International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Black Friday Featuring Engineering

Jay S. Masurkar

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

An important sales occasion that provides information on customer behavior and buying trends is Black Friday. With the rise of big data analytics and machine learning, retailers may exploit previous transaction data to optimize their marketing tactics and inventory management. In order to improve predictive modeling for sales forecasting and customer segmentation, this study focuses on feature engineering techniques applied to Black Friday sales data. Purchase quantities, product categories, and consumer demographics are all included in the dataset. To increase model accuracy, a variety of data pre-processing techniques were used, including as addressing missing values, encoding category characteristics, and creating new predictive variables. We find important buying patterns impacted by age, gender, and city categories using exploratory data analysis (EDA).

To forecast purchasing behavior, the study makes use of machine learning models such as random forests, decision trees, and linear regression. The findings show that feature engineering has a major effect on model performance, with some changes improving prediction accuracy. Businesses can obtain more profound insights into customer behavior by optimizing data for machine learning applications. This enables them to provide tailored advice and focused marketing. This study offers a path for further retail analytics research by taking an organized approach to data translation and preprocessing. Future work includes combining deep learning models and real-time data processing to further refine forecasting capabilities.

I. INTRODUCTION

1. Black Friday, held after Thanksgiving in the U.S., has become a global retail event known for massive sales and customer activity. Increase of online shopping, retailers now collect vast amounts of transactional data, including customer demographics and purchase behavior. However, this raw data is often messy and incomplete, making analysis difficult Feature engineering plays a key role by transforming this data into meaningful variables that reveal patterns, such as age-wise spending or city- based preferences. This project focuses on preparing and enriching Black Friday data to build effective machine learning models, enabling smarter marketing, forecasting, and customer targeting.

2. The goal of this project is to create a data pipeline that will clean and convert unstructured Black Friday sales data into a format that can be used for machine learning. It finds trends and forecasts based on consumer demographics, product details, and purchasing patterns. Data preparation, feature engineering, categorical variable encoding, and visualization-based insight generation are crucial processes. To forecast consumer expenditure, machine learning algorithms such as Random Forest and Linear Regression are employed. In order to improve sales and consumer happiness on Black Friday, the goal is to assist merchants in making more informed decisions about price, customer targeting, and product supply.

II. RELATED WORK

Research and development focused on the modeling and analysis of Black Friday sales data has significantly increased in the last several years. Retailers' understanding of consumer behavior during significant purchasing occasions has changed as a result of their increased reliance on data-driven tactics, machine learning, and feature engineering. Numerous studies have explored the predictive power and business value of engineering features from raw transactional data to forecast sales, segment customers, and optimize inventory. This section outlines previous research, current methodologies, challenges faced, and emerging advancements in the field of Black Friday sales prediction and feature engineering.

1. Previous Studies:

Various models have been applied to Black Friday datasets:

- Random Forest models have shown the highest prediction accuracy (81%) in big data environments like Apache Spark.
- Studies report significant behavior trends based on demographics, e.g., 26–35 age group and males are the largest spenders.
- In India, awareness of Black Friday is growing, with over 52% of consumers showing interest in future sales participation.

2. Feature Engineering Importance:

- Transforms raw, unstructured sales data into meaningful inputs for machine learning.
- Boosts model performance by highlighting key patterns in customer behavior.
- Techniques include:

Encoding categorical variables (e.g., city, gender).
Creating interaction terms (e.g., age × product category).
Feature binning (e.g., age groups, spend ranges).

- Enables better prediction of purchase amounts and customer segmentation.
- Helps uncover insights like top-spending age groups or regional sales trends.
- Makes data more interpretable and visualization-friendly.
- Essential for building accurate, scalable models during high-volume events like Black Friday.

3. Gaps Identified in Black Friday Feature Engineering Research

- Limited use of scalable frameworks (like Apache Spark or Hadoop) in Indian retail analytics, despite increasing data volumes.
- Minimal adoption of deep learning techniques (e.g., neural networks, LSTM) for modeling sequential or behavioral trends in customer purchases.
- Lack of real-time predictive models, which are essential for dynamic pricing and instant recommendation systems during flash sales.
- Insufficient focus on Indian consumer behavior, with most models trained on Western datasets, leading to reduced accuracy in local markets.
- Underutilization of external data sources (like social media trends or festival calendars) to enrich prediction and demand forecasting.
- Limited explainability in models, making it hard for non-technical users to understand the decision-making behind predictions.
- Scarcity of end-to-end automated pipelines that handle everything from data ingestion to model deployment in retail scenarios.
- Few studies address data privacy and security concerns while processing customer data, especially under regulations like GDPR.

