Machine Learning and Deep Learning to Analyse and Forecast an Operational Loss in E-Commerce Logistics

Acknowledgment

The research of *Machine Learning and Deep Learning to Analyse and Forecast an Operational Loss in E-Commerce Logistics* contributes to the Master of Science degree in Business Analytics which is taken by Mr. Nattawat Apichitpitipong, a master's degree student at the University of Kent, in the academic year of 2023/2024. The process of researching and writing this thesis has been both challenging and offering the author the opportunity to deeply engage with the loss in e-commerce logistics.

The author's personal working experiences in the e-commerce logistics industry have inspired the author to research the loss that occurred in the e-commerce logistics company. The objective is to explore the factors that could potentially impact the loss in e-commerce logistics activities, along with creating a machine learning or deep learning model to forecast the loss in e-commerce logistics. This research has broadened my understanding of the loss of the e-commerce logistics industry and the application of both machine learning and deep learning in e-commerce logistics.

The author is deeply grateful to Dr. Mahdi Shavarani, the supervisor of the author who provided beneficial suggestions on the research.

The author is delighted to share the findings of the research with others in terms of knowledge-sharing. The author hopes that the findings will contribute to the ongoing discussion on the loss in e-commerce logistics and inspire further research in this vital field.

Nattawat Apichitpitipong
12th August, 2024
University of Kent

Table of Contents

Table of Figures

Table of Tables

Abstract

E-commerce logistics has presently played a significant role in the world's economy. To maintain growth, defining the risk factor that potentially leads to a loss in e-commerce logistics is one of the important milestones. This research has utilised the e-commerce logistics observations recorded by the Brazilian marketplace for analysing the factors that influence a loss and creating an artificial model to forecast the loss in e-commerce logistics. The correlation matrix, Variance Inflation Factor (VIF), and Principal Component Analysis (PCA) have been conducted to reduce the multi-collinearity, dimensionality, and noise in the dataset. The Random Forest (RF) and FeedForward Neural Network (FNN) have been selected to forecast the loss in e-commerce logistics. The GridSearchCV method has been applied to find the best parameters for both RF and FNN models. The outcome has found that the RF model performs better performance than the FNN model. The Mean Decrease Accuracy (MDA) method has been utilised to prioritise the significance of the factors. The results emphasises that the customer zip code is the most influential factor in loss in e-commerce logistics activities, followed by product height or a matter of product size, product description which is located in the shipping label, and order item ID which is considered as the least influential factor in loss in e-commerce logistics activities.

1. Introduction

1.1 What is E-Commerce Logistics?

Nowadays, E-commerce plays a crucial role in the world economy. E-commerce changes how businesses operate and interact with customers online. E-commerce allows the business to expand beyond local boundaries to worldwide, as well as increase business sales and profits due to the additional revenue channels. The online platform industry also enhances customer satisfaction since the customer could complete the transaction easily compared to the offline purchase. Additionally, the growth of the e-commerce market leads to the n new business opportunity that allows the seller to sell the products on the additional online channel (Kim, 2019). The growth also directly affects the digital economy where e-commerce plays a key role in expanding the e-sale of the digital economy (Falk & Hagsten, 2015). Those impacts imply that the growth of the economy or the Gross Domestic Product (GDP) of the country has a significant positive correlation with the growth of e-commerce.

E-commerce logistics (ECL) has emerged as a critical component of the fast-growing e-commerce sector worldwide. The rapid growth of online retail, exampled by China's 32% surge in e-commerce sales to over $1 trillion in 2019, has correspondingly driven a 28% increase in logistics orders to 40 billion. This data underscores the symbiotic relationship between e-commerce and logistics. Companies that effectively manage their logistics operations often gain a competitive edge. However, ECL is full of challenges, including fluctuating demand, inventory management issues, and unforeseen disruptions like accidents or natural disasters (Xu, et al., 2019). These factors can significantly impact a logistics company's service quality or reduction in value or usability of goods, which leads to customer dissatisfaction (Larson, 1992).

1.2 What is Operational Loss in E-Commerce Logistics?

Operational losses could occur randomly and are difficult to predict due to many different factors that could influence them. In traditional ways of managing goods and deliveries, the biggest risk comes from the trouble of moving and distributing goods (Zheng, et al., 2020). The problems could be caused by transportation risks and disruptions caused by natural disasters, labour disputes, terrorist attacks, and infrastructure failures can lead to

delivery delays and loss of goods. These delays and losses can then result in lost sales and harm to the company's reputation to customers (Paul, et al., 2019). Apart from losing quantity, there's also a risk of quality loss. Poor quality in products or services can lead to wasted efforts in logistics and increase costs. To be precise, if a product is delivered to the customer in a low-quality and fails to satisfy a customer, the customer might refuse to receive the package. The company then ends up storing these unwanted products, leading to extra storage costs (Larson, 1992). Furthermore, in the e-commerce industry, low-quality goods can also lead to reverse logistics, which is the return of the product to the seller. These additional costs to re-deliver the products will eventually become an unnecessary cost for the ECL company. Reverse logistics isn't only for low-quality products, but also for poor services, such as late deliveries, which necessitate further actions to maintain customer satisfaction and loyalty (Jalil, 2019).

To mitigate customer dissatisfaction, many supply chain companies decide to compensate customers to restore their satisfaction (Yu, et al., 2019). However, this compensation can also be seen as a financial loss for the company. Since compensating isn't always beneficial financially, the company should carefully consider the costs of these recovery services to maintain the appropriate level since it could negatively impact the business's financial health (Grewal, et al., 2008).

With the fast growth of the online shopping industry, there are both big opportunities and risks (Yu, et al., 2021). Reducing operational losses could help these companies improve how they operate and their finances. But, to cut losses, it's crucial to identify the factors that contribute to these losses. The main objective of this study is to identify the factors that could potentially affect operational loss and build the forecasting model for predicting operational loss in ECL by applying both machine learning and deep learning techniques, including Principal Component Analysis (PCA), Random Forest (RF) algorithm, and Feedforward Neural Network (FNN).

2. Literature Review

2.1 The E-Commerce Logistics Risk Factors Analysis

2.1.1 The Risk Factors in E-Commerce Logistics

In e-commerce logistics, potential risks that could lead to operational losses are varied. Therefore, the empirical interview is conducted at Amazon and Flipkart warehouses in India to identify the risk event. The result of the interview shows that the risk event includes:

- Supply disruption risks: A leader in supply might face interruptions from uncontrollable natural disasters.
- Demand volatility risks: Unexpected spikes in demand that exceed supply can lead to significant sales losses and customer dissatisfaction.
- Legislation and compliance risks: There are various restrictions and laws related to cross-border transportation, as well as local transportation rules, labeling, and packaging.
- Information security risks: Protecting customer personal data against leaks, unauthorized access, and misuse is crucial for e-commerce companies.
- Contract obligation and legal risks: Adhering to quality standards, shipment terms, and payment terms as stipulated in contracts is necessary.
- Employee and outsource party fraud risks: The risk of internal and third-party information leaks can damage a company's reputation and benefits.
- Reverse logistics risks: The process that allows customers to return products can also lead to a high level of returned goods, which adversely affects the business.

According to the interviews, creating a clear risk mitigation strategy is challenging. an ANOVA test is applied to determine the significance of differences between these solutions, as same as, a 2-sample t-test, which is to evaluate the effectiveness of specific risk mitigation strategies. The findings suggest practical approaches for risk reduction including:

- Implementing technology such as GPS for tracking shipments, ERP or CRM systems for real-time information sharing, screening technologies to prevent customer fraud, and networks or databases for data backup and system failure prevention (Dutta, et al., 2019).

- Maintaining a flexible supplier and seller network along with reliable insurance to handle unforeseen shortages of products or services (Dutta, et al., 2019).

Beyond the above approaches, there is also the integration of risk control, which is mentioned in the existing empirical study to mitigate the risks and enhance the system's reliability. The study emphasises that the risks of logistic accidents, internal and external service quality, high cost and long-term payback, distribution risk, and risk-sharing and information management are required control from the ECL company to reduce the risk event in ECL. Additionally, the integration of third-party logistics also improves logistics capabilities and reduces logistic risk at the same time. As well as the application of Information Technology (IT) facilitates the information system for real-time data (Xiaoqiong, 2019).

