PRODUCT DEMAND FORECASTING VIA MACHINE LEARNING

CHEW WAY YAN

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF DATA SCIENCE

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITI MALAYA

KUALA LUMPUR

2021

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate : Chew Way Yan

(I/C) No. : 970905-14-5149

Registration No. :

Name of Degree: Master in Data Science

Title of Research Report/Dissertation/Thesis ("this Work"): Product Demand Forecasting via Machine Learning

Field of Study: Machine Learning

do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the Universiti of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work. I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate Signature

Date 28 January 2022

Submitted and sincerely declared before:

Witness Signature

Date 28/01/2022

DR KASTURI DEWI A/P VARATHAN
DEPT OF INFORMATION SYSTEMS
FACULTY OF COMPUTER SCIENCE &
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
50603 KUALA LUMPUR

Name: Dr.Kasturi Dewi Varathan
Designation: Senior Lecturer

ABSTRACT

Product demand has high volatility where there are full of uncertainties. Hence, an accurate forecast model would be essential to predict the future product demand. Various statistical approaches have been developed to forecast demands; however, their performances were not desirable due to limitations to linear relationships. Hence, Artificial Intelligence (AI) technology emerges where machine learning (ML) models could achieve better performance where they exhibit better forecasting power. Despite the excellent performance exhibits by ML models, there are still rooms for improvement. This study aims to (1) development of ML models, (2) parameter optimization of ML models as well as (3) ML model evaluation and assessment. This study has developed six (6) ML models – Support Vector Regression (SVR), nonlinear regressive neural network (NARNN), Gradient Boosting Regression Tree (GBRT), Long Short-Term Memory (LSTM) and AdaBoost Regressor (ABR). Neural Network Time Series package application in MATLAB and Python coding program version 3.8.8 were used to preprocess the dataset and develop ML models in this study. Subsequently, the performance of the models would be evaluated via root mean square error (RMSE) and mean absolute percentage error (MAPE). It is found that SVR incorporate with polynomial kernel exhibits the best performance where it achieved RMSE and MAPE at 0.0565 and 0.83% for product demand forecasting.

Keywords:

Machine learning, Demand Forecast, Support Vector Regression

ABSTRAK

Permintaan produk mempunyai turun naik yang tinggi di mana terdapat penuh ketidakpastian. Oleh itu, model ramalan yang tepat adalah penting untuk meramalkan permintaan produk masa depan. Pelbagai model statistik telah digunakan untuk meramalkan permintaan, tetapi hasil keputusan mereka tidak mencapai tahap yang memuaskan kerana terhad kepada hubungan linear. Oleh itu, teknologi *Artificial Intelligence (AI)* berkembang kerana model *Machine Learning (ML)* dapat mencapai prestasi yang lebih baik di mana mereka mempamerkan kuasa ramalan yang lebih baik. Walaupun model *ML* mencapai ramalan cermerlang, model *ML* masih mempunyai ruang untuk penambahbaikan. Kajian ini bertujuan untuk (1) melaksanakan model *ML*, (2) mengoptimumkan parameter model *ML*, dan (3) penilaian dan perbandingan model *ML*. Kajian ini telah melaksanakan enam (6) model *ML – Support Vector Regression (SVR), nonlinear regressive neural network (NARNN), Gradient Boosting Regression Tree (GBRT), Long Short-Term Memory (LSTM)* dan *AdaBoost Regressor (ABR)*. Pakej aplikasi bernama *Neural Network Time Series* dalam MATLAB dan aplikasi Python versi 3.8.8 telah digunakan untuk pra-pemproses set data dan melaksanakan model *ML* dalam kajian ini. Seterusnya, prestasi model akan dinilai melalui *root mean square error (RMSE)* dan *mean absolute percentage error (MAPE)*. Berdasarkan hasil kajian, didapati bahawa *SVR* menggabungkan dengan kernel *polynomial* mempamerkan prestasi model yang terbaik di mana ia mencapai *RMSE* dan *MAPE* pada 0.0565 dan 0.83% untuk ramalan permintaan produk.

Kata kunci:

Pembelajaran mesin, ramalan permintaan, *Support Vector Regression*

ACKNOWLEDGEMENTS

First and foremost, I would like to use this opportunity to express my deepest gratitude to University of Malaya (UM) for providing the platform for the students to expose in completing this Data Science Research Project with excellent outcomes and achievements.

Furthermore, I would like to express my appreciation to my supervisor, Dr. Kasturi Dewi A/P Varathan for her patient guidance and support on this research project throughout the semester. In spite of his heavy workload, she continually led me on this research project to ensure everything to be on track. Moreover, she has imparted me with the knowledge relevant to the project to strengthen my fundamental and concepts regarding the research. Her attentive supervision and guidance are much appreciated.

Many people, especially our classmates and seniors have made valuable comments and suggestions on this research which gave me an inspiration to improve the research project. Hence, I would like to thank all people for their help directly and indirectly in this research project.

Last but not least, I would like to thank my loved ones, who have supported me with moral support throughout the process, both in holding me harmonious and helping me putting pieces together. I will be thankful for their affection.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABR	AdaBoost Regressor
AI	Artificial Intelligence
A-MAPE	Alternative Mean Absolute Percentage Error
ANN	Artificial Network
ARIMAX	Autoregressive Integrated Moving Average Exogenous Factor
CART	Classification and Regression Tree
Crisp-DM	Cross Industry Standard Process for Data Mining
DNN	Deep Neural Network
GBRT	Gradient Boosting Regression Tree
KNN	K-nearest Neighbor
LR	Linear Regression
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLR	Multilinear Regression
MSE	Mean Square Error
NB	Naïve Bayes

NARNN	Nonlinear Autoregressive Neural Network
R^2	Coefficient of Determination
RBF	Radial Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
XGBoost	Extreme Gradient Boosting

CHAPTER 1: INTRODUCTION

1.1 Overview of Product Demand Forecast (Business Understanding)

Manufacturing is the process of converting raw resources into intermediate or final commodities that fulfil the specification needs of the client. Each manufacturing company would have their respective standard operating procedure to produce their goods. In general, several lines of business processes are engaged in manufacturing companies, including procurement, in and out inventories, production processes, administration, accounting and finance, and sales and marketing, where these processes are chained to one another.

The overall timeframe of manufacturing production and delivery time ranged from weeks or even months to progress in order to meet the needs of market demand depending on the product complexity and shipping distances. Demand forecast may come in handy as it provides the vision for future market demand to the manufacturers based on training via historical sales data. Hence, it enables the manufacturers adapt to demand in time in order to ensure continuous supply to feed the market demand, as well as maximizes the company revenues (Tirkeş, Güray & Çelebi, 2017).

In fact, market demand is full of uncertainties and huge fluctuation in demand owing to the global competition have resulted in more erroneous demand forecasts, as are the costs of those errors associated with them, causing in less profit or even loss (Ren, Chan & Siqin, 2019). It varies occasionally which mainly contributed by seasonality and promotional impacts, competition activities and low product sales conversion (Kantasa-ard, Nouiri, Bekrar, El Cadi & Sallez, 2020; Khan, et al., 2021; Priyadarshi, Panigrahi, Routrov & Garg, 2019; Ren, Chan & Siqin, 2019). Inaccurate forecast demand would disrupt the availability of goods in storage where the commodities supply may not be able to keep up with the market demand at times, resulting in shortage.

Thus, accurate demand forecast is essential to provide critical information to the manufacturer to do preplanning on production and delivery process. It is critical for the manufacturers as it is the preliminary stages for all production planning activities and execution procedures, regardless of whether a supply chain system is push or pull (Merkuryeva, Valberga & Smirnov, 2019). It enables manufacturers to rethink and revamp not just the layout of their supplier networks, timetables for creating and delivering raw materials, transportation, and the quantity and placement of warehouses but also the designs of their goods. Besides, high demand forecast accuracy may aid the formulation of a market strategy, resulting in increased stock turnover, reduction in supply chain costs, and an improvement in customer satisfaction and market competition (Khan, et al., 2020; Loureiro, Miguéis & da Silva, 2018; Priyadarshi, et al., 2019; Wu & Chen, 2021).

Demand predictions can be created on a monthly, weekly, daily, or even hourly basis to assist various planning processes and business choices, but very detailed forecasts are always incredibly helpful (Mor, Jaiswal, Singh & Bhardwaj, 2018). When it comes to goods with short shelf life, the advantages of a granular forecast are evident, since their short shelf life may necessitate intra-day predictions at the product-location level to avoid wastage and spoilage (Afifi, 2020). Thus, manufacturers may require multiple demand forecasts with varying levels of granularity to examine various time periods to discover critical information for their production and storage planning.

Over the years, statistical approaches such as moving averages models, Holt-Winters method, regression models and so forth were traditionally being used to anticipate demand (Kantasa-ard, et al, 2020; Mor, et al., 2018; Tirkeş, Güray & Çelebi, 2017). These statistical approaches are widely utilized due to their simplicity and speed, as well as the fact that they produce excellent results in a wide range of forecasting applications. However, it was discovered that these statistical approaches are restricted to linear relationships, which greatly influence the performance of demand forecasting models (Afifi, 2020). To address the issue of linearity, the statistical approaches were replaced with machine learning to anticipate demand.

Besides of the improvement of demand forecasting accuracy via machine learning approaches, it also offers automation of planning labor and processing of massive amounts of datasets (Afifi, 2020; Bousqaoui, Achchab, & Tikito, 2017). Furthermore, a wide range of variables such as weather and geographical data can be taken into consideration in machine learning algorithms to create a robust system to improve adaptation to changes while also improving demand forecasting (Huber & Stuckenschmidt, 2020; Tsoumakas, 2018).

Among the machine learning models, regression analysis is primarily used for predictions, as well as inferring the causal relationships between the independent and dependent variables making the system scalable and robust. Hence, machine learning regression analysis would be performed in this study and the models would be assessed and identify the optimal machine learning regression model for demand forecasting.

1.2 Problem Statement and Background Study

For the time being, the accuracy of demand forecasting models remains a major problem (Merkuryeva, et al, 2019). Due to its high processing speed and simplicity, conventional demand forecasting was done using statistical approaches such as regression models, Bayesian analysis, and so on. However, it was discovered that it was restricted to linear relationships, which has a major impact on the accuracy of demand forecasting models (Afifi, 2020).

Artificial intelligence (AI) methods such as neural networks, support vector machines, and others, on the other hand, have recently gained popularity since they cope well with complexity and ambiguity while providing higher forecast accuracy (Bousqaoui, Achchab, & Tikito, 2017). Despite AI approaches obtained better performance with a forecast accuracy up to 92.38% but there is room for improvement to better predict the future demand trend (Khan, et al., 2020).

1.3 Research Objectives

The main goals of this study are:

- To optimize the parameters of ML models for product demand forecast
- To develop machine learning (ML) models based on the optimized parameters to forecast product demand
- To evaluate and compare the performance of ML models for product demand forecast

1.4 Content of Chapter

The brief introduction, background studies, problem statements as well as the research objectives are all included in Chapter 1.

The machine learning regression models are demonstrated in Chapter 2 along with the dataset taken from earlier work in tabular form, as well as the gaps in previous studies that are being addressed in this study.

The methodology, which comprises the Crisp-DM Framework, data description and preprocessing, machine learning models, and performance evaluation metrics to be employed in this study, is detailed in Chapter 3.

Furthermore, the outputs of each regression model with optimal parameters will be explored in depth in Chapter 4, along with the overview of the research outcomes.

Last but not least, Chapter 5 summarizes the entire study, in the comparison of research goals and objectives.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The product demand defines the tendency of a product to be desired during a certain time period. The demand fluctuates throughout time as a result of known trends such as seasonality factors or unanticipated external causes such as economic disruptions and natural catastrophes (Kantasa-ard, et al., 2020; Khan, et al., 2021; Priyadarshi, et al., 2019; Ren, Chan & Siqin, 2019). In fact, a larger volume of transactions of the specified product spikes the demand which might lead to price increases, since the greater the product demand, the more consumers are willing to pay. As vice versa, when demand falls, the price tends to reduce along with it.

By observing current and historical market demand can show trends in the organization's sales data to impact the future purchase intention and minimizes the issues of understocking or out-of-stock situations when the manufacturers organize their inventory management and replenishment schedules. Thus, various research has been done to discover the potential of machine learning models for product demand forecast.

This chapter discusses the past studies related to product demand forecast which uses machine learning models. This section covers the product demand forecast techniques and performances, as well as the dataset used from the past studies to identify the gaps of research, thus address them in this study.

2.2 Product Demand Forecast Techniques and Performances

2.2.1 Techniques for Product Demand Forecast

According to past studies, supervised learning algorithms are often employed for product demand forecasting (Afifi, 2020; Khan, et al., 2020; van Nguyen, et al., 2020). Labeled datasets are used to train algorithms that categorize data or predict outcomes in supervised machine learning models. There are two types of problems in supervised learning: regression and classification. Regression techniques predict continuous values, whereas classification algorithms predict discrete values into categories. Table 1 shows the techniques used in estimating product demand based on past studies

Table 1: Techniques and Model Used for Product Demand Forecast

No	Reference	Techniques	Model Used
1	Wu & Chen (2021)	Regression	NARNN, Prophet
2	Afifi (2020)	Classification & Regression	OneR, Naïve Bayes, KNN, RIPPER, K-means clustering, RULES-6
3	Kantasa-ard, et al. (2020)	Regression	SVM, MLR, ARIMAX, LSTM
4	Khan, et al. (2020)	Classification	RNN
5	van Nguyen, et al. (2020)	Regression	Linear Regression, CART, M5 Model Tree, ANN, RF
6	Priyardarshi, et al. (2019)	Regression	LSTM, SVR, RF, GBR, XGBoost, ARIMA
7	Kaya & Turkyilmaz (2018)	Regression	ANN, SVM, C5, Croston
8	Loureiro, et al. (2018)	Regression	Deep Neural Network, Decision Tree, SVM, ANN, Linear Regression
9	Mor, et al. (2018)	Regression	MA, Holtz-Winter, Linear Regression, Multiple Regression
10	Tirkeş, et al. (2017)	Regression	Trend Analysis, Decomposition, Holtz-Winter
11	Ferreira, et al. (2016)	Regression	Regression Tree + Bagging

According to Table 1, the majority of previous research employed regression models to anticipate product demand since demand data is often continuous and changes over time. Thus, only regression model would be considered for demand forecasting in this study.

2.2.2 Performance Metrics Used for Product Demand Forecast

Performance assessment is a vital element of machine learning as it measures the effectiveness of ML models. To evaluate performance of these ML models, various evaluation metrics included RMSE, MAPE, MAE, R^2, accuracy, etc., were used for product demand forecast. Table 2 lists the models along with their respective performance based on the past studies.

