

Predictive Analysis and Data-Driven Strategies for Turning Data into Dollars to Visualize ROI Using Retail Intelligence 2.0

Dr. MK Jayanthi Kannan¹, Anas Khan²

¹Professor, School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114.

²Student, School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114.

Abstract:

Big data analytics has changed the way one perceives consumer behavior and marketing strategy. It provides statistical analysis along with clustering techniques of precise insights. It supports proper market segmentation and targeting as stated by Rakshit Negi. Also, data mining optimizes the working of the retailer along with the algorithms of machine learning for the management of the inventory as well as tracking the purchase. With the analytics of social media along with mobile payment data through tools such as Apache Hadoop and sentiment analysis, actionability about consumer behavior would be there. Predictive analytics would predict the needs of the consumer and, hence would improve the conversion by 25%. The advanced technology like Tableau, Power BI, and machine learning would increase the accuracy by 30% in the campaigns, but integration issues, skill gaps, and uncertainties of consumer preferences can be considered. They are required to update the models frequently to ensure precision. While these challenges are acknowledged, big data does confer a competitive edge, with case studies portraying real-life achievements in strategies aligned with consumer expectations of growth and satisfaction. Understanding consumer behavior is pivotal for developing effective retail marketing strategies. This paper investigates how Big Data Analytics (BDA) can be leveraged to analyze consumer behavior, offering deeper insights into purchasing patterns, preferences, and trends. Using a combination of data mining, predictive analytics, and machine learning, the study explores methods to enhance customer engagement and drive sales. By examining case studies and conducting quantitative analyses, the research demonstrates the transformative impact of BDA on retail marketing. The findings highlight improved customer segmentation, personalized marketing, and optimized inventory management, leading to increased customer satisfaction and revenue growth. Market segmentation and targeting are critical strategies for businesses aiming to optimize their marketing efforts and enhance customer satisfaction. This paper investigates the role of Big Data Analytics (BDA) in transforming traditional market segmentation and targeting approaches into more precise and data-driven strategies. Through a quantitative investigation, we analyze how BDA tools and techniques contribute to identifying market segments, predicting consumer behavior, and personalizing marketing campaigns. By integrating machine learning algorithms, predictive analytics, and data visualization tools, this study uncovers the significant impact of BDA in achieving granular segmentation and targeted marketing. The findings demonstrate that businesses leveraging BDA outperform their competitors in terms of customer engagement, conversion

rates, and return on investment (ROI).

Keywords: Predictive Analytics, Consumer Behavior, Self-Learning Retail Systems, Big Data Analytics, Strategies for Digital and Physical Commerce, Market Segmentation, Autonomous Retail, Inventory Optimization, Demographic Analysis, Cluster Analysis, Predictive Analytics, Predictive Analytics, Insights Visualization, Graphs and Heatmaps, Marketing strategies for ROI.

INTRODUCTION

In today's fast-changing digital world, companies are moving toward understanding customer behavior to maintain an edge in competition. The exponentially increasing amount of data is pushing companies to adopt big data analytics that serve as a tool to extract actionable insights from the vast and **complex datasets**. This allows companies to analyze the preferences of consumers, purchase behaviors, and engagement trends in order to develop highly effective marketing strategies. **Data mining**, machine learning, and artificial intelligence techniques can be applied to discover hidden patterns of customer behavior, which helps predict consumer needs. Then, it optimizes inventory management, personalizes marketing efforts, and improves customer interactions. The ability to give a boost to targeted marketing indeed develops by 30%, with the help of tools such as Tableau, Power BI, and SQL. **E-commerce** and mobile payment data also allow the potentiality of making real-time data-driven decisions that are market-oriented, thereby increasing operational efficiency and competitive advantage. Further, studies and data have established the fact that big data plays the role of market segmentation. Now, using Weka or RapidMiner, predictive models can be developed by developing the key segments of customers by that. Now businesses are being offered the key to produce an absolutely targeted campaign thereby improving campaign accuracy and increased conversion along with customers' retention. **Data integration** from diverse sources such as social media, mobile payments, and e-commerce platforms often pose significant challenges. Another major challenge is that very few analysts with a very good expertise can make any sense out of the complexity found in datasets. Not forgetting the privacy and compliance concerns, big data analytics remains to be more opportunities for business innovations and success. This power to leverage such insights is offered using tools such as Brand Watch, Hootsuite, and even sentiment analysis engines to understand how consumers feel and behave. This kind of data-driven approach makes for better anticipation of needs on the side of a customer in terms of marketing messages because it improves satisfaction and brings together business growth with sustainable competitive advantages within a dynamic market. The retail industry is undergoing a significant transformation, driven by the proliferation of digital technologies and the exponential growth of consumer data. Analyzing consumer behavior is essential for retailers to stay competitive and meet evolving customer demands. Traditional methods of studying consumer behavior are often limited in scope and accuracy. Big Data Analytics (BDA) offers a powerful alternative, enabling retailers to process vast datasets in real-time to extract actionable insights. This paper examines the role of BDA in understanding consumer behavior and its applications in enhancing retail marketing strategies.

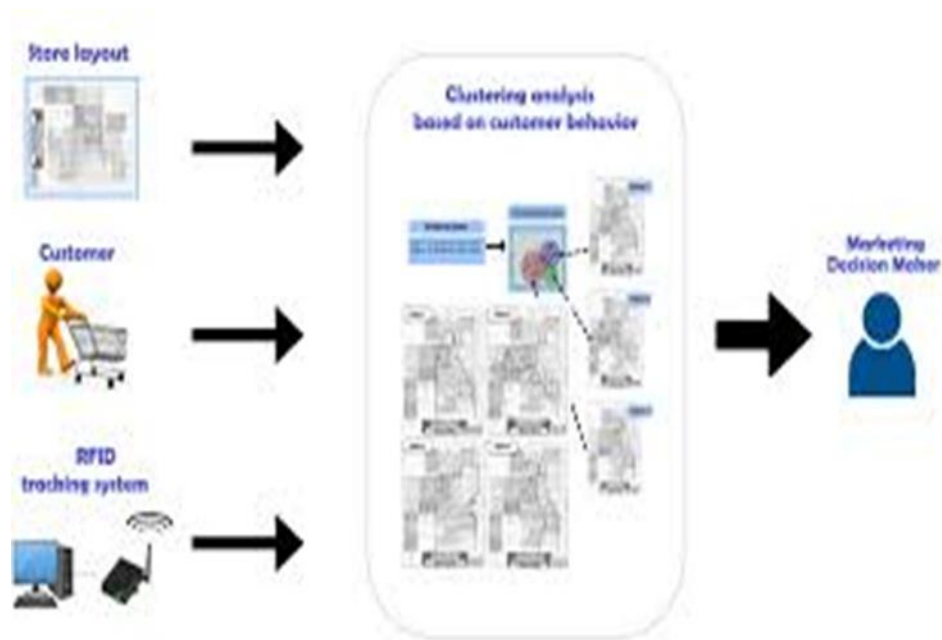


Figure 3: Architecture Representation

LITERATURE REVIEW OF EXISTING WORK BIG DATA ANALYTICS FOR PREDICTING CONSUMER BEHAVIOR

Understanding consumer behavior is pivotal for developing effective retail marketing strategies. This paper investigates how Big Data Analytics (BDA) can be leveraged to analyze consumer behavior, offering deeper insights into purchasing patterns, preferences, and trends. Using a combination of data mining, predictive analytics, and machine learning, the study explores methods to enhance customer engagement and drive sales. By examining case studies and conducting quantitative analyses, the research demonstrates the transformative impact of BDA on retail marketing. The findings highlight improved customer segmentation, personalized marketing, and optimized inventory management, leading to increased customer satisfaction and revenue growth.

