APPLICATION OF MACHINE LEARNING ALGORITHMS IN PREDICTING CUSTOMER LOYALTY TOWARDS GROCERY RETAILERS

Jelena Franjkovic[1],
Ivana Fosic[2],
Ana Zivkovic[3]

Abstract: Retailers strive for customer loyalty in the sense of repeat purchases, but also as a high proportion of purchases (compared to competitors) and willingness to recommend to other customers. This paper examines customer loyalty in the grocery sector as a three-dimensional construct and shows how machine learning techniques can be useful in its study. Price characteristics of the retailer (price level, value for money, price dynamics, price communication and price dispersion) and non-price characteristics of the retailer (general product range, retailer's private label product range, store design and atmosphere, service level and location) are included in the model as predictor variables. Using the data collected through the primary research conducted in Croatia, 433 samples were divided into 10 independent predictor variables and one dependent variable (customer loyalty), a prediction was created using supervised machine learning classification algorithms. The Random Forest classifier proves to be the best choice overall, with ROC_AUC value of 0.790, a high accuracy of 0.915 and an F1 score of 0.954, reflecting both precision and responsiveness. The application of the SHapley Additive exPlanations analysis additionally enables the interpretation of the results, highlighting the influence of features on the accuracy of the prediction. The results indicate that price dynamics and service level are the most important features for the model predictions, followed by value for money and price communication.

[1]Department of Marketing, Faculty of Economics and Business in Osijek, Osijek, Croatia, e-mail: jelena.franjkovic@efos.hr, ORCID: 0000-0001-7725-3098

[2] Department of Management, Organization and Entrepreneurship, Faculty of Economics and Business in Osijek, Osijek, Croatia, e-mail: ivana.fosic@efos.hr, ORCID: 0000-0003-4893-3654

[3] Department of Management, Organization and Entrepreneurship, Faculty of Economics and Business in Osijek, Osijek, Croatia, e-mail: ana.zivkovic@efos.hr, ORCID: 0000-0002-6469-4377

Key words: customer loyalty, grocery retail, supervised machine learning, prediction, price dynamics
JEL: M30, C53.
DOI: https://doi.org/10.58861/tae.bm.2025.2.05

Introduction

Every company strives for customer loyalty, for a long-term relationship with the consumer. The specificity of grocery and food retailing lies in a wide range of factors, including the price and non-price characteristics of the retailer, which can influence the loyalty of customers towards the retailer and its brand as a company. The approach to customer loyalty in retail is more sophisticated and comprehensive than loyalty to a single manufacturer brand. Retailers advertise on price and product availability while providing a certain level of service that consumers continually evaluate. Retailers collect data on consumers through POS systems and loyalty programs, which encourages the use of machine learning (ML) approaches to data analysis. It is also even easier for online retailers to use AI and ML tools (Gauri et al., 2021). The topic is also gaining attention among academics who are exploring the role and possibilities of machine learning, AI and IoT (Popova, & Petrova, 2024; Popova et al., 2024; Petrova et al., 2022) and its predictive capabilities with the aim of increasing customer satisfaction and loyalty (e.g. Andresini et al., 2023; Wang et al., 2021; Rane et al., 2023; Islamgaleyev et al., 2020; Arefin et al., 2024). The aim of this paper is to evaluate the application of ML algorithms to predict loyalty towards retailers in traditional grocery retailing, focusing on the high accuracy and reliability of the prediction results. Particular attention will be paid to analyzing the factors and identifying the key price and non-price factors that contribute to customer loyalty. The aim is to enable more precise targeting and adaptation of marketing strategies. The ML algorithms aim to gain a more detailed and accurate understanding of the factors that influence customer loyalty to grocery retailers. The application of SHapley Additive exPlanations (SHAP) analysis additionally enables the interpretation of the results by highlighting the influence of individual characteristics on the accuracy of the prediction. The research problem includes the following aspects: the identification of key factors, the questions of how accurately different ML algorithms can predict customer loyalty based on the identified factors, how the individual factors behave within the model and how much they contribute to the final prediction. This approach improves prediction performance and provides valuable insights for developing effective strategies to create and maintain customer loyalty.

1. Literature Review

1.1. Customer loyalty and influencing factors

Customer loyalty towards the retailer can encompass more dimensions than the repeat purchase itself, even though it is very often seen as loyalty. Attracting customers to the store is only one step towards the goal of creating a loyal customer. In this paper, customer loyalty is examined as a multidimensional construct, taking into account not only purchase intention but also purchase share and willingness to recommend. Purchase intention refers to the intention to buy again and mainly concerns retailers with which consumers have already had experience. Purchase share is an important dimension as most consumers make their purchases at multiple retailers and through multiple channels (Flavian et al., 2020). The focus of this research is on the share of purchases from different grocery retailers and their physical stores, as online sales of food and fast moving consumer goods (FMCG) still lag behind other retail sectors. Willingness to recommend is an emotional dimension of loyalty and one of the effects of overall consumer satisfaction (Kumar et al., 2017).

