Grenze International Journal of Engineering and Technology, January Issue

Integrating Machine Learning Insights into Business Intelligence Dashboards for Enhanced Marketing Strategy Analysis

Sanjana V[1], Amirtha Varshini M[2], Chitra A D[3] and Umarani M[4]
[1-4]Coimbatore Institute of Technology/Department of Software Systems, Coimbatore, India
Email: vsanjana666@gmail.com, amirthavarshini.0102@gmail.com, umarani@cit.edu.in, dchitra@cit.edu.in

***Abstract*— This paper presents an innovative approach to integrating machine learning (ML) insights with Business Intelligence (BI) dashboards to enhance marketing strategy analysis. By processing and encoding customer data, and applying a Random Forest Classifier for predictive analysis, we derive key metrics such as 'Profit per Item' and 'Recency.' These insights are visualized through interactive Power BI dashboards, which include bar charts for gender-based marketing, scatter plots for VIP programs, and line graphs for time-based trends. Integration allows for dynamic data exploration, providing actionable insights and enabling more targeted marketing strategies, thus facilitating improved decision-making and business performance.**

***Index Terms*— Artificial Intelligence, Business Intelligence, Machine Learning, Predictive Analysis, Random Forest Classifier, Marketing Strategy Analysis.**

I. INTRODUCTION

In today's competitive market landscape, leveraging advanced technologies to enhance marketing strategies is crucial for business success. This project presents a comprehensive approach to integrating machine learning (ML) and Business Intelligence (BI) tools to refine and optimize marketing efforts.

The project employs a Random Forest Classifier to analyze customer data, focusing on several key marketing strategies, including gender-based marketing, VIP programs based on sales and profit, and time-based marketing initiatives.

The process begins with detailed data preprocessing to handle missing values, convert categorical data into numerical formats, and engineer features that are crucial for effective marketing analysis. Specifically, the 'Profit per Item' and 'Recency of Purchase' features are created to provide valuable insights into customer value and engagement. These features are pivotal in identifying high-value customers and understanding recent purchasing behavior.

Using the Random Forest algorithm, the project models these customer metrics to uncover patterns and trends that inform targeted marketing strategies. The results are then visualized using Power BI, a powerful BI tool that facilitates the creation of interactive and dynamic dashboards. These dashboards include various chart types, such as bar charts for gender-based marketing analysis, scatter plots for VIP program insights, and line graphs or heatmaps for time-based marketing trends.

The integration of ML and BI in this project enables a deeper understanding of customer behavior, allowing for the development of personalized marketing strategies that drive customer engagement and business growth.

Grenze ID: 01.GIJET.11.1.544

Through interactive filters and slicers in Power BI, users can explore different marketing strategies and gain actionable insights. This approach not only enhances the decision-making process but also provides businesses with a competitive edge by aligning marketing efforts with data-driven insights. The ultimate goal of this project is to demonstrate how combining ML and BI can lead to more effective marketing strategies and improved business outcomes.

II. Literature Survey

Market segmentation, targeting, and positioning (STP) are key marketing methods that enable organizations to identify distinct consumer groups and customize their efforts accordingly. Camilleri [1] emphasizes STP's role in increasing consumer satisfaction and revenues by segmenting markets based on demographic, psychographic, geographic, or behavioral criteria. This segmentation allows businesses to focus on the most receptive client segments, resulting in targeted marketing that resonate with their target market. Businesses may improve the likelihood of purchases and consumer loyalty by integrating real-time data and predictive analytics.

Ali et al. [2] highlight the importance of Random Forests and Decision Trees in predictive modeling, discussing how Random Forests use ensemble approaches to reduce overfitting and improve accuracy. Their findings highlight the importance of proper model selection in improving predictive accuracy.

Burke [3] investigates knowledge-based recommender systems, demonstrating how contextual information can deliver individualized recommendations despite little user history. This paper addresses the cold-start problem, which is critical for developing adaptive recommendation strategies.

Felfernig et al. [4] present a framework for creating knowledge-based recommender systems that emphasizes the integration of user models and domain information to improve suggestion relevance.

