USING SUPERVISED MACHINE LEARNING METHODS FOR RFM SEGMENTATION: A CASINO DIRECT MARKETING COMMUNICATION CASE

KORIŠTENJE NADZIRANIH METODA STROJNOGA UČENJA ZA RFM SEGMENTACIJU: SLUČAJ IZRAVNE MARKETINŠKE KOMUNIKACIJE KASINA

Market-Tržište
Vol. 35, No. 1, 2023, pp. 7-22
UDK 658.8:005.52:004.021
DOI http://dx.doi.org/10.22598/mt/2023.35.1.7
Original scientific paper

Danijel Bratina[a], Armand Faganel[b]

[a] University of Primorska, Faculty of Management, Izolska vrata 2, 6000 Koper, SLOVENIA, e-mail: danijel.bratina@fm-kp.si

[b] University of Primorska, Faculty of Management, Izolska vrata 2, 6000 Koper, SLOVENIA, e-mail: armand.faganel@fm-kp.si

Abstract

Purpose – This paper explores various supervised machine learning algorithms as an additional classification method to RFM (recency, frequency, and monetary) models with the aim of improving the accuracy in predicting target groups of customers for direct marketing response campaigns conducted by a casino. The purpose of this paper is twofold – first, to test how the addition of demographic variables increases the accuracy of the basic RFM model and second, to assess if and how machine learning algorithms improve the initial model. Ultimately, we propose a model for direct marketing response at individual level using RFM scores and customer demographic and behavioral data as endogenous variables to be used by the company. The findings can be used as an alternative to the simpler RFM model when approaching customer response modeling for large datasets and can be generalized to other industries.

Design/Methodology/Approach – Our research employed supervised machine learning methods tuned on historical responses to a casino's direct marketing activities to improve the company's RFM segmentation model.

Sažetak

Svrha – Rad istražuje različite nadzirane algoritme strojnoga učenja kao dodatne metode klasifikacije RFM-a (recentnost, učestalost i monetarnost) modelima da bi se poboljšala točnost predviđanja ciljanih skupina potrošača za kampanje izravne marketinške komunikacije za kasino. Svrha je rada dvostruka. Prvo, ispitati kako dodavanje demografskih varijabli povećava točnost osnovnog RFM modela i, drugo, procijeniti poboljšavaju li i kako algoritmi strojnog učenja početni model. U konačnici, predlažemo model za izravni marketinški odgovor na individualnoj razini koristeći RFM rezultate te demografske i podatke o ponašanju potrošača kao endogene varijable za korištenje od strane poduzeća. Rezultati se mogu koristiti za pristup modeliranju potrošačkih odgovora za velike skupove podataka i kao alternativa jednostavnijem RFM modelu te se mogu generalizirati na druge industrije.

Metodološki pristup – Istraživanje koristi nadzirane metode strojnoga učenja usklađene s povijesnim odgovorima na aktivnosti izravne marketinške komunikacije kasina kako bi se poboljšao RFM segmentacijski model

Demographic variables were also included with the aim of improving the power of the models employed. Finally, we attempted to improve the best-performing model by hypertuning its algorithm parameters.

Findings and Implications – The best and most intuitive model was found to be that using decision trees with Recency (from RFM) together with age and the awarded amount (from the demographic element) as independent variables. Surprisingly, the company's own RFM segmentation was also found to perform well.

Limitations – Not all machine learning methods used for classification were included in our research nor did we use ensemble methods to improve the models' power. While all models developed are applicable to similar data, they could lose their accuracy when applied to data from a different industry. The company's own RFM model was not analyzed but was included in the model as is. Further insight could be gained by determining its optimal parameters.

Originality – This study contributes to the existing literature by showing how direct marketing efficiency modeling using standard RFM could be improved with the addition of a company's customer property. It also provides insight into how classification algorithms perform on a casino database of direct marketing activities.

Keywords – RFM segmentation, machine learning algorithms, decision trees, support vector machines, naïve Bayes algorithm, logistic regression

poduzeća. Uključene su i demografske varijable radi poboljšanja snage modela. Konačno, pokušavamo poboljšati model s najboljom izvedbom hiperpodešavanjem parametara njegova algoritma.

Rezultati i implikacije – Otkriveno je da su najbolji i najintuitivniji model stabla odlučivanja s recentnošću (iz RFM-a) zajedno s dobi i dodijeljenom nagradom (iz demografskih podataka) kao neovisnim varijablama. Iznenađujuće, RFM segmentacija poduzeća također pokazuje dobre rezultate.

Ograničenja – Rad ne uključuje sve metode strojnoga učenja koje se koriste za klasifikaciju niti pokušava koristiti skupne metode za poboljšanje snage modela. Svi razvijeni modeli primjenjivi su na slične podatke, ali mogu izgubiti točnost kada se koriste na podacima iz druge industrije. RFM model poduzeća nije analiziran i uključen je u model kakav jest. Potreban je dodatni uvid radi određivanja njegovih optimalnih parametara.

Doprinos – Rad pridonosi postojećoj literaturi pokazujući kako se modeliranje učinkovitosti izravnoga marketinga korištenjem standardnog RFM-a može poboljšati dodatkom informacija o potrošačima koje posjeduje poduzeće. Isto tako, pruža uvid u to kako se algoritmi za klasifikaciju ponašaju u bazi podataka kasina o aktivnostima izravnog marketinga.

