Predicting Online Shopper Behavior: Machine Learning Approaches for Enhanced E-Commerce Insights

[1]Dannala Appaji Sesha Sai Kumar, [2]Dr. M.Veera Bhadrarao, [3]P.Sujana Priya Dharshinni, [4]K.V.V.Ramana [5]M.Jyothi

[1]P.G Student, [2]Professor, [3,4,5]Assistant Professor

[1,2,3,4,5]Department of Computer Science,

[1,2,3,4,5]Pydah College of Engineering

Abstract—The ability to predict online shopper behavior is essential for e-commerce platforms striving to enhance customer satisfaction and optimize operational efficiency. This paper presents a machine learning-based approach to forecasting user preferences, engagement patterns, and purchasing likelihood. Leveraging the UCI Online Shoppers Purchasing Intention Dataset, we evaluate various algorithms, including Random Forest, Logistic Regression, and Naive Bayes, to determine the most effective predictor of shopper behavior. Exploratory Data Analysis (EDA) highlights key features influencing customer decisions, such as session duration and bounce rates. Our findings underscore the potential of predictive analytics in driving personalized marketing, improving inventory management, and increasing conversion rates. The Random Forest algorithm outperforms other models in accuracy and generalization, demonstrating the transformative role of machine learning in the digital retail landscape.

Index Terms—Online Shopper Behavior, Machine Learning, Predictive Analytics, Random Forest, Logistic Regression, Naive Bayes, E-commerce, UCI Online Shoppers Dataset, Customer Engagement, Session Duration, Bounce Rate, Personalization, Conversion Rate, Inventory Management, Marketing Optimization, Exploratory Data Analysis (EDA), Behavior Prediction.

I. Introduction

E-commerce has revolutionized the retail industry, providing consumers with unparalleled convenience and access to a wide range of products and services. With the rapid growth of online shopping, e-commerce platforms are now essential for businesses to stay competitive. These platforms are not only offering products but also creating sophisticated shopping experiences through personalized recommendations, tailored advertisements, and dynamic pricing strategies. However, as the e-commerce landscape evolves, understanding and predicting consumer behaviour has become a complex and critical challenge.

Online shoppers exhibit diverse preferences and purchasing habits that vary widely across demographics, product categories, and even individual purchasing cycles. This variability makes it difficult to predict future actions, such as whether a shopper will complete a purchase, abandon their cart, or engage with a particular product. Additionally, the vast amounts of data generated by online interactions – including clicks, page views, session times, and product searches – require advanced methods to analyse and extract meaningful insights. This wealth of information, while valuable, introduces complexities in managing and processing large datasets, especially when it comes to real-time analysis.

Another challenge lies in the dynamic nature of online shopping behaviour. Market trends, seasonal events, promotions, and even external factors like economic shifts or social influences can all impact a shopper's decision-making process. For example, a customer’s likelihood to purchase may vary depending on whether they are shopping during a major sales event, like Black Friday, or browsing casually on a weekday. These fluctuations further complicate the task of building reliable predictive models.

In order to overcome these challenges, e-commerce businesses need to leverage advanced technologies, such as machine learning and predictive analytics. These technologies allow companies to better understand and anticipate consumer needs, optimize marketing efforts, and ultimately enhance the shopping experience, driving increased engagement and sales.

The significance of accurately predicting online shopper behaviour extends far beyond simply understanding consumer preferences; it plays a crucial role in shaping the strategies of e-commerce businesses. As online platforms become increasingly competitive, the ability to anticipate customer actions empowers businesses to make data-driven decisions that drive growth, efficiency, and customer satisfaction.

One of the most critical applications of shopper behaviour prediction is **targeted marketing**. By identifying patterns in consumer behaviour, businesses can tailor their marketing strategies to specific customer segments, enhancing engagement and increasing the likelihood of conversions. For example, predicting which customers are likely to purchase certain products allows companies to deliver personalized advertisements, promotions, and product recommendations at the right time, leading to more effective marketing campaigns. Personalized offers can increase a customer's sense of value and relevance, fostering stronger brand loyalty and higher lifetime customer value.

Another significant advantage lies in **inventory management**. Predicting purchasing trends allows businesses to optimize their stock levels and reduce the risk of overstocking or stockouts. For instance, if predictive models forecast a surge in demand for a specific product, businesses can adjust their inventory strategies accordingly, ensuring they have the right products available at the right time. Conversely, by predicting items that are unlikely to sell, businesses can reduce unnecessary stock, thus improving their cash flow and minimizing storage costs.