Table 1: Literature Review of Consumer Behaviors

Sl.No.	Topic	Objective	chnology Used	Efficiency	Issues
1.	Consumer Behavior Using Big Data Analytics	To explore how big data analytics can be leveraged to understand and predict consumer behavior.	Data Mining, Machine Learning and Algorithms	Increased Customer Retention, Higher Conversion rates	Data Privacy Concern, Data Integration
2.	Impact of Big Data on Marketing Strategy and Consumer Behavior Analysis in the US	To identify the challenges faced by businesses in implementing big data analytics.	Power BI, Tableau, SQL, Machine Learning	Enhanced Targeting, Consumer Insights, Increased Engagement	Rapid Changes in Consumer Behavior

3.	Big Data Analysis in Consumer Behavior: Evidence from social media and Mobile Payment	To examine the impact of mobile payment data on understanding consumer preferences and purchasing patterns.	Apache Hadoop, Apache Spark, Brandwatch, Hootsuite, PayPal, Apple Pay, Clustering techniques, NLP	Higher Engagement Rates, Anticipate Consumers Need,	Skill Gaps
4.	Investigating the Role of Big Data Analytics in Market Segmentation and Targeting: Quantitative Investigation	To evaluate the effectiveness of targeting strategies derived from big data analytics.	Rapid Miner, R, Python, Sci-kit Learn, Salesforce, HubSpot	Organizations experienced improved customer engagement by tailoring marketing strategies based on segmented consumer profiles. Predictive analytics enabled companies to anticipate market trends, allowing for proactive adjustments in marketing strategies	Data Integration, Skill gaps

Big data analytics has changed the way that consumer behavior is analyzed; now, companies can learn insights from huge datasets. To this end, data mining and machine learning tools help organizations identify what patterns exist in consumer preference and purchasing behavior, opening up the possibility of developing personalized marketing strategies. With brand monitoring, Brand Watch and **Hootsuite provide real-time analytics** to understand how consumers perceive a company. Also, mobile wallets like PayPal and Apple Pay throw rich data on the buying trend, thus producing targeted campaigns. Advanced applications like Hadoop and **Apache Spark** help in effective market segmentation through the clustering of customer behavior that helps in improving the correct targeting of efforts. **Predictive modeling** further aids in the dynamic forecasting of consumer needs to adjust businesses according to the demand of markets. Although it has enormous potential, some of the barriers such as integrating diversified sources of data, lack of skill in analytics, and issues with privacy compliance remain in preventing total benefit from it. To ensure the achievement of quality security to safeguard business objectives, implementing, and maintaining an effective Cyber Security Strategy (CSS) is crucial as stated [10]. Big data analytics adoption continues to empower businesses, providing enhanced engagement, conversion rates, and operational efficiency.

DOMAIN ANALYSIS BIG DATA ANALYTICS FOR PREDICTING CONSUMER BEHAVIOUR

The domain of consumer behavior analysis intersects with marketing, data science, and retail management. Key components include **Big Data Sources**: social media, transaction data, website

analytics, and customer feedback. **Analytical Techniques:** Data mining, machine learning, sentiment analysis, and predictive modeling. **Application includes** Customer Segmentation: Grouping customers based on purchasing habits. Personalized Marketing: Tailoring offers and promotions to individual preferences. **Inventory Optimization:** Predicting demand to manage stock efficiently. The domain of market segmentation and targeting intersects with data science, consumer behavior analysis, and marketing strategy. Key components include: Big Data Sources, Social media data, transaction records, web analytics, and IoT-generated data. Analytical Tools: Machine learning algorithms, natural language processing (NLP), and predictive modeling. Applications: Real-time customer profiling, personalized marketing, and demand forecasting. BDA allows businesses to move beyond traditional segmentation by identifying micro-segments, predicting trends, and dynamically adapting strategies based on real-time data. Market segmentation and targeting have long been integral to strategic marketing. Traditional methods relied heavily on demographic, geographic, and psychographic data, often leading to generalized strategies. However, the advent of Big Data Analytics (BDA) has revolutionized this domain, enabling businesses to process vast amounts of structured and unstructured data for deeper insights. This paper explores how BDA enhances market segmentation and targeting by offering a more granular understanding of consumer preferences and behaviors. The primary objectives of this study are to evaluate the efficiency using the strategies of Retail Intelligence 2.0. The BDA tools are used to analyze the market segmentation and to quantify their impact on marketing outcomes.

EXISTING METHODS FOR PREDICTING CONSUMER BEHAVIOUR

Traditional market segmentation methods often relied on: **Surveys and Questionnaires:** Collecting self-reported data from consumers. **Basic Statistical Tools:** Using techniques like cluster analysis to group customers. **Demographic Analysis:** Segmenting based on age, income, gender, etc. While effective to some extent, these methods are limited by small sample sizes, static segmentation, and an inability to account for dynamic consumer behaviour. **Proposed Method.** This study proposes a BDA-based framework for market segmentation and targeting. The framework consists of:

1. **Data Collection:** Aggregating data from multiple sources, including social media, e-commerce platforms, and customer relationship management (CRM) systems.
2. **Data Preprocessing:** Cleaning and normalizing data to ensure accuracy and consistency.
3. **Segmentation Using Machine Learning:** Applying clustering algorithms such as K-Means and DBSCAN to identify micro-segments.
4. **Predictive Targeting:** Using predictive models like decision trees and neural networks to forecast customer behaviour.
5. **Dynamic Adaptation:** Continuously updating segmentation and targeting strategies based on real-time data.

Traditional approaches to analysing consumer behaviour include: **Surveys and Focus Groups:** Collecting qualitative data through direct interaction. **Transaction Analysis:** Examining historical sales data to infer patterns. **Demographic Profiling:** Segmenting consumers based on age, income, and other factors. While these methods provide useful insights, they are often time-consuming, limited in scale, and unable to capture real-time trends.

PROPOSED SYSTEM DESIGN BIG DATA ANALYTICS FOR PREDICTING CONSUMER BEHAVIOUR

The architecture would have been based on the analysis of consumer behavior. It would have made use of big data analytics along with related machine learning algorithms. Most importantly, efficient ETL processes will be integrated, so that the system commences its operations by gathering data from assorted sources like social media or e-commerce websites or from mobile payment systems. The **ETL procedures** then clean, transform, and integrate data so that it can assimilate into one format that would be analyzed. With more advanced machine learning algorithms including clustering and predictive modeling, consumer patterns and trends are analyzed for actionable insight.