Ključne riječi – RFM segmentacija, algoritmi strojnog učenja, stabla odlučivanja, metoda potpornih vektora, naivni Bayesov algoritam, logistička regresija

1. INTRODUCTION AND THEORETICAL FRAMEWORK

The informatization of data and transaction tracking has led to an increase in the amount of consumer information, allowing companies to create better decision-making models in all areas of business. Effective customer segmentation can be a source of competitive advantage (Pavlić, Vojvodić & Puh, 2020; Zaheer & Kline, 2018). In marketing, the practice of database marketing is typically used in the process of customer segmentation for direct marketing activities. As the volume of collected data is growing exponentially, marketers are faced with the challenge of spending their marketing communication budget appropriately on the most promising customers. The RFM (Recency, Frequency, Monetary) framework (Blattberg, Kim & Neslin, 2008), introduced in 1960, has been used extensively as a segmentation tool for direct marketing activities in almost all industries (including B2B). The main idea behind RFM is to segment on three variables related to past customer behavior: (1) time since their last purchase (i.e., recency), (2) number of purchases in a given period (month, year – i.e., frequency), and (3) value of past purchases (i.e., monetary). The rationale is that customers who have made a purchase recently, frequently, and spent relatively large amounts of money are more likely to respond to a direct marketing activity (Bult & Wansbeek, 1995; Gönül, Kim & Shi, 2000; Wei, Lin & Wu, 2010). The aim of RFM is to identify such segments, thus allowing marketers to target these potentially profitable customers.

1.1. Overview of RFM

Despite being easy to implement (i.e., inexpensive, due to the use of internally available data) and understandable, RFM has limited differentiation ability and cannot be used for new customers (Yang, 2004). The original RFM splits the customer transaction database by its three dimensions into quintiles (20% groups), generating a 5x5x5 (125) segments framework. Two major scoring methods have been proposed – customer and behavior quintile. The former ensures equally sized segments, whereas the latter splits the groups based on parameter value. Both scoring systems are sensitive to frequency, which could lead to the grouping of customers with considerably different behavior (Miglautsch, 2000). Improvements with regard to weighting and averaging methods have been proposed by Liu and Shih (2005a, 2005b) and Yang (2004). The number of segments has been reduced by summing the R, F, and M into a single score (Hughes, 1994) – assuming all three variables have identical weights. Miglautsch (2000) proposed weighting the factors before aggregating, and Tsai and Chiu (2004) added a criterion that sums all weights to 1 while allowing for different weights for different product categories. An analytical hierarchical process of calculating weights has also been proposed (Liu & Shih, 2005a, 2005b).

RFM has been widely used in many areas, including the financial sector (Sohrabi & Khanlari, 2007), telecommunications (Li, Shue & Lee, 2008), and marketing (Gustriansyah, Suhandi & Antony, 2019). Moreover, several customer lifetime value (CLV) models have been derived from the RFM (Liu & Shih, 2005b; Sohrabi & Khanlari, 2007).

Critics of RFM models argue that RFM aims to identify valuable customers only while providing little or no meaningful scoring on recency, frequency, and monetary value when a company's customer base does not make frequent purchases, when it spends small amounts on such purchases, and/or has made their last purchase long ago. Such customers fall into the scoring of recency, frequency, monetary (1,1,1), which is the lowest possible score. They are usually identified as the group largest in size and with the greatest potential (Miglautsch, 2002). Other critics argue that RFM is unable to account for customer heterogeneity (Suh, Noh & Suh, 1999) or focus on existing customers (McCarty & Hastak, 2007) while also failing to address multicollinearity among factors (Bult & Wansbeek, 1995; Chan, 2005). In addition, RFM is inconsistent when it comes to assessing the importance of R, F, and M across industries (Yeh, Yang & Ting, 2009).

In order to address RFM weaknesses, several models that incorporate other variables have been developed with the aim of improving predictability. Bucklin and Van Den Poel (2005) found evidence of the importance of demographic variables and the length of customer relationships when building an attrition model. Hosseini, Maleki, and Gholamian (2010) combined RFM variables with product loyalty to yield better prediction, while Yeh et al. (2009) added time since first purchase and churn probability in creating a model named RFMTC, which is able to estimate the probability that a customer will make a purchase next time as well as the expected number of times the customer will make a purchase in the future.

1.2. Alternative methods to RFM for direct marketing activities in the machine learning framework

When modeling direct marketing activities, the aim is to engage only with historically or potentially profitable customers with the goal of maximizing returns on investments. Faganel and Costantini's (2020) study suggests that CRM projects help develop long-term relationships with casino customers and provide continuing benefits (retention rates) due to higher gambler satisfaction. Improved RFM models such as AID (Automated Interaction Detection) and CHAID (Chi Square AID) (Kass, 1980), the general purchase model (Bauer, 1988), and gains chart analysis (Bult & Wansbeek, 1995) have been proposed as improvements to the RFM. With increasing computer power and storage space available for analysis, more computationally intensive models have recently been advanced from the ML supervised learning set, namely logistic regression (McCarty & Hastak, 2007; Tabaj Pušnar & Bratina, 2018), support vector machines – SVM (Kim, Chae & Olson, 2013; Rahim, Mashafiq, Khan & Arain, 2021), and k-nearest neighbors – kNN (Bing, Xin-xingi & Ke, 2006).

2. GENERAL SUPERVISED ML FRAMEWORK AND MODELS USED

We fit several machine learning models to our research data. Advantages and disadvantages for each of them are shown in Table 1.

