DATA ANALYTICS FOR OPTIMAL CONTAINER SHIPPING

Zisis Konstantinos

Athens, 17/7/24

School of Mechanical Engineering,
Sector of Industrial Management & Operational Research
Thesis Supervisor: A. Rentizelas, Associate Professor NTUA

Acknowledgement

I need to thank all the members of Sappi's digital transformation team as well as the faculty of the industrial engineering section of the Mechanical Engineering School of NTUA. More specifically I would like to thank Dr. Khan and Dr. Rentizelas for giving me this opportunity to develop my career.

Table of Contents

Table of figures

Εκτεταμένη περίληψη

Σε αυτή την διπλωματική εργασία παρουσιάζεται η ανάπτυξη και η δημιουργία ενός μοντέλου πρόβλεψης καθυστερήσεων εμπορευματοκιβωτίων (Containers) σε μεταφορές από λιμάνι σε λιμάνι με την χρήση τρίτων εταιριών μεταφοράς. Εξετάστηκε αν είναι πιθανό να προβλεφθούν αυτές οι καθυστερήσεις και με τι ακρίβεια. Έγινε ερεύνα για τις μεγαλύτερες αιτίες αυτών των καθυστερήσεων και συγκρίθηκαν συγκεκριμένοι αλγόριθμοι μεταξύ τους.

Η παγκόσμια εφοδιαστική αλυσίδα έχει επηρεαστεί αρνητικά ως προς την σταθερότητα της λόγω διάφορων πρόσφατων συγκυριών (π.χ. πανδημία, γεωπολιτικές εξελίξεις), με αποτέλεσμα οι εταιρίες να αναγκαστούν να αναζητήσουν ή να δημιουργήσουν εργαλεία για την καλύτερη διαχείριση της. Λύση σε αυτό το πρόβλημα έρχεται να δώσει ο ψηφιακός μετασχηματισμός, μέσω του οποίου μπορούν να δημιουργηθούν εφαρμογές για καλύτερη περιγραφή, διάγνωση και πρόβλεψη των διαδικασιών και των προβλημάτων της εφοδιαστικής αλυσίδας.

Σε αυτή την εργασία, περιγράφονται οι διαδικασίες, η μεθοδολογία και τα αποτελέσματα της διπλωματικής εργασίας, η οποία συνδυάστηκε με μια πρακτική άσκηση διάρκειας 4 μηνών σε μια πολυεθνική εταιρία παραγωγής χαρτιού διάφορων τύπων - την Sappi - με παραρτήματα σε Ευρώπη, βόρεια Αμερική και Αφρική. Όπως είναι λογικό μια εταιρία τέτοιου βεληνεκούς αντιμετωπίζει πολλά προβλήματα γύρω από την εφοδιαστική αλυσίδα της. Επομένως το πρώτο βήμα για την εκκίνηση του project ήταν να γίνουν κάποιες συναντήσεις με εκπροσώπους από το εργαστήριο επιχειρησιακής έρευνας του ΕΜΠ με εργαζόμενους από την Sappi που ειδικεύονται στην εφοδιαστική αλυσίδα. Από αυτές τις συναντήσεις προέκυψε ότι έπρεπε να γίνει μια αναλυτικότερη περιγραφή των απαιτήσεων (need-finding) των εργαζόμενων πάνω στην διαχείριση της εφοδιαστικής αλυσίδας. Για αυτόν τον λόγο, ξεκίνησε το πρώτο μέρος του project, που αποτελούταν από έξι συνεντεύξεις με ερωτήσεις ανοικτού τύπου και μια θεματική ανάλυση που προέκυψε από αυτές, όπως αναφέρεται στο κεφάλαιο 5.

Σε αυτή την φάση του έργου, για τις συνεντεύξεις, επιλέχθηκαν εργαζόμενοι, ειδικευμένοι πάνω στην διαχείριση της εφοδιαστικής αλυσίδας και από τα 3 παραρτήματα της Sappi (Ευρώπη, Αφρική και Βόρεια Αμερική), με σκοπό να αναδειχθούν τα βασικότερα διεθνή προβλήματα. Από τις βιντεοσκοπήσεις και τα πρακτικά των συνεντεύξεων, επιλέχθηκαν οι σημαντικότερες απαντήσεις και συσπειρώθηκαν σε ομάδες με βάση την θεματολογία τους. Τα σημαντικότερα συμπεράσματα που προέκυψαν ήταν:

1. Η εξυπηρέτηση πελατών και ο προγραμματισμός των μεταφορών θα βελτιωθούν περισσότερο με την χρήση ψηφιακών εργαλείων για έλεγχο της εφοδιαστικής αλυσίδας.
2. Υπάρχει ανάγκη για νέα προγνωστικά και διαγνωστικά υλικά στην Sappi.
3. Η ποιότητα των διαθέσιμων δεδομένων ενδέχεται να δημιουργήσει προβλήματα στην αξιόπιστη λειτουργία των ψηφιακών εργαλείων.
4. Είναι θεμιτό να εξεταστεί η πιθανότητα να δοθεί πρόσβαση στους πελάτες της Sappi σε συγκεκριμένα δεδομένα που αφορούν την παραγγελία τους.
5. Τα ψηφιακά εργαλεία θα έπρεπε να περιέχουν στοιχεία που θα βοηθούν την αειφόρα (sustainable) διαχείριση της εφοδιαστικής αλυσίδας.

Ορμώμενοι από τα αποτελέσματα των συνεντεύξεων, προτάθηκε από την ομάδα διαχείρισης της εφοδιαστικής αλυσίδας, να δημιουργηθεί ένα εργαλείο το οποίο θα προέβλεπε καθυστερήσεις στις θαλάσσιες μεταφορές. Προέκυψε ότι η Sappi πλήρωνε δυσανάλογα ψηλά ποσά σε σχέση με τις υπόλοιπες εταιρίες για πρόστιμα λόγω υπεραναμονής και κρατήσεως (demurage & detention) των εμπορευματοκιβωτίων. Μετά από συζητήσεις μεταξύ εμπείρων μελών της ομάδας ψηφιακού μετασχηματισμού και της ομάδας διαχείρισης εφοδιαστικής αλυσίδας, αποφασίστηκε ότι το καταλληλότερο project θα ήταν να δημιουργηθεί ένας αλγόριθμος δυαδικής κατηγοριοποίησης, ο οποίος προβλέπει την πιθανότητα να καθυστερήσει μια παραγγελία εμπορευματοκιβωτίων. Με αυτό τον τρόπο, τα αποτελέσματα γίνονται απλούστερα γεγονός που τα καθιστά ευκολότερα κατανοητά από τους διαχειριστές των ομάδων, που έχουν την δύναμη της απόφασης για την επέκταση και εφαρμογή των ψηφιακών εργαλείων.

Για την ανάπτυξη του μοντέλου χρειάζονται 4 βασικά στάδια. Το στάδιο περιγραφής, πρόβλεψης, διάγνωσης και προτροπής.

Στο στάδιο της περιγραφής γίνεται η εξόρυξη δεδομένων (Data mining), κατά το οποίο δημιουργούνται πολλά διαγράμματα και εκτυπώνονται πολλές πληροφορίες γύρω από το σετ των δεδομένων. Αυτή είναι μια χρονοβόρα διαδικασία, που απαιτεί υποκειμενική κρίση από τον προγραμματιστή ως προς το ποια διαγράμματα και εκτυπώσεις θα δώσουν αρκετή πληροφορία ώστε να ανακαλυφθούν τα πρότυπα και τα μοτίβα που ακολουθούν τα δεδομένα του συγκεκριμένου προβλήματος.

Στο επόμενο στάδιο, της πρόβλεψης δημιουργείται το μοντέλο. Πριν ξεκινήσει η δημιουργία των μοντέλων, πρέπει να γίνει η διαδικασία της προετοιμασίας των δεδομένων. Για την προετοιμασία των κατηγορικών δεδομένων χρησιμοποιήθηκαν, και έγινε σύγκριση μεταξύ, 2 ειδών κωδικοποιήσεων, η τυχαία κωδικοποίηση (label encoding), και η κωδικοποίηση με βάση τον στόχο (target encoding). Εφόσον τα δεδομένα ήταν έτοιμα, δοκιμάστηκαν τα μοντέλα με τους αλγορίθμους random forest, xgboost, deep neural networks, logistic regression καθώς και ένα μοντέλο το οποίο ήταν βασισμένο σε κανόνες που διατυπώθηκαν από την ομάδα διαχείρισης εφοδιαστικής αλυσίδας (knowledge based models). Η διαδικασία ανάπτυξης ενός μοντέλου, είναι κυκλική καθώς μετά την δημιουργία του μοντέλου, εξετάζονται κάποιοι βασικοί δείκτες, όπως το precision, το accuracy, το F1-score και σε περίπτωση που είναι ικανοποιητικοί σταματάει εκεί η διαδικασία, αλλιώς επαναλαμβάνεται η προετοιμασία των δεδομένων με διαφορετικό τρόπο. Πριν την ολοκλήρωση κάθε κύκλου γίνεται αναζήτηση των βέλτιστων παραμέτρων του μοντέλου.

Αφού ολοκληρωθεί η παραπάνω διαδικασία, ξεκινάει το διαγνωστικό κομμάτι της μεθοδολογίας. Αυτό γίνεται με την προσπάθεια να εξηγηθούν τα αποτελέσματα των παραπάνω μοντέλων και ο τρόπος με τον οποίο καταλήγουν σε αυτά τα αποτελέσματα. Για την περιγραφή των αποτελεσμάτων των αλγορίθμων που χρησιμοποιούν δέντρα αποφάσεων (decision trees), χρησιμοποιούνται μετρικές όπως η εντροπία των δεδομένων πριν και μετά από έναν κόμβο και η μέση μείωση του θορύβου μεταξύ αυτών των σετ δεδομένων. Για το μοντέλο logistic regression, χρησιμοποιούνται οι τιμές των συντελεστών για κάθε χαρακτηριστικό (feature) dataset, καθώς και τα επίπεδα εμπιστοσύνης για τον καθένα από αυτούς. Τα νευρωνικά δίκτυα δεν είναι εύκολο να εξηγηθούν ως προς τον τρόπο που παίρνουν αποφάσεις.

Για το τελευταίο στάδιο, το στάδιο της προτροπής, δημιουργούνται κανόνες μέσω των οποίων μπορούν να χρησιμοποιηθούν τα αποτελέσματα. Σε αργότερο στάδιο, και εφόσον αποδειχθεί η αξία των παραπάνω μοντέλων θα είναι εφικτό να αυτοματοποιηθούν κάποιές διαδικασίες.

Σχετικά με τις διαφορετικές κωδικοποιήσεις των κατηγορηματικών μεταβλητών, καλύτερη τεχνική αποδείχτηκε η κωδικοποίηση με βάση τον στόχο, όπως φαίνεται στον πίνακα 1. Η διαφορά στην απόδοση των δυο κωδικοποιήσεων είναι σημαντική, και επιβεβαιώνεται και από την βιβλιογραφία.

Πίνακας 1: Σύγκριση μεταξύ της κωδικοποίησης ως προς τον στόχο (target) και την τυχαία κωδικοποίηση (label).

KPI's	**F1-score**		**Accuracy**		**Precision**	
Model	label	target	label	target	label	target
Random Forest	0.83	0.87	0.87	0.92	0.84	0.91

XGBoost	0.81	0.87	0.83	0.91	0.89	0.90
DNN	0.66	0.69	0.78	0.81	0.76	0.77

Για την σύγκριση ως προς τα μοντέλα παρουσιάζεται ο πίνακας 2. Την καλύτερη απόδοση την έχουν τα μοντέλα που χρησιμοποιούν decision trees, με τον αλγόριθμό του random forest να είναι λίγο καλύτερος. Αν παρατηρήσουμε και στον πίνακα 1, και για την τυχαία κωδικοποίηση, παρατηρείται το ίδιο. Το γεγονός αυτό προμηνύει ένα dataset με μη επιθυμητό θόρυβο σύμφωνα με την βιβλιογραφία, γεγονός που επιβεβαιώνεται και από τους εμπείρους υπαλλήλους της Sappi. Έπειτα τα μοντέλα logistic regression και νευρωνικά δίκτυα, παρουσιάζουν σημαντικά υποδεέστερα αποτελέσματα. Τέλος την χειρότερη απόδοση έχει το μοντέλο που βασίστηκε σε κανόνες που προέκυψαν από την γνώση των εργαζόμενων.

Πίνακας 2 : Σύγκριση μεταξύ των μοντέλων με κωδικοποίηση ως προς τον στόχο.

Model	F1-score	Accuracy	Precision
Random Forest	0.87	0.92	0.91
XGBoost	0.87	0.91	0.90
Logistic Regression	0.69	0.81	0.77
Neural Network	0.69	0.79	0.71
Transhipment model	0.62	0.71	0.56

Σχετικά με τα διαγνωστικά αποτελέσματα του κυρίαρχου μοντέλου προέκυψε ότι οι σημαντικότερες πέντε μεταβλητές ώστε να παρθεί η απόφαση για την πιθανότητα καθυστερήσης είναι:

1. Ο μήνας άφιξης των εμπορευματοκιβωτίων
2. Το λιμάνι άφιξης των εμπορευματοκιβωτίων
3. Η συνηθισμένη διάρκεια της διαδρομής
4. Η προγραμματισμένη διάρκεια της διαδρομής
5. Η εταιρία μεταφοράς.

Ο προτεινόμενος τρόπος χρήσης ενός τέτοιου μοντέλου, είναι να χωριστεί κάθε πρόβλεψη σε 3 ζώνες, την ζώνη πιθανοτήτων για τα εμπορευματοκιβώτια που θα φτάσουν στην ώρα τους (π.χ. 0-33%), την ζώνη πιθανοτήτων των εμπορευματοκιβώτιων που δεν είναι εύκολο να βγει κάποια ικανοποιητική απόφαση και χρειάζονται στενή παρακολούθηση (π.χ. 33-66%) και τέλος την ζώνη πιθανοτήτων για τα εμπορευματοκιβώτια που θα καθυστερήσουν (66-100%).

Η συγκεκριμένη μελέτη περίπτωσης ξεκινάει χρησιμοποιώντας στοιχεία της νοητικής εργονομίας που μπορούν να χρησιμοποιηθούν στον ψηφιακό μετασχηματισμό και καταλήγει να συμβάλει στο πεδίο της εφοδιαστικής αλυσίδας επιδεικνύοντας τις δυνατότητες της μηχανικής μάθησης για την έγκαιρη πρόβλεψη και αντιμετώπιση των καθυστερήσεων των εμπορευματοκιβωτίων στις θαλάσσιες μεταφορές από λιμάνι σε λιμάνι. Τα αποτελέσματα και η συζήτηση γύρω από αυτά είναι χρήσιμα για πολλές εταιρίες οι οποίες θέλουν να ψηφιοποιήσουν και να βελτιστοποιήσουν την εφοδιαστική τους αλυσίδα, μειώνοντας την συχνότητα αλλά και την σημασία των καθυστερήσεων. Επιπλέον η γενικότερη μεθοδολογία για την ανάπτυξη των μοντέλων μπορεί να εφαρμοστεί και σε άλλους τομείς εκτός από την εφοδιαστική αλυσίδα.

Summary

This thesis investigated the feasibility of accurately predicting delays in container shipping for big multinational companies. Two projects were completed, different in scope but related by their common field. Both these projects were completed with the help and the data of Sappi, which is a big multinational company focusing on forestry products and papers. The first project was a user experience (UX) project on a supply chain control tower in order to find out the needs and requirements of supply chain employees and managers. The main result of this project was that Sappi's supply chain needed more predictive analytics capability; hence in the second project a predictive model was developed to predict whether a container shipment would be late.

The research employed a four-stage methodology: describe, predict, diagnose, and prescribe. In the describing stage, relevant factors influencing container shipping delays were identified. The predicting stage focused on developing machine learning models to predict these delays. Target encoding and label encoding were compared for representing categorical features, and the performance of various algorithms (XGBoost, Random Forest, Neural Networks, Logistic Regression, as well as a knowledge-based model) was evaluated. The diagnosing stage aimed to understand the root causes of delays based on model predictions. Finally, the prescribing stage is done by incorporating some rules as to how these tools should be used. Creating a fully automated supply chain control tower is beyond the scope of this thesis.

The key findings of the research were that machine learning models can effectively predict delays in container shipping for big multinational companies with the random forest model achieving the highest accuracy in predicting delays. Moreover, as expected, target encoding showed a slight advantage over label encoding for representing categorical features in this specific case. Unfortunately, the analysis revealed that most of the features that are most useful in predicting delays are not directly affected by Sappi. However, Sappi can use these findings as negotiating tools, both with customers (e.g. timing their orders more efficiently) and with their partners (e.g. negotiating prices based on performance). To conclude, this thesis' key outputs are the predictive model that can predict delayed containers and insights as to why containers are getting delayed.

This research contributes to the field of logistics by demonstrating the potential of machine learning and predictive algorithms for proactive delay management in container shipping. The findings provide valuable insights for companies that rely on port-to-port shipping seeking to optimize their supply chains and mitigate the negative impacts of delays. The methodology employed in this thesis also holds promise for adaptation in other domains, laying the groundwork for future research exploring its transferability to diverse areas.

1. Introduction

1.1. Context

The recent disruptions caused by the COVID-19 pandemic and other unforeseen events have delivered a significant blow to global supply chains, with impacts felt across various industries. A study by McKinsey & Company found that 90% of companies experienced disruptions in their supply chains due to the pandemic and they are trying to find ways to digitalize their supply chain. Production, storage, and transportation were tremendously affected, highlighting the vulnerabilities of global trade networks. Digitization priorities are slowly shifting from visibility to demand planning (Alicke et al., 2022).

One primary challenge lies in the reliance on ships for moving goods across vast distances. This reliance has exposed numerous challenges and disruptions. A 2021 report by the International Chamber of Commerce (ICC) identified port congestion as a major bottleneck, with average waiting times for container ships tripling at major ports compared to pre-pandemic levels (Laloum Melanie, 2023). These delays can be attributed to several factors.

- Port closures due to lockdowns or outbreaks.
- Labor shortages impacting port operations and container handling.
- Logistical hurdles caused by changes in customs regulations and safety protocols.
- Geopolitical contingencies disrupting trade routes and availability of resources.

Recent global disruptions have crippled supply chains across industries. From production and storage to transportation, vulnerabilities in today’s interconnected trade system are exposed. Reliance on maritime shipping for long distances reveals challenges like port congestion, labor shortages, and logistical hurdles. Delays at ports, compounded by safety measures, worsen backlogs and shipping times. A shortage of containers and unpredictable demand further complicate matters, impacting everything from production costs to consumer prices. These disruptions highlight the need to strengthen supply chain resilience and develop robust contingency plans for future challenges. Furthermore, they have underscored the vulnerability of global trade networks to unforeseen events. As a result, there's a growing emphasis on reevaluating supply chain resilience and developing robust contingency planning strategies to mitigate future disruptions.

1.2. Problem Statement

Optimizing complex logistics systems like port-to-port container shipping requires a clear understanding of the most pressing needs within the organization. To address this, a comprehensive user experience (UX) project was undertaken, detailed in Chapter 3. This project played a crucial role in identifying the most valuable tools for Sappi. The findings revealed that implementing predictive analytics would significantly benefit various aspects of their supply chain.

Predicting delays in port-to-port container shipping is of paramount importance due to several compelling reasons. Firstly, it directly impacts customer service, as accurate predictions allow for proactive communication with customers regarding potential delays. This transparency fosters trust and satisfaction, ultimately strengthening overall customer relationships. Moreover, by anticipating delays, transportation planning costs can be significantly reduced. With advanced knowledge of potential disruptions, companies can optimize routes, adjust schedules, and allocate resources more efficiently, thereby minimizing unnecessary expenses. Especially when some stages of the transportation are handled by third party transportation companies, it is essential to inform them about delays as early as possible.

The supply chain sector, susceptible to various contingencies such as adverse weather conditions, port congestion, or geopolitical events, has suffered throughout the last years from unpredicted delays. By implementing predictive models, the supply chain becomes more resilient, enabling stakeholders to anticipate and mitigate potential disruptions effectively. In essence, predicting delays not only enhances customer service and reduces transportation costs but also fortifies the resilience of the entire supply chain ecosystem. Finally, this predictive tool can easily gain diagnostic power just by a few adjustments of the model. By diagnosing the delays, companies can tackle the bottlenecks of their supply chain directly but also use this information to gain negotiation power. Negotiating is very important, especially for companies that outsource their port-to-port logistics. Using the diagnosing models, useful insights are emerging regarding carrier and port performance, which can be used to gain a negotiating advantage.

The aim of this Master thesis is to dive into the field of supply chain and specifically to use data analytics in order to optimize container shipping. The scope of this thesis was not specified from the start as it is an outcome from a 4-month internship at Sappi.

1.3. Research Questions

The scope of this thesis was influenced by the results of the UX project. While conducting the interviews and creating the thematic analysis around Supply Chain Control Towers, the research questions of this thesis emerged. These research questions are shown below:

RQ1 Is it possible to accurately predict delays in container shipping of a big multinational company?

RQ1.1 What are the main causes of delays in overseas container shipping?
RQ1.2 Which prediction algorithm works best for this task?

The research questions above aim to investigate the accuracy of delay prediction in container shipping for a prominent multinational corporation, this thesis endeavors to delve deeper into the underlying factors contributing to delays in overseas container shipping operations. By identifying the primary causes of delays, this study further aims to provide valuable insights that can inform strategic decision-making and operational improvements within the company's supply chain. Finaly, the investigation extends to evaluating various prediction algorithms to determine which one yields optimal performance for the task at hand.

1.4. Research Boundaries

In this case study, the researchers focus on predicting port-to-port container shipping delays. While a study around final product delivery lead time would be interesting, the scope of this thesis is focused on marine transportation, as it adds business value as explained in chapter 3. Moreover, this study does not take into account the return of the empty container, as from Sappi's point of view this is not an important concern.

Essentially this research uses data from the inttra status messages which are referring only to one sender - Sappi- and are updated from the 11 carrier companies that Sappi was using. There is information about the booking confirmation, upon which the prediction of delays takes place. The quality of the data used in this study is limited by the cooperation and information sharing of the carrier company. This means that some information was missing from the booking confirmation, and this was dealt with in the way that is described in chapter 4.4.2.

The data used in this study originates from a collection process initiated in October 2021. While the initial months inevitably suffered from data quality issues, a turning point came in May 2022. A shift in the company's data collection philosophy led to the creation of a more robust database designed to enhance analytical capabilities. It's important to note that all data employed in this research predates September 2023, as the analytics part of this study ended then. This is nicely illustrated by figure 1, where the points before May 2022 have very few observations compared to the rest of the dataset.

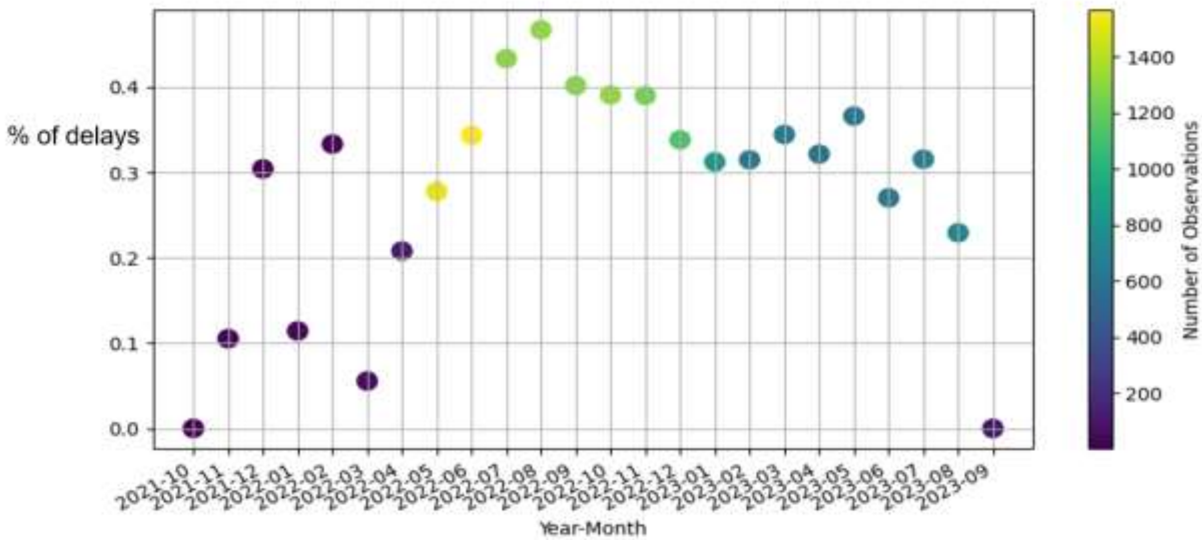

Figure 1: On the y-axis is the mean target which means percentage of delays, on the x-axis is the month. The color of the point contains information about the number of valid entries we have in this dataset for the specific month.

2. Literature Review

This section serves to explore existing literature concerning optimal container shipping and reliability estimation, aligning with the research questions of this thesis. It aims to distill insights, methodologies, and findings to inform the theoretical framework on the research questions of this thesis. Key areas of focus include predictive analytics, optimization techniques, and technological advancements in the field of supply chain, logistics and specifically container shipping. By combining and comparing academic knowledge this review seeks to identify gaps and opportunities for enhancing container shipping efficiency and resilience.

2.1. Methodology

To conduct the literature review, the approach to identify and select relevant research papers was employed. The primary databases utilized were Scopus and Google Scholar, renowned for their comprehensive coverage of scholarly literature across various disciplines. The search was conducted using the following keywords: "Data Analytics container shipping," "Machine Learning Supply Chain," "Logistics Data Analytics," and "Supply Chain Control Towers."

Through this search strategy, a total of 32 research papers were initially identified. These papers were subsequently screened based on predefined inclusion and exclusion criteria. The inclusion criteria encompassed papers published after 2010 to ensure the incorporation of recent advancements in the field. Additionally, relevance to the overarching themes of data analytics, machine learning in supply chain management, and container shipping logistics was paramount for inclusion.

Upon screening, 24 papers met the inclusion criteria and were retained for further analysis. Excluded papers predominantly comprised duplicates, non-English or non-Greek language publications, and those lacking substantial relevance to the scope of the review.

2.2. Supply Chain Control Tower

Supply Chain Control Towers (SCCTs) have emerged as vital tools for enhancing visibility and efficiency in complex supply chain networks. As global markets become increasingly dynamic and interconnected, organizations are turning to SCCTs to provide real-time insights and facilitate proactive decision-making. This review aims to synthesize existing research, identify key success factors, and address barriers hindering the full realization of SCCTs, ultimately informing strategic decision-making and advancing supply chain management theory and practice.

