To catch a customer: Comparing two models of classification and clustering-then-classification in predicting online customers based on browsing behaviors.

Khanh Chu, Stijn Huitenga and Jean Huijnen

Erasmus University Rotterdam
Burgemeester Oudlaan 50, 3062 PA
Rotterdam, the Netherlands

{554897cn, 588352sh, 547401jh}@student.eur.nl

Abstract. The uprising of E-commerce allows for much data about customer behaviors – which businesses can use to devise strategies that encourage customer loyalty. Our research aims to optimize and compare the performance of two classification models: A general random forest algorithm and an algorithm that categorizes visitors into clusters and then applies random forest. Both strive to predict whether a visitor (visit session) ended up as a customer with purchases, emphasizing good accuracy, precision rate, and F1-measure. Results show that the cluster-then-classify model performed marginally worse than the generic classifying model on all metrics, but still comparable to previous studies. Clustering-then-classify model also offers more insights into visitors than the generic model and is more applicable in realistic situations, considering data concerns. The generic model is excellent for predicting customers' purchase intentions quickly, especially with unseen records.

Keywords: customer segmentation, customer relationship management (CRM), feature selection, models comparison, clustering-then-classify algorithm.

1 Introduction

With the development of the Internet in recent years, more and more customers are turning to online shopping for convenience. A new world of shopping with many advantages opens up online – the Internet made it easier to find information about products, compare them for cheaper options, and visit many outlets. As such, electronic commerce (E-commerce) has become one of the largest fields in the electronics and Marketing industries. This is evident in an extreme case recently: while COVID-19 prohibited offline shopping, E-commerce thrived and sustained the commerce industry, jumping from 15% to 21~22% of total sales over two years [10].

E-commerce, along with computer science advancements, notably data mining, has allowed businesses to capture explicit and implicit customer data – something impossible to obtain with only human labor. With the data, enterprises are informed about devising strategies that best serve their purpose, whether acquiring or retaining customers - the basis of customer relationship management (CRM). A company that captures customer data well to inform strategies can easily gain advantages over its competitor [11]. As such, improving ways of capturing customer data – partly consisting of improving algorithms that gather insights – is a continuous task.

Various data mining techniques have long been applied to gather customer data, including classifying and clustering models. However, few, if any, have tried to combine both tasks. Since customer grouping might be able to help with the profitability prediction, we are curious whether an algorithm that cluster-then-classify a visitor could enhance this prediction. It would only be worth building and applying this model if it performs marginally better than a pure classification model. Within the scope of our research, we will conduct an experiment where the performance of two models on the same dataset is compared in an E-commerce context. Our research questions will be: What are the performances of two classification models - one general random forest and one clustering-then-random forest - in predicting E-commerce customer purchases?

Through each step of our research, we will conduct a data mining process – some of which also yield insights for the business. Namely, through the clustering information, we can build personas that represent our target customer groups [2]. When optimizing random forests, the attribute selection process also suggests the customers' most representative attributes (or customer groups). The results of these tasks will be the answers to our sub-questions: based on our dataset, we want to know what the groups of customers are, and which attributes best provide insights about them.

Through our research, we hope to offer an example of best practice for (e)-CRM managers when building a model to capture customer data. We also want to add to the discourse around clustering-then-classification machine learning models, which have potential but are rarely mentioned before. Finally, specifically in the business context, we hope our findings can help marketers understand their customers better, to build campaigns aimed at building loyalty [7].

The structure of the paper will be organized as follows: Sect. 2 discuss the previous work, related literature, and possible gaps that our paper can fill in. After an extensive discussion of our methodology in Sect. 3, we discuss obtained results in Sect. 4. This is also where we analyze the results, along with implications in the business context, as well as our shortcomings when conducting the research. Finally, we conclude our findings and map out possible areas for future research in Sect. 5.

