Work

7.1 Conclusion

This research shows the benefits of applying ML and DL in estimating the predicted and actual duration of shipping. This research addresses the gap in predicting the both scheduled and actual shipping durations rather than just predicting last delivery risk. This research shows that conventional supervised ML models such as Random Forest and XGBoost achieve slightly better predictive performance than DL models using tabular data from the DataCo SMART SUPPLY CHAIN dataset when preprocessed properly and with the best hyperparameters for feature engineering and model selection. These conclusions show the necessity of the model selection depending on the data features, focusing on the capacity of tree-based algorithms in processing intricate feature interactions. The findings also include predictors of delays and offer insights for enhancing logistics to reduce such consequences while enriching literature on supply chain predictive analysis.

7.2 Future Work

The potential work for future research can include more data like weather conditions, transport conditions, etc., for better model prediction. The combination of Machine Learning and Deep Learning techniques was expected to enhance performance because the two techniques have their strengths. Further, constructing data models that are optimized for real-time prediction will help with changes in shipping requirements and disturbances. Explaining such models could also enhance understanding of specific factors affecting shipping durations and optimally adjust business strategies.

References

Abosuliman, S. S. and Almagrabi, A. O. (2021). Computer vision assisted human-computer interaction for logistics management using deep learning, *Computers & Electrical Engineering* **96**: 107555.

Aljohani, A. (2023). Predictive analytics and machine learning for real-time supply chain risk mitigation and agility, *Sustainability* **15**(20): 15088.

Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms, *Artificial Intelligence Review* **54**: 1937–1967.

Breiman, L. (2001). Random forests, *Machine learning* **45**: 5–32.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Chin, S. H., Lu, C., Ho, P. T., Shiao, Y. F. and Wu, T. J. (2021). Commodity anti-counterfeiting decision in e-commerce trade based on machine learning and internet of things, *Computer Standards & Interfaces* **76**: 103504.

Issaoui, Y., Khiat, A., Bahnasse, A. and Ouajji, H. (2021). An advanced lstm model for optimal scheduling in smart logistic environment: E-commerce case, *IEEE Access* **9**: 126337–126356.

Jana, A. K. (2021). Optimization of e-commerce supply chain through demand prediction for new products using machine learning techniques, *J Artif Intell Mach Learn & Data Sci* **1**(1): 565–569.

Kalkha, H., Khiat, A., Bahnasse, A. and Ouajji, H. (2023). The rising trends of smart e-commerce logistics, *IEEE Access* **11**: 33839–33857.

Mittal, U. and Panchal, D. (2023). Ai-based evaluation system for supply chain vulnerabilities and resilience amidst external shocks: An empirical approach, *Reports in Mechanical Engineering* **4**(1): 276–289.

Sayyad, N. and Varshney, G. (2024). Ai-based hybrid ml models for supply chain optimization, *Artificial Intelligence in Supply Chain Management* .

Sharma, M., Sharma, V. and Kapoor, R. (2022). Study of e-commerce and impact of machine learning in e-commerce, pp. 1–22.

Tuan, H. D. (2022). Hybrid decision-tree model for predictive supply chain management with an application in cross-border logistics, *Computational Economics* **60**: 337–358.

Tuli, S. and Gupta, N. (2024). Leveraging machine learning techniques for effective inventory management in e-commerce, *International Journal of Information Management* **58**: 102234.

Wisetsri, T. and Senarat, P. (2022). Supply chain resilience: Case study of e-commerce company, *International Journal of Supply Chain Management* **11**(5): 23–30.

Yu, B., Zhang, Y. and Zhang, S. (2024). Deep learning for logistics optimization: Real-time task scheduling and demand prediction using lstm, *Future Generation Computer Systems* **39**: 1–10.

Zargar, S. (2021). Introduction to sequence learning models: Rnn, lstm, gru, *Department of Mechanical and Aerospace Engineering, North Carolina State University* .