University for the Creative Arts

BERLIN SCHOOL OF BUSINESS & INNOVATION

**Dissertation Title: EXPECTED TRANSACTION VOLUME ESTIMATE FROM MEMBER BUSINESSES**

**Master title: MSC DATA ANALYTICS**

**Name: EMRE ÖZYÜREK**

**Year: 2025**

## ABSTRACT

While the adoption of machine learning technologies creates enormous opportunities for SMEs, it also engenders very significant challenges. The thesis examined in detail the feasibility of the use of machine-learning models in the predictions of transaction volume for SMEs through case studies using the Instacart dataset and a survey of SMEs operating under resource constraints. The four major themes it addresses are developing ML models appropriate for SMEs, integrating real-time data analytics, integrating behavioral economics, and the evaluation of leading-edge technologies like AutoML and federated learning. Key findings from the survey reveal that recent machine-learning techniques are little known; only 10% of the SMEs adopt AutoML, and 70% are unaware of federated learning. Logistic Regression, being the baseline model, performed really well in terms of accuracy, around 90%, but failed to perform well on class balance, specifically in the prediction of reordered products. AutoML and federated learning appear as two promising solutions to address challenges in resource constraints, class imbalance, and data privacy concerns. Real-time analytics and behavioral economics are very important, both for better forecasting accuracy and for inventory management, yet their application is still very limited. The similarities in the challenges faced by SMEs and those that arose during model implementation were also very evident-computational constraints and feature engineering in big datasets. Future research should focus on sector-specific applications, hybrid models, and explainable AI that can increase the adoption of ML in SMEs.



Keywords

Machine Learning, SMEs, AutoML, Federated Learning, Transaction Forecasting, Real-Time Analytics, Behavioral Economics, Class Imbalance, Instacart Dataset, SME Survey

# CONTENTS

ABSTRACT ........ 2

CONTENTS ........ 3

ACKNOWLEDGEMENTS ........ 7

INTRODUCTION ........ 9

BACKGROUND/ BRIEF ANALYSIS OF THE RESEARCH TOPIC ........ 9
TRIGGER AND RATIONALE FOR RESEARCH TOPIC SELECTION ........ 10
RESEARCH QUESTIONS ........ 10
RESEARCH OBJECTIVES ........ 11
METHODOLOGY ........ 12
SYNOPSIS OF THE CHAPTERS ........ 13

CHAPTER ONE – LITERATURE REVIEW I ........ 14

1.1 INTRODUCTION ........ 14
1.2 DEFINITION AND SCOPE ........ 14
1.3 THE IMPORTANCE OF VOLUME FORECAST ........ 14
1.4 TRANSACTION VOLUME ESTIMATION IN LITERATURE ........ 15
1.5 CHALLENGES AND OPPORTUNITIES ........ 15
1.6 TRADITIONAL MODELS AND METHODS ........ 16
1.7 MACHINE LEARNING APPROACHES ........ 16
1.8 DATA QUALITY AND PROCESSING CHALLENGES ........ 17
1.9 SEASONALITY AND REGIONAL DIFFERENCES ........ 18
1.10 PERFORMANCE MEASURES IN FORECAST MODELS ........ 18
1.11 SAMPLE APPLICATION AREAS FOR TRANSACTION VOLUME FORECASTING ........ 20
1.12 FUTURE RESEARCH AREAS AND INNOVATION ........ 21
1.13 ADVANCED MACHINE LEARNING TECHNIQUES AND TRANSACTION VOLUME PREDICTION ........ 21
1.14 DATA SECURITY AND ETHICAL ISSUES ........ 22
1.15 BIG DATA AND REAL-TIME ANALYTICS ........ 23
1.16 SECTORAL APPLICATION EXAMPLES ........ 23

**1.17 Literature Gaps** ........ **24**
**1.18 Incorporating Theories** ........ **25**
**1.19 Summary** ........ **25**

**CHAPTER TWO – LITERATURE REVIEW II** ........ **27**

**2.1 Introduction** ........ **27**
**2.2 Characteristics of SMEs and Their Challenges in Volume Forecasting** ........ **27**
**2.2.1 Overview of SMEs in the Modern Economy** ........ **27**
**2.2.2 Transaction Volume Forecasting in SMEs** ........ **28**
**2.3 Machine Learning for SMEs** ........ **28**
2.3.1 Overview of Machine Learning Approaches ........ 29
2.3.2 SME-Friendly Machine Learning Models ........ 30
2.3.3 Improvements aimed at SMEs ........ 31
**2.4 Real-Time Analytics for SMEs** ........ **31**
**2.5 Addressing Data Quality Issues** ........ **32**
2.5.1 Challenges in Data Quality for SMEs ........ 32
2.5.2 Methods for Improving Data Quality ........ 32
**2.6 Behavioral Economics and Forecasting** ........ **33**
2.6.1 Behavioral Factors Influencing Transaction Volumes ........ 33
**2.7 Ethical Considerations and Data Privacy** ........ **35**
2.7.1 Importance of Ethical Forecasting ........ 35
2.7.2 Strategies for Ethical Implementation ........ 35
**2.8 Literature Gaps** ........ **35**
2.8.1 Lack of Context-Specific ML Solutions for SMEs ........ 35
2.8.2 Integration of Real-Time Analytics with ML in SME Forecasting ........ 36
2.8.3 Insufficient Exploration of Behavioral Economics in ML Forecasting Models ........ 36
2.8.4 Challenges in Data Quality and Standardization ........ 36
2.8.5 Limited Emphasis on Explainability and Usability of ML Models for SMEs ........ 36
2.8.6 Sector-Specific Adaptations of ML Models ........ 37
2.8.7 Lack of Studies on Collaborative ML Approaches for SMEs ........ 37
2.8.8 Minimal Focus on AutoML for SME Applications ........ 37
2.8.9 Real-World Case Studies and Best Practices ........ 37
**2.9 Summary** ........ **37**

**CHAPTER THREE – METHODOLOGY** ........ **39**

**3.1 INTRODUCTION** ........ **39**

**3.2 RESEARCH DESIGN** ........ **39**
**3.3 RESEARCH METHOD** ........ **40**
**3.4 DATA COLLECTION** ........ **41**
3.4.1 PRIMARY DATA ........ 41
3.4.2 SECONDARY DATA ........ 42
**3.5 SAMPLING PROCESS** ........ **42**
3.5.1 SAMPLING STRATEGY ........ 43
3.5.2 SAMPLE CHARACTERISTICS ........ 43
**3.6 DATA ANALYSIS** ........ **44**
3.6.1 QUANTITATIVE ANALYSIS ........ 44
3.6.2 QUALITATIVE ANALYSIS ........ 44
3.6.3 SYNTHESIS OF FINDINGS ........ 45
**3.7 RESEARCH LIMITATIONS** ........ **45**
**3.8 ETHICAL CONSIDERATIONS** ........ **46**
**3.9 SUMMARY** ........ **46**

**CHAPTER FOUR – FINDINGS / ANALYSIS / DISCUSSION** ........ **48**

**4.1 INTRODUCTION** ........ **48**
**4.2 FINDINGS** ........ **49**
4.2.1 Exploratory Data Analysis (EDA) ........ 49
4.2.2 Machine Learning Statistical Tests and Results ........ 56
4.2.3 Summary of Findings ........ 65
**4.3 ANALYSIS** ........ **66**
4.3.1 Objective 1: Investigating the development of ML models for SMEs under resource constraints ........ 66
4.3.2 Objective 2: Integration of real-time data analytics with ML models ........ 66
4.3.3 Objective 3: Incorporating behavioral economics into ML forecasting models ........ 67
4.3.4 Objective 4: Evaluating the feasibility of AutoML and federated learning for SMEs ........ 67
**4.4 DISCUSSION** ........ **69**
4.4.1 Beneficiaries of the Research ........ 69
4.4.2 Implications ........ 69
4.4.3 Reflections and Future Directions ........ 70
**4.5 SUMMARY** ........ **71**

**CONCLUDING REMARKS** ........ **72**

**Key Findings of the Study** ........ **72**

SUPPORT OF THE LITERATURE REVIEW .......... 72
CONTRADICTIONS TO EXISTING LITERATURE .......... 73
LIMITATIONS AND PROBLEMS FACED .......... 73

BIBLIOGRAPHY .......... 75

LIST OF FIGURE .......... 83

## ACKNOWLEDGEMENTS

This motivation and knowledge have been an uphill but very rewarding journey in coming to this dissertation, and I'm really indebted to those who guided me through.

Above all, I would like to extend my deep and profound gratitude to my thesis supervisor, Dr. Anuj Batta, who was indispensable to this research with his expertise, patience, and encouragement. His sagacious guidance and constructive criticisms have shaped not only this dissertation but deeply enriched my academic and professional growth.

I will never forget to thank the Berlin School of Business and Innovation for the course of MSc Data Analytics, which provides enabling environments for academic excellence and innovation. It is due to the resources, knowledge, and support acquired from this institution that this work was able to reach fruition.

To the SMEs and other respondents who took my survey, thank you for your generous investment of time and insight. Your contributions were instrumental in helping me embed this research in reality and ensure it was relevant and practical.

I want especially to thank my family and friends for their love, encouragement, and understanding. You all have been my strength, keeping me grounded in difficult times and cheering for me when all was going right.

Thanks to everyone big or small, whoever made this possible: thanks for making this milestone a reality.

## Statement of compliance with academic ethics and the avoidance of plagiarism

I honestly declare that this dissertation is entirely my own work and none of its part has been copied from printed or electronic sources, translated from foreign sources and reproduced from essays of other researchers or students. Wherever I have been based on ideas or other people texts I clearly declare it through the good use of references following academic ethics.

(In the case that is proved that part of the essay does not constitute an original work, but a copy of an already published essay or from another source, the student will be expelled permanently from the postgraduate program).

Name and Surname (Capital letters):

EMRE ÖZYÜREK

Date: 20/01/2025

## INTRODUCTION

***Background/ Brief Analysis of the Research Topic***

Small and medium enterprises (SMEs) represent more than 90% of businesses and 50% of employment worldwide. Formal SMEs contribute up to 40% of national income (GDP) in emerging economies (World Bank, 2021). However, one of the critical problems faced by them is the difficulty in accurately predicating their transaction volumes. This challenge affects inventory management and resource allocation for financial stability. With more data-driven approaches, machine learning has emerged as one of the prime transformative elements in transaction volume prediction. Despite the potential, the adoption of ML by small businesses is limited due to resource constraints, lack of technical expertise, and fragmented data systems (Iyelolu et al., 2024).

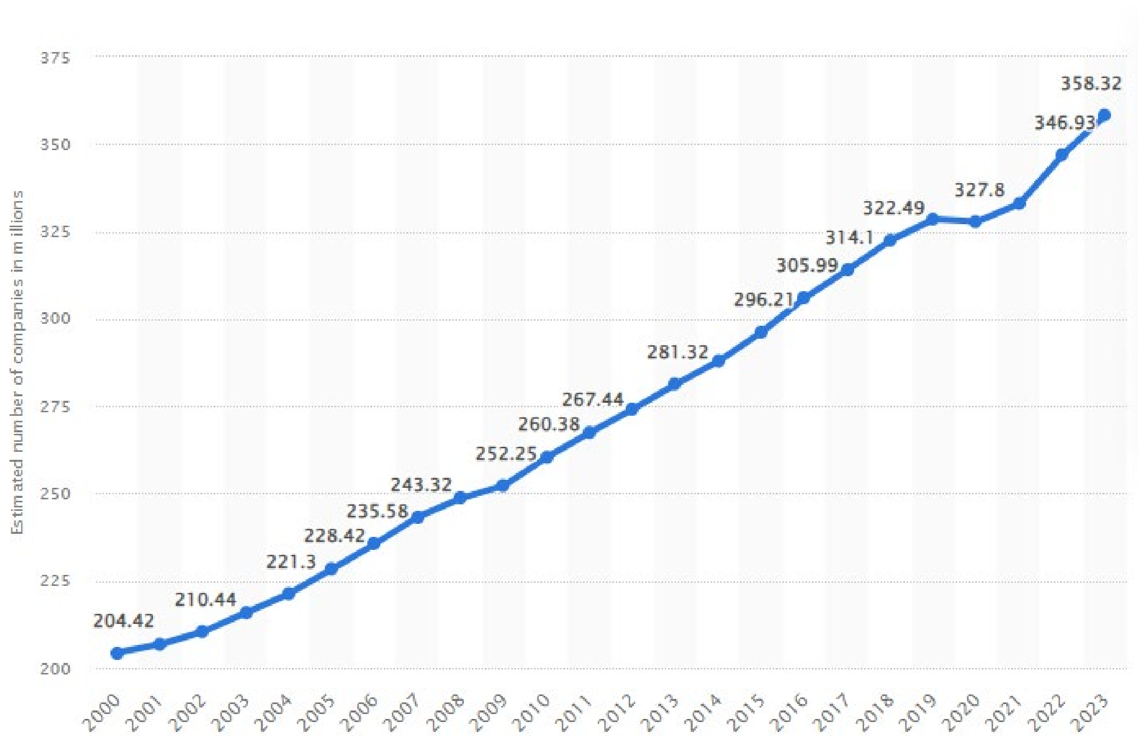


**Figure 1.Estimated number of small and medium sized enterprises (SMEs) worldwide from 2000 to 2023 in millions (Source: Statista, Published by Einar H. Dyvik, Aug 6, 2024)**

### *Trigger and Rationale for Research Topic Selection*

The rationale behind this study is informed by the observation that small businesses mostly rely on rudimentary means of forecasting transactions; hence, inefficiencies abound, coupled with missed opportunities (Chong et al., 2017). While large-scale enterprises have benefited from ML, its application in small business is still underexplored. This research is informed by the need to bridge this gap through the development of appropriate ML solutions that respond to peculiar challenges that beset small enterprises. Secondly, there is a great opportunity for enhancing the accuracy and relevance of the forecasts, which has not been fully realized by integrating real-time analytics and behavioral economics into ML models (Giglio et al., 2021).

### *Research Questions*

Building on the identified gaps in the literature, the following are the guiding research questions for this study:

- How can machine learning models be adapted to meet the specific challenges and resource constraints faced by SMEs in transaction volume forecasting?

The most of the literature that does exist on machine learning adoption does tend to focus on large-enterprise applications, with limited attempts having been made to adapt these models into the SME context. SMEs often encounter obstacles such as limited technological infrastructure and perceived costs when considering AI adoption (Al-Okaily et al., 2023).

- What challenges and benefits are there related to the integration of real-time data analytics into the machine learning models of the SMEs for the forecasts?

Real-time data enhances the precision and agility of forecasting systems (Wasserbacher and Spindler, 2021). However, the majority of the literature has not addressed issues related to IT infrastructure limits and fragmentation in the context of SMEs adopting real-time analytics. SMEs often face limitations due to constrained human and financial resources (Dörr et al., 2023). However, new technologies are anticipated to profoundly affect SMEs and society at large, which could ameliorate some constraints encountered by SMEs (Ameen et al., 2022).

- How would any insights from behavioral economics about consumer preferences and market behaviors be integrated into machine learning models for better forecasting accuracy by an SME?

While behavioral economics does give very important insights into the consumers that could add depth to the forecasting models of Giglio et al. (2021), the area of practical research on how these insights apply to the SME forecasting models is minimal and should, therefore, be an area of interest.

- To what extent can AutoML or federated learning helps make machine learning more usable and effective for SMEs, and how can they be used in practice?

AutoML systems automate the democratization of machine learning; users can design and deploy autonomously. On the other hand, federated learning enables users to collaborate and share data while preserving privacy (Zhang et al., 2021). However, their practical applications for SMEs are relatively under-explored and need examination.