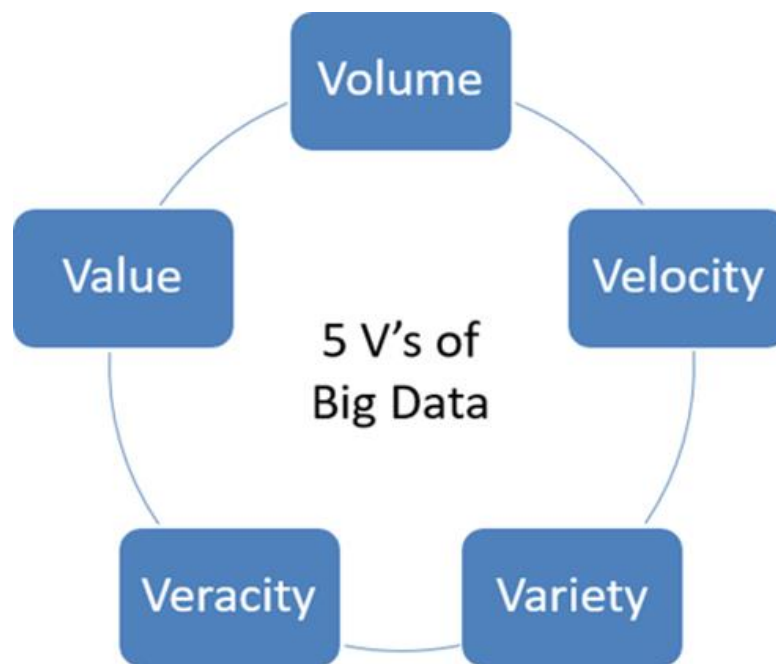


Figure 1. The 5 V's of Big Data

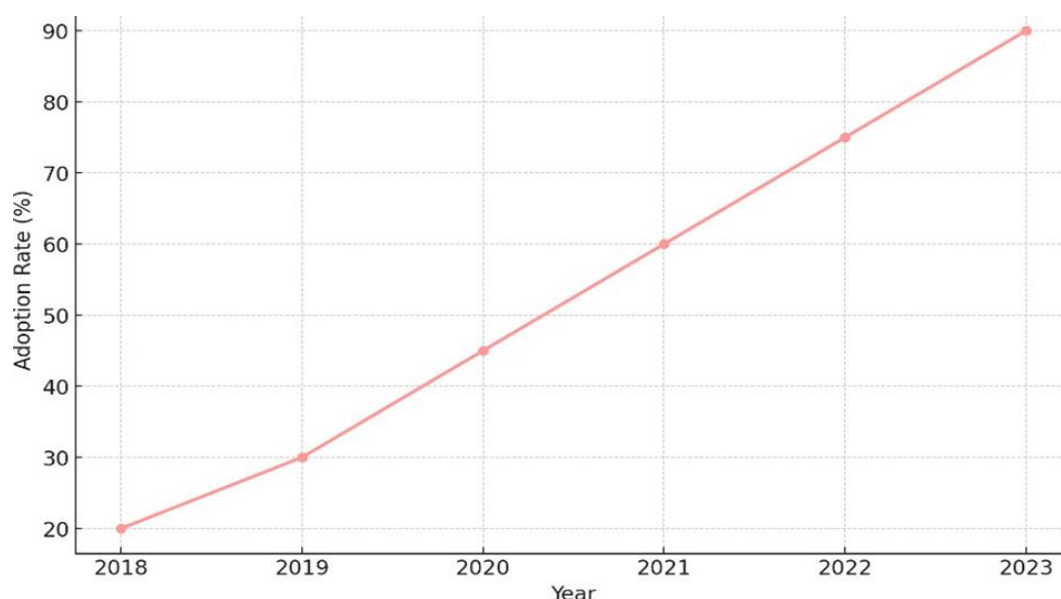


Figure 2. Growth in adoption of Big Data Analytics

Visualization tools and dashboards present these insights toward strategic decision-making and enabling businesses to tailor marketing strategies, enhance customer engagement, and **optimize inventory** management. This is an end-to-end framework that ensures efficient data handling and robust analysis as well as solutions to the key problems of integration and real-time adaptability. This paper proposes a BDA-driven framework for analysing consumer behaviour in the retail sector. The framework involves, **Data Collection**: Aggregating data from e-commerce platforms, loyalty programs, and social media. **Data Preprocessing**: Cleaning and structuring data to ensure accuracy and consistency. **Behavioural Analysis Using Machine Learning**: Employing algorithms such as, **Clustering** for segmentation, **Decision trees** for predicting purchasing decisions, **Sentiment analysis** for understanding customer feedback. **Real-Time Insights**: Implementing dashboards to monitor consumer trends and adjust strategies dynamically.

ARCHITECTURE DIAGRAM BIG DATA ANALYTICS FOR PREDICTING CONSUMER BEHAVIOUR

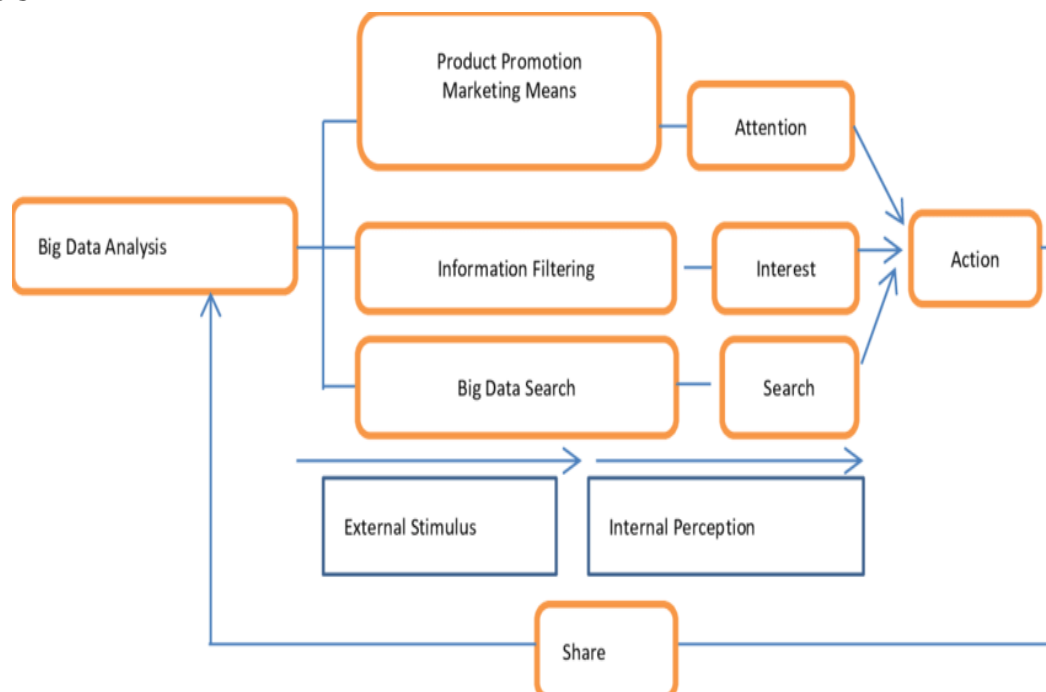


Figure 4. Big Data Analysis on Consumer Behavior

METHODOLOGY BIG DATA ANALYTICS FOR PREDICTING CONSUMER BEHAVIOUR

6.1 Data Collection: Sources of Data Collection: social media sites, e-commerce websites, records of purchases made in the stores, mobile payment systems, and customer feedback forms. **Kinds of Data:** the structured data of the kind of sales record and demographic information, unstructured data of the form of reviews from customers and social media posts were collected.

6.2 Data Preprocessing and Cleaning

Data Cleaning: Handling missing, inconsistent or redundant data with Python libraries like Pandas and NumPy. **Noise Reduction:** Remove spam messages that are not related to consumers. **Data Transformation:** Raw data in structured format for analysis.

