Use of Machine Learning on Predicting Online Shoppers' Purchasing Intentions

Ahmet Sefa Güren
Department of Civil Engineering
Middle East Technical University
Ankara, Türkiye
sefa.guren@metu.edu.tr

Erkan Yetüt
Department of Mathematics
Middle East Technical University
Ankara, Türkiye
e249984@metu.edu.tr

***Abstract*—E-commerce has transformed the retail landscape by providing unparalleled convenience and accessibility. However, low conversion rates remain a persistent challenge, driven by complex customer behaviors that are difficult to interpret in digital environments. Predicting online shoppers' purchase intentions is critical for addressing these challenges and improving conversion rates. This study focuses on the widely used dataset introduced by Sakar et al., which reflects the imbalanced nature of purchase intention data. To address this imbalance, we employed the Synthetic Minority Oversampling Technique (SMOTE) to enhance model training. Various machine learning models, including Logistic Regression, Random Forest, Stacking, and XGBoost, were compared using metrics such as F1-Score, Matthews Correlation Coefficient (MCC), auROC, and auPR.**

Our results demonstrate that XGBoost outperforms other models, achieving an F1-Score of 0.8956, MCC of 0.5988, and auROC of 0.9283. The combination of feature selection and SMOTE significantly improved classification performance, underscoring the critical role of preprocessing in e-commerce prediction tasks. By leveraging ensemble methods and advanced evaluation metrics, this study highlights the potential of machine learning in building robust predictive models to analyze customer behavior and improve e-commerce platform performance.

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

E-commerce has rapidly transformed the retail landscape, offering convenience, accessibility, and an unparalleled range of options for customers worldwide. Despite its widespread adoption and growth, one of the persistent challenges faced by e-commerce platforms is the issue of low conversion rates—only a small fraction of website visitors complete a purchase [1]. Unlike physical retailing, where the loss of a customer can often be traced to direct, observable factors such as poor in-store experience or product unavailability, the digital environment poses unique challenges [2]. Here, customer drop-off may occur at any point, from browsing to checkout, leaving little tangible feedback for businesses to address and optimize their processes. To combat this issue, detection and behavioral analysis systems have emerged as crucial tools for understanding customer interaction and decision-making patterns. These systems leverage vast amounts of user data, such as clickstreams, search queries, and cart abandonment rates, to uncover insights into customer behavior and potential pain points [3, 4]. In recent years, the introduction of machine learning and deep learning has revolutionized these systems, enabling the development of predictive and adaptive models that can identify underlying patterns and trends with remarkable accuracy [5, 6]. These technologies have not only improved the ability to detect customer disengagement early but also allowed for personalized interventions to increase the likelihood of conversion.

In this paper, we focus on predicting online shoppers' purchase intentions using the benchmark dataset introduced by Sakar et al. [6], which has been extensively used in related studies [7–16]. Our approach emphasizes the challenges posed by the dataset's imbalanced nature, as only a small percentage of records correspond to the "Buy" class. To address this, we utilize performance metrics such as F1 score, Matthews Correlation Coefficient (MCC), area under the Receiver Operating Characteristic curve (auROC), and area under the Precision-Recall curve (auPR), following the recommendations for imbalanced datasets [15]. These metrics provide a more reliable evaluation of model performance compared to traditional accuracy-based measures. Unlike some prior studies that directly use all dataset features for analysis [11, 14], we first conduct a feature selection process to identify the most relevant attributes and eliminate redundancy. This ensures a more efficient and interpretable training process. Furthermore, we experiment with various classification algorithms, enabling a thorough comparative analysis of their performance. Our results are evaluated against those reported in the literature to highlight the strengths of our approach. By building on established methods and addressing common drawbacks, such as the over-reliance on confusion-matrix-based metrics, this study aims to contribute to the development of more robust models for predicting online shoppers' purchasing intentions.