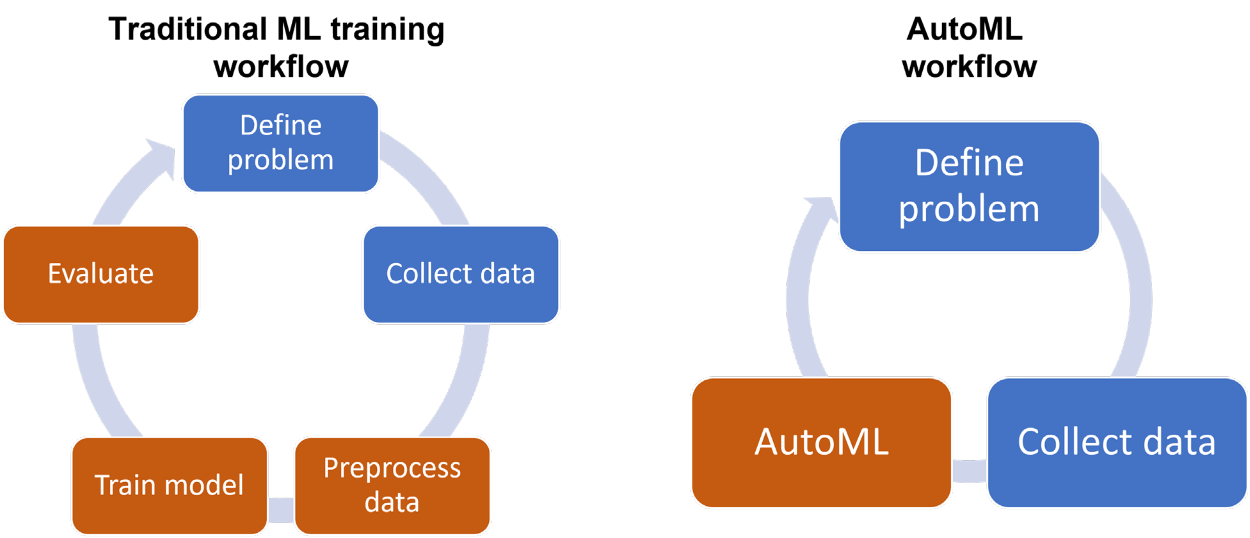


**Figure 2. How does AutoML work? (Source: learn.microsoft, 2024)**

***Research Objectives***

This paper intends to fill knowledge gaps that have been identified through the literature review, specifically:

- To investigate how machine learning models can be developed to address the unique challenges and resource constraints of SMEs in transaction volume forecasting.

The aim is to provide pragmatic insights on how machine learning models could be adapted to suit operational constraints of SMEs, considering their data, budgetary restrictions, and lack of technical expertise (Akter & Wamba, 2016).

- Investigate the potential of integrating real-time data analytics with machine learning models for improving forecasting accuracy in SMEs.

This objective will address how the integration of real-time transaction data and market signals into forecasting models can enhance their accuracy and responsiveness, given the challenges SMEs face in implementing real-time analytics (Wasserbacher & Spindler, 2021).

- Improving volume forecasts for SMEs may require integrating machine learning models with behavioral economics concepts such as consumer preference and market dynamics.

Second, behavioral economics can also provide insights into market behavior that can fine-tune forecasting models to better target consumer behavior and volumes (Giglio et al., 2021). This will look at how these lessons can be applied to the machine learning models for the SMEs. Evaluate the feasibility and effectiveness of AutoML or federated learning in making machine learning more accessible to users and improving their predictive capabilities. The proposed objective will improve predictions for users by testing the feasibility of using AutoML and federated learning to overcome both the technical expertise barrier and limitations in data sharing (Zhang et al., 2021).

***Methodology***

This research applies the mixed-methods approach by combining quantitative and qualitative research methods. The methodology will be conducted in the following manner:

• Primary Research: Case studies and interviews with SME owners on current practices in forecasting, challenges, and preparedness for the adoption of ML technologies.

• Secondary Research: Extensive literature review about state-of-the-art applications of ML, focusing especially on model forecasting, challenges, and their solutions in SMEs and using a data representing SME, performing predictive and behavioral analysis.

• Analytical Tools: These include the adaptation of statistical modeling and validation of the ML forecasting techniques to suit the small business needs in combination with qualitative analysis of interview data contextualizing the findings.

The disadvantages of this approach are that high-quality, real-world data from small businesses may be hard to come by, and that the models proposed can generalize across industries and geographical regions.

***Synopsis of the Chapters***

The dissertation is organized in the following manner:

- Chapter One: Literature Review I is devoted for the basic theorical concepts and empirical research on forecasting transaction volume, specifically on the shortcomings and developments of the attempted forecasting models.

- Chapter Two: Literature Review II broadens the scope of the literature on the use of machine learning methods in forecasting, with special emphasis on the use case focuses problems, potential opportunities, and gaps in the literature that are evident.

- Chapter Three: Methodology describes the research design including the ways of data collection, analytical methods, and limits of the research.

- Chapter Four: Findings – Analysis – Discussion: The research results are presented, analyzed and discussed in the context of the most important theoretical and practical insights.

# CHAPTER ONE – LITERATURE REVIEW I

***1.1 Introduction***

The last few years of the digital age and technological progress in data analytics have positioned the ability to predict transaction volume of member companies as one of the key elements of financial management. Transaction volume estimation is a valuable instrument for companies to understand their financial performance, optimize sales tactics, and better manage liquidity. Accurately predicting transaction volumes is of great importance, especially for member businesses operating in sectors sensitive to seasonality, consumer behavior, and economic fluctuations (González-Sánchez, 2021).

This literature review evaluates current approaches to estimating the expected transaction volume of member businesses. A wide range of methods, from traditional statistical methods to modern approaches such as machine learning and big data analytics, are analyzed. In addition, the limitations of existing studies, difficulties encountered in processing data, and areas foreseen for future research are discussed.

***1.2 Definition and Scope***

Transaction Volume is the total number of dealings transacted by any enterprise or network in any period (Law Insider, n.d.). In businesses operated on membership subscription systems, retail chains, and financial services, transaction volume is used as a tool for assessing consumer engagement and health in operations.

On the other hand, volume estimates incorporate the fundamental activities of any resource allocation, inventory management, and liquidity planning.

***1.3 The Importance of Volume Forecast***

It allows the company to be financially stable; and forecasting of transaction volume has the following advantage:

Operational efficiency ensures appropriate resource allocation and maximizes the processes of the company.

### *1.4 Transaction Volume Estimation in Literature*

In the literature, predictions of volume are normally handled in two ways:

Traditional Statistical Models: Regression analysis along with time series models (such as ARIMA) were employed. In traditional methods, volume is estimated by making inferences from past data (González-Sánchez, 2021).

Machine Learning and Big Data Approaches: More sophisticated methods for analyzing complex nonlinear interactions have included Random Forest, Gradient Boosting, and artificial neural networks. These are particularly adept at working with huge, diverse data sets (Akter, 2016).

Volume estimation has been a wide area of research in literature, as evident across financial markets, e-commerce, and other retail industries. In most of the research works, time series analysis, regression models, and machine learning techniques have come into the main discussion. Though traditionally statistical models like ARIMA and SARIMA have been used, deep learning techniques are becoming more utilized in solving more complex and nonlinear relationships. The accuracy of the estimation of trading volume depends on the scope of the dataset and the structure of the model used (Swedroe, 2024).

### *1.5 Challenges and Opportunities*

The following are the key issues encountered during the process of volume forecasting:

Data Quality and Access: Partial or inaccurately handled data reduces forecast accuracy.

Model Complexity: Advanced models require a lot of computational power and skill.

Seasonality and Regional Differences: These would form the major variables for consideration in volume forecasting (Akter, 2016).

The most significant challenge to volume forecasting would be inconsistent, uncertain data quality. Additionally, incomplete data with misplaced information and high noise levels typically affects the performance of a forecasting model. Big data and advanced machine learning models, however, can be a solution by which most of the problems are solved. Real-time data and AI models also enable cross-industry volume predictions to be made more accurately (Kotios, 2022)

### *1.6 Traditional Models and Methods*

Regression Models:

Regression analysis is the most standard method for trading volume forecasting. In this method, the volume of trading is taken as the dependent variable, and the forecasted outcome of it depends on various independent variables, such as market trend, earlier volume of trading, or seasonal impact.

Advantages: Easy to apply and the results are normally interpretable.

Disadvantages: Not very good with complex and nonlinear relations (González-Sánchez, 2021).

Time Series Models:

Time series models comprise ARIMA and SARIMA, which try to predict the next level of trading volume using past data. These can be particularly effective if there is a seasonal pattern.

Traditional estimation methods in trading volume estimation mostly lie with the statistical time series models. Using techniques such as ARIMA and Exponential Smoothing, these are performed on historical data to predict future trading volumes. These perform highly optimally with linear and deterministic sets of data. In the case of large and complex datasets, their accuracy is a bit restricted since it would require a more sophisticated approach (Silva, 2024).

### *1.7 Machine Learning Approaches*

Predicting trading volume has been revolutionized using machine learning. Traditional models are incapable of handling the strength of much more complex data sets and capturing nonlinear relationships better than these methods.

This is because machine learning can learn such complicated relationships and hidden patterns in the data. SVM, Random Forest, XGBoost, and LSTM model the nonlinear and time-varying feature of the data better. In principle, the advantage of these methods is that they increase accuracy by learning hidden information from the data set. In recent times, deep learning techniques are increasingly favored in the forecast of trading volume, while financial markets have become an important research focus (Wasserbacher, 2021)..

Random Forest and Gradient Boosting:

These methods of algorithms are ideal to model non-linear patterns and explain complex variable-variable interactions.

Advantages: Resistant to errors and reduce, in general the risk of over-fitting.

Example Application: It can be used for the prediction of daily trading volume by member businesses (Buntak, 2021).

Artificial Neural Networks:

Neural networks yield very high performance, especially for big datasets. This leads to a very high capability for learning higher-order data patterns and their relationships.

Area of Use: It can be used to estimate daily or hourly transaction volume of e-commerce platforms.

Limitations: The training period consumes much time and requires a very high processor requirement (Buntak, 2021).

| Aspect | Traditional Models | Machine Learning Models |
|---|---|---|
| Data Dependency | Historical, structured data | Big, diverse datasets |
| Relationship Type | Linear relationships | Nonlinear relationships |
| Interpretability | High interpretability | Medium interpretability |
| Computational Need | Low computational demand | High computational demand |
| Example Methods | ARIMA, Regression | Random Forest, Neural Networks |
| Advantages | Simplicity, explainability | Higher accuracy, adaptability |

**Figure 3. Comparison Between Traditional Statistical Models and Machine Learning Models**

### *1.8 Data Quality and Processing Challenges*

Transaction volume estimation is very much depends on the quality of the data. Poor data, noisy data, inconsistencies in data can directly affect the accuracy of forecast models. Moreover, the variability in data over time can reduce accuracy on the part of the forecast. Data cleaning and preprocessing is very important for the success of a model. On the other hand, diversifying sources of data and using advanced techniques of data processing gives an opportunity to increase forecast accuracy (Wasserbacher, 2021).

Missing Data Problem: Data deficiencies are more common, especially in small businesses. It significantly impacts the forecast models' accuracy. Imputation techniques can be applied as a remedy.

Outliers and Noise: The forecast may be weakened by transaction data outliers or random variance. Processes for cleansing data are consequently crucial (Akter, 2016).

### *1.9 Seasonality and Regional Differences*

Members' trading volumes vary significantly by region and industry:

Seasonal Variation: The retail industry is heavily dependent on holidays for trading volume.

Regional Factors: The local economy, population density, and cultural traditions of that region play a significant role.

These can be modeled to provide an increase in forecast accuracy. For example, regional trends can be considered with SARIMA models in retail chains (Buntak, 2021).

That is to say, one and the same economic recession influences the volume of transactions differently in another region. Machine learning algorithms should provide better modeling for seasonality and regional variations, hence more accurate forecasting (Silva, 2024).

### *1.10 Performance Measures in Forecast Models*

The metrics of performance help in benchmarking and comparing the various performances of predictive models, be it in wide areas of machine learning, statistics, among others. Below are key performance measures that are common and widely used for assessing forecasting models:

1. Mean Absolute Error (MAE)

MAE calculates the average of the absolute differences between the predicted and actual values. It is easy to interpret, as the error is expressed in the same units as the target variable. MAE treats all errors equally, making it suitable for applications where over- and under-predictions are equally important. However, it does not emphasize larger errors, which could be significant in some contexts (Draxler, 2014).

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m} |X_i - Y_i|$$

**Figure 4**

2. Mean Squared Error (MSE)

MSE is the average of the squared differences between predicted and actual values. Because of the squaring, larger errors get more weight; hence, MSE is useful for applications where the reduction of large deviations is a priority. Its output, is in squared units, which complicates its interpretation. It is usually used as a predecessor in deriving RMSE (Draxler, 2014).

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m} (X_i - Y_i)^2$$

**Figure 5**

3. Root Mean Squared Error (RMSE)

The RMSE is the square root of MSE, returning the metric to the same units as the data. It is sensitive to large errors but more interpretable than MSE. The RMSE is widely used in regression tasks and finds its most effective use in model performance evaluation when large deviations from actual values are critical (Draxler, 2014).

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (X_i - Y_i)^2}$$

**Figure 6**

4. Coefficient of Determination ($R^2$)

$R^2$: This explains the proportion of variance in the dependent variable explained by the model. It ranges from 0 to 1, and higher values indicate that the model performed well.

Unlike the MAE or MSE, $R^2$ has no units. That makes this metric appropriate for comparing performances across datasets.

That is usually adopted in regression tasks to check for model fitness (Chicco, 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (X_i - Y_i)^2}{\sum_{i=1}^{m} (\bar{Y} - Y_i)^2}$$

**Figure 7**

Which metric to use would depend on the purpose of the forecasting task. MAE is suitable for a general view of the error, while MSE and RMSE are more appropriate for penalizing large errors. $R^2$ gives an intuitive measure of the model's explanatory power. Many times, several metrics have to be used to get an appropriate overall view of the model performance.

### *1.11 Sample Application Areas for Transaction Volume Forecasting*

Retail Industry:

Because of seasonality and promotional periods, transaction volume estimation is most used in retail. For example, ARIMA or neural network models for higher transactions during Black Friday or the holiday seasons.

Financial Services:

Estimating the volume of transactions is highly relevant in services like credit card usage, ATM transactions, or digital payments for the purpose of risk management and infrastructure planning. Machine learning algorithms can model transaction density and user behavior (Akter, 2016).

E-commerce:

Transaction volume estimation in e-commerce is an important means for reflecting the effectiveness of inventory management, logistics, and improving customer satisfaction. Big data and real-time analytics are highly used in this industry.

### *1.12 Future Research Areas and Innovation*

Some opportunities for the next research could be forecast methods that are limited today. The real-time processing of transaction volume data brings great advantages to many sectors today, especially in finance and e-commerce.

Blockchain and Distributed Ledger Technology: The data from blockchain can be used in estimating transaction volume in a more reliable and traceable way.

Models for Small Businesses: The limitation of resources constrains the small businesses from accessing advanced forecasting facilities. These models should be more affordable and easier to use.

Quantum Computing: Provides opportunities for faster predictions on large data sets.

Customized Models: Specific prediction algorithms can be developed for sectoral needs.

Deep Learning and Hybrid Methods: Provide innovation for more complex data structures (Kotios, 2022).

These innovations remain critical issues to be focused on by both academic and applied studies (Buntak, 2021).

### *1.13 Advanced Machine Learning Techniques and Transaction Volume Prediction*

Machine learning has facilitated the diversification and development of algorithms for volume estimation. Advanced models outperform the classical techniques for the more challenging datasets.

Transformers:

While Transformers have become popular in the NLP domain, it has also grown strong in time series forecasting. Primarily, it is found that for large datasets and long-term dependency, transformers are useful in volume estimation through their attention mechanisms.

Example: To predict future volumes based on the long-term analysis of user behavior in e-commerce platforms (Buntak, 2021).

AutoML (Automatic Machine Learning):

AutoML automates model development and parameter optimization processes, enabling the user to develop powerful forecasting models with minimum technical knowledge.

Pros: It allows small businesses or teams with limited technical people to have an easy solution.

Cons: The model customization option may be limited (Buntak, 2021).

Reinforcement Learning:

Particularly effective in dynamic market conditions (Silva, 2024).

### *1.14 Data Security and Ethical Issues*

Most of the data applied to volume estimation contains customer behavior and transactions. This naturally justifies a number of data privacy and ethical issues.

GDPR and Data Privacy:

Laws like the European Union General Data Protection Regulation put boundaries on the processing of data of customers. Businesses developing forecasting models must comply with these rules.

Challenges: GDPR allows only those data to be processed which are necessary; hence richness of the data may be at stake in the models (Buntak, 2021).

Ethical Use:

Prediction models process sensitive information to predict user behavior. It should be implemented properly to safeguard business reputation and gain more consumer trust.

Recommendation: Development and Implementation of guidelines on ethical use will help businesses in the longer term (Akter, 2016).

### *1.15 Big Data and Real-Time Analytics*

Volume estimation using big data and real-time analytics tools is on the rise at an alarming rate. The approaches have a broad range of benefits, particularly for those industries involving high volumes of transactions.

Big Data Analytics:

Traditional methods cannot compare with big data tools since they operate on huge data sets. Hadoop and Spark are examples of tools that can even combine and analyze large-scale data sources that affect transaction volumes.

Application: Integration and analysis of data from several stores in the retail sector (Kotios, 2022).

