

**INTELLIGENT DATA-DRIVEN METHODS FOR  
DEMAND AND PRICE PREDICTION IN THE  
SHIPPING INDUSTRY**

**Ayesha Ubaid**

**School of Computer Science**

**Faculty of Engineering and Information Technology**

**University of Technology Sydney, NSW, Australia**

**A Thesis submitted in fulfilment**

**Of the requirements of the degree of**

**Doctor of Philosophy**

May 10<sup>th</sup> 2021

## **Certificate of Original Authorship**

I, ***Ayesha Ubaid*** declare that this thesis, is submitted in fulfilment of the requirements for the award of ***Doctor of Philosophy, in the FEIT at the University of Technology Sydney***.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:  
Signature removed prior to publication.

Date: 10th May 2021

# SYNOPSIS

Machine Learning has found its applications in many industrial and commercial domains. However, there are few ignorant industries that still lack digitization and require fusion of AI into their processes. Once such industry is the container-shipping industry. The global supply chain is complex, with cargo volumes that are highly seasonal driven by events, consumer related, Christmas and Chinese New Year, agriculture harvests further impacted by extreme weather occurrences and changes to the Geo, Political regulatory environment affecting trade. In contrast, supply, the shipping capacity is fixed in the short run; this results in periods of mismatched supply and demand and therefore shipping price volatility. The Shipping lines are trapped by the current spot market pricing practice of using date validity and not the vessel voyage. This causes a disconnect between price, demand and supply, a problem that is compounded by the shipping line's enterprise systems. Entrenched operational silos within the lines are resulting in missed revenue opportunities through a lack of real time visibility into the availability of equipment and vessel space which are key inputs to price decisions. Forty percent of all shipping containers moved around the world are purchased on the short-term spot market; their commercial terms are set manually with emails and phone calls. Spot market price should be a product of supply and demand. However, the industry as a whole has little visibility into the state of the market in real time, and carriers are making sub optimal pricing decisions and their customer procurement decisions because of this.

Australian Linear Shipping industry is lacking digitization hence the visibility into the industry statistics is missing. Without real time visibility of the market as a whole, future spot pricing decisions are based mainly on a carrier's internal assessment of current and future booking build up and net contribution targets. This can, and usually does, present a different and misleading picture of market conditions. With the limited market wide

information and inability to use vessel voyage specific price as a lever to steer the cargo opportunity to the optimum vessel sailing, opportunity cost materializes.

In this research, we want to empower Australian Container Shipping Industry with machine learning capabilities to provide a market wide view of current and future demand (both for short-term and long-term), optimal spot pricing model and machine learning based prices prediction model that can predict spot pricing based on current demand and available capacity (supply).

To do so, the first step is to explore and model the relationship between demand, supply, and price. This can provide the current market scenario in which industry is operating. Based on the results inferred, a model is designed to set optimal spot pricing. In order to price effective spot pricing model, statistical analysis is done to discover the relationship between demand, capacity and price. Based on inferred results, historic data and spinning companies' limitations for price quotation, a novel mathematical model is designed that can calculate optimal pricing based on the mentioned factors. Finally, for making contract pricing more effective, the demand forecast performed earlier is used with the available capacity to predict optimal container pricing based on demand and supply using a regression based multivariate machine learning model. This predicted price will shorten the gap between contract and spot pricing in the container shipping industry. Hence, this research provides visibility into future demand and will allow shipping lines and their customers to make better-informed pricing and procurement decisions.

## **ACKNOWLEDGEMENTS**

I would like to acknowledge the blessing of Almighty Allah, the most gracious and beneficent for giving me this opportunity to complete my Doctoral Dissertation under their supervision of Associate Professor Farookh Khadeer Hussain and Associate Professor Asif Qamar Gill.

Additionally, I would like to acknowledge the efforts of my family specially my husband Ubaid and mother, Shaheen without whose support and prayers I would have never been the person I am today. A special thanks to my kids, Suleman and Mustafa who despite their quite young age of 5 and 3 were able to understand that their mum has to study. Their positive and helping attitude made things easy for me. To all of my family, I hope this thesis completion will bring positive changes in our lives.

Moreover, I would like to thank my supervisors whose trust and faith in me made me achieve this goal quite comfortably. The never-ending support, compassion, encouragement and superb guidance are the key to my success. This thesis is as much of their work as mine. Finally, I would like to thank the Centre of Artificial Intelligence (CAI) for supporting me throughout this whole three-year tenure.