In this paper, the prediction of customer loyalty is examined on the basis of price and non-price characteristics of the grocery retailer, which are described very briefly below. Price level (PL) refers to a certain image of how cheap or expensive the retailer is in customers' minds (Graciola et al., 2018). Value for money (VM) can be seen as the result of an appropriate price and quality evaluation (Zielke, 2006), not only of the perceived price-quality ratio of the individual product but of the retailer's overall service. Price dynamics (PDY) refers to frequent and short-term price changes, usually in the form of price promotions (Cacchiarelli i Sorrentino, 2019). Its main purpose is to attract customers into stores. Price communication (PC) can influence customers' perception of price image (Grandi & Cardinali, 2022) and refers to the general information provided to consumers about prices through various communication channels, including the store. Price dispersion (PD) stands for the dispersion of prices within the retailer's range, which largely depends on the size and depth of the range (Hamilton

& Chernev, 2013). As one of the most important non-price characteristics of the retailer that expresses the retailer's identity (Graciola et al., 2018), the total product range (PR) stands out, "a set of products and/or services that a retailer offers to consumers" (Kumar et al., 2017). To further differentiate themselves from the competition (Kumar & Kim, 2014), grocery retailers are intensively developing their private label product range (PLPR). Store design and atmosphere (SDA) refer to factors that influence consumers' shopping mood (Kumar et al., 2017). Retail service level (SL) is a very broad concept. This paper focuses on in-store staff and their contribution to a positive customer experience (O'Cass & Grace, 2008) as well as the overall impression of service level, e.g. product availability. Location (L) remains one of the most important attributes of physical stores. Bell (2014) points out that even in modern retail, despite the development of e-commerce, location is still "everything".

1.2. Supervised machine learning and classification (prediction) algorithms

ML methods can be categorized according to the degree of human control or supervision of the learning process, including supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning aims to develop a model based on labeled data to enable the prediction of future data, with the two most common tasks being classification and regression. These methods form the basis of modern ML thanks to their precision and wide application, and are crucial for the development of advanced analytical and predictive models (Fosic et al., 2022). The Random Forest (RF) algorithm consists of multiple decision trees applied to different subsets of a dataset to improve prediction accuracy over the original dataset. RF combines the predictions of all trees and determines the final result based on a majority decision (Hasan et al., 2014). Support Vector Classifier (SVC) is an advanced ML algorithm used in classification tasks and is based on defining a boundary between different datasets that can be categorized into different classes (Sen et al., 2020). Gaussian Naive Bayes (GNB) supports continuous data that follows a Gaussian normal distribution and relies on Naive Bayes (NB), derived from Bayes' theorem. NB is based on the assumption that features are independent of each other (Naiem et al., 2023). The Stochastic Gradient Descent Classifier (SGD) is a variant of the gradient algorithm used for optimization. SGD uses only a random sample to estimate the gradient in each iteration. This reduces computational complexity and makes the learning process faster and more efficient, especially for large datasets, although gradient estimation may be less precise (Tian et al., 2023). K-nearest neighbors (KNN) is a supervised ML algorithm that classifies results based on the majority class of its nearest neighbors. Its goal is to classify new entries based on samples from the training set. It is based on finding the minimum distance between the observed sample and the samples from the training set to identify the nearest neighbors. The class of the majority of these nearest neighbors is used to predict the class of the observed sample (Hu et al., 2016). A Decision Tree (DT) is a non-parametric supervised learning method suitable for classification and regression. It predicts the value of the target variable by learning simple decision rules based on data features. The DT Classifier takes two data as input: an array of X shapes (n_samples, n_features) containing the training samples and an array of Y integer values (n_samples) containing the class labels for these samples (Decision Trees, n.d.). Gradient Boosting Classifier (GBC) is an algorithm that gradually builds an additive model and enables the optimization of different differentiable loss functions. It efficiently models complex relationships between features and targets, improving accuracy (Czakon, 2024). Classification models have two main goals: 1) they should perform well, i.e. they should predict the output data for new input features as accurately as possible; 2) they should be interpretable, i.e. provide an understanding of the relationship between input features and output data. The most common analysis of classification models is via feature importance, i.e. how important a particular feature is to the performance of a classification model. This is a measure of the individual contribution of the corresponding feature to a particular classifier, regardless of the form or direction of the feature influence. The importance of the features of the input data depends on the corresponding classification model and a feature that is important for one model may be unimportant for another model (Saarela & Jauhiainen, 2021). Shapley Additive Explanation (SHAP) belongs to the family of additive, model-independent feature mapping techniques. It uses game-theoretic concepts to calculate the importance of features in two steps: 1) by training a classification model with all features; 2) by calculating the SHAP value for each feature and ranking it to identify the most important features. The average SHAP value indicates the typical influence of each feature on the predictions of the model over the entire data set, while the absolute SHAP value represents the importance of the feature regardless of its direction (positive or negative). By sorting the features according to their average absolute SHAP values, features with higher SHAP values are identified as more influential on the model predictions (Wang et al., 2024).