The logistic risks are quantitatively estimated and evaluated by the Analytic Hierarchy Process (AHP) and Fuzzy Analysis. These methods refine the evaluation index system and weight system, making the risk assessment more objective and reasonable (Xiaoqiong, 2019).

2.1.2 The Risk Factor in Cold Chain and Fresh E-Commerce Logistics

The aforementioned risk events are identified in the general ECL. The specific ECL section, such as fresh ECL, has a different risk factor added from the general ECL. In the fresh e-commerce supply chain, where goods are perishable, not only are uncertainties in demand, supply disruptions, and coordination challenges important, but the perishability of goods can also lead to significant losses (Bai, et al., 2022).

Research in the fresh ECL has explored both centralized and decentralized decision-making models, assessing how sensitivity to freshness and variations in quantity lost affect optimal decisions and profits. It also involves comparing outcomes from both models and conducting a numerical analysis of cost coefficients in the supply chain (Bai, et al., 2022).

Given the critical nature of perishability and freshness, these factors directly influence the pricing of merchandise. Retailers adjust prices based on the perishability and freshness of products, emphasizing the importance of maintaining freshness in the fresh e-commerce supply chain (Bai, et al., 2022).

The other study, which focuses on the cold chain ECL, also find additional risk factors that consisted of supply risk, circulation processing risk, warehousing risk, transportation risk, distribution risk, and external risk. The circulation processing risk is found to be a distinct factor compared to risk events from general ECL and fresh ECL. The circulation processing risk refers to the potential hazards and uncertainties that arise during the handling and processing stages of fresh agricultural products within the cold chain logistics system. For instance, temperature monitoring or quality inspection risk (Zhao, et al., 2023).

The study is conducted using data collected from 128 questionnaires, with 104 valid responses. The study involved participants from various sectors, including academia, logistics enterprises, and fresh food platforms. The study analyses the risk factor by constructing a risk evaluation model using the Analytic Hierarchy Process (AHP) and Particle Swarm Optimisation (PSO) methods (Zhao, et al., 2023).

The result of the study ranks the potential of six risk factors, which the distribution risk is the most potential risk, followed by transportation risk, warehousing risk, supply risk, circulation processing risk, and external risk respectively. The main factors affecting distribution risk are delivery timeliness and delivery quality. For transportation risk, transportation timeliness and transportation information circulation are crucial. Storage turnover and temperature and humidity control are the main factors affecting warehousing risk. The quality inspection risk is found as critical in circulation and processing risk. In terms of external risk, the market environment and macro policy are identified as the greatest external risk of the cold chain ECL (Zhao, et al., 2023).

2.2 Machine Learning and Deep Learning Application in E-Commerce Logistics

The implementation of machine learning and deep learning in the e-commerce logistics industry has been applied for various purposes, including delivery date prediction, capacity planning, demand and supply forecasting, route optimisation, risk analysis, fraud detection, etc. Nevertheless, the application of machine learning for ECL operational loss forecasting seems to not be explicitly conducted in the ECL industry. This study then aims to conduct the analysis and forecasting on the topic of ECL operational loss by applying Variance Inflation Factor (VIF), eigenvalues, correlation matrix, and the PCA to analyse the relevant factors that could lead order to the operational loss, RF to forecast the operational loss with the machine learning classification and regression methods, and FNN to predict the operational loss based on the artificial neural network having a multilayer perceptron to analyse the data. The study also intends to find the best model of RF and FNN, and compare the results to find the best accurate model of forecasting the operational loss in ECL.

The VIF, eigenvalues, and correlation matrix are the methodologies to clean and prepare the data to obtain the most stability and interpretation of the model coefficient. VIF is widely used to identify collinearity in regression models. High VIF values indicate that a predictor has a strong linear relationship with other predictors. Those high VIF values are recommended to be removed from the model since the high collinearity could inflate the variance of the coefficient estimate and make the model unstable (Ekiz, 2021). Eigenvalues are a fundamental concept in linear algebra and are associated with matrices and linear transformations. Eigenvalues play a crucial role in PCA, a technique used for dimensionality reduction. The eigenvalues of the covariance matrix of the data indicate the amount of variance captured by each principal component (Howard & Robert, 1963). A correlation matrix is a table showing correlation coefficients between variables, where the values with 1 indicate a perfect positive correlation, -1 indicate a perfect negative correlation, and 0 indicate no correlation (Abbood, et al., 2021).

PCA has been widely used for data reduction, which the model reduces the dimensionality of large datasets by transforming them into a smaller set of principal components. This helps in simplifying the data while retaining most of the original variability. PCA also helps in noise reduction and handling multi-collinearity, by focusing on the principal

components that explain the most variance, PCA helps in reducing noise, irrelevant information, and removing highly inter-correlated variables in the datasets. This leads to cleaner and more interpretable datasets. These benefits help in interpreting complex datasets, and understanding of the contribution of different variables and observations (Abdi & Williams, 2010).

RF has been used for classification and regression tasks. The model combines several randomized decision trees and aggregates its predictions by averaging, which has shown excellent performance in settings where the number of variables is much larger than the number of observations. RF also can rank the importance of the variables by using Mean Decrease Accuracy (MDA) measurement. MDA is based on the idea that if a variable is not important, rearranging its values should not degrade prediction accuracy. RF is also capable of handling unbalanced datasets by down-sampling the majority class and growing each tree on a more balanced dataset. This makes them suitable for applications where one class is significantly underrepresented (Biau & Scornet, 2016). The RF is then chosen to predict the operational loss in ECL in this study.

FNN is a type of artificial neural network where connections between the nodes do not form a cycle. They are widely used due to their simplicity and effectiveness in various applications, such as prediction, optimization, etc. One of the key strengths of FNNs is the ability to generalize from the training data to unseen data. This creates effectiveness in applications where the model needs to perform well on new, unseen examples. FNN also has the ability to model a complex relationship and tabular data (Razavi & Tolson, 2011). These benefits support a decision to apply FNN to predict the operational loss in ECL.

To achieve the most accurate model of RF and FNN, appropriately adjusting the parameters of both models is strongly required. GridSearchCV is then selected to play a key role of finding the best parameters. GridSearchCV helps in finding the optimal hyperparameters for a machine learning model. It could find the nest settings of the model by exhaustively searching over a specified parameter grid and evaluating each combination using cross-validation (Gill & Rupesh, 2023).

2.2.1 Fraudulent Detection with Random Forest and Neural Networks

There are several existing studies on the application of machine learning or deep learning in ECL, such as demand forecasting, route optimising, etc. Fraud detection is also famous as the application of machine learning. The study of ECL fraud detection presents the comparison of effectiveness in forecasting fraudulent activities in ECL between machine learning and deep learning. The data used in for training the model is collected between 2018-2023. Many machine learning and deep learning models are implemented to compare and find the out-performed model, including decision tree, random forest, support vector machine (SVM), convolutional neural network (CNN), k-nearest neighbor (KNN), bidirectional long short-term memory (BiLSTM), bidirectional gated recurrent unit (BiGRU), naïve Bayes, etc (Zhang, et al., 2023).

The results show that deep learning methods do not consistently outperform traditional machine learning methods in fraud detection. This inconsistency is often due to the large quantities of data required for deep learning models to fully learn features. For instance, in a comparative analysis of credit card fraud detection, a random forest model showed slightly better accuracy than a deep neural network. However, in another study, BiLSTM and BiGRU outperformed naïve Bayes, Adaboost, random forest, decision tree, and logistic regression (Zhang, et al., 2023).

The study goes through the challenge of an imbalanced dataset, in which fraudulent transactions are typically much less frequent than legitimate ones. This imbalance can lead to biased models with poor performance in the minority class. These challenges could be addressed by various methods, such as re-sampling techniques and weighted training could be applied to enhance the model performance and achieve more accurate prediction (Zhang, et al., 2023).

