Proceedings of CONF-APMM 2026 Symposium: Applied Mathematics and Quantitative Modeling
DOI: 10.54254/2753-8818/2026.PU37983

Comparison of Machine Learning Models under the Background of the 2013 University of California, Irvine Wholesale Customers Dataset

Peilin Lyu

Department of Statistic, University of Wisconsin- Madison, Madison, USA
plyu22@wisc.edu

Abstract. Customer spending patterns can provide useful information for identifying the channels of wholesale customers, supporting customer management and marketing decisions. This study examines whether six annual spending categories can be used to classify customers by sales channels. Using the University of California, Irvine (UCI) Wholesale Customers dataset, which contains 440 customers, four classification models, including Logistic Regression, K-Nearest Neighbors, Random Forest, and Gradient Boosting, were compared using Accuracy, Precision, Recall, F1-score, and Receiver Operating Characteristic - Area Under the Curve (ROC-AUC). Stratified five-fold cross-validation was used to report mean and standard deviation. Feature subsets, Channel and Region prediction, and K-Means clustering were also compared. Logistic Regression, Random Forest, and Gradient Boosting showed similar performance, and K-Nearest Neighbors (KNN) performed worse. Using two or three important features produced performance close to that of all six features. Channel was easier to predict than Region, and K-Means showed moderate agreement with the true Channel labels (ARI = 0.5135). These findings show that a small number of spending features can provide useful information for customer channel classification.

Keywords: customer Channel Classification, wholesale Spending, feature Importance, model Comparison, cluster Validation

1. Introduction

1.1. Research background and research gap

Customer spending patterns can provide useful information for understanding customer groups and supporting marketing and customer management. The UCI Wholesale Customers dataset has been used in previous studies for customer segmentation and machine learning analysis. For example, previous work has applied K-Means clustering to this dataset to identify customer groups based on purchasing patterns, while other work has focused on predictive analysis of wholesale customer purchasing patterns using machine learning methods [1, 2]. These studies show that the dataset can be used for both customer segmentation and prediction. However, existing studies often focus on

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.PU37983

one main analytical approach rather than combining several types of analysis in one study. Therefore, there is limited research that compares several classification models for Channel prediction, tests whether a small number of important features can provide similar performance to all features, directly compares Channel and Region prediction, and validates unsupervised clusters against the actual Channel labels. This study addresses this gap by combining these analyses in the same dataset.

1.2. Research questions and methods

This study uses the Wholesale Customers dataset from the UCI Machine Learning Repository. The dataset contains 440 customers from a wholesale distributor and records annual spending in six product categories: Fresh, Milk, Grocery, Frozen, Detergents_Paper, and Delicassen. The dataset has no missing values [3]. The main target is Channel, which contains two groups, Horeca (Hotel, Restaurant, and Café) and Retail. Region, consisting of Lisbon, Oporto, and Other, is used as a second prediction target.

The six spending variables are right-skewed, so log1p transformation is applied before analysis. The Channel classes contain 298 Horeca and 142 Retail customers. Class weights are used to address this mild imbalance, avoiding oversampling or SMOTE. Four models are compared: Logistic Regression, KNN, Random Forest, and Gradient Boosting. Logistic Regression and KNN are standardized using a training-only procedure.

The four models are evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Both a stratified 80/20 train-test split and stratified five-fold cross-validation are conducted. Random Forest feature importance is then used to select the top two and top three features and compare them with the full six-feature model. The same six features are also used to predict Region and are compared with a majority-class baseline. Finally, K-Means with two clusters is applied without using Channel labels, and the resulting clusters are compared with the true Channel labels by reporting Adjusted Rand Index (ARI).

1.3. Research objectives and significance

This study has three main objectives. First, it compares Logistic Regression, KNN, Random Forest, and Gradient Boosting for Channel classification using five evaluation values and cross-validation. Second, it measures whether two or three important spending categories can achieve performance close to the model with all six features. Third, it compares the predictability of Channel and Region using the same features and uses K-Means to examine whether customer spending patterns naturally form two groups that correspond to the actual Channel labels. Together, these analyses provide a more complete view of how the wholesale spending dataset can be used for customer classification and segmentation.

2. Methodology

2.1. Data

The dataset used in this study is the Wholesale Customers dataset from the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/292/wholesale+customers, DOI: 10.24432/C5030X). This dataset contains real annual spending records collected from a Portuguese wholesale distributor

DOI: 10.54254/2753-8818/2026.PU37983

[3]. It includes 440 customer records and contains no missing values. Each record represents one customer's yearly spending behavior across different product categories.

2.2. Variables

Here are 6 numerical variables and 2 categorical variables in this dataset. The dataset describes customers based on their annual spending amount in six product categories, which were used as the input features: Fresh, Milk, Grocery, Frozen, Detergents_Paper, and Delicassen. Delicassen is the original column name used in the UCI dataset and is commonly written as "Delicatessen." These six numerical variables describe customers' purchasing patterns. Table 1 gives more useful statistics of these features, such as median and standard deviation. For Channel classification, only the six annual spending variables were used as input features. Region was not used as an input feature because it was treated as a separate target variable in the additional Region classification analysis.

Table 1. Useful statistics of features

Category	Minimum	Maximum	Mean	Median	Std. Deviation
Fresh	3	112151	12000.30	8504.00	12647.33
Milk	55	73498	5796.27	3627.00	7380.38
Grocery	3	92780	7951.28	4755.50	9503.16
Frozen	25	60869	3071.93	1526.00	4854.67
Detergents_Paper	3	40827	2881.49	816.50	4767.85
Delicassen	3	47943	1524.87	965.50	2820.11

Data from: UCI Machine Learning Repository [3]

The main target variable in this study is Channel. It represents two types of customers: Horeca (Hotel/Restaurant/Café) customers and Retail customers. In this dataset, Channel was recorded as a binary variable, where Horeca was assigned as 1, and Retail was assigned as 2. The Channel variable contains 298 Horeca customers and 142 Retail customers, which is slightly imbalanced.

