



HAL
open science

Customer Churn Prediction in FMCG Sector Using Machine Learning Applications

S. Nazlı Günesen, Necip Şen, Nihan Yıldırım, Tolga Kaya

► **To cite this version:**

S. Nazlı Günesen, Necip Şen, Nihan Yıldırım, Tolga Kaya. Customer Churn Prediction in FMCG Sector Using Machine Learning Applications. 8th IFIP International Workshop on Artificial Intelligence for Knowledge Management (AI4KM), Jan 2021, Yokohama, Japan. pp.82-103, 10.1007/978-3-030-80847-1_6 . hal-04041361

HAL Id: hal-04041361

<https://inria.hal.science/hal-04041361v1>

Submitted on 22 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Customer Churn Prediction in FMCG Sector Using Machine Learning Applications

S. Nazlı Günesen¹[0000-0002-9960-5370], Necip Şen¹[0000-0003-4201-6505], Nihan Yıldırım¹[0000-0002-6279-3849] and Tolga Kaya¹[0000-0003-4841-482X]

¹ İstanbul Technical University, Maslak, 34467 Sarıyer/İstanbul, Turkey
gunesen16@itu.edu.tr, senn15@itu.edu.tr, yildirimni@itu.edu.tr,
kayatolga@itu.edu.tr

Abstract. Non-contractual setting and many brands and alternative products make customer retention relatively more difficult in the FMCG market. Besides, there is no absolute customer loyalty, as most buyers split their purchases among several almost equivalent brands. Thereby, this study aims to probe the contribution of various machine learning algorithms to predict churn behaviour of the most valuable part of the existing customers of some FMCG brands (detergent, fabric conditioner, shampoo and carbonated soft drink) based on a real dataset obtained in the Turkish market over the two successive years (2018 and 2019). In this context, exploratory data analysis and feature engineering are carried out mostly to build many predictive models to reach consistent and viable results. Further, RFM analysis and clustering techniques with K-Means clustering are employed to generate meaningful insights for business operations and marketing campaigns. Lastly, revenue contributions of improved customer retention can be achieved, utilising actionable intelligence created by the churn prediction.

Keywords: Machine Learning, Business Intelligence, FMCG, Churn Prediction, Customer Retention, Customer Loyalty, RFM Analysis, K-Means Clustering.

1 Introduction

The competitive landscape of today's marketplaces entails companies, regardless of industries they are in, to emphasize retaining their existing customers. FMCG companies face extra inconvenience in managing customer retention due to the non-contractual customer relationship setting in retail markets. Unsurprisingly, the inclination to churn is directly proportional to the switching cost. The lesser the switching cost, the higher the tendency to churn [1]. Therefore, FMCG firms experience difficulties in predicting customer churns and in taking preventive actions accordingly. Consequently, they lose their hard-gained customers to competitors since they fail to respond proactively.

A modest improvement in customer churn projection precision can yield a significant positive impact on a company's profitability. The cost of acquiring a new customer is 5 to 6 times higher than retaining the existing one [2]. There are much more substantial shreds of evidence in the relevant literature for cost rate of new customer acquisition

to customer retention, such as 12-fold [3], 20 times [4] and even 25 times [5]. So, it is an indisputable fact that gaining a new customer is much more expensive than keeping an existing one. Moreover, only a 5% improvement in customer retention rate would result in a 25% to 95% profit increase [5].

In today's complex FMCG distribution environment, there are numerous factors causing dissatisfaction resulting in a customer churn. For example, quality defects, the inefficiency of distribution channels, new product offerings, inter-customer dynamics through social media, aggressive competitive campaigns, regulations, price changes etc. [4]. Accordingly, identifying churners who behave similarly and relevant root causes of their action enables companies to improve customer retention [6]. That's why creating accurate and comprehensible churn prediction models providing actionable business intelligence by detecting early symptoms to identify customers with a high probability to leave is a must-have practice for FMCG companies [4]. Thus, they can revise their marketing strategies proactively to address identified customers' concerns worth keeping to improve their bottom-line profits [7].

2 Literature Review

2.1 General Overview

In the last decade, the ingenuity of processing raw data to reveal hidden value inside providing actionable intelligence has been in humankind's spotlight on a global scale. The impact of analytics that is the engine of this initiative has already become mainstream and changed societies' way of living and the business world's capabilities, irrespective of the industry they are in [8].

On the other hand, the fierce competition of today's multichannel, digitalized markets entails FMCG companies to be more attentive to retaining their hardly gained existing customers to stay ahead of the game [9]. Moreover, customers can reach many alternative products offered by several competitive companies and easily switch from one to another or partition their purchases amongst a selected group of producers in a non-contractual nature of FMCG markets [10]. For this reason, they aim to make optimum use of their limited sales and marketing resources to attain the purpose of elongating the duration of their customers' engagement with their products.

However, many companies do not emphasize retaining customers adequately in their strategies and still perform traditional marketing planning processes prioritizing acquiring new customers and increasing market share. Further, they discern an increase in their churn rates 6 to 8 months later than they actually dissatisfy the customer and consequently, they are six months late to improve their performance [5]. Companies practising relationship marketing to maintain long-term relationships with existing customers, thus increasing their revenue stream, will positively impact the high customer retention rate on profitability [11].

In this context, creating actionable customer churn insight from raw data by predictive analytics based on machine learning techniques is one of the leading domains that have been frequently attempted by many researchers [12]. Nevertheless, there is still

room for further research into accurate and comprehensible customer churn prediction models to interpret customers' purchasing behaviour more precisely [13].

2.2 Definition of Customer Churn

Customer churn is the customers' act of abandonment of the existing relationship with the company they currently purchase from [14]. It merely is customer movement from one company/brand to another [15]. In other words, customer churn is a customer's action discontinuing purchasing a brand in a definite time interval. It reveals an indication concerning the duration of customer's engagement with product/s of a company and, thereby, the customer's lifetime value (LV) that is projected net present value of total spending of a customer for the product during his/her entire relationship with the company [16]. LV propounds that all customers are not equally important in terms of profit potential in future. High LV customers bringing along more significant profit opportunities have to be detected, and resources need to be allocated to retain them [17].

