

SALES FORECASTING MODELS FOR DIRECT SELLING BUSINESS: A
DATA-DRIVEN APPROACH TO PREDICTIVE ANALYTICS

SIVARAJAN A/L S.ESVARAN

UNIVERSITI TEKNOLOGI MALAYSIA

CHAPTER 5

DISCUSSION AND FUTURE WORKS

1.1 Introduction

This final chapter concludes the comprehensive data science project for sales forecasting in direct selling business by presenting a summary of significant findings, critical observations of model performance, and strategic business recommendations derived from rigorous analysis of Amway sales data. As the culmination of an extensive application of machine learning and time-series forecasting models, including LSTM, Random Forest, ARIMA, and Linear Regression, this chapter synthesizes the contributions of the study, acknowledges its limitations, and outlines potential directions for future research and model refinement. The findings not only provide insights into predictive capabilities for sales trends but also establish a foundation for data-driven strategic business decisions in the direct selling ecosystem.

1.2 Summary

The project started with a comprehensive dataset of 553,542 sales transactions that took place over two years, from April 2023 to April 2025. This data came from Amway distributors and was real market data. The dataset was of very high quality, with 100% completeness across 12 core features. This gave it a strong base for advanced analytics and modelling. A lot of exploratory data analysis showed important information about how customers act, how well products perform in different situations, and how sales change over time. This laid the groundwork for strategic business intelligence.

One of the main things that came out of the analysis was that premium products were the main source of revenue. The Atmosphere Sky™ Air Treatment System brought in RM 44.5 million and the eSpring Water Purifier brought in RM 37.5 million in total sales. The analysis of the customer demographic showed that the average age of the customers was 47.8 years and that they had a lot of buying power, as shown by the fact that the average transaction value was RM 606.31 and that the average order size was 3 units.

The temporal analysis showed that the business was growing steadily and that demand patterns were stable over time, which means that the business had sustainable momentum. Advanced preprocessing and feature engineering expanded the dataset from 12 to 28 variables through the creation of temporal features, customer behavioral metrics, pricing indicators, and purchase recency measures. This enrichment facilitated the development of four distinct predictive models: LSTM Neural Networks, Random Forest, ARIMA, and Linear Regression. However, the model evaluation revealed significant challenges, with three machine learning models (LSTM, Random Forest, Linear Regression) achieving identical performance metrics - R^2 of 0.964, RMSE of RM 355.61, MAE of RM 112.49, and MAPE of 52.68%. Despite strong explanatory power, the high MAPE severely limited practical forecasting utility, with only 25.04% of predictions meeting acceptable accuracy thresholds.

ARIMA presented a contradictory performance profile, demonstrating poor overall metrics (negative R^2 of -0.106) while paradoxically achieving 100% custom accuracy and the lowest scaled RMSE (5.26%). This inconsistency highlighted fundamental evaluation methodology issues requiring further investigation.

Throughout the project, detailed visualisations helped each analysis phase by turning complicated technical outputs into easy-to-use business intelligence tools that showed patterns in revenue distribution, seasonal trends, customer demographics, and model performance comparisons. The project showed that traditional statistical

measures can be misleading, but that careful analysis of multiple evaluation criteria is necessary to create reliable forecasting systems for direct selling.

1.3 Future Work

While meeting the main research objectives, various important improvements and lines of future research have been recognized in order to further develop the forecasting capacities and business value of the analytical framework:

Model Architecture Refinement and Optimization: Tackle this issue as you would with the same one across other machine learning models for potential data leakage, pre-processing inconsistencies or feature engineering issue. Investigate more advanced LSTM architectures in search of attention mechanisms or bidirectional processing or even GRU-based alternatives to boost the temporal pattern capturing ability. Build ensemble techniques that incorporate multiple models and take the advantage of the strong sides of individual models and relieve the problem of the weaknesses of them.

Enhancement of the Evaluation Framework: Disentangle the opposite values of ARIMA performance metrics by performing detailed methodological reviews of the evaluation process, using common cross-validation methods, and following similar accuracy measurement schemes. Establishing business-sensitive scoring measures to meet the forecasting needs of direct selling operations.

Customer-Level Predictive Analytics: Further model the aggregate forecasting method at the customer level through prediction of individual customer behavior based on customer demographics, transaction history and behavioral segmentation. Develop customized predictive models that can forecast customer lifetime value, purchase propensity, and product preference trends to better manage customer relationships.

Real-Time Integration and Automation: Enable automated pipelines to retrain models as new data comes in via streaming transactions, so forecasts remain accurate as business evolves. Enable live dashboard systems to monitor sales flows in real time, alert for insane pattern formation and give inventory advice automatically.

Geographic And Channel Level Analysis: Extend the model to support regional sales performance analysis, channel effectiveness measurement, and market penetration optimization. Information forecasting models by geographical location to inform expansion strategy and resource allocation.

In summary, although the present study lays a good foundation for data driven sales forecasting in direct selling businesses, these future improvements are expected to further evolve the analytical framework into a complete business intelligence ecosystem that supports strategic decision making, operational optimization, and sustainable growth in competitive market environments. The advancements achieved will close the gap between the scientific academic research and business applications and provide value-added analytics for the modern direct selling companies.