# analisePDFs_artigos

March 6, 2021

## 1 PDFs analysis

[1]:
```
caminho=!pwd
```

[2]:
```
caminho[0]
```

[2]:
```
'/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles'
```

[3]:
```
import pdfplumber
```

[4]:
```
from transformers import pipeline
```

The framework to use, either "pt" for PyTorch or "tf" for TensorFlow. The specified framework must be installed.

[5]:
```
summarization = pipeline("summarization",framework='pt')
```

```
Downloading:   0%|          | 0.00/1.62k [00:00<?, ?B/s]

Downloading:   0%|          | 0.00/899k [00:00<?, ?B/s]

Downloading:   0%|          | 0.00/456k [00:00<?, ?B/s]

Downloading:   0%|          | 0.00/26.0 [00:00<?, ?B/s]

Downloading:   0%|          | 0.00/1.22G [00:00<?, ?B/s]
```

[6]:
```
caminho = caminho[0]
```

[7]:
```
caminho
```

[7]:
```
'/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles'
```

[8]:
```
import torch
torch.cuda.is_available()
```

[8]:
```
False
```

[9]:
```
import os,re
```

```python
for root, dirs, files in os.walk(caminho):
    for file in files[0:1]:
        if file.endswith(".pdf"):
            #ficheiros.append(os.path.join(root, file))
            file = os.path.join(root,file)
            #Normaliza path
            file = os.path.normpath(file)

            pdf=pdfplumber.open(file)
            all_text = ''
            resumo=''
            print(file)
            for page in pdf.pages:
                page_text = page.extract_text()
                try:
                    resumo=summarization(page_text)
                    resumo=resumo[0]['summary_text']#obter só o texto do resultado
                    print(page,resumo)
                except:
                    print(page,'erro')
                #print(page_text)
                all_text = all_text + '\n' + resumo
```

```
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11391/Hung et al_2006_Applying data mining to telecom churn
management.pdf
<Page:1>  This material is brought to you by the Pacific Asia Conference on
Information Systems (PACIS) at AIS Electronic Library (AISeL) It has been
accepted for inclusion in PACIS 2004 Proceedings by an authorized administrator
of AIS . For more information, please contact contact
contactelibrary@aisnet.org.au .
<Page:2>  Taiwan deregulated its wireless telecommunication services in 1997 .
Churn management becomes a major focus of mobile operators to retain customers .
Figure 1 suggests that Asian telecom providers face a more challenging challenge
than those in the other parts of the world . Study compares various data mining
techniques that can assign a "propensity-to-churn" score periodically to each
and every  subscriber of a mobile operator .
<Page:3>  This paper shares the result of the research. Section 2 defines some
basic concepts (and rationale) that we use in the research . Section 3 describes
our research methodology, and Section 4 presents the findings. Section 5
concludes our presentation. Section 1 defines some of the basic concepts and the
findings .
<Page:4>  Data mining techniques most commonly used include clustering,
associations, rule induction, genetic algorithm, decision tree, and neural
network . Table 1 summarizes some data mining functionalities, techniques, and
applications in CRM space . Churn Prediction Data Mining Assessment Methodology
```

assesses performance of various data mining techniques .
<Page:5>  Research selected Decision Tree, Neural Network and K-means cluster as data mining techniques to build predictive models or segment customers . Churner is defined as a subscriber  who is voluntary to leave; non-churner  is the subscriber who is still using this operator's service . We used latest six months' transactions of each subscriber to  predict customers' churn probability of the following month .
<Page:6>  Models were built by Decision Tree (C5.0) and Back .Propagation Neural Network (BPN) techniques . Data Preprocessing, Variable Analysis and Selection, and Data Extraction, is a formalized system integration process to ensure data quality and code optimization . We took two approaches to assess how models built using Decision Tree and Back.0 techniques perform .
<Page:7>  In Approach 1, we used K-means clustering methods to segment customers into 5 clusters . Then we create a Decision Tree model in each cluster (see Approach 1 in Figure 3) This is to assess if the churn behaviors are different in different "Value-Loyalty's" segments . LIFT is a measure of productivity with modeling .
<Page:8>  A wireless telecom company in Taiwan provides their customer related data . The data source includes data of about 160,000 subscribers, including 14,000 churners, from July 2001 to June 2002 . We got possible variables from other researches and telecom experts' interviews . We then analyzed these variables with z-test from four dimensions .
<Page:9>  To segment customers by loyalty, contribution, and usage, we selected Bill Amount, Tenure, MOU, MTU, and Payment Rate as variables . We used K-Means to model the customers into 5 clusters . To generate roughly the same number of subscribers in each of the 5 clusters, we divided the customers equally into three segments .

Token indices sequence length is longer than the specified maximum sequence length for this model (1530 > 1024). Running this sequence through the model will result in indexing errors

<Page:10> erro
<Page:11>  We use the same training set for BPN as for Decision Tree . Table 6 shows the results, in which model  N18-R6, for example, uses 18 neurons in the hidden layer with 0.6 learning rate . Figure 5 shows that all the models demonstrate stable accuracy in the first 6 months . However, there a significant degradation occurs in the month of February 2002, regardless  of modeling techniques .
<Page:12>  Table 7 lists T-test results:  The performance of decision tree model without segmentation is better than that with segmentation . Table 7 shows that the performance of BPN is better on both  hit ratio and capture rate . Table 8 shows that neural network models are better than decision tree models without segmentations .
<Page:13>  The mobile service provider only budgeted this study at the population of about 160,000 customers, and the associated monthly churn rate was only 0.71% . The data size was not sufficient to build a good predictive model by each customer segment because we could not explore real significant information from only few churners .

<Page:14>  Churn prediction and management is critical in liberalized mobile telecom markets . Mobile service providers must be able to predict possible churners and take proactive actions to retain valuable customers . Data mining techniques can be applied in many fields in CRM space, such as credit card fraud detection and credit score .
<Page:15>  Thearling, Kurt "A Introduction of Data Mining", Direct Marketing Magazine, Feb 1999 . Setiono, Rudy, Liu, Huan "Neural-Network feature selector', IEEE transaction on neural . network, Vol. 8(3), 1997, pp654-661 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11393/Dierkes et al_2011_Estimating the effect of word of mouth on churn and cross-buying in the mobile.pdf
<Page:1>  The results provide evidence that word of mouth has a considerable impact on customers' churn decisions and also on the purchase decisions, leading to a 19.5% and 8.4% increase in sensitivity of pre-dictive models . The results show that information on the churn of network neigh-bors has a significant positive impact on the predictive accuracy and in particular the sensitivity of churn models .
<Page:2>  The churn rate refers to the proportion of contractual customers or subscribers who leave a service pro-uablyvider during a given time period . It is a possible indicator of customer dissatisfaction, cheaper and/or bet-ter offers from the competition, or reasons related to the customer life cycle . The accuracy of churn prediction models matters and that just using one method rather than another  can easily amount to changes in profit in the hundreds of thousands of dollars .
<Page:3>  We use customers' anonymized call detail records as a way of modeling WOM . In contrast to traditional classification methods, we take into account the infor-heticalmation about who a customer calls - i.e., a customer's neighbors in the communication graph derived  from the call detail record data . We interpret these graphs as social networks, which can be stored in a relational data model .
<Page:4>  Statistical relational learning is a relatively young field and there is still only limited empirical evidence on the performance of respective learners . There are two fundamentally different approaches to analyzing multi-relational data . Markov logic networks (MLNs) have recently been suggested as a significant step forward in this field .
<Page:5>  The prediction of customers' churn or buying decisions is important to marketers . In this paper, we analyze whether WOM has an impact on customer behavior or not . The analysis is based on anonymized calling data from a telecom provider . We found the churn behavior of a customer's neighbor has a significant positive impact on predictive accuracy of churn models .
<Page:6>  Churn management is to determine the reasons for churn and to predict the potential churners . In the following section we discuss churn and the related literature on churn and WOM literature . There are different strands of literature that are relevant to this paper . There is also a relation between customer tenure and the tendency of customers to engage in word of mouth .
<Page:7>  A huge body of literature has emerged on the analysis of social networks . We focus on post-paid customers, which allows us to leverage information about potential churners . The authors study the evolution of

churners in an operator's network of pre-paid subscribers and the pro-verselypensity of a subscriber to churn out of a service provider's network depending on the number of ties .
<Page:8> Statistical Relational Learning (SRL) has been an emerging research topic in the data mining community in recent years . We understand propositionalization as a transformation of multi-relational learning prob-lems into attribute-value representations . We discuss this approach in more detail in the next section .
<Page:9> ILP has traditionally dealt with multi-relational data . ILP tools can be applied directly to multi-referred data to find first-order rules from relational data . Markov logic networks (MLNs) have become very popular in statistical relational learning recently . In our analysis, we use Alchemy (http://alchemy.cs.washington.edu/), an open source software tool for learning MLNs from data .
<Page:10> There are two main approaches to propositionalization in the literature, logic-oriented and database-oriented literature . Business databases present different challenges than those found in the classical showcase areas of ILP and logic-based propositionalizations . Relational databases are usually structurally simpler .
<Page:11> A churner is defined as a customer who gives notice about their intent to cancel the contract and does not re-evaluate his decision by extending his contract at some point afterwards . A non-churner (or negative) is a customer who does not give notice at any time . In our data, 6,800 customers told the phone provider that they wanted to cancel their contract (notification) Roughly 1,000 revoked their decision afterwards by extending their contracts .
<Page:12> Figure 1 (a) and (b) shows a visualization of connections be-tween positives and customers with a game download in the test data set . The graphs could suggest that game downloading is contagious because there are many connections between persons downloading games, while this is less so in the case of churn customers .
<Page:13> We split the set of 2,645 customers in training and test datasets such that it was stratified with re-naissancespect to the number of positives and edge counts . Training and test data contained 1325 and 1320 custom-ishlyers, respectively . Positives were assigned according to their notification date; customers who notified before July 1, 2008 were assigned to the training data .
<Page:14> Among the selected 3,000 customers there were 7,950 edges - 1,865 among positives, 2,127 among negatives, and 3,950 between both groups . From the 70 available customer attributes we selected the 32 best ones based on information gain with respect to the target variable .
<Page:15> MLN, classification is the problem of inferring the truth value of C(x,v) for all x and v of interest . In relational learning problems, dependencies between objects can be represented by relational predicates . In our example, churn of customer x would be considered independent of the churn of other custom-crafteders . We model an influence relationship between connected customers' urchurn behavior saying that customer x is likely to churn if customer x did already .
<Page:16> A fundamental problem with database-oriented propositionalization has

been referred to as degree disparity [25]. It describes the systematic variation in the distribution of the degree with respect to the target variable . A customer with a large number of neighbors would also have more churn neighbors than a customer with just a few neighbors . The difference among MLN settings is again the way in which connec-propriitionalization approaches are different .
<Page:17> The resulting models for the logistic regression (benchmark) as well as for all propositionalization settings showed that three groups of customer attributes were especially important . In the logit model, all but one variable about a customer's utilized products were highly significant ( <0.001) In contrast, for propositionalized settings T2.1-to-T2.4 we found only five significant customer attributes in the respective logit models .
<Page:18> Propositionalization settings with churn aggregates (T2.1 to T2.4) dominate all other settings, apologetic and MLN settings . Sensitivity measures the proportion of true positives that are cor-rectly recognized as true positives . Specificity measures proportion of false positives correctly recognized as false positives . Table 1 presents the results for churn prediction .
<Page:19> The best propositionalization approach outperforms the best MLN model . Figure 2 shows the ROC curve of the three MLN settings and the logistic regression . The ROC is based on the proportion of true positives (TPR) vs. false positives (FPR) for every possible cutoff .
<Page:20> Propositionalization with churn rate aggregates (T2.1 to T2.4) has the highest overall sensitivity and accuracy, but MLNs yield comparable results for smaller samples . Both propositionalization and MLN clearly outperform the baseline model .
<Page:21> We were interested in ways how information about neighbors can help predict alternative target variables . We looked at game download, since people with game downloads appear to be well connected . Game download showed exceptionally many connections among the positives, suggesting that there might also be an influence of customers on each other .
<Page:22> The overall accuracy and precision were again highest for the propositionalization settings T2.1 to T.2.4 and the three MLN settings T3.1 . All nine settings performed better than the benchmark logistic regression, except for one attribute about the duration of the present contract .

Your max_length is set to 142, but you input_length is only 140. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:23> Both propositionalization and MLN outperform the baseline model . The number of calls and voice minutes weighted did not have a strong effect . The impact of relational attributes was highly significant for predicting game download as well, but had a lower effect on sensitivity and the ROC curve compared to churn prediction .
<Page:24> Figure 5: Visualization of connections a) between negatives (without churn), b) from positives with game (squares) to negatives (triangles) and c (without game download) Figure 5 shows the connections between negatives and negatives (with churn) and negatives without churn . Figure 4: Churn: positives to negatives, negatives to negatives; game download: negatives only .

<Page:25> Churn positives are less common than customers with a game download in the sample . Overall, positives are much less connected in the churn sample than in the game download sample . If there was churn in the neighborhood of a cus-culartomer, this event was a powerful predictor for churn of this customer, as compared to game download .
<Page:26> Traditional discrete choice models do not allow the influence of peers through a social network to be modeled . We developed an MLN for churn prediction based on the anonymized data set of a mobile phone provider . We found this approach to provide even better results than MLNs in terms of increasing sensitivity of the benchmark logit model .
<Page:27> The book was written by W.-H. Au, K.C. Chan, X. Yao and S. Dzeroski . We would like to thank Florian Wangenheim, the anonymous associate editor and the reviewers for their valuable comments .
<Page:28> L. Getoor, B. Taskar, D. Mayzlin and D. Godes have written a number of articles on the topic of customer churning . Churn data mining techniques have been used to model churning patterns in the past . The study has been published in the European Symposium on Artificial Neural Networks in Bruges .
<Page:29> Decision Support Systems, 49(4) (2010) 474-485. [24] O. Hinz, M. Spann, Managing information diffusion in Name-Your-Own-Price auctions . [25] D. Jensen, J.-U. Kietz, J. Neville and M. Hay, Avoiding Bias when Aggregating Relational Data with Degree Disparity, in: Twentieth International Conference on Machine Learning (ICML-2003), (Washington, DC, USA, 2003).
<Page:30> Customers Churn: Stop it Before it Starts, Mercer Management Journal, 17(2004). [36] M. Kon, M.-A. Krogel, S. Rawles, F. Zelezny, P. Flach, N. Lavrac and S. Dzeroski .
<Page:31> P.A. Schweidel, P.S. Fader, E.T. Bradlow and J.R. Quinlan have written a number of articles on machine learning and customer retention models . P.J. Shaw, C. Subramaniam, G.W. Tan, M.E. Welge and M.C. Mozer have written numerous articles on the topic of machine learning .

Your max_length is set to 142, but you input_length is only 120. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:32> P. Getoor, B. Taskar Eds. Introduction to statistical relational learning, (MIT Press, Cambridge, Mass., ,) 2007), pp. 93-129 . P. Sutton, A. McCallum: An Introduction to Conditional Random Fields for Relational Learning .
<Page:33> Figure 7: ROC curves of three MNL settings and the logistic regression for game download . Figure 8: Roc curves of propositionalization settings . Figure 9: ROC curve of MLNs, propositionalized settings and logistic regressions . Figure 10: The logistic reparation for game download is based on a logistic model .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11397/Liu et al_2020_Micro- and macro-level churn analysis of large-scale mobile games.pdf
<Page:1> Mobile gaming has emerged as a promising market with billion-dollar revenues . A critical challenge is to understand churn behavior in mobile games . Churn behavior usually involves churn at micro level (between an app and a speci c user) and macro level . We model this two functions by deep neural networks with a unique edge embedding technique .

<Page:2> The experimental results with this data demonstrate the superiority of our proposed models against existing state-of-the-art methods . $70.3 billion revenue was generated by mobile games in 2018 according to Newzoo's Global Games Market Report [1] Mobile games are expected to generate $106 billion revenue in 2021, account for more than half of the overall game market .
<Page:3> A recommender can slightly adjust the position of recommended games before delivering to pre-emptive users . The macro-level churn ranking problem is intimately related to the micro-scale churn prediction problem . The method is fully automatic and can be easily integrated into existing mobile platforms .
<Page:4> The solution has been tested in Samsung Game Launcher, one of the largest commercial mobile gaming platforms . We propose a simple method SimSum and adapt two link analysis algorithms to solve the macro-level churn ranking problem . The model is able to embed new users or games not used in training . The paper mainly applies the solution to churn analysis in other contexts .
<Page:5> The relationship between players and games can be represented by an at-heticaltributed bipartite graph as illustrated in Fig.1 . In the sequel, we use the terms player and user, and game andapp interchangeably . We note that uninstall is di erent from churn .
<Page:6> We propose a novel inductive semi-supervised model in dynamic graphs that jointly learns the prediction functionf and the embedding function g . The architecture of the proposed DNN is presented in Fig. 2, which consists of three parts . In the mobile game industry, play-goers and games change very quickly .
<Page:7> The overall objective function of our model consists of four parts: the unsupervised component, the supervised component, Part II, Part III and Part III . We propose a novel random walk to sample similar edges as contexts (see Section 3.4 for details). Part I handles the graph dynamics in training . Part III ful lls the supervised churn prediction task from embedding featurevectors .
<Page:8> L denotes the unsupervised loss, which comes from failures of U.U.U-U-centric context inference and will be addressed in Section 3.2.1 . L is the temporal lossT thatconsistsoftwoparts:temporalsmoothness andtemporaldynamics, andwill be explained in section 3.1.2 . L represents the regularization of a regularization term, and ( , , ) are trade-o weights .
<Page:9> erro
<Page:10> The longer a relationship exists, the longer the relationship, the more likely the user is to churn the game . This is because the content of a mobile game is usually somewhat somewhat somewhat xed . 95% of user-game play relationships end after 40 days, which well justi es our findings .
<Page:11> Average mobile game retention rate as a function of time of time, weomitthe gurehere . It followssthatthetopologyandattributevaluesof the attributed bipartite graph mostly evolve smoothly attwo consecutive times-tamps, resulting in similar contexts for a given edge at consecutive timestamps .
<Page:12> An illustration of censored data by three randomly sampled player-game-game portions becomes unknown . Since the existence of edge (u,v) after t is unconfirmed, the existence is unknown . The temporal loss L is expressed as: (cid:107)hlp(z(i+1) + (12)glyglyglyT uv uv e uv 2glyi=t0(u, v) GlyglyTglyL: (glyglytgly) +glyglyL = (glygglytlyl) Glycglyglycglyclytglycryclyl, (glycilytglyxglycyl
<Page:13> A simple topology-based random walk may return two adjacent edges

having the same player or the same game while ignoring the similarity of the similarity . In contrast, attributed random walk measures such similarity by attributes and allows to transit to similar nodes even if they're not connected .