Table 2: Performance of ML models for Product Demand Forecast of past studies

No	Reference	Proposed model	Performance				
			RMSE	MAPE (%)	MAE	R^2	Accuracy (%)
1	Wu & Chen (2021)	**NARNN**	**0.5865**	-	-	-	-
		Prophet	0.7305	-	-	-	-
2	Afifi (2020)	OneR	84.3	5.2	70.7	-	-
		NB	78.9	4.5	63.2	-	-
		KNN	66.8	3.7	51.3	-	-
		RIPPER	21.1	2.9	16.4	-	-
		K-means clustering	12.9	2.3	9.7	-	-
		RULES-6	**7.3**	**1.8**	**5.9**	**-**	**-**
3	Kantasa-ard, et al. (2020)	SVM	151.14	11.1	98.35	0.91	-
		MLR	150.16	10.97	99.56	0.91	-
		ARIMAX	150.15	11	99.75	0.91	-
		LSTM	149.24	10.97	98.05	0.92	-
4	Khan, et al. (2020)	RNN	-	-	-	-	83.9
5	van Nguyen, et al. (2020)	Linear Regression	0.142	-	0.0108	0.15	-
		CART	0.148	-	0.0111	0.14	-
		M5 Model tree	0.154	-	0.0110	0.12	-
		ANN	0.147	-	0.0115	0.13	-
		RF	0.129	-	0.0097	0.29	-

Table 2: Demand forecast techniques and performances of past studies (continued)

No	Reference	Proposed model	Performance				
			RMSE	MAPE (%)	MAE	R^2	Accuracy (%)
6	Priyardarshi, et al. (2019)	LSTM	3.75	0.18	3.28	-	-
		SVR	6.28	4.88	6.05	-	-
		RF	7.05	4.80	4.87	-	-
		GBR	6.57	4.62	4.74	-	-
		XGBoost	6.92	5.60	5.57	-	-
		ARIMA	12.61	10.39	10.39	-	-
7	Kaya & Turkyilmaz (2018)	ANN	2.66	-	2.17	-	-
		SVM	**1.35**	**-**	**0.67**	**-**	**-**
		C5	2.93	-	2.34	-	-
		Croston	4.38	-	3.16	-	-
8	Loureiro, et al. (2018)	Deep Neural Network	820.1	14.7	-	-	-
		Decision Tree (DT)	937.4	11.7	-	-	-
		RF	983.9	14.1	-	-	-
		SVM	**618.5**	**6.9**	**-**	**-**	**-**
		ABB	1045.6	20.8	-	-	-
		Linear Regression	1696	38.7	-	-	-
9	Mor, et al. (2018)	Moving Average (MA)	1500.0	114.0	-	-	-
		Holtz-Winter	550.5	18.8	-	-	-
		Linear Regression	1225.3	105.8	-	-	-
		Multiple Regression	321.4	15.0	-	-	-
10	Tirkeş, et al. (2017)	Trend Analysis	-	2.47	-	-	-
		Decomposition	-	1.35	-	-	-
		Holtz-Winter	-	1.64	-	-	-
11	Ferreira, et al. (2016)	Regression Tree + Bagging	-	-	-	-	-

According to Table 2, it is observed that the common performance evaluation metrices are root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), coefficient of determination (R^2) and accuracy. In this study, MAPE would be used for the comparison as most of the past studies were using MAPE to evaluate their model performance. According to findings of Lewis (1982), models which achieved MAPE lower than 10 percent are considered as highly accurate forecasting model. Thus, the best model from each literature and MAPE value of 10 percent would be used as the bench line for the selection of models.

Long Short-Term Memory (LSTM), Nonlinear autoregressive neural network (NARNN), Support Vector Regression (SVR), Gradient Boosting Regression Tree (GBRT) and RULES-6 technique were chosen with the criteria of best model from the past studies as well as the 10 percent MAPE benchmark. Out of the five (5) selected models, RULES-6 is classed as a rule induction algorithm which is not regression. Hence, RULES-6 algorithm would be neglected whereas LSTM, NARNN, SVR and GBRT which uses regression approaches would be developed and examined in this study.

Amongst the chosen regression models, GBRT was not the best model from any past studies for product demand forecasting, however, it exhibits performance which is slightly poorer as compared to to SVR (Priyadarshi, 2019). In fact, the GBRT model has only used least squared error loss function and decision tree as their base model for model assessment. The potential of GBRT might be undiscovered as no parameter was tuned in the study. On the other hand, it is deemed that GBRT provides a more promising results for regression forecasting as compared with support vector machines and neural network models based on the past studies. (Cai, Xu, Zhu, Hu & Li, 2020; Gong, Bai, Qin, Wang, Yang & Wang, 2020; Li, Li & Xu, 2018). Thus, GBRT with various parameter settings would be developed in this study and their performance to discover the potential of GBRT for product demand forecasting.

On top of that, AdaBoost Regressor (ABR) has never been developed for product demand forecasting. Various studies have developed ABR and found that ABR provide promising results in regression forecasting (McGlynn, Coleman, Kerr & McHugh, 2018; Pan, Shi, Wang & Li, 2017; Taufiqurrahman, Putrada & Dawani, 2020). Since ABR has never been tested before and create a research gap in product demand forecasting, ABR is proposed to be developed in this study and assess its performance to identify the potential of ABR for product demand forecast.

2.3 Data Understanding of Product Demand Forecast

2.3.1 Product Demand Forecast Datasets

Besides techniques and performance evaluation of the results, dataset is also essential to gather information and insights for the research for product demand forecast. To avoid the concept of 'garbage in, garbage out', the previous research would be studied in advance to have a better understanding on the dataset being used for product demand forecast. The past studies could provide information and insights on the dataset or even data features to be used to predict product demand with high forecasting power. To look into the dataset that being used from the past studies, Table 3 shows the sources and domains of dataset for product demand forecast.

Table 3: Datasets used for Product Demand Forecast based on past studies

No	Reference	Domain	Dataset		
			No. of instances	Source	Status
1	Wu & Chen (2021)	Manufacturing	1,048,576 (6 years)	Kaggle	Open
2	Afifi (2020)	Electronics accessories	507 products	2B Egpyt	Closed
3	Kantasa-ard, et al. (2020)	Agricultural products	3000 observation days	OAE Thailand	Closed
4	Khan, et al. (2020)	Electronics gadgets	-	Not mentioned	Closed
5	van Nguyen, et al. (2020)	Remanufacturing products	-	Amazon Electronics	Closed
6	Loureiro, et al. (2018)	Fashion	684 types of women bags	Fashion company (Spain & Portugal)	Closed
7	Mor, et al. (2018)	Dairy products	9 dairy products	Milk Processing Unit (Northern India)	Closed
8	Tsoumakas (2018)	Food	(4 years data)	Not mentioned	Closed
9	Tirkeş, et al. (2017)	Food	-	Tarihi Yudumla	Closed
10	Ferreira, et al. (2016)	Fashion	2 observation years	Rue La La	Closed

According to Table 3, the dataset covers various domains including manufacturing, electronic products, fashion products, etc., for product demand forecast. In general, product demand forecast usually took demand data for long period (several years) into consideration. As more data was used, the more the data would be used to train the ML models, the better performance of the ML models. Besides, all datasets being used from the past studies were having closed source of dataset except for the research paper authored by Wu and Chen (2021). Hence, the open-source dataset would be utilized in this project where the results obtained from the research paper of Wu and Chen (2021) would also be used for benchmarking to compare and evaluate the performance of each tested models in this study.

2.3.2 Data Feature Analysis for Product Demand Forecast

To further study the dataset being used in past studies for product demand forecast, the attributes are studied. Table 4 lists the data features used in the past studies for product demand forecast.

Table 4: Data Features Used in Past Studies for Product Demand Forecast

No	Reference	Features			
		Sales	Inventory	Calendar	Weather
1	Wu & Chen (2021)	✓			
2	Afifi (2020)	✓			
3	Kantasa-ard, et al. (2020)	✓			
4	Khan, et al. (2020)	✓	✓	✓	
5	van Nguyen, et al. (2020)	✓			
6	Loureiro, et al. (2018)	✓			
7	Mor, et al. (2018)	✓			
8	Tsoumakas (2018)	✓		✓	✓
9	Tirkeş, et al. (2017)	✓			
10	Ferreira, et al. (2016)	✓			

Based on Table 4, various data features include sales, inventory, calendar, and weather data were used to perform demand forecast. However, the most commonly data features used are sales data. As the research by Wu and Chen (2021) would be used as benchmarking, hence only sales feature would be taken into consideration in this study.

2.4 Conclusion

Various studies were examined in this chapter in order to have a better knowledge of the art of product demand forecasting. In this study, only regression model will be used to forecast future product demand. Various machine learning models, namely SVR, NARNN, GBRT, LSTM, and ABR, will be develop with various parameter settings and their performance will be evaluated using RMSE and MAPE. In this study, the similar dataset as the benchmarking research done by Wu and Chen (2021) would be utilized in this study as the outcomes of the study would be compared, with just demand or sales data features being included for model predictors.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter covers the formal methodology for product demand forecast, then evaluate the new solution to predict the future product demand. In this chapter, theoretical framework, detailed description of demand data set, data preparation processes, machine learning models and performance evaluation approaches would be discussed and explained.

3.2 Methodology Framework – Crisp-DM

Cross Industry Standard Process for Data Mining (Crisp-DM) is the standard methodology used for data mining projects. It is widely used analytic methodology as it provides the common structure to guide the execution of all data mining projects ranges from data simulation to narrative exploration (Martinez-Plumed, Contreras-Ochando, Ferri, Hernandez-Orallo, Kull, Lachiche, Ramirez-Quintana, Flach, 2021). Crisp-DM model is a life cycle model which consists of six (6) phases, namely business understanding, data understanding, data preparation, modelling, evaluation, and deployment which was illustrated in Figure 1.

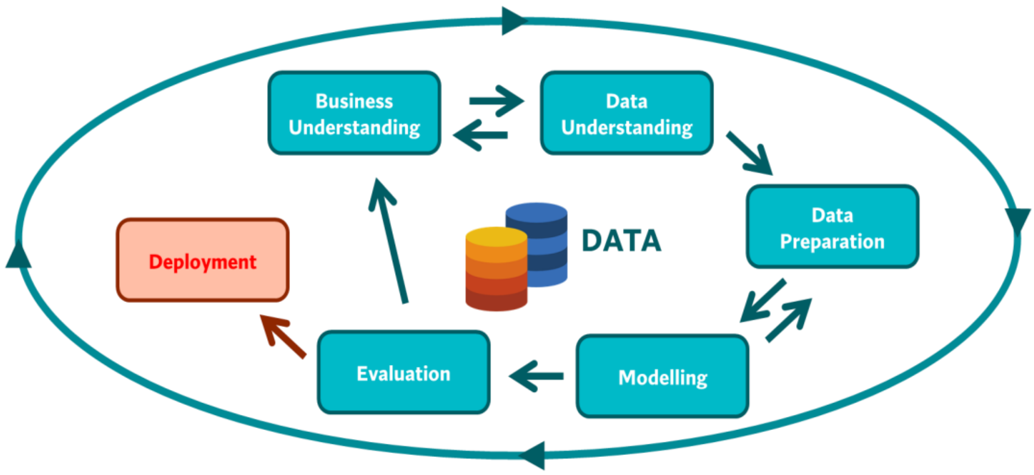

Figure 1: Diagram of Cross Industry Standard Process for Data Mining (Crisp-DM)

The Figure 1 the arrows depict the frequent dependencies between one phase to another. Stage one, the business understanding phase, evaluates the organization's existing condition, including resource fact-finding, limitations, assumptions, and other variables, before identifying business objectives to be met, which will affect project outcomes and es4lish desired outputs. In this study, business understanding phase would be covered in Chapter 1 (Introduction) and Chapter 2 (Literature review) which the research objectives are identified with the common approaches to predict product demand along with their performance are assessed. The second stage necessitates data collection for description, exploration, and quality assurance would be covered in subsection 3.3 (Dataset). Proceed to stage three, data preparation is required, which includes data cleansing, transformation, and integration to verify that the data structure is suitable for modelling. In subsection 3.4 (Data Preparation), the detailed data preprocessing including data cleaning and resampling, data transformation as well as outlier removal would be described. On the other hand, the model techniques along with the parameter settings and particular assumptions about the data, would be chosen for the modelling step would be discussed in subsection 3.5 (Machine Learning Models). As for evaluation stage, it assesses the degree to which the model meets the business objectives and review the thorough process of the data mining engagement as it covers in Chapter

4 (Results and Discussion). Based on the evaluation and process review, decision would be made either to initiate further iterations or proceed to deployment. The approach in this article ends at the assessment step, when the models' performance was evaluated in order to find the best machine learning model for forecasting product demand.

3.3 Tools Used

Two (2) software were employed in this study. For data preparation phase, the coded Python program were utilized, which compromised data transforming, filtering, and training-testing data separation functions. For model execution, both MATLAB and Python software were utilized, with the NARNN model being predicted using the *Neural Network Time Series package* application in MATLAB and the SVR, GBRT, LSTM and ABR models were being predicted using a coded Python program.

3.4 Dataset

The target variable of this study is the product market demand of a manufacturing company. This project considered the historical product demand data which recorded from year 2011 to 2017 of the manufacturing firm which involved global operation was extracted from a platform for data science competition named Kaggle (https://www.kaggle.com/felixzhao/productdemandforecasting). There are 1,048,576 instances in the dataset, each having five (5) variables: date, product code, product category, storage location, and order demand. Based on the dataset, it distributes in four (4) warehouses along with 33 product categories and 2160 items.

In this study, the metadata for this study are shown in Table 5. A total of three features were removed as the regression model for this research focuses on the overall demand regardless the variety of product or distribution of warehouse. Thus, only both date and order demand attributes would be retained.

Table 5: Metadata

No	Feature Name	Description	Type	Decision
1	Product_Code	The unique code of product	String	Removed
2	Warehouse	Warehouse to be distributed	String	Removed
3	Product_Category	Product category for each product	String	Removed
4	Date	The date customer needs the product	String	Retained
5	Order_Demand	Quantity of the specified product in single order	Numeric	Retained

3.5 Data Preparation

The data preparation phase is critical for ensuring the dataset's quality by screening data that is frequently missing, inconsistent, or contains outliers. This section presents the data preparation process which includes data cleaning and resampling, followed by data transformation as well as outlier removal in order to prepare for modelling.

3.5.1 Data Cleaning and Resampling

Due to human statistical mistakes and a lack of on-time records, the data collection contains uncounted product demand data. We keep the original data with relatively complete records and discard the instances which contain missing values for order demand modelling later.

