Dairy product Sales Forecasting based on Customer Segmentation, Demographics and Purchase patterns

Analyzing Patterns and Influences in Retail Data

Team Members

PRN No.	Name	Department	Email ID
122B1F047	Disha Kamalaskar	Information Technology	disha.kamalaskar22@pccoepune.org
122B1F058	Anushka Korde	Information Technology	anushka.korde22@pccoepune.org
123B2F149	Revati Keskar	Information Technology	revati.keskar23@pccoepune.org

Guided By

Dr.Harsha Bhute

Information Technology,PCCoE

Abstract

This case study investigates a data-driven approach to a crucial role in the dairy industry, where daily consumption and perishable products demand accurate inventory management and demand prediction. The constant nature of sales in this sector makes it an ideal domain for forecasting models, as consumer behavior can often be predicted based on historical data,demographic trends, and purchasing patterns. The motivation behind choosing this domain lies in its accessibility—understanding the factors affecting dairy product sales does not require specialized expertise and relies on common knowledge related to everyday consumption habits.

Keywords

1.Sales Forecasting

2.Customer Segmentation

3.Demographics

4.Purchasing Patterns,

5.Consumer Behavior

6.Predictive Analytics

Introduction

By applying machine learning techniques to analyze customer segmentation, demographics, and purchase patterns, businesses can better anticipate demand and optimize their operations. This case study aims to showcase how foundational methods, such as K-means clustering for customer segmentation and linear regression for sales forecasting, can be effectively applied to real-world business scenarios.

Forecasting dairy product sales by analyzing customer segmentation, demographics, and purchasing patterns. By leveraging basic machine learning techniques, K-means clustering is employed to identify customer groups with similar behaviors, while linear regression is used to model sales trends. The study focuses on uncovering patterns in customer data, such as demographic influences on purchase frequency and seasonal variations in demand. The methods used are straightforward, aiming to provide practical insights for inventory and sales management. Designed as an introductory exploration into data science, this case study demonstrates how even simple analytical techniques can lead to meaningful conclusions and offer value for businesses looking to optimize their sales strategies. The findings emphasize the importance of customer segmentation and trend analysis in better understanding market dynamics within the industry.

Several studies highlight the importance of advanced sales forecasting techniques, providing valuable insights into demand prediction. For instance, **Intelligent Sales Forecasting and Optimization** discusses methods for enhancing product demand prediction and strategic planning. Similarly, **AI-Driven Probabilistic Models for Sales Forecasting** introduces a multi-modal framework, emphasizing the flexibility and adaptability of machine learning in predicting sales for industries like motherboard manufacturing. In addition, **Leveraging Online Consumer Reviews for Sales Forecasting** explores how unstructured data from consumer reviews can improve sales predictions, which suggests a broader application of AI-driven models.

This case study, though basic in its scope, demonstrates how even simple machine learning techniques can yield valuable results for sales **Leveraging Online Consumer Reviews for Sales**

Literature Review

Effective sales forecasting is critical for dairy product manufacturers and retailers due to the challenges posed by dynamic consumer behavior, seasonal trends, and market fluctuations. Insights from various studies underscore the importance of understanding seasonal variations and time lags in sales data, as these factors significantly impact forecasting accuracy[1]. This is particularly relevant for the dairy industry, where demand can spike during specific seasons or events, necessitating precise forecasting to manage inventory effectively.

The use of models that can handle limited data, such as NPFS (New Product Forecasting Systems), is essential for predicting sales of new dairy products that may lack historical data[2]. Implementing a combination of classic and heuristic forecasting methods can help adapt to changing consumer preferences, improving sales predictions for dairy products over time.

Moreover, empirical methods that prioritize simplicity often yield better results than overly complex models. Tailoring forecasting approaches to account for specific dairy product characteristics can enhance the predictive power of these models[3]. This can involve considering customer segmentation and demographics, which are crucial in understanding consumer purchasing patterns in the dairy market.

Incorporating product descriptions and analyzing consumer feedback into forecasting models has been shown to significantly improve accuracy[4]. For the dairy industry, this means that understanding how consumers perceive different dairy products—such as taste preferences or nutritional value—can lead to better sales forecasts. A study involving over 10,000 fashion items demonstrated that integrating detailed product features into forecasting models surpasses existing baselines, suggesting a similar approach could benefit the dairy sector[5].

Utilizing all available data modalities, including sales trends, customer demographics, and product reviews, is confirmed to enhance performance[6]. In the context of dairy, this integrated approach can provide insights into customer preferences and market demands, leading to more accurate forecasting outcomes.

Lastly, the PoissonGP model has been shown to outperform multiple baseline models in key forecasting metrics, effectively managing uncertainty[7]. This capability is particularly valuable in the dairy industry, where market fluctuations can create unpredictable demand patterns. By applying such advanced models, dairy manufacturers and retailers can better navigate uncertainties and improve their sales forecasting processes.

