

ARIMA Forecasting with LLM-Powered Multi-Agent Coordination for Omnichannel Retail KPIs

1st Darian Reyes

Lero—Science Foundation Ireland

Research Centre for Software

University of Galway

Galway, Ireland

Darian.Reyes@universityofgalway.ie

2nd Pierangelo Rosati

Lero—Science Foundation Ireland

Research Centre for Software

University of Galway

Galway, Ireland

Pierangelo.Rosati@universityofgalway.ie

Abstract—This paper presents a novel artificial intelligence (AI) agent-based framework for forecasting key performance indicators (KPIs) in omnichannel retail environments. Traditional retail forecasting often relies on historical or budgeted KPIs, which limits the ability to respond proactively to market dynamics. We address this gap by integrating time series machine learning (ML) models—specifically, ARIMA—with large language model (LLM)-powered agents within a cloud-based architecture. Our approach enables automated, accurate forecasting and the seamless delivery of predictive insights directly into retailers’ operational workflows. The methodology is validated using real-world omnichannel retail data, with performance evaluated at the store and department levels. The system further leverages AI agents for data analysis and reporting, offering actionable recommendations to enhance model accuracy and business outcomes. This work demonstrates the potential of combining advanced ML techniques with agentic reasoning to support data-driven decision-making, improve inventory management, and optimise customer experience across multiple retail channels.

Index Terms—AI agent, omnichannel, retail, time series analysis.

I. INTRODUCTION

The global retail landscape is increasingly characterised by a powerful combination of technology, sustainability, and personalised experiences, with omnichannel shopping becoming the norm [1]. Over the past decade, the retail industry has undergone a profound transformation, driven by rapid advancements in digital technologies and shifting consumer expectations. Customers now demand seamless, flexible, and personalised shopping experiences that transcend traditional channel boundaries. As a result, retailers are increasingly adopting omnichannel strategies to remain competitive, foster customer loyalty, and drive growth. This shift has been further accelerated by global events, such as the COVID-19 pandemic, which have reshaped consumer behaviour and underscored the importance of digital readiness in retail operations.

Omnichannel retail represents the seamless integration of all online and offline retail channels within an organisation to create a unified customer experience. This approach fundamentally differs from multichannel retail, where channels

function as separate entities without explicit coordination. An omnichannel strategy requires the careful coordination of multiple touchpoints through effective marketing-operational interaction. While this creates operational and strategic challenges for organisations, it also represents a significant opportunity for the development of innovative systems that leverage large language models (LLMs) to reduce the technical complexities related to multi-source data aggregation and analysis for business decision-makers. By combining traditional machine learning (ML) tasks, such as time series forecasting, with the dynamic workflows of AI agents, these systems can enable more intelligent, adaptive, and context-aware decision-making.

In the retail sector, Key Performance Indicators (KPIs)—such as transaction value, quantity sold, customer visits, and inventory levels—are captured and reported by aggregating hourly or daily data stored as decimal numbers to ensure precision in measurement and analysis. Accurate forecasting of KPIs is fundamental for optimising inventory management, enhancing customer satisfaction, and maximising profitability. In an omnichannel context, forecasting becomes more complex due to the interplay of multiple sales channels and the influence of external factors such as seasonality, promotions, and market trends. Advanced forecasting systems empower retailers to anticipate demand fluctuations, allocate resources efficiently, and proactively address potential disruptions.

Recent breakthroughs in artificial intelligence (AI), particularly the advent of LLMs, have opened new possibilities for retail analytics. LLMs can process unstructured data, generate actionable insights, and facilitate natural language interactions with business users. By combining traditional ML tasks, such as time series forecasting, with the dynamic workflows of AI agents, these systems can enable more intelligent, adaptive, and context-aware decision-making. When combined with traditional ML techniques, LLM-powered agents can automate complex analytical workflows, support scenario planning, and democratise access to advanced analytics across the organisation. Traditional forecasting typically operates on siloed data

from individual channels, failing to capture the interconnected nature of omnichannel customer behaviour. This creates blind spots when customers research online but purchase in-store, or vice versa, leading to inaccurate demand predictions across channels [1].

The main objective of this paper is to present an overview of a software system for KPI forecasting in omnichannel retail, which combines time series analysis with AI agents powered by LLMs. The reports that use actual figures are populated with these predicted figures, which enables stakeholders to better understand the trends [2].

Business Intelligence (BI) and analytics in retail face challenges. First, data integration and harmonisation: each sales channel may rely on a different data source, requiring transactions to be standardised before they can be analysed. Second, dynamic KPI calculation: since figures cannot always be precomputed, metrics must be generated in real time for end-users. Third, flexibility and rapid adaptation: companies often need to exclude a department or a store from KPI calculations, so accommodating such adjustments must be seamlessly incorporated into ongoing deployments. Data quality is another challenge. For example, in one data source, a store identifier may be reassigned to a different store while the original location is temporarily inactive. When integrating this data source, the retail company must correct the store ID to account for this temporary reassignment.

