Expert Systems with Applications 38 (2011) 12293–12301

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction

Koen W. De Bock [a,b], Dirk Van den Poel [a,*]

[a] Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, B-9000 Ghent, Belgium
[b] IESEG School of Management – Université Catholique de Lille (LEM, UMR CNRS 8179), Department of Marketing, 3 Rue de la Digue, F-59000, Lille, France

ARTICLE INFO

Keywords:
CRM
Database marketing
Customer churn prediction
Ensemble classification
Rotation-based ensemble classifiers
RotBoost
Rotation Forest
ICA
AUC
Lift

ABSTRACT

Several studies have demonstrated the superior performance of ensemble classification algorithms, whereby multiple member classifiers are combined into one aggregated and powerful classification model, over single models. In this paper, two rotation-based ensemble classifiers are proposed as modeling techniques for customer churn prediction. In Rotation Forests, feature extraction is applied to feature subsets in order to rotate the input data for training base classifiers, while RotBoost combines Rotation Forest with AdaBoost. In an experimental validation based on data sets from four real-life customer churn prediction projects, Rotation Forest and RotBoost are compared to a set of well-known benchmark classifiers. Moreover, variations of Rotation Forest and RotBoost are compared, implementing three alternative feature extraction algorithms: principal component analysis (PCA), independent component analysis (ICA) and sparse random projections (SRP). The performance of rotation-based ensemble classifier is found to depend upon: (i) the performance criterion used to measure classification performance, and (ii) the implemented feature extraction algorithm. In terms of accuracy, RotBoost outperforms Rotation Forest, but none of the considered variations offers a clear advantage over the benchmark algorithms. However, in terms of AUC and top-decile lift, results clearly demonstrate the competitive performance of Rotation Forests compared to the benchmark algorithms. Moreover, ICA-based Rotation Forests outperform all other considered classifiers and are therefore recommended as a well-suited alternative classification technique for the prediction of customer churn that allows for improved marketing decision making.

1. Introduction

Customer retention refers to the degree to which a company is able to satisfy and retain its current customers, and is generally perceived as a cornerstone of successful customer relationship management (CRM) (Payne & Frow, 2005; Reinartz, Krafft, & Hoyer, 2004; Winer, 2001). An important instrument in customer retention is customer churn prediction, aimed at the identification of customers with a high probability to attrite (Neslin, Gupta, Kamakura, Lu, & Mason, 2006). A typical churn prediction model generalizes the relationship between churn behavior on the one hand, and customer characteristics and behavior based on historical data on the other hand in such a way that a company is able to use it to produce fair predictions about future behavior of its customers. Effective churn prediction has a beneficial impact upon firm profitability in several ways. First, identification of potential churning customers allows marketing decision makers to target marketing actions in a cost-effective manner. Retention campaigns can be limited to a selection of customers but cover a large proportion of all customers with an actual intention to attrite. Second, high customer retention eases the pressure to attract a substantial number new customers every period. It has been shown that the acquisition of new customers generally comes at higher costs than keeping the existing customer base satisfied (Reinartz & Kumar, 2003).

The prediction of customer churn is generally approached as a problem of binary classification. In this context, companies typically apply data mining techniques to conduct customer churn analysis (Xie, Li, Ngai, & Ying, 2009). Algorithms for binary classification are suited to generalize the relationship between the outcome, i.e., the question whether a person is a churner or not, and a range of predictor variables that describe the characteristics and the behavior of the customer. The quality of a customer churn prediction model is directly influenced by two important factors: the available input data and the data mining algorithm used to model customer churn. The first factor involves the data that is available to describe customers and their relationship with the company. Relevant predictive information includes customer

* Corresponding author. Tel.: +32 9 264 89 80, fax: +32 9 264 42 79.
E-mail addresses: k.debock@ieseg.fr (K.W. De Bock), Dirk.VandenPoel@UGent.be (Dirk Van den Poel).
URL: http://www.crm.UGent.be (Dirk Van den Poel).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2011.04.007

demographics (e.g., Burez & Van den Poel, 2007; Lemmens & Croux, 2006; Xie et al., 2009), historical transactional data (e.g., Glady, Baesens, & Croux, 2009), financial information (Larivière & Van den Poel, 2005), textual information from customer e-mails (Coussement & Van den Poel, 2008b), longitudinal data (Van den Poel & Larivière, 2004) and so on. A second factor involves the classification technique that is used to model churn. Neslin et al. (2006) point out that the choice of the modeling technique has a significant impact upon the return on investment of customer churn prediction efforts and they emphasize the importance of comparing alternative algorithms in search for optimal model performance. This suggestion is reflected in the variety of algorithms that have been suggested in customer churn literature, including logistic regression (Smith, Willis, & Brooks, 2000), artificial neural networks (Pendharkar, 2009; Tsai & Lu, 2009), survival analysis (Van den Poel & Larivière, 2004), Markov chains (Burez & Van den Poel, 2007), support vector machines (Coussement & Van den Poel, 2008a), generalized additive models (Coussement, Benoit, & Van den Poel, 2010), decision trees (Lemmens & Croux, 2006; Smith et al., 2000), naive Bayes classifiers (Buckinx, Baesens, Van den Poel, Van Kenhove, & Vanthienen, 2002), K-nearest neighbor classifiers (Ruta, Nauck, & Azvine, 2006), Random Forests (e.g., Larivière & Van den Poel, 2005), cost-sensitive classifiers (Glady et al., 2009) and evolutionary algorithms (Au, Chan, & Yao, 2003).