In addition, the dataset contains another categorical variable, Region, which includes three groups: Lisbon (77 customers), Oporto (47 customers), and Other regions (316 customers). Region is used as a secondary comparison target in this study for comparison, which helps to examine whether the same spending features can predict where customers come from and to compare the difficulty between predicting customer type and customer region.

The original spending variables have large differences in scale and highly skewed distributions. As shown in Figure 1, several spending categories contain extreme values because some customers spend much more than average customers. Therefore, additional preprocessing steps are required before model training.

DOI: 10.54254/2753-8818/2026.PU37983

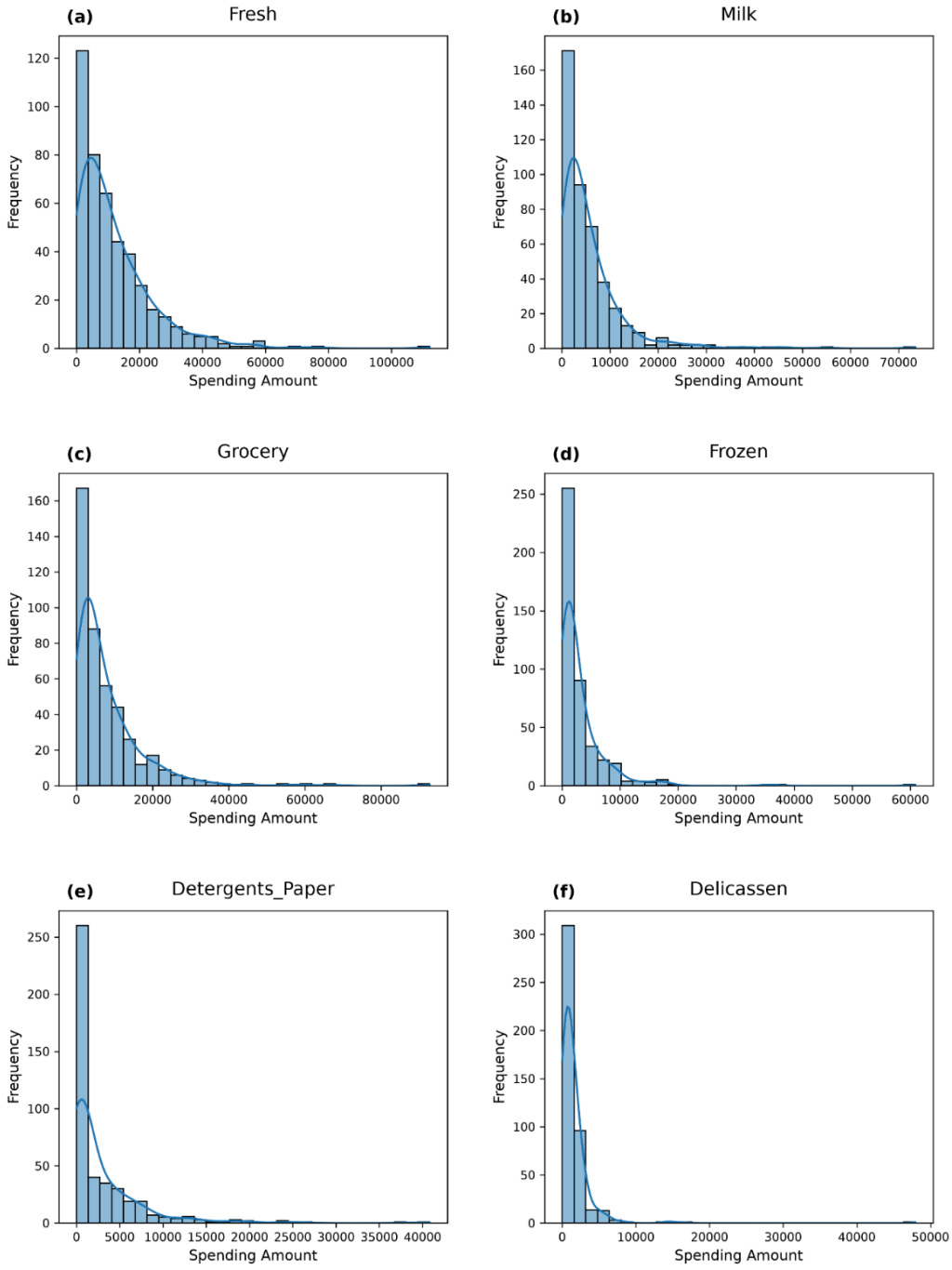

Figure 1. Frequent graph of distribution of features with fitted lines (a) fresh spending; (b) milk spending; (c) grocery spending; (d) frozen spending; (e) detergents_paper spending; (f) delicassen spending (picture credit: original. Data from: UCI Machine Learning Repository [3])

