Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 31s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A Study on Sales Forecasting Through Social Media Trends Using Power Bi

Tharik Azeez R[1], Dr. K. Selvavinayaki[2], Dr. M Sheela Newsheeba[3], Suhas K S[4], Sandeep Kumar[5], Vimal Raj G[6], Shan Raymond B[7], Kailas[8]

[1]Project Student.

[2]Head, Department of Computer Applications.

Email ID: nascselvavinayaki@nehrucolleges.com

[3]Assistant Professor.

Email ID: sheela79.rincy@gmail.com

[4,5,6,7,8]Project Student.

Cite this paper as: Tharik Azeez R, Dr. K. Selvavinayaki, Dr. M Sheela Newsheeba, Suhas K S, Sandeep Kumar, Vimal Raj G, Shan Raymond B, Kailas, (2025) A Study on Sales Forecasting Through Social Media Trends Using Power Bi, *Journal of Neonatal Surgery*, 14 (31s), 548-553

ABSTRACT

In today's fast-paced and data-driven world, consumers seek innovative tools to make informed purchasing decisions. This project introduces a Sales Forecasting Application that empowers users to predict sales trends, identify the best purchase opportunities, and make smarter shopping choices. By integrating Artificial Intelligence (AI), machine learning, and data analytics, the application offers consumers a personalized experience to optimize budgets and leverage market trends.

The application utilizes time-series forecasting models (e.g., ARIMA, Prophet) to analyse historical data and predict future sales patterns. It incorporates real-time social media trends and customer sentiment analysis using Natural Language Processing (NLP) to assess public perception and its impact on product demand. This enhances forecast accuracy and keeps users updated on trending products.

Another key feature of the application is its ability to perform price recognition using Optical Character Recognition (OCR) technology. This enables users to extract and analyse price data directly from product images, such as photos of labels, receipts, or promotional banners. The extracted data is seamlessly integrated into the application's analysis, providing users with insights into pricing fluctuations and helping them identify the best deals available in the market.

The application's integration with Power BI takes its functionality to the next level by offering users highly interactive and visually appealing dashboards. These dashboards are designed to provide an intuitive experience, allowing consumers to:

- Track and visualize historical sales data with future forecasts.
- Correlate social media sentiment trends with product sales and pricing.
- Identify seasonal patterns and trends for specific products or categories.
- Customize analyses by time periods, brands, or product categories.
- Access real-time, dynamically updated visualizations for accurate insights.

By delivering actionable insights in an easy-to-understand format, the Power BI integration ensures that even non-technical users can benefit from the application's features. Additionally, the application provides options for users to export data and reports, enabling them to share insights or further analyse the results independently.

This Sales Forecasting Application bridges the gap between AI-driven analytics and consumer decision-making. It empowers users to make smarter purchases, save money, and stay ahead of trends. With advanced forecasting, intuitive dashboards, and real-time insights, it is a vital tool for navigating today's dynamic marketplace.

1. INTRODUCTION

In today's competitive retail environment, accurate sales forecasting is crucial for effective inventory management and strategic planning. This project proposes a comprehensive sales forecasting model that integrates autoregressive integrated moving average (ARIMA) with sentiment analysis and optical character recognition (OCR) for enhanced predictive

accuracy. The study focuses on a dataset comprising sales data for ten distinct products from various stores, employing the ARIMA model to predict future sales trends based on historical data. To further enrich the forecasting process, sentiment analysis is conducted using Twitter data related to these products, enabling the identification of consumer sentiment and its potential impact on sales. A machine learning model is trained to classify sentiments as positive, negative, or neutral, allowing for a nuanced understanding of public perception.In addition, this project incorporates a hybrid OCR extraction technique to gather real-time price information from images of product labels or advertisements. The OCR system captures textual data, extracting prices to facilitate sentiment prediction through contextual analysis. By combining sales data with consumer sentiment and price information, the model aims to produce more reliable sales forecasts. Graphical representations of the forecasted sales trends, along with sentiment classifications, will be visualized to provide a clear understanding of how external factors influence product performance. This multidisciplinary approach not only leverages statistical modeling and machine learning but also highlights the importance of market sentiment and visual data extraction in developing robust sales forecasting methods