In addition, by identifying patterns in **cart abandonment**, e-commerce platforms can take proactive steps to reduce revenue loss. Understanding the triggers behind cart abandonment – such as shipping costs, website friction, or long checkout processes – enables businesses to implement strategies that mitigate these issues, increasing conversion rates. Similarly, by forecasting purchase trends based on customer behaviour, businesses can fine-tune their **supply chain management**, ensuring that products are stocked in alignment with future demand and minimizing the costs associated with excess inventory.

Ultimately, predicting online shopper behaviour allows businesses to not only improve operational efficiency but also create more meaningful and satisfying customer experiences, leading to greater customer retention and long-term profitability.

This study is focused on leveraging machine learning to effectively predict online shopper behavior by analyzing various types of data, including **demographic**, **transactional**, and **behavioral** information. Understanding consumer behavior in the digital space is a complex task due to the sheer volume and diversity of data generated by online shoppers. However, by using advanced machine learning algorithms, this study aims to provide deeper insights into customer decision-making processes, ultimately helping e-commerce businesses enhance their strategies.

Demographic data such as age, gender, location, and income play a critical role in shaping a consumer's preferences and purchasing patterns. For example, younger shoppers may be more inclined to purchase trendy products, while older demographics may prefer more traditional items. Location can also affect purchase decisions due to cultural preferences, regional product availability, or even local promotions. By incorporating demographic factors, machine learning models can identify trends and preferences specific to different consumer segments, allowing businesses to target marketing efforts more effectively.

Transactional data provides insights into a shopper's previous purchases, frequency of transactions, and average order value. This type of data is especially useful for predicting future purchasing behavior. For instance, if a customer has consistently purchased a certain product category, the model can predict their future likelihood of purchasing similar items. By analyzing transactional data, businesses can also identify high-value customers and tailor loyalty programs to retain them.

Finally, **behavioral data** such as browsing patterns, session duration, clickstream data, and cart abandonment rates are essential in predicting purchase intent. Shoppers who spend more time on a product page or visit a site multiple times may be more likely to make a purchase. By analyzing this data, machine learning models can assess the likelihood of a visitor converting into a buyer or abandoning their cart, providing businesses with opportunities to intervene and optimize the shopping experience.

By integrating all these data types, this study aims to provide **actionable insights** that will guide e-commerce decision-makers in refining their marketing strategies, inventory management, and overall customer experience, ultimately boosting sales and customer retention.

II. Literature Review

Sakar et al. (2018) - Dynamic Pricing and Customer Segmentation

Sakar et al. (2018) conducted a study focusing on dynamic pricing and customer segmentation using **session data**. This research utilized data generated during users' browsing sessions to predict their behaviour and segment customers accordingly. **Session data** includes valuable interaction information such as the number of pages visited, time spent on each page, and the sequence of actions performed by a visitor on an e-commerce website.

- **Dynamic Pricing**: Dynamic pricing involves adjusting the price of products in real-time based on factors like demand, competitor pricing, and customer behaviour. Sakar et al. used session data to predict when a shopper is likely to purchase and adjusted prices accordingly. This approach helps businesses maximize their revenue by offering personalized prices that appeal to the shopper's behaviour and intent.
- **Customer Segmentation**: The study also explored customer segmentation by identifying groups of shoppers with similar behaviours. Using machine learning techniques, such as clustering and classification, Sakar et al. were able to identify patterns in the data that allowed them to categorize customers into meaningful segments. These segments could then be targeted with specific marketing strategies, enhancing customer engagement and retention.

The predictive models built in this study achieved **significant predictive accuracy**, demonstrating the potential of using session data for more granular customer insights and personalized strategies.

Machine Learning Methods: Random Forest and Logistic Regression

Building on Sakar et al.'s work, several machine learning algorithms, such as **Random Forest** and **Logistic Regression**, have been widely adopted in similar domains for **classification tasks**. These models are designed to classify data into distinct categories, making them well-suited for predicting shopper behaviour, such as whether a shopper will make a purchase or abandon their cart.

Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting. It's particularly useful when dealing with large datasets with many features, making it ideal for e-commerce scenarios where there are complex relationships between various shopper attributes (e.g., demographic details, and session behaviour). Random Forest has been successful in predicting various outcomes in e-commerce, such as conversion rates, purchase likelihood, and customer segmentation.

