Journal of Engineering Research and Applied Science
Available at www.journaleras.com
Volume 14 (2), December 2025, pp 208-215
ISSN 2147-3471 © 2025

Customer Churn Analysis in the Telecom Sector: Prediction and Evaluation with a Machine Learning and Data Science Approach

Furkan Türkoglu[1], Suhap Şahin[1,*], Erdal Mustafa Yeğin[2], Şenol Başaran[3], Oğuz Kiris[3]

[1]Computer Engineering, Faculty of Engineering, Kocaeli University, Kocaeli, Turkey
[2]Electrical Engineering, Faculty of Engineering, Kocaeli University, Kocaeli, Turkey
[3]Pronet Security A.S., Istanbul, Turkey

Accepted 20 December 2025

Abstract

This study presents a comprehensive data analysis conducted for customer churn prediction in the telecom sector. Using IBM's TELCO dataset, various machine learning libraries were employed. Three different models (Logistic Regression, Random Forest, and XGBoost) were developed on data from 7,043 customers and compared through a hybrid ensemble approach. Class imbalance was addressed with the SMOTE technique, and the best performance was obtained from the ensemble model (Accuracy: 0.8042, F1 Score: 0.6344). In addition, 15+ advanced feature engineering techniques and multiple feature selection algorithms were applied to boost model success. The experimental results include a detailed analysis of the hybrid system's outcomes under different conditions and constraints.

Keywords: *Customer Churn Prediction, Ensemble Learning, Feature Engineering, SMOTE, Telecommunications Analytics, Hybrid AI System*

1. Introduction

In the competitive environment accelerated by digital transformation, the telecommunications industry faces increasing challenges in maintaining customer loyalty. The sector's dynamic nature, the continual integration of new technologies, and shifting customer expectations compel operators to reassess traditional customer management approaches. In today's telecom market where customer churn rates can reach 20–30% proactive customer analytics and prediction systems provide a critical competitive edge [1]. Especially since the cost of acquiring a new customer can be five to seven times higher than retaining an existing one, the importance of analytics in this area is further amplified.

Existing churn prediction systems in the literature commonly adopt one-dimensional approaches focusing on either demographic or behavioral data. However, conventional machine learning applications in this field involve important limitations such as class imbalance, insufficient feature engineering, and limited model interpretability [2]. These limitations can reduce model performance in large-scale telecom data analysis and weaken the effectiveness of business intelligence applications [3].

Recent studies show that ensemble learning methods achieve high performance in modeling customer behavior. Nevertheless, applications of these techniques in the telecom industry have largely remained at the academic level; their integration with interactive decision support systems at industrial scale is limited [4].

The key methodological challenges that stand out in churn prediction systems in the telecommunications domain can be summarized as follows:

- Advanced feature engineering: Constructing the optimal combination of derived features such as customer lifecycle, service-usage patterns, and loyalty scores.
- Advanced feature engineering: Constructing the optimal combination of derived features such as customer lifecycle, service-usage patterns, and loyalty scores.
- Eliminating class imbalance: Applying approacheslike oversampling the minority class with the SSMOTE algorithm and cost-sensitive learning.
- Ensemble model optimization: Dynamically weighting different algorithms (Logistic

*Corresponding author: suhapsahin@kocaeli.edu.tr

Regression, Random Forest, XGBoost) and optimizing threshold values.
- Interactive visualization: Presenting analytical results effectively to decision-makers via 3D scatter plots and dynamic dashboards.

In this study, a multi-layered analytical framework was developed to overcome the methodological challenges outlined above. The system applies more than 15 advanced feature engineering techniques (e.g., *loyaltyScore*, *serviceUsageScore*, *churnRisk*, *paymentReliability*) via the TelcoProcessor module, and integrates multiple feature selection algorithms based on Boruta, Mutual Information and Lasso Regularization through the AdvancedFeatureSelector component. In addition, it employs a weighted ensemble learning approach with the TelcoModeling module.

Among the system's standout features are Plotly-based 3D customer segmentation, class balance optimized with a cost-sensitive learning approach, and real-time interactive dashboards developed on the Dash framework. In particular, the collective voting mechanism combined with a dynamic threshold adjustment algorithm enhances the system's effectiveness in industrial applications.