Supply chain control tower is a name that is widely used in business for different specific purposes. Most often, this tool provides increased visibility, agility, and supply chain collaborations. A control tower usually provides support in decision making. However, to build a successful Control Tower for supply chain management there needs to be collaboration between two departments with different expertise, logistics and IT. With IT solutions improving day by day, it is extremely important to continuously develop the Supply Chain Control Tower (Trzuskawska-Grzesińska, 2016).

Integrated planning and execution, logistics visibility, procurement advancements, smart warehousing, efficient spare parts management, and autonomous logistics are integral components of the digital supply chain. Consultations with business leaders and subject matter experts have identified future developments such as open innovation, distributed manufacturing, and new collaboration models, emphasizing the importance of embracing innovative paradigms for sustainability (Lehmacher et al., 2017).

Moreover, the fourth industrial revolution is expected to positively impact sustainability efforts within supply chains (Lehmacher et al., 2017). On the other hand, discussions around blockchain technology emphasize its potential benefits, including time-stamping tracking, automating transactions, and establishing proof of quality provenance (Ganeriwalla et al., 2018). Blockchain's ability to minimize intermediaries, set up self-executing contracts, and facilitate onboarding of new vendors and partners through digital IDs is noteworthy (Ganeriwalla et al., 2018).

Both blockchain technology and digital SCCTs promise end-to-end visibility and task automation, albeit with differing considerations. While blockchain is beneficial when trust is paramount, its implementation should align with the value proposition of the supply chain (Ganeriwalla et al., 2018). This underscores the need for strategic decision-making regarding the adoption of blockchain within SCCTs.

A typical SCCT usually offers 5 types of analytics, planning analytics, descriptive analytics, predictive analytics, diagnostic analytics, and prescriptive analytics. With the predictive power of the SCCT it is possible to manage the company’s processes a lot better (e.g., Inform the customer about a late delivery). A SCCT enhances process automation by having an organized automated schedule, which is tested and can run without much human intervention. In 2021 Patsavellas et al., proposed an End-to-End SCCT, in which all stakeholders of the supply chain would exchange information and doing so, it would lower the bullwhip effect costs. They also conducted some interviews in order to see how different employees perceive the SCCT tool (Patsavellas et al., 2021). In the interview they mostly used closed-ended questions. However, in their analysis, they ignore the sustainability factor.

In this article, it is mentioned that SCCTs should be considered a socio-technical tool and not just a technical tool (Vlachos, 2023). A supply chain control tower can have vision over: 1. Demand forecasting, 2. Production and procurement planning, 3. Supplying raw materials, 4. Product manufacturing 5. Distribution planning (Chae, 2009).

In conclusion, the literature underscores the importance of integrating innovative technologies and collaboration models in SCCTs to enhance supply chain performance and sustainability. Strategic considerations, including the adoption of blockchain, should be informed by the specific needs and context of the supply chain ecosystem.

2.3. AI and predictive algorithms in supply chain and logistics

This subsection serves the literature review by exploring the integration of artificial intelligence (AI) and predictive algorithms within supply chain management. It delves into the significant advancements and applications of AI technologies in optimizing supply chain operations. By harnessing predictive algorithms, organizations can proactively address challenges such as demand forecasting, inventory management, and transportation optimization.

In recent years, researchers have grappled with a fundamental question: whether to rely on traditional methods such as multiple regression etc. or embrace machine learning for optimizing supply chain management. This dilemma underscores the ongoing debate within the field, as both approaches offer distinct advantages and challenges. Traditional methods provide familiarity and established frameworks, while machine learning promises enhanced predictive capabilities and adaptability to complex data patterns. In 2020, Mohamed-Iliasse et al. mention that machine learning applications have greater predictive potential due to the fact that they integrate information from many different sources. On the other hand, it is mentioned that the traditional methods are a lot cheaper, since they usually require less data storage and computational power (Mohamed-Iliasse et al., 2020). The same concept is found in the research conducted by Wisestri et al. While the author highlights the tremendous benefits that ML provides by identifying patterns in different stages of supply chain and emphasizes that ML is giving the opportunity to organizations to best control and coordinate all those stages, also mentions that each of these applications necessitates significant investments in digital transformation, such as developing infrastructure for data storage, implementing AI technologies, and facilitating data mining processes (Wisetsri et al., 2022). However, companies are realizing that data are going to play a crucial role in the future. It is a relatively new field, especially for the corporate environment. As the supply chain is flooded with data, more and more ML applications are going to be used. This thesis provides an example of that.

Supply Chain Management is a very well researched field, with most of the recent research focused on the implementation of technologies associated with the fourth industrial revolution. Machine learning applications are especially interesting to researchers and journals. It is evident that the number of articles around machine learning applications in demand forecasting is on the rise. Specifically for the 5 years from 2016 to 2020, 60% of the articles published and available in “Scopus” and “ScienceDirect” were published in 2019 and in 2020 (Filali et al., 2021). Demand forecasting is the subfield of supply chain that is most often mentioned with machine learning. According to Mohamed-Iliasse et al., approximately 32% of papers regarding ML applications in SC are about demand planning. While Production planning is almost 28%, management and planning of inventories is 25% and management and planning of transportation only 15% as shown in figure 2 (Mohamed-Iliasse et al., 2020). It seems to be that the objectives of this thesis are the least researched application of machine Learning in Supply chain.

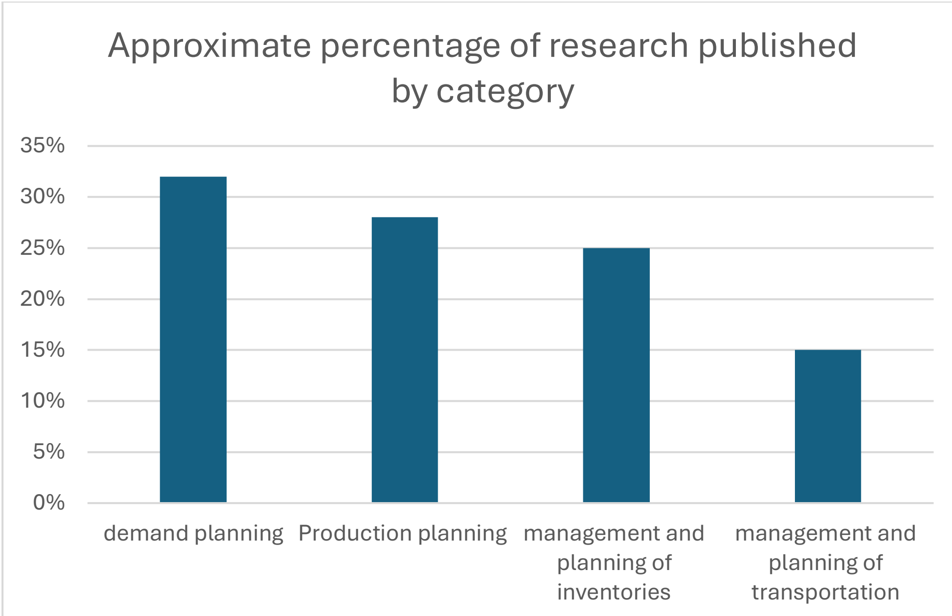

Figure 2: Percentage of articles regarding ML applications in SCM by category.

Final price prediction holds significant importance in the realm of supply chain management (SCM), despite not being directly affiliated with SCM itself. This aspect frequently surfaces in literature concerning machine learning (ML) applications within SCM. The reason for this attention lies in the intricate relationship between pricing dynamics and various SCM functions. Accurate price prediction enables organizations to optimize inventory management, demand forecasting, and strategic decision-making processes. While Amellal et al., focus on the most researched topic, demand forecasting, they also address price prediction and customer sentiment prediction. They compare the performance of the BERT transformer Model, Gated Recurrent Unit and Bayesian network models in each of the categories mentioned above (Amellal et al., 2024). In their paper He et al, use an edge computing-oriented mechanism to predict future prices and crop yields, while also creating a fuzzy neural network as a decision support tool to optimize decisions across supply chain stages (He et al., 2024).

2.4. Predicting container shipping delays

Predicting delays in container shipping is crucial for several reasons, all of which converge on the overarching goal of ensuring smooth and efficient supply chain operations. Firstly, predicting delays allows for proactive measures to be taken to mitigate potential disruptions in the supply chain (Viellechne & Spinler, 2020). By identifying possible bottlenecks or issues in advance, companies can strategize and allocate resources more effectively, minimizing the impact of delays on both costs and delivery schedules.

Moreover, accurate predictions enable better communication and coordination among all stakeholders involved in the shipping process. Whether it's manufacturers, freight forwarders, shipping lines, or port authorities, having foresight into potential delays allows for timely adjustments to be made, reducing confusion and optimizing overall logistics operations (Pani, 2012). This is of significant importance for companies like Sappi that are outsourcing the transportation of their products.

Additionally, predicting delays fosters trust and reliability in the shipping industry. Consistently delivering goods on time is crucial for maintaining customer satisfaction and loyalty. A good predicting model can be transformed into a diagnostic model, and then container carriers would be held accountable for their services. By providing accurate estimates of arrival times, companies can manage expectations and uphold their commitments to clients, thereby enhancing their reputation and competitiveness in the market.

Furthermore, in today's interconnected global economy, where just-in-time production and inventory management are prevalent, any disruptions in shipping schedules can have far-reaching consequences, leading to stock shortages, production delays, and increased costs (Schimi-Levi et al., 2009). Thus, the ability to predict delays becomes not only important but indispensable for the efficient functioning of supply chains worldwide.

Prediction of arrival times is of upmost importance for container terminals as well. Essentially, in 2011 Fancelo et al. in their report are making predictions of arrival times from the point of view of the container terminal. Later they use the predictive algorithm as part of a broader project that is going to be implemented in developing a Decision Support System (DSS) for optimal allocation of human resources. Even though this is somewhat different from the scope of this thesis, the prediction methodology is of some relevance. Due to the complexity and the irregular timeseries of arrivals the prediction was done by using Neural networks (Fancello et al., 2011).

In recent years, there has been growing recognition of the importance of enhancing the resilience of supply chains to mitigate disruptions and uncertainties. One approach to achieving this is through leveraging advanced data analytics techniques and satellite data sources. For instance, Viellechne & Spinler, 2020Viellechne & Spinler, 2020Viellechne & Spinler, 2020utilized satellite data from the Automatic Identification System (AIS), a non-public information utilized by ports and maritime authorities to track vessels, to enhance the resilience of supply chains. Focusing on 54,908 shipments connecting European, African, or Asian ports, their methodology involved three stages.

Firstly, they conducted data cleaning by capturing data from the 100 busiest ports and identified 2,954 shipments with high data quality. Subsequently, they incorporated various variables such as weather conditions, demographics, chokepoint congestions (e.g., Gibraltar), port congestion, travel details, piracy risk, vessel characteristics, seasonality, port size, and shipper/service information. Through variable reduction techniques, they were able to reduce the initial 315 variables to 166, further employing ML-based Lasso Regression to refine the model.

The performance of their models was assessed using Key Performance Indicators (KPIs) such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for regression, and RMSE, Accuracy, and sensitivity for classification. Notably, their findings indicated that Support Vector Machine (SVM) polynomial and neural networks were the best performing models for regression tasks. However, for classification, the Random Forest algorithm exhibited the highest accuracy at 81%, with Neural Networks also performing well (Viellechne & Spinler, 2020).

Furthermore, research by (Salleh et al., 2017) focused on predicting the punctuality of liner container shipping using a decision-making technique, the Fuzzy Rule-Based Bayesian Network. Their dataset comprised information about ports and vessels, with criteria derived from these data. By assigning a scale from 1 to 5 for delays, they produced probabilities indicating the likelihood of a shipment falling into one of these delay categories. This tool was found to be particularly beneficial for shipping companies and port terminals, offering insights into punctuality prediction Salleh et al., 2017).

Similarly, Hathikal et al. (2020) concentrated on predicting ocean freight international transport, considering both product and shipment delays. They employed algorithms such as multinomial logistic regression, decision tree, support vector machine, Naïve Bayes, and K-nearest neighbors to analyze and predict delays. Their research contributes to the broader understanding of predictive analytics in the context of supply chain resilience and transportation logistics (Hathikal et al., 2020). Additionally, their contribution lies in the development of a model that estimates shipment lead time for ocean import shipments, taking into account different stakeholder interests. They evaluated the performance of machine learning models by varying the class label threshold value to classify shipments into different classes and assessed the method's performance with different treatments of categorical features. The different encoding methods they used for categorical variables included converting them to binary variables and ordinal variables. Their best-performing model was multinomial logistic regression, followed by decision trees.

2.5. Novelty of this study

Though container shipping predictions have been done many times in the past as shown in the sub section above, the novelty of this thesis is in the processing of the dataset. While Hathical et al, have done a very similar study, comparing different models and different ways to encode the categorical variables, they didn’t use target encoding.

3. Case study description

Sappi, a global leader in renewable resource utilization, operates across three continents with a diverse product portfolio. Their core business centers on woodfibre-based materials, transforming this sustainable resource into high-quality dissolving pulp, packaging and specialty papers, and even biomaterials. They are a diversified company, positioned at the forefront of innovation within the renewable materials sector.

This section focuses on the importance of the project and the business value it contains. Given Sappi's annual production of approximately 3.7 million tons of paper and 1.12 million tons of pulp, logistics management presents a substantial challenge. The Digital Transformation (DT) team, tasked with enhancing data analytics and tool provisions for other departments, plays a pivotal role in addressing these challenges. To understand the work being done in this thesis is very important to understand the logistics process of Sappi. Sappi does not own a fleet of ships and uses contractors for most of the goods’ transportation. This study is solely focused on predicting delays in port-to-port container shipping, as the main problem of Sappi there is identified. Sappi, is having unreasonably high costs for demurrage and detention

The way Sappi deals with container shipping companies are by negotiating every year the freight price and the volume with each carrier. Relying solely on one contractor is not an optimal practice as it is easy to lose negotiation power. By allocating percentages of the product volume in multiple carrier companies, it incentivizes the carriers to perform better.

However, the container shipping industry was an outdated industry without much regard for customer service, with some efforts of digitization being made in recent years (Toygar et al., 2023). Despite being a crucial component of global trade, the industry has traditionally lagged in terms of customer-centric practices. This lack of focus on customer service often results in poor communication regarding delays, leaving customers frustrated and uninformed. Additionally, inefficiencies such as overbooking, inaccurate scheduling, and insufficient infrastructure worsen delays and disrupt the smooth flow of goods. These challenges not only hinder timely deliveries but also contribute to increased costs and diminished reliability within the supply chain. Therefore, addressing these issues and implementing predictive models to anticipate delays is essential for modernizing the container shipping industry and improving its overall efficiency and customer satisfaction.

3.1. Importance of predicting lead time

Predicting lead time is very important for improving customer and partner relations, especially in port-to-port logistics where timing is critical. For companies outsourcing their logistics services, like Sappi, accurate lead time predictions are essential for maintaining trust and satisfaction among customers and partners. By reliably estimating the time it takes for goods to move from one port to another, businesses can effectively manage expectations and ensure timely delivery, thereby enhancing their reputation and strengthening relationships.

Additionally, having precise lead time predictions provides companies with greater negotiating power when dealing with logistics providers, enabling them to secure favorable terms and maintain competitive advantages in the market. This strategic approach not only fosters collaboration but also drives efficiency and profitability throughout the supply chain. By accurately diagnosing the causes of delays, such as congestion at ports or transportation bottlenecks, companies can proactively address these issues during negotiations with carriers when outsourcing logistics services. Armed with insights into potential challenges and solutions, Sappi can advocate for more flexible terms, improved service levels, and cost-effective arrangements, ultimately ensuring smoother operations and enhanced customer satisfaction.

In the context of Sappi's operations, accurately predicting lead times would be a game-changer for managing demurrage and detention charges. By having a clearer picture of when deliveries will arrive, Sappi can optimize scheduling and avoid exceeding allotted unloading or container return times, which can incur significant demurrage and detention fees. This improved lead time visibility would ultimately lead to cost savings and a more efficient supply chain.

4. Methodology

In this section, a comprehensive overview of the methodologies employed to address the research questions is provided. The first step was to define the general scope of the project. This general scope was defined in the first meeting between the author of this thesis, the thesis advisor and the internship coaches. The general scope was to develop a digital tool or an algorithm that could optimize supply chain operations. After defining the scope of the project, it was decided to move forward with a need finding test for the supply chain control tower that Sappi was trying to build. Step 2 was to conduct some interviews with the supply chain employees. From these interviews, a thematic analysis was created, part of which is presented in chapters 4 and 5. The thematic analysis led to cementing a specific scope for the internship project and the specific research questions for this thesis. After that it was time to create the model according to the technical specification that derived from the need finding. This process is shown in figure 3.

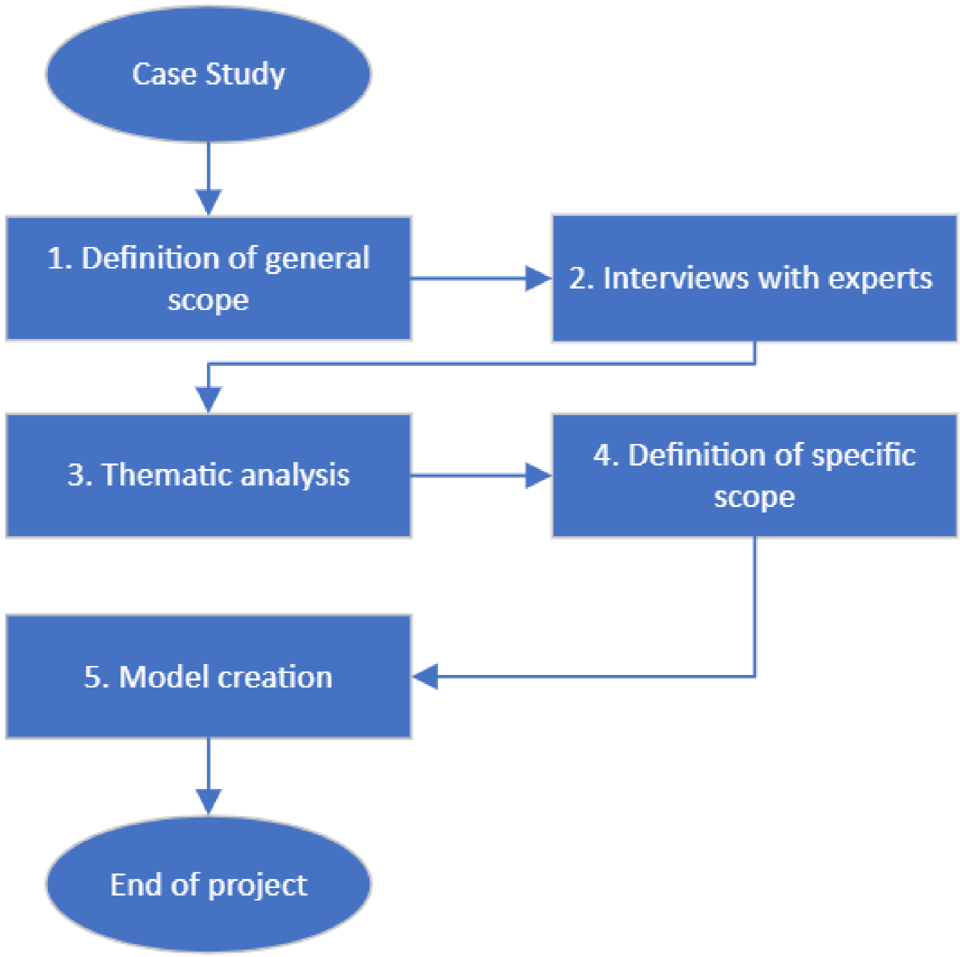

Figure 3: The Methodology of this case study on a very high level.

4.1. Supply chain control tower UX project

Optimizing complex logistics systems necessitates a deep understanding of user needs and main points within the organization. To address this crucial aspect, a comprehensive user experience (UX) project was undertaken. This chapter delves into the details of this project, outlining its objectives, the research methodology employed, and the key insights gleaned from the user experience research.

4.1.1. Project objectives

The objectives of this project were to gather insights, strengths, weaknesses, and opportunities in Sappi's efforts to build a Supply Chain Control Tower (SCCT). The aim of this initiative was to enhance the efforts to create SCCTs by cementing the steps to get to the desired tools. The deliverables of this project were a thematic analysis and a survey. From the results of the thematic analysis, emerged the project about predicting delays in port-to-port container shipping.

4.1.2. Methodology

In order to achieve these objectives, six people were interviewed, four people working in Sappi EU, and one from Sappi NA and one from Sappi SA. More information about them is shown in table 1.

Table 1:This table presents the interviewees, their organization and their role in supply chain as described by them.

Interviewees	Organization	Role
Interviewee 1	Sappi EU	Manager Strategic Logistics Design
Interviewee 2	Sappi EU	Strategic Logistics design Manager
Interviewee 3	Sappi EU	Manager of Analytics & Automation
Interviewee 4	Sappi EU	Managing director
Interviewee 5	Sappi SA	Export Supply Chain manager
Interviewee 6	Sappi NA	Senior Manager

The interviews employed were in semi-structured format, and were recorded, to gather in-depth qualitative data from key stakeholders within Sappi's supply chain department. These interviews focused on exploring a range of topics relevant to the challenges and opportunities surrounding supply chain control towers and how they can improve Sappi's supply chain. The interview themes included:

- Day to Day Operations
- Information sharing
- Sustainability
- Data analytics

The full questionnaire can be found on the appendix. The approximate time for the interviews was 30 to 45 minutes, and all interviewees managed to give a lot of useful information without taking more time than needed.

Thematic analysis was employed to extract key insights from the interviews conducted during the project. This process involved a meticulous review of the interview transcripts, focusing on identifying recurring themes within the participants' responses. Whenever a significant quote was discovered, it was noted in a virtual post it note and saved into a whiteboard. After gathering and analyzing all those quotes, they were organized into themes. These themes represent the core issues, challenges, and opportunities surrounding data analytics in Sappi's supply chain operations and are described in chapter 5.

4.2. Model development

4.2.1. How to *develop* models?

In this section, the methodology of how to develop models is presented. The process commenced with a descriptive exploration of the data, accompanied by the creation of descriptive plots to foster a comprehensive understanding. Subsequently, predictive power was harnessed through the development of a model. Beyond its predictive capabilities, the tool derives diagnostic strength through a detailed elucidation of the underlying model. Although the current culmination of the internship project lies here, an avenue for further enhancement emerges. By incorporating pertinent business logic into the model, it evolves beyond mere prediction, offering valuable prescriptive suggestions. This augmentation not only expands the tool's utility but also positions it as a strategic asset in decision-making processes.

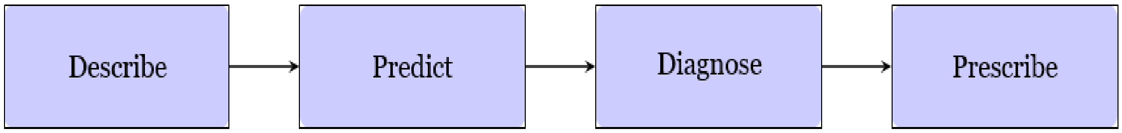

Figure 4: The methodology for creating models.

With a focus on utilizing machine learning and other algorithms to predict port-to-port logistics delays, the approach encompasses several key components aimed at ensuring transparency, reproducibility, and reliability of results. The section begins by detailing the methodology employed for developing the models on a high level. Moreover, the model evaluation metrics chosen to assess predictive performance are discussed, offering insights into how these metrics inform the interpretation of the results. Subsequently, the specifics of the machine learning algorithms employed are delved into, delineating their parameters, hyperparameter tuning processes, and rationale behind their selection. Preprocessing techniques, such as data cleaning and feature engineering, as well as the dataset utilized for training and testing the machine learning models, including its sources, size, and pertinent features are then elucidated to underscore the steps taken to prepare the data for analysis.

As shown in figure 5, the methodology for creating the model begins by importing the data, and then cleaning them. After that the processes of data mining and feature engineering take place. Then there is the stage of training and testing the model. And finally there is the stage where the model needs to be explained to produce diagnostic results.

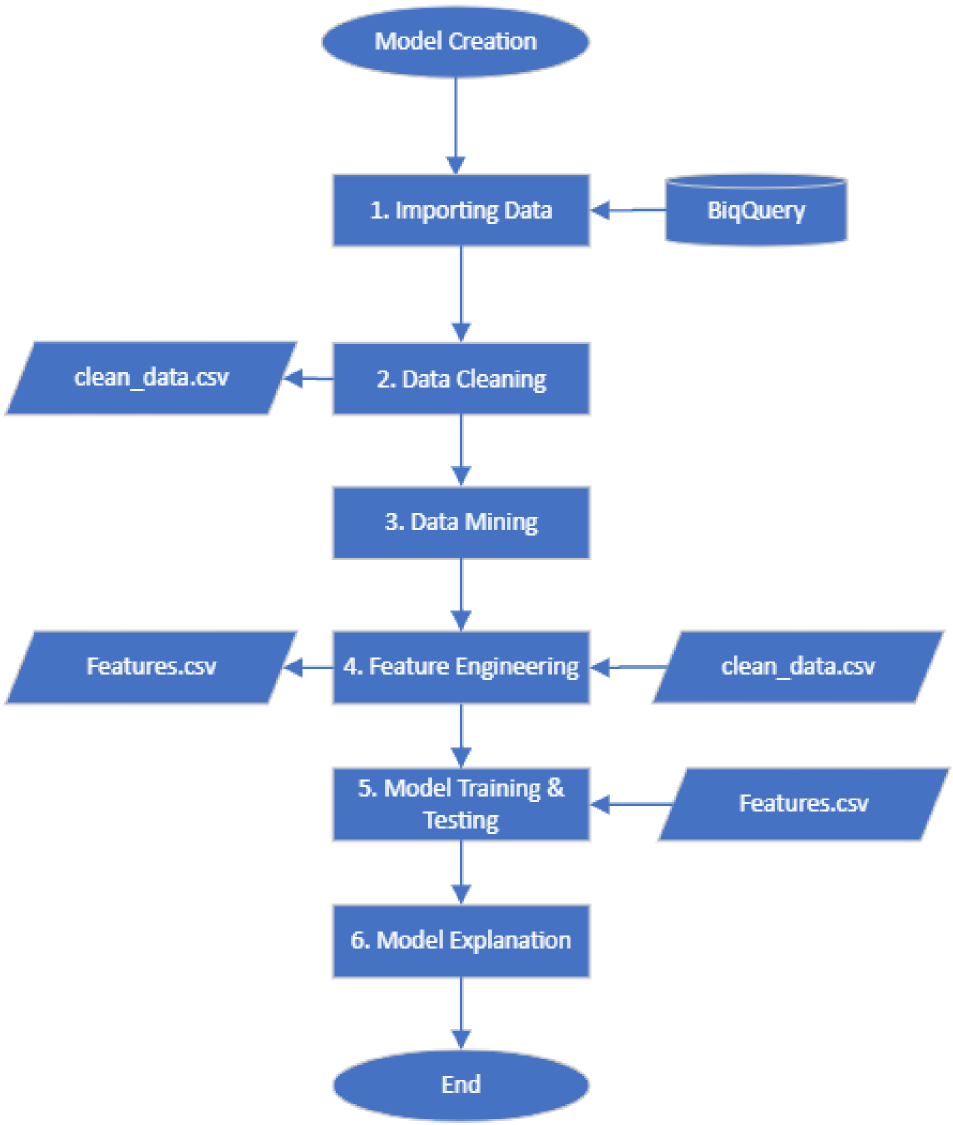

Figure 5: The Methodology of this case study on a very high level.