Based on the cleansed dataset, there are instances with zero demand which these records may bring challenges to establish the model. Thus, data resampling was carried out which converts the data frequency to another rate. Data downsampling processes aggregate the data and reduce the frequency of dataset. In this study, the dataset was transformed by downsampling the frequency of order demand from daily to weekly basis.

The Figure 4 shows the product demand data before resampling which is in daily basis.

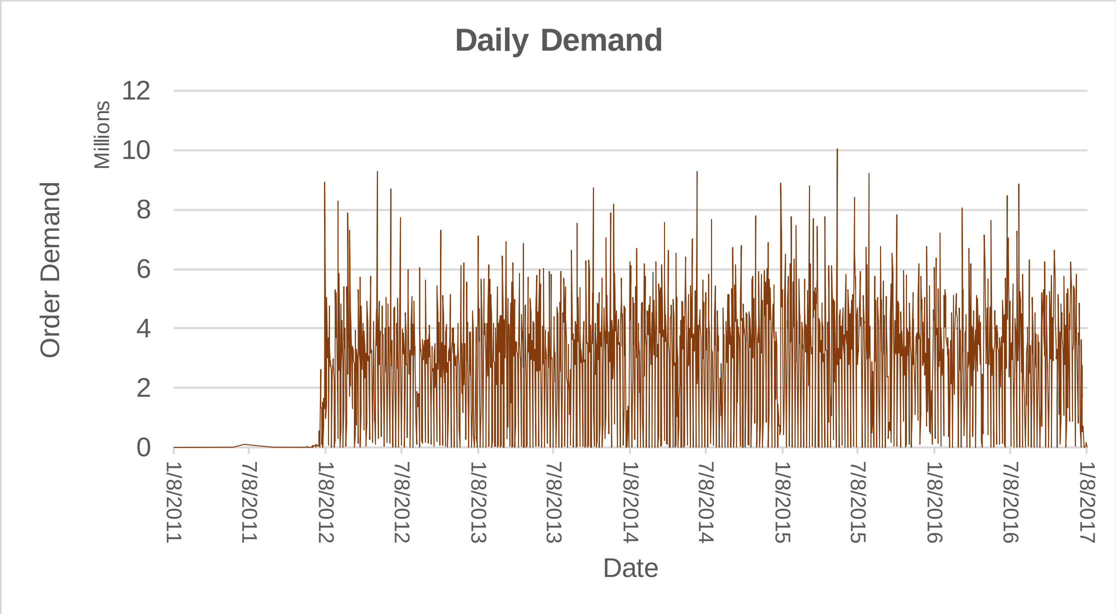

Figure 2: Product demand data before resampling (Daily basis)

The Figure 5 illustrates the product demand data after resampling which is in weekly basis.

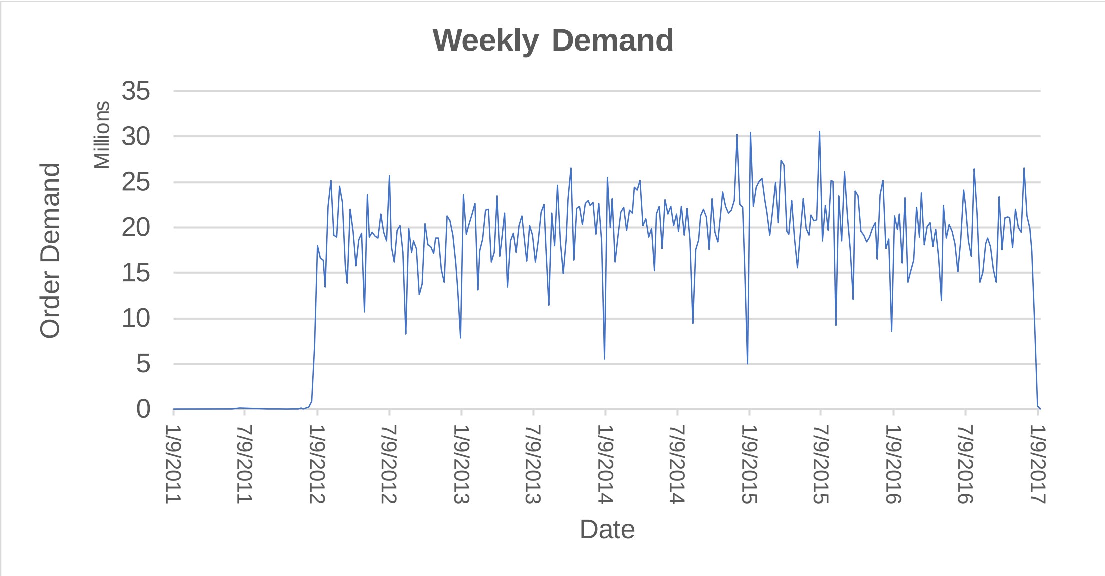

Figure 3: Product demand data in weekly basis

Illustration as shown in Figure 4 was relatively dense whereas Figure 5 depicts the data after downsampling as the weekly demand data is aggregate which greatly reduces the influence of zero values in the dataset.

3.5.2 Data Transformation

Based on Figure 5, it is observed that the weekly order demand contains noises produced by erroneous recording and measurement techniques, as well as impact of sudden changes (Wu & Chen, 2021). These data noises may cause the data difficult to fit the model, leading to reduction of forecasting power of models with high errors and residuals in the prediction. Thus, the noise of the weekly demand data was reduced by using logarithm of 10. The logarithm formula is shown as follow in Equation 1.

$$D_S' = \log(D_S) \tag{1}$$

Where D_s denotes the weekly demand and $D_s{}'$ denotes the logged weekly demand.

The weekly product demand is depicted in Figure 6 for both original and logarithmic data. It is noticed that the logarithmic results transformation has considerably decreased data fluctuation when compared to the original data, as well as noise.

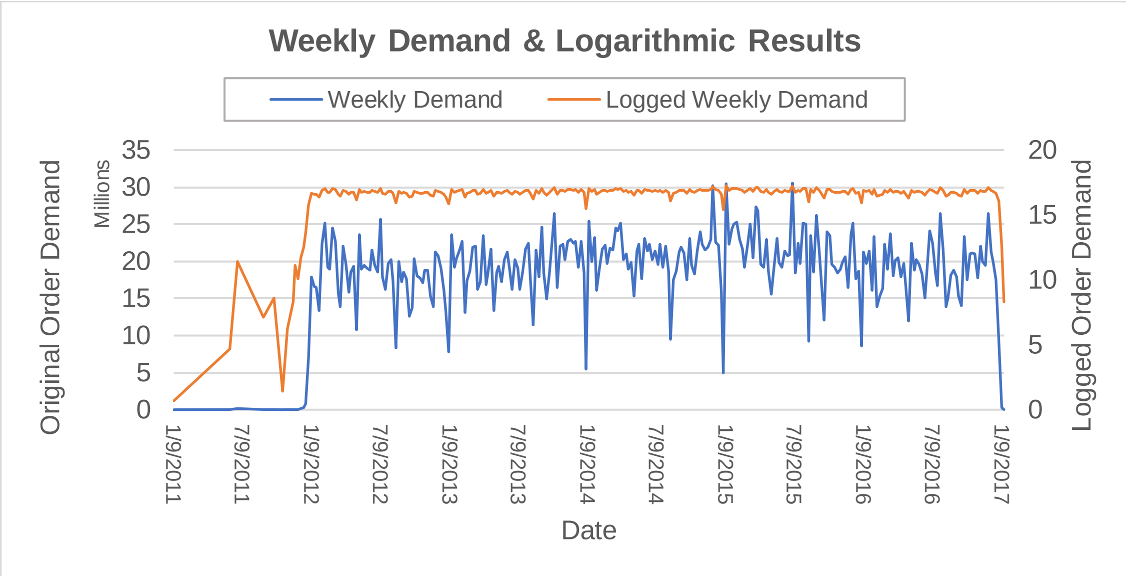

Figure 4: The original and logarithmic results

3.5.3 Outliers Removal

As further look into the data, outliers that may impair the effectiveness of forecast models may be seen in Figure 6 when looking at the logarithmic weekly demand. The logged weekly demand recorded in 2011 and 2017 was lower than the rest of the data, which is suspected to be outliers. Thus, as depicted in Figure 7, a box plot of the logged weekly demand was created to observe the weekly demand via their quartiles to identify the existence of outliers.

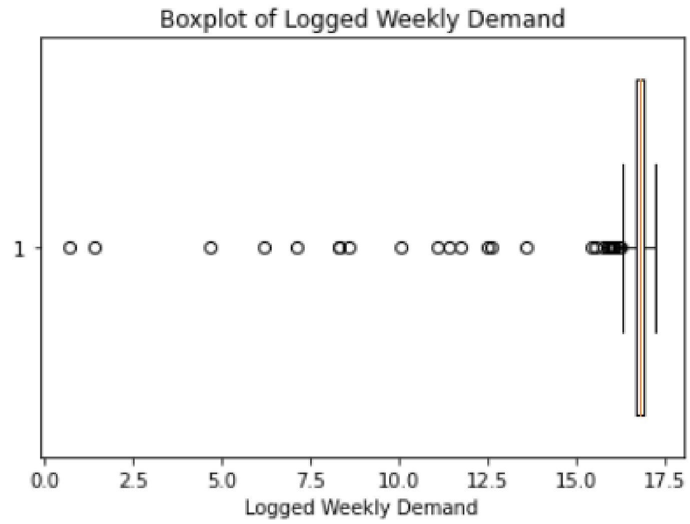

Figure 5: Boxplot of Logged Weekly Demand before outlier removal

The box plot as shown in Figure 7 shows that the data dispersion scattered towards the left which cause data skewness. There are also extreme outliers which that the data points are greatly distanced from the lower whiskers of the box plot. The dispersion of data points and existence of extreme outliers could pose a great impact on the performance of model. Thus, the instances for year 2011 and 2017 is removed, then box plot is constructed again to check on the presence of extreme outliers.

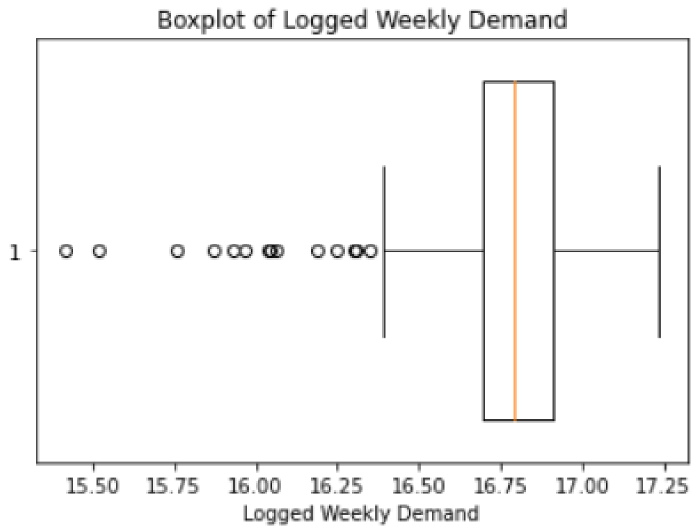

Figure 6: Boxplot of Logged Weekly Demand after outlier removal

After the instance removal for year 2011 and 2017, only logged weekly demand more than 15 were retained. It is observed that the extreme outliers are considerably reduced which is shown in Figure 8 as compared to Figure 7. This increased the data quality by reducing the data skewness as well as the extreme outliers to increase the reliability of the data thus lead to better performance of forecast model. Figure 9 depicts the logged weekly demand after the outlier removal which the noise are greatly reduced.

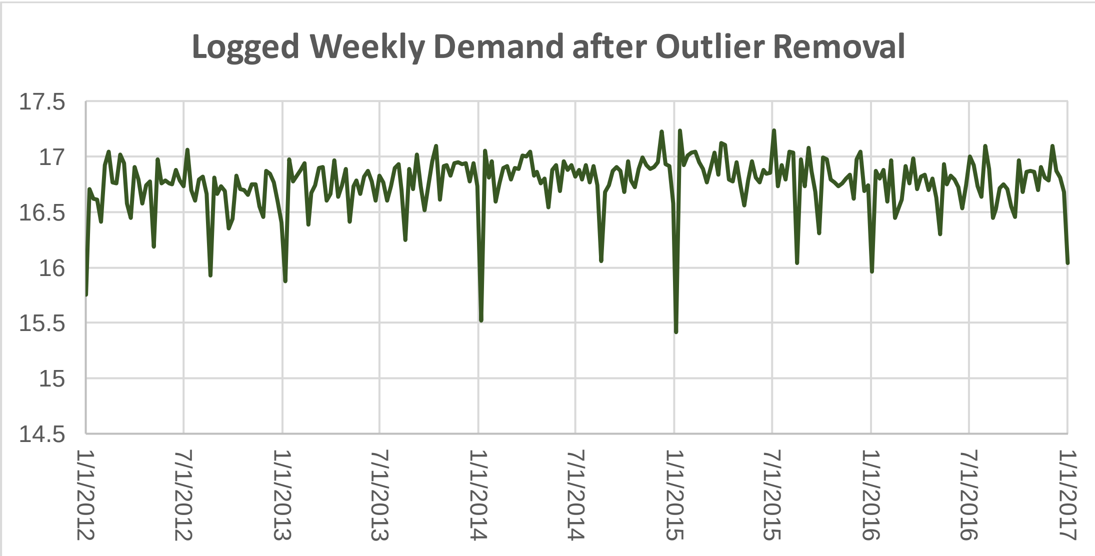

Figure 7: Logged Weekly Demand after outlier removal

Furthermore, the logarithmic weekly demand data would then be transformed into year and week where the regression model would only consider numeric data as input instead of dates input.

To train, validate and test the data, it was separated into three (3) sections in the following order: training set, validation set, and testing set. The training, validation, and testing sets were divided into 70%, 25%, and 5% of the total product demand data. The training set comprises 194 instances, the validation set contains 69 instances, and the test set contains the remaining 14 cases, according to the dataset.

3.6 Machine Learning Models

As mentioned in Chapter 2 (Literature Review), the machine learning models to be used in this study are Support Vector Regression (SVR), nonlinear regressive neural network (NARNN), Gradient Boosting Regression Tree (GBRT) and AdaBoost Regressor (ABR). A brief introduction as well as the parameters to be optimized of each of these models would be discussed in this subsection.