Standardising data across a retail business is challenging, largely because it requires coordination among different teams. For example, a team may want to exclude a given department from the KPI calculations for certain reports, while another may need to monitor changes in those KPIs with all departments.

The remainder of the paper is structured as follows. Section II establishes connections between this study and prior literature. Section III presents an in-depth exposition of the proposed methodology. Following that, Section IV details the experimental results, while Section V concludes and outlines potential avenues for future research.

II. RELATED WORK

The omnichannel literature is still quite nascent and has increasingly focused on the symbiotic relationship between advanced forecasting methodologies and intelligent system architectures [3], [4]. While early studies conceptualised omnichannel retailing merely as an extension of multichannel retailing, contemporary research demonstrates that successful omnichannel implementation requires a fundamental paradigm shift [5]. Indeed, the multidimensional nature of contemporary retail ecosystems necessitates approaches that transcend traditional statistical models while addressing the operational complexities introduced by interconnected sales channels [6]. This transformation requires organisations to develop novel capabilities for resource reconfiguration through advanced technologies such as AI and data analytics, in order to deliver truly seamless customer experiences [7], [8].

A. Time Series Forecasting Architectures in Multichannel Environments

Time series analysis has been widely used in retail for forecasting key performance indicators (KPIs) such as revenue, inventory levels, and customer demand. The evolution from traditional statistical autoregressive integrated moving average (ARIMA) implementations to hybrid machine learning frameworks demonstrates substantial progress in handling the non-linear dynamics of retail data [9], [10]. While Box-Jenkins methodologies retain utility for baseline predictions in stable product categories, their limitations in modelling promotional uplifts and cross-channel demand interactions have driven the adoption of long short-term memory (LSTM) networks, which are capable of processing multivariate temporal dependencies [11]. Contemporary implementations increasingly combine these approaches—using ARIMA for initial trend decomposition before applying deep learning architectures to residual patterns—achieving error reductions of 18–22% in perishable goods forecasting compared to standalone models [12].

Hierarchical reconciliation techniques have emerged as critical tools for maintaining coherence between stock keeping unit (SKU)-level predictions and enterprise-wide inventory strategies [13]. The Demetra framework, for example, exemplifies this evolution by integrating bottom-level LightGBM forecasts with top-down temporal adjustments through differential ensembling, achieving a 14% improvement in accuracy over traditional reconciliation methods in multinational retail deployments [14]. Such systems address the critical challenge of aligning granular product forecasts with aggregate financial planning, which is particularly vital for organisations managing more than 10,000 SKUs across hybrid physical-digital infrastructures.

B. Cognitive Architectures for Operational Decision Support

Multi-agent systems have emerged as a leading paradigm in AI to address this complexity, combining perception, cognition, and action capabilities enhanced by large language models (LLMs) [15]. The application of AI agents in retail is an emerging area that focuses on leveraging intelligent systems to support decision-making, automate operations, and enhance customer engagement [7], [16]. Additionally, AI agents have shown promise in tasks such as personalised recommendations, automated customer service, and dynamic pricing [17]. The integration of LLM-powered agents into forecasting workflows represents a paradigm shift in retail analytics [18]. Modern implementations extend beyond basic report generation to encompass closed-loop control systems, where predictions directly trigger replenishment orders and pricing adjustments [19]. The application of the LangGraph framework in multi-agent retail environments demonstrates how cascading neural networks can coordinate markdown optimisation, warehouse allocation, and campaign budgeting through shared temporal embeddings.

A multi-agent system is particularly advantageous for omnichannel retail KPI forecasting due to its deep understanding

of channel-specific patterns, customer behaviours, and seasonal variations. This specialisation enables more accurate modelling than a monolithic system, which attempts to handle all channels uniformly. Agents can share insights and coordinate forecasts through negotiation protocols. For example, if the e-commerce agent predicts high online demand, it can communicate with the inventory and logistics agents to ensure adequate stock allocation and fulfilment capacity. Recent developments have enabled AI agents not only to interpret data but also to reason and take action within human-in-the-loop workflows [20].

Retrieval-augmented generation (RAG) [21], [22] has been widely used for enhancing customer support [23]. Efficient RAG-based information extraction is essential for deriving valuable insights for decision-making [24], [25]. In the latest developments, [26] released an open-source RAG system to enhance user engagement in retail.

This combination of temporal forecasting rigour with agentic system flexibility establishes new benchmarks for retail operational intelligence [6]. As validation trials progress across multinational retail cohorts, the focus shifts to implementation challenges—particularly data quality standardisation and computational resource allocation—that will determine real-world adoption rates. The architectural principles demonstrated in current research nonetheless chart a viable path toward truly adaptive retail ecosystems capable of thriving in increasingly volatile markets. None of these research papers specifically study a multi-agent approach combined with KPI forecasting for analytics and reporting in omnichannel retail, which is the specific contribution of this article.