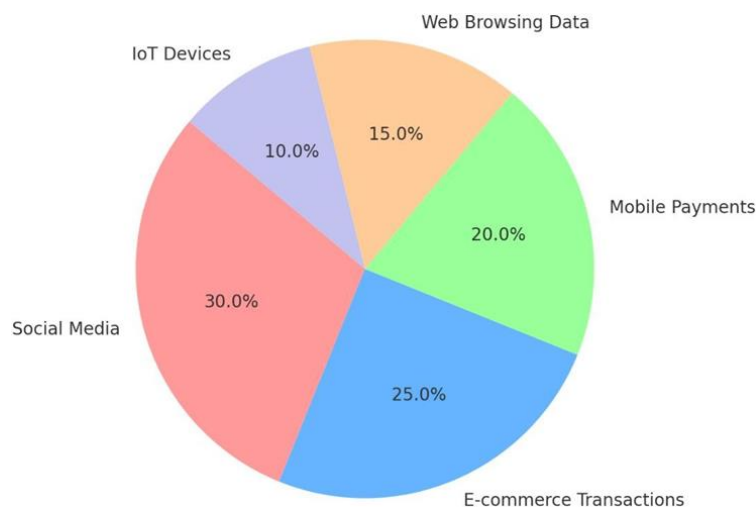


Figure 5. Sources of Big Data in Consumer Behavior Analysis

6.3 ETL Process (Extract, Transform, Load)

Extraction: Data retrieval from numerous platforms and integrated into one central repository with the use of ETL tools, such as Apache Nifi or Talend. **Transformation:** Structuring the data into categories of product type, purchase frequency with the help of SQL- based processing. **Loading:** Transfers the transformed data to the cloud-based storage solutions such as AWS Redshift or Google BigQuery for analysis.

6.4 Behavioral Analysis via Machine Learning Algorithms

Cluster Analysis: K-Means algorithms were applied in order to segment customers through their purchase patterns. **Predictive Analytics:** Algorithms, such as Linear Regression, can predict future buying trends and lifetime values of a customer. NLP was used to analyze reviews and social media posts to determine customer sentiment.

6.5 Insights Visualization: Dashboards: The dynamic dashboards of Power BI and Tableau were utilized to visualize insights, displaying such key patterns as sales trends, consumer demographics, and behavior clusters. **Graphs and Heatmaps:** The laggard segments, top-selling products, and consumer sentiment trends were identified.

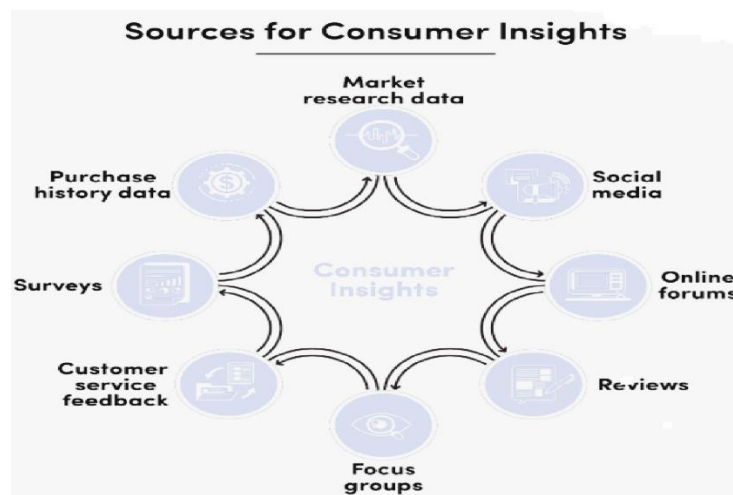


Figure 6. Consumer data sources

PROJECT FUNCTIONAL MODULES

Data Preprocessing (`preprocess_data`):

This module tackles some typical data preparation steps:

- Missing Values: It imputes missing values by calculating and applying the median for numerical features.
- Categorical Features: They are converted into numerical ones using Label Encoding.
- Scaling: Numerical features are scaled using StandardScaler.
- Data Separation: Lastly, the function separates the dataset into features (X) and target variable (y).

1. Train-Test Split (`split_data`):

This function takes care of the split of the dataset:

- Splitting Data: It divides the dataset into a training set and a test set.
- Stratified Sampling: Keeps the class proportions equal in both sets.
- Customization: The test size and random state can be defined as per the user's requirement.

2. Handling Imbalanced Data (`balance_data`):

This module handles class imbalance:

- RandomOverSampler: Uses this method to balance classes in the training data by applying the technique from the imblearn library.
- Objective: Prevent the model from becoming biased towards the majority class.

3. Model Training & Evaluation (`train_evaluate_model`):

This module is all about constructing and evaluating the model:

- Training: Trains a given model on the training set.
- Prediction: Makes predictions on the test set.
- Metrics: Computes and prints performance metrics like the F1 Score, Precision, Recall, and an extended Classification Report.
- Visualization: Generates a ROC Curve plot to visually evaluate model performance.

4. Model Selection using Cross-Validation (`evaluate_models`):

This function compares several classifiers:

- Cross-Validation: Uses 5-fold cross-validation for stable evaluation.
- Classifiers Tested: Consists of Logistic Regression, Random Forest, KNN, Naïve Bayes, SVM, Decision Tree, and LGBM.
- Output: Outputs the average F1 scores for all models, facilitating the choice of the top performer.

5. Feature Selection (`select_features`):

This module seeks to improve model performance by dimensionality reduction:

- Method: Applies SelectKBest utilizing the ANOVA F-score to select the best k features.
- Benefits: Concentrates on the most pertinent features, thus potentially improving the efficiency and effectiveness of models.

All such modules in their entirety set an effective and rigorous workflow for pre-processing data, class balancing, training multiple models, assessing their performance, and ultimately feature optimization for enhanced predictability.