2. Research Methodology

2.1. Methodology and research sample

For the empirical study of customer loyalty towards grocery retailers, a survey method was used. It was conducted both web-based with Google Forms and with a questionnaire using pen and paper in 2022. The survey instrument was a highly structured questionnaire based on previous research in the field of retailing and the behavioral aspects of retailing. The respondents were customers over the age of 18 who buy groceries and everyday products for their household in Croatia. A 5-point Likert scale was used for the scales of all variables of the model, which allowed respondents to indicate their level of agreement. 435 questionnaires were collected, two of which were excluded from further processing. 69% of respondents were women, mostly in the age groups 18 – 31 (39%) and 46 – 59 (27%), with a college degree (50%) and employed (62%). Using the collected data consisting of 433 samples divided into 10 independent predictor variables and one dependent variable, a prediction was generated using seven supervised machine learning classification algorithms (RF, SVC, GNB, SGDC, KNN, DT and GBC). The predictor variables were determined by grouping questionnaire questions according to predefined characteristics: retailer price characteristics (PL, VM, PDY, PC and PD) and retailer non-price characteristics (PR, PLPR, SDA, SL and L).

2.2. Data preparation

The data preparation involved several steps. First, the data was cleansed to remove incomplete or incorrect data records. This included the identification and elimination of rows with missing values and the correction or removal of anomalies that could affect the quality of the results. The features of the data set (10 independent features) were functionally determined as the average values of the respondents' statements in this Likert scale for specific questionnaire questions. The dependent variable, representing customer loyalty to retailers, was also determined by calculating the average of respondents' answers to the retailer loyalty questions. For the classification or prediction of the dependent variable by machine learning, the dependent variable was converted to binary values 0 or 1, depending on the average of the values from which the dependent variable was determined. The method of converting the dependent variable into binary form is such that average values less than or equal to 3.0 are labelled with zero (0), while average values of the dependent variable greater than 3.0 are labelled with one unit (1). In this way, a dependent variable in binary form was obtained, where the value 1 indicates the loyalty of customers towards retailers and the value 0 means the lack of loyalty towards retailers. This process of data preparation resulted in an extremely unbalanced data set from the aspect of the dependent variable, where 88% of the dependent variable represented the value 1, i.e. the majority class, and the rest of the 12% of the data represented the minority class, i.e. the value 0 in the dependent variable. To further segment the respondents' loyalty, three dependent sub-variables were identified to determine the specifics of customers' loyalty towards retailers. The determination process was the same, i.e. the average of certain variables, i.e. questions about respondents' loyalty, was calculated. The binary values are also determined according to the same principle as the global dependent variable. This dependent variable also determined the data set as a very unbalanced data set. The process of segmenting the dependent variable was carried out with three groups of questions on customer loyalty towards the retailer, and three different reduced data sets were obtained from the aspect of the dependent variable. After the preparation and creation of four data sets: the base data set, where the dependent variable is represented by the average of all values on the loyalty questions, and three reduced data sets in terms of how the "clustered" dependent variable was created, the data was split into a training and a test data set in a 70:30 ratio to allow the evaluation of model performance on unseen data and to reduce the possibility of overfitting.