Moreover, due to continuously increasing product choices and digital channels, well-informed and empowered consumers of today have ceased following the marketing funnel concept where they enter with a shortlist of brands in their minds and assess them to narrow options. At the same time, they move through and leave with the one brand they decided to buy [18].

On the other hand, defining the condition that classifies a customer as a churner is not easy, particularly in the FMCG sector's non-contractual setting, as most consumers purchase from several suppliers concurrently. Usually, a threshold based on the customer purchasing activity is set by a business preference, and any customer with lower purchasing activity is considered as a churner [19]. Hence, some cases categorize customers with no activity in the given time as churners [20], while others accept customers lowering their purchasing activity below the threshold as churners.

Some other studies focus only on the churning likelihood of loyal customers who look promising in their potential of future revenue contribution [21]. In this context, they evaluate partial defections in loyal customers as the early signals of total churns and focus on behaviourally loyal ones whose purchase frequency is above average and the standard deviation to mean for their successive purchases below average [22].

However, significant differences concerning volume and evolution of spending in loyal customers group and the idea of defining them as a whole, as the most valuable group based on the duration of their lifecycle (LC), are only questionable. For this reason, churn analysis should also consider the customers who are not classified as loyal since they purchased a large number of products over a short period in the past, but then reduced or discontinued their purchasing (hence, denominated as "butterfly") because they have a high monetary value in the past [23].

Consequently, estimating a recently acquired customer's future LC slope to predict future spending evolution based on the initial purchase information provides valuable qualitative insights to reshape marketing strategy [23]. Classifying churners with a threshold is not always meaningful, and the evolution of customer purchasing patterns is also a significant factor [19].

2.3 Substantiality of Customer Churn

Today's consumers are more demanding and need to be responded with a more customized treatment [24]. Companies spend significant effort to attract new customers. However, the total spending of an average new customer within the first year or two of his/her LC with the company just offsets the acquisition cost. On the contrary, a 5% increase in customer retention rate adds at least 3 years to the average LC of customers. Besides, only a 2% improvement in customer retention improves overheads by 10% [25].

The positive impacts of customer retention on business performance are notably referred to in the literature. A compilation of precedents to give an idea are as follows: companies that succeeded in retaining their customers do not have to spend much effort to acquire new customers, and thereby they can concentrate on concerns of their existing customers to develop stronger relations [26], cost of acquiring new clients is at least 5 to 6 times higher than that of retaining existing customers [27; 28], higher profits are generated out of customers with long LC with the brand [21], loyal customers bring along new customers by word of mouth in favour of the brand they have purchased, while churned customers might negatively influence potential customers with their biased product comments [21; 20], long-term customers can tolerate price differences and are less responsive to competitive marketing campaigns [28; 12], even a small increment in customer retention can culminate in a remarkable profit increase [26].

For an appropriate customer retention strategy, it is needed to analyse the factors motivating customers to churn. Those who are profitable to retain and continue to serve must also be identified [25]. Therefore, viable, conceivable, and accurate predictive churn models are essential to detect customers' intentions to leave and identify causes behind their churn behaviour [34]. Some recent studies performed for the FMCG sector in this context are summarized in Table 1.

Table 1. Some recent studies at a glance [relevant to customer churn prediction in the FMGC sector].

Reference	Models Used	Dataset	Remarks
Buckinx et al. (2002) Using machine learning techniques to predict defection of top clients.	Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, C4.5, Neural Networks, Naive Bayes.	Private Dataset: Individual records of 158,884 customers with a loyalty card (85 per cent of total customers) within ten months (# of features: 32).	Partial defection is the initial signal of total churn in future. Neural network outperforms other classification techniques in terms of both PCC and AUC metrics. The number of past shop visits and the time between past shop incidences are best performing variables concerning predictive power.
Buckinx et al. (2005) Customer base analysis: partial defection of behaviourally loyal cli-	Logistic regression, Automatic Relevance Determination (ARD) Neural Networks and Random Forest.	Private Dataset: Individual records of 158,884 customers with a loyalty card (85 percent of total	PCC and AUC are used to evaluate classifier performance of predicting partial defection. No remarkable performance differences are observed among alternative classification

ents in a non-contractual FMCG retail setting [29].		customers) within ten months (# of features: 32).	methods used. RFM variables are the best predictors to be used for partial customer defection.
Lessmann & Voss (2008) Supervised Classification for decision support in customer relationship management [30].	N/A	N/A	Extraordinary approaches (such as learning from imbalanced data and cost-sensitive learning) and algorithms are necessary to overcome complications caused by imbalanced class and cost distributions. Basic linear models may fall short of explanatory power in complex CRM settings.
Burez & Van den Poel (2009) Handling class imbalance in customer churn prediction [31].	Gradient Boosting Machine and Weighted Random Forest (with random and advanced under-sampling), CUBE Sampling Technique, Random Forest, Logistic Regression.	Private Dataset: Purchase data of 32,371 customers with a 25.15% churn rate (#of features: 21).	Under-sampling can lead to improved prediction accuracy, primarily when evaluated with AUC. The advanced sampling technique CUBE does not increase predictive performance. Weighted random forest performs noticeably better than the random forest.
Miguéis et al. (2013) Customer attrition in retailing: An application of multivariate adaptive regression splines [32].	Logistic Regression and Multivariate Adaptive Regression Splines (MARS).	Private Dataset: Purchase data and demographics of 130,284 customers with a loyalty card within two years.	MARS performs better than Logistic regression when variable selection procedures are not used. However, MARS loses its superiority when logistic regression is conducted with stepwise feature selection.
Coussement (2014) Improving customer retention management through cost-sensitive learning.	Logistic Regression Classifying Technique and Non-Sampling Based Cost-Sensitive Learning Techniques [Direct Minimum Expected Cost (DMECC), Metacost, Thresholding and Weighting].	Private Dataset: Records of 100,000 customers including transactional, socio-demographical and client/company interaction information (with 30.86% churn incidence).	Total misclassification cost, as a churn prediction evaluation measure, is a crucial metric. Cost-sensitive learners are benchmarked for different cost and churn levels in terms of indicating churn prediction reliability.
Jahromi et al.(2017) Managing B2B customer churn, retention and profitability [33].	Decision Tree, Logistic Regression, Adaptive Boosting.	Private Dataset: Transactional records of 11,021 business customers in a year.	Adaptive boosting considerably outperforms the decision tree model concerning prediction accuracy. However, its accuracy performance superiority is negligible compared to that of logistic regression.
Calciu et al. (2015) Recognising dangerous drop out incidents as opposed to accidents to improve the	Stochastic Model (Pareto NBD).	Private Dataset: 199,352 randomly selected customers from the top four	Drop Out Incidents (DOIs) of customers are used to estimate probability of being active and high Duration between DOIs (DDOI) are considered as the signals of potential customer