II. Related Work

Several works in the literature have explored the problem of predicting customers' purchasing intentions on e-commerce platforms. For example, Esmeli et al. [17] proposed a machine learning model that predicts purchasing intention within a session by dynamically creating features. Using classifiers such as Decision Tree, Random Forest, and Naive Bayes, they achieved high performance, with the Decision Tree obtaining 97% auROC. In another study [18], the authors have introduced binary classification over user sessions in an online store, which are two types: buying sessions and browsing sessions. They explored the utility of a K-NN classifier, determining the optimal value of K to be 11, which yielded a sensitivity of 87.5% and an impressive accuracy of 99.85%. On the behavioral side, Houda et al. [19] conducted an empirical study with 147 students, emphasizing that cognitive variables like perceived usefulness and ease of use significantly influence online shopping attitudes. In another interesting approach, Chung et al. [20] analyzed the impact of input devices on purchasing decisions, finding that shoppers using touch interfaces demonstrated higher engagement and were more likely to choose hedonic over utilitarian options. These insights highlight the interplay between user experience design and purchasing behavior. Moreover, studies like that of Rita et al. [21] explored e-service quality and its influence on consumer behavior, identifying website design, privacy, and fulfillment as critical factors affecting customer trust. This focus on consumer-centric design is echoed by Dabbous et al. [22], who examined the role of social media stimuli, including content quality and brand interaction, in shaping brand awareness and offline purchase intentions. Notably, Martins et al. [23] highlighted the role of smartphone advertising, combining web advertising models with flow experience theory to show how interactive web designs and brand recognition positively influence purchasing behavior.

III. Methods

In this section, we briefly describe the dataset, explain how our code works, and introduce the model used for classification. First, we discuss the dataset's features and their meanings. Next, we cover the necessary preprocessing steps for model construction. Finally, we outline the learning algorithms applied to the dataset.

A. Dataset Description

In this study, the purchasing intention model is formulated as a binary classification problem, focusing on predicting whether a user intends to complete a transaction. The goal is to offer tailored content to users likely to make a purchase while avoiding irrelevant recommendations for others. Numerical and categorical features used for predicting purchasing intention are detailed in Tables 1 and 2.

We downloaded the dataset from UCI Machine Learning Repository [24]. It comprises feature vectors for 12,330 sessions, each corresponding to a unique user over a one-year period. This design prevents biases associated with specific campaigns, special days, user profiles, or time periods. Among these sessions, 84.5% (10,422) belong to the negative class (sessions that did not result in a transaction), while the remaining 15.5% (1,908) are positive class samples (sessions that ended with a purchase), which is shown as categorical feature of "Revenue" in Table 2.

Table 1 highlights the numerical features along with their statistical summaries. These include variables such as Administrative, Administrative Duration, Informational, Informational Duration, Product Related, and Product Related Duration, which quantify the number and total time spent on various page categories during a session. These values are extracted in real time from the URLs visited by the user and are dynamically updated based on user actions, such as navigating between pages.

Additional numerical features—Bounce Rate, Exit Rate, and Page Value—are metrics obtained via Google Analytics. Specifically, Bounce Rate represents the percentage of visitors who leave the site after visiting a single page without triggering any additional interactions. Exit Rate refers to the percentage of sessions ending on a specific page, while Page Value reflects the average value of pages visited before a completed transaction. These metrics are stored in the application database and updated automatically.

The Special Day feature, as shown in Table 1, captures the temporal proximity of a session to significant events (e.g., Mother's Day, Valentine's Day) that might influence purchasing behavior. Its value ranges from 0 to 1, with nonzero values assigned during relevant periods leading up to the event. For instance, in the case of Valentine's Day, this feature takes nonzero values from February 2 to February 12, peaking at 1 on February 8.

TABLE I
Numerical Features provided in the dataset

Feature Name	Feature Description	Min. Value	Max. Value	SD
Administrative	Number of pages visited by the visitor about account management	0	27	3.32
Administrative duration	Total amount of time (in seconds) spent by the visitor on account management related pages	0	3398	176.70
Informational	Number of pages visited by the visitor about website, communication, and address information of the shopping site	0	24	1.26
Informational duration	Total amount of time (in seconds) spent by the visitor on informational pages	0	2549	140.64
Product related	Number of pages visited by the visitor about product related pages	0	705	44.45
Product related duration	Total amount of time (in seconds) spent by the visitor on product related pages	0	63,973	1912.25
Bounce rate	Average bounce rate value of the pages visited by the visitor	0	0.2	0.04
Exit rate	Average exit rate value of the pages visited by the visitor	0	0.2	0.05
Page value	Average page value of the pages visited by the visitor	0	361	18.55
Special day	Closeness of the site visiting time to a special day	0	1.0	0.19