2 Previous work

Most previous papers addressing data mining as a tool to support marketer's decision focus on one of the two following tasks [11]. *Firstly*, data mining is used for customer segmentation - the idea that customers are not homogenous, but can be divided into customer groups with some shared commonalities. Many data about customers can be found from their device, provided information, tracking behavior and more, allowing us to effectively group them [3, 11]. The personas that are derived from aggregated data allow marketers to treat a big population of audience as a single person [2], so they can prioritize customer needs that are encompassing and important when building campaigns. *Secondly*, and this is usually a build-up from the first task, data mining is used to predict customer profitability *(our focus for this study)*. Some customer data, and understanding behaviors of different customer typologies, can signal marketers whether a specific visitor is likely to purchase on the website. Since building customer loyalty is always more worth than chasing new customers [7], data mining can point marketers to the ones that are worth investing in, as demonstrated in [9].

As example, the first part of the original research that spawned our dataset focused on predicting in real-time visitor abandonment via various classification techniques – to make automated offers for those who didn't abandon [12]. In a way, they are rewarding potential visitors that can become customers. Testing with various classification techniques, including random forests, support vector machines and neural network, they concluded that neural network performed the best, at >87% accuracy and 0.96 True Negative Rate. Another study [6] performed prediction on customers using different attributes, but still for the purpose of having the algorithm as an automated salesman that can calculate human statistics for each data entry.

With these examples, clustering is usually not considered alongside classification. Clustering can also be considered a classification, as it can categorize data based on attributes. It is not a supplementary step, but rather a weak alternative to other classification methods – weak since it is not very predictive, and categorize without considering the class. Additionally, clustering is not very applicable in an automated algorithm situation. K-means neighbor algorithms can only be applied retroactively, *after* a customer has finished their visit session on a website [12]. Therefore, clustering in an online, automated prediction models does not make sense.

For automated predicting process, clustering is rendered useless. However, for informing marketers about customers to create strategies, clustering can still be valuable. I.e., when managing relationships with customers. While classification models can be reactive to customer needs, sustaining and turning regulars into loyal customers still needs human intervention. This is because customers have various motivations to purchase online, some of which machine learning is not intuitive enough to address – the very same reason why it cannot replicate a salesman in online environment. Therefore, may it be improving website design or create strategies, the data mining processes are included in those processes to inform human decisions. This is when clustering becomes valuable, since it can perform customer segmentation for business on a much larger scale. It can also be combined to a classification model to point to the most profitable and predictable customer groups, thus enhancing the classifying task with more insights for humans to create strategies.

Because of clustering's potential usefulness in classification process, we are not only looking at metrics when evaluating performance of both models. Instead, we also consider its application in a business context: *When used to inform human strategies, what does the model add to collected insights? How easy, effective, and scalable can the* *model be?* Model recommendation for CRM from the results of this study is informed by both statistics *and* practical applications.

3 Methodology

3.1 Dataset & Preprocessing

For this research, we are going to build two random forest classification models to try predicting a binary outcome on whether a customer's visit ended up with or without a purchase. The dataset is the Online Shoppers Purchasing Intentions Dataset, donated to the UCI respiratory in 2018 [8]. It contains 12,330 instances of customer sessions in the span of one year on an e-commerce bookstore, with the class value "Revenue". For preliminary selection, we only included attributes that have clear description and/or transformation documentation. Table 1 summarizes attributes included in this research:

Table 1. Descriptions of attributes from the dataset

Attribute	Description	Type
Administrative	Number of pages visited by the visitor about account management	Numerical
Administrative duration	Amount of time (in seconds) spent by the visitor on account management related pages	
Informational	Number of pages visited by the visitor about Web site, communication, and address information of the shopping site.	
Informational duration	Amount of time (in seconds) spent by the visitor on informational pages	
Product-related	Number of pages visited by visitor about product related pages	
Product-related duration	Amount of time (in seconds) spent by the visitor on product related pages	
Page value	Average page value of the pages visited by the visitor	
Special day	Closeness of the site visiting time to a special day	
Visitor Type	Visitor type as "New", "Returning", and "Other"	Categorical
Weekend	Whether the date of the visit is weekend	Boolean
Season	Aggregated months of the visit: Spring [Feb -> May]; Summer [June -> Sep]; Winter [Remaining]	Categorical

