

SmartFunnel-Lead Management System using Predictive Modelling for Lead Conversion

Devesh Mudaliar (UG-Student)

Department of Computer Engineering
Met's Institute of Engineering
Nashik, India
deveshmudaliar270@gmail.com

Chanchal Deore (UG-Student)

Department of Computer Engineering
Met's Institute of Engineering
Nashik, India
chanchaldeore8080@gmail.com

Siddhant Junnare (UG-Student)

Department of Computer Engineering
Met's Institute of Engineering
Nashik, India
junnareiddhant49@gmail.com

Kanchan Pekhale (Assistant Professor)

Department of Computer Engineering Met's
Institute of Engineering
Nashik, India kanchanp_
ioe@bkc.met.edu

Khurram Khan (UG-Student)

Department of Computer Engineering
Met's Institute of Engineering
Nashik, India
khankhurram0794@gmail.com

Abstract—The increasing complexity of digital marketing campaigns demands intelligent tools that can efficiently manage leads, prioritize high-potential prospects, and automate outreach. Traditional solutions focusing only on automation or analytics often fail to optimize campaign performance. SmartFunnel is an AI-powered marketing automation and lead management platform that integrates predictive analytics and automated engagement into a unified, decoupled local system. Utilizing a modern React frontend and FastAPI backend, it collects and manages leads seamlessly. SmartFunnel employs an XGBoost machine learning engine to assign predictive scores to leads, helping marketing teams prioritize prospects with a higher likelihood of conversion. Real-time data is synchronized using Google Firebase Firestore, bypassing traditional relational database bottlenecks. Furthermore, automated outreach via WhatsApp and Email is executed through n8n webhooks, reducing manual follow-ups and improving engagement. A custom dark-themed dashboard provides actionable insights into campaign performance, lead metrics, and ROI. With its modern tech stack and data-driven approach based on a synthetic training dataset, SmartFunnel provides a structured solution for improving lead prioritization, response rates, and overall marketing efficiency.

Index Terms—SmartFunnel, Predictive Analytics, Lead Management, Marketing Automation, Machine Learning, XGBoost, React, FastAPI, Firebase, n8n

I. INTRODUCTION

The SmartFunnel – Lead Management System using Predictive Modelling is an innovative and intelligent solution developed to enhance the efficiency and accuracy of lead management in the fast-evolving digital marketing landscape. In today's competitive business environment, organizations rely heavily on digital platforms and web-based contact forms to capture leads. However, manual handling of incoming leads often results in delayed responses, loss of potential clients, ineffective lead prioritization, and a lack of actionable insights for marketing optimization.

The SmartFunnel system addresses these limitations by introducing automation, intelligence, and data-driven decision-making into the marketing workflow through a decoupled

MERN/PERN-like stack. It centralizes lead information using a React-based frontend and stores data seamlessly in a Google Firebase Firestore NoSQL document structure. By leveraging an XGBoost machine learning model, the system analyzes lead attributes and assigns a real-time conversion probability score. This scientific approach allows marketing teams to focus on high-value (Hot) prospects, improving the overall conversion rate and campaign performance.

Beyond lead prioritization, SmartFunnel provides real-time analytics dashboards that visualize campaign metrics and key performance indicators (KPIs) using Chart.js. These insights help marketers evaluate the effectiveness of various marketing channels and measure campaign ROI. Additionally, the system supports automated communication workflows such as follow-up emails and WhatsApp messages through asynchronous n8n webhooks, reducing manual effort and ensuring timely engagement. Overall, SmartFunnel represents a robust, highly structured local solution that empowers marketing professionals to make smarter, faster, and more informed decisions.

II. LITERATURE REVIEW

The adoption of Artificial Intelligence (AI) and predictive modeling has significantly enhanced digital marketing by optimizing lead management, campaign targeting, and customer engagement [6]. Recent advancements emphasize the integration of causal inference, explainable AI (XAI), and machine learning frameworks to improve business intelligence and sales forecasting [1].