TABLE 1: Supervised machine learning algorithms (Source: Crisci et al., 2012)

Model	Advantages of the model	Disadvantages of the model
Naïve Bayes	Simple	Naïvety cannot hold (multicollinearity)
	Better for categorical data	Black box
	No distribution requirements	
Logistic Regression	Most widely used still	Only suitable if linearity is assumed
	Easy to interpret	
Support vector machines	Better suited for nonlinear problems	Black box
	No distribution requirements	Slow to train
		Not very good for high-quantity data
Decision trees	Easy to interpret and explain	Good for categorical data
	Non-parametric	

Supervised machine learning techniques fit output (Y – dependent) variables to a set of input (X – independent) variables using optimization algorithms for a cost (L – loss) function:

$$f^* = \underset{f \in F}{\operatorname{argmin}}\ L(F, X, Y)$$

where F is a set of all possible solution functions. The manner in which the loss function is operationalized for each method is shown in Table 2.

TABLE 2: Loss functions for ML algorithms

Method	Loss function construction
Naïve Bayes	Uses Bayes' rules instead of cost functions
Logistic regression	Maximum likelihood – $$l(\beta; y, X) = \sum_{i=1}^{N}[-\ln(1+\exp(x_i\beta)) + y_i x_i \beta$$
Support vector machines	$$Cost(h_\theta(x,y) = \begin{cases}\max(0,1-\theta^T x)\ if\ y=1\\ \max(0,1+\theta^T x)\ if\ y=0\end{cases}$$ $$J(\theta) = \sum_{i=1}^{m} y^{(i)} Cost_1(\theta^T(x^{(i)}) + (1-y^{(i)})Cost_0(\theta^T(x^{(i)})$$
Decision trees	Information-gain-based – $$Entropy(p) = -\sum_{i=1}^{n} p_i log_2(p_i)$$ $$Gini(P) = 1 - \sum_{i=1}^{n}(p_i)^2$$

2.1. Decision trees

The main idea behind decision trees is to split the sample into homogeneous groups based on a selection of independent variables. The most suitable variable for the split is selected by entropy decrease (or information gain) (Leeflang, Wieringa, Bijmolt & Pauwels, 2017), Gini impurity (Kuhn & De Mori, 1995) or variance reduction (Gascuel, 2000) for the class of models named Classification and Regression Trees (CART). The latter two are used in the case of a continuous dependent variable.

Alternatively, Chi-squared Automatic Interaction Detector (CHAID) models, introduced by Kass (1980), use Chi-square tests (for binary dependent variables) and the F-test (for continuous variables) as splitting criteria, or QUEST (Quick, Unbiased, Efficient, Statistical Tree) relying on linear discriminant analysis (Loh & Shin, 1997).

Decision trees are prone to overfitting. To overcome this, a variety of pruning techniques can be used (Rokach & Maimon, 2007) to reduce the number of end nodes (final split criterion). Aggregating decision rules across multiple models (i.e., ensembling – Random Forest, Boosting, Bagging) can be used to further address overfitting issues (Leeflang et al., 2017).

As trees are easy to understand and visualize, they are very popular in practice. The fact that they, by their very nature, address interaction effects as well as nonlinear phenomena is another positive feature to be noted.

2.2. Support vector machine

Support vector machines, developed as early as 1962 but reaching practical application only in 1992, can be considered an extension of linear discriminant analysis (Vapnik & Cortes, 1995). The aim of the support vector machine (SVM) algorithm is to determine a hyperplane that separates two (or more) classes and concerns quadratic optimization (Flach, 2001). The original model, allowing for linear hyperplane separation of positive and negative cases, was soon extended by means of nonlinear-mapping (kernel-based) support vector machines (Boser, Guyon & Vapnik, 1992). Kernels (Gaussian, sigmoid, radial, polynomial, etc.) are functions that transform the original predictors to address inseparable sets. For 2–3-dimensional predictor spaces one can select the appropriate kernel function by visually inspecting the graphical representation of classified data in the 2D or 3D space of the predictor variables. For higherorder space statistics such as ROC, accuracy helps to determine the most appropriate kernel function.

2.3. Naïve Bayes

The Bayesian classification is known for being computationally inexpensive, especially when large numbers of parameters are used in the model (Hastie, Tibshirani & Friedman, 2009). The naïvety of the model comes from the assumption that all predictor variables are independent, yielding independent marginal classification probabilities for each variable allowing the use of Bayes' rule. Bayes' classifier calculates probability *P* for class *y* (took the offer, did not take the offer) given the vector *x* of predictors:

$$P(Y = y|x = (x_1, x_2, \dots, x_n)),$$

The class with the highest probability is chosen using Bayes' classifier. Applying Bayes' rule, we calculate $P(Y|X) = \frac{P(X|Y)*P(X)}{P(Y)}$.

3. DATA AND OPERATIONALIZATION OF VARIABLES

The main aim of this research is to determine whether the addition of demographic variables and the use of machine learning algorithms yields better predictions than the company's own RFM model when determining which customers should be provided with promotional communication. Our dataset represents 302,000 direct communications sent by a casino venue via SMS to 39,400 casino customers over a period of 3 years. In each individual case, all recipients received a message inviting them to visit the casino which contained a code for a gift amount expiring within 14 days. The average number of messages received per database member was 7.65 (sd = 9.03), with a maximum number of 49. The gift amount was determined based on average spending in the past. On average, the voucher amounted to EUR 17.75 (sd = EUR 31.54), while the average amount spent was EUR 16.56 (sd = EUR 30.91). In both cases, the maximum amount was EUR 2,500. The total amount of voucher redemptions was 184,400 (61% of all vouchers sent, with all of them redeemed in full value). Apart from the customers' R, F, and M scores, the dataset includes customer ID, status (allowing access to complimentary services), birthdate (age at the time of communication), address, gender, preferred game played at the casino, amount gifted/claimed, and the month of direct communication activity. During this period, the company undertook no other communication activities.