Real-time Analytics:

It allows businesses to make rapid decisions based on transaction volume data. Forecasting transaction density at any ATM allows a financial service provider to optimize the maintenance or cash replenishment operation.

Challenges: Require high processing power and increased costs due to infrastructure requirements (Akter, 2016).

### *1.16 Sectoral Application Examples*

Applications of Predictive Models in various sectors enable us to comprehend the success and benefit of the model better.

Healthcare:

Transaction volume prediction in the healthcare sector is used in fields such as hospital density, management of drug inventories, and organizing patient appointments. Time series models and neural networks are widely known to work well in this sector.

Example: Forecasting how seasonal flu outbreaks affect patient density.

Energy Sector:

In the energy sector, forecasting models allow electricity distribution companies to predict energy demand and thus prevent outages. This also aids in economizing on operation, besides outage avoidance.

Example: How Cold Waves Drive Energy Demand by (Buntak, 2021).

Education and Public Sector:

Forecasting transaction volume is vital for addressing citizen requests in education and the public sector. The models could be used in education to develop better student registration and examination processes.

### *1.17 Literature Gaps*

Despite tremendous improvements in volume forecasting, the literature contains many gaps:

Integration of Small Businesses: Many studies (e.g., Akter & Wamba, 2016; Kotios et al., 2022) focus on big data and machine learning for larger enterprises. However, little research has been conducted on tiny enterprises that do not have access to high-quality data or computing resources. The development of models tailored to such entities is, therefore, crucial.

Cross-Sector Analysis: Many studies, such as those by González-Sánchez et al. (2021) and Silva et al. (2024), focus on very specific industries, such as financial services or trading. Cross-industry frameworks that could be useful across many industries, especially retail and energy, have not been explored in great depth. Future research could focus on adaptive models that account for the industry diversity.

Incorporating Real-Time Analytics: While Wasserbacher and Spindler (2021) and Tran et al. (2023) underline the relevance of real-time data, little is said about how to integrate it with hybrid models. Several practical problems with data latency and costs have not yet been resolved.

Ethical and Privacy Concerns: Although most studies acknowledge data privacy, such as González-Sánchez et al. (2021), in practice, not very few ethical frameworks among AI-assisted volume forecasts are identified. Clear rules for ensuring GDPR and other regulatory compliance are crucial for research.

Most of the literature did without the theoretical groundwork and concentrated on empirical and algorithmic models. This implies that decision theory or behavioral finance principles applied to model transaction volume is seldom used.

### *1.18 Incorporating Theories*

Theoretical approaches can explain transaction volume estimation in greater detail:

1. Decision Theory: This theory can underpin models of transaction volume by investigating how businesses make decisions under uncertainty. It corresponds to González-Sánchez et al. (2021), who analyze market risks and support the frameworks for risk assessment in fluctuating transaction environments.

2. Resource Dependence Theory: Buntak et al. (2021) indirectly refer to the challenges that businesses face when dependent on external data or computational resources. Applying this theory might create premises upon which intervention could be made to reduce dependence on costly analytics platforms.

3. Behavioral Economics: Understanding consumer behavior, as Tran et al. (2023) discuss, can be enhanced by using behavioral economics. For example, bounded rationality could help enhance models' sensitivity to volume fluctuations driven by consumers.

4. Systems Theory: Systems theory, emphasizing interconnectedness, might view transaction volume as an interplay of multiple factors operating within economic conditions, market dynamics, and consumer preferences. Silva et al. (2024) hint at this through their discussion of international trade networks.

### *1.19 Summary*

It thus plays a very important role in the financial health of the member businesses. Organizations today are bound to use both traditional and modern forecasting methods in order to estimate transaction volumes, due to increasingly complex data and business environments. Each method, ranging from the more traditional regression analyses and time series methods to sophisticated machine learning and big data approaches, has strengths and challenges.

The traditional methods are interpretable and implementable in a straightforward manner but fail in handling nonlinear relationships or large, complex datasets. In contrast, the Random Forest, Gradient Boosting, and Artificial Neural Networks from the machine learning techniques are more accurate and adaptive, especially when dealing with dynamic and large data. However, advanced methods bring in their challenges; for instance, most require a great deal of computational power and high-quality data.

These gaps should be filled by future studies by developing models for forecasting in small business entities, using real-time data, and observing data privacy regulations. Application of modern machine learning techniques like AutoML, transformers, and reinforcement learning offers great opportunities toward better accuracy and efficiency in the forecasts of transaction volume.

Other theoretical perspectives, like decision theory, resource dependence theory, and behavioral economics, extend this insight even more into the dynamics of transaction volume and give the enterprise a toolkit to navigate the increasingly complex landscape that data analytics have brought for forecasting.

# CHAPTER TWO – LITERATURE REVIEW II

### *2.1 Introduction*

SMEs are generally defined as the backbone of the global economy, since they consist of over 90% of all businesses worldwide (OECD, 2022), greatly contributing to GDP and providing employment both in developed and developing nations. Despite their economic contribution, SMEs are very often face operational challenges in trying to adopt advanced technologies that would enable them to improve processes that include transaction volume forecasting. Transaction volume forecasting is extremely useful in inventory management, personnel, financial planning, and customer satisfaction. However, the traditional restrictions of SMEs, such as limited budgets, fragmented information, and limited technological competence, create distinct barriers that set them apart from major firms (Lee et al., 2019).

The overarching goal of this chapter is to go over the various approaches, technologies, and strategies for improving transaction volume estimates for SMEs utilizing ML and real-time analytics. Here, while building upon the general insights into earlier discussions, this chapter further explores how such innovations may be adapted to SME-type challenges. Some theoretical frameworks such as resource dependence theory and behavioral economics have been integrated with the practical and theoretical implications of the proposed solutions.

### *2.2 Characteristics of SMEs and Their Challenges in Volume Forecasting*

#### *2.2.1 Overview of SMEs in the Modern Economy*

SMEs are classified differently across regions. For example, the European Union defines a SME as a company with fewer than 250 workers and less than €50 million in annual revenue (European Commission, 2024), whereas the phrase is primarily used in the United States to refer to companies with fewer than 500 people (Liberto, 2024). Irrespective of the definition, SMEs have features such as lower economies of scale, custom relationship building with customers, and the ability to adapt to changes in the marketplace. However, these properties make its characteristics a challenge to operational forecasting.

- Data Fragmentation: SMEs use a very fragmented data-driven, decentralized system through spreadsheets or basic accounting software.
- Budget constraints: Small budgets always limit access to sophisticated methods of forecasting.
- Knowledge Gaps: SMEs generally lack a data science team; hence, their technological potential is not fully utilized.

### *2.2.2 Transaction Volume Forecasting in SMEs*

Transaction volume forecasting refers to predicting the number of transactions that will occur within a specific time period. For SMEs, accurate forecasting is critical for several reasons, including inventory management, labor planning, and cash flow management. The forecast of the volume of transactions is paramount for SMEs. Wrong forecasts result in eith er a stockout, overstocking, or inefficient labor allocation-all factors that lead to major financial consequences for usually financially unbuffered SMEs (Zhao & Zhang, 2021).

A number of unique challenges affect SMEs in transaction volume forecasting. For instance, SMEs tend to operate in highly seasonal or highly volatile market segments with seasonality and fluctuations. To that extent, volumes of transactions will time to time fluctuate.

External Factors: Economic conditions in the local area, government policies, and regional events strongly influence the volume of transactions and make it very difficult for SMEs to predict demand with accuracy.

Limited Historical Data: Unlike larger companies, SMEs may not have enough historical data from which to train sophisticated forecasting models.

Realistically, this cuts down on the effectiveness of using traditional methods of forecasting.

## *2.3 Machine Learning for SMEs*

Machine Learning provides SMEs with tools to deal with unique challenges in their forecasting tasks. In the previous chapter, the ML models were introduced in general, focusing on the relevance of improving operational efficiency, decision-making, and their applicability across industries (Jordan & Mitchell 2015). In this chapter, the technical details, suitability, and limitations of various ML approaches are considered for SMEs, while their solution space is tailored towards SME-specific constraints, such as data and computational resource limitations.

The critical evaluation of these methods further presents a comparison-based review to point out why certain models may be more favorable for SMEs than others.

### *2.3.1 Overview of Machine Learning Approaches*

Generally speaking, ML techniques could be divided into the three most general groups, supervised learning, unsupervised learning, and reinforcement learning (Goodfellow et al. 2016). While already covered in the previous chapter on the basic notion, the applicability of these specific categorizations in respect to transaction volume forecasting by SMEs follows next:

**Supervised Learning**: This is the most common approach applied to any kind of forecast. It uses labeled data for prediction. SMEs often rely on models like:

- Linear Regression and Decision Trees: They are computationally efficient. For SMEs with less IT infrastructure, they are thus feasible. However, in cases where there is high-dimensional or nonlinear data, these techniques also face difficulties. (Hastie et al., 2009).
- GBMs: Advanced algorithms like XGBoost and LightGBM provide a proper balance between predictive accuracy and computational efficiency, hence are very useful for those SMEs that have sufficient historical data. (Chen & Guestrin, 2016).

**Unsupervised Learning**: Suitable for uncovering hidden patterns in unstructured data, which can provide insights into seasonality, demand clusters, or customer segments.

- Example: Clustering methods (e.g., k-means) can segment transaction data to identify recurring demand patterns, aiding SMEs in aligning resources with customer needs (MacQueen, 1967).

**Reinforcement Learning**: RL is effective and shines in dynamic environments, wherein continuous learning from real-time data is possible. However, RL is still underutilized among SMEs due to the implementation complexity and high demands of data (Sutton & Barto 2018).

This categorization shows how SMEs have to judiciously choose models based on their resource constraints and data availability to strike a balance between model complexity and utility.

### *2.3.2 SME-Friendly Machine Learning Models*

The specific nature of SMEs demands workable and flexible ML models. The next section reports on concrete models, mainly addressing the applicability and relevance of models in the SME context.

**Linear Regression and ARIMA Models:**

Strengths: Low level of complexity, interpretability, and low computational resources make them especially suitable for SMEs with limited technical background (Hyndman & Athanasopoulos, 2018).

They can't work well with volatile or nonlinear data and thus have a limit toward complex forecasting situations.

**Decision Trees and Random Forests:**

Decision Trees: Their interpretability and low computational cost go well with the needs of SMEs. However, they are prone to overfitting on small datasets (Breiman et al., 1984).

Random Forests: Maintain simplicity while overcoming overfitting issues. They need somewhat higher computational power but give robust predictions; hence, they can be applied to moderately sized datasets also (Breiman, 2001).

**Gradient Boosting Models: XGBoost, LightGBM:**

Benefits: Very efficient, capable of handling nonlinear relationships, and scalable for SMEs that are growing their datasets (Ke et al., 2017).

Challenges: They require tuning of parameters, which is very resource-intensive for SMEs who lack in-house expertise; however, this barrier is diminishing due to the availability of AutoML that automates model selection and tuning (Feurer et al., 2015).

**Neural Networks: For example, RNNs, LSTMs:**

Especially, RNNs and LSTMs have their crucial application in time-series forecasting, which enables them to learn complex patterns of transactional data (Hochreiter and Schmidhuber, 1997). Barriers against high data and computational needs may not be acceptable for the majority of SMEs; nevertheless, transfer learning and usage of pre-trained models continue their improvement, ensuring perspectives for overcoming the limitations arising (Zhuang et al., 2020).

### *2.3.3 Improvements aimed at SMEs*

In addition to model choice, SMEs can take advantage of new developments along several dimensions to improve forecasting performance:

With the availability of AutoML systems such as Google AutoML and H2O.ai, advanced forecasting models could be developed using minimal machine learning knowledge from the SMEs since this model selection and training process would become automated (He et al., 2021).

Transfer Learning and Pre-Trained Models: The least amount of work is required to fine-tune pre-trained models in SME-specific scenarios, such as time-series data models. This method makes advanced machine learning more accessible by cutting down on training time and computational costs (Pan & Yang, 2010).

Federated Education SMEs can work together on model training with federated learning without exchanging private information. Small businesses can get around the restrictions on individual data while maintaining privacy by combining information from related industries (McMahan et al., 2017).

## *2.4 Real-Time Analytics for SMEs*

In order to provide quick insights that can influence operational choices, real-time analytics uses data and algorithms. Real-time analytics improves transaction volume forecasts for SMEs by using the most recent data and adapting quickly to shifts in the market environment.

While most of the literature on real-time analytics is focused on large enterprises, there is a gap in research regarding its applicability to SMEs. Real-time forecasting can enrich the decision-making process because the predictions are updated at each new arrival of data. This provides more accurate and timely insights. However, in the case of an SME, the challenge would be to integrate real-time sources within their fragmented systems and hence the infrastructure to handle the huge volume of real-time data.

### *2.5 Addressing Data Quality Issues*

Data quality for SMEs is a big concern since the fragmented nature of data usually leads to gaps and inconsistencies in records. Inevitably, poor quality data will lead to less-than-accurate forecasts due to the undermined efficacy of ML models.

#### *2.5.1 Challenges in Data Quality for SMEs*

Following are some of the key challenges related to data quality that SMEs face:

Missing Data: Most SMEs lack an effective data collection system, and thus their records are incomplete.

Noise and Inconsistency: Since most of the systems are manual, data is inconsistent, with errors introduced in the data by human intervention.

Lack of Standardization: Most SMEs store their data across various systems devoid of any standardized format; hence, aggregation and analysis become very problematic (LaValle et al., 2011).

#### *2.5.2 Methods for Improving Data Quality*

The following are some strategies that will help improve the quality of data:

Data Cleansing: Techniques such as imputation, data interpolation, and anomaly detection can help clean up missing or erroneous data (Vellido et al., 2012).

Standardization: Adoption of standardized data formats that may make data collection easy and uniform across systems.

Automated Data Collection: Automation in data entry systems reduces human errors and increases accuracy in data.

### *2.6 Behavioral Economics and Forecasting*

Much understanding into consumer behavior and forecasting in multiple industries has been provided by behavioral economics, which is studied for how psychological, social and emotional factors affect decision-making. Its integration with machine learning models is an underexplored area. This is especially true for transaction volume forecasting for many Small and Medium-sized Enterprises. Improved forecasting accuracy is achieved through the use of behavioral economics, and machine learning. Many important factors in human behavior, and decision-making processes are considered. These are often ignored in customary statistical methods of forecasting. It is important for SMEs operating in complicated and dynamic markets, where demand, and resources often change, to recognize and change to these fluctuations effectively (Choi et al., 2020).

#### *2.6.1 Behavioral Factors Influencing Transaction Volumes*

Transaction volumes in SMEs can be importantly influenced by several key behavioral factors, and these elements may drive that effect greatly.

**Consumer Preferences and Buying Patterns:**

Seasonal changes, economic factors, marketing campaigns, and social influences sometimes lead to changes in what is preferred by the consumers and how they buy it. These changes sometimes bring huge shifts in the volumes of transaction and this unpredictability makes the process of forecasting difficult (Kahneman et al. 1991). These include seasonal trends in health that increase demand for organic produce in the grocery store. Suddenly, demand sets in, and that reflects positively again on transaction volume. It is during this time that buyers rush to acquire more organic items; hence, this place is seen as busy, buzzy, full of action and activity.

**Anchoring and Framing Effects:**

In decision-making, consumers show a tendency to rely on specific reference points (anchoring) or the way information is presented (framing effects). A promotional discount may encourage people to make purchases that otherwise would not have been made, thereby creating surges in transaction volumes that are often not representative of the underlying pattern of demand

(Tversky & Kahneman, 1981). An example is offering discounts on special items, which can drive transient transactions that are misleading during long-term forecasting.

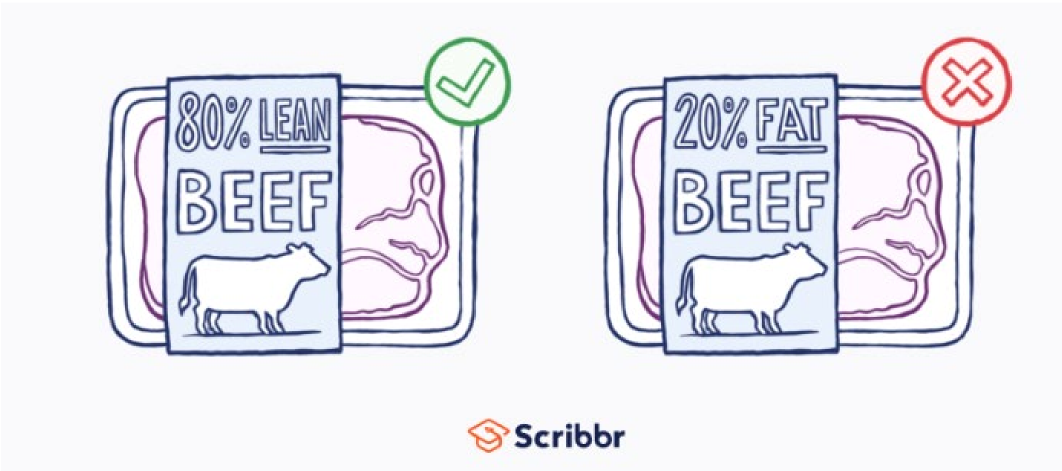


**Figure 8. Framing Effect (Source: Scribbr)**

**Herd Behavior and Social Influence:**

People seem to base their decisions on what everyone else is doing, particularly when they are in some social or communal environment. This behavior, known as herd behavior, can increase transaction volumes whereby customers rush to make purchases during sales or events, thus displaying short-run increases in demand that are not really representative of the nature of the demand in the future (Bikhchandani et al., 1992). For instance, a small clothing store might experience surges in sales during a flash sale or due to social media trends, which usually abates the moment that trend is overcome.