efficiency of triggers reducing customer churn [34].		RFM segments within 3,5 years.	churns. Allocating marketing resources on active users with risk of churn is suggested.
Sulistiani & Tjahyanto (2017) Comparative analysis of feature selection method to predict customer loyalty [35].	Random Forest.	Private Dataset: Consumer questionnaire (# of respondents: 386, # of features 26).	Feature selection methods can considerably affect the prediction accuracy of random forest algorithm. The Chi-square feature selection method improved random forest accuracy up to 83.2%.
Shoab (2018) Customers' churn prediction in a retail store through machine learning algorithms [36].	Logistic Regression, KNN, Decision Tree, Random Forest, XG Boost.	Private Dataset: 880,000 transactions of 180,000 customers within two months.	XG boost, random forest and decision tree are the best performing classifiers, respectively. K-means clustering technique is used for segmentation. Churn probability is found higher for the high-value customer segment.

3 Data Preparation

3.1. Dataset

A real-life dataset obtained through field research for four selected FMCG brands (detergent, fabric conditioner, shampoo and carbonated soft drink) over the two successive years [2018 (P1 period) and 2019 (P2 period)] was used. These nineteen features are included: age, gender, marital status, location, education, job status, region, duration of consumption, point of purchase, basket value, the basket of goods, purchase value of the brand, purchase value of the product category, the quantity of the brand purchased, the quantity of the product category purchased, frequency of buying, recency of buying, number of equivalent brands consumed, and purchasing points used.

3.2. Sampling

Instead of focusing on the entire customer base, we selected the most profitable (valuable) and behaviourally loyal customers. For determining this segment for each brand, we designated the first 2/3 of customers by their purchasing value in P1 and P2 periods.

Afterwards, we identified the customers-in-common included in the selections of both successive years, and we nominated them as the valuable customer base. For example, the detergent brand number of the first 2/3 of customers in terms of purchasing value is 8414 out of 12622 in P1 and 8738 out of 13105 in P2, and the number of customers who are common in both periods is 3936. In this way, we selected the heavy buyers in common in the top 2/3 of customers regarding purchasing value for the P1 and P2 periods to sample the detergent brand's valuable customers.

3.3. Loyalty Definition

Loyalty was defined based on the threshold of the ratio of the brand's quantity purchased to the quantity of the product category purchased. The valuable customers whose loyalty threshold are higher than 0.33 in the P1 period were accepted as loyal. By using the loyalty threshold for elimination, 616 detergent customers out of 3936, valuable customers were identified as loyal. In this way, 466 laundry softener customers, 937 shampoo customers, and 2200 carbonated soft drink customers were selected as samples representing loyal customers.

3.4. Churn Definition

In a non-contractual setting of the FMCG market, customers can easily switch from one brand to another or partition their purchases amongst a selected group of products. For this reason, churn must be defined carefully by paying regard to purchase tendency. In this context, we divided loyal customers into two groups: (1) Customers whose loyalty ratio is between 0.33 and 0.5 in the P1 period are accepted as churn if their loyalty ratio is 0 (no purchase from that brand) in the P2 period. (2) Customers whose loyalty ratio is higher than 0.5 in the P1 period are accepted as churn if their loyalty ratio in the P2 period is at least 0.5 less than their P1 loyalty ratio. To exemplify, 0.2 and lower P2 loyalty ratio will be accepted as a churn while greater than 0.2 P2 loyalty ratio is taken as no-churn for a P1 loyalty ratio of 0.7.

4 Method

The procedures explained in this section were applied for all selected FMCG brands (detergent, fabric conditioner, shampoo and carbonated soft drink). However, only the procedures conducted for the shampoo brand have been presented.

4.1. RFM Analysis

At first, the shampoo customers were grouped by a cluster dendrogram after scaling the relevant data. The following output indicates eight different groups.

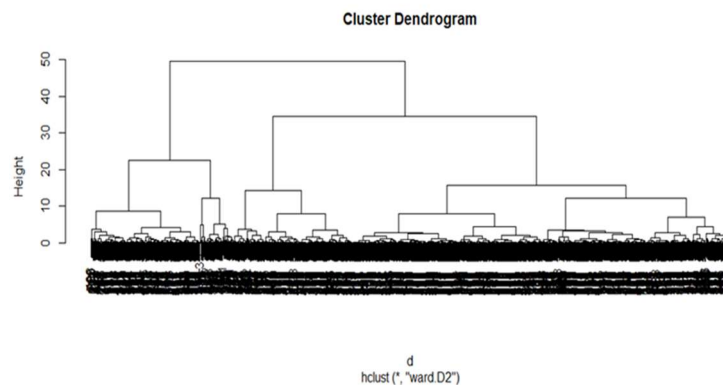


Fig. 1. Cluster dendrogram.

These groups are illustrated in the following 3D RFM scatter plot where X represents Frequency, Y represents the Recency, and Z represents the Monetary.

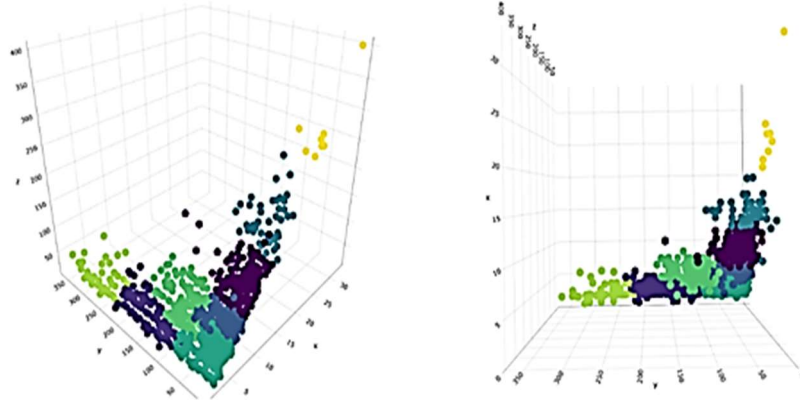


Fig. 2. 3D RFM.