III. METHODOLOGY

In this paper, we propose a multi-agent workflow designed specifically for retail environments to achieve true omnichannel integration. Our system architecture is tailored to meet generic requirements of a retail company that sells both online (through its own e-commerce website, social media channels, and third-party online marketplaces) and offline through geographically dispersed physical stores. The system consists of a series of goal-oriented agents dedicated to specific channels (e.g., in-store, social media) and functional domains (e.g., warehouse operations, KPI reporting), all coordinated by a central learning agent. While individual agents optimise for their specific tasks, the central agent ensures that the system avoids local optima that benefit individual agents but harm overall performance (global optima). When agents have competing goals or resource requirements, the central agent mediates using learned preferences, priority systems, and negotiation strategies.

Each channel agent maintains a unique perspective on incoming data, while cognitive processes operate at the marketing–operational interface. The system incorporates strategic human-in-the-loop oversight, providing retailers with critical checkpoints before task execution to ensure both the benefits of automation and appropriate human supervision. For the

purpose of this paper, we focus on the Reporting & KPIs agent, as illustrated in Figure 1 (see green box).

The Reporting and KPIs agent is supported by Data Analysis agents. Figure 2 visualises the corresponding workflow for report generation. Starting with a scheduled trigger, the agent calculates the required KPIs. An overview generator produces a human-friendly performance summary, while the next node identifies possible explanations for past performance and recommends actions based on internal documentation and external tools to understand current business conditions (e.g., events, weather, seasonality, etc.) using Retrieval-Augmented Generation (RAG). The primary strategy for detecting hallucinations is to compare the generated content with its sources to identify unsupported statements or misrepresentations of facts. A human-in-the-loop checkpoint (grey) is implemented before the report is sent for final verification. The outcome at this point serves as a draft of the report. A human-in-the-loop checkpoint allows the user to verify and customise the final version of the report even further.

Each agent operates through specialised workflows designed for its particular responsibilities. For instance, Figure ?? illustrates the Social Media Agent’s process for post generation, which can be initiated by the Omnichannel Agent when launching new discount campaigns. Post content is informed by RAG, while two human-in-the-loop checkpoints ensure proper review and approval before publication. The workflow further connects to external services through dedicated Application Programming Interfaces (API), enabling the execution of specific actions.

The advanced features of retail KPI dashboards elevate analytics from simple reporting to strategic decision support tools [27]. Dynamic benchmarking allows performance comparison against targets and peers at different levels of granularity, providing crucial context for metrics interpretation. Anomalies in the data, or discrepancies between the target and actual values of a given metric, are highlighted using conditional formatting, and automated alerts are triggered to proactively flag potential outliers or concerning trends [28]. Cross-department impact analysis creates visibility into how metrics in one area affect others, enabling users to trace cause-effect relationships across the retail ecosystem. Time intelligence features maintain contextual awareness during period comparisons, with rolling calculations that preserve seasonal contexts and trend visibility. Finally, natural language Q&A (Question and Answer) capabilities democratise data access by allowing non-technical stakeholders to conduct ad-hoc analyses using natural language. This removes the technical barrier between business users and their data insights [29].

Our proposed model includes the following key features:

- Single source of truth: All departments reference the same core data models in a data warehouse.
- Consistent time periods: Clear and shared reporting periods across all departments, using, for example, a finance calendar where week 1 starts on the last Monday of January—a common practice in the retail sector.

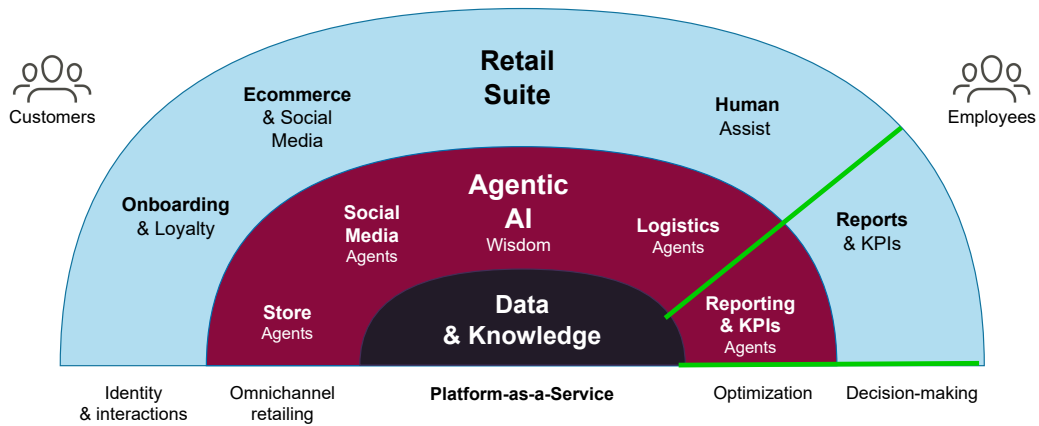


Fig. 1: Retail suite general framework overview. Note: The target area of this study is highlighted in green, the reporting and KPIs area is highlighted in red, and the employees' use of the information area is shown in blue.