DOI: 10.54254/2753-8818/2026.PU37983

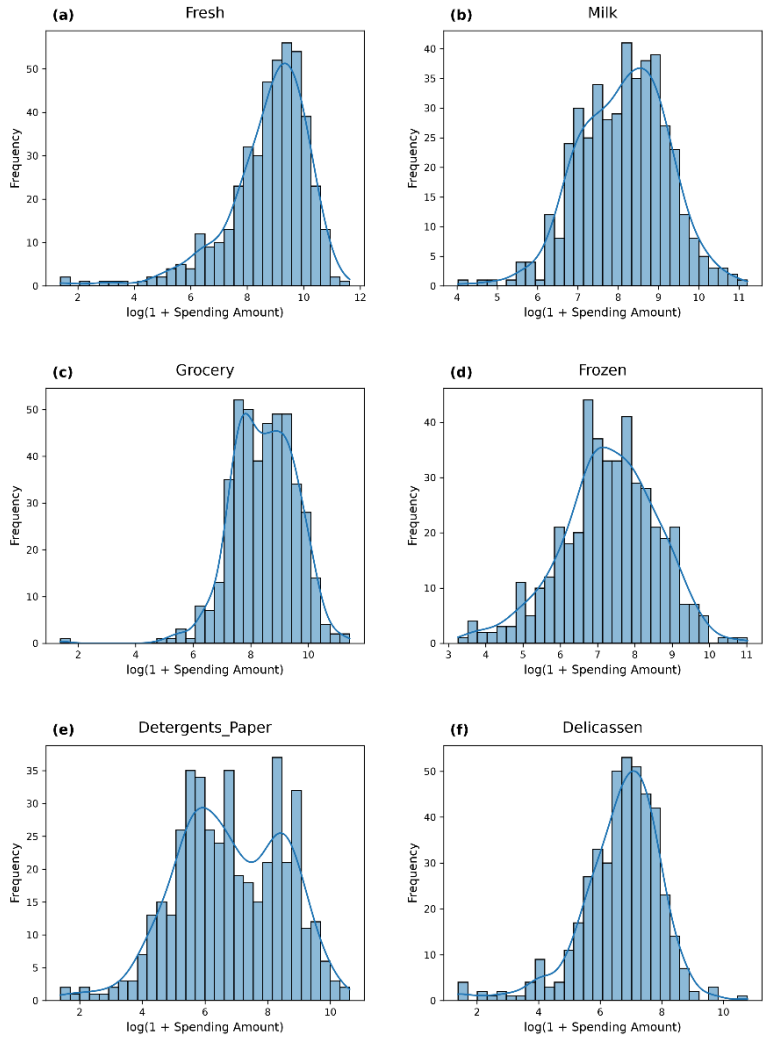

Figure 2. Frequent graph of distribution of features with fitted lines (after log- transformation) (a) log-transformed fresh spending; (b) log-transformed milk spending; (c) log-transformed grocery spending; (d) log-transformed frozen spending; (e) log-transformed detergents_paper spending; (f) log-transformed delicassen spending (picture credit: original. Data from: UCI Machine Learning Repository [3])

The spending categories also showed different levels of correlation with each other. In particular, as shown in Figure 3, Grocery and Detergents_Paper had a very strong positive correlation of 0.92, indicating that these two variables contained highly overlapping information. This figure was calculated using the original spending amounts before the log transformation. It was used to describe the original relationships among the six spending categories rather than the feature values directly used in model training. This strong correlation suggested possible multicollinearity, which was considered when interpreting the Logistic Regression results.

DOI: 10.54254/2753-8818/2026.PU37983

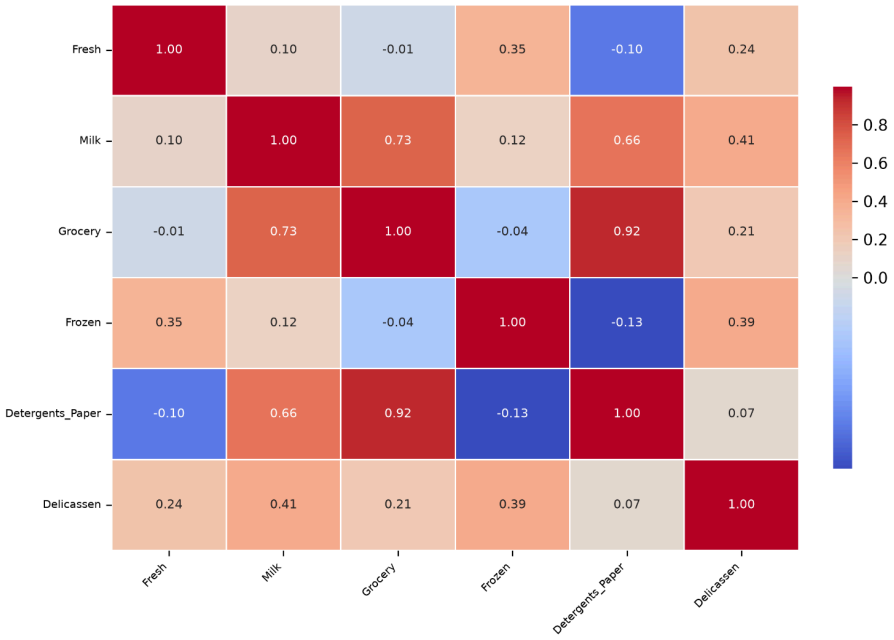

Figure 3. Correlation heatmap of spending categories (picture credit: original. Data from: UCI Machine Learning Repository [3])

Moreover, as shown in Figure 4, the spending patterns differed between Horeca and Retail customers across several product categories. These differences provided an initial indication that the six spending variables could contain useful information for distinguishing the two customer groups.

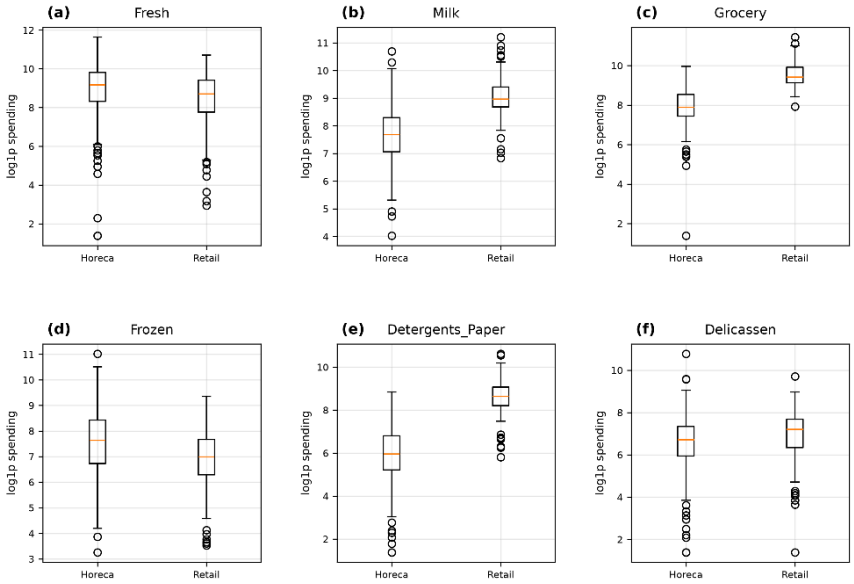

Figure 4. Box graph of log-transformed spending distribution for different channels, with outliers (a) fresh spending by channel; (b) milk spending by channel; (c) grocery spending by chenal; (d) frozen spending by channel; (e) detergent_paper spending; (f) delicassen spending by channel (picture credit: original. Data from: UCI Machine Learning Repository [3])