3.6.1 Support Vector Regression (SVR)

SVM is a supervised algorithm which derived from machine learning theory which originally meant to handle classification problems, then extended to address regression and time-series forecasting problems. It divides the input data into two categories, namely upper and lower scale. Then, using three parallel lines to the created hyperplane, an optimum hyperplane would be formed to encompass the two categories of input data. In fact, the more the hyperplanes overlap the margins, the lower the error in generalization would be obtained. Thus, several hyperplanes are created in infinite-dimensional space for regression, with these hyperplanes encompassing the most space around the nearest training data points, causing SVM to be capable of forecasting.

In this study, SVR is adopted. SVR is a subset of SVM which aims to map the original data into a high dimensional feature space to locate the optimal regression hyperplanes, allowing the value of dependent variables to be estimated. According to Priyadarshi, et al. (2019), the standard procedure of SVR is as follows: (1) data input, (2) data training, (3) incorporation of kernel functions, (4) data separation into upper and lower scale, (5) hyperplane(s) building and hyperparameter settings, (6) regression modelling, (7) forecast result.

To convert the linear model into nonlinear model, SVR implements the kernel functions which are commonly used for trend analysis. Kernel function assists SVM mathematical functions in providing a window to alter input data in order to translate a linear equation into a larger number of dimensional spaces. In this study, four (4) kernels would be examined, and their performance would be evaluated, including linear, polynomial, radial basis function (RBF) and sigmoid. Linear kernels are frequently utilized as if input data is separable in low-dimensional space; polynomial kernels are extensively used for image processing; RBF kernels are a general-purpose kernel with high local performance; and sigmoid kernels are a surrogate for neural networks.

The formulas of the selected kernel functions are illustrated in Table 6 below:

Table 6: Table of kernel functions

No	Kernels	Functions	Parameters
1	Linear	$k(x_i, x_j) = (x_i \cdot x_j)$	-
2	Polynomial	$k(x_i, x_j) = (x_i \cdot x_j + c)^d$	c, d
3	Radial Basis Function	$k(x_i, x_j) = \exp\left(\frac{-\lVert x_i - x_j \rVert^2}{2\sigma^2}\right) = \exp\left(-\gamma \lVert x_i - x_j \rVert^2\right)$	σ or γ
4	Sigmoid	$k(x_i, x_j) = \tanh(\gamma(x_i \cdot x_j) + c)$	γ, c

3.6.2 Nonlinear Autoregressive Neural Network (NARNN)

Artificial Neural Network (ANN) is a network that replicates the nervous system of the human brain and performs particular activities. It consists of node layers which includes of an input layer, one or more hidden layers and an output layer. The neural network is resilient as the nodes mimic the synaptic connections of the human brains and interconnect to one node and other in each layer with an associated weight and threshold value (Wu & Chen, 2021).

In this study, nonlinear autoregressive (NAR) neural network model was adopted. NAR model is similar as recurrent neural network (RNN) which it has feedback connections enclosing several layers of the network, in which the predicted value may serve as an input for a new prediction at more advanced points in time. NAR model employs itself as the regression variable and uses a linear combination of many temporal variables to explain the random variable at a certain moment throughout the observation period.

In order to achieve the best forecast performance, an ideal number of hidden layer neurons is required to be discovered. According to Wu and Chen (2021), five (5) hidden layers neurons were selected and justified as the optimal number to provide strong generalization in the product demand forecast as it obtained the lowest RMSE approximately 0.53 as compared with other numbers of hidden layer neurons.

Several training methods, namely Levenberg-Marquardt, Bayesian regularization, and scaled conjugate gradient backpropagation were examined to determine the best training algorithms for NARNN in this study. Levenberg-Marquardt training optimizes the model training by optimizing the weight and bias values. Among the selected training algorithms, it has rapid training speed, but it also stores the most memory. It is most commonly used to solve nonlinear least square problems by minimizing the function of sum of squares. Similar to Levenberg-Marquardt, Bayesian regularization optimizes model training and then discovers the optimum combination of squared errors and weights in order to generate the proper combination that generalizes well. It is better to generalize noisy or small dataset but having drawbacks of longer processing time. For scaled conjugate gradient backpropagation, it optimizes the weight and bias value based on scaled conjugate gradient approach. It is generally used for big problems as it is memory efficient due to the employment of gradient computations as compared with Jacobian calculations used by both Levenberg-Marquardt and Bayesian regularization training algorithms.

3.6.3 Gradient Boosting Regression Tree (GBRT)

Gradient Boosting model is an ensemble machine learning approach which combines multiple simple models or also known as weak learners into a single composite model. As the combination of simple models increased, the accuracy and resilience of the final model improves, making it a better predictor. In this study, decision trees were used as the base learner. Decision trees transform the data into tree representation to solve machine learning problems. Each leaf node of the tree representation signifies a class label, and each branch denotes an attribute.

In this study, the gradient boosting model begins by creating a single leaf regression tree, then the tree was built iteratively and divides the data into nodes or branches into smaller pieces. The data are then distributed into two (2) partitions, using every conceivable split on available predictor based on the given threshold value. Gradient boosting improves the overall prediction performance of the model by training the base model to acquire a base residual, then adding another weak model to the model while maintaining the present tree unchanged. The tree will continue to divide until it meets the user-defined limit, such as maximum tree depth or minimum final node size. The methodology of GBRT is as follow: (1) selection of base learner, (2) add-ins of additive model, (3) define loss function, (4) minimizes loss function.

Numerous parameters such as loss function and learning rate were tuned in this study. Loss functions includes ‘least absolute deviation’, ‘least square’ and ‘Poisson’ were implemented for GBRT to find the minimum loss, then learning rate of 0.1, 0.01 and 0.001 were used to find the optimal parameter. While moving towards the minimum loss function, the learning rate sets the step size at each iteration.

3.6.4 Long Short-Term Memory (LSTM)

LSTM neural network is a functional RNN with short-term and long-term memory which developed by Hochreiter and Schmidhuber in 1997. It has the capabilities to learn long-term dependencies via backpropagation approach while avoid noise by filtering the gradients simultaneously (Priyadarshi, et al., 2019).

As compared to neural networks, LSTM networks feature memory blocks that are linked by layers instead of neurons. A block has elements that make it smarter than a traditional neuron, as well as a memory for recent sequences. Each block has gates that control the state and output of the block. Each of the gates inside a block employs the activation units to regulate whether or not they are activated, making the change of state and addition of information flowing through the block conditional.

In this study, LSTM network would be adopted with a visible input layer, a hidden layer, and an output layer. In term of activation function, sigmoid function would be used for the LSTM blocks. A batch size of one is utilized to train the network for 100 epochs.

Based on Kantasa-ard, et al. (2020), the number of memory blocks could impact the performance of the LSTM model. Thus, different number of memory blocks would be assigned to the hidden layer and tested to identify the optimal number of neurons to be used for LSTM. In this study, number of memory blocks in the hidden layer would be tuned at 5, 10, 15, 20, 25 and 30 and the respective performance would be assessed.

3.6.5 AdaBoost Regressor (ABR)

AdaBoost, also known as adaptive boosting, is a statistical classification meta-algorithm that may be used to improve the performance of many other types of learning algorithms. It started out as an iterative approach that primarily focused on classification problems, but after a number of years of research, it was expanded to include regression (Sun & Gao, 2019).

AdaBoost's learning mechanism is similar to GBRT where it is a method of combining numerous basic algorithms (weak learners) to create a single optimal prediction algorithm. The output of each of these weak learners is blended into a weighted sum that reflects the boosted classifier's final output. This improves the reliability of the boosted model while also improving forecast performance by reducing bias and variance for supervised learning.

In this study, decision tree regressor would be used as the base estimator. Various parameter settings were done to identify the optimal performance for ABR. Loss functions includes 'linear', 'square' and 'exponential' were used for ABR to identify the minimum loss, whereas learning rate at 0.1, 0.01 and 0.001 were used to identify the optimal parameters by determining the least error.

3.7 Performance Evaluation

This paper uses two (2) evaluation metrices – root mean square error (RMSE) and mean absolute percentage error (MAPE) to monitor and assess the model in the context of forecasting the time series data of the weekly product demands.

3.7.1 Root Mean Square Error (RMSE)

The standard deviation of residuals between the data points and the regression line is measured by root mean square error, which shows the concentration of data points around the projected regression line. The formula of root mean square error is shown as in Equation (3).

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{t=1}^{n} (y_t - \overline{y_t})^2} \tag{3}$$

Where the y_t denotes the true score of time series and $\overline{y_t}$ denotes the predicted score of the data point.

3.7.2 Mean Absolute Percentage Error (MAPE)

Mean absolute percentage error measures the accuracy of the forecast system in percentage where it calculates the average absolute residuals of the time series. The formula of mean absolute percentage error is depicted in Equation (4).

$$MAPE(\%) = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \overline{y_t}}{y_t} \right| \times 100\% \tag{4}$$

In fact, Lewis (1982) has come up with a table of evaluation criteria as shown in Table 7 on the forecasting power of regression model based on the measurement of MAPE.

Thus, the interpretation of Lewis (1982) would be considered to justify the forecasting power of the tested models based on Table 7 in this study.

Table 7: Interpretation of forecasting power based on evaluation of MAPE

MAPE	Forecasting power
Less than 10%	Highly accurate forecasting model
Range between 10% to 20%	Good forecasting model
Range between 20% to 50%	Reasonable forecasting model
More than 50%	Inaccurate forecasting model

3.8 Conclusion

To summarize the methodology, this study utilized the Crisp-DM framework as the guideline to forecast the product demand. The demand data used was extracted from a manufacturing company which only date and demand features were considered for the forecast model. Before modelling, the demand data were preprocessed with resampling, logarithmic transformation, and removal of outliers. SVR, NARNN, GBRT, LSTM, and ABR were developed in this work, along with parameter tuning, to assess their efficacy in forecasting product demand. Both RMSE and MAPE assessment metrics were used to assess the models' performance.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

The results and discussion of the study are presented in this chapter. The subsection 4.2 covers the exploratory analysis of the dataset. Parameter optimization for all the tested models: SVR, NARNN, GBRT, LSTM and ABR is covered in subsection 4.3, followed by an assessment of each model's performance depending on its parameter values. The following subsection 4.4 evaluates the tested models based on their optimum parameter settings, and then finds the best model for forecasting product demand.

4.1 Data Exploratory Analysis

An initial exploration analysis has been performed on the historical product demand dataset to study the relationships of variables within the dataset. Figure 2 presents the top five (5) product categories while Figure 3 illustrates the top ten (10) products based on the total order demand of the manufacturing firm.

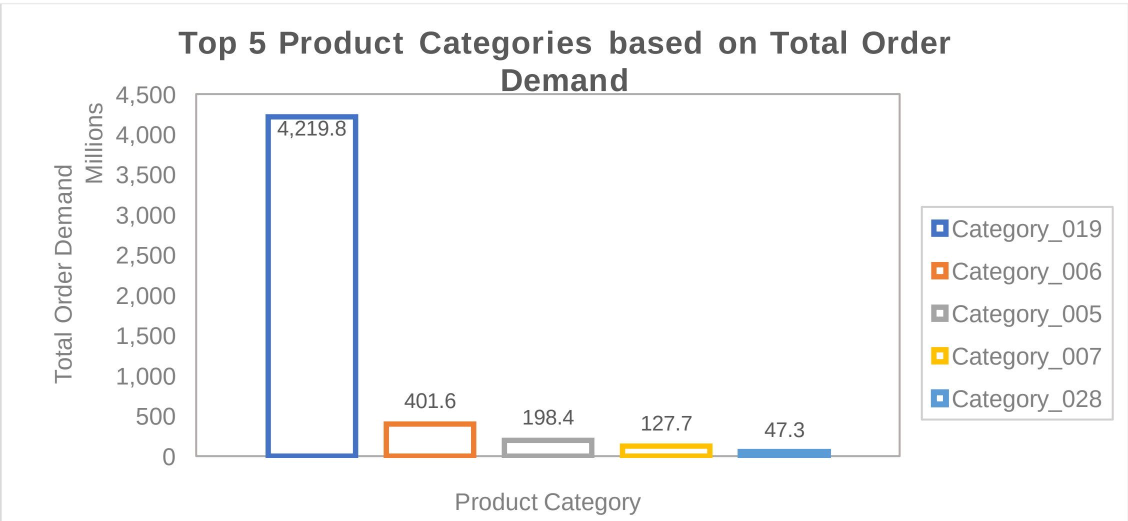

Figure 8: Top five (5) product categories based on Total Order Demand

Based on Figure 2, product category 019 had the highest demand of all product categories, accounting for about 82.2 percent of the overall demand of the manufacturing business.

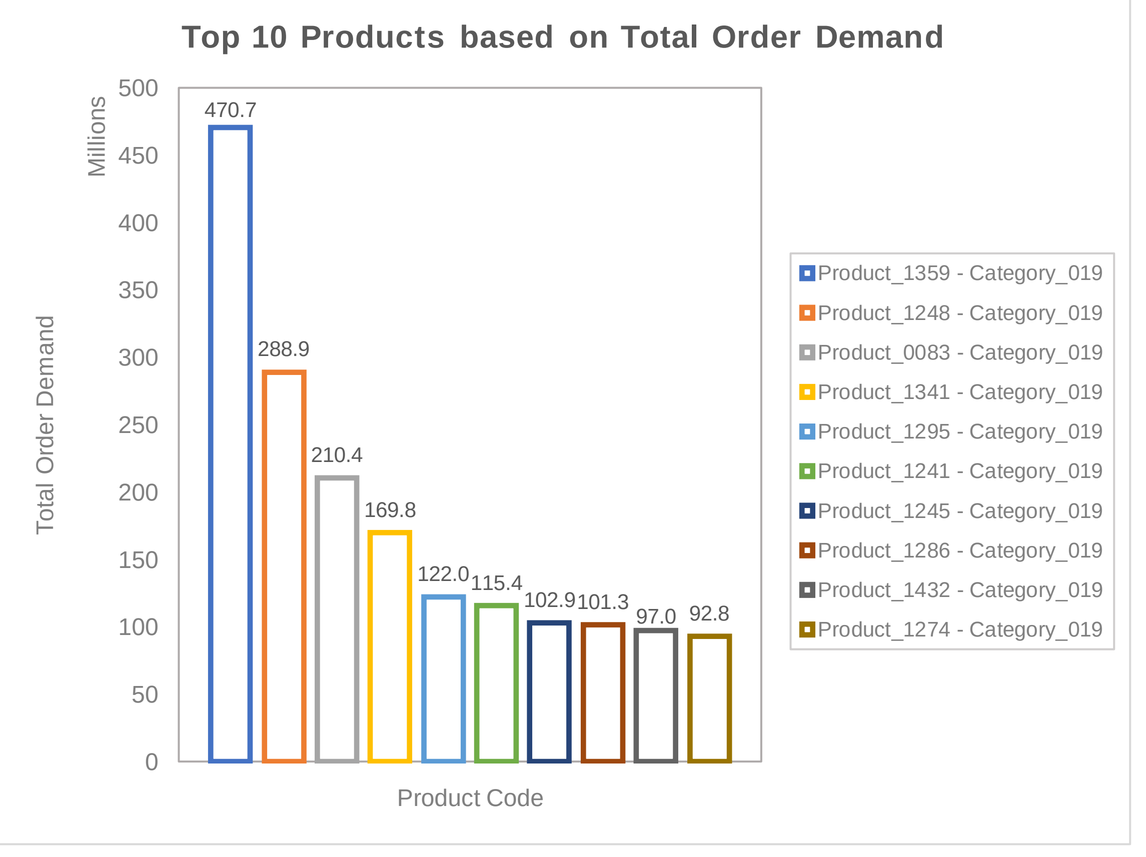

Figure 9: Top ten (10) products based on Total Order Demand

According to the statistics shown in Figure 3, it is observed that product with code of 1359 was the most popular among the products in the dataset. Besides, the top ten products are all from category 019, implying that this is the manufacturing firm's primary goods.