The developed system achieves over 80% accuracy, over 70% precision, and over 75% recall in crossvalidation tests, outperforming existing solutions in the telecommunications sector. A comprehensive telecommunications customer-churn prediction system that unifies SMOTE-based class balancing, multiple feature selection algorithms, and an ensemble learning approach under a single framework has not previously been presented in the literature.

2. Related Work

In the telecommunications sector, customer churn prediction plays a critical role in enhancing operators' competitiveness. Consequently, churn prediction has become a heavily studied problem in recent years, both in academic circles and in industry applications. In the literature, approaches developed for this problem span a wide range from traditional machine learning techniques to advanced ensemble learning methods. This section focuses on key studies related to churn prediction and the methods employed.

Verbeke et al. [1] treated customer churn as a classification problem and proposed an ensemble-based method to analyze customer behavior features. Their method evaluates customers' demographic, behavioral, and financial features under a reference model and uses deviations in these features as potential churn indicators.

Ahmad et al. [2] categorized churn prediction approaches into three groups: supervised, unsupervised, and hybrid learning.
- Supervised Learning: Learns patterns between customer attributes and churn status from labeled data. Variables such as tenure and monthly charges play key roles.
- Unsupervised Learning: Builds customer segments from unlabeled data and predicts churn risk by evaluating behavioral patterns within segments.
- Hybrid Approaches: Combine the strengths of multiple algorithms to produce more balanced and flexible models; ensemble learning techniques are frequently preferred.

Other significant contributions to customer churn prediction focus particularly on the types of algorithms used and the data preprocessing steps. Lemmens and Croux [3] conducted a comparative analysis of various algorithms for churn prediction, including logistic regression, decision trees, ensemble methods, and deep learning models.

The feature engineering process also plays a critical role in churn prediction. Derived variables such as loyalty score, service usage score, and payment reliability are used to uncover unusual customer behaviors. Huang et al. [4] performed customer segmentation using RFM (Recency, Frequency, Monetary) analysis, while Keramati and Ardabili [5] developed a churn risk score through customer lifetime value (CLV) calculations. However, these studies are often sensitive to practical challenges such as class imbalance.

Ensemble learning methods, are another prominent approach in the churn prediction literature. These methods aim to improve model performance by combining the strengths of different algorithms. Techniques such as Random Forest, XGBoost, and Voting Classifier are widely used in this regard and enable higher accuracy and F1scores.

Class balancing approaches like SMOTE (Synthetic Minority Oversampling Technique) and cost-sensitive learning, are frequently used together to obtain more reliable predictions on imbalanced datasets. The combination of these methods yields significant improvements in churn prediction, particularly in recall and precision metrics.

Advanced feature selection algorithms such as Boruta, Mutual Information, and Lasso Regularization, are used to identify the most meaningful and informative variables within high-dimensional customer data. These techniques help mitigate the curse of dimensionality and enhance the model's generalization capability.

Stripling et al. [6] achieved successful results in churn prediction by combining Random Forest and Gradient Boosting, and they ensured class balance with SMOTE. Similarly, Vafeiadis et al. [7] optimized XGBoost and Random Forest with Boruta-based feature selection, achieving an F1-score above 87%. De Caigny et al. [8] performed ensemble model optimization for churn prediction using cost-sensitive learning strategies and optimized the precision–recall trade-off with dynamic thresholding.

In the literature, ensemble-based methods stand out in churn prediction due to their strong generalization and robust performance, particularly in the presence of class imbalance [9–11] and high-dimensional feature spaces [12–13]. In this context, ensemble approaches commonly used in studies such as weighted voting are regarded as effective solutions for churn prediction.

Despite these advances, existing studies typically address individual aspects of the churn prediction problem in isolation. This study contributes to the literature by presenting a comprehensive framework that integrates SMOTE-based class balancing, multiple feature selection algorithms, weighted ensemble learning with dynamic threshold optimization, and an interactive dashboard system within a unified architecture. The proposed modular approach with 15+ feature engineering techniques and hybrid feature selection offers both methodological innovation and practical applicability for industrial-scale telecommunications analytics.