DOI: 10.54254/2753-8818/2026.PU37983

2.3. Processing

For more convenient modeling, the variable Channel is recoded, where Retail is coded as 1, and Horeca is coded as 0.

First, the six spending variables were transformed using the np.log1p() function, which calculates log(1+x). As shown in Figure 1, the original spending distributions are strongly right-skewed, while the transformed distributions become more balanced (see Figure 2). Log transformation reduces the influence of extremely large spending values and allows models to better capture general customer patterns. Instead of directly making log calculation, the function np.log1p ( ) was selected because it can safely handle zero values in the dataset while maintaining information from small spending values.

Second, the channel variable is slightly imbalanced. Instead of generating artificial samples using oversampling methods such as SMOTE, class_weight="balanced" was used for Logistic Regression and Random Forest. This is because the imbalance is not extremely severe, and make adjustment on the importance of each class is sufficient (more importance for Channel = Retail, which is the smaller class). This approach avoids creating artificial customer records and keeps the model training process based only on real observations from the original dataset.

Third, the dataset was divided into training and testing sets using an 80/20 stratified split. The training set contained 352 observations, while the test set contained 88 observations. The stratify parameter was used during splitting to maintain approximately the same proportion of Horeca and Retail customers in both sets as in the original dataset

Fourth, feature standardization was applied to the models that were sensitive to feature scales, including Logistic Regression and K-Nearest Neighbors. The mean and standard deviation used for standardization were only calculated based on the training set. These statistics were then used in the standardization of both the training set and the testing set. Therefore, both sets were standardized without using any information from the test set. This procedure prevented information from the test set from being used during model training and avoided data leakage.

2.4. Models

Four supervised machine learning models were selected for Channel classification: Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and Gradient Boosting.

Logistic Regression is used as the basic linear classification model. It provides a simple baseline to test whether the relationship between spending features and customer Channel can be separated by a linear boundary. Logistic Regression is commonly used in customer classification problems. For example, Zhu applied Logistic Regression and other models to customer churn prediction and demonstrated that this is the model that performs the best [4].

KNN is a distance-based model that classifies customers by comparing them with their closest neighbors. This is widely used in customer classification. Othman, for example, applied KNN to customer segmentation in a telecommunications company and demonstrated that customer characteristics could be used to classify customers into different segments [5].

Random Forest is an ensemble learning method that combines multiple decision trees and makes predictions through voting. It can capture nonlinear relationships between customer spending patterns and classification results. In addition, Random Forest provides feature importance values, which help identify the product categories that contribute most to customer classification. Random Forest has been widely applied in customer analytics. For example, Boozary et al. demonstrated that

DOI: 10.54254/2753-8818/2026.PU37983

tree-based ensemble models can improve customer classification performance compared with traditional methods [6].

Gradient Boosting is another tree-based ensemble method. Unlike Random Forest, which builds trees independently, Gradient Boosting creates trees sequentially, where each new tree focuses on correcting the errors of previous models. This allows the model to learn complex patterns in customer data. Gradient Boosting has been successfully applied in customer behavior prediction tasks. Lessmann et al. compared multiple machine learning methods for credit scoring and showed that boosting-based models achieved strong predictive performance in customer-related classification problems [7].

In addition, K-Means clustering is used as an unsupervised learning comparison method. Because the Channel variable contained two customer groups, K-Means was set to 2. The clustering was performed using the standardized log-transformed values of the six spending features. Unlike supervised classification models, K-Means does not use Channel labels during training. Instead, it automatically groups customers based on similarities in their spending behaviors. Customer segmentation studies frequently apply K-Means because it can discover natural customer groups without predefined labels. Wu and Lin applied clustering methods to customer segmentation and showed that unsupervised learning can reveal meaningful customer patterns [8].

For the additional Region classification analysis, Random Forest was selected from the four supervised models because it can capture nonlinear relationships and is less sensitive to feature scaling. This experiment was designed to examine whether the six spending variables could also predict customer region. The comparison provided additional evidence about whether spending patterns were more useful for predicting customer type or geographic region.

2.5. Evaluation

The performance of the supervised classification models was evaluated using five metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC. Retail was treated as the positive class, while Horeca was treated as the negative class. Accuracy measured the overall proportion of correctly classified customers. Precision measured the proportion of customers predicted as Retail who were actually Retail, while Recall measured the proportion of actual Retail customers that were correctly identified. F1-score provided a balance between Precision and Recall. ROC-AUC measured the model's ability to distinguish between Retail and Horeca customers across different classification thresholds.

As for the unsupervised model, K-means clustering, the resulting two clusters were compared with the actual Horeca and Retail groups using the Adjusted Rand Index (ARI). An ARI of 1 indicates complete agreement between the clustering and the true grouping, while an ARI around 0 indicates agreement similar to that expected from random grouping.

In addition to the single 80/20 train-test split, stratified 5-fold cross-validation was performed to obtain a more stable comparison of the four supervised models. In each fold, the training portion was used to fit the preprocessing steps and the model, while the testing portion was used only for evaluation. Standardization was included inside a Pipeline so that the mean and standard deviation were calculated separately from the training portion for each fold and then used to transform both the training and testing portions. Standardizing the full dataset before cross-validation would allow information from the validation folds to enter the preprocessing step and could lead to data leakage and overly optimistic performance estimates. Therefore, each fold was fit separately, using only its own information. For each model, the mean and standard deviation of Accuracy, Precision, Recall, F1-score, and ROC-AUC across the five folds were reported.

