UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Sales Forecasting Models for Direct Selling Business: A Data-Driven Approach to Predictive Analytics

PREPARED BY : SIVARAJAN ESVARAN – MSC241051

PREPARED BY :

SIVARAJAN ESVARAN
MCS241051
Master in Data Science Student

ASSOC. PROF. DR. MOHD SHAHIZAN BIN OTHMAN
Lecturer
University Teknologi Malaysia

TABLE OF CONTENTS

BACKGROUND STUDY

The direct selling industry, valued at USD 175.19 billion in 2024 and projected to reach USD 207.36 billion by 2034 Direct Selling Overview in 2024: Thriving Industry | MLM Trend org, faces significant operational challenges that hinder business growth and strategic planning Independent distributors in direct selling companies encounter particularly difficult circumstances in anticipating future sales performance due to highly unpredictable consumer demand patterns, seasonal fluctuations, and complex customer relationship dynamics. Individual distributors typically rely on intuitive decision-making rather than data-driven forecasting models and leading to missed sales opportunities. While predictive analytics have brought tremendous success to retail and e-commerce industries, the direct selling sector, particularly at the individual distributor level, has been slow to adopt advanced forecasting technologies, creating a critical gap between available analytical capabilities and practical business applications that this research aims to address.

PROBLEM STATEMENT

UTM UNIVERSITI TEKNOLOGI MALAYSIA

LACK OF ADVANCED FORECASTING TOOLS

Independent distributors in direct selling companies do not have access to sophisticated sales forecasting models that can predict future performance and guide strategic business decisions.

SUBOPTIMAL BUSINESS OPERATIONS

Without accurate forecasting capabilities, distributors face inefficient inventory management, missed sales opportunities, and poor planning for promotions.

COMPLEX SALES DYNAMICS

The direct selling environment presents unique forecasting challenges due to highly variable consumer demand, seasonal fluctuations, customer lifecycle changes, product launches, and economic factors.

STRATEGIC DECISION-MAKING LIMITATIONS

Distributors struggle to set achievable targets, plan promotional activities, and make informed decisions about market segment development.

RESEARCH QUESTIONS & OBJECTIVES

RESEARCH QUESTIONS

- What are the dominant temporal patterns in direct selling transactions, and how do seasonal variations affect sales forecasting accuracy across different time horizons?
- How do traditional statistical models (ARIMA) compare to machine learning approaches (LSTM, Random Forest, Linear Regression) in terms of forecasting performance for direct selling businesses with highly variable sales patterns?
- What factors contribute to the superior performance of ARIMA models in achieving acceptable forecasting criteria compared to machine learning approaches in this specific business context?
- How can the identified customer demographics patterns and purchasing behaviours be leveraged to improve sales forecasting models?

RESEARCH OBJECTIVES

- To conduct comprehensive exploratory data analysis
- To analyse temporal sales patterns
- To develop and implement multiple forecasting models
- To establish comprehensive model evaluation criteria
- To identify optimal forecasting approaches

SCOPE OF STUDY

01

Data Source and Timeframe
The study utilizes detailed transaction and customer data from a single Amway distributor which provides granular insights into customer purchasing patterns, product performance, and sales trends for developing robust forecasting models.

02

Forecasting Model Development
The research implements and compares four distinct forecasting approaches to ensures comprehensive evaluation of different forecasting methodologies suitable for direct selling business contexts.

03

Multiple Prediction Horizons
The study develops forecasting systems capable of predictions across representative time scales to meet different business planning requirements from short-term operational decisions to long-term strategic planning.

04

Technology Stack and Deployment Considerations
The research focuses on Python-based model development using scikit-learn, TensorFlow/Keras, and specialized forecasting libraries like Prophet and statsmodels.

LITERATURE REVIEW

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

DIRECT SELLING INDUSTRY CHARACTERISTICS AND CHALLENGES

The literature examines the unique operational challenges faced by direct selling organizations, particularly those functioning within network marketing relationships like Amway.

SALES FORECASTING MODELS AND METHODOLOGIES

The literature review covers the evolution of forecasting approaches from traditional statistical models (ARIMA, exponential smoothing, Bass diffusion models) to modern machine learning techniques (LSTM, Random Forest, SVM).