By capturing user-item associations, Hofmann's [5] latent semantic models for collaborative filtering address data sparsity and enhance suggestion quality.

The significance of dynamic user profiling is emphasized by Pazzani and Billsus [6], who point out that in quickly evolving digital environments, the efficiency of recommendations is improved by continuously updating user profiles.

According to Tso and Schmidt-Thieme's [7] evaluation of recommender system algorithms, customized techniques can greatly improve performance by taking into account the differences in dataset features.

Finally, Schmidt-Thieme [8] addresses compound classification models and shows that, in comparison to single models, integrating multiple classifiers enhances recommendation accuracy.

III. Types of Personalized Marketing

Personalized marketing is an advanced strategy that enables businesses to tailor their promotions and communication according to individual customer characteristics, enhancing customer engagement and driving sales. Machine learning plays a pivotal role in identifying patterns within customer data, allowing companies to craft targeted campaigns. This section describes several types of personalized marketing strategies that can be effectively implemented using predictive analytics.

A. Gender-Based Marketing

Gender-based marketing focuses on creating customized marketing strategies based on the gender of the customer. By analyzing purchasing behavior and preferences across gender segments, businesses can provide more relevant product recommendations, discounts, and services. For example, a clothing retailer might detect that women are more likely to purchase certain types of accessories during the holiday season, while men tend to buy high-value items like electronics during specific sale events. Targeted gender-based promotions, like offering a discount on women’s fashion during a known buying peak, can significantly increase sales conversion. Gender-based marketing is especially useful in industries like fashion, technology, and cosmetics, where preferences and buying patterns often vary significantly between genders. Machine learning models help automate this segmentation by analyzing sales data and customer demographics, ensuring timely and personalized outreach.

B. Sales and Profit-Based VIP Programs

Sales and profit-based VIP programs are designed to reward the highest-value customers, those who consistently contribute the most to a company’s bottom line. By using predictive models, businesses can identify customers who have a history of high-value purchases or frequent buying patterns. These customers are then offered exclusive rewards, discounts, or priority services that incentivize them to continue their purchasing behaviors.

Machine learning models such as Random Forest can help segment customers based on both their historical sales data and predicted future purchases, allowing businesses to maintain a VIP tier that feels exclusive and valued. Additionally, loyalty programs tied to sales volume or profit margins ensure that the most lucrative customers receive personalized attention, enhancing their brand loyalty and lifetime value.

C. Priority Orders

Priority order programs are geared towards customers who frequently place high-value or time-sensitive orders. For businesses that deal in fast-moving consumer goods or urgent services, identifying customers who regularly place priority orders allows for improved resource allocation and customer satisfaction. For example, a logistics company could use machine learning to flag customers whose orders are consistently marked as urgent, allowing the company to automatically prioritize shipping and fulfillment for these customers. This type of marketing ensures that priority customers are consistently satisfied, which in turn encourages repeat business. Predictive models play a crucial role here by identifying customers whose purchasing behavior indicates the need for quicker processing, thus reducing potential delays and improving overall service levels.

D. Discount and Reward Programs for Price-Sensitive Customers

Price-sensitive customers, those who are more likely to make purchases when offered significant discounts, form a critical segment for discount and reward programs. Machine learning can identify these customers by analyzing historical purchasing data, focusing on variables like discount usage, frequency of purchases during sales events, and responses to price changes. By flagging customers who only purchase when offered a discount of 20% or more, for instance, businesses can develop targeted promotions aimed at this segment. Price-sensitive customers can be sent personalized offers or included in limited-time discount programs that encourage them to make a purchase. This strategy helps businesses maintain profitability while catering to a segment that is highly responsive to price reductions.