4.2.2. Model's KPI's

Predictive KPI's

Various machine learning models were created in order to predict whether a shipment will be delayed more than the 8-day threshold. To calculate the performance of the models, the metrics below were used:

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

$$MAE = 1 - Accuracy$$

Key Performance Indicators (KPIs) serve as crucial metrics to evaluate the performance of models and systems across various domains. Accuracy, a fundamental KPI, measures the proportion of correctly classified instances among the total. Precision reflects the proportion of true positive predictions among all positive predictions, emphasizing the model's ability to avoid false positives. Recall, also known as sensitivity, measures the proportion of true positive instances that the model correctly identifies, highlighting its ability to capture all positive instances. F1-score, a harmonic means of precision and recall, offers a balanced evaluation of a model's performance, particularly useful in scenarios with imbalanced classes. Additionally, Mean Absolute Error (MAE) quantifies the average absolute difference between predicted and actual values, providing insights into the model's predictive accuracy. These KPIs collectively offer comprehensive insights into the effectiveness and reliability of models, guiding decision-making processes and facilitating continuous improvement efforts.

Due to the nature of the case, precision is the most important indicator. This decision was made alongside domain experts, and the logic behind it is that a false positive will increase detention & demurrage costs. In essence to maintain a better relationship with partners the most important indicator is precision, while to increase customer service recall would be the indicator to choose by. However, both recall and precision can be manipulated by defining positives, negatives and neither, depending on probability. This technique is explained in the results section in further detail. To provide a holistic assessment of model performance, it's imperative to consider the additional metrics mentioned above. These metrics offer complementary insights into the model's effectiveness and robustness, ensuring a comprehensive evaluation that aligns with the specific objectives and requirements of the case at hand.

Diagnostic KPI's

Diagnostic Key Performance Indicators (KPIs) play a pivotal role in evaluating the performance and interpretability of models, especially in domains requiring in-depth analysis and understanding of model behavior. Mutual Information (mi) and Mean Decrease in Impurity (MDI) are among the crucial diagnostic metrics employed to gain insights into model feature importance, contribution, and decision-making processes.

Mutual Information quantifies the amount of information obtained from one variable through another variable. It measures the dependency between variables, particularly useful for feature selection and understanding variable relationships. In this specific case mutual information was used to check each feature's relevance with the target.

$$MI(X;Y) = H(X) - H(X|Y) \qquad (1)$$

Where:

$H(X) = -\sum_{x \epsilon X} P(x) * \log_2 P(x)$: the entropy of Feature X in the dataset

$H(X|Y) = -\sum_{y \epsilon Y} \sum_{x \epsilon X} P(x,y) * \log_2 \frac{P(x,y)}{P(y)}$: is the conditional Entropy of Feature X for the dataset given the target Y

x,y are all possible values for features X and the target Y respectively

The mean decrease in impurity (MDI) is a metric used to assess the importance of features in decision tree-based models, such as Random Forests. It quantifies the contribution of each feature to the reduction in impurity achieved by splitting on that feature across all trees in the ensemble.

4.2.3. Descriptive analytics

In the process of data mining, the data scientist is trying to understand the patterns, correlations, anomalies, and trends within the dataset. Descriptive analytics often serves as the underlying technique for uncovering patterns and insights within historical data. By using data mining techniques, analysts can identify important trends, relationships, and patterns in data, which can then be used to generate descriptive insights.

More specifically in this process, several plots and tables are created. Soe of the descriptive analytics that were done to complete the needs of this case are presented in chapter 7.1.

4.2.4. Predictive analytics

Predictive methods encompass a wide range of techniques utilized across various fields to forecast future outcomes based on historical data patterns. These methods serve as invaluable tools for decision-making, risk assessment, and resource optimization. One common approach, which is also applied here is regression analysis, which models the relationship between independent variables and a continuous dependent variable, allowing for the prediction of future numerical values. Another widely employed technique is classification, which assigns discrete labels to data points based on their characteristics, enabling the categorization of new instances into predefined classes. Time series analysis focuses on predicting future values of a variable based on its past behavior, often used in financial forecasting, weather prediction, and demand forecasting. Machine learning algorithms, such as decision trees, and neural networks, offer powerful predictive capabilities by identifying complex patterns and relationships within data, thereby enabling accurate predictions across diverse domains. Ensemble methods, like random forests and gradient boosting, combine multiple models to improve prediction accuracy and robustness. Overall, predictive methods provide valuable insights and foresight into future trends and outcomes, empowering organizations to make informed decisions and mitigate risks effectively.

The methodology for crafting predictive models remains steadfast and unwavering across the spectrum, delicately adorned with nuanced differentiations distinguishing network models, regression analyses, and decision tree models. Nevertheless, the fundamental essence remains unaltered and is illustrated in figure 6.

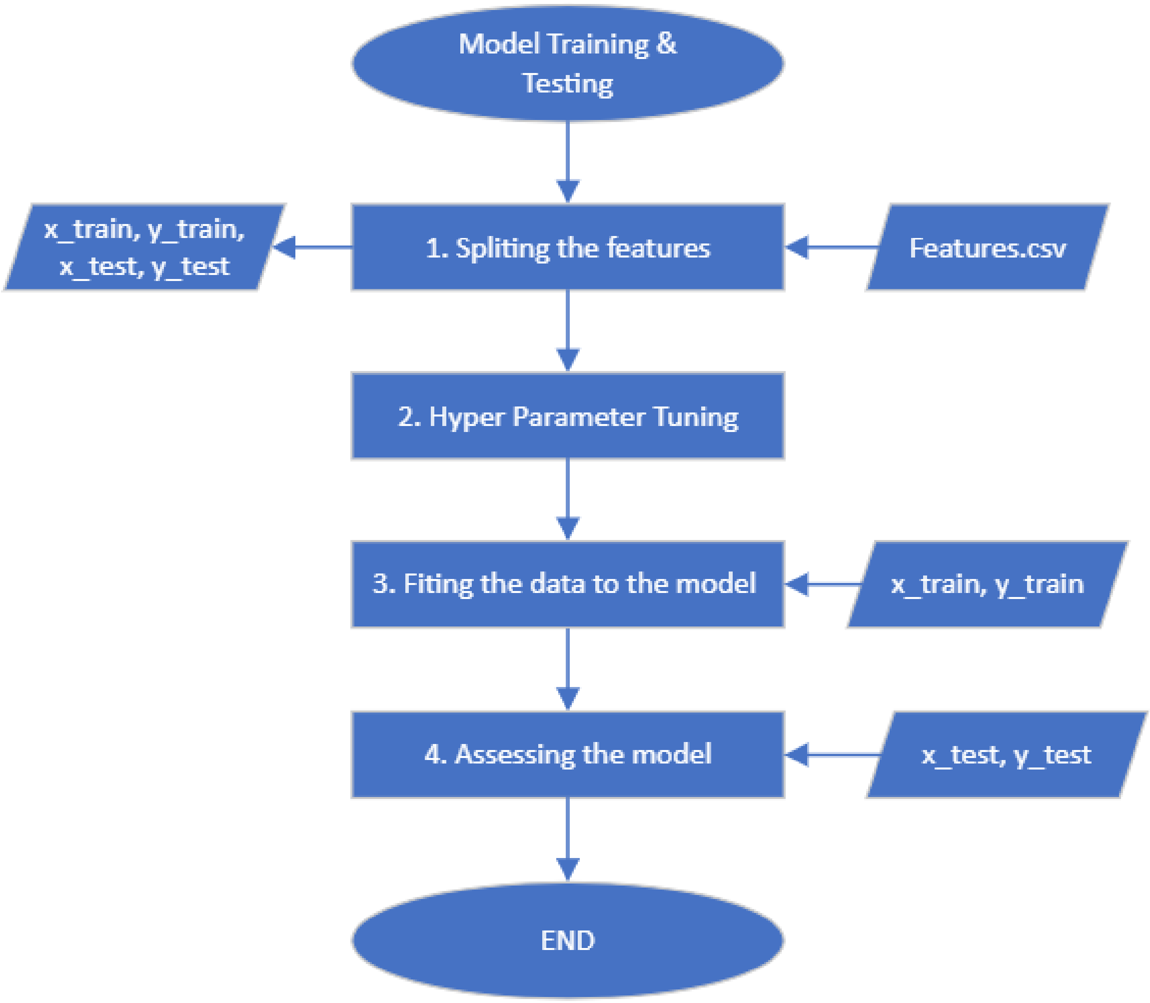

Figure 6: The methodology to create predictive models.

Splitting the features

In this stage the training dataset, comprised of the features and the target are split into two categories, training data and testing data. In this specific case 25% of the dataset was used for the assessment while 75% was used for training the model.

Hyper parameter tuning

Hyperparameter tuning is the process of fine-tuning the hyperparameters of a machine learning model to optimize its performance. Hyperparameters are configurations that are set before the learning process begins, unlike model parameters, which are determined during training.

The process typically involves selecting a range of values for each hyperparameter and then systematically exploring different combinations to find the configuration that yields the best performance according to a chosen evaluation metric (such as accuracy, precision, recall, etc.).

In this case, grid search was employed for hyperparameter tuning while the negative mean squared error was used for scoring. Grid search exhaustively explores a predefined subset of the hyperparameter space by systematically trying every combination of hyperparameter values. Each combination is evaluated using cross-validation to determine the optimal configuration that maximizes the chosen evaluation metric. While grid search can be computationally expensive, it ensures thorough exploration of the hyperparameter space and will most probably identify the best-performing configuration within the specified range of values.

Fitting the data to the model

For this critical step only the training dataset is needed. The algorithm chosen each time (e.g. random forest, deep neural networks, etc.), absorbs these data and adjusts the parameters to achieve maximum compatibility between predicted and actual values. This is the stage in which the model recognizes and discovers relationships between the variables and the target.

Assessing the model

After training the model, it is important to evaluate its performance using the testing data. Common evaluation metrics for classification problems, such as in this case study, include accuracy, precision, recall, F1 score, ROC curve, and confusion matrix.

4.2.5. Diagnostic analytics

Explaining machine learning models and their predictions is crucial for fostering trust, understanding, and acceptance among stakeholders. One commonly used technique is “feature importance analysis”, which identifies the most influential features in the model's decision-making process. By elucidating which features have the most significant impact on predictions, stakeholders gain insights into the underlying mechanisms driving model behavior. Moreover, model interpretation methods, such as partial dependence plots and SHAP (SHapley Additive exPlanations) values, provide intuitive visualizations that depict how individual features affect predictions across different scenarios. These tools facilitate the understanding of complex models by illustrating the relationship between input variables and output predictions. Beyond machine learning models, explaining other types of models, such as statistical models or simulation models, involves similar principles of transparency and clarity. Detailed documentation, model documentation, and clear communication of assumptions and limitations are essential aspects of effectively explaining models to stakeholders. By employing these strategies, organizations can promote transparency, enhance understanding, and foster trust in the models and their predictions, ultimately facilitating informed decision-making processes.

The methodology for crafting diagnostic models upholds a steadfast approach, imbued with nuanced differentiations to distinguish between various techniques such as probabilistic modeling, anomaly detection, and decision support systems. Despite these variations, the fundamental essence of diagnostic modeling remains unaltered, elegantly depicted in the accompanying figure below. Through meticulous analysis and interpretation of data patterns, diagnostic models aim to uncover insights that facilitate precise identification of underlying issues, guiding informed decision-making processes in diverse domains.

4.2.6. Prescreptive analytics

In the domain of prescriptive analytics, the approach to crafting models follows a sturdy framework, relying on advanced algorithms and optimization techniques. Unlike predictive and diagnostic analytics, prescriptive analytics focuses on offering practical recommendations and optimizing decisions. Though the details may differ, the main goal remains the same: providing the best possible actions based on data-driven insights. While delving deep into prescriptive analytics might be beyond the scope of this case study, its importance in guiding smart decision-making is undeniable.

4.3. How do the models work?

This section elucidates the linear algebraic processes underlying the model's operations, providing a comprehensive examination of the mathematical procedures involved. It offers readers a detailed insight into the intricacies of matrix manipulations, vector transformations, and fundamental mathematical concepts employed within the model. From basic principles to advanced methodologies, this section serves as a foundational resource for understanding the model's mathematical framework, enabling researchers and practitioners to grasp its operational mechanisms and optimization techniques. Through clear elucidations and illustrative examples, it highlights the significance of linear algebra in shaping the model's functionality and performance, empowering individuals to harness its insights for further exploration and practical applications.

4.3.1. XGBOOST

XGBoost stands for "Extreme Gradient Boosting". This algorithm uses Boosted Trees in a forest Regression model. The way this works is with Classification and Regression Trees (CARTs). Each tree has leaves, and in each leaf is assigned a value. If one observation falls into a leaf in every tree, then the value is added to it.

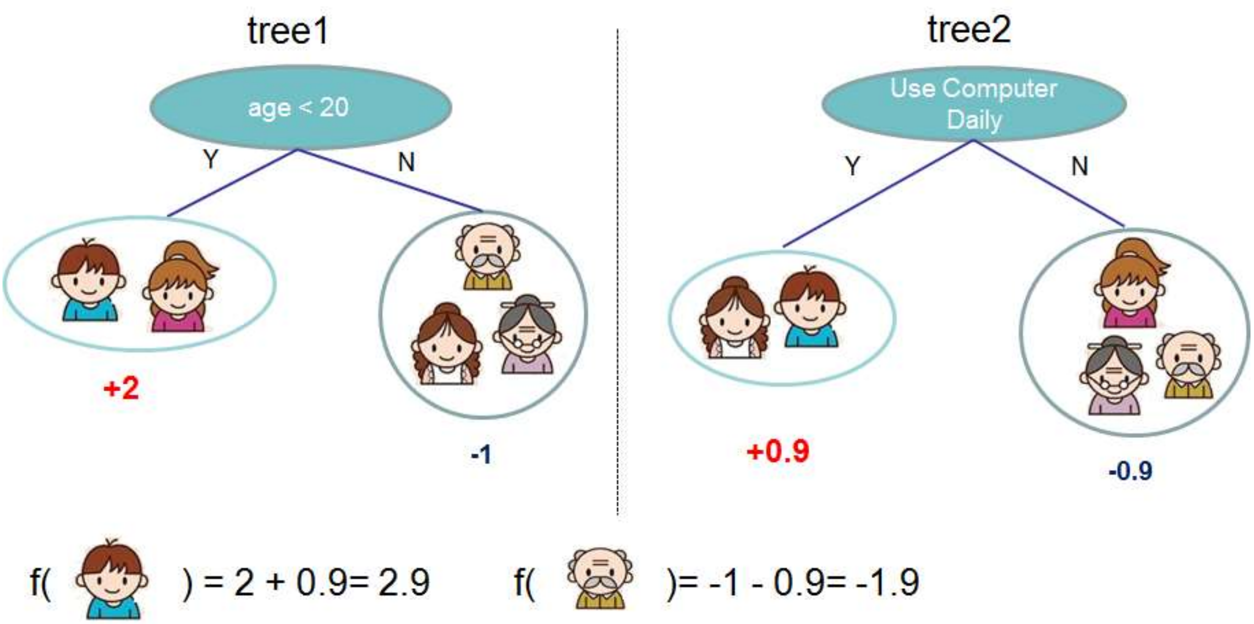

Figure 7: Example of decision trees. In this example the algorithm is trying to predict whether someone will like a videogame.

Figure 8: A specific tree from this case, produced by the xgboost model.

Mathematically this can be written as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \qquad (2)$$

$$obj(\theta) = \sum_i^n l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \omega(f_k) \qquad (3)$$

Where K is the number of trees, f_k is a function in the functional space F which is the set of all possible CARTs and $\omega(f_k)$ is the complexity of the tree f_k and l is a loss function of our choosing.

The way tree boosting works is with additive training. It is not easy to optimize all the trees at once. Moreover, this problem is harder than any other gradient based optimization problem because we cannot just take the gradient. The way around this problem is by adding a tree for every step t of the algorithm.

$$y_i^{(0)} = 0$$

$$y_i^{(1)} = f_1(x_i) = y_i^{(0)} + f_1(x_i)$$

$$y_i^{(2)} = f_1(x_i) + f_2(x_i) = y_i^{(2)} + f_2(x_i)$$

....

$$y_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = y_i^{(t-1)} + f_t(x_i) \qquad (4)$$

The question of how to choose which tree to add remains. The answer is simple, the tree that minimizes the Objective function.

$$Obj^{(t)} = \sum_i^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_{t=1}^{t} \omega(f_t) \quad (5)$$

After applying Taylor's expansion:

$$Obj^{(t)} = \sum_i^n \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t) + c \quad (6)$$

Where $g_i = \frac{\partial l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}}$ and $h_i = \frac{\partial^2 l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial^2 \hat{y}_i^{(t-1)}}$.

The reason it is preferable to use this format, is because our decision variable $f_t(x_i)$ is appearing only in first and second order terms regarding the loss function. The complexity function can be written in the same format as the loss function.

The decision variables need to change now because it is easier to define the complexity function with the leaf values w.

$$f_t(x) = w_{q(x)}, w \in R^T, q \in R^d \rightarrow \{1, 2, \dots, T\} \quad (7)$$

$$\omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \quad (8)$$

Where $q(x)$ is a function that assigns each data point to a leaf. By replacing f into the loss function, the final objective function is.

$$Obj^{(t)} = \sum_{i=1}^{n} \left[l\left(y_{i,} \hat{y}_i^{(t-1)}\right) + g_i w_{q(x)} + \frac{1}{2} h_i w_{q(x)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \; (9)$$

After keeping only what is affected by our design variables.

$$Obj^{(t)} = \sum_{i=1}^{N} \left[g_i w_{q(x)} + \frac{1}{2} h_i w_{q(x)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2$$

$$Obj^{(t)} = \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j}^{N} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j}^{N} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (10)$$

Where $I_j = \{i \,|q(x_i) = j\}$. If we assume that $G_i = \sum_{i \in I_j}^{N} g_i$and $H_i = \sum_{i \in I_j}^{N} h_i$. This quadratic objective function is minimized when:

$$w_j = -\frac{G_i}{H_i + \lambda} \quad (11)$$

$$Obj^* = -\frac{1}{2} \sum_{j=1}^{T} \left[\frac{G_j^2}{H_j + \lambda} \right] + \gamma * T \quad (12)$$

4.3.2. Random forests

This subsection on random forests aims to elucidate the operational mechanisms of this widely utilized ensemble learning technique. Random forests operate by constructing multiple decision trees during the training phase.

Figure 9: a tree produced from the random forest model.

Each decision tree is built on a random subset of the training data and features. Throughout the construction of each tree, a random selection of features is considered at each node, with the best feature among them chosen based on a specified criterion, typically maximizing information gain or minimizing impurity. This injected randomness into the tree-building process helps to decorrelate the individual trees, rendering the ensemble more robust and less prone to overfitting, which is a problem in noisy datasets such as in this case. Once the ensemble of decision trees is constructed, predictions are made by aggregating the predictions of all individual trees. For classification tasks, the class with the most votes among the trees (mode) is assigned as the predicted class, while for regression tasks, the average of the predictions across all trees is taken as the final prediction. The random forest algorithm offers several advantages, including its ability to handle high-dimensional data, manage missing values, and provide estimates of feature importance, which can aid in model interpretation. This is of significant importance for the specific project.

Decision trees are particularly useful in scenarios where noisy datasets are present. In such cases, decision trees leverage metrics like entropy to discern meaningful patterns amidst the noise, effectively filtering out irrelevant information and providing robust predictions. This adaptability to noisy environments underscores the utility of decision trees in real-world applications, complementing the broader capabilities of random forests in addressing diverse machine learning challenges. Overall, decision trees and random forests represent potent and adaptable machine learning techniques suitable for a myriad of applications, as expounded upon in the thesis.

4.3.3. Logistic regression

This subsection aims to elucidate the operational mechanisms of multiple logistic regression analysis, a cornerstone technique in statistical modeling and predictive analytics. Multiple logistic regression analysis involves modeling the relationship between a dependent binary variable and two or more independent variables. At its core, the process entails estimating the coefficients of each independent variable to determine their respective contributions to the variation in the dependent variable. This is achieved through fitting a linear equation to the data, where the coefficients represent the slope of the relationship between each independent variable and the dependent variable. The multiple regression model accounts for the combined effects of all predictors, allowing for a comprehensive understanding of how changes in the independent variables impact the outcome of interest. Key to the multiple regression analysis is the interpretation of regression coefficients, which elucidate the strength and direction of the relationships between variables. Additionally, logistic regression facilitates hypothesis testing, enabling researchers to assess the significance of the relationships and the overall fit of the model. Through thorough exploration and interpretation of the regression output, practitioners can gain valuable insights into the underlying dynamics of the data and make informed decisions based on the predictive capabilities of the model.

4.3.4. Deep Neural Networks

This subsection delves into the operational intricacies of deep neural networks (DNNs), a powerful class of machine learning models that have revolutionized various fields, including computer vision, natural language processing, and reinforcement learning. Deep neural networks consist of multiple layers of interconnected neurons, organized in a hierarchical fashion. Each neuron receives inputs from the neurons in the preceding layer, performs a weighted sum of these inputs, applies an activation function to introduce non-linearity, and passes the result to the neurons in the subsequent layer. During the training phase, DNNs learn to map inputs to outputs by adjusting the weights and biases of the neurons through an optimization process known as backpropagation. This process involves iteratively updating the model parameters to minimize a predefined loss function, typically using gradient descent or its variants. The depth of the network, characterized by the number of layers, allows DNNs to learn complex hierarchical representations of the input data, enabling them to capture intricate patterns and dependencies. Additionally, techniques such as dropout regularization, batch normalization, and various activation functions help prevent overfitting and improve the generalization ability of the model. Despite their computational complexity and the need for large amounts of data, DNNs offer state-of-the-art performance across a wide range of tasks, making them indispensable tools in modern machine learning and artificial intelligence research. Through comprehensive exploration of the architecture, training process, and optimization techniques, this subsection aims to provide a deep understanding of how DNNs operate and their significance in advancing the frontiers of AI technology.

4.3.5. Knowledge Based Models

This subsection explores knowledge-based models, which utilize structured representations of domain-specific knowledge to perform reasoning and decision-making tasks. These models encode knowledge into formal representations like ontologies or knowledge graphs, enabling inference and decision-making based on logical rules and relationships. By employing reasoning algorithms such as deductive or probabilistic reasoning, they derive conclusions from encoded knowledge, offering transparency and interpretability in decision-making. Despite their reliance on domain expertise, knowledge-based models excel in various domains, including expert systems and semantic web applications.

4.4. Feature Engineering

This section is essential to understand the dataset and the features that derived from it. Feature engineering is the art and science of selecting, transforming, and creating new features from raw data to improve the performance of machine learning models. It plays a pivotal role in the success of a predictive model, as the quality and relevance of features directly influence its ability to capture patterns and make accurate predictions. Effective feature engineering involves domain knowledge, creativity, and a deep understanding of the data at hand. By carefully crafting features, one can enhance model interpretability, reduce overfitting, and uncover hidden relationships within the data, ultimately leading to more robust and reliable predictions. In essence, feature engineering serves as a bridge between raw data and predictive models, empowering data scientists to extract meaningful insights and drive actionable decisions from complex datasets.

This process, despite not being a part of the development or the tuning of the model, needs to be done with attention to detail as it can be the main driver of the performance of the models. For this case, feature engineering is challenging due to the nature of the data used. As shown in section 4.4.24.4.2 there are a lot of categorical data that need to be transformed. In this thesis, the target encoding technique is implemented, as it adds value to the prediction of this specific model and is recognized for its capacity to enhance model outcomes and performance in general. Overall, meticulous feature engineering is crucial for unlocking the full potential of predictive models.

4.4.1. Basic principles

Feature engineering involves several basic principles that aim to enhance the predictive power of machine learning models. Firstly, it begins with understanding the domain and the problem at hand, which helps in identifying relevant features. Next, feature extraction involves transforming raw data into meaningful features that capture essential information. This can include techniques like scaling, normalization, or polynomial transformations. Additionally, feature selection focuses on choosing the most informative features while discarding irrelevant or redundant ones to improve model efficiency and interpretability. Another key principle is handling categorical variables, which often involves techniques such as one-hot encoding, label encoding, or target encoding. Finally, domain-specific knowledge and creativity play a crucial role in crafting new features that capture complex relationships within the data. Overall, these principles guide the process of feature engineering towards building robust and effective machine learning models.

In this case the methodology employed for feature engineering is simple yet effective as shown in chapter 5. The initial step was to create the knowledge date variable. This variable is not used to train the model, but it is needed to make the features target encoded. Essentially it is the date that the information needed for the prediction to be made exists, without knowing the result. After that, the features used for training are created. Then there are the steps of treating datetime variables and target encoding, which are explained later in this chapter. After those steps are completed, model training takes place and then the evaluation. If the model performance is deemed acceptable, then there is no need for more feature engineering and the model development stops there. If it is not, then new features are created in order to improve performance, as shown in figure 10.