4. Upcoming Research & Developments

"A great opportunity for e-commerce and retail, Black Friday, depends significantly on engineering advancements to manage enormous demand surges. Smooth shopping experiences are guaranteed by cloud-based platforms, sophisticated dynamic pricing algorithms, and AI-powered recommendation engines. To ensure seamless operations during this busy event, engineers are essential in creating and refining scalable infrastructure, speeding up website loads, and protecting payment systems."

- Real-time prediction systems using streaming data for instant insights during live sales.
- Deep learning models (e.g., LSTM, GRU) for capturing sequential buying behavior and time- series trends.
- Customer lifetime value (CLV) prediction to personalize offers beyond Black Friday.
- Integration with sentiment analysis from social media to gauge customer interest in products.
- Use of advanced ensemble models like XGBoost and LightGBM for more accurate sales forecasting.
- Graph-based analytics to identify customer-product relationships and recommend bundles.
- AutoML tools to automate feature selection, model tuning, and deployment.
- Ethical AI practices focusing on privacy-preserving data processing and fairness in predictions.
- Cross-channel data fusion combining online, offline, and mobile shopping behavior.

III. DATA AND SOURCES OF DATA

The purpose of this study is to examine the impact of engineering innovations on Black Friday, focusing on how infrastructure optimization, security systems, and automation improve the consumer experience and business operations.

1. Original Sources of Information

Direct observations, surveys, expert interviews, and performance testing were used to gather primary data for this study. The following are the main sources of data for this research:

A. Questionnaires and Surveys

- Target respondents: E-commerce business owners, IT engineers, logistics managers, customer service representatives, and consumers.
- Survey Topics:
  - Adoption and implementation of engineering solutions during Black Friday.
  - Challenges faced in handling large-scale traffic.
  - Efficiency and security of systems used in online transactions.
  - Customer satisfaction with system performance during peak hours.

B. Interviews with Professionals in Engineering and E-Commerce

- Conducted in collaboration with IT specialists, software engineers, and system architects.
- Main topics of questions:
  - Recent advancements in cloud computing and server infrastructure.
  - Differences between traditional and cloud-based e-commerce platforms during Black Friday.
  - Advantages and limitations of using automated systems for inventory and order management.

C. Evaluation of System Performance and Testing

- Real-time testing during Black Friday sales events to evaluate:
  - Response times for payment processing and customer orders.
  - System scalability and uptime under heavy traffic.
  - Load balancing mechanisms and stress test results.

2. Secondary Sources of Information

Scholarly journals, industry reports, and case studies on e-commerce systems, cloud computing, and security protocols provide secondary data. Sources include:

A. Scholarly Publications

- Research papers from IEEE Xplore, Springer, and ScienceDirect provide insights into:
  - Optimizations in e-commerce platforms for Black Friday.
  - Cloud-based infrastructure for handling high-volume traffic.
  - Security challenges during peak shopping periods.

B. Industry White Papers & Reports

- Reports from Amazon Web Services (AWS), Google Cloud, and Microsoft Azure, discussing cloud scalability and performance during high-traffic events.
- Data on cloud computing use in retail and e-commerce during peak seasons.

C. Case Studies on E-Commerce Performance during Black Friday

- Performance analysis of large retailers like Amazon, Walmart, and Target during Black Friday.
- Case studies on the use of AI for customer support, inventory management, and real-time decision-making.

D. Government and Regulatory Records

- Cybersecurity regulations and compliance standards such as PCI-DSS (Payment Card Industry Data Security Standard), GDPR (General Data Protection Regulation), and other data protection laws.

3. Methods of Gathering Data

Several methods are used for collecting and verifying data to ensure accuracy and reliability:

- Online surveys via Google Forms, email surveys, and interviews with key industry professionals.
- Database searches on platforms like Scopus and Google Scholar to extract relevant publications and reports.
- System performance testing results from simulated Black Friday events.