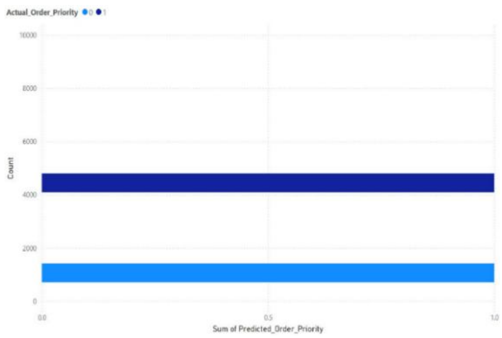

Figure 1. Priority Based Ordering

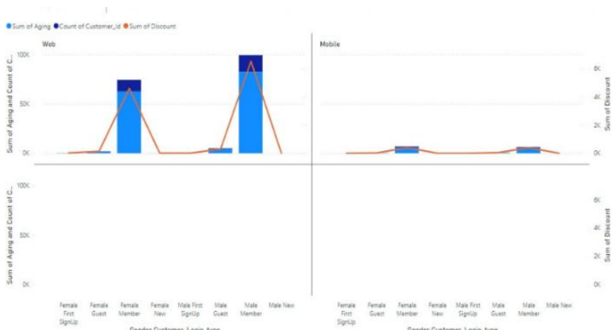

Figure 2. Price Sensitive Customers

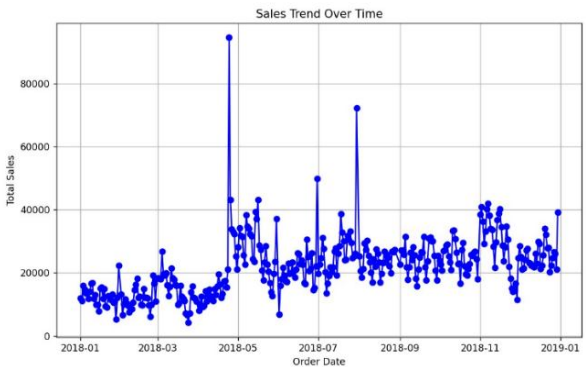

Figure 3. Price Sensitive Customers

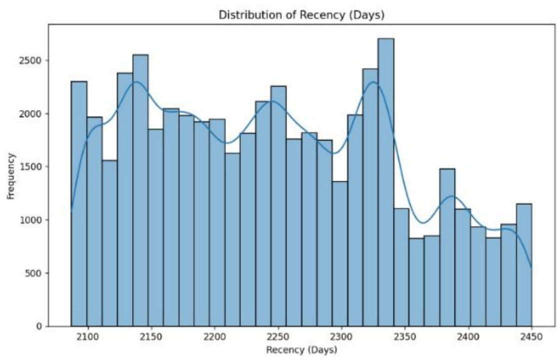

Figure 4. Distribution of Recency

E. Payment Method-Based Marketing

Payment method-based marketing involves personalizing offers based on the customer's preferred method of payment. Machine learning models can analyze customer payment data, allowing businesses to identify patterns,

such as a preference for digital wallets, credit cards, or installment payment plans. By recognizing these preferences, companies can offer promotions such as cash-back rewards or discounts for customers who use a specific payment method.
For example, if a segment of customers frequently uses mobile payment platforms, the business can provide exclusive mobile wallet offers to encourage higher adoption. This type of marketing can also help businesses forge partnerships with payment providers, offering joint promotions that benefit both the business and the payment service. Fig 3 represents the total sales made for each month.

F. Time-Based Marketing

Time-based marketing segments customers based on the recency of their purchases, focusing on how long it has been since a customer last interacted with the business. For this type of strategy, the concept of recency plays a crucial role, where customers who recently made a purchase might be targeted with offers to encourage repeat purchases, while those who have been inactive for a longer period might receive re-engagement campaigns. Machine learning models can calculate the recency by analyzing the last purchase date of each customer and segmenting them into active, lapsed, or dormant customers. These insights can then be used to deliver time-sensitive promotions, such as flash sales or reminder emails, to bring customers back to the store. Time-based marketing ensures that businesses maintain a strong connection with their customers, particularly those who are at risk of churning. Fig 4 represents the frequency of customers with respect to the number of days visited.

