Application of Transformer Models for Demand Forecasting in FMCG Industry

Vineeth Pydi*
Daniels School of
Business
Purdue University
West Lafayette, IN
vpydi@purdue.edu

Pratiksha Gupta
Daniels School of
Business
Purdue University
West Lafayette, IN
gupt1018@purdue.edu

Nikolai Saporoschetz
Daniels School of
Business
Purdue University
West Lafayette, IN
nsaporos@purdue.edu

Liana Simopoulos
Daniels School of
Business
Purdue University
West Lafayette, IN
lsimopou@purdue.edu

Nicholas Zimmerman
Daniels School of
Business
Purdue University
West Lafayette, IN
zimmern@purdue.edu

Abstract—**The dynamic Fast-Moving Consumer Goods sector necessitates advanced demand forecasting methods that adapt to evolving market conditions. This study introduces an innovative approach, leveraging iTransformer models, distinguished for their efficiency in processing multivariate time series data, alongside demand-sensing techniques. By incorporating sentiment analysis derived from Amazon reviews, Google Trends, and competitive analysis, this methodology extends beyond traditional forecasting models, providing a nuanced understanding of market behaviors and trends. The integration of DistilBERT for sentiment analysis further enhances the model, offering depth in consumer sentiment insights. The findings reveal this comprehensive model significantly outstrips conventional forecasting methods in accuracy and reliability, marking substantial progression. The application of these technologies offers actionable insights for FMCG companies, enabling them to navigate demand forecasting with unprecedented precision.**

Keywords—PyTorch, iTransformer, TimesNet, Demand Forecasting, Time-Series Forecasting)

I. INTRODUCTION

In today's highly competitive market, the ability to accurately forecast demand for fast-moving consumer goods (FMCG) offers companies a distinct advantage. FMCG markets are characterized by their vast product range, high turnover rates, and changing consumer preferences. Traditional forecasting methods like linear regression are static and struggle to adapt to this market's dynamic nature. Consequently, businesses face challenges with improper inventory levels and reduced profitability.

The need to improve demand forecasting in the FMCG industry is increasingly recognized as a critical concern. A 2022 article on Wolters Kluwer's, titled "Supply Chain Uncertainty: A fundamental challenge for supply chain planners" discusses the complexities inherent in supply chain planning. It emphasizes the volatility and unpredictability in consumer demand, which are exacerbated by factors like promotional activities [1][2]. Demand uncertainty, coupled with the fallout from COVID-19, has made it clear that traditional forecasting methods are no longer sufficient.

Our research is focused on addressing these challenges. We pose two fundamental questions: How can the implementation of advanced transformer models in analyzing time series data enhance long-term demand forecasting accuracy? Additionally, how do external factors like social media trends, promotional events, competitor activities, and complementary product dynamics influence demand patterns?

To address these questions, we utilized a two-tiered approach by integrating transformer models with demand-sensing techniques. Our research extends beyond traditional data points, incorporating an analysis of external factors such as social media sentiment and promotional activities. This holistic approach acknowledges the multifaceted influences of consumer demand, providing a more realistic and comprehensive forecasting model.

This paper is structured to demonstrate our process and insights, and to bring a new perspective to demand forecasting. The following section offers an extensive review of existing literature on the application of transformers and demand-sensing. This sets the stage for our proposed methodology where we discuss the formulation of criteria that enabled our analysis. The next section is dedicated to explaining how we formulated and tested various models, followed by a review of their performance, evaluated against traditional forecasting methods. Finally, we conclude the paper with a comprehensive discussion of the implications of our study and directions for future research.

With this structured approach, we not only aim to offer practical solutions to an industry struggling with the complexities of modern demand forecasting. Our research strives to bridge the gap between theoretical models and real-world applications, offering a unique perspective on the integration of advanced analytics in the FMCG sector.

II. LITERATURE REVIEW

Accurate demand forecasting in the fast-moving consumer goods (FMCG) market is critical, meaning that innovative approaches are imperative. Advanced analytical models like transformers show promise in meeting this challenge. Originally designed for natural language processing, transformers excel in capturing temporal dependencies in time series data, as highlighted by Kambale et al. (2023) and further explored by Rao et al. (2023) in handling multivariate scenarios. Their research underscores transformers' ability to capture intricate relationships between variables, which is crucial in the dynamic FMCG market [3][4]. Building on this, Vallés-Pérez et al. (2022) demonstrate the effectiveness of deep learning techniques, particularly transformer models, in sales forecasting [5]. Their study showcases transformers' ability to improve forecast accuracy significantly, reinforcing the practical benefits for FMCG companies striving for more precise demand forecasting.