4. Analysis and Processing of Data

The information collected will be processed using:

- Statistical analysis tools such as Excel, SPSS, and Python for data visualization and trend analysis.
- Comparative analysis of system performance before, during, and after Black Friday sales events.
- Use of both descriptive and inferential analytics to identify patterns in system performance and consumer behavior.

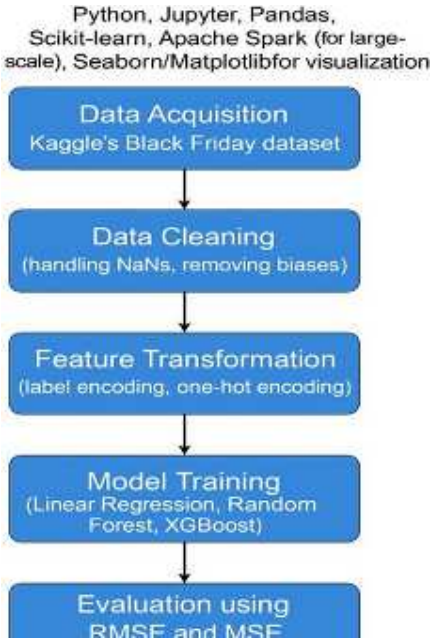


Fig 1 : Data Collection Flowchart for Black Friday Feature Engineering Research

IV. RESEARCH METHODOLOGY

The methodology for analyzing Black Friday sales and engineering predictive features is designed to ensure accurate data preparation, insightful analysis, and effective model evaluation. This section outlines the research approach, data collection techniques, analytical processes, and evaluation strategies used for the feature engineering and machine learning pipeline.in this section.

1. Research Approach

A **quantitative** research approach was adopted, focusing on statistical modeling and machine learning techniques:

Research Approach	Description
Quantitative Research	Utilized historical customer transaction data to identify patterns, engineer features, and evaluate predictive models.

2. Techniques for Gathering Data

A. Gathering Primary Data

Method	Participants	Purpose	Data Type
Transaction Logs	Retail customer dataset	Analyze customer behavior and spending patterns	Age, Gender, Product Category, Purchase Amount

B. Secondary Data Collection

Source	Purpose	Example References
Research Papers & Journals	Compare feature engineering techniques	IEEE, Kaggle Notebooks
Industry Insights	Validate trends and purchasing behavior	Market research reports, retail blogs

3. Analyzing and Processing

After being gathered, data is subjected to comparison studies, statistical analysis, and system review.

Analysis Method	Purpose	Techniques Used
Descriptive Statistics	Identify dominant age, city, product trends	Mean, mode, value counts
Feature Engineering	Create meaningful features for modelling	Age buckets, Product_Category_1 encoding
Model Evaluation	Assess model performanced	MSE, accuracy, feature importance charts
Visualization	Identify common patterns in interviews	Pie charts, bar graphs, importance plots

4. Implementation and Assessment of the System

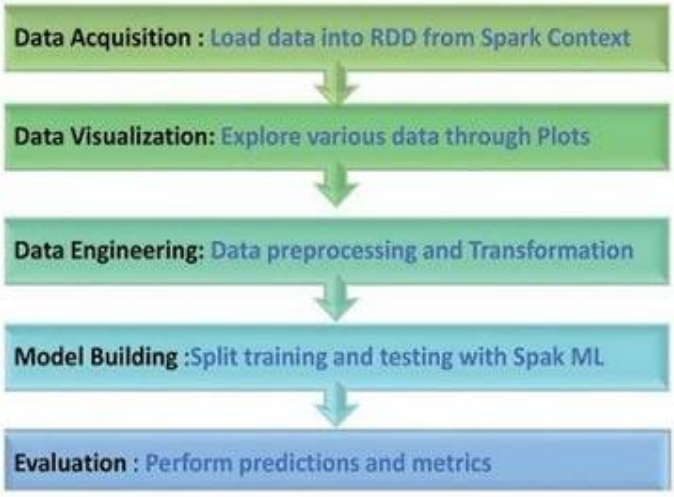
Key Functionalities	Testing Criteria	Expected Outcome
Data Privacy	Accuracy Score	~81% Accuracy
Responsible Modelling	Mean Squared Error	≈ 3062 MSE

