

**RETAIL SALES FORECASTING IN THE PRESENCE
OF PROMOTIONS: COMPARISON OF STATISTICAL
AND MACHINE LEARNING FORECASTING
METHODS**

HALGAMUWE HEWAGE HARSHA RUWAN CHAMARA

208068F

Degree of Master of Science

Department of Transport and Logistics Management

University of Moratuwa

Sri Lanka

July 2022

**RETAIL SALES FORECASTING IN THE PRESENCE
OF PROMOTIONS: COMPARISON OF STATISTICAL
AND MACHINE LEARNING FORECASTING
METHODS**

HALGAMUWE HEWAGE HARSHA RUWAN CHAMARA

208068F

Thesis/Dissertation submitted in partial fulfilment of the requirements for the degree
of Master of Science in Supply Chain and Data Science

Department of Transport and Logistics Management

University of Moratuwa

Sri Lanka

July 2022

DECLARATION OF ORIGINALITY

I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

I hereby also grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or another medium. I retain the right to use this content in whole or part in future works (such as articles or books).

UOM Verified Signature

Signature:

Date: 08/07/2022

(H.H.H.R. Chamara)

STATEMENT OF THE SUPERVISOR

The above candidate has carried out research for the Degree of Master of Science under my supervision.

Signature of the Supervisor: *UOM Verified Signature* Date: 08/07/2022.....

Dr. H.N. Perera,
Senior Lecturer,
Dept. of Transport & Logistics Management,
Faculty of Engineering,
University of Moratuwa.

Abstract

Retail sales forecasting is the process of estimating the number of future sales for a specific product or products. However, producing reliable and accurate sales forecasts at a product level is a very challenging task in the retail context. Many factors can influence observed sales data at the product level, such as sales promotions, weather, holidays, and special events, all of which causes demand irregularities. Sales promotions are one of the salient drivers in generating irregular sales patterns. Sales promotions confound retail operations, causing sudden demand changes not just during the promotion period, but also throughout the demand series. As a result, three types of periods are relevant for sales promotions: normal, promotional, and a post-promotional. However, previous research has mostly focused on promotional and normal (i.e., non-promotional) periods, often neglecting the post-promotional period. To address this gap, we explore the performance of comprehensive methods, namely gradient-boosted regression trees, random forests, and deep learning in all periods. Moreover, we compare proposed approaches with conventional forecasting approaches in a retail setting. Our results demonstrate that machine learning methods can deal with demand fluctuations generated by retail promotions while enhancing forecast performance throughout all time periods. The base-lift model outperformed machine learning methods, although with more effort necessary to cleanse sales data. Our findings indicate that machine learning methods can automate the forecasting process and provide significant performance even with the standard approach. Hence, our research demonstrates the way retailers can successfully apply machine learning methods in forecasting sales.

Keywords: Forecasting, Promotions, Retail supply chain, Post-promotional effect, Machine learning

ACKNOWLEDGMENTS

I would like to take this opportunity pen my appreciation and heartfelt gratitude heartfelt gratitude to each and every one of them.

The first person I am extremely grateful and indebted to is Dr. H. Niles Perera, my supervisor and mentor. During the last two years, he gave me his continued guidance and support and also provided me with lots of freedom. Your kind support has been priceless to me. I am extremely grateful to my external advisor Dr. Kasun Bandara of The University of Melbourne, Australia for his guidance and support on forecasting methodologies.

I would also wish to extend my sincere gratitude to Dr. T. Sivakumar, postgraduate research coordinator of the Department of Transport and Logistics Management, for his guidance and support during the degree program progress reviews. I also wish to express my heartiest gratitude to Dr. Charith Chitraranjan for his extensive evaluations and feedback during the progress reviews.

I am immensely grateful to Senior Prof. Amal S. Kumarage, former Head and Founder of the Department of Transport and Logistics Management at the University of Moratuwa, and Prof. A.A.D.A.J. Perera, the current Head of the Department, for providing me with the opportunity to pursue the degree program.

I am extremely grateful to Dr. Amila Thibbotuwawa for evaluating my work and providing feedback to further improve it. My special thanks are due to the Center for Supply Chain Operations, and Logistics Optimization (SCOLO) fraternity for always being there to discuss anything and everything. I would also like to extend my gratitude towards academic and non-academic staff members at the Department of Transport and Logistics Management, for their support. I am thankful to The Senate Research Committee of the University of Moratuwa and the grant bearing ID SRC/LT/2020/20 which enabled this thesis and must be mentioned with immense gratitude.