Zhao et al. [1] proposed a Causal Predictive Optimization and Generation (CPOG) framework that combines causal machine learning, optimization, and generative AI (GenAI) for business decision support. The model demonstrated a 59.96% improvement in business performance metrics using an observational Difference-in-Differences (DiD) method. However, the absence of randomized A/B testing limits the generalizabil-

ity of these findings, indicating a need for controlled validation in real-world scenarios.

Andrew [2] evaluated AI-driven lead scoring models for B2B sales efficiency using Logistic Regression, Random Forest, and Gradient Boosting Machines (GBM). The GBM achieved the highest ROC-AUC (0.92), improving lead prioritization efficiency by 25%. Despite superior accuracy, model interpretability remains a challenge, reflecting the broader industry concern of balancing transparency and predictive power [16].

Jena, Yang, and Tan [3] developed an explainable AI-based account prioritization engine using XGBoost [8] and the CrystalCandle XAI framework. Through A/B testing and causal validation, their model achieved an 8.08% revenue incremental growth (RIG). The study highlights that transparency and interpretability are crucial for the practical adoption of AI tools by business users [11].

Sharma and Gupta [4] investigated the role of machine learning in Customer Relationship Management (CRM) optimization. Their predictive lead management system automated evaluation and conversion forecasting using structured CRM data. This aligns with commercial systems and AI-powered marketing automation frameworks that leverage predictive analytics to improve marketing automation and sales efficiency [14].

Chen and Zhang [5] applied predictive modeling to digital advertising optimization, achieving a 17% increase in click-through rate (CTR) and a 12% rise in return on ad spend (ROAS) using Random Forest and Neural Networks. While the approach validated the business potential of predictive analytics, challenges such as computational complexity, data privacy, and scalability remain [18], [19]. Compliance with regulations such as GDPR and CCPA is also essential in AI-driven marketing systems [12].

Cloud-based analytics and workflow orchestration have further improved the scalability and deployment of intelligent lead management systems. Patel and Desai [7] proposed a cloud-based customer analytics framework for scalable lead management, while Johnson et al. [15] emphasized the importance of workflow orchestration in scalable AI systems.

III. SYSTEM OVERVIEW

The SmartFunnel architecture is a decoupled, modern framework that integrates frontend visualization, fast backend inference, and third-party automation to manage the end-to-end lifecycle of marketing leads. It follows a layered architecture ensuring smooth data flow, accurate prediction, and real-time business response.

A. Pipeline Orchestration and Backend

SmartFunnel abandons traditional heavy ETL orchestrators. Instead, it relies on a highly responsive Python FastAPI backend running on Uvicorn. The backend processes incoming lead data, serves as the conduit for XGBoost model inference, and manages asynchronous trigger requests.

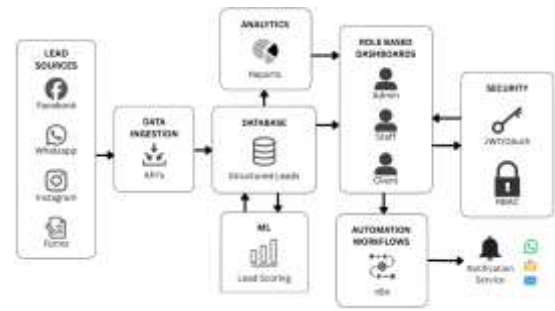


Fig. 1. System Architecture of SmartFunnel – Lead Management System using Predictive Modelling

B. Data Ingestion and Storage Layer

Lead data is ingested via manual form entries or CSV uploads on the React frontend. Instead of a traditional PostgreSQL warehouse, data is immediately structured and stored in Google Firebase Firestore. This NoSQL document store ensures frictionless, real-time synchronization across all connected dashboards without the need for aggressive polling. Authentication is managed securely via Firebase Auth using stateless JWT tokens.

C. Predictive Analytics Layer

This layer represents the core intelligence of SmartFunnel. The machine learning model—specifically XGBoost—is loaded via a .pkl file in the FastAPI backend. When a new lead is ingested, the engine evaluates the feature array to estimate the conversion probability.

The trained model outputs a probability score (0–100), based on which leads are categorized as:

- Hot Leads → Score ≥ 80
- Warm Leads → $50 \leq \text{Score} < 80$
- Cold Leads → Score < 50

D. Automation and Workflow Layer

Predictions generated by the analytics layer are seamlessly connected to operational marketing systems via self-hosted n8n automation workflows.