An additional category of classifier algorithms that have been applied successfully to customer churn prediction are ensemble classifiers, or multiple classifier systems (MCS). In an ensemble classifier, several classifier models are combined into one aggregated classifier, and their predictions are combined into one aggregated outcome using a fusion rule (Kuncheva, 2004). Several studies have demonstrated that ensembles of classifiers often demonstrate superior performance over single classification models (e.g., Bauer & Kohavi, 1999; Breiman, 1996; Dietterich, 2000). Similar findings are found in studies on customer churn prediction (e.g., Larivière & Van den Poel, 2005; Lemmens & Croux, 2006). In this study, rotation-based ensembles are proposed as classification algorithms for churn prediction. Rotation-based ensembles are ensembles that apply rotations on the input data through linear feature extraction algorithms and have been found to demonstrate superior performance over both single and ensemble benchmark classifier algorithms in terms of accuracy (Rodríguez, Kuncheva, & Alonso, 2006; Zhang & Zhang, 2008). Two algorithms are considered: Rotation Forest (Rodríguez et al., 2006) and RotBoost, which is a combination of Rotation Forest and AdaBoost (Zhang & Zhang, 2008). In an experimental validation, Rotation Forest and RotBoost are applied to four real-life customer churn applications from various industries, and their performance is compared to three well-known ensemble classifiers (Bagging (Breiman, 1996), the random subspace method (RSM) (Bryll, Gutierrez-Osuna, & Quek, 2003; Ho, 1998), and Random Forests (Breiman, 2001)), and two decision tree algorithms (C4.5 (Quinlan, 1993) and CART (Breiman, 1984)). Moreover, results from rotation-based ensembles based on three alternative feature extraction algorithms (principal component analysis (PCA) (Jolliffe, 2002), independent component analysis (ICA) (Comon, 1994) and sparse random projections (SRP) (Kuncheva & Rodríguez, 2007)) are compared. Rotation Forest and RotBoost have, to the best of our knowledge, never been applied to customer churn prediction. Moreover, their performance has never been evaluated in terms of AUC or top-decile lift. Finally, the influence of using alternative feature extraction algorithms for RotBoost has not been investigated yet.

The rest of the paper is organized in the following manner. In Section 1, related literature is discussed. This includes an introduction to ensemble classification and an overview of applications to customer churn prediction. Section 2 focuses on methodology with a presentation of the Rotation Forest and RotBoost algorithms, and the performance criteria that are used in the experimental validation. Section 3 presents the data, experimental conditions and results of an experimental comparison of classifier performance. Finally, a conclusion is formulated and limitations to the study and directions for future research are provided.

2. Related literature

2.1. Ensemble classification

During the last decade, ensemble classification has become a popular field of research, both in methodological (e.g., Bauer & Kohavi, 1999; De Bock, Coussement, & Van den Poel, 2010; Dietterich, 2000; Kuncheva, 2004) and applied literature (e.g., Kim, 2006; van Wezel & Potharst, 2007). Numerous theoretical and empirical studies demonstrate how the practice of combining classification models into one aggregated model can significantly improve classification accuracy. An ensemble classification model typically consists of the following elements: a base or member classification algorithm, a fusion rule to combine the outputs of the constituent ensemble members and a heuristic for injecting diversity into the ensemble (Kuncheva, 2004). Diversity is an important concept within ensemble classification theory. It is generally perceived that the best performing ensemble classifiers combine high accuracy of member classifiers with a maximum disagreement (hence, diversity) among the ensemble members.

The most well-known ensemble classifiers are Bagging and Boosting. In Boosting, ensembles are built in an iterative manner. A prominent algorithm in this category is AdaBoost (Freund & Schapire, 1996). Using weight manipulation or resampling, misclassified instances are attributed higher importance in the training data over consecutive iterations to force classifiers to concentrate on instances that are hard to classify correctly. In Bagging (Breiman, 1996), each member classifier in the ensemble is built upon a bootstrap sample, i.e., a sample taken with replacement and usually of the same size as the training data set. Base classifier outputs, in general from decision trees, are combined using majority voting. Successful variations upon Bagging are the random subspace method (RSM) (Bryll et al., 2003; Ho, 1998) and Random Forests (Breiman, 2001). In RSM, each member in an ensemble of decision trees is trained on a random selection of features of a specified size, and member classifiers' outputs are averaged. In Random Forests, Bagging is applied to randomized trees, i.e., CART decision trees whereby random feature selection is performed at each tree node split.

2.2. Applications to customer churn prediction

Several successful applications of ensemble classification in customer relationship management can be found in literature. An overview is included in Table 1.

Most applications involve customer churn prediction. The first application of an ensemble method to customer churn prediction, to the best of our knowledge, is found in Mozer, Wolniewicz, Grimes, Johnson, and Kaushansky (2000). Using data from a major US wireless carrier, customer churn is predicted using logistic regression, C5.0 trees, neural networks, AdaBoost and Boosting with neural networks. Lift curves reveal favorable performance for boosted neural networks and AdaBoost. In Hu (2005), customer churn at a retail bank is analyzed. The study compares performance of a decision tree, a boosted naive Bayesian network, a selective Bayesian network, a neural network and a hybrid ensemble of all these classifiers. The authors conclude that the ensemble classifier outperforms all individual classifiers in terms of lift. Kim (2006) constructs ensembles of neural networks and logistic

Table 1
Journal articles on customer churn prediction using ensemble classifiers.

Study	Application	Ensemble classifiers used	Number of datasets	Industry
Mozer et al., 2000	Churn	AdaBoost, ANN Boosting	1	Telecom
Hu, 2005	Churn	Boosted Naive Bayesian Networks, hybrid ensemble	1	Bank
Larivière and Van den Poel, 2005	Churn & customer profitability	Random Forests, Regression Forests	1	Bank
Kim, 2006	Churn	Logit and ANN ensembles	1	Telecom
Lemmens and Croux, 2006	Churn	Bagging, Stochastic Gradient Boosting	1	Telecom
Burez and Van den Poel, 2007	Churn	Random Forests	1	Pay TV
van Wezel and Potharst, 2007	Customer choice modeling	Bagging, LogitBoost, MultiBoost	2	US household data
Burez and Van den Poel, 2008	Churn	Random Forests	1	Pay TV
Prinzie and Van den Poel, 2008	Cross-sell	Random Forest and Random Multinomial Logit (RMNL)	1	Home appliances retailer
Bose and Chen, 2009	Churn	C5.0 Boosting	1	Telecom
Burez and Van den Poel, 2009	Churn	Random Forests, weighted Random Forests, Gradient Boosting Machine	6	Bank, Telecom, Pay TV, Supermarket, Newspaper
Coussement and Van den Poel, 2009	Churn	Random Forests	1	Newspaper subscription
Glady et al., 2009	Churn	AdaCost	1	Bank
Xie, et al., 2009	Churn	Improved Balanced Random Forests (IBRF)	1	Bank