3. Results

3.1. Implementation and evaluation of prediction

Python libraries such as scikit-learn and pandas were used to implement the above algorithms. The data was split into a training set and a test set in a 70/30 ratio to enable validation of model performance. Model evaluation was performed using various metrics, including classical accuracy, AUC and F1-score. The success of an ML algorithm is evaluated using various metrics. Classification evaluation is usually done using a single evaluation number that attempts to summarize the specifics of the model. For imbalanced datasets, a common mistake is to rely on the classical classification accuracy score, which can be misleading. The classification of imbalanced datasets refers to the classification problem when the number of samples of the different categories in the dataset is unequal. The solution to such problems depends on the method and the use of data processing algorithms (Zhou et al., 2023). Therefore, the F-beta score and the AUC score were used in this study because they allow a more accurate assessment of model performance in imbalanced datasets. The F-1 score considers the balance between precision and responsiveness, while the AUC score measures the model's ability to discriminate (Fosic et al., 2022). The F-1 score is calculated as 2 * (accuracy * response) / (accuracy + response). It treats accuracy and responsiveness as equally important. One of the most commonly used techniques for evaluating classifiers is the ROC (Receiver Operating Characteristic) curve, a graphical representation of the response measure versus the false positive rate. The classification performance information in the ROC curve can be summarized in a result known as the AUC (Area Under the Curve). This measure is less sensitive to imbalances in the class distribution and represents a compromise between responsiveness and specificity of results. The AUC values range from 0 to 1. An excellent model has an AUC close to 1, which means that it can separate the classes well. A poor model has an AUC value close to 0, which means that it separates the classes very poorly, i.e. it predicts the opposite of the actual values. For a model that cannot separate the classes at all, the AUC value is equal to 0.5. Classical accuracy is measured as the ratio of correct predictions relative to the total number of predictions and is often used as a basic measure of a model's success. However, for imbalanced datasets, classical accuracy can be misleading as it can give high values even if the model does not correctly classify less represented classes (Czakon, 2024; Stapor, 2018). Pseudocode algorithm 1 was used for prediction and classification to find the optimal classifier for the most successful prediction of store loyalty. The results obtained by implementing Algorithm 1 in the Python programming language are shown in Table 1.

Table 1.
Implementation of Algorithm 1 in the Python programming language

MODEL	ACCURACY	ROC_AUC	F1 SCORE
RF	0.915	0.79	0.954
SVC	0.9	0.625	0.946
GNB	0.885	0.691	0.933
SGDC	0.908	0.6	0.95
KNN	0.915	0.662	0.954
DT	0.885	0.674	0.936
GBC	0.915	0.749	0.953

Source: authors' work

Based on the performance metrics shown in Table 1, the Random Forest (RF) classifier proves to be the best choice overall. Not only does it exhibit the highest ROC_AUC value of 0.790, indicating an excellent ability to discriminate between classes, but it also achieves a high accuracy of 0.915 and an F1 score of 0.954, reflecting both precision and responsiveness. Although the Support Vector Classifier (SVC) and Stochastic Gradient Descent Classifier (SGDC) have high accuracy and F1 score, they are less reliable due to their significantly lower ROC_AUC values (0.625 and 0.600). The Gaussian Naive Bayes (GNB) classifier, despite its high ROC_AUC value of 0.691, lags behind the best results in terms of accuracy and F1 score. Even the Decision Tree (DT) classifier with a ROC_AUC value of 0.674 does not achieve the high accuracy and F1 score of RF, KNN and Gradient Boosting Classifier (GBC). Therefore, if the ROC_AUC value is the main focus, RF is the best classifier. If the main focus is on accuracy and F1 score, RF, KNN and GBC are almost equal in their prediction. Based on the capabilities of the RF classifier, which is also the best classifier according to the obtained results shown in Table 1, it can be determined which feature of the dataset is most significant in predicting the results considering the entire dependent variable. The analysis shows that PDY and SL (Figure 1a) are the most important features for the model predictions. VM and PC also contribute significantly, but to a slightly lesser extent. The features PD, PR, SDA and PL are of medium importance, while PLPR and L are the least important.