The dataset is divided into 80% training and 20% testing, used the same by both models. In the first model, the sample population will go through one overarching random forest. In the second model, each record will be categorized first into a cluster (customer group) built from training data, then go through the random forest algorithm built specifically for that cluster. The second model will produce *k* clusters, and correspondingly k random forest algorithms. The weighted average result of these algorithms is the performance of the second model. Performance of both models are compared and tested for significance to draw conclusions.

With numerical attributes, since we partly use a distance measure to calculate their similarity (to be discussed later in Sect **3.3**), they first went through a normalization process. Most numerical attributes were normalized with linear scaling:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

However, three duration attributes are normalized through the logarithmic scale. Purchase intentions scales exponentially with durations like time spent on a website [1], with the upper-bound limit of 1. As an example to visualize, the growth of purchase intentions between two people spending 0 and 1 hours (respectively) on a website will be more significant than two people spending 3 and 4 hours. Therefore, we use the logarithm to the base of two before normalizing these attributes: $X' = log2(X)$.

3.2 Classification

Both models utilize random forests, with 5-folds cross validation, to classify customers on their purchase intentions. Decision trees classification is computationally inexpensive and produce comparable results to other classification techniques [13]; random forest – an ensemble machine learning algorithm built based on many decision trees – inherits these advantages while having more confident generalization and less prone to noises [13]. Since the study requires k + 1 classification models to be built, random forests are easier and more efficient to construct, allowing us to optimize them well. For each model in this research that is evaluating a sample size N, the following steps apply:

1. Each tree in the random forest model draws 90% (parameter-tuned number) from the sample size N, through sampling with replacement.
2. Each tree in the model evaluates each instance of own drawn samples using a random set of attributes, outputting a class prediction for that instance.
3. Using majority voting, whichever class is voted the most will become the model's prediction for that instance.

The algorithms are further fine-tuned by allowing each tree in the random forest to be full-grown, pick $log2(number\ of\ selected\ attributes) + 1$ random attributes to consider at a specific node & include >500 individual trees in its prediction.

Since the dataset is heavily imbalanced (15% of instances were of minority class - purchased), the classifiers are cost-sensitive, meaning that they will try to minimize the cost of misclassification as specified by user. The cost matrix is based on a popular idea within the CRM domain: Keeping a customer loyal is 5 times as valuable as acquiring a new customer [7]. The costs are assigned to reward the algorithm for correctly identify a potential customer, and punished for misclassification based on the business context. A False Positive affects our whole perception of customer base, while a False Negative means we missed a potential customer. Table 2 demonstrates the final cost matrix used for this study, after going through adjustments that account for classifier's role in keeping customers:

Table 2. Defined cost-matrix for the study

	Customer	Non-customer
Customer	-1	2
Non-customer	3	0

Before constructing the random forests, we also conduct an attribute selection for each of the algorithms. Using the wrapper method, we evaluate subsets of attributes when they are applied to our cost-sensitive random forest algorithm, adding attributes to the subset as we go using the greedy stepwise search method. The cutoff point is when the F1-measure of the positive class decrease.

3.3 Clustering

Specifically applied to the second model, we are performing mixed-data type clustering using Partitioning Around Medoids (PAM). [14] This has been a popular method, being less sensitive to noises and outliers compared to familiar k-means technique. It also allows more flexible distance metrics, going beyond Euclidean or Manhattan distances. More importantly, PAM allows us to identify clusters via medoids – real instances from the dataset [14]. This eases the work of identifying a persona, since the medoids of each cluster can be considered a representative persona for that cluster/customer group.