<Page:14> The macro-level churn ranking is to provide a ranked list of games based on their total numbers of users to churn in the near future . We use the SimSumSum method to rank all games with an unbiased estimation of the ground-level Churnprobability of the game v .

<Page:15> erro

<Page:16> In the algorithm, players with distinct tendencies of churn contribute di erently to the computation of the game’s ranking score . The score will be propagated along with a dampingfactor from onegametoanothergame . The algorithm is weighted by the churn probabilities, where the players with larger churn probabilities are more likely to contribute .

<Page:17> We conduct a comprehensive experimental evaluation over the large-scale real data collected from the Samsung Game Launcher platform . We compare our semi-supervised model with the state-of-the-art models for mobile game churn prediction: LR: logistic regression based solution used in [17, 13, 9]- RS: supervised variant of our model, DT: decision tree based solution, RF: random forests based solution .

<Page:18> The SimSum, PageRank and HITS methods are evaluated on the test datasets . 18 X.Liuetal-Georgian-Semi-supervised model along with SimSum and PageRank are evaluated . The regularization parameters { }4 are all set to be 1.2.4aresettobe1,1,and0.05,respectively . The maximal number of iterations M for the PageRank is set to 100. The Dampingfactor for PageRank issettobe0.85.

<Page:19> We use three widely-used evaluation metrics to compare the performance of micro-level churn prediction models . The most important metric with respect to the business goals is the area under the ROC curve (AUC) Our model achieves the best AUC and recall on both datasets .

<Page:20> SS outperforms RS in general under di erent numbers of epochs and for both Korean users and USA users . SS is neither over-over-supervised nor unsupervised, we expose more details on how we choose the parameters and train the model . We experimentally test learning rates between 0.1 and 0.001 to converge slowly to the optimal point .

<Page:21> We try two training methods: co-train and alternative train . Co-train means that we simulta-ishly train the supervised loss function and the unsupervised loss function . Alternative train is a widely-used training method for similar structures[21,20]

<Page:22> We use Kendall’s Tau correlation coe cient, weighted rankings and Spearman correlation . We compare the performance of these two methods to the results of micro-level churn-prediction . Each performance metric is experimentally evaluated on the same test datasets described in Section 4.1 .

<Page:23> Figure 10. Comparison of Kendall’s Tau correlation coe cients in testing under the macro-level churn ranking and micro-level . churn prediction methods . Table 10.1 shows the results of the study using Kendall's Tau correlation ranking and a page-rank ranking .

<Page:24> SimSum Weighted Kendall’s Tau on
rea.u.uu.C.Kendall’STauonKo- Researchers. 24 X.Liuetal X. 24 . Liuetal: Charts

of churn ranking and micro-level churn prediction methods .

Your max_length is set to 142, but you input_length is only 124. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:25> Spearman correlation coe cientonKorea tests under di er-ishly-regulated macro-level churn ranking and micro level churn prediction methods .
PageRank SpearmanCoe cientonUSA rea. rea rea
Churn Analysis of Large-Scale
Mobile Games 25 "Churn
analysis of large-scale mobile games" is based on Spearman.
<Page:26> Figure 13. 26 X.Liuetal: PageRank Average Precision at K on (f)PageRankAveragePrecisionatKonKo-USA rea.u.KonKorea (d)HITS AveragePrecision atKon.USA) PageRank average Precision at . K in testing under . di erent macro-level churn ranking ranking and micro .level churn prediction methods .
<Page:27> We evaluate the performance of di erent churn ranking methods by using metrics in the recommendation domain . Precision at K corresponds to thepercentage of relevant results in the top K games of the ranked list . A game at a certain position i in a ranked list is considered relevant if it is in a top i game . Table 5 shows the results based on the methods in terms of MAP .
<Page:28> There are several recent deep-learning-based studies [33, 34, 35] for non-game churn-prediction problems, which report better performance . We propose to reuse the estimated churn probabilities of the micro-level churn prediction task to reduce the overhead . While being a generic solution, our model is able to accommodate the unique characteristics of mobile gaming . We provide a comparisonison betweenallexistingworksandours in Table6.
<Page:29> In our problem new users and new games appear continuously; new relationships between existing users and games mayform any time at any time . We propose a novel inductive semi-supervised model for large-scalemicro-level gamechurnprediction . The model captures graph dynamics by simultaneously capturing contexts and graph dynamics . We modeled the prediction function and the embedding function by deep-scale networks .
<Page:30> Mobile revenues account for more than 50% of the global games market as it reaches 137.9 billion in 2018 . The findings could be modeled in a similar way, for instance, customer disengagement prediction in membership business (e.g., Apple Music, Costco, and insurance companies) and interest group unsubscription prediction in social networks .
<Page:31> Aims to measure player retention and monetization in free-to-play games with highly biased data . "Asemi-supervised learning with graph embeddings" is discussed in the journal of the ACM (JACM) 46.5 (1999) "Authoritative sources in a hyperlinked environment". "SEANO: "Semi-Supervisedembeddinginattributednetworkswithoutliers"
<Page:32> Researchers from Texas A&M.University and Samsung Research.com discuss deep-learning machine learning and deep-reinforcementlearning. The study was published in the journal International Research.JournalofEngineeringandTechnology(IRJET)4.3(2017),pp.1846-1854 .
<Page:33> Xidao Wenisa Ph.DcandidateatPITTComputationalSocialSci- Researchers at University of Pittsburgh . His research interests include data mining, machine learning and machine learning . He has published more than 40 technical

papers in top venues, including CSUR, VLDBJ, PVLDB, and IEEETKDE . He received the ICDM-2011 Best Research Paper Award, Excellence in Academic Research at Rutgers University in 2013 .
<Page:34>  Nick Duwice received a BA in Natural Sciences in 1982 and an MMath in 1983 from the University of Cambridge, UK, and a PhD in Mathematical Physics in 1987 . Na Wang received her M.S. and Ph.D. degrees in Information Sci-ences and Technology from the Pennsylvania State University . She is currently a Software Engineer at Samsung Research America .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11402/Kaya et al_2018_Behavioral attributes and financial churn prediction.pdf
<Page:1>  Erdem Kayaetal. and ZiaowenDong, Xiaoowen Dong, YoshihikoSuhara, SelimBalcisoy,Selim Balcisroy, BurcinBozkaya and Alex "Sandy'Pentland's'. E.Kayaetel.""Churnprediction" has attracted greatattention from both the businessand academic worlds .
<Page:2> erro
<Page:3> erro
<Page:4>  EPJDataScience: The spatio-temporal and choice models utilized in this paper can also be used to churn prediction problems of other domains such as telecommunication in-repredustries . The paper provides insights into the psychology and economics of credit card transactions, money transfers, and electronic fund transfers .
<Page:5> erro
<Page:6> erro
<Page:7>  There exists numerous de nitions of customer churn in the literature . By adopting the bank's de-nition of churn, we developed a set of de-churner-sponding labels . For example, a cusus-based order of the inactive order of each customer based on a non-pertaining to the rules of the customer .
<Page:8>  EPJ DataScience: The results reported in this study are based on the churn de nition inac-full . The data sets were so large that we decided to use 8-fold cross-validationforevaluation . We used SVM-SMOTE[39] withtheratio of 0.25, the average credit card transaction was $1.25 .
<Page:9>  Around 40-.50thousandcustomersalongwith1.9to3.3milliontransactions have been considered . The resultsremainsigni cant even for11dif-ferentde nitionsofchurn, andvariousversionsofthedatasets .
<Page:10>  EPJDataScience: Areaunder ROCcurvemetriccomparison ofdemographicfeatures, STCbehavioralfeatures, and the.ypecombinationofbothfeaturesetsforeachofthedatasetversions . The.comparisons performedforthelabelabelinac-full. The lengthoftheerrorbarcorrespondsto1.0 standarddeviation.p** . The results were very similar .
<Page:11>  EPJDataScience: FeatureImportanceAnalysis. It is notablethattheeducationalstatuses, high, andmiddleschool were foundtobeimportant in this order . The average number of college students who have higher-than-despirited grades and high-preferred college degrees is found to be higher than the average college student .
<Page:12>  EPJ DataScience: Churnpredictionperformancebasedongenderandagegroups . College, and masters,comprising about12-13% of allthecustomers, are not considered . College and masters students are considered in the study .

Churnersseemtohaveincreasingloyaltyandde-
recreasingdiversitytrendtowardsthetimetheydecidetochurn .
<Page:13> erro
<Page:14> erro
<Page:15> erro
<Page:16> erro
<Page:17> erro
<Page:18>  EPJ data science: EpJ: Thecurse ofimbalanced-learn:aPythontoolboxtota
cklethecurseofimbalanceddatasetsinmachinelearning.coRRabs/1609.06570http://arxiv
.org/abstract.com/gourere/Goure.org: The case is presented in EpJ DataScience .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11406/Mitrovic et al_2018_On the operational efficiency of
different feature types for telco Churn.pdf
<Page:1>  On the Operational E ciency of Di erent Feature Types for implementing
Telco Churn Prediction: FeatureEngineering, FeatureTypeClassi cation,OptimalFeat
ureTypeCombinations,OperationalE ciency,ChurnPrediction,Pareto.-Multi-
ObjectiveOptimization, Paretomulti-criteriaoptimization .
<Page:2>  Churnprediction(CP) is probably the
mostfrequentlytackledpredictivetaskinthetelecom-municationindustry . Many
di erent approaches have been used in recent studies . The mostimportant driver
for this study is that none of the existing studies discusses the resource ef-
privilegedtodataavailability,datacollection, feature engineering,
featureengineering, time and model evaluation .
<Page:3>  Onlinereal-time CP is becoming more and more important (Diaz-Aviles et
al., 2015). The OE/PP trade-o  becomes imperative . The key contributions are
threefold: a new feature type classi cation; 2) a novel reusable methodology for
determining optimal feature type combinations .
<Page:4>  The paper is organized as follows: in Section 2, we provide an
overview of related work . Section 3 introduces the methodology with Section 4
detailing the experimental experimentalsetup . In Section 5,wepresent results,
whicharediscussedin Section6, before the paper is concluded in Section7 .
<Page:5>  There are several studies that compare di erent feature types on PP .
However, the comparison scope is typically limited to a strict comparison
between feature types or their combination as well . Other studies disagree on
the importance of feature types' importance . This is an excellent motivation
for analyzing the impact of feature type on PP at a higher level .
<Page:6> erro
<Page:7>  The most thorough study on the OE/PP trade-o  is, looking into several
PPMeasuressuchasaccuracy,AUC,precision,preision,recall,meanabsoluteerror, and
test time . However, the study does not take into account resources needed for
feature engineering (OEOE)
<Page:8>  A number of alternatives to Pareto optimization have been proposed as
well . These include the lexicographic method, where each of the objective
functions is optimized separately (one at a time) and thecalarization method .
We deem the strategies inconvenient for CT/AUC trade-o  since they not only
require an a priori assumption of a priora assumption of objective preference
and a convex search space, but also provide a single solution that cannot
properly balance di erent objectives .

<Page:9> A distinction is made between features using the most recent data and features using older data . Local and network features are exploited, distinguishing between local and network . We further categorize network features into direct (simple) and indirect (complex) features . Since bothnodede-walletgreeinformationandRFM(Recency-Frequency-Monetary)features(Huges,1994)features are considered indirect features .
<Page:10> With RFM, we account for recency of calls, number of calls and monetary value of calls per customer within a certain period of time . In order to capture more detailed infor- mation, in the nal list of features each of the RFM (and degree-related) features areassessedinseveraldi erent avors,alongfollowingdimensionsfortheobservationalversion .
<Page:11> Thede nitioncolumnexplainsthenatureand purposefully originofdi erentfeaturetypesandassuchcan beappliedtoanytypeofdataset(evenbeyondthe typeofdatatomain) Feature Type De Name: Local Features that characterize indi- gender; number of reloads; handset characteristics (e.g. traditional) National Features calculated from the number of incom-insured direct Engi- customer ego-network (1-level ing/outgoing toward a neighbourhood) across di-erent home operatorcalls in di Aerent homeoperatorcallsin various dimensions/granularities month
<Page:12> The main goal of this work is to develop a method that can identify feature types(orcombinations) whichprovid the bestCT/AUC trade-o . Pareto multi-objective optimization takes into account the trade-O between the trade and objectives of the task .
<Page:13> Givenobjectivefunctions , - , ; ( ) is called the multi-objective optimization task . Wesay that asolution isadominating reformative solutionover 2 if "betterthanitinatleastone'sobjective and not worse in all other objectives." The set of allParetooptimal solutionsiscalled theParetoset .
<Page:14> The algorithm is based on a collection of di erent feature type sets . It starts from thenon-dominatedsolutionsthatarefoundinpreviousiterations . The cardinality of feature types per iteration is exactly one feature type higher than the current one .
<Page:15> Every feature type combination is evaluated in terms of CT and AUC . Logisticregression(LR)without regularizationand RandomForests(RF)are used formodelconstruction . Both LR and RFare well establishedmeth-insuredods and have been successfully applied in previous studies on CP .
<Page:16> erro
<Page:17> Data and tools wereperformed using two data sets . CDRscon-taintain information about customer calls (no SMS or other usage types), in the form of: caller,callee,date/time,callduration . For prepaid customers, we have information about the lastrecharge and the amountspentonvoiceandSMS during theobservedmonth . For postpaid, themonthlysnapshotforported-outcustomers is provided as well .
<Page:18> Theobtainednetworkisusedtocalculatedirectnetworkfeatures . Edgeweights areassigned basedonthetotalnumberofcalls between two nodes, for the rst, andtotalduration(thatis, totalnumberofseconds)ofthecalls . The network is weighted networks with approximately.5millionedges and 2.8millionofnodes .

<Page:19> Wealsocalculated(twoversionsof) personalizedPageRankscores based on onexponentialtimedecays(tofavormorerecentchurndatesfromthe,)lessrecentones) We are trying to predict who will churn in month M for the customer base of month M . The exact numberoffeaturesperfeaturetypeperdatasetcanbeseeninTable3 .
<Page:20> To avoid multi-collinearity, Chi-square and Spearman correlation tests were applied to all pairs of categorical and continuous features . As a result, several features (mostly related to handset char-otypesacteristics, butalsopersonalizedPageRankscores) wereeliminated .
<Page:21> RF and LR provide di erent results in terms of AUC per combination . Optimalsolutions for LR and RF share the following three-problems: L, ND-E, HLNT-NE, HNNTD-E and HLT-E . However, therearenolineardependencies between the two solutions .
<Page:22> Figure4: Theshortlistedsolutionsfortheprepaiddataset and theirperformanceinCT(relatively)to the CT obtained using all features . AUC scores of other optimal solutions at the 95% con dence level (see) Table C.10 and Table C .11 in Appendix C . Similar as with prepaid, a mixture of localandobservationaldirectnetworkfeatures scores better in AUC .
<Page:23> Figure5: Theshortlistedsolutionsforthepostpaiddataset and theirperformanceinCT obtained using all features; -axis) and AUC ( ) obtained applying LR (top) and (bottom) Figure 5: Theoptimalsolutionsaremarkedwithred lledcircles/squares whilethedominatedsolutionsareblueemptycirclecles/sqares for LR/RF,respectively . Figure 6: Figure 6 for prepaid RF and Figure 7 for postpaid RF(duetobetterAUCscoresweexplainRF resultsin more detail)
<Page:24> Only ten di erent features are retained for all four Pareto optimaloptimalsets . Local and direct network features seem to be of crucial importance for prepaid RF . Only features of four feature types remain in the nally retained feature sets: L, HLNT-NE,ND-E, HNNTD-E .
<Page:25> For postpaid RF, in total 18 di erent features are retained, of which: six local-centric, threelocalhistorical, sixdirectnetwork(observational) and sixdirect network(observed) The feature lifetime in days (ɷ) appears in all of the feature features .
<Page:26> We basedourapproachonRFM(already successfullyappliedintheCP domain(Benoit&VandenPoel, 2012;Coussement.& De Bock, 2013; Modani et al., 2013) Page Rank scores (also used in (Huang.e.etal., 2015) and other measuresinspiredbyBaesensetal.(2015) However, we are aware of several allimitations whichledtolowerAUCscoresthanusually reported in the CP studies .
<Page:27> We propose novel, reusable method for determining optimal solutions, based on Pareto multi-objective optimization . The method re-quiresnoaprioripreference betweencon ictingobjectives, while it stillallowsfor making aninformeddecisionbasedonthePareto-optimalsolutions . The obtained results demon-ishlystrate that the choice of modelling technique matters .
<Page:28> The right approach for choosing feature types for CP would be to start small, using (good quality) local features and the least complex network features . We observe that investing in certain more complex feature types like trends and indirectnetworkfeatures does not payo intermsofPP .
<Page:29> Researchers: Customer churn prediction in gambling industry: The bene cial e ect of ensemble learning . The research was published in the journal

of the European Journal of Operational Research, which is published by Springer . The study was published by the journal ACM.com .
<Page:30>  Diaz-Aviles, E., Pinelli, F., Lynch, K., Gkoufas, Y., Bouillet, E. Bouillet and Bouillet . Cal-abrese,F.,Coughlan,E.,Holland,P., & Salzwedel,J. (2015) Towardsreal-time customer experience prediction for telecommunication operators. In Big Data(BigData),2015IEEEInternationalConferenceon(pp.1063-1072).
<Page:31>  Analysing customer attribute and social network mining for prepaid mobile churnprediction . In Proc. the 23rd Annual Belgian Dutch Conference on Machine.Learning(BENELEARN) (pp.50-58) The PageRank citation citation ranking: Bringingordertotheweb.. TechnicalReportStanfordInfoLab.
<Page:32>  Verbeke, W., Dejaeger, K., Martens, D., & Baesens, B.G. (2010). Customer churn-drivenprediction: doestechniquematter? InProceedingsoftheJointStatisticalMeet-Likeing,JSM2010,Vancouver, Canada .
<Page:33> erro
<Page:34>  TableA.6: Resultsfor ‘shortlisted’featuretypecombinationsobtainedwith LRforpostpaiddataset.computership computation has computation has been used for two 64-bit processors at two 64GB of Intel processors working at two.2, 3.3GHz withcoresores .