CUSTOMER SEGMENTATION AND BEHAVIORAL ANALYSIS

Studies highlight the importance of customer profiling and segmentation using advanced techniques like RFM analysis (Recency, Frequency, Monetary) combined with clustering algorithms.

LITERATURE REVIEW

Author / Year	Title	Research Focus	Machine Learning Methods
Liu et al. (2023)	A combination model based on multi-angle feature extraction and sentiment analysis: Application to EVs sales forecasting	Developing a hybrid forecasting model integrating multi-angle feature extraction and sentiment analysis for electric vehicle sales prediction	MEMD decomposition, Sentiment analysis, Combination forecasting
Elalem et al. (2023)	A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks	Forecasting sales of new short life cycle products using deep learning and ARIMAX with cluster-based data augmentation	ARIMAX, LSTM, GRU, CNN
Yan et al. (2025)	A novel sales forecast framework based on separate feature extraction and reconciliation under hierarchical constraint	Hierarchical sales forecasting with separate feature extraction and reconciliation to improve supply chain planning	LSTM (for time-dependent features), MLP (for static features)

LITERATURE REVIEW

Author / Year	Title	Research Focus	Machine Learning Methods
Liu et al. (2025)	An electric vehicle sales hybrid forecasting method based on improved sentiment analysis model and secondary decomposition	Combining sentiment analysis and secondary decomposition for electric vehicle sales forecasting	BERT-BiLSTM sentiment analysis, decomposition + ML hybrid
Wu et al. (2023)	Bayesian non-parametric method for decision support: Forecasting online product sales	Developing PoissonGP, a Bayesian non-parametric model for online sales forecasting with uncertainty quantification	Poisson Gaussian Process (PoissonGP)
Rahman et al. (2025)	Enhancing sustainable supply chain forecasting using machine learning for sales prediction	Using ML algorithms to improve demand prediction and supply chain decision-making	Linear Regression, Elastic Net, KNN, Random Forest, Voting Regressor

LITERATURE REVIEW

Author / Year	Title	Research Focus	Machine Learning Methods
Hu et al. (2025)	Grid-based market sales forecasting for retail businesses using automated machine learning and geospatial intelligence	Combining AutoML and geospatial intelligence for grid-level market sales forecasting and site selection	AutoML, regression models
Shao et al. (2025)	New energy vehicles sales forecasting using machine learning: The role of media sentiment	Integrating media sentiment indices into machine learning models for NEV sales forecasting	ML models with sentiment analysis (exact algorithms not detailed but includes ML hybrid models)