4.2. Segmentation with Clustering

Three different approaches are used to determine an optimal number of clusters. According to the total within-cluster sum of squares graph, the slope starts to flatten after 8 groups. The gap statistics method predicates that at least 6 clusters would be sufficient. As a final check, the average silhouettes graph remarks that 8-9 clusters would be appropriate for customer segmentation (See Fig. 3).

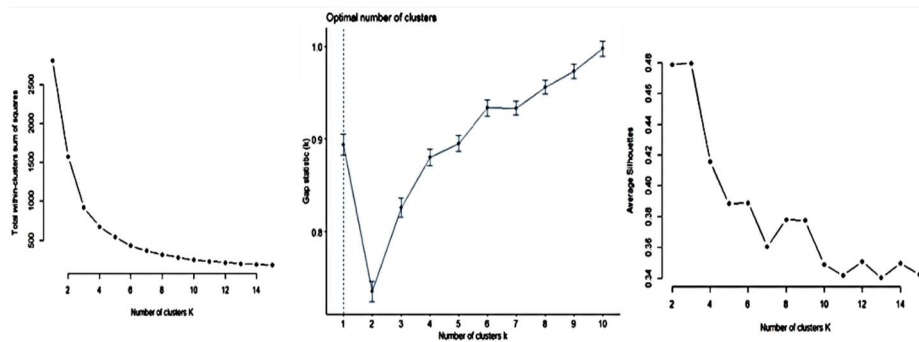


Fig. 3. The total within-clusters sum of squares graph (left), gap statistics graph (middle), and average silhouettes graph (right).

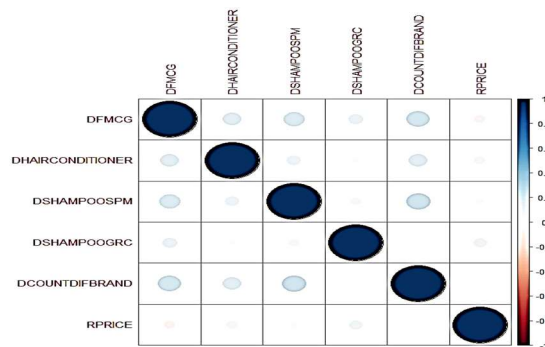
4.3. Exploratory Data Analysis, Visualisation and Feature Engineering

After trying various combinations of features, the best-performing feature set was defined as shown in Table 2.

Table 2. Table of features used.

LABEL	VARIABLES FORMULA	CATEGORY
SES_GROUP	SES_GROUP	FACTOR
PAX	PAX	FACTOR
KENT_KIR	KENT_KIR	FACTOR
PS_GENDER	PS_GENDER	FACTOR
EDU_LEVEL	EDU_LEVEL	FACTOR
AGE	AGE	FACTOR
REGION	REGION	FACTOR
DHAIRCONDITIONER	[HAIRCONDITIONER(P2)-HAIRCONDITIONER(P1)*1.2]	NUMERIC
DFMCG	[FMCG_TOTAL(P2)-FMCG_TOTAL(P1)*1.2] [FMCG_TOTAL(P1)*1.2]	NUMERIC
DSHAMPOOSPM	[SHAMPOO_SPM(P2)-SHAMPOO_SPM(P1)*1.2]	NUMERIC
DSHAMPOOGRC	[SHAMPOO_GRC(P2)-SHAMPOO_GRC(P1)*1.2]	NUMERIC
RPRICE	[SHAMPOOPRICE(P2)-SHAMPOOPRICE(P1)*1.2]	NUMERIC
DCOUNTDIFBRAND	[DCOUNTDIFBRAND(P2)-DCOUNTDIFBRAND(P1)]	NUMERIC

1. With `as.factor()` function, all character columns were transformed into factor columns.
2. `{sum(is.na())}` function is used to look for sum of “NA” numbers in dataset.
3. `summary()` shows the information about numeric and factor variables [202 CHURN, 735 NO CHURN labels in the dataset]
4. `Unselect` function is used to eliminate all columns, including factors to show correlations and densities of all numeric variables with `ggpairs()` function (see Fig. 5).
 - DSHAMPOOGRC feature’s distribution is not expanded because the 0 values have dominated the whole column, so these columns seem like such. Kurtosis of these features’ distributions is positive and higher than the normal distribution’s kurtosis.
 - The Farness of the correlations shown above from 1 and -1 indicates that there is no correlation among features. The correlation value between RPRICE and DCOUNTDIFBRAND is so close to 0. This finding is surprising because they are expected to be parallel.
 - The scatters between each pair show the distribution of the observations among them, and the line on them indicates the direction of observations.
5. The graph in Figure 4 is built with the `corplot()` function showing the correlations among features after scaling. `Ggpairs` plot is also presented in Figure 5.

**Fig. 4.** Correlation plot.

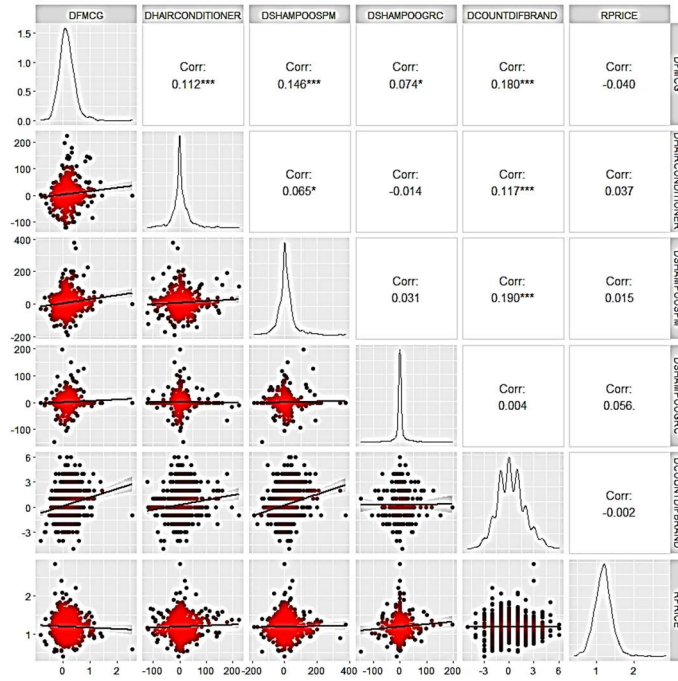


Fig. 5. Ggpairs plot.

