IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⨳ Peer-reviewed & Refereed journal ⨳ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13457

Data-Driven Business Intelligence and Prediction System using Customer, Local and Demand Analysis

Dr. N. A Ghodichor[1], Mr. Prathmesh Bijwe[2], Ms. Yashasvi Vairagade[3], Mr. Prakhar Jais[4], Mr. Nishant Shende[5], Ms. Sejal Bhajipale[6]

Assistant Professor, Computer Technology, Priyadarshini College of Engineering, Nagpur, India[1]

UG Student, Computer Technology, Priyadarshini College of Engineering, Nagpur, India[2-6]

Abstract: The "Smart Market Intelligence & Opportunity Detection System" project aims to support small businesses and entrepreneurs by providing data-driven insights into market gaps, consumer behavior, and competitive dynamics. Leveraging Machine Learning and Data Science techniques, this system analyzes diverse data sources such as ecommerce platforms, regional economic indicators, and online consumer sentiment to identify unmet product demand and emerging opportunities. It utilizes clustering algorithms, predictive models, sentiment analysis, and scoring mechanisms to evaluate feasibility, assess competition, and recommend strategic actions. By integrating features like market gap detection, price intelligence, and risk analysis, the system enables informed decision-making, reduces market uncertainty, and promotes sustainable business growth.

Keywords: Market intelligence, e-commerce analysis, sentiment analysis, business opportunity, small business, machine learning

I. INTRODUCTION

The rapid expansion of digital commerce has significantly transformed the way buying and selling activities are conducted. Consumers today rely on online platforms to compare products, evaluate prices, check reviews, and make informed purchasing decisions. This shift has created a highly competitive and fast-paced marketplace that extends beyond geographical boundaries. While this evolution benefits consumers, it has also created challenges for local businesses, which often struggle to compete with large e-commerce platforms due to limited access to market insights and data-driven strategies.

Traditional decision-making in small businesses is often based on intuition or limited information, which increases the risk of investment in products that may not align with actual market demand. Additionally, regional markets frequently face imbalances where some products are oversupplied while others remain unavailable. This leads to a phenomenon known as e-commerce leakage, where consumer demand is fulfilled by external or online vendors instead of local sellers, affecting regional economic growth and sustainability.

With advancements in technology, Machine Learning and Data Science provide powerful tools to address these challenges. These technologies enable the analysis of large-scale data from e-commerce platforms, regional economic indicators, and online consumer sentiment. By applying clustering algorithms, predictive modeling, and sentiment analysis, it becomes possible to identify market gaps, forecast demand trends, and evaluate consumer preferences more accurately.

This project presents a Smart Market Gap and Local Business Predictor System designed to assist entrepreneurs, policymakers, and small business owners. The system analyzes multiple data sources to identify business opportunities, assess competition, and evaluate market feasibility. It includes features such as market gap detection, price intelligence, risk analysis, and business expansion planning, all integrated into an interactive dashboard.

In conclusion, the proposed system aims to improve decision-making in local businesses by providing data-driven insights and predictive capabilities. It reduces uncertainty, promotes efficient resource utilization, and supports the growth of sustainable and competitive local markets through the use of modern Machine Learning techniques.

II. LITERATURE REVIEW

In recent years, significant advancements have been made in the field of sentiment analysis, market prediction, and data-driven decision-making using Machine Learning techniques.Traditional approaches relied heavily on manual

DOI: 10.17148/IARJSET.2026.13457

analysis and product demand relied on manual interpretation and limited datasets, which are time-consuming, subjective, and often inaccurate. To overcome these limitations, researchers have explored automated systems for extracting insights from large-scale e-commerce data such as product reviews and pricing trends. Xuechun Li, Xueyao Sun, Zewei Xu, and Yifan Zhou used a Bi-LSTM model with an attention mechanism to analyze sentiment in Amazon reviews, achieving high accuracy and improved interpretability using TF-IDF. Shashank D Shetty analyzed reviews using TextBlob and VADER, showing a trade-off between accuracy and speed. Nishit Shrestha and Fatma Nasoz used a GRU-based model with SVM to detect mismatches between review text and ratings.