The algorithm to perform PAM is as follows:

Algorithm. Partitioning Around Medoids (PAM)

select K random medoids β from the dataset
for point $\alpha \notin \beta_i$ of the dataset
  calculate distance to β
  if distance to medoid β_i is smallest
    add α to $cluster\ \beta_i$
  end if
end for
calculate SSE of the population
for medoid β_i
  for point $\alpha_i \notin \beta_i$ of the dataset
    calculate distance to (α_i and $\beta \neg \beta_i$) medoids
    add points based on distance
    calculate SSE of the change
    if $SSE_{population} > SSE_{change}$
      set $\beta_i = \alpha_i$
    end if
  end for
end for

Since attributes used for clustering is mixed (both numerical and categorical), distance between any two points is measured by Gower's distance. It equals to the average of dissimilarities of two points across individual dimensions. For two points $x_1,\ x_2$ that has p dimensions, their Gower's distance is:

$$D_{Gower}(x_1, x_2) \ = \ 1 \ - \ (\frac{1}{p}\sum_{j\ =\ 1}^{p} s_j(\,x_1, x_2))$$

In which the similarity of x_1, x_2 in a specific dimension is calculated by:

$$s_j(x_1, x_2) \ = \ |y_{1j} \ - \ y_{2j}|$$

if the values of x_1, x_2 in this dimension is numerical, and:

$$s_j(x_1, x_2) \ = \ \begin{cases} 1 \ if \ x_1 \ = x_2 \\ 0 \ if \ x_1 \ \neq x_2 \end{cases}$$

if the values are categorical.

The K-number of medoids is decided using the elbow method. We conducted the PAM algorithm with K within range [2;10] and calculated the Within Cluster Sum of Squared errors (WCSS) of the iteration:

$$WCSS(K) = \sum_{j=1}^{k} \sum_{\alpha_i \in cluster\ \beta_i} ||\,\alpha_i \ - \beta_i||^2$$

3.4 Evaluation

Both classification models are evaluated on how well they capture potential customers, via the same test set (20% of the dataset, sampled with replacement) on the following metrics:

- Accuracy (How well it identifies customers purchases in general)
- Precision for positive class (How well it captures true customers)
- F1-score (Good performance regarding customers overall)

The first model uses its results metrics, while the second model uses the average performance of individual algorithms evaluating each cluster, weighted by cluster size. Metric performances are tested for significance [13] at 95% confidence level. Within the same metric, let p_1 be performance of first model, p_2 be performance of second model, and p be the average of p_1 and p_2. The difference is significant if:

$$Z \ = \left| \frac{p_1 \ - \ p_2}{\sqrt{\frac{2p\ (1-p)}{N}}} \right| \geq \ 1.96$$

4 Evaluation

4.1 Clustering results

Figure 2 shows the Within-cluster Sum of Squared errors with different number of K. The drop-off is significant until K = 4, at which point further increase of K does not drastically reduce the errors anymore. Therefore, for this study we will have 4 clusters.

Fig. 1. Elbow chart

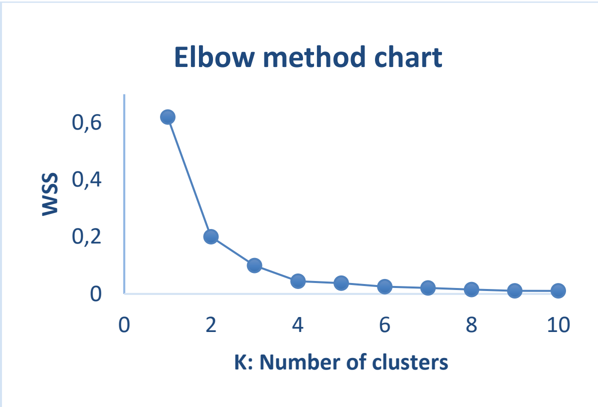


Table 3 shows the result when applying the PAM algorithm to the dataset with K = 4, including number of instances in each cluster in the training set and asymmetric attributes of their representative instance:

Table 3. Summary of clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
# of instances	**2096**	**2169**	**1984**	**3615**
Administrative	2	1	7	
Administrative duration (s)	48.83	20	228.25	
Informational	6			
Informational duration (s)	370.875			
Product	22	21	39	10
Product duration (s)	526.02	2179.66	968.09	130.5
Page Value	23		7	
Special Day	Near			
VisitorType	Returning Visitor			
Weekend		Yes		
Season	Spring	Spring	Winter	Winter

Based on asymmetric features, we categorize the customers in the dataset into the following categories (although the clusters are very open for different interpretations):

- **Cluster 1: Seasonal engaged shoppers**. Engage near the seasonal holidays, needing to go through website thoroughly (hence the statistics on informational and administrative pages), prioritize product pages with high value.
- **Cluster 2: Weekend shoppers.** Regular visitor, have an account on the website and occasionally visit to browse books, especially in the weekend.
- **Cluster 3: Engaged shoppers.** Visit a lot of pages about account, either to check the status of their product or because of an incentive when completing account information. Visitors from this group is most likely to purchase among the four groups (>20% positive class)
- **Cluster 4: Casual browsers.** Do not engage with creating accounts & information about the page, rather only browse through product pages. Quickly leave if the products are not what they are expecting of.

We want to note in this section that before settling on the PAM algorithm, we tried a different approach to clustering using Hartigan's algorithm [13]. Starting with random medoids, we group them into clusters if their distance is lower than a specified threshold – otherwise, they become a new medoid themselves, and points are grouped into this new medoid. This process is repeated with L different user-specified lengths, with WCSS calculated for each iteration. While this did not result in a feasible indication of the optimal intra-cluster length (they fluctuate with different lengths), we have noticed that small clusters with <1% of the population still exist in some iterations with moderate lengths. Since distance-based clustering is prone to noises [13], being able to identify and eliminate noise to increase cluster quality is valuable. Due to the small scope of the research, we have not been able to eliminate noises in the dataset using Hartigan's algorithm; future research can investigate this algorithm to increase cluster quality.

4.2 Classifying

When the classification model for the sample is at best performance, the following list of attributes were used to predict (for both population & individual cluster models):

Table 4. Attributes used in best predicting classification model for the sample

	total	1	2	3	4
Administrative	x	x	x	x	
Administrative duration		x	x	x	
Informational	x	x	x	x	x
Informational duration		x		x	x
Product	x		x	x	
Product duration	x	x		x	
Page Value	x	x	x	x	x
Special Day	x			x	x
VisitorType		x		x	x
Weekend				x	
Season	x	x		x	

While most subsets of best attributes to classify are normal, the fourth cluster (casual browsers) used suspecting attributes to predict. For these attributes, the majority of instances (≥97%) register the value of 0. Other attributes have higher variance, and using them to predict this customer group yields unsatisfactory result. This reflects the unpredictability of casual browsers, in which their browsing behavior have weak or no correlations to their purchase intentions.

Fig. 2. Distribution of attributes used to predict the 4th cluster (blue = purchased)