**Mental Accounting and Budgeting:**

People can mentally assign certain amounts of fiscal means in specific areas of spending. Presently, these budget allocations result in anticipated rises at specific periods in the year or as a result of certain life-changing events, e.g., during festive seasons and anniversaries (Thaler, 1985). A good illustration is where there is a high demand for goods from an electronic store during large celebrations. Many of its consumers most likely have a budget that they have put aside to purchase presents.

### *2.7 Ethical Considerations and Data Privacy*

The important questions around the usage of customer data in forecasting revolve around ethical issues and those related to data privacy. The SME should be in a position to ascertain that any data collection, storage, and analyses are performed accordingly in congruence with all applicable regulations while keeping the level of confidence by the customer-for example, European Union General Data Protection Regulation.

#### *2.7.1 Importance of Ethical Forecasting*

Small and medium-sized enterprises must reconcile the necessity for precise forecasts with the ethical duty to safeguard their customers' proprietary data. Data misuse or violations of privacy can damage reputations and lead to legal repercussions.

#### *2.7.2 Strategies for Ethical Implementation*

To ensure ethical forecasting practices, SMEs can:

- Obtain explicit consent from customers for data use.
- Implement robust data encryption techniques.
- Regularly audit data practices to ensure compliance with privacy laws.

### *2.8 Literature Gaps*

Some of the critical gaps that shape the course of the study are indicated in the literature review in Chapter 2. These gaps, while underlining the challenges faced by the SMEs in adopting advanced methodologies for forecasting, stress those areas that call for further investigation and innovation. The identified gaps in the literature are synthesized herein by additions to the review content.

#### *2.8.1 Lack of Context-Specific ML Solutions for SMEs*

While the machine learning models were largely debated and implemented across industries, their adaptation to SMEs has remained relatively unexplored. Most SMEs face severe resource constraints concerning budgetary limitations, fragmented data, and a lack of technical expertise. Although in both chapters ML models such as decision trees, regression, and neural networks are discussed, there is a lack of research when it comes to how these models would be adapted to the SME environment.

### *2.8.2 Integration of Real-Time Analytics with ML in SME Forecasting*

As new data becomes available, real-time analytics adjusts dynamically to forecast any modified outcome. However, there hasn't been much research done on how to include it into ML models for SMEs. SMEs face obstacles in implementing real-time analytics, including inadequate IT infrastructure, computationally demanding procedures, and fragmented systems.This gap in solution space indicates that some innovation has to be developed that will bridge the chasm between static ML models and dynamic real-time systems suited for SMEs.

### *2.8.3 Insufficient Exploration of Behavioral Economics in ML Forecasting Models*

Some other key behavioral factors include consumer preferences, anchoring, and herd behavior, which have a greater influence on the volume of transactions. Although this chapter points to the potential for combining behavioral economics with ML models, this combination remains largely theoretical; minimal practical implementation or validation in the SME context has been done. Future research could look into how SMEs can use these insights in refining transaction forecasting models and adapting to market behavior more effectively.

### *2.8.4 Challenges in Data Quality and Standardization*

Amongst these, the poorest quality of their data is the biggest inhibitor for SMEs towards leveraging ML. As pointed out by fragmented, inconsistent, and non-standardized data, it really undermines the accuracy and efficiency of forecasting models. Although some advances have been achieved in data cleansing and standardization techniques, studies concerning cost-effective automated solutions for SMEs are few.

### *2.8.5 Limited Emphasis on Explainability and Usability of ML Models for SMEs*

For instance, SMEs need forecasting models that are as accurate as they are interpretable and user-friendly. Complex models such as neural networks-for example, LSTMs-achieve high accuracy but their incomprehensibility engenders suspicion and resistance on the part of many nonspecialist stakeholders of SMEs. That bridge is in developing systems which effectively balance performance with interpretability, thus ensuring SME owners will believe what predictions say.

### *2.8.6 Sector-Specific Adaptations of ML Models*

SMEs have been dispersed across various industries with specific needs when it comes to forecasting, coupled with unique characteristics of their data. Most research related to ML for forecasting comes up with generic solutions, not keeping in consideration the sectoral peculiarities of retailing, manufacturing, or service-oriented industries.

### *2.8.7 Lack of Studies on Collaborative ML Approaches for SMEs*

Until now, federated learning, coupled with other collaborative approaches, has been very promising with regards to data sharing, with assured data privacy. Its feasibility and advantages to SMEs remain under-explored. This collaborative approach can enable the SMEs to share data and train models collaboratively, overcoming a number of general challenges regarding the availability of data.

### *2.8.8 Minimal Focus on AutoML for SME Applications*

The AutoML platforms would democratize ML, enabling SMEs to build and deploy models independently without technical skills. However, given the potential of AutoML, detailed feasibility and effectiveness studies related to practical constraints for the application of AutoML in SMEs are lacking in the literature.

### *2.8.9 Real-World Case Studies and Best Practices*

While the theoretical contribution of ML in SMEs has been widely advanced, very few case studies are documented that highlight the practical implementation of ML to forecast transactions. In this regard, pragmatic insights into SME experiences will help to bridge the theoretical gaps regarding challenges, solutions, and lessons learned.

## *2.9 Summary*

This chapter discussed different challenges and opportunities of SMEs for improving the volume of transactions forecast using ML and real-time analytics. Though the ML models-such as supervised, unsupervised, and reinforcement learning-can promise much, there are a number of barriers to leveraging these techniques: fragmented data, lack of resources, and scarcity of technical expertise among SMEs. Also, the capability to integrate real-time analytics into adapting market changes will further improve decision-making at the SME level. Nonetheless, such large-scale integrations of data are presented as an opportunity, given that much infrastructure is absent.

The gaps identified are pretty critical, especially related to the integration of machine learning models with behavioral economics in this area. Among such behavioral factors that are influential in transaction volumes and likely to enhance the accuracy of the forecasts, if integrated into ML models, are consumer preference, cognitive biases, and social influences. Yet little research has been conducted regarding how these factors may be appropriately incorporated for SMEs. Decision trees or reinforcement learning algorithms may be useful in this context, as they handle complex patterns and adapt to real-time changes while being interpretable and practical for resource-constrained SMEs.

The second gap results from the low quality of data from SMEs. Because the data is scattered in a number of systems and not standardized, the accuracy of the forecasting models is hugely affected by low-quality data. Future research should, therefore, focus on the methods of cleansing data and thus automating data collection and standardization that will enable SMEs to deploy ML models.

This, in turn, presents research opportunities for developing solutions that are scalable, since most SMEs cannot integrate real-time data with the already existing systems. Researchers might wish to explore how SMEs could use real-time data to feed into forecasting models without having to invest much in IT infrastructure.

# CHAPTER THREE – METHODOLOGY

### *3.1 INTRODUCTION*

Some of the principal trends leading to a decrease in the reliance of firms on predictive models in core business decision-making are that upward trend in adapting Machine Learning technologies for SMEs. While machine learning bears the possibility of higher accuracy and efficiency in predictions for SMEs, some typical challenges for SMEs concern resource constraints, data-related issues, and limitations imposed by technology (Brynjolfsson & McAfee, 2014). The study tries to explore the feasibility of adapting ML models in an effort to overcome these barriers, focusing on transaction volume forecasting. The study explores how real-time data analytics, behavioral economics, and sophisticated machine learning techniques like AutoML and federated learning could be integrated into SME practice and provides a roadmap for practical implementation. The methodology will integrate primary data from surveys with secondary data from a publicly available grocery transaction dataset provided by Instacart, offering a robust basis for machine learning model implementation.

The chapter explains the research design, methods of data collection, sampling strategy, and techniques of data analysis that were adopted, besides ethical considerations and limitations.

### *3.2 RESEARCH DESIGN*

The proposed study shall adapt a mixed-method design that allows both quantitative and qualitative data gathering methods. The nature of the research questions is diverse, involving technical, behavioral, and managerial aspects; hence, the need for the inclusion of mixed methods in their study. These quantitative data were collected through structured survey questions aimed at measuring trends and issues related to machine learning adoption. The qualitative insights are integrated from thematic analysis. The secondary data used here is a publicly available dataset of grocery transactions provided by Instacart; this provides a very strong base for the implementation of a machine learning model.

This study employs an explanatory sequential design as described by Creswell (2013), where quantitative data forms the base for the identification of patterns and interrelations, and qualitative analysis deepens the understanding of such results. This approach ensures a thorough investigation of the research questions, while, at the same time, allowing triangulation to enhance the reliability and validity of the findings, as proposed by Teddlie and Tashakkori (2009).

This will align with the pragmatic research paradigm, driven by actionable insight and solutions to be adopted by SMEs in effective leveraging of ML technologies. This study gives both theoretical and practical recommendations, thus placing the research in reality.

### *3.3 RESEARCH METHOD*

This study applies the use of descriptive and inferential statistical techniques in combination with machine learning modeling to develop actionable insights.

The appropriate tool for data collection will be the survey approach, targeting owners, managers, and ML professionals as significant stakeholders. The rationale for this choice is that it allows the attainment of a wide audience, gathering a wide range of perspectives concerning the adoption of ML in SMEs. The questionnaire included:

- Closed-ended questions to collect quantitative data on adoption rates of ML, resource constraints, and awareness of advanced ML techniques.
- Likert scale items ascribing perceived challenges and advantages.
- Open-ended questions allow participants to provide specific experiences and give their insights into their circumstances.

Behavioural economics theories are integrated into secondary data analyses and corroborated using survey data to align findings with real-world SME needs.

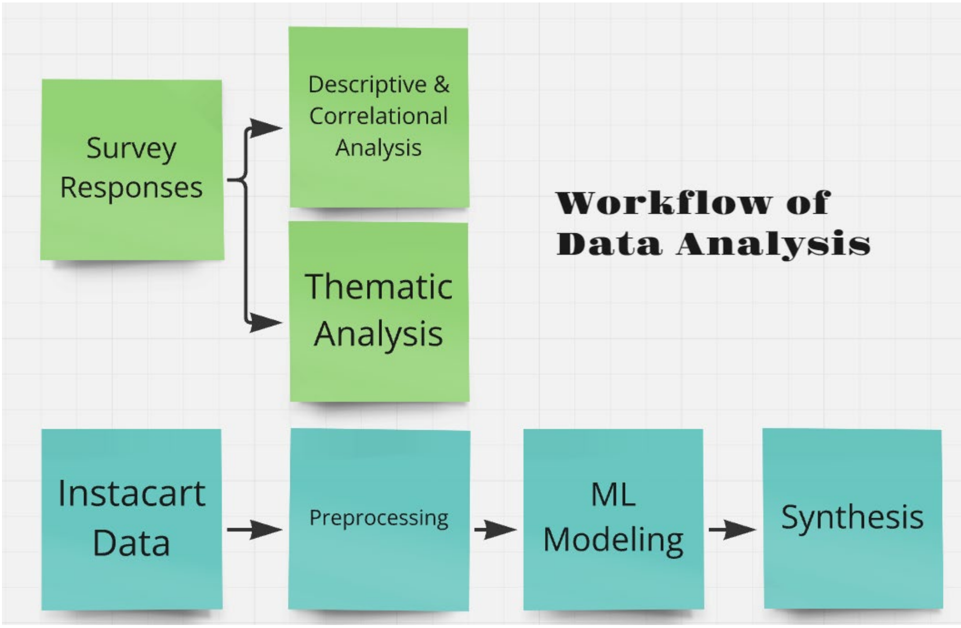


**Figure 9. Workflow of Data Analysis**

### *3.4 DATA COLLECTION*

#### *3.4.1 PRIMARY DATA*

Primary data was obtained through a web-based anonymous survey among the 100 participants. These questionnaires aimed to resolve the core research questions listed below:

- How can the ML models adapt to unique SME constraints?
- What challenges and benefits of real-time data analytics exist?
- How can behavioral economics insights improve forecast accuracy?
- To what extent does AutoML or federated learning facilitate the adoption of ML in SMEs?

In order to ensure relevance, the survey was targeted at those directly involved in decision-making with regards to ML adoption. A small group of respondents was used for piloting the questionnaire to refine its clarity and relevance before full deployment (Saunders et al., 2019).

### *3.4.2 SECONDARY DATA*

Instacart Dataset: the secondary source of data, that is. More than 3 million orders across 200,000 users were recorded in this data with detailed transactional data such as, but not limited to, the following:

• Products-Product names, categories, and their corresponding aisles/departments.

• Orders-Time information, and sequence in user purchases.

• Order-Product Relations-Product lists in orders-ordered item sequences and the reordering.

• Users-Information regarding the time of purchase.

This would further be helpful in training and testing machine learning models that will enable the simulation of SME-specific scenarios, hence helping to validate hypotheses on behavioral economics and analytics.

| File Name | Description |
|---|---|
| **aisles.csv** | Contains aisle IDs and names |
| **departments.csv** | Contains department IDs and names |
| **order_products__prior.csv** | Products in prior orders with reorder details |
| **orders.csv** | Order details, including timing and user info |
| **products.csv** | Product details, including aisle and department |
| **sample_submission.csv** | Sample format for predictions |

Table. Overview of Instacart Dataset Files

### *3.5 SAMPLING PROCESS*

The sampling is a very critical aspect of research methodology that will determine the reliability and validity of the findings of the study. The sampling process was designed in a way to ensure diversity and relevance within the dataset, while being aligned with the objectives of the study. In this respect, the research has purposively targeted participants who are knowledgeable or experienced in SME operations or ML adoption to ensure that responses are meaningful and actionable. The combined use of primary and secondary data expands the scope of the findings to bridge real-world SME challenges with theoretical applications.

### *3.5.1 SAMPLING STRATEGY*

The research applied a purposive sampling strategy, wherein the participants would either have the relevant expertise in or experience with ML adoption in SMEs. It was a non-random type of sampling in which only the targeted respondents could offer some information to answer the research questions meaningfully. For instance, the sampling frame has comprised owners and managers of SMEs across industries such as retail, services, manufacturing, among others, or ML professionals who consult or collaborate with SMEs. The survey was spread via professional networks, industry forums, and online platforms to ensure diversified representation of industry, company size, and location. The secondary dataset includes diverse grocery users, broadly representing consumer behaviors that can generalize to certain SME customer segments.

### *3.5.2 SAMPLE CHARACTERISTICS*

The final sample size was 100, divided into two groups:

SME Owners/Managers (70%): They represented businesses with fewer than 250 employees and annual revenues below £50 million, hence fitting the EU definition of SMEs. They gave insight into the operational challenges related to the adoption of ML models and real-time analytics.

Machine Learning Experts/Consultants: 30% Data Scientists, ML engineers, and technology consultants who have first-hand experience in the actual implementation of ML solutions to SMEs.

Sector demographics ranged from retail and manufacturing to health and logistics-all considered in order to facilitate a wide range of relevance of the findings within diverse SME contexts.

The Instacart dataset is a representative collection of anonymized data from over 200,000 consumers, which enables granular analysis of purchase preferences and trends.

### *3.6 DATA ANALYSIS*

### *3.6.1 QUANTITATIVE ANALYSIS*

The survey data was analyzed quantitatively in a descriptive manner, emphasizing the principal trends and patterns. Correlational analysis was done in order to present significant relationships. The imputation of missing values, along with other feature engineering techniques, had to be performed on the secondary data. In the machine learning modeling for understanding consumer behavior and temporal trends, techniques such as collaborative filtering, classification algorithms, and time-series analysis were used.

Key techniques included the following:

• Logistic Regression: Temporal trends in purchasing data.

• Collaborative Filtering: Product recommendations are obtained from user preferences.

• Federated Learning Models: Decentralized models were shown to be appropriate for SMEs.