TABLE II
Categorical Features provided in the dataset

Feature Name	Feature Description	Number of Categorical Values
OperatingSystems	Operating system of the visitor	8
Browser	Browser of the visitor	13
Region	Geographic region from which the session has been started by the visitor	9
TrafficType	Traffic source by which the visitor has arrived at the website (e.g., banner, SMS, direct)	20
VisitorType	Visitor type as "New Visitor," "Returning Visitor," and "Other"	3
Weekend	Boolean value indicating whether the date of the visit is weekend	2
Month	Month value of the visit date	12
Revenue	Class label indicating whether the visit has been finalized with a transaction	2

By combining these carefully selected features, the dataset enables a robust exploration of user behavior and purchasing patterns for binary classification modeling.

B. Data Preprocessing

We conducted our analyses using the Jupyter Notebook environment, leveraging libraries such as pandas and numpy for numerical analyses and scikit-learn for implementing classifier algorithms.

Initially, we examined the dataset's structure and confirmed the absence of missing values (NaN). The dataset contained 18 features, including both numerical and categorical ones. While there were 125 duplicate records, this did not pose a concern, as Sakar et al. [6] confirmed that the sessions were independent and represented distinct users.

To better understand the categorical features, we used countplots and observed that most features were non-uniformly distributed across instances. For the target variable, Revenue, we noted that only 15% of the samples belonged to the 'buy' class, indicating class imbalance.

We explored the relationships between the target variable and other features. For instance, sessions in May and November exhibited the highest revenues. Using boxplots and violin plots, we identified patterns in the numerical features. High Page Value counts were associated with the 'buy' class, while a high Exit Rate was more prevalent in the 'no-buy' class.

Additionally, a correlation heatmap revealed significant correlations among many numerical features. This suggested the potential for optimizing feature selection based on inter-feature relationships, as highlighted by Rizal et al. [13].

To prepare the data for machine learning algorithms, we split the dataset into 80% training and 20% testing subsets. For encoding, we applied one-hot encoding to categorical features and scaled numerical columns using a Robust Scaler. The robust scaling technique mitigates the impact of outliers, ensuring compatibility with various learning algorithms while preserving the relationships within the data. After preprocessing, the dataset expanded to 78 features, primarily due to one-hot encoding.

C. Handling Data Imbalance

To address the class imbalance in our dataset (with only 15% positive samples), we employed the Synthetic Minority Oversampling Technique (SMOTE) to balance the training data [25]. The process was carried out as follows:

- Reserved 30% of the dataset (12,330 samples) for testing.
- Applied SMOTE to the remaining 70% of the dataset for training.
- SMOTE was used to oversample the minority class (positive samples), generating synthetic samples.
- The synthetic samples helped balance the class distribution, enabling the model to better learn patterns related to purchasing behavior.
- This approach ensured a balanced dataset for training while addressing the skewed class distribution.

D. Feature Selection

After balancing the dataset, we focused on selecting the most impactful features for the purchasing intention prediction model. To achieve this, we utilized a Random Forest Classifier, an ensemble learning method based on decision trees. Random forests are widely used in predictive modeling due to their high accuracy, as noted in the literature [6, 11, 15, 17, 25].

Beyond prediction, random forests also provide feature importance scores, helping identify the features most critical to predicting shopper behavior. Based on these importance scores, we defined a threshold value to select features for the learning phase.

We found that limiting the dataset to the top 25 features,which can be seen on Figure 1, (out of 78) yielded the best cross-validation training score of 92% F1-score. This score was prioritized over accuracy, given the dataset's imbalanced nature. Accurately identifying potential buying customers (positive class) was deemed more important, even at the cost of a slight reduction in overall accuracy.

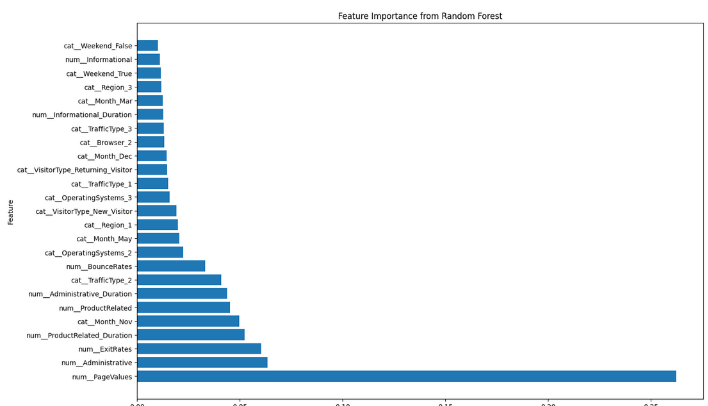

Fig. 1. Feature Importance calculated by Random Forest Classifier.