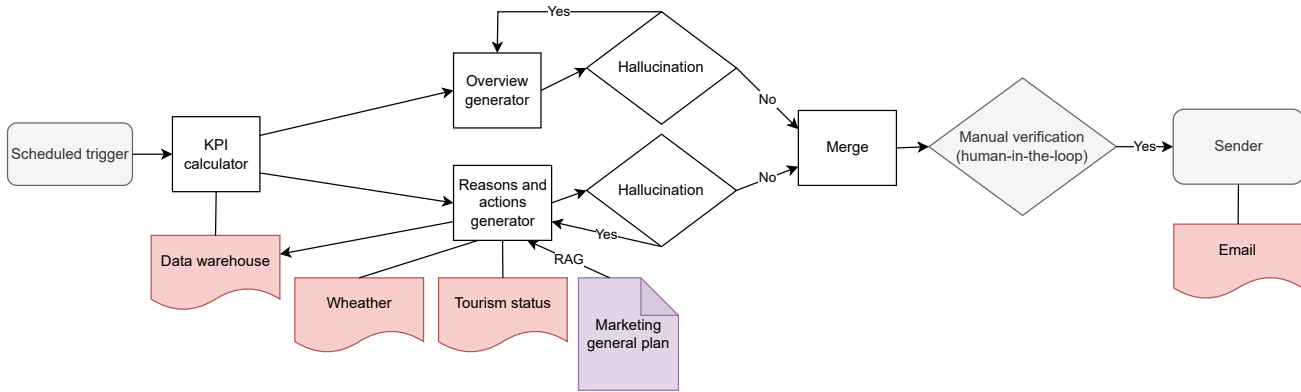


Fig. 2: Reporting & KPIs agent workflow for report generation.

- Visual hierarchy: PowerBI dashboards with higher-level KPIs presented first, but with the option to drill down into each for more detailed insights.
- Actionable insights: Target values and variance indicators are displayed using different colours based on how far the target value of a given metric is from its actual value, providing performance at a glance.
- Mobile optimisation: Reports are accessible on mobile devices for business decision-makers.
- Automated alerts: These are triggered when a given KPI falls outside acceptable ranges.

In the context of our initial development, the most important time periods typically used to analyse various KPIs are: yesterday, this week, last week, this month, last month, this year, and last year. These are based on historical data. Other time periods included in the reports are based on forecast values: today (end of day), tomorrow, and next week. Forecast values are also used in conjunction with actual values of a given KPI to calculate aggregated future figures at the end of a given period. For example, the forecast level of revenue at the end of a given month for a specific sales channel combines historical revenue data from the beginning of the

month up to the day before reporting, with forecasted revenue for the remaining days in the month. The forecast value is updated daily using a rolling time window. This approach is also applied to other time periods, such as end of the year, end of the week, etc. This key combination of actual past values with predicted future values provides a significant advantage to users by enabling the use of time filters in the reporting interface.

The system presented in this paper is designed to run on a cloud infrastructure (coded via Terraform). In the specific case for which the model has been designed and tested, the system is hosted in Google Cloud Platform (GCP) using Cloud Run for serverless Docker containers. Each agent is encapsulated as a piece of software in a backend microservice [30], implemented with the following technologies: the FastAPI framework in Python, the LangGraph framework for large language model operations (LLMOPs) with cycled graphs and RAG. LangGraph offers precise control over the processes and states of the application, allowing the development of Artificial Intelligence (AI) agents with advanced human-computer interaction and memory features [31], [32]. Additionally, Chroma database for vector embedding and the model GPT-4o, as well

as pylint, pytest, and GitHub workflows for Continuous Deployments / Continuous Integration (CD/CI). FastAPI provides high-performance API development with automatic OpenAPI documentation, built-in data validation using Pydantic models, and excellent asynchronous support. LangGraph excels at orchestrating complex AI workflows through its cyclic graph architecture, enabling sophisticated agent behaviours with state persistence and memory capabilities. Chroma offers efficient semantic search capabilities through vector embeddings, enabling fast retrieval of relevant context for RAG operations. GPT-4o provides state-of-the-art language understanding and generation capabilities with multimodal support, enabling sophisticated reasoning and natural conversation flows within AI agents.

In terms of the computational requirements, lightweight agents or event-driven tasks can be deployed using Cloud Run or Cloud Functions, reducing the need for manual infrastructure management (serverless). RAG or Search-Based Agents: Require vector databases (e.g., Vertex AI Matching Engine, Pinecone on the GCP marketplace) and sufficient RAM for embeddings. Persistent logging and monitoring provided by GCP are recommended.

A. KPIs forecasting

This section presents the end-to-end lifecycle of a ML model developed to forecast key KPIs commonly used in the retail sector, and illustrates its implementation in Google BigQuery, leveraging standard SQL queries. Time series forecasting involves using historical observations to project future trends. BigQuery’s ARIMA model—referred to as ARIMA PLUS—incorporates a robust suite of automated preprocessing steps, including:

- Imputation of missing or duplicate data points.
- Smoothing of irregularities, such as outliers or abrupt spikes and dips in historical patterns.
- Adjustments for temporal anomalies related to holidays.