2.2.2 Capacity Planning with Gaussian Process Regression

Capacity planning is one of the most crucial activities in ECL. Since the demand and volume of ECL have increased exponentially, the number of deliveries in ECL also increased simultaneously. This surge makes it difficult to meet delivery deadlines, which subsequently impacts customer satisfaction and logistics performance, necessitating more

effective capacity planning to plan both manpower and resources appropriately. The study focuses on the fleet's capacity that the route commences on the cross-dock depot in a particular time slot (Kup, et al., 2023).

the paper utilizes a comprehensive dataset that includes daily aggregated deliveries, encoded categorical values, and various raw and statistical features. The continuous update with real-time shipment data ensures that the model remains adaptable and accurate in predicting delivery capacities. The Gaussian Process Regression (GPR), Extreme Gradient Boosting (XGBoost), Linear Regression (LR), Random Forest (RF), Multilayer Perceptron (MLP), and Support Vector Regressor (SVR) have been used for developing and evaluating. The study compared the results of each model to find the best model for predicting the daily delivery and capacity of the fleet at the cross-dock depot in each time slot (Kup, et al., 2023).

The GPR model outperforms other state-of-the-art regression methods and is updated daily using real-time shipment data. This ensures adaptability to unexpected events and special occasions, maintaining accurate and efficient daily delivery operations (Kup, et al., 2023).

Regarding the observed literature, there are numerous studies of applying machine learning or deep learning in the ECL industry. Nonetheless, there are no explicit studies of operational loss in ECL. Most of the existing studies are either on the topic of risk factors analysis or the prediction of ECL components, such as demand, capacity, sales, etc. The study of fraudulent activity detection is the most similar to the study of operational loss prediction since the dependent variable is binary and the topic of both studies is the lost perspective of ECL. From the risk factor perspective, the existing studies focus on the risk event rather than the operational factors. To illustrate, those studies analysed and found that the risk events that could impact the ECL performance are demand volatility, supply disruption, fraud risk, warehousing risk, and others. These risks seem to be uncontrollable since the risk possibly implies an error in processing, which could occur by accident or unexpected incident. Furthermore, the solutions to mitigate these risks are also not solid. More precisely, some of the given solutions from existing papers recommend applying the Internet of Things (IoT) to the ECL system to track and share real-time data for

increasing the accuracy of processes in ECL or building strong connections with third-party pools for preventing a supply disruption. These solutions are great for risk mitigation, but they do not clearly identify the root cause of the problem. Consequently, this paper aims to identify the explicit operational factor, including the customer's area, product's price, freight value, product's size, product's review, etc., that could potentially lead to the operational loss in ECL by implementing PCA, and forecast the operational loss by comparing and finding the best model of RF and FNN to mitigate the risk of operational loss and prepare for the best-lost prevention solution.

3. Methodology and Findings

3.1 Data Source

The data used in this study is secondary data which is provided by the largest department store in the Brazilian marketplace, Olist, on the Kaggle website. The data is gathered between 2016 and 2018. The data contains 99,441 observations and 52 variables with 8 different datasets, including review dataset, orders dataset, payment dataset, product dataset, item dataset, seller dataset, customer dataset, and geolocation dataset (Olist & Sionek, 2018). The reason for choosing Olist's dataset is because of a multiple dimension of order aspect. Since the study aims to research the factors affecting lost order, numerous perspectives of the order are essential to analyse the factor, which also leads to higher accuracy in the lost order prediction. All of the 8 different datasets have different foreign keys that could merge all datasets. The following figure represents the data schema of the datasets.

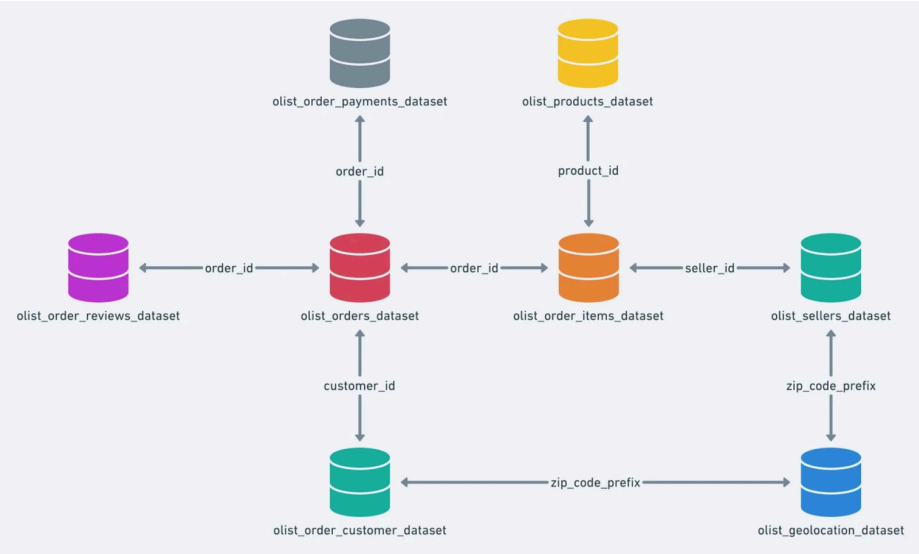

Figure 1: Data Schema of datasets (Olist & Sionek, 2018)

3.2 Data Cleaning and Preprocessing

3.2.1 Merging Data

All datasets are filtered and merged to create the main dataset for the study. The main dataset consists of customer zip code, order item ID, price, freight value, payment value, product name length, product description length, product weight (g), product length (cm), product height (cm), product width (cm), seller zip code, review score, order status, payment channel, and product category. After merging all datasets, the observations increased to 119,142 entries due to the fact that one purchased order could have several products, as well as different payment channels. To illustrate, the order could contain product A which is paid by credit card, and product B which is paid by voucher. To conduct a deep analysis of a factor influencing the loss, the data is supposed to be most detailed.

After merging the datasets, the order status is treated as the dependent variable, while other columns are treated as independent variables. The order status has been explored and found that the “delivered” status has a significant portion of the order status variable, 94,678 observations. The bar plot showing the count of order status is displayed in the below figure.

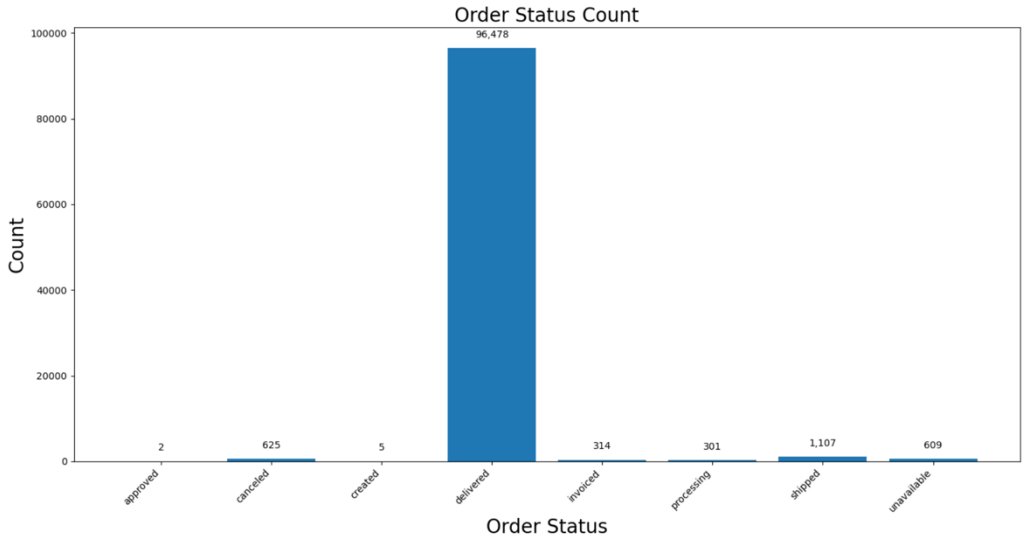

Figure 2: Order Status Count

3.2.2 Preparing Data

In order to clean the database for machine learning and deep learning models, the observations with null values have been removed, and the additional variables that are duplicated and irrelevant to the objective, such as timestamps, review's title, seller city, etc., are also deleted. The variables with categorical data or string also require to be converted to nominal data due to both RF and FNN requiring mathematical operations so that the algorithm could understand and process effectively (Anitha, et al., 2023). The columns that are converted to nominal data are order status, payment channel, and product category. The classification of the converted nominal data is shown in the following tables.