Your max_length is set to 142, but you input_length is only 88. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:35>  FigureB.8: Retainedfeatures(intheLRmodel)forParetooptimalfeaturetypecombinationsfortheprepaiddataset . The feature typecombinationsaresortedbyincreasing order of AUCperformance(fromlefttoright)
<Page:36>  Preferred with LR, prepaid with RF, prepaid with RF and postpaid with LR and RF . Postpaid withLR and RF were pre-loaded with RF . Preferred LR was preloaded with LR; RF was paid with RF; prepaid with LR . Preloaded withLR, RF was pre-paid and pre-depended with RF. Prepaid withRF was prepaid withLR; Prepaid and RF with RF were prepaid .
<Page:37> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11408/Oskarsdottir et al_2018_Profit-Based Model Selection for Customer Retention Using Individual Customer.pdf
<Page:1>  Pro t based model selection based on model selection using individual customer lifetime values . Maria Oskarsdo´ttir, Bart Baesens and Jan Vanthienen are the authors of the study at KU Leuven and the University of Southampton .
<Page:2>  The goal of customer retention campaigns, by design, is to add value and en-hance the operational e ciency of businesses . For organizations that strive to retain their customers in saturated, and sometimes fast moving, markets such as telecommunications and banking industries .
<Page:3>  In the case of customer churn prediction (CCP), including a person who is not likely to churn in a retention campaign, will not a ect the company very much, while failing to identify a potential potential churner will causelosses . The state-of-the-art Maximum Pro t (MP)3 and Expected Maximum Pro-
<Page:4>  Customer Lifetime Value (CLV) is de ned as the present value of all the future cash attributed to a customer’s relationship with an organization . The EMP measure, as proposed by Verbraken et al., assumes a  xed and equal CLV for all customers . In this paper, we introduceceanewwayofincorporating

customerheterogene-genicity in the earlier introduced EMP measure by allowing the CLV to vary on a subject basis .

<Page:5>  We apply boot-strap techniques to help distinguish between good and bad models . We apply our techniques to two real life datasets and benchmark datasets using six distinct classi cation techniques . We discuss the usefulness of our approach, compared to the standard EMP measure . Since our method takes into account the variability of the customer base, it has the ad-vantage over the traditional EMP measure to provide a range in performance .

<Page:6>  Pro t based model selection with individual CLV models . Table 1 shows a confusion matrix resulting from suchaclassi  er, withacuto t. Inthismatrix, N denotes the population size,  0 and  1 the prior probabilities of classes 0 and 1 and the cumulative distribution functions of the scores for both classes. Then,inthematrix,N 0F0 representssthenumber of actual churners and N 1F1(t) the number of actual.non-churners classi.i ed incorrectly as churners .

<Page:7>  Pro t based model selection with individual CLVthattheAUCisanincoherentmeasure ofaggregatedclassi cationperformanceperformance . Hand proposed the H-measure, which minimizes the ex-expectedlossofaclassi  er,ortheaverageclassi-cationloss, given by the function 'glyglyglyphobicQ(t,c,b)')'H'measure is an alternative to the H measure . Verbeke et al. proposed the maximum pro t measure as an alternative .

<Page:8>  Pro t based model selection with individual CLVvalue (CLV) The probability that the retention o er has a negative e-ect is considered negligible . N is the total number of customers and A the cost of administrative costs . The value of EMP can be computed using an empirical convex hull .

<Page:9>  Pro t based model selection with individual CLV customer churn is given by a pro-propro-profusion factor . The fraction is an advantageous side-o product of the EMP measure . The top decile lift is commonly used for customer churn models as it compares the ratio of churners in the the 10% of customers with the highest predicted probabilities to the ratio in the actual customer base .

<Page:10>  There are numerous challenges of computing and using CLV, with many issues and various components that a ect those issues . McCarthy et al. proposed a novel way to derive, predict and validate the variance of CLV using a combi-centricnation of stochastic models .

<Page:11>  The type of customer base we consider in this study is contractual and con-tinuous and the relationshipisfurthermoreviewed as ‘lost-for-good’ The most common way to compute CLV is by using variables of Recency-Frequency-Monetary (RFM) variables .

<Page:12>  In the EMP measure,   represents the fraction of customers who accept the retention o er . We use the latter understanding of the parameter   to .derive a distribution of EMP values . Let CLV be a vector of N lifetime values of customers of a given company .

<Page:13>  We compute separate EMP values for each instance in thevectoroftheCLV . Each individual value is not meaningful, since EMPisameasure of the classi er’s performance,buttogainfurtherunderstanding of the EMP values . The distribution of EMP values can be studied using either the maximum likelihood method or the beta distribution .

<Page:14> To obtain a vector of CLV for the customers, we draw a sample of size N. from the distribution 0( , ) distribution . This sample represents the customer base as a whole, not each individual in the dataset . The bootstrap methods can be used to estimate con dence statistics of the EMP vector .

<Page:15> We build churn prediction models following the binary classi ers logistic regression (LR), deci-ariesion trees (DT) and random forests (RF) These classi-referred methods were chosen because of their popularity in both academia and industry .

<Page:16> The Bank and Telco datasets contain rich enough information to estimate CLV and distribution parameters . The Telcodatasets were viewed as the historical information about the customers and used as attributes to predict churn in the last three months . When applicable, models were trained using a 10-foldcrossvalidationonthetrainingsettotuneparameters .

<Page:17> In the case of Telco, the CLV was computed with data from the last three-month contract information from the telecommunication provider . For post-paid contracts, the monthly subscription fee is e15, and includes un-insurednumberoftextmessages and120minutesof phonecalls .

<Page:18> The parameter estimates can be used as a reference by telecommunication providers that wish to evaluate their churn prediction models using EMP . In general, thereislesstra cintheprepaidcasewhich explains the di erence in the estimate for .

<Page:19> Table 5: Comparison of the performance measures. Pro t based model selection with individual CLV . We used the computed vector of CLV to compute EMP and retrieve EMP and its mean and median value, as seen in the fth and sixth columns of the table. Thevariousperformancemeasuresintable5donotagreeonthebestmodel.

<Page:20> Pro t based model selection with individual CLV is based on model selection . Table 6: Comparison of measures when EMP is applied on new datasets . The method can still be applied in cases when CLV cannot be computed, for example when the appropriate data is not available . The highest value for each performance measure within the dataset is under the best one, at least the 95% level .

<Page:21> XGB seems to perform the best overall but the ranking of the methods beyond that is not consistent . The EMP values tend to show very little discrimination, especially in datasets D1,D2,D3andD5 . Figure 3 shows a combination of a box and scatterplot for ve of the six performance measures in table 6 .

<Page:22> The EMP measure provides a way to assess the pro tability of a retention campaign, but with the disad-uvevantage of assuming equal customer lifetime values . Figure 3: Box- and scatterplot showing the correlation among the performance metrics . Managerial implications for businesses that target the most likely churners are an essential part of their operations .

<Page:23> erro

<Page:24> Pro t based model selection with individual CLV overlap with the limits of the LR and DT models . We conclude that RF model performs better than the other two models . Although LR performs badly, the performance is not necessarily di erent from the other three models .

<Page:25> Figure 4: Con dence intervals for median 0 together with model

performance metrics measured in AUC (stars) and top decile lift (triangles) for data set D5.6 Conclusion:
Measuringtheperformanceofcustomerchurnpredictionmodelsisanimportant task .
<Page:26>  An extension of measuring EMP can be used to distinguish actual separation in perfor-orative separation between two models . It can aid in selecting the best performing model for deployment in customer retention campaigns . By taking into account the variability in CLV, it focusesontheheterogeneity of customersasiscompliant with modern businessanalyt-centricics .
<Page:27>  The datasets do unfortunately not contain ground truth about the estimates, it is di cult to estimate their accuracy . The addition of such information wouldbeaninterestingextensionofthisresearchandprovidevalu-
able insights to the model selection process . Pro t based model selection with individual CLV models .
<Page:28>  Pro-pro t based model selection with individual CLVsures," European Journal of Operational Research, vol. 238, no. 2, pp. 505-513, 2014. Pro-Pro-pro-propro-ejective-based feature selection and svm classi-classi-cation in credit scoring,' Deci-Review: Available at SSRN 2739475 .
<Page:29>  Pro-pro t based model selection with individual CLV model selection . Pro-Pro-pro-proper analysis of customer lifetime value models. Pro-performable models for customer-base analy-driven models. A. Fader and G. Hardie, "Probability models for customers-base analysis." Journal of interactive marketing, vol.16, no.2, 2002. P.H.Davenport, J.Harris, and J.Shapiro, 'Competingontalentanalytics.'

Your max_length is set to 142, but you input_length is only 30. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:30>  Pro t based model selection with individual CLV models . B. Jackson, B. Efron and R. J. Tibshirani, C. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services .
<Page:31>  Pro t based model selection with individual CLVsestraat 69, 3000 Leuven, Belgium.31.31, is a model of a car with a price tag of up to $1,800 . The model is based on the size of the car, with prices starting at around $2,800 a year .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11411/Zhu et al_2018_Benchmarking sampling techniques for imbalance learning in churn prediction.pdf
<Page:1>  Many data-level sampling solutions have been developed to deal with this issue . In this paper, we comprehensively compare the performance of several state-of-the-art sampling techniques in the context of churn prediction . The impact of sampling methodsdependsontheusedevaluationmetric and the impactpatternisinterrelated with the classi ers .
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6>  Figure 1 presents the dynamical process of customer churn and retention . In the retention process, a fraction aof the current customers with

the highest churn propensities are targeted and offered some incentive . There are true and false would-be churners within the target customers . Figure 1: The dynamic process of launching a retention program is in uenced by the customers' churn propensity .
<Page:7> erro
<Page:8> Features with highest Fisher scores on the four above mentioned data sets are re-tained and thenumber ofvariablesinthesedatasets is reduced to 30. jx¯ (cid:0) jFisherScore= √c nc (3)
<Page:9> erro
<Page:10> erro
<Page:11> We used the Holm's post-hoc procedure to further explore the statistical difference . When MP is used, the LB strategy signif-uablyicantly outperforms the other two strategies . When AUC and top-decile are considered, LB sampling outperformsthe PBstrategy .
<Page:12> The non-sampling strategy achieves the best results with both AUC and top-decile lift . SMOTE-ENN may haveasmallimprovementwhentheMPmeasureisconsidered. The results indicate that sampling.121212 is not the best method for the top ranking methods .
<Page:13> We further investigate the results of each individual measure . ROS is the best performing sampling technique when using AUC . For top-decile lift, SMOTE-Tomek and ROS are the top two methods . Random under-sampling methods show their superiority when using the MP measure .
<Page:14> An important novelty of our paper is that we introduce the MP measure into our experimental comparison . This metric gives us signi cant rankings of sampling methods . C4.5 decision tree presents totally different reaction patterns with the three evaluations measures . Mostsamplingtechniquesimprovetheperformancemeasured by AUCandRUSisthe best optioninthissituation .
<Page:15> erro
<Page:16> erro
<Page:17> erro
<Page:18> erro
<Page:19> erro
<Page:20> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11415/Garcia et al_2017_Intelligent data analysis approaches to churn as a business problem.pdf
<Page:1> The use of intelligent data-based analysis,ordatamining,fortheanalysis of marketsurveyed information can be of greatassistance to churn management . This is preceded by an in-depthdiscussion of churn within the context of customer continuity management . The research was partially supported by SpanishMINECOTIN2012-31377 research project .
<Page:2> DavidL.Garcia: Churn management may be a competitive advantage in some markets . He offers a review of churn analysis applications of IDA techniques reported in recent academic literature . He also provides an overview of the four stages of the mining process for building predictive models of abandonment . The management of the management of customer loyalty and customer continuity management is discussed .

<Page:3> Many companies shift target from capture of new customers to the preservation of existing ones . This struggle for achieving customer loyalty collides with the grinding exposure to competitors . Customers' market awareness is constantly on the increase and, as a result, so are their expectations . The creationofloyaltybondsincustomers requiresasystematicapproach toitsmanagement .
<Page:4> erro
<Page:5> It is not possible, given the high cost involved, to ask all customers for their opinions on the service they are being o ered and/or their level of bonding . Companies must have a reliable prediction model (adapted to market research and based on behavioural information gathered by the company) that allows them to identify -with enough anticipation- those clientsthat showsymptoms of propensity to churn .
<Page:6> The designand development of a predictive model of supplier abandonment (customer abandoning the supplier) will then be summarily organized in Section 3.2 and the Appendix in the form of tables created according to two main grouping criteria . The last three stages of this process form a cycle that is completed only if and when adequate prediction results are achieved . We will now take a closer look at each of these stages in turn .
<Page:7> Decision trees (DT), regression analysis and neural networks (ANN) are the most commonly used modelling techniques used in the area of abandonment prediction . In more recent years new methods such as support vector machines (SVM) have proven their adequacy [17, 13, 106] The next stage of predictive model development involves the choice of the most suitable methods and techniques for building such model .
<Page:8> An important factor when considering the practical use of ANN is that they do not necessarily uncover patterns in an easily understandable and interpretable form . Support Vector Machines: This ML method, based on statistical learning theory, is able to optimally separate two class of objects (e.g., churners and retained customers) through the generation of a multivariate maximally separating hyperplane . SVMs have been widely used in recent studies due to a lower number of controlling parameters and good generalization capability [12, 40]
<Page:9> We have compiled the reviewed literature in a number of detailed summary tables: 2 to 14 . They list the main references in recent literature (roughly over the last 15 years) and including mainly peer-reviewed journal publications . The main criteria are organized according to the main criteria: According to the predictive methods used: 2, 3 and 4 for standard techniques used; 5, 4, 7 and 8 for alternative ones; 9 and 11 (banking, telecommunications, and other areas of application)
<Page:10> A selection of di erent data requirements and motifs for the analysis of churn can be drawn from recent literature . Data on customer usage have also been used to identify the behaviour of website-using customers [46] and to predict repeat purchasing by mail [94]
<Page:11> erro
<Page:12> The investigation of stage 2 of the prediction model building process provides a mixed picture in the reviewed literature . Many studies do not even con-ishly consider attribute selection, while others justify the selection not on the quantitatively demonstrated impact of the data on the prediction, but,

instead, on domain knowledge .
<Page:13> The most recent publications of churn analysis in banking seem to be paying adequate attention to this issue . There are still too many papers with either no attribute selection at all or a selection-based only on expert and domain knowledge . The majority of the reviewed publications have used quantitativequantitativeselectionmethods .
<Page:14> Churn Analysis Modelling and Prediction (CHAMP) also uses DTs to predict customer churn in the telecommunications industry . Logistic regression and DTs were later used in [76] to forecast credit card customer defection, reporting a better performance of DTs . DTs have also been successfully applied in recent years to problems such as email users and broadband internet users churn .
<Page:15> Standard methods are, by far, the most popular ones in recent literature: DTs and Regression Analysis . Computational Intelligence methods are also reasonably well accepted but still a novelty despite their long-standing record in many other application elds, including business [96, 64].
<Page:16> Random Forest (RF):ItisacombinationofBagging[9], RandomSubspaceMethod[37]and CART DTs [5] RFs solve instability that hampers the use of DTs . Random Forest outperformed SVMs and logistic regression, and its usefulness was recently con rmed when predicting abandonment in the online gaming industry .
<Page:17> Many studies do not seem to include any form of validation explicitly, something that should be compulsory if we aim to assess the ability of the model to generalize its results with unseen data . The use of cross-validation is the most favoured strategy, but many studies only report a single training/test data split, which is a sub-optimal validation strategy .
<Page:18> Thispaperhassurveyed recent literature in which the use of IDA has been proposed for the problem of churn analysis . It is clear that no particular IDA method has the upper hand in terms of results, which means that the choice of method is very problem-dependent . Thereviewedstudies are scattered in a large number of international journals (over 20)
<Page:19> erro
<Page:20> erro
<Page:21> Kim K, Jun C, Lee J (2014) Improved churn prediction in telecommunication industry by analyzing a large network . KimMK, ParkMC, JeongDH (2004) Thee ectsofcustomersatisfactionandswitchingbarrier oncustomerloyaltyin Korean mobile telecommunication .
<Page:22> erro
<Page:23> VellidoA,LisboaPJ,VaughanJ(1999) Neuralnetworksin business:asurveyofapplications(1992-1998). Expert Systems with Applications 17(1):51-70) VerbekeW,DejaegerK,MartensD,HurJ,BaesensB(2011) Newinsightsintochurnpredic-turned-scientists: Apro tdrivendataminingapproach .

Your min_length is set to 56, but you input_length is only 23. You might consider decreasing min_length manually, e.g. summarizer('…', min_length=10)
Your max_length is set to 142, but you input_length is only 23. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:24> The gure shows the generated value -ordinate axis- of three

illustrative customer pro les -gold, silver, bronze- during their time of relationship with the company . It also shows the stages of customer-company interactions and the basic commercial aspects to solve in each one of the stages .

Your min_length is set to 56, but you input_length is only 17. You might consider decreasing min_length manually, e.g. summarizer('…', min_length=10)
Your max_length is set to 142, but you input_length is only 17. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:25>  Figure 1. IntelligentDataAnalysis.ApproachestoChurnasaBusinessProblem:aSurvey 25. Figure 1 . Figure 1 is a survey of 25,000 people who participated in a survey . Figure 25 is a result of an analysis of the data collected by IntelligentData Analysis.

Your min_length is set to 56, but you input_length is only 23. You might consider decreasing min_length manually, e.g. summarizer('…', min_length=10)
Your max_length is set to 142, but you input_length is only 23. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:26>  Figure 1: DavidL.Garc´ıaetal. 26: DavidGarc GarcâÂÂÂ: Figure 2: Figure 3: Figure 4: Figure 5: Figure 8: Figure 1. Figure 4. Figure 5. Figure 8. Figure 10: Figure 6: Figure 7: Figure 9: Figure 10 . Figure 10. Figure 11: Figure 13: Figure 12: Figure 14: Figure 11. Figure 13. Figure 14 .

Your min_length is set to 56, but you input_length is only 17. You might consider decreasing min_length manually, e.g. summarizer('…', min_length=10)
Your max_length is set to 142, but you input_length is only 17. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:27>  Figure 3. IntelligentDataAnalysis.ApproachestoChurnasaBusinessProblem:aSurvey 27. Figure 3 . Figure 3 is a survey of 27,000 people who participated in a survey . Figure 4 is a result of an analysis of the data collected by IntelligentData Analysis.

Your min_length is set to 56, but you input_length is only 23. You might consider decreasing min_length manually, e.g. summarizer('…', min_length=10)
Your max_length is set to 142, but you input_length is only 23. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:28>  Figure 4. 28 DavidL.Garc´ıaetal.Figure 4. Figure 4 . Figure 4: Figure 4 is a representation of David Garcia . David Garcia is a member of the Garcia family . David García is a former member of García's family of the García family .

Your min_length is set to 56, but you input_length is only 17. You might consider decreasing min_length manually, e.g. summarizer('…', min_length=10)
Your max_length is set to 142, but you input_length is only 17. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:29>  Figure 5. IntelligentDataAnalysis.ApproachestoChurnasaBusinessProblem:aSurvey 29. Figure 5 . Figure 5 is a survey of 29 people who participated in the survey . Figure 4.