BigQuery provides native ARIMA functionality, eliminating the need to export data to external tools. This keeps the entire analytical pipeline within the GCP ecosystem. Furthermore, the model employs Seasonal-Trend decomposition based on Loess (STL), which enables the isolation of underlying temporal structures and enhances both the interpretability and accuracy of forecasts.

B. Data Preparation and Structuring

Historical transaction data have been systematically recorded over four years in a PostgreSQL data warehouse. The initial step involves identifying the relevant features required for model training. One of the key advantages of the ARIMA PLUS implementation in BigQuery is its ability to autonomously perform data cleaning, anomaly detection, and holiday adjustment, thereby significantly reducing the burden of manual pre-processing.

The primary task at this stage is the consolidation of transactional sales data into a structured format suitable for ML modelling. Specifically, a table or view should be created

in BigQuery that aggregates daily sales volumes for each channel and store. This can be achieved using standard SQL aggregation functions. The final dataset consists of one record per date and store, including fields such as total quantity sold, profit, tax, transactions, item department, item category, item supplier, and so on. The most important aggregation logic from the raw transaction data is the sum of sales, profit, and tax per day per store in terms of forecasting and final reports.

Training, validation (Figure 3), and forecasting are executed daily by means of a set of scheduled jobs, as illustrated in Figure 4.

```
01 | CREATE OR REPLACE TABLE testing_data AS
02 | SELECT * FROM ML.EVALUATE(MODEL
    |     arima_model_items,
03 |     (SELECT * FROM testing_data),
04 |     STRUCT(TRUE AS perform_aggregation, 7 AS
    |         horizon))
```

Fig. 3: Scheduled SQL query for validation.

IV. RESULTS

We first present the model’s performance based on forecast accuracy, segmented by store and item department. Second, we present the insights and recommendations generated by a Reporting & KPI Agent, based on its analysis of validation outcomes.

A. Performance

This work is centred on the time series domain with decimal numerical data; for this reason, we select Root Mean Square Error (*RMSE*) as the fitness function:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - t_i)^2} \quad (1)$$

where n is the number of training cases, p the vector of predictions, and t the target vector.

The forecasting covers the following metrics: sales, tax, profit, quantity, and footfalls. From these metrics, other KPIs are calculated, such as Average Transaction Value (ATV), Items Per Transaction (IPT), and Conversion Rate. The final report compares the forecasted figures with both the budgeted figure and the actual figure from the previous year. The end user can dynamically adjust the comparison metrics as needed.

Figure 5 shows the *RMSE* by location and Figure 6 shows the *RMSE* by item department in terms of profit values. Both cases are evaluated over one week in the finance calendar using the validation subset.

These measurements highlight specific stores and departments exhibiting poor performance, indicating atypical behaviour relative to the overall trend and suggesting the need for further in-depth analysis. These results fall within an acceptable error margin for business purposes. For other KPIs, the empirical analysis will be conducted using alternative fitness functions like Mean Absolute Error (MAE), Mean

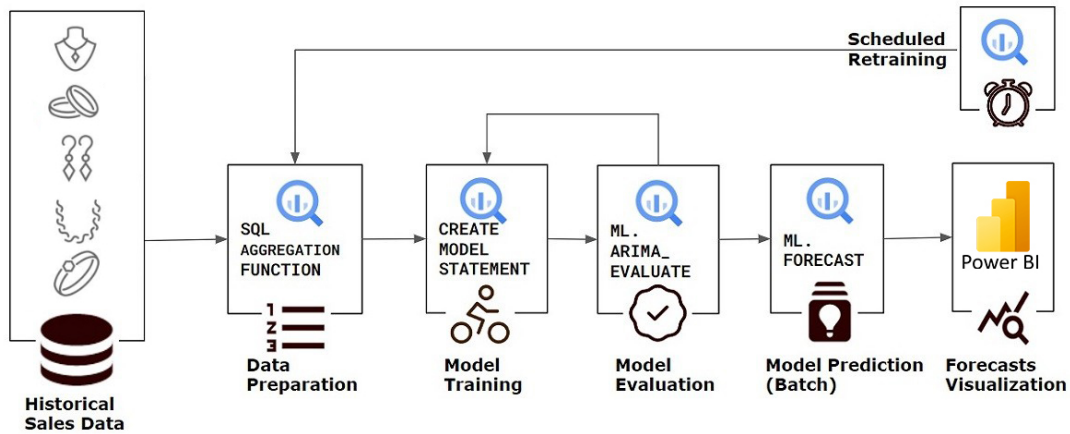


Fig. 4: Step-by-step pipeline including BigQuery and PowerBI to generate the forecasting.

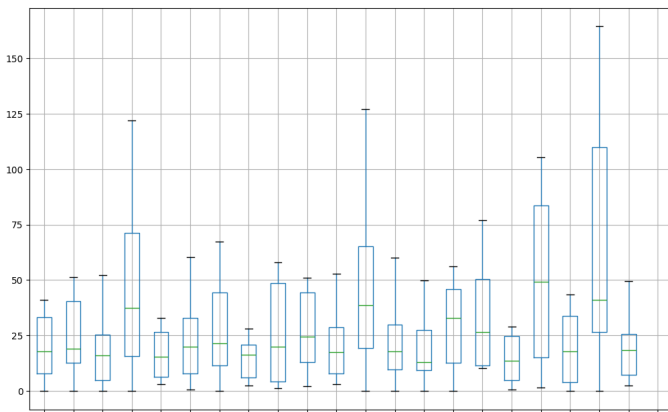


Fig. 5: Profit $RMSE$ by stores across different channels for one month in the finance calendar.