Insights on seasonal variations, time lags, and dynamic product interactions are vital for sales forecasting, with NPFS effectively handling limited data for new products [8]. The system combines classic and heuristic forecasting methods, adapting to new data for improved accuracy [9]. Simpler empirical methods often outperform complex models, with tailored approaches yielding better results [10]. Incorporating product descriptions into forecasting models enhances accuracy, as 44% of the terms significantly impact sales forecasts [11]. A study on over 10,000 fashion items showed that the proposed method surpasses existing baselines in new product sales forecasting [12]. Utilizing all data modalities leads to optimal performance, confirmed by an ablation study [13]. The PoissonGP model outperformed 11 baseline models, achieving significant improvements in key metrics and effectively managing uncertainty [14].

[15]This study examines online reviews' impact on sales for new products (electronics, video games) using Amazon.com panel data. It finds review valence and page views boost sales for "search products," while review volume matters more for "experience products." Marketing strategies are advised, though limited to Amazon. [16]This research evaluates how review quality, exposure, and timing affect sales on Amazon. High-quality reviews have the greatest impact early in a product's lifecycle, with diminishing effects over time as other information sources emerge. [17] Introduces New Product Sales Forecasting Procedure (NPSFP) using a decision-support system (NPFS). The system employs classic and heuristic methods (e.g., Exponential Smoothing) to enhance forecasting accuracy for new products.
[18]Examines time-series forecasting methods (e.g., ARMA, Exponential Smoothing) and Causal Factor Forecasting. The ARMAV with Linear Trend model, incorporating sales inputs, provides the highest accuracy.[19] Integrates data science and machine learning (e.g., clustering, regression) to improve e-commerce sales on social web platforms by analyzing customer

behavior and boosting engagement and conversion rates. [20] Uses big data architecture and neural networks to predict online sales based on review characteristics. Reviewer helpfulness and sentiment strength (especially negative) are key predictors.
[21] Employs XGBoost and feature engineering to forecast dairy sales, showing improved accuracy over other models by leveraging historical data and enhancing predictive capabilities.

Sr.no	Title	Author	Methodology	Observations	Advantages	Disadvantages	Keywords
1	An Intelligent Model for predicting the sales of product.	Avinash Kumar Sharma, Neha Goel,Jatin Rajput,Mohd. Bilal.	Use of K-means Clustering Algorithm for current product analysis. Random Forest tree Algorithm For monitoring product change. Artificial Neural network(ANN) used for new products with lesser data availability.	The study notes that the use of multiple decision trees in the Random Forest reduces the chances of overfitting, compared to using a single decision tree. This is important for producing generalized models that perform well on new, unseen data.	This model can be applied across different products and industries, making it a flexible tool for various sales prediction tasks. Random Forest is capable of capturing non-linear relationships in data, which is essential in complex sales environments.	While Random Forest provides accurate results, the time taken for prediction can be slow, especially with larger forests or datasets. The performance of the model heavily depends on the quality and quantity of the input data, and it may struggle with imbalanced datasets.	Present scenario of product;Future forecast;Prediction;Random Forest Algorithm

2	Analysis and Optimization of Online Sales of Products	Dr.Zainab Pirani,Anuja Marewar, Zainab Bhavnagarwala,Madhuri Kamble	Affinity Analysis: Helps discover co-occurrence relationships . Linear Regression: Provides financial forecasting. Based on the analysis, the tool offers strategies like recommending products, offering discounts, or adjusting prices to increase sales.	Affinity Analysis helps in categorizing fast-selling and slow-selling products. Logistic Regression: Provides predictions on whether products will be sold.Linear Regression aids in strategic planning for future sales.	The tool automates the extraction of sales patterns, making it easier for organizations to manage product inventory and optimize sales. Provides valuable insights into fast-selling products and customer preferences. Managing inventory by identifying popular products and restocking accordingly.	The system's performance depends heavily on the quality and completeness of input data. Limitations of Existing Tools: The study notes that tools using Big Data can be imprecise and prone to data breaches, which could affect performance.	Sales Analytics Tool Data Mining Affinity Analysis Logistic Regression Linear Regression Merchandise Planning
3	Model of the New Sales Planning Optimization and Sales Force Deployment ERP Business Intelligence Module	Marko Velić, Ivan Padavić, and Zrinka Lovrić	The core of the methodology involves: Addressing the traveling salesman problem (TSP) for route optimization. Accumulating sales potentials over	The study presented a case from the Croatian telecommunications market, analyzing client data distribution and potential sales opportunities. It was observed that	The model is self-learning and becomes more accurate over time as more data is fed into the system. It is flexible and adaptable to various industries that deal with recurring	The model does not provide automatic route generation; instead, the sales manager must manually select the routes. The system is based on the assumption	ERP, CRM, direct sales, business intelligence, data mining, traveling salesman problem, sales optimization, call center, decision support system, telecommunication sales, customer intelligence