My family has been the extremely encouraging. Last but not least, I'd like to express my heartfelt gratitude to my parents and sister, whose love and wisdom are with me in whatever I do. I could not have done this without you!

TABLE OF CONTENTS

DECLARATION OF ORIGINALITY	i
STATEMENT OF THE SUPERVISOR.....	ii
Abstract	iii
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF EQUATIONS	x
LIST OF ABBREVIATIONS	xi
1 INTRODUCTION	1
2 LITERATURE REVIEW.....	5
2.1 Retail Supply Chain.....	5
2.2 Retail Sales Promotions.....	6
2.3 Supply Chain Forecasting	7
2.3.1 Demand forecasting	7
2.3.2 Retail sales forecasting.....	7
2.4 Retail Sales Forecasting Methods	8
2.4.1 Human factor in retail sales forecasting.....	8
2.4.2 Incorporating sales promotions in retail sales forecasting	9
2.5 Problem Description.....	10
2.5.1 Research problem derivation	10
2.5.2 Hypothesis development	11
3 METHODOLOGY.....	13
3.1 Data and Input Features.....	13
3.2 Data Pre-processing.....	16

3.3	Benchmark Model	17
3.4	Forecasting methods	17
3.4.1	NAIVE and SNAIVE	18
3.4.2	ARIMA	18
3.4.3	ETS and ETSX	18
3.4.4	Gradient-boosted Regression Trees	20
3.4.5	Random Forest	20
3.4.6	DeepAR and WaveNet	20
3.4.7	Overview of the Candidate Models	21
4	ANALYSIS AND RESULTS	23
4.1	Magnitude and Sign of Post-promotional Effect	24
4.2	Comparison of Forecast Performances	26
4.2.1	Forecast performance during the normal period	27
4.2.2	Forecast performance during the promotional period	27
4.2.3	Forecast performance during the post-promotional period	28
4.3	Forecast improvement under compared methods	29
5	DISCUSSION	34
5.1	Findings	34
5.2	Managerial and Practical Implications	36
5.3	Limitations and Future Directions	36
6	CONCLUSION	38
	REFERENCES	39

LIST OF FIGURES

Figure 1-1 Variations in demand in retail sales promotions	2
Figure 3-1 Category distribution of the SKUs	14
Figure 3-2 Weekly sales by category	14
Figure 3-3 Average weekly sales by promotional period	15
Figure 3-4 Correlation matrix for features	16
Figure 4-1 Distribution of the magnitude of post-promotional dip.....	25
Figure 4-2 (a) sMAPE values in normal period; (b) MASE values in normal period	27
Figure 4-3 (a) sMAPE values in promotional period; (b) MASE values in promotional period.....	28
Figure 4-4 (a) sMAPE values in post-promotional period; (b) MASE values in post-promotional period	29

LIST OF TABLES

Table 3-1 Descriptive summary of the dataset	13
Table 3-2 Selected input features	15
Table 3-3 Overview of the candidate models.....	22
Table 4-1 Descriptive summary of the magnitude and sign of the post-promotional dip	25
Table 4-2 Forecast accuracy for each forecasting method	26
Table 4-3 FVA value comparison for normal period	31
Table 4-4 FVA value comparison for promotional period.....	32
Table 4-5 FVA value comparison for post-promotional period.....	33

LIST OF EQUATIONS

Equation 3-1 Post-promotional effect calculation	16
Equation 3-2 Base-lift estimation calculation.....	17
Equation 3-3 ARIMA (p,d,q) model.....	18
Equation 3-4 ETS model	19
Equation 3-5 ETSX model.....	19
Equation 4-1 Post-promotional effect calculation	23
Equation 4-2 sMAPE calculation	23
Equation 4-3 MASE calculation	23
Equation 4-4 Forecast value added calculation	24

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
ANN	Artificial Neural Networks
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
BL	Base-Lift
BT	Boosted Trees
DL	Deep Learning
ETS	Exponential Smoothing
ETSX	Exponential Smoothing with Exogenous Variable
FSS	Forecasting Support System
FVA	Forecast Value Added
GBRT	Gradient-Boosted Regression Trees
LGB	LightGBM
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
ML	Machine Learning
SNAIVE	Seasonal NAIIVE
NN	Neural Network
RF	Random Forest
RT	Regression Trees
SKU	Stock Keeping Unit
sMAPE	Symmetric Mean Absolute Percentage Error
SVR	Support Vector Machines
TPR	Temporary Price Reductions
XGB	xgBoost