To briefly explore some of the literature sourced in Appendix A, the iTransformer model, featured in "Transformer: Inverted Transformers are Effective for Time Series Forecasting," revolutionizes time series analysis by prioritizing variate tokens over temporal tokens, significantly enhancing its capacity to handle multivariate data [6]. In contrast, the Temporal Fusion Transformers (TFT) model, described in "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting," emphasizes interpretability and diverse data integration [7]. While TFT provides comprehensive insights, the iTransformer's streamlined approach and computational efficiency offer a scalable solution for the dynamic FMCG sector [6] [7]. Additionally, the "Adversarial Sparse Transformer for Time Series Forecasting" introduces sparse attention and adversarial training, enhancing forecasting accuracy [8] [9]. Despite its promise, the iTransformer's simplicity may offer quicker adaptability for FMCG workflows. "Timenet: Temporal 2D-Variation Modeling for General Time Series Analysis" excels in capturing complex patterns but lacks the iTransformer's flexibility [11]. Lastly, "A Time Series is Worth 64 Words: Long-Term Forecasting with Transformers" explores transformers' potential for long-term forecasting, yet the iTransformer's architecture strikes a balance between long-term forecasting and nuanced understanding crucial for FMCG applications [10].

Building on the exploration of transformer models for time series forecasting in the FMCG sector, we now move into a crucial aspect of our research which is demand sensing— the integration of external factors such as social media data. This section highlights the importance and implications of leveraging external data sources to enrich demand forecasting models.

Utilizing social media data for demand sensing provides invaluable real-time insights into consumer sentiments and emerging trends, crucial for FMCG companies to stay ahead. F, Dachyar, M., Taurina, Z., & Qaradhawi, Y. (2021) stress the importance of adapting to rapidly changing consumer behaviors[12] [13]. Social platforms serve as rich sources of consumer insights, aiding in forecasting by capturing market trends and reactions to brand initiatives. Integrating social data enhances predictive accuracy, allowing companies to anticipate demand shifts and align production accordingly. However, this integration poses challenges in data collection and analysis due to the diverse nature of sources. Transformer models offer a robust solution, adept at processing and integrating varied data inputs, thus enhancing demand sensing systems' adaptability and effectiveness [14].

In conclusion, the iTransformer model's unique approach to handling multivariate time series data, coupled with its computational efficiency and superior forecasting capabilities, justifies its selection as the optimal model for time series forecasting in the FMCG industry. Its comparison with models discussed in Table 1 describes its advantages in addressing the specific challenges of FMCG demand forecasting, making it a valuable tool for improving the accuracy and reliability of forecasting processes in this sector.

III. METHODOLOGY

Our research applies the iTransformer model for demand forecasting in the FMCG sector, aiming to enhance the performance over the existing LSTM model. We integrated client data, Google Trends, and social media data, ensuring alignment by week, year, product name, and country. This integration involved standardizing and incorporating sentiment scores from social media and customer reviews from platforms like Amazon, crucial for understanding product sentiment.

The datasets underwent rigorous cleaning to handle missing values, correct outliers, and normalize features, pivotal for the iTransformer model's effectiveness. Feature engineering included calculating rolling means and consumption rates, alongside competitor analysis via Google Trends data.

Model training involved selecting relevant features, tuning the iTransformer model to our dataset, and testing its efficacy against a TimesNet model, using metrics like MSE and MAE. This comparison helped refine our model by identifying performance gaps and areas for improvement.

For demand-sensing, five separate models were implemented, and the code was designed to choose the best model for the specific forecast for predictions. The five models implemented were Gradient Boosting, XG Boosting, Light GBM, Random Forest, and a stacking ensemble method of the four mentioned models. These models were compared using the RMSE (root mean squared error) and implemented in such a way that the best model for each forecast was chosen to perform the predictions. Training these models involves splitting the data into sets for training, validation, and testing, using techniques like cross-validation to optimize hyperparameters and prevent overfitting.

During the initial phases of training our iTransformer and TimesNet models with the client-provided datasets, it became apparent that the volume of data was insufficient to fully leverage the potential of these advanced transformer models. To address this limitation, we augmented our dataset size by a factor of 46, which significantly improved the models'

training process. This experience underscores a critical insight into the operational requirements of transformer models: the substantial volume of data necessary to harness their predictive power effectively. In scenarios where data availability is constrained, the efficacy of these models may be compromised, highlighting a potential limitation in their application to small datasets.