	for Direct Sales of the Products and Services with Temporal Characteristics		time to determine optimal sales periods. Creating visualizations to assist in decision-making regarding sales routes and customer engagement timing	many geographical areas were underserved, especially remote locations. The model's application enabled better sales route planning and the identification of optimal times for sales force deployment. Over time, the system becomes more accurate as it learns from historical data	services and products. The visual approach and decision support system help sales managers make informed decisions about deployment strategies and customer engagement. Helps optimize travel routes and sales schedules, reducing operational costs	that geographical and sales data are consistently available and updated, which may not always be the case. The lack of automatic route calculations could be a limitation in more complex or larger-scale operations	
4	A Product Recommendation Method by Analyzing Sales Volume, Sales Period, and User Satisfaction	Haoyang Xia, Yuanyuan Wang	The Google Cloud Natural Language API was used to analyze sentiment in product reviews and calculate satisfaction scores. The number of reviews is used as a proximate for	Products with higher satisfaction, larger sales volumes, and longer sales periods are ranked higher in recommendations. The experimental results showed that	The method integrates multiple factors (sales, user satisfaction, and product lifespan), which results in a more well-rounded product ranking system. It accounts	Sales volume is estimated from the number of reviews, which may not always accurately reflect actual sales. Limited Personalization: The method assumes equal weighting of	e-commerce, product recommendation, sales volume, sales period, user satisfaction

			sales volume, while timestamps of reviews estimate the sales period. Products are assigned scores based on these factors.	this method works well but has limitations in reflecting user preferences, as different users prioritize different factors.	for real user experiences through sentiment analysis making it more user-focused. This approach is adaptable for various product categories and can handle large datasets	all features, which may not align with individual user preferences, leading to suboptimal recommendations in some cases.	
5	Time Series Forecasting of Agricultural Products Sale Using Deep Learning	Md. Touhidur Rahman, Mst. Eshita Khatun, Archina Mahmuda Asha, Mohammad	Collected 832 entries with 13 attributes such as month, product type, and price. Exploratory Data Analysis: Statistically analyzed the data to identify the most sold products by region and size. Multi-Step LSTM was used for time series forecasting to predict sales for the next month.	"AIM Goal" was the highest-selling product, while "Gourlin 30 EC" was the least sold. Bogura was identified as the highest-selling region. Model Performance: After training the LSTM model with 20 epochs, the final prediction showed good results.	The use of Multi-Step LSTM provides an effective solution for time series forecasting. The statistical analysis helps companies identify the best-selling products, regions, and optimal package sizes, supporting better business planning.	The dataset is restricted to one agricultural company and a short time frame, which may affect the generalizability of the model.	Product Sale Forecasting, Agriculture Product, Time Series Models, LSTM

6	Research on Commodity Sales Forecasting Based on Combination Model	Shasha Mo, Luqiao Zhang, Yi Xiang, Xingyu Lu, Chunling Li	The sales data is preprocessed.PCA is applied to reduce the dimensionality of the input features. The results of the LSTM and XGBoost models are combined using a linear weighted combination approach, where the weights are determined by the entropy method.	The combination model (PCA_LSTM-XGBoost) achieves better predictive performance than individual models. The PCA helps reduce the complexity of the LSTM model by eliminating low-correlation features, which prevents overfitting. The experimental results show that the combination model reduces the Mean Absolute Percentage Error by 8% compared to traditional methods like SARIMA and Prophet.	The combination model significantly improves prediction accuracy.PCA reduces computational complexity by lowering the number of input features. XGBoost includes regularization to prevent overfitting, and PCA helps reduce unnecessary feature noise.	The combination of PCA, LSTM, and XGBoost increases the complexity of implementation. Training an LSTM and XGBoost model requires significant computational power.	Sales forecasting;XGBoost model; LSTM neural network;Combination prediction model; PCA (Principal Component Analysis)

7	A Comparison of Data Mining Approaches for Forecasting Sales of FMCG Food Products	Milan Doshi, Atmiya University, India	Regression Analysis to identify relationships between dependent and independent variables. K-Means Clustering helps to forecast demand by identifying patterns in customer groups. Tracking Patterns: This technique identifies trends and recurring behaviors in sales data. Visualization is further done.	Regression Analysis: can be influenced by noisy data. Tracking Patterns: This technique performed the best as it effectively captured recurring trends in the sales data. Visualization helped communicate findings and results more clearly.	High accuracy in identifying trends and forecasting sales. K-Means Clustering segments customers into groups that are useful for targeted marketing and inventory management. Visualization: Makes data trends easier to understand, aiding in decision-making.	Regression Analysis and K-means algorithm are Sensitive to noisy data, which can result in inaccurate predictions.The model is less effective during unforeseen events like floods or earthquakes, which disrupt normal sales patterns.	Data Mining FMCG (Fast-Moving Consumer Goods) Food Products Sales Forecasting K-Means Clustering Regression Analysis Tracking Patterns Visualization Techniques
8	Demand Planning and Sales Forecasting for Motherboard Manufacturers	C.-H. Wang and Y. Yun	This research paper focuses on improving how motherboard manufacturers predict future demand and sales using a mix of advanced methods. It	Their insights on capturing seasonal variations, time lags, and dynamic interactions between products and competitors as general principles	**1.Dynamic Interactions**(Compitetor interactions) **2.Improved accuracy**.(Hybrid Model) **3.Sensitivity analysis** : The model allows for sensitivity	1.Traditional qualitative methods rely heavily on expert opinions. 2.Some quantitative methods fail to capture dynamic interactions.	Demand Planning, Sales Forecasting, Time-Series Models, Dynamic Interactions, Hybrid Models, Supply Chain Analytics.