Further research focused on improving classification performance using machine learning and deep learning techniques. Tanjim Ul Haque, Nudrat Nawal Saber, and Faisal Muhammad Shah found that Support Vector Machines perform best for large-scale sentiment classification. Sarthak Choubey, Uma Bharathi G V, and Priya Prathala showed that deep learning models outperform traditional methods using word embeddings. Anushka Narula and Sukhleen Bindra identified Random Forest as the most accurate classifier among multiple models. Adarsh Godia and L. K. Tiwari applied TF-IDF with Logistic Regression for Flipkart reviews, achieving strong results.

Apart from sentiment analysis, research has also explored pricing and demand prediction. Mohit Apte, Ketan Kale, Pranav Datar, and Dr. P. R. Deshmukh developed a Reinforcement Learning-based pricing model that adapts to demand and market changes. A. A. Patel, A. R. Bhandare, A. S. Chauhan, S. V. Ghume, and T. V. Rajarshi built a system to track real-time price variations. Dr. Navneet Kaur evaluated regression models for price prediction, while Priyam Ganguli and Isha Mukherjee used machine learning to forecast retail demand.

In addition, studies have examined risk and strategic decision-making. Pankaj Dutta, Pravin Suryawanshi, Pallav Gujarathi, and Arnab Dutta analyzed supply chain disruptions. Seyed Parsa Parvasi, Ata Allah Taleizadeh, and Park Thaichon used game theory for competitive pricing analysis. Riana Steen proposed a model for evaluating launch risks, while Rajeev Pandey, Shiv Shankar Pandey, Muqbil Burhan, Moon Moon Haque, and Rupak Bhattacharjee applied ISM and Bayesian methods for risk analysis. Rajalakshmi, M., Saravanan, V., Arun Prasad, V., Romero, C. T., Khalaf, O. I., and Karthik applied machine learning for process optimization.

Recent work also includes large-scale sentiment and behavior analysis. Hashir Ali, Ehtesham Hashmi, Sule Yildirim Yayilgan, and Sarang Shaikh compared machine learning, deep learning, and transformer models. Dr. V. Vasudevan and Nandini S used Naïve Bayes for real-time sentiment analysis.

III. METHODOLOGY

The methodology of the Smart Business Intelligence & Local Business Predictor System is designed to automate data-driven decision-making by integrating data collection, preprocessing, machine learning, and result generation into a structured pipeline. The system follows an evidence-based approach, combining quantitative data such as pricing and product statistics with qualitative insights like public sentiment and user behavior. It is based on the CRISP-DM framework and enhanced with real-time data processing to improve adaptability and performance over time.

Initially, data is collected from multiple sources including e-commerce platforms such as Amazon and Flipkart, social media platforms, and user inputs. The collected data exists in formats like CSV, JSON, HTML, and text, requiring preprocessing and standardization. Data cleaning includes handling missing values using mean, median, or mode imputation. Textual data is processed using stop word removal, stemming, and lemmatization. Outlier detection is performed using IQR and Isolation Forest techniques to remove abnormal values. Feature scaling using Min-Max normalization ensures compatibility with machine learning models, while sentiment data is processed using tokenization, noise filtering, and vectorization techniques such as TF-IDF or embeddings.