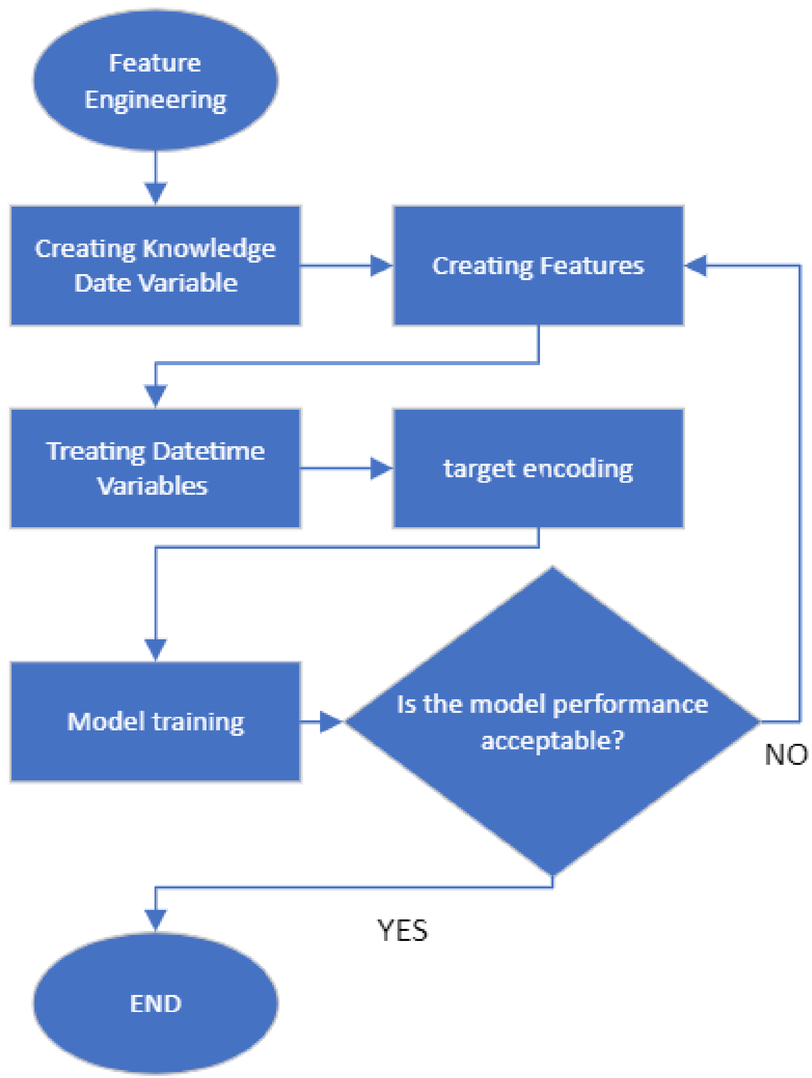

Figure 10: The methodology applied for feature engineering in this case.

4.4.2. The case dataset

This section provides the necessary background for this case. Other than the basic principles of feature engineering, in order to have a better understanding of which techniques were adopted or rejected and the logic behind these decisions, it is important to comprehend the nature of the data in hand.

The dataset used in this study was sourced from the status messages given by INTTRA, albeit with minor logical filters applied to eliminate blatantly inaccurate observations. INTTRA is an ocean booking platform that provides shippers with data sent by the carriers or the logistics service providers. This way they are able to improve their planning and decision making.

These status messages were put into a BigQuery table with some SQL processing to make sure that there is high data quality. From there utilizing a query in Vertex AI, the data is transformed into a pandas DataFrame. At this stage the data is ready for further processing and can be saved as a csv file. This procedure is shown in figure 11.

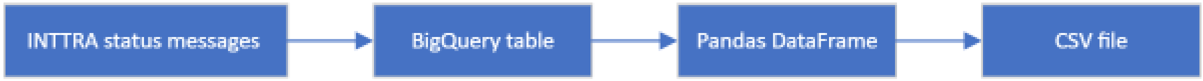

Figure 11: The preprocessing of the case dataset.

The data presented and described in table 2, are part of the initial dataset after the, and were retrieved from BigQuery using a SQL query. The data in this case study range from October 2021 until September 2023.

Table 2: Dataset columns, their description, and missing values.

Column	Description	missing values
Booking number	This Column is used to identify each shipment	0
Equipment identifier	This Column is used to identify each container	0
Last BC message	The last date-time we received a booking confirmation message	18631
Carrier	This Column describes the carrier code of the carrier that promised to deliver the containers	0
Port of Origin Datetime	This Column is the promised date-time that the containers are going to depart from the port	12028
Port of Origin	This Column is the promised port that the containers are going to be departed from	18757
Port of Discharge datetime	This Column describes the datetime that the shipment is going to be discharged	12254
Port of Discharge	This Column is the promised port that the containers are going to be discharged at	18757
Transshipment	This Column is the first port of transshipment the vessel arrived at	na
Vessel Arrival Datetime	This Column is the actual date-time that the vessel arrived in the appropriate port	5270
Vessel Discharge Datetime	This Column is the actual date-time that the containers were discharged in the appropriate port	4896
Vessel Departure Datetime	This Column is the actual date-time that the vessel departed from the appropriate port	6075
Vessel Load Datetime	This Column is the actual date-time that the containers were loaded in the vessel	5996
First Vessel Name	This Column is the name of the first vessel	5275
Last Vessel Name	This Column is the actual date-time that the containers were loaded in the vessel	6075

To summarize the information shown in table 2, the data quality of this dataset is poor. Considering the nature of the specific case, missing values on Port_of_Dicharge_time are useless in the creation of the model, so they are dropped. Following the consultation of domain experts, where Vessel_Discharge_time was missing it was replaced by Vessel_Arrival_Datetime. However, the rows where both these columns were null, were dropped.

These columns were transformed into features. This is done for two reasons, on one hand is impossible for the machine learning model to understand categorical data that are not encoded. On the other hand, by using feature engineering techniques, the columns that are transformed into features provide more useful information for the model.

Table 3: Feature columns description.

Feature	Description
Scheduled duration	This Feature is the difference in days between the scheduled datetime of departure and the scheduled datetime of arrival (e.g. 3 days)
Usual duration	This Feature is the usual duration for a specific route
Carrier	This Feature is a number that represents a Carrier or a missing value
Carrier Performance	This Feature is percentage of delays for entries before the prediction date of this entry for the specific carrier
Port of Origin	This Feature is the promised port that the containers are going to depart from
Port of Origin Performance	This Feature is percentage of delays for entries before the prediction date of this entry for the specific port of origin
Port of Discharge	This Column is the promised port that the containers are going to be discharged at
Port of Discharge Performance	This Feature is percentage of delays for entries before the prediction date of this entry for the specific port of discharge
Countries of Discharge	This Column is the promised country that the containers are going to be discharged in
Transshipment	This Feature is a Boolean value that indicates whether the shipment has a transshipment
Vessel Discharge Month	This Feature is the scheduled month that the containers are to be discharged in the appropriate port
Vessel Departure Month	This Feature is the that the containers are to depart in the appropriate port
Order Size	This Feature is the number of containers that are going to be loaded into the ship
Vessel Discharge Month Performance	This Feature is the Percentage of delays that happened in this month of Discharge until the current date
Vessel Departure Month Performance	This Feature is the Percentage of delays that happened in this month of Departure until the current date

4.4.3. Encoding of Categorical Variables

This step is very important to the methodology, as it is impossible to use categorical variables as training data. The three techniques tried in this case study are one-hot encoding, label encoding and target encoding.

Table 4: Categorical Variables and the number of their unique values

Categorical Variable	Distinct Values
Port of Load	19
Port of Discharge	147
Carrier	12
Transshipment Port	110

The first three variables mentioned in table 4, which presents categorical variables and the number of their unique values, were encoded through the traditional methods. However, the Transshipment port was encoded in a different way explained below.

One hot encoding

One-hot encoding converts categorical variables into binary vectors where each category is represented by a binary value (0 or 1). Each unique category gets a separate binary column, indicating its presence or absence. This technique enables machine learning algorithms to handle categorical data effectively.

This technique failed to provide good results due to the nature of the dataset. Even though it is the most popular method to encode categorical variables, it does not work well when there are a large number of categories. As shown in table 4 there are 147 unique values for the port of discharge. This would create a very large Dataframe, that the computing machine could not handle.

Label encoding

Label encoding converts categorical variables into numerical labels. Each unique category is assigned a unique integer value. This method creates ordinal relationships between categories, which might not be appropriate for all algorithms. However, it is useful when the categorical variable has an inherent order or ranking.

This technique was implemented, and it produced respectable results as shown in section 5.

Target encoding

Target encoding, also known as mean encoding, replaces categorical variables with the mean of the target variable for each category. It transforms categorical variables into continuous values based on the target variable's distribution within each category. This encoding method can help capture valuable information from categorical variables, especially in regression tasks, but it may lead to overfitting if not applied carefully, particularly when dealing with high-cardinality categorical variables.

This technique produced the best possible results.

Transshipment

The way the transshipment column was dealt with was by binarizing whether there was a transshipment or not. The value 0 was added to the column wherever Transshipment was None and 1 wherever Transshipment had any other value other than none.

4.4.4. Encoding of Datetime values

There are a few ways to encode datetime value. In this case studies the way those columns were created was by categorizing them by the month and then using the target encoding and label encoding technique for testing our models. In the case of Vessel Departure month, missing information was labeled 0.

5. Results

In this section, the presentation focuses on the results obtained from the UX project as long as the evaluation of various models and the two encoding techniques that were employed. Through meticulous analysis and experimentation during the data mining stage, key findings emerge to shed light on the efficacy and performance of these models. Moreover, there are findings around the dataset discovered during the data mining phase. Finally, results from the diagnostic inquiries of this thesis are presented below.

5.1. Supply Chain Control Tower

Leveraging the insights gleaned from the UX project, this sub-section explains how the project ended up as being the development of a predictive model for container shipping delays

The key takeaways from the thematic analysis are presented below:

1. Customer Service and Transport planning will improve the most with the SCCT.

Almost all the interviewees mentioned that customer service and transport planning would improve the most by creating a SCCT. Having visibility in all logistics and supply chain operations would improve communication with the client regarding the final Estimated time of Arrival (ETA). Moreover, it would improve transparency from Sappi's side and trust from the client's side.

On a slightly different note, an interviewee mentioned that the clients only care about the ETA not deviating from what was agreed, and no other information around the product's route. However, having visibility over the product throughout all logistics operations would be beneficial for better transport planning. Sappi would have more data on why contingencies happen and would be more responsive to them.

2. There is a need for more predictive and diagnostic tools in Sappi's supply chain.

Sappi's supply chain managers communicate effectively and often with the digital transformation team, facilitating the development of descriptive tools. These tools harness the power of bigQuery, a powerful data warehouse and analytics platform, and Looker, an intuitive business intelligence tool, to analyze and visualize data insights efficiently. This sets a concrete base for developing predictive and diagnostic tools. One interviewee mentioned that there has been some initiative with the project44 tool, which is a high velocity supply chain platform, however more personalized tools could be created by the digital transformation team in order to perfectly represent Sappi's supply chain.

3. Data quality is going to be an issue while building the SCCT.

All interviewees mentioned the poor quality of data sent by the carrier companies. Container carriers use very traditional methods to operate. Moreover, the industry standards are very low, so they do not need to digitalize and offer better information sharing as it was not heavily demanded by their clients.

4. The SCCT should have restricted views for the customer where they could see the condition and the location of their product.

One of the interviewees mentioned that the client should have a role or a view in the supply chain control tower. When asked about it directly all of the interviewees agreed to it. It is a good idea to have the client informed from an application. It would cut down on unnecessary conversations and save time.

5. Sappi's digital tools should have some sustainability elements.

Sappi operates in a sustainable way. Some interviewees mentioned that they use EcoVadis to rate supplier's operation from a green perspective. Sustainability is not an option anymore. It is time for big companies to make decisions and make their operations eco friendlier.

5.2. Data mining

The data mining process is a pivotal component in extracting valuable insights from vast datasets, enabling informed decision-making, and uncovering hidden patterns within complex data structures. In this section, the results obtained through rigorous data mining techniques applied to our dataset are presented. The findings offer a comprehensive understanding of the dataset, shedding light on both anticipated and unforeseen relationships, thus paving the way for further investigation and practical applications in the predictive and diagnostic phases of the development.

5.2.1. Carriers

The exploration of carriers and their interplay with Sappi holds significant promise in unraveling intricate dynamics within the transportation ecosystem. In this investigation, the aim is to dissect the symbiotic connections between Sappi and carriers, scrutinizing factors contributing to timely deliveries and potential bottlenecks leading to delays. This endeavor not only offers a deeper understanding of the intricate network of relationships within the transportation domain but also equips stakeholders with actionable intelligence to mitigate delays, bolster collaboration with carriers, and elevate the overall service quality provided by Sappi.

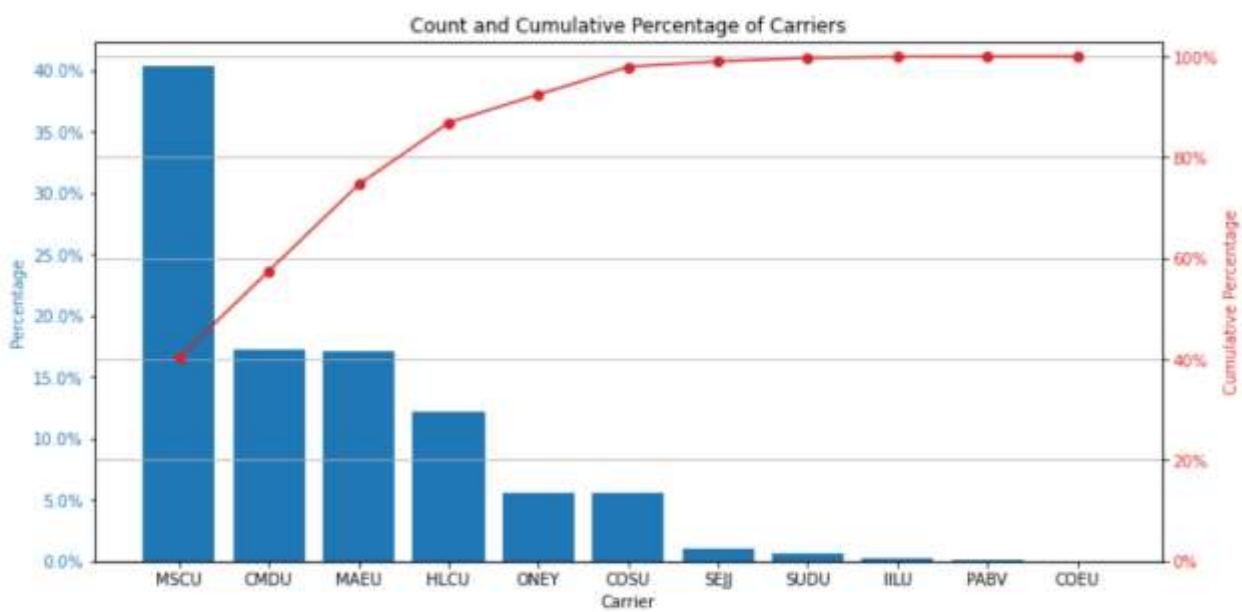

Figure 12: a bar chart of the most used carriers for Sappi and a line plot with the cumulative percentage of them.

Figure 12 shows how the pareto principle is evident in Sappi's use of carriers. Just 27% of the carriers used comprise nearly 75% of all shipments. Of course, this is done to achieve better prices with the carriers they use the most. It seems like the Mediterranean Shipping Company (MSCU) is the most commonly used with 40% of total shipments, followed by the Compagnie Maritime d Affretement Compagnie Generale Maritime (CMDU) and the Maersk Line (MAEU) with 17% of all shipments each. Fourth most used carrier is the Hapag Lloyd Conntainer line (HLCU) with 12% of total shipments. The rest of the carrier SCAC codes are explained in Table 5.

Table 5: The rest of the SCAC codes for the carriers sappi use.

SCAC code	Carrier Line
ONEY	Ocean Network Express
COSU	COSCO Container Lines
SEJJ	Seago Line
SUDU	Hamburg Sud
IILU	Independent Container Line
PABV	Pacific International Lines
COEU	Cosco Shipping Lines (Europe)

While it would make sense that Sappi would focus on the few carriers they prefer to achieve better prices through bulk shipments, this is not a favorable practice. The way they negotiate their prices per container prices is by yearly meetings with the transporters. If Sappi gave more than 50% of their product to just one carrier, that would give that company tremendous negotiating power. By spreading their shipments across 11 carrier companies they achieve better partner relationships while they are not giving away full negotiating power to them.

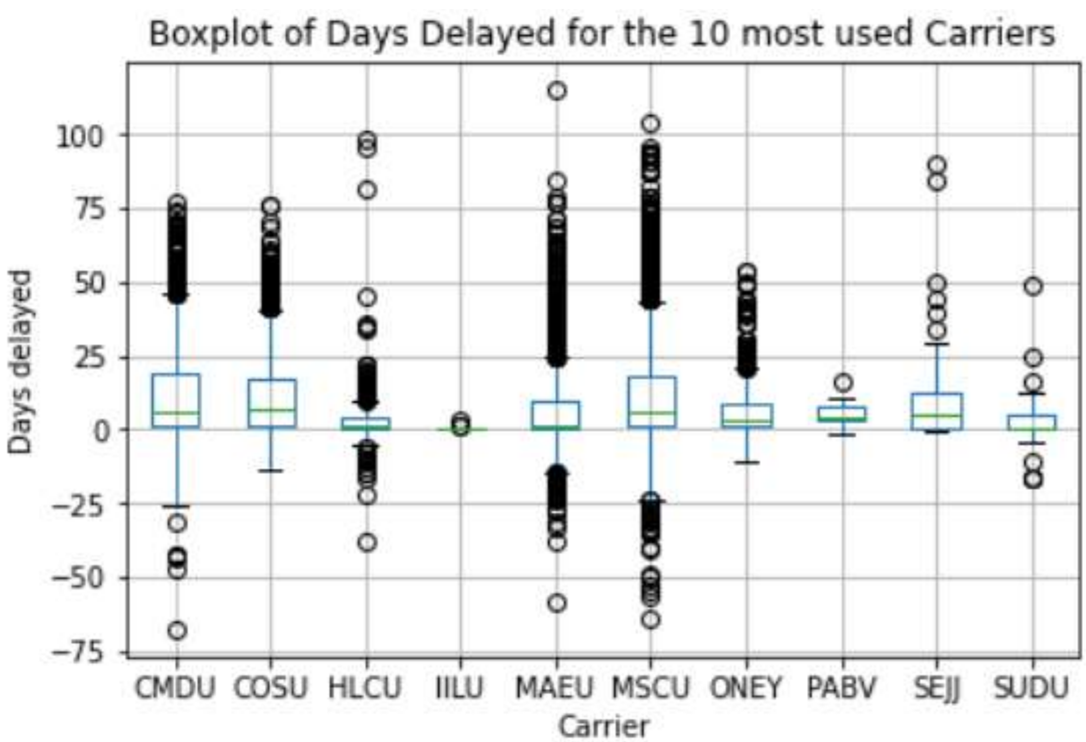

Figure 13: A boxplot showcasing the behavior of Sappi's most used carriers regarding lead time.

Figure 13 not only highlights the consistency of the most used carriers but also shows that there are a lot of outliers in the datasets. These outliers produce valuable information as to the prediction of late shipments and they are not to be removed from the dataset.

Table 6: This table is showing in a more concise and clear manner the 25% percentile (Q1), the median and the 75th percentile (Q3), for each carrier

SCAC	Q1	Median	Q3
CMDU	0.9	5.6	19.0
COSU	1.7	6.5	17.3
HLCU	0.5	1.6	4.3
IILU	0.3	0.3	0.5
MAEU	0.0	1.6	9.8
MSCU	1.5	6.0	18.4
ONEY	1.1	3.6	9.1
PABV	2.8	3.9	7.5
SEJJ	0.3	5.3	12.4
SUDU	0.2	0.8	5.5

Table 6 provides information that already exists in figure 13 but in a more concise and clear manner. As expected, the most used carriers have the most outliers and the widest ranges of observations. Due to the amount of volume, they handle, the most often used carriers produce more robust ranges. This can’t be used to compare carriers with each other, as they have different sample sizes. However, this boxplot can be used as an argument in the yearly negotiations of the freight prices. Out of the 4 most used carriers it seems that HLCU has the most desirable behavior. The median of those observations is close to 2, which means that almost as many shipments were late as early. It also has one of the best interquartile ranges with most of the observations nearly on time. However, outliers are not absent. MAEU also has a median value around 2. The quartile range is also acceptable, however there are extreme outliers both for late and early containers. CMDU and MSCU have quite similar behavior regarding delays. Their median values are quite low, and their quartiles are the biggest of the whole group. However, it is not accurate to compare their performance with the others due to the different sample sizes.

5.2.2. Scheduled duration and transshipment

In the domain of data mining, the examination and interpretation of distinct features within datasets play a pivotal role in uncovering valuable insights and patterns. Among these features, "Scheduled duration" and "transshipment" stand prominently, initially perceived, by experts to bear close resemblance and to harbor significant mutual information, however this was not the case. This subsection endeavors to illuminate the data mining process as applied to these two features, dissecting their purported similarities, and elucidating the underlying mechanisms that distinguish them within the context of the broader dataset analysis. Through systematic exploration and analysis, the aim is to unravel the intricate relationships and unique attributes inherent to each feature, thereby contributing to a deeper understanding of their respective roles in the dataset and their potential implications for decision-making processes.

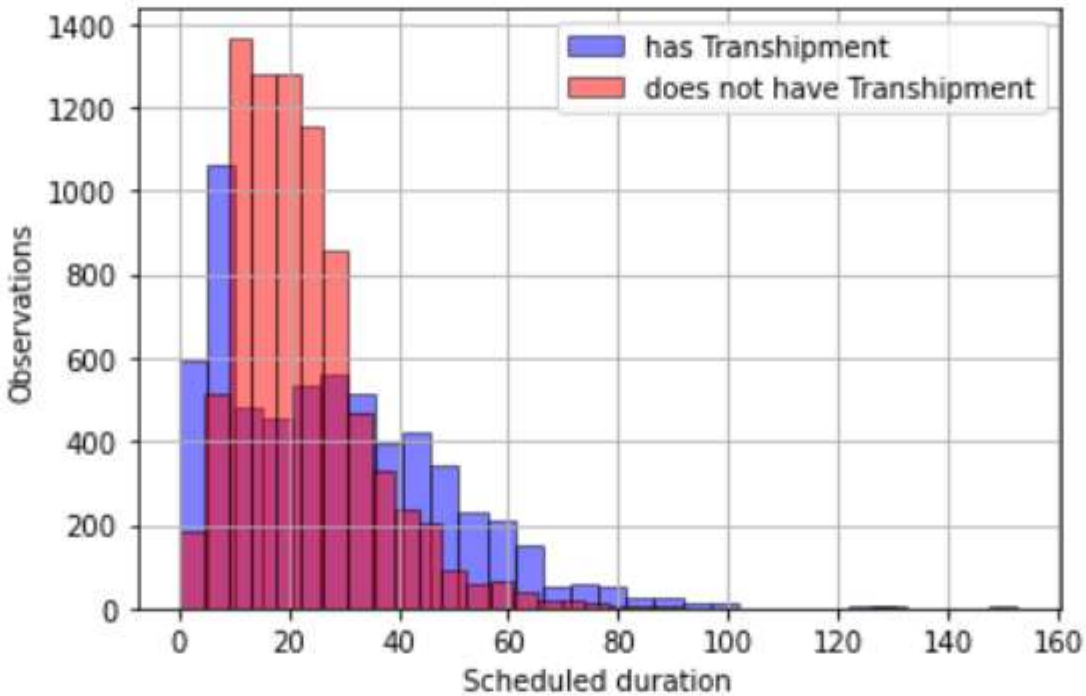

Figure 14: On the y-axis is the number of entries that fall into each bin. On the x-axis there is the scheduled duration. There are two plots; the blue distribution of scheduled duration in case there was a transshipment, and the red is the same plot but in case there was no transshipment.

Figure 14 illustrates that the absence of a transshipment is associated with a shorter scheduled duration. In contrast, the presence of a transshipment does not provide a clear conclusion regarding the duration. Specifically, the data indicates that the median for cases without a transshipment is 20.35 days, whereas for cases with a transshipment, the average duration increases to 25.06 days. This suggests that transshipments tend to lengthen the overall duration, but other factors also influence this outcome, preventing a definitive conclusion.

5.2.3. Departure month

The Month of Departure serves as a critical feature in various analyses, contributing significant insights into patterns and trends within this specific data set.

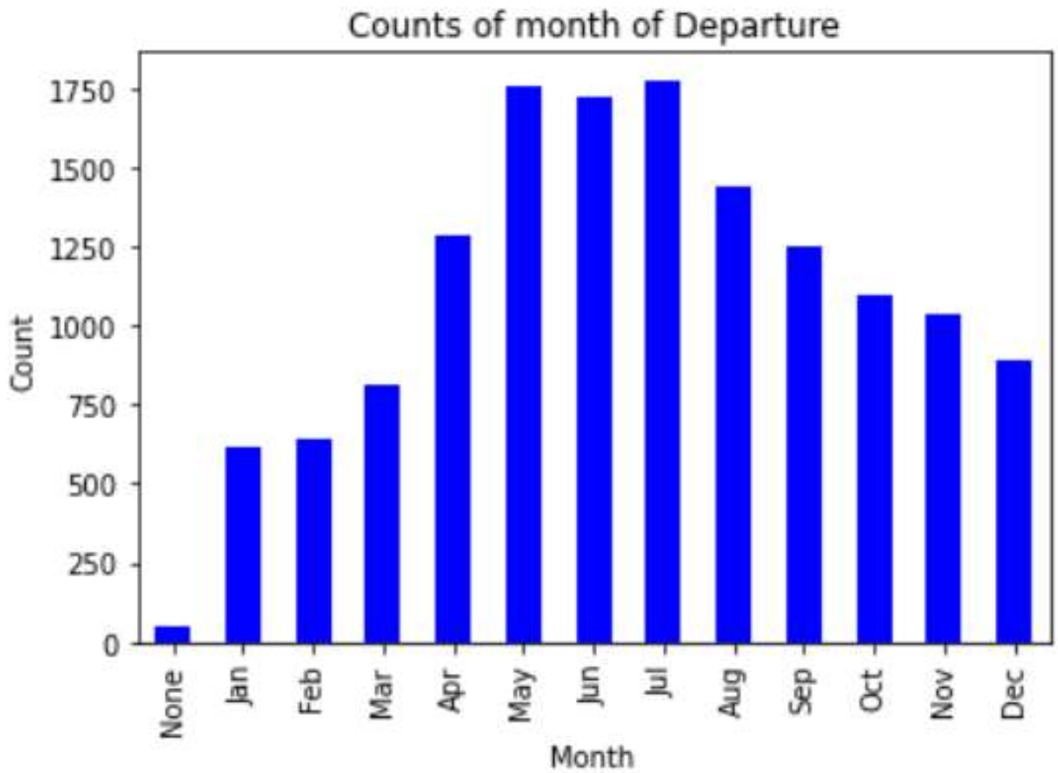

Figure 15: Figure of departure month bar chart. None is in every case that the month was missing.