Order Status Variable

Order Status	Labels	Meaning
Delivered	1	Successful Delivered
Shipped	0	Unsuccessful Delivered
Invoiced	2	Non-logistics status
Processing	2	Non-logistics status
Canceled	0	Unsuccessful Delivered
Unavailable	2	Non-logistics status
Approved	2	Non-logistics status

Table 1: Order status nominal data conversion

Payment Channel Variable

Payment Channel	Labels
Credit Card	1
Voucher	2
Boleto	3
Debit Card	4

Table 2: Payment channel nominal data conversion

The product category variable is also converted to nominal data, which the numbers from 1-71 are labels representing each category, and are presented in the below table.

Product Category Variable

Category	No.	Category	No.	Category	No.
housewares	1	fashion_underwear_beach	25	fashion_male_clothing	49
perfumery	2	christmas_supplies	26	cine_photo	50
auto	3	fashion_bags_accessories	27	furniture_living_room	51
pet_shop	4	musical_instruments	28	art	52
stationery	5	construction_tools_lights	29	food_drink	53
furniture_decor	6	books_technical	30	tablets_printing_image	54
office_furniture	7	costruction_tools_garden	31	fashion_sport	55
garden_tools	8	home_appliances	32	la_cuisine	56
computers_accessories	9	market_place	33	flowers	57
bed_bath_table	10	agro_industry_and_commerce	34	computers	58
toys	11	party_supplies	35	home_comfort_2	59
construction_tools_construction	12	home_confort	36	small_appliances_home_oven_and_coffee	60
telephony	13	cds_dvds_musicals	37	dvds_blu_ray	61
health_beauty	14	industry_commerce_and_business	38	costruction_tools_tools	62
electronics	15	consoles_games	39	fashio_female_clothing	63
baby	16	furniture_bedroom	40	furniture_mattress_and_upholstery	64
cool_stuff	17	construction_tools_safety	41	signaling_and_security	65
watches_gifts	18	fixed_telephony	42	diapers_and_hygiene	66
air_conditioning	19	drinks	43	books_imported	67
sports_leisure	20	kitchen_dining_laundry_garden_furniture	44	fashion_childrens_clothes	68
books_general_interest	21	fashion_shoes	45	music	69
small_appliances	22	home_construction	46	arts_and_craftmanship	70
food	23	audio	47	security_and_services	71
luggage_accessories	24	home_appliances_2	48		

Table 3: Product category nominal data conversion

After converting data, the converted order status which has transformed to 2 or non-logistics status is subsequently removed from the dataset since the study focuses on the loss during the logistics-related activities, which are the statuses with labels 1 and 0.

3.3 Correlation Matrix

The correlation matrix is used to find the correlation between each independent variable. The correlation matrix is visualised in the following heatmap for easy understanding and interpretation.

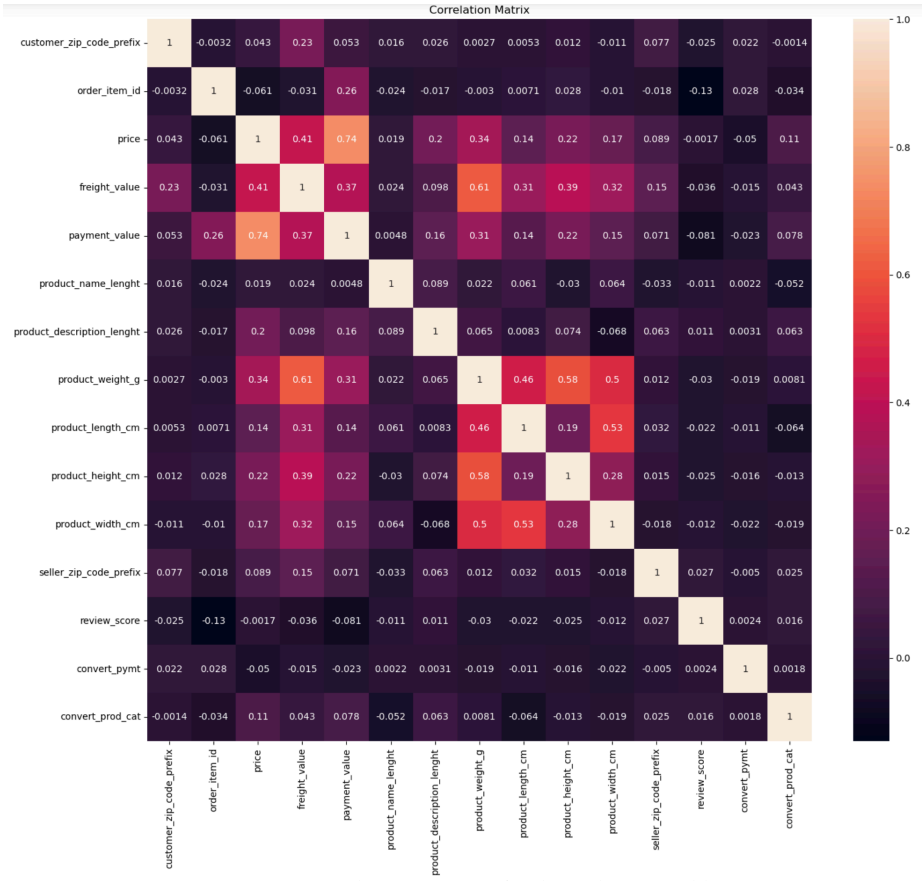

Figure 3: Correlation Matrix of Independent Variables

The visualisation is coloured in a coolwarm palette, where purple represents the lower value, orange indicates the near 0 value, and white or lighter colour represents the higher value. The heatmap presents that most of the variables have a slight negative correlation, which is represented in dark purple colour. However, there are some moderate to strong positive correlations between the metrics that indicate the size of the product, such as product weight, product height, product length, and product width. These variables also have a positive correlation with price, freight value, and payment value. Particularly, the correlation of product weight and freight value, which achieves a 0.61 score, implies a strong positive correlation with each other. Similar to the correlation between price and payment value metrics, which obtains a 0.74 score, represents the strongest positive correlation across the dataset.

3.4 Multi-Colinearity, Dimensionality and Noise Reduction

3.4.1 Variance Inflation Factor (VIF)

Even though the dataset has been cleaned and prepared for being trained in the machine learning and deep learning models, there are still unnecessary duplications in the dataset. For instance, dimensionality, multi-collinearity variables, and noise in the dataset. The VIF is applied to the dataset, which only contains the independent variables to find the collinearity in the dataset. The VIF scores are displayed in a below bar plot.

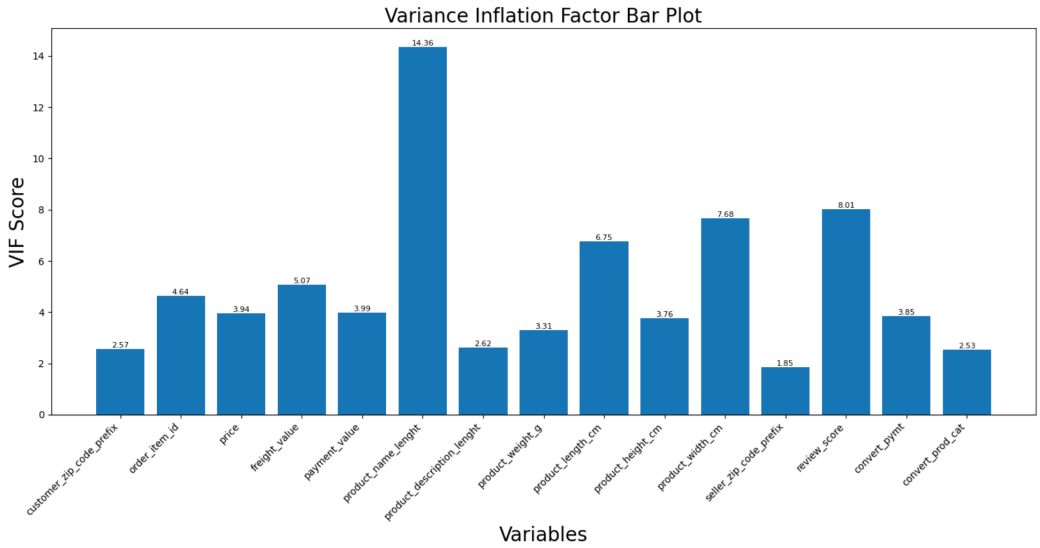

Figure 4: Variance Inflation Factor (VIF) Scores

Regarding the VIF methodology, the variable with a VIF value greater than 10 is considered as strong multi-collinearity in the dataset, a VIF value between 5 to 10 represents a potentially problematic level of multi-collinearity, a VIF value between 1 to 5 indicates a moderate correlation, and VIF value of 1 implies no correlation (Garcia, et al., 2015). From the above figure, it is obvious that the product name length has a significantly high VIF score of 14.36 score, the highest score in the dataset. The second highest VIF score is the review score variable, which obtains an 8.01 VIF score, followed by product width, product length, and freight value, which have 7.68, 6.75, and 5.07 scores respectively. The results present that the product name length has a strong multi-collinearity in the dataset and requires removal to stabilize the model. Those review score, product width, product length, and freight value variables also present that they would have a potentially problematic multi-collinearity in the dataset. These potentially

problematic variables require further analysis and supportive information to finalise the decision of whether to remove or keep them in the dataset.