IV. PREDICTIVE ANALYSIS

Predictive analysis uses historical data and machine learning techniques to forecast future customer behaviors, enabling businesses to optimize their personalized marketing efforts.
By building models that predict outcomes like customer purchases, churn rates, and responses to promotions, businesses can create highly targeted marketing campaigns. In this section, predictive analysis is broken down into several critical steps, each playing a vital role in building accurate customer models.

A. Data Preparation

Data preparation is the foundational step in predictive analysis, ensuring that the dataset is clean and structured for model building. Key actions include:

- Handling Missing Values: Missing data can skew predictions, so it is critical to fill these gaps. In this project, the Shipping_Cost column had missing values, which were handled by using a mean imputation strategy through the SimpleImputer. This ensures no records are lost while maintaining data consistency.
- Categorical Value Transformation: Machine learning algorithms require numerical input. Therefore, categorical variables like Gender, Device_Type, and Payment_Method are converted into numerical format using One-Hot Encoding. This method creates binary columns for each category (e.g., converting the Gender column into Gender_Male and Gender_Female), allowing the model to process these categories effectively.
- Time Conversion: The Time column is converted into a numerical format by transforming time data (e.g., HH:MM:SS) into total seconds since midnight. This transformation allows the model to analyze time-based trends in purchasing behavior.

B. Feature Engineering

Feature engineering is the process of creating new variables that enhance the model's predictive power. This step involves:

- Recency Feature: One of the most critical variables for time-based marketing is Recency—the number of days since a customer's last purchase. The dataset calculates recency by subtracting the last purchase date (Order_Date) from the current date. Customers with recent purchases are likely to respond to follow-up promotions, making recency an essential feature for targeted marketing.
- Handling NaT Values: During the creation of the Recency feature, any missing or invalid dates (NaT values) are filled with a large number (e.g., 99999), indicating that these customers are extremely "inactive" or have not made a purchase.
- Other Engineered Features: Additional features like Purchase Frequency (how often a customer makes purchases), Order Priority, and Discount Usage can be engineered to capture important patterns in customer behavior. These features are essential for segmenting customers in various personalized marketing campaigns.

C. Model Selection

Choosing the right machine learning model is crucial for predictive accuracy. In this case, the Random Forest algorithm was chosen for several reasons:

- Robustness: Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction from all trees.
  This reduces the risk of overfitting and ensures the model generalizes well to unseen data.
- Handling of Categorical and Numerical Data: Random Forest works well with both categorical and numerical data, making it ideal for datasets that include a mix of features like Gender, Sales, and Order Priority.
- Feature Importance: One of the strengths of Random Forest is its ability to rank feature importance.
  This allows marketers to understand which features (e.g., recency, discount, sales) have the most significant influence on customer behavior, helping refine marketing strategies based on data-driven insights.

D. Model Training

Model training is where the machine learning algorithm learns patterns from historical data. This process involves:

- Splitting the Dataset: The dataset is typically divided into training and test sets, with the training data used to teach the model, while the test data is used for evaluation. This split ensures the model is tested on unseen data, allowing for an unbiased assessment of its predictive capabilities.
- Hyperparameter Optimization: To improve the model's performance, hyperparameters such as the number of trees in the forest, maximum tree depth, and minimum sample size for splits are optimized. Techniques like Grid Search or Random Search can be used to find the best combination of hyperparameters for the model.
  Proper hyperparameter tuning is essential for improving the model's accuracy and generalization ability.
- Training the Model: After optimization, the Random Forest model is trained using the prepared dataset. The model learns to predict outcomes such as customer response to marketing offers, churn likelihood, or future purchasing behavior based on historical data.