3.1. The company's RFM model

The company's RFM model follows behavioral criteria for all three determinants, splitting its customer database into five dimensions for Monetary, five for Frequency, and three for Recency, thus creating a 75-segment matrix. Monetary criteria are determined based on average spending per visit over the last year, Frequency by the number of visits in the last year, and Recency by the time since the last visit. Scores are dynamically changed throughout the period of research (3 years). Details are provided in Table 3.

TABLE 3: Proprietary RFM model

Casino's RFM model criteria selection	R (time since the last visit)	F (number of visits over the last year)	M (average amount spent per visit in the last year, in EUR)
Rank	1 – more than 6 months	1 – 0	1 – less than 150
	2 – more than 1 month and less than 6 months	2 – 1-2	2 – less than 1,000
	3 – less than 1 month	3 – 3-5	3 – 1,001+
		4 – 6-10	
		5 – 11+	

The rationale for using the above criteria and their suitability have not been disclosed to the authors and are beyond the scope of this paper. The company has sent direct messages to all 75 segments. The profitability of each segment (*j*) for a given marketing communication activity could be calculated using the following formula:

$$\pi_{t,j} = \sum_{i=1}^{n}(spent_i - won_i) - n * MC_{cost} \text{ ,}$$

where represents the amount spent on a particular occasion by segment member *i* and the amount won on that occasion, all summed over the segment members and deducted by the total cost of the communication (MC_{cost}l. Consequently, the total profitability of a segment would be:

$$\pi_j = \sum_{t=1}^{m}\pi_{t,j} = \sum_{t=1}^{m}\sum_{i=1}^{n} (spent_i - won_i) - n * m * MC_{cost} \text{ ,}$$

where $\sum_{t=1}^{m}\pi_{t,j}$ represents aggregation over time. The below contingency table shows frequencies of voucher redemptions per different RFM segments of the company. Birthdate has been converted to the age at the time of the communication and address to the distance from the casino, while data have been summarized by customer ID by averaging the gift voucher amount granted to and consumed by each recipient.

Therefore, the final models use distance from the venue, age, average bet, coupon amount, and typical game played as independent variables. The binary dependent variable indicates whether the customer cashed in the voucher or not.

TABLE 4: Contingency table for proprietary RFM model

Recency score	Frequency score	Monetary score	Took Offer	
			No	Yes
1			949	12,418
2	3		1,058	13,919
3			1,738	22,086
1			12	122
2	2	5	7	64
3			35	293
1			1	18
2	1		0	2
3			6	51
1			496	6,707
2	3		494	6,515
3			824	11,108
1			34	317
2	2	4	69	454
3			101	799
1			8	42
2	1		11	77
3			18	135
1			912	11,435
2	3		891	11,783
3			1,524	19,372
1			446	4,410
2	2	3	314	3,199
3			644	6,378
1			222	2,171
2	1		127	1,199
3			305	2,942
1			76	1,066
2	3		110	1,252
3			177	2,284
1			61	821
2	2	2	63	681
3			122	1,350
1			97	995
2	1		60	620
3			84	934
1			5	43
2	3		6	93
3			17	145
1			2	56
2	2	5	3	50
3			8	105
1			0	10
2	1		8	75
3			6	60

RFM scores, along with the normalized aggregated RFM score, have been used as independent variables in our direct marketing response modeling and resulting models were compared with those that incorporate customer demographic data.

3.2. The logit models

The ratio of amount claimed and amount gifted has been recoded to a categorical variable representing whether the recipient took the offer or not (all cases of offers claimed resulted in the full claim of the gifted amount). Distance from the venue, age, gifted amount, and average bet per game have been included in the logit model as predictors. The dataset has been split into train and test sets using bootstrapping (Zoubir & Iskandler, 2007) with 1,000 replications. Model estimates are shown in the table below. Specifically, three different models were developed – one using exclusively RFM scores, one using demographic data only, and the third using both sets of independent variables.

Estimates are coefficients (β) in the logit probability equation:

$$p = \frac{e^{\left(\sum_{i=0}^{k} \beta_i x_i\right)}}{1 + e^{\left(\sum_{i=0}^{k} \beta_i x_i\right)}}$$

Applying the rule "recipient took offer" for $p > 0.5$ and "recipient did not take offer" for $p < 0.5$, predictions are obtained from the model, after which the confusion matrix discussed in the next chapter is calculated.

4. RESULTS AND DISCUSSION

The aim of the research was, first, to determine whether using demographic data alone or in combination with RFM model variables would in any way improve direct marketing response predictions and, second, to assess which supervised machine learning algorithm is best for prediction.