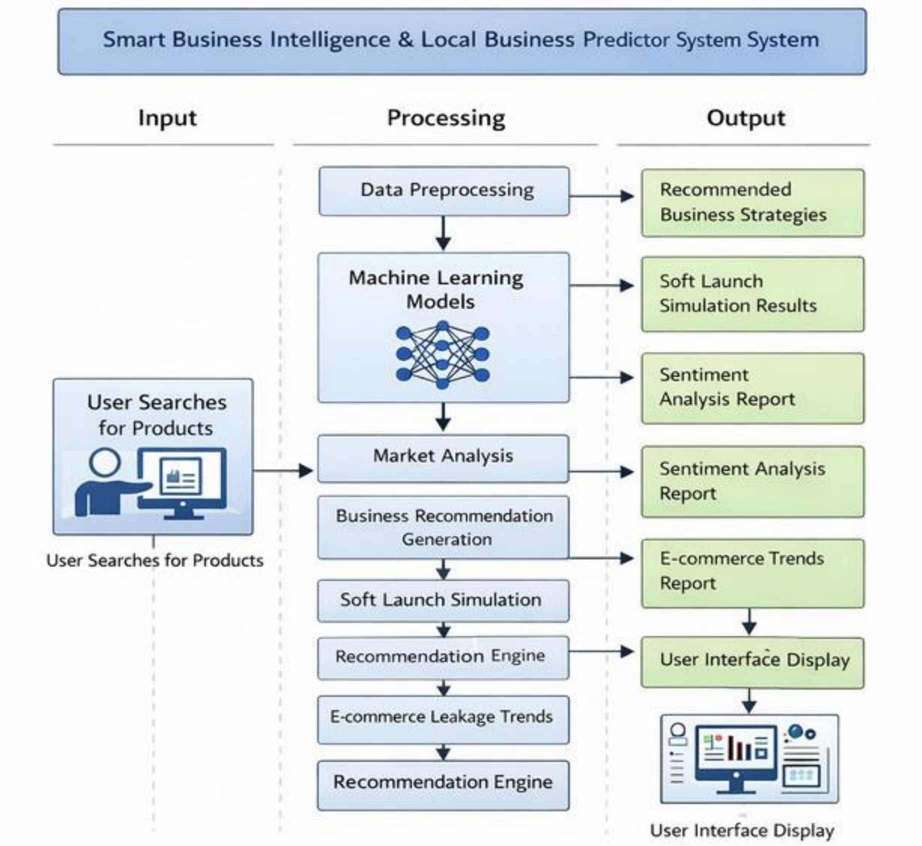
After preprocessing, the system applies multiple machine learning models for analysis. Clustering techniques such as K-Means and DBSCAN are used to group products and patterns based on demand, pricing, and competition. Classification models including Logistic Regression, Random Forest, and XGBoost are used to predict the probability of business success. An E-commerce Leakage Classifier evaluates dependency on external sellers by analyzing supply differences. A Risk and Simulation module estimates uncertainty and potential outcomes using scoring methods. The system also includes a recommendation engine that suggests business ideas, pricing strategies, and expansion plans based on combined analysis of demand, sentiment, and competition.

The system is implemented using a scalable architecture where the backend is built using Flask or Node.js, and

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026
DOI: 10.17148/IARJSET.2026.13457

machine learning models are deployed as services. Data is stored using PostgreSQL and MongoDB, and the frontend is developed using React with visualization support. Model performance is evaluated using metrics such as accuracy, precision, recall, F1-score, RMSE, and silhouette score. The overall architecture follows a layered design including data processing, machine learning, and visualization components, ensuring scalability, efficiency, and real-time decision support.

IV. BLOCK DIAGRAM



V. RESULT & DISCUSSION

Results

The proposed Smart Business Intelligence & Local Business Predictor System demonstrated effective performance in analyzing consumer behavior and generating data-driven business recommendations. The system was evaluated using standard machine learning approaches for classification, clustering, and forecasting tasks. It successfully identified demand patterns, analyzed pricing variations, and captured user sentiment across different conditions, providing meaningful insights for decision-making. Clustering techniques helped in grouping similar products and trends, enabling better segmentation and understanding of market behavior, while classification models supported prediction of business feasibility. To improve performance, preprocessing techniques such as normalization, outlier removal, and text cleaning were applied. These methods enhanced the system's ability to generalize and improved the consistency and reliability of results across datasets. Another important component of the system is the scoring and ranking mechanism, which filters and prioritizes recommendations based on multiple factors. This results in clear, structured, and interpretable outputs, improving overall usability, accuracy, and effectiveness of the system in real-world applications.

Discussion

The results obtained from the proposed system demonstrate the effectiveness of machine learning in supporting intelligent business decision-making. Traditional methods rely heavily on manual judgment and limited data, which can lead to inconsistencies, delays, and inaccurate conclusions. In contrast, the developed system provides a structured and automated approach that improves reliability, efficiency, and consistency in analysis. One of the key strengths of the system is its ability to process large amounts of data and generate insights in near real-time, making it suitable for

DOI: 10.17148/IARJSET.2026.13457

web-based applications and practical deployment. Users can analyze trends, understand consumer preferences, and receive business recommendations efficiently. The use of machine learning enables the system to capture complex relationships between factors such as demand, pricing, and sentiment. However, certain limitations exist, including dependency on data quality, availability of real-time data, and challenges in handling sudden market changes or unpredictable external factors. Despite these limitations, the system provides a scalable and practical solution for improving business planning, strategic analysis, and informed decision-making.