To overcome the data insufficiency, we tested the architecture using the ETT-m1 dataset, providing insights into its ability to handle different data granularities and forecasting lengths. More details on the dataset used for iTransformer and TimesNet (Benchmark model for iTransformer) can be found in Table 9 in the Appendix.
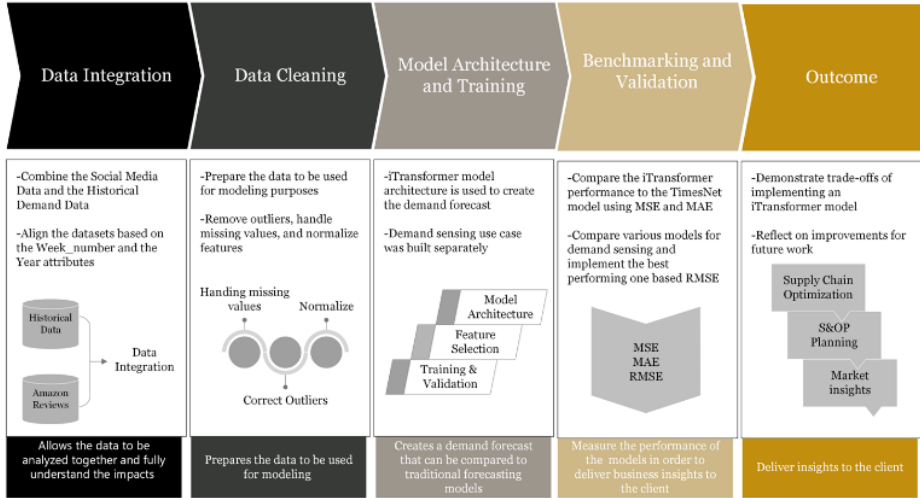
IV. DATA

A. Client Data

The datasets supplied by the client are organized into four main categories, detailed in Appendix B tables 2 through 5, and are designed to aid in the analysis of shipment and consumption metrics for specific brand products. These datasets are structured to support the training and testing of transformer models, which excel in processing sequential time-series data. With attributes like week numbers, years, product identification, and country information, these datasets provide the temporal and categorical context needed for forecasting future trends in both shipments and consumption.

Utilizing these datasets, transformer models can be trained to predict key outcomes such as shipment and consumption volumes, using historical data to gauge the model's accuracy and effectiveness. This allows for performance tuning and validation by comparing predicted values against actual outcomes. Further, testing these models against existing predictive tools can highlight potential improvements in accuracy and efficiency, making transformer models highly beneficial for enhancing predictive capabilities in supply chain and inventory management scenarios.

B. Google Trends Data

The Google Trends dataset, a vital tool for demand forecasting, records search interest over time for various brands along with data completeness, providing insights into consumer behavior patterns. An upward trend in searches often suggests a rise in demand, guiding businesses to adjust their production and inventory levels, while a decline may indicate waning consumer interest, prompting a reassessment of marketing tactics. The cyclical patterns captured by this dataset reveal seasonality and event-driven fluctuations in demand, enabling companies to optimize planning for these peaks, thereby maintaining robust brand health and consumer perception which correlates with demand levels.

Furthermore, analyzing Google Trends data in comparison to competitors can shed light on market positioning and unearth strategic opportunities by identifying strengths and areas for improvement. This data, when combined with other metrics such as sales figures, social media sentiment, and economic indicators, offers a comprehensive view of the market dynamics, substantially enhancing the precision of demand forecasts. The dataset's "isPartial" indicator helps distinguish between established trends and emerging patterns, allowing companies to quickly adapt to changing market conditions and leverage real-time insights to capitalize on new opportunities or mitigate risks, making it indispensable for strategic planning in a competitive market landscape. Additional details on this data set can be found in table 6 of Appendix B.

C. Amazon Reviews Data

The dataset from Amazon reviews, as detailed in the data dictionary and comprising elements such as review titles, bodies, sentiment labels, and sentiment scores, offers invaluable insights for refining demand forecasting strategies. Beyond its obvious application in sentiment analysis, this data can crucially predict product demand by analyzing consumer reviews to understand preferences and trends. Businesses can identify key product features that resonate with customers, with positive reviews indicating strengths and negative feedback highlighting areas for improvement. Seasonal mentions and event-specific comments within these reviews further guide predictions on when product demand might peak, aiding in more accurate planning and inventory management.

Sentiment labels categorize consumer feedback into positive or negative, providing an overview of market sentiment toward a product. This data, coupled with sentiment scores, allows for a nuanced, quantitative analysis of consumer emotions over time, enhancing the precision of demand forecasts. By correlating these sentiment scores with actual

sales data, companies can craft sophisticated models to anticipate demand fluctuations based on evolving consumer sentiments. For instance, high sentiment scores often correlate with spikes in demand, signaling businesses to adjust their production strategies accordingly. Integrating these insights with traditional metrics like sales history enriches demand prediction models, equipping businesses with a comprehensive view of market dynamics and enabling proactive strategy adjustments in alignment with consumer demand trends. Further details about the dataset can be found in table 7 of Appendix B.

D. ETT-m1 Data

The Electricity Transformer Temperature (ETT) dataset, essential for the long-term functionality of electric power systems, incorporates high-resolution, 15-minute interval data over two years from distinct regions in China. Identified as part of the ETT-small collection, it comprises datasets from two separate stations, each labeled ETT-small-m1 and ETT-small-m2, reflecting the specific region of data collection. Each data set aggregates over a million observations per series, meticulously recording eight unique features that are pivotal for analyzing and understanding the operational dynamics of electrical transformers. These features encapsulate a detailed timestamp, the primary predictive measure of transformer oil temperature, and six additional variables indicative of various external power load characteristics, offering a comprehensive dataset for time-series forecasting.