It is evident from figure 15 that most of the shipments depart from the ports in late spring and early summer. This is an indication of how demand fluctuates during the year. The information gained by this graph could aid in better decision making. Sappi has a lot of bulk customers that have a specific reordering strategy. If those customers were eager to change their reordering strategy to cater to Sappi's transportation capacity, this might lessen the delayed shipments by making this distribution looking from normal to uniform.

5.2.4. Delayed shipments

In this subsection, the focus shifts towards illuminating the intricate dynamics surrounding delayed shipments, with a particular emphasis on lead time analysis. Delays in shipments not only impact logistical operations but also reverberate through supply chains, affecting inventory management, customer satisfaction, and overall business performance. Through a comprehensive examination of lead time, this section aims to unravel the underlying factors contributing to shipment delays, identify potential bottlenecks in the logistical process, and propose strategies for mitigation. By delving into the nuances of delayed shipments, valuable insights emerge, enabling stakeholders to proactively address challenges and enhance the efficiency and resilience of supply chain operations.

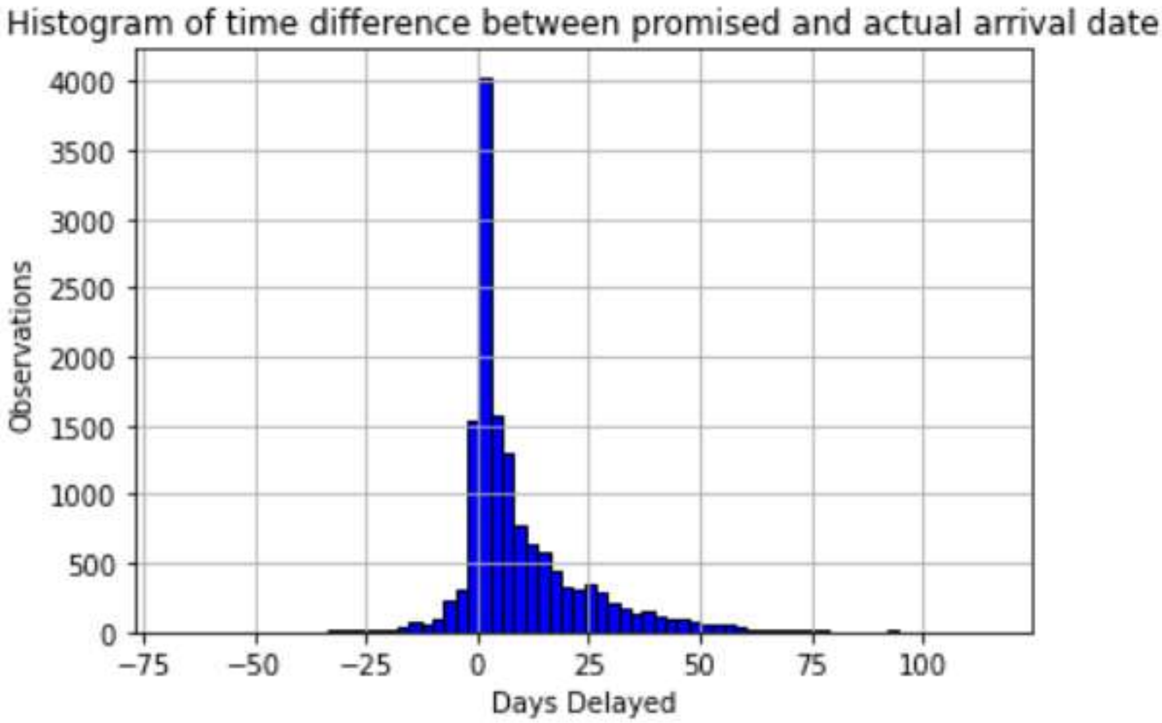

Figure 16: This figure showcases the observations of the total dataset. Essentially it contains information about every container that arrived before September 2023.

The shipments that were delayed more than 8 days accounted for 37% of the total dataset. From figure 16, it is shown that 75.7% of shipments are approximately between 10 days earlier and 15 days later. Those numbers are typical considering the nature of the container shipping industry. While the median value of the dataset falls into the non-delayed containers being 4.1 days, the mean value is outside of that range being 9.2 days. This indicates that there are a lot of observations with high values pulling the mean outside of the 8-day cutoff. The 25th and 75th percentile are 0.7 and 13.7 days respectively. Out of the 14381 observations only 82 observations are outliers from arriving too early, while 1045 are outliers for arriving a lot later than expected.

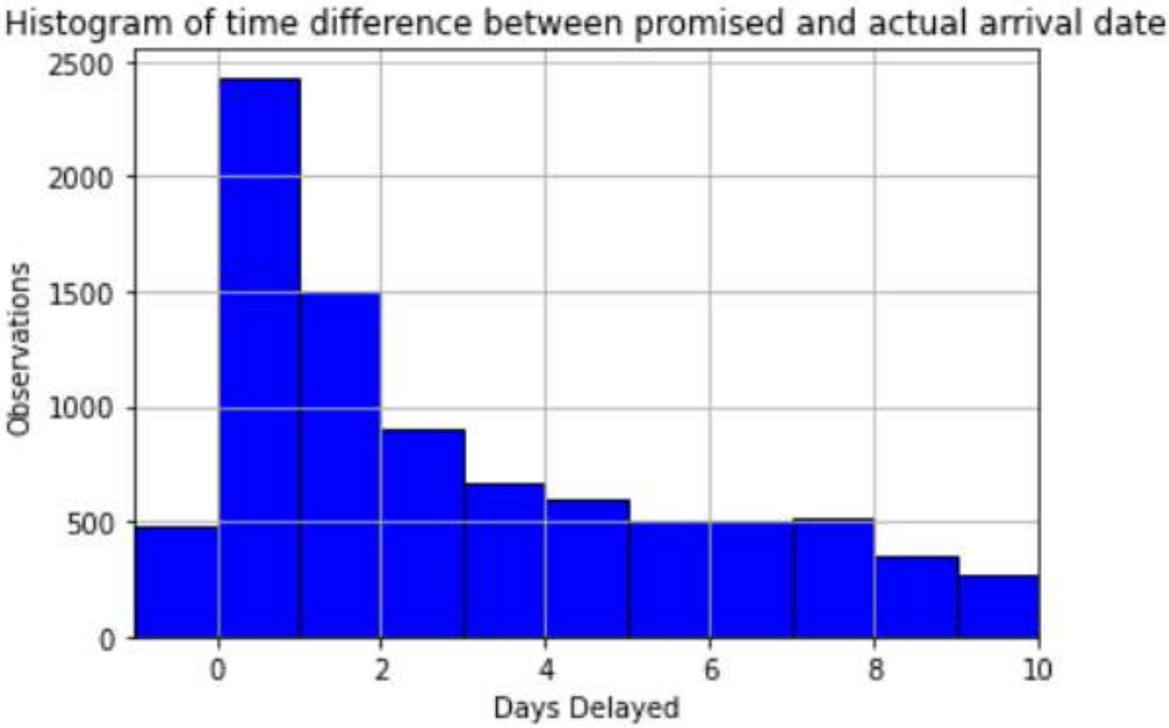

Figure 17: This figure showcases the observations out of the total dataset that arrived from 1 day earlier and 10 days later.

Figure 17 is focused on the shipments that arrived approximately on time. Each bar represents the timespan of a day, starting with 1 day earlier than expected on the left and 10 days delayed on the right. What is worth noticing is the drop in probability after 8 days. This indicates that binary models will perform better if they have an 8-day cutoff than a 9-day cutoff for example. Despite that this is coincidental to the fact that the best business decision for Sappi was 8 days, models are most likely to distinguish differences between observations that arrived a little earlier and a little later than 8 days, respectively, than the same scenario but with a 9-day or 7-day cut-off.

5.3. Target Encoding

In this section, readers will find a comprehensive summary of the outcomes stemming from the extensive model testing conducted. These findings represent the culmination of meticulous analysis and experimentation, offering insights into the effectiveness and performance metrics of the models examined. This section serves as a pivotal moment in the narrative, encapsulating the endpoint of the developmental process and emphasizing the critical role of comprehensive evaluation in predictive modeling.

In table 7 the results from the evaluation of the models are presented.

Table 7: Key performance indexes of the final approach for each model.

Model	F1-score	Accuracy	Precision
Random Forest	0.87	0.92	0.91
XGBoost	0.87	0.91	0.90
Logistic Regression	0.69	0.81	0.77
Neural Network	0.69	0.79	0.71
Transhipment model	0.62	0.71	0.56

5.3.1. Random forest

The main outcome of this project is the random forest model. Some results are shown in table 8 listed with their booking numbers. These predictions refer to some of the shipments that were booked to be shipped in September 2023. The prediction was based on the features from table 2. It seems that the results are between 52% and 74%. This is not the case for the total testing dataset as is shown in figure 18. Most of the predictions are skewed to the edges, either with a very low or a very high probability of being delayed more than 8 days. In the diagnostic section of the results there are reasons listed as to why the results from the shipments of September produce different probabilities for the shipments than the majority of the previous.

Table 8: Results from shipments of September by booking number and probability of being delayed more than 8 days.

Booking Number	Probability (>8days)	Booking Number	Probability (>8days)
230275481	64%	230860651	64%
230823689	64%	CPTD26743900	62%
29764047	55%	FRA0487571	52%
230196751	64%	HAMD66790700	55%
230228897	64%	459IHA0953424	74%
230812949	64%	HAMD73737600	52%
230838209	64%	231046489	63%

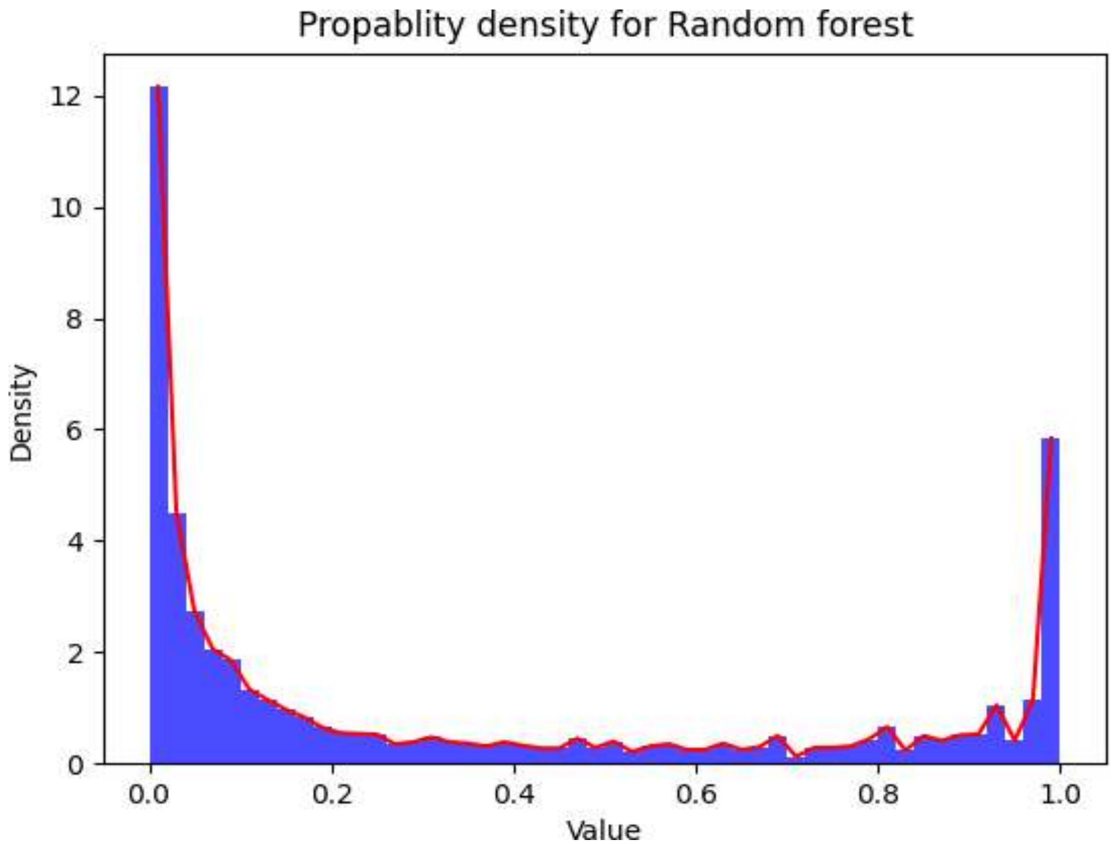

Figure 18: This figure shows the probability density of the predictions in the testing set. It is observed that most probabilities are either very high or very low.

The model could be used by assuming that shipments with probability lower than 30% are safe to arrive in time, shipments with probability higher than 70% are going to be delayed and anything between 30-70% needs close monitoring, however this is the minority of the shipments. How accurate are those probabilities determined by the calibration curve are shown in figure 19.

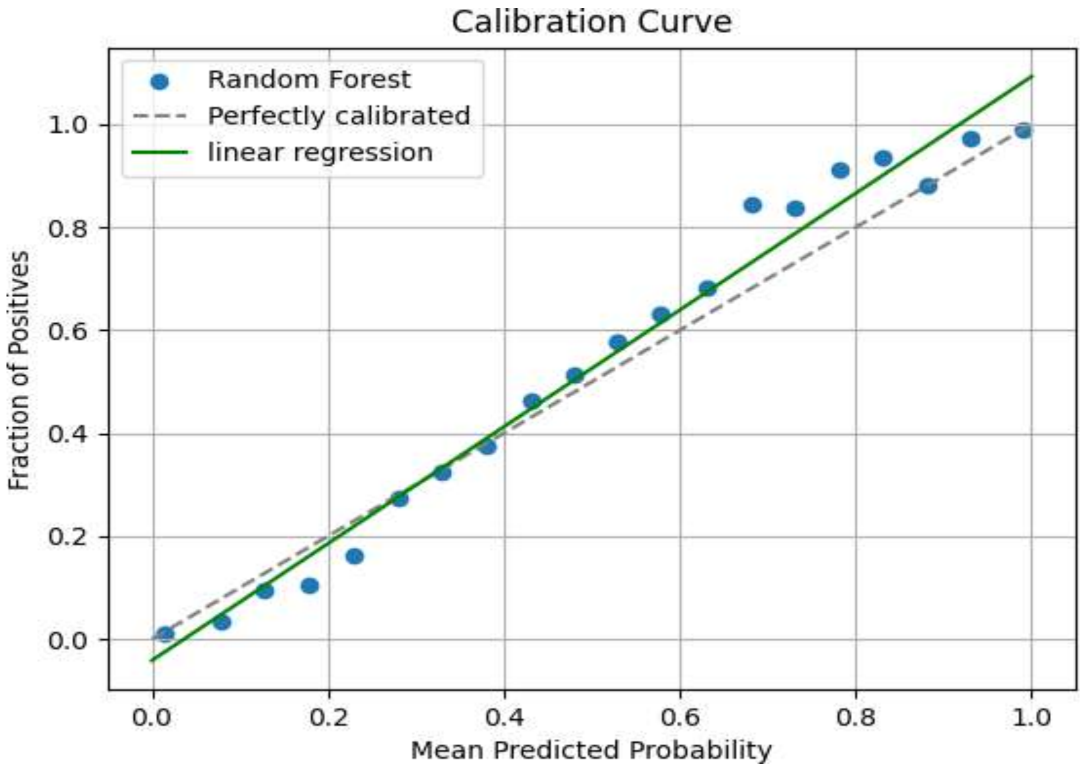

Figure 19: This calibration curve showcases that the model is well calibrated, therefore the probabilities produced by the model are right. On the y- axis is the percentage of delayed containers and on the x-axis the probability output of the model.

On the y-axis there is the fraction of positives which means the percentage of delayed containers. The green line is a linear regression of the blue dots, which is the actual data. The model is well calibrated, but it seems that when the probability is low (lower than 30%) it is even less probable than predicted, while if it's high (higher than 50%) the probability is even higher than predicted. This is not necessarily bad, in fact if the model is used as described above, it performs better this way. While it would be normal to wrongly predict 20% of the observations that have 20% probability of being late, the model actually wrongly predicts less than 20%. Similarly, while it is expected that the model would predict 20% of the observations with 80% probability of being delayed wrong, it exceeds the expectations, in this case as well.

Confusion Matrix

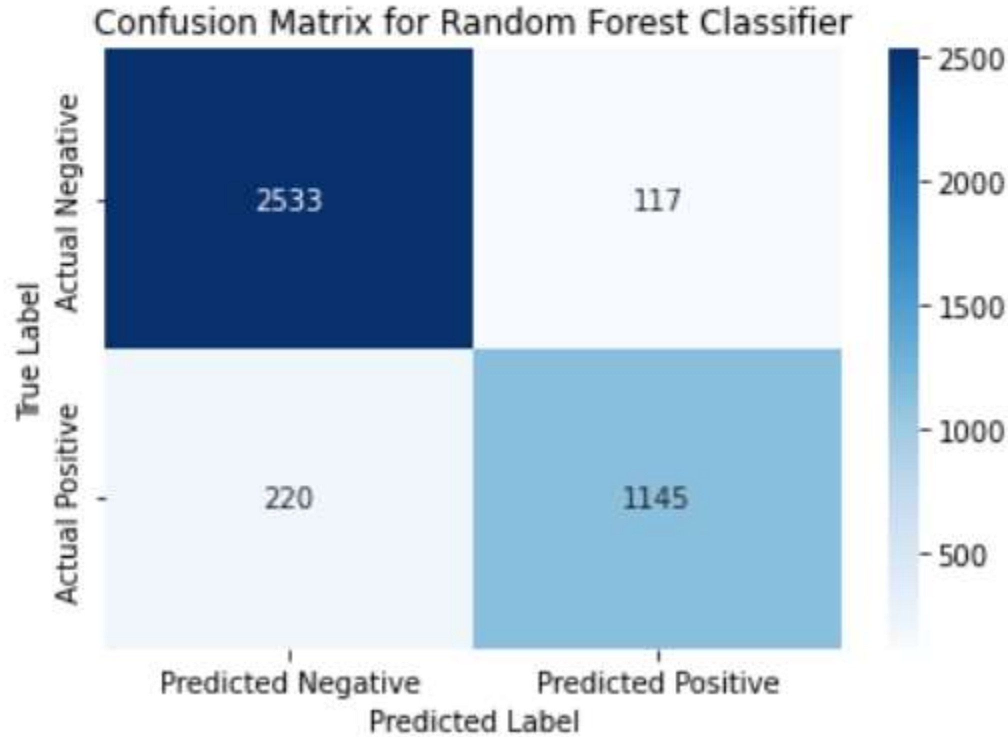

Figure 20: The confusion matrix for random forest classifier.

The confusion matrix provides a detailed breakdown of the model's performance on the 4,015 observations in the testing dataset. Ideally, the matrix would show a high concentration of observations along the main diagonal (from top left to bottom right), indicating correct predictions, and few observations on the off diagonal, indicating errors. Out of the 1,365 delayed observations, the model correctly predicted 84% as delayed, but incorrectly predicted 16% as on time. Conversely, of the 2,650 on-time deliveries, 96% were correctly predicted, while the remaining 4% were incorrectly predicted as delayed. This distribution highlights the model's overall accuracy and areas for improvement. There is a clear difference between predicting positive and predicting negative results. A predicted delay is less robust than a predicted on-time delivery.

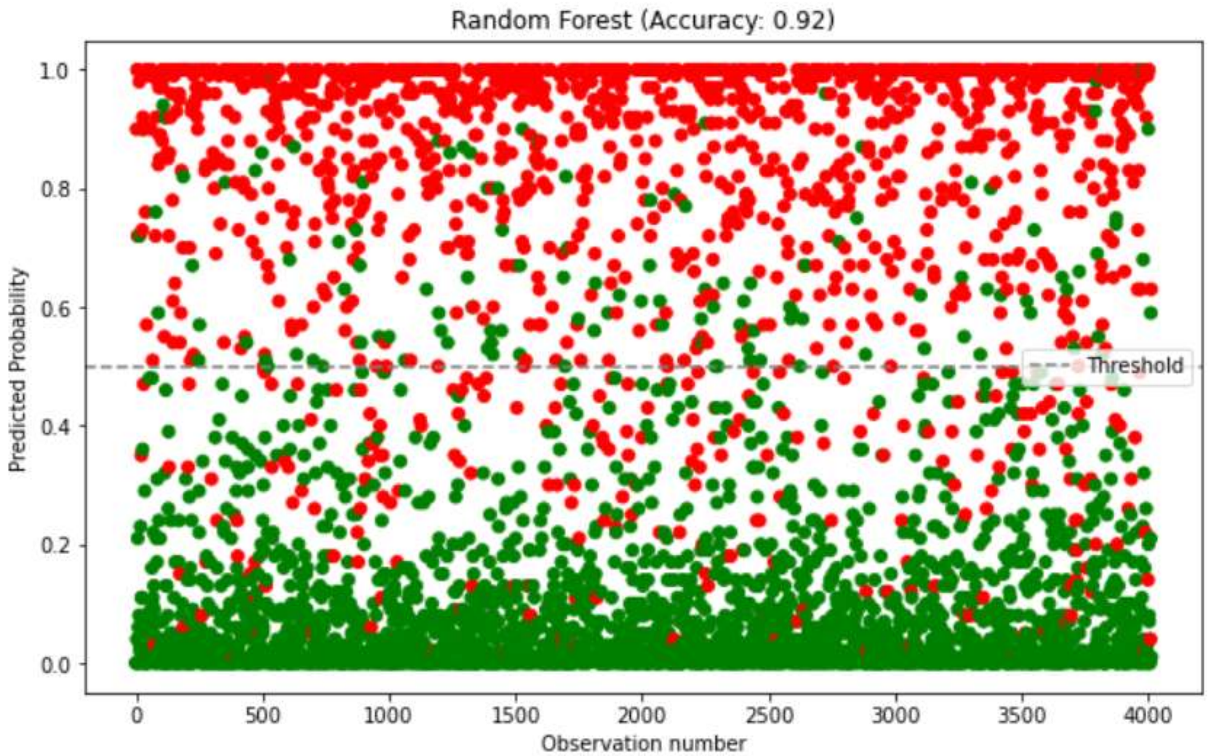

Figure 21: The predicted probability for each observation in the testing dataset. The red points represent actual delayed shipments while the green points represent on time deliveries. Their position in regard to the y axis is the predicted probability of the model. The dotted line is the proposed threshold (=0.5).

Figure 21 presents the same information as the confusion matrix, and possibly more, but in a more complex manner. The green dots represent observations that were delivered on time, while the red dots represent observations that were delivered with a delay. The dashed line indicates the prediction threshold: points above this line are predicted to be delayed, and points below it are predicted to be on time. This visualization provides a clear depiction of the model's classification boundary and the distribution of predictions relative to the actual outcomes.

Diagnostic Results

There is a graph shown below in figure with the importance of all the features. Using a computational method called mean decrease in impurity, explained in section 4, the dataset is broken down and reveals the features that have the highest mutual information paired with the target (the delays). The most important features are not something that Sappi has control over. However, it is still useful for customer advising and service. The most probable reason that the probabilities of the predicted shipments in table 4 are all indecisive is that the month where the containers are discharged plays a crucial role in creating the prediction. The most important features are shown below mentioned by importance.

Table 9: The features ranked by importance on the outcome of the model.

Features
Discharge Month
Discharge Port
Usual Duration
Scheduled Duration
Carrier

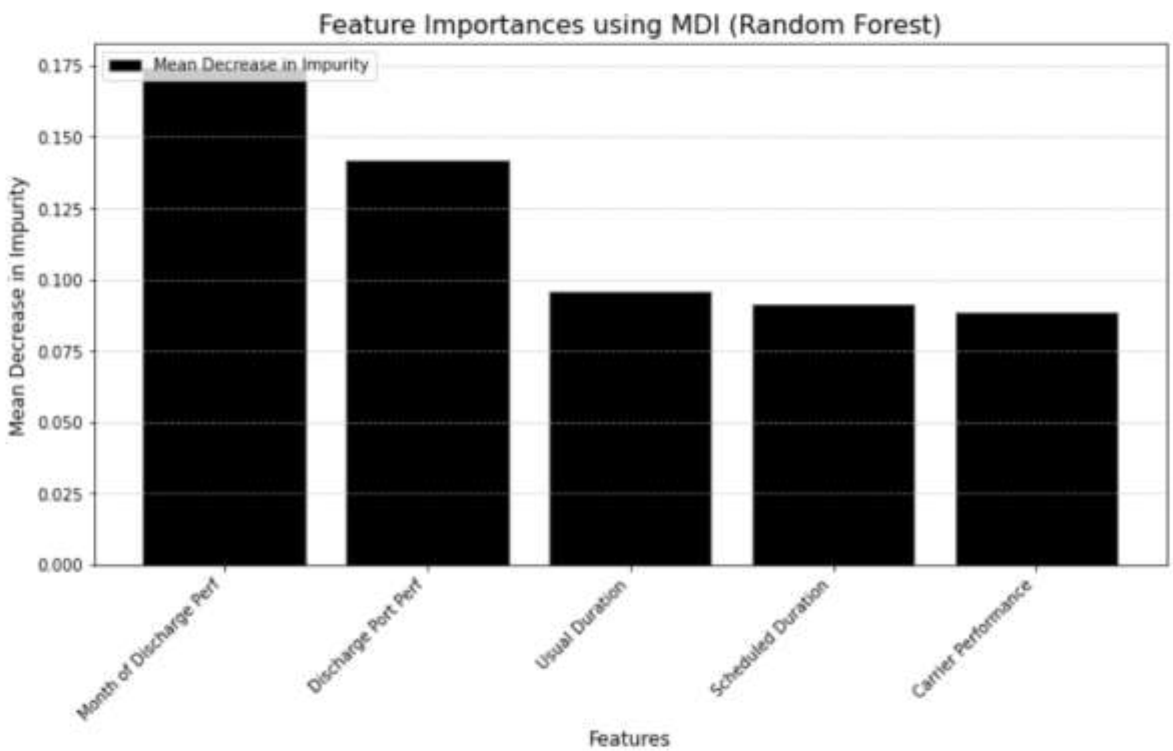

Figure 22: a bar chart of mean decrease in impurity for the top 5 features.