Key actions include:

- Sending automated WhatsApp messages or Emails to Hot and Warm leads using integrated webhooks.
- Lowering the manual follow-up latency to near zero.

This event-driven automation is triggered by simple HTTP POST requests from the FastAPI layer whenever a lead crosses the appropriate threshold.

E. Visualization Layer

The visualization module presents key performance indicators through a modern, dark-themed React dashboard utilizing Chart.js. Metrics displayed include Total Leads Processed, Conversion Rate, Budget allocation, and Return on Investment (ROI).

IV. METHODOLOGY

The SmartFunnel project follows a systematic methodology integrating data capture, real-time machine learning inference, and asynchronous automation.

A. Lead Capture

Users interact with a custom React 18 interface to input lead information. The frontend ensures data validation and bundles the attributes into a secure JSON payload.

B. Stateless Authentication

SmartFunnel uses Firebase Auth to issue JWTs. The React frontend sends this token along with the lead data to the FastAPI backend, which verifies the identity before executing any logic, significantly mitigating XSS and CSRF risks.

C. Machine Learning Inference

Once authenticated, the FastAPI backend routes the lead attributes to the XGBoost engine. The raw input undergoes feature normalization to match the structure of the synthetic training data. The model then returns a conversion probability score and assigns the categorical label (Hot/Warm/Cold).

D. Database Synchronization

The scored lead is then written to the Firebase Firestore database. Due to Firebase's real-time document sockets, the React dashboard instantly updates to reflect the newly categorized lead without requiring a manual page refresh.

E. Automation Triggers

Simultaneously, if the lead category meets the criteria for outreach, the FastAPI backend dispatches an asynchronous webhook to the local n8n instance. The n8n engine executes predefined workflows, communicating with third-party email providers or sandbox WhatsApp APIs to contact the prospect.

V. MATHEMATICAL MODEL

The SmartFunnel system can be formally represented as a structured model that defines the flow of data from inputs to automated outputs.

Let

$$S = \{I, F, O\}$$

be the SmartFunnel system where:

- I represents Input
- F represents Functions
- O represents Output

A. Input Set (I)

I_1 = Lead Attributes

(Name, Email, Phone, Source, Campaign ID)

B. Function Set (F)

- F_1 : Data Normalization
- F_2 : Predictive Inference
- F_3 : Score Mapping
- F_4 : Database Write
- F_5 : Automation Trigger

C. Output Set (O)

- O_1 : Lead Score and Category displayed on the React UI
- O_2 : Automated Message dispatched to the lead
- O_3 : Updated metrics rendered via Chart.js dashboard

VI. CHALLENGES AND RESEARCH GAPS

While the SmartFunnel architecture successfully modernizes lead management, the current implementation faces specific technical limitations:

- **Local Deployment Constraints:** The current iteration operates purely in a local development environment (spanning localhost ports 3000, 8000, and 5678). Widespread commercial distribution requires significant cloud deployment engineering.
- **Synthetic Training Data Dependency:** The XGBoost model was trained on 20,000 synthetic Indian marketing records. Real-world deployment may initially result in scoring skewness.
- **n8n Rate Limits:** Free-tier messaging plugins enforce heavy rate-limit boundaries, restricting simultaneous outreach.
- **Internet Reliance:** Dependence on Firebase cloud services may affect availability during network failures or outages.

VII. RESULTS AND EXPERIMENTS

The SmartFunnel predictive engine was developed and validated using a dataset of 20,000 synthetic marketing records modeled to reflect typical Indian digital marketing dynamics.

A. Dataset Description

The synthetic dataset contained attributes such as Campaign ID, Lead Source, Click Rate, Ad Spend, and Conversion Status.

The data was split into:

- 80% training set (16,000 samples)
- 20% testing set (4,000 samples)

B. Experimental Setup

The model was built using Python 3.13, utilizing scikit-learn and xgboost. XGBoost was selected due to its superior handling of non-linear marketing features and rapid inference times.