Figure 4 is a result of a survey by IntelligentData Analysis. Approarliest Data Analysis.
<Page:30> Figure 6: 30 DavidL.Garc´ıaetal. 30: Figure 6. Figure 7: Figure 8: David Garcia, 30: David García, 30. Figure 8. Figure 9: David Garcia, 30, David Garcia . Figure 10: Figure 7 . Figure 8 . Figure 9 : David Garcia. 30 David.
<Page:31> Table 1 is structured by application area and then by category of predictive model . References are listed in the right-handcolumn . The "mixed" category includes studies in which methods from di erent categories have been used . Table 1 includes Tables 2 to 14, located in the Appendix section .
<Page:32> erro
<Page:33> erro
<Page:34> 34 DavidL.Garc´ıaetal. 3.3.3-esque(rejected) standardmeth Validationmethod Accuracy,sensitivity,specificity and AUC. Accuracy,AUC,Precision,Recall,MeanAbsoluteErrorandcomputingtime. accuracy,precision,recall,accuracyandF-measure .
<Page:35> Validationmethod LiftChartandcross-validation. Accuracyonaseparatevalidationdataset. Leastaverageerror,leastrootmeansquareerrorandaccuracy. NotSpecified. Lift chartandAccuracyvalues. LiftChart andAccuracy.values.
<Page:36> 36 DavidL. Garcia´aetal.Garc´ı aetals.u.g.u: ‘a.u’ (‘g) ‘g’: ‘i’, ‘u.h’; ‘v’ ‘p’. ‘e.u.: ‘I’ve got the data from the BusinessIntelligenceCup,Univ.ofChile. Company database. Datasets from UCIMLRepositoryand theannualDMCup.com .
<Page:37> The IntelligentDataAnalysisApproachestoChurnasaBusinessProblem:aSurvey 37.3.3 (%), aSurvey of 37,000 people . Data from 13consecutivemonths. Notspecified. Data from13consecutives months. Notdeclared 1-yeardata, from September 2004 to August 2005 .
<Page:38> 38 DavidL.Garc´ıaetal.2, trulyof.truly-reformed methods . Validationbasedontherealeffectsoftheimprovedsegmentationonthecustomerbase. AccuracyandAUC,usingseparatevalidationdatasets. Fractionalerror,weightedabsolutevalueandweightedsumofdifferencesontheleadtimes.
<Page:39> The IntelligentDataAnalysisApproachestoChurnasaBusinessProblem:aSurvey 39.g.u.n.i.u’s (39.g) andco Validationmethods. Based on the BusinessIntelligenceCup,Univ.ofChile. Database, availableinpreviousstudies. DatasetfromtheBusinessIntelligence.Cup. Database. Company database. Datatype Purchasebehaviour, seenasasequenceofthehistoricalpurchased products(groupedin9setsof products)
<Page:40> erro
<Page:41> IntelligentDataAnalysisApproachestoChurnasaBusinessProblem:aSurvey 41. p.
"IntelligentDataAnalysis’s. accuracy,precision,recall,accuracyandF-measurewithMonteCarlo-basedcross-validation. Accuracy. accuracy .
<Page:42> erro
<Page:43> AUC,AUC,TopDecileLiftandLiftIndex, TopDecile Lift and 10-foldcross-validation. Validationfocussedonthegoodnessofthesegmentation. Accuracy andAUC .

Validationbasedonimprovementofcustomerse segmentation.
<Page:44>  44 DavidL.Garc´ıaetal: Theoreticalstudy.
Fractionalerror,weightedabsolutevalandweightedsumofdifferencesontheleatimes.
Accuracy,AUC,TopDecileLiftandLiftIndex .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11432/Ballings Van den Poel_2012_Customer event history for churn prediction.pdf
<Page:1>  Study: How long should the length of customer event history for churn prediction be for customer churn prediction? The key question of this study is how long is long enough? Time window optimization with respect to predictive performance . The practical implication is that analysts can substantially decrease data-related burdens, such as data storage, preparation and analysis .
<Page:2>  From an algorithm-based perspective, CRM has evolved from using RFM (Recency, Frequency and Monetary) models, to more recent advanced techniques such as random forests (Larivière & Van den Poel, 2005), neural networks (Zahavi & Levin, 1997) and support vector machines (Shin & Cho, 2006) The data and algorithm dimensions are very important, but they constitute only two of three components of the modeling problem . The third component, the time window, remains thus far under-researched (see fig. 1)
<Page:3>  We provide a formalization of different types of windows . A window type can be defined by the variability of the length of the window: constant (C) or variable (V) Given the three time windows in one configuration, the predictors period, the operational period and the response period, there are theoretically 8 possible configurations .
<Page:4>  We analyze the entire customer database of a newspaper company . The churn rate for the estimation and validation is respectively 11.15% and 11.47% . Churn prediction involves predicting whether  the customer will or not renew his or her subscription in the four-week period following the end of the subscription . The performance of the two techniques is very similar and depends on a multitude of factors such as the normality of the data .
<Page:5>  Logistic regression and classification trees are widely used by practitioners and academics . Bagging is a simple approach to increase the predictive performance of classification techniques . We used CHAID (Chi-squared automatic interaction detection) and not CART (Breiman,  Friedman, Olshen & Stone, 1984) as the method of classication tree construction .
<Page:6>  Frequent and heavier buyers are more likely to display loyal behavior . The more money a customer spends with the company, the higher the repurchase likelihood . In addition to RFM variables, length of relationship (LOR) also is a top predictor . Table 1 provides an overview of the included variables in this study .
<Page:7>  We use AUC instead of accuracy (Percentage of correctly classified, PCC) because AUC is not sensitive to the cut-off value of the ‘a posteriori’ probabilities . Figure 3 shows the predictive performance, in terms of AUC, across the different lengths of the  customer event history .
<Page:8>  All three classifiers show a logarithmic increase in performance when the length of the independent period increases . After the fifth year, the increase in predictive performance seems to level off for two  out of three classifiers (trees  and trees + bagging) The difference with the minimum length

of one year is not improving . The third dimension of predictive modeling, the time window, remains under researched .

<Page:9>  In a world where data is growing at an exponential rate, companies are especially looking for efficiency . We have used logistic regression, classification trees and classification trees to study relation between length of customer event history and classification performance . We conclude that the length of the predictors period is logarithmically related to  classification performance .

<Page:10>  Researchers from Van den Poel and Baecke (P., P., D. P., P. & D. B. A. Bhattacharya (1998) When customers are members: Customer retention in paid membership  contextss. (1998). When customers were members, when customers were paying members, they were not paying members .

<Page:11>  Researchers from the Academy of Marketing Science and the University of Stanford University discuss big data data mining . They use Random Forests and Regression Forests techniques to predict customer loyalty and customer retention . The study was published in the Journal of Marketing Research, 17(2), 212-220 .

<Page:12>  The case against accuracy estimation for comparing uctiveinduction algorithms is discussed in Proc. Morgan Kaufman, San Francisco, CA . The case is being discussed at the 15th International Conference on Machine.Learning.  In: J. J. Shavlik (Ed.),  Proc.  of  15th  International Conference  on Machine Researchers.  on  Machine

/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11440/Ballings et al_2012_Improving Customer Churn Prediction by Data Augmentation Using Pictorial.pdf

<Page:1>  The purpose of this paper is to determine the added value of pictorial stimulus-choice data in customer churn prediction . The practical implication is that companies should start mining pictorial data from social media sites (e.g. Pinterest) or use pictures instead of verbal scales in surveys .

<Page:2>  Even small increases in retention can have substantial impact on a company's results (Gupta,  Lehmann, &  Stuart, 2004) Even an increase of 1% in retention in a company can have a dramatic influence on contributions . The internal transactional database is considered the baseline for database enhancement because it contains the  top predictors in extant database marketing modeling: recency, frequency and  monetary value .

<Page:3>  Pictures contain a massive amount of information and user actions could potentially reveal part of that information . Social media platforms and the rise of picture centered social networks (e.g., Pinterest) new opportunities for data augmentation are created in the realm of pic-torial content . The study aims to assess the added value of picorial stimulus choice data to customer churn prediction over and above traditional customer data .

<Page:4>  Traditional surveys are too long asking the re-spondent too much of his/her time . Social barriers, language barriers and memory-based barriers can be overcome . Pictures alleviate these problems in that they function as an enabling technique . Picture association is a task in which the respondent is asked to choose a picture or image out of multiple possibilities as re-request to a specific question .

<Page:5>  An alternative response method (pictures instead of scales) can be

conceived of as being the source of multiple advantages and would there-ishlyfore be an attractive means of data augmentation . The study aims to fill this gap in literature by evaluating the added value of  pictorial stimulus-choice data in customer intelligence over and above the tradi-centric data sources discussed above .

<Page:6>  Table 2 displays the sample characteristics (5 times 2 fold cross-validation) Table 2: Sample characteristics . Figure 1 displays the time-window to analyze churners in order to discriminate churners from non-churners . We were able to assess the added value of pictorial stimulus-choice data . The model is built in the first step, both the independent and dependent period can be shifted forward .

<Page:7>  Administrative data represents all information regarding agreements made be-tween the customer and the company at the time of the purchase decision . Operational data holds the entire customer history (at the subscriber's level) Data also contains socio-demographic data, and data about suspensions, forward interruptions, credit handling and marketing actions .

<Page:8>  Complaints data contains information about the number and topic of complaints and the solution and answer given . Survey data can be conceived of as resulting from a company-initiated feedback process . Mindset variables (e.g., purchase intentions, commitment, product rec-ommendations) are impossible to collect from internal processes .

<Page:9>  We used Random Forest (Breiman, 2001) to create the churn model because of  multiple reasons . We opted for an ad-hoc approach here (an online survey) while companies could go online and mine social network data . Figure 2 represents an example such as the picture sets we used (a woman's facial expressions)

<Page:10>  AUC is argued to be an objective criterion for classifier performance by several authors . We use AUC instead of accuracy (Percentage of correctly classified, PCC) because AUC,  in contrast to PCC, is the insensitive to the cut-off value of the 'a posteriori' prob-abilities . AUC ranges from .5, if the predictions are not better than random, to 1, if the model predicts the behavior perfectly .

<Page:11>  The Wilcoxon signed ranks test (Wilcoxon, 1945) ranks, per data set, the differences in performance of two classifiers . We follow the recommendation of (Demšar, 2006) to use the Wilcox signed rank test . Figure 3 presents the added value per data type . Results indicate that the addi-generation of the operational data accounts for the biggest increase in AUC .

<Page:12>  When only administrative data is analyzed pictorials add 0.9% to the predictive performance . When all data are modeled pictorial stimulus- choice data adds 0.1% . The same conclusion can be drawn for models based on administrative, operational, complaints and survey da-centricta ($p > .10$)

<Page:13>  A picture contains a huge amount of information, user actions (e.g. 'liking' a pic-insuredture on a social media platform) could possibly say a great deal about future behavior . Future research could use a structural approach by mining online social networks .

<Page:14>  It might prove valuable to adapt the question to gauge a more proxi-centricmal attitude to behavior (e.g., Which picture best represents your commitment to  grotesquely staying with the newspaper? The Belgian government

grant No. P7/06 of the Belgian government (Belgian Sci-Review Policy)
<Page:15> Van den Poel, D. D. and J. Vanthienen (2012) use kinship network information to improve retention of customers . Vanden: Churn prediction in subscription services: An application of support vector machines while comparing two parame-privatter-selection techniques . Vanne: "How long is long enough?"
<Page:16> Researchers have found a way to reduce social bias in marketing . The study was published in the journal Journal of Marketing Research, 47, 1, 14-27 . The author has published a number of articles on the topic of marketing and customer satisfaction in the U.S. and Europe .
<Page:17> A.Gounaris, SP. (2005). Trust and commitment influences on customer retention: Journal of Business Research. 58 (2) 126-140. (2009). Understanding the psychological process underlying customer satisfaction and retention in a relational service, Journal of
(2009)
<Page:18> The case of financial services is the result of using survival analysis and choice modeling to predict customer retention and prof-ishlyitability . The C-OAR-SE procedure for scale development in marketing has been used in marketing . The case against accuracy estima-heticaltion for comparing induction algorithms is discussed in the ICML-1998 .
<Page:19> Researchers from Van den Poel and Thorleuchter predict e-commerce company's success by mining the text of its publicly-accessible website . They also use a customer attrition analysis for financial services using proportional hazard models . The research was published in the journal International Journal of Production Economics .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11453/Burez Van den Poel_2009_Handling class imbalance in customer churn prediction.pdf
<Page:1> The author of this article is Dirk Van den Poel, Professor of Marketing Modeling / Analytical Customer Relationship at Ghent University, Faculty of Economics and Business Administration, Department of Economics . The author is a researcher at the University of Tweekerkenstraat 2, B-9000 Gent, Belgium .
<Page:2> Customer churn is often a rare event in service industries, but of great interest and great value . Until recently, class imbalance has not received much attention in the context of data mining (Weiss, 2004) In this study, we investigate how we can better handle class imbalance in churn prediction . Using more appropriate evaluation metrics (AUC, lift), we investigated the increase in performance of sampling and two specific modelling techniques .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11456/Burez Vandenpoel_2008_Separating financial from commercial customer churn.pdf
<Page:1> A modeling step towards resolving the conflict between Sales And Credit departments . Separating Financial From Commercial Customer Churn: a Modeling Step Towards Resolving The Conflict . The paper was published in the journal of Marketing Modeling, Analytical Customer Relationship Management, Ghent University .
<Page:2> The paper shows that the two different processes mentioned can be

separated by using information from the internal database of the company . Previous bad-payment behavior is more important as a driver for financial churn than for commercial churn . The impact of ‘loyalty’ actions is far greater with potential commercial churners as compared to financial churners .

<Page:3>  General Motors led the field during the 1920s with their new focus on advertising, installment financing, and the art of styling low-priced automobiles . GM faced tension to the extent that one party’s profit came at the expense of another party . Subsequently, we will build a case about churn definitions, and hypothesize on churn definitions . After a methodological part, the case study at the pay-TV company is described, and results are reported .

<Page:4>  A subscription renewal decision is a type of repeat buying . CRM in subscription services comes down to "attracting new customers" and "keeping  the existing customers’’, known as defensive marketing . The marketing department tries to convince as many customers to renew their contract as possible .

<Page:5>  Based on the way a customer terminates a contract, we will show that there exist three different types of churn: involuntary churn, financial churn and commercial churn . We also include variables often used in credit scoring, but not (often) used in churn prediction . Other variables that can be added to the model include detailed credit bureau reports (e.g., Overstreet and Bradley, 1994)

<Page:6>  Previous bad-payment behavior is more important in the financial churn prediction model than in the commercial churn predictions model . Different types of interventions can have different impacts on customers depending on their characteristics (e.g., De Wulf et al. 2001) Targeting financial defection-prone customers worsens customer retention . Targeting with relationship-oriented marketing interventions will hence improve retention .

<Page:7>  Breiman (2001) introduced a solution to the previously mentioned problem: Random Forests . This technique uses a subset of m randomly chosen predictors to grow each tree on a bootstrap sample of the training data . After a large number of trees is generated, each tree votes for the  most popular class . By aggregating these votes over the different trees, each case is predicted a class label .

<Page:8>  The aim of the definition is to quantify the instantaneous risk that the event will occur at time t . The Kaplan-Meier estimator (also known as the Product Limit Estimator) provides an estimate of the survival function from life-time data (Kaplan and Meier, 1958) In customer churn prediction, one measures the length of  time customers remain with the company .

<Page:9>  At this European pay-TV company, all customers have a 12-month subscription . Cancelling within that period is not allowed, nor is prematurely reporting that a subscription will not be renewed . In almost all cases, this is due to bad payment behavior: a customer cannot or does not want to pay his subscription any more .

<Page:10>  Different subsets of the training set will be used while modeling financial churn . This is done to investigate the effects of leaving some customers out of the data set, on the predictive power of the model . Data is extracted for this study from the data warehouse of the pay-TV company . The static model had over 100,000 observations, of whom we collected 171 independent variables .

<Page:11>  The survival curve of the customers of the pay-TV company (see Figure 3) clearly confirms the subscription renewal process . After one year, a huge drop can be noticed in the survival curve . Commercial churners do not churn during their one-year subscription; they do it at the end of their subscription .

<Page:12>  The biggest part of commercial churn happens after one year of subscription: almost 20% of the customers leave at that point . Financial churners leave at any point in their relationship with the company . The curve decreases less and less, meaning that a customer is more at risk in the first few years of subscription .

<Page:13>  Table 1: AUC performance results (shown in %) from the different models . Financial churn is a lot easier to predict than commercial churn because on average all financial churn models score about 86% on the out-of-period dataset compared to 68% for commercial churn models .

<Page:14>  Table 2 shows the most important variables for the financial churn prediction model . The last three columns represent respectively the z-value, the ranking, and the significance . For those variables, we also include in columns 2-5 the importance of those variables in  commercial churn prediction models . The table clearly shows that we can confirm hypothesis 2 .

<Page:15>  Avant-premiere questionnaire group was targeted with different relationship marketing interventions (RMIs) Commercial churn (not renewing a subscription) and financial churn (no longer paying invoices of a current subscription) were defined as commercial churn . Financial churn could not be reduced; on the contrary, the churn rate slightly increased  compared to the control group .

<Page:16>  Financial churn is easier to predict, commercial churn is much easier to prevent . A field experiment pointed out that you can convince commercially defection-prone customers to stay at your company . Further research could inquire into the cost-effectiveness of different incentives for different types of churn .

<Page:17>  Au, T., Li, S. and Ma, G. Ma (2003) Applying and Evaluating Models to Predict Customer Attrition Using Data Mining Techniques . The author also discusses the role of credit scoring models in improving cash flow and collections . References include: A.u, W., Chan, Chan, K.C. and Yao, X.Y. and Liu .

<Page:18>  Researchers from the Academy of Marketing Science (Academy of Marketing) and the University of Virginia (2005) have published a book on how to measure customer loyalty and customer lifetime value . The book is entitled, "Predicting customer retention and profitability by using random forests" and regression forests techniques .

<Page:19>  The impact of sample bias on consumer credit scoring performance can be seen in the European Journal of Operational Research, 157(1), 196-217. (2005). (2005) (The impact of the sample bias is still unknown) (G.C. Thomas, L.C., Oliver, R.W. & Hand D.J.J.) (2004) (A.Y.M. Therneau, T.M., Grambsch, P.M.) (2000). Modeling Survival Data: Extending the Cox Model, Springer, New York . (2000) A survey of credit and behavioural scoring: forecasting financial risk of lending to . consumers .