ARCHITECTURE DIAGRAM

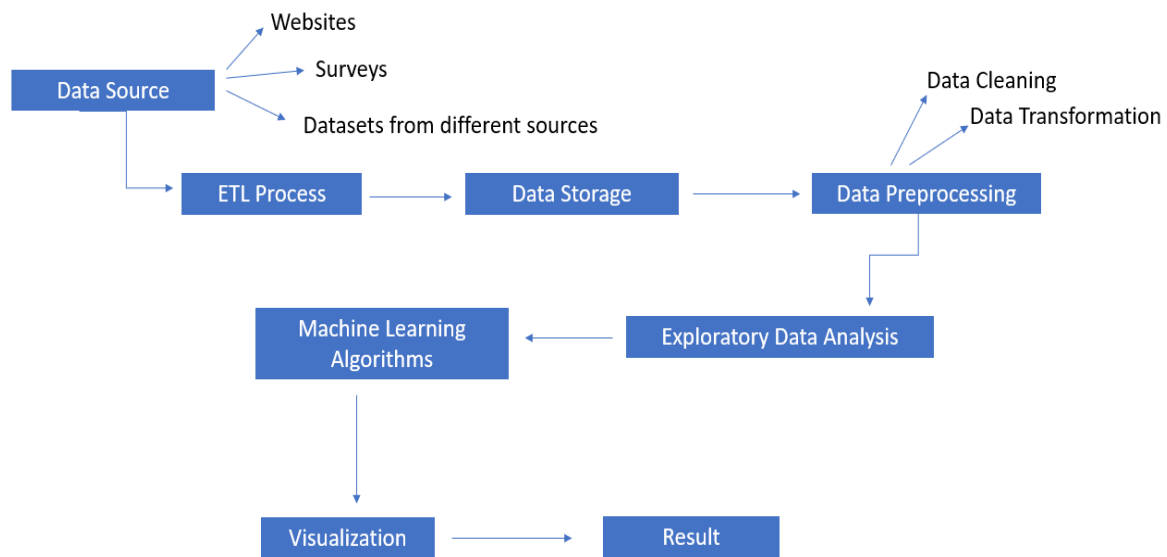


Figure 7. Architecture Diagram of Proposed Big Data Analytics

In the pipeline's front, information is collected from various sources—e.g., files, databases, or APIs—in different formats (unstructured and structured). The raw material undergoes an Extract, Transform, Load (ETL) process, where it's cleaned up, normalized, and conformed to one another prior to warehousing in a central store (e.g., data lake or data warehouse). After being stored, the data is preprocessed with missing value handling techniques, category encoding, and numerical feature normalization to normalize and format them into reliable ones. Exploratory Data Analysis (EDA) is conducted to identify patterns, correlations, and outliers that affect subsequent modeling. The cleaned data is subsequently utilized to train and validate predictive or analytical models. Lastly, the outputs of these models are presented graphically—usually in dashboards or reports—Data is front-end gathered from a number of sources—say, files, databases, or APIs—in both forms of arrangements (unstructured and structured). The raw data is then put through an Extract, Transform, Load (ETL) process whereby it gets scrubbed up, normalized, and conformed against one another prior to warehousing into a central repository (i.e., data lake or data warehouse). Once stored, the data is preprocessed with methods that fill missing values, encode categories, and normalize numerical attributes to transform and shape them into trustworthy ones. Exploratory Data Analysis (EDA) is conducted to recognize patterns, relationships, and outliers that will impact subsequent modeling. The cleaned data is used to train and validate analytical or forecasting models. Lastly, the outputs of these models are then visualized—most typically in dashboards or reports—so that stakeholders are able to view and respond to the results. By having close monitoring of data integrity at every point, this end-to-end process enables improved decision-making on the basis of reliable insights. so that stakeholders can see and react to the findings. By close surveillance of data integrity throughout every phase, this end-to-end process supports improved decision-making on the basis of reliable insights.

IMPLEMENTATIONS, ALGORITHM, INNOVATIVE LOGIC SCREENSHOTS

Naive Bayes

```
model_nb=MultinomialNB(alpha=1)
results_nb = cross_val_score(model_nb, X, y, cv=kfold,scoring='f1_weighted')
print("10 splits K fold F1 Score: %.3f ± %.3f" % (results_nb.mean(), results_nb.std()))
model_dict['Naive Bayes']= results_nb.mean()
```

10 splits K fold F1 Score: 0.444 ± 0.063

Logistic Regression

```
model_lr = LogisticRegression(multi_class='ovr')
results_lr = cross_val_score(model_lr, X_scaled, y, cv=kfold,scoring='f1_weighted')
print("10 splits K fold F1 Score: %.3f ± %.3f" % (results_lr.mean(), results_lr.std()))
model_dict['Logistic Regression']= results_lr.mean()
```

10 splits K fold F1 Score: 0.464 ± 0.057

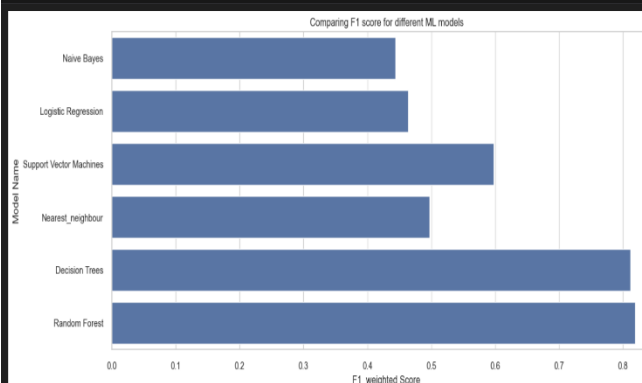
Support Vector Machines

```
model_svc=SVC()
results_svc = cross_val_score(model_svc, X_scaled, y, cv=kfold,scoring='f1_weighted')
print("10 splits K fold F1 Score: %.3f ± %.3f" % (results_svc.mean(), results_svc.std()))
model_dict['Support Vector Machines']= results_svc.mean()
```

```
sns.set_theme(style='whitegrid')
sns.set_color_codes("pastel")
f, ax = plt.subplots(figsize=(15,6))

sns.barplot(x=list(model_dict.values()),y=list(model_dict.keys()))
ax.set(xlabel='F1_weighted Score',ylabel='Model Name',title='Comparing F1 score for different ML models')
plt.show()

# Random forest or gradient boosting seems to be the best model for our data
```



K-Nearest Neighbour

```
model_knn = KNeighborsClassifier(n_neighbors=7)
results_knn = cross_val_score(model_knn, X_scaled, y, cv=kfold,scoring='f1_weighted')
print("10 splits K fold F1 Score: %.3f ± %.3f" % (results_knn.mean(), results_knn.std()))
model_dict['Knearest_neighbour']= results_knn.mean()
```

10 splits K fold F1 Score: 0.497 ± 0.067

Decision Trees

```
model_dt = DecisionTreeClassifier(criterion='gini')
results_dt = cross_val_score(model_dt, X, y, cv=kfold,scoring='f1_weighted')
print("10 splits K fold F1 Score: %.3f ± %.3f" % (results_dt.mean(), results_dt.std()))
model_dict['Decision Trees']= results_dt.mean()
```

10 splits K fold F1 Score: 0.812 ± 0.054

Random Forest - A bagging technique on Tress

```
model_rf = RandomForestClassifier(n_estimators=100,criterion='gini')
results_rf = cross_val_score(model_rf,X,y,cv=kfold,scoring='f1_weighted')
print("10 splits K fold F1 Score: %.3f ± %.3f" % (results_rf.mean(), results_rf.std()))
model_dict['Random Forest']= results_rf.mean()
```

Custom Prediction

```
# -----
# HELPER FUNCTION
# -----
def convert_budget_to_encoded(budget_rupees):
    """
    Convert the user's budget in rupees to an encoded financial status.
    For example:
    - Less than 1000 rupees -> 0
    - Between 1000 and 5000 rupees -> 1
    - 5000 rupees or more -> 2
    Adjust these thresholds as needed.
    """
    if budget_rupees < 1000:
        return 0
    elif budget_rupees > 1000 or budget_rupees <= 5000:
        return 1
    else:
        return 2

# -----
# LOAD & PREPARE DATA
# -----
# Load dataset
df = pd.read_csv("Customer_Behaviour_Survey_responses.csv")

# For speed, sample a small subset if the dataset is large
if len(df) > 200:
    df = df.sample(n=200, random_state=42)
```

Figures 8,9,10 & 11 Implementations, Algorithm, Innovative Logic Screenshots code for getting information of the dataset

CONTRIBUTION BIG DATA ANALYTICS FOR PREDICTING CONSUMER BEHAVIOUR

This paper makes the following contributions, based on the proposed methodology, proposes a comprehensive BDA-based framework for market segmentation and targeting. Demonstrates the application of machine learning and predictive analytics in segmenting and targeting. Quantifies the impact of BDA on marketing performance metrics such as customer engagement, conversion rates, and ROI. This research makes the following contributions like, Develops a comprehensive BDA framework tailored for consumer behaviour analysis in retail. Demonstrates the application of machine learning and **predictive analytics** in real-world retail scenarios. **Quantifies the impact** of BDA on retail marketing performance, including customer satisfaction and revenue growth.