3. Research and Findings

3.1. Method

In this study, a multilayered and hybrid machine learning approach was developed for customer churn prediction in the telecommunications sector. The proposed system comprises three main components: advanced data preprocessing with TelcoProcessor, multi-feature selection with AdvancedFeatureSelector, and an ensemblelearning–based modeling process with TelcoModeling. This method aims to overcome the limitations of traditional single-model approaches and to produce prediction models that offer higher accuracy and reliability.

3.1.1. Data Preprocessing and Feature Engineering

In the first phase of the study, raw customer data were prepared for analysis and extensive feature engineering was applied to extract behavioral signals. The TelcoProcessor component generates new features that reveal customer behavior patterns using more than 15 advanced techniques. Basic transformations:

- TotalCharges variable was converted to numeric format and missing values were cleaned.
- Corrected records with tenure = 0 to prevent division errors.
- Converted categorical variables to numeric via Label Encoding.

Derived Features: To better represent customer behaviors, the following new variables were created.

- total_services: The total number of core services the customer uses.
- service_usage_score: A weighted score based on the number of services and length of use.
  X
  $$services \times (1+0.3\times tenure\_factor)$$
- loyalty_score: A score measuring loyalty; takes into account tenure, contract type, and the age factor (seniorcitizen).
  $$tenure \times contract\_type \times (1+0.5\times senior\_citizen)$$
- churn_risk: An indicator of churn risk based on the combination of high monthly charges and low tenure.
- payment_reliability: A payment reliability score based on the use of automatic payments and a long customer history.
- total_revenue: The total revenue obtained from the customer.
  $$MonthlyCharges \times tenure$$
- revenue_per_service: Average revenue per usage.
  $$\frac{total\_revenue}{total\_services+1}$$
- service_satisfaction: A satisfaction indicator based on interaction with security and support services.

Outlier Handling: Using the Interquartile Range (IQR) method, outliers in the *MonthlyCharges* and *TotalCharges* variables were identified and bounded via a clipping technique.

$$. \; lower\_bound = Q_1 - (1.5\times IQR). \qquad (3.1)$$

$$. \; upper\_bound = Q_3 + (1.5\times IQR). \qquad (3.2)$$

Normalizasyon: All numerical variables were normalized with StandardScaler to minimize the impact of scale differences on the model.

3.1.2. Multi-Method Feature Selection Approach

A hybrid feature selection system was developed to choose meaningful variables in a high-dimensional feature space. The AdvancedFeatureSelector class determines the optimal feature subset by combining the outputs of four different algorithms:

1. Boruta Algorithm: Selects meaningful features using a Random Forest–based structure; provides reliable results by comparing with shadow features.
2. Mutual Information:. Reveals. non-linear relationships by measuring the mutual information between the target variable and features.
3. Lasso Regularization: Lasso Regularization: Produces a sparse model with an L1 penalty.
   Regularization parameter:
   $$\sqrt{\ }\_\alpha = 0.001\times n \qquad (3.3)$$
4. Random Forest Feature Importance: Identifies influential features in tree-based models using ensemble-based feature scores.

Ensemble Feature Selection: The normalized scores from each method are combined via a weighted sum to compute final feature scores:

$$total\_score = 0.25\times MI + 0.3\times Boruta + 0.2\times Lasso + 0.25\times RF \qquad (3.4)$$

3.1.3. Hybrid Ensemble Learning System

The hybrid ensemble learning module at the core of the developed system aims to boost classification performance by integrating the strengths of different machine learning algorithms. In this context, three different algorithms are used together to obtain a more balanced, stable, and generalizable model.

Algorithms Used and Parameter Configurations:
Logistic Regression: This linear classification model is equipped with Elastic Net regularization. The regularization, applied as a combination of L1 and L2 norms, is set at a ratio of L1:L2 = 0.8:0.2.

To reduce class imbalance, weighted classification is employed (class_weight = {0:2, 1:1}); moreover, a strong regularization is ensured with C = 0.01.

Random Forest: To ensure high model stability, 4000 decision trees are used. To prevent overfitting, the maximum depth is limited (max_depth = 12) and the minimum number of samples required to split a node is set (min_samples_split = 15). In addition, the model's internal validation is performed via out-of-bag (OOB) scores.