<Page:20>  The static churn prediction model is based on the number of

subscriptions the customer has on pay-TV . The number of products the customer now has depends on the type of service the customer uses . Churns are calculated by the product type and the technology used to predict the customer's churn .
<Page:21> Customer status of the customer (0 = Me, Ml, SM; 1 = Ccivil
: , ; : :
, ,
=
January 2001: February 2011: January/2012: February .
<Page:22> Length_subs Continuous Length of relationship at 28/2/2002 . Length of actual subscriptions (lor - gaps) is length of relationship . Markov value of the fourth order (no distinction is made between analogue or digital) Markov2_dummy Continuous for this Markov chain . Noc2 Continuous Absolute number of contracts relative to the lor .
<Page:23> The number of days that elapsed since filling in the questionnaire was the number of generic letters of a certain type received from a certain pay-TV company . In_call_A Continuous Number of times customer called the company, (relative to length of relationship for the relationship for the company) concerning: the last 3,5 years .
<Page:24> After_call_C__ Continuous Continuous "After call work" on the subscriber (types as above, X - else) ‘After’ is ‘after’ on subscriber . ‘Call_Calls_pos_pos’ and ‘Stand_contact_contact’ are ‘stand’s contacts between customer and Stand .
<Page:25> The number of churns in commercial churns is based on a churning model . The churns model has a churn rate of 38 per cent . The model is a churn model with churn rates of 1,000 churns per year . The average churn rate is 1,600 churns and churns are 1,800 churns .
<Page:26> Pind_subs 11,926 51 0 ep_mv 11,406 52 0
ev 11,351 53 0 in_call_D 11,221 54 0 og_6 10,897 56 0 rec_eg 10,877 57 0 noc 7,993 73 0
canc_78P 7,875 74 0 , , ;
; ;; candidate candidates were selected for categorized with qualities
<Page:27> The average number of objects is 4,268,000, with a score of 5,268 and 4,257,000 . The average value of a person is 1,000 points . The number of items in the table was 1,100 . The value of an object is 1-1,000. The average amount of objects in the list is 1.5 points .
<Page:28> The churn model is based on a churning model with a churn rate of 1% . The churn rate rate is 1% higher than the average churn rate in the industry . The rate rate rate of churning is 1 per cent . The number of churns in the churning process is 1,000,000. The churning rate rate has been set at 1,500. The rate is 2,000 per churn rate .
<Page:29> The study was designed to measure the importance of a person's presence in a state . The study is divided into categories of importance and importance . The average number of words used in the study is 9,571 and 9,000 words per word . The score is 1-1 and the average number is 1,600 words per person .
<Page:30> The average number of words is 4,069-111 . The average value of a word is 1,000. The average word was 1,500 . The number of items in the table is

1.1,000 . The value of each item is based on a value of 1,100. The number is 1-100 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11459/Burez Van den Poel_2007_CRM at a pay-TV company.pdf
<Page:1>  At a Pay-TV Company: UsinAg Analytical Models to Reduce  ‘Customer Attrition by Targetedw  M0arketing for Subscription Services . M’ s Marketing, Ghent University, Hoveniersberg 24, B-9000 Gent, Belgium .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11461/Wai-Ho Au et al_2003_A novel evolutionary data mining algorithm with applications to churn prediction.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6>  AUetal.:NOVELEVOLUTIONARYDATAMININGALGORITHM 537: 537. Thereproducefunction. The genetic operators used by DMEL are imple- (b) progressivelymented in the reproduce function shown in Fig. 4 . The crossover-1 operator allows the crossover points to occur between two rules only .
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
<Page:11>  Liftcurves for DMEL, C4.5,andneuralnetworkunderdifferentmonthlychurnratesaveragedovertenruns . TableV shows that it is difficult, if not impossible, to decode execution times for the DMEL and neural networks under different monthly churn rates .
<Page:12>  The experimental results showed that a subscriber churns if he/she lives in the data mining task faster than neural networks . Of the three KualaLumpur,isofagebetween36and44,and paidbillsusingapproaches,C4.5 required theleastexecutiontimetocomplete cash with weight of evidence of 1.20. The domain expert found this rule useful because it helps retain subscribers .
<Page:13> erro
<Page:14> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11467/De Bock Van den Poel_2010_Ensembles of Probability Estimation Trees for Customer Churn Prediction.pdf
<Page:1>  The potential of using probability estimation trees (PETs) instead of standard decision trees as base classifiers is investigated . The effect of the proposed strategies heavily depends on the chosen ensemble algorithm in which they are implemented . The results demonstrate the value of  using PETs over standard decision Trees in order to increase lift .
<Page:2>  An effective Customer Relationship Management strategy is of the  most important aspect of CRM is customer retention . In churn prediction, information from customers that is available in the company database is used to determine their proneness to leave the company . Once built, these models can be used to predict the future behavior of customers and to target targeting information for churn-preventing marketing campaigns .

<Page:3> erro
<Page:4>  Lift focuses on the segment of customers with the highest risk to the company . The definition of lift depends upon the percentage of riskiest customers one is considering for a retention campaign . Churn data sets are typically characterized by high dimensionality, both in terms of number of features and number of instances . Another issue is the class imbalance of the data .
<Page:5> erro
<Page:6> erro

Your max_length is set to 142, but you input_length is only 41. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:7> erro
<Page:8>  The authors thank the reviewers for reviewing the paper and Ghent University for funding the PhD project of Koen W. De Bock . Acknowledgements are made in this article . The authors are happy to provide an overview of the work done in the book . The author and the author of the book are open to the public .
<Page:9>  Using analytical models to reduce  customer attrition by targeted marketing for subscription services, CRM at a pay-TV company: using analytical models . Using AdaBoost in customer churn prediction, AdaBoost could reduce attrition by targeting targeted marketing . The study was published in the journal of the International Conference on Service Systems and Service Management .
<Page:10>  Researchers: Learning probabilistic decision trees for AUC . They also use random forests to test accuracy of PETs on imbalanced datasets when training and testing . Researchers: Bagging, boosting, and variants. An empirical comparison of voting classification algorithms. A. Bauer, E., Kohavi, R., Bauer, R.E.: A decision-theoretic generalization of on-line learning and an gorithmic generalization .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11477/Antipov Pokryshevskaya_2010_Applying CHAID for logistic regression diagnostics and classification accuracy.pdf
<Page:1>  A CHAID-based approach to detecting classi  cation accuracy heterogeneity across segments of observations is proposed . The approach was applied to churn data from the UCI Repository of Machine  Researchers develop accuracy across segments . Different segments may have absolutely different churn  predictors .
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5>  The logistic regressions built on three segments  revealed with the help of the CHAID decision tree are presented in Table 3 . Table 2: Parameter estimates of model 1 and 2 models for four large segments of data are presented . Table 3: The parameter estimates for Model 2 and Model 1 are presented here .
<Page:6> erro
<Page:7>  CHAID decision tree: Accuracy of Model 2 (training sample) Churned: 1 681   36   1101   32    D  id not churn: D id did not churn . Table 4: Predicted

category: Churned  223   60, 60   168, 168, 32, and D id didn't churn .
<Page:8> erro
<Page:9> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11479/Al-Molhem et al_2019_Social network analysis in Telecom data.pdf
<Page:1>  Many systems can be represented as networks or graph collections of nodes joined by graph theory . Social structures in these networks can be investigated using graph theory through social network analysis (SNA) In this paper, networks and SNA concepts were applied using Telecom data such as call detail records (CDRs) and cus-tomers data to model our social network and to construct a weighed graph .
<Page:2>  In Telecom domain influence sub-scribers are usually considered to be well-connected to other subscribers in network . This good connection guarantees low risk for churn but high potential for diffusion of products and services [6] Finding the actual influencers in Telecom field is a chal-ophobiclenging topic and it depends largely on intuition .
<Page:3>  Al-Molhemet et al. J Big Data    (2019) 6:99:99 . We used k-shell values to calculate the influence score for each node in the network . The most influential nodes are the highest in score . Multi-SIM subscribers with different operators have a higher potentiality to churn than ordinary subscribers . Detecting the multi-SIMs across different operators allows for even more usage profiling .
<Page:4>  Onnela et al. [12] analyzed a weighted call graph  by examining its degree, strength, weight distributions,                  clustering and weighted clustering, together with correlations between these quan-centricities . Nanavati et al. [13] analyzed the graph properties such degree distribution and neighborhood distribution over time of calls and SMS networks . They identified the most influential customers who can  spread positive or negative messages through the network using PageRank algorithm .
<Page:5>  The paper proposes a novel approach to detect influence subscribers in the Telecom social network . The new approach is more accurate and efficient than traditional methods that using only centrality measures . We have chosen Hortonworks Data Platform (HDP)3 as a big data platform to install and  use in the study .
<Page:6>  Al-Molhemet al. al. J Big Data: The tel-profit social network was built with 10 million customers with their data and about one billion records of calls between customers . We used Spark tools for processing data, building the tel-repreecom social network and calculating SNA features . We stored Data in HDFS as a spark DataFrame9 format which is a Dataset organized into columns .
<Page:7>  Al-Molhemet al. J Big Data    (2019) 6:99  Page 7 of 17 has been published in the journal J Big Book, Big Data . We used detailed data to build the social network for 3 months . Figure 2 visualizes a sample of our social network where size and color of nodes express ranking degrees and lines between nodes express rankings .
<Page:8>  Table 4 shows sample of Telecom social network data . We analyzed our social network and calculated centrality measures for each node . In our network, calls duration was considered a little bit more important than calls

number so we selected  0.6 and the calculated ogleweight must be 0 <1 .
<Page:9>  The calculated SNA features were used to enhance the churn prediction models that used in the Telecom company by adding social network features on top of the traditional churn predictors . The calculated measures are normalized by dividing to the max value of each measure over all the graph .
<Page:10>  Al-Molhemet et al.Wang et.al. [18] proposed an Influence Capability measure based on k-shell values and the iteration information in the decomposition process to distinguish nodes with the geysame ks values . Table 6 shows a sample of calculated EV and IC measures for all nodes in network and normalized values .
<Page:11>  The next step in our Multi-SIM subscribers' model was calculating two types of SNA similarity measures and SNA behavioral measures for each pair of nodes . Table 7 shows a sample of calculated similarity SNA measures . The similarity score plays a main role to detect pairs that have high prob-generationability to be similar and exclude ones with low probability .
<Page:12>  Al-Molhemet et al. al. J Big Data (2019) 6:99:99  Page 12 of 17 of 17. The solution was designed to deliver high performance and speed, especially with ETL activities and SNA operations . The HDP framework was installed and customized with a variety of systems and tools such as Hadoop, Spark, Yarn and Zeppelin .
<Page:13>  Spark is very useful for ETL processing and analytics because of its ability to per-form calculations in-memory . Spark abilities were used to build the social network of 3 months CDRs provided by the Telecom company . Figure 4 presents the fre-uvequency distribution of in-degree, out-degree and degree, where frequency distribution is the fraction of nodes in the network with different types of degrees .
<Page:14>  The results of Multi-SIM detection model contained more than 1.5 million records . The model can be tested in two ways: first by making direct calls to subscrib-rousers in previous groups with a questionnaire about the number of lines that subscriber has . Al-Molhemet al. J Big Data .
<Page:15>  Traditional methods used in detecting Multi-SIM subscribers were able to detect subscribers only in the same opera-tor based on customer data . Traditional methods had a success rate between 30 and 40% . By using our multi-SIM detection model we have achieved a better results within the same operator and across different operators .
<Page:16> erro
<Page:17> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11487/Gok et al_2015_A Case Study for the Churn Prediction in Turksat Internet Service Subscription.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4>  The raw data was obtained from the  Postgres database, which is used by the corporation . Cluster size is determined by trial and error method by using various cluster sizes . The performance of the two phased solution models has been given in Table 2 . The k-means algorithm is 2 percent more accurate than a hierarchical clustering algorithm .

<Page:5> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11488/Azeem et al_2017_A churn prediction model for prepaid customers in telecom using fuzzy.pdf
<Page:1>  A.C.M.Fong, MuhammadUsman, A.Azeem and A.ShaheedZulchurners . Churnprediction has been compared with fuzzy classi ers to highlight the ability of a model to correctly classify the percentage of churners as part of CRM systems .
<Page:2> erro
<Page:3> erro
<Page:4>  Ensemble do not give signi cantoptimallysampledandreducedfeaturesdataset improvement . In ensemble clas-Predictive performance of Naïve Bayesian . Bayesian Bayesian classi-cation si  ers,Stacking(SVM)gavebetterperformanceintermsof the performance of feature selection .
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9>  Achurnpredictionmodelforprepaidcustomersintelecomusingfuzzyclassi ers 611.11 metrics . Training and testingratioforeachdatasethasbeenconsis-tent through implementation and that is 80:20 respectively .
<Page:10> erro
<Page:11> erro
<Page:12>  A.C.M Fong is an associate professor in the Department of ACMSIGMOD International Conferenceonmanagementofdata of Computer Science . Muhammad Usman has a PhDinComputer&Infor-Programmed Sciences from Auck-Grabland University of Technology, New Zealand . MuhammadUsman has published in international journals and conferenceproceedings,and reviewed for anumber ofpremier journals and conferences .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11491/Agrawal et al_2018_Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11492/Amornvetchayakul Phumchusri_2020_Customer Churn Prediction for a Software-as-a-Service Inventory Management.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin

al_dataset/files/11493/Benedek et al_2014_The Importance of Social Embeddedness.pdf
<Page:1> The Importance of Social Embeddedness: Churn Models at Mobile Providers is discussed . The article uses data from regional snowball sampling to identify groups with different network topological properties . Clear evidence indicates that individual network characteristics (node-level metrics) have considerable impact on churning .
<Page:2> Mobile providers all over the world face the phenomenon of customer churn . Attracting new customers is more expensive than retaining old ones . Mobile companies must identify their risk customers and target them with marketing campaigns . Churn modelstypically includes personaldetails, typeofcustomerpackage,telephone usage patterns, andinforma-centricmarketing,asexplanatoryvariables .
<Page:3> Models incorporating network measures can identify churning customers with higher precision, allowing mobile providers to gain-and maintain-their competitive edge ef ciently . The organization of the article is as follows. The next section provides theoretical foundations of customer churn and surveys the empirical literature . We describe the sampling method, data, and network characteristics . The last section includes the methodology of formulating de-cision rules based on topological properties .
<Page:4> The importance of social embeddedness and churning has not been systematically investigated yet . The most commonly used data mining techniques for churn prediction are decisiontrees, logisticregressions, andneuralnetworks . The majority of churning prediction models use contractual data (eg., contract type service, type of service, and type of payment type) as predictors .
<Page:5> The authors of this article opted for regional (geographic location-based) snowball sampling . The sampling tech-niques used in onlinesocialcommunitiescaneasilybeadaptedfortelecommuni-cationnetworks . The largest mobile company of a Central European country provided the data, and the algorithm was performed with multiple random seeds .
<Page:6> Table 1 summarizes the effect of marketing on marketing on Core 1 customers . Data set includes callers' and recipients' phone numbers, the month of transaction, the number and duration of calls, and the number of SMS calls . Table 1: Customersinthecoresandontheperiphery.
<Page:7> Asextremeoutliers donotre ecttruetrue socialrelationships, visualizations,descriptivestatistics,andboxplots were used . Table2summarizesthecallandSMSrecords, where records indicate veryweaksocialrelationship .
<Page:8> The average out-degreewas91 contacts, themedianout-deewas61, and themodewas25. The lowest out-degree was 993, with 70 customers phoning only one number (at least twice) within the period under analysis . The highestout-degree of 2007 was 994, with three socially active customers phoned 993 contacts within that period .
<Page:9> This article de nes churn probability as the likelihood of a customer switching providers, and churn ratio as the number of churning customers over the total number of customers . It offers network-topological insights focusing on the question whether customers can be mean-riddenfully segmented into groups, with signi-cantly different churn ratios .

<Page:10>  Figure 3 summarizes the calculations of this network measure . A Core 1 customer makes phone calls to three Core 2 customers (Friend 1, Friend 2, and Friend 3) with proportions of in-network phone calls of 100%, 75%, and 33.3%,respectively .
<Page:11>  The churn risk assessment models offer better, longer-term predictive validity . Call pattern analysis is a particularly popular method of predicting churners within three months . Risk-riskassess-based models may predict churners even a year in advance . Table 3: Churn risk assessment is a six-stage process, it determines the network topologicalpropertiesofcustomersthroughinvestigationofcallandSMSrecords over a 6-month period .
<Page:12>  The churn ratio for customers with seven or more in-network relations is 28.1% . Thislatter gure is76.9% higherthanthechurnratio ofcustomerswithsevenormorein-network-relations (49.7%)
<Page:13>  The most signi cant difference between the two subsegments is observed at a threshold of 12 relations . Figure 4(b) shows that if the in-network degree of a nontargeted customer is less than 12 relations, then their rate of call duration is higher . Table3: Thechurnratios of customers are based on the number of relations (de-gres) of customers .
<Page:14>  The churn probability is signi cantly lower if the proportion of in-networkcalldurationishigherthan31%,fortargetededcustomers, and higherthan.ophobic41% for nontargeted customers . Product customization in the form of directmarketingcampaigns creates value, but there are exceptions to the power and importance of tailored marketing actions .
<Page:15>  The two degrees of separation metric shows counterintuitive results that can be explained by theinsigni canceofindirectin uenceexerted by friends, authors say . The setup of the ‘simulationmodel’mimicked these segmentationexercisedescribed in this article .
<Page:16>  The study used SNA to explain why customers with identical mobile contracts may switch providers . The research relied on real-life call and SMS records of approximately 26,000 customers calling or texting almost 800,000 people withina6-month periods . The studysegmented the customers into two distinct groups with signi cantly different churn ratios .
<Page:17>  Figure 5: Samplerobustness . (a)Simulation results for a randomly selected balanced sample . (c) Comparison with a comparison of the simulation results . (b) The simulations were simulated using a random sample of randomly selected human samples . The results were compared to a simulated simulation of a balanced sample with a random random sample .
<Page:18>  192 SocialEmbeddednessandChurn.comparisons and riskassessmentmod-elsemphasizeddifferences in howfarahead they can predict churn . Customers with the highest number of connections may be exploited as an asset to understand this segment of the mobile company .
<Page:19>  The snowball sampling algorithm resulted in a single component, disregarding the other smaller, connected customers . The number of disconnected customers is fairly low (Onnelaet al., 2007a, 2007b; Dong et al., 2009). The snowball sample algorithm was limited by the scope of the dataset, but our results are promising, but limited .
<Page:20>  SocialEmbeddednessandChurnipientvariables were the number of in-

network relations and weighted embeddedness . This suggests that network variables bring marginal value in improved prediction performance relative to variables already employed in standard churnpredictionmodels . To our knowl-ishlyedge, fromthedataathand,theextentoftheimprovedationallyprediction performance could not be extracted .
<Page:21>  Researchers: How does the data sampling strategy impact the discovery of informationdiffusioninsocialmedia?
DeChoudhury,M.,Lin,Y.R.,Sundaram,H.,Candan,K.S.,Xie,L., &Kelliher,A.A. (2010) How does data sampling . impact the . discovery of the . data . sampling strategy? Proceedingsofthe4thInternational.AAAI Conference on Weblogs and Social Media, Washington, DC. Menlo.
<Page:22>  An approach to correct biases by snowball sampling is based on snowball sampling . Churn model accuracy improved by 47% with In niteInsightTM, available at http://www.kxen.com/ (2012) Churn models improved by . 47% in 2012 with In Injective-social.social.com .
<Page:23>  The impact of prepaid churn prediction for mobile telecommunications: What to setup in prepaid churn predictions for mobile users's choice . The impact is similar to the impact of a prepaid churning model in the U.S. market . The study is published in the New Year's edition of New Year' s Bestseller: The Bestseller, by Edward Elgar .
<Page:24>  The SocialEmbeddednessandChurn and churning analyses have been used to predict customer churn in the telecommunication sector . The social network effect is a result of the network effect . The research is published by the Center for Network Science, Central European Univer-College and Maven Seven Network Research Ltd.
<Page:25>  Network theory has a long tradition in social sciences, especially among sociol-orientedogists, and outside,amongphysiologists . The most popular graph traversal techniques are .breadth- rst search, depth- . search, forest re, and snowball sampling .
<Page:26>  The de nition of snowball sampling varies from study to study . The algorithm requires three basic parameters:
thenumberofverticesselectedasseeds,thenumberofneighborspicked,and the number of iterations . The location-based sampling technique performs reasonably well .
<Page:27>  She is investigating the role of socio-demographic and network topological characteristics of doctors in professional interactions between general practitioners and specialists . Her research areas are financial stability, network theoryineconomics, telecommunication, and networks in healthcaresystems . She received a two-year postdoctoral fellowship from the AXA Research Fund in 2011 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11494/Esteves-Mendes-Moreira_2016_Churn perdiction in the telecom business.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro

/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11495/Farquad et al_2009_Data Mining Using Rules Extracted from SVM.pdf
<Page:1> Data Mining Using Rules Extracted from SVM: An Application to Churn Prediction in Bank Credit Cards . The data set analyzed in this paper is about churn prediction in bank credit cards . The proposed hybrid SVM+NBTree yielded the best classifier compared to other classifiers .
<Page:2> Churn management consists of developing techniques that enable firms to keep their profitable customers and aims at increasing customer loyalty [8] Churn prediction and management is one of the important activities of Customer Relation-ship Management . Rule Extraction from SVM recently became one of most popular classification methods .
<Page:3> The dataset is from a Latin American bank that suffered from an increasing number of churns with respect to their credit card customers . It consists of 14814 records, of 13812 are nonchurners and 1002 are churners . The dataset comprises 22 variables, with 21 predictor variables and 1 class variable . Two groups of variables are available for each customer: sociodemographic and behavioural data, which are described in Table 1 .
<Page:4> The data set used in this study is highly unbalanced but we did not employ any balancing technique to balance the data . The hybrid approach presented here is different from [28, 29] in the following ways: Dealing with unbalanced large scale data set. Using the predictions of support vectors using SVM model i.e. Case-SP to generate rules with NBTree [12].
<Page:5> 70% of the data is then used for 10-Fold Cross Validation (10-FCV) and 30% is named as validation set . The accuracy and validity of the rules are then tested against the validation set. The class distribution in the train-generation and validation data sets is as same as that in the original data i.e. 93.11% for loyal customers and 6.89% for churned customers .
<Page:6> The proposed hybrid SVM+NBTree using Case-P and Case-SP are the best performers compared to other classifiers evaluated in this study . The number of rules extracted using our ap-grotesqueproach i.e. SVM plus NBTree is very much less and rule length is smaller when compared to those of Kumar and Ravi [4]. The number extracted is 67.8% less than the actual num-agyber of training instances .
<Page:7> erro
<Page:8> Researchers use fuzzyARTMAP for churn churn prediction in bank credit cards . They also use fuzzyARMAP for fuzzyARTMap for churning churn predictions in credit card data . Researchers have published numerous papers on the topic of machine learning and pattern recognition algorithms in the past and present at various conferences across the world .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11496/Jafari-Marandi et al_2020_Optimum profit-driven churn decision making.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> The role of a model that includes pro tattheleveldataanalytic hidden

layers is to  nd nonlinear relationships between them is necessary . This paper exclusively (predicting)attributeshastenneuronsinthe inputlayerand (an outputlayer) has inculcated a pro-t drive at the level of ANN learning . For a churn prediction task, thereisonlyoneddependent(targeting)attribute, thereislyonedependent(targeted)attribute .

<Page:6> erro

<Page:7> erro

<Page:8> erro

<Page:9>  NeuralComputingandApplications(2020)32:14929-14962 14937 . The results are based on a map of decisions between RE and SR . The analysis was carried out using an algorithm called an ANH-NNFlowchart foroptimumpro t-drivenSOEDchurnpredictionsystem .

<Page:10>  Figure 4a represents the output from Step-1 of the proce- employed to generate the average monthly revenue of a given cluster . Figure 4b presents the estimated customer revenue for all the clusters with only churn customers and all of the clusters without churn customers . The average customer lifetime value (CLV) is based on certain variables revenues for all clusters .

<Page:11> erro

<Page:12>  The MLP’s prediction for each customer is based on the line adjustment procedure: MOD such that the misclassi cation costs of the classi-cation costs will be minimum . The line adjustment process is called ‘lineformsbysequentiallyconnecting’ to one ‘procedureusesthesetwopiecesofinformationtosegment of the points in the next situation .

<Page:13> erro

<Page:14>  Figure 9 is related to experiments similar to that of values of SR, RER, and RSE lead to different dividing lines . Increasing the rate to values although not completely, showed similar behavior . Figure 9 (SR = 0.6) represents that increasing retention makes the dividing line push out in some spots . Figure 10a, b shows the optimum-tention steady expenditure-RSE .

<Page:15>  The proposed MLP-based cost-sensitive classi cation algorithms are compared with the recent state-of-the-art rithms, has been adaptededtotakeadvantage of thresholding (cost-sensitive) techniques and the recent [47]andresamplingstrategies[20] to lead to four different churn decision-making efforts .

<Page:16>  The methods use the same proportions of the data for accuracy,F-score,andmisclassi cationcosts . Train set (70%) and cost-blind, class-imbalance-blind metric are used to train and evaluate each algorithm . The computed comparison between the pre-problems and the actual churn occurrence based on the performance of each method is based on metrics .

<Page:17> erro

<Page:18>  The training for best-of-the-range neural-computingand-apparents was published in January 2019 . The results show that the training process is based on a network of networks with parameters and parameters . The training process has been described as 'proveable' and 'preparative'

<Page:19> erro

<Page:20> erro

<Page:21>  The average of AD SOED performs 0.72% better in the classi cation of

```
472 classi-churn misclassi-classifications . Table 2 validates recent pro-driven
churn decision-driven decision-making efforts [6, 16, 54] The one best
performance of AD MLP may have been the making efforts .
<Page:22> erro
<Page:23>  Figure 13 illustrates the misclassi cations of all the cost- higher
customer values as opposed to colors white and costly methods in Table 2 . The
color of the clusters, fully captures the importance and the essence of cost-
centric methods . Figure 13 shows yellow which show lower customer revenues .
<Page:24> erro
<Page:25> erro
<Page:26> erro

Your max_length is set to 142, but you input_length is only 114. You might
consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:27> erro
<Page:28>  14956 NeuralComputingandApplications(2020)32:14929-14962 . See Tables
6, 7, 8, 9, 10, and 11, respectively, represent-ingthe20validationruns for
CSMLP1-4, CSDT,and CSABoost (Table 12)
<Page:29>  NeurocomputingandApplications(2020)32:14929-14962 14957.
CognitiveResearcher: SOMhitrateexamplesforTable4experiments . Neuroscientists:
10% TD 10%TD B1 B2 and 10% CDT: 1/2 .
<Page:30> erro
<Page:31> erro
<Page:32> erro
<Page:33> erro
<Page:34>  A churn prediction model using a random forest remains neutral with
regard to machine learning techniques for churn prediction techniques . An
empirical model of churn prediction can be used to assess the impact of derived
behavior information on cus- churnprediction in the  nancial service industry.
An empirical study of churn predictions has been published by Springer Nature .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11497/Jahromi et al_2016_Customer Churn Models.pdf
<Page:1>  The study aims to compare the performance of probability and data
mining models of customer churn . The results suggest that the decision tree
model with cost sensitive learning has the upper hand in identifying the true
churners . Probability models are the most well-known and recommended stochastic
methodologies to recognize customer churn as well as predicting future sales in
non-contractual settings .
<Page:2>  The data for this study comes from customer transactional records of
the online CD retailer CDNOW, in a period between January 1997 and June 1998 .
For the model building purposes this time window has been broken into two equal
‘calibration’ and ‘validation’ periods of 39 weeks .
<Page:3> erro
<Page:4>  The study investigated the performance of existing customer churn
modeling approaches . Three predictive models were developed and compared using
Pareto/NBD model from probability modeling stream and decision tree from data
mining stream . Results revealed that although the Pare to/nBD model shows a
slightly better performance in terms of general accuracy, the decision tree
```

model with cost sensitive learning has the upper hand in terms .
<Page:5> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11498/Mitrovic et al_2017_Scalable RFM-enriched Representation Learning for Churn Prediction.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11499/Huang et al_2009_Customer Churn Prediction for Broadband Internet Services.pdf
<Page:1>  Customer Churn Prediction for Broadband. Internet Services. Thechurnofcustomerscauses a huge lossoftelecommunicationservice anditbecomesa very serious problem . Until now, churn prediction has been focused on voice services available over mobile and  xed-line networks .
<Page:2>  Until now very little churn-prediction has been carried out on broadband Internet services over  xed-revenue networks . The literature of churn prediction in telecommunication does not provide the details of methodologies for churn prediction using broadband-internet services . This paper presents a new set of features with four modelling techniques for customer churn prediction . The mod-uctiveelling techniques used to predict churns are LR, DT, ANN and Support VectorMachines (SVM)
<Page:3>  The accuracy of true churn (TP) is de ned as the proportionof churn cases that were classi-formed correctly . The false churn rate (FP) is the proportion of nonchurn cases that are incorrectly classed as churn . The proposed churn prediction system for broadband Internet consists of sam-pling data, preprocessing, and classi.-prediction/classi-cation/prediction .
<Page:4>  The available demographic information for this research is gender and country . Account information includes account status, creation date, the bill frequency, service usage information, the number of calls, the standard prices and the fees payed in 30 days of the most recent bill . These information may be useful for predicting the fur-ther behaviour of a customer . For example, a customer with a disability or over 80 are more unlikely to churn from that service .
<Page:5>  The algorithm of Henley segmentation splits cus-tomers and potential customers into di-erent groups or levels according to characteristics, needs, and commercial value . The number of broadband Internet and telephone lines, the number of telephone lines and the voice mail service indicator are selected as part of new features .
<Page:6>  Some new features should be extracted from the usage informa-tion ofbroadbandInternet forchurnpredictionintelecommunicationservice . The ratio

between the total sizes of information downloaded/uploaded and the duration of online broadband Internet for month i is "R GB ONTi" The ratio can be calculated by equation (8)

<Page:7> Customer Churn Prediction for Broadband Internet Services 235aresegmented into 15-dayperiod, then numberofcalls,durationandfees of each customer are aggregated for each customer . For a segment i of a customer'scalldetails,letthe aggregatednumber of calls, duration and fees will be"CALL Ni","DURi"and"COSTi',respectively. The value of a feature was rewritten into binary strings . The values of each of these features (e.g the number of lines, the sizes of information downloaded/uploaded).

<Page:8> In this study, r is set by one parameter which is de ned by a user . The valuesofthesefeaturescan bebenormalised into a similar range by Equation 11.(cid:3) Logistic Regressions: Logistic regression [9] is a widely used statistical mod-reatingtechniquefordiscriminativeprobabilistic classi-classi-cation. The model can be written as:. (Cid:2)

<Page:9> The Back-Propagation(BP) orquick-propagation learning algorithms would be used to train MLP . The more details with learning algorithm can be found on [14]. An SVM classi er can be trained by nding amaximalmarginhyper-planeintermsofalinearcombinationofsubsets(support-vectors)ofthetrainingset .

<Page:10> The broadband monthly usage information for a number of months is formed using the current months data in addition to all previous months data e.g. the 3-month data subset contains the dataformonth3,2,and1 and the 7-monthdatasubsetcontainstheinformation for month 7, 6, 5, 4, 3, 2 and 1 . Four predictionmodelling techniques (LR, DT, MLP, MLPandSVM) were used for each subset of features .

<Page:11> Based on the extracted and normalised features, each SVM was trained to maximise the separation of the decision hyper-plane that maximises the margin of the clas-centricsi-si ed training data . 289 combi-nations of C and 2 with 3 folds of cross-validation were used for training each.SVM . The optimalparametersets (C, 2)yielding a maximumclassi cationac-rophiccuracy of standardSVMswere(2−6,28)foreachsetofexperiments .

<Page:12> Researchers from B.Q. Huang,M-T. Kechadi, and B. Buckley. Buckley have published a new study of broadband usage data . The results were based on the use of various datasets to predict broadband usage patterns . The study was conducted by BQ Huang, M-Kechadi and Buckley, with the help of B-Buckley .

<Page:13> Figures 1(a) and 1(b) show that: The number of months of broadband usage information is between 3 to 9 to obtain better prediction rates . Figure 2 shows that:. the DT and SVM would get lower prediction rates (FP) than the SVM and MLP .

<Page:14> Four modelling techniques (LR, DT, MLP and SVM) were used for customer churn predictions . The most expensivecomputationalcost was spent on using the MLP, the computational cost of using the SVM is more expensive . The prediction rates (TP, FP) obtained on the information without broad-genreband information are also high (about 71% and 1.1%)

<Page:15> Customer Churn Prediction for Broadband Internet Services 243 . Because the imbalance classi cationproblemtakes place in this application,the.ap ologeticmethodsofimbalanceclassi.cationsshouldbefocusedinthefuture . This research was partly supported by Eircom of Ireland .

/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11500/Idris et al_2013_Intelligent churn prediction in telecom.pdf
<Page:1> Churn prediction in telecom has recently gained on the standard telecom datasets . We propose an intelligent churn prediction system inachieving higher accuracy by employing ef cient feature extraction tech- stances . We have observed ality of the telecom datasets. The telecom operators realize the importance of retaining the customers instead of adding new customers .
<Page:2> erro
<Page:3> An ensemble approach handles large dimensionality and typically, ensemble methods are considered a better producesimprovedpredictionperformance . We have adopted 10-fold cross validation to assess the effectiveness of ensemble methods for predicting churn prediction in various areas [7] A simulation based study is performed to analyze the capabilities of high performing ensembles methods and feature extraction methods .
<Page:4> The maximum relevance is sort out by searching 2.3 F-score-based featureselection . Given training vectors, the features have higherde- xk,k=1…,m, m,ifthenumberofinstances ofchurnerand the higher total accuracy . Random Forest minimizes the overall error rate .
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
<Page:11> IntelligentchurnpredictionintelecomːemployingmRMRfeatureselectionandRotBoostbasedensemble 669.6 Performancecomparison of RandomForest,RotationForest, DECORATE and RotBoostonCell2Celldataset 669 .
<Page:12> erro
<Page:13> erro
<Page:14> Yeon Soo Lee studied Metallurgy and Mechanical engineering in South Korea . AdnanIdris received his M.S. and Ph.D. courses at the University of Gwangju Institute of Science and Technology, South Korea, in 2002 . He has more than 15 years of research experience and is working as Asso- enceproceedings,Paris,France,June28,2009,vol28, 2009, . He is a professor in DepartmentofComputerandInformationSciences in the Department of Mechatronics .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11501/Mohanty-Rani_2015_Application of Computational Intelligence to Predict Churn and Non-Churn of.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11502/Wang-Xiao_2011_Transfer Ensemble Model for Customer Churn Prediction with Imbalanced Class.pdf