- **Logistic Regression**: Logistic Regression is a simpler yet effective classification algorithm, especially for binary classification tasks, such as predicting whether a customer will buy a product (yes/no). In e-commerce, it has been widely used to model the relationship between independent variables (e.g., session duration, product views) and the probability of an event occurring (e.g., making a purchase). While less complex than Random Forest, Logistic Regression is fast, interpretable, and works well for linear relationships.

Ensemble Methods and Feature Engineering

Recent studies have highlighted the **potential of ensemble methods** and **feature engineering** in improving prediction reliability and accuracy.

- **Ensemble Methods**: Ensemble methods combine multiple machine learning models to improve prediction performance. By aggregating the predictions of several individual models (like decision trees in Random Forest), ensemble methods reduce variance and bias, leading to more robust and accurate results. In e-commerce, these methods are particularly effective in handling the vast variability of customer behavior and can be used to create more reliable predictions regarding purchases, churn, or engagement.
- **Feature Engineering**: Feature engineering involves creating new input features or transforming existing ones to improve the performance of machine learning models. In e-commerce, this can include combining session data with demographic details, product interaction metrics, and time-based features (e.g., time of day, seasonality). Feature engineering allows models to capture more complex patterns in shopper behavior, leading to better predictive accuracy. For example, creating new features that capture customer engagement levels (e.g., time spent per page or number of interactions) can significantly enhance a model's ability to predict whether a shopper will make a purchase.

Integration of Advanced Preprocessing, EDA, and Real-Time Deployability

This study builds upon the existing research by integrating **advanced preprocessing techniques**, **comprehensive Exploratory Data Analysis (EDA)**, and **real-time deployability** into a single framework for predicting online shopper behaviour.

- **Advanced Preprocessing**: Data preprocessing is a critical step in machine learning, as raw data often contains noise, missing values, or irrelevant features. This study employs advanced techniques like **handling missing data**, **scaling numerical features**, and **encoding categorical variables** to prepare the data for modelling. These preprocessing steps ensure that the models receive clean, relevant data, leading to better performance.
- **Exploratory Data Analysis (EDA)**: EDA helps understand the underlying patterns in the data. By visualizing data distributions, detecting outliers, and assessing feature relationships, EDA provides valuable insights that inform feature engineering and model selection. This step is crucial in identifying key predictors of online shopper behaviour, such as session duration, bounce rates, and product views.
- **Real-Time Deployability**: Unlike some studies that focus on offline analysis, this work emphasizes the importance of deploying predictive models in **real time**. By integrating the trained models into e-commerce platforms, businesses can predict shopper behaviour as users interact with the website, providing opportunities for real-time interventions. For example, businesses can trigger personalized recommendations or promotions based on the likelihood of purchase, improving user engagement and conversion rates.

By combining these advanced techniques, the study provides a comprehensive and actionable framework for predicting online shopper behaviour, which can be deployed in real-world e-commerce platforms to optimize marketing, inventory, and customer engagement strategies.

III. Methodology

The methodology of this study is centred around the application of machine learning techniques to predict online shopper behavior effectively. It involves the systematic collection, preprocessing, and analysis of data, followed by model selection and evaluation to achieve the best predictive performance.

3.1 Data Collection

The dataset used in this study is sourced from the **UCI Online Shoppers Purchasing Intention Dataset**, which is widely recognized for its comprehensive attributes relevant to e-commerce behavior analysis. The dataset consists of **numerical features**, such as Session Duration (time spent by a user on the website) and Page Value (an indicator of the financial value of visited pages before purchase). Additionally, it includes **categorical features**, such as Visitor Type (returning or new visitor) and Weekend Flag (indicating if the session occurred on a weekend). These attributes provide a diverse range of information, enabling a holistic analysis of shopper behavior.

3.2 Data Preprocessing

Effective preprocessing is crucial for ensuring high-quality inputs to the machine learning models:

- **Handling Missing Values**: Missing data in numerical features was addressed using **median imputation**, a robust method that minimizes the impact of outliers.
- **Encoding Categorical Variables**: Features like Visitor Type were transformed into **binary flags** through one-hot encoding, making them suitable for machine learning algorithms.
- **Normalization**: Numerical features were scaled to ensure uniform value ranges, preventing dominance by features with larger scales and enhancing model performance.