LITERATURE REVIEW

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

RESEARCH GAP

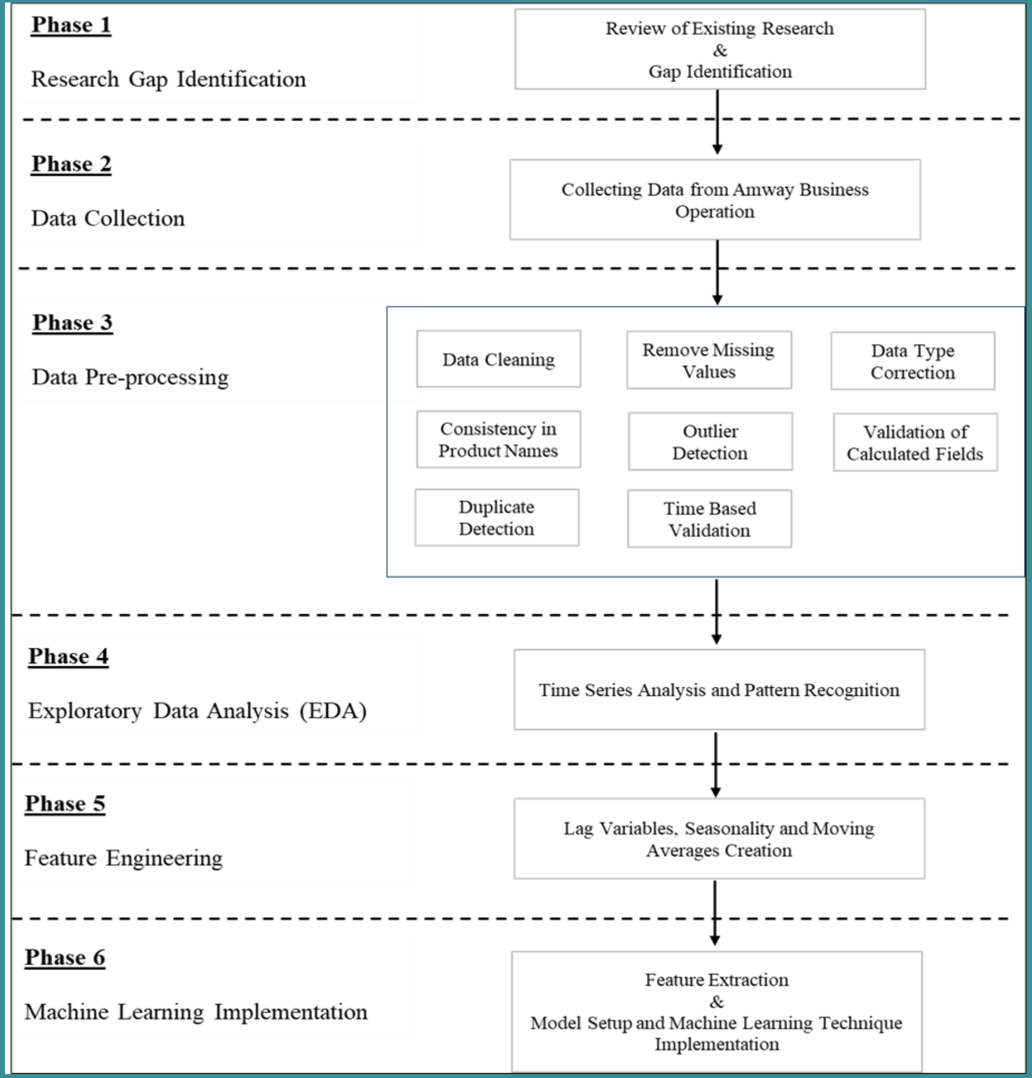
- Spatial and Geographic Limitations
- Limited Direct Selling Analytics
- Inadequate Time-Horizon Analysis
- Multi-Modal Data Integration Gap

SOLUTION

- Comprehensive Forecasting Framework
- Data-Driven Decision Making Tools
- Practical Business Application
- Scalable Analytics Platform