SYSTEM ANALYSIS

1.2.1 Existing System

The existing systems that consumers rely on for making purchasing decisions are fragmented and lack the advanced features necessary for accurate sales forecasting and real-time insights. Below is an evaluation of the key existing systems:

Price Comparison Websites

Price comparison websites, such as Google Shopping or PriceGrabber, allow users to compare product prices across multiple online retailers. These platforms provide quick comparisons of prices and help users identify cheaper options for specific products. However, they lack forecasting capabilities to predict future price trends or sales patterns. Additionally, they do not integrate with social media sentiment, which can significantly influence product demand. Price comparison websites also focus primarily on major e-commerce sites, excluding local retailers and niche markets, which limits their overall effectiveness.

Social Media Analytics Tools

Social media analytics tools like Hootsuite or Brandwatch specialize in analysing social media trends and public sentiment. These tools offer valuable insights into trending products, hashtags, and discussions, making them useful for businesses to track brand performance. However, they are not designed for individual consumers and do not integrate pricing or sales data. Their complex interfaces and high subscription fees reduce accessibility for everyday users, further limiting their potential as a standalone tool for consumers looking to make informed purchasing decisions.

Manual Forecasting or Shopping Practices

Many consumers rely on intuition, historical experiences, or well-known seasonal sales patterns for purchasing decisions. This approach is simple and does not require technical expertise. However, it lacks data-driven insights, which often leads to less accurate decisions. Manual forecasting is prone to human error and personal biases and cannot account for unforeseen market changes or real-time influences, making it less reliable compared to data-backed predictive models.

Receipt or Label Scanning Apps

Mobile applications, such as Expensify or Shoeboxed, use Optical Character Recognition (OCR) to scan and extract price data from receipts and product labels. These tools automate the process of extracting data from physical receipts or images, which can be useful for expense tracking and personal budgeting. However, they do not integrate with sales forecasting or social sentiment analysis. Furthermore, they have limited functionality, focusing solely on recording prices rather than offering insights or actionable recommendations to help consumers make better purchasing decisions.

Limitations of the Existing System

Despite offering specific benefits, the existing systems fail to provide a comprehensive solution for consumers. Some key limitations include:

1. **Lack of Predictive Analytics**: Existing tools do not leverage historical data and trends to predict future sales or price fluctuations, leaving consumers without a reliable way to forecast market behaviour.
2. **Fragmented Tools**: Consumers are forced to use multiple tools, such as price comparison websites, social media monitoring platforms, and receipt trackers, without a centralized system offering consolidated insights, making the process time-consuming and inefficient.
3. **No Real-Time Updates**: Most existing systems lack dynamic updates based on live data, such as real-time social media sentiment or price changes, which diminishes their relevance and accuracy.
4. **Limited Accessibility for Consumers**: Many tools are designed for business use and are too complex for individual shoppers. The high costs of advanced analytics platforms also make them prohibitive for everyday consumers.

5. **Absence of Personalization**: Existing systems do not offer tailored recommendations or insights based on individual purchasing preferences or habits, which limits their usefulness for consumers who need customized advice.