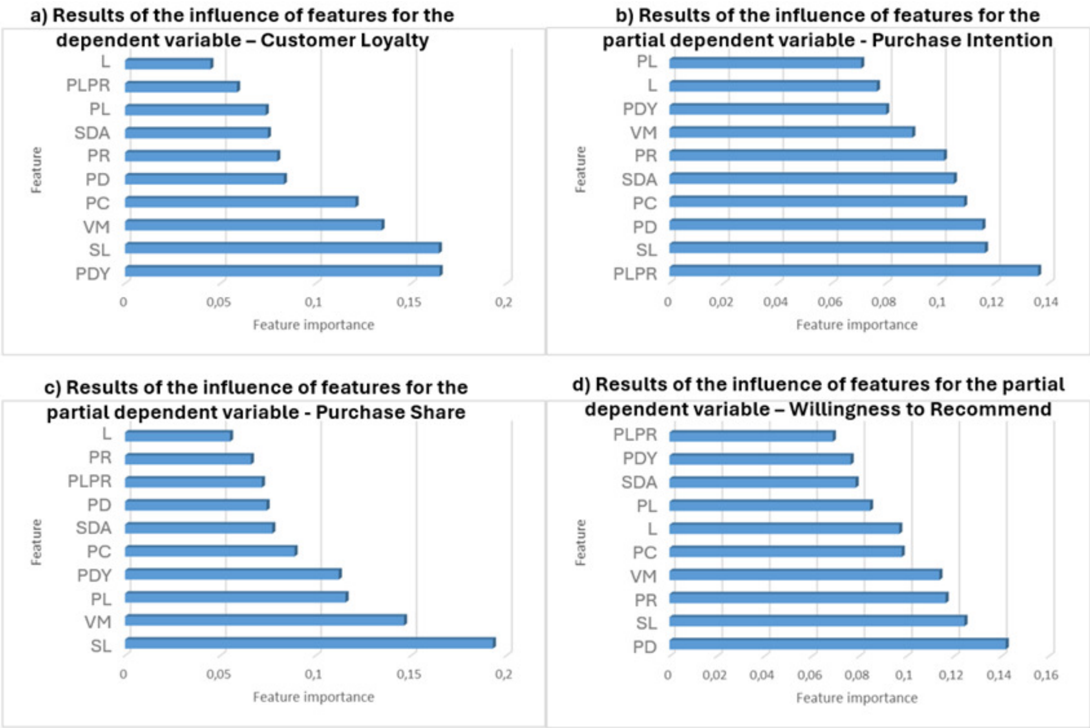


Figure 1. Results of the influence of features for the (partial) dependent variable

Source: authors' work

Since the dependent variable was segmented in the data preparation for a more detailed analysis of respondents' loyalty, three different reduced data sets were obtained from the aspect of the dependent variable. The analysis shows that PLPR is the most important feature for the Purchase Intention (Figure 1b), followed by SL and PD, which are also crucial for the predictions of the model. PC and SDA are moderately important features with a slightly lower influence. PR, VM, PDY, L and PL are less important, with PL being the least important. For the Purchase Share (Figure 1c), the feature importance analysis shows that SL is the most influential feature that significantly affects the model predictions. VM is the second most important feature, while PL and PDY also make an important contribution. PC and SDA are moderately important features, i.e. they play a significant but minor role compared to the most important features. PD, PLPR and PR have lower, but still significant importance, while L is the least important feature and has minimal impact on the model. Finally, for the Willingness to recommend (Figure 1d), the analysis shows that PD is the most influential feature and has the largest impact on the model's predictions. SL, PR and VM also play an important role and contribute significantly to model performance. PC and L have moderate importance, i.e. they are significant but less important than the main features. PL contributes, but to a lesser extent. SDA, PDY and PLPR are the least important features with minimal impact on model predictions. It is recommended to ensure high data quality and representation for PD, SL, PR and VM. Further research on PC and L can help to understand and improve their role. A review of SDA, PDY and PLPR can shed light on whether they add noise or can be improved.

3.2. SHAP analysis

The SHAP analysis was performed for the total and partial dependent variables. In Figure 2, the Y-axis shows features such as VM, SL, PC, PD, PDY, SDA, PL, PR, PLPR and L. The X-axis shows the SHAP values, which represent the influence of each feature on the model output. Positive SHAP values mean that the feature increases the prediction, while negative values mean that it decreases it. For the overall dependent variable (Figure 2a), features such as PC and PD have a wide range of SHAP values, indicating a significant and diverse influence on model performance. These features can both increase and decrease predictions depending on their values. In contrast, features such as L and PLPR have a narrower range of SHAP values, indicating a more consistent and less variable influence on model predictions.