This expansive dataset serves as a critical asset in training the iTransformer and TimesNet models, showcasing their capability to effectively forecast future scenarios based on intricate patterns and trends observed in historical data. The ETTm1 subset, specifically, with its seven-dimensional feature space and designated prediction lengths of 48, 96, and 192 intervals, which spans dataset sizes as mentioned in Table 9 Appendix B, ensuring a robust framework for evaluating the performance of advanced forecasting models. The frequent, 15-minute collection interval provides a granular view into the operational status and trends within the electric power system, enabling the iTransformer model to leverage its unique architecture for precise, long-term forecasting across a wide array of use cases within the sector. More details about the data and the attributes can be found in table 8 Appendix B.

V. MODELS

A. iTransformer Modeling and Architecture

The core of our study leverages the iTransformer, an innovative adaptation of the traditional Transformer model, specifically engineered for time series forecasting. This model is built upon the premise that inverting the dimensions of the Transformer architecture can significantly enhance its efficacy in multivariate time series analysis, particularly for the Fast-Moving Consumer Goods (FMCG) sector.

The iTransformer distinguishes itself by embedding each time series as variate tokens and employing the self-attention mechanism to ascertain multivariate correlations. This approach allows for a more nuanced understanding of the interplay between different variables over time. Each variate is processed through a series of Transformer blocks, which consist of layer normalization and a feed-forward network, facilitating the learning of complex series representations from historical data.

Our choice to employ the iTransformer model is grounded in its demonstrated capability to handle the volatile nature of the FMCG sector. By incorporating external variables such as sentiment analysis, Google Trends data, and competitor analytics, the iTransformer provides a comprehensive framework for predicting demand.

1) Pros and Cons

The model excels in capturing complex dynamics of multivariate time series, improving forecasting accuracy and adaptability. It efficiently manages large lookback windows, avoiding computational issues typical in traditional models. The complexity of the model and the novel approach may present a steep learning curve. Performance is highly dependent on the quality of input data.

2) Tuning Parameters

The iTransformer model's performance can be optimized through adjustments to several key parameters, including the number of Transformer blocks (L), the dimensionality of the embedded tokens (D), and the settings for the self-attention mechanism. Fine-tuning these parameters allows for a balance between computational efficiency and forecasting accuracy.

3) TimesNet as a Benchmark Model

TimesNet, a novel approach to time series analysis, shifts 1D time series data into a 2D space to capture temporal variations effectively. It utilizes multi-periodicity, transforming time series into 2D tensors and applying 2D kernels, which enhances its ability to model complex temporal dynamics. The TimesNet architecture features adaptive discovery of time series multi-periodicity and extraction of temporal variations through its TimesBlock module and inception block, providing a nuanced modeling of temporal variations.

B. Demand-Sensing Modeling

The FMCG industry is highly volatile, and demand forecasts can be influenced by external factors. For our study, we implemented various models using the external data that was collected from Google Trends and Amazon. This enabled us to measure the sentiment surrounding the brand and competitors utilizing a pretrained Transformer model DistilBERT from the PyTorch library. The Google Trends interest over time attributes described earlier were used to measure consumer engagement with that of the client and the competing brands. Companies often struggle to measure external influences on demand forecasts, but our approach offers a solution by using publicly available data. Using the actual shipment and consumption values 5 separate models were implemented: Gradient Boosting, XGBoosting, Light GBM, Random Forest, and an ensemble stacking method of

the four models. These results are discussed later in the Results section.

1) Pros

The analysis process offers numerous advantages, including comprehensive data collection and preprocessing, demonstrated by meticulous web scraping and thorough preprocessing efforts, such as addressing missing values and encoding categorical variables. This foundational work ensures the data's quality and readiness for subsequent analysis and modeling, setting a strong basis for accurate insights. The use of a broad spectrum of regression models facilitates an extensive comparison of different algorithms, highlighting the multi-model approach's value in identifying the most suitable model for the data's characteristics. Additionally, the application of the Root Mean Square Error (RMSE) metric provides a quantitative basis for model comparison, while feature scaling enhances model performance, especially for algorithms sensitive to input variable magnitudes.