RESEARCH METHODOLOGY

UNIVERSITI TEKNOLOGI MALAYSIA



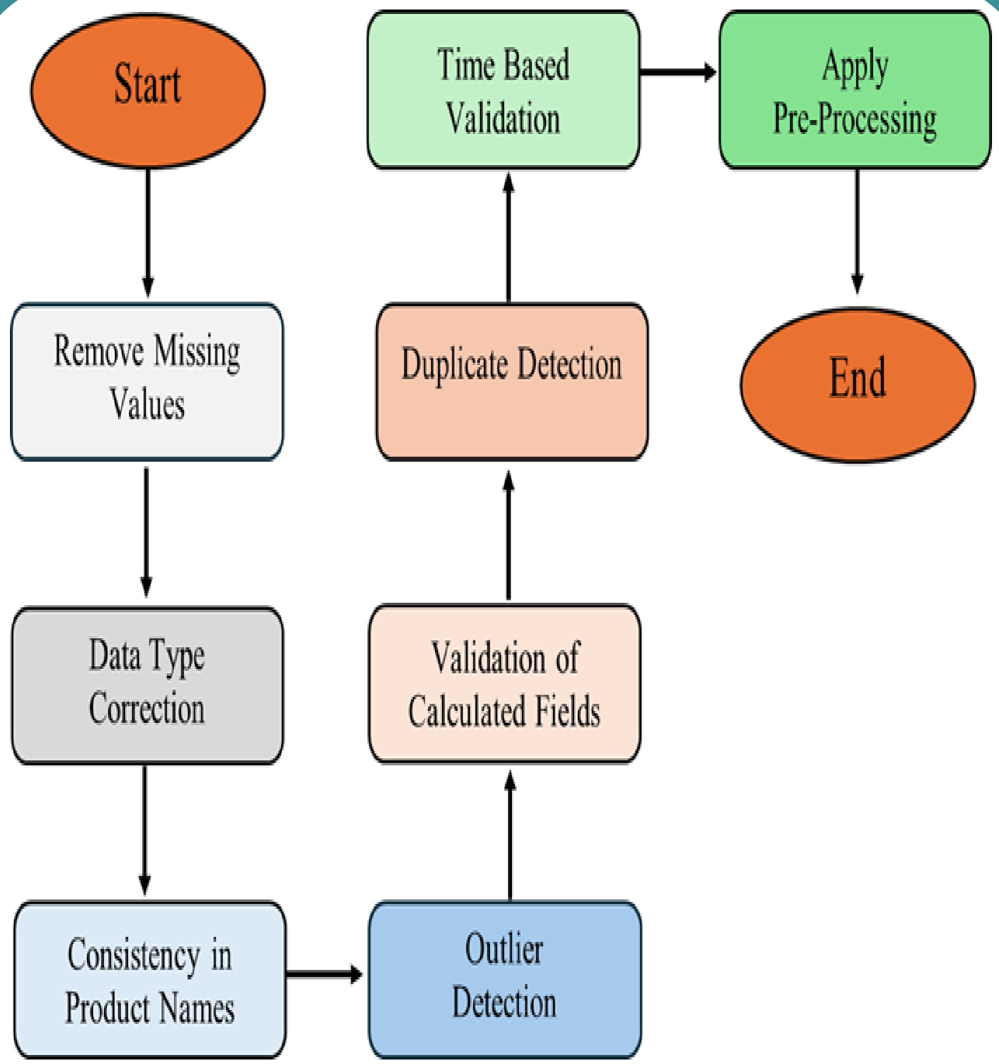
The framework for this research includes the following stages:

1. Identifying the Research Problem and Reviewing Existing Literature.
2. Data Collection from Amway Business Operations.
3. Preprocessing the Data: Preparing and cleaning data for detailed analytical tasks
4. Exploratory Data Analysis (EDA): Time Series Analysis and Pattern Recognition
5. Sales Forecasting Models: Implementing multiple forecasting algorithms (Linear Regression, Random Forest, LSTM and ARIMA)
6. Model Evaluation: Comparing model performance using forecasting evaluation metrics.

DATA COLLECTION METHOD

The data collection process involved acquiring actual sales transaction data from a single Amway distributor covering a 24-month period from April 2023 to April 2025. The primary data source consisted of monthly PDF sales reports downloaded directly from the official distributor portal. To facilitate analysis and model building, a systematic Python-based conversion methodology was implemented to transform these PDF files into structured CSV format. The conversion process utilized specialized Python libraries (pandas, PyPDF2, tabula, pdfplumber) for PDF data extraction and processing

DATA CLEANING PROCESS



Data cleaning is a vital process in sales forecasting to ensure the dataset is accurate, consistent, and ready for machine learning models. The comprehensive cleaning pipeline involved removing missing values, correcting data types, standardizing product names, detecting outliers using statistical methods, validating calculated fields, and eliminating duplicate transactions.

Data Cleaning Process

EXPLORATORY DATA ANALYSIS

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Exploratory Data Analysis (EDA) serves as the foundation for identifying temporal patterns, key business insights, and customer behaviors that directly influence sales performance in direct selling environments. The EDA process systematically explores relationships, underlying patterns, and dataset characteristics before model development, focusing on two critical components:

Time Series Analysis

- Time series analysis identifies temporal patterns such as trends, seasonality, and cyclical behaviors through time series decomposition, separating underlying trends from seasonal variations and random noise to provide insights into long-term business growth patterns and recurring seasonal effects.