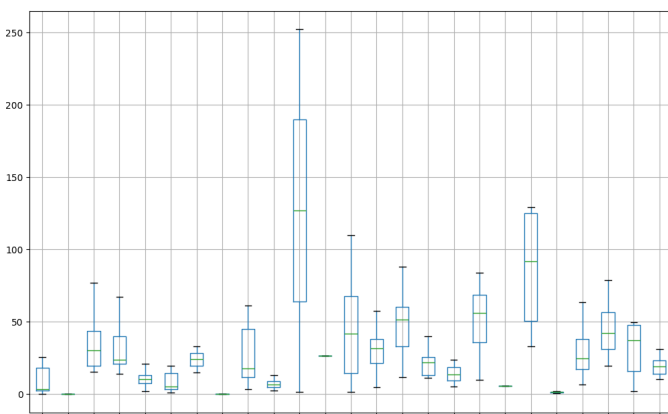


Fig. 6: Profit $RMSE$ by item departments for one month in the finance calendar.

Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE).

B. Recommendations from the AI Agent

The following recommendations identify and implement opportunities to enhance the accuracy of ML models. They are focused on:

- Addressing supplier and location anomalies with unusually high error rates.
- Targeting departments with disproportionate error distributions.
- Improving consistency at locations with high variability.
- Managing departments with problematic error distributions.
- Leveraging correlation patterns between different error metrics.

Each recommendation includes specific actions the AI Agent could take automatically to improve model accuracy. Below are three key examples (Figures 7, 8, and 9).

```

01 | WITH DepartmentAvgRMSE AS
02 | (SELECT department_name, AVG(
03 |   root_mean_square_error) AS avg_rmse
04 | FROM testing_data GROUP BY department_name)
05 | SELECT t1.location_id, t1.supplier_name, t1.
06 |   department_name, t1.
07 |   root_mean_square_error, t2.avg_rmse
08 | FROM testing_data AS t1
09 | JOIN DepartmentAvgRMSE AS t2 ON t1.
10 |   department_name = t2.department_name
11 | WHERE t1.root_mean_square_error > 2 * t2.
12 |   avg_rmse;

```

Fig. 7: Identify locations and suppliers with unusually high $RMSE$ compared to the average $RMSE$ for their department.

Targeted intervention recommendations were autonomously identified as follows:

```

01 | SELECT department_name FROM testing_data
    | GROUP BY department_name
02 | HAVING MAX(root_mean_square_error) > 2 * AVG
    | (root_mean_square_error);

```

Fig. 8: Identify departments where the maximum *RMSE* exceeds twice the average *RMSE*.

```

01 | WITH LocationRmseStats AS (SELECT
    | location_id,
02 | AVG(root_mean_square_error) AS avg_rmse,
    | STDDEV(root_mean_square_error) AS
    | stddev_rmse
03 | FROM testing_data GROUP BY location_id)
04 | SELECT location_id, avg_rmse, stddev_rmse,
    | stddev_rmse / avg_rmse AS
    | coefficient_of_variation FROM
    | LocationRmseStats
05 | WHERE stddev_rmse / avg_rmse > 1.0
06 | ORDER BY coefficient_of_variation DESC;

```

Fig. 9: Identify locations with a high Coefficient of Variation (CoV) in *RMSE*, indicating inconsistent forecasting accuracy.

- Location and Supplier Anomaly Remediation: Automatically detect and flag suppliers and locations exhibiting abnormally high *RMSE* relative to department averages. Implement adaptive correction factors for these outliers and continuously monitor improvement trajectories. Recommend targeted data collection strategies for problematic location-supplier combinations.
- Department-Level Precision Enhancement: Proactively identify departments where maximum *RMSE* exceeds twice the average *RMSE*. Deploy specialised model variants for high-variance departments with automated hyperparameter optimisation. Establish feedback loops to incrementally refine prediction accuracy in atypical departments.
- Location-Specific Consistency Optimisation: Systematically identify locations with elevated Coefficient of Variation (CoV) in *RMSE*. Deploy localised ensemble techniques that adapt to inconsistency patterns. Implement progressive learning mechanisms to prioritise stability across prediction cycles.
- Error Distribution Analysis and Correction: Monitor departments with wide *RMSE* percentile gaps (90th–10th percentile > threshold). Implement automated outlier detection and handling tailored to departmental error distributions. Suggest data preprocessing adjustments to normalise error distributions where significant gaps exist.

Implementation Strategy:

- Continuously monitor these metrics using real-time dashboards.
- Automatically enact corrective actions when thresholds

are exceeded.