regression models. An adapted version of Bagging is used: member classifiers are trained on random samples taken without replacement and with a size equal to half the number of instances in the training data set. Predicted posterior class membership probabilities are averaged over the ensemble members. The results indicate that the ensemble of neural networks demonstrates great improvement over a single neural network and that improvements are more modest for logit ensembles. In Lemmens and Croux (2006), Bagging and Stochastic Gradient Boosting are applied to customer churn prediction in a US wireless telecommunications company. They find that both algorithms perform comparably when evaluated in terms of top-decile lift and Gini coefficient, and that both improve performance substantially over a logistic regression. Several studies have suggested the use of Random Forests for customer churn prediction in financial services (Larivière & Van den Poel, 2005), pay TV (Burez & Van den Poel, 2007; Burez & Van den Poel, 2008) and newspaper subscription (Coussement & Van den Poel, 2009). In each of these studies, Random Forests demonstrated superior classification performance over benchmark algorithms. Bose and Chen (2009) apply hybrid models, consisting of clustering and boosted C5.0 decision trees to churn prediction for a mobile telecommunications operator. Other studies suggest the use of classification techniques that deal with the problem of class imbalance in customer churn prediction. Burez and Van den Poel (2009) compare several strategies to deal with class imbalance and advise the use of weighted Random Forests. Xie et al. (2009) propose Improved Balanced Random Forests (IBRF) as a variation of Random Forests that demonstrates competitive performance on data from a Chinese bank. Finally, Glady et al. (2009) apply AdaCost, a cost-sensitive version of AdaBoost, to customer churn prediction for a European bank.

Finally, two related applications are found in customer choice modeling. Van Wezel and Potharst (2007) compare several classification algorithms for the construction of next-product-to-buy (NPTB) models. While no dominance of any method is observed, ensemble methods consistently outperform individual classifiers. Prinzie and Van den Poel (2008) propose a new ensemble classifier for multi-class classification, the Random Multinomial Logit (RMNL) and apply it to NPTB modeling.

3. Methodology

3.1. Rotation-based ensemble classifiers

In this study, rotation-based ensemble classifiers are evaluated for customer churn prediction. Two algorithms are considered: Rotation Forest (Rodríguez et al., 2006) and RotBoost (Zhang & Zhang, 2008). Consider the following notations. Let T be a training data set with $T = \{(x_i, y_i)\}_{i=1}^{n}$ consisting of n observations. An instance (x_i, y_i) consists of a vector of input feature values x_i and a response y_i. Note that for customer churn prediction, only binary classification is considered, so $y_i \in \{0, 1\}$ where class 1 represents the churn event. Further, T can be decomposed into X and Y, where X is the input vector, an $n \times p$ matrix containing feature values for all n instances, and Y with dimensionality $n \times 1$ contains class labels. F is the set of predictive features; $F = \{X_1, \ldots, X_p\}$. In the Rotation Forest algorithm, an ensemble classifier C of m decision trees is constructed, $C = \{C_1, C_2, C_3, \ldots, C_m\}$, whereby the training data for each base classifier is rotated using a (linear) feature extraction algorithm $\mathscr{E}$. More specifically, for each base classifier C_j a rotation matrix R_j^a is constructed by randomly taking s subsets from F (or dividing F into feature subsets of size r), for each subset performing feature extraction algorithm $\mathscr{E}$ on a bootstrap sample of X, with a size of 75% of X, and rearranging the coefficients. The training data for C_j is then obtained by rotating the input vector X using R_j^a and combining the result with Y. To combine the member classifiers' outputs, predictions are averaged. In RotBoost, the base classifier C_j in Rotation Forests is replaced by an AdaBoost classifier (as described by Freund and Schapire (1997)). The detailed pseudocodes of Rotation Forest and RotBoost can be found in Rodríguez et al. (2006) and Zhang and Zhang (2008), respectively.

Rotation Forests demonstrated superior accuracy over Bagging, AdaBoost and Random Forests on a broad range of data sets in Rodríguez et al. (2006). The strong performance is attributed to a simultaneous improvement of (i) diversity within the ensemble, obtained by the use of feature extraction on training data and the use of decision trees, known to be sensitive to variations in the training data, as base classifiers, and (ii) accuracy of the base classifiers, by keeping all extracted features in the training data. Zhang and Zhang (2008) introduced RotBoost as a combination of Rotation Forest and AdaBoost and their experiments indicated superior accuracy performance of RotBoost over Bagging, AdaBoost and MultiBoost and a slight improvement over Rotation Forests. Applications of Rotation Forest and RotBoost are rather scarce, and both methods have, to the best of our knowledge, never been implemented in a context of customer churn prediction.

Feature extraction plays an important role in rotation-based ensemble classifiers. In the original Rotation Forest algorithm, feature extraction through principal component analysis is applied (Rodríguez et al., 2006). In Kuncheva and Rodríguez (2007), experiments are conducted with alternative feature extraction algorithms, i.e., non-parametric discriminant analysis (NDA), random projections (RP) and sparse random projections (SRP). In the latter, a simulated rotation matrix is constructed by sampling all non-zero elements from a standard normal distribution. Experimental results demonstrate good performance of Rotation Forests based on NDA and SRP, but both are outperformed by PCA. In an application of Rotation Forests to cancer classification, Liu and Huang (2008) also investigate the value of alternative feature extraction methods. They find that independent component analysis (ICA) (Comon, 1994) is a valuable alternative to PCA that can further improve the accuracy of Rotation Forests over PCA and random projections. The influence of alternative feature extraction techniques in RotBoost has not been studied so far. Given the importance of the choice regarding the algorithm for feature extraction in Rotation Forests, several alternatives are considered in this study. Three alternative feature extraction algorithms, i.e., principal component analysis, independent component analysis and sparse random projections are compared for both Rotation Forest and RotBoost in the experimental evaluation.