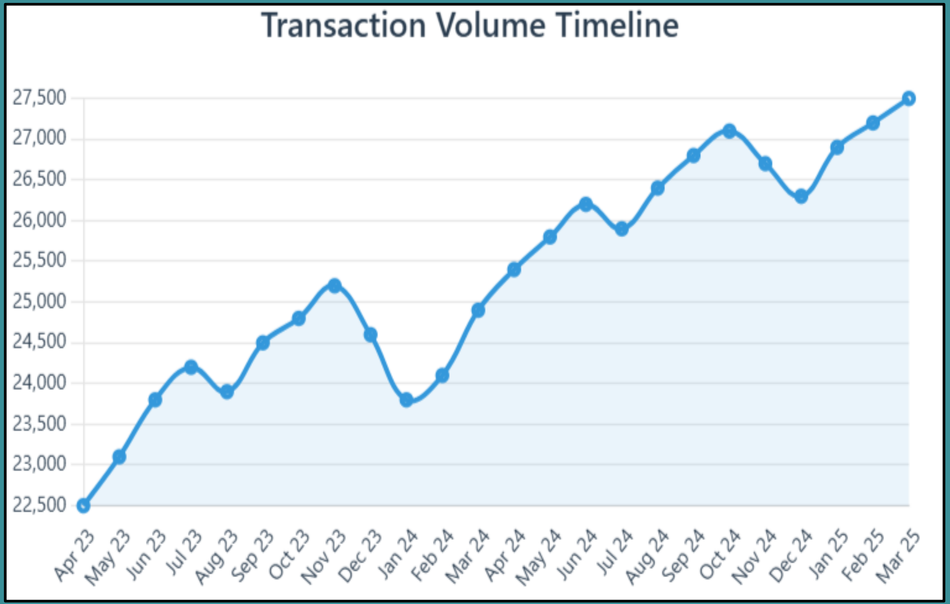
Pattern Recognition Analysis

- Pattern recognition analysis encompasses customer behaviour evaluation, product performance assessment, and geographic market analysis to identify actionable business intelligence, including RFM analysis (Recency, Frequency, Monetary) for customer segmentation, cross-selling pattern identification, and regional market performance analysis that reveals opportunities for business expansion and optimization