E. Classification Algorithms

In this section, we detail the learning algorithms used to represent the generalizations inherent in the shopper's intention dataset. The goal was to achieve the best predictions while understanding the dataset's patterns.

1) Random Forest: To identify the most relevant features for our subsequent learning steps, we employed a **Random Forest (RF)** classifier. Random Forest, as an ensemble learning algorithm, is efficient for both feature selection and predictive tasks. It combines the outputs of multiple decision trees, reducing the risk of overfitting and increasing robustness. Moreover, it provides *feature importance scores*, which guided our feature selection process.

The Random Forest parameters were optimized, resulting in:

- **n_estimators**: 130
- **max_depth**: 18

These parameters align closely with those reported in [16], where **n_estimators = 100** and **max_depth = 20** yielded high performance on a similar dataset. The optimized RF model demonstrated the ability to generalize well, which justified its use in feature extraction for this dataset [11].

2) XGBoost: For predictions, we utilized **eXtreme Gradient Boosting (XGBoost)**, a gradient boosting algorithm widely used in the literature for its strong performance in structured datasets [16]. XGBoost builds trees sequentially, minimizing errors of the previous trees, and effectively handles non-linear relationships and class imbalance.

The best parameters obtained for XGBoost were:

- **learning_rate**: 0.01 (controls step size for gradient updates)
- **max_depth**: 8 (maximum depth of each tree)
- **n_estimators**: 150 (number of boosting rounds)
- **subsample**: 0.5 (fraction of samples used for training each tree)

The parameter **max_depth = 8** may seem large for a weak learner, but it is reasonable given the dataset contains **25 selected features**, ensuring each tree has sufficient flexibility to capture patterns. This configuration yielded a **Best Cross-Validation F1 Score** of **0.909**, demonstrating its effectiveness on the imbalanced dataset.

3) Support Vector Machine (SVM): We also experimented with **Support Vector Machine (SVM)** models, specifically using the **RBF kernel** to handle non-linear decision boundaries. SVMs are known for their effectiveness in smaller feature spaces and are used extensively in classification problems.

Our parameter optimization resulted in:

- **Kernel**: RBF (Radial Basis Function)
- **C**: 10 (regularization parameter, balancing margin width and misclassification tolerance)

These values slightly differed from those reported in [16], where **C = 7** provided the best results. The higher value of **C** in our model indicates a preference for correctly classifying the training data, even if it sacrifices margin width.

4) Stacking Model: Finally, we implemented a **stacking model** to combine the predictive power of multiple algorithms. The stacking ensemble consisted of:

- **Base learners**: XGBoost, SVM, and Random Forest (using their respective best parameters).
- **Meta-learner**: Logistic Regression.

While training the stacking model, we encountered convergence issues with Logistic Regression at low maximum iterations (e.g., **300**). Increasing the maximum iterations resolved the issue and allowed the meta-learner to converge properly.

The minimal improvements in metrics suggest that overfitting may be occurring. However, when the dataset, conditions, and models are carefully selected, this approach can still be viable, leveraging the strengths of each base learner to achieve better overall performance.

F. Performance Metrics

In this study, we evaluate the performance of our models using a variety of metrics commonly employed in binary classification tasks. These metrics are calculated based on the confusion matrix, which provides the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Additionally, we include threshold-independent metrics such as the area under the receiver operating characteristic curve (auROC) and the area under the precision-recall curve (auPR), which are especially useful for imbalanced datasets. Below, we explain each metric in detail.