			combines time-series forecasting (like ARIMA), machine learning models (like Random Forest and Gradient Boosting), and a biological model (Lotka-Volterra) to capture changes over time.	applicable to sales forecasting in your industry.	analysis, helping companies see how changes in factors (e.g., market conditions) can affect forecasts, aiding better decision-making	3.Limited consideration of external factors in demand planning.	
9	New Product Sales Forecasting Procedure (NPSFP)	C. Ching-Chin et al.	The NPFS employs a structured **five-step process:** 1. Data Collection: Gather historical sales data and related information. 2. Parameter Determination: Identify parameters for various forecasting methods. 3. Sales Forecast Calculation: Use selected methods to	1.The NPFS effectively handles **limited data scenarios** typical for new products. 2.It combines both classic and heuristic forecasting methods to enhance accuracy. 3.The system adapts to new data inputs, refining forecasts over time.	1.Improved forecasting accuracy for new products. 2.Flexibility in applying various methods based on available data. 3.Reduces reliance on human judgment.	1.Limited historical data can still affect forecast reliability. 2.Complexity in selecting the best method from multiple options.	New Product Sales Forecasting, Time Series Methods, Heuristic Methods, Forecasting Accuracy, Data Handling.

			generate forecasts. 4. Subjective Adjustments: Allow experienced managers to adjust forecasts based on market insights. 5. Evaluation: Assess forecast performance using metrics like Mean Absolute Percentage Error (MAPE).				
10	Probabilistic Forecasting Methods for Product Sales: Insights from the M5 Competition	Evangelos Spiliot, Spyros Makridakis b, Anastasios Kaltsounis a, Vassilios Assimakopoulos	1.Evaluates various forecasting methods using M5 competition data, including **empirical and statistical techniques.** 2.Applies the Multiple Comparisons with the Best (MCB) test to assess performance across quantiles.	1.Simpler empirical methods often outperform complex models in accuracy and efficiency. 2.Performance varies by quantile, with no single best method for all scenarios. 3.Tailored methods for specific data characteristics yield better	1.Provides insights into the most effective forecasting methods for different data series. 2.Highlights the performance of simpler empirical methods over complex machine learning models. 3.Offers a	1.Performance varies significantly by quantile, indicating no universally superior method. 2.Some methods may require extensive parameter tuning, complicating their application. 3.Results may not be	Probabilistic forecasting, product sales, M5 competition, empirical methods, stock-control simulation, quantile regression, forecasting performance.

				results.	comprehensive evaluation of forecasting accuracy and computational efficiency.	generalizable across all types of sales data.	
11	Which product description phrases affect sales forecasting? An explainable AI framework by integrating WaveNet neural network models with multiple regression	Shan Chen, Shengjie Ke, Shuihua Han, Shivam Gupta, and Uthayasankar Sivarajah	1.The study proposes an explainable AI framework that combines text mining, WaveNet neural networks, multiple regression, and SHAP model to analyze the impact of product descriptions on sales forecasting. 2.The framework consists of **four steps:** data processing, WaveNet model construction, multiple regression analysis, and SHAP model interpretation. 3.The study uses a dataset	1.The study finds that adding product descriptions to the forecasting model improves forecasting accuracy. 2.The results show that approximately 44% of the terms used in product descriptions can significantly influence the level of sales forecasts. 3.The study identifies 28 BOW vectors with significant effects on sales forecasting and adds them to the WaveNet	1.The proposed framework provides an explainable approach to sales forecasting, allowing companies to understand why product descriptions can influence sales forecasting and how to improve product descriptions to drive sales. 2.The study demonstrates the effectiveness of using text mining and deep learning techniques in sales forecasting. 3.The framework can help	1.The study notes that the BOW vector obtained by text mining of product descriptions had a smaller impact on product sales compared to historical product sales and price variables. 2.The study suggests that future research should explore the reasons for this and extend the research to consider inter-SKU factors in the predictor variables.	Sales forecasting, product descriptions, text mining, WaveNet neural networks, multiple regression, SHAP model, explainable AI, e-commerce, cross-border trade.