VI. CONCLUSION

The research presents a robust and efficient Smart Business Intelligence & Local Business Predictor System that utilizes Machine Learning and Data Science techniques to support informed business decision-making. The system successfully automates the analysis of consumer behavior, pricing trends, and business feasibility, overcoming the limitations of traditional methods that rely on intuition and limited data. The overall performance demonstrates consistency, reliability, and the ability to generate meaningful insights across different scenarios. The integration of sentiment analysis, predictive modeling, and recommendation mechanisms makes the system suitable for real-world applications in business planning and strategy development. The proposed solution not only reduces uncertainty but also empowers users by providing clear, data-driven insights for decision-making. It enables entrepreneurs and analysts to evaluate opportunities effectively without heavy dependence on manual analysis. This represents a significant step toward intelligent and user-centric digital solutions in the business domain. In conclusion, the adoption of AI-driven analytical systems can transform business planning by improving accuracy, reducing risks, and enabling more efficient and informed decision-making.

VII. ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Rashtrasant Tukdoji Maharaj Nagpur University for providing the academic environment and resources necessary for the successful completion of this research. Special thanks are extended to the faculty members of the Computer Technology Department for their continuous guidance, valuable insights, and technical support throughout the project. We also acknowledge the open-source community and various platforms whose tools, frameworks, and documentation played a crucial role in the development and implementation of this system. The availability of publicly accessible datasets, e-commerce platforms, and research publications significantly contributed to the success of this work. No external financial support was received for this research, and the project was carried out as part of academic research and development activities. Finally, we thank our peers and collaborators for their constructive feedback, support, and encouragement during the research process.

REFERENCES

[1] Li, X., Sun, X., Xu, Z., & Zhou, Y. (2021). Explainable sentence-level sentiment analysis for Amazon product reviews. November 2021.

[2] Shetty, S. D. (2023). Analyzing Amazon product review sentiment. International Research Journal of Modernization in Engineering and Technology (IRJMET), May 2023.

[3] Shrestha, N., & Nasoz, F. (2019). Deep learning sentiment analysis of Amazon.com reviews and ratings. International Journal of Soft Computing and Artificial Intelligence Applications (IJSCAIA), Vol. 8, February 2019.

[4] Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. IEEE International Conference on Innovative Research and Development (ICIRD), May 2018.

[5] Choubey, S., Bharathi, U. G. V., & Prathala, P. (2023). Deep learning techniques for text-based sentiment analysis of Amazon product reviews. May 2023.

[6] Narula, A., & Bindra, S. (2023). Sentiment analysis of Amazon product reviews. February 2023.

[7] Godia, A., & Tiwari, L. K. (2025). Sentiment analysis and classification of product reviews using NLP and machine learning techniques. April 2025.

[8] Apte, M., Kale, K., Datar, P., & Deshmukh, P. R. (2024). Dynamic retail pricing via Q-learning: A reinforcement learning framework for enhanced revenue management. November 2024.

[9] Dutta, P., Suryawanshi, P., Gujarathi, P., & Dutta, A. (2020). Managing risk for e-commerce supply chains: An empirical study. IFAC Conference on Manufacturing, Modelling, Management and Control, Vol. 52, February 2020.

[10] Parvasi, S. P., Taleizadeh, A. A., & Thaichon, P. (2024). Price optimization for manufacturers in a competitive retail market: Imported products and online crowdfunding option. Journal of Revenue and Pricing Management, Vol. 24, February 2024.

DOI: 10.17148/IARJSET.2026.13457

[11] Steen, R. (2023). A risk assessment approach to support the launching of new products, services or processes. International Journal of Business Continuity and Risk Management, Vol. 6, January 2023.

[12] Pandey, R., Pandey, S. S., Burhan, M., Haque, M. M., & Bhattacharjee, R. (2020). Evaluation of risk in new product launch for an automotive company: Interpretive structural modeling and Bayesian belief network approach. May 2020.

[13] Rajalakshmi, M., Saravanan, V., Prasad, V. A., Romero, C. T., Khalaf, O. I., & Karthik, C. (2022). Machine learning for modeling and control of industrial clarifier process. Intelligent Automation & Soft Computing, Vol. 32, January 2022.