3.2. Performance criteria

This study uses three performance criteria for the evaluation of classifier performance: accuracy, AUC and top-decile lift. Accuracy (also referred to as percentage correctly classified or PCC) is the dominant performance criterion in machine learning and ensemble classification literature (e.g., Rodríguez et al., 2006; Zhang, Zhang, & Wang, 2008), while AUC and top-decile lift are well established performance measures in churn literature (e.g., Bose & Chen, 2009; Burez & Van den Poel, 2009; Lemmens & Croux, 2006). While accuracy assumes the transformation of posterior class membership probabilities, produced by a classification model, to class predictions based on the fixation of a threshold value, AUC (AUROC), or the area under the receiver operating characteristics curve is not influenced by this threshold value and thus is a more objective performance criterion (Provost, Fawcett, & Kohavi, 2000). AUC summarizes the performance of a classifier represented by a ROC curve, which plots, for every possible threshold value, the true positive ratio (or sensitivity) versus the false positive ratio (equivalent to one minus the specificity). It takes a value between 0.5 and 1, where larger values represent stronger performance. AUC is universally recognized as an objective performance criterion, well-suited for the comparison of classification models (Langley, 2000; Provost et al., 2000).

Several studies underline the importance of top-decile lift for the evaluation of customer churn prediction models (e.g., Lemmens & Croux, 2006; Pendharkar, 2009). Top-decile lift refers to the ratio of the percentage of actual churners in the top ten percent of the highest predicted churn probabilities, and the percentage of actual churners in the total data set. It concentrates on the segment of riskiest customers and focuses on the essence of customer churn prediction models, i.e., their ability to identify the group of customers most likely to churn so that retention campaigns can be targeted at a fraction instead of all customers and still reach a majority of all potential churners.

4. Experimental evaluation

4.1. Data

To evaluate the performance of Rotation Forest and RotBoost, experiments are conducted on data sets from four real-life customer churn prediction projects in large European companies. For reasons of confidentiality, company names are not disclosed. The characteristics of these data sets are summarized in Table 2.

These data sets have a number of common features. First, they all (with the exception of the first data set) exhibit large dimensionalities, both in terms of number of instances and the number of descriptive features. Second, they are characterized by considerable class imbalance, most notable the data sets originating from a bank, a European telecommunications operator and a do-it-yourself (DIY) hardware store chain. Predictive features among these data sets capture information on customer demographics, historical transactional data and financial information.

To deal with class imbalance, which is known to distort classifier performance for classification algorithms that are not particularly designed to deal with this problem, undersampling is applied, as suggested by Weiss (2004) and applied to customer churn prediction by Burez and Van den Poel (2009). Undersampling involves randomly removing instances from the majority class from the training data until both classes are balanced.

4.2. Experimental conditions

Based on four data sets from real-life customer churn prediction projects, classification performance of Rotation Forest and RotBoost is compared to five benchmark algorithms: Bagging, Random Forests, the random subspace method (RSM), CART and C4.5. Moreover, for both Rotation Forest and RotBoost, three alternative versions are included, based on feature extraction through PCA, ICA and SRP. As outlined earlier, classification performance is evaluated in terms of three performance metrics: accuracy, AUC and top-decile lift.

All variations of Rotation Forest and RotBoost are programmed in Matlab and implement PCA and ICA using the Matlab Toolbox for Dimensionality Reduction (van der Maaten, 2007). Bagging and Random Forest results are obtained using the adabag (Alfaro, Gámez, & García, 2006) and randomForest (Liaw & Wiener, 2002) packages in R (R Development Core Team, 2009). C4.5 results are based upon the J4.8 classifier in WEKA (Frank, Holmes, Pfahringer, Reutemann, & Witten, 2009). Parameter settings for the algorithms are based on default or recommended values. Random feature subsets in Random Forests are equal to the square root of the number of features in the data set, as suggested by Breiman (2001). This setting is also used for RSM. Ensemble sizes of RotBoost, Rotation Forest, Bagging, RSM and Random Forest are set to 100 constituent members per ensemble. All ensemble classifiers are combinations of CART base classifiers. Finally, the number of features per feature subset for both Random Forests and RotBoost is set to 3, as suggested by Rodríguez et al. (2006). All ensemble classifiers are combinations of unpruned decision trees, while the results for the individual classifiers originate from pruned C4.5 and CART classifiers.

Experimental results are all based upon five times twofold cross-validation (5 x 2cv), as recommended by Dietterich (1998). In twofold cross-validation, instances in the data set are randomly assigned to two parts of equal size. One part is once used as training data for a classifier and the performance is calculated for the other part, acting as a test set. This process is then repeated, switching the roles of the two data set parts. In order to test for significant differences among classifiers' results, one-tailed paired

Table 2
Data set properties.

Data set	Instances	Number of features	Minority class percentage
DIY supplies	3827	15	28.14
Bank	20,456	137	5.99
Telecom	35,550	529	2.76
Mail-order garments	43,305	244	1.76

t-test are performed with significance level $\alpha = 0.05$, as for example applied by Zhang and Zhang (2008).

4.3. Results

This section presents the results of the experimental comparison of Rotation Forest and RotBoost to a selection of benchmark algorithms, for data sets from four real-life customer churn prediction projects. Tables 3–5 report result averages and standard errors of results in terms of accuracy, AUC and top-decile lift respectively based on runs from a five times twofold cross-validation (5×2cv). The best and second best results per data set are indicated in bold and italic fonts, respectively.

A first consideration involves the comparison of performance among the alternative rotation-based ensemble classifiers. Table 6 summarizes these results by means of counts of wins, losses and ties, both for absolute figures and based on significance tests, as described earlier. The following observations are made from these results. First, RotBoost clearly outperforms Rotation Forest in terms of accuracy. This holds for all three variations of the proposed algorithms. This confirms the findings of Zhang and Zhang (2008). In their experiments, RotBoost is found to outperform Rotation Forest in terms of accuracy, based on an experimental comparison on 36 UCI data sets. However, in terms of AUC and top-decile lift, Rotation Forests have a clear advantage over RotBoost. As AUC and top-decile lift are the most relevant performance metrics for customer churn prediction, it is found that, based on these results, Rotation Forests is best suited for the prediction of customer churn. Second, there are considerable differences among the alternative variations. Moreover, these differences are not consistent over all three performance measures. The value of considering alternative feature extraction algorithms is most visible when considering Rotation Forests. Rotation Forests based on ICA are superior to the variations based on both PCA and SRP and are found to demonstrate the best AUC and top-decile lift results among all rotation-based ensemble classifiers. This partially extends findings in Liu and Huang (2008), in which ICA was found to improve the performance of Rotation Forests, measured in accuracy, over standard PCA.