1.3 PROPOSED METHODOLOGY& ARCHITECTURE

The proposed Sales Forecasting Application aims to bridge the gaps present in existing systems by combining predictive analytics, price recognition, and social media sentiment analysis into a unified, user-friendly platform. This application will provide consumers with real-time insights, allowing them to make informed purchasing decisions, track sales trends, and optimize their shopping behaviour. Below is an in-depth overview of the key features and functionalities of the proposed system:

1. Sales Forecasting and Predictive Analytics
2. Social Media Sentiment Analysis Integration
3. Price Recognition Using OCR Technology
4. Power BI Integration for Interactive Dashboards
5. Real-Time Data and Dynamic Insights
6. User-Friendly Interface and Personalization
7. Mobile and Web Accessibility

Advantages of the Proposed System

- Combines social media, sales trends, and pricing data into a single platform.
- Provides accurate forecasts using AI-driven predictive models.
- Enhances decision-making with real-time, actionable insights.
- Empowers users with interactive and customizable Power BI dashboards.
- Saves time and improves budgeting for everyday consumers.

This comprehensive solution bridges the gaps in existing systems, offering a powerful and user-friendly tool for smarter consumer decision-making.

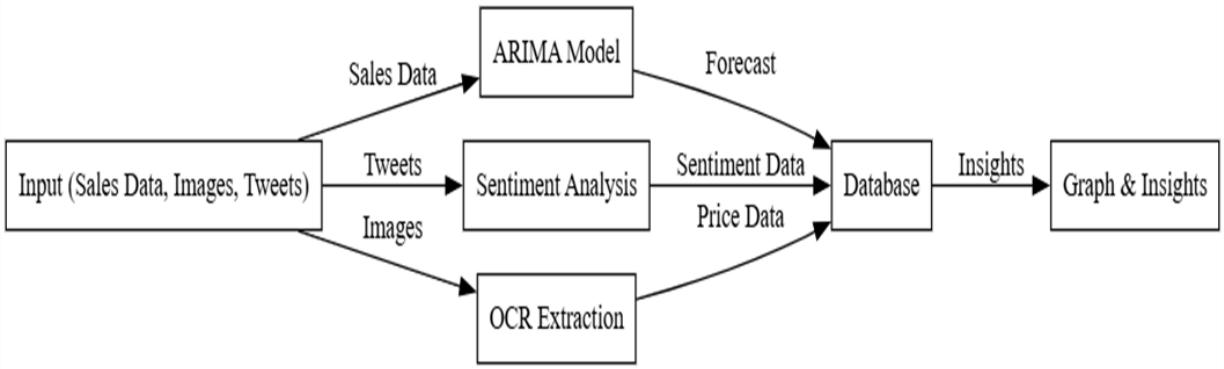


Fig 1: Proposed System Architecture

1.3.2 Input Design

The sales forecasting system combines ARIMA-based predictive analytics, sentiment analysis of social media data, and hybrid OCR extraction. The system consists of three modules: Sales Prediction, Sentiment Analysis, and Price Extraction.

1. **Sales Prediction Module**: Historical sales data from 10 store products is collected. Using the ARIMA model, the system identifies patterns and seasonality to forecast future sales. The output includes graphical representations of trends, aiding in inventory and marketing strategies.
2. **Sentiment Analysis Module**: Twitter data is collected and preprocessed to remove noise. A machine learning model is trained to classify sentiments (positive, negative, neutral). Sentiment trends are correlated with sales patterns to provide actionable insights.

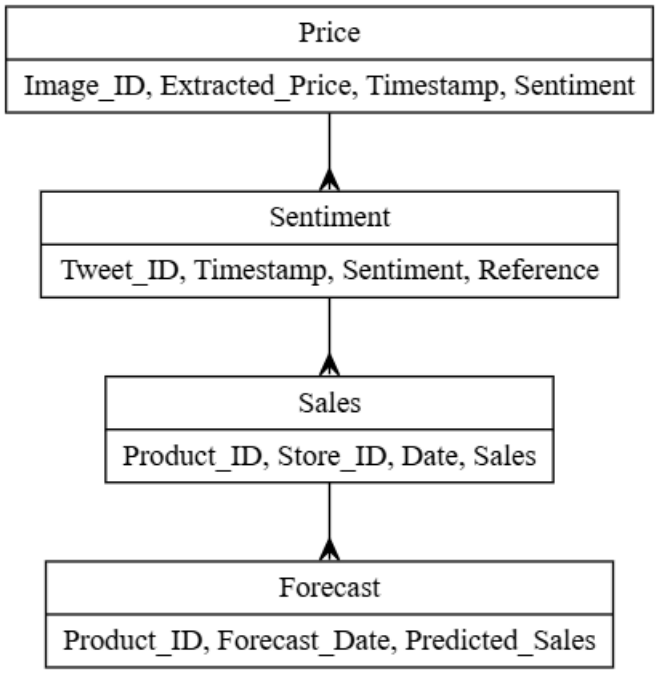
3. **Price Extraction Module**: Images of price tags or product labels are processed using Python OCR libraries (e.g., Tesseract). Extracted price data is analyzed for anomalies or trends and classified for sentiment prediction using machine learning models.