After preprocessing of the demand data, the logarithmic weekly demand data with outliers removed was then described statistically which are shown in Table 8.

Table 8: Statistical Description of Logarithmic Weekly Demand

	Value
Mean	16.7644
Standard Deviation	0.2470
Minimum Value	15.4157
1^{st} Quartile (25%)	16.6959
Median	16.7924
3^{rd} Quartile (75%)	16.9135
Maximum value	17.2353

Based on Table 8, the logarithmic weekly demand has a minimum and maximum value at 15.42 and 17.24 respectively. As the first and third quartile values were revealed to be 16.70 and 16.91, respectively, the skewness of logarithmic demand was altered to be balanced after the outliers were removed, reducing the model's negative impacts.

4.2 Parameter Optimization

This subsection describes the parameter to be optimized for all machine learning models that have been tested, as well as identifying the best parameter to utilize for each machine learning model for parameter assessment. The validation set would be used to find the optimal parameter for each of the machine learning model to forecast product demand via evaluation of RMSE and MAPE.

4.2.1 Support Vector Regression (SVR)

To identify the optimal kernel to be used, implementation of SVR with several kernels namely linear, polynomial, radial basis function (RBF) and sigmoid kernels, have been tested in this study. The resultant performance of the SVR integrated with each kernel is depicted in Figure 10.

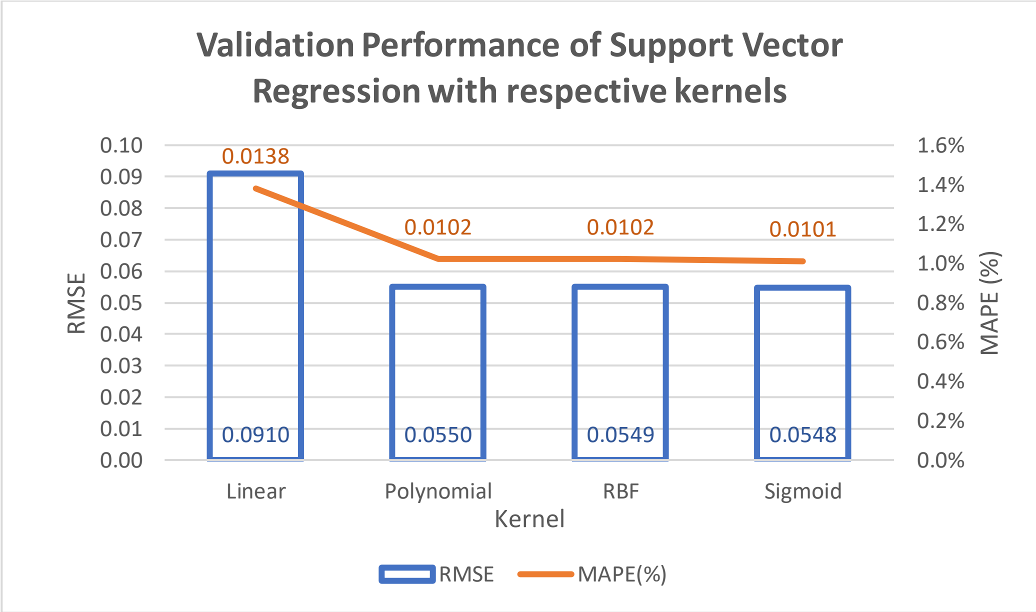

Figure 10: Validation Performance of Support Vector Regression with respective kernels

Based on Figure 10, it is observed that the RMSE has a positive correlation with MAPE for SVR model. SVR with linear kernel has the highest RMSE and MAPE, followed by polynomial, RBF, and sigmoid kernel. Among these tested kernels, SVR incorporated with sigmoid kernel with lowest values of RMSE and MAPE of 0.0548 and 1.01 percent respectively, would be chosen.

4.2.2 Nonlinear Autoregressive Neural Network (NARNN)

For parameter optimization of nonlinear autoregressive neural network (NARNN), different training algorithms, namely Levenberg-Marquardt, Bayesian regularization and scaled conjugate gradient approaches were tested. Only the RMSE assessment measure was presented in the MATLAB software for the NARNN model, hence only RMSE would be examined. Figure 11 shows the performance of NARNN with various training algorithms.

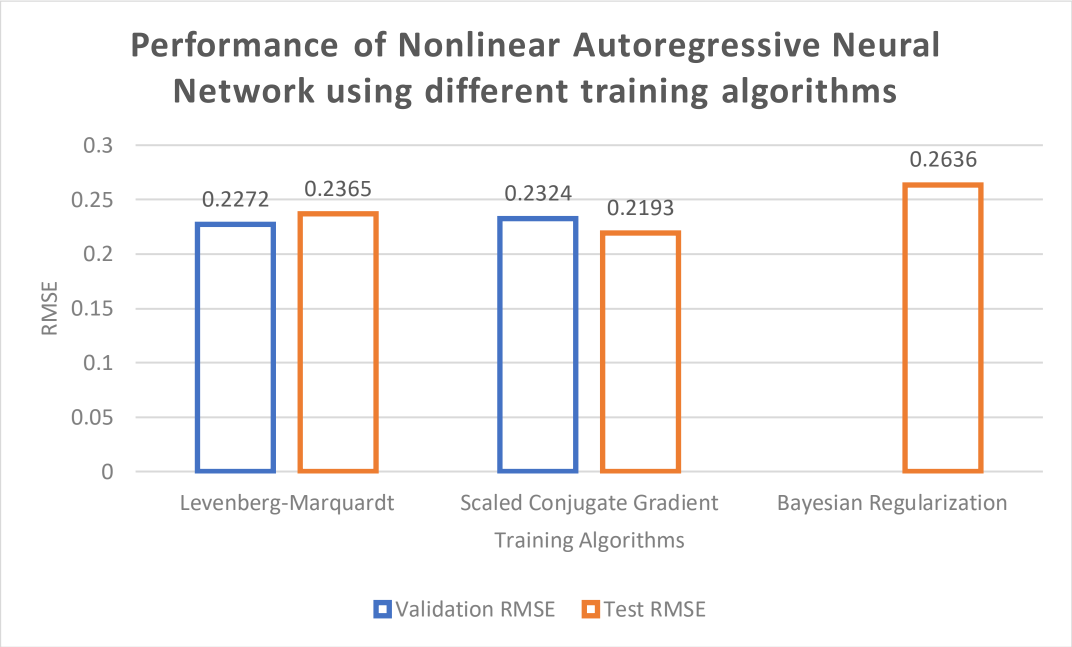

Figure 11: Performance of Nonlinear Autoregressive Neural Network using different training algorithms

According to Figure 11, only NARNN models with Levenberg-Marquardt and scaled conjugate gradient training algorithms have taken the validation dataset into consideration for parameter optimization. Thus, RMSE of the testing set would also be considered in the parameter optimization of NARNN.

By comparing the RMSE of validation set, NARNN model using Levenberg-Marquardt training algorithm achieved the lowest RMSE of 0.2272. On the other hand, scaled conjugate gradient with RMSE of 0.2193 was determined as the optimal training algorithm, whereas Bayesian regularization training approach was deemed as the worst for testing set of NARNN model. In short, Bayesian regularization training algorithm could be omitted as its performance is incompetence as compared to the other training algorithms.

4.2.3 Gradient Boosting Regression Tree (GBRT)

GBRT was developed for demand forecasting in the previous studies where it exhibits performance slightly inferior as compared to SVR (Priyadarshi, 2018). However, no parameter tuning was done in the research where the model could be improvised. Therefore, various parameters – loss function and learning rate, were tuned in this study to identify the optimal parameters for GBRT.

Figure 12 shows the validation set performance of each GBRT model with different parameters.

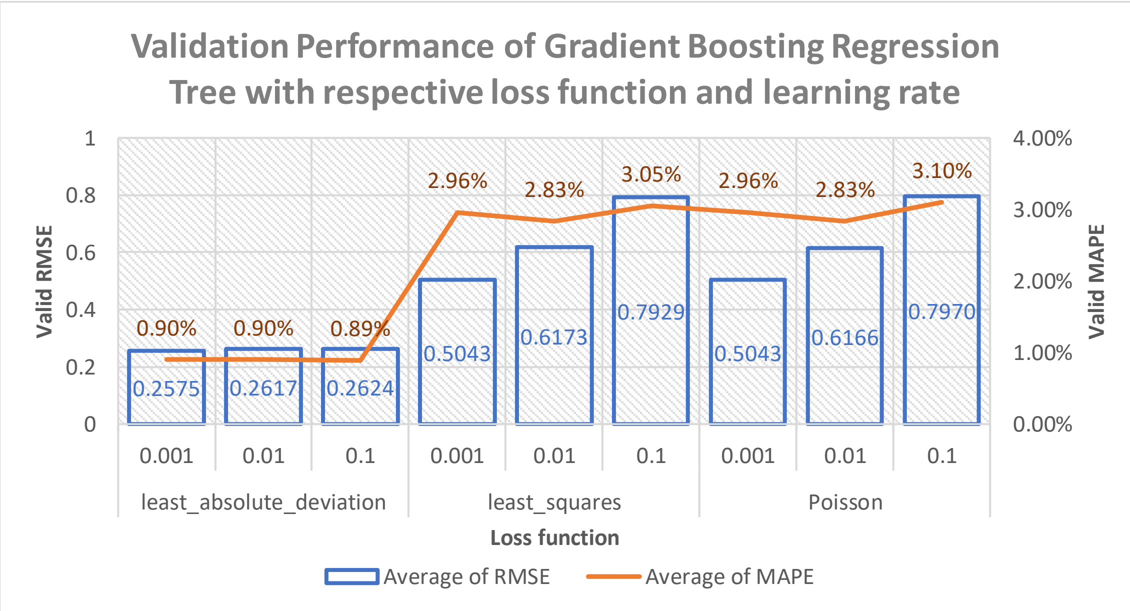

Figure 12: Validation Performance of Gradient Boosting Regression Tree (GBRT) with respective loss and learning rate

Based on Figure 12, it is observed that the RMSE values decreases as the learning rate decreases, whereas ‘least absolute deviation’ was deemed to have the best performance as compared with other loss functions. Thus, ‘least absolute deviation’ loss function and 0.001 learning rate are the optimal parameters for GBRT as it achieved RMSE and MAPE of 0.2575 and 0.9 percent.

4.2.4 Long Short-Term Memory (LSTM)

LSTM was adopted with different number of memory blocks embedded in the hidden layers. The validation performance of the LSTM model with different number of memory blocks are illustrated in Figure 13.

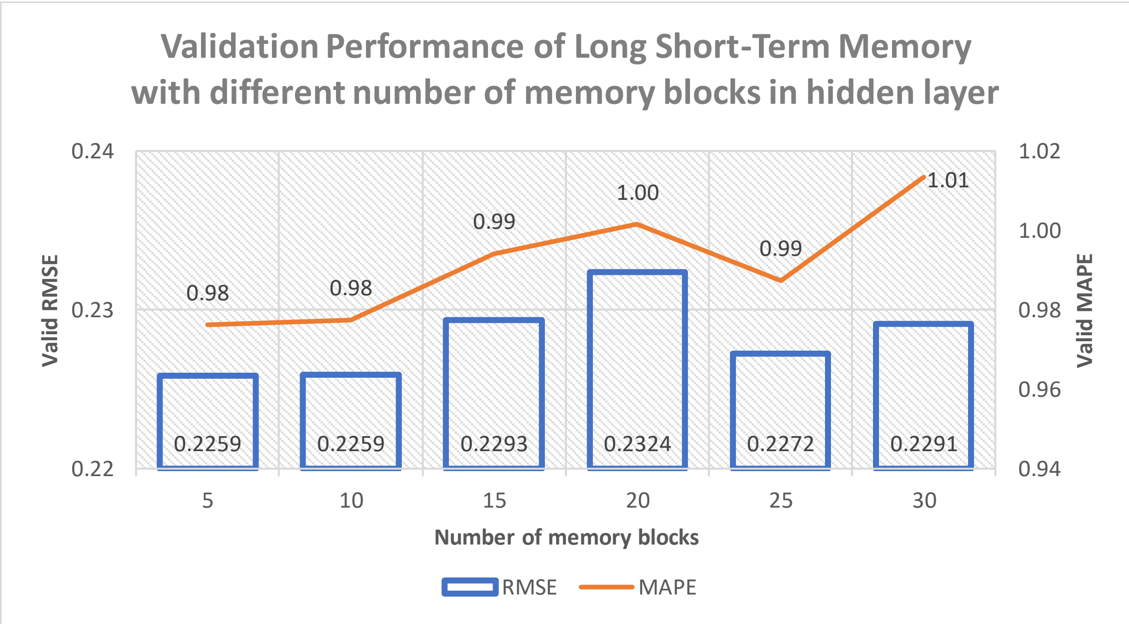

Figure 13: Validation Performance of Long Short-Term Memory with different number of memory blocks in hidden layer

According to Figure 13, it is observed that the LSTM with memory blocks of 5 and 10 gives the lowest RMSE and MAPE of 0.2259 and 0.98 percent respectively. However, the valid RMSE and MAPE of LSTM ranged within 0.2259 to 0.2324 and 0.98 to 1.01 percent respectively which does not show distinct difference. Thus, the test performance of LSTM would be further look into to define the optimal number of memory blocks embedded in the hidden layer where it is shown in Figure 14.