I would like to dedicate this thesis to my supervisor Dr. Farookh Khadeer Hussain and my beloved husband Ubaid-UR-Rehman who are my real mentors.

# **LIST OF PUBLICATIONS**

## **JOURNAL PAPER**

- Modeling Shipment Spot Pricing in the Australian Container Shipping Industry: Case of ASIA-OCEANIA trade lane Research Question 1
- Container Shipment Demand Forecasting in the Australian Shipping Industry: Case Study of Asia-Oceania Trade Lane Research question 2 and 3 (under review)

## **CONFERENCE PAPERS**

- Framework for feature selection in health assessment systems. Research Question 1.
- Machine Learning-Based Regression Models for Price Prediction in the Australian Container Shipping Industry: Case Study of Asia-Oceania Trade Lane. Research Question 4.

# TABLE OF CONTENTS

1	Introduction .....	1
1.1	Introduction: .....	1
1.2	Background and Motivation: .....	2
1.3	Factors Affecting Shipment Pricing: .....	5
1.4	Gaps in the Current Shipping Pricing Methods: .....	7
1.5	Research/Thesis Scope: .....	7
1.6	Significance of intelligent data driven model for the shipping industry: .....	9
1.6.1	Visibility into industry Stats: .....	10
1.6.2	Future Planning:.....	10
1.6.3	Substantial financial impact:.....	10
1.6.4	Productivity: .....	10
1.6.5	Improved consistency and accuracy:.....	11
1.6.6	Faster and Assured decisions:.....	11
1.6.7	Consistent pricing:.....	11
1.7	Thesis Objectives:.....	11
1.8	Significance of the Thesis:.....	12
1.8.1	Novel Feature Selection Framework: .....	12
1.8.2	Modelling of the relationship between price, demand, and supply: 12	
1.8.3	Model to forecast short-term demand in the Australian Shipping Industry:.....	13
1.8.4	Model to forecast long-term demand in Australian Shipping Industry:.....	13

1.8.5	Model to predict the price based on current demand and available supply: .....	13
1.9	Innovation and Research Contribution: .....	13
1.10	Stakeholders: .....	14
1.11	Thesis Organization: .....	14
1.12	Summary:.....	15
1.13	References:.....	16
2	Literature Review.....	18
2.1	Introduction: .....	18
2.2	Classification of Literature Review:.....	18
2.3	Time Series Forecasting: .....	19
2.4	Forecasting Process: .....	20
2.5	Existing Time Series Forecasting Models: .....	22
2.5.1	Statistical Models:.....	23
2.5.2	Hybrid Models: .....	24
2.5.3	Deep Learning-Based Models: .....	25
2.5.4	Comparative Analysis of existing time series techniques: .....	25
2.6	Feature Selection Techniques:.....	26
2.6.1	Filtration based methods: .....	28
2.6.2	Wrapper Methods:.....	28
2.6.3	Embedded Methods:.....	30
2.6.4	Classification Methods: .....	30
2.6.5	Comparative Analysis of existing feature selection techniques:.....	30
2.7	Optimal Price Calculation Techniques:.....	31
2.7.1	Scenario-based Pricing:.....	32



2.7.2	Algorithmic Pricing:.....	33
2.7.3	Comparative Analysis of existing spot pricing techniques: ....	33
2.8	Machine Learning-based prediction Models: .....	34
2.8.1	Regression-based models: .....	35
2.8.2	Classification-based models: .....	36
2.9	Comparative Analysis of existing feature selection techniques:.....	36
2.10	Overall identified research gaps: .....	37
2.11	Summary:.....	38
2.12	Reference: .....	38
3	Problem Definition and Overview .....	43
3.1	Introduction: .....	43
3.2	Key Concepts:.....	43
3.2.1	Spot pricing: .....	43
3.2.2	Contract pricing: .....	44
3.2.3	Cargo demand:.....	44
3.2.4	Time series forecasting:.....	44
3.2.5	Prediction: .....	44
3.2.6	Multivariate prediction model: .....	44
3.2.7	Relationship:.....	45
3.2.8	EDA: .....	45
3.2.9	Correlation:.....	45
3.2.10	Model:.....	45
3.2.11	Machine learning model:.....	45
3.2.12	Mathematical model: .....	45
3.3	Problem Overview and Problem Definition: .....	46
3.4	Research Issues: .....	47