It is shown, from figure 22, that the month of discharge is the most influential feature on the decision of the model. This explains why all of the predictions shown in the table 9 are all around the same area of probability, while the rest of the dataset has either really high or really low probability.

5.3.2. XGBoost

This section presents the results from the XGBoost model. From Table 10, it appears that the XGBoost model is more certain about its predictions compared to the random forest model, despite having slightly lower overall performance. This certainty is reflected in the tighter confidence intervals and more decisive classification outcomes, indicating a higher confidence level in the predictions made by the XGBoost model.

Table 10: Results from shipments of September by booking number and probability of being delayed more than 8 days.

Booking Number	Probability (>8days)	Booking Number	Probability (>8days)
230275481	95%	230860651	95%
230823689	95%	CPTD26743900	91.2%
29764047	35.6%	FRA0487571	56.9%
230196751	95%	HAMD66790700	83.4%
230228897	95%	459IHA0953424	93.6%
230812949	95%	HAMD73737600	88.5%
230838209	95%	231046489	94.7%

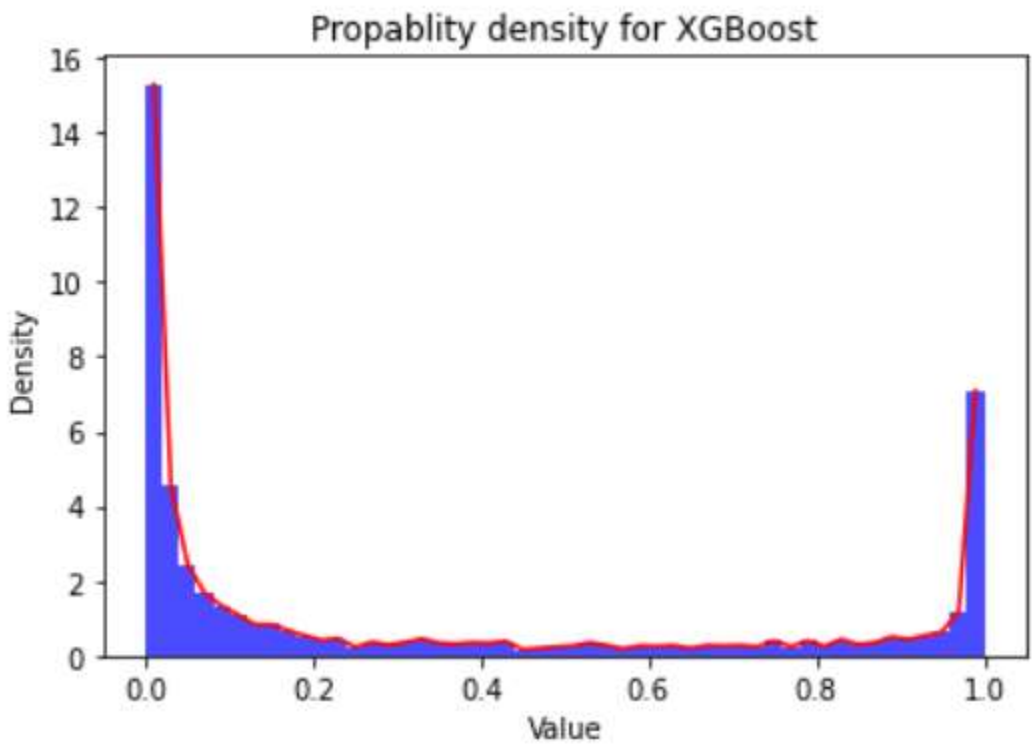

Figure 23: This figure shows the probability density of the predictions in the testing set. It is observed that most probabilities are either very high or very low.

This graph confirms what is indicated in the table above: the predicted probabilities of the XGBoost model are more skewed towards the edges compared to the random forest model, even though the random forest also has a strong majority of its observations at the edges. This skewness suggests that the XGBoost model assigns higher confidence to its predictions, resulting in more decisive classification outcomes. This is also shown by the axes as well, the bars on the edges reach higher than the bars on the edges in figure 18.

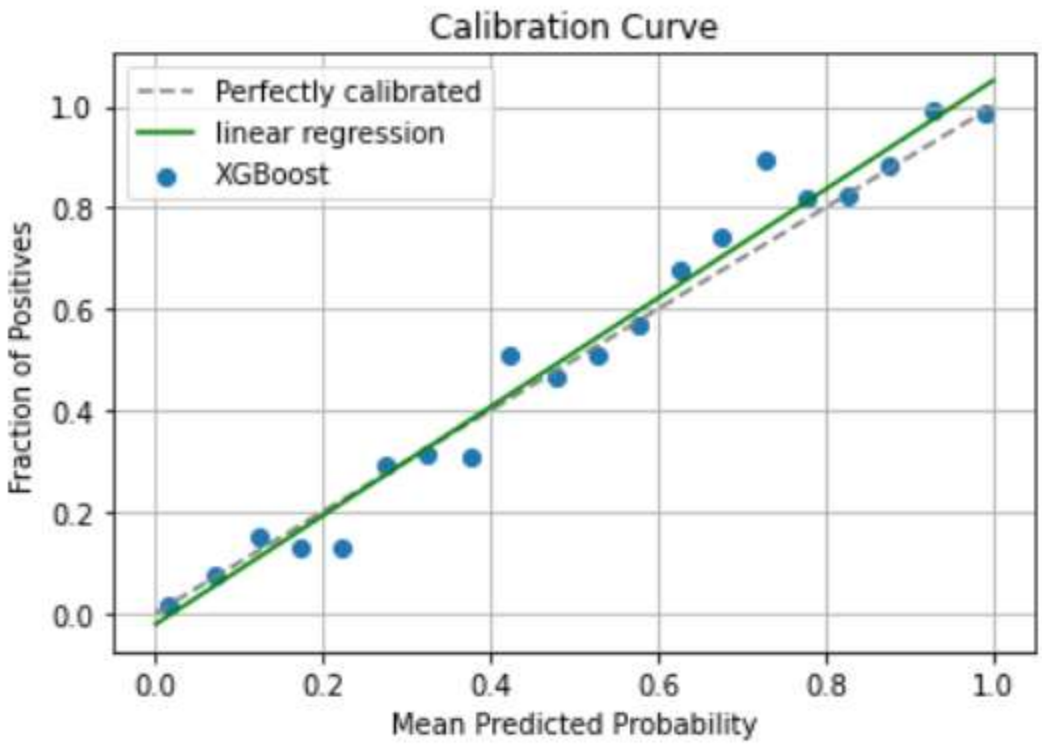

Figure 24: This calibration curve showcases that the model is well calibrated, therefore the probabilities produced by the model are right. On the y- axis is the percentage of delayed containers and on the x-axis the probability output of the model.

It appears that this model is better calibrated than the random forest model, as evidenced by the linear regression analysis. However, as discussed above, this improved calibration does not translate into better overall KPIs, such as accuracy and precision. This discrepancy arises because the high predicted probabilities from this model are well-calibrated, whereas the random forest model tends to underestimate these probabilities. Conversely, the lower probabilities in this model are more accurate, whereas the random forest model tends to overestimate them.

Confusion Matrix

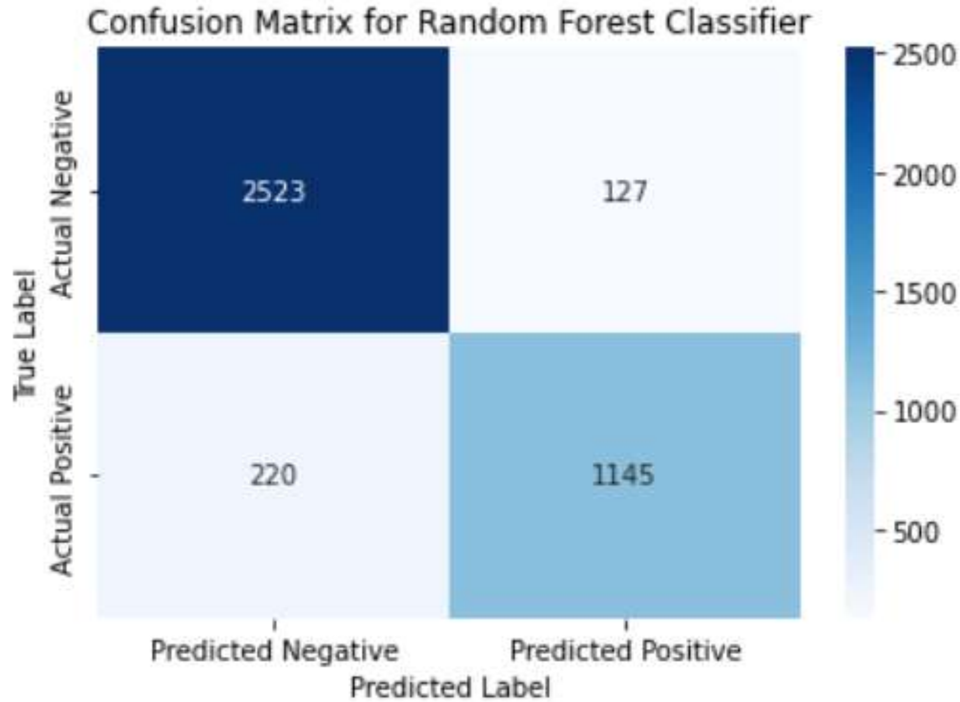

Figure 25: The confusion matrix for the XGBoost classifier.

The confusion matrix, presented in figure 25, provides a detailed breakdown of the model's performance on the 4,015 observations in the testing dataset. Ideally, the matrix would show a high concentration of observations along the main diagonal, indicating correct predictions, and few observations on the off diagonal, indicating errors. Out of the 1,365 delayed observations, the model correctly predicted 84% as delayed, but incorrectly predicted 16% as on time. Conversely, of the 2,650 on-time deliveries, 95% were correctly predicted, while the remaining 5% were incorrectly predicted as delayed. This distribution highlights the model's overall accuracy and areas for improvement. While the percentages of the delayed observations stayed the same as with the random forest model, the observations that were predicted as on time belie the lesser performance of this model.

Compared to the random forest model, it is evident that the accuracy is lower. However, this decrease in accuracy primarily affects the precision, while the recall remains unaffected. This can be observed in the confusion matrix where only the top half, representing the negatives, has changed, while the bottom half, representing the positives, remains unchanged. This indicates that the model's ability to identify positives (recall) is consistent, but its precision in distinguishing negatives has decreased.

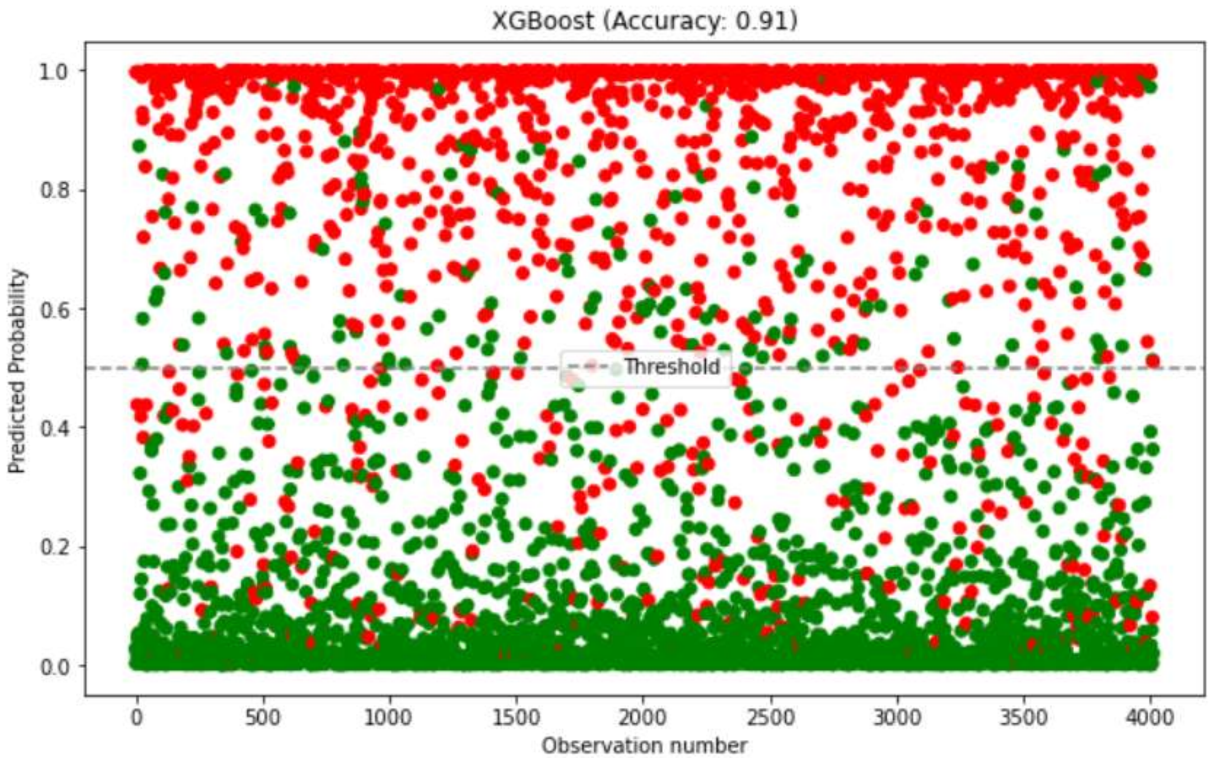

Figure 26: The predicted probability for each observation in the testing dataset. The red points represent actual delayed shipments while the green points represent on time deliveries. Their position in regard to the y axis is the predicted probability of the model.

Comparing this diagram with the one generated by the random forest model, it appears that despite having lower accuracy, the area near the threshold is less densely populated. This indicates that the XGBoost model is more confident about its predictions, even though it makes slightly more mistakes overall. The reduced density near the threshold suggests that the model's predictions are more decisive, with fewer observations falling into the uncertain, borderline range.

Diagnostic Results

There is a graph shown below in figure 27 with the importance of all the features. Using the same computational method called mean decrease in impurity the dataset is broken down and reveals the features that have the highest mutual information paired with the target (the delays). The most important features are not something that Sappi has control over. However, it is still useful for customer advising and service.

Table 11: The features ranked by importance on the outcome of the XGBoost model.

Features
Transshipment
Discharge Port
Discharge Month
Departure month
Origin Port

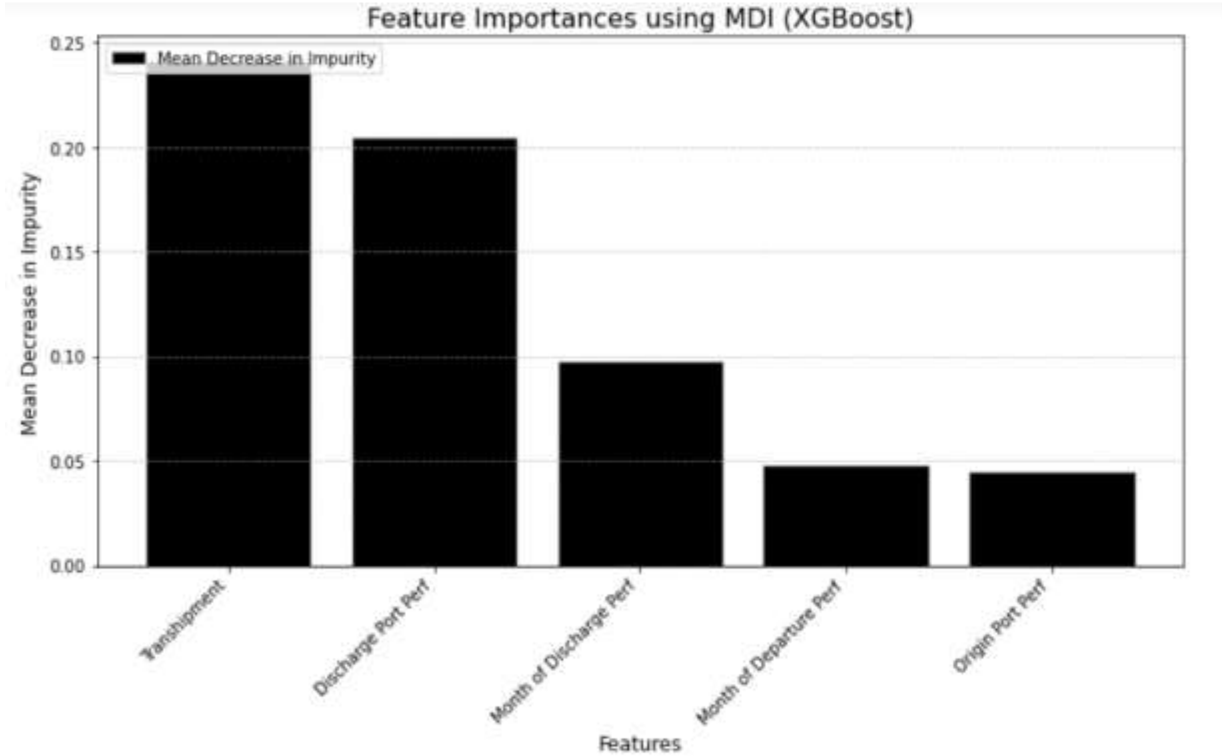

Figure 27: a bar chart of mean decrease in impurity for the top 5 features.

It is worth mentioning that XGBoost has chosen different features as the most useful. It is closer to the guesses of the interviewees of the UX project, having transshipment as the most influential feature for the prediction of a delay. After that is the target encoded feature of the discharge port followed by the discharge month, which was first in the random's forest tree explanation. Finally, there has been given similar importance to the month of departure and the origin port which ranked 4th and 5th respectively. It is worth noting that most of the features mentioned as important from both the random forest and the XGBoost model were target encoded.

5.3.3. Logistic Regression

This section presents the results from the Logistic Regression model. From Table 12, it appears that this model is more optimistic about its predictions compared to the last two models, despite having slightly lower overall performance. Contrary to what has been shown in this thesis so far, this model predicts that most of the shipments of September are going to be delivered on time.

Table 12: Results, produced by the multiple regression model, from shipments of September by booking number and probability of being delayed more than 8 days.

Booking Number	Probability (>8days)	Booking Number	Probability (>8days)
230275481	19.9%	230860651	19.9%
230823689	19.9%	CPTD26743900	18.2%
29764047	35.5%	FRA0487571	33.6%
230196751	19.9%	HAMD66790700	24.8%
230228897	19.9%	459IHA0953424	71.4%
230812949	19.9%	HAMD73737600	25.1%
230838209	19.9%	231046489	19.0%

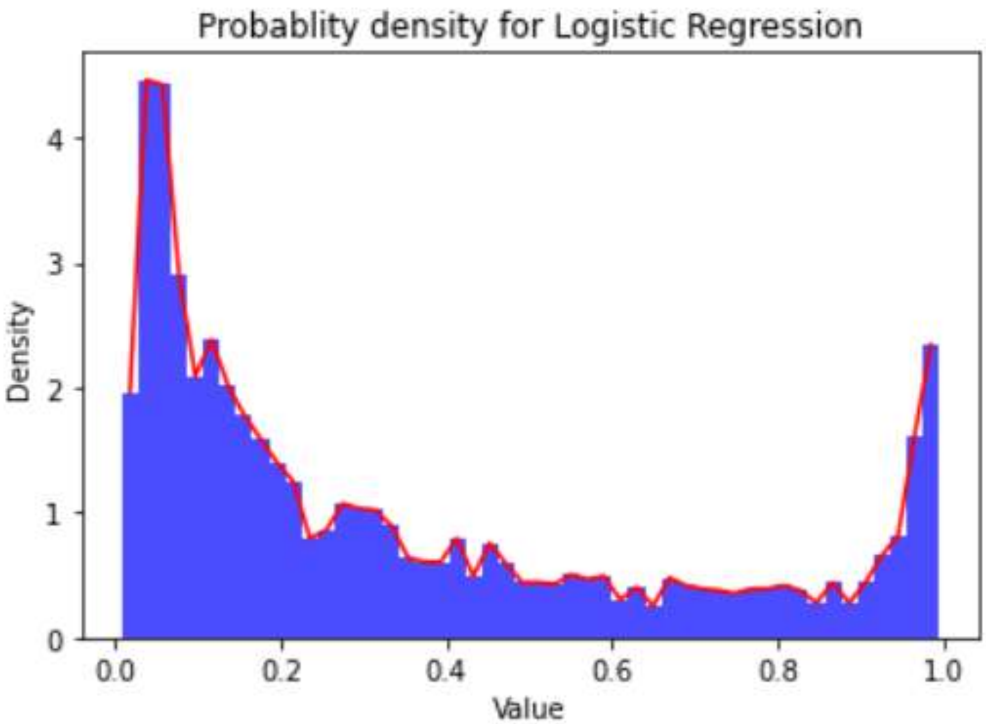

Figure 28: This graph showcases the Probability density for logistic regression.

This model, as shown in figure 28, demonstrates a broader prediction range compared to the previously shown models. Additionally, a qualitative examination of the curve's shape reveals a greater concentration of observations within the middle probability range. This indicates a more dispersed distribution of predicted probabilities, suggesting a wider range of potential outcomes.

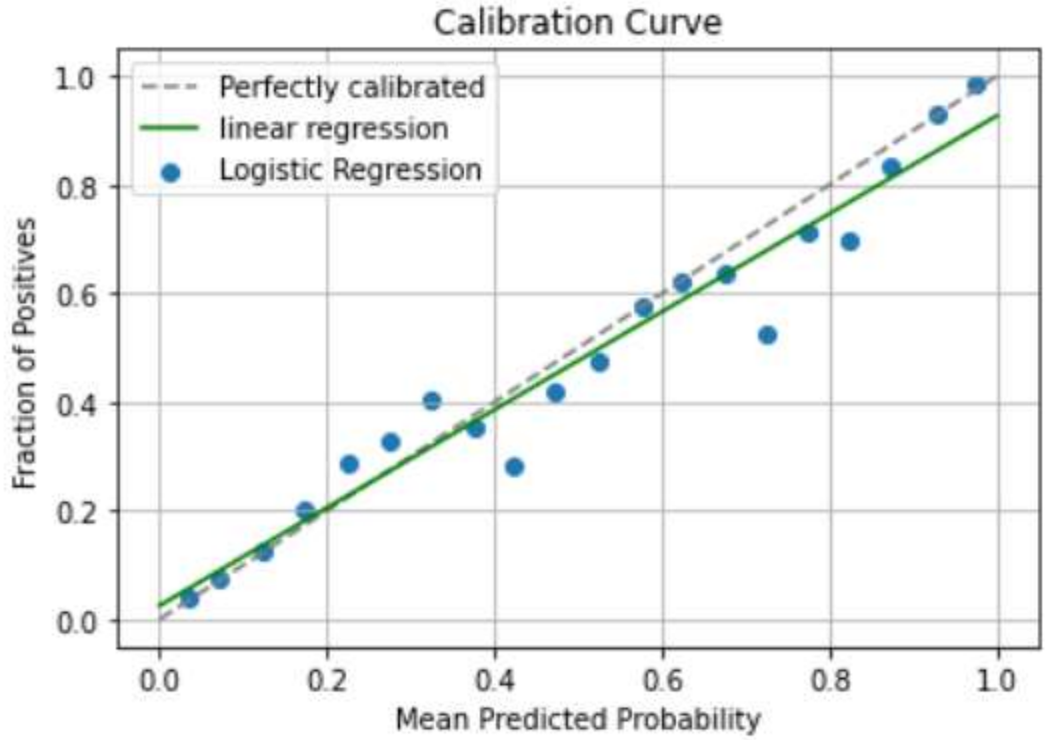

Figure 29: This calibration curve showcases that the model is well calibrated, therefore the probabilities produced by the model are right. On the y- axis is the percentage of delayed containers and on the x-axis the probability output of the model.

It appears that this model is also well calibrated. However, as discussed above, this improved calibration does not translate into better overall KPIs, such as accuracy and precision. In fact, combining the information from figure 25 and figure 24, it is eminent that the KPI's are lower than the rest of the two models. Having a larger number of observations in the mid ranges of probability, as shown in figure 24 and having a well calibrated model converses into having more wrongly predicted observations. Contrary to the previous models, it appears to be that when the predicted probability is low, the actual probability is higher, and vice versa.

Confusion Matrix

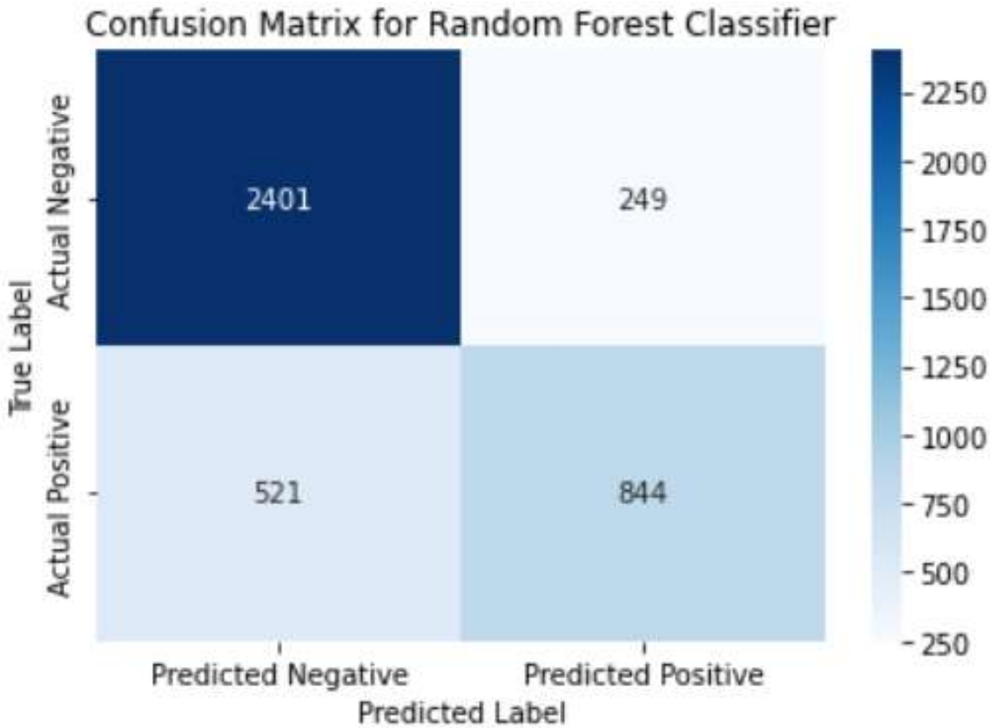

Figure 30: The confusion matrix for the Logistic Regressor classifier.