1) Accuracy: Accuracy represents the percentage of correctly predicted instances over the total number of instances. It provides an overall measure of model performance but may be less informative in cases of imbalanced datasets. The formula for accuracy is given in Equation 1:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2) Precision: Precision measures the proportion of correctly predicted positive instances to the total predicted positive instances. A high precision value indicates that the model produces fewer false positives. The formula for precision is shown in Equation 2:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

3) Recall (True Positive Rate): Recall, also known as sensitivity or true positive rate (TPR), is the proportion of true positive predictions to the total number of actual positive instances. A high recall value indicates that the model captures most of the positive instances. The formula is given in Equation 3:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

4) F1-Score: The F1-Score is the harmonic mean of precision and recall. It is a balanced metric that provides a single value reflecting both the precision and recall of the model. The formula is shown in Equation 4:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

5) Matthews Correlation Coefficient (MCC): The Matthews Correlation Coefficient (MCC) is a comprehensive metric that takes into account all four quadrants of the confusion matrix. It provides a balanced measure of model performance, even for imbalanced datasets. The formula for MCC is given in Equation 5:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

6) Area Under the ROC Curve (auROC): The area under the receiver operating characteristic curve (auROC) represents the trade-off between the true positive rate (TPR) and the false positive rate (FPR) across different thresholds. It is a threshold-independent metric and is particularly suitable for evaluating the discriminatory ability of a classifier on imbalanced datasets.

7) Area Under the Precision-Recall Curve (auPR): The area under the precision-recall curve (auPR) evaluates the trade-off between precision and recall across varying thresholds. It is more informative than auROC when dealing with highly imbalanced datasets, as it focuses on the performance of the model in predicting the minority class.

These metrics collectively provide a comprehensive evaluation of the models' performance, capturing both their overall accuracy and their ability to handle imbalanced data effectively.

IV. Results and Discussions

In this section, we report the results of our experiments using the aforementioned performance evaluation metrics.Also, we compare our results with other papers in the literature. The implementation of all algorithms was carried out using the Scikit-learn library [26] within the Jupyter notebook environment. The dataset was divided into training and testing sets with an 80-20 split. Feature selection and oversampling techniques were applied exclusively to the training set to prevent any potential overfitting during the experiments.

1) Addressing Class Imbalance with SMOTE: The impact of class imbalance was evident in preliminary experiments, where models trained on the imbalanced dataset demonstrated poor classification performance. Table III, adapted from Baati [11], highlights the detrimental effect of imbalance on Random Forest, which achieved an F1-Score of only 0.10. This aligns with prior findings that emphasize the importance of addressing class imbalance to improve predictive performance.

TABLE III
Results obtained on class imbalanced dataset, adapted from Baati [11].

Classifier	**Accuracy (%)**	**F1 Score**
Random Forest	83.64	0.10

To address this issue, SMOTE was employed to oversample the minority class in the training set. This technique effectively improved model performance across all metrics, as summarized in Table IV. By generating synthetic examples, SMOTE ensured a balanced dataset and allowed models to learn equally from all classes, resulting in improved generalization and predictive accuracy.

TABLE IV
Performance metrics for various models trained on the balanced dataset.

Model	**F1**	**Accuracy**	**Precision**	**Recall**	**MCC**	**auROC**	**auPR**
Logistic Regression (LR)	0.8761	0.8669	0.8939	0.8669	0.5664	0.8871	0.6394
Random Forest (RF)	0.8925	0.8914	0.8936	0.8914	0.5786	0.9217	0.6959
Stacking	0.8930	0.8906	0.8961	0.8906	0.5875	0.9233	0.6960
XGBoost	**0.8956**	**0.8931**	**0.8990**	**0.8931**	**0.5988**	**0.9283**	**0.7140**
Support Vector Classifier (SVC)	0.8783	0.8710	0.8909	0.8710	0.5595	0.8958	0.6350

2) Performance Analysis of Learning Algorithms: Table IV presents the performance metrics for various models on the balanced dataset. Among the evaluated models, XGBoost consistently demonstrated superior performance across all metrics. It achieved the highest F1-Score (0.8956) and accuracy (0.8931), indicating its effectiveness in providing balanced predictions and overall classification performance. XGBoost also recorded the highest Matthews Correlation Coefficient (MCC) value (0.5988), confirming its ability to minimize false positives and negatives effectively.

The gradient-boosting mechanism of XGBoost, which sequentially optimizes weak learners, enables it to excel in handling complex patterns and imbalanced datasets. Its tree-based architecture also contributed to the highest auROC (0.9283) and auPR (0.7140) values, showcasing its robustness in distinguishing between classes and maintaining high precision and recall.

Random Forest also performed strongly, achieving an MCC of 0.5786 and an F1-Score of 0.8925. Its ensemble-based bagging approach minimized overfitting and variance, making it a reliable classifier. However, its performance slightly lagged behind XGBoost in all metrics.