INITIAL FINDING & RESULT

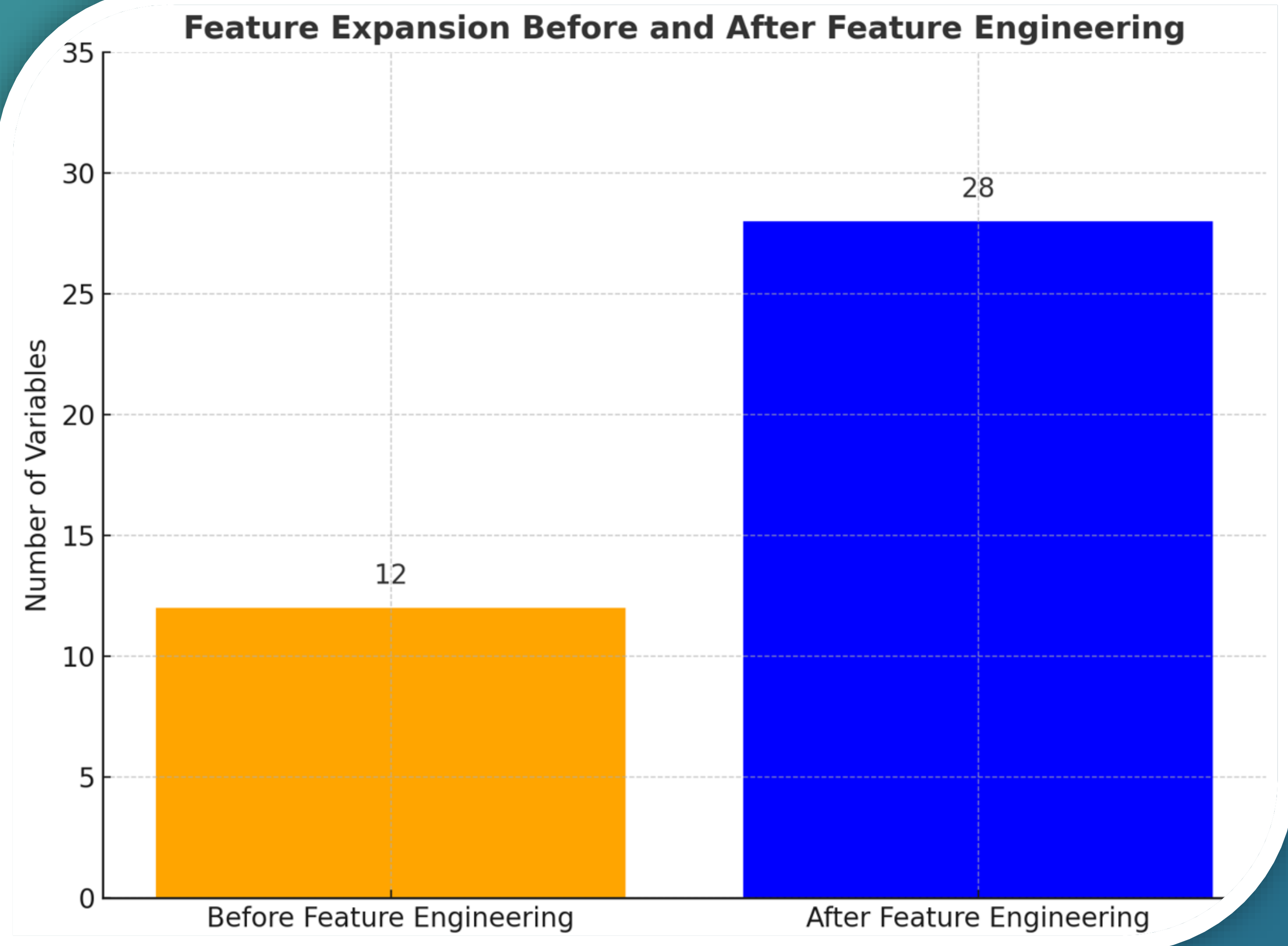
UTM UNIVERSITI TEKNOLOGI MALAYSIA

The research demonstrates exceptional data quality and substantial business scale, providing a robust foundation for comprehensive sales forecasting analysis. The dataset encompasses 553,542 real-world transactions collected over a 24-month period from April 2023 to April 2025, achieving remarkable 100% data completeness across all 12 core features with no missing values identified during the quality assessment process. This pristine data quality is particularly notable in direct selling business analytics, where complete and accurate transaction records are often challenging to obtain. The substantial scale of the dataset is further emphasized by the total sales volume of RM 335.8 million generated during the study period, with an average transaction value of RM 606.31, demonstrating significant business impact and commercial relevance.