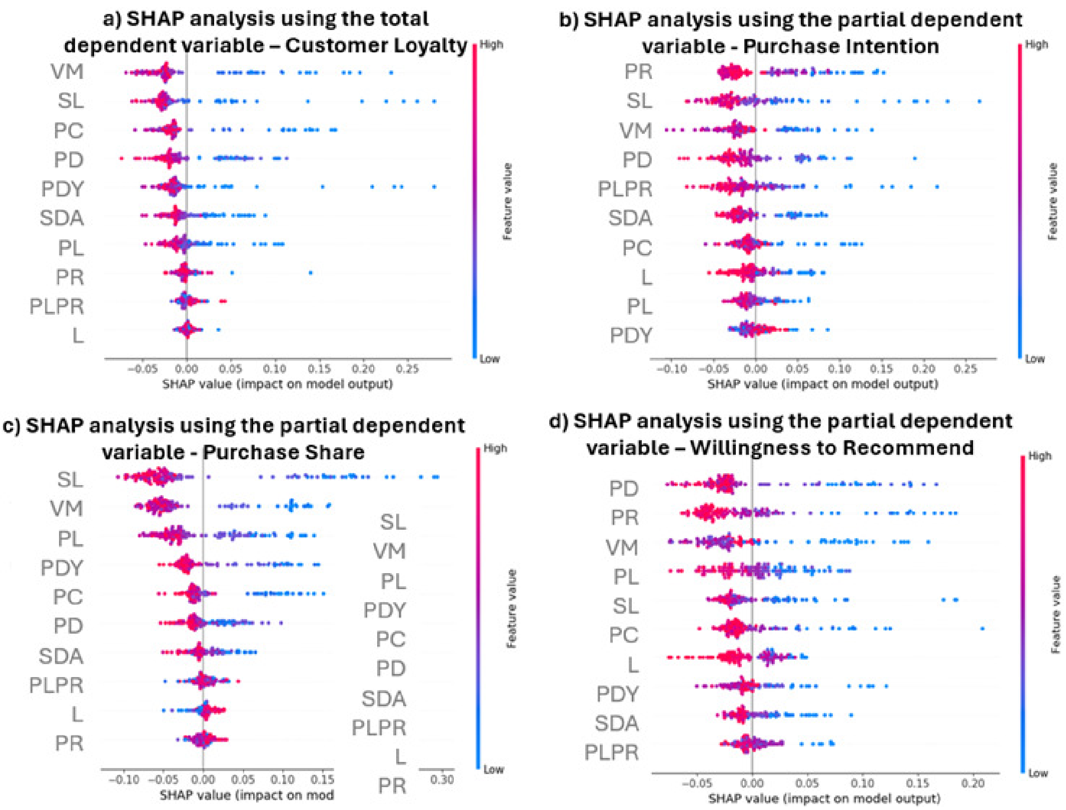


Figure 2. SHAP analysis using the total and partial dependent variable

Source: authors' work

The color of the dots varies from blue to red and stands for features values, with blue for low and red for high values. In the case of PC, for example, high feature values (red dots) are generally associated with positive SHAP values, which means that higher values of this feature improve the prediction of the model. On the other hand, low feature values (blue dots) are associated with negative SHAP values, meaning that lower values of this feature decrease the prediction of the model. In general, features with a wider range of SHAP values, both positive and negative, have a greater impact on model predictions. For example, PC and PD are highly influential due to the wide range of SHAP values. In contrast, features with SHAP values that cluster around zero, such as L, have less influence on model predictions. The color coding helps you understand the relationship between the feature values and their influence because it shows how different values of each feature affect the model output. When classifying (predicting) with the partial dependent variable Purchase Intention (Figure 2b), features with a wider range of SHAP values such as PR, SL, VM and PD have a greater influence on the model predictions. Features PR and SL have a wide range of SHAP values, indicating a significant and diverse influence, while features such as PDY and PL have a narrower range, indicating a more consistent and smaller influence. For the partial dependent variable Purchase Share (Figure 1c), the trait SL has a wide range of SHAP values, indicating a significant and diverse influence on the model output. In contrast, features such as PR and L have a narrower range, indicating a more consistent influence. Finally, for the partial dependent variable Willingness to Recommend (Figure 1d), features such as PD and PR have a wide range of SHAP values, indicating that they have a significant and varied influence on the model output. In contrast, features such as SL and PLPR have a narrower range of SHAP values, suggesting a more consistent influence. High PC and PD values (red dots) increase predictions, while low values (blue dots) decrease predictions.

4. Discussion

If we analyze the importance of the features, we come to the conclusion that PDY and SL are crucial for the predictions of the model, while VM and PC also contribute, but to a lesser extent. The features PD, PR, SDA and PL have a medium importance, while PLPR and L are less important. It is recommended to focus on optimizing PDY and SL to improve data quality, as these features have the greatest impact on the model. Price dynamics confirm the need and impact of keeping customers interested in retailers' offers and creating short-term and frequent excitement through price promotions. The service level, reflected in the availability of products and contact with in-store staff, stands out as an important factor. The importance of value for money highlights that customers are evaluating the overall package of the grocery retailer's offering and efforts - what they get for what they pay. In the context of customer loyalty, being fair is much more important than being cheap (price level). Transparency, clarity and visibility of price communication build on these customers' expectations of being fair. Price dispersion is closely related to product range, brand architecture and variety of offerings - customers prefer a diverse range of products in one place, ranging from economy to premium prices and satisfying different needs.

The SHAP analysis provided an insight into the importance and influence of individual features on the model predictions and highlighted attributes such as PC and PD as highly influential due to the wide range of SHAP values. For features such as PLPR and L, further investigation is recommended to identify their potential negative impact on the model or opportunities for improvement. The private label product range is found to be an important attribute for the partially dependent variable Purchase Intention, but less important for Purchase Share and Willingness to Recommend, suggesting that a deeper customer relationship with the retailers' brands is still developing. Examining interactions between the most important features and other variables can also significantly improve the predictive power of the model. This should be prioritized in future analysis to achieve better accuracy and performance of the model. Based on this analysis, it is possible to improve the prediction by:

• data optimization: focusing on improving the quality of PDY and SL features, as these have the greatest impact on the model;

• investigating less important features: further investigating features such as PLPR and L to determine their potential negative influence on the model.

• improving predictive power: investigate interactions between key features and other variables to improve model accuracy and performance.