3.5 Research Scope: .....	48
3.5.1 Determine relationship between shipment demand, available shipping capacity and shipment pricing to determine optimal shipment spot pricing:.....	49
3.5.2 Forecasting short-term container demand using historical demand data incorporating holiday and seasonality effects in the Australian shipping industry:.....	49
3.5.3 Forecasting long-term container demand using historical demand data incorporating holiday and seasonality effects in the Australian shipping industry:.....	50
3.5.4 Developing ML based price prediction model:.....	50
3.6 Research approach to problem solving: .....	50
3.6.1 Choice of Research Method:.....	51
3.6.2 Evaluation of Research Issues 4: .....	54
3.7 Summary:.....	55
3.8 References: .....	55
4 Solution Overview .....	58
4.1 Introduction: .....	58
4.2 Overview of the Solution for implementing intelligent data driven methods for demand and price prediction in the shipping industry: ....	58
4.3 Proposed solutions:.....	59
4.4 <i>Solution Overview for RQ 1: How to model the relationship between demand, supply and price in the Australian container shipping industry?</i> .....	60
4.5 Solution overview RQ 2 AND RQ3: How to forecast short-term demand using historical demand data which incorporates holiday and seasonality effects?.....	62

4.6	Solution Overview for RQ 4: How to build reliable price prediction model given actual shipment demand and available shipping capacity?.....	63
4.7	Solution Overview of RQ5: Evaluation of RQ1, RQ2, RQ3 and RQ4:..	63
4.8	Summary:.....	64
5	Domain Driven Selective Wrapping .....	65
5.1	Introduction .....	65
5.2	Data Set Sourcing: .....	65
5.2.1	Capacity (Supply) Dataset Sourcing: .....	65
5.2.2	Shipment Price Dataset Sourcing:.....	66
5.2.3	Demand Dataset Sourcing .....	68
5.3	Feature Extraction Process: .....	71
5.3.1	Phase 1: Domain Knowledge-Based Filtration: .....	72
5.3.2	Phase 2: Domain Knowledge-Based Correlation Analysis and Wrapping:.....	73
5.4	Application of DSW on the shipment demand dataset:.....	75
5.5	Data Pre-processing and Cleansing:.....	82
5.5.1	Horizon Selection:.....	83
5.5.2	Missing values .....	83
5.5.3	Outlier Detection and removal:.....	86
5.5.4	Data Distribution and Regression Plot:.....	89
5.5.5	Test Train Split: .....	92
5.6	Summary:.....	93
5.7	Reference:.....	93
6	Opti Price-An Optimal Shipping Price Calculation Model .....	96
6.1	Introduction: .....	96

6.2 Research Settings: .....	97
6.3 Relationship Modeling: .....	98
6.4 Optimal Price Calculation:.....	100
6.5 Results and Evaluations: .....	102
6.6 Summary:.....	106
6.7 References: .....	107
7 Container shipment Demand Forecasting for the Australian shipping industry.....	108
7.1 Introduction: .....	108
7.2 Research Design:.....	109
7.2.1 Data Sourcing and Cleansing: .....	109
7.2.2 Missing Value Handling: .....	112
7.2.3 Test Train Split: .....	113
7.3 Forecasting Models: .....	114
7.3.1 SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (SARIMA):.....	114
7.3.2 SEASONAL HOLT WINTERS' METHOD: .....	116
7.3.3 FACEBOOK PROPHET: .....	117
7.4 Results: .....	119
7.5 Evaluation: .....	128
7.6 Summary:.....	130
7.7 References: .....	131
8 Price Prediction Models for The Australian shipping Industry .....	133
8.1 Introduction: .....	133
8.2 Background and Motivation: .....	133
8.3 Experiment Design: .....	134