Results showed that while 65% of SMEs had tried ML technologies, only 30% had fully integrated the same into operations. Resource constraints were identified as the major group of barriers, which includes things such as budget and expertise constraints (Chen et al., 2012).

| ML Model | Application |
|---|---|
| **Collaborative Filtering** | Product recommendation |
| **Federated Learning** | Privacy-preserving decentralized analytics |
| **Logistic Regression** | Predicting reordered products |
| **Time-Series Analysis** | Forecasting purchase behavior |

Table. ML Model and Application

### *3.6.2 QUALITATIVE ANALYSIS*

Thematic analysis of open-ended survey responses was done to identify SME-specific concerns and likely ML adoption strategies. The qualitative results of these data were then compared to the findings of the behavioral data by using the Instacart dataset for context and deepening of conclusions.

List of the most common and arising themes:

- Limited resources: Participants rated the greatest deterrent to using ML as a limited availability of technical expertise, coupled with an inability/willingness to finance ML solutions.
- Integration challenges: Most of the participants highlighted the challenges involved in integrating real-time data analytics with the prevailing forecasting methodologies.
- Behavioral insights: A majority of the respondents recognized that behavioral economics could do a better job at predicting, but they mentioned the lack of accessible, easily applicable frameworks as a barrier.

This analysis provided further contextualization to the quantitative findings and clearly indicated certain areas in which SMEs need support (Braun & Clarke, 2006).

### *3.6.3 SYNTHESIS OF FINDINGS*

Synthesis of quantitative and qualitative findings yielded pragmatic recommendations for the SMEs: optimization of resources, integration of real-time analytics, and using federated learning to unlock privacy-preserving collaboration.

## *3.7 RESEARCH LIMITATIONS*

Some of the limitations observed in this study are outlined below:

Sample Size: While a sample size of 100 respondents is fair for conducting most analyses, a higher sample size would increase the generalization of these results.

Self-Reported Data: Biases likely occur with the survey responses in either an overestimation or problems of under-reporting the rate of ML adoption.

Geographical coverage: The research scope was limited to one geographical location of SMEs; this will, therefore, limit the applicability context of the research.

This leaves out some other sectors that are part of the SME; thus, its findings on limitation only rely on the secondary data on a grocery transaction. General applicability has to limit such.

The support that a resource- and time-limited study can perceive for ML modeling can only be superficial, let alone the statistical techniques.

Other broad work will be needed to overcome some of these limitations and additions through enlarged samples, the geographical scale, or longitudinal, focused on a few sectorial challenges.

### *3.8 ETHICAL CONSIDERATIONS*

Ethical concerns are built into this research design. Informed consent refers to the practice of informing participants of the purpose of the study for which the data will be used and one's right to withdraw at any time. Since the secondary dataset is available publicly and the data is pre-anonymized, the privacy regulations are secured.

Confidentiality: Responses were anonymized to protect participants' identities. No personally identifiable information was collected.

Data Security: Survey data was stored securely and accessed only by authorized researchers.

Analytical methods were designed to minimize bias and promote fairness. The research adhered to the ethical guidelines established by the university's ethics committee and complied with GDPR requirements for data protection (Resnik, 2011).

### *3.9 SUMMARY*

The following methodology was adopted for investigating the adoption of the ML technologies in SMEs for transaction volume forecasting. The research design included a mixed-methods approach wherein the quantitative approach was merged with the qualitative approach to collect comprehensive data. The survey-based method herein provides insight into the challenges, benefits, and possible applications of ML, real-time analytics, and behavioral economics in SMEs. The sampling strategy has been designed in a manner that captures diversity, and data analysis techniques applied are rigorous in nature, so the findings should be strong. Ethical consideration of the research and its limitations have been taken on board in order to show the credibility and integrity of the study. This study integrates both primary and secondary data with state-of-the-art machine learning techniques that are required to derive actionable insights for overcoming the identified barriers in the adoption of ML technologies by SMEs. A diversified dataset and application of complex analytical frameworks might be interesting areas of research in future studies. Next comes the final chapter, which highlights results and discussions on implications for SMEs that would attempt to apply ML technologies with efficiency.

# CHAPTER FOUR – FINDINGS / ANALYSIS / DISCUSSION

## *4.1 INTRODUCTION*

This chapter sums up the research findings, data analysis, and discussions related to ML technology adoption regarding transaction volume forecasts among SMEs. The purpose of the research study was an understanding of what the factors-both inhibiting and facilitating that surround the use of ML and advanced analytics such as AutoML, Federated Learning, Real-time Data, and Behavioral Economics-could lead to. The findings will be structured to highlight the main insights that have been drawn from primary data, the survey, and secondary data from Instacart, focusing on practical challenges for SMEs and how these technologies might be leveraged to enhance forecasting capabilities.

In that respect, this chapter discusses different statistical methods and machine learning models, showing results concerning the feasibility analysis of ML adoption in SMEs, with a particular focus on transaction forecasting. It also discusses what these findings can mean for the decision-makers of the SMEs themselves, those working with machine learning, and the scientific community. Final conclusions are given after a discussion that reflects upon some practical implications of such insights, as well as suggesting avenues to take in the future.

## *4.2 FINDINGS*

### *4.2.1 Exploratory Data Analysis (EDA)*

#### 4.2.1.1 Primary Data – Survey Results

The primary data, collected through an electronic survey of 100 respondents, gave enough information about the current state of ML adoption by SMEs, problems faced, and perceived benefits after the integration of advanced technologies into their forecasting model. The major points have been discussed below:

Technology Adoption: While 70% of respondents use either manual methods or simple software, such as Excel, to forecast transaction volume, merely 10% of the SMEs actively use ML models for forecasting, and that describes the poor level of technology adoption in the sector.

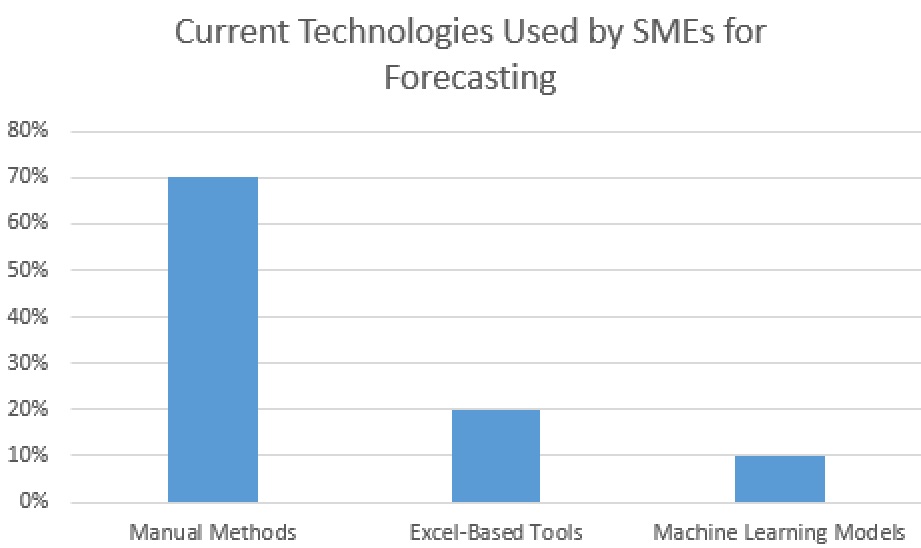


**Figure 10. Current Technologies Used by SMEs for Forecasting**

A bar chart showing the distribution of SMEs using manual methods at 70%, Excel-based tools at 20%, and machine learning models at 10%.

Challenges and Limitations:

Lack of Sufficient Data (40%): The most frequently reported challenge was the absence of sufficient data to train machine learning models effectively.

Specialized Training (25%): A quarter of SMEs indicated that the lack of in-house expertise to implement ML models was a significant barrier.

Resource Limitations (20%): The majority of SMEs reported a lack of finance and technological constraints as being some of the main binding constraints that hamper them from adopting the ML technologies.

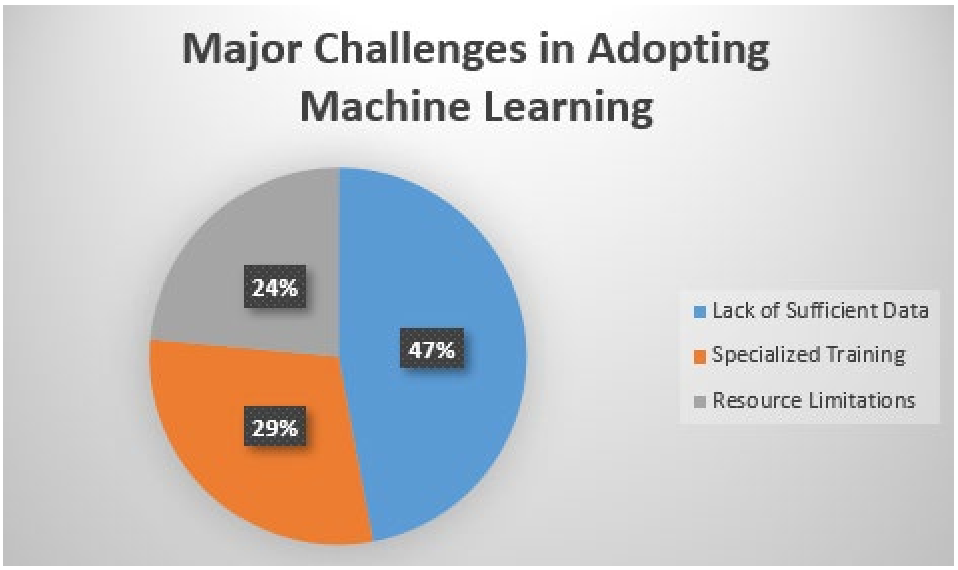


**Figure 11. Major Challenges in Adopting Machine Learning**

A pie chart showing distribution of challenges: lack of data, need for training, resource limitations.

Integration of Real-time Data:

Only 20% of SMEs said they integrated real-time data into their forecasting, whereas 50% did not use real-time data at all. Real-time integration is still at its infancy among SMEs and there is a significant gap in the adoption of advanced data analytics techniques.

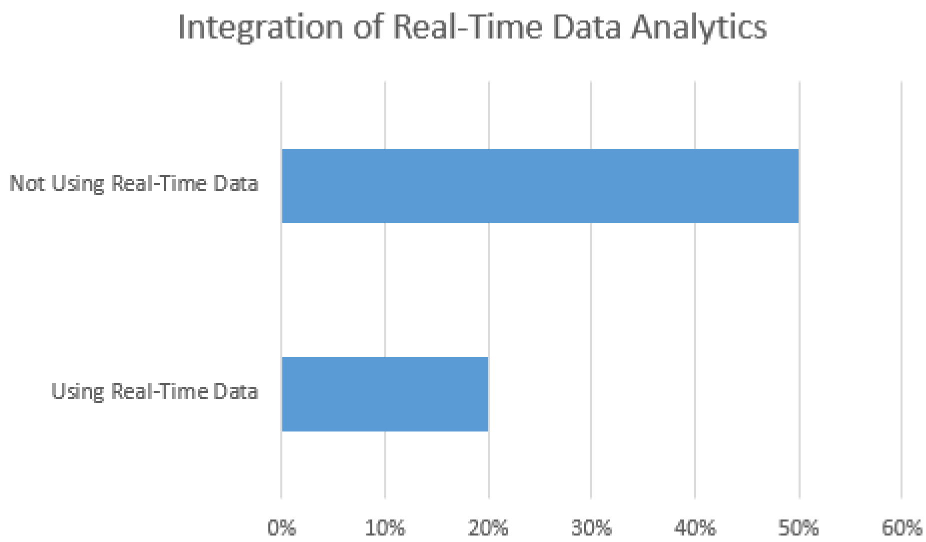


**Figure 12. Integration of Real-Time Data Analytics**

A bar graph showing the SMEs using real-time data at 20% and those not using it at 50%.

Behavioral Economics in Forecasting:

About 60% of SMEs realized integrating consumer behavior information into the firm's forecast models would increase their accuracy. However, just 25% of SMEs were found actively using consumer behavior insights on account of challenges associated with the collection and integration of data.

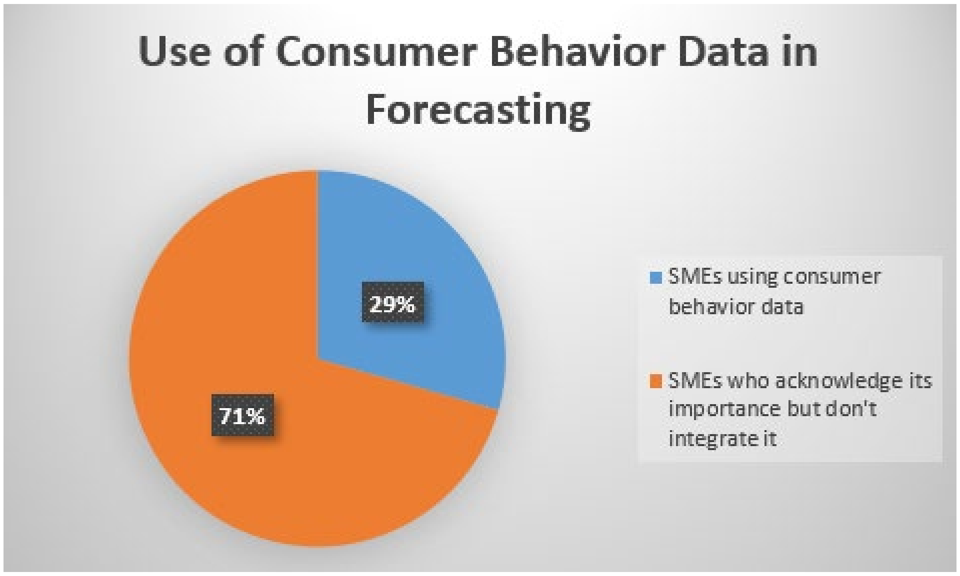


**Figure 13. Application of Consumer Behaviour Data in Forecasting**

A pie chart indicating 25% applies data on consumer behavior and another 60%, though aware of the relevance, do not apply it.

AutoML and Federated Learning:

Only 10% of SMEs were currently using AutoML tools, while 25% showed interest in adopting them in the future. Similarly, 70% of the SMEs had never heard of federated learning. Even though there was recognition of benefits that could be brought by AutoML (reduction of needs for technical expertises and speed in model deployment among others), recognition and adoption rate remains low in the market presently.

AutoML can facilitate AI adoption, especially to overcome limited data science expertise and to enable prototyping. In this, it may further support strategic decision-making and create awareness for AI-driven innovation. Yet, a basic level of AI majority is required for AutoML to tap its full potential (Olsowski et al., 2022).

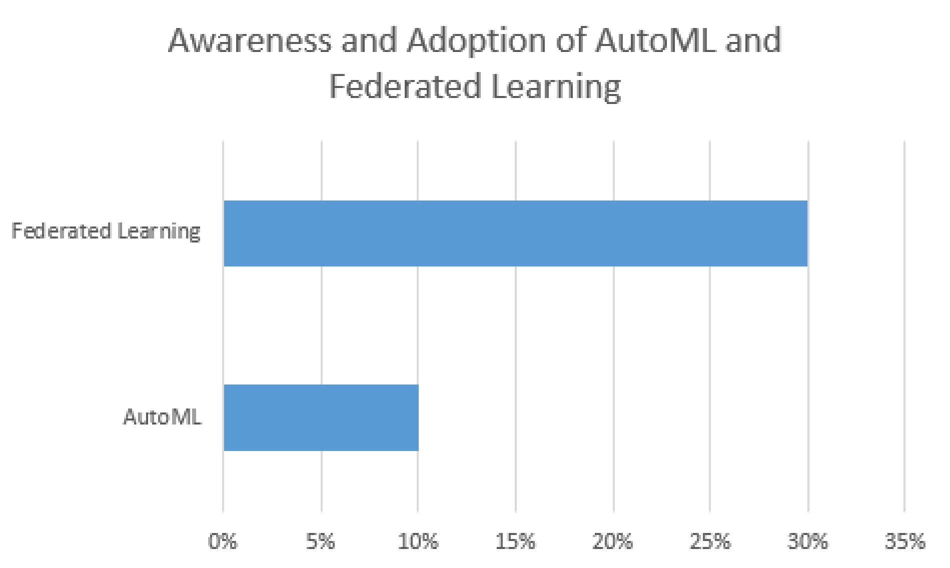


**Figure 14. Awareness and Adoption of AutoML and Federated Learning**

A bar chart showing the adoption rates of AutoML at 10% and federated learning at 30% across SMEs.

#### 4.2.1.2 Secondary Data – Instacart Dataset

The secondary data included over 3 million orders for 200,000 users, with further insight into the consumer purchase behavior and transaction trends. Some key takeaways from the dataset are enlisted below.