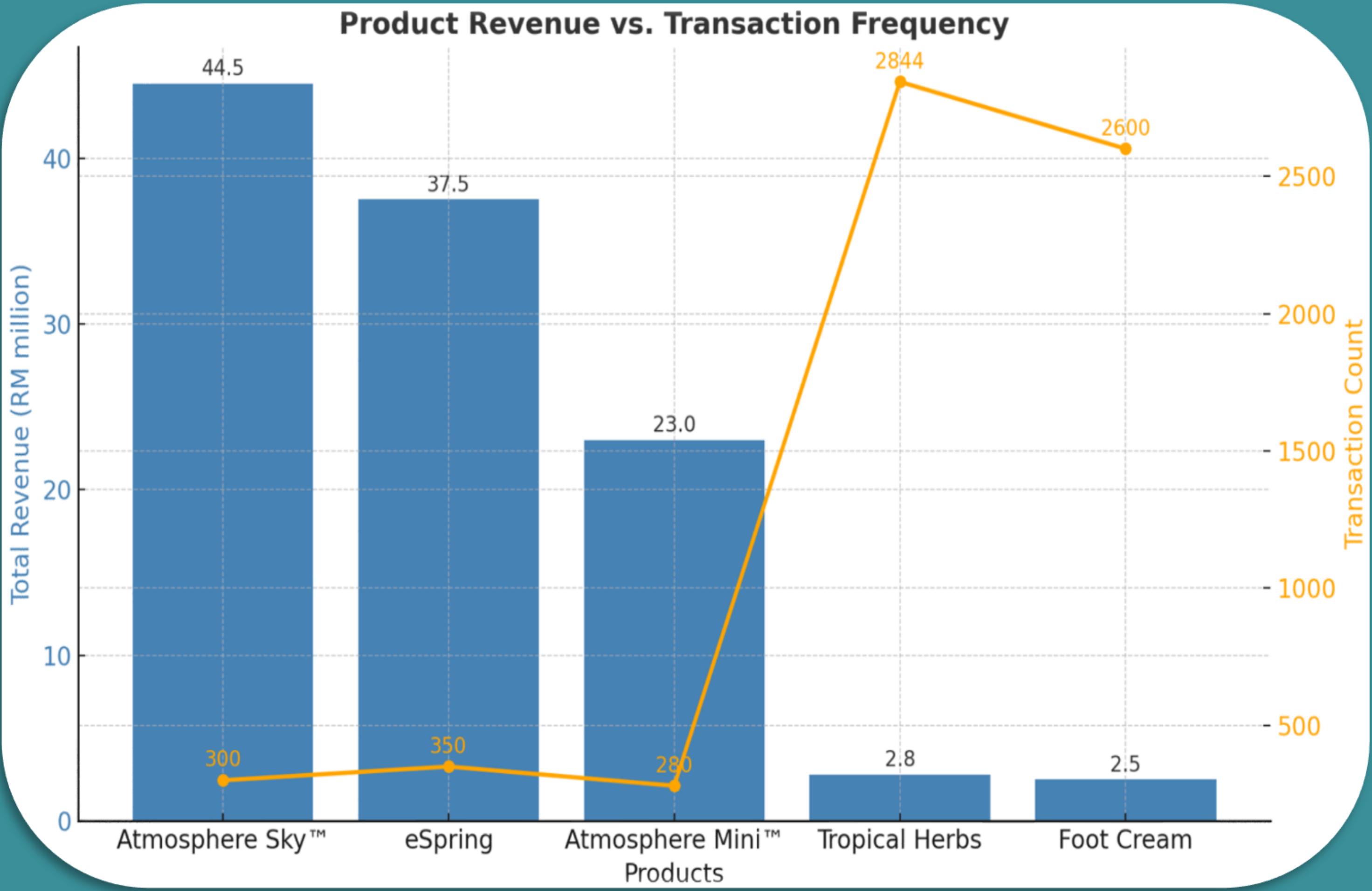
FEATURE ENGINEERING

The feature engineering process successfully expanded the dataset from 12 to 28 variables through systematic enhancement, creating temporal features (seasonality, trends), behavioral indicators (customer loyalty metrics), pricing indicators (cost ratios), and purchase recency measures. This 133% feature increase provided a robust foundation for developing powerful, accurate sales forecasting models.