3.4.2 Principal Component Analysis (PCA) and Eigenvalues

According to the previous section, further analysis of multi-collinearity reduction is necessary to finalise the training dataset for the machine learning and deep learning models. Consequently, PCA is conducted to reduce the dimensionality, noise, and multi-collinearity variables which could decrease the training performance of the models. The scree plot and eigenvalues interpretation have been used for interpreting the results of PCA, where the scree plot and eigenvalues are shown in the following.

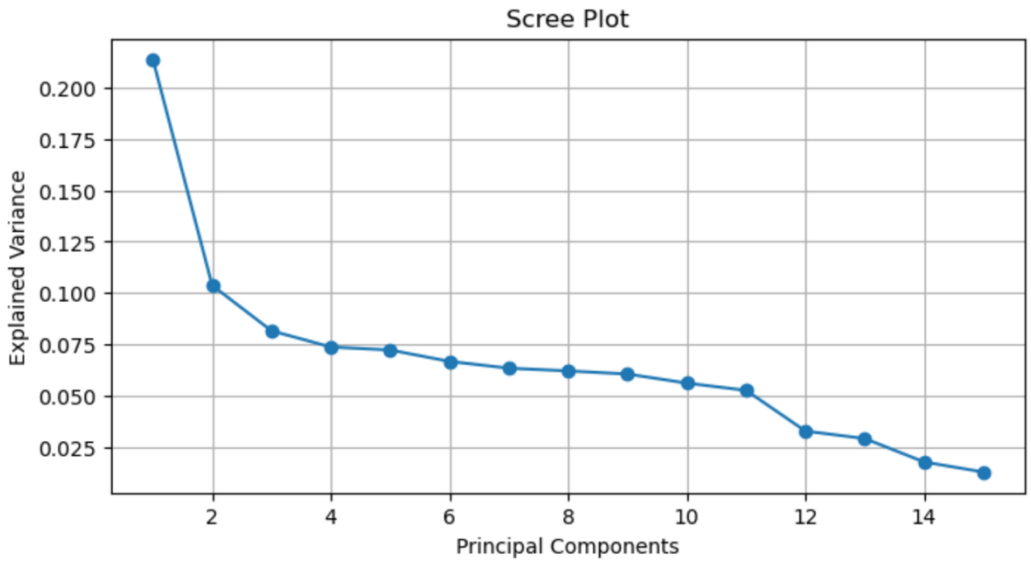

Figure 5: Scree Plot of PCA Model

According to the scree plot, the x-axis presents the number of principal components or independent variables, 1 to 15. While the y-axis shows the explained variance ratio for each principal component. This ratio indicates how much of the total variance in the dataset each component accounts for. The plot's shape starts high, which is common in PCA. The sharp significantly drops after the first few components suggest that these components capture the most significant patterns in the data. Using the Elbow point method, the plot starts to flatten after the fourth component. The elbow point is located approximately at the fourth component. The result conveys that 4 components or independent variables could capture the most significant aspects of the data's variability.

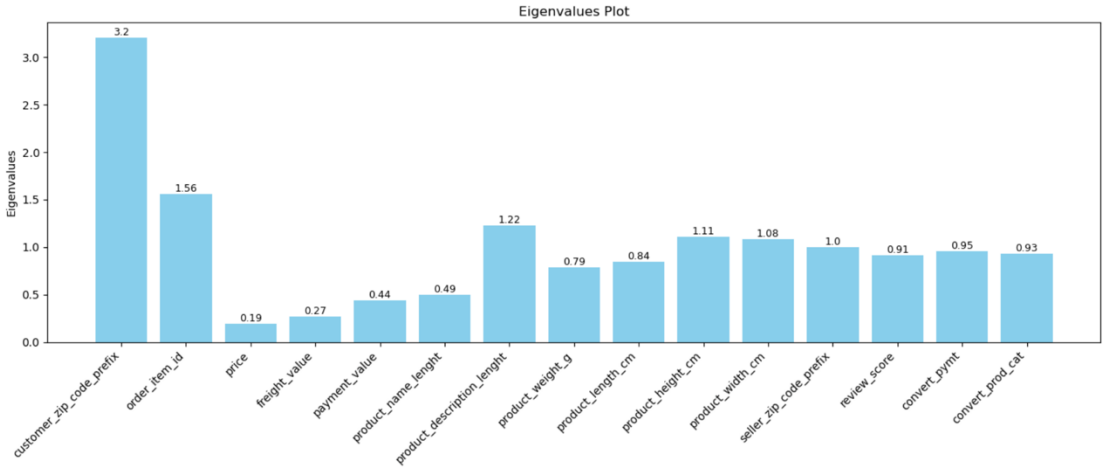

Figure 6: Eigenvalues Bar Plot of PCA Model

Regarding the eigenvalues bar plot, there are 5 independent variables that have an eigenvalue greater than 1, including customer sip code prefix, order item ID, product description length, product height, and product width. These independent variables achieve 3.2, 1.56, 1.22, 1.11, and 1.08 of eigenvalues respectively. These results also say that these 6 variables should be retained in the dataset, following a common rule of thumb mentioning that the factor with eigenvalues greater than 1 contributes significantly to explaining the variance in the dataset (Howard & Robert, 1963).

The results of the VIF score, PCA scree plot, and eigenvalues could be combined and interpreted to retain the 4 independent variables, which follows the outcome of PCA having the elbow point at the fourth component in the scree plot. According to the eigenvalues, the 4 variables that are supposed to be retained are customer zip code prefix, order item ID, product description length, and product height, since they are the 4 highest eigenvalues in the dataset. The conclusion to retain the 4 aforementioned variables also accords with the result of VIF scores, which emphasizes that the product name length variable should be removed, as well as the review score, product width, product length, and freight value, which have potentially problematic level of multi-collinearity. The training dataset is subsequently consists with customer zip code prefix, order item ID, product description length, and product height as the independent variables. This final training dataset will be utilised for training RF and FNN which will be conducted in the next section.

3.5 Balancing Data

The dataset found that there are 113,209 observations that are marked as a successfully delivered order and 1,674 observations marked as unsuccessfully delivered orders. This massive different size of data could be biased and pose a serious challenge to the machine learning model. The balancing method is necessary to be applied to the dataset before being used for training the model. Since the gap is very high, conducting the oversampling, which increases the number of unsuccessfully delivered orders, might lead the machine learning model to be overfitted as the model would learn the exact same instances multiple times (Ali, et al., 2024). In the opposite of oversampling, conducting undersampling, decreasing the number of successfully delivered orders, could potentially discard useful data or instances. Hence, the combination of over and undersampling is chosen to balance the skewness of the data. Consequently, the number of successfully delivered orders decreased to 70,755 observations and the number of unsuccessfully delivered orders increased to 56,604 observations, 127,359 observations in total, after finishing the combination of over and undersampling methods.