DOI: 10.54254/2753-8818/2026.PU37983

A single 80/20 train-test split was not considered sufficient for comparing model performance because the results could depend strongly on the particular 88 observations included in the test set. The stratified 5-fold cross-validation results therefore provided an additional and more stable basis for model comparison by evaluating each model across multiple training and validation splits. The mean performance and standard deviation across the five folds were used to assess both predictive performance and variation across folds.

In addition, a feature subset comparison was conducted using the Random Forest feature importance results. The models were retrained using the top three and top two most important spending variables, respectively, and their performance was compared with the model using all six spending variables. The comparison focused on how much classification performance changed when less important features were removed.

For the additional Region classification analysis, Accuracy was reported because the three Region classes were strongly imbalanced. The dataset contained 316 customers from Other regions, compared with 77 from Lisbon and 47 from Oporto. Therefore, a majority-class baseline was also calculated by assigning every customer to the Other class. This baseline achieved an Accuracy of approximately 72%. Comparing the classification models with this baseline showed whether the models provided useful predictive information beyond simply predicting the largest class.

2.6. Reproducibility

To improve reproducibility, the same random seed was used for the train-test split, K-Means, Random Forest, and Gradient Boosting models. The random state was set to 42. The main model hyperparameters were also recorded. Logistic Regression used max_iter= 1000 (since the dataset is relatively small). The number of neighbors for KNN is 3. An odd value was selected to avoid ties between the two classes in binary classification. The value of k was selected as a commonly used small odd value rather than being optimized through an additional tuning procedure. The number of trees for Random Forest is 100 (also due to the small dataset). K-Means used n-cluster = 2 since there are 2 channels, and the same random state of 42. The Python version is 3.12.10, and the scikit-learn version is 1.9.0.

3. Results

3.1. Descriptive statistics

The correlation heatmap in Figure 3 shows the relationships among the six spending categories. Among these variables, Grocery and Detergents_Paper show a very strong positive correlation, indicating that their spending values tend to increase together. The other spending categories show weaker relationships compared with these two variables.

Figure 4 presents the log-transformed spending values for the six product categories, grouped by Channel. The differences between Horeca and Retail customers appear more noticeable for Grocery and Detergents_Paper than for the other categories. In both of these categories, the Retail group generally has higher values than the Horeca group. The distributions also show some variation within each customer group (the two groups having different ranges and central patterns).

DOI: 10.54254/2753-8818/2026.PU37983

3.2. Four model comparison

Table 2 compares the performance of the four classification models using Accuracy, Precision, Recall, F1-score, and ROC-AUC. The results from the single 80/20 train-test split and the stratified 5-fold cross-validation are presented side by side. Based on the results, Logistic Regression, Random Forest, and Gradient Boosting show similar overall performance, while K-Nearest Neighbors has the lowest performance.

Table 2. Resulting values of four models (single 80/20 split)

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.920	0.839	0.929	0.881	0.952
KNN	0.909	0.857	0.857	0.857	0.932
Random Forest	0.932	0.893	0.893	0.893	0.946
Gradient Boosting	0.943	0.897	0.929	0.912	0.935

Data from: UCI Machine Learning Repository [3]

Table 3. Resulting values of four models (stratified five-fold cross-validation, mean ± standard deviation)

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.909 ± 0.024	0.814 ± 0.050	0.937 ± 0.033	0.870 ± 0.030	0.965 ± 0.009
KNN	0.905 ± 0.018	0.828 ± 0.045	0.894 ± 0.045	0.858 ± 0.024	0.935 ± 0.016
Random Forest	0.930 ± 0.013	0.870 ± 0.040	0.923 ± 0.025	0.895 ± 0.017	0.957 ± 0.006
Gradient Boosting	0.927 ± 0.014	0.881 ± 0.049	0.902 ± 0.034	0.890 ± 0.016	0.954 ± 0.014

Data from: UCI Machine Learning Repository [3]

Figure 5 presents the 5-fold cross-validation results with error bars showing the standard deviation across the five folds. The F1-score error bars of all four models overlap, indicating that the differences in F1-score are relatively small and the four models have similar overall classification performance. For ROC-AUC, the error bars of Logistic Regression, Random Forest, and Gradient Boosting overlap with each other, while the KNN error bar is lower than those of Logistic Regression and Random Forest. This suggests that KNN has a weaker ability to distinguish between the two Channel groups than the other models.

The model ranking is consistent between the single train-test split and the 5-fold cross-validation results. KNN has the lowest F1-score and ROC-AUC in both evaluations. This consistency suggests that the comparison is not strongly affected by the particular train-test split.