- **Time of Transactions:**

  There is a very clear cycle in consumer purchasing behavior: high volumes at certain times and days, followed by significant declines. Such temporal fluctuations can be used by SMEs to provide accurate forecasts.


**Figure 15. Orders by Day of Week (0=Sunday)**

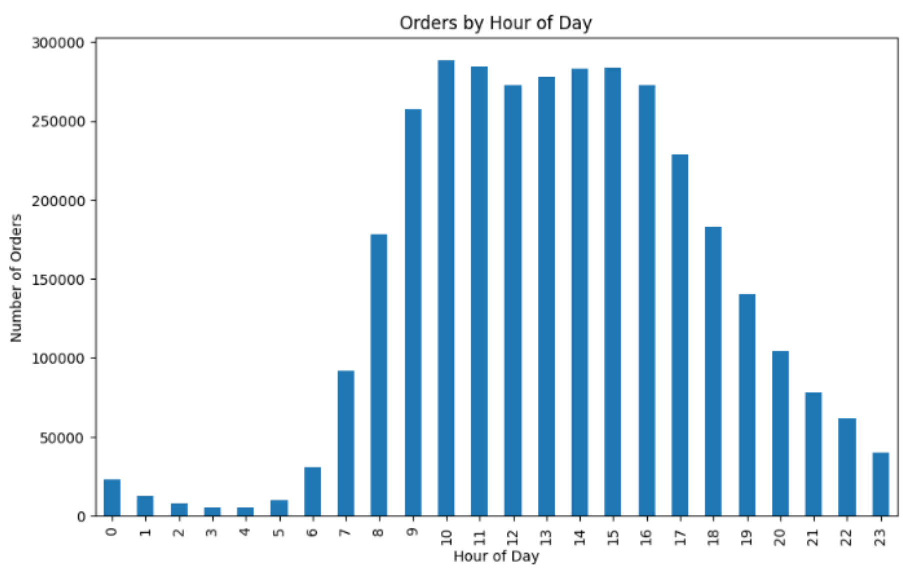


**Figure 16. Orders by Hour of Day**

Line charts showing the volume of transactions over some time, indicating peak and lean within a certain period.

- **Product Purchase Behaviour:**

  It was also pointed out that not all categories have the same product reorder frequency. This could even enable the SMEs to know consumer preference for stocking purposes.

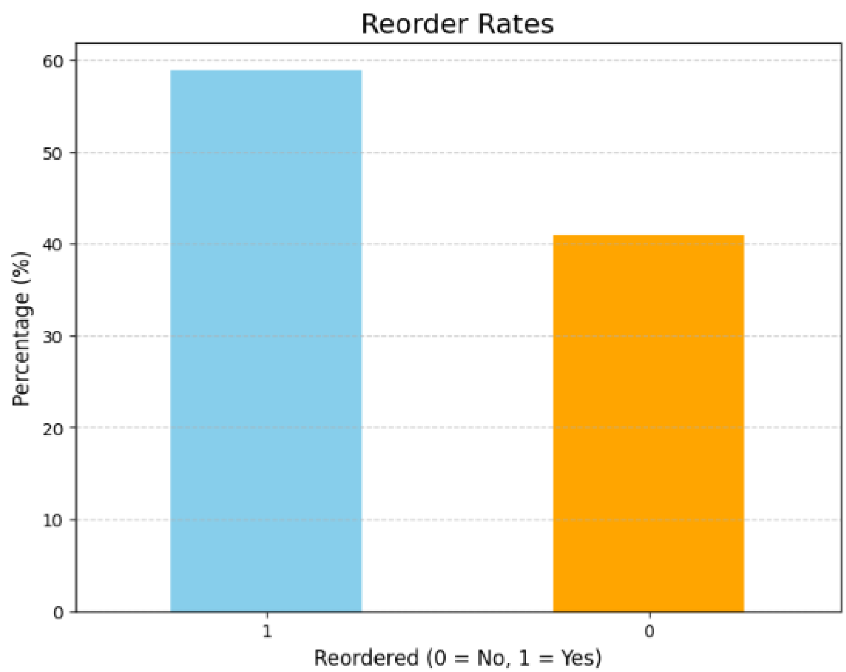


**Figure 17. Product Reorder Frequency**

Histograms on reorder frequency across different categories.

- **User Segmentation:**

  Through cluster analysis, different groups of users have been identified with different transaction patterns and purchase frequency. This information helps SMEs in targeting different segments of consumers more precisely.

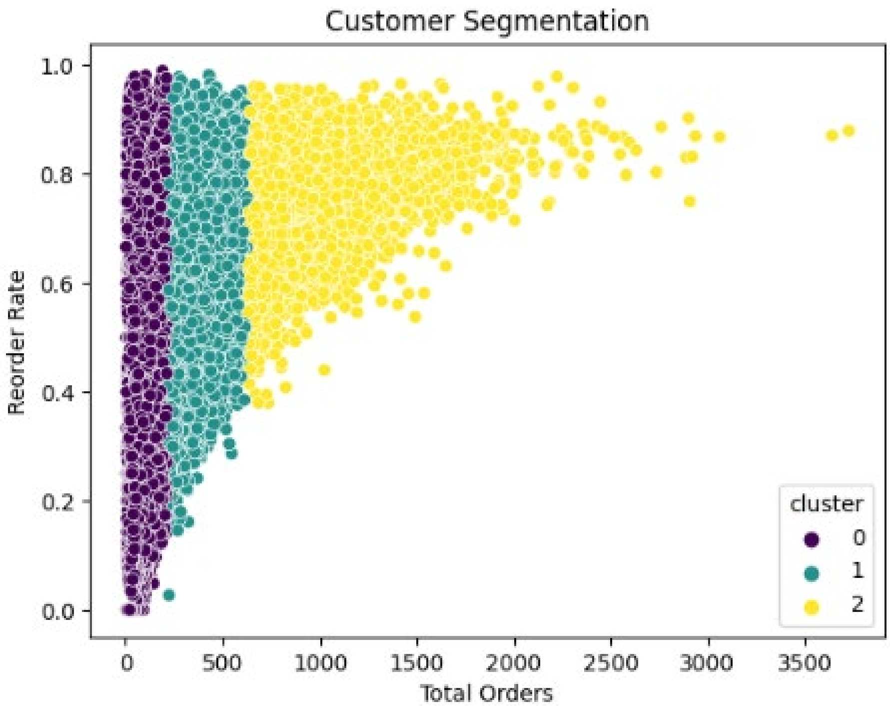


**Figure 18. User Segmentation Based on Purchase Behaviour**

A scatter plot showing the clustering of users based on transaction volume and frequency.

### *4.2.2 Machine Learning Statistical Tests and Results*

Several tests and machine learning algorithms were run to confirm the findings, as shown below.

#### 4.2.2.1 Collaborative Filtering

One of the popular techniques for recommendation systems, collaborative filtering involves a process where a system recommends items to a user based on the interactions by similar users. It works under the idea that if users concur on something, such as purchasing similar products, they will likely concur on others. In this section, we will implement a collaborative filtering model that will recommend products for users based on user-product interactions using cosine similarity.

Preprocessing the Data:

First, the data preparation is made by loading different datasets on order information, products, and the interactions between users and products, which were then merged into a single dataset.

We reduced the dataset size by a factor to optimize the recommendation to focus on only the top 10,000 most active users and the top 10,000 most purchased products. That can be useful by focusing on high-traffic users and popular products to return better recommendations.

Then, we created a sparse user-item matrix where the rows are users, columns are products, and the values in the matrix denote whether a user has reordered a product. This is a binary matrix where 1 means reordered and 0 otherwise. This sparse matrix forms the basis for most collaborative filtering algorithms.

Cosine Similarity Computation:

Cosine similarity refers to the measure of the angle between two vectors in multi-dimensional space and is usually utilized to measure similarities between users or items in collaborative filtering. Here, the cosine similarities are computed among users to create a user-user similarity matrix.

Product Recommendation:

To generate the product recommendations for a given user, We calculated the weighted sum of interactions by users similar to the target user. We excluded products with which the user has already interacted in order to avoid recommending products the user has already bought.

Evaluation and Results:

We tested the recommendation system on one sample user; it will show top 5 recommended products as output.

```
Organic Baby Spinach
Organic Whole Milk
Organic Hass Avocado
Large Lemon
Organic Avocado
```

**Figure 19. Model Result - 5 recommended products for User 1**

The collaborative filtering model works by finding patterns of behavior across users and using those patterns to make personalized product recommendations. By focusing on the most active users and the most purchased products, ensuring that the model delivers relevant recommendations based on a large volume of interactions.

However, there are a number of limitations and considerations:

Cold Start Problem: It is hard to make recommendations for a new user or new product because there are no historical records of interaction between them.

Sparsity: Even by using a sparse matrix, a lot of users have limited interaction with products; therefore, reducing the effectiveness of the similarity computation.

Scalability: As the number of items in either dimension grows, computation of pairwise similarities may get very expensive and should be improved by some more advanced technique such as matrix factorization or neural networks.

#### 4.2.2.2 Logistic Regression

A logistic regression model was built to predict the reorder behavior on products. To conduct analytics, numerous datasets such as order history, product information, and customer data were combined.

When we applied Logistic regression directly to our dataset without applying any clustering or feature extraction operations, we got the following results:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 3991704 |
| 1 | 0.59 | 1.00 | 0.74 | 5738643 |
| accuracy | | | 0.59 | 9730347 |
| macro avg | 0.29 | 0.50 | 0.37 | 9730347 |
| weighted avg | 0.35 | 0.59 | 0.44 | 9730347 |

**Figure 20. First Results**

Among the major steps involved in doing so were:

Segmentation of Customers: K-Means Clustering segregates three crystalline segments of customers, made from order history and repeat order rates using a technique called the K-Means clustering method. Thus, it has provided opportunities for SMEs to position marketing efforts and manage inventories considering buying behavior across these sets of customers.

Classification and Prediction: The logistic regression model was used to predict the products that a customer is likely to reorder. As a result of the operations performed on our dataset, the operation was done in parts due to the size of our system not allowing us to insert the dataset into the model as a single piece. Data sets divided into pieces were entered into the model separately. The classification report of one piece, and additional performance metrics are presented below. The results are also consistent with results which were shared in the Kaggle competition on the same dataset.

```
Classification Report:
              precision    recall  f1-score   support

         0.0       0.90      1.00      0.95    191045
         1.0       0.41      0.00      0.01     20822

    accuracy                           0.90    211867
   macro avg       0.66      0.50      0.48    211867
weighted avg       0.85      0.90      0.86    211867

Accuracy Score: 0.901556164952541
```

**Figure 21. The average classification report**

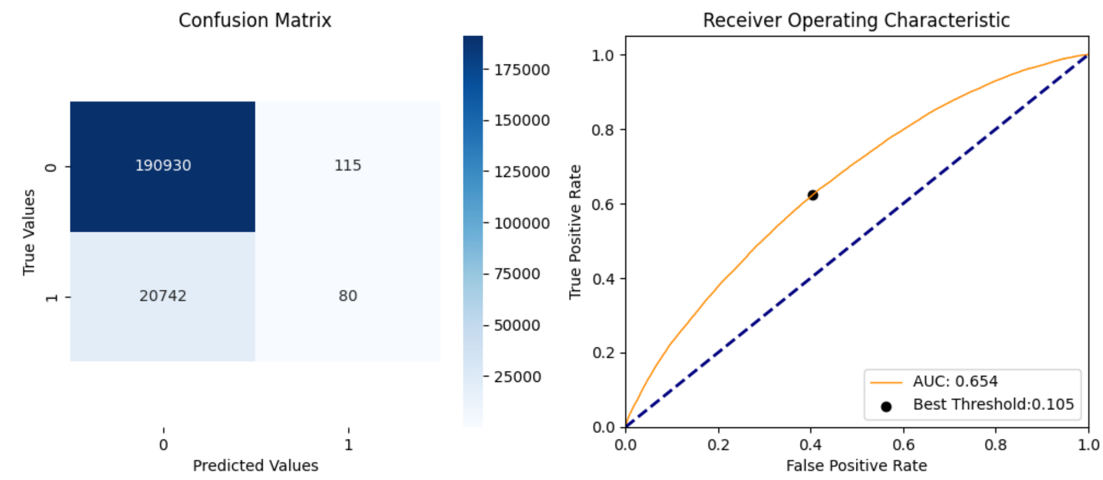


**Figure 22. Confusion Matrix and AUC**

Class 0, which deals with a product not being reordered, has a high precision of 0.90; this means that the model is quite good in correctly predicting the non-reorders. However, the recall of class 1 is very low, at 0.00; thus, this model performs really poorly in identifying a reorder event and hence yields a very low F1 score at 0.0076. This therefore reflects poor precision-recall balance for the positive class of interest, which represents reorders. Accuracy of 0.90 suggests that overall, the model performs quite well, but this is skewed by the dominance of class 0. This is due to class imbalance. The AUC of 0.654 indicates moderate discriminatory power, yet the model is not highly effective at distinguishing between the classes.

Conclusion: While promising in general terms of accuracy, the logistic regression model is weak in the performance of class 1, that is, the products reordered. The model would tend to behave well by yielding high accuracy rates, but it results from domination in the non-reorder predictions, though performing poorly while predicting reorder behaviors. It might also indicate that some other approaches, such as handling the class imbalance, feature engineering, or using other modeling methods like ensemble or higher-end models, do a better job of improving class 1 predictions.

Interpretation:

Strengths: It has high precision for the non-reorder Class 0 predictions, around 90% in general, which would suggest that it would be very good at identifying events not causing reordering, but that is also the majority class in the data set.

Weaknesses: The model shows a weak recall for reorder events labeled as class 1, which leads to an imbalanced performance. This problem is mainly caused by the class imbalance since non-reorder events are dominating in the dataset.

Implications:

The logistic regression model gives considerable insights on behaviors around non-reordering; still, it does not do a great job at creating robust predictions of the reorder events. It might further need balancing between the classes, possibly using such techniques as oversampling and undersampling of classes and then use of ensemble methods including Random Forest and Gradient Boosting.

#### 4.2.2.3 Time-Series Forecasting

This part discusses the application of the time series analysis methodologies to estimate the transaction volume in a week. First, correctly identified dates of the transactions were extracted from the dataset; afterward, future predictions using the ARIMA model have been made. Performance of the model and prediction results have been supported by the visuals.

Data Preparation:

The process first grouped transaction volumes on a weekly basis, created transaction dates from the day_since_prior_order column as a Unix timestamp date; then regularly, the ARIMA model processes the time series in a periodical trend at a week frequency.

Time Series Analysis by ARIMA Model:

The time series analysis was done using the ARIMA model. In estimating the parameters of the ARIMA model, the estimates came out to be (5, 1, 0), and the model was trained on data for the weekly transaction volume.

Model Evaluation:

The performance of the in-sample prediction of the model was checked using the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics.

For the in-sample predictions of the model, the following results were obtained:

Mean Absolute Error (MAE): 374.486

Root Mean Squared Error (RMSE): 476.171

These values give the ground to assess the accuracy of the model's predictions on historical data. The higher value of RMSE may indicate that the model is limited in predicting large fluctuations.

Forecasting:

The model was used to predict 12-week future trading volume.

The future predictions of the model are that the trading volume is going to continue with fluctuations. In the predicted trading volume, values approximately ranged from 400,000 to 600,000 on a weekly basis.