The system integrates these modules into a Python-based application, with a user-friendly interface for data input, visualization, and decision-making. Databases store historical sales data, OCR-extracted prices, and sentiment-labelled tweets.

Table Design

1. **Sales Table**:
   - Product_ID: Unique identifier for each product.
   - Store_ID: Store identifier.
   - Date: Timestamp for the sales record.
   - Sales: Units sold.
2. **Sentiment Table**:
   - Tweet_ID: Unique identifier for the tweet.
   - Sentiment: Sentiment classification (Positive, Negative, Neutral).
3. **Price Table**:
   - Extracted_Price: Numeric value of the extracted price.
   - Sentiment: Sentiment classification for price data.
4. **Forecast Table**:
   - Forecast_Date: Date for the forecasted value.
   - Predicted_Sales: Predicted units.

UML DIAGRAM:



Input Design

The system accepts three types of input:

1. **Sales Data**: Historical sales data of 10 products in tabular format (CSV/Excel) containing product IDs, store IDs, dates, and sales figures.
2. **Twitter Data**: Live or pre-collected tweets related to the products or stores. This data is fed into the sentiment

analysis module for classification.

3. **Images**: Price tags or product labels captured through a camera. These are uploaded to the OCR module for price extraction.

Each input is validated for format and accuracy before processing. Errors, such as invalid dates or non-numeric sales data, trigger alerts for correction.

Implementation

The system is implemented using Python, leveraging libraries like pandas, stats models, and matplotlib for sales forecasting, NLTK/TextBlob for sentiment analysis, and Tesseract for OCR. A Flask-based web interface integrates all modules, providing an interactive platform for data input and visualization.

- **ARIMA Model**: Decomposes sales data into trend, seasonality, and residuals. Hyperparameters are tuned using AIC/BIC scores.
- **Sentiment Analysis**: Tweets are tokenized, cleaned, and transformed into feature vectors. A classifier (e.g., SVM, Naïve Bayes) predicts sentiment.
- **OCR Integration**: Extracts numeric data from images and converts it into a structured format.

Testing

The system is tested through unit testing, integration testing, and user acceptance testing.

1. **Unit Testing**: Ensures the accuracy of each module (e.g., ARIMA predictions, sentiment classification, OCR output).
2. **Integration Testing**: Validates the seamless flow of data between modules, ensuring consistency.
3. **Performance Testing**: Assesses the system's ability to handle large datasets, OCR images, and live tweet streams.
4. **User Acceptance Testing**: Involves end-users to confirm usability, correctness, and relevance of insights.

Testing metrics include accuracy, recall, precision, and runtime efficiency. Errors are logged and debugged iteratively for improvement.