			of 192,876 sales records from a cross-border e-commerce firm and extracts product descriptions from the e-commerce platform.	model to form a new model (Model C).	companies identify key phrases that impact sales forecasting and optimize product description information on e-commerce platforms.		
12	Multi-Modal Transform-Based Fusion Model for New Product Sales Forecasting	Xiangzhen Li, Jiaxing Shen, Dezhi Wang, Wu Lu, Yuanyi Chen	1The authors propose a multi-modal transform-based fusion model (M2TFM) for new product sales forecasting. 2.M2TFM integrates multiple data sources, including product images, attributes, text descriptions, and context factors like holidays, weather, and trends. 3.The model	1.The authors conducted a comprehensive evaluation on a large e-commerce dataset with more than 10,000 fashion items. 2.The results demonstrate that the proposed method is more effective than existing state-of-the-art baselines for new product sales forecasting. 3.The authors performed an ablation study	1.M2TFM can effectively capture visual and textual signals from product images and attributes. 2.The model facilitates cross-modal interactions, allowing for a bidirectional exchange of semantic information between different modalities. 3.The incorporation of temporal	1.The performance of the M2TFM model may be affected by incomplete or missing data. 2.The model has high complexity, which may lead to difficulties in its interpretation and explanation. 3.M2TFM requires a large amount of labeled data for training, which can be a challenge in	multi-modal fusion, transform-based fusion model, new product sales forecasting, diffusion embedding, transformer self-attention mechanism, e-commerce dataset, ablation study, temporal context modeling, seasonality, consumer preferences.

			leverages diffusion embedding to fuse heterogeneous data modalities into a unified representation that models their complex interactions. 4.M2TFM uses a transformer self-attention mechanism to extract nuanced signals across modalities to make more accurate new product sales forecasts.	to evaluate the impact of different feature types on product sales prediction. 4.The study found that using all data modalities (text, images, attributes, temporal signals, and contextual data) achieves the best performance.	context modeling provides an additional layer of information that is crucial for sales forecasting. 4.M2TFM can capture the seasonality and evolving consumer preferences within different product categories.	practical application scenarios.	
13	A multiple long short-term model for product sales forecasting based on stage future vision	Daifeng Li, Xuting Li, Kaixin Lin, Jianbin Liao, Ruo Du, Wei Lu, Andrew Madden	1.Proposes a novel model called SLST-PKNet (Stage future-vision-based multiple Long Short-term model with prior knowledge) for product	1.Tested on two industrial datasets (Galanz and Cainiao) and two public datasets (Traffic and Exchange-Rate) 2.Outperformed 11	1.Considers different types of **time series data** 2.Models **future unknown dependencies** effectively 3.Captures both linear	1.Higher computational complexity than some simpler models 2.Running speed is not very high 3.May still be	Sales forecasting, time series prediction, long short-term memory, convolutional neural networks, prior knowledge, multivariate time series

	with prior knowledge		sales forecasting 2.Uses three sub-models to consider different types of time series data 3.Incorporates a two-layer convolutional neural network (TLCNN) and two-stage LSTM (TSLSTM) to model future dependencies 4.Includes a dynamic co-integration (DCI) mechanism to capture linear correlations 5.Integrates prior knowledge (PK) about seasonal and promotional influences	state-of-the-art baseline models on various metrics 3.Achieved significant improvements in RMAE, RRSE and CORR metrics compared to baselines 4.Ablation studies showed the contributions of different components	and non-linear patterns 4.Incorporates domain knowledge to improve predictions 5.Performs well on both high and low sales products	influenced by noisy information in sparse time series	

14	Bayesian non-parametric method for decision support: Forecasting online product sales	Ziyue Wu, Xi Chen, Zhaoxing Gao	1.Proposes PoissonGP, a novel Bayesian model that uses a non-homogeneous Poisson process with a Gaussian process prior for sales prediction 2.Combines Poisson intensity and Gaussian process prior 3.Uses additive property of Gaussian process prior to make disentangled predictions without explicitly decomposing time series data 4.Provides prediction intervals in addition to point forecasts	**1.PoissonGP outperformed existing approaches on synthetic and empirical datasets** 2.Performed well on data with multiple trends and distribution shifts **3.Provided prediction intervals to assess uncertainty** 4.Avoided error accumulation caused by separate decomposition and prediction	1,Flexible in dealing with **complex, multi-trend time series data** 2.Manages distribution shifts caused by changes in long-run sales 3.Incorporates forecast uncertainty 4.Provides interpretability for decision support systems 5.Does not require explicit decomposition of time series	1.Requires specifying priors and setting Gaussian process components based on assumptions or prior knowledge 2.May be computationally intensive compared to simpler models	Time series analysis, Gaussian process, Bayesian non-parametric method, Online product sales, Forecasting, Decision support systems