TABLE 5: Logit model using demographic data; AIC = 12,788

Predictor	Estimate ()	P value (sig < 0.05)	Wald
(Intercept)	*0.059*	*0.000*	*0.000*
Distance from venue	*0.020*	*0.000*	*0.000*
Age	*-0.050*	*0.000*	*0.003*
Gifted amount	*0.020*	*0.000*	*0.003*
Country is Italy	*-0.030*	*0.008*	*0.002*

TABLE 6: Logit model using RFM score only; AIC = 13,859

Predictor	Estimate ()	P value (sig < 0.05)	Wald
(Intercept)	*0.205*	*0.008*	*0.008*
Weighted RFM score	*-0.127*	*0.006*	*0.006*

TABLE 7: Logit model using both data types; AIC = 12,480

Predictor	Estimate ()	P value (sig < 0.05)	Wald
(Intercept)	*0.800*	*0.000*	*0.000*
Age	*-0.040*	*0.000*	*0.003*
Gifted amount	*-0.050*	*0.000*	*0.003*
Country is Italy	*-0.030*	*0.000*	*0.002*
RFM	*-0.083*	*0.000*	*0.003*

4.1. Testing the models – confusion matrices and AUC comparison

The confusion matrix shows correctly or wrongly predicted true (took offer) and negative (did not take offer) values for all models. Four metrics – accuracy $(\frac{TP+TN}{ALL})$, precision $(\frac{TP}{TP+FP})$, specificity $(\frac{TN}{TN+FP})$, and recall or sensitivity $(\frac{TP}{TP+FN})$ are typically calculated from the confusion matrix. Depending on the cost of a wrong prediction, a single metric is selected as relevant. In our case, the cost of a customer not taking advantage of a voucher and thus returning to the casino (false negative) far outweighs the costs of failing to detect a non-taker (false positive), so the relevant metric in this instance is recall. Table 8 shows the confusion matrix of the logit model for the 4-independent-variables set: (1 = aggRFM) aggregated RFM score as the only predictor (calculated as the weighted average of all three RFM factors, assuming equal importance, as $RFM_{agg} = \frac{R}{R_{ss}} + \frac{F}{F_{ss}} + \frac{M}{M_{ss}}$), where *agg* stands for aggregated and for scale size (in our case, 3 for R,F and 5 for M); (2 = RFM) individual R,F, and M scores as predictors; (3 = D) demographic variables only as predictors; and (4 = FULL) full model using RFM and demographic variables as predictors.

Table 8: Confusion matrices for the logit model

aggRFM		
	Reference	
Prediction	0	1
0	3,423	1,577
1	1,320	3,680
Accuracy		0.71
Precision		0.74
Sensitivity		0.70
Specificity		0.72
AUC		0.71

RFM		
	Reference	
Prediction	0	1
0	3,651	1,349
1	1,124	3,876
Accuracy		0.75
Precision		0.78
Sensitivity		0.74
Specificity		0.76
AUC		0.75

D		
	Reference	
Prediction	0	1
0	3,915	1,085
1	1,120	3,880
Accuracy		0.78
Precision		0.78
Sensitivity		0.78
Specificity		0.78
AUC		0.78

FULL		
	Reference	
Prediction	0	1
0	4,150	850
1	811	4,189
Accuracy		0.83
Precision		0.84
Sensitivity		0.83
Specificity		0.84
AUC		0.83

The addition of independent variables increases the model's performance. Sensitivity with respect to the FULL model increased by roughly 0.12 points or 17% compared to the basic aggRFM. The improvement from using the aggRFM as opposed to RFM is negligible.

The results obtained by using the same four sets of independent variables in the confusion matrices for naïve Bayes, decision tree, and SVM models are shown in Table 9, Table 10, and Table 11, respectively.

Table 9: Confusion matrices for naïve Bayes models

aggRFM		
	Reference	
Prediction	0	1
0	3,205	1,795
1	2,150	2,850
Accuracy		0.61
Precision		0.57
Sensitivity		0.61
Specificity		0.60
AUC		0.61

RFM		
	Reference	
Prediction	0	1
0	3,360	1,640
1	2,030	2,970
Accuracy		0.63
Precision		0.59
Sensitivity		0.64
Specificity		0.62
AUC		0.63

D		
	Reference	
Prediction	0	1
0	3,450	1,550
1	1,950	3,050
Accuracy		0.65
Precision		0.61
Sensitivity		0.66
Specificity		0.64
AUC		0.65

FULL		
	Reference	
Prediction	0	1
0	3,571	1,429
1	1,635	3,365
Accuracy		0.69
Precision		0.67
Sensitivity		0.70
Specificity		0.69
AUC		0.69

TABLE 10: Confusion matrices for decision trees

aggRFM		
	Reference	
Prediction	0	1
0	2,750	2,250
1	2,360	2,640
Accuracy		0.54
Precision		0.53
Sensitivity		0.54
Specificity		0.54
AUC		0.54

RFM		
	Reference	
Prediction	0	1
0	2,976	2,024
1	2,112	2,888
Accuracy		0.59
Precision		0.58
Sensitivity		0.59
Specificity		0.58
AUC		0.59

D		
	Reference	
Prediction	0	1
0	3,150	1,850
1	2,024	2,976
Accuracy		0.61
Precision		0.60
Sensitivity		0.62
Specificity		0.61
AUC		0.61

FULL		
	Reference	
Prediction	0	1
0	3,216	1,784
1	1,841	3,159
Accuracy		0.64
Precision		0.63
Sensitivity		0.64
Specificity		0.64
AUC		0.64

TABLE 11: Confusion matrices for support vector machines

aggRFM		
	Reference	
Prediction	0	1
0	3,346	1,654
1	1,440	3,560
Accuracy		0.69
Precision		0.71
Sensitivity		0.68
Specificity		0.70
AUC		0.69

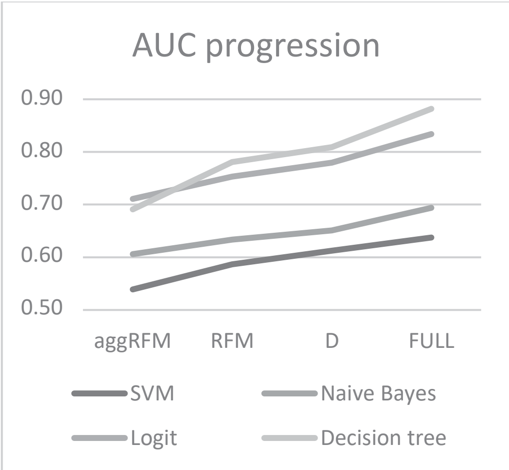
RFM		
	Reference	
Prediction	0	1
0	3,800	1,200
1	995	4,005
Accuracy		0.78
Precision		0.80
Sensitivity		0.77
Specificity		0.79
AUC		0.78