4.4. Application of Models

In this study, various prediction models based on eight different machine learning algorithms (logistic regression, logistic regression with k fold, boosted logistic regression, linear discriminant analysis, quadratic discriminant analysis, decision tree, random forest, and extreme gradient boosting) were built, tested and optimized. Ultimately, we found the best performing machine learning technique with the highest meaningful accuracy for each FMCG brand. The accuracy comparison of all models used for each brand is given in Table 13.

The first applied method was decision tree, and before pruning; the training set accuracy, sensitivity and specificity were 81%, 21% and 98%, respectively. Upon pruning, test set accuracy was 77%, sensitivity was 10%, and specificity was 96%. The final decision tree generated after pruning was as follows:

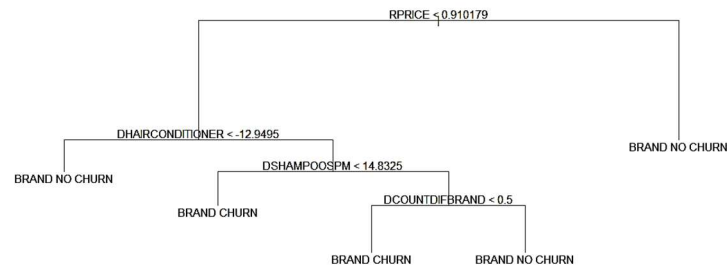


Fig. 6. Decision tree

The most dominating variable was the RPRICE, and its terminal value was calculated as 0.91. After pruning, the test set accuracy was 77%, sensitivity was 13%, and the specificity was 95%.

Table 3. Decision tree test set confusion matrix after pruning

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	8	11
BRAND NO CHURN	53	210

Application of logistic regression made it possible to understand churn behaviour better by showing variables' contributions to the churn prediction (Table 4);

Table 4. Logistic regression variable contribution

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.474	0.620	-0.765	0.444
PS_GENDERFemale	0.547	0.342	1.598	0.110
PS_GENDERMale	1.770	0.624	2.834	0.005 **
REGIONBlacksea	0.766	0.512	1.495	0.135
REGIONCentral	0.081	0.341	0.239	0.811
REGIONEast and Southeast	0.099	0.343	0.287	0.774
REGIONMarmara	0.374	0.295	1.266	0.206
REGIONMediterranean	0.739	0.349	2.114	0.034 *
DFMCG	0.365	0.373	0.979	0.328
DHAIRCONDITIONER	-0.009	0.003	-2.926	0.003 **
DSHAMPOOSPM	0.005	0.002	2.223	0.026 *
DSHAMPOOGRC	0.015	0.006	2.522	0.012 *
RPRICE	0.753	0.414	1.819	0.069 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

PS_GENDER, REGION, DFMCG, DHAIRCONDITIONER, DSHAMPOOSPM, DSHAMPOOGRC, RPRICE were the most contributing variables. PS_GENDERERKEK, REGIONMediterranean, DFMCG, DSHAMPOOSPM, DSHAMPOOGRC and PRICECHANGE have positive coefficients. The findings can be interpreted as all else being equal; (1) males, people living in the Mediterranean

region, (2) customers with increased total FMCG expenditures, (3) customers with increased supermarket shampoo expenditures, and (4) customers with increased grocery store shampoo expenditures are more likely to churn. Moreover, RPRICE has a positive coefficient which shows that all else being equal; when the average shampoo category price increases, customers are more likely to churn. DHAIRCONDITIONER has a negative coefficient, which indicates: all else being equal; customers with increased hair conditioner expenditures are less likely to churn. From these variables, the most critical churn factor is the increase in the product category's average price and customer's region since their coefficients are the biggest ones. The training set accuracy was 78%, sensitivity was 0.05%, and specificity was 98%. The test set accuracy was 77%, sensitivity was 0%, and specificity was 78%.

Table 5. Logistic regression test set confusion matrix

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	0	61
BRAND NO CHURN	2	219

The prediction power of the algorithm is also illustrated with a ROC curve;

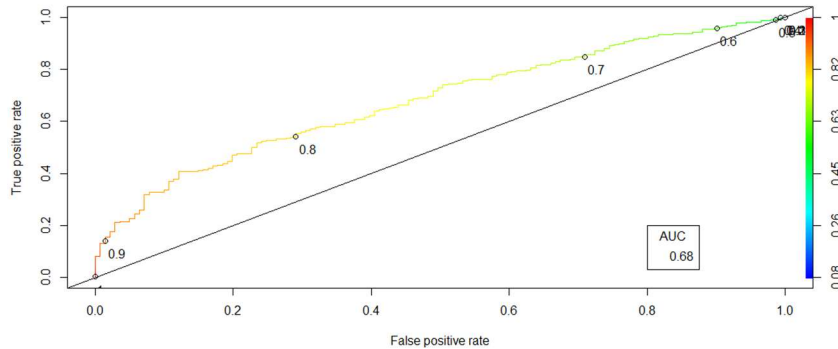


Fig. 7. ROC Curve

To apply the random forest technique, the dataset was divided into test and training parts same as the decision tree method. However, in the random forest algorithm `SmoteClassif()` function was used to create artificial observations that belong to the "CHURN" label. When the model was performed, k-fold helped to build it with the value of 5. The other crucial concern in applying the method was to determine optimum `n`tree and `m`try values.