3.3 Exploratory Data Analysis (EDA)

EDA revealed significant insights into shopper behavior:

- **Returning visitors** were three times more likely to complete a purchase compared to new visitors, highlighting the importance of customer retention.
- **Session Duration** was positively correlated with purchase likelihood, indicating that engaged users were more likely to convert.
- **Bounce Rates and Exit Rates** emerged as critical predictors of disengagement, providing actionable insights for improving website design and user experience.

3.4 Model Selection

To identify the most effective predictor of online shopper behavior, three models were evaluated:

1. **Random Forest**: This ensemble method builds multiple decision trees and aggregates their predictions. It is particularly effective for handling **imbalanced datasets** and capturing **non-linear relationships** in the data.
2. **Logistic Regression**: Known for its efficiency and simplicity, Logistic Regression is ideal for **binary classification tasks**. However, it is less adept at capturing complex patterns in high-dimensional data.

3. **Naive Bayes**: This probabilistic classifier assumes independence among features. While it is computationally efficient, its performance diminishes when feature interdependencies are significant.

The insights gained from EDA and the strengths of the selected models enable a comprehensive approach to predicting shopper behaviour. The combination of these methodologies provides actionable insights, enhancing both predictive accuracy and practical applicability.

IV. EXPERIMENTAL SETUP

The experimental setup outlines the process for training, testing, and optimizing machine learning models to ensure their effectiveness and robustness in predicting online shopper behavior. This section describes the data split strategy, the use of cross-validation, and the tuning of hyperparameters for each model.

4.1 Training and Testing Split

The dataset was divided into two subsets: 80% for training the models and 20% for testing their performance. This split ensures that the models are trained on a substantial portion of the data while reserving an independent set for evaluation, thus avoiding overfitting.

To further enhance the reliability of the results, a 5-fold cross-validation technique was employed. Cross-validation involves partitioning the training data into five subsets (or folds), training the model on four folds, and validating it on the fifth. This process is repeated five times, with each fold serving as the validation set once. The final performance metric is the average across all folds. Cross-validation provides the following benefits:

Reduces the risk of overfitting by exposing the model to diverse training and validation subsets.

Ensures that the evaluation metric reflects the model's generalization ability across different data splits.

4.2 Hyperparameter Tuning

Hyperparameters are configuration settings that influence how a machine learning algorithm learns and generalizes. Each model requires careful tuning to maximize its performance.

Random Forest

Random Forest, being an ensemble learning algorithm, has several critical hyperparameters:

Number of Trees: Determines how many decision trees are built. More trees often improve accuracy but increase computation time. The optimal number was identified using grid search, a systematic approach to testing various values.

Maximum Depth: Specifies the maximum number of splits in each tree. Deeper trees capture more complexity but risk overfitting. By testing depths ranging from shallow to deep, an optimal balance between bias and variance was achieved.

Logistic Regression

Logistic Regression depends heavily on its regularization parameter:

Regularization (C): Controls the penalty for large coefficients to prevent overfitting. A smaller value of C increases regularization strength, leading to simpler models, while a larger value reduces regularization, capturing more complexity.

Through grid search, different values of C were tested, balancing the bias-variance trade-off. This ensured the model was neither too simplistic nor overly complex.

Naive Bayes

Naive Bayes assumes feature independence, which simplifies the model but can limit its flexibility:

Unlike Random Forest and Logistic Regression, Naive Bayes has fewer hyperparameters to tune. Its primary configuration lies in choosing the type of distribution for the input data (e.g., Gaussian, Multinomial, or Bernoulli). Based on the data characteristics, the appropriate variant was selected.

Since Naive Bayes inherently assumes independence among features, further tuning was minimal, making it computationally efficient.

4.3 Benefits of This Experimental Setup

Robustness: The combination of an 80-20 data split and 5-fold cross-validation ensures models generalize well to unseen data.

Efficiency: Hyperparameter tuning using grid search systematically explores parameter combinations to optimize model performance without overfitting.

Flexibility: Different optimization strategies are applied to each algorithm, leveraging their unique strengths while addressing limitations.

This setup ensures the chosen models are not only accurate but also resilient to variations in input data, making them well-suited for real-world e-commerce applications.

V. Results and Discussion

This section presents the performance of the evaluated models, highlights key insights gained from the analysis, and discusses the implications of these findings for e-commerce businesses.