The confusion matrix, presented in figure 30, provides a detailed breakdown of the model's performance on the 4,015 observations in the testing dataset. Ideally, the matrix would show a high concentration of observations along the main diagonal, indicating correct predictions, and few observations on the off diagonal, indicating errors. Out of the 1,365 delayed observations, the model correctly predicted 62% as delayed, but incorrectly predicted 38% as on time. Conversely, of the 2,650 on-time deliveries, 91% were correctly predicted, while the remaining 9% were incorrectly predicted as delayed. This distribution highlights the model's overall accuracy and areas for improvement.

It is evident that this model is lacking in accuracy compared to the other two models. Both precision and recall have decreased. This can be shown by the last row and the last column.

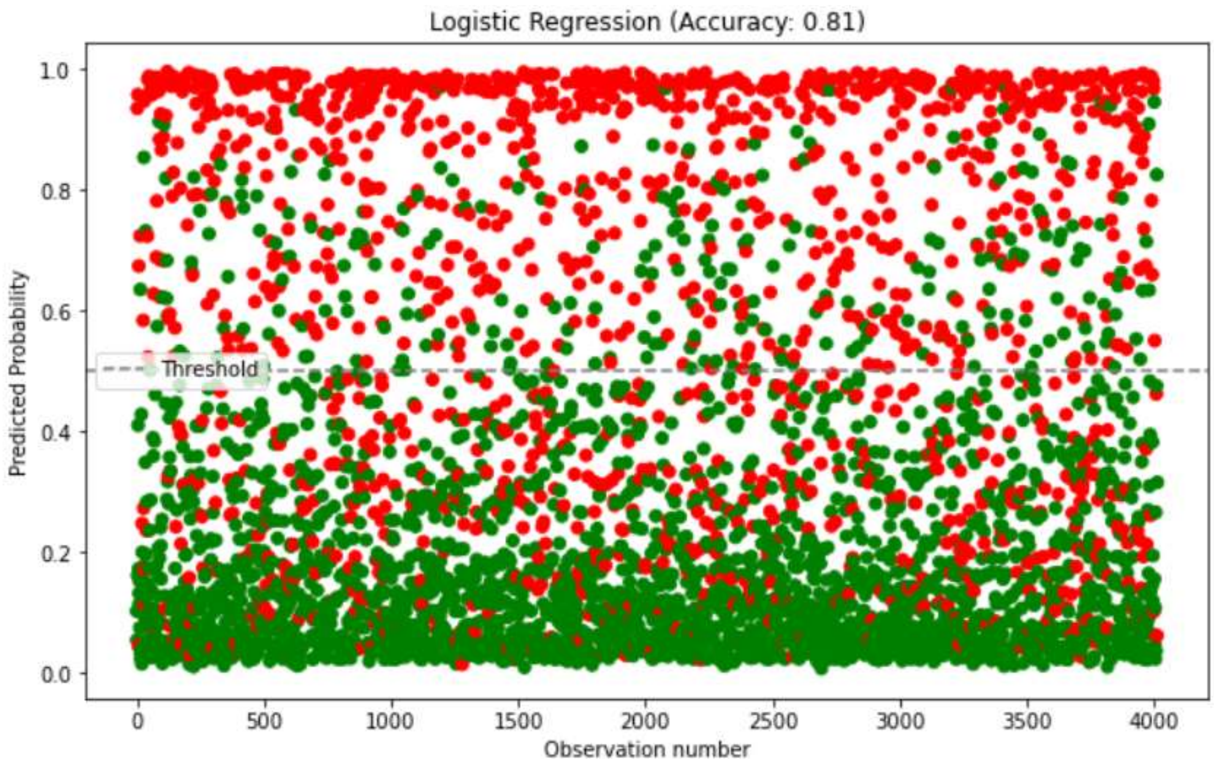

Figure 31: The predicted probability for each observation in the testing dataset. The red points represent actual delayed shipments while the green points represent on time deliveries. Their position in regard to the y axis is the predicted probability of the model.

Figure 31 confirms the information gained by the other plots so far. Comparing this diagram with the two previous that were generated by the forest models, it appears that this model has lower accuracy and naturally the area near the threshold is far more densely populated. This indicates that the Logistic regression model has more uncertainty around its predictions. The increased density near the threshold suggests that the model's predictions are more indecisive, with fewer observations falling into the uncertain, borderline range.

Diagnostic Results

The logistic regression model, presented through the summary provided in table 13, offers a distinct approach to understanding the relationships between predictor variables and the binary outcome of interest. Unlike the mean decrease in impurity method utilized in the previous models, which are typically associated with decision tree-based algorithms such as Random Forests, logistic regression operates within a probabilistic framework to estimate the probability of occurrence for the outcome variable.

While the mean decrease in impurity method assesses the importance of features based on their contribution to the reduction of impurity within decision trees, logistic regression estimates the impact of each predictor variable on the log-odds of the outcome. Consequently, the significance of variables in logistic regression is determined by their coefficients, representing the change in log-odds associated with a one-unit change in the predictor, as demonstrated in the summary below.

This alternative approach allows for a more direct interpretation of the relationship between predictors and the outcome, facilitating the identification of key variables influencing the probability of occurrence. Through the following summary, we delve into the specifics of the logistic regression model, highlighting its unique characteristics and providing insights into its performance and predictive capability.

Table 13: The summary table for logistic regression.

Independent variable	coef	std err	z	P>|z|	[0.025	0.975]
Const	-4.72	0.186	-25.393	0	-5.084	-4.356
Scheduled_duration	-0.0297	0.002	-12.788	0	-0.034	-0.025
Transhipment_binary	1.5812	0.06	26.258	0	1.463	1.699
Carrier_Performance	1.7452	0.208	8.38	0	1.337	2.153
Origin_Performance	0.6241	0.233	2.681	0.007	0.168	1.08
Discharge_Performance	3.3094	0.138	23.963	0	3.039	3.58
LastUpdate_SCAC	-0.0403	0.011	-3.611	0	-0.062	-0.018
BC_MainCarriage_PortOfDischarge_LocationCode	0.042	0.006	6.776	0	0.03	0.054
Countries_Of_Discharge	-0.0742	0.012	-5.949	0	-0.099	-0.05
Usual_duration	0.005	0.003	1.511	0.131	-0.001	0.012
BC_MainCarriage_PortOfLoad_LocationCode	0.0582	0.005	12.361	0	0.049	0.067
Order_size	-0.0234	0.003	-6.88	0	-0.03	-0.017
Departure_month	0.0209	0.012	1.779	0.075	-0.002	0.044
Discharge_month	-0.0106	0.011	-0.959	0.338	-0.032	0.011
Departure_month_Performance	2.4583	0.313	7.853	0	1.845	3.072
Discharge_month_Performance	4.1924	0.293	14.311	0	3.618	4.767

Table 13 summarizes the results of a logistic regression analysis. For better understanding, there follows a short breakdown of the columns:

- **coef (coefficient)**: This column shows the estimated coefficient for each predictor variable in the model. However, unlike linear regression, these coefficients don't directly represent the change in the probability of the event. Instead, they represent the log-odds change in the probability of the event occurring for a one-unit increase in the predictor variable, holding all other predictor variables constant.
    - Positive coefficients indicate that a higher value of the predictor variable is associated with an increased log-odds of the event happening. In other words, for every one-unit increase in the variable, the log of the odds of the event goes up.
    - Negative coefficients suggest a higher value of the predictor variable is associated with a decreased log-odds of the event. The log of the odds of the event goes down for every one-unit increase in the variable.

- **std error (standard error)**: This reflects the precision of the coefficient estimate. A smaller standard error suggests a more precise estimate, meaning the coefficient is less likely to vary if you were to fit the model to a different sample of data.
- **z (z-statistic)**: This column shows the z-statistic, which is calculated by dividing the coefficient estimate by its standard error. It's used to assess whether the coefficient is statistically significant, meaning it's different from zero at a certain level of confidence.
- **P>|z| (p-value):** The p-value associated with the z-statistic represents the probability of observing a z-statistic as extreme as the one calculated, assuming the null hypothesis is true (i.e., the coefficient is truly zero). A low p-value (typically less than 0.05) indicates that the coefficient is statistically significant.
- **[0.025, 0.975] (confidence interval)**: These columns represent the confidence interval for the coefficient estimate. It's a range of values that likely holds the true population coefficient with a certain level of confidence (in this case 95%).

Sorting the top 5 features by the absolute value of their coefficient, table 14 is created.

Table 14: The features ranked by importance on the outcome of the Logistic Regression model.

Features
Discharge Month
Discharge Port
Departure Month
Carrier
Transshipment

This table provides insights as to what are the most important features of the model. It is not surprising that all these features have been mentioned at least once in the similar tables for the tree models. This means that more or less the models are in agreement as to the most important feature when it comes to identifying the most prominent causes for a shipment to be delayed more than 8 days.

5.3.4. Deep Neural networks

The performance of Deep Neural Networks (DNN) is analyzed in this section. Table 15 (or relevant table number) details its accuracy and offers insights into the model's behavior. While potentially more accurate due to its ability to capture complex relationships, DNNs can be less interpretable compared to simpler models. This chapter explores this accuracy-interpretability trade-off.

Table 15: Results, produced by the DNN model, from shipments of September by booking number and probability of being delayed more than 8 days.

Booking Number	Probability (>8days)	Booking Number	Probability (>8days)
230275481	98.1%	230860651	98,1%
230823689	98.1%	CPTD26743900	98.1%
29764047	96.0%	FRA0487571	95.8%
230196751	98.1%	HAMD66790700	96.9%
230228897	98.1%	459IHA0953424	99.4%
230812949	98.1%	HAMD73737600	98.4%
230838209	98,1%	231046489	98.0%

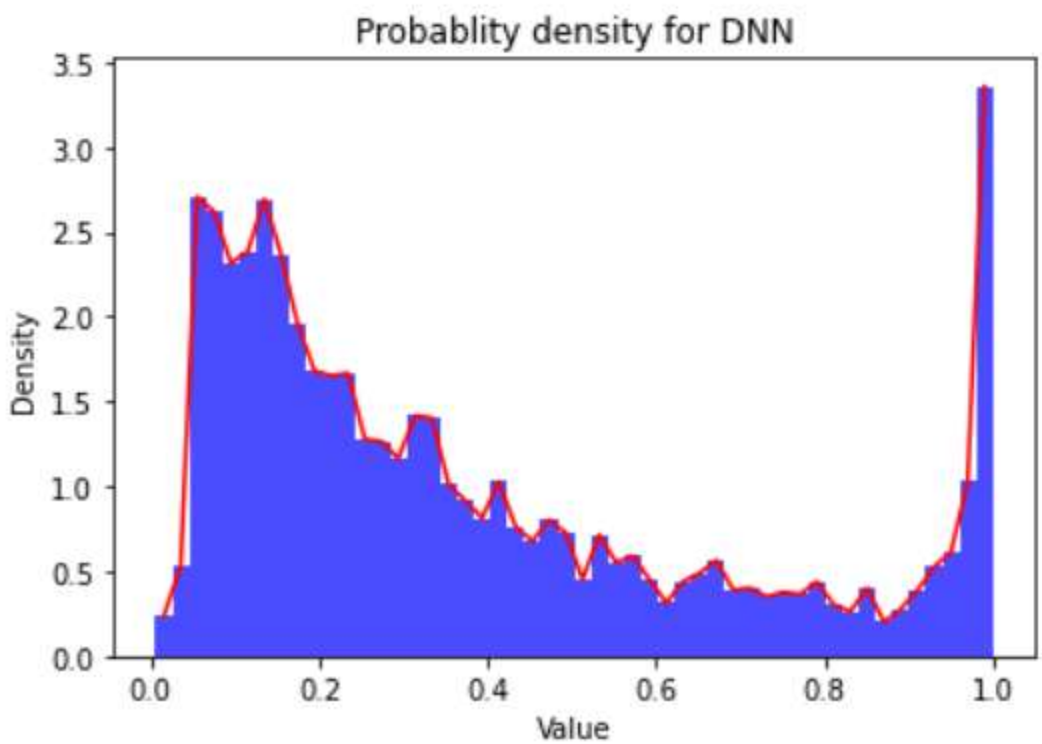

Figure 32: This graph showcases the Probability density for DNN.

This model demonstrates broader predictions compared to the previously shown models, even to the logistic regression, as shown in figure 32. Additionally, a qualitative examination of the curve's shape reveals a greater concentration of observations within the lower middle probability range. This indicates a more dispersed distribution of predicted probabilities, suggesting a wider range of potential outcomes.

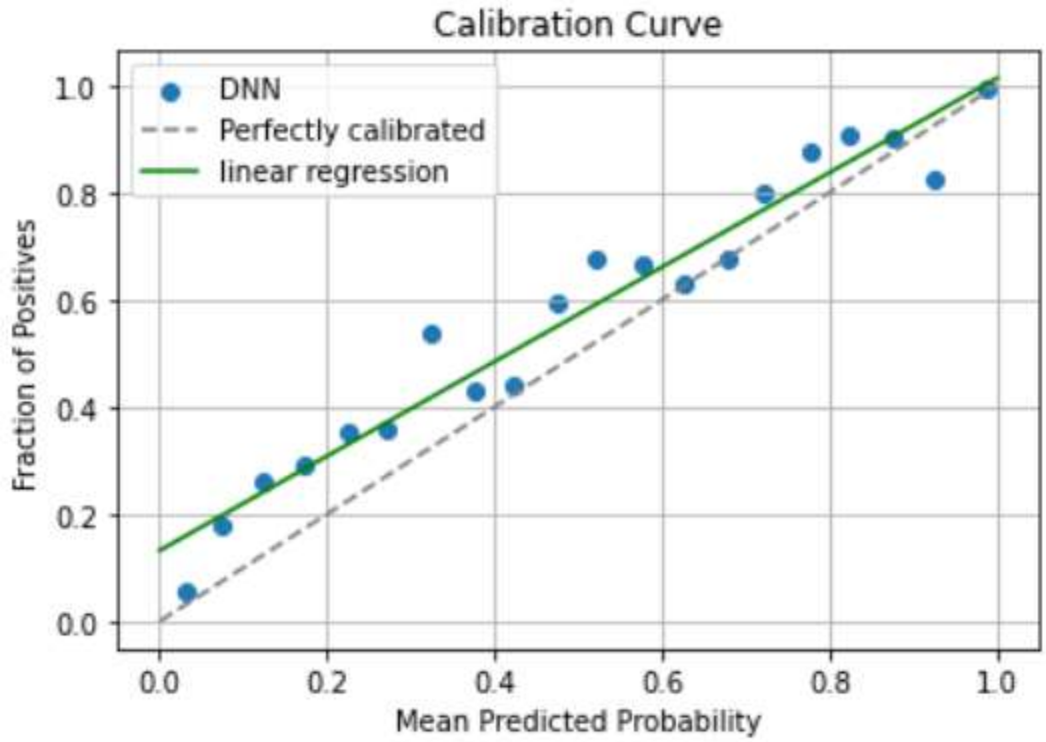

Figure 33: This calibration curve showcases that the model is well calibrated, therefore the probabilities produced by the model are right. On the y- axis is the percentage of delayed containers and on the x-axis the probability output of the model.

While other models show a better alignment between their predicted probabilities and the actual outcomes (represented by the diagonal y=x line), the DNN struggles in this regard, as shown in figure 33. This is especially true for predictions with low probabilities. In these cases, the DNN tends to be overly confident, meaning its assigned probabilities are often higher than the actual likelihood of the event occurring. However, for predictions with high probabilities, the DNN performs better, and its assigned probabilities become more accurate in reflecting the true outcome.

Confusion matrix

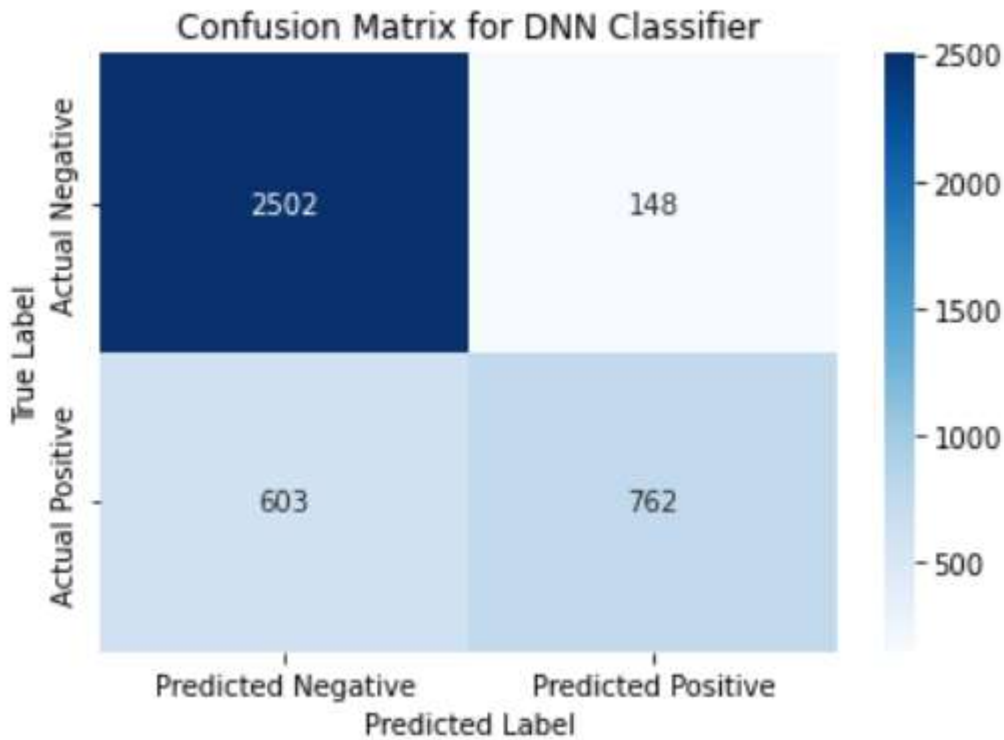

Figure 34: The confusion matrix for the DNN classifier.

The confusion matrix provides a detailed breakdown of the model's performance on the 4,015 observations in the testing dataset. Ideally, the matrix would show a high concentration of observations along the main diagonal, indicating correct predictions, and few observations on the off diagonal, indicating errors. Out of the 1,365 delayed observations, the model correctly predicted 56% as delayed, but incorrectly predicted 44% as on time. Conversely, of the 2,650 on-time deliveries, 94% were correctly predicted, while the remaining 6% were incorrectly predicted as delayed. This distribution highlights the model's overall accuracy and areas for improvement.

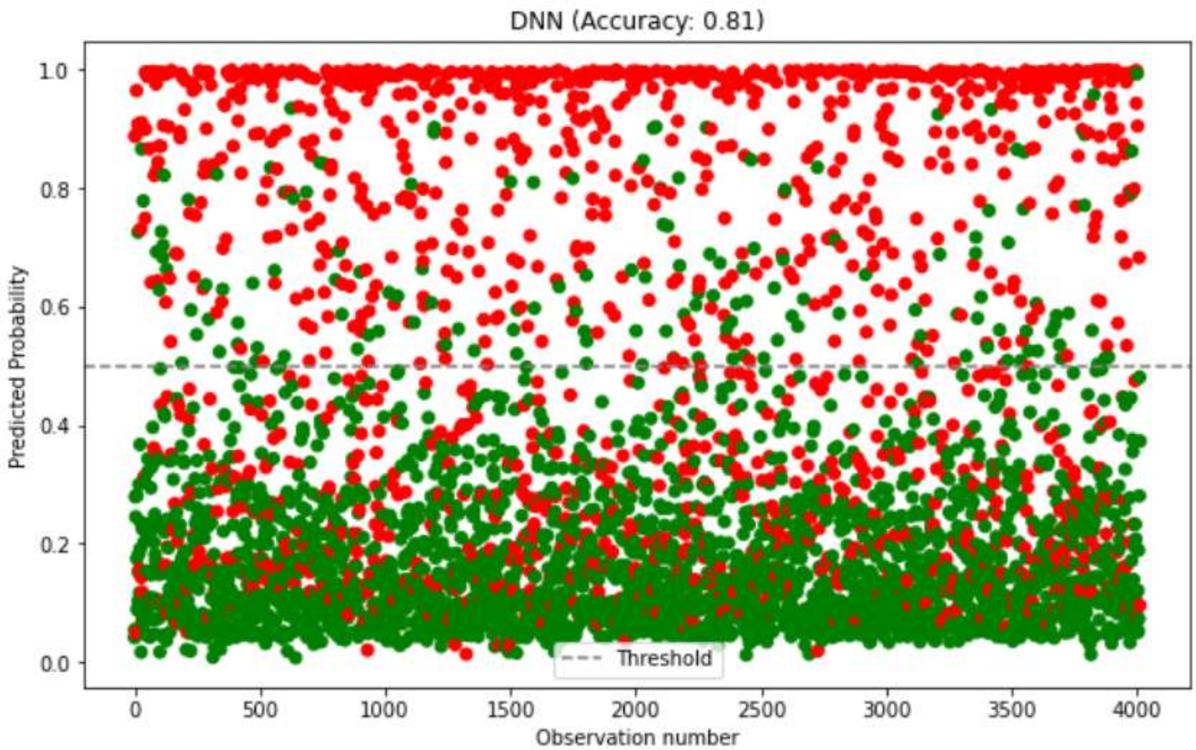

Figure 35: The predicted probability for each observation in the testing dataset. The red points represent actual delayed shipments while the green points represent on time deliveries. Their position in regard to the y axis is the predicted probability of the model.

Figure 35 confirms the information gained by the other plots so far. Comparing this diagram with the two previous that were generated by the forest models, it appears that this model has lower accuracy and there are a lot of observations that were delayed predicted on time, as showcased in figure 34. This is also indicated in figure 32 that the DNN model predicts more often that the shipment will be on time.

Diagnostic results

Unlike simpler models that reveal their decision-making process through coefficients or rules, DNNs are opaque. Their power lies in their complex web of interconnected neurons, making it challenging to pinpoint why they arrive at a specific prediction. This lack of transparency can be frustrating for users who want to understand the rationale behind the model's decisions, especially in critical applications.

5.3.5. Transshipment model

This model is the simplest of the ones presented here, and it falls into the category of knowledge-based models (or expert-based models). During the user experience project, a strong belief was discovered around the transshipment and its relationship with the delays. Most participants in the interviews and experts in meetings had mentioned that transshipment would be the most dominant feature for predicting delays. Because of that there was a simple model created.

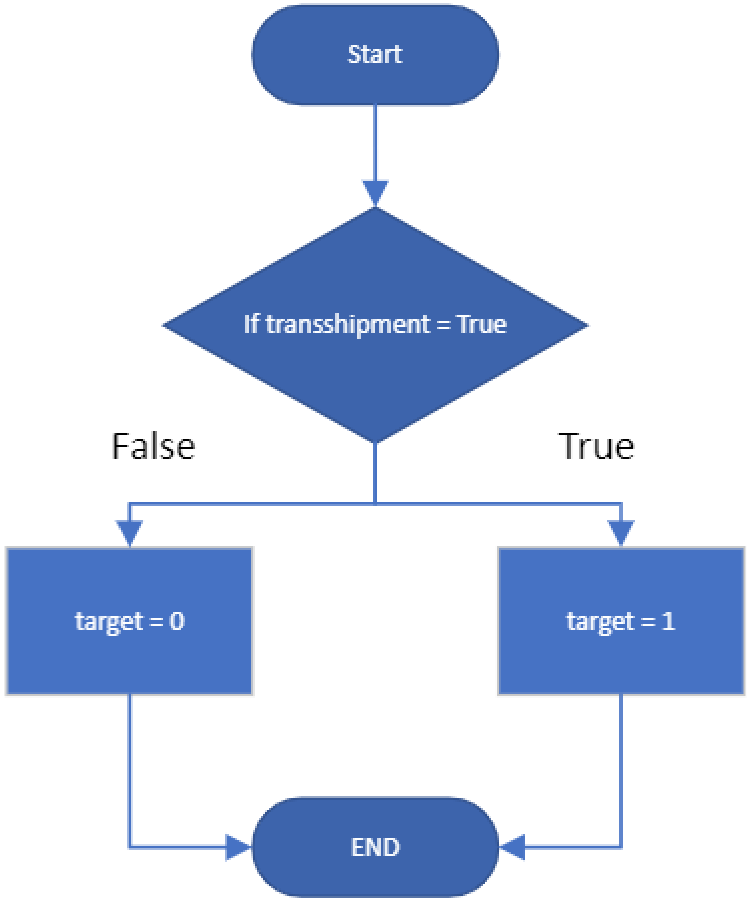

Figure 36: The transshipment model explained

Essentially, this model produces the same output as the transshipment model (target = Transshipment_binary). If a shipment goes through a transshipment port, then it is predicted to be delayed and if does not have a transshipment then it is predicted as on time. Some information around that model is also shown in figure 14. From that plot, it is shown that the model is not going to be extremely accurate.

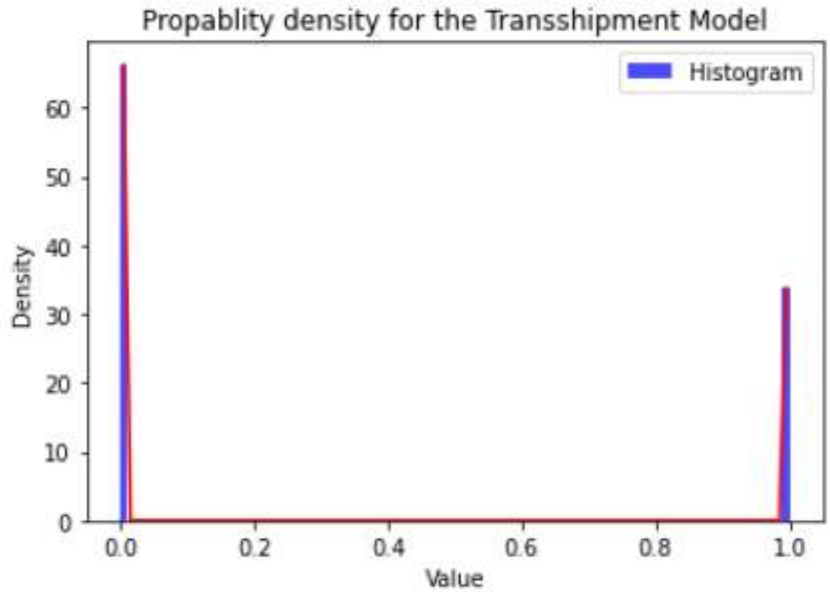

Figure 37: The probability density for the transshipment model

Since this model doesn't output a probability just a definite yes or no, about the delay of the shipments, it is obvious that the shape of the graph is going to be like this. Having 42.7% of the observations going through a transshipment port, translates that 42.7% of the observations are going to be predicted as delayed, while the rest 57.3%, that did not have a transshipment, were predicted as not delayed.