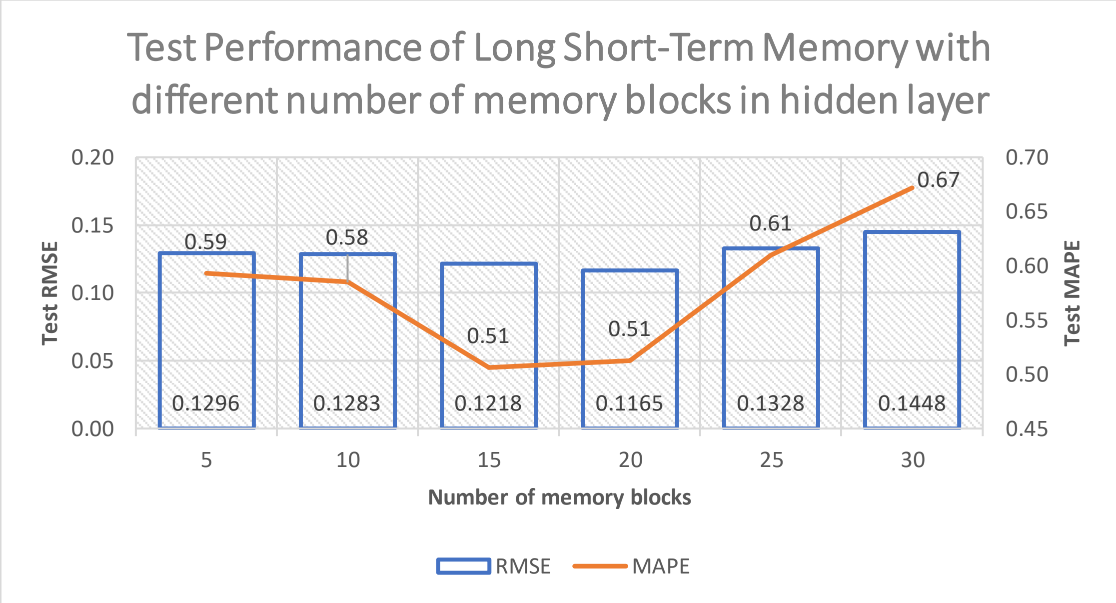

Figure 14: Test Performance of Long Short-Term Memory with different number of memory blocks in hidden layer

As shown in Figure 14, it is found that LSTM exhibits the lowest test RMSE and MAPE of 0.1165 and 0.51 percent respectively at memory blocks of 20. Thus, LSTM with memory blocks of 20 would be selected as the best parameter for model assessment.

4.2.5 AdaBoost Regressor (ABR)

Based on the past literature, AdaBoost Regressor (ABR) was not being tested for product demand forecast. Thus, ABR would be implemented and assessed in this study. In order to identify the optimal parameters for ABR, two (2) parameters – loss function and learning rate has been tuned where the validation performance are shown in Figure 15.

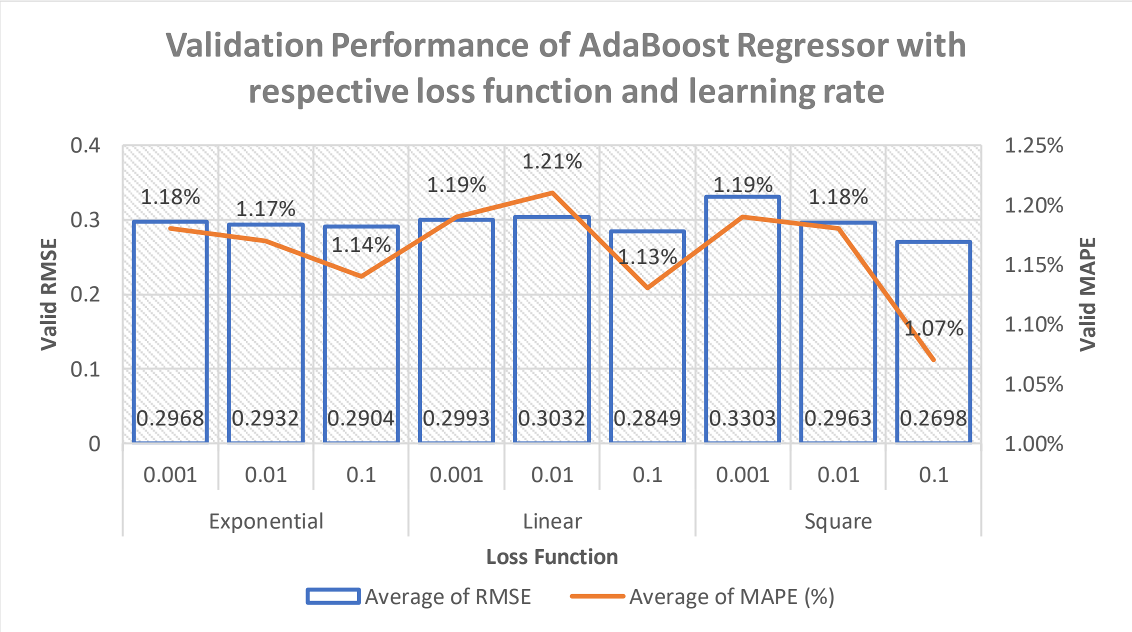

Figure 15: Validation Performance of AdaBoost Regressor (ABR) with respective loss function and learning rate

According to illustration of Figure 13, it is observed that both RMSE and MAPE exhibits lowest error for learning rate at 0.1. As can be observed, ABR with square function loss at 0.1 learning rate was judged to be the best, with RMSE and MAPE of 0.2698 and 1.07 percent, respectively.

4.3 Performance for Product Demand Forecast

In this section, the outcomes of all tested models are discussed. Table 9 shows the detailed summary of tested model performance for both validation and testing set.

Table 9: Summary of Model Performance for Product Demand Forecast

<table>
<tr><th rowspan="3">Model</th><th rowspan="3" colspan="2">Parameters tuned</th><th colspan="4">Performance</th></tr>
<tr><th colspan="2">Validation set</th><th colspan="2">Testing set</th></tr>
<tr><th>RMSE</th><th>MAPE (%)</th><th>RMSE</th><th>MAPE (%)</th></tr>
<tr><td rowspan="5">SVR</td><td colspan="2">Kernel</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Linear</td><td>0.0910</td><td>1.38</td><td>0.1067</td><td>1.39</td></tr>
<tr><td colspan="2">Polynomial</td><td>0.0550</td><td>1.02</td><td>0.0565</td><td>0.83</td></tr>
<tr><td colspan="2">Radial basis function</td><td>0.0549</td><td>1.02</td><td>0.0567</td><td>0.83</td></tr>
<tr><td colspan="2">Sigmoid</td><td>0.0548</td><td>1.01</td><td>0.0568</td><td>0.83</td></tr>
<tr><td rowspan="4">NARNN</td><td colspan="2">Training Algorithm</td><td colspan="4"></td></tr>
<tr><td colspan="2">Levenberg-Marquardt</td><td>0.2272</td><td>-</td><td>0.2365</td><td>-</td></tr>
<tr><td colspan="2">Bayesian Regularization</td><td>-</td><td>-</td><td>0.2636</td><td>-</td></tr>
<tr><td colspan="2">Scaled Conjugate Gradient</td><td>0.2324</td><td>-</td><td>0.2193</td><td>-</td></tr>
<tr><td rowspan="10">GBRT</td><td>Loss Function</td><td>Learning rate</td><td colspan="4"></td></tr>
<tr><td rowspan="3">Least Square</td><td>0.1</td><td>0.2305</td><td>1.06</td><td>0.2602</td><td>0.94</td></tr>
<tr><td>0.01</td><td>0.2345</td><td>1.04</td><td>0.2410</td><td>0.84</td></tr>
<tr><td>0.001</td><td>0.2288</td><td>0.99</td><td>0.2348</td><td>0.88</td></tr>
<tr><td rowspan="3">Least Absolute Deviation</td><td>0.1</td><td>0.2513</td><td>1.13</td><td>0.2748</td><td>1.03</td></tr>
<tr><td>0.01</td><td>0.2439</td><td>1.06</td><td>0.2495</td><td>0.86</td></tr>
<tr><td>0.001</td><td>0.2325</td><td>1.00</td><td>0.2373</td><td>0.84</td></tr>
<tr><td rowspan="3">Poisson</td><td>0.1</td><td>0.2307</td><td>1.06</td><td>0.2605</td><td>0.94</td></tr>
<tr><td>0.01</td><td>0.2344</td><td>1.04</td><td>0.2409</td><td>0.84</td></tr>
<tr><td>0.001</td><td>0.2288</td><td>0.99</td><td>0.2348</td><td>0.88</td></tr>
<tr><td rowspan="7">LSTM</td><td colspan="2">Number of memory blocks</td><td colspan="4"></td></tr>
<tr><td colspan="2">5</td><td>0.2259</td><td>0.98</td><td>0.1296</td><td>0.59</td></tr>
<tr><td colspan="2">10</td><td>0.2259</td><td>0.98</td><td>0.1283</td><td>0.58</td></tr>
<tr><td colspan="2">15</td><td>0.2293</td><td>0.99</td><td>0.1218</td><td>0.51</td></tr>
<tr><td colspan="2">20</td><td>0.2324</td><td>1.00</td><td>0.1165</td><td>0.51</td></tr>
<tr><td colspan="2">25</td><td>0.2272</td><td>0.98</td><td>0.1328</td><td>0.61</td></tr>
<tr><td colspan="2">30</td><td>0.2291</td><td>1.01</td><td>0.1448</td><td>0.67</td></tr>
<tr><td rowspan="10">ABR</td><td>Loss Function</td><td>Learning rate</td><td colspan="4"></td></tr>
<tr><td rowspan="3">Linear</td><td>0.1</td><td>0.2849</td><td>1.13</td><td>0.2191</td><td>0.79</td></tr>
<tr><td>0.01</td><td>0.3032</td><td>1.21</td><td>0.2182</td><td>0.79</td></tr>
<tr><td>0.001</td><td>0.2993</td><td>1.19</td><td>0.2375</td><td>0.84</td></tr>
<tr><td rowspan="3">Square</td><td>0.1</td><td>0.2698</td><td>1.07</td><td>0.2080</td><td>0.77</td></tr>
<tr><td>0.01</td><td>0.2963</td><td>1.18</td><td>0.2411</td><td>0.85</td></tr>
<tr><td>0.001</td><td>0.3303</td><td>1.19</td><td>0.2344</td><td>0.84</td></tr>
<tr><td rowspan="3">Exponential</td><td>0.1</td><td>0.2904</td><td>1.14</td><td>0.2184</td><td>0.79</td></tr>
<tr><td>0.01</td><td>0.2932</td><td>1.17</td><td>0.2387</td><td>0.83</td></tr>
<tr><td>0.001</td><td>0.2968</td><td>1.18</td><td>0.2316</td><td>0.83</td></tr>
</table>

For SVR, the test RMSE values for linear, polynomial, radial basis function, and sigmoid kernels are 0.1067, 0.0565, 0.0567, and 0.0568, respectively, as shown in Table 9. The test MAPE values for linear, polynomial, radial basis function and sigmoid kernels are 1.39%, 0.83%, 0.83%, and 0.83%. SVR using polynomial kernel has the lowest RMSE for the testing set, however SVR incorporate with sigmoid kernel has the lowest RMSE for the validation set. Despite validation RMSE of SVR incorporated with sigmoid kernel showed the lowest value, however SVR with polynomial kernel would be chosen as the final model as its test performance are the best among all tested kernels.

The performance results as shown in Table 9 for NARNN model exhibits test RMSE values of 0.2365, 0.2636, and 0.2193 for NARNN incorporate with Levenberg-Marquardt, Bayesian regularization and scaled conjugate gradient training approaches. The lowest RMSE for the testing set is NARNN with scaled conjugate gradient training method, whereas the lowest RMSE for the validation set is NARNN with Levenberg-Marquardt training algorithm. Hence, NARNN with scaled conjugate gradient training algorithm would be chosen.

For GBRT, Table 9 depicts the performance results for various parameter settings. For GBRT incorporate with least squares loss function, it exhibits test RMSE and MAPE values of 0.2602, 0.94%, 0.2410, 0.84%, 0.2348 and 0.88% at learning rate of 0.1, 0.01 and 0.001 respectively. For GBRT incorporate with least absolute deviation loss function, it exhibits test RMSE and MAPE values of 0.2748, 1.03%, 0.2495, 0.86%, 0.2373 and 0.84% at learning rate of 0.1, 0.01 and 0.001 respectively. For GBRT incorporate with Poisson loss function, it exhibits test RMSE and MAPE values of 0.2605, 0.94%, 0.2409, 0.84%, 0.2348 and 0.88% at learning rate of 0.1, 0.01 and 0.001 respectively. GBRT with function loss of least square and Poisson have the lowest RMSE for both validation and testing set at learning rate of 0.001.

For LSTM, LSTM network with different number of memory blocks embedded in hidden layers were adopted where the respective performances are shown in Table 9. The test RMSE values are 0.1296, 0.1283, 0.1218, 0.1165, 0.1328 and 0.1448; whereas the test MAPE values 0.59%, 0.58%, 0.51%, 0.51%, 0.61% and 0.67% respectively for LSTM with memory blocks of 5, 10, 15, 20, 25 and 30 in the hidden layer. Despite LSTM with memory blocks of 5 and 10 shows the lowest RMSE for validation set, LSTM with memory blocks of 20 would be chosen as the final model as it exhibits the lowest RMSE for testing set.

Various parameter settings were tuned and tested for ABR where the validation and test performance are shown in Table 9. For incorporation of linear function loss for ABR, it exhibits test RMSE and MAPE values of 0.2191, 0.79%, 0.2182, 0.79%, 0.2375 and 0.84%, for learning rate of 0.1, 0.01 and 0.001 respectively. For incorporation of squared function loss for ABR, it exhibits test RMSE and MAPE values of 0.2080, 0.77%, 0.2411, 0.85%, 0.2344 and 0.84%, for learning rate of 0.1, 0.01 and 0.001 respectively. For incorporation of exponential function loss for ABR, it exhibits test RMSE and MAPE values of 0.2184, 0.79%, 0.2387, 0.83%, 0.2316 and 0.83%, for learning rate of 0.1, 0.01 and 0.001 respectively. ABR with squared function loss has the lowest test RMSE and MAPE at learning rate of 0.1.

In overall, RMSE and MAPE of testing set for all models is lower as compared to validation set in this study. Besides, all models exhibit MAPE lower than 10 percent which is considered as high accurate forecasting model. Amongst the model, SVR with polynomial kernel would be chosen as this model achieved the lowest RMSE and MAPE values of 0.0565 and 0.83 percent respectively.