5. Conclusion

The application of ML algorithms is superior to standard statistical methods in several key aspects, including predictive accuracy and the ability to process complex data. The limitations of this study arise from the geographically limited area from which the respondents are drawn and the non-representative sample. However, the algorithms used in this paper can significantly improve the understanding of consumer behavior patterns and optimize marketing strategies in the context of predicting customer loyalty. In addition, the research has identified the key predictor variables that have the greatest impact on prediction, which can be very helpful in developing effective strategies to promote and maintain customer loyalty. This paper contributes to the literature by demonstrating how ML techniques can be effective in predicting loyalty towards grocery retailers and by highlighting potential areas for future research and improvement.

References

Czakon, J. (2024). *24 Evaluation Metrics for Binary Classification. (n.d.).* Retrieved July 18, 2024, Available online: https://neptune.ai/blog/evaluation-metrics-binary-classification

Arefin, S., Parvez, R., Ahmed, T., Ahsan, M., Sumaiya, F., Jahin, F., & Hasan, M. (2024). Retail Industry Analytics: Unraveling Consumer Behavior through RFM Segmentation and Machine Learning. *2024 IEEE International Conference on Electro Information Technology (eIT)*, Eau Claire, WI, USA, 2024, pp. 545-551, doi: 10.1109/eIT60633.2024.10609927.

Andresini, G., Appice, A., Ardimento, P., Brunetta, A. A., Doronzo, A. G., Ieva, G., & Pasquadibisceglie, V. (2023). CENTAURO: An Explainable AI Approach for Customer Loyalty Prediction in Retail Sector. In: Basili, R., Lembo, D., Limongelli, C., Orlandini, A. (eds) AIxIA 2023 – Advances in Artificial Intelligence. AIxIA 2023. Lecture Notes in Computer Science(), vol 14318. Springer, Cham. doi: 10.1007/978-3-031-47546-7_14.

Bell, D. R. (2014). *Location is (still) everything: The surprising influence of the real world on how we search, shop, and sell in the virtual one*. Houghton Mifflin Harcourt.

Cacchiarelli, L., & Sorrentino, A. (2019). Pricing Strategies in the Italian Retail Sector: The Case of Pasta. *Social Sciences*, 8(4), 113.

Gauri, D. K., Jindal, R. P., Ratchford, B., Fox, E., Bhatnagar, A., Pandey, A., & Howerton, E. (2021). Evolution of retail formats: Past, present, and future. *Journal of Retailing*, 97(1), pp.42-61. doi:10.1016/j.jretai.2020.11.002

Grandi, B. & Cardinali, M. G. (2022). Colours and price offers: How different price communications can affect sales and customers' perceptions. *Journal of Retailing and Consumer Services*, 68, 103073. doi:10.1016/j.jretconser.2022.103073

Flavian, C., Gurrea, R., & Orus, C. (2020). Combining channels to make smart purchases: The role of webrooming and showrooming. *Journal of Retailing and Consumer Services*, 52, 101923. doi:10.1016/j.jretconser.2019.101923

Fosic, I., Zagar, D., & Grgic, K. (2022). Network traffic verification based on a public dataset for IDS systems and machine learning classification algorithms. 2022 45th *Jubilee International Convention on Information, Communication and Electronic Technology* (MIPRO), pp.1037–1041. doi: 10.23919/MIPRO55190.2022.9803674

Hamilton, R., & Chernev, A. (2013). Low prices are just the beginning: Price image in retail management. *Journal of Marketing*, 77(6), pp.1-20. doi:10.1509/jm.08.0204

Hasan, Md. A. M., Nasser, M., Pal, B., & Ahmad, S. (2014). Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS). *Journal of Intelligent Learning Systems and Applications*, 06(01), pp.45–52. doi:10.4236/jilsa.2014.61005

Hu, L.-Y., Huang, M.-W., Ke, S.-W., & Tsai, C.-F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*, 5(1), 1304. doi:10.1186/s40064-016-2941-7

Islamgaleyev, A., Petrova, M., Kurenkeyeva, G., Shalbayeva, Sh., Kadirbergenova, A. 2020. Increasing customer focus in metal trading. *Entrepreneurship and Sustainability Issues,* 8(1), 604-617. http://doi.org/10.9770/jesi.2020.8.1(42)

Kumar, V., Anand, A., & Song, H. (2017). Future of retailer profitability: An organizing framework. *Journal of Retailing*, 93(1), pp.96-119. doi:10.1016/j.jretai.2016.11.003

Kumar, A. i Kim, Y. K. (2014). The store-as-a-brand strategy: The effect of store environment on customer responses. *Journal of Retailing and*