<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5>  Using a flow network graph to predict customer churn in credit card data engineering, "." Expert Systems with Applications, vol. 38, 2011. "Predicting customer retention and  predicting customer churn," "TrBagg: a simple  simple  tool's 'Tragg'
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11503/Xiao et al_2015_Feature-selection-based dynamic transfer ensemble model for customer churn.pdf
<Page:1>  The study proposes a feature-selection-based dynamic transfer ensemble (FSDTE) model that aims to introduce transfer learning theory for uti-reprelizingthecustomerdatain . The modelmainlyconductsatwo-layer featureselection.Inthe rstlayer, an initialfeaturesubsetisselected by GMDH-typeneuralnetwork onlyinthetarget domain .Inthe secondlayer, severalalappro-priatepatterns fromthesourcedomaintotargettrainingset are selected, and then,wetrainabaseclassi erisselecteddynamicallyforeachtestpatterns .
<Page:2>  Customer churn is de ned as the propensity of customers to cease doing business with a company in a competitive market . To support the enterprises and reduce customer churn rate, we need to identify the customers that are at high risk of churn and optimizeizemarketinginterventionresourcetoretain more customers .
<Page:3> erro
<Page:4>  Traditional machinelearning methods usuallysupposethatthetrainingdatasetandthe test data . DCE strategies containdynamicclassi erselection[22,26] anddynamic classi erensembleselection(DCES) strategies . The mainideaoftransferlearningistoutilizethedata ofrelatedtaskstoassistinmodelingoftargettask[16]
<Page:5> erro
<Page:6>  The algorithm determines the input variables, structure, and parameters of the model automatically, and can .accomplish the process of self-organizing modeling, and . also can avoid over- tting [37]. 34 J.Xiaoetal.Xiang: Algorithm generates candidate models, select and reserve; generate candidate models and select again; …… Continue the above process till getting the optimal model .
<Page:7>  Feature-selection-baseddynamictransferensemblemodel 35. The FSDTE model is based on a GMDH-type model . The target domain T and the source domain S contain m and m patterns, respectively, and they are subject to different distributions . T is divided into two subsets: target.training set T and target test set T . And there are m patterns in T . Repeating N times in the second layer, N new feature subsets are obtained, and then, a base classi er is trained in each subset .
<Page:8> erro
<Page:9>  Feature-selection-baseddynamictransferensemblemodel 37.2. CombineeverytwoinitialmodelstoobtainC2middlecandidatemodelsinthe rstlayer (seeFig.2),estimatethemodelparametersbyleast-square(LS)insetA) and thencompute

theexternalcriterion(theGMDH-
typeneuralnetworkhasanexternalcriterionsystem)valueforeachmodelinset B;.
SelectQL(QL-1) models withsmallerexternalcriterionvaluestoenterthenextlayer,.and
thencompute.theexternalcrit
<Page:10>  Thereare 3,333 patterns, among which 2,850belongtonon-churncustomers
and483belong to churncustomers;Theratiois 5.9006, and the
classdistributionishighlyimbalanced. Thedataset includes 20
features,namely,phonenumber(uniqueforeach)andareacode,aredeemedirrelevant .
<Page:11>  Feature-selection-baseddynamictransferensemblemodel 39.1 . Table2
Attributed Description of "China-churn"dataset . In this study, we adopted the
random sampling without the . estrainreplacement method to select 30% of the .
random sampling . Table 2: We need to partition the target domain Tinto target
set Tinto Tinto training and Tinto test set T
<Page:12>  In this study, we choose support vector machine (SVM) for its
popularity and immense success in various customer classi cation tasks .
Bagging, Bagg-OT, TFS, TrBagg, and TrAdaBoost models all collectively considered
to be the best performance models . To ensure the fairnessofcomparison, we
balanced the classdistribution ofdatabyusingtheover-sampling technique .
<Page:13>  The proposed FSDTE model has four parameters: the number of nearest
neighbors, number of nearby neighbors, rpercentpatternsselected, ppercent
patterns, and ppercentfeaturesselected from theremainingfeaturesubsetF-F . The
averagevalueoftheresultsof10experimented is the largest, followed by those of
over-sampling and over the targettrainingset .
<Page:14> erro
<Page:15>  Feature-selection-baseddynamictransferensemblemodel 43.00 1.00 (a)
(b) The impactofparameterK on the performance.aPerformancein.bPerformancein
"churn"dataset. Theoptimalvalue of K
maybedifferentfordifferentensemblestrategies. We experimented withsevendifferent
values of K:3,5,7,9,11,13, and15 .
<Page:16>  The FSDTE model with p = 70 shows the best performanceinthe
‘churn’dataset because,inthiscase,the TypeIaccuracyandAUC ‘values reach their
maxima’ The model performance when the value of p varies from 10 to 100 is shown
in Fig.6 . The performance of the model with P=70isalsothebestinthe’s ‘China-
churn.’
<Page:17>  Feature-selection-baseddynamictransferensemblemodel 45.00 . FSDTE
Bagging Bagg-OT TFS TrBagg TrAdaBoost . We conducted the ‘reachesthemaximum’ of
each row [44] to determine whether the proposed model can signi cantly
‘outperformtheother’ models .
<Page:18> erro
<Page:19>  Feature-selection-baseddynamictransferensemblemodel 47.5 Conclusions:
FSDTE outperforms two traditional churn prediction strategies . TrBagg and
TrAdaBoost showcomparableperformanceinthe’s "China-churn"dataset .
<Page:20> erro
<Page:21>  Feature-selection-baseddynamictransferensemblemodel 49.7.8.9.5.5 .
Xiao received his PhD. degree from Business School of S.ichuan University,
S.A.University, Chengdu, in China, in 2010 . Xiao is currently a post-doctoral
research assistant at Business School .
<Page:22>  50 J.Xiaoetal.YiXiaoreceivedhis Ph.D.degreefrom

SchoolofInformationMan- agement, Central China Normal University, Wuhan, China, in 2009 . Currently, he is an associateprofessoratManagementFaculty,Chengdu UniversityofInformation-Technology. His research interest includes achievements transforma-repretion,thealliancesbetweenindustry,academia,andtheresearchcom-com-munity.
<Page:23> ShouyangWang received the Ph.D. degree from Institute of Systems Science, Chinese Academy of Sciences (CAS), in 1986 . He is currently the President of International Society ofKnowledgeandSystemsSciences . He has published 18 books and over 200 papers in leading journals .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11504/Vijaya-Sivasankar_2019_An efficient system for customer churn prediction through particle swarm.pdf
<Page:1> An ef cient system for customer churn prediction through particleswarm optimization based feature selection model with simulatedannealing . J.Vijaya1 · E.Sivasankar1: Churn prediction in telecom has gained a huge 1 Introduction . The paper presents a technique today. This is caused in part due to increased competition.
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> S10766 ClusterComput(2019)22:S10757-S10768: S10757-20: S10768 . ROC curve and PR curve for different churn predictor (PSO,PSO-FS, PSO-SA, PSo-FSSA,DT, NB, KNN, SVM, RF, K- Means-DT,WK-FOIL,ANN-MLR)onorangedatasetwith5000customers.aROCPLOT(Orange-5000),bPRPLOT
<Page:11> erro
<Page:12> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11505/Semrl-Matei_2017_Churn prediction model for effective gym customer retention.pdf
<Page:1> erro
<Page:2> erro
<Page:3> We will look at whether consistency in behaviour, from one week to the next, is a predictor of engagement . The models wedeveloped onbothplatforms performbetter than random or mean, so they can already provide a business-oriented business . The study is published on November 02, 2020 at 1749:55 UTC from IEEE Xplore .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11506/Shao et al_2008_Construction of Bayesian Classifiers with GA for Predicting Customer Retention.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro

<Page:5> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11507/Sivasankar-Vijaya_2019_Hybrid PPFCM-ANN model.pdf
<Page:1>  The data mining approaches can aid in the prediction of churn behavior of consumers . The proposed hybrid PPFCM-ANN model provides maximum accuracy when compared to any single model . The retention of customers depends on cus- peripheral changes and retain the customers and income . The mobile mobileshift from one supplier to another heavily damages business .
<Page:2> erro
<Page:3> erro
<Page:4>  Figure 1 shows how the proposed prediction system (HMM) and comparisons were made with the per- canbeprocessed. dataset (DS1, DS2, DS3 and DS4) Among the four con-sidered datasets, DS1 and DS2 were tested using the pro- intentionofthisresearchistodesignCCPmodel based on a Markov sequence alignment (MSV), random model on probabilistic fuzzy C-means and arti cial neural net- induced neural net .
<Page:5> erro
<Page:6>  In hybrid models, unsupervised learning techniques play a vital role for predicting better results . The clusters are made based on the 3.2 Prediction based on artificial neural network. The ANN indicates the biologically aggravated type of a type of type of evaluation. This procedure is rehashed until the topredictthechurningcustomer.
<Page:7>  In this article, the data part of cali-reprocessing was forecasted using the most appropriate classi er, where the bration set is taken . There are 100,000 samples and
172neighboringclusterresemblesthetesthetestinformationbasedon attributes with1churnattributepresentinthisdataset . Out of 172 attributes, 137 attributes are deemed to be numerical attributes and 35 are found to be non-churn . Data preprocessing is the most essential and fundamental and fundamental .
<Page:8>  The unwanted features are eliminated from the dataset following Eqs. (14)-(18) and reduced to 170features(136numericand 34nominal) The missing data present in the numerical attributes are FC¼ Z21 ð15Þ- (21) and the mean value of the particular attributes . Performance metrics were calculated using Eq. (13) and to test how likely the observed distribution of data is with the distribution that is expected .
<Page:9> erro
<Page:10> erro
<Page:11> erro
<Page:12> erro
<Page:13>  Theneuralnetworklearningprocessgeneratedbysingle.bysingleANN and the proposed PPFCM-ANN is shown in Fig.5a, . The sample S10 is found to have better accuracythan other 9 samples . Table 5, 6, 7 and 8 illustrate the numerical performance of classi cation performance of projected clas-si-si .
<Page:14>
TheneuralnetworklearningprocessgeneratedbysbysbysingletheANNandtheproposedPPFCM-ANN-ANN. Figure 5: 5: 1: 7194 NeuralComputingandApplications(2019: 31:7181-7200) Figure 7:5 Comparativestudyof.Sample Performancemetric(%) .
<Page:15> erro

<Page:16> erro
<Page:17> erro
<Page:18> The bestprediction methodhasa point inupper left corner. The ROC curve ofthe proposed hybridfuzzyclustering and thearti cialneuralnetworkare present in upper left corner . Table 14 shows the prediction functions of all samples on the basis of altered number of cluster (U)
<Page:19> erro
<Page:20> 29. YeshwanthV,RajVV,SaravananM (2011) and Yeshwawanth V (2011): Evolutionarychurn . Bose I,ChenX (2009)Hybrid models usingunsupervised clus- 38. WuX,KumarV,QuinlanJR,GhoshJ,YangQ,MotodaH,ZhouZHtering for prediction of customer churn . J Organ Comput Electr (2008)Top10algorithmsindatamining.KnowlInfSyst14(1-37)
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11508/Ullah et al_2019_Churn Prediction in Banking System using K-Means, LOF, and CBLOF.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11509/Perianez et al_2016_Churn Prediction in Mobile Social Games.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> There is a higher concentration of data at the beginning of the study . This is due to the fact that we work with censored data and do not follow a normal distribution . The longerthetimeofstudygrows, as there are many whales who have not experienced the event yet because they are still connecting to the game .
<Page:7> We use the same algorithm of conditional in-condition in-ference ensembles, the outcome differs . The binary response model provides useful insight for a very short-termprediction . We train the binary model with several sets of features to obtain the nal list of attributes shown in Fig. 4 . We compare our results with other classi cation methods .
<Page:8> The darkbluedotscorrespondtoshorterlylifetimes(indays) of players, softbluedotsre ectplayerswithlongerlifetimes . Conditional inference survival.ensembles were evaluated to this purpose and compared with traditional survival methods like Cox regression . The results directly impact the game business, the community and the business of the authors .
<Page:9> erro
<Page:10> A.Saas,A.Guitart,andA.Perianez.Discoveringplayingpatterns:Time.series clusteringoffree-to-playgamedata. ComputationalIntelligence.andGames(CIG), 2016IEEEConferenceon, 2016. by j. ross quinlan. morgan kaufmann publishers,

inc., 1993. MachineLearning,1:6,1994.
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11510/Prasasti-Ohwada_2014_Applicability of machine-learning techniques in predicting customer defection.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> The convenience of using each model is represented by the ease of tuning the parameters before proceeding with the algorithm . Figure 5 presents the output of SMO on the customer defection data . The SMO algorithms implement the sequential minimal-centric method used for the customer-defection problem . The study was published on November 02,2020 at 17:50:35 UTC from IEEE Xplore .
<Page:5> erro
<Page:6> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11511/Routh et al_2020_Estimating customer churn under competing risks.pdf
<Page:1> Journal of the Operational Research Society: Estimating customer churn under competing risks . Pallav Routh, Arkajyoti Roy & Jeff Meyer published online: 06 Aug 2020. The article is published by the journal's online version of the journal Tandfonline.com .
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> The PDP for peak-season fre-insured customers predicted to resign, a marketer must con- quency show an increasing and then plateauing rela-quency with the risk of LOA . For seasonal interpurchase times, the risk from LOAcompute an overall importance of variables to increases and then plateaus . This can help marketers frequency. Customers who transact more or less are at risk of resigning .
<Page:11> Figure 3. 10 P.ROUTHETAL. PDP for important covariates for leave of absence . For tree based tree based on CR-RSF, there is an increase in risk of leaving . Figure 5 shows that risk of LOA increases for customers with transaction frequency ranges between (0,41] to (41,82)
<Page:12> 100% of customers who opt for LOA or resign have transactions in areas_dining during peak-seasons . The effect of years of membership, therefore, plays a role in decreasing the values for CIF within this sea-driven CR-RSF . Senior customers who have been with the firm longer than 40years had less frequent purchases .
<Page:13> erro
<Page:14> This study uses a competing risk random survival forest forest-forest-forest method to estimate the churn probabilities of cus-centrictomers . The method can be readily extended to any number of competing events under a

business setting . For example, satellite providers and other telecommunication-providers allow customers to go on a leave of absence and pay a substantially reduced fee for the leave-of- absence .
<Page:15> erro

Your max_length is set to 142, but you input_length is only 50. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:16> erro
<Page:17>  16 P.R.ROUTHETAL. CIFs are active, LOA or resign . CIFIs are active or resigning in the event of an active, resigning or a resignation . The CIFC is based on a correlation matrix of covariates . The correlation matrix is a matrix of correlation and F1 scores .
<Page:18>  The CR-RSF provides estimates of churn for a customer's total expenditure or average expenditure if customer for all time-periods (months) starting from 2009 and ending in 2016 . We constructed an alternative 2009 and 2016 models for both events: active, resign and leave . We computed multinomial logit model provides a single churn prediction . The results were based on 100 bootstrap cross validations .
<Page:19>  CR-RSF outperforms Mlogit for all val-consistent comparison, we only use probabilities of resign ues of threshold for predicting resign and LOA . For higher values of threshold, CR- is on churn prediction via multiple avenues. For higher . values of . churn prediction . CR- outperforms CR- RSF for churning events only: resign and leave .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11512/Naik-Reddy_2017_An innovative optimized model to anticipate clients about immigration in.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5>  Downloaded on November 02,2020 at 17,52:20 UTC from IEEE Xplore . Restrictions apply . (cid:55)    (cids:55)(cid;72)(cids;). (Cid:70)  -   -
 -- - - Downloaded from November 2, 2020 at 17:52  20 UTC . (Cids:44) -  - - . - - - + . - + + + . + .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11514/Zhang et al_2012_Predicting customer churn through interpersonal influence.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11517/Xie et al_2009_Customer churn prediction using improved balanced random forests.pdf

<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> We apply IBRF to a set of churn data in a bank as described above . To test the performance of our proposed method, we run several comparative experiments . We also compare our method with other random algorithms . The results turn out to be insensitive to the value of certain variables .
<Page:5> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11520/Xiao et al_2012_Dynamic classifier ensemble model for customer classification with imbalanced.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11525/Wei-Chiu_2002_Turning telecommunications call details to churn prediction.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11528/Shirazi-Mohammadi_2019_A big data analytics model for customer churn prediction in the retiree segment.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> The attrition rate of Mass Affluent clients vs. non-affluent clients has remained at 5% flat, year by year . Mass-retail clients' accounts for 1%, versus that of 7% for non-mass affluent clients . We investigate the following question: The current research is being designed to address this objective by answering the main question .
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> F.Shirazi,M.Mohammadi International Journal of Information Management

48 (2019) 238-253.1. The study used big data analysis of clients with allpossiblesplittingvariablestotheroot by selecting the most different starting and ending MI balances between $4M and $10M are cate- according to the study .
<Page:10> erro

Your max_length is set to 142, but you input_length is only 77. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:11> erro
<Page:12> erro

Your max_length is set to 142, but you input_length is only 67. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:13> F. F.Shirazi,M.Mohammadi International Journal of Information Management . 48 (2019) 238-253: ChurnRateafterRetirement . Churn pattern of mass-Affluent vs. non-mass-affluent is similar to that of the past .
<Page:14> F. F.Shirazi,M.Mohammadi International Journal of Information Management 48 (2019) 238-253 . ChurnRateandOnlineBehavior: F. Shirazi and M. Shiraazi . Shiraaz: Churn RateandOnline Behavior: A. B. A. Churn Rates and Behaviour: F F. Shirazi, M.Shiraazi, F. Shah, F., M. Shah: F.,M. Shah .
<Page:15> erro
<Page:16> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11531/Coussement et al_2017_A comparative analysis of data preparation algorithms for customer churn.pdf
<Page:1> erro
<Page:2> erro
<Page:3> The churn model was created by K. K.Coussementetal/DecisionSupportSupportSystems95(2017)27-36 29 . The study uses either a remapping strategy for categorical variables or manycategoriessimultaneously. Thedecisiontree-basedremapping methodsforcontinuousvariables. process startswiththeentirecustomerdataset,ortherootnode, then. then.splitsthedataintosmallersubsetsor internalnodeson.1.1 .
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11534/De Bock-Poel_2011_An empirical evaluation of rotation-based ensemble classifiers for customer.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro

<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11537/Verbeke et al_2011_Building comprehensible customer churn prediction models with advanced rule.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7>  The reported measures are the average out-of-sample perforfor-out-Of-sampleperformancegainforAntMiner+ usingoversampling . The results were based on ten random 70/30 split ups of the dataset in training and test sets . Early stopping is applied since the dataset is rela-phthalphthalphthalic0 87.66 1 16.34 .
<Page:8> erro
<Page:9>  AntMiner+ and RIPPER clearly induce much less rules than C4.5, iments.Thiscon rmsprevious results(Martensetal., 2006; van-ishly.TheissuefacedbyC4 .5isitsgreedy decruys et al. et al., 2008).
<Page:10> erro
<Page:11> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11540/Kianmehr-Alhajj_2009_Calling communities analysis and identification using machine learning.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  K.Kianmehr,R.Alhajj/ExpertSystemswithApplications36(2009)6218-6226 6225 . We demonstrated how cluster analysis can be used to identify calling communities . The order of the classi ers do not change . We have proposed a similarity-likelihood measure that combines both the rst-and-second-orderdistances .
<Page:9> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11543/Coussement-Van den Poel_2008_Churn prediction in subscription services.pdf
<Page:1>  Churn prediction in subscription services: An application .of support vector machines while comparing twoparameter-selection techniques . We show that only when the optimal parameter-selection procedure is applied, support vector . machines perform well when applied to noisy marketing data . The parameter optimization procedure plays an important role in the predictive performance .