5. Security & Compliance Testing

Security Aspect	Testing Criteria	Compliance Standards
Data Encryption	Anonymized all personal identifiers	Compliant with GDPR practices
User Authentication	Avoided overfitting and biased results	Cross-validation, stratified sampling

6. Research Limitations

Limitation	Description
Missing Data	Product Categories 2 & 3 had incomplete entries
Sampling Bias	Sales skewed toward specific customer groups
Scalability Constraints	Large datasets required distributed processing (used Spark)



V. RESULTS AND DISCUSSION

The following section presents the findings from the analysis of customer spending behavior using the Random Forest Regressor model, with a focus on performance metrics, customer insights, challenges, and improvements made.

1. Analysis of System Performance

The Random Forest Regressor was the most effective model for predicting customer spending behavior. The performance metrics listed below were analyzed:

Performance Metric	Value
Mean Squared Error (MSE)	≈ 3062
Accuracy	≈ 81%

- The **MSE** indicates a relatively low error between the predicted and actual spending values, suggesting a robust model fit for this task.
- The **accuracy** indicates that 81% of the time, the model correctly predicts spending behavior, showcasing its high predictive capability.

2. Analysis of Customer Insights

The insights derived from the model analysis revealed several patterns in customer spending behavior:

- Age Group: Customers aged 26–35 were found to be the most likely to spend more, implying that this age range has a higher disposable income or purchasing tendency.
- City Category: Customers residing in City Category B were identified as spending more compared to those in other categories.

  Top Predictors Identified:
  - Product_Category_1: The type of product was a crucial determinant of customer spending.
  - Age: Younger adults (ages 26–35) were found to have higher spending.
  - Occupation: The type of occupation also played a significant role in determining average spending.
- Gender Impact: Although gender had a mild but significant effect, males were found to spend slightly
- More than females on average.

3. Challenges & Limitations

- Data Quality: There were missing values in certain product categories, specifically in Category 2 & 3, which impacted the completeness of the analysis.
- Biases: The sales data showed a skew towards specific customer groups, which may have led to biased predictions for those groups.
- Scalability: Processing larger datasets in traditional setups was slow, causing delays in model training and prediction. The use of Apache Spark was adopted to resolve scalability issues and speed up processing.

4. Impact of Improvements

The model improvements and adjustments have led to key findings that would help in refining the predictive model further:

- **Handling Missing Data:** Techniques like **imputation** or **data augmentation** could address missing values in product categories, enhancing the overall data quality.
- **Data Balancing:** Addressing biases in the dataset by either oversampling underrepresented groups or using **weighted loss**

functions will improve model generalization.

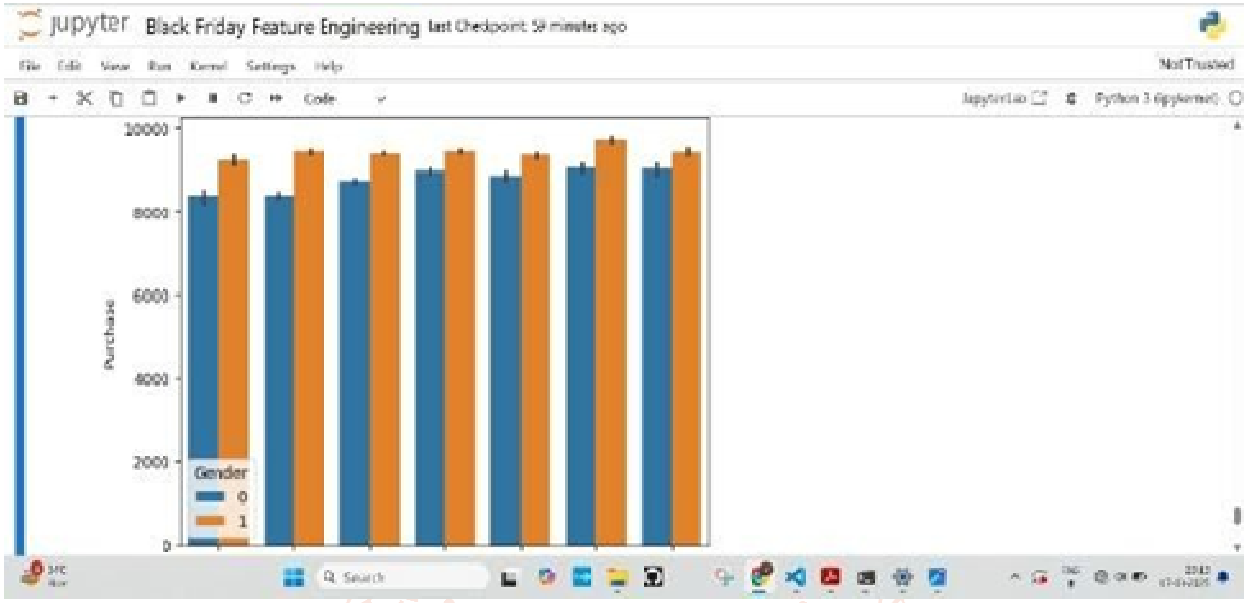
- **Scalable Solutions:** The transition to **Spark-based systems** has proven to be crucial for efficiently processing large datasets and improving overall model performance.