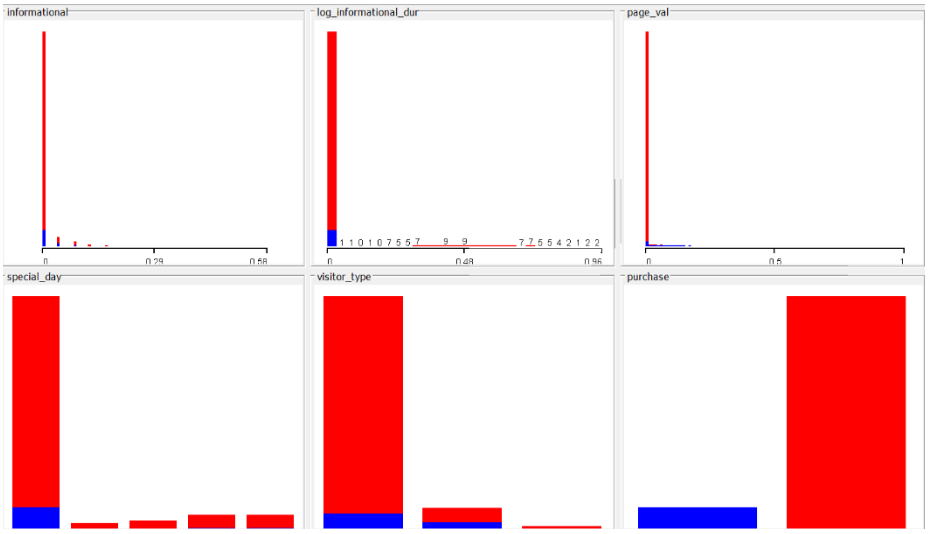


Table 5 describes the performance of the two classification models, with regards to performance metrics mentioned in Sect. 3.4:

Table 5. Performance on a supplied test set

		Accuracy	Precision	F1-measure	ROC Area
Cluster-then-classify model	1 [484]	84.3%	0.600	0.832	0.712
	2 [525]	89.3%	0.628	0.868	0.766
	3 [479]	79.2%	0.571	0.774	0.670
	4 [977]	95.9%	0.822	0.950	0.810
	Aggregated	**87.9%**	**0.688**	**0.875**	
Classifying model [2465]		96.1%	0.917	0.961	0.901

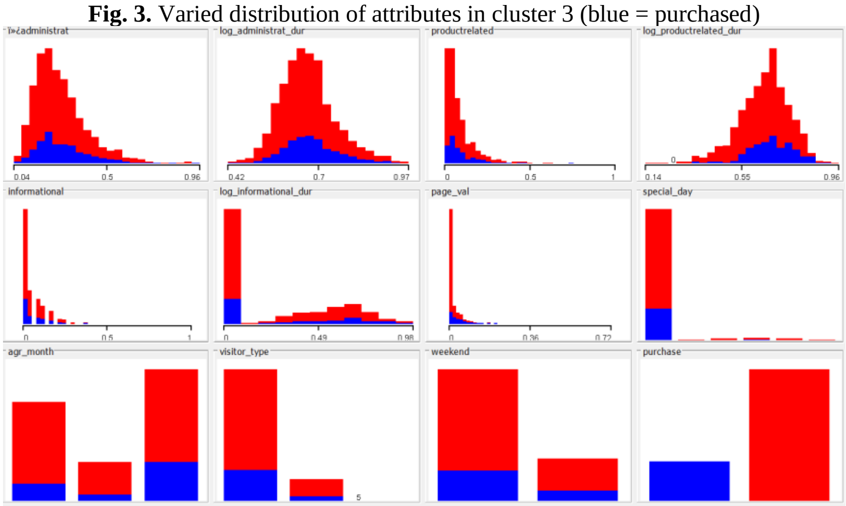
With regards to the dataset size N = 12330, the differences in all metrics at significant at 95% confidence level.

While the cluster-then-classify model performed marginally worse than the generic classifying model, it still produced a model that was comparable to previously built models [5, 12]. These comparisons have taken into account that the second model was used in a different context compared to [12], and has to be optimized for different metrics rather than purely accuracy in [5]. For our purpose of capturing potential customers, the model does an acceptable job and can be applied to predict where customers are, so marketers can build strategies that encourage their loyalty.