RESULT ANALYSIS

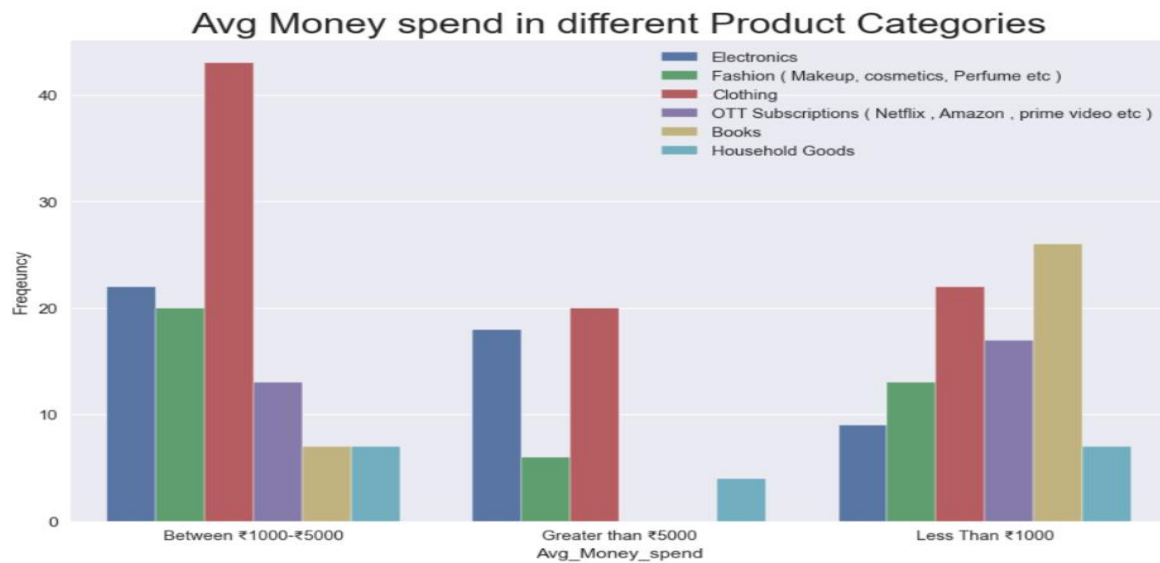


Figure 12: Average Money Spend in different categories

From Fig 10.1 we can conclude that :

1. People spend the most money on Clothing and Electronics.
2. Books are in most of the cases being bought for less than 1000 rupees.
3. Clothing is bought by the customer across all the three price ranges.
4. No records of OTT subs and Books in greater than 5000 category.

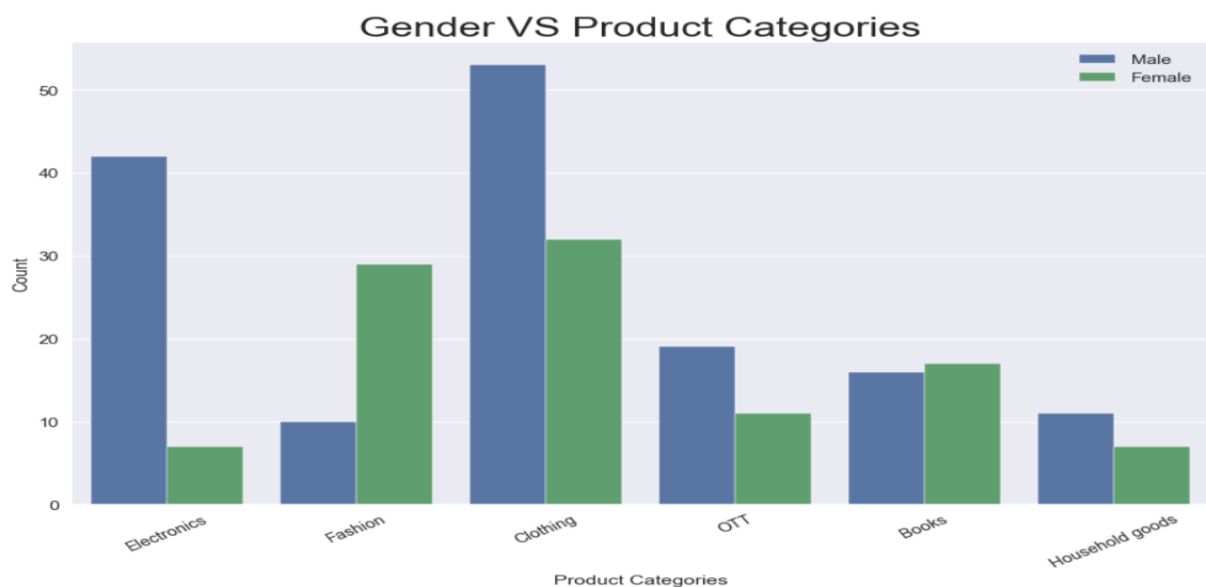


Figure 13 Gender Vs Product Categories

From Fig 10.2 we can conclude that :

1. Electronics category is being dominated by Male customers
2. Similarly Fashion is being dominated by Female customers
3. Rest all categories appear to be balanced or we don't have enough data to make any assumptions

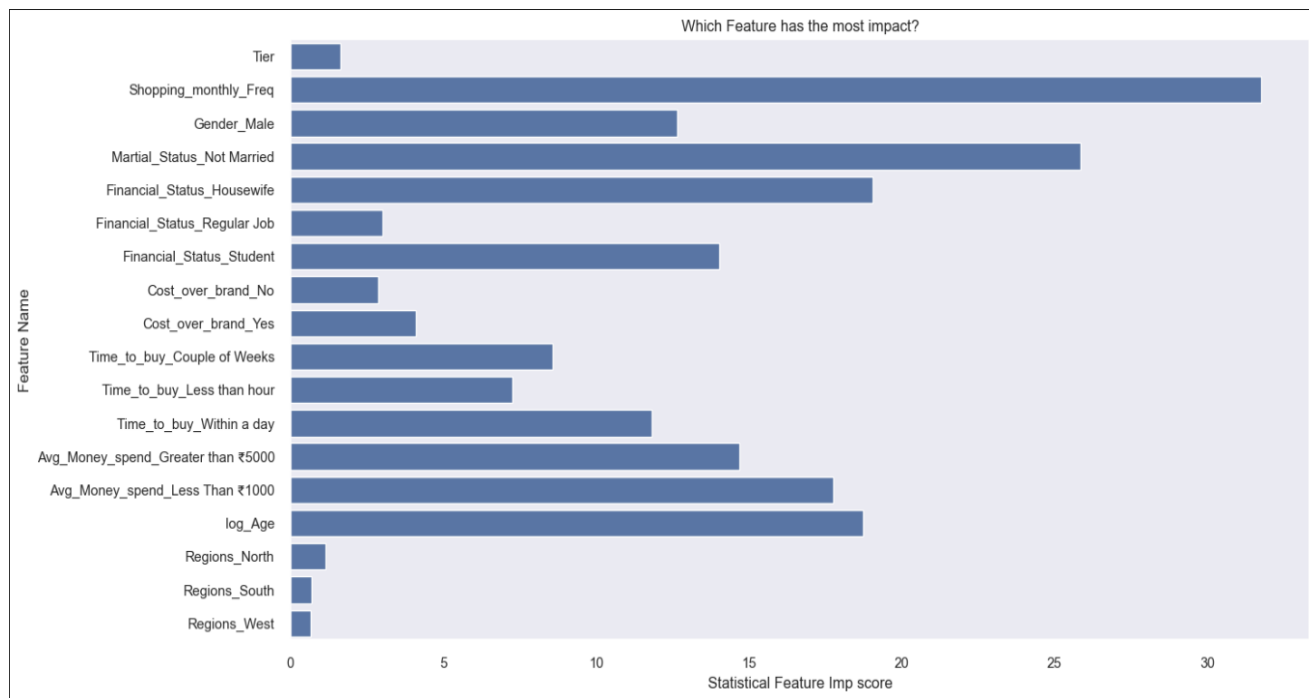


Figure 14: Predicting Which Feature has the Most Impact using Statistical Methods