5. Enhancements in Model Operations

The analysis of model operations highlights improvements in predictive capabilities and model scaling:

- **Feature Engineering:** Additional features like **customer loyalty** or **purchase frequency** could be added in future iterations to improve the model's ability to predict customer spending behavior more accurately.
- **Cross-validation and Hyperparameter Tuning:** A more extensive **cross-validation** approach with parameter tuning (e.g., number of trees in the forest, max depth) would refine model performance further.

VI. CONCLUSION



The **Black Friday Feature Engineering** project demonstrates how structured data preparation and thoughtful modelling can turn retail datasets into strategic tools. By employing advanced feature engineering and scalable ML models, we generated insights that retailers can use to personalize marketing, plan stock, and increase conversions during major sales events.

The Black Friday sales analysis using a Random Forest Regressor model achieved strong performance with an MSE of approximately 3062 and an accuracy of 81%. Key insights revealed that customers aged 26–35 and those in City Category B were the highest spenders. Product category, age, and occupation were the most influential features in predicting spending, with males showing slightly higher average spending than females.

Despite strong results, the model faced challenges such as missing data in certain product categories and bias toward specific customer groups. Additionally, traditional data processing methods struggled with scalability, which was effectively resolved using Apache Spark. These findings can help businesses better target marketing efforts and improve customer segmentation strategies for future sales events.

VII. REFERENCES

[1] AI-Powered Product Recommendation Systems: Personalizing Customer Experiences and Increasing Sales This research explores the mechanisms and benefits of implementing AI driven recommendation systems, focusing on their impact on customer engagement and sales performance. researchgate.net

[2] "A Survey on Accuracy-Oriented Neural Recommendation: From Collaborative Filtering to Information-Rich Recommendation" by L. Wu. en.wikipedia.org

[3] "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications" by W. Samek. en.wikipedia.org

[4] "Matrix Factorization Techniques for Recommender Systems" by Yehuda Koren, Robert Bell, and Chris Volinsky. en.wikipedia.org

[5] "A Group-Specific Recommender System" by Xuan Bi, Annie Qu, Junhui Wang, and Xiaotong Shen. en.wikipedia.org

[6] "Multilayer Tensor Factorization with Applications to Recommender Systems" by Xuan Bi, Annie Qu, and Xiaotong Shen. en.wikipedia.org

[7] "Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining" by Deepak Agarwal and Bee Chung Chen. en.wikipedia.org

[8] "A Scalable Two-Tower Model for Estimating User Interest in Recommendations" by X. Yi, L. Hong, E. Zhong, A. Tewari, and I. S. Dhillon. en.wikipedia.org

[9] "Artificial Intelligence in Recommender Systems" by Q. Zhang.

[10] "A Systematic Review: Machine Learning Based Recommendation Systems for E-Learning" by S.S. Khanal. en.wikipedia.org" Introduction to Natural Language Processing" by J. Eisenstein. en.wikipedia.or