2) Cons

Although there are many advantages to our approach, the analysis would benefit from incorporating explicit model validation and overfitting mitigation strategies such as cross-validation or regularization to ensure models' generalizability to unseen data. While preprocessing effectively addresses many data issues, further attention to error handling and validation within the modeling phase is crucial for safeguarding against unexpected data anomalies, thereby enhancing the analysis's reliability. Moreover, the exploration of advanced models without corresponding hyperparameter tuning might not fully exploit these models' capabilities. Systematic tuning stands to significantly improve performance, underscoring the potential for optimizing model outcomes through careful adjustment of model parameters.
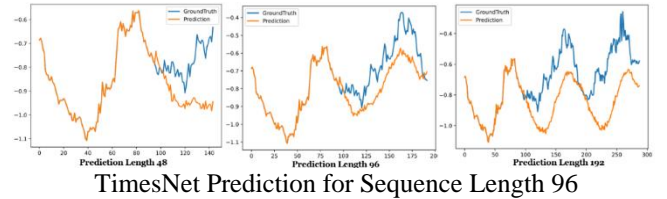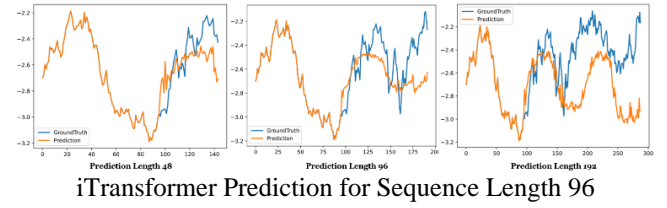
VI. Results

A. iTransformer Results

The results of the iTransformer and TimesNet models, as applied to the ETTm1 dataset, are summarized in the table and corresponding time series forecasts. For prediction lengths of 48, 96, and 192, the iTransformer consistently outperformed TimesNet, as evidenced by lower Mean Squared Error (MSE) and Mean Absolute Error (MAE) values across all prediction horizons. Specifically, the iTransformer achieved MSEs of 0.31299, 0.34189, and 0.38293 for the respective prediction lengths, whereas TimesNet produced higher MSEs of 0.46972, 0.54550, and 0.63605. Similarly, MAEs for the iTransformer were 0.35610, 0.37674, and 0.39570, compared to TimesNet.

The Graphs of the forecasted time series against the ground truth for both models further illustrate the iTransformer's superior performance. Notably, the iTransformer's forecasts are closer to the ground truth, especially for shorter prediction lengths, demonstrating its robustness and precision in modeling time series data. These results validate the iTransformer's efficacy in time series forecasting and

underscore its potential in improving demand forecasting accuracy in the FMCG sector. The findings will be instrumental in guiding future model optimizations and implementations.

iTransformer Prediction for Sequence Length 96

TimesNet Prediction for Sequence Length 96

Dataset	Prediction Length	iTransformer		TimesNet	
		MSE	MAE	MSE	MAE
ETTm1	48	0.31299	0.35610	0.46972	0.49384
	96	0.34189	0.37674	0.54550	0.53269
	192	0.38293	0.39570	0.63605	0.57069

B. Demand-Sensing Results

The results of our demand-sensing modeling results provide insight into the effectiveness of utilizing different factors. Here are our top-3 performing models in terms of MSE:

Brand	Shipment			Consumption		
	Ensemble	Gradient Boosting	XGBoost	Ensemble	Gradient Boosting	XGBoost
Brand 1	3,196.67	3,290.58	3,440.99	3,430.34	4,298.22	4,749.20
Brand 2	15,834.61	15,322.67	16,890.92	11,768.22	12,841.09	15,788.81

Utilizing Google Trends and Amazon data, coupled with a pretrained DistilBERT model, the study evaluated the impact of public sentiment on demand forecasts for Brand 1 and Brand 2.

For Brand 1, the ensemble model yielded an MSE of 3,196.665 for shipments and 3,430.344 for consumption, indicating relatively precise predictions. Gradient Boosting reported slightly higher errors, while XGBoost had the highest MSE among the three evaluated models.

In the case of Brand 2, the MSEs were generally higher across all models compared to Brand 1. The ensemble model again performed the best for shipments with an MSE of 15,834.61 and for consumption with an MSE of 11,768.22. Gradient Boosting and XGBoost followed, with the latter showing the highest error rates again, indicating a trend consistent with Brand 1's results.

VII. Conclusion

In conclusion, this study successfully demonstrated the effectiveness of transformer models and demand-sensing

techniques for demand forecasting in the fast-moving consumer goods (FMCG) industry. By leveraging the innovative iTransformer model and integrating external factors such as social media sentiment and competitive analysis, we significantly improved forecasting accuracy compared to traditional methods. The iTransformer's unique architecture and computational efficiency, paired with comprehensive sentiment analysis using DistilBERT, provided a robust framework for understanding market trends and consumer behaviors.

The results of this research highlight the potential of advanced analytics in enhancing demand forecasting precision, offering actionable insights for FMCG companies to navigate market volatility with increased confidence. The successful application of these models not only underscores their practical relevance but also opens up avenues for future research to further refine and adapt these techniques to other sectors and industries, contributing to a broader understanding of advanced forecasting methodologies.

REFERENCES

[1] CCH® Tagetik. (2022, April 15). Supply Chain Uncertainty: A fundamental challenge for supply chain planners. Wolters Kluwer. Retrieved from [Wolters Kluwer](https://www.wolterskluwer.com/en/expert-insights/supply-chain-uncertainty-a-fundamental-challenge-for-supply-chain-planners).