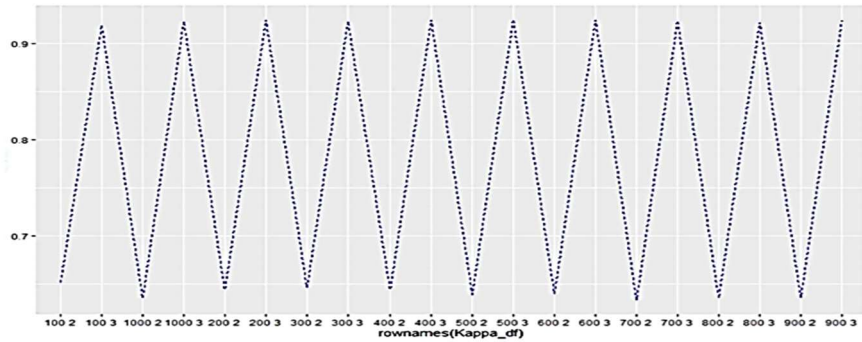


Fig. 8. Random forest classifier plot

Tuning parameter 'mtry' was held constant at a value of 3 because of the highest accuracy rate. The ntree parameter also was chosen as 1000.

Accuracy, sensitivity and specificity rates were calculated as 74%, 15% and %91, respectively.

Table 6. Random classifier confusion matrix

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	9	20
BRAND NO CHURN	52	201

The random forest technique gives insights about churn behaviour because it orders variables according to their contribution.

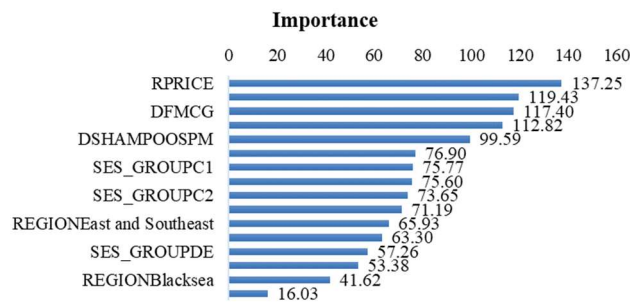


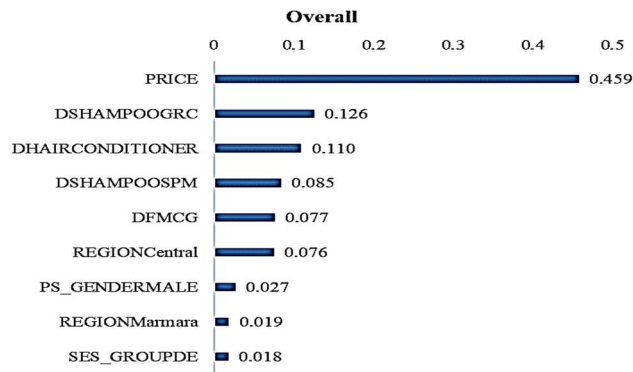
Fig. 9. Random classifier variable contribution

XG boost algorithm was another applied method, and accuracy, sensitivity and specificity rates were calculated as 87%, 17% and %93, respectively.

Table 7. XG boost confusion matrix

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	4	21
BRAND NO CHURN	19	273

XG Boost provides insights concerning churn behaviour as it puts variables in order according to their contribution.

**Fig. 10.** XG boost variable importance

RPRICE (change in average product category price), DSHAMPOOGRC (change in grocery store product category expenditure), DHAIRCONDITIONER (change in hair conditioner expenditures) were the most important factors.

For boosted logic; the test set accuracy was 87%, sensitivity was 39% and specificity was 91%.

Table 8. Boosted logic confusion matrix

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	9	27
BRAND NO CHURN	14	267

QDA model calculated the test set accuracy, sensitivity and specificity as 70%, 48% and 72%, respectively.

Table 9. QDA confusion matrix

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	11	82
BRAND NO CHURN	12	212

LDA was another applied model and it calculated the test set accuracy, sensitivity and specificity as 87%, 0% and 94%, respectively.

Table 10. LDA confusion matrix

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	0	17
BRAND NO CHURN	23	277

Finally, logistic regression with k fold was applied, and the test set accuracy, sensitivity and specificity were calculated as 77%, 0% and 98%, respectively.

Table 11. K-fold logistic regression confusion matrix

	BRAND CHURN	BRAND NO CHURN
BRAND CHURN	0	11
BRAND NO CHURN	141	503

A summary table of the models applied for the shampoo category is illustrated in the following table.

Table 12. Model comparison for shampoo category

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Error Rate (%)	Precision (%)	Negative Predictive Value (%)
Decision Tree	77.305	13.115	95.023	22.695	42.105	79.848
Logistic Regression with k fold	76.794	0.000	97.860	23.206	0.000	78.106
Logistic Regression	77.660	0.000	78.214	22.340	0.000	99.095
Random Forest	74.468	14.754	90.950	25.532	31.034	79.447
XG Boost	87.382	17.391	92.857	12.618	16.000	93.493
Boosted Logic	87.066	39.130	90.816	12.934	25.000	95.018
LDA	87.382	0.000	94.218	12.618	0.000	92.333
QDA	70.347	47.826	72.109	29.653	11.828	95.848

4.5. Comparison of Each Model for 4 FMCG Product Categories

Accuracies of the models applied with the base rates (“NO CHURN” ratio) for all product categories are demonstrated in the following table.