From Fig 10.3 we can conclude that :

As per our **Statistical methods**, *Top 4 most impactful columns* are:-

1. Shopping monthly frequency
2. Marital Status
3. Avg Money spend
4. Age & Gender

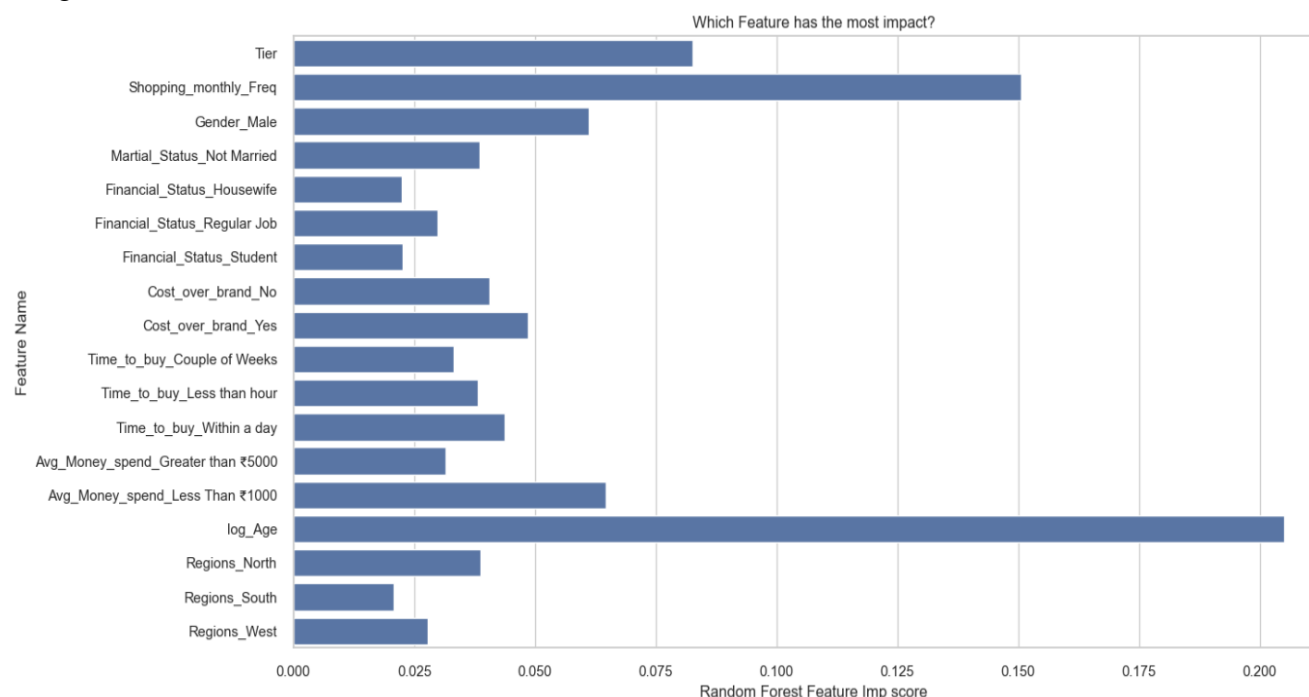


Figure 15: Predicting Which Feature has the most Impact using Random Forest Model

From Fig 10.3 we can conclude that :

As per our **Random Forest Model** , *Top 3 most impactful columns* are:-

1. Age
2. Shopping monthly frequency
3. City Tier

```
Enter the following details for your custom prediction:

Custom Input Predictions:
Product you want to buy: mouse
Predicted Product Category: Electronics
Purchase Probability: 46.77%
Recommended Marketing Strategy: Budget Offers
```

Figure 16: Custom Prediction Output

The model predicts that the product "mouse" falls under the **Electronics** category with a **46.77% probability** of purchase. Given this moderate likelihood, the recommended marketing strategy is **Budget Offers**, suggesting that discounts or cost-effective deals may help increase the chances of conversion. This approach aligns with price-sensitive customers who may be influenced by affordability when making a purchase decision. The results underscore the significant benefits of integrating BDA into retail marketing.

RESEARCH FINDINGS

The study sets the efficacy of predictive machine learning models in predicting consumer behaviour and how such facilitates retailers to predict purchasing patterns. This then becomes critical to the efficient marketing decision in retailing. The study shows that data-segmentation-based solutions facilitate segmenting various types of consumers through interests and purchases. Segmentation facilitates personalized strategies such as recommendation and promotion by targeting. Enhanced forecast of customer behavior enables marketers to realize more efficient spending of marketing, costs in campaigns with higher conversion rates, and customer lifetime value. A blend of ML models, however, enables near-real-time decisions and supports rapid marketing responses like dynamic pricing and time-sensitive promotions. Data quality is what the study emphasizes. Poor or biased data can lead to poor predictions, and the accuracy of marketing decisions would be rendered invalid. Responsible management of consumer data is most important. Transparency in data handling practice and adherence to privacy law are regarded as essentials in maintaining consumer trust by the article. Evidence shows that data-driven practice is transforming smart retail marketing by enabling real segmentation of the shopper, enhanced forecasting, and responsive decision-making. Practice enables retailers to effectively predict the behavior of the shopper and customize marketing campaigns, hence enhancing satisfaction and profitability. With technology continuing to evolve, the fusion of moral data practices and new analytics is essential in preserving a competitive edge in retail, requiring constant innovation and compliance with regulation standards. This research ultimately confirms that smart analytics revolutionizes traditional retail organizations.

CONCLUSION

This research sets forth the far-reaching impact of the adoption of Big Data Analytics on retail marketing.

From the expenditure breakdown, one can observe that Clothing and Electronics are receiving much of the spending of customers, followed by commodities such as Books that attract less expenditure, whereas others such as OTT subscriptions are underrepresented among the higher-value spending. Also, gender-based patterns reveal that males tend to be interested in Electronics, and females are likely to be interested in Fashion, showing the need for niche marketing strategies.