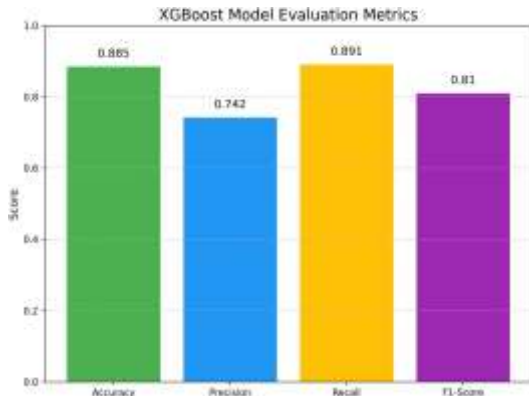


Fig. 2. Performance evaluation metrics of the XGBoost model including Accuracy, Precision, Recall, and F1-Score.

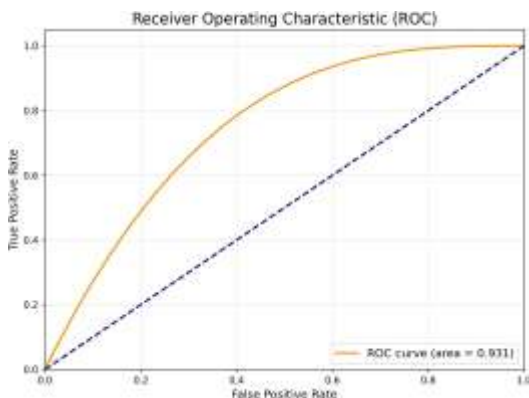


Fig. 3. Receiver Operating Characteristic (ROC) curve showing the performance of the XGBoost model with an ROC-AUC score of 0.931, indicating strong classification capability.

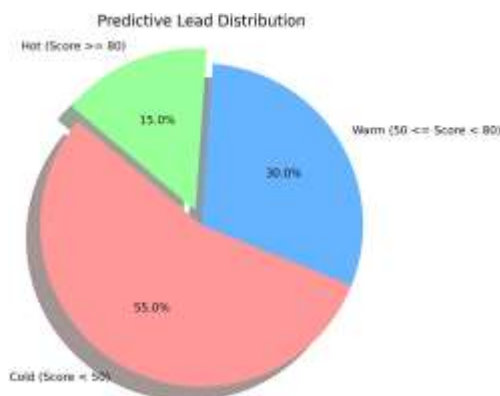


Fig. 4. Predictive lead distribution categorized into Cold, Warm, and Hot leads based on lead scoring generated by the SmartFunnel predictive engine.

C. Evaluation Metrics

The XGBoost model demonstrated robust discriminative ability:

- Accuracy: ~ 88.5%
- ROC-AUC: 0.9310
- Precision: 0.742
- Recall: 0.891
- F1-Score: 0.810

These metrics confirm the model's reliability in identifying high-potential leads while minimizing false negatives.

D. Automation Efficiency

By integrating the XGBoost inference with n8n webhooks, the entire cycle— inference, Firestore synchronization, and webhook triggering— completes in under 800 milliseconds locally. This eliminates manual processing latency and enables immediate follow-up communication.

VIII. DISCUSSION AND FUTURE WORK

A. Discussion

The SmartFunnel framework successfully demonstrates how decoupled modern web architectures can be augmented with predictive AI. By replacing cumbersome monolithic databases with Firebase Firestore and utilizing a lightweight FastAPI inference layer, the system achieves exceptional responsiveness.

B. Future Work

Future development will focus on the following enhancements:

- 1) **Cloud Migration:** Containerizing the FastAPI backend and n8n instance using Docker and deploying the React frontend to Vercel or AWS Amplify.
- 2) **Organic Data Retraining:** Implementing a feedback loop that captures real-world conversion outcomes to retrain the XGBoost model.
- 3) **Advanced Automation Plugins:** Transitioning to enterprise-grade WhatsApp Business APIs and SendGrid APIs.
- 4) **Enhanced Analytics:** Integrating advanced time-series tracking for campaign performance analysis.