Visualization:

Following is the graph showing the observed data side by side against the ARIMA model's predicted future trading volumes:

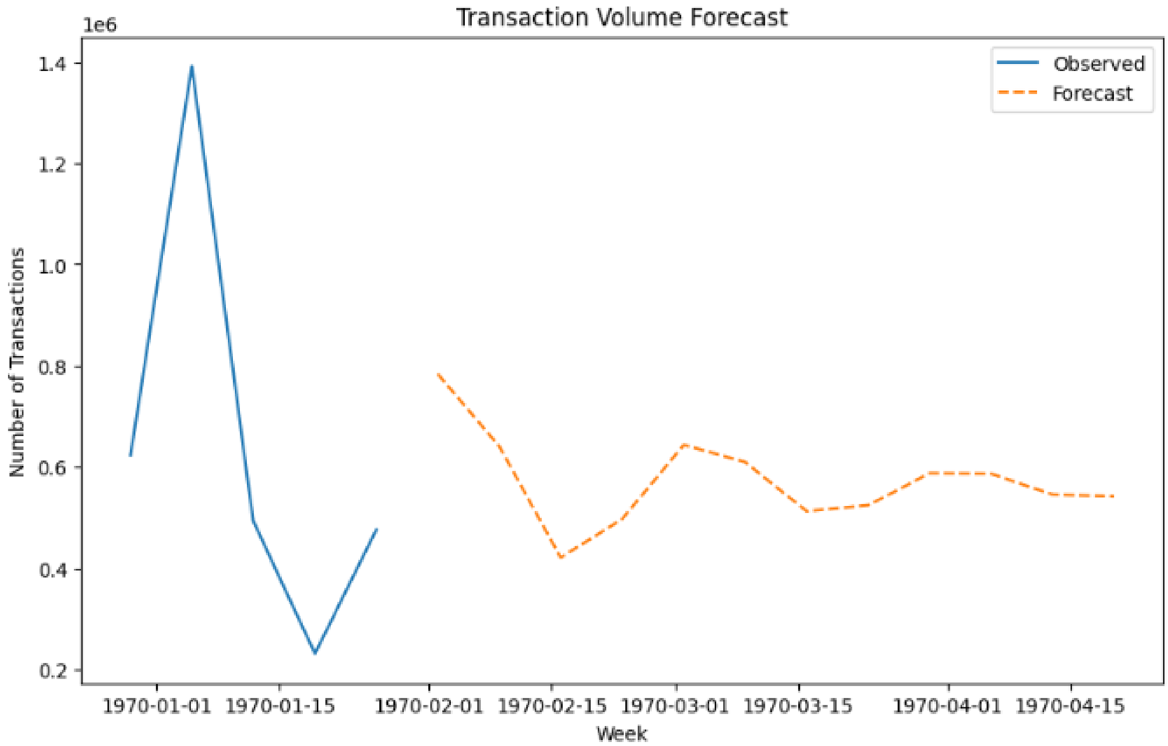


**Figure 23. Transaction Volume Forecast**

Blue line: Actual weekly trading volumes.

Orange dashed line: ARIMA model predicted values.

The graph shows that the model can capture general trends in short-term forecasts but is limited in capturing sudden fluctuations in trading volumes. The predicted trading volume values tend to be generally stationary.

Evaluation

The following findings were recorded regarding the performance of the model:

In-sample Performance: The MAE of 374.486 and RMSE of 476.171 values indicate reasonable accuracy in which the model operates; however, it may see a decline in the performance of the model, especially during periods of high volatility.

Forecast Accuracy: The forecasted values capture the general trends quite successfully. However, it was hard to predict short-run sudden ups and downs.

Conclusion

Therefore, the ARIMA model was chosen as appropriate for the forecast of weekly trading volumes. From the graph, the accuracy of the forecasts shows that, in general, the model correctly predicts the trend in trading volume. However, the following steps can be considered further to enhance the performance:

Model Optimization: Better performance could be achieved by optimizing ARIMA parameters (p, d, q).

Seasonality Components: Inclusion of possible seasonal effects on volume, like holiday seasons, may give better estimates.

Other advanced time series models such as SARIMA or Prophet can be tried. In the last, although ARIMA is a suitable model to capture the general trend, more complex models may be needed for short-run fluctuations.

#### 4.2.2.4 Federated Learning

One of these recent approaches to model training is distributed across devices; federated learning allows it without the transmission of data to a centralized server. Especially for SMEs' sales volume forecasting, it offers the big advantages in terms of ensuring data privacy and security while performing local training business. This approach is particularly useful for SMEs in diverse geographical locations and across various business scales, as each entity can use its own data without relying on a central server. One of the key benefits of FL is that data owners can share only model parameters while keeping their data confidential (Konečný,2015).

In this work, the federated learning process was simulated by partitioning the data from different clients. Each client's data was divided into predetermined partitions, with parts of the model being locally trained. After each round, model parameters were aggregated in a central server to achieve better prediction power. Thus, this solution was effective for those SMEs who were concerned with data privacy: securely using their data for sales predictions. The results imply that FL can be applied to small and medium-sized businesses for data-driven predictions with privacy preservation.

### *4.2.3 Summary of Findings*

This research focused on the application of different machine learning methods for sales volume forecasting in SMEs. The results show the relevance of traditional and modern approaches during the prediction process.

Class imbalance was one of the challenges that the logistic regression model faced when used to predict the product reorder behavior. While it showed high accuracy, it performed poorly in predicting reorder events, hence the need for advanced methods. This therefore, brings into focus the need for SMEs to consider class imbalance when making predictions.

The ARIMA model performed well in the forecast of the sales volume and reflected the general trend of the short-term trading volume. However, it proved to be weak in capturing sudden fluctuations. This model might be appropriate for long-term sales forecasting in SMEs, but improvements are essential for short-term predictions.

The significant advantage of Federated Learning is that it can ensure data privacy and security since the data processing is performed locally on the devices without sending the data to a central server. This approach allowed SMEs to leverage data from different regions and datasets in a secure way, hence more accurate forecasts.

In general, machine learning and times-series analysis show their promising results on the sales volume forecasting of SMEs. At the same time, every single model has challenges that it can't avoid. Besides, SMEs should pay great attention to some issues while using them for better performance: data privacy, class imbalance, model optimization.

## 4.3 ANALYSIS

### *4.3.1 Objective 1: Investigating the development of ML models for SMEs under resource constraints*

The results show that, even though the current adoption rate of ML within SMEs, especially for transaction volume forecasting, is relatively low, there is substantial potential to overcome some key barriers: data scarcity, limited resources, and lack of technical expertise. Only 10% of the SMEs had adopted AutoML, while 70% had never heard of federated learning.

AutoML or federated learning are promising technologies to decrease the inherent complexity and resource demands of traditional ML models. These findings raise the importance of adapting these technologies to the peculiar needs of SMEs. For example, in federated learning, the data may be processed on the client devices, which guarantees data privacy and reduces infrastructure costs. If these technologies are done right, they might provide high-quality, accurate forecasting models that SMEs deserve despite resource limitations.

Logistic Regression analysis therefore shows that SMEs can still use simple and applicable ML models despite their resources. That is quite a successful model with a very overall good accuracy rate of 90%, showing even a basic model can be enough to suit the initial basis for most of the SMEs. However, this created a serious problem in terms of class imbalance, especially in predicting reordered products (Class 1). The low values of precision (0.41) and recall (0.00) are indicative that for Class 1, the performance is not sufficient, and probably more complex solutions will be needed for this class. In such scenarios, class weight, data balancing techniques, or a more complex model such as ensemble methods should be considered by the SMEs.

### *4.3.2 Objective 2: Integration of real-time data analytics with ML models*

The real-time analytics, which have enormous potential for improving forecast accuracy, remain largely unexploited. The results of the survey indicate that though SMEs are able to recognize the true value of real-time data, there is a considerable gap in terms of infrastructure and quality. For example, 60% of the respondents felt that consumer behavior data were significant, but only 25% were actually working on integrating this data into their forecast models.

The logistic regression model, while efficient in the prediction of non-reorder events, had a hard time predicting reorder events due to class imbalances, which is often the case with SMEs. This could be improved by resampling techniques or feature engineering and the inclusion of ensemble methods. These insights are important for SMEs in their effort to optimize inventory management and waste reduction through the accurate prediction of consumer behavior.

#### *4.3.3 Objective 3: Incorporating behavioral economics into ML forecasting models*

The real-time analytics, which have enormous potential for improving forecast accuracy, remain largely unexploited. The results of the survey indicate that though SMEs are able to recognize the true value of real-time data, there is a considerable gap in terms of infrastructure and quality. For example, 60% of the respondents felt that consumer behavior data were significant, but only 25% were actually working on integrating this data into their forecast models.

The logistic regression model, while efficient in the prediction of non-reorder events, had a hard time predicting reorder events due to class imbalances, which is often the case with SMEs. This could be improved by resampling techniques or feature engineering and the inclusion of ensemble methods. These insights are important for SMEs in their effort to optimize inventory management and waste reduction through the accurate prediction of consumer behavior. The integration of behavioral economics with ML models can probably present a more formidable approach toward accounting for market fluctuations and consumer psychology.

#### *4.3.4 Objective 4: Evaluating the feasibility of AutoML and federated learning for SMEs*

AutoML or federated learning are promising technologies to decrease the inherent complexity and resource demands of traditional ML models. These findings raise the importance of adapting these technologies to the peculiar needs of SMEs. For example, in federated learning, the data may be processed on the client devices, which guarantees data privacy and reduces infrastructure costs. If these technologies are done right, they might provide high-quality, accurate forecasting models that SMEs deserve despite resource limitations.

While Logistic Regression can be considered an entry-level solution for SMEs, limitations regarding this model, for example, class imbalance or low AUC (0.654), also pointed to the need to move on to more advanced approaches such as AutoML or federated learning. Poor performance using Logistic Regression in this context underlines the help of AutoML tools for

SMEs. In the case of AutoML, it automatically does complex hyperparameter tuning of SME and helps resolve problems such as class imbalance. Moreover, Federated learning gives an opportunity to improve model performances on larger data while preserving the privacy of data.

## *4.4 DISCUSSION*

### *4.4.1 Beneficiaries of the Research*

SMEs into retail, manufacturing, and services shall be the major beneficiaries of this research because they will immensely benefit from adopting the technologies of machine learning. Through leveraging appropriate ML models to their constraints, the companies can elevate their forecasting, efficiently manage their inventory, and effectively align operations to customer demand. For instance, embedding behavioral economics and real-time analytics within ML forecasting models can indeed provide SMEs with the much-needed agility to dynamically respond toward changes in the market with reduced levels of overstock or understock.

This research would also benefit consultants in machine learning and data scientists who use such findings to develop custom-built solutions to address unique challenges faced by SMEs, such as class imbalance and resource scarcity. The integration of behavioral economics into ML models is indeed a valuable addition to academia, helping to enhance the knowledge of both fields and, therefore, enticing more interdisciplinary research that may come up with new forms of forecasting methods.

### *4.4.2 Implications*

Better Forecasting Performance:

Highlighted is how the study, if an SME, adopts an appropriate model under given constraints- the forecast about volume is rendered quite correct. For instance, embedding real-time data analytics and behavioral economics in ML Forecasting is therefore paramount for SME to respond towards dynamically changed markets with not just better inventories of commodities, goods and lessening wastages thereof.

Class Imbalance Challenges:

The challenges witnessed in Logistic Regression performance, especially for the prediction of reordered products, Class 1, raise the need for techniques to handle class imbalance effectively. Approaches such as oversampling, threshold tuning, and ensemble methods will yield SMEs better forecasts about the minority classes, hence giving more practical utility to their forecasting models.

Accessibility and Awareness:

These results highlight the dire need for focused training and awareness programs in introducing SMEs to modern machine learning techniques, such as AutoML and federated learning. Simplified tools and frameworks should be developed in order to bridge the gaps created by resource limitations and a lack of technical expertise. For example, AutoML can automate complex hyperparameter tuning, thus making the adoption of ML easier for SMEs.

Privacy-Preserving Solutions:

Federated learning provides a very feasible solution for SMEs sensitive to data privacy. Allowing decentralized model training, federated learning enables SMEs to collaborate on the forecast without actually sharing sensitive data. This builds trust and allows for data-driven decision-making with no compromise on security; hence, this could be particularly helpful for SMEs in data-sensitive industries.

### *4.4.3 Reflections and Future Directions*

The present study, though it throws much light on barriers and benefits of adopting ML amongst SMEs, it opens up avenues for future research. Long-term studies may be carried out in regard to sector-specific application of machine learning by studying cultural and regional factors affecting ML adoption amongst SMEs.

The results of this study are thus conducted on a sample of SMEs and a single dataset, Instacart, which reduces the generalizability of the conclusions. In addition, further research should be done by considering various industries and datasets to establish the validity and extension of the findings. Besides, exploring hybrid models and emerging technologies such as explainable AI could further enhance the capabilities of SMEs in improving their forecasting, presenting interpretable and reliable solutions to overcome the limitations identified in simpler models like Logistic Regression.

Through continuous attention to these aspects, future research will be able to give even more remarkable support to SMEs in overcoming obstacles to ML adoption; thus, they are able to take data-driven decisions and bring efficiency into their operations.

## *4.5 SUMMARY*

The present chapter has identified findings from the survey and the Instacart dataset, analyzed them against the background of the objectives of the study, and reflected on their larger implications for SMEs. Conclusively, the results indicated that while SMEs face many key challenges in the adoption of machine learning technologies-which include data quality, resources, and skill constraints-benefits arise when real-time data is integrated with behavioral economics and cutting-edge ML techniques such as AutoML/federated learning. These results also provide crisp actionable insights to SMEs aiming to improve both the accuracy of forecasts and operational efficiency. The next chapter gives practical recommendations to SMEs, policymakers, and other stakeholders who want to implement ML technologies and overcome the barriers identified in this study.

# CONCLUDING REMARKS

## *Key Findings of the Study*

The current study has indeed been a great source of understanding various challenges and opportunities linked to machine learning adoption in small and medium-sized enterprises. Though resource-constrained, data-scarce, and with limited technical expertise, SMEs hold substantial potential for enhancing forecasting accuracy and operational efficiency by the adoption of ML models. Results have shown that even simple models, like logistic regression, work quite well if the conditions are constrained but the problems of class imbalance need to be sorted out. Other sophisticated techniques involve resampling, ensemble methods, AutoML, and federated learning that promise overcoming some of these challenges toward enhancing the forecasting models of SMEs.

The integration of real-time analytics with data and behavioral economics in creating ML models will further impart dynamism in the way SMEs respond to fluctuating market conditions. Equally, the role of AutoML and federated learning for simplification in model building and tuning-with added benefits of reduced infrastructural requirements-can facilitate the application of ML in SMEs.

## *Support of the Literature Review*

These findings correspond to the available literature in respect of challenges for SMEs in the adoption of ML technologies. Results from studies by Narasimha Rao Vajjhala, 2024, indicated that SMEs face multiple challenges in the areas of resource limitation, lack of expertise, and access to technology. Additionally, Shad and Doris (2025) and Davy Preuveneers (2023) prove that such technologies as AutoML and federated learning may seriously reduce these barriers by automating complex processes and maintaining data privacy. The same goes for the research of Soria-Olivas et al. (2022), which underlines the possibility to embed behavioral economics within ML algorithms with the purpose of enhancing prediction accuracy and market response-a finding with which the conclusion of this study aligns.

## *Contradictions to Existing Literature*

While literature indicates that SMEs are increasingly willing to adopt advanced technologies like AutoML and federated learning, the actual rate of adoption for AutoML stood at only 10%, with 70% of the responding SMEs not even aware of the existence of federated learning. This would highlight the difference between the availability and actual usage of these technologies and thus is indicative of various barriers in depth, such as problems with awareness, perceived complexity, and limitations of resources.

Additionally, while existing studies claim that real-time data analytics can significantly improve forecasting accuracy in SMEs (vorecol.com&APICS, 2022), only 25% of SMEs in this research were actually working on integrating real-time data into their forecasting models. This gap underscores the infrastructural and data quality challenges that SMEs face, which hinder them from fully exploiting the potential of real-time analytics.

## *Limitations and Problems Faced*

Several limitations hampered the key findings of the present study. These are as follows:

- Sample Size and Data Set Limitations: The study is based on only one dataset, Instacart; therefore, conclusions cannot be generalized to other industries or geographical locations. These features also demand further validation in future research that will consider an ensemble of industries with a wide dataset for enhanced outcomes.

- Data Quality and Availability: Poor was the quality of data collected through survey forms. Similarly, poor was the data acquired on the transactions side. Most SMEs are not equipped to manage any infra needed for collecting or cleaning data in bulk, and directly influencing an ML model. In fact, the data challenges faced in this study are representative of those SMEs commonly encounter.

- Technical Expertise Limitations: Most SMEs do not have the in-house technical expertise that complicates the adoption of more advanced ML techniques. This was also echoed in the survey responses, where SMEs wished for simpler tools, training, and support that would make the adoption of ML easier.

- Computational Bottleneck: The size of the Instacart dataset, after feature engineering, increased tremendously and presented huge computational limitations in model training.These reflect common problems most SMEs are burdened with whenever they work on huge and raw datasets for any ML use case.
- Time Constraints: SMEs often prioritize day-to-day tasks at the expense of long-term technological adoption. Pressed for time, it is quite hard for any small business to completely merge ML solutions. It was an issue while integrating ML into the speed of a small business generally.
- Biased Response: The survey might have been biased by drawing a sample of SMEs already familiar with technology to a certain extent; this may overestimate the readiness of SMEs to adopt machine learning solutions. This could be because there is an overrepresentation of more tech-savvy companies, which biased the results.

These limitations could be addressed in future research for further insights to support SMEs in overcoming these barriers and leveraging ML technologies more fully.