[2] CCH® Tagetik. (2021, June 29). Forecasting tips that reduce Supply Chain uncertainty. Wolters Kluwer. Retrieved from [Wolters Kluwer](https://www.wolterskluwer.com/en/expert-insights/forecasting-tips-that-reduce-supply-chain-uncertainty)

[3] Kambale, W. V., Kadurha, D. K., El Bahnasawi, M., Al Machot, F., Benarbia, T., & Kyamakya, K. (2023). Transformers in Time Series Forecasting: A Brief Transfer Learning Performance Analysis. In 2023 27th International Conference on Circuits, Systems, Communications and Computers (CSCC) (pp. 212-217). Rhodes (Rodos) Island, Greece. https://doi.org/10.1109/CSCC58962.2023.00041

[4] Rao, G. R. K., Pujari, J. J., Daniel, R., Rama Krishna, S. V., & Hema, C. (2023). Transformer Architecture-based Multivariate Time Series Prediction. In *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)* (pp. 1691-1696). Trichy, India. https://doi.org/10.1109/ICAISS58487.2023.10250751

[5] Vallés-Pérez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications, 201*, 116993. https://doi.org/10.1016/j.eswa.2022.116993

[6] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106-11115. https://doi.org/10.1609/aaai.v35i12.17325

[7] Lima, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2020). Temporal fusion transformers for interpretable multi-horizon time series forecasting. University of Oxford & Google Cloud AI.

[8] Wu, S., Xiao, X., Ding, Q., Zhao, P., Wei, Y., & Huang, J. Adversarial Sparse Transformer for Time Series Forecasting. Conference/Journal Name, Volume(Issue), Page numbers. Tsinghua University/Peng Cheng Laboratory/Tencent AI Lab/University of Texas at Arlington.

[9] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. arXiv. https://arxiv.org/abs/2211.14730

[10] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2023). iTransformer: Inverted transformers are effective for time series forecasting [arXiv preprint arXiv:2310.06625].

[11] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. arXiv. https://arxiv.org/abs/2210.02186

[12] F, F., Dachyar, M., Taurina, Z., & Qaradhawi, Y. (2021). Disclosing fast moving consumer goods demand forecasting predictor using multi linear regression. Engineering and Applied Science Research, 48(5), 627–636. Retrieved from https://ph01.tci-thaijo.org/index.php/easr/article/view/242407

[13] Kazmi, S. W., & Ahmed, W. (Date). Understanding dynamic distribution capabilities to enhance supply chain performance: A dynamic capability view. Department of Business Administration, IQRA University, Karachi, Pakistan.

[14] Nozari, H., Szmelter-Jarosz, A., & Ghahremani-Nahr, J. (2022). Analysis of the Challenges of Artificial Intelligence of Things (AIoT) for the Smart Supply Chain (Case Study: FMCG Industries). Sensors, 22(8), 2931. https://doi.org/10.3390/s22082931

Table 1: Transformer Models Literature Review Comparison

Transformer Architecture	No of Layers	Challenges	Benefits	Loss Function
Informer	Informer contains a 3-layer stack and a 1-layer stack (1/4 input) in the encoder, and a 2-layer decoder.	Managing the computational complexity introduced by long sequences and ensuring the model's scalability and efficiency without sacrificing prediction accuracy.	Ability to handle very long sequences efficiently, its improved performance compared to existing models, and its innovative approach to reducing computational and memory requirements.	MSE
TFT	The number of layers in TFT is not specified explicitly, but the model architecture includes multiple components like gating mechanisms, variable selection networks, and several layers within the temporal fusion decoder, including sequence-to-sequence layers and multi-head attention mechanisms.	Ensuring interpretability without sacrificing performance. Complexity of integrating various input types	The ability to handle a wide range of forecasting scenarios with different types of inputs, and new forms of interpretability that offer insights into the temporal dynamics learned by the model.	Jointly minimizing the quantile loss summed across all quantile outputs:
AST	Encoder-decoder models consist of multiple layers in both the encoder and decoder, and the number can vary based on the implementation specifics. Discriminator is added	Managing adversarial training complexity and ensuring sparse attention accurately identifies relevant temporal dependencies	Improved forecasting accuracy, especially for long sequences, and enhanced model robustness	Combination of quantile loss for the generator (to capture the overall pattern of time series and align predictions with ground-truth) and adversarial loss (to regularize the prediction from a global perspective and improve sequence-level accuracy). This dual-loss approach leverages the benefits of both worlds: accurate step-level forecasting and global sequence fidelity.
PatchTST	The model experiments with varying the number of Transformer layers ($L=\{3,4,5\}$) and model dimensions ($D=\{128,256\}$), with the inner-layer of the feed-forward network being $2D$. This setup indicates a flexible architecture that can be adjusted based on the complexity of the time series data and the specific requirements of the forecasting task.	Model's adaptability to different types of multivariate time series data, Adjusting the model's parameters, including the number of Transformer layers and model dimensions, to optimize performance without overfitting or underfitting can be complex.	Reduced computational complexity, ability to learn from longer historical data, and enhanced representation learning capabilities.	MSE
iTransformer	Model architecture does not specify the exact number of layers or parameters. It mentions the use of multiple Transformer blocks (TrmBlock) stacked together, each comprising layer normalization, a feed-forward network, and a self-attention module, suggesting a multi-layered approach	Adapting the Transformer architecture without modifying its core components for time series data.	Improved performance on time series forecasting tasks by effectively capturing multivariate correlations. Enhanced interpretability of the model's predictions due to the novel application of self-attention mechanisms. Flexibility and generalization capability, making it suitable for a wide range of forecasting applications.	MSE, MAE
TimesNet	Does not explicitly enumerate a total count of layers within the TimesNet model, it provides detailed insights into its architecture, highlighting the core component known as the TimesBlock.	Model complexity, computational resources, and the need for fine-tuning parameters to specific dataset characteristics	Good with mixed dataset performance. Results also verify the potential of TimesNet in performing as the general-purpose backbone for large-scale pre-training in time series forecasting.	MSE