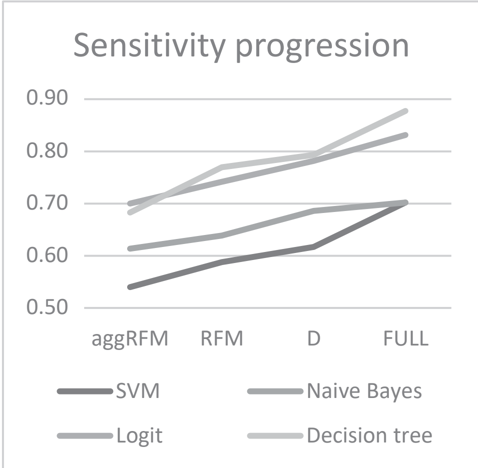
D		
	Reference	
Prediction	0	1
0	3,915	1,085
1	832	4,168
Accuracy		0.81
Precision		0.83
Sensitivity		0.79
Specificity		0.82
AUC		0.81

FULL		
	Reference	
Prediction	0	1
0	4,380	620
1	562	4,438
Accuracy		0.88
Precision		0.89
Sensitivity		0.88
Specificity		0.89
AUC		0.88

Figure 1 shows the improvement of all four methods with the addition of independent variables. Naïve Bayes was found to be the least successful prediction method, followed by SVM and logit methods. The most successful method, yielding an AUC of 0.88 and recall of 88%, is the decision-tree method.

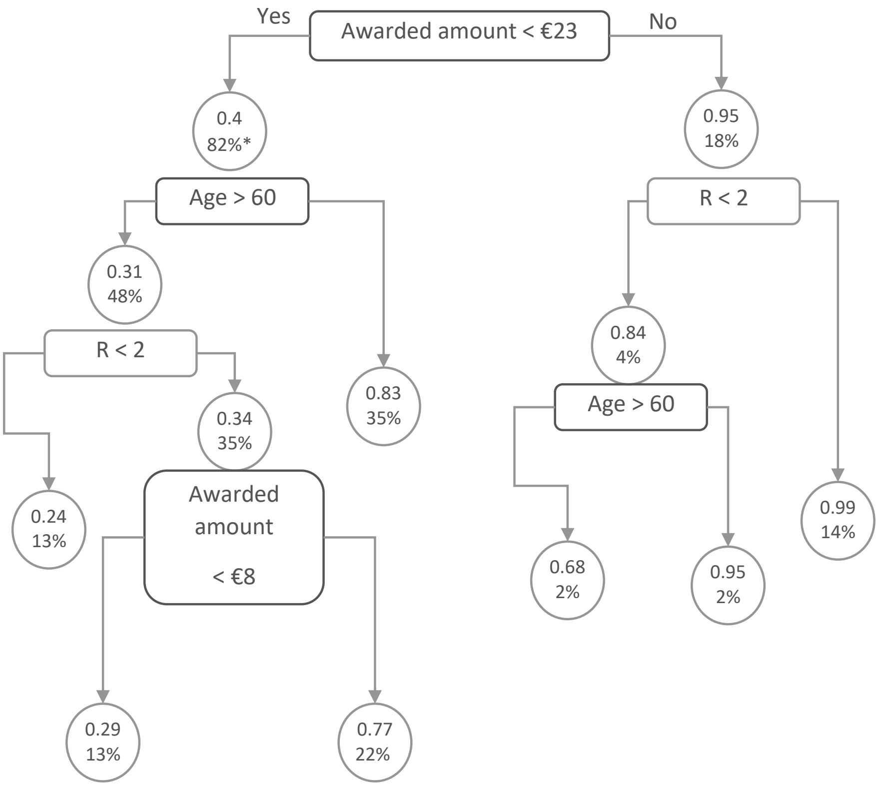
FIGURE 1: AUC and Sensitivity graphs for all models





The SVM model is computationally most expensive, resulting in considerably longer creation times. The decision-tree model that yields the best accuracy is depicted in Figure 2.

Figure 2: Final decision-tree model yielding best accuracy



* The top number represents the proportion of those who took advantage of the offer (e.g., 0.4 means that 40% of those who were given an amount less than €23 took advantage of the offer; there were 82% of them (bottom number)).

Decision-tree models can be interpreted very intuitively. The final model uses only three variables from the initial set – that including R, F, M, the awarded amount, distance, age, and the most played game as independent variables – namely, recency, age, and the awarded amount. The first split shows that the awarded amount is the best splitting parameter, resulting in 95% (sd = 15% over 100 samples of 5,000 events with an even split of 2,500 that took the offer and 2,500 that did not) of customers who received gifts of over EUR 23 taking advantage of the offer (99%, sd = 15%, where the Recency score is over 2). For amounts less than EUR 23 the prediction accuracy is lower. However, for customers older than 60 years an 83% success rate is predicted, and if younger than 60, where the amount is higher than EUR 8, a 77% (sd = 21%) success rate is predicted.