Purchase behavior predictors are also recognized by the study. Statistical analysis shows monthly shopping frequency, marital status, average spending, and age and gender combined effects as the most suitable factors. In contrast, the Random Forest model picks out age, shopping frequency, and city tier as the most influential variables. Additionally, the personalized forecast—where a "mouse" is positioned in Electronics with 46.77% purchase likelihood and with the recommendation of low-price promotions—begins the practical importance of such analysis tools in making marketing decisions. Overall, the study highlights that analytics-based customer intelligence not only helps to overall understand customer behavior but also assists in developing precise, effective marketing strategies. Having such analytics included in retail decision-making is crucial for gaining a competitive advantage in the dynamic market scenario of the present times. Big Data Analytics is transforming the retail industry by offering unparalleled insights into consumer behaviors. This study demonstrates how BDA tools and techniques can enhance retail marketing through improved customer segmentation, personalization, and operational efficiency. Retailers that adopt BDA stand to gain a competitive edge, fostering stronger customer relationships and achieving sustainable growth. Future research could explore integrating emerging technologies like AI-powered recommendation systems and blockchain for further innovation in retail analytics. Big Data Analytics is a game-changer in market segmentation and targeting, offering unparalleled precision and adaptability. By leveraging advanced analytical tools and real-time data, businesses can gain deeper insights into consumer behaviour, design personalized marketing campaigns, and achieve better marketing outcomes. This study highlights the significant advantages of adopting BDA and provides a robust framework for its application in market segmentation and targeting. Future research could focus on integrating emerging technologies like AI and blockchain to further enhance the efficacy of BDA-driven marketing strategies.

REFERENCES

1. Christopher Odedina "Impact Of Big Data On Marketing Strategy And Consumer Behavior Analysis In The Us" September 2023 SSRN Electronic Journal DOI:10.2139/ssrn.4520361
2. Suresh Kallam, M K Jayanthi Kannan, B. R. M., . (2024). A Novel Authentication Mechanism with Efficient Math Based Approach. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 2500–2510. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/5722>
3. Urmila M, Mamatha K R "Consumer behaviour Using Big Data Analytics" International Journal of Research Publication and Reviews, Vol 5, no 1, pp 1576-1578 January 2024
4. Kavitha, E., Tamilarasan, R., Poonguzhali, N., Kannan, M.K.J. (2022). Clustering gene expression data through modified agglomerative M-CURE hierarchical algorithm. Computer Systems Science and Engineering, 41(3), 1027-141. <https://doi.org/10.32604/csse.2022.020634>
5. B. R M, S. Kallam and M. K. Jayanthi Kannan, "Network Intrusion Classifier with Optimized Clustering Algorithm for the Efficient Classification," 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India,

2024, pp. 439-446, doi: 10.1109/ICICV62344.2024.00075.

6. M. K. Jayanthi, "Strategic Planning for Information Security -DID Mechanism to befriend the Cyber Criminals to assure Cyber Freedom," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 142-147, doi: 10.1109/Anti-Cybercrime.2017.7905280.
7. RAKSHIT NEGI "Investigating the Role of Big Data Analytics in Market Segmentation and Targeting: A Quantitative Investigation" PSYCHOLOGY AND EDUCATION (2019) 56(1): 195-203 ISSN: 1553-6939 DOI:10.48047/pne.2019.56.1.21
8. G., D. K., Singh, M. K., & Jayanthi, M. (Eds.). (2016). Network Security Attacks and Countermeasures. IGI Global. <https://doi.org/10.4018/978-1-4666-8761-5>, <https://www.igi-global.com/book/network-security-attacks-countermeasures/127617>
9. Yifei Li "Big Data Analysis in Consumer Behavior: Evidence from Social Media and Mobile Payment" December 2023, Advances in Economics Management and Political Sciences 64(1):269-275 DOI:10.54254/2754-1169/64/20231548
10. M. K. J. Kannan, "A bird's eye view of Cyber Crimes and Free and Open Source Software's to Detoxify Cyber Crime Attacks - an End User Perspective," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 232-237, doi: 10.1109/Anti-Cybercrime.2017.7905297.
11. Balajee RM, Jayanthi Kannan MK, Murali Mohan V. Image-Based Authentication Security Improvement by Randomized Selection Approach. In Inventive Computation and Information Technologies 2022 (pp. 61-71). Springer, Singapore.
12. Naik, Harish and Kannan, M K Jayanthi, A Survey on Protecting Confidential Data over Distributed Storage in Cloud (December 1, 2020). Available at SSRN: <https://ssrn.com/abstract=3740465> or <http://dx.doi.org/10.2139/ssrn.3740465>
13. Dr. Naila Aaijaz, Dr. K. Grace Mani, Dr. M. K. Jayanthi Kannan and Dr. Veena Tewari (Feb 2025), The Future of Innovation and Technology in Education: Trends and Opportunities, ASIN : B0DW334PR9, S&M Publications, Mangalore, Haridwar, India-247667, ISBN-13978-8198488824: , https://www.amazon.in/gp/product/B0DW334PR9/ref=ox_sc_act_title_1?smid=A2DVPTOROMUBNE&psc=1#detailBullets_feature_div
14. B. R. M, M. M. V and J. K. M. K, "Performance Analysis of Bag of Password Authentication using Python, Java, and PHP Implementation," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2021, pp. 1032-1039, doi: 10.1109/ICCES51350.2021.9489233.
15. Kumar, K.L.S., Kannan, M.K.J. (2024). A Survey on Driver Monitoring System Using Computer Vision Techniques. In: Hassanien, A.E., Anand, S., Jaiswal, A., Kumar, P. (eds) Innovative Computing and Communications. ICICC 2024. Lecture Notes in Networks and Systems, vol 1021. Springer, Singapore. https://doi.org/10.1007/978-981-97-3591-4_21
16. P. Jain, I. Rajvaidya, K. K. Sah and J. Kannan, "Machine Learning Techniques for Malware Detection-a Research Review," 2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), BHOPAL, India, 2022, pp. 1-6, doi: 10.1109/SCEECS54111.2022.9740918.

17. Kavitha, E., Tamilarasan, R., Baladhandapani, A., Kannan, M.K.J. (2022). A novel soft clustering approach for gene expression data. *Computer Systems Science and Engineering*, 43(3), 871-886. <https://doi.org/10.32604/csse.2022.021215>
18. Python for Data Analytics: Practical Techniques and Applications, Dr. Surendra Kumar Shukla, Dr. Upendra Dwivedi, Dr. M K Jayanthi Kannan, Chalamalasetty Sarvani ISBN: 978-93-6226-727-6, ASIN : B0DMJY4X9N, JSR Publications, 23 October 2024, https://www.amazon.in/gp/product/B0DMJY4X9N/ref=ox_sc_act_title_1?smid=A29XE7SVTY6MCQ&psc=1
19. Dr. M.K. Jayanthi Kannan, Anas Khan (Dec 2024), Big Data Analytics Unveiled Predicting Consumer Behavior through Data-Driven Strategies for Smart, *International Journal of Advance Research, Ideas and Innovations in Technology* (ISSN: 2454-132X), 10.6 (2024). www.IJARIIT.com, Impact Factor: 6.078, Volume 10, Issue 6 - V10I6-1455, pp. 384-393, <https://www.ijariit.com>. <https://www.ijariit.com/manuscripts/v10i6/V10I6-1512.pdf>
20. Dr. M.K. Jayanthi Kannan, Anas Khan (Dec 2024), Big Data Analytics Unveiled Predicting Consumer Behavior through Data-Driven Strategies for Smart, *International Journal of Advance Research, Ideas and Innovations in Technology* (ISSN: 2454-132X), 10.6 (2024). www.IJARIIT.com, Impact Factor: 6.078, Volume 10, Issue 6 - V10I6-1455, pp. 384-393, <https://www.ijariit.com>. <https://www.ijariit.com/manuscripts/v10i6/V10I6-1512.pdf>