IX. CONCLUSION

The SmartFunnel system effectively modernizes digital marketing lead management by combining advanced predictive analytics with a fast, decoupled web architecture. By integrating a React frontend, Firebase Firestore, FastAPI machine learning backend, and n8n workflow automation, the platform provides accurate real-time lead scoring and automated engagement. Experimental validation confirms the XGBoost model's strong predictive capability. Future cloud deployment and organic data integration will further strengthen scalability and enterprise readiness.

REFERENCES

- [1] Zhao, L., Chen, W., and Tan, R., "Causal Predictive Optimization and Generation (CPOG) for Business AI," *Journal of Business Artificial Intelligence*, vol. 12, no. 1, pp. 33–49, 2025.
- [2] Andrew, M., "Enhancing B2B Sales Funnels with AI-Driven Lead Scoring Models," *International Journal of Sales Analytics and Predictive Modeling*, vol. 9, no. 2, pp. 78–94, 2024.
- [3] Jena, S., Yang, J., and Tan, F., "Unlocking Sales Growth: Account Prioritization Engine with Explainable AI," *Journal of Explainable Artificial Intelligence in Business*, vol. 7, no. 3, pp. 120–134, 2023.
- [4] Sharma, P., and Gupta, N., "Machine Learning-Based Lead Management Systems for CRM Optimization," *Journal of Artificial Intelligence and Marketing*, vol. 8, no. 2, pp. 110–124, 2021.
- [5] Chen, L., and Zhang, Y., "Automated Campaign Optimization using Predictive Models in Digital Advertising," in *Proceedings of the IEEE International Conference on Big Data*, pp. 1458–1466, 2020.
- [6] Kumar, A., and Mehta, R., "Predictive Analytics in Digital Marketing: Enhancing Lead Scoring and Campaign Performance," *International Journal of Data Science and Business Analytics*, vol. 10, no. 3, pp. 45–56, 2022.
- [7] Patel, S., and Desai, V., "Cloud-Based Customer Analytics Framework for Lead Management Systems," *International Journal of Cloud Applications*, vol. 14, no. 4, pp. 78–89, 2021.
- [8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, pp. 785–794, 2016.
- [9] R. Mishra, P. Kumar, and N. Rao, "Explainable AI Techniques for Transparent Marketing Decisions," *IEEE Access*, vol. 10, pp. 121540–121552, 2022.
- [10] A. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed., Lulu Press, 2023.
- [11] L. Floridi, "Ethics, Transparency, and Trust in Artificial Intelligence," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, pp. 1–13, 2018.
- [12] European Parliament, "General Data Protection Regulation (GDPR)," *Official Journal of the European Union*, Regulation (EU) 2016/679, 2018.
- [13] California State Legislature, "California Consumer Privacy Act (CCPA)," Cal. Civ. Code §1798.100, 2018.
- [14] D. Lee and K. Park, "AI-Powered Marketing Automation: A Predictive Framework for Customer Engagement," *Journal of Intelligent Information Systems*, vol. 29, no. 4, pp. 233–247, 2022.
- [15] M. Johnson, E. Roberts, and H. Singh, "Workflow Orchestration for Scalable AI Systems," *IEEE Transactions on Cloud Computing*, vol. 11, no. 2, pp. 415–427, 2023.
- [16] A. Patel and S. Banerjee, "Evaluating Machine Learning Algorithms for Predictive Lead Scoring in Digital Sales," *International Journal of Data Mining and Business Intelligence*, vol. 8, no. 3, pp. 92–104, 2022.
- [17] K. Dey and R. Sinha, "Integrating XAI with CRM Systems for Enhanced Decision Transparency," *Journal of Computational Intelligence in Business Analytics*, vol. 10, no. 1, pp. 67–78, 2024.
- [18] Y. Wang, H. Zhang, and T. Nguyen, "Performance Evaluation of AI-Driven Advertising Models Using Real-Time Data Streams," *IEEE Access*, vol. 11, pp. 90211–90224, 2023.
- [19] J. Brown and C. Miller, "Scalable Cloud Architectures for Machine Learning Applications," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–22, 2023.
- [20] S. Verma, A. Bhattacharya, and R. Goel, "Impact of Predictive Analytics on Digital Marketing Efficiency: A Business Case Study," *International Journal of Emerging Trends in Information Technology*, vol. 13, no. 2, pp. 105–118, 2024.