Table 2: Shipment Forecast Using Best Model

Column Name	Description
Week_number	Integer used to store the calendar week for the forecasted shipment
Year	Calendar year for the target variable

Product_name	String value for the product name of the specific brand
Product_ID	Unique number value for the product identification
Country	String value for country where the shipment is being forecasted
Forecasted_shipment_value	Numeric value that is the target variable and predicted shipment forecast

Table 3: Actual Shipment Demand

Column Name	Description
Week_number	Integer used to store the calendar week for the actual shipment
Year	Calendar year for the target variable
Product_name	String value for the product name of the specific brand
Product_ID	Unique number value for the product identification
Country	String value for country where the shipment value
Actual_shipment_value	Numeric value that is the target variable and actual shipment value

Table 4: Consumption Forecast Using Best Model

Column Name	Description
Week_number	Integer used to store the calendar week for the predicted consumption
Year	Calendar year for the target variable
Product_name	String value for the product name of the specific brand
Product_ID	Unique number value for the product identification
Country	String value for country where the consumption is being forecasted
Forecasted_consumption_value	Numeric value that is the target variable and predicted consumption forecast

Table 5: Actual Consumption Demand

Column Name	Description
Week_number	Integer used to store the calendar week for the actual consumption
Year	Calendar year for the target variable
Product_name	String value for the product name of the specific brand
Product_ID	Unique number value for the product identification
Country	String value for country where the consumption value
Actual_consumption_value	Numeric value that is the target variable and actual consumption value

Table 6: Google Trends Data – Over Time

Column Name	Description
Date	The date for the data point, typically weekly.
Brand Name	An integer representing the search interest for Brand on that date.
isPartial	A Boolean indicating whether the data for the given period is complete or preliminary.

Table 7: Amazon Reviews

Column Name	Description

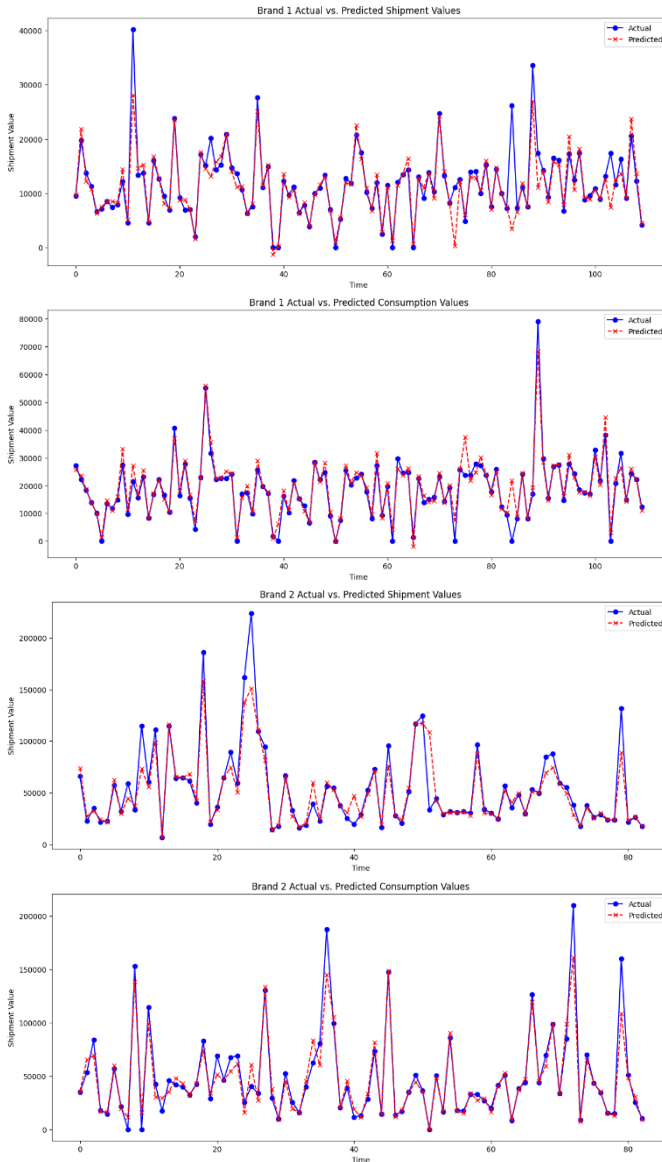| title	The title of the review provided by the customer. It summarizes the customer's experience or opinion about the product.
body	The full text of the review written by the customer. It contains the customer's detailed experience, opinions, and comments about the product.
sentiment_label	A label indicating the overall sentiment of the review as interpreted by sentiment analysis algorithms. It classifies the sentiment of the review into categories.
sentiment_score	A numerical score that quantifies the sentiment of the review. The score typically ranges from -1 to 1.