DOI: 10.54254/2753-8818/2026.PU37983

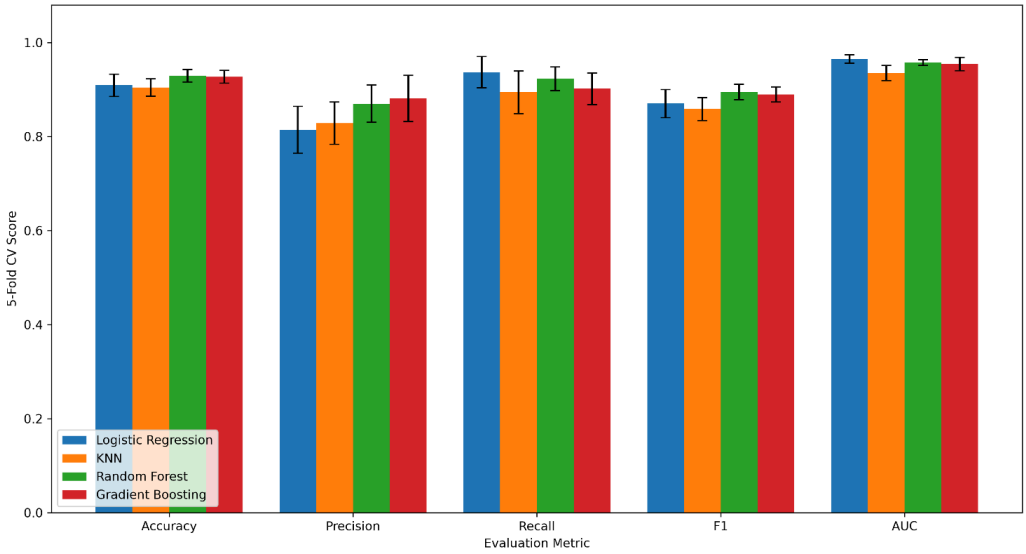

Figure 5. Bar graph of resulting values of four models, with error bars (picture credit: original. Data from: UCI Machine Learning Repository [3])

3.3. Feature subsets and dual-target comparison

Figure 6 shows the feature importance results from the Random Forest model. The three most important features are Detergents_Paper, Grocery, and Milk, respectively. Figure 7 compares the performance of Random Forest models using all six spending features with models using only the top two and top three features. The results show that reducing the number of features to two or three causes little change in model performance. The error bars for the three feature sets overlap across all of the values, suggesting that their performance is similar. Therefore, the reduced feature sets can achieve performance close to that of the model using all six features.

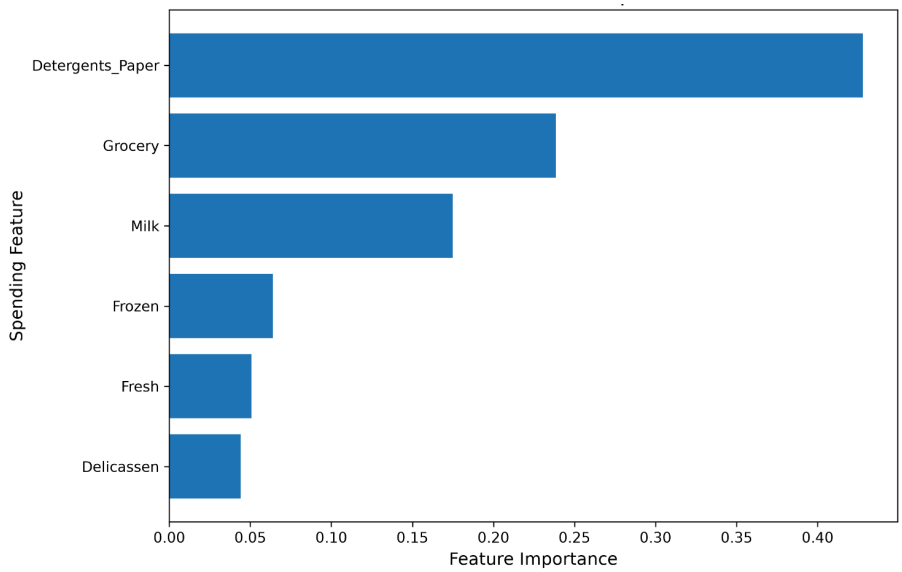

Figure 6. Bar graph of feature importance from Random Forest (picture credit: original. Data from: UCI Machine Learning Repository [3])

DOI: 10.54254/2753-8818/2026.PU37983

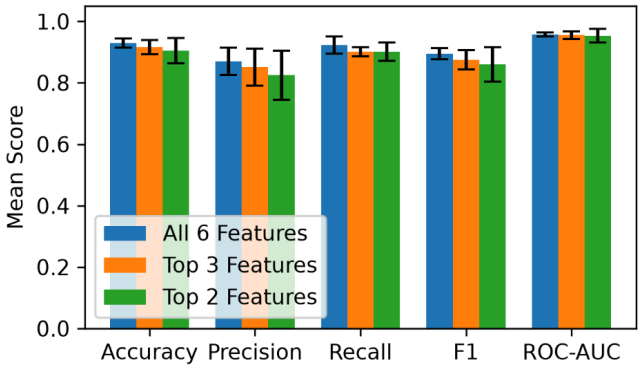

Figure 7. Bar graph of resulting values of four models with different number of features (four models with all features, top 3 important features, and top 2 important features), with error bars (using Random Forest) (picture credit: original. Data from: UCI Machine Learning Repository [3])

Figure compares the Channel classification accuracy, Region classification accuracy, and the Region majority-class baseline. Using the same spending features, the model achieves higher accuracy when predicting Channel than when predicting Region, indicating that Channel is easier to predict from the consumption data. The mean accuracy of the Region model, 0.689, is only slightly lower than the majority-class baseline, which is 0.718. The difference between the two means is approximately one standard deviation (SD = 0.029), showing that the Region model's prediction is even worse than the result when simply predicting the majority class.