8.4 Result and Evaluation: .....	137
8.5 Summary: .....	141
8.6 References: .....	141
9 Illustrative Examples of Predicting Shipment Demand and Shipment Price Prediction in Container Shipping Industry.....	143
9.1 Introduction: .....	143
9.2 Intelligent Data Driven Methods for shipment demand and price prediction:.....	144
9.3 Prediction of Short-term and Long-term Shipment Demand in the Container Shipping Industry: .....	145
9.4 Prediction of Shipment Price in the Container Shipping Industry: .....	145
9.5 Abstract Flow Chart of the intelligent data driven methods of demand and price prediction in the Australian shipping industry: .....	154
9.5.1 Flowchart of Short-term and Long-term demand prediction: .....	156
9.5.2 Flowchart of shipment price prediction: .....	157
9.6 Mizzen Software User Interface: .....	158
9.7 Summary: .....	159
10 Conclusion and Future Work.....	160
10.1 Introduction:.....	160
10.2 Thesis Conclusion: .....	160
10.3 Future Research Directions: .....	162
Appendix- A .....	164
Research papers .....	164

# Table of Figures

Fig. 1-1 Price Quotation Process in the Shipping Industry .....	5
Fig. 1-2 Thesis Scope.....	9
Fig. 2-1 Literature Review Classification .....	19
Fig. 2-2 Schultz Forecasting Model .....	20
Fig. 2-3 Shima and Siegel’s forecasting Model.....	21
Fig. 2-4 Brockwell and Devis’ forecasting model .....	22
Fig. 2-5 Classification of time series forecasting models .....	23
Fig. 2-6 Generic feature selection process .....	27
Fig. 2-7 Classification of feature selection methods.....	28
Fig. 2-8 Classification of shipment pricing methods .....	32
Fig. 2-9 ML base Prediction Models .....	35
Fig. 4-1 Solution Overview of Intelligent Data Driven Models for shipment demand and price prediction in the Australian Shipping Industry.....	59
Fig. 5-1 Sea Intelligence weekly capacity dataset snapshot[3]] .....	66
Fig. 5-2 Snapshot of Price dataset (a) SCFI price summary (b) Mizzen price summary .....	68
Fig. 5-3 Demand Source for Shipment Dataset .....	68

Fig. 5-4 Snapshot of Port Botany trade summaries (a) empty and full container demand (b) region-based demand (c) commodity-based demand.....	70
Fig. 5-5 Trade Summaries (a) Fremantle Port (b) Port of Melbourne (c) Flinders Port (d) Port of Brisbane.....	71
Fig. 5-6 Domain Driven Selective Wrapping (a) Phase-1 (b) Phase-2.....	72
Fig. 5-7 DSW phase-1 .....	76
Fig. 5-8 DSW Phase-2 .....	78
Fig. 5-9 Monthly demand for Asia-Oceania trade lane in the Australian shipping industry.....	81
Fig. 5-10 Weekly demand for Asia-Oceania trade lane in the Australian shipping industry [3].....	82
Fig. 5-11 Data Cleansing Process.....	83
Fig. 5-12 Shipment demand dataset visual.....	84
Fig. 5-13 Data description (a) demand data (b) available shipping capacity (supply) data (c) price data .....	85
Fig. 5-14 Snapshot of the Australian shipping dataset for Asia-Oceania trade lane .....	86
Fig. 5-15 Missing Value Checking.....	86
Fig. 5-16 Box plot analysis for dataset variables (a) Capacity (supply) (b) Demand (c) Pricee.....	89
Fig. 5-17 Density distribution and regression plots (a) Shipment Demand and prices (b) Shipment available capacity (supply) and prices .....	90

Fig. 5-18	Scatter matrix of dataset .....	90
Fig. 5-19	Regression plot between Demand, capacity (supply) and prices. .....	91
Fig. 5-20	Correlation heat map between demand, shipping capacity (supply) and prices.....	91
Fig. 5-21	Test train split of demand dataset (70-30).....	92
Fig. 6-1	Density Plot (a) Demand and Price (b) Available shipping capacity (supply) and Price.....	99
Fig. 6-2	Scatter plot and Regression analysis Price vs Demand and Available shipping capacity (supply) .....	99
Fig. 6-3	Density Plot (a) Demand and Price (b) Available shipping capacity (supply) and Price.....	103
Fig. 6-4	Scatter plot and Regression analysis Optimal Price vs Demand and Available shipping capacity (supply) .....	104
Fig. 6-5	Shipment optimal pricing (a) Traditional price calculation (b) Proposed optimal price calculation.....	105
Fig. 6-6	Correlation heat maps (a) current spot pricing (b) optimal spot pricing .....	106
Fig. 7-1	Data Sources for Shipment Demand Dataset.....	110
Fig. 7-2	Overview if the Data Cleansing Process .....	110
Fig. 7-3	Shipment Demand Dataset Visual.....	112
Fig. 7-4	Data description of demand dataset [8] .....	113
Fig. 7-5	Test train split of demand dataset (70-30).....	114