- Generate periodic comparative analyses to track improvement trajectories.
- Develop a prioritisation framework balancing potential accuracy gains against implementation complexity.

V. CONCLUSIONS AND FUTURE WORK

A. Conclusions

This study focuses on the practical problem of forecasting in the context of omnichannel retail, with a particular focus on integrating ARIMA time series models into an agent-based system. Drawing on four years of transactional records from a medium-sized Irish retail business, we found that the resulting framework, while not without its shortcomings, proved capable of generating forecasts for key retail indicators such as sales, profit, and footfall. These outputs, in turn, allowed for the calculation of composite metrics like Average Transaction Value (ATV), Items Per Transaction (IPT), and Conversion Rate, which are widely used in day-to-day retail management.

One of the more tangible benefits observed was the system’s ability to highlight persistent forecasting discrepancies at store and department level. This feature is quite useful in practice. For instance, when the model repeatedly overestimated demand in certain niche categories, these errors are surfaced promptly, prompting further scrutiny and, in some cases, may trigger changes to inventory allocation and promotional tactics. It is worth noting, however, that the ARIMA approach was most effective in departments with relatively stable sales patterns. In departments with higher volatility—often due to unpredictable shifts in consumer behaviour or the influence of other factors such as a campaign or adverse weather conditions—the model’s limitations became apparent. In these instances, the agentic layer’s anomaly detection and reporting functions provided some mitigation, but could not fully compensate for the lag in model adaptation.

B. Future Work

This study presents an initial version of the proposed architecture and demonstrates that, even though it represents a good starting point, there is scope for improvement, particularly in the direction of hybrid modelling. The combination of traditional statistical methods with more flexible neural network approaches—such as LSTM or Transformer models—may offer a way to address the shortcomings observed in more volatile retail segments. These alternative algorithms can be seamlessly incorporated into the approach proposed in this paper by substituting ARIMA for them, as they share the same inputs, outputs, and objectives. Additionally, moving towards real-time data integration, though technically demanding, would likely improve the system’s responsiveness to sudden market changes. There is also a case to be made for expanding the agentic framework to handle more nuanced decision-support tasks, including scenario planning and ethical considerations, which are becoming increasingly relevant as automation becomes more prevalent in retail.

Although this study examines an Irish retail company, the proposed approach is generic, scalable and can be readily applied to retail organisations of any size, location, or product category.

To conclude, whilst the integration of ARIMA forecasting within an agent-based architecture has clear operational advantages, it is equally clear that no single approach will suffice for all retail contexts. The work presented here provides a basis for further experimentation and adaptation, and it is hoped that others in the field will build upon these findings to develop more robust and versatile forecasting solutions.

As future work, we plan to implement these predictive models for other KPIs using stock and customer data merged with transaction data. For example, we aim to build a predictive model for the Sell-Through Conversion Rate using stock data, and a predictive model for customer segmentation using customer data. Lastly, we are keen to measure the impact of the AI Agent recommendations on model accuracy and business outcomes.

VI. ACKNOWLEDGMENTS

This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094_P2 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Science Foundation Ireland Research Centre for Software (www.lero.ie).