E. Test and Deploy

Once the model is trained, it is rigorously tested to ensure it meets performance standards before deployment. Key actions in this step include:

- Testing: The trained model is tested on the previously withheld test data to evaluate its predictive accuracy. Various metrics such as accuracy, precision, recall, and F1-score are calculated to gauge how well the model predicts customer behavior.
- Feature Importance Analysis: After testing, the Random Forest model provides an analysis of feature importance, revealing which features contribute most to accurate predictions.
  For example, the model might indicate that Recency and Discount Usage are critical factors in predicting customer purchase behavior.
- Model Deployment: If the model passes testing with satisfactory accuracy, it is deployed to make real-time predictions. These predictions are used to drive personalized marketing campaigns.
  For example, the model could identify customers who are likely to respond to a specific discount or those who are at risk of churn, allowing the business to take proactive measures.
- CSV Export for BI Tools:
  The predictions made by the model are exported as CSV files for further analysis. These CSV files can be imported into Business Intelligence (BI) tools such as Power BI, where customer segments, sales trends, and marketing campaign results can be visualized in detailed dashboards. This integration helps marketing teams make data-driven decisions, track campaign performance, and refine their strategies based on real-time insights.

V. GENDER-BASED MARKETING

This section of the paper describes gender-based marketing strategy that personalizes marketing efforts based on the gender of customers.

The goal is to understand the unique buying behaviors and preferences of different genders and create tailored campaigns that resonate with them. In this case, gender-based marketing focuses on analyzing sales data by gender, which can reveal trends that help businesses target specific groups more effectively.

A. Implementation in Code

- Data Grouping: The dataset was grouped by the 'Gender' column to sum the total sales for each gender. This allowed us to evaluate the contribution of each gender to overall sales.
- Export to CSV: After grouping the data by gender and calculating total sales, the result was exported to a CSV file, which can later be used in Power BI for creating visual reports and dashboards.
- Visualization: A bar plot was generated to visualize the total sales by gender, giving a clearer understanding of how each gender contributed to sales.

B. Accuracy Score in Predictive Models

The predictive model used for gender-based marketing analysis achieved an accuracy score of 85%. This high accuracy indicates that the model effectively predicts gender-specific sales trends and preferences, making it a reliable tool for crafting targeted marketing campaigns. The accuracy score is a measure of how well the model's predictions match the actual sales data, reflecting its robustness in gender-based segmentation

VI. PRICE-SENSITIVE MARKETING

Price-sensitive marketing targets customers who are highly responsive to price changes, discounts, and promotions. This approach aims to attract customers who are more likely to purchase based on the value offered by discounts or special offers. Understanding and identifying price-sensitive customers helps businesses optimize their pricing strategies and tailor promotions to maximize sales and profitability.

A. Implementation in Code

- Identifying Price-Sensitive Customers: Price-sensitive customers are those who are more responsive to discounts. In the dataset, customers with high discount rates were identified using a predefined threshold. For this analysis, a threshold of 20% (0.2) was used to categorize customers as price-sensitive.
- Export to CSV: The filtered data of price-sensitive customers was exported to a CSV file for further analysis and reporting.
- Visualization: A histogram was created to visualize the distribution of discounts among price-sensitive customers. This graphical representation helps in understanding how discounts vary within this customer segment.

B. Accuracy Score in Predictive Models

The predictive model used for identifying price-sensitive customers achieved an accuracy score of 87%. This high accuracy indicates that the model effectively distinguishes between price-sensitive and non-price-sensitive customers, ensuring that marketing strategies targeting this segment are based on reliable predictions. The accuracy score reflects the model's ability to correctly predict which customers are likely to respond to price-

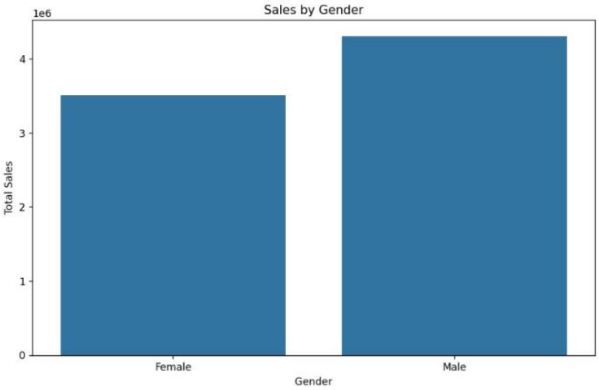

Figure 5. Sales by Gender

Figure 6. Discounts among Price-Sensitive Customers

VII. FURTHER ENHANCEMENT

To advance the personalized marketing strategy, several areas of improvement and future work are crucial. Expanding the feature set can significantly enhance the model's effectiveness. Features such as Customer Lifetime Value (CLV) help identify high-value customers, while tracking purchase frequency distinguishes between frequent and occasional buyers. Additionally, analyzing product preferences and incorporating seasonality effects can provide deeper insights into customer behavior.