3.6 Random Forest Algorithm (RF)

RF is selected to predict the operational loss of the ECL dataset due to its high performance in forecasting the unbalanced dataset in classification and regression tasks (Biau & Scornet, 2016). Nevertheless, finding the best setting for the RF is a challenge. To figure out the best setting for RF, the GridSearchCV technique has been conducted to optimise the hyperparameters of RF. The parameters of GridSearchCV are also selected to 5 cross-validation, which means the data will be split into 5 folds and trained and tested 5 times, value of 2 for verbose for displaying some progress information while processing the operation, and a value of 1 for *"n_jobs"* parameter, which means telling the model to use all processors for speeding up the execution by performing multiple operations in parallel. The hyperparameters set for finding the best model of RF by implementing GridSearchCV are presented in the following table.

Parameters	Settings	Description
n_estimator	50, 100, 200	Number of trees in the forest
max_features	Auto, sqrt, log2	Number of features to consider at every split
max_depth	None, 10, 20, 30	Maximum number of levels in tree
min_samples_split	2, 5, 10	Minimum number of samples required to split a node
min_samples_leaf	1, 2, 4	Minimum number of samples required at each leaf node

Table 4: GridSearchCV parameters for the Random Forest model

After computing the GridSearchCV, the result mentions that the below table is the best setting of the RF model in forecasting the operational loss in ECL.

Parameters	Best Settings
n_estimator	200
max_features	sqrt
max_depth	None
min_samples_split	2
min_samples_leaf	1

Table 5: The best parameters for the Random Forest model

The value of 200 in *"n_esetimator"* means the number of trees in the forest. More trees in the forest generally provide a more robust and stable prediction as they reduce variance. The maximum number of features or *"max_features"* considered for splitting a node is the square root of the total number of features, as the "sqrt" is chosen to be the best setting of this parameter. This is a common choice to add randomness to the model while ensuring each tree in the forest is not identical. Since the *"max_depth"* is set to *"None"*, The model does not have a limit on the depth of the trees. This means trees in the forest can grow until all leaves are pure or until all leaves contain less than *"min_samples_split"* samples, which the *"min_samples_split"* is adjusted to a value of 2 meaning a node will be split if it contains at least two samples. This is the smallest split possible, which allows for very detailed learning from the data. The setting of 1 in *"min_samples_leaf"* conveys that A leaf node will be created only if it contains one sample.

Regarding the result of the RF model which is presented below, this model achieves 95.97% overall accuracy with 96% accuracy in both successfully and unsuccessfully delivered

orders as shown in the classification report. The results imply that the RF model has performed well on this classification task, with strong metrics across precision, recall, and F1-score for both classes.

```
Accuracy: 0.9597204773869347
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.95      0.95     11396
           1       0.96      0.97      0.96     14076

    accuracy                           0.96     25472
   macro avg       0.96      0.96      0.96     25472
weighted avg       0.96      0.96      0.96     25472
```

Figure 7: The Results of the Final RF Model

Moreover, the Mean Decrease Accuracy (MDA) is also computed to find and prioritise the importance of the variables in the final dataset. The below plot of MDA presents that the product height is the most important variable with a 0.26 score. The customer zip code prefix achieves the second most vital features in the model which obtain approximately 0.22 score. The third most crucial variable is product description length having a 0.17 score. The order item ID variable is the least important variable in the training dataset with a 0.07 score. The result emphasises that the product height and customer zip code prefix are important to the RF operational loss in ECL prediction model.

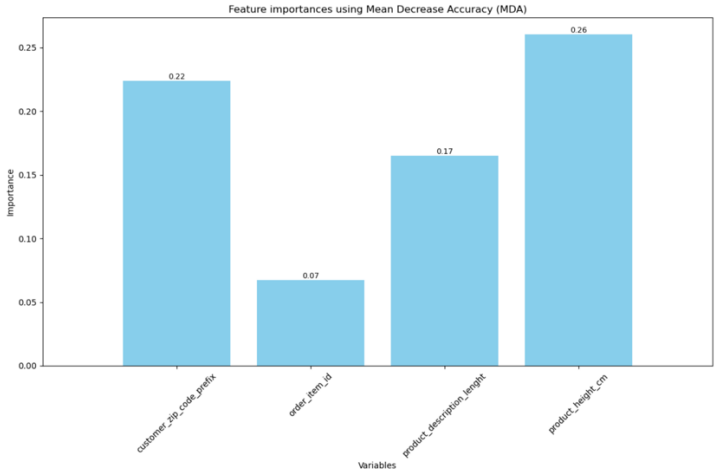

Figure 8: The Mean Decrease Accuracy Bar Plot of Random Forest Model

3.7 FeedForward Neural Network (FNN)

FNN is also chosen to forecast the operational loss in ECL and compare the result with RF. The FNN is selected because of its simplicity and effectiveness, as well as the ability to generalise the training data to unseen data (Razavi & Tolson, 2011). As same as RF model,

the FNN requires the adjustment of parameters to find the best model for the classification task. Since FNN is a multilayer perceptron model, the model is selected to have 3 layers, including the input layer, hidden layer, and output layer. The number of neurons is set to have 128, 64, and 1 in the input, hidden, and output layers respectively. The sequential type of model is chosen for easily sequentially adding the layer in the model. The first 2 layers, input and hidden layers, have the activation function of Rectified Linear Unit (ReLu), which is good for non-linear transformations and avoiding vanishing gradient issues and the saturation problem (Jackson, 2023). The output layer has a different activation function, sigmoid, since the dependent variable is binary classification, the sigmoid is necessary to be utilised for squashing the output between 0 or 1. To avoid the overfitting of the model, the *"L2 regularizer"* with a lambda of 0.001 is applied in the *"kernel_regularizer"* and *"bias_regularizer"* parameters to prevent overfitting by penalising large weights. The model uses Adaptive Moment Estimation (Adam) as the optimiser since Adam has ability to make each parameter's step size independent of the norms of the other parameters. Adam also helps the deep learning model to gain more efficiency and robustness in the training session by handling sparse gradients and adaptive learning rate adjustments (Li, et al., 2023). The *"binary_crossentropy"* is used in *"loss"* parameter according to the binary classification dependent variable. During the training and testing, the accuracy is used for evaluating the model, which is a common metric for classification tasks. There are some parameters that require cross-validation for determining the best settings. The GridsearchCV is subsequently applied to seek the best parameters of FNN, where the parameters needed to be sought are mentioned in Table 5.

Parameters	**Settings**	**Description**
kernel_initializer	he_uniform, he_normal	Initial weights of the layer in neural network
batch_size	10, 20, 40, 60, 80, 100	Number of training iteration in noe epoch
epochs	10, 20, 50, 70, 100	Number of one complete pass of the training dataset through the algorithm

Table 6: GridSearchCV parameters for the FeedForward Neural Network model

Typically, the *"kernel_initializer"* has more various options than only *"he_uniform"* and *"he_normal"*. Nonetheless, since ReLu is selected to be the activation function of the

model, the appropriate initialiser is then either *"he_uniform"* or *"he_normal"* (Telgarsky, 2023).

The following table presents the best result of FNN's settings for forecasting the operational loss in ECL which is figured out by the implementation of GridSearchCV.

Parameters	Settings
kernel_initializer	he_uniform
batch_size	100
epochs	50

Table 7: The best parameters for the FeedForward Neural Network model

The *"he_uniform"* initialiser performs the most accurate model than *"he_normal"* initialiser. The *"he_uniform"* initializer draws weights from a uniform distribution. The range of this distribution is determined based on the number of input units in the weight tensor. 100 value of batch size indicates that the best performance was achieved when the model processed 100 samples at a time to update weights. 50 in the epochs parameter implies The optimal number of complete passes through the entire training dataset is 50.

These settings of the FNN model result in the performance of 63.38% of the best cross-validation accuracy and 63.55% of test set accuracy. The results of cross-validation accuracy and test set accuracy are relatively close, indicating that the model is not excessively overfitting to the training data, and effectively generalise to unseen data

```
Best cross-validation accuracy:  0.6338296107349551
Test set accuracy:  0.6354820979899497
```

Figure 9: The Results of the Final FNN Model

4. Findings and Discussion

According to the result of the RF model, which achieves 95.96% accuracy, the RF model explicitly performs a great performance and high accuracy. Nevertheless, there is a concern about overfitting in the model. This RF model is advisable to be validated on a completely separate validation set or through k-fold cross-validation to ensure that these results are robust and stable, as well as affirming that the model is not overfitting. Also, checking the model's performance in

terms of other metrics, such as ROC-AUC, could provide deeper insights, especially in imbalanced datasets.