2. CONCLUSION

In this project, we have successfully developed a comprehensive sales forecasting system that integrates the ARIMA model with sentiment analysis and hybrid OCR technology. By collecting data from ten store products and analyzing sales patterns, we utilized the ARIMA model to predict future sales effectively. The results from the model were visualized through accurate graphical representations, which facilitated a clear understanding of projected sales trends. Furthermore, sentiment analysis was incorporated by leveraging Twitter data to gauge public opinions and sentiments towards the products in focus. This additional layer of insight proved to be invaluable, as it enabled us to connect consumer sentiment with sales performance, thus enriching our forecasting model. The integration of OCR technology allowed us to capture product images and extract relevant data, including pricing information, which further enhanced the accuracy of our sentiment classification and subsequent sales predictions. Overall, our hybrid approach demonstrates a significant advancement in sales forecasting through a multi-faceted analysis that combines traditional statistical methods with modern machine learning and information extraction techniques.

3. FUTUREENHANCEMENTS

The future scope of this project is vast and presents numerous avenues for enhancement and expansion. One potential direction is to broaden the dataset beyond ten products and incorporate a wider array of store locations to improve the robustness of the ARIMA model. Additionally, exploring alternative forecasting models such as LSTM neural networks could provide deeper insights, especially in handling nonlinear patterns and time series anomalies. Enhancing sentiment analysis could also involve integrating data from various social media platforms, thereby capturing a more comprehensive view of consumer opinions. Furthermore, refining the OCR process to recognize diverse formats and textures in images could lead to more accurate data extraction regarding prices and promotional offers, allowing for real-time optimization of pricing strategies. Implementing a feedback loop where predictions are updated regularly based on real sales data could also improve accuracy over time. Lastly, incorporating external factors such as seasonality and economic indicators into the forecasting model could lead to better-informed decisions for sales strategies and inventory management in a dynamic market landscape

REFERENCES

[1] ..Zixuan Huo,"Sales Prediction based on Machine Learning",2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)

[2] Akshay Krishna,AkhileshV,AnimikhAich,ChetanaHegde,"Sales-forecasting of Retail Stores using Machine Learning Techniques",2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)

[3] FatmaAbubaker,Ala'Khalifeh,"Sales Forecasting Based on Big Data and Machine Learning Analysis",2023 9th International Conference on Control Decision and Information Technologies (CoDIT)

[4] Rameshwaram Sai Mallik,RAbhiram,SeguruRithvikReddy,R M Jagadish,"A Comprehensive Survey on Sales Forecasting Models Using Machine Learning Algorithms",2022 Fourth International Conference on Emerging Research in Electronics Computer Science and Technology (ICERECT)

[5] .Khushi Pai,MoonisahBatool,NityaRavi,RuchiraSurendra,A N RamyaShree,S J Meenakshi,"Enhancing Sales Forecasting and Prediction with Cutting-Edge Machine Learning Methods",2024 Second International Conference on Data Science and Information System (ICDSIS)

[6] Sunitha Cheriyan,ShanibaIbrahim,SajuMohanan,SusanTreesa,"Intelligent Sales Prediction Using Machine Learning Techniques",2018 International Conference on Computing Electronics & Communications Engineering (iCCECE)

[7] Akanksha,DeveshYadav,DeepakJaiswal,AshwaniAshwani,AshutoshMishra,"Store-sales Forecasting Model to Determine Inventory Stock Levels using Machine Learning",2022 International Conference on Inventive Computation Technologies (ICICT)

[8] Ryali Ajay,R. Sabin Joel,P. G. Om Prakash,"Analyzing and Predicting the Sales Forecasting using Modified Random Forest and Decision Tree Algorithm",2023 8th International Conference on Communication and Electronics Systems (ICCES)

[9] Balpreet Singh,PawanKumar,NonitaSharma,K P Sharma,"Sales Forecast for Amazon Sales with Time Series Modeling",2020 First International Conference on Power Control and Computing Technologies (ICPC2T)

[10] S. M. Aqil Burney,SyedMubashirAli,"Sales Forecasting for Supply Chain Demand Management - A Novel Fuzzy Time Series Approach",2019 13th International Conference on Mathematics Actuarial Science Computer Science and Statistics (MACS).