Table 13. Accuracy comparison of the models applied

Model	Detergent Accuracy (%)	Shampoo Accuracy (%)	Carbonated Soft Drink Accuracy (%)	Fabric Softener Accuracy (%)
Decision Tree	68.280	77.305	96.369	68.794
Logistic Regression with k fold	61.628	76.794	96.491	56.308
Logistic Regression	63.441	77.660	96.067	65.957
Random Forest	62.903	74.468	93.041	64.539
XG Boost	75.000	87.382	98.951	45.098
Boosted Logic	55.435	87.066	96.970	49.020
LDA	66.304	87.382	93.939	45.098
QDA	52.174	75.393	98.718	61.765
Base	0.646	0.784	0.965	0.631

4.6. Impact of Customer Retention Improvement on Revenue

The shampoo brand has approximately 2,100,000 customers, and the number of loyal customers was estimated with the help of the method used for sampling previously. Thus, the impact of the increase of retained loyal customers' percentage was estimated for each churn improvement scenario as follows;

Table 14. Reflection of churn improvement on customer retention

Churn Improvement	Retention Rate	# of Loyal Customers* Retained					
		Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
0,00%	78,44%	333.270	261.417	205.055	160.846	126.167	98.966
5,00%	83,44%	333.270	278.080	232.030	193.606	161.545	134.793
10,00%	88,44%	333.270	294.744	260.672	230.538	203.888	180.318
15,00%	93,44%	333.270	311.407	290.979	271.891	254.055	237.389
20,00%	98,44%	333.270	328.071	322.953	317.915	312.956	308.073

of Total Customers: 2.100.000 [estimated min.]

*Loyal Customers: Customers who constitute the segment realizing the top 2/3 of annual transaction value among total customers and also whose brand volume/product category volume ratio is greater than 0,33.

To estimate the impact of improved loyal customer retention on the revenue over the next five years, we multiplied the number of loyal customers obtained above by the loyal customers' average revenue contribution in the first year (which is TL 73.16 ₺). We then corrected the amount by the annual average CPI inflation rate for the following five years. Cumulative revenues were projected and plotted for 10% and 20% customer retention improvement scenarios in the following graph (Fig.11).

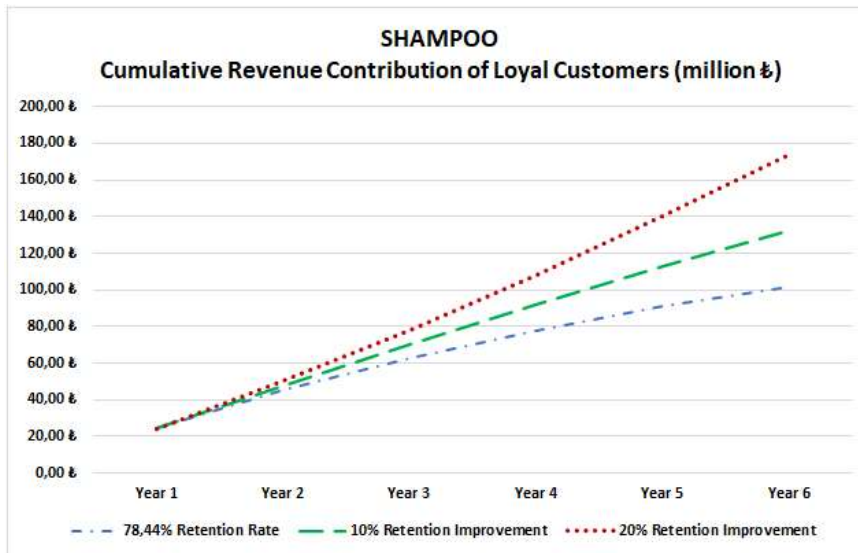


Fig 11. Impact of retention improvement on cumulative revenue contribution of loyal customers

5 Findings and Conclusion

5.1. Findings

Findings regarding predicting customer churn for the selected detergent, shampoo, fabric conditioner, and carbonated soft drink brands are as follows:

- For most of the product categories, customers with a relatively higher number of preferred brands are less likely to churn. This might indicate strong brand positioning of the selected brands in the conscious consumer segment.
- The category price change is the most critical factor referring to increased churn potentiality for all FMCG product categories.
- Changes in the discount store, grocery store and supermarket expenditures might give initial warning signals of churn for most product categories.
- In most of the product categories, being located in the Mediterranean region has a remarkable impact on churn potentiality. This might indicate a structural difference in distribution channels as the region is a leading touristic destination. Particularly for the shampoo brand, the Mediterranean region variable is significant, together with the change in grocery store expenditures. This designation supports this argument.

5.1.1 Detergent

- Customers with larger households are more likely to churn. A deeper analysis of this segment might be instrumental for the detergent brand to improve customer retention. Targeted marketing and advertising efforts can bring along significant returns. For instance, bundling the product with baby detergents or other child products might be considered.

5.1.2 Shampoo

- When hair conditioner expenditure increases, customers are less likely to churn. This indicates that complimentary personal care products together with the shampoo should also be taken into consideration.
- Customers with increased supermarket and grocery store expenditures are more likely to churn, possibly due to the increased number of available substitute products, distribution problems, or the store's wrong shelf location.

5.1.3 Fabric conditioner

- Discount markets, supermarkets and grocery stores are the most determinant purchasing channels, although channel effects differ according to the product's quality perception. Changing the preferred channel is decisive for churn. It may be useful to review the efficiency of product distribution channels.
- When there is a price increase, customers are less likely to churn. This might indicate that price was increased less than the customers' expectations or price increase is not a factor influencing loyal consumer behaviour for the products addressing upper segments. Moreover, this may be caused by the price-performance perception of upper segment products from the customer perspective.

- Customers located in the Central Anatolian region are more likely to churn. However, customers located in the Mediterranean region are less likely to churn. This finding might point out that fabric conditioner brand suffers from product distribution problems.
- Upper-class customers are found as more likely to churn for the fabric conditioner brand. Underlining cause may be the availability problem of the product or increased existence of substitute products.

5.1.4 Carbonated soft drink (CSD)

- When the alcoholic drinks expenditures increase, customers are more likely to churn. This might indicate that consumers perceive some alcoholic drinks as a substitute for CSD products. Thus, when investigating an edible product, identifying substitute products is critically important.
- Customers with increased salt expenditures are less likely to churn for the CSD brand. This indicates that analyzing the products that reflect a similar lifestyle (i.e., both salt and CSD are signs of an unhealthier lifestyle) can help foresee customers' behavioural changes.
- Upper-class customers are more likely to churn for the CSD brand. This can also be associated with the tendency of switching to a healthier lifestyle.