REFERENCES

- [1] R. Rooderkerk, S. de Leeuw, and A. Hübner, "Advancing the marketing-operations interface in omnichannel retail," *Journal of Operations Management*, vol. 69, no. 2, pp. 188–196, 2023. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/joom.1241>
- [2] G. Nunnari and V. Nunnari, "Forecasting monthly sales retail time series: A case study," in *2017 IEEE 19th Conference on Business Informatics (CBI)*, vol. 01, 2017, pp. 1–6.
- [3] N. Sharma and N. Dutta, "Omnichannel retailing: exploring future research avenues in retail marketing and distribution management," *International Journal of Retail & Distribution Management*, vol. 51, no. 7, pp. 894–919, 2023.
- [4] S. Y. Liew, M. E. Rana, V. A. Hameed, and S. Safavi, "Navigating the retail 4.0 landscape: The transformative impact of cloud computing," in *2023 IEEE 21st Student Conference on Research and Development (SCoReD)*, 2023, pp. 499–507.
- [5] W. Piotrowicz and R. C. and, "Introduction to the special issue information technology in retail: Toward omnichannel retailing," *International Journal of Electronic Commerce*, vol. 18, no. 4, pp. 5–16, 2014.
- [6] R. Fildes, S. Ma, and S. Kolassa, "Retail forecasting: Research and practice," *International Journal of Forecasting*, vol. 38, no. 4, pp. 802–808, 2022.
- [7] P. Thaichon, S. Quach, M. Barari, and M. Nguyen, "Exploring the role of omnichannel retailing technologies: Future research directions," *Australasian Marketing Journal*, vol. 32, no. 2, pp. 162–177, 2024.
- [8] K. Oosthuizen, E. Botha, J. Robertson, and M. Montecchi, "Artificial intelligence in retail: The ai-enabled value chain," *Australasian Marketing Journal*, vol. 29, no. 3, pp. 264–273, 2021. [Online]. Available: <https://doi.org/10.1016/j.ausmj.2020.07.007>
- [9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.
- [10] A. R. Chowdhury, R. Paul, and F. Z. Rozony, "A systematic review of demand forecasting models for retail e-commerce enhancing accuracy in inventory and delivery planning," *International Journal of Scientific Interdisciplinary Research*, vol. 6, no. 1, p. 01–27, Mar. 2025. [Online]. Available: <https://ijsir.org/index.php/IJSIR/article/view/1>
- [11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [12] F. Petropoulos, D. Apiletti, V. Assimakopoulos *et al.*, "Forecasting: theory and practice," *International Journal of Forecasting*, vol. 38, no. 3, pp. 705–871, 2022.
- [13] S. Ma and R. Fildes, "Retail sales forecasting with meta-learning," *European Journal of Operational Research*, vol. 288, no. 1, pp. 111–128, 2021.
- [14] K. Oosthuizen, E. Botha, J. Robertson, and M. Montecchi, "Artificial intelligence in retail: The ai-enabled value chain," *Australasian Marketing Journal*, vol. 29, no. 3, pp. 264–273, 2021.
- [15] Y. Yang, H. Chai, Y. Song, S. Qi, M. Wen, N. Li, J. Liao, H. Hu, J. Lin, G. Chang, W. Liu, Y. Wen, Y. Yu, and W. Zhang, "A survey of ai agent protocols," 2025. [Online]. Available: <https://arxiv.org/abs/2504.16736>
- [16] N. Kshetri, Y. K. Dwivedi, T. H. Davenport, and N. Panteli, "Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda," *International Journal of Information Management*, vol. 75, p. 102716, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S026840122300097X>
- [17] Z. Li, Y. Wang, H. Chen, and X. Liu, "Generative ai in e-commerce: Applications and challenges," *arXiv preprint arXiv:2304.06747*, 2023.
- [18] A. Zhou, D. Kim, and L. Tran, "Llm4retail: Leveraging large language models for retail decision-making," *arXiv preprint arXiv:2310.02345*, 2023.
- [19] S. Gallino and R. P. Rooderkerk, "New product development in an omnichannel world," *California Management Review*, vol. 63, no. 1, pp. 81–98, 2020.
- [20] S. Yao, J. Zhao, D. Yu, W. Du, M. Yu, and S. Sun, "React: Syn-ergizing reasoning and acting in language models," *arXiv preprint arXiv:2210.03629*, 2023.
- [21] A. Asai, Z. Wu, Y. Wang, A. Sil, and H. Hajishirzi, "Self-rag: Learning to retrieve, generate, and critique through self-reflection," 2023.
- [22] S. Jeong, J. Baek, S. Cho, S. J. Hwang, and J. C. Park, "Adaptive-rag: Learning to adapt retrieval-augmented large language models through question complexity," 2024.
- [23] D. T. Thoutam and T. S. Jalasri, "Enhancing customer support in the telecommunications industry through ai-driven chatbots : A telecom-specific approach," Master's thesis, , Faculty of Computing, 2024.
- [24] M. Arslan and C. Cruz, "Business-RAG: Information Extraction for Business Insights," in *21st International Conference on Smart Business Technologies*. Dijon, France: SCITEPRESS - Science and Technology Publications, Jul. 2024, pp. 88 – 94. [Online]. Available: <https://hal.science/hal-04862172>
- [25] M. Arslan, S. Munawar, and C. Cruz, "Exploring business events using multi-source rag," *Procedia Computer Science*, vol. 246, pp. 4534–4540, 2024, 28th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES 2024). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050924023226>
- [26] B. A. T. de Freitas and R. de Alencar Lotufo, "Retail-gpt: leveraging retrieval augmented generation (rag) for building e-commerce chat assistants," 2024. [Online]. Available: <https://arxiv.org/abs/2408.08925>
- [27] S.-W. Lin, E. Y. Huang, and K.-T. Cheng, "A binding tie: why do customers stick to omnichannel retailers?" *Information Technology & People*, vol. 36, no. 3, pp. 1126–1159, 2023.
- [28] J. Xu and L. Cao, "Optimal in-store inventory policy for omnichannel retailers in franchising networks," *International Journal of Retail & Distribution Management*, vol. 47, no. 12, pp. 1251–1265, 2019.
- [29] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Pearson, 2016.
- [30] J. D. Joshua and M. Z. C. Candra, "Omnichannel system development for integration with various marketplace using event-driven architecture," in *2023 IEEE International Conference on Data and Software Engineering (ICoDSE)*, 2023, pp. 191–195.
- [31] A. M. Easin, S. Sourav, and O. Tamás, "An intelligent llm-powered personalized assistant for digital banking using langgraph and chain of thoughts," in *2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, 2024, pp. 625–630.
- [32] H. Chen and Y. Ding, "Implementing traffic agent based on langgraph," in *Fourth International Conference on Intelligent Traffic Systems and Smart City (ITSSC 2024)*, H. Chen and W. Shangguan, Eds., vol. 13422, International Society for Optics and Photonics. SPIE, 2025. [Online]. Available: <https://doi.org/10.1117/12.3050639>