The Stacking model, which integrates predictions from multiple base learners, achieved an MCC of 0.5875, slightly outperforming Random Forest. However, the marginal improvement suggests that the stacking model's ensemble structure did not significantly enhance performance compared to its base learners, such as Random Forest.

Support Vector Classifier (SVC) and Logistic Regression demonstrated balanced yet comparatively lower performance. Their MCC values of 0.5595 and 0.5664, respectively, indicate limitations in capturing complex interactions within the dataset, which impacted their overall predictive ability.

3) MCC as a Robust Performance Metric: The Matthews Correlation Coefficient (MCC) scores, depicted in Figure 2, provide critical insights into the models' balanced performance. As MCC accounts for all entries in the confusion matrix, it is particularly well-suited for evaluating imbalanced datasets. XGBoost achieved the highest MCC score, reaffirming its balanced predictions for both positive and negative classes.

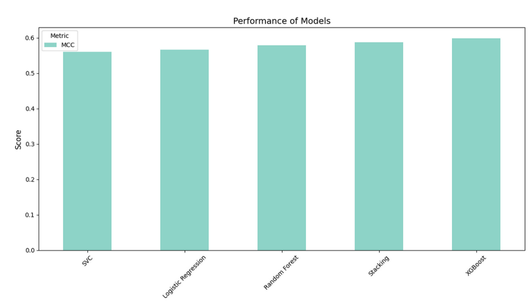

Fig. 2. MCC values for all models on the balanced dataset.

Random Forest and Stacking also achieved high MCC scores, reinforcing their robustness as ensemble methods. However, their performance was slightly lower than XGBoost, highlighting the gradient-boosting technique's superior ability to optimize predictions. Logistic Regression and SVC, while providing reasonable MCC scores, were limited by their simpler structures and inability to handle complex patterns effectively.

A. Significance of SMOTE and Gradient Boosting

The experimental results underscore the importance of addressing class imbalance through techniques such as SMOTE. By balancing the dataset, SMOTE significantly improved the models' predictive performance, as evidenced by the enhanced F1-Scores, MCC values, and auROC scores. The findings also highlight the strength of gradient boosting methods, particularly XGBoost, in handling complex classification tasks and imbalanced datasets. These results emphasize the need for careful preprocessing and model selection to achieve optimal performance in real-world applications.

V. Conclusion

E-commerce platforms face the pressing challenge of low conversion rates, which hinder profitability and customer retention. This study aimed to address this issue by developing machine learning models to predict online shoppers' purchase intentions. By utilizing a benchmark dataset and tackling its inherent class imbalance with SMOTE, we enhanced model training and enabled a more accurate prediction of purchase behaviors.

Our findings demonstrate the superior performance of ensemble models, particularly XGBoost, which achieved the highest F1-Score, MCC, and auROC among all tested algorithms. The feature selection process further improved model interpretability and efficiency by isolating the most relevant attributes of shopper behavior. These results emphasize the importance of preprocessing and algorithm choice in tackling the complex challenges of e-commerce data.

This work contributes to the growing field of predictive analytics in e-commerce by demonstrating the potential of machine learning to optimize customer interaction and decision-making processes. Future research could explore hybrid oversampling techniques, deep learning architectures, and real-time behavioral analysis systems to build even more accurate and adaptive models. By addressing the unique pain points of digital retail environments, these advances hold promise for significantly improving e-commerce conversion rates and customer experiences.

References

[1] Rajamma RK, Paswan AK, Hossain MM (2009) Why do shop pers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience. J Prod Brand Manag 18(3):188–197

[2] Moe WW (2003) Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream. J Consum Psychol 13(1–2):29–39

[3] Monica Law, Ron Chi-Wai Kwok, Mark Ng, An extended online purchase intention model for middle-aged online users, Electron. Commer. Res. Appl. 20 (2016) 132–146. [5]

[4] C. Ruiz-Mafe, K. Chatzipanagiotou, and R. Curras-Perez, (2018). The role of emotions and conflicting online reviews on consumers' purchase intentions. Journal of Business Research, 89, 336-344.

[5] Grazyna Suchacka, Magdalena Skolimowska-Kulig, Aneta Potempa, Classification of e-customer sessions based on support vector machine, ECMS 15 (2015) 594–600.