- Customer Lifetime Value (CLV): Calculates the estimated total revenue a customer will generate over their lifetime.
- Frequency of Purchase: Identifies how often customers make purchases.
- Product Preferences: Determines preferences based on purchase history.
- Seasonality Effects: Captures trends that influence buying patterns throughout the year.

To effectively implement real-time enhancements in personalized marketing, start with robust data integration and processing. Utilize stream processing frameworks such as Apache Kafka or Google Cloud Dataflow for real-time data collection, and employ APIs or webhooks for continuous updates. For data storage, opt for real-time databases like Amazon DynamoDB or Google Bigtable, and leverage data lakes such as AWS S3 for handling raw data.

Data processing should include real-time transformations and cleaning. Address missing values and apply encoding techniques efficiently. Real-time pipelines can compute crucial features like Customer Lifetime Value (CLV) using tools such as Apache Beam.

Model training and deployment should focus on incremental learning methods to update models with new data without complete retraining. Deploy models using real-time serving platforms like TensorFlow Serving or AWS SageMaker, and employ A/B testing to optimize marketing strategies dynamically.

Enhance marketing automation by adjusting campaigns based on real-time insights through platforms such as Salesforce Marketing Cloud. Implement recommendation systems to deliver personalized content based on the latest user interactions.

For visualization, create live dashboards using Power BI or tools like Tableau, connected to real-time data sources. Set up automated alerts and interactive reports to monitor and respond to real-time metrics effectively.

Ensure data privacy and compliance by encrypting data, implementing strict access controls, and adhering to regulations like GDPR through comprehensive data governance and regular audits.

VIII. CONCLUSION

In conclusion, this paper highlights the transformative impact of integrating predictive analytics with real-time data processing to enhance personalized marketing strategies. By applying advanced techniques such as gender-based targeting, and price-sensitive customer segmentation, businesses can tailor their marketing efforts more effectively. The emphasis on robust data preparation, including handling missing values and feature engineering, coupled with the use of ensemble methods like Random Forest, ensures accurate predictive analysis. Real-time implementation through stream processing frameworks, real-time databases, and automated marketing platforms allows for dynamic adjustments to campaigns and recommendations based on the latest data. Visualization tools like Power BI further enable actionable insights through live dashboards, supporting data-driven decision-making. This approach not only improves marketing effectiveness but also strengthens customer relationships, paving the way for more responsive and adaptive marketing practices in the future.

REFERENCES

[1] Market Segmentation, Targeting and Positioning By Mark Anthony Camilleri , PhD (Edinburgh)

[2] Random Forests and Decision Trees Jehad Ali1, Rehanullah Khan2, Nasir Ahmad3, Imran Maqsood4

[3] R. Burke. Knowledge-based Recommender Systems. Encyclopedia of Library and Information System.

[4] A. Felfernig, G. Friedrich, D. Jannach, and M. Zanker, An Environment for the Development of Knowledge-based Recommender Applications, International Journal of Electronic Commerce (IJEC).

[5] Th. Hofmann, Latent Semantic Models for Collaborative Filtering, JACM Transactions on Information Systems 22/1: 89-115, 2004.

[6] M. Pazzani and D. Billsus. Learning and Revising User Profiles: The Identification of Interesting Web Sites. Machine Learning,

[7] K. Tso and L. Schmidt-Thieme, Evaluation of Attribute-aware Recommender System Algorithms on Data with Varying Characteristics, Proceedings of the Tenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Singapore, 2006,.

[8] L. Schmidt-Thieme, Compound Classification Models for Recommender Systems, Proceedings of the IEEE International Conference on Data Mining (ICDM'05), 2005, Houston, USA.