Interestingly enough, while they shared some attributes (*Page value* and *Informational*), each cluster predicting model has to use more attributes to predict purchase intention, and the set of attributes vary from being very simple to a full set of attributes. The insight implies that groups of customers are not the same and should not be predicted in the same manner, using the same statistics. This is further doubled down by the fact that in recent times, customers are less likely to give away their information online. Online tracking such as cookies have become optional – and realistically, session data will not be as coherent as this study, instead plagued with missing values. Customer concerns over their data have already fueled research on how to keep algorithms as lean as possible, such as the research in [4]. In the long term, having smaller models that need few attributes to predict customer groups - even if they can only produce acceptable results - will be more feasible in the business context than a generic algorithm that must handle missing values of all customers. Within the research scope, given that the dataset has no missing value, individual models of the clustering-then-classify model is undesirable; however, it shows promises when applying in the real-world context.

Putting the business context aside, the intricacies of models increase, and performance subsequently decrease, when the cluster has more purchasers (and vice versa). Model for cluster 3 had to use all attributes included in the study for best performance, and even then, it was still the worst model. At the same time, customers within this cluster have the highest purchase rate among the four clusters, at over 20% as mentioned above. An opposite situation exists in the 4th model, where it used mostly irrelevant attributes to predict with excellent performance, for a group of visitors that rarely purchased from the website. This showcases how hard it is to realistically capture customers using mined data about their behavior – since their behavior varies from one to another, and are harder to predict.

Fig. 3. Varied distribution of attributes in cluster 3 (blue = purchased)



In an imbalanced class problem such as with this research, focusing on the minority class does not yield a great result – especially when we are bisecting them further into groups of customers with changing distribution. A better approach would have been to focus on eliminating *non-customers* instead, effectively turning the problem into building the best majority class classifier. [12] has already approached their problem in a similar fashion: They have the algorithm recognize non-customers and ignore them, and give encouragement to the remaining ones (now considered potentially loyal customers). From a logical perspective, building a classifier to identify the majority class – given that the misclassification to the minority class is costly but not extreme – is a lot easier as well. We would recommend future research focusing on the minority class to consider the proxy approach through identifying the majority class instead.

Despite the challenges of identifying a minority class in an imbalanced dataset, the generic classifying model in our research performed *exceptionally well* with unseen records. While both models had an average-to-good performance with the training set, the generic model jumped up over 9 percent point in test set accuracy, with other metrics rising to around 0.9. We have tested the generic model with two additional sets using both sampling with and without replacement, changing the random seed in both cases, only to receive similar results. This showcases the effectiveness of random forest algorithm as a tool to inform human strategies, with easy setup, quick and reliable results.

The difference between the two models, and the jump between training and test of the generic model, can be explained through the number of instances used by each model in training. With 5-fold cross validation, each fold of the generic model has >7800 instances, whereas for mini-models of the second model that number ranges from 1500-3000. A higher number of instances for training, in a high-variance dataset, produces a better result when testing.

Additionally, the generic model performs well given how little time is needed to build it, compared to the clustering-then-classify model. For the latter model, the time complexity needed was $O(4 * (12330 - 4)^2)$ – which was computationally expensive. It also requires four attribute selection processes, and four random forests to be built – as opposed to one from the generic model. When a new record is registered, there needs to be a K-nearest neighbor algorithm designated to assign. Therefore, for quick and efficient results, the first, generic model is more desirable.

Finally, while both models have performed well on the task with varying degrees of success, a question remains: *How applicable is the result when replacing the classification technique?* For this research, random forest is the most common technique to apply and enhance [5, 12] due to its easy construction. However, other classifying techniques can perform better, such as neural network and long-short term memory (LSTM) in the research of [12]. Therefore, research replicating ours with a different classification method will be highly valuable, both in comparing different techniques and assessing the generalization of our findings.

5 Conclusion

The research question – *What are the performances of two classification models - one general random forest, and one clustering-then-random forest - in predicting E-commerce customer purchases?* – can be answered as follows.