4.2. Improving the decision-tree model with hyperparameter tuning

Furthermore, we have experimented with several hyperparameter tuning variants, adjusting for the maximum number of leaves, minimum split samples, and cost of complexity. Other methods have also been suggested to optimize the algorithm, including the selection of the splitting mechanism at each node – the best, determined by Gini impurity or entropy, or the random, where the best evaluates all splits before splitting, while the random uses a uniform function (Shokeen, Yadav & Kumar Singhal, 2018). The best is computationally more expensive but yields better results. For advanced trees with many splits and several features, the random could represent a viable alternative if computational speed is a factor (Kumar, 2022).

The theoretical maximum of leaves is determined as one less than the number of training samples or until all leaves are pure. Deeper trees give rise to more complexity and result in overfitting the model on the training sample, which usually leads to poor performance on the test sets. The minimum samples parameter determines the least amount of data to be present in each node if a new split is to be created. This parameter enables smoothing and decreases overfitting by allowing each leaf to have impurity greater than 0.

The cost of complexity is determined as a penalization of the sum of squares formalized as

$$SSE = \alpha|T|$$

where T denotes the number of terminal nodes and the penalization parameter. Values of α are typically below 0.1. Large complexity values result in smaller trees. For our best decision-tree model, we have experimented with complexity ranging from 0.1 to 0.0001, resulting in the decision rules below.

Table 12: Cost of complexity analysis

SSE (or cost of complexity)	Decision rules	Accuracy
> 0.03	Awarded amount < 23 No: 95% chance of taking offer	72%
0.025	Above criteria + If awarded amount < 23 and age > 60, 83% chance of taking offer	74%
0.001	Rules shown in Figure 2	88%
0.0006	10 rules (overcomplex)	88% (difference in optimal model below 0.1%)

Calculating the relative X error for various costs of complexity results in an optimum of 8 splits at cost of complexity 0.00068 with 10+ rules. As the confusion matrix is only marginally better than the decision tree with 5 nodes, the latter is preferred.

Our findings show a moderate improvement of the RFM when additional (demographic) variables are included, thus confirming Bucklin and Van Den Poel's (2005) findings. In their analysis, they argue that demographics (as an extension of the RFM) have only been used due to convenience and that other customer variables should be included to improve the models' performance. Such variables would require primary research which would considerably increase the cost and time needed for such an analysis. Hosseini et al. (2010) and Yeh et al. (2009) also confirmed that adding customer characteristics yields better model performance.

5. CONCLUSION

5.1. Limitations and further research

In our analysis, we used test sampling from the initial database of 5,000 entities split into 2,500 positive (took offer) and 2,500 negative (did not take offer) cases. Averaging multiple models created from several instances of random sampling could improve the models and is computationaly feasible for logit, decision trees, and naïve Bayes models. SVM models are computationally more expensive (around 1,000 times with respect to our data) and would require much more computing power.

Our research shows decision trees to be the best decision-modeling algorithms for the given setup, ensuring average accuracy of 88%, followed by logit (max. of 83%), naive Bayes (max. of 69%), and SVM (max. of 59%). All algorithms perform best using RFM in combination with demographic variables. Although all demographic variables are statistically significant, the most explanatory ones are the awarded amount (gift granted) and age of the participant. Interestingly, Frequency and Monetary are not present in the best-performing model. Essentially, this indicates that consumers should not be differentiated by the amount spent or by repetition of purchase, but rather by the time elapsed since their last purchase.This suggests that frequent communications to recent customers would yield the best return on direct communication investments.

Our research also shows that the basic R+F+M segmentation that the company is using can generate good results (78% accuracy), which in turn suggests that the use of demographic data (if expensive) could be unnecessary, depending on the opportunity cost of a lost customer and the cost of communication distribution and creation.

As RFM methods for scoring and the number of clusters are not strictly predefined, one could determine the optimal number of segments and the manner in which the customer base should be scored for R, F, and M using analytical algorithms (supervised and unsupervised – Gustriansyah et al., 2019). Such an analysis would optimize the (usually arbitrarily predefined) scoring and number of segments to the company's data.

Furthermore, our research is limited to the classification methods analyzed. Further studies should include additional ML methods, including but not limited to random forest and neural networks. The methods presented in this paper (support vector machines, decision trees, logistic regression and naïve Bayes) could be further improved using ensemble methods (i. e., bootstrap aggregating – bagging, boosting, Bayesian model averaging, bucketing, etc. – Dietterich & Oregon, 2000).

5.2. Concluding remarks

Extending a model's input set of variables can lead to significant improvements in performance. If such data is also readily available to the researcher (from the company's internal data), the cost of such operation is negligible. The models we have presented in our research are aplicable to the data of the company examined and can thus not be generalized (i.e., we cannot claim that the decision-tree method is generally better than SVM). However, the models can easily be adjusted to a company's RFM database if other customer data is available to be used. Every company which monitors the efficency of its marketing communication activities using its customer database can benefit from our research. The present research, however, does not address any of the more general pitfalls of RFM modeling, such as omission of non-customers, bias, etc.