4.4 Benchmarking with Previous Studies on Product Demand Forecast

4.4.1 Dataset

As mentioned in Chapter 2 (Literature Review), only research by Wu and Chen (2021) uses an open-source dataset for demand forecasting. Thus, similar demand data were used in this study. To further analyses the dataset, visualization of daily demand and weekly demand of dataset were depicted in Chapter 3.4 (Methodology – Data Preprocessing).

As compared with the data used to forecast product demand by Wu and Chen (2021), it is suspected that the dataset used by Wu and Chen (2021) may focus on a specified product or product category on demand forecasting. Besides, the original data retrieved from Kaggle platform was given from 2011 to 2017. However, the visualization provided by Wu and Chen (2021) only shows only demand data ranged from year 2012 till 2017 where the demand data for year 2011 was neglected. The author was contacted to provide clarification on data preparation, but no response was received. Hence, there is no information to be used as guideline for data preprocessing. In this study, only demand data ranges from year 2012 to end of 2016 were considered as there is extreme outliers and zero values which could greatly impact the model performance.

4.4.2 Performance Results of Support Vector Regression from Past Studies for Product Demand Forecasting

For SVR, various past studies have used SVR for product demand forecasting to determine its performance results in term of RMSE and MAPE from the past studies which has shown in Table 9.

Table 10: Performance results of SVR from past studies for product demand forecasting

No	Source	Performance	
		RMSE	MAPE (%)
1	This study	0.0565	0.83
2	Priyadarshi, et al. (2019)	6.28	4.88
3	Loureiro, Miguéis & da Silva (2018)	0.923	6.90
4	Kaya, & Turkyilmaz (2018)	0.67	21.0 (A-MAPE)
5	Villegas, et al. (2018)	0.099 – 0.471	-

Based on Table 10, it lists out the performance of SVR from the past studies which involved in product demand forecasting. Based on previous studies, Priyadarshi, et al. (2019) has developed SVR incorporate with RBF kernel which that the lowest RMSE and MAPE obtained are 6.28 and 4.88 percent respectively. Loureiro, Miguéis and da Silva (2018) has tried on two different kernels – linear and radial basis function which their SVR exhibits the best performance of RMSE and MAPE value of 0.923 and 6.90 percent respectively. Kaya and Turkyilmaz's (2018) SVR exhibits RMSE of 0.923 and A-MAPE of 21 percent, however its parameter setting for SVR was not informed in their paper. SVR developed by Villegas, et al. (2018) uses gaussian kernels, which also known as radial basis function has achieved RMSE as low as 0.099.

In this study, SVR with four (4) kernels were developed where their performance results were shown in Table 8. It is found that SVR incorporate with polynomial kernel achieved the lowest RMSE and MAPE as compared with SVR with other kernels which the past studies only focused on SVR with linear and radial basis function. To compare the SVR performance, SVR incorporated with polynomial kernel overtakes the performance in terms of both RMSE and MAPE as compared with SVR from the past studies for demand forecasting.

4.4.3 Performance Results of Nonlinear Autoregressive Neural Network from Past Studies for Product Demand Forecasting

To assess the performance of NARNN model, performance results from past studies were used to compare. Wu and Chen (2021) have performed NARNN to forecast product demand where the performance results are shown in Table 11.

Table 11: Performance results of NARNN from past studies for product demand forecasting

No	Source	Performance	
		RMSE	MAPE (%)
1	This study	0.2193	-
2	Wu & Chen (2021)	0.5865	-

Based on Table 11, Wu and Chen (2021) has developed NARNN with various hidden layer neurons to identify the optimal parameter for better performance. Five (5) hidden layer neurons were determined to be the optimal number which gives the best performance results where their NARNN model achieved overall RMSE of 0.5865. However, no information on training algorithms were depict in the research done by Wu and Chen (2021). In this study, five (5) hidden layer neurons were used to develop NARNN model where it extended to test on different training algorithms to identify the optimal training algorithms which could achieve the lowest error. Based on Table 9, NARNN model with five hidden layer neurons and utilized scaled conjugate gradient training algorithms exhibits the lowest test RMSE of 0.2193. The findings for NARNN in this study was found to have better performance as it shows lower RMSE as compared to NARNN model done by Wu and Chen (2021) as same dataset were used.

4.4.4 Performance Results of Gradient Boosting Regression Tree from Past Studies for Product Demand Forecasting

For GBRT, Priyadarshi, et al. (2019) have developed GBRT to forecast product demand. The performance results of GBRT from previous studies are shown in Table 12.

Table 12: Performance results of GBRT from past studies for demand forecasting

No	Source	Performance	
		RMSE	MAPE (%)
1	This study	0.2348	0.88
2	Priyadarshi, et al. (2019)	6.57	4.62

According to research done by Priyadarshi, et al. (2019), they have performed GBRT model which uses least square error loss function where their model achieved RMSE and MAPE as low as 6.57 and 4.62 percent respectively. To further investigate the potential of GBRT for product demand forecasting, GBRT with different loss functions were performed at different learning rate to identify the optimal parameters for best performance results. As shown in Table 7, it is observed that both least square error and Poisson loss functions at learning rate of 0.001 were the optimal parameter settings for GBRT as it exhibits the lowest RMSE and MAPE values of 0.2348 and 0.88 percent respectively.

4.4.5 Performance Results of Long Short-Term Memory from Past Studies for Product Demand Forecasting

LSTM model was used in few past studies to predict product demand. Table 13 shows the performance results of LSTM from previous research for product demand forecasting.

Table 13: Performance results of LSTM from past studies for demand forecasting

No	Source	Performance	
		RMSE	MAPE (%)
1	This study	0.2348	0.88
2	Kantasa-ard, et al. (2020)	149.24	10.97
3	Priyadarshi, et al. (2019)	6.57	4.62

Kantasa-ard, et al., (2020) adopted LSTM in product demand forecast where their LSTM model exhibits RMSE and MAPE of 149.24 and 10.97 percent respectively. Besides, LSTM which developed by Priyadarshi, et al., (2019) exhibits RMSE and MAPE of 6.57 and 4.62 percent respectively. Based on the performance as shown in Table 7, LSTM with memory block of 20 exhibits the lowest RMSE and MAPE of 0.2348 and 0.88 percent respectively as compared to LSTM with other parameter settings.

4.4.6 Performance Results of AdaBoost Regressor from Past Studies for Product Demand Forecasting

There has never been any research that used ABR to estimate product demand, however, there have been several studies that used ABR to forecast energy demand which was mentioned in Chapter 2.1 (Product Demand Forecast Techniques and Performance). In this study, ABR was performed with various parameter settings – linear, square and exponential loss functions as well as learning rate at 0.1, 0.01 and 0.001. Based on the performance results and shown in Table 7, ABR with square loss function and learning rate of 0.1 gives the lowest RMSE and MAPE at 0.2080 and 0.77 percent for product demand forecasting.

4.4.7 Summary of Performance Results for All Tested Models

In this section, summary of performance for all tested models in this study would be discussed. Validation performances were used for parameter testing whereas the test performance would be used to assess the generalization error where the lowest error in test performance would be chosen as the final model. Table 14 lists the test performance results for SVR, NARNN, GBRT, LSTM and ABR with optimal parameter settings which have been conducted in this study for demand forecasting.

Table 14: Summary of performance results in this study for product demand forecast

Model	Performance	
	RMSE	MAPE (%)
SVR	0.0565	0.83
NARNN	0.2193	-
GBRT	0.2348	0.88
LSTM	0.1165	0.51
ABR	0.2080	0.77

Based on Table 14, LSTM showed the lowest MAPE at 0.51 percent. However, all models depict MAPE lower than 10 percent. Thus, it can be deemed that all models exhibit highly accurate forecasting power according to theory given by Lewis (1982) illustrated in Table 7. On the other hand, SVR showed the best performance in term of RMSE as it exhibits RMSE of 0.0565 which are the lowest among the tested models. In overall, SVR is deemed as the best model for product demand forecasting in this study.

4.5 Conclusion

In this chapter, numerous ML models have been conducted for product demand forecast. SVR models were developed with different kernels, namely linear, polynomial, radial basis function and sigmoid; NARNN were trained with three training algorithms, namely Levenberg-Marquardt, Bayesian regularization and scaled conjugate gradient; GBRT models were developed with learning rate ranged from 0.001 to 0.1 and loss function of least square, least absolute deviation and poisson; LSTM were done with memory blocks ranged from 5 to 30; whereas ABR models were constructed with learning rate ranged from 0.001 to 0.1. Amongst the tested model, SVR incorporated with polynomial kernel would be chosen as it achieved the lowest RMSE value of 0.0565.

CHAPTER 5: CONCLUSION

5.1 Introduction

In this chapter, the summary for the findings in this research would be discussed. Section 5.2 states the contribution of this study; section 5.3 discusses the limitation of the study, whereas section 5.4 discusses the recommendation for future work for product demand forecasting.

5.2 Contribution

In this study, regression models – SVR, NARNN, GBRT, LSTM and ABR were developed to predict product demands. These models with various parameter settings to identify the optimal parameter for performance assessment of the models. SVR was developed with different kernels, namely linear, polynomial, radial basis function and sigmoid, SVR using polynomial kernel was chosen as it exhibits the lowest RMSE and MAPE; NARNN was developed with three training algorithms – Levenberg-Marquardt, Bayesian regularization and scaled conjugate gradient, where scaled conjugate gradient training approach gives the lowest RMSE and MAPE; LSTM was adopted with different number of memory blocks which are 5, 10, 15, 20, 25 and 30 to be embedded in its hidden layer; whereas GBRT and ABR were developed with different loss functions (least square, least absolute deviation and Poisson for GBRT; linear, square and exponential for ABR) at learning rate of 0.1, 0.01 and 0.001. Based on the performance as shown in Table 7, both least square and Poisson loss functions were chosen at learning rate of 0.001 for GBRT, whereas square loss function at learning rate of 0.1 was selected for ABR.

Amongst the tested models, SVR model was found to be the best model for product demand forecast in this study which it achieved the lowest RMSE which are highlighted in Table 14. SVR algorithm which uses polynomial kernel are determined to meet the needs/demand patterns. Thus, SVR model is suggested for use since they have shown to be effective in dealing with non-linear data.

5.3 Limitations

In this study, demand data was taken from 2012 to the end of 2016 where most of the outliers are removed to decrease the skewness of data to reduce the impact on inaccurate forecasting on product demand. The demand data was resampled into weekly data where the daily data might be omitted during the data collection process. Hence, the incomplete dataset might affect the performance of the model.

Besides, this study only considered the total demand where different data levels were neglected. Hence, further discovery on the model performance on different data granularity such as product categories or even product to understand the correlation of demand over time, and aid in better decision making.

5.4 Recommendations

In future work, there is still much room for improvement on product demand forecasting for manufacturing industry. Various aspects could be further look into:

- **Data collection**

  Normalization of data collection process by recording the data with high quality and more granularity segment to make the forecasting model more resilience.

- **Multivariate data input**

  Different data features such as calendar and location data could be added into consideration for data input, hence understand the effect of seasonality, geography needs and so forth.

- **Parameter tuning of models**

  More parameter settings could be done to SVR to discover the model's potential in demand forecasting despite only different kernels were incorporated.

- **Granularity of forecasting**

  Modelling on different data levels required to be tested and optimized to have a better understanding on the distribution of production.

CHAPTER 6: REFERENCES

Afifi, A. A. (2020). Demand Forecasting of Short Life Cycle Products Using Data Mining Techniques. IFIP Advances in Information and Communication Technology, 151–162. https://doi.org/10.1007/978-3-030-49161-1_14

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019). Sales Demand Forecast in E-commerce Using a Long Short-Term Memory Neural Network Methodology. Neural Information Processing, 462–474. https://doi.org/10.1007/978-3-030-36718-3_39

Bousqaoui, H., Achchab, S., & Tikito, K. (2017). Machine learning applications in supply chains: An emphasis on neural network applications. 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), 1–7. https://doi.org/10.1109/cloudtech.2017.8284722

Cai, J., Xu, K., Zhu, Y., Hu, F., & Li, L. (2020). Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. Applied Energy, 262, 114566. https://doi.org/10.1016/j.apenergy.2020.114566

Chen, J. M., & Baker, A. (2020). Forecasting Mortgage Demand: An Application of Traditional Methods, Machine Learning, and Neural Networks. SSRN Electronic Journal. Published. https://doi.org/10.2139/ssrn.3656924

Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2016). Analytics for an Online Retailer: Demand Forecasting and Price Optimization. Manufacturing & Service Operations Management, 18(1), 69–88. https://doi.org/10.1287/msom.2015.0561

Gong, M., Bai, Y., Qin, J., Wang, J., Yang, P., & Wang, S. (2020). Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin. Journal of Building Engineering, 27, 100950. https://doi.org/10.1016/j.jobe.2019.100950

Guegan, D., & Iacopini, M. (2018). Nonparametric Forecasting of Multivariate Probability Density Functions. SSRN Electronic Journal. Published. https://doi.org/10.2139/ssrn.3192342

Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. International Journal of Forecasting, 36(4), 1420–1438. https://doi.org/10.1016/j.ijforecast.2020.02.005

Kantasa-ard, A., Nouiri, M., Bekrar, A., Ait El Cadi, A., & Sallez, Y. (2020). Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand. International Journal of Production Research, 1–25. https://doi.org/10.1080/00207543.2020.1844332

Kaya, G. O., & Turkyilmaz, A. (2018). Intermittent demand forecasting using data mining techniques. Applied Computer Science, 14(2). http://dx.doi.org/10.23743/acs-2018-11

Khan, M. A., Saqib, S., Alyas, T., Ur Rehman, A., Saeed, Y., Zeb, A., Zareei, M., & Mohamed, E. M. (2020). Effective Demand Forecasting Model Using Business Intelligence Empowered With Machine Learning. IEEE Access, 8, 116013–116023. https://doi.org/10.1109/access.2020.3003790

Kilimci, Z. H., Akyuz, A. O., Uysal, M., Akyokus, S., Uysal, M. O., Atak Bulbul, B., & Ekmis, M. A. (2019). An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain. Complexity, 2019, 1–15. https://doi.org/10.1155/2019/9067367

Lee, M., Mu, X., & Zhang, Y. (2020). A Machine Learning Approach to Improving Forecasting Accuracy of Hotel Demand: A Comparative Analysis of Neural Networks and Traditional Models. Issues In Information Systems, 21(1). https://doi.org/10.48009/1_iis_2020_12-21

Lewis, C. D. (1982). Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. Butterworth-Heinemann.