In terms of the FNN performance, the model obtains 63.38% accuracy of the prediction. While the accuracies are moderate, the model may or may not be sufficient depending on the complexity of the problem and the benchmark standards for similar tasks. The study of Long Short-Term Memory (LSTM) in logistics demand forecasting and the study of Red-Green-Blue Depth (RGBD) in saliency detection of logistics packages could be utilised as a benchmark accuracy of the deep learning model prediction in ECL. The study of LSTM reveals that LSTM achieves approximately 96% accuracy (Lu, 2023), which is similar to the study of RGBD which achieves 96.92% of detection accuracy (Jiang, et al., 2022). Comparing the accuracy of FNN operational loss in ECL prediction, which obtains 63.68%, could emphasise that the accuracy of FNN is unacceptable, and requires a reconciliation of whether to adjust the parameters or utilise the other deep learning model for the operational loss in ECL prediction.

In comparison, the results of both RF and FNN for operational loss in ECL prediction are explicit. Since the FNN model has an unacceptable accuracy, the RF model is consequently the best and recommended model for forecasting the operational loss in ECL. Nonetheless, the RF model is recommended to be applied to a different dataset to affirm the robustness and stability of the model. The application of other machine learning or deep learning is also suggested to be conducted and compared to the results of this study. There is a possibility that the RF and FNN could perform differently in a different dataset and settings. The RF might not be the best model for forecasting the ECL operational loss in different datasets.

From the operational factors that could affect the operational loss in ECL, the outcomes of VIF, PCA, and MDA emphasise that the customer zip code prefix, product length, product description length, and order item ID could potentially influence the parcel to operational loss. The results could be interpreted to accord with a realistic ECL situation. Regarding the MDA score, the customer zip code prefix is the most important factor in the operational loss prediction. In reality, the customer’s zip code conveys the destination of the parcel, where the area is responsible for last-mile operation (Niu, et al., 2024). This could be interpreted that there are some operational issues in the last-mile operation. The operational loss could happen for many reasons, including

the delay of delivery that is overdue the approximate delivery date and causes the logistics company to compensate the recipient, the loss of parcel which the parcel is literally lost and could not be found, and damaged parcel which could not be delivered or get rejected by recipients. There is a possibility to have a backlog in the last-mile operations causing a bottleneck in delivery and leading to the delay in delivery. This situation could happen in resource shortage, such as manpower shortage or lack of vehicles. The geographical challenge also could impact delivery performance and lead to a bottleneck or backlog in the last-mile operations. For example, the destination is located on the mountain or in the forest which is difficult to transport with 4-wheeled vehicles, or the destination with a high risk of disaster such as floods or landslides. Another possible scenario is there might be fraud in last-mile operations that lead to the loss of the parcel. This could significantly cause the high importance and correlation of customer zip code and operational loss orders. These aforementioned situations could be tracked by monitoring the performance of the last-mile operation. Additionally, the monitoring of each last-mile hub operation performance individually could foreshadow a possible backlog and uncommon situation in each last-mile area as well. The error of the sorting machine should also be considered. Typically, the parcel will be sent to the sorting centre before being distributed to the destination area (Werners & Wulfing, 2010). The parcel might be sorted to the wrong destination or stuck in the sorting machine. This could also be the reason for the correlation between customer zip codes and operational loss orders. The problem of miss-sorting or parcels stuck in the machine could be noticed by monitoring the sorting machine's performance. The number of parcels scanned at the entrance and exit should be equal to each other. Tracking the parcel by scanning the parcel several times while on the sorting machine could find where the parcel is stuck or lost on the sorting machine as well.

The second most influential factor of operational loss in ECL is product height. The parcel with a tall height might be stuck in the machine or might be manually sorted due to the height exceeding the limit of the machine and causing a longer processing time. In contrast, a parcel with a short height could also lost in the sorting centre due to a thin and light parcel can easily be stuck on the sorting machine's belt or bounce off the sorting machine's belt. Enhancing the capability of the sorting machine to receive a large size of parcel would reduce the manual processing time, increase accuracy in processing, and make the parcel traceable. Plus, the monitoring of parcels by tracking the last scan on the sorting machine would help find a trouble spot on the machine. In

case of fraud, a small parcel is explicitly easier to steal than a large parcel due to the ease of picking and portability to carry out of the working area. This fraudulent activity could be preliminarily prevented by improving fraudulent policies and penalties, such as providing a special financial incentive for a fraudulent reporter. There is a study emphasising that enhancing the severity of penalties and monetary rewards could encourage employees to report fraudulent activity (Fiorin, 2019).

The product description length is the third most important factor of operational loss in ECL. This factor is different from previous factors since previous factors involve more operational activities. To demonstrate, the performance of last-mile delivery or the flow of sorting machines. This factor focuses on fraudulent activities since the reason for a correlation between product description length and operational loss is the shipping labels. Due to the fact that, normally the delivery person could not know what the product is inside the package. The only fact that tells about the product is the product description on the shipping label. The product description could tell what product is inside the parcel, and could also imply the worth of the product. This possibly leads to a fraudulent incident which the delivery person aims to steal a worthy product and report it as a loss of parcel. The solution to this problem requires further study on whether to keep or remove the product description on the label. The removal of the product description might impact customer satisfaction because they would not know what the product is inside the package. The preliminary risk mitigation way of fraud is the improvement of fraudulent policies and penalties as mentioned in the previous paragraph.

The product item ID is ranked as the least important factor that affects the operational loss in ECL, which does make sense. Since the product item ID is the distinct code of the product, which is generally used for identifying the order being tracked on the system. In other words, it implies nothing to the delivery person neither what is inside the parcel nor the value of the product. For the reason why product item ID is retained in the final dataset is due to the non-collinearity between itself and other variables. Nevertheless, the correlation between a pattern of product item ID and loss order should be studied to affirm the assumption that there is no significant correlation between them.

The findings and interpretations of this study would be beneficial to the ECL company across the world, particularly, the ECL company in Brazil. Since the dataset used in the study is the real observations obtained from the Brazilian ECL company, this study could represent the behaviour of the ECL operational activities in Brazil. However, both Brazilian and non-Brazilian ECL companies are recommended to use this study as a benchmark and baseline model, and conduct the analysis on the different datasets, especially, the non-Brazilian ECL company according to the different behaviour of ECL processing regarding the dissimilar region. The study of different datasets is required to be conducted and compared the result of this study to affirm the robustness and effectiveness of the findings.

5. Conclusion

The study aims to analyse the factors that could lead to an operational loss in ECL and create the best machine learning or deep learning model that could effectively and accurately forecast the operational loss in ECL. The datasets of Olist, the Brazilian marketplace store, are used for the analysis. The dataset has removed the null value and converted the categorical data to nominal data. The observations with non-logistic status have also been removed from the dataset. The status order has been treated as the binary dependent variable, which is classified into successfully delivered status and unsuccessfully delivered status.

The correlation matrix has been conducted to find the correlation between independent variables. The correlation matrix presents that most of the independent variables have slight to moderate negative correlation, except the variables of sizes of parcel and product value, which mostly have a moderate to strong positive correlation.

The Variance Inflation Factor (VIF) has been used to analyse the multi-collinearity in independent variables. The VIF scores show that the product name length has a high multi-collinearity, while the review score, product width, and product height have a potentially problematic multi-collinearity in the dataset. Consequently, those 4 mentioned variables have been removed from the dataset to ensure the effectiveness and robustness of the model. The Principal Component Analysis (PCA) and eigenvalues have also been applied to reduce the collinearity, dimensionality, and noise in the dataset. The PCA's result presents that having 4 components is the most effective number to capture the most significant pattern and variance in the data. The

eigenvalues scores also reveal that the customer zip code prefix, order item ID, product description length, and product height achieve the 4 highest eigenvalues among the dataset. Hence, the customer zip code prefix, order item ID, product description length, and product height have been retained in the dataset, while the rest variables have been removed from the dataset. The dataset which contains 4 selected independent variables and the order status variable, the dependent variable, has found a significant imbalance in the dependent variable. The combination of over and undersampling methods has subsequently been applied to balance the dataset.