Table 8: ETTml Dataset

Column Name	Description
date	The recorded date
HUFL	High UseFul Load
HULL	High UseLess Load
MUFL	Middle UseFul Load
MULL	Middle UseLess Load
LUFL	Low UseFul Load
LULL	Low UseLess Load
OT	Oil Temperature (target)

Table 9: ETT-ml Overview

Dataset	Dimensions	Sequence Length	Frequency	Prediction Length	Dataset Split (Train, Validate, Test)
ETT-m1	7	96	15min	48	(34,417, 11,473, 11,473)
				96	(34,369, 11,425, 11,425)
				192	(34,273, 11,329, 11,329)

APPENDIX C: RESULTS

Brand 1 Actual vs. Predicted Shipment Values

Brand 1 Actual vs. Predicted Consumption Values

Brand 2 Actual vs. Predicted Shipment Values

Brand 2 Actual vs. Predicted Consumption Values

APPENDIX D: AI RESEARCH TOOLS

As part of our IP project, we tried an extensive literature review to anchor our research in the current scientific discourse on transformers and demand sensing. The sheer volume of existing literature initially seemed daunting. However, leveraging several AI tools transformed this challenge into a manageable, even enjoyable task. Here's a breakdown of the tools that were particularly instrumental to our success:

- **Semantic Scholar**: Our first stop was Semantic Scholar. Its AI-driven search and summarization capabilities were invaluable. Not only did it help us find relevant papers quickly, but its smart summaries allowed us to gauge a paper's relevance without reading it in full. This tool was particularly helpful in narrowing down our initial pool of literature.

- **Google Scholar**: Though not an AI tool, Google Scholar served as our gateway to the vast world of academic literature. Its powerful search algorithms helped us pinpoint research papers, theses, books, and conference papers relevant to our topic. What made Google Scholar stand out was its ability to show highly cited papers and the references within each article, providing us with a roadmap of influential works in our area of study. The simplicity of Google Scholar, combined with its depth of resources, made it an indispensable tool for our literature review.

- **Scite**: Scite was helpful in assessing the reliability and impact of the papers we considered. Its unique approach to categorizing citations (supporting, mentioning, disputing) gave us a clear picture of the consensus around specific findings, which helped in refining our bibliography with authoritative sources.

- **Connected Papers**: Understanding the relationship between different studies was crucial for our review. Connected Papers provided a visual representation of how each paper was related, helping us identify key papers and themes we might have otherwise missed. It was like having a map of the literature landscape, guiding us to crucial connections and insights.

- **Zotero**: Managing our references was a breeze with Zotero. Its ability to automatically extract metadata and organize citations saved us countless hours. Moreover, Zotero's integration with word processors meant that formatting our bibliography was no longer a tedious task but a seamless part of our writing process.

- **ChatGPT:** Whenever we hit a roadblock, whether understanding complex concepts or drafting sections of our review, ChatGPT was there to help. It provided quick summaries, explained methodologies, and even suggested how to structure our arguments. Though we always double-checked its outputs against primary sources, ChatGPT was like having an ever-present tutor guiding us through the process.

- **PubTrawlr**: This tool was particularly useful later in our review process. As we started synthesizing our findings, PubTrawlr's AI-driven summaries and analysis helped us identify trends and gaps in the literature, ensuring our review was both comprehensive and insightful.

- **Turnitin**: Before final submission, we ran our review through Turnitin. While primarily a plagiarism checker, it helped us ensure that all citations were properly attributed and that our work maintained academic integrity. It was the final step in our process, giving us confidence in the originality and rigor of our review.

- **Claude.ai:** proved invaluable for quickly understanding complex texts and generating concise summaries, enhancing our efficiency in digesting vast amounts of research. Its AI-driven insights helped us identify key themes and draft parts of our review with greater ease.

- **Typeset.io:** was instrumental in the final stages, simplifying the process of formatting our document according to academic standards. It ensured our review was professionally presented, with correctly formatted citations and references, saving us significant time and effort.