[6] C. Sakar, S. Polat, M. Katircioglu and Y. Kastro, "Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks", Neural Computing and Applications, vol. 31, no. 10, pp. 6893-6908, 2018. Available: 10.1007/s00521-018-3523-0.

[7] J. Adhikari, "Online Shoppers' Purchase Intention using Ensemble Learning Approach", International Journal of Next-Generation Computing, 2023. Available: https://api.semanticscholar.org/CorpusID:266249313.

[8] C. Yap and K. C. Khor, "Utilising Sampling Methods to Improve the Prediction on Customers' Buying Intention", in Proc. 2022 International Conference on Decision Aid Sciences and Applications (DASA), 2022, pp. 352-356. doi: 10.1109/DASA54658.2022.9765293.

[9] G. Sang and S. Wu, "Predicting the Intention of Online Shoppers' Purchasing", in Proc. 2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), 2022, pp. 333-337. Available: https://api.semanticscholar.org/CorpusID:253555650.

[10] M. A. Mostafa, W. S. Al-Dayyeni, A. H. Fadel, A. H. Abbas, M. M. Jaber, and R. R. Ali, "A Classification Technique for Online Shoppers' Purchasing Intention", in Proc. 2024 1st International Conference on Logistics (ICL), 2024, pp. 1-6. doi: 10.1109/ICL62932.2024.10788661.

[11] Karim Baati, Mouad Mohsil, Real-time prediction of online shoppers' purchasing intention using random forest, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 2020, pp. 43–51.

[12] Peiyi Song, Yutong Liu, An xgboost algorithm for predicting purchasing behaviour on e-commerce platforms, Teh. Vjesn. 27 (5) (2020) 1467–1471.

[13] Rizal Dwi Prayogo, Siti Amatullah Karimah, Feature selection and adaptive synthetic sampling approach for optimizing online shopper purchase intent prediction, in: 2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), IEEE, 2021, pp. 1–5.

[14] Md Rayhan Kabir, Faisal Bin Ashraf, Rasif Ajwad, Analysis of different predicting model for online shoppers' purchase intention from empirical data, in: 2019 22nd International Conference on Computer and Information Technology (ICCIT), IEEE, 2019, pp. 1–6.

[15] Faqih Hamami, Ahmad Muzakki, Machine learning pipeline for online shopper intention classification, in: AIP Conference Proceedings, vol. 2329:1, AIP Pub lishing LLC, 2021, p. 050014.

[16] A.-A. Tanvir, I. A. Khandokar, A. K. M. M. Islam, S. Islam, and S. Shatabda, "A gradient boosting classifier

for purchase intention prediction of online shoppers," Heliyon, vol. 9, no. 4, p. e15163, 2023.

[17] Ramazan Esmeli, Mohamed Bader-El-Den, Hassana Abdullahi, Towards early purchase intention prediction in online session based retailing systems, in: Electronic Markets, 2020, pp. 1–19.

[18] Grazyna Suchacka, Magdalena Skolimowska-Kulig, Aneta Potempa, A k-nearest neighbors method for classifying user sessions in e-commerce scenario, J. Telecommun. Inf. Technol. (2015).

[19] Houda Zarrad, Mohsen Debabi, Online purchasing intention: factors and effects, Int. Bus. Manag. 4(1) (2012) 37–47.

[20] Sorim Chung, Thomas Kramer, Elaine M. Wong, Do touch interface users feel more engaged? The impact of input device type on online shoppers' engagement, affect, and purchase decisions, Psychol. Mark. 35 (11) (2018) 795–806.

[21] Paulo Rita, Tiago Oliveira, Almira Farisa, The impact of e-service quality and customer satisfaction on customer behavior in online shopping, Heliyon 5 (10) (2019) e02690.

[22] Amal Dabbous, Karine Aoun Barakat, Bridging the online offline gap: assessing the impact of brands' social network content quality on brand awareness and purchase intention, J. Retail. Consum. Serv. 53 (2020) 101966.

[23] José Martins, Catarina Costa, Tiago Oliveira, Ramiro Gonçalves, Frederico Branco, How smartphone advertising influences consumers' purchase intention, J. Bus. Res. 94 (2019) 378–387.

[24] URL https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset

[25] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002)

[26] URL https://scikit-learn.org/stable/