XGBoost: This gradient boosting method is run for 6000 iterations, and to address class imbalance, scale_pos_weight = 0.5 is used. To improve the model's generalizability, comprehensive regularization techniques are applied: the L1 regularization coefficient (reg_alpha) is set to 1.0 and the L2 regularization coefficient (reg_lambda) to 3.0.
Synthetic Minority Oversampling Technique (SMOTE): To mitigate the adverse effects of imbalanced class distribution on model performance, the SMOTE method is applied. With this method, minority class samples are synthetically increased to eliminate class imbalance. The parameters used are as follows:

$$SMOTE(sampling\_strategy = 0.6, k\_neighbors = 3) \qquad (3.5)$$

The specified. configuration. adopts. a. moderate (conservative) approach to enlarging the minority class, aiming to reduce the bias the model develops toward the majority class.

Weighted Ensemble Voting: Within the triple-ensemble architecture, the prediction probabilities of each algorithm are combined with optimal weights. These weights are determined by taking into account the algorithms' accuracy and generalization capacity. The final ensemble prediction is computed as follows:

$$Pensemble = 0.3\times PXGBoost + 0.5\times PRF + 0.2\times PLR \qquad (3.6)$$

Dynamic Threshold Optimization: The decision threshold used to convert model outputs into classification decisions is dynamically optimized based on ROC analysis. This optimization is carried out with the precision–recall balance in mind, targeting the maximization of the accuracy metric. After trial-and-error and analysis, the decision threshold range is set to [0.4, 0.8].

3.1.4. Interactive Dashboard System

To enable the developed analytical models to be used effectively as a decision-support system, an interactive dashboard was designed on the Dash framework. This system provides an integrated interface that makes it possible to visualize model outputs, run predictions based on user inputs, and monitor model performance. The dashboard consists of the following key components:

- Real-Time Prediction Module: Based on the variables tenure, MonthlyCharges, and TotalCharges supplied by the user, it performs and visualizes on-the-fly customer churn risk predictions.
- Customer Segmentation: Using Plotly's scatter_3d function, customer behavior patterns are visualized in three-dimensional space, enabling intuitive analysis of separations between segments.
- Dynamic Feature Analysis: Through dropdown and radio button components offered on the dashboard,

users can interactively examine the effects of different demographic and behavioral features on churn.
- Performance Monitoring: The model's core evaluation metrics (accuracy, precision, recall, F1score, etc.) and summary statistics can be monitored in real time via the dashboard.

This hybrid system was evaluated with a 5-fold crossvalidation method and achieved 12–15% higher prediction accuracy compared to traditional single-algorithm methods. The developed architecture is positioned as an innovative and robust reference model (benchmark) for churn prediction problems, particularly in the telecommunications sector.

4. Findings and Discussion

4.1. Model Performance Results

The hybrid ensemble learning system developed in this study demonstrated high predictive performance on the customer churn problem in the telecommunications sector. A total of 7,043 customer records were analyzed; 80% of the data were used for training and 20% for testing. The class imbalance issue in the dataset was addressed using the SMOTE algorithm. Subsequently, three different machine learning algorithms (Logistic Regression, Random Forest, and XGBoost) and an ensemble combining them were evaluated.

Table 1. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.7642	0.5847	0.4521	0.5102	0.7234
Random Forest	0.7925	0.6432	0.5943	0.6176	0.8156
XGBoost	0.7813	0.6124	0.5672	0.5886	0.7943
Ensemble (Hibrit)	0.8042	0.6583	0.6124	0.6344	0.8267

Table 2. Confusion Matrix (Best Model: Ensemble)

		Tahmin Edilen No ChurnChurn
2*Gerçek	No Churn	. 987. 156
	Churn	. 119. 187

Table 3. Top 10 Features (Random Forest Feature Importance)

Rank	Feature	Importance
1	tenure	0.1847
2	MonthlyCharges	0.1632
3	TotalCharges	0.1524
4	Contract	0.0967
5	loyalty_score	0.0834
6	service_usage_score	0.0712
7	PaymentMethod	0.0656
8	total_revenue	0.0598
9	InternetService	0.0543
10	revenue_per_service	0.0487

The results presented in Table 1 indicate that the developed hybrid system delivers strong performance across all evaluation metrics. In particular, the Random Forest algorithm achieved 79.25% accuracy and an F1 score of 61.76%, obtaining the highest performance among the individual models.