# BIBLIOGRAPHY


1. González-Sánchez, M., Ibáñez Jiménez, E.M. and Segovia San Juan, A.I. (2021). Market and Liquidity Risks Using Transaction-by-Transaction Information. Mathematics, 9(14), p.1678. doi:https://doi.org/10.3390/math9141678.

2. Akter, S. and Wamba, S.F. (2016). Big Data Analytics in E-commerce: a Systematic Review and Agenda for Future Research. *Electronic Markets*, 26(2), pp.173–194.

3. Buntak, K., Kovacic, M. and Mutavdzija, M. (2021). APPLICATION OF ARTIFICIAL INTELLIGENCE IN THE BUSINESS. *International Journal for Quality Research*, [online] 15(2), pp.403–416. doi:https://doi.org/10.24874/ijqr15.02-03.

4. Tran Tri Dang, Khang Nguyen Hoang, Long Bui Thanh, Nguyen, T. and Cuong Nguyen Quoc (2023). Constructing and Understanding Customer Spending Prediction Models. *SN Computer Science*, 4(6). doi:https://doi.org/10.1007/s42979-023-02284-0.

5. Swedroe, L. (2024). *Using Trading Volume to Optimize Portfolio Construction and Implementation*. [online] Alpha Architect. Available at: https://alphaarchitect.com/2024/11/trading-volume/ [Accessed 19 Nov. 2024].

6. Silva, T.C., Wilhelm, P.V.B. and Amancio, D.R. (2024). *Machine learning and economic forecasting: the role of international trade networks*. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2404.08712.

7. Kotios, D., Makridis, G., Fatouros, G. and Kyriazis, D. (2022). Deep learning enhancing banking services: a hybrid transaction classification and cash flow prediction approach. *Journal of Big Data*, 9(1). doi:https://doi.org/10.1186/s40537-022-00651-x.

8. Wasserbacher, H. and Spindler, M. (2021). Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digital Finance*, [online] 4. doi:https://doi.org/10.1007/s42521-021-00046-2.

9. Chicco, D., Warrens, M.J. and Jurman, G. (2021). The Coefficient of Determination R-squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Computer Science*, 7(5), p.e623. doi:https://doi.org/10.7717/peerj-cs.623.

10. Chai, T. and Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, [online] 7(3), pp.1247–1250. doi:https://doi.org/10.5194/gmd-7-1247-2014.

11. Das, P., Jha, G.K., Lama, A. and Parsad, R. (2023). Crop Yield Prediction Using Hybrid Machine Learning Approach: A Case Study of Lentil (Lens culinaris Medik.). *Agriculture*, 13(3), p.596. doi:https://doi.org/10.3390/agriculture13030596.

12. Liberto, D. (2024). *Small and mid-size Enterprise (SME)*. [online] Investopedia. Available at: https://www.investopedia.com/terms/s/smallandmidsizeenterprises.asp.

13. European Commission (2024). *SME definition - European Commission*. [online] single-market-economy.ec.europa.eu. Available at: https://single-market-economy.ec.europa.eu/smes/sme-fundamentals/sme-definition_en.

14. Anderson, J.C., Narus, J.A., Das Narayandas and D V R Seshadri (2012). *Business market management (B2B) : understanding, creating and delivering value*. New Delhi: Pearson.

15. Bhardwaj, S. (2022). Data Analytics in Small and Medium Enterprises (SME). *Information Resources Management Journal*, 35(2), pp.1–18. doi:https://doi.org/10.4018/irmj.291691.

16. Chen, T. and Guestrin, C. (2016). XGBoost: a Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pp.785–794. doi:https://doi.org/10.1145/2939672.2939785.

17. Breiman, L. (2001). *Random Forests*. [online] Available at: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf.

18. Chapter 11 Classification Algorithms and Regression Trees. (n.d.). Available at: https://rafalab.dfci.harvard.edu/pages/649/section-11.pdf.

19. Feurer, M., Klein, A., Eggensperger, K., Springenberg, J.T., Blum, M. and Hutter, F. (2019). Auto-sklearn: Efficient and Robust Automated Machine Learning. *Automated Machine Learning*, pp.113–134. doi:https://doi.org/10.1007/978-3-030-05318-5_6.

20. Goodfellow, I., Bengio, Y. and Courville, A. (2016). Deep Learning. [online] Cambridge, Massachusetts: The MIT Press. Available at: https://www.deeplearningbook.org/.

21. Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting : Principles and Practice*. 2nd ed. Heathmont, Vic.: Otexts.

22. Barriers to Adopting AI Technology in SMEs A Multiple-Case Study on Perceived Barriers Discouraging Nordic Small and Medium-sized Enterprises to Adopt Artificial Intelligence-Based Solutions. (2019). Available at: https://research-api.cbs.dk/ws/portalfiles/portal/60704162/790410_Aarstad_Saidl_Barriers_to_Adopting_AI_Technology_in_SMEs.pdf.

23. Schwaeke, J., Peters, A., Kanbach, D.K., Kraus, S. and Jones, P. (2024). The new normal: The status quo of AI adoption in SMEs. *Journal of Small Business Management*, pp.1–35.

24. ResearchGate. (n.d.). *(PDF) Defining Small and Medium Enterprises: a critical review*. [online] Available at: https://www.researchgate.net/publication/276294683_Defining_Small_and_Medium_Enterprises_a_critical_review.

25. Brandy, S. (2023). Overcoming Challenges and Unlocking the Potential: Empowering Small and Medium Enterprises (SMEs) with Data Analytics Solutions. *International Journal of Information Technology and Computer Science Applications*, 1(3), pp.150–160. doi:https://doi.org/10.58776/ijitcsa.v1i3.47.

26. Behavioral Economics and Its Positive Impact on Overcoming the Corona Pandemic Among Owners of Small and Medium Enterprises. (2023). *Information Sciences Letters*, [online] 12(10), pp.2795–2805. doi:https://doi.org/10.18576/isl/121028.

27. Konečný, J., McMahan, B. and Ramage, D. (2015). Federated Optimization:Distributed Optimization Beyond the Datacenter. *arXiv:1511.03575 [cs, math]*. [online] Available at: https://arxiv.org/abs/1511.03575.

28. Wang, H., Xie, F., Duan, Q. and Li, J. (2022). Federated Learning for Supply Chain Demand Forecasting. *Mathematical Problems in Engineering*, [online] 2022, pp.1–8. doi:https://doi.org/10.1155/2022/4109070.

29. Accounting Insights. (2024). *ARIMA Models: Financial Forecasting and Advanced Techniques*. [online] Available at: https://accountinginsights.org/arima-models-financial-forecasting-and-advanced-techniques/ [Accessed 12 Jan. 2025].

30. Team, A. (2024). *A Guide to Adopting AI Solutions for Small Business Owners*. [online] Act! Available at: https://www.act.com/en-gb/blog/adopting-ai-for-small-businesses-key-challenges-and-best-practices [Accessed 12 Jan. 2025].

31. ProfileTree Web Design and Digital Marketing. (2024). *Unlocking Advanced Machine Learning Techniques for SMEs*. [online] Available at: https://profiletree.com/advanced-machine-learning-techniques-for-smes/.

32. Roa Baez, J. and Igbekele, R. (2021). *Challenges of AI Adoption in SMEs Insights from the Swedish AI Ecosystem*. [online] Available at: https://www.diva-portal.org/smash/get/diva2:1591640/FULLTEXT01.pdf.

33. ProfileTree Web Design and Digital Marketing. (2024). *Overcoming Challenges in AI Adoption for SMEs: A Practical Guide to Embracing Technology | ProfileTree*. [online] Available at: https://profiletree.com/overcoming-challenges-in-ai-adoption-for-smes/.

34. Bauer, M., Van Dinther, C., Kiefer, D., Kiefer and Daniel (n.d.). Machine Learning in SME: An Empirical Study on Enablers and Machine Learning in SME: An Empirical

Study on Enablers and Success Factors Success Factors. [online] 12, p.0. Available at: https://core.ac.uk/download/pdf/326836032.pdf [Accessed 3 Sep. 2022].

35. Olsowski, S., Schlögl, S., Richter, E. and Reinhard Bernsteiner (2022). Investigating the Potential of AutoML as an Instrument for Fostering AI Adoption in SMEs. *Communications in computer and information science*, pp.360–371. doi:https://doi.org/10.1007/978-3-031-07920-7_28.

36. Iyelolu, V., Ebele, E., None Courage Idemudia and Ignatius, T. (2024). Driving SME innovation with AI solutions: overcoming adoption barriers and future growth opportunities. *International Journal of Science and Technology Research Archive*, 7(1), pp.036–054. doi:https://doi.org/10.53771/ijstra.2024.7.1.0055.

37. Zhang, J., Li, M., Zeng, S., Xie, B. and Zhao, D. (2021). A survey on security and privacy threats to federated learning. doi:https://doi.org/10.1109/nana53684.2021.00062.

38. World Bank (2021). Small and Medium Enterprises (SMEs) Finance. [online] World Bank. Available at: https://www.worldbank.org/en/topic/smefinance.

39. Cleary, P., Quinn, M., Rikhardsson, P. and Batt, C. (2022). Exploring the Links Between IT Tools, Management Accounting Practices and SME Performance: Perceptions of CFOs in Ireland. Accounting, Finance & Governance Review, 28. doi:https://doi.org/10.52399/001c.35440.

40. Vera, D.D. (2024). Financial Forecasting for Small Businesses: Why It Matters and How to Do It. [online] CFO Consultants, LLC | Trusted Financial Consultants. Available at: https://cfoconsultants.net/financial-forecasting-for-small-businesses-why-it-matters-and-how-to-do-it/ [Accessed 14 Jan. 2025].

41. Cleary, P., Quinn, M., Rikhardsson, P. and Batt, C. (2022). Exploring the Links Between IT Tools, Management Accounting Practices and SME Performance: Perceptions of CFOs in Ireland. Accounting, Finance & Governance Review, 28. doi:https://doi.org/10.52399/001c.35440.

42. Giglio, Stefano, Matteo Maggiori, Johannes Stroebel and Stephen Utkus. 2021. "Five Facts about Beliefs and Portfolios." American Economic Review, 111 (5): 1481–1522.

43. Kelly, B.T. and Dacheng Xiu (2023). Financial Machine Learning. Social Science Research Network. doi:https://doi.org/10.2139/ssrn.4519264.

44. Barberis, N. and Jin, L.J. (2023). Model-free and Model-based Learning as Joint Drivers of Investor Behavior. SSRN Electronic Journal. doi:https://doi.org/10.2139/ssrn.4331775.

45. Katsinis, A., Lagüera-González, J. and Bella, D. (2024). EUROPEAN COMMISSION. [online] Available at: https://single-market-economy.ec.europa.eu/document/download/2bef0eda-2f75-497d-982e-c0d1cea57c0e_en?filename=Annual%20Report%20on%20European%20SMEs%202024.pdf.

46. Chong, A.Y.L., Ch'ng, E., Liu, M.J. and Li, B. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. International Journal of Production Research, 55(17), pp.5142–5156. doi:https://doi.org/10.1080/00207543.2015.1066519.

47. Zhang, S., Gong, C., Wu, L., Liu, X. and Zhou, M. (2023). AutoML-GPT: Automatic Machine Learning with GPT. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2305.02499.

48. Samer Abaddi (2024). Factors and moderators influencing artificial intelligence adoption by Jordanian MSMEs. Management & Sustainability An Arab Review. doi:https://doi.org/10.1108/msar-10-2023-0049.

49. Schwaeke, J., Peters, A., Kanbach, D.K., Kraus, S. and Jones, P. (2024). The new normal: The status quo of AI adoption in SMEs. Journal of Small Business Management, pp.1–35.

50. Molete, Olebogeng B, Mokhele, Selloane E, Ntombela, Somquba D and Thango, Bonginkosi A (2025). The Impact of IT Strategic Planning Process on SME Performance:

A Systematic Review. Businesses, [online] 5(1), p.2. doi:https://doi.org/10.3390/businesses5010002.

51. Davy Preuveneers (2023). AutoFL: Towards AutoML in a Federated Learning Context. Applied sciences, 13(14), pp.8019–8019. doi:https://doi.org/10.3390/app13148019.

52. Narasimha Rao Vajjhala (2024). Exploratory Review of Applications of Machine Learning for Small- and Medium-Sized Enterprises (SMEs). *Smart innovation, systems and technologies*, pp.261–270. doi:https://doi.org/10.1007/978-981-99-7711-6_21.

53. Shad, R. and Doris, L. (2025). The Role of Federated Learning in Protecting Data Privacy Across IT Networks. *INFORMATION TECHNOLOGY EDUCATION*. [online] Available at: https://www.researchgate.net/publication/387721836_The_Role_of_Federated_Learning_in_Protecting_Data_Privacy_Across_IT_Networks.

54. Soria-Olivas, E., Gisbert, Cardeñosa, Regino Barranquero and Gomez, Y. (2022). Integration of Behavioral Economic Models to Optimize ML performance and interpretability: a sandbox example. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.2205.01387.

55. vorecol.com (2022). *The Role of RealTime Data Analytics in Enhancing Supply Chain Efficiency*. [online] Vorecol.com. Available at: https://vorecol.com/blogs/blog-the-role-of-realtime-data-analytics-in-enhancing-supply-chain-efficiency-163715.

56. Johnson, A. (2023). *Best Recommendation Engine with Collaborative Filtering in Python*. [online] wowPython. Available at: https://www.wowpython.com/how-to-build-a-recommendation-engine-with-collaborative-filtering/ [Accessed 20 Jan. 2025].

57. Chen, C.-H., Jeng, S.-Y. and Lin, C.-J. (2021). Prediction and Analysis of the Surface Roughness in CNC End Milling Using Neural Networks. *Applied Sciences*, 12(1), p.393. doi:https://doi.org/10.3390/app12010393.

58. malcom (2023). *Mastering Linear Regression: From Basics to Testing the Model - Adventures in Machine Learning*. [online] Adventures in Machine Learning. Available at:

https://www.adventuresinmachinelearning.com/mastering-linear-regression-from-basics-to-testing-the-model/ [Accessed 20 Jan. 2025].

59. Fasinu, D. (n.d.). *Collaborative Entrepreneurship and Building Small Business Collaborative Entrepreneurship and Building Small Business Communities in The Nigerian Fashion Design Industry Communities in The Nigerian Fashion Design Industry*. [online] Available at: https://core.ac.uk/download/322557597.pdf [Accessed 20 Jan. 2025].

60. A MIXED METHODS STUDY OF AFFECTIVE WELL-BEING AMONG SME EMPLOYEES IN MALAYSIA YAP WAI MENG B. Psych (Hons) (First Class Honours). (n.d.). Available at: https://core.ac.uk/download/603239562.pdf [Accessed 20 Jan. 2025].

61. Bissessur, S. (n.d.). *UvA-DARE (Digital Academic Repository) Earnings quality and earnings management : the role of accounting accruals*. [online] Available at: https://core.ac.uk/download/489747035.pdf [Accessed 20 Jan. 2025].

## List of Figure

Figure 1.Estimated number of small and medium sized enterprises (SMEs) worldwide from 2000 to 2023 in millions (Source: Statista, Published by Einar H. Dyvik, Aug 6, 2024) ...................... 9

Figure 2. How does AutoML work? (Source: learn.microsoft, 2024)........................................ 11

Figure 3. Comparison Between Traditional Statistical Models and Machine Learning Models.. 17

Figure 4 .......................................................................................................................... 19

Figure 5 .......................................................................................................................... 19

Figure 6 .......................................................................................................................... 19

Figure 7 .......................................................................................................................... 20

Figure 8. Framing Effect (Source: Scribbr) .................................................................... 34

Figure 9. Workflow of Data Analysis............................................................................ 41

Figure 10. Current Technologies Used by SMEs for Forecasting .............................................. 49

Figure 11. Major Challenges in Adopting Machine Learning.................................................... 50

Figure 12. Integration of Real-Time Data Analytics ............................................................ 51

Figure 13. Application of Consumer Behaviour Data in Forecasting ......................................... 52

Figure 14. Awareness and Adoption of AutoML and Federated Learning ................................. 53

Figure 15. Orders by Day of Week (0=Sunday)................................................................. 54

Figure 16. Orders by Hour of Day ................................................................................ 54

Figure 17. Product Reorder Frequency ........................................................................... 55

Figure 18. User Segmentation Based on Purchase Behaviour.................................................... 56

Figure 19. Model Result - 5 recommended products for User 1 ................................................ 57

Figure 20. First Results ............................................................................................. 58

Figure 21. The average classification report.................................................................... 59

Figure 22. Confusion Matrix and AUC .......................................................................... 60

Figure 23. Transaction Volume Forecast.......................................................................... 63