5.2. Conclusion

FMCG companies deal with difficulties to manage customer retention because the non-contractual setting of the FMCG market enables customers to switch to other brands easily due to the availability of many alternatives in various channels. For this reason, understanding the underlying factors of customer churn behaviour and identifying its early signals is critically essential. Although market research companies collect customer data by conducting surveys, transforming them to actionable business intelligence practices is challenging and time-consuming for FMCG companies. Needless to mention that, providing FMCG companies with solutions empowered by machine learning techniques to predict customer churn and divide customers into segments to handle accordingly will be a competitive differentiator. As a result of our analysis; analyzing customer demographics, changes in expenditures of different channels, and change in purchasing of complementary or substitute products reveals valuable insights concerning future consumer behavior and uncovers the brands' strengths and weaknesses. To sum up, using machine learning applications would enable FMCG companies to generate the best of breed customer-centric strategies and targeted marketing efforts.

References

1. Lipowski, M. (2018). Customer Churn as a Purchasing Journey. 32nd IBIMA Conference: 15-16 November 2018, Seville, Spain. Maria Curie-Sklodowska University, Poland.

2. Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3), 2354–2364.
3. Abbasimehr, H., Setak, M. & Tarokh, M. J. (2011). A Neuro-Fuzzy Classifier for Customer Churn Prediction. *International Journal of Computer Applications (0975 – 8887)* Volume 19– No.8, April.
4. Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9.
5. Gallo A. (2014). The Value of Keeping the Right Customers. *Harvard Business Review*, October 29.
6. Figalist, I., Elsner, C., Bosch, C., 2, & Olsson, H. H. (2020). Customer Churn Prediction in B2B Contexts. 10th International Conference, ICSOB 2019, Jyväskylä, Finland, November 18–20, 2019, Proceedings. https://www.researchgate.net/publication/338950901_Customer_Churn_Prediction_in_B2B_Contexts.
7. Sulistiani, H. & Tjahyanto, A. (2017). Comparative Analysis of Feature Selection Method to Predict Customer Loyalty. *IPTEK, Journal of Engineering*, Volume:3, No:1.
8. Isson, J. P. (2018). *Unstructured Data Analytics: How to Improve Customer Acquisition, Customer Retention, and Fraud Detection and Prevention*. John Wiley & Sons, Inc., Hoboken, New Jersey, USA. ISBN: 9781119129752.
9. Shetty, P. P., Varsha, CM., V., Vadone, V. D., Sarode, S., & Pradeep, K. D. (2019). Customers Churn Prediction with Rfm Model and Building a Recommendation System using Semi-Supervised Learning in Retail Sector. *International Journal of Recent Technology and Engineering (IJRTE)*, Volume:8 Issue:1, May 2019, 3353–3358.
10. Buckinx, V., Baesens, B., Van den Peel, D., Van Kenhovel, P., & Vanthienen, J. (2002). Using machine learning techniques to predict defection of top clients. *Data Mining III*. ISBN 1-85312-925-9.
11. Ahmad, R., & Buttle, F. (2001). Customer retention: a potentially potent marketing management strategy. *Journal of Strategic Marketing*, 9(1), 29–45.
12. Coussement, K. (2014). Improving customer retention management through cost-sensitive learning. *European Journal of Marketing*, 48(3/4), 477–495.
13. Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy* 30, 552–568.
14. Miguéis, V. L., Van den Poel, D., Camanho, A. S., & Falcão e Cunha, J. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, 39(12), 11250–11256.
15. Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, Issue 6: Article 28.
16. Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. *Journal of Marketing Research*, 43(2), 204–211.
17. Griffin, J., & Lowenstein, M. W. (2001). *Customer WinBack: How to Recapture Lost Customers and Keep Them Loyal*. Jossey-Bass Inc., San Francisco, USA.
18. Court, D., Elzinga, D., Mulder, S., & Vetvik, O. J. (2009). *The Consumer Decision Journey*. McKinsey Quarterly, June.
19. Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1), 402–411.

20. Lariviere, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29(2), 472–484.
21. Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the Customer Base of Service Providers: An Examination of the Differences Between Switchers and Stayers. *Journal of Marketing*, 64(3), 65–87.
22. Buckinx, 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.
23. Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., & Vanthienen, J. (2004). Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *European Journal of Operational Research*, 156(2), 508–523.
24. Meister, S. (2012). *Brand Communities for Fast Moving Consumer Goods - An Empirical Study of Members' Behavior and the Economic Relevance for the Marketer*. Springer Gabler Verlag, Hamburg.
25. Murphy, J.A. (2001). *The Lifebelt: The Definitive Guide to Managing Customer Retention*. John Wiley & Sons Ltd, Chichester, West Sussex, UK.
26. Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157(1), 196–217.
27. Athanassopoulos, A. D. (2000). Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior. *Journal of Business Research*, 47(3), 191–207.
28. Colgate, M. R., & Danaher, P. J. (2000). Implementing a Customer Relationship Strategy: The Asymmetric Impact of Poor versus Excellent Execution. *Journal of the Academy of Marketing Science*, 28(3), 375–387.
29. Buckinx, 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.
30. Lessmann, S. & Voss, S. (2008). Supervised Classification for Decision Support in Customer Relationship Management. *Intelligent Decision Support: Current Challenges and Approaches* (pp.231-253), Gabler.
31. Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626–4636.
32. Miguéis, V. L., Camanho, A., & Falcão e Cunha, J. (2013). Customer attrition in retailing: An application of Multivariate Adaptive Regression Splines. *Expert Systems with Applications*, 40(16), 6225–6232.
33. Jahromi, A. T., Stakhovych, S., & Ewing, M. (2017). The impact of personalised incentives on the profitability of customer retention campaigns. *Journal of Marketing Management*, 1–21.
34. Calciu, M., Crie, D., & Micheaux, A. (2015). Recognising Dangerous Drop Out Incidents as opposed to accidents to improve the efficiency of triggers reducing customer churn. Application to RFM customer segments of a Fast Moving Consumer Goods retail chain. *Proceedings International Marketing Trends Conference 2015*.
35. Sulistiani, H. & Tjahyanto, A. (2017). Comparative Analysis of Feature Selection Method to Predict Customer Loyalty. *IPTEK, Journal of Engineering*, Vol. 3, No. 1.
36. Shoaib, T. (2018). Customers Churn Prediction in Retail Store. doi: 10.13140/RG.2.2.30545.38242.