Table 7 provides wins, losses and ties counts, both for absolute figures and based on significance tests, for comparisons between the rotation-based ensemble classifiers, and all benchmark algorithms. First, results for accuracy reveal no clear dominance for any of the algorithms. RotBoost demonstrates the strongest accuracy performance among the rotation-based ensembles when compared to the benchmark algorithms. However, none of the algorithms is superior over all others. The strong performance of the individual trees C4.5 and CART versus the rotation-based ensemble classifiers raises questions upon their ability to generate improvements in accuracy in a churn prediction context. Bagging, RSM and CART appear to be the strongest performing benchmarks. Among the Rotation Forest variations, Rotation Forests based on ICA demonstrate the strongest performance, but are outperformed by RSM.

Second, AUC results are better for Rotation Forests than for RotBoost. The best results are observed for Rotation Forests based on ICA and SRP. Both outperform Bagging, RSM, CART and C4.5 and perform comparably to Random Forest. Overall, all variations of both Rotation Forests and RotBoost outperform both individual classifiers, CART and C4.5. From this finding, it is concluded that rotation-based ensembles provide a viable strategy to increase AUC performance over single classifiers. Further, PCA-based Rotation Forest outperforms Bagging, while ICA-based Rotation Forest outperforms Bagging and RSM and performs comparably to Random Forests.

Third, in terms of top-decile lift performance, two conclusions emerge. First, all three Rotation Forest algorithms demonstrate

Table 3
Experimental results: accuracy (average and standard error).

Data set	Algorithm										
	Bagging	Random Forest	RSM	CART	C4.5	Rotation Forest (PCA)	Rotation Forest (ICA)	Rotation Forest (SRP)	RotBoost (PCA)	RotBoost (ICA)	RotBoost (SRP)
DIY supplies	0.67958 (0.038408)	0.645909 (0.013682)	0.667025 (0.047758)	0.671941 (0.049643)	**0.700129 (0.019735)**	0.646432 (0.009015)	0.66102 (0.010544)	0.648939 (0.013673)	0.655999 (0.008319)	0.667188 (0.012593)	0.653907 (0.011945)
Bank	0.761262 (0.023503)	0.741257 (0.017151)	**0.794456 (0.028423)**	0.761593 (0.042192)	0.699556 (0.018432)	0.739721 (0.010459)	0.754114 (0.013706)	0.745531 (0.013073)	0.762816 (0.014914)	0.767919 (0.013969)	0.760119 (0.015856)
Telecom 1	0.627368 (0.139022)	0.611189 (0.012581)	**0.697064 (0.126918)**	0.581751 (0.153487)	0.574666 (0.168184)	0.617108 (0.017847)	0.621165 (0.019457)	0.618172 (0.013381)	0.65585 (0.012307)	0.65909 (0.010704)	0.659862 (0.010045)
Mail-order garments	0.757386 (0.036968)	0.767542 (0.011721)	**0.788453 (0.040276)**	0.728223 (0.075765)	0.704761 (0.014948)	0.752971 (0.01593)	0.759178 (0.01593)	0.759675 (0.01474)	0.78565 (0.010461)	0.78035 (0.008644)	0.782176 (0.007353)

Table 4
Experimental results: AUC (average and standard error).

Data set	Algorithm										
	Bagging	Random Forest	RSM	CART	C4.5	Rotation Forest (PCA)	Rotation Forest (ICA)	Rotation Forest (SRP)	RotBoost (PCA)	RotBoost (ICA)	RotBoost (SRP)
DIY supplies	**0.751875 (0.017471)**	0.715235 (0.005956)	0.735385 (0.015932)	0.68534 (0.011471)	0.704201 (0.024349)	0.714694 (0.007432)	0.728061 (0.006869)	0.719185 (0.006339)	0.712851 (0.007259)	0.716544 (0.006464)	0.711165 (0.006486)
Bank	0.783767 (0.016323)	**0.80794 (0.013705)**	0.781244 (0.014906)	0.711671 (0.014935)	0.704596 (0.018448)	0.802956 (0.013703)	0.800644 (0.013190)	0.788139 (0.02004)	0.796827 (0.017953)	0.788204 (0.016830)	0.773731 (0.022324)
Telecom 1	0.617815 (0.018505)	0.627711 (0.014645)	0.613174 (0.019734)	0.580664 (0.028621)	0.569797 (0.015564)	**0.632488 (0.013966)**	0.63141 (0.014942)	0.624575 (0.019047)	0.616719 (0.012719)	0.608667 (0.017075)	0.615776 (0.012624)
Mail-order garments	0.813901 (0.006573)	0.837894 (0.005450)	0.833482 (0.006506)	0.751118 (0.009261)	0.69375 (0.030787)	0.830779 (0.006997)	**0.837978 (0.004546)**	0.83486 (0.005076)	0.833589 (0.006002)	0.831862 (0.006753)	0.831904 (0.006627)

Table 5
Experimental results: top-decile lift (average and standard error).

Data set	Algorithm										
	Bagging	Random Forest	RSM	CART	C4.5	Rotation Forest (PCA)	Rotation Forest (ICA)	Rotation Forest (SRP)	RotBoost (PCA)	RotBoost (ICA)	RotBoost (SRP)
DIY supplies	1.231839 (0.675133)	1.89094 (0.109593)	1.595527 (0.52256)	1.61893 (0.128876)	1.71041 (0.164107)	1.89467 (0.114842)	**1.97283 (0.074294)**	1.94865 (0.080356)	1.64344 (0.254098)	1.49081 (0.18594)	1.51686 (0.18349)
Bank	3.32787 (0.566213)	4.17382 (0.167603)	3.85036 (0.193287)	2.610721 (0.987598)	1.94886 (0.412066)	4.05457 (0.157386)	**4.18526 (0.198798)**	4.02352 (0.185257)	3.77194 (0.342562)	3.44686 (0.298847)	3.47137 (0.208837)
Telecom 1	1.88947 (0.287528)	2.04901 (0.159907)	2.05103 (0.112790)	1.440157 (0.423809)	1.454928 (0.4527)	**2.15117 (0.206861)**	2.14091 (0.204642)	2.02857 (0.172071)	1.45461 (0.204492)	1.37258 (0.323750)	1.49932 (0.328532)
Mail-order garments	4.48775 (0.286863)	4.94664 (0.202670)	4.75512 (0.291992)	3.918638 (1.071530)	1.67002 (0.333009)	4.82608 (0.231544)	**4.97811 (0.155906)**	4.90222 (0.257748)	4.63743 (0.292847)	4.47486 (0.186871)	4.51686 (0.297776)

Table 6
Performance comparison: wins-losses-ties counts among rotation-based ensemble classifiers. Results are presented for both significance tests and absolute figures (between brackets).