15	The Effect of Online Consumer Reviews on New Product Sales	Geng Cui, Hon-Kwong Lui, and Guo	Utilized panel data of 332 new products from Amazon.com over nine months. Employed fixed effects models with lagged variables to assess the impact of online reviews on sales.	Volume and valence of online reviews significantly influence consumer purchasing decisions. Negative reviews spread faster and have a more substantial impact than positive reviews. The effect of reviews diminishes over time after a product's launch.	Provides a comprehensive framework for understanding the role of eWOM in product sales. Offers insights into the differences between experience and search products. Highlights the importance of early consumer feedback for new product launches.	Data collected from only one online retailer (Amazon.com), limiting generalizability. Findings may not apply to other e-commerce platforms or product categories. Potential biases in consumer reviews and ratings not fully addressed.	Online Consumer Reviews eWOM (Electronic Word of Mouth) New Product Sales Experience Products Search Products Marketing Strategies Negativity Bias
16	Do Online Reviews Affect Product Sales?	Nan Hu Æ Ling Liu Æ Jie Jennifer Zhang	Data collected from Amazon.com's Web Service (AWS) focusing on books, DVDs, and videos. Utilized panel data to analyze the dynamics of online	Online reviews significantly influence sales, with varying effects based on product age and reviewer characteristics.	Provides a comprehensive understanding of how online reviews affect sales. Highlights the importance of both quantitative	Results may not be generalizable across different product categories (e.g., books vs. DVDs). Does not consider the textual content or length of	Online reviews, product sales, consumer behavior, reviewer quality, transaction cost economics, panel data, marketing.

			reviews and their impact on sales over time. Considered factors such as reviewer quality, exposure, product coverage, and product age.	Consumers utilize both quantitative and qualitative information from reviews. The market responds more to reviews from reputable reviewers and less to extensively covered items.	and qualitative aspects of reviews. Offers insights for marketers on leveraging online reviews effectively.	reviews, which may also affect review quality.	
17	Designing a decision-support system for new product sales forecasting	Chern Ching-Chin*, Ao Ieong Ka Ieng, Wu Ling-Ling, Kung Ling-Chieh	The NPFS employs a four-step process: Collect and Analyze Data, Determine Parameters for Forecasting Methods, Calculate Sales Forecast, and Adjust Results Subjectively. It utilizes classic methods (Moving Average, Exponential Smoothing) and heuristic	The NPFS demonstrated improved forecasting accuracy compared to traditional methods, particularly in scenarios with limited sales data. Variations caused by seasonality, promotional events, and other external factors were identified and minimized to enhance	Increased forecast accuracy by reducing data noise. Utilizes both classic and heuristic forecasting methods. Provides a structured approach to sales forecasting.	Requires sufficient historical data for effective forecasting. May still be influenced by unpredictable external factors. Complexity in parameter determination and method selection.	New Product Forecasting, Sales Forecasting, NPFS, Exponential Smoothing, Moving Average, Heuristic Methods, Forecast Accuracy, Data Variations.

			methods (Sales Index, Diffusion Model) to improve forecast accuracy by reducing data variations.	results.			
18	Data Mining Algorithms and Statistical Analysis for Sales Data Forecast	Lin Wu; JinYao Yan; YuanJing Fan	The study compares various forecasting models, including Exponential Smoothing, Holt's Linear Method, ARMA, and ARMAV with linear trend, to analyze their accuracy in predicting sales data for new consumer electronics products.	ARMAV with linear trend provided the best forecasting accuracy and lowest residual sum of squares (RSS), while traditional methods like Exponential Smoothing tended to underestimate increasing trends.	ARMAV with linear trend incorporates both trend and input factors, leading to improved accuracy. Exponential Smoothing is easy to understand and implement.	Exponential Smoothing is limited to historical data and cannot incorporate explanatory factors. ARMA models are complex and require programming for implementation.	Forecast; Time-Series Forecasting; Causal Factor Forecasting; ARMA; ARMAV; Sales Data; Consumer Electronics.

19	DATA SCIENCE AND MACHINE LEARNING APPROACH TO IMPROVE E-COMMERCE SALES PERFORMANCE ON SOCIAL WEB	Hussain Saleem Khalid Bin Muhammad Altaf Hussain Nizamani Samina Saleem Jamshed Butt	The study employs data analytics tools, A/B testing, and customer behavior analysis to develop hypotheses for improving e-commerce sales.	The research identifies key factors affecting customer engagement and conversion rates, emphasizing the importance of understanding user behavior through data.	Improved customer experience through targeted strategies. Enhanced decision-making based on data-driven insights. Increased sales conversion rates via optimized processes.	Challenges in logistics and payment security. Potential for low sales conversions despite analytics efforts. Dependence on accurate data collection and analysis.	E-commerce, Data Science, Machine Learning, Customer Behavior, A/B Testing, Conversion Rates, Social Web.
20	Understanding and Predicting Online Product Sales	Fangfang Hou, Boying Li, Alain Yee-Loong Chong, Natalia Yannopoulou & Martin J. Liu	Sentiment analysis of online reviews. Neural network analysis to predict sales based on review characteristics.	Helpful votes and reviewer pictures significantly influence sales. Sentiment strength and polarity are important predictors.	Provides insights for online sellers to manage businesses effectively. Utilizes big data architecture applicable to various research contexts.	Limited to data with review comments (only 6,000 out of a larger dataset). May not account for all variables affecting sales.	Big data, neural network, online reviews, product demands, online marketplace, reviewer characteristics.