The prepared dataset has been utilised to train both the Random Forest (RF) classification model and the FeedForward Neural Network (FNN) model. To find the best parameters of both RF and FNN, the GridSearchCV method has then been applied. The results have provided the best settings of each hyperparameter of the RF and FNN model with the accuracy score of the predictions. The outcomes present that the best performance model is the RF model, which achieves 95.97% accuracy, while the FNN achieves only 63.38%. The Mean Decrease Accuracy (MDA) has also been analysed to find the most important variables that potentially affect the dependent variable. The MDA scores emphasise that customer zip code is the most important feature, followed by product height, product description length, and order item ID respectively. The results could be interpreted as that there is a possibility of having a last-mile operational issue, a problem with the sorting machine, and potential fraud in the operational activities in the ECL, which could potentially affect the operational loss in ECL.

The results help the ECL company to notice the factors that could affect or lead to operational loss in ECL. Particularly, the Brazilian ECL company which is located in the same region as the training dataset. Nonetheless, further study of the application of the RF model on different datasets is required to affirm the effectiveness, stability, and robustness of the model since the different datasets might have a dissimilar factor and behaviour which could impact to the model's performance. This study would be a benchmark and baseline model for the ECL company to conduct the research on the operational loss in ECL, and compare the result for ensuring the outcomes of the study.

6. Appendices

The Python code of data preprocessing:
https://github.com/NattawatApi/The-Analysis-and-Forecasting-of-the-Operational-Loss-in-E-Commerce-Logistics/blob/main/Data_Processing.ipynb

The Python code of data analysis and forecasting:
https://github.com/NattawatApi/The-Analysis-and-Forecasting-of-the-Operational-Loss-in-E-Commerce-Logistics/blob/main/ML_DL_Analysis.ipynb

7. References

Abbood, A. D., Hasan, A. & Attea, B., 2021. Pearson coefficient matrix for studying the correlation of community detection scores in multi-objective evolutionary algorithm. *Periodicals of Engineering and Natural Sciences (PEN).*

Abdi, H. & Williams, L. J., 2010. Principal component analysis. *WIREs Computational Statistics,* 2(4), pp. 387-515.

Ali, A. H., Charfeddine, M., Ammar, B. & Hamed, B. B., 2024. Intrusion Detection Schemes Based on Synthetic Minority Oversampling Technique and Machine Learning Models. *2024 IEEE 27th International Symposium on Real-Time Distributed Computing (ISORC),* pp. 1-8.

Anitha, M., Nicholas, S., Bhanu, M. S. & Gayathiri, S., 2023. Unlocking the Potential of Weight of Evidence and Entity Embedding Encoding for Categorical Data Transformation in Medical Datasets: An Innovative Approach to Enhance Classification Accuracy.

Bai, S., Lv, Y. & Liu, Z., 2022. Optimal Decision and Coordination of Fresh E-Commerce Supply Chain Considering Double Loss. *Discrete Dynamics in Economic and Business Systems,* Volume 2022, pp. 1-13.

Biau, G. & Scornet, E., 2016. A random forest guided tour. *TEST,* Volume 25, pp. 197-227.

Dutta, P., Suryawanshi, P., Gujarathi, P. & Dutta, A., 2019. Managing risk for e-commerce supply chains: an empirical study. *IFAC-PapersOnline,* 52(13), pp. 349-354.

Ekiz, O. U., 2021. An improved robust variance inflation factor: Reducing the negative effects of good leverage points. *Kuwait Journal of Science.*

Falk, M. & Hagsten, E., 2015. E-commerce trends and impacts across Europe. *Int. J. Production Economics,* Volume 170, pp. 357-369.

Fiorin, S., 2019. Reporting Peers' Wrongdoing: Experimental Evidence on the Effect of Financial Incentives on Morally Controversial Behavior.

Garcia, C. B., Garcia, J., Martin, M. M. & Salmeron, R., 2015. Collinearity: revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics,* Volume 42, pp. 648-661.

Gill, K. S. & Rupesh, G., 2023. Chronic Kidney Disease Detection Using GridSearchCV Cross Validation Method.. *2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON),* pp. 318-322.

Grewal, D., Roggeveen, A. L. & Tsiros, M., 2008. The Effect of Compensation on Repurchase Intentions in Service Recovery. *Journal of Retailing,* December, 84(4), pp. 424-434.

Howard, K. I. & Robert, A. G., 1963. Empirical Note on the "Number of Factors" Problem in Factor Analysis. *Psychological Reports,* Volume 12, pp. 247-250.

Jackson, J., 2023. An Isometric Stochastic Optimizer. *ArXiv,* Volume abs/2307.12979.

Jalil, E. E. A., 2019. Customer Satisfaction and Reverse Logistics in E-Commerce: The Case of Klang Valley. *9th International Conference on Operations and Supply Chain Management.*

Jiang, G., Zhang, W., Wang, W. & Sun, X., 2022. International Conference on Artificial Intelligence, Virtual Reality, and Visualization. *International Conference on Artificial Intelligence, Virtual Reality, and Visualization,* pp. 85-92.

Kim, S. H. B. J. H. &. J. H. M., 2019. Continuous Intention on Accommodation Apps: Integrated Value-Based Adoption and Expectation–Confirmation Model Analysis. *Sustainability,* 11(6), p. 1578.

Kup, B. U. et al., 2023. Capacity Planning in E-Commerce Logistics Using a Hybrid Machine Learning Model. *2023 Innovations in Intelligent Systems and Applications Conference (ASYU),* pp. 1-6.

Larson, P. D., 1992. Business Logistics and the Quality Loss Function. *Journal of Business Logistics,* 13(1), p. 125.

Li, J. et al., 2023. Generalized Activation via Multivariate Projection. *ArXiv,* Volume 2309.17194.

Lu, H., 2023. Logistics demand forecasting method based on deep learning. *International Conference on Artificial Intelligence, Virtual Reality, and Visualization.*

Niu, H., Jia, J. & He, Y., 2024. Research on the Optimization Strategy of Last-Mile Distribution under the E-Commerce Logistics Model. *Frontiers in Business, Economics and Management.*

Paul, S. K., Asian, S., Goh, M. & Torabi, S. A., 2019. Managing sudden transportation disruptions in supply chains under delivery delay and quantity loss. *Annals of Operations Research,* 23 October, Volume 273, pp. 783-814.

Razavi, S. & Tolson, B. A., 2011. A New Formulation for Feedforward Neural Networks. *IEEE TRANSACTIONS ON NEURAL NETWORKS,* 22(10), pp. 1588-1598.

Telgarsky, M., 2023. Feature selection and low test error in shallow low-rotation ReLU networks. *International Conference on Learning Representations.*

Werners, B. & Wulfing, T., 2010. Robust optimization of internal transports at a parcel sorting center operated by Deutsche Post World Net. *Eur. J. Oper. Res.,* Volume 201, pp. 419-426.

Xiaoqiong, Z., 2019. E-commerce Logistics Wealth Risk Control. *2019 5th International Conference on Economics, Management and Humanities Science (ECOMHS 2019).*

Xu, G. et al., 2019. Data-driven operational risk analysis in E-Commerce Logistics. *Advanced Engineering informatics,* April, Volume 40, pp. 29-35.

Yu, H. et al., 2021. Research on the financing income of supply chains based on an E-commerce platform. *Technological Forecasting and Social Change,* Volume 169.

Yu, W.-H., Chiu, S.-K. & Tung, C. M., 2019. The study of evolution among logistic service quality, service compensation and long-term cooperation commitment. *Procedia Manufacturing,* Volume 39, pp. 1493-1500.

Zhang, X. et al., 2023. A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research. *Journal of Theoretical and Applied Electronic E-Commerce Research,* Volume 18, pp. 2188-2216.

Zhao, J., Ye, F. & Li, S., 2023. Research on Cold Chain Logistics Risk Control of Fresh E-commerce under New Retail. *2023 the 7th International Conference on Management Engineering, Software Engineering and Service Sciences.*

Zheng, C., Peng, B. & Wei, G., 2020. Operational risk modeling for cold chain logistics system: a Bayesian network approach. *Kybernetes,* 50(2), pp. 550-567.