Algorithm	Criterion	Benchmark				
		Rotation Forest (ICA)	Rotation Forest (SRP)	RotBoost (PCA)	RotBoost (ICA)	RotBoost (SRP)
Rotation Forest (PCA)	Accuracy	0/3/1 (0/4/0)	0/1/3 (0/4/0)	0/4/0 (0/4/0)	0/4/0 (0/4/0)	0/4/0 (0/4/0)
	AUC	0/2/2 (2/2/0)	2/2/0 (2/2/0)	3/0/1 (3/1/0)	2/0/2 (2/2/0)	2/0/2 (3/1/0)
	Top-decile lift	0/3/1 (1/3/0)	1/1/2 (2/2/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)
Rotation Forest (ICA)	Accuracy	–	**2/0/2 (3/1/0)**	**0/3/1 (1/3/0)**	**0/3/1 (0/4/0)**	**1/2/1 (1/3/0)**
	AUC	–	**4/0/0 (4/0/0)**	**3/0/1 (4/0/0)**	**4/0/0 (4/0/0)**	**4/0/0 (4/0/0)**
	Top-decile lift	–	**2/0/2 (4/0/0)**	**4/0/0 (4/0/0)**	**4/0/0 (4/0/0)**	**4/0/0 (4/0/0)**
Rotation Forest (SRP)	Accuracy	–	–	0/3/1 (0/4/0)	0/4/0 (0/4/0)	0/3/1 (0/4/0)
	AUC	–	–	1/0/3 (3/1/0)	1/0/3 (3/1/0)	3/0/1 (2/2/0)
	Top-decile lift	–	–	4/0/0 (4/0/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)
RotBoost (PCA)	Accuracy	–	–	–	1/2/1 (1/3/0)	0/0/4 (3/1/0)
	AUC	–	–	–	2/0/2 (3/1/0)	2/0/2 (4/0/0)
	Top-decile lift	–	–	–	2/0/2 (4/0/0)	3/0/1 (3/1/0)
RotBoost (ICA)	Accuracy	–	–	–	–	2/0/2 (2/2/0)
	AUC	–	–	–	–	2/1/1 (2/2/0)
	Top-decile lift	–	–	–	–	0/1/3 (0/4/0)

Table 7
Performance comparison: wins-losses-ties counts of Rotation Forest and RotBoost versus benchmark algorithms. Results are presented for both significance tests and absolute figures (between brackets).

Algorithm	Criterion	Benchmark				
		Bagging	Random Forest	RSM	CART	C4.5
Rotation Forest (PCA)	Accuracy	0/2/2 (0/4/0)	0/1/3 (2/2/0)	0/3/1 (0/4/0)	0/0/4 (2/2/0)	2/1/1 (3/1/0)
	AUC	3/1/0 (3/1/0)	0/3/1 (1/3/0)	2/2/0 (2/2/0)	4/0/0 (4/0/0)	3/0/1 (4/0/0)
	Top-decile lift	4/0/0 (4/0/0)	1/2/1 (1/3/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)
Rotation Forest (ICA)	Accuracy	**0/0/4 (1/3/0)**	**3/1/0 (3/1/0)**	**0/3/1 (0/4/0)**	**0/0/4 (3/1/0)**	**2/1/1 (3/1/0)**
	AUC	**3/1/0 (3/1/0)**	**2/1/1 (3/1/0)**	**3/0/1 (3/1/0)**	**4/0/0 (4/0/0)**	**4/0/0 (4/0/0)**
	Top-decile lift	**4/0/0 (4/0/0)**	**2/0/2 (4/0/0)**	**3/0/1 (4/0/0)**	**4/0/0 (4/0/0)**	**4/0/0 (4/0/0)**
Rotation Forest (SRP)	Accuracy	0/2/2 (1/3/0)	1/1/2 (3/1/0)	0/3/1 (0/4/0)	0/0/4 (3/1/0)	2/1/1 (3/1/0)
	AUC	1/1/2 (3/1/0)	1/1/2 (1/3/0)	3/0/1 (3/1/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)
	Top-decile lift	3/0/1 (4/0/0)	0/1/3 (1/3/0)	2/0/2 (3/1/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)
RotBoost (PCA)	Accuracy	2/1/1 (3/1/0)	4/0/0 (4/0/0)	0/3/1 (0/4/0)	1/0/3 (3/1/0)	2/1/1 (3/1/0)
	AUC	2/1/1 (2/2/0)	0/1/3 (0/4/0)	1/1/2 (3/1/0)	4/0/0 (4/0/0)	4/0/0 (4/0/0)
	Top-decile lift	1/1/2 (3/1/0)	0/4/0 (0/4/0)	0/1/3 (1/3/0)	2/0/2 (4/0/0)	2/0/2 (4/0/0)
RotBoost (ICA)	Accuracy	1/0/3 (3/1/0)	4/0/0 (4/0/0)	0/1/3 (1/3/0)	1/0/3 (3/1/0)	2/1/1 (3/1/0)
	AUC	1/2/1 (2/2/0)	0/3/1 (1/3/0)	0/1/3 (1/3/0)	4/0/0 (4/0/0)	3/0/1 (4/0/0)
	Top-decile lift	0/1/3 (2/2/0)	0/4/0 (0/4/0)	0/3/1 (0/4/0)	1/1/2 (2/2/0)	2/1/1 (2/2/0)
RotBoost (SRP)	Accuracy	1/1/2 (2/2/0)	4/0/0 (4/0/0)	0/1/3 (0/4/0)	1/0/3 (3/1/0)	2/1/1 (3/1/0)
	AUC	1/1/2 (1/3/0)	0/3/1 (0/4/0)	0/1/3 (0/4/0)	4/0/0 (4/0/0)	3/0/1 (4/0/0)
	Top-decile lift	0/1/3 (2/2/0)	0/4/0 (0/4/0)	0/3/1 (0/4/0)	1/1/2 (3/1/0)	2/1/1 (3/1/0)