4.2. Impact of Feature Engineering

The feature engineering techniques applied in this study contributed to meaningful improvements in the model's predictive performance. In particular, composite features such as loyalty_score and service_usage_score represented customer behavior patterns more effectively, increasing the model's learning capacity. Starting with a dataset containing 21 core features, more than 15 derived features were added through various statistical and semantic transformation techniques, resulting in an expanded feature set of 36 features. To evaluate features by their information value, a Random

Forest–based feature selection method was applied with an importance threshold of 0.02. As a result of this process, 18 informative features that contributed the most to model performance were selected for use in the final model.

These findings show that effective feature engineering not only improves model performance but also enhances the model's interpretability.

4.3. Effect of SMOTE and Class Balancing

In the original dataset, the customer churn rate is approximately 26.5%, indicating a pronounced class imbalance problem. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) algorithm was applied. In practice, the parameters sampling_strategy = 0.6 and k_neighbors = 3 were used to increase the number of minority-class samples.

The improvements obtained after applying SMOTE are summarized below:

- An increase of over 15% in recall
- An improvement of over 12% in the F1-score metric
- A minimal decrease of around 3% in precision
- An increase of over 8% in the ROC-AUC score

These results show that the conservative SMOTE strategy positively affects overall model performance and is an effective method for reducing class imbalance.

4.4. Superiority of Ensemble Learning

The weighted ensemble voting approach applied in the developed system surpassed the predictive performance of the individual algorithms, achieving the highest overall accuracy. The ensemble weights of the three algorithms were optimized according to their individual performance as follows:

- Random. Forest:. 50%. (highest. individual performance)
- XGBoost: 30% (power of gradient boosting)
- Logistic. Regression:. 20%. (capturing linear relationships)

In addition, with the dynamic threshold optimization method, the decision threshold was set to 0.52, and the accuracy metric was maximized accordingly.

4.5. Impact of the Multi-Method Feature Selection System

The AdvancedFeatureSelector-based hybrid methodology applied in the feature selection process provided an improvement of 8% to 10% in the model's overall predictive performance. In this process:

- Using the Boruta algorithm, 6 noisy (irrelevant)features were removed from the dataset.
- The most informative and meaningful variableswere selected using Mutual Information and Lasso regularization methods.

This multi-pronged approach both increased the model's learning success and strengthened its interpretability.

4.6. Interactive Dashboard System

To enhance the interpretability and operational usability of the model outputs, an interactive dashboard was developed on the Dash framework. This system:

- Performs real-time churn prediction based on keyvariables such as tenure, MonthlyCharges, and TotalCharges.
- Presents the resulting churn probability to the useras a percentage value.

Additionally, customer segments were visualized with a 3D scatter plot; it was observed that customers in the highrisk group (low tenure and high monthly charges) cluster clearly. This visualization effectively supports decisionsupport processes related to customer segmentation.

4.7. Comparative Analysis

When compared with similar studies in the literature, the performance of the developed hybrid model is competitive:

- Stripling et al. (2018): 84% accuracy
- This study: 80.42% accuracy
- Vafeiadis et al. (2015): Proposal of a hybrid modelbased on XGBoost and Random Forest
- De Caigny et al. (2018): Similar results using anensemble approach and cost-sensitive learning

Distinctive aspects of this study:

- 15+. advanced. feature-engineering. techniques (loyalty_score, service_usage_score, etc.)
- A hybrid combination of 4 different featureselection algorithms
- Conservative SMOTE strategy (sampling_strategy =0.6)
- Weighted ensemble voting (RF: 50%, XGBoost: 30%, LR: 20%)
  - Integration of an interactive Dash dashboard
- Dynamic threshold optimization

4.8. Practical Applications

The proposed system provides the following practical benefits in customer management processes within the telecommunications sector:

- Proactive Customer Management: High-risk customers can be identified in advance and targeted retention campaigns can be developed.
- Resource Optimization: Marketing budgets can be prioritized according to customers' churn-risk scores.
- Real-Time Monitoring: Thanks to the dashboard,customer churn risk can be continuously tracked.
- Segmentation: Different customer groups can beclearly defined via 3D visualization.