Confusion matrix

Unlike the rest of the models, this classifier does not require training, so a testing dataset does not exist. This means that the confusion matrix is fuller, as it contains observations from the whole dataset.

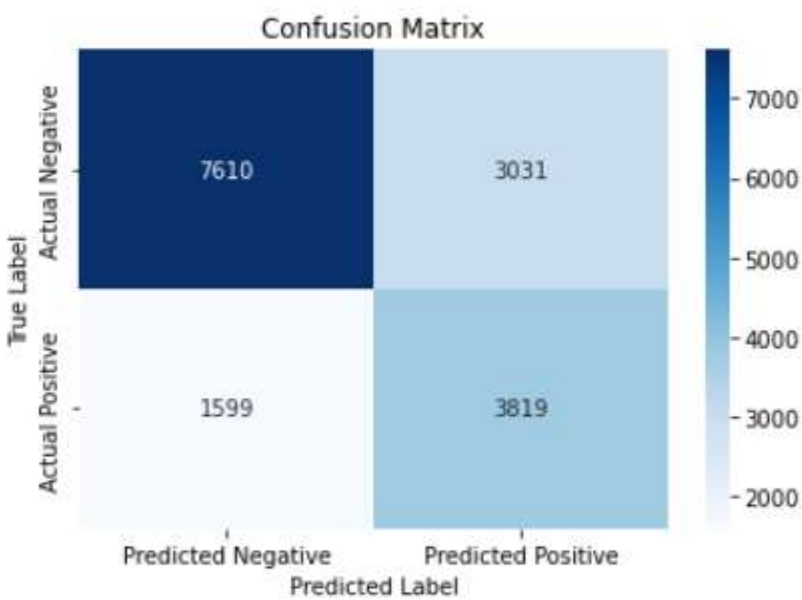

Figure 38: The confusion matrix for the Transshipment based classifier.

The confusion matrix provides a detailed breakdown of the model's performance on the total 16,059 observations in the testing dataset. Ideally, the matrix would show a high concentration of observations along the main diagonal, indicating correct predictions, and few observations on the off diagonal, indicating errors. Out of the 5,418 delayed observations, the model correctly predicted 70% as delayed, but incorrectly predicted 30% as on time. Conversely, of the 10,641 on-time deliveries, 72% were correctly predicted, while the remaining 28% were incorrectly predicted as delayed. This distribution highlights the model's overall accuracy and areas for improvement.

5.4. Label Encoding

In this approach, the features utilized for model creation were not target encoded. Nonetheless, as the initial methodology employed, they provide valuable insights into the developmental process and underscore the importance of target encoding. Through this discussion, readers gain a deeper understanding of the evolution of the modeling process and the significance of incorporating target encoding techniques.

Table 16: Performance of the models with label encoding

Model	F1-score	Accuracy	Precision
Random Forest	0.83	0.87	0.84
XGBoost	0.81	0.83	0.89
Neural Network	0.66	0.78	0.76

It is evident that label encoding is underperforming in all models tried in this case study. However, it is interesting how the encoding technique changes the performance of the models. If the ranking is done by accuracy, then the sorting by performance stays the same. However, XGBoost surpasses the random forest in precision, meaning it has fewer false positives observations, which is ideal for the current case study.

6. Discussion

This concluding section presents the culmination of the investigation into delay prediction in container shipping operations for a multinational corporation, such as Sappi. The analysis undertaken has focused on key methodological aspects, including the evaluation of the feasibility of accurately predicting delays, uncovering underlying causes, and assessing the efficacy of prediction algorithms in optimizing logistical operations.

By examining these methodological elements, the study offers valuable insights into the complexities of container shipping logistics and their implications for strategic decision-making within Sappi or other multinational companies outsourcing their port-to-port logistics.

6.1. Binning Decision

This sub section delves deeper into the structure of the classification problem and the 8-day cut-off. This thesis focuses on predicting delays exceeding eight days in container shipping, which significantly impacts supply chain efficiency and incurs additional costs. These delays can lead to demurrage charges (when containers linger within the terminal beyond a free period) and detention charges (when containers are held outside the terminal past the allotted free time). Despite literature suggesting that demurrage & detention (D&D) costs are relatively low compared to other transportation costs (Storms et al., 2023), Sappi noticed that these costs were disproportionately high. This was revealed in the first stages of the methodology where a lot of meetings were taking place with a logistics specialist based in Sappi EU but having a lot of information on the logistics practices of Sappi global.

Accurately predicting such delays allows stakeholders to proactively manage these costs and optimize shipping schedules. Turning this problem into a binary classification has a lot of benefits in the impact it could have on the company. Primarily, it produces more accurate results compared to a regression analysis, therefore it gains the trust of middle management, which is very difficult to achieve and very important in order for digital tools to be implemented (Bagrationi & Thurner, 2023). Trust is also gained by the simplicity of the model, as it is easier for middle managers to comprehend the prediction inside two classes and the probabilities produced rather than the regression algorithm. It has a direct business outcome, where a container is either on time or delayed, which ultimately, saves time and it is more likely to be correct, if the model has a third option for inconclusive observations, which as discussed in chapter 5, would not be a lot if any of the best-performing tree model is used. This way, only the observations with high probability of being delayed are going to be marked as delayed, while the observations with medium probability are going to be marked indecisive, improving the overall accuracy of the model.

The 8-day period when a shipment was considered delayed was decided after having meetings with the logistics specialist in Sappi Europe. Coincidentally it also aids the model to distinguish better observations near the border, which is also shown and explained in figure (put number).

6.2. Target vs label encoding

Label encoding offers simplicity and efficiency in converting categorical variables into numerical form, making it easy to implement and understand. However, it may introduce unintended ordinality or hierarchy, potentially misleading machine learning algorithms. On the other hand, target encoding provides a powerful way to capture the relationship between categorical variables and the target variable, especially in high-cardinality scenarios. Yet, without proper regularization techniques, it runs the risk of overfitting, particularly when dealing with categories with few observations. Thus, while both encoding methods have their advantages, careful consideration of the dataset's characteristics and modeling goals is essential to choose the most appropriate approach.

However, literature suggests when dealing with high cardinality features such as the ports used by the carriers, target encoding is usually the best performing approach (Pargent et al., 2022). In this thesis, it was decided, after consulting the digital transformation team, that both techniques would be applied, and they produced the results shown in tables 16 and 7. Due to the fact that there were a lot of data and a lot of columns with categorical variables it was essential to find the best way to encode the variables. In this specific case target encoding was the correct choice, as expected. It produced 5% more accurate and 7% more precise results in the best performing model.

6.2.1. Comparison of encoding types by models

In this section these techniques are presented and compared model by model.

Table 17: This table showcases the KPI's of encoding techniques by models.

KPI’s	**F1-score**		**Accuracy**		**Precision**	
Model	label	target	label	target	label	target
Random Forest	0.83	0.87	0.87	0.92	0.84	0.91
XGBoost	0.81	0.87	0.83	0.91	0.89	0.90
DNN	0.66	0.69	0.78	0.81	0.76	0.77

As expected, target encoding outperforms label encoding in every KPI and in every model. It seems that accuracy was the metric that improved the most across all KPI's. In the best performing model – Random Forest – Precision increased the most, which means that the model has more robust predictions of delayed shipments, meaning that if a shipment is predicted as delayed is more likely to be delayed. However, both accuracy and F1 score increased by quite a lot. In XGBoost, precision didn't have such a great increase in Precision, however there were significant increases in the other 2 KPI's. The same phenomenon was observed for neural networks as well.

6.2.2. Feature importance

Moreover, there are other observations that show how important target encoding is for the model. Table 11, figure 22 and figure 27, have all chosen as the most useful features, the features that have been target encoded. For example, the departure or arrival month or port. This means that the models extract most of the information needed to make a prediction by the target encoded features.

6.3. Model comparison

To answer the main research question of this thesis, it is indeed possible to predict delayed shipments with an accuracy and precision exceeding 90%. As demonstrated in table 17, the forest models emerged as the top performers in this task. However, it was notable that Gradient Boosted Trees (XGBoost) exhibited slightly inferior performance compared to random forests. This discrepancy may indicate that the dataset contains significant noise or variability, complicating the model's ability to discern patterns effectively (Tian et al., 2021).

What is notable is how the forest models tend to be surer of their predictions having most of their predictions laying near 0% or near 100% probability, while both the logistic regression and the neural networks classifier had more well distributed probabilities for the observations in the testing set. In most theoretical studies xgboost is outperforming random forests. However, neural networks are performing as expected. Despite being one of the most popular deep learning algorithms, they do not perform that well as there are a lot of features with high cardinality and high dimensionality.

7. Conclusions

7.1. Research significance

In today's globalized world, efficient container shipping is crucial for multinational companies. This research tackles a major challenge: predicting delays. By accurately forecasting disruptions, businesses can optimize logistics and minimize supply chain issues.

This research demonstrates that accurate prediction of delays in container shipping for big multinational companies is achievable. By leveraging machine learning models, companies can gain valuable foresight into potential disruptions, allowing them to optimize logistics and mitigate negative impacts on their supply chains. This study further contributes by comparing the effectiveness of different encoding techniques (target vs label) and various machine learning algorithms (XGBoost, Random Forest, Neural Networks, Logistic Regression) for this specific task. The findings provide valuable insights for future research and practical implementation in the container shipping industry.

The uniqueness of this specific case study, and the noisy dataset offer an opportunity to test these algorithms in application concerning the industry of container shipping outside a purely theoretical academic environment. This is the reason why a comparison of the performance of the models is important. Because of the authenticity of the data, this study belies the conventional hierarchy of predictive models.

Furthermore, the comparison of encoding techniques also provides value, for the same reasons. However, this time, the case study confirms the expectations from theory, as target encoding performs better than label encoding.

7.1. SCCT discussion

The findings from the UX project conducted at Sappi underscore the critical need for more predictive and diagnostic tools within the company's supply chain operations. As revealed through the thematic analysis, supply chain managers at Sappi recognize the importance of leveraging digital transformation initiatives to enhance operational efficiency and decision-making processes. While the company has made strides in developing descriptive tools utilizing platforms like BigQuery and Looker, there remains a clear gap in the availability of personalized predictive and diagnostic tools tailored to Sappi's unique supply chain requirements.

This finding resonates deeply with the objectives outlined in the research questions of this master thesis, particularly addressing the effectiveness of predicting delays in container shipping for a multinational company. The demand for more sophisticated tools aligns with the overarching goal of improving supply chain resilience and efficiency, where accurate prediction algorithms play a pivotal role in mitigating disruptions and optimizing resource allocation.

By incorporating advanced predictive analytics techniques into Sappi's supply chain control tower (SCCT), the company can proactively anticipate potential delays and contingencies, thereby enhancing customer service, transport planning, and overall operational performance. This integration aligns with the broader industry trend towards digitalization and data-driven decision-making, empowering organizations to stay ahead in an increasingly complex and dynamic business environment.

Furthermore, the insights gleaned from the UX project highlight the importance of addressing data quality issues, an aspect crucial to the successful implementation of predictive tools. Despite the challenges posed by poor data quality from carrier companies, investing in data cleansing and enrichment strategies becomes imperative to ensure the reliability and accuracy of predictive models.

In essence, the findings from the UX project provide valuable insights into the practical considerations and challenges faced by Sappi in optimizing its supply chain operations. By leveraging these insights and aligning them with the objectives of this thesis, there is a contribution made to the development of effective predictive models tailored to Sappi's specific supply chain dynamics, ultimately enhancing its competitiveness and sustainability in the global marketplace.

7.2. Future work

This section delves into alternative approaches worth considering for future investigations, beyond the scope of this work.

7.2.1. Demurrage and detention

This thesis focused on predicting delays in container shipping, which significantly impacts demurrage and detention (D&D) costs. However, the current model only considers containers experiencing delays. To achieve a more comprehensive understanding of D&D costs, future research should explore incorporating early arrivals into the analysis.

Here's a suggested approach:

- Three-class classification: Develop a model that classifies container arrivals into three categories: early, on-time, and delayed. This will enable a more nuanced understanding of how various factors influence D&D costs across the entire arrival spectrum.
- Improved binning strategy: Investigate alternative binning methods for D&D costs. The current approach might not fully capture the complexities of these costs, which can vary significantly within the "delayed" category. Exploring techniques like quantile-based binning or clustering could provide a more refined cost distribution.

By incorporating early arrivals and refining the binning strategy for D&D costs, future research can provide a more holistic view of the factors impacting these critical expenses. This can lead to the development of even more comprehensive prediction models that consider the full range of container arrival times and their associated D&D costs.

7.2.2. One hot encoding

This thesis primarily focused on label encoding and target encoding for representing categorical features in the prediction model. While these methods proved valuable, future research could investigate the effectiveness of one-hot encoding for this specific task.

One-hot encoding creates a new binary feature for each category within a categorical variable. These features indicate the presence (1) or absence (0) of a particular category for each observation. Here's why it might be beneficial to explore in the future:

- Improved Interpretability: Compared to target encoding, one-hot encoding can offer greater interpretability. The individual feature coefficients in the model can directly reflect the impact of each category on the predicted delays.
- Handling New Categories: One-hot encoding can gracefully handle new categories that weren't present in the training data. This can be advantageous in real-world scenarios where new data points might contain categories not encountered previously.

However, it's important to consider potential drawbacks, in this specific case:

- Increased Dimensionality: One-hot encoding can significantly increase the data's dimensionality, especially with features containing a large number of categories, like in this case dataset. This can lead to computational challenges and potentially overfit models.

To mitigate these drawbacks, future research could explore techniques for dimensionality reduction alongside one-hot encoding. Feature selection methods or Principal Component Analysis (PCA) could be used to identify the most informative one-hot encoded features, maintaining model performance while reducing complexity.

By incorporating one-hot encoding and exploring strategies to address its limitations, future research can broaden the exploration of different encoding techniques for container shipping delay prediction. This can provide valuable insights into the optimal approach for representing categorical features in this specific domain.

7.2.3. Transferring the model and the methodology

This thesis explored a methodology based on four key stages: describe, predict, diagnose, and prescribe. While applied here to container shipping delay prediction, this framework holds promise for adaptation in various domains. Future research could investigate the transferability of this methodology to other sectors. For instance, in finance, it could be used to analyze historical data to predict market trends and prescribe investment strategies. Similarly, the healthcare field could leverage this approach to analyze patient data, predict disease risk, and recommend personalized treatment plans. By demonstrating the effectiveness of this methodology in container shipping, this research lays the groundwork for its exploration and potential optimization across diverse areas.

7.3. Personal Remarks

This internship was great for my career and my expertise. It was very educating working with the digital transformation team but more than that, this experience shaped my personal and professional advancement in other areas. It is evident that my data science skills have improved by working on a real case study. I am more confident that my data science projects can add business value to a company. Moreover, I had the opportunity to organize and run a UX project which not only introduced me to a new field (UX) but also strengthened my knowledge around logistics and the technologies that can improve it.

However, my soft skills improved the most in this three-month period. I learned how to ask for and conduct a meeting by having 6 interviews in the UX project but also by asking people for information I needed for Pandora’s model. Moreover, I have had many presentations which were skills I had not acquired through my academic studies. Working in an international environment, although challenging at first, gave the opportunity for my communication skills to flourish.

Finally, I want to thank the whole digital transformation team that made sure to give me enough guidance without restricting my ideas. Special thanks to Javed, my internship coach, and Dr. Rentizelas, my thesis supervisor who made sure I had enough resources to complete my projects.

8. References

Alicke, K., Bariball, E., Foster, T., Mauhourat, J., & Trautwein, V. (2022). *Taking the pulse of shifting supply chains Our latest survey shows global supply chains on the move*.

Amellal, I., Amellal, A., Seghiouer, H., & Rida Ech-Charrat, M. (2024). An integrated approach for modern supply chain management: Utilizing advanced machine learning models for sentiment analysis, demand forecasting, and probabilistic price prediction. *Decision Science Letters*, *13*, 237–248. https://doi.org/10.5267/dsl.2023.9.003

Bagrationi, K., & Thurner, T. (2023). Middle Management's Resistance to Digital Change. *Foresight and STI Governance*, *17*(2), 49–60. https://doi.org/10.17323/2500-2597.2023.2.49.60

Chae, B. (2009). Developing key performance indicators for supply chain: An industry perspective. *Supply Chain Management*, *14*(6), 422–428. https://doi.org/10.1108/13598540910995192

Fancello, G., Pani, C., Pisano, M., Serra, P., Zuddas, P., & Fadda, P. (2011). Prediction of arrival times and human resources allocation for container terminal. *Maritime Economics and Logistics*, *13*(2), 142–173. https://doi.org/10.1057/mel.2011.3

Filali, A. E. L., Ben Lahmer, E. H., & Filali, S. E. L. (2021). Exploring applications of Machine Learning for supply chain management. *2021 3rd International Conference on Transportation and Smart Technologies, TST 2021*, 46–52. https://doi.org/10.1109/TST52996.2021.00015

Ganeriwalla, A., Casey, M., Shrikrishna, P., Bender, J. P., & Gstettner, S. (2018). *DOES YOUR SUPPLY CHAIN NEED A BLOCKCHAIN?*

Hathikal, S., Chung, S. H., & Karczewski, M. (2020). Prediction of ocean import shipment lead time using machine learning methods. *SN Applied Sciences*, *2*(7). https://doi.org/10.1007/s42452-020-2951-5

He, Q., Zhao, H., Feng, Y., Wang, Z., Ning, Z., & Luo, T. (2024). Edge computing-oriented smart agricultural supply chain mechanism with auction and fuzzy neural networks. *Journal of Cloud Computing*, *13*(1). https://doi.org/10.1186/s13677-024-00626-8

Laloum Melanie. (2023). *2023 Trade report*. https://iccwbo.org/news-publications/policies-reports/icc-2023-trade-report-a-fragmenting-world/.

Lehmacher, W., Betti, F., Beecher, P., Grotemeier, C., & Lorenzen, M. (2017). *Impact of the Fourth Industrial Revolution on Supply Chains*. www.weforum.org

Mohamed-Iliasse, M., Loubna, B., & Abdelaziz, B. (2020, October 28). Is machine learning revolutionizing supply chain? *Proceedings - 2020 5th International Conference on Logistics Operations Management, GOL 2020*. https://doi.org/10.1109/GOL49479.2020.9314713

Pani, C. (2012). *MANAGING VESSEL ARRIVAL UNCERTAINTY IN CONTAINER TERMINALS: A MACHINE LEARNING APPROACH*.

Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics*, *37*(5), 2671–2692. https://doi.org/10.1007/s00180-022-01207-6

Patsavellas, J., Kaur, R., & Salonitis, K. (2021). Supply chain control towers: Technology push or market pull—An assessment tool. *IET Collaborative Intelligent Manufacturing*, *3*(3), 290–302. https://doi.org/10.1049/cim2.12040

Salleh, N. H. M., Riahi, R., Yang, Z., & Wang, J. (2017). Predicting a Containership's Arrival Punctuality in Liner Operations by Using a Fuzzy Rule-Based Bayesian Network (FRBBN). *Asian Journal of Shipping and Logistics*, *33*(2), 95–104. https://doi.org/10.1016/j.ajsl.2017.06.007

Schimi-Levi, D., Kaminsky, P., & Schimi-Levi, E. (2009). *Designing and managing the supply chain: concepts, strategies, and case studies*.

Storms, K., Sys, C., Vanelslander, T., & Van Deuren, R. (2023). Demurrage and detention: from operational challenges towards solutions. *Journal of Shipping and Trade*, *8*(1). https://doi.org/10.1186/s41072-023-00132-1

Tian, R., Chen, F., & Dong, S. (2021). Compound Fault Diagnosis of Stator Interturn Short Circuit and Air Gap Eccentricity Based on Random Forest and XGBoost. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/2149048

Toygar, A., Yildirim, U., & Nart, S. (2023). Understanding the impact of global challenges on container shipping: a qualitative study in the COVID-19 era. *WMU Journal of Maritime Affairs*. https://doi.org/10.1007/s13437-023-00327-0

Trzuskawska-Grzesińska, A. (2016). Control towers in supply chain management – past and future. *Journal of Economics and Management*.

Viellechne, r A., & Spinler, S. (2020). *Novel Data Analytics Meets Conventional Container Shipping: Predicting Delays by Comparing Various Machine Learning Algorithms*.

Vlachos, I. (2023). Implementation of an intelligent supply chain control tower: a socio-technical systems case study. *Production Planning and Control*, *34*(15), 1415–1431. https://doi.org/10.1080/09537287.2021.2015805

Wisetsri, W., Donthu, S., Mehbodniya, A., Vyas, S., Quiñonez-Choquecota, J., & Neware, R. (2022). An Investigation on the Impact of Digital Revolution and Machine Learning in Supply Chain Management. *Materials Today: Proceedings*, *56*, 3207–3210. https://doi.org/10.1016/j.matpr.2021.09.367

9. Appendix

The Receiver Operating Curve

The Receiver Operating Characteristic (ROC) curve serves as a fundamental tool in evaluating the performance of binary classification models. It visually depicts the trade-off between the true positive rate and the false positive rate across different threshold values. A typical ROC curve plots the true positive rate on the y-axis against the false positive rate on the x-axis, illustrating the model's ability to distinguish between positive and negative instances across varying decision boundaries. The area under the ROC curve (AUC) quantifies the model's overall performance, with a higher AUC indicating superior discrimination ability.

The ideal ROC curve has an area under it equal to 1. That is equivalent to the existence of a threshold that can accurately separate all the on-time deliveries from the delayed. However, in this specific case the area is equal to 0.97, which implies that there are some imperfect thresholds that provide good results. The importance of the threshold can also be shown in the figure above.

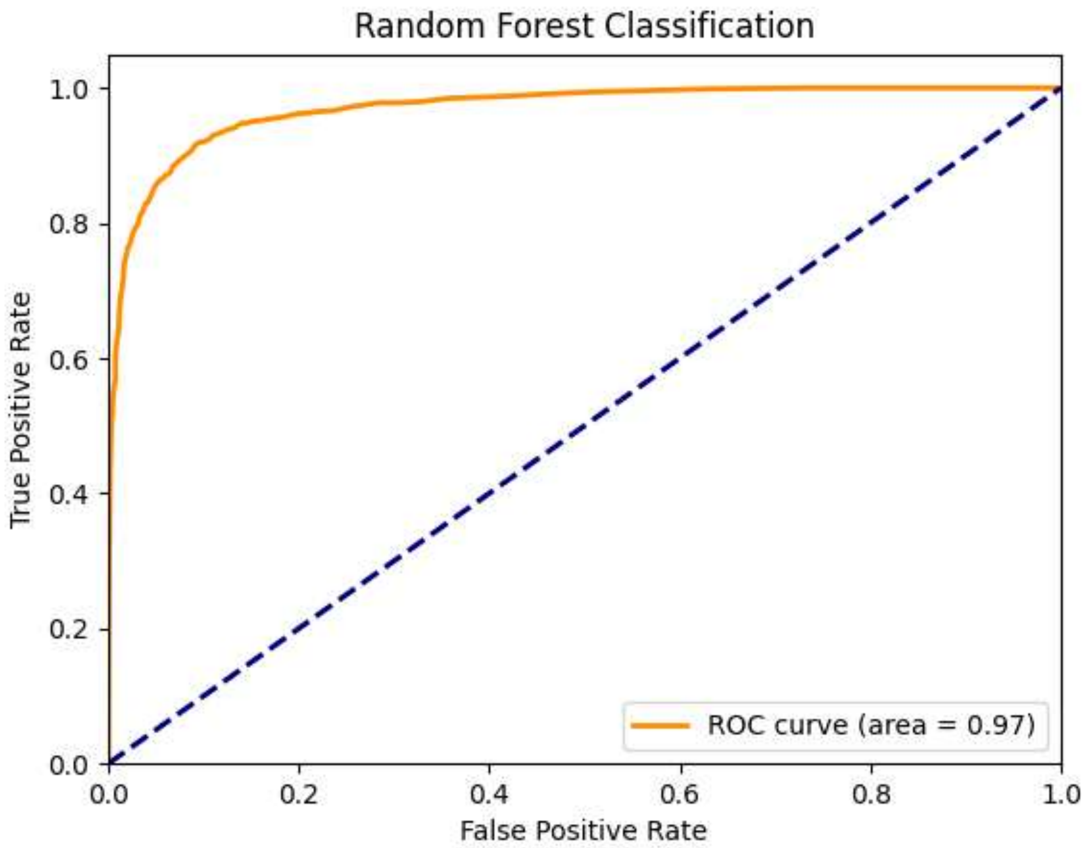

Figure 39: The ROC curve, which showcases the balance between precision and recall.

UX project questionnaire

Operations

1. What digital tools are you currently using to monitor and manage the supply chain?
    1.1. In what supply chain operations do you use decision support system?
    1.2. What do you like about them? What do you hate about them?
2. What according to you should be the capabilities of such a control tower?

    2.1. What decisions do you wish to have support on by this system?
    2.2. How do you take these decisions now?
3. What do you anticipate as core challenges in implementation of Control Towers in your Supply Chain?
    3.1. How important is financing in adoption of Control Towers for you?
    3.2. What are some apprehensions you have regarding adoption of SCCTs?
4. Are autonomy and exception management something you are targeting in the near future?

Information sharing

1. In literature, academics are referring to a SCCT which offers visibility across and between stakeholders. How open is Sappi to information sharing, and how open are Sappi's partners?
    1.1. Can you give some examples?
2. How difficult is getting collaboration from trading partners for such a technology tool?
3. Would you like outside partners to have limited access to the SCCT?

Sustainability

1. How sustainable is Sappi's supply chain?
    1.1. In what ways?
    1.2. What metrics do you use?
2. What digital tools support the sustainable operations of Sappi's Supply chain?
3. Are we using reverse logistics? For what products?
4. Should the SCCT have some sustainability elements? Do you plan on expanding our supply chain to include reverse logistics?

Predictions

1. What do you think affects the delivery date of a container? What is the main driver of delays in container shipping?
2. Question about 4 types of analytics.