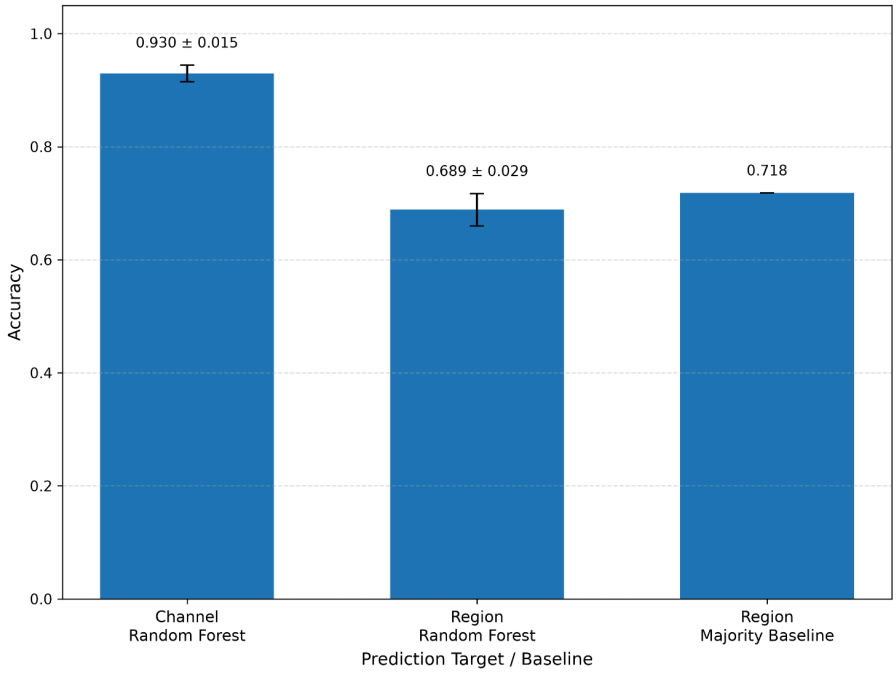

Figure 8. Bar graph of accuracy of models with different target and a majority baseline, with error bars (using Random Forest) (picture credit: original. Data from: UCI Machine Learning Repository [3])

Figures 9 and 10 present the PCA-based two-dimensional visualization of the customer data before and after the K-means classification. Figure 9 shows the original data in the two-dimensional PCA space, while Figure 10 shows the data after classification. The results suggest that the

DOI: 10.54254/2753-8818/2026.PU37983

consumption structure can naturally form two groups even without using the Channel labels. The two groups also show a general correspondence with the actual Channel categories. In particular, there is a somewhat clear difference between the two groups in the middle area of the scatter plot. The K-Means clustering results give an Adjusted Rand Index (ARI) of 0.5135 when compared with the true Channel labels. This indicates a moderate level of agreement between the two unsupervised clusters and the actual Channel categories.

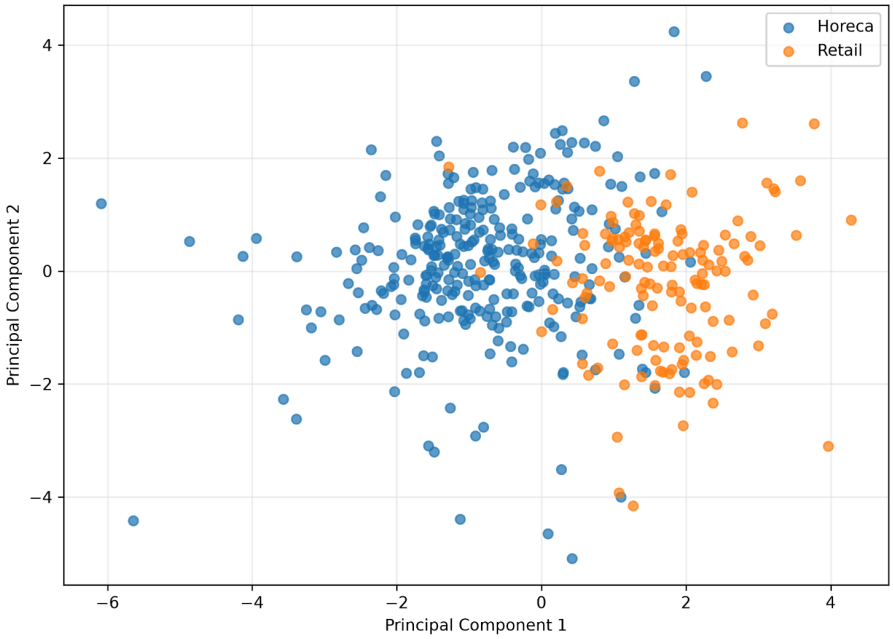

Figure 9. PCA dot plot of original data (after log-transformation) (picture credit: original. Data from: UCI Machine Learning Repository [3])

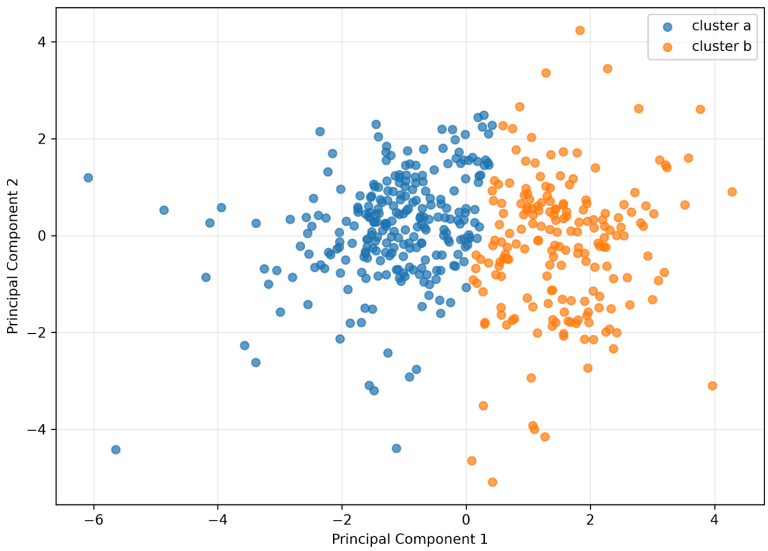

Figure 10. PCA dot plot of result of K-means (picture credit: original. Data from: UCI Machine Learning Repository [3])

DOI: 10.54254/2753-8818/2026.PU37983

4. Discussion

4.1. Interpretation of results

The results suggest that Logistic Regression, Random Forest, and Gradient Boosting are suitable models for the Channel classification task. The similar performance of the three models suggests that a more complex nonlinear model does not provide a large additional benefit for this dataset. This is probably because there is no complex nonlinear information in this dataset, which leads simpler linear models such as logistic regression and more complex models to have similar performance. Also, based on Figures 9 and 10, it is convincing that KNN does not perform well in distinguishing the two channels, since KNN is a neighbor-based model and there is clear overlap between groups in both PCA scatter plots. However, the strong correlation between Grocery and Detergents_Paper indicates that some features contain overlapping information. Therefore, individual Logistic Regression coefficients should not be interpreted as direct evidence that one product category is more important than another.