Consumer services, 21(5), pp.685-695. doi: 10.1016/j.jretconser.2014.04.008

Naiem, S., Khedr, A. E., Idrees, A. M., & Marie, M. I. (2023). Enhancing the Efficiency of Gaussian Naïve Bayes Machine Learning Classifier in the Detection of DDOS in Cloud Computing. *IEEE Access*, 11, pp.124597–124608. doi:10.1109/ACCESS.2023.3328951

O'Cass, A., & Grace, D. (2008). Understanding the role of retail store service in light of self-image–store image congruence. *Psychology & Marketing*, 25(6), pp.521-537. doi:10.1002/mar.20223

Popova, P., Petrova, M. (2024). The Role of AI for Smart Environments based on Big Data and IoT Applications," *2024 5th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)*, Veliko Tarnovo, Bulgaria, 2024, pp. 1-8, doi: 10.1109/CIEES62939.2024.10811252

Popova, P., Popov, V., Marinova, K., Petrova, M. and K. Shishmanov. (2024). The Digital Platform — new opportunities and implementation strategy. 1*6th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Iasi, Romania, 2024, pp. 1-9, doi: 10.1109/ECAI61503.2024.10607401

Petrova, M., Popova, P., Popov, V., Shishmanov, K. and K. Marinova. (2022). Potential of Big Data Analytics for Managing Value Creation. 2022 International Conference on Communications, Information, Electronic and Energy Systems (CIEES), 2022, pp. 1-6, DOI: 10.1109/CIEES55704.2022.9990882

Rane, N. (2023). Enhancing customer loyalty through Artificial Intelligence (AI), Internet of Things (IoT), and Big Data technologies: improving customer satisfaction, engagement, relationship, and experience. *Internet of Things (IoT), and Big Data Technologies: Improving Customer Satisfaction, Engagement, Relationship, and Experience* (October 13, 2023). doi:10.2139/ssrn.4616051

Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3(2). doi: 10.1007/s42452-021-04148-9

Scikit-learn. (n.d.). *Decision Trees*. Retrieved July 18, 2024, Available online: https://scikit-learn.org/stable/modules/tree.html

Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised Classification Algorithms in Machine Learning: A Survey and Review. *In: Mandal, J., Bhattacharya, D. (eds) Emerging Technology in Modelling and*

Graphics. Advances in Intelligent Systems and Computing, vol 937. Springer, Singapore. doi: 10.1007/978-981-13-7403-6_11.

Stapor, K. (2018). Evaluating and Comparing Classifiers: Review, Some Recommendations and Limitations. *In: Kurzynski, M., Wozniak, M., Burduk, R. (eds) Proceedings of the 10th International Conference on Computer Recognition Systems CORES 2017. CORES 2017. Advances in Intelligent Systems and Computing*, vol 578. Springer, Cham. https://doi.org/10.1007/978-3-319-59162-9_2.

Tian, Y., Zhang, Y., & Zhang, H. (2023). Recent advances in stochastic gradient descent in deep learning. *Mathematics*, 11(3), 682. doi: 10.3390/math11030682

Wang, H., Liang, Q., Hancock, J.T. & Khoshgoftaar, T.M. (2024). Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*,*11*(1), 44. doi:10.1186/s40537-024-00905-w

Wang, X. S., Ryoo, J. H. J., Bendle, N., & Kopalle, P. K. (2021). The role of machine learning analytics and metrics in retailing research. *Journal of Retailing*, 97(4), pp.658-675. doi: 10.1016/j.jretai.2020.12.001

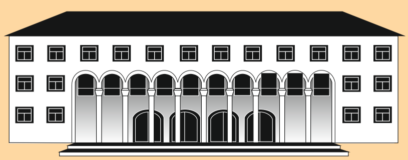
Zhou, Q., Qi, Y., Tang, H., & Wu, P. (2023). Machine learning-based processing of unbalanced data sets for computer algorithms. *Open Computer Science*, 13(1). doi: 10.1515/comp-2022-0273

Zielke, S. (2006). Measurement of retailers' price images with a multiple-item scale. *The International Review of Retail, Distribution and Consumer Research*, 16(3), pp. 297-316. doi:10.1080/09593960600696990

ISSN 0861 - 6604
ISSN 2534 - 8396

BUSINESS management

BUSINESS management 2/2025



PUBLISHED BY
D. A. TSENOV ACADEMY
OF ECONOMICS - SVISHTOV

2/2025

The printing of the issue 2-2025 is funded with a grand from the Scientific Research Fund, Contract KP-06-NP6/29/04.12.2024 by the competition "Bulgarian Scientific Periodicals - 2025".

Submitted for publishinq on 27.06.2025, published on 30.06.2025, format 70x100/16, total print 80

BUSINESS management

D. A. Tsenov Academy of Economics, Svishtov

Year XXXV * Book 2, 2025

CONTENTS

MANAGEMENT practice