performance that is at least as good as the benchmark algorithms. Second, Rotation Forests based on ICA demonstrate superior performance over all benchmark algorithms. Top-decile lift measures observed for Rotation Forests based on ICA are the highest among all compared algorithms for three out of four data sets, and second highest for one data set. Random Forests and RSM are the strongest performing benchmark algorithms for this performance measure. This leads to the conclusion that ICA-based Rotation Forest is a well-suited algorithm for customer churn prediction that has the potential to result in higher top-decile lift performance than many well-established classification algorithms, in particular Bagging, RSM, Random Forests, CART and C4.5.

5. Conclusions, limitations and directions for future research

In applications of customer churn prediction, classification performance has a substantial impact upon customer retention and firm profitability. For this reason, classification algorithm choice is an important topic in literature on customer churn prediction. In classification literature, ensemble learning has received a lot of attention in recent years. An ensemble classifier is a combination of several member classifier models into one aggregated model, including a fusion rule to combine member classifiers' outputs. Several studies have indicated that ensemble classifiers substantially improve classification performance in a variety of domains and in churn prediction in particular. In this study, rotation-based ensemble classifiers are evaluated for the prediction of customer defection. In rotation-based ensembles, feature extraction algorithms are applied to rotate the training data that is presented for training member classifiers in the ensemble. In Rotation Forests, the feature set is randomly divided in subsets and a feature extraction algorithm is applied to each subset. The resulting coefficients are rearranged in a rotation matrix that is used to rotate the training data for a base classifier. In RotBoost, the base classifier in Rotation Forest algorithm is replaced with AdaBoost.

This study provides the following contributions to literature on customer churn prediction: (i) it presents a synthesis of literature on the use of ensemble classifiers for churn prediction; (ii) it compares two ensemble-based ensemble algorithms, i.e., Rotation Forest and RotBoost, to a set of often used benchmark algorithms, in terms of accuracy, AUC and top-decile lift on four real-life customer churn prediction applications; and (iii) it compares the influence of the use of three alternative feature extraction algorithms, i.e., principal component analysis (PCA), independent component analysis (ICA) and sparse random projections (SRP) on classification performance of both RotBoost and Rotation Forest.

The main conclusions that are derived from the results are the following. First, a mutual comparison of the rotation-based ensembles demonstrates that Rotation Forests outperform RotBoost in terms of AUC and top-decile lift, while RotBoost demonstrates higher accuracy than Rotation Forests. Second, considerable differences are introduced by implementing three alternative feature extraction algorithms within RotBoost and Rotation Forest. Within RotBoost, ICA and PCA outperform SRP, while within Rotation Forest, a clear dominance is observed for ICA. Overall, both AUC and top-decile lift of Rotation Forest based on ICA are the highest among all rotation-based ensembles. Third, the dominance of ICA-based Rotation Forests for AUC and top-decile lift is also observed in a comparison to the benchmark algorithms Bagging, Random Forests, the random subspace method (RSM), and pruned CART and C4.5 classifiers. In terms of accuracy, none of the proposed algorithms offers a clear advantage over the benchmark algorithms. However, results in terms of AUC and top-decile lift clearly show the competitive nature of Rotation Forests compared to other well-known ensemble algorithms. Moreover, ICA-based Rotation Forests outperform all benchmark algorithms included in the experiments when considering top-decile lift. In summary, this study demonstrates the value of rotation-based ensembles for customers churn prediction and ICA-based Rotation Forests in particular for marketing decision makers who are interested in optimizing AUC, and especially top-decile lift.

Finally, some limitations of this study and directions for future research can be identified. First, the study involves using recommended and default values rather than performing optimization for algorithm parameters. Fine-tuning all algorithm parameters is infeasible in the context of comparison between several algorithms on multiple datasets and is also unrealistic in the time-constrained business context in which customer churn prediction is often applied. However, we agree that optimization of algorithms might have an impact. Second, the impact of alternative strategies to deal with class imbalance (instead of undersampling) is not considered in the present study. Future work could investigate the influence of such strategies upon the performance on Rotation Forest and RotBoost. Moreover, Rotation Forest and RotBoost could be adapted to deal with class imbalance directly, while eliminating the need for additional data pre-processing such as under- or oversampling.

Acknowledgements

The authors thank all former and current Ph.D. researchers at the modeling cluster of the Department of Marketing who contributed the real-life business data sets they gathered and processed during their Ph.D.'s, the reviewers for reviewing the paper and Ghent University for funding the Ph.D. project of Koen W. De Bock. Further, the authors acknowledge the IAP research network Grant No. P6/03 of the Belgian government (Belgian Science Policy). Finally, we would like to express our gratitude to Chuanxia Zhang for the provision of the RotBoost code, the developers of WEKA, R, the randomForest and adabag packages and the Matlab Toolbox for Dimensionality Reduction.

References

Alfaro, E., Gámez, M., & García, N. (2006). *adabag: Applies Adaboost.M1 and Bagging*. R Package version 1.1.

Au, W. H., Chan, K. C. C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation, 7*(6), 532–545.

Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning, 36*(1–2), 105–139.

Bose, I., & Chen, X. (2009). Hybrid models using unsupervised clustering for prediction of customer churn. *Journal of Organizational Computing and Electronic Commerce, 19*(2), 133–151.

Breiman, L. (1984). Classification and regression trees. In L. Breiman, J. H. Friedman, R. A. Olshen, & C. J. Stone (Eds.), *The Wadsworth statistics/probability series* (pp. 358). Belmont, CA: Wadsworth International Group.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Bryll, R., Gutierrez-Osuna, R., & Quek, F. (2003). Attribute bagging: Improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition, 36*(6), 1291–1302.