Fig. 7-6 SARIMA parameters used in research (a) Grid search (b) Grid search results snapshot (c) Fitted Model .....	116
Fig. 7-7 Parameters for Holts' Winter Model Training .....	117
Fig. 7-8 FACEBOOK PROPHET Parameters used in this research (a) custom defined holidays (b) model definition for container shipment demand forecasting. ....	119
Fig. 7-9 Shipment Demand Dataset Decomposition.....	119
Fig. 7-10 Demand Forecast by SARIMA (a) Test forecast (b) Train test forecast (c) Short-term forecast (d) Long-term forecast .....	122
Fig. 7-11 Demand Forecast by Seasonal Holt-Winters' (a) Test forecast (b) Train test forecast (c) Short-term forecast (d) Long-term forecast	125
Fig. 7-12 Demand Forecast by PROPHET (a) Test forecast (b) Train test forecast (c) Short-term forecast (d) Long-term forecast .....	128
Fig. 7-13 <b>Error! Use the Home tab to apply 0 to the text that you want to appear here.</b> Comparative RMSE of forecasting models.....	130
Fig. 7-14 Comparative MAPE values of SARIMA, Prophet and Seasonal Holt-Winters .....	130
Fig. 8-1 Demand, supply and pricing data distribution .....	138
Fig. 8-2 Comparison results for models (a) RMSE (b) R2 Score (c) Accuracy .....	140
Fig. 9-1 Forecasted Shipment Demand (a) Short-term shipment demand forecast (b) Long-term shipment demand forecast.....	147
Fig. 9-2 Shipment spot pricing tool (a) Initial UI (b) UI showing the entered shipment demand and capacity (c) UI showing calculated shipment price based on manual pricing dataset. ....	151

Fig. 9-3	Shipment spot pricing tool (a) Initial UI (b) UI showing the entered shipment demand and capacity (c) UI showing calculated shipment price based on Opti-price.....	154
Fig. 9-4	Abstract Flowchart of Intelligent Data Driven Methods for Shipment Demand and Price Prediction in the Australian Shipping Industry .....	155
Fig. 9-5	Flow chart for Demand Forecasting Models.....	157
Fig. 9-6	Flow chart for Price Prediction Model.....	158
Fig. 9-7	Mizzen Software Design.....	159

# Table of Tables

Table 2-1 Summary of Time Series Models Application in the Container Shipping Industry.....	26
Table 2-2 Summary of Feature Selection Method's Application in the Container Shipping Industry .....	31
Table 2-3 Summary of spot pricing techniques designed so far in the Container Shipping Industry .....	34
Table 2-4 Summary of regression- based machine learning model's application in the Container Shipping Industry .....	37
Table 4-1 Features used in the research .....	61
Table 5-1 List of Demand Dataset Features [3] .....	80
Table 7-1 List of Demand Dataset Features .....	111
Table 7-2 Demand Dataset Feature for Asia Oceania Imports .....	111
Table 8-1 List of features selected from the demand dataset.....	136
Table 8-2 List of features selected from supply dataset. ....	136
Table 8-3 List of features from pricing dataset.....	136
Table 8-4 Dataset used in research (import price prediction for Asia-Oceania trade lane in the Australian shipping industry). ....	136
Table 8-5 Data summary of key variables.....	138
Table 8-6 Accuracy percentage of regression models.....	140

# LIST OF ACRONYMS

ML	Machine learning
UI	User interface
DS	Dataset
FS	Feature selection
LR	Linear R
NNs	Neural networks
AR	Autoregressive
SVR	Support vector regressor
SVM	Support vector machines
GBR	Gradient boosted regression
RFR	Random forest regressor
RMSE	Root mean squared error
R2 SCORE	R square score
EDA	Exploratory data analysis
MAPE	Mean absolute percentage error
LSTM	Long-term Short-term memory
ARIMA	Autoregressive moving average
SARIMA	Seasonal Autoregressive moving average
PACF	Partial auto correlation function