5.1 Model Performance

The performance of the machine learning models was evaluated using three key metrics: Accuracy, F1-Score, and ROC-AUC (Receiver Operating Characteristic - Area Under Curve). These metrics provide a comprehensive understanding of each model's predictive power.

Model	Accuracy	F1-Score	ROC-AUC
Random Forest	94%	0.91	0.94
Logistic Regression	87%	0.84	0.88
Naive Bayes	81%	0.79	0.83

Random Forest outperformed the other models across all metrics. Its accuracy of 94% and F1-score of 0.91 indicate a strong ability to correctly classify shopper behaviour while balancing precision and recall. Its ROC-AUC score of 0.94 reflects excellent discrimination between classes (e.g., purchase vs. no purchase).

Strengths: Random Forest's ensemble approach allows it to capture non-linear relationships and handle feature interactions effectively. It is also robust against outliers, which are common in shopper data.

Logistic Regression, while simpler, achieved 87% accuracy and a 0.88 ROC-AUC, showing reasonable predictive power. However, it struggled with capturing complex patterns in the data, leading to lower performance compared to Random Forest.

Naive Bayes achieved the lowest performance, with an accuracy of 81% and an F1-score of 0.79. Its assumption of feature independence limited its ability to model complex dependencies, making it less suitable for this dataset.

5.2 Insights

The analysis revealed several key insights that can inform e-commerce strategies:

Bounce Rates and Cart Abandonment:

Bounce rates (percentage of visitors leaving the site after viewing one page) emerged as a significant factor influencing cart abandonment. High bounce rates often indicate dissatisfaction with the website, unclear navigation, or lack of engaging content. Addressing these issues through better website design and user experience improvements can reduce cart abandonment and increase conversions.

Session Duration and Page Values:

Shoppers who spent longer session durations on the website or engaged with high-value pages were more likely to complete a purchase. These metrics indicate strong interest and intent to buy, suggesting that businesses should focus on strategies that keep visitors engaged, such as dynamic content, detailed product descriptions, and live chat support.

Seasonal Trends:

The analysis uncovered spikes in traffic during holiday seasons, highlighting the importance of contextual marketing. Special events like Black Friday or Christmas generate increased interest, making them ideal times for targeted promotions, discounts, and product launches.

5.3 Implications

The findings of this study have significant implications for e-commerce platforms:

Personalized Recommendations:

By leveraging insights such as session duration and page engagement, businesses can tailor product recommendations to match shopper preferences. For example, returning visitors with high engagement in specific product categories can be targeted with personalized emails or promotions.

Optimized Advertisements:

Bounce rate analysis can help refine marketing campaigns by identifying underperforming landing pages. Adjustments to ad copy, visuals, and call-to-action elements can significantly improve engagement and conversion rates.

Seasonal Campaign Planning:

Seasonal traffic spikes emphasize the need for well-planned marketing strategies during peak shopping periods. Predictive insights can guide inventory stocking, promotional schedules, and dynamic pricing to maximize sales during these high-demand periods.

VI. Conclusion

This study robustly validates the efficacy of machine learning techniques in predicting online shopper behavior, a critical capability for e-commerce businesses seeking to enhance customer experiences and operational performance. Through a comprehensive analysis of various algorithms, the Random Forest model demonstrated outstanding performance metrics, making it the most effective algorithm for this particular application.

Effectiveness of Machine Learning in Predictive Analytics

Machine learning, particularly through algorithms like Random Forest, offers significant advantages in handling complex datasets inherent to e-commerce. Unlike traditional statistical methods, machine learning models can automatically detect patterns and relationships within large volumes of data, allowing businesses to make data-driven decisions quickly. This predictive capability is essential in the fast-paced e-commerce environment, where consumer preferences and market conditions can change rapidly.

Random Forest Model

The Random Forest algorithm stood out due to its high accuracy (94%) and robustness in predicting shopper behavior. It excels at managing feature interactions and non-linear relationships, which are common in consumer behavior data. Furthermore, its ability to mitigate overfitting—by averaging the results of numerous decision trees—ensures reliable performance even in diverse scenarios. The model's interpretability also allows stakeholders to understand which features significantly impact predictions, enhancing transparency in decision-making processes.

Integration of Predictive Insights

By integrating these predictive insights into their operations, e-commerce businesses can implement several strategic enhancements:

Personalized Marketing Strategies: Leveraging the insights gained from predicting online shopper behavior enables businesses to tailor their marketing strategies. For example, identifying customer segments that are likely to convert can lead to targeted promotions and personalized recommendations, significantly improving conversion rates.