Buckinx, W., Baesens, B., Van den Poel, D., Van Kenhove, P., & Vanthienen, J. (2002). Using machine learning techniques to predict defection of top clients. In A. Zanasi, C. A. Brebbia, N. F. F. Ebecken & P. Melli (Eds.), *Proceedings of the 3rd international conference on data mining methods and databases* (pp. 509–517).

Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications, 32*(2), 277–288.

Burez, J., & Van den Poel, D. (2008). Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department. *Expert Systems with Applications, 35*(1–2), 497–514.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications, 36*(3), 4626–4636.

Comon, P. (1994). Independent component analysis, a new concept. *Signal Processing, 36*(3), 287–314.

Coussement, K., Benoit, D. F., & Van den Poel, D. (2010). Improved marketing decision making in a customer churn prediction context using generalized additive models. *Expert Systems with Applications, 37*(3), 2132–2143.

Coussement, K., & Van den Poel, D. (2008a). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications, 34*(1), 313–327.

Coussement, K., & Van den Poel, D. (2008b). Integrating the voice of customers through call center emails into a decision support system for chum prediction. *Information & Management, 45*(3), 164–174.

Coussement, K., & Van den Poel, D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications, 36*(3), 6127–6134.

De Bock, K. W., Coussement, K., & Van den Poel, D. (2010). Ensemble classification based on generalized additive models. *Computational Statistics & Data Analysis, 54*(6), 1535–1546.

R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna, Austria.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation, 10*(7), 1895–1923.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In J. Kittler & F. Roli (Eds.), *Proceedings of the 1st international workshop on multiple classifier systems (MCS 2001). Lecture Notes in Computer Science* (Vol. 1857, pp. 1–15). Berlin-Heidelberg: Springer-Verlag.

Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations, 1*(1).

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In L. Saitta (Ed.), *Proceedings of the thirteenth international conference on machine learning (ICML 1996)* (pp. 148–156). San Francisco, CA: Morgan Kaufmann.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119–139.

Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research, 197*(1), 402–411.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 832–844.

Hu, X. H. (2005). A data mining approach for retailing bank customer attrition analysis. *Applied Intelligence, 22*(1), 47–60.

Jolliffe, I. T. (2002). *Principal component analysis*. Berlin: Springer-Verlag.

Kim, Y. S. (2006). Toward a successful CRM: Variable selection, sampling, and ensemble. *Decision Support Systems, 41*(2), 542–553.

Kuncheva, L. I. (2004). *Combining pattern classifiers: Methods and algorithms*. Hoboken, New Jersey: John Wiley & Sons.

Kuncheva, L. I., & Rodriguez, J. J. (2007). An experimental study on rotation forest ensembles. In M. K. J. R. F. Haindl (Ed.), *Proceedings of the 7th international workshop on multiple classifier systems (MCS 2007). Lecture Notes in Computer Science* (Vol. 4472, pp. 459–468). Berlin-Heidelberg: Springer-Verlag.

Langley, P. (2000). Crafting papers on machine learning. In P. Langley (Ed.), *Proceedings of the 17*th *international conference on machine learning (ICML 2000)* (pp. 1207–1216).

Larivière, 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.

Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research, 43*(2), 276–286.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18–22.

Liu, K. H., & Huang, D. S. (2008). Cancer classification using Rotation Forest. *Computers in Biology and Medicine, 38*(5), 601–610.

Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks, 11*(3), 690–696.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J. X., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research, 43*(2), 204–211.

Payne, A., & Frow, P. (2005). A strategic framework for customer relationship management. *Journal of Marketing, 69*(4), 167–176.

Pendharkar, P. C. (2009). Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Systems with Applications, 36*(3), 6714–6720.

Prinzie, A., & Van den Poel, D. (2008). Random forests for multiclass classification: Random MultiNomial Logit. *Expert Systems with Applications, 34*(3), 1721–1732.

Provost, F., Fawcett, T., & Kohavi, R. (2000). The case against accuracy estimation for comparing induction algorithms. In J. Shavlik (Ed.), *Proceedings of the 15th international conference on machine learning (ICML 1998)* (pp. 445–453). San Francisco, CA: Morgan Kaufmann.

Quinlan, R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kauffman Publishers.

Reinartz, W. J., Krafft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of Marketing Research, 41*(3), 293–305.

Reinartz, W. J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing, 67*(1), 77–99.

Rodríguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(10), 1619–1630.

Ruta, D., Nauck, D., & Azvine, B. (2006). K nearest sequence method and its application to churn prediction. In E. Corchado, H. Yin, V. Botti & C. Fyfe (Eds.), *Proceedings of the 7th international conference on intelligent data engineering and automated learning (IDEAL 2006). Lecture Notes in Computer Science* (Vol. 4224, pp. 207–215). Berlin-Heidelberg: Springer-Verlag.

Smith, K. A., Willis, R. J., & Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: A case study. *Journal of the Operational Research Society, 51*(5), 532–541.

Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications, 36*, 12547–12553.

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.

van der Maaten, L. (2007). *An introduction to dimensionality reduction using matlab*. Technical report MICC 07-07. Maastricht, The Netherlands: Maastricht University.

van Wezel, M., & Potharst, R. (2007). Improved customer choice predictions using ensemble methods. *European Journal of Operational Research, 181*(1), 436–452.

Weiss, G. M. (2004). Mining with rarity: A unifying framework. *SIGKDD Explorations, 6*(1), 315–354.

Winer, R. S. (2001). A framework for customer relationship management. *California Management Review, 43*(4), 89–108.

Xie, Y. Y., Li, X., Ngai, E. W. T., & Ying, W. Y. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications, 36*(3), 5445–5449.

Zhang, C. X., & Zhang, J. S. (2008). RotBoost: A technique for combining Rotation Forest and AdaBoost. *Pattern Recognition Letters, 29*(10), 1524–1536.

Zhang, C. X., Zhang, J. S., & Wang, G. W. (2008). An empirical study of using Rotation Forest to improve regressors. *Applied Mathematics and Computation, 195*(2), 618–629.