21	Sales Forecasting Using XGBoost	Yiyang Niu	The study employs the XGBoost machine learning algorithm for sales forecasting, utilizing feature engineering and data mining techniques to predict sales of products and commodities.	The study employs the XGBoost machine learning algorithm for sales forecasting, utilizing feature engineering and data mining techniques to predict sales of products and commodities.	Improved accuracy in sales forecasting. Effective feature ranking for better model performance. Utilizes advanced machine learning techniques.		Sales Forecast, XGBoost, Machine Learning Algorithms, Data Mining.

Conclusion:

In conclusion, this case study demonstrates the value of employing straightforward machine learning techniques for sales forecasting in the dairy industry. By analyzing customer segmentation, demographics, and purchasing patterns, businesses can gain valuable insights into consumer behavior and demand fluctuations. The literature highlights the importance of capturing seasonal variations and dynamic interactions while showcasing the effectiveness of both classic and heuristic forecasting methods. Simplified approaches often yield superior results compared to complex models, underscoring the significance of tailoring methods to specific data characteristics. Furthermore, the incorporation of additional features, such as product descriptions and multiple data modalities, enhances forecasting accuracy, as evidenced by studies across various industries. Ultimately, this exploration

References:

1.A. K. Sharma, N. Goel, J. Rajput and M. Bilal, "An Intelligent Model For Predicting the Sales of a Product," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 341-345, doi: 10.1109/Confluence47617.2020.9057880. keywords: {Cloud computing;Data science;Conferences;Forestry;Prediction algorithms;Production;Organizations;Present Scenario of product;Future forecast;Prediction;Random Forest Algorithm}, 2.Z. Pirani, A. Marewar, Z. Bhavnagarwala and M. Kamble, "Analysis and optimization of online sales of products," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICIIECS.2017.8276165. keywords: {Tools;Data mining;Market research;Organizations;Databases;Clothing;Sales Analytics tool;data mining;affinity analysis;logistic regression;linear regression;merchandise planning},

3.M. Velić, I. Padavić and Z. Lovrić, "Model of the new sales planning optimization and sales force deployment ERP business intelligence module for direct sales of the products and services with temporal characteristics," Proceedings of the ITI 2012 34th International Conference on Information Technology Interfaces, Cavtat, Croatia, 2012, pp. 91-96, doi: 10.2498/iti.2012.0379. keywords: {Marketing and sales;Cities and towns;Planning;Companies;Telecommunications;Sorting;ERP;CRM;direct sales;business intelligence;data mining;travelling salesman problem;sales optimization;call centre;decision support system;telecommunication sales;customer intelligence},

4.H. Xia and Y. Wang, "A Product Recommendation Method by Analyzing Sales Volume, Sales Period, and User Satisfaction," 2024 12th International Conference on Information and Education Technology (ICIET), Yamaguchi, Japan, 2024, pp. 407-411, doi: 10.1109/ICIET60671.2024.10542781. keywords: {Sentiment analysis;Electronic learning;Correlation;Reviews;Education;Electronic commerce;Reliability;e-commerce;product recommendation;sales volume;sales period;user satisfaction},

5.M. T. Rahman, E. Khatun, A. M. Asha and M. Jahangir Alam, "Time Series Forecasting of Agricultural Products Sale Using Deep Learning," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10307538. keywords: {Deep learning;Machine learning algorithms;Computational modeling;Time series analysis;Companies;Production;Predictive models;Product Sale Forecasting;Agriculture Product;Time Series Models;LSTM},

6.S. Mo, L. Zhang, Y. Xiang, X. Lu and C. Li, "Research on commodity sales forecasting based on combination model," 2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA), Dalian, China, 2022, pp. 124-128, doi: 10.1109/ICDSCA56264.2022.9988375. keywords: {Solid modeling;Computational modeling;Time series analysis;Neural networks;Predictive models;Market research;Data models;sales forecasting;XGBoost model;LSTM neural network;combination prediction model},

7.M. Doshi, "A Comparison of Data Mining Approaches for Forecasting Sales of FMCG Food Products," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi:

10.1109/ICCCNT56998.2023.10307432. keywords: {Measurement;Supply chains;Clustering algorithms;Companies;Prediction algorithms;Food products;Data mining;Data Mining;FMCG (Fast Moving Consumer Goods);Food Products;Sales Forecasting;Techniques;K-Means Clustering;Tracking Patterns;visualization techniques;Regression analysis techniques},

8.A two-stage prediction model based on behavior mining in livestream e-commerce
Qinping Lin , Ning Jia , Liao Chen , Shiquan Zhong , Yuance Yang , Tong GaoCollege of Management and Economics, Tianjin University, Tianjin, 300072, China

bLaboratory of Computation and Analytics of Complex Management Systems (CACMS), Tianjin University, Tianjin 300072, China
cHIAI Co. Ltd., Tianjin 300072, China