Li, X., Li, W., & Xu, Y. (2018). Human Age Prediction Based on DNA Methylation Using a Gradient Boosting Regressor. Genes, 9(9), 424. https://doi.org/10.3390/genes9090424

Loureiro, A., Miguéis, V., & da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. Decision Support Systems, 114, 81–93. https://doi.org/10.1016/j.dss.2018.08.010

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048–3061. https://doi.org/10.1109/tkde.2019.2962680

McGlynn, D., Coleman, S., Kerr, D., & McHugh, C. (2018). Day-Ahead Price Forecasting in Great Britain's BETTA Electricity Market. 2018 IEEE Symposium Series on Computational Intelligence (SSCI). Published. https://doi.org/10.1109/ssci.2018.8628744

Merkuryeva, G., Valberga, A., & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. Procedia Computer Science, 149, 3–10. https://doi.org/10.1016/j.procs.2019.01.100

Mor, R. S., Jaiswal, S. K., Singh, S., & Bhardwaj, A. (2018). Demand Forecasting of the Short-Lifecycle Dairy Products. Understanding the Role of Business Analytics, 87–117. https://doi.org/10.1007/978-981-13-1334-9_6

Pan, K., Shi, W., Wang, X., & Li, J. (2017). A short-term marginal price forecasting model based on ensemble learning. *2017 International Conference on Progress in Informatics and Computing (PIC)*. Published. https://doi.org/10.1109/pic.2017.8359519

Priyadarshi, R., Panigrahi, A., Routroy, S., & Garg, G. K. (2019). Demand forecasting at retail stage for selected vegetables: a performance analysis. Journal of Modelling in Management, 14(4), 1042–1063. https://doi.org/10.1108/jm2-11-2018-0192

Ren, S., Chan, H. L., & Siqin, T. (2019). Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. Annals of Operations Research, 291(1–2), 761–777. https://doi.org/10.1007/s10479-019-03148-8

Suma, V. (2020). Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics. Journal of Soft Computing Paradigm, 2(2), 101–110. https://doi.org/10.36548/jscp.2020.2.007

Sumaiya Farzana, G., & Prakash, N. (2020). Machine Learning in Demand Forecasting - A Review. SSRN Electronic Journal, 37. https://doi.org/10.2139/ssrn.3733548

Sun, W., & Gao, Q. (2019). Exploration of energy saving potential in China power industry based on Adaboost back propagation neural network. Journal of Cleaner Production, 217, 257–266. https://doi.org/10.1016/j.jclepro.2019.01.205

Taufiqurrahman, A., Putrada, A. G., & Dawani, F. (2020). Decision Tree Regression with AdaBoost Ensemble Learning for Water Temperature Forecasting in Aquaponic Ecosystem. 2020 6th International Conference on Interactive Digital Media (ICIDM). Published. https://doi.org/10.1109/icidm51048.2020.9339669

Tirkeş, G., Güray, C., & Çelebi, N. (2017). Demand forecasting: a comparison between the Holt-Winters, trend analysis and decomposition models. Tehnicki Vjesnik - Technical Gazette, 24(Supplement 2). https://doi.org/10.17559/tv-20160615204011

Tsoumakas, G. (2018). A survey of machine learning techniques for food sales prediction. Artificial Intelligence Review, 52(1), 441–447. https://doi.org/10.1007/s10462-018-9637-z

van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2020). Predicting customer demand for remanufactured products: A data-mining approach. European Journal of Operational Research, 281(3), 543–558. https://doi.org/10.1016/j.ejor.2019.08.015

Villegas, M. A., Pedregal, D. J., & Trapero, J. R. (2018). A support vector machine for model selection in demand forecasting applications. Computers & Industrial Engineering, 121, 1–7. https://doi.org/10.1016/j.cie.2018.04.042

Wu, P., & Chen, Y. (2021). Product Demand Forecasting in Ecommerce Based on Nonlinear Autoregressive Neural Network. https://doi.org/10.21203/rs.3.rs-194285/v1

Xenochristou, M., Hutton, C., Hofman, J., & Kapelan, Z. (2020). Water Demand Forecasting Accuracy and Influencing Factors at Different Spatial Scales Using a Gradient Boosting Machine. Water Resources Research, 56(8). https://doi.org/10.1029/2019wr026304

Yoon, J. (2020). Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. Computational Economics, 57(1), 247–265. https://doi.org/10.1007/s10614-020-10054-w

CHAPTER 7: APPENDICES

7.1 Detailed Results Performance of Validation set for Nonlinear Autoregressive Neural Network (NARNN) Model

Table 15: Validation performance of NARNN model using different training algorithms (10 runs)

Training Algorithms	Runs	MSE	RMSE
Levenberg-Marquardt	1	0.0447	0.2114
	2	0.0546	0.2337
	3	0.0577	0.2402
	4	0.0572	0.2392
	5	0.0632	0.2514
	6	0.0349	0.1868
	7	0.0636	0.2522
	8	0.0316	0.1778
	9	0.0460	0.2145
	10	0.0700	0.2646
Bayesian Regularization	1		
	2		
	3		
	4		
	5		
	6		
	7		
	8		
	9		
	10		
Scaled Conjugate Gradient Backpropagation	1	0.0855	0.2924
	2	0.0384	0.1960
	3	0.0392	0.1980
	4	0.0647	0.2544
	5	0.0693	0.2632
	6	0.0556	0.2358
	7	0.0594	0.2437
	8	0.0451	0.2124
	9	0.0600	0.2449
	10	0.0337	0.1836

7.2 Detailed Results Performance of Testing set for Nonlinear Autoregressive Neural Network (NARNN) Model

Table 16: Test performance of NARNN model using different training algorithms (10 runs)

Training Algorithms	Runs	MSE	RMSE
Levenberg-Marquardt	1	0.0390	0.1975
	2	0.0789	0.2809
	3	0.1477	0.3843
	4	0.0678	0.2604
	5	0.0498	0.2232
	6	0.0499	0.2234
	7	0.0481	0.2193
	8	0.0211	0.1453
	9	0.0599	0.2447
	10	0.0348	0.1865
Bayesian Regularization	1	0.0492	0.2218
	2	0.0583	0.2415
	3	0.0235	0.1533
	4	0.1913	0.4374
	5	0.0386	0.1965
	6	0.1744	0.4176
	7	0.0482	0.2195
	8	0.0635	0.2520
	9	0.1174	0.3426
	10	0.0236	0.1536
Scaled Conjugate Gradient Backpropagation	1	0.0312	0.1766
	2	0.0747	0.2733
	3	0.0308	0.1755
	4	0.0506	0.2249
	5	0.1480	0.3847
	6	0.0688	0.2623
	7	0.0407	0.2017
	8	0.0442	0.2102
	9	0.0248	0.1575
	10	0.0159	0.1261

S2023355_Dissertation

DR KASTURI DEWI A/P VARATHAN
DEPT OF INFORMATION SYSTEMS
FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
50603 KUALA LUMPUR

ORIGINALITY REPORT

12%	10%	6%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Malaya Student Paper	1 %
2	studentsrepo.um.edu.my Internet Source	1 %
3	assets.researchsquare.com Internet Source	1 %
4	Submitted to Swinburne University of Technology Student Paper	1 %
5	link.springer.com Internet Source	<1 %
6	www.hindawi.com Internet Source	<1 %
7	www.journaltocs.ac.uk Internet Source	<1 %
8	Submitted to University of Northumbria at Newcastle Student Paper	<1 %
9	mdpi-res.com	

Internet Source <1%

10	Sepp Hochreiter, Jürgen Schmidhuber. "Long Short-Term Memory", Neural Computation, 1997 Publication	<1%
11	www.mdpi.com Internet Source	<1%
12	Submitted to University of Northampton Student Paper	<1%
13	docplayer.net Internet Source	<1%
14	dergipark.org.tr Internet Source	<1%
15	Submitted to MODUL University Vienna Student Paper	<1%
16	Submitted to Segi University College Student Paper	<1%
17	dokumen.pub Internet Source	<1%
18	www.researchgate.net Internet Source	<1%
19	ebin.pub Internet Source	<1%

20	repositorio.iscte-iul.pt Internet Source	<1 %
21	www.efka.utm.my Internet Source	<1 %
22	d-nb.info Internet Source	<1 %
23	Submitted to Coventry University Student Paper	<1 %
24	Submitted to Universiti Putra Malaysia Student Paper	<1 %
25	onlinelibrary.wiley.com Internet Source	<1 %
26	Submitted to The University of Manchester Student Paper	<1 %
27	eprints.usm.my Internet Source	<1 %
28	ir-library.ku.ac.ke Internet Source	<1 %
29	researchbank.ac.nz Internet Source	<1 %
30	scholarworks.bridgeport.edu Internet Source	<1 %
31	uwspace.uwaterloo.ca Internet Source	<1 %

32	Submitted to Boston University Student Paper	<1 %
33	Mohammad Atrees, Ashraf Ahmad, Firas Alghanim. "Enhancing Detection of Malicious URLs Using Boosting and Lexical Features", Intelligent Automation & Soft Computing, 2022 Publication	<1 %
34	Submitted to Hellenic Open University Student Paper	<1 %
35	umpir.ump.edu.my Internet Source	<1 %
36	research-information.bris.ac.uk Internet Source	<1 %
37	www.acreinc.net Internet Source	<1 %
38	www.sepln.org Internet Source	<1 %
39	Submitted to Birla Institute of Technology and Science Pilani Student Paper	<1 %
40	Daniel McGlynn, Sonya Coleman, Dermot Kerr, Catherine McHugh. "Day-Ahead Price Forecasting in Great Britain's BETTA Electricity Market", 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018	<1 %

Publication

41 Seunghoon Baek, Younghoon Cho, Jih-Sheng Lai. "Average Periodic Delay-Based Frequency Adaptable Repetitive Control With a Fixed Sampling Rate and Memory of Single-Phase PFC Converters", IEEE Transactions on Power Electronics, 2021 <1%
Publication

42 Shiliang Zhang, Cong Liu, Hui Jiang, Si Wei, Lirong Dai, Yu Hu. "Nonrecurrent Neural Structure for Long-Term Dependence", IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017 <1%
Publication

43 Submitted to Saint Joseph's University <1%
Student Paper

44 Submitted to University of Leicester <1%
Student Paper

45 Www.mdpi.com <1%
Internet Source

46 dspace.plymouth.ac.uk <1%
Internet Source

47 ijsrcseit.com <1%
Internet Source

48 tutorsonspot.com <1%
Internet Source

49	A.L.D. Loureiro, V.L. Miguéis, Lucas F.M. da Silva. "Exploring the use of deep neural networks for sales forecasting in fashion retail", Decision Support Systems, 2018 Publication	<1 %
50	purehost.bath.ac.uk Internet Source	<1 %
51	"Cloud Computing and Big Data: Technologies, Applications and Security", Springer Science and Business Media LLC, 2019 Publication	<1 %
52	1library.net Internet Source	<1 %
53	Submitted to Universidade do Porto Student Paper	<1 %
54	research.chalmers.se Internet Source	<1 %
55	researchspace.ukzn.ac.za Internet Source	<1 %
56	www.science.gov Internet Source	<1 %
57	"Modelling, Computation and Optimization in Information Systems and Management Sciences", Springer Science and Business Media LLC, 2022	<1 %

Publication

58	Qi Wang, Xiaodan Wang, Fengyi Xia. "Integration of Grey Model and Multiple Regression Model to Predict Energy Consumption", 2009 International Conference on Energy and Environment Technology, 2009 Publication	<1 %
59	kansai-u.repo.nii.ac.jp Internet Source	<1 %
60	rgu-repository.worktribe.com Internet Source	<1 %
61	www.pi.events Internet Source	<1 %
62	Afonso Silva. "Estimating Fuel Consumption from GPS Data", Repositório Aberto da Universidade do Porto, 2014. Publication	<1 %
63	Chih-Hsuan Wang, Jen-Yu Chen. "Demand forecasting and financial estimation considering the interactive dynamics of semiconductor supply-chain companies", Computers & Industrial Engineering, 2019 Publication	<1 %
64	Ivana P. Milovanovic, Dejan B. Popovic. "Mapping of sensory representation of walking and EMG of prime joint movers: Control of functional electrical stimulation",	<1 %

10th Symposium on Neural Network Applications in Electrical Engineering, 2010
Publication

---

65 Kishan Bhushan Sahay, Suneet Sahu, Pragya Singh. "Short-term load forecasting of Toronto Canada by using different ANN algorithms", 2016 IEEE 6th International Conference on Power Systems (ICPS), 2016 <1%
Publication

---

66 Marco A. Villegas, Diego J. Pedregal, Juan R. Trapero. "A support vector machine for model selection in demand forecasting applications.", Computers & Industrial Engineering, 2018 <1%
Publication

---

67 Submitted to University of Westminster <1%
Student Paper

---

68 Y.W. Lee, K.G. Tay, Y.Y. Choy. "Forecasting Electricity Consumption Using Time Series Model", International Journal of Engineering & Technology, 2018 <1%
Publication

---

69 archive.org <1%
Internet Source

---

70 energsustainsoc.biomedcentral.com <1%
Internet Source

---

eprints.usq.edu.au

71	Internet Source	<1%
72	f1000research.com Internet Source	<1%
73	ikee.lib.auth.gr Internet Source	<1%
74	polen.itu.edu.tr Internet Source	<1%
75	publications.polymtl.ca Internet Source	<1%
76	repository.tudelft.nl Internet Source	<1%
77	savoirs.usherbrooke.ca Internet Source	<1%
78	www.tandfonline.com Internet Source	<1%
79	Daming Zhou, Fei Gao, Elena Breaz, Alexandre Ravey, Abdellatif Miraoui. "Degradation prediction of PEM fuel cell using a moving window based hybrid prognostic approach", Energy, 2017 Publication	<1%
80	Jaehyun Yoon. "Forecasting of Real GDP Growth Using Machine Learning Models:	<1%

Gradient Boosting and Random Forest Approach", Computational Economics, 2020
Publication

81 Vatani, A., K. Khorasani, and N. Meskin. "Health Monitoring and Degradation Prognostics in Gas Turbine Engines Using Dynamic Neural Networks", Volume 6 Ceramics Controls Diagnostics and Instrumentation Education Manufacturing Materials and Metallurgy Honors and Awards, 2015.
Publication
<1%

82 "Intrusion detection by machine learning = Behatolás detektálás gépi tanulás által", Corvinus University of Budapest, 2020
Publication
<1%

83 Fernando Martinez-Plumed, Lidia Contreras-Ochando, Cesar Ferri, Jose Hernandez Orallo et al. "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories", IEEE Transactions on Knowledge and Data Engineering, 2020
Publication
<1%

84 Jie Zhang, Souma Chowdhury, Achille Messac. "Uncertainty Quantification in Surrogate Models Based on Pattern Classification of Cross-validation Errors", 12th AIAA Aviation Technology, Integration, and Operations
<1%