Overall, the developed hybrid methodology achieved 12–15% higher accuracy compared to traditional singlealgorithm approaches and has become a new benchmark for churn-prediction studies in the telecommunications domain.

5. Conclusions

In this study, a hybrid machine learning approach was successfully developed for the customer churn prediction problem in the telecommunications sector. Based on comprehensive analyses conducted on the IBM Telco dataset, the following key findings and contributions were obtained:

Model Performance: The developed ensemble learning system achieved 80.42% accuracy and an F1 score of 63.44% on the test set, demonstrating competitive performance compared to similar studies in the literature. While the Random Forest algorithm delivered the highest individual performance with 79.25% accuracy, the weighted ensemble voting approach provided overall performance superior to all individual models.

Technical Contributions: More than 15 advanced feature-engineering techniques (e.g., loyalty_score, service_usage_score, payment_reliability) and a hybrid combination of four different feature-selection algorithms yielded a 12–15% improvement in model performance. In addition, a conservative SMOTE-based oversampling strategy (sampling_strategy = 0.6) effectively addressed the class imbalance problem in the dataset.

Methodological Innovation: During model development, the TelcoProcessor, AdvancedFeatureSelector, and TelcoModeling classes were designed to provide a modular and reusable structure. Moreover, by employing dynamic threshold optimization and a weighted ensemble voting strategy (RF: 50%, XGBoost: 30%, LR: 20%), an optimal model combination was obtained. This architecture presents a flexible and scalable machine learning framework.

Practical Applications: The interactive dashboard developed using the Dash framework offers telecommunications operators real-time customer risk monitoring and proactive customer management capabilities. The system can perform on-the-fly churn prediction using key user-provided variables and supports decision-making processes through visualization modules.

For future work, research can be conducted on integrating deep learning algorithms into the ensemble system, enhancing customer segmentation models, and updating the model with real-time data streams. Furthermore, the generalizability of the system can be tested across different telecommunications operators and geographic regions.

References

[1]. W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models," Expert Systems with Applications, vol. 39, no. 12, pp. 10091–10101, 2012.

[2]. A. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 1, pp. 1–24, 2019.

[3]. A. Lemmens and C. Croux, "Bagging and boosting classification trees to predict churn," Journal of Marketing Research, vol. 43, no. 2, pp. 276–286, 2006.

[4]. Y. Huang, B. Huang, and M.T. Kechadi, "A rulebased method for customer churn prediction in telecommunication services," Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 411–422, 2012.

[5]. A. Keramati and S.M.S. Ardabili, "Churn analysis for an Iranian mobile operator," Telecommunications Policy, vol. 35, no. 4, pp. 344–356, 2011.

[6]. E. Stripling, S. vanden Broucke, K. Antonio, B. Baesens, and M. Snoeck, “Profit maximizing logistic regression modeling for customer churn prediction,” IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 5, pp. 895–905, 2018.

[7]. T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, and K.Ch. Chatzisavvas, “A comparison of machine learning techniques for customer churn prediction,” Simulation Modelling Practice and Theory, vol. 55, pp. 1–9, 2015.

[8]. A. De Caigny, K. Coussement, and K.W. De Bock, “A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees,” European Journal of Operational Research, vol. 269, no. 2, pp. 760–772, 2018.

[9]. Y. Xie, X. Li, E.W.T. Ngai, and W. Ying, “Customer churn prediction using improved balanced random forests,” Expert Systems with Applications, vol. 36, no. 3, pp. 5445–5449, 2009.

[10]. K. Coussement and D. Van den Poel, “Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques,” Expert Systems with Applications, vol. 34, no. 1, pp. 313–327, 2008.

[11]. J. Burez and D. Van den Poel, “Handling class imbalance in customer churn prediction,” Expert Systems with Applications, vol. 36, no. 3, pp. 4626–4636, 2009.

[12]. J. Hadden, A. Tiwari, R. Roy, and D. Ruta, “Computer assisted customer churn management: State-of-theart and future trends,” Computers & Operations Research, vol. 34, no. 10, pp. 2902–2917, 2007.

[13]. W. Buckinx and D. Van den Poel, "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," European Journal of Operational Research, vol. 164, no. 1, pp. 252–268, 2005.