9.Integrating human judgement into quantitative forecasting methods: A review
Authors :Meysam Arvan, Behnam Fahimnia, Mohsen Reisi, Enno Siemsen Institute of Transport and Logistics Studies, The University of Sydney, Darlington, NSW 2000, Australia
B Wisconsin School of Business, University of Wisconsin, 975 University Ave, Madison, Wisconsin 53706, USA

10.M5 accuracy competition: Results, findings, and conclusions
Spyros Makridakis b, Evangelos Spiliotis a, Vassilios Assimakopoulos a 2022, International Journal of Forecasting

11.The value of data, machine learning, and deep learning in restaurant demand forecasting: Insights and lessons learned from a large restaurant chain
Bongsug (Kevin) Chae a, Chwen Sheu a, Eunhye Olivia Park b
Department of Management, College of Business, Kansas State University, 1301 Lovers Lane, Manhattan, KS 66506, United States of America
b Department of Food Nutrition, College of BioNano Technology, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, South Korea

12.Introduction to the special issue on recent advances on digital economy-oriented artificial intelligence
Yu-Lin He, Philippe Fournier-Viger, Sebastian Ventura, Ling Zhang
Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen, 518107, China
Big Data Institute, College of Computer Science & Software Engineering, Shenzhen University, Shenzhen, 518060, China
Department of Computer Science and Numerical Analysis, University of Cordoba, Spain
School of National Finance, Guangdong University of Finance, Guangzhou, 510521, China

13.Multi-modal transform-based fusion model for new product sales forecasting
Xiangzhen Li, Jiaxing Shen , Dezhi Wang, Wu Lu, Yuanyi Chen
Institute of Digital Finance, Hangzhou City University, Hangzhou, 310015, Zhejiang, China

b Hangzhou City University, Hangzhou, 310015, Zhejiang, China
C Lingnan University, 8 Castle Peak Road, Tuen Mun, New Territories, Hong Kong
D State Key Laboratory of Public Big Data, Guizhou University, 550025, Guizhou, China

14.The value of data, machine learning, and deep learning in restaurant demand forecasting: Insights and lessons learned from a large restaurant chain
Bongsug (Kevin) Chae a, Chwen Sheu a, Eunhye Olivia Park b
Department of Management, College of Business, Kansas State University, 1301 Lovers Lane, Manhattan, KS 66506, United States of America
b Department of Food Nutrition, College of BioNano Technology, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, South Korea

15.Cui, Geng & Lui, Hon-Kwong & Guo, Xiaoning. (2012). The Effect of Online Consumer Reviews on New Product Sales. International Journal of Electronic Commerce. 17. 39-58. 10.2307/41739503.

16.Hu, Nan & Liu, Ling & Zhang, Jie. (2008). Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects. Information Technology and Management. 9. 201-214. 10.1007/s10799-008-0041-2.

17.Chern Ching-Chin, Ao Ieong Ka Ieng, Wu Ling-Ling, Kung Ling-Chieh,
Designing a decision-support system for new product sales forecasting,
Expert Systems with Applications,
Volume 37, Issue 2,
2010,
Pages 1654-1665,
ISSN 0957-4174,
https://doi.org/10.1016/j.eswa.2009.06.087.
(https://www.sciencedirect.com/science/article/pii/S0957417409006034)

18.L. Wu, J. Y. Yan and Y. J. Fan, "Data Mining Algorithms and Statistical Analysis for Sales Data Forecast," 2012 Fifth International Joint Conference on Computational Sciences and Optimization, Harbin, China, 2012, pp. 577-581, doi: 10.1109/CSO.2012.132. keywords: {Marketing and sales;Predictive

models;Forecasting;Data models;Mathematical model;Analytical models;Accuracy;Time-Series Forecasting;Causal Factor Forecasting;ARMA;ARMAV},

19.Saleem, Hussain & Muhammad, Khalid & Nizamani, Altaf & Saleem, Samina & Butt, Jamshed. (2021). Data Science and Machine Learning Approach to Improve E-Commerce Sales Performance on Social Web. INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ENGINEERING & TECHNOLOGY. 12. 401-424. 10.34218/IJARET.12.4.2021.040.

20.Fangfang Hou, Boying Li, Alain Yee-Loong Chong, Natalia Yannopoulou & Martin J. Liu (2017) Understanding and predicting what influence online product sales? A neural network approach, Production Planning & Control, 28:11-12, 964-975, DOI: 10.1080/09537287.2017.1336791

21.Y. Niu, "Walmart Sales Forecasting using XGBoost algorithm and Feature engineering," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Bangkok, Thailand, 2020, pp. 458-461, doi: 10.1109/ICBASE51474.2020.00103. keywords: {Measurement;Machine learning algorithms;Software algorithms;Machine learning;Predictive models;Prediction algorithms;Data models;Sales Forecast;XGBoost;Machine learning algorithms;Data Mining},