Inventory Management Optimization: Understanding purchasing patterns allows businesses to optimize inventory levels, reducing instances of stockouts or overstock situations. By aligning inventory with predicted demand, companies can enhance operational efficiency and reduce holding costs.

Improving Customer Satisfaction: By utilizing predictive analytics, businesses can enhance the overall shopping experience. For instance, addressing factors that lead to cart abandonment—such as optimizing website navigation and reducing load times—can significantly improve customer satisfaction and loyalty. Providing personalized experiences based on shopping behavior insights fosters a sense of value and connection with the brand.

Dynamic Pricing Strategies: Predictive models can assist in developing dynamic pricing strategies that respond to real-time changes in consumer demand and competition. By analyzing shopper behavior and market trends, businesses can adjust prices to maximize revenue without alienating customers.

Future Implications

The successful implementation of machine learning in predicting online shopper behavior opens the door to further advancements in e-commerce analytics. Future research can explore the integration of additional data sources, such as social media interactions and customer sentiment analysis, to refine predictions. Furthermore, the potential of deep learning techniques can be investigated to enhance predictive accuracy, particularly in analyzing sequential data from user interactions.

In conclusion, the study underscores the transformative role of machine learning, particularly the Random Forest algorithm, in driving improvements in customer engagement, operational efficiency, and overall business performance in the competitive landscape of e-commerce. By embracing these technologies, businesses can position themselves for sustained growth and success in an increasingly data-driven world.

VII. Conclusion

This study has demonstrated the significant potential of machine learning in accurately predicting online shopper behavior, a vital capability for e-commerce platforms striving to stay competitive. The research aimed to identify and implement the most effective predictive models using a robust methodology, ultimately validating the efficacy of these techniques in real-world applications.

References

[1] T.Yoshida,M.Hasegawa,T.Gotoh,H.Iguchi,K.Sugiokaand K. Ikeda,, "Consumer behavior modeling based on social psychology and complex networks," The 9th IEEE International," ConferenceECommerce Technology , no. 6th edition, 2020.

[2] H. Valecha, "Prediction of Consumer Behaviour using Random Forest Algorithm," 2018[Online]. Available: https://ieeexplore.ieee.org/abstract/document/8597070

[3] A.H.Kumar, F.S.John, & S. Senith, "A Study on factors influencing consumer buying behavior in cosmetic Products," International journal of Scientific and research, vol. 4, no. 9, pp.pp. 1-6, 2015.

[4] H. T. Cheng, L. Koc, J. Harmsen, et al, "Wide and deep learning for recommender systems,"2016. [Online]. Available: ACM Workshop on Deep Learning for Recommender Syst., Boston, MA,USA,

[5] S. K. Singh, "A Machine Learning Model for Effective Consumer Behaviors Prediction,"2021. [Online].Available:https://ieeexplore.ieee.org/document/9702495

[6] Kwanruan Rusmee, Narumol Chumuang, "Predicting System for the Behavior of Consumer Buying Personal Car Decision by Using SMO," vol. 2, no. 3, pp. 10-15, 2021

[7] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for YouTube recommendations,"

[8] ACM Conf. Recommender Syst., Boston, MA, USA,, pp. pp. 191-198, 2016

[9] X. Zhai, "Prediction Model of User Purchase Behavior Based on Machine Learning," 2020.[Online]. Available: https://ieeexplore.ieee.org/document/9233677

[10] K. Surendro, "Predictive Analytics for Predicting Customer Behavior," 2019. [Online].

[11] Available: https://ieeexplore.ieee.org/document/8834571

[12] V. Martin, "Predicting the french stock market using social media analysis,"

[13] International workshop on sentimentic, pp. pp/3-7, 2020

[14] N. Kapoor, "A Review on Consumer Behavior Prediction using Data Mining Techniques," 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8991415.

[15] K. Yada, "Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques," 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7941935/similar#similar

[16] S.R. Sujithra; R.K. Kavitha, "Consumer Online Buying Behaviour - A Perspective Study Analysis using R," 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9675566

[17] Muqaddas Gull; Arshi Pervaiz, "Customer Behavior Analysis Towards Online Shopping using Data Mining," 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8467262.

[18] M. Kosinski, S. C. Matz, S. D. Gosling, "Facebook as a research tool for the social sciences: Opportunities, challenges