REFERENCES

1. Bauer, C. L. (1988). Customer Purchase Model. *Journal of Direct Marketing*, *2*(3), 16-24.
2. Bing, F., Xin-Xingi, L. U. O., & Ke, G. (2006). A Novel Approach to Customer Classification Based on Reliability. *Proceedings of International Conference on Management Science and Engineering*, 1035-1040.
3. Blattberg, R. C., Kim, B-D., & Neslin, S. A. (2008). *Database Marketing*. Boston, MA: Springer.
4. Boser, E. B., Guyon, M. I., & Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*.
5. Bucklin, W., & Van Den Poel, D. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*(1), 252-268.
6. Bult, J. R., & Wansbeek, T. (1995). Optimal Selection for Direct Mail. *Marketing Science*, *14*(4), 378-394.
7. Chan, C-C. H. (2005). Online auction customer segmentation using a neural network model. *International Journal of Applied Science and Engineering*, *3*(2), 101-109.
8. Dietterich, T. G., & Oregon. (2000). Ensemble methods in machine learning. *Oncogene*, *12*(2), 1-15.
9. Faganel, A., & Costantini, M. (2020). CRM adoption factors in the gaming industry. *Management*, *25*(1), 1-23.
10. Flach, P. A. (2001). On the state of the art in machine learning: A personal review. *Artificial Intelligence*, *131*(1/2), 199-222.
11. Gascuel, O. (2000). On the optimization principle in phylogenetic analysis and the minimum-evolution criterion. *Molecular Biology and Evolution*, *17*(3), 401-405.
12. Gönül, F. F., Kim, B. D., & Shi, M. (2000). Mailing smarter to catalog customers. *Journal of Interactive Marketing*, *14*(2), 2-16.
13. Gustriansyah, R., Suhandi, N., & Antony, F. (2019). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(1), 470-477.
14. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Boston, MA: Springer.
15. Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*(7), 5259-5264.
16. Hughes, A. M. (1994). *Strategic database marketing*. London: Probus Publishing.
17. Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, *29*(2), 119.
18. Kim, G., Chae, B. K., & Olson, D. L. (2013). A support vector machine (SVM) approach to imbalanced datasets of customer responses: Comparison with other customer response models. *Service Business*, *7*(1), 167-182.
19. Kuhn, R., & De Mori, R. (1995). The Application of Semantic Classification Trees to Natural Language Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(5), 449-460.

20. Kumar, M. (2022). *How to tune a decision tree?* Retrived from https://towardsdatascience.com/how-to-tune-a-decision-tree-f03721801680 (accessed March 2022).
21. Leeflang, P. S. H., Wieringa, J. E., Bijmolt, T. H. A., & Pauwels, K. H. (2017). *Advanced methods for modelling markets: International series in quantitative marketing*. Boston, MA: Springer.
22. Li, S. T., Shue, L. Y., & Lee, S. F. (2008). Business intelligence approach to supporting strategy-making of ISP service management. *Expert Systems with Applications*, *35*(3), 739-754.
23. Liu, D. R., & Shih, Y. Y. (2005a). Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. *Journal of Systems and Software*, *77*(2), 181-191.
24. Liu, D. R., & Shih, Y. Y. (2005b). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information and Management*, *42*(3), 387-400.
25. Loh, W. Y., & Shin, Y. S. (1997). Split selection methods for classification trees. *Statistica Sinica*, *7*(4), 815-840.
26. McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, *60*(6), 656-662.
27. Miglautsch, J. R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, *8*(1), 67-72.
28. Miglautsch, J. R. (2002). Application of RFM principles: What to do with 1-1-1 customers? *Journal of Database Marketing & Customer Strategy Management*, *9*(4), 319-324.
29. Pavlić, I., Vojvodić, K., & Puh, B. (2020). Consumer segmentation in food retailing in Croatia: A latent class analysis. *Market-Trziste*, *32*(Special Issue), 9-29.
30. Rahim, M. A., Mushafiq, M., Khan, S., & Arain, Z. A. (2021). RFM-based repurchase behavior for customer classification and segmentation. *Journal of Retailing and Consumer Services*, *61*, 102566.
31. Rokach, L., & Maimon, O. (2007). *Data mining with decision trees* (2nd ed.). London: World Scientific.
32. Shokeen, V., Yadav, M., & Kumar Singhal, P. (2018). Comparative Analysis of FLR Approach Based Inverse Tree and Modern Fisher-Yates Algorithm Based Random Interleavers for IDMA Systems. *Proceedings of the 8th International Conference Confluence 2018 on Cloud Computing, Data Science and Engineering, Confluence 2018*, 447-452.
33. Sohrabi, B., & Khanlari, A. (2007). Customer lifetime value determination based on RFM model. *Marketing Intelligence and Planning*, *34*(4), 446-461.
34. Suh, E. H., Noh, K. C., & Suh, C. K. (1999). Customer list segmentation using the combined response model. *Expert Systems with Applications*, *17*(2), 89-97.
35. Tabaj Pušnar, T., & Bratina, D. (2018). Improving Direct Marketing Activities Effectiveness Using Analytical Models: RFM vs. Logit Model on a Casino Case. *Management*, *13*(4), 323-334.
36. Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications*, *27*(2), 265-276.
37. Vapnik, V., & Cortes, C. (1995). Support-Vector Networks. *Machine Learning*, *20*, 273-297.
38. Wei, J., Lin, S., & Wu, H. (2010). A review of the application of RFM model. *African Journal of Business Management*, *4*(19), 4199-4206.
39. Yang, A. X. (2004). How to develop new approaches to RFM segmentation. *Journal of Targeting, Measurement and Analysis for Marketing*, *13*(1), 50-60.
40. Yeh, I. C., Yang, K. J., & Ting, T. M. (2009). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, *36*, 5866-5871.
41. Zaheer, N., & Kline, M. (2018). Use of lifestyle segmentation for assessing consumers' attitudes and behavioral outcomes towards mobile advertising. *Market-Trziste*, *30*(2), 213-230.
42. Zoubir, A. M., & Iskandler, D. R. (2007). Bootstrap methods and applications. *IEEE Signal Processing Magazine*, *24*(4), 10-19.

MARKET TRŽIŠTE