The feature subset results provide further evidence that much of the information needed for Channel classification is concentrated in a small number of features. Detergents_Paper, Grocery, and Milk were the three most important features in the Random Forest model, and the Top 2 and Top 3 feature models achieved performance close to that of the model using all six features. This suggests that the remaining features add relatively limited predictive information for this classification task. This finding is consistent with previous purchase-based segmentation research showing that product-specific purchasing and monetary information can be useful for identifying customer groups [9].

Channel was easier to predict than Region using the same spending features. This may be because Channel is more directly related to purchasing behavior, whereas Region may depend on additional customer or geographic characteristics that are not included in the dataset. The Region model also did not perform better than the majority-class baseline, suggesting that the six spending variables contain limited information for distinguishing the three regions.

The K-Means result provides another perspective. The ARI of 0.5135 indicates moderate agreement between the two unsupervised clusters and the actual Channel labels. This suggests that some Channel-related structure exists naturally in the spending data, even without using the Channel labels. Similar customer segmentation studies have shown that purchasing behavior can be used to identify meaningful customer groups, while adding other customer characteristics can further improve segmentation [10].

4.2. Limitation and further work

This study has several limitations. First, the dataset comes from only one wholesale distributor and represents only one year of customer spending. Therefore, the results may not generalize to other companies, industries, or time periods. Second, the dataset contains only 440 customers, which limits the amount of information available for model training and validation. Third, the Region classes are highly imbalanced, with 316 customers in Other compared with only 47 in Oporto, making Region classification more difficult to evaluate fairly. Finally, the dataset contains only six spending categories, and some other important customer characteristics may not be included.

Future research could add more customer attributes, such as demographic information, purchasing frequency, and other behavioral variables. Previous research has found that combining purchasing information with demographic attributes can improve customer segmentation results

DOI: 10.54254/2753-8818/2026.PU37983

[10]. Future studies could also validate the models on multiple retail or wholesale datasets to test whether the observed patterns remain consistent across businesses and markets. Purchase-based segmentation research has shown the value of using detailed transaction information for identifying customer groups [9]. Finally, future research could use more detailed definitions of customer channels rather than the current two-channel classification. Research on customer segmentation has shown that purchase channels can be combined with demographic information to identify more detailed customer segments [11].

5. Conclusion

Logistic Regression, Random Forest, and Gradient Boosting performed similarly in predicting customer channels, while KNN showed weaker performance across the evaluation metrics. The similar results among the three stronger models suggest that the relationship between customer spending patterns and customer groups may not require a highly complex classification method to achieve good predictive performance. In contrast, the lower performance of KNN indicates that a distance-based method may be less effective for this dataset, possibly because some customer groups have overlap on spending patterns. Feature selection also showed that using fewer variables did not substantially reduce predictive performance. In particular, using the top two or three features, Detergents_Paper, Grocery, and Milk, achieved results close to those obtained using all six spending features. This suggests that these features contain important information for distinguishing customer groups and that a smaller feature set may be sufficient for practical classification.

The results also showed that Channel was easier to predict than Region. This indicates that the spending variables contained clearer information about the two Channel groups than about the different Region groups. In addition, K-Means clustering showed moderate agreement with the real Channel groups, with an ARI of 0.5135. This result suggests that unsupervised clustering could identify some of the same customer patterns as the known Channel labels, although the agreement was not strong enough to fully reproduce the original groups. Overall, these findings can help wholesalers better understand and manage different customer groups based on their spending patterns. Future studies could improve the analysis by including more detailed customer information, using larger and more diverse datasets, and developing more detailed Channel categories rather than relying only on the current two-channel classification.

References

[1] Phudjashakty, N. T., Firmansyah, H., Asriyani, W., & Sofyan, A. (2026). Segmentasi pelanggan grosir menggunakan K-Means: Analisis outlier dan ketidakseimbangan data. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan, 4*(3), 3057–3066.

[2] Ranu, R., Pal, R., & Mishra, N. K. (2025). Predictive analysis of wholesale customer purchases pattern using machine learning techniques. *International Journal of Applied and Behavioral Sciences, 2*(1), 352–365.

[3] Cardoso, M. (2013). Wholesale customers [Dataset]. *UCI Machine Learning Repository*. https://doi.org/10.24432/C5030X

[4] Zhu, Z. (2023). Customer churn prediction based on big data and machine learning approaches. In *Proceedings of the 8th International Conference on Financial Innovation and Economic Development* (pp. 4–15).

[5] Othman, I. A. (2022). A machine learning model for customer segmentation in a telecom company using K-Nearest Neighbor (KNN). *International Journal of Computer Applications, 184*(31), 36–42.

[6] Boozary, P., Sheykhan, S., GhorbanTanhaei, H., & Magazzino, C. (2025). Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction. *Intelligent Systems with Applications, 27*, 200529.

[7] Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124–136.

DOI: 10.54254/2753-8818/2026.PU37983

[8] Wu, J., & Lin, Z. (2005). Research on customer segmentation model by clustering. *Proceedings of the 7th International Conference on Electronic Commerce (ICEC '05)*, 316–318.
[9] Tsai, C.-Y., & Chiu, C.-C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications, 27*(2), 265–276.
[10] Sarvari, P. A., Ustundag, A., & Takci, H. (2016). Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. *Kybernetes, 45*(7), 1129–1157.
[11] Kondo, F. N., & Okubo, T. (2022). Understanding multi-channel consumer behavior: A comparison between segmentations of multi-channel purchases by product category and overall products. *Journal of Retailing and Consumer Services, 64*, 102792.