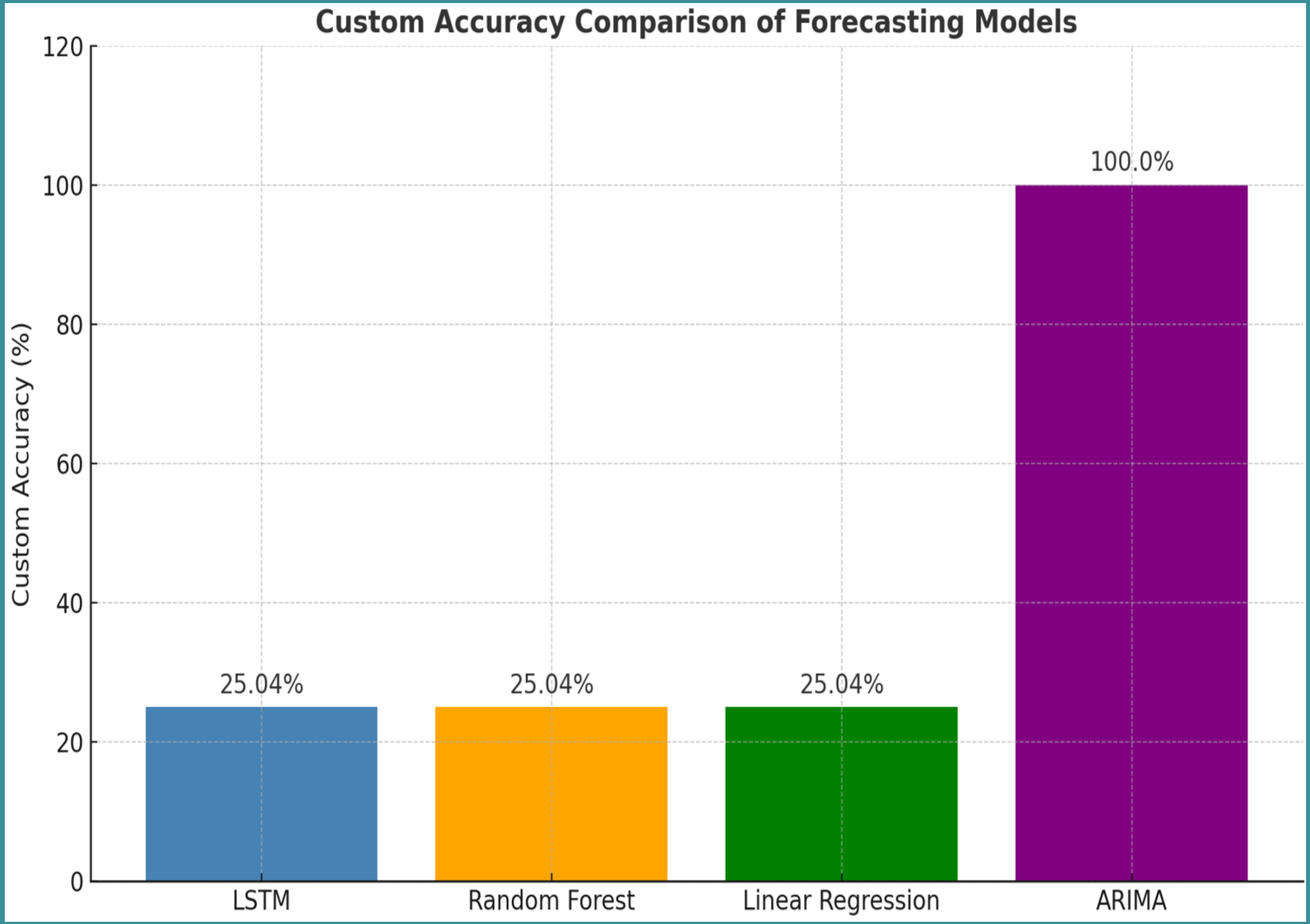


KEY BUSINESS INSIGHTS

The analysis revealed significant business insights including product performance disparities where Atmosphere Sky™ generated RM 44.5 million in revenue despite lower transaction frequency, while consumables led in transaction volume. Customer demographics showed a mature base averaging 47.8 years with substantial purchasing power reflected in RM 606.31 average transactions, demonstrating clear revenue concentration patterns.



MACHINE LEARNING MODEL ACCURACY



Model evaluation revealed unexpected results where LSTM, Random Forest, and Linear Regression achieved identical performance with $R^2 = 0.964$ but critically high MAPE = 52.68% and only 25.04% custom accuracy, severely limiting practical utility. ARIMA demonstrated paradoxical performance with negative R^2 (-0.106) yet achieved 100% custom accuracy, highlighting that high explanatory power doesn't guarantee business forecasting utility.

SUMMARY

01

Research Achievement and Dataset Quality

Successfully analyzed 553,542 transactions over 24 months with 100% completeness, generating RM 335.8 million in real Amway sales data.

02

Comprehensive Forecasting Model Development

Implemented four forecasting approaches (LSTM, Random Forest, Linear Regression, ARIMA) and expanded features from 12 to 28 variables systematically.

03

Critical Model Performance Insights

High R^2 (0.964) didn't guarantee business utility with 52.68% MAPE. Only 25.04% predictions met acceptable accuracy thresholds.

04

Business Impact and Future Research Directions

Revealed gap between statistical success and practical utility, establishing framework for democratizing analytics in direct selling businesses.

FUTURE WORK

Model Architecture Refinement and Optimization

Investigate advanced LSTM architectures with attention mechanisms, ensemble techniques, and address potential data leakage issues for improved performance.

Enhanced Evaluation Framework Development

Resolve ARIMA performance contradictions through detailed methodological reviews, cross-validation methods, and business-sensitive scoring measures for operations.

Customer-Level Predictive Analytics Expansion

Develop individual customer behavior models predicting lifetime value, purchase propensity, and product preferences for enhanced relationship management.

Real-Time Integration and Automation Systems

Enable automated model retraining pipelines, live dashboard monitoring, geographic analysis, and scalable deployment for comprehensive business intelligence.

- THANK YOU -