Our research constructed two models to predict whether a visitor on an E-commerce website became a purchased customer or not. The first is a generic random forest model, which performed well on the training set and exceptionally well on unseen records. For the business object of capturing customers retroactively so marketers can devise strategies, this model can quickly accomplish the task and easy to be put into practice.

The second model first group visitors into four groups using the PAM algorithm with Gower's distance, then have the respective classifying model built for that group to predict purchase intentions. This model, when aggregated, performed significantly worse than the first model across all metrics in the test set, even though on par when training. However, it offered insights into customers visiting the website and their behavior. Namely, for this E-commerce website, we have seasonal shoppers, weekend shoppers, engaged shoppers and casual browsers. For some mini-models, the attributes they used for prediction are fewer than the generic model – which is preferred considering the realistic context of data privacy and safeguarding. While the model didn't surpass the standard generic model, it showed potential in the future as businesses strive to capture customers without much data available.

From a business perspective, our study showed the effectiveness of data mining models in predicting E-commerce customers to inform marketers on how to construct strategies for customers. For research purposes, it also opens many suggestions for the future:

- There can be a closer look into using algorithms such as Hartigan's in effectively removing noises and outliers from the dataset.
- Research can replicate our settings with a different approach to class and techniques to generalize, compare, or improve our findings.
- More grounded and business-oriented research can investigate attributes or indicators that predict online customers, even when many tracking methods are restricted in this age of data privacy and consumer concerns.

References

1. Ahsain S, M'Hamed AK. Predicting the client's purchasing intention using Machine Learning models. E3S Web of Conferences. 2022; 351:01070.
2. An J, Kwak H, Jung S-g, Salminen J, Jansen BJ. Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. Social Network Analysis and Mining. 2018;8(1):54.
3. Bounsaythip C, Rinta-runsala E. Overview of Data Mining for Customer Behavior Modeling. 2002.

4. Bourtoule L, Chandrasekaran V, Choquette-Choo CA, Jia H, Travers A, Zhang B, et al., editors. Machine Unlearning. 2021 IEEE Symposium on Security and Privacy (SP); 2021 24-27 May 2021.
5. Kabir MR, Ashraf F, Ajwad R. Analysis of Different Predicting Model for Online Shoppers' Purchase Intention from Empirical Data2019. 1-6 p.
6. Larivière B, Van den Poel D. Predicting customer retention and profitability by using random forests and regression forests techniques. Expert Systems with Applications. 2005;29(2):472-84.
7. Level-up customer engagement & loyalty [Internet]. Gent (BE): StriveCloud; [date unknown]. Learn your customer lifetime value first, if you want to improve loyalty; [date unknown] [cited 2022 Oct 12]. Available from: https://strivecloud.io/blog/fan-engagement/cltv-improves-loyalty/
8. Machine Learning Repository [Internet]. [place unknown]: ICU; 201. Available from: https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset
9. Moe WW. Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. Journal of Consumer Psychology. 2003;13(1):29-39.
10. Morgan Stanley [Internet]. New York City (USA): date unknown. Here's Why E-Commerce Growth Can Stay Stronger for Longer; 2022 Jun 14 [cited 2022 Oct 10]. Available from: https://www.morganstanley.com/ideas/global-ecommerce-growth-forecast-2022
11. Rygielski C, Wang J-C, Yen DC. Data mining techniques for customer relationship management. Technology in Society. 2002;24(4):483-502.
12. Sakar CO, Polat SO, Katircioglu M, Kastro Y. Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. Neural Computing and Applications. 2019;31(10):6893-908.
13. Tan PN, Steinbach M, Karpatne A, Kumar V. Introduction to Data Mining. Second Edition. Pearson; 2018. *169-170, 290-294, 203, 564.*
14. Van der Laan M, Pollard K, Bryan J. A new partitioning around medoids algorithm. Journal of Statistical Computation and Simulation. 2003;73(8):575-84.