<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  SVM outperforms SVM on the ‘natural’ distribution3 (11.14% churners) tribution . Table 5 reveals that on all test sets that contain the arti cial ones (50%, 40%, 30%, 20%, 20, 18%, 18, 16%, 16%, 14%) SVM has a higher top-decile lift compared to SVM . This gap increases when deviating from the origi-genicauc acc-nal training distribution .
<Page:9> erro
<Page:10> erro
<Page:11> erro
<Page:12>  The number of renewal points, what product the subscriber has, the length of the current subscription, the month of contract expiration and the number of days a week the newspaper is delivered . The average suspension length (in number of day) and how many days the previous subscriptions are renewed .
<Page:13>  The average cost of a complaint (in terms of compensation Subscription) The number of days a week the newspaper is delivered Subscription(intensity indication) is available Subscriber's phone number (telephone, mobile number, Subscription,) is available (subscription) Whether the previous subscription was renewed before the Subscription’expiry date was renewed . Subscription
X.Coussement,D.VandenPoel/ExpertSystemswithApplications34(2008)313-327 325. 325
<Page:14> erro
<Page:15>  Coussement,D.VandenPoel/ExpertSystemswithApplications34(2008)313-327 327. (2008) Swets, J. A., A., & Pickett, R. M. (1982). Evaluation of diagnostic systems: Weiss, G., & Provost, F. K. (2001). The e ect of class distribution on the eect of .Methodsfromsignaldetectiontheory.NewYork:AcademicPress. classi er learning.
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11546/Farquad et al_2014_Churn prediction using comprehensible support vector machine.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11549/Nie et al_2011_Credit card churn forecasting by logistic regression and decision tree.pdf
<Page:1> erro

```
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10>  15282 G.Nieetal./ExpertSystemswithApplications38(2011)15273-15285
CarbuncleAppendix1(continued) Appendix1: Appendix 1: Valuable assets (cid:2:2)
Averageageofthe 4 2286 L X88 Theamountofloan 0 37,977,770 8.52 X89 X89 Thetimes
ofloanforcash 0 392 16.02 X90 X90s X89s X91s X96s X100 RatioofDebittransactionas
0 1 2.1801 1 4.82s X99 RatioofCredittransactionsas (Cid
<Page:11>  G.Nieetal./ExpertSystemswithApplications38(2011)15273-15285
15283.1(continued)Appendix1(Continued) Var. Description MIN MAX VIF .
Dummy,takes1if 1 3 1.80 X107 RatioofCredittransactionvia 0 2.62 14.87.87
Thelongestlength 0 3E+06 3.79 X108 RatioofDebittransactionsvia 0 1 1 1.31
Betweenissue and telephone/allchannel.com/Allchannel/comporporporations X41
Theshortestlength . X43 Theintervallength 0 2013 7.
<Page:12> erro
<Page:13>  G. G.Nieetal./ExpertSystemswithApplications38(2011)15273-15285 (2011)
15285 (2003) 15284 (2004) 15289 (2003), 15287 (2005) 15287 . (2005). Customer
churn prediction using an Ahybrid model .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11552/Coussement-De Bock_2013_Customer churn prediction in the
online gambling industry.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11555/Amin et al_2017_Customer churn prediction in the
telecommunication sector using a rough set.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  Aminetal. Neurocomputing 237 (2017) 242-254. A.neurocomputing. 237
(17) published: "We have evaluated four di erent algorithms for rules
generation" The GA for computing reducts has a probability of 60%, starting with
an initialpopulationof 50chromosomes . The GA converges in the span of 100 .
```

<Page:9>  Aminetal. Neurocomputing 237 (2017) 242-254.Aminetals.computing . The EA method performed better than NC and CA algorithms (i.e. LA) in terms of coverage, recall, and F-C 483 483 1,000 measures .
<Page:10> erro
<Page:11> erro
<Page:12> erro
<Page:13> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11558/Delen et al_2020_Development of a Bayesian Belief Network-based DSS for predicting and.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4>  In this study, an experimental design was designed to illustrate the importance of variable selection and data-balancing . In Step 1, the combined and consolidated institutional data was analyzed and compared within the BBN . Step 4, a sensitivity study used the heuristic data balancing technique, SMOTE, is designed and executed to discover and show the level  was shown to have the best performance among others in the same domain .
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  A confusion matrix for a two-class (i.e., binary) classi ca-
 probability of root variable and conditional probability of all  tion model contains four populated cells: True positives (TP), True Negatives (TN), False Positives (FP), True Negative (FN) and True Positive (TN) Cells from the upper left to lower-right contain the accurately predicted cases . All other cells contain cells containing the incorrectly pre-  "true performance" of two class response variable models .
<Page:9> erro
<Page:10>  Can be of great use to practitioners (i.e., administrators and man-
probability of attrition in educational institutions) because it offers a holistic view of student who would drop out of the university, if the values of all relationships are different . In fact, with this net-  this study, a 10
-fold cross validation methodology was employed;  it is possible to calculate the student-speci-c risk .
<Page:11> erro
<Page:12> erro
<Page:13> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11561/Alboukaey et al_2020_Dynamic behavior based churn prediction in mobile telecom.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6>  RFM-basedModelArchitecture.com proposes a differentarchitecturefor

CNN-based model . The CNN model is composed of tomer from his daily behavior. We employed CNN to learnrepresentativefeaturesofthecus-cantakeinan input2Dor3Dmatrix .
<Page:7> CNN-basedModelArchitecture. The one-dimensional convolutional layer in the CNN model connectedlayerseachwith128 units and 20% dropout . Each kernelcorrespondstoamoving output layer of one sigmoid node . The all parameters and layers'weighted sum function with learnable weights shared over time .
<Page:8> In churn prediction domain, thereare just few works that investigated this type of deep learning tech-nique on monthlybehavioralfeatures . In this paper we used LSTM and fed it with daily behavioral fea-centrictures . In each time step(day), all behavioral features follow throughout the four gates of the 4 gates of (LSTMcell) Thesegatesconsti-receive information throughouttheprocessingof buildamodel named RF-Daily.
<Page:9> erro
<Page:10> erro
<Page:11> N.Alboukaeyetal./ExpertSystemswithApplications162(2020)113779 11: RF-Monthly LSTM-monthly RFM-basedModel LSTm-based model . The Area under the Curve (AUC) summarizes the overall performance in all possible cutoffs .
<Page:12> Our dataset consists in Eq. (3) This dataset has been split into two datasets: training and usage tables from March 2019 to August dataset of 1.35 million observations . The detailed usage tables contain daily customers' usage amount thetesting daysconstitutethirtytestingdatasetseachof 50thou-insured calls, SMS, data, and services .
<Page:13> N.Alboukaeyetal./ExpertSystemswithApplications162(2020)113779 13 wereimplemented in python3usingscikit-learn0.22.0,andkeras dowisfrom29thJuly-to27thAugust . The mean and the standard 2.2.2-2.4 withtensor ow2.1.0backend was performed on a PC with an Intet Core i7 CPU@2.8 GHz, the testpredictionwindowsarelistedinTables4and5 .
<Page:14> DeepRFM-based model is faster to train and predict than Statistics-learningmodelist . CNN-based features don'tcontributeetotheaccuracyatall. However, Statistics-based models are more accurate than CNN models . Daily models need to train more time to train than monthly models, and that they vary in dailyandthemonthlymodelson .
<Page:15> There are statistical differences in perfor- doso,wecomparedthe performancebydayofthetwobestdaily-performance . The RF-Daily model signi cantly outperforms the Statistics-based . CNN-based models and RF-daily model always outperform the dynamic monthly .
<Page:16> In this paper, our goalistopredictchurnata is to improve the efirir of retention marketing marketing . Previous works have lookedatchurnasastaticpredictionproblem and the quality the representation learned by the proposed model . The output of the t-SNE algorithm is shown in Fig.9 and Fig.10 .
<Page:17> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11564/Gur Ali-Aniturk_2014_Dynamic churn prediction framework with more effective use of rare event data.pdf
<Page:1> erro
<Page:2> erro

<Page:3>  Table 1: Table1: Sample ofchurnpredictionpapers since 2004 . Table2: Table3: Table4: Table5: Table6: Table7: Table8: Table9: Table10: Tabletable1: Table 4: Table 5: Table 6: Table 7: Table 2: Table 3: Table 8: Table 1; Table 4; Table 5 : Table 5; . Table 6 : Table 7;. Table 10: Table 9: Table 10 . Table 10; . . Table 7.: Table 11: Table 12: Table 13: Table .
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
<Page:11> erro
<Page:12> erro
<Page:13>  Ö.GürAli,U.Arıtürk/ExpertSystemswithApplications41(2014)7889-7903 7901 . The average and standard error of the AUC and TDL measures persist across time periods - we see that the lines in average differences across lead times are not cross, with the exception .
<Page:14> erro
<Page:15> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11567/Schaeffer-Rodriguez Sanchez_2020_Forecasting client retention - A machine-learning approach.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4>  S.E.SchaefferandSV.V.RodriguezSanchez Journal of Retailing and Consumer Services 52 (2020) 101918 200 200 200 150. 150. We seek to train a supervised machine-learning algorithm to dis-proportion of lost clients and retained clients .
<Page:5> erro
<Page:6> erro
<Page:7>  Journal of Retailing and Consumer Services 52 (2020) 101918Table4. WehighlighttheA,Se,Seingreenwhen 0.80; thecombinationsof
, ,and thathaveanadequateclassifierare
highlightedinblue,nbeingthenumberoftimeseriesnineachdataset.
<Page:8> erro
<Page:9> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11570/Keramati et al_2014_Improved churn prediction in telecommunication industry using data mining.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7>  The Turkish telecommunications industry has been described as a

‘telecommunication industry industry’s ‘Telecommunicationindustryindustry of Turkey’ The study was published on November 4, 2013 . The study is entitled "Ethiopia’ s Ethiopia," and ‘Ethiopian’. The Turkish study is published on December 4, 2014, and 2015 .
<Page:8>  Theproportio n ofnon-voic ecalls/proportionofcalls duringthedaytime (2) and number of calls (e.g., international or local calls) is a factor . Table1(Continued) shows Table1, Table2, Table3, Table4, Table5, Table6, Table7, Table8, Table9, Table10, Table11, Table12, Table13, Table14, Table15, Table1: Table4: Table3: Table1.
<Page:9> erro
<Page:10> erro
<Page:11> erro
<Page:12> erro
<Page:13> erro
<Page:14> erro
<Page:15> erro
<Page:16> erro
<Page:17> erro
<Page:18> erro
<Page:19> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11572/Coussement et al_2010_Improved marketing decision making in a customer churn prediction context using.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7>  2138 K.Coussementetal./ExpertSystemswithApplications37(2010)2132-2143 (2010) 2138 C. C.C. Coussementeetal/Expert SystemswithAppendixA. The client/company-interactionvariables. The average positioningofthecomplaintsinthecurrentsubscription,. Theaveragesuspensionlength(innumberofdays), Theaveragenumberofdaystheprevioussubscriptionsarerenewedbeforeexpirydate, Theaverageumberof.days,. Elapsedtimesincelastconversionindistribution.channel,paymentmethod&edition,
<Page:8>  K. K.Coussementetal./ExpertSystemswithApplications37(2010)2132-2143 2139.2139.25 churnprobability0.45 0.45.45 ..0 0.15.1 0.02.52 .
<Page:9>  C.Coussementetal./ExpertSystemswithApplications37(2010)2132-2143 . Churn probability00.182 churn probability00..001.182 . The churn probability of churning is 0.0.182, with churning probability of up to 1,000,000 .
<Page:10>  K. K.Coussementetal./ExpertSystemswithApplications37(2010)2132-2143 2141.2141.1.2.1 (continued) Churn probability00..000.681 churnprobability0.81.81 (churn probability)
<Page:11> erro
<Page:12>  Coussementetal./ExpertSystemswithApplications37(2010)2132-2143 2143 . Coussementsmentetals/Expert systemswith applications37(1). 42(3).

Coussemenetal.C. J. Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, E. E. K. (2001) Knowledge ofthehousinganddevelopment board,Singapore.
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11576/Coussement-Poel_2009_Improving customer attrition prediction by integrating emotions from.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7>  We would like to thank the anonymous Belgian company for providing us with data for testing our research questions . There is a wide variety of data Ghent University for funding the Ph.D. project . C.-C. Chang and C.-J. Lin for sharing their SVM-toolbox, LIBSVM .
<Page:8> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11579/Benoit-Van den Poel_2012_Improving customer retention in financial services using kinship network.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  VandenPoel/ExpertSystemswithApplications39(2012)11435-11442
D.F.Benoit,D.Vandenpoel,D., &Peppard,J.
(2000)CustomerRelationshipManagement(CRM)in CRM) in CRM .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11582/Moeyersoms-Martens_2015_Including high-cardinality attributes in predictive models.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11585/Glady et al_2009_Modeling churn using customer lifetime value.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro

<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  N.Gladyetal./EuropeanJournalofOperationalResearch197(2009)402-411
409.409.489.488.488 . Table5: Performanceofclassi erswithh¼3,fortwodifferentcut-
offs, the PCC, theneuralnetwork, thedecisiontree, the cost-sensitivetree,
AdaCostclassi.73 96.46 0.13 93.83 84.77 97.77 . Table 5: PCC(%) Truepositives(%)
s L3(%) PCC) Truepositive(%) S PCCS(%) L3S) PCC S(S) L
<Page:9> erro
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11587/Verbeke et al_2012_New insights into churn prediction in
the telecommunication sector.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  The general process model of the development of a customer-driven
algorithm is described in the European Journal of Operational Research
218(2012)211-229. 218 W.Verbekeetal/EuropeanJournalofOperationalResearch.218
(2012)Table1 describes the methodology that is followed in preprocessing the raw
data sets .
<Page:9> erro
<Page:10> erro
<Page:11>  W.Verbekeetal./EuropeanJournalofOperationalResearch218(2012)211-229
221 . The percentage of churners typically lies within a range of 1-10% of the
entire customer base . The benchmarking study is to learn about how customers
can be retained (Bolton et al. 2006)
<Page:12> erro
<Page:13>  W.Verbekeetal/EuropeanJournalofOperationalResearch218(2012)211-229
223/211/229 223 . The returnoninvestment of
oversamplingontheperformanceofacustomerchurnpredic-
toimprovedataqualitydependsonthestructureoftheprocesses- tion model .
<Page:14> erro
<Page:15> erro
<Page:16> erro
<Page:17> erro
<Page:18> erro
<Page:19> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11594/De Bock-Van den Poel_2012_Reconciling performance and
interpretability in customer churn prediction using.pdf
<Page:1> erro
<Page:2> erro

<Page:3> erro
<Page:4>  Generalized feature ting and marketing objectives of retention-increasing efforts . Permutation accuracy importance scores are calculated using evaluationcriterionPC.AseverymembertreeF,j=1…,M.VandenPoel/ExpertSystemswithApplications39(2012)6816-6826 6819 .
<Page:5> erro
<Page:6> erro
<Page:7> erro

Your max_length is set to 142, but you input_length is only 114. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:8> erro
<Page:9>
Bootstrapcon denceintervalsandaveragetrendsforaselectionofpredictivefeatures. 6824 K.W.DeBock,D.VandenPoel/ExpertSystemswithApplications39(2012)6816-6826 .
<Page:10> erro
<Page:11>  The case against accuracyestimation for
Strobl,C.,Boulesteix,A.L.,Zeileis,A., &Hothorn,T. (2007)
Biasinrandomforest.comparinginductionalgorithms. 6826
K.W.DeBock,D.VandenPoel/ExpertSystemswithApplications39(2012)6816-6826 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11597/Verbeke et al_2014_Social network analysis for customer churn prediction.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8> erro
<Page:9> erro
<Page:10> erro
<Page:11> erro
<Page:12> erro
<Page:13> erro
<Page:14> erro
<Page:15> erro
<Page:16> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11600/Risselada et al_2010_Staying Power of Churn Prediction Models.pdf
<Page:1> erro
<Page:2> erro
<Page:3>  The study Literaturestream Comparedmethods were conducted by 200 H.Risseladaetal./JournalofInteractiveMarketing24(2010)198-208/2010 . The study was conducted in conjunction with the University of Minnesota's Marketing Institute of Minnesota. The results were published in the U.S. market.

<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  There are three possible explanations for the changes in the model: multicollinearity,omittedvariables, and Staying Power: Top-decile Lift . The most plausible explanation is that the two lines overlap in Fig.4 . However, we period t and t+1, but performs slightly better than the other possible explanations .
<Page:9> erro
<Page:10> erro
<Page:11> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11603/Tsai-Chen_2010_Variable selection by association rules for customer churn prediction of.pdf
<Page:1> erro
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5>  2010
C.-F.Tsai,M.-Y.Chen/ExpertSystemswithApplications37(2010)2006-2015)
New.icioTraining.icioModels data is based on the original 22 variables .
The.icioPredictionperformancebythetesting.dataset(22variables)
<Page:6> erro
<Page:7> erro
<Page:8>  C.-F.Tsai,M.-Y.Chen/ExpertSystemswithApplications37(2010)2006-2015 2013 . Table 10: Predictionperformancebythevalidationdataset(12variables) Model Status Precision(%) Recall(%) Accuracy(%) F-measure.1: Precision of DT and NN using the 22 and 12 variables .
<Page:9>  2014
C.-F.Tsai,M.-Y.Chen/ExpertSystemswithApplications37(2010)2006-2015 . The entropy method can provide the best prediction performance, i.e. to delete the rule that contains less than 100 cases . ThenewDTmodel contains only 8 levels of hierarchy and 43nodes . That is,thedarknodes mean higher proportion of customer churn .
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11667/Liao-Chueh_2011_Applying Fuzzy Data Mining to Telecom Churn Management.pdf
<Page:1>  Customers tend to change telecommunications service providers in pursuit of more favorable rates . How to avoid customer churn is an extremely critical topic for the intensely competitive telecommunications industry . Study used fuzzy data mining to determine effective marketing strategies by analyzing the responses of customers to various marketing activities . These techniques can help tele-centriccommunications service providers determine the most appropriate marketing opportunities and methods for different customer groups .
<Page:2>  The study used fuzzy data mining techniques to analyze the re-sponses of customers to various marketing activities and thus determine effective marketing strategies . The study also confirmed that pre-use association rules

to carry out factor analysis of customer churn predictive model selection accuracy of the analysis of association rules is better than no prediction model . The main purpose of the study was to build an effective churn prediction model to forecast which customers are likely to churn .

<Page:3>  There are three basic operators of fuzzy sets: union, Intersection, and complement . In this study, the fuzzy sets can be used to assist in dealing linguistic means and avoiding the boundary shape problem . Data mining is defined as use of automated or semi-automated method from a large number of data collections to extract potential, meaningful and useful information or patterns .

<Page:4>  400 customers whose contracts were due to expire in June and July 2008 were randomly selected from each of the following groups: customers with monthly bills of NT$ 0 ~ NT$300 . Customers with monthly bill amount ranges were divided into two subgroups of 200 customers each . Customer retention marketing programs were implemented by sending direct mail (DM) and through telemarketing .

<Page:5>  Telecom marketing model uses fuzzy data mining techniques to analyze past records of results of various market-related activities to establish a marketing model . The degree of effective marketing is 0.4; if mail DM is used, then it is completely not effective marketing . The proposed marketing model can provide companies on determining the best marketing strategies for different customer groups .

<Page:6> erro

/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11668/Lee-Jo_2010_Bayesian Network Approach to Predict Mobile Churn Motivations.pdf

<Page:1>  Bayesian Network Approach to Predict Mobile Churn Motivations: Emphasis on General Bayesian network, Markov Blanket, and What-If Simulation . The annual churn rate ranges from 20% to 40% for most global mobile telecommunications service companies . Reducing churn is important because acquiring new customers is more expensive than retaining existing customers .

<Page:2>  Bayesian Network Approach to Predict Mobile Churn Motivations  305 customers . Churn occurs when a customer terminates the use of a service from the service provider . In the telecommunication market, churn can be  measured as the cancelation rate in a certain period of time .

<Page:3>  In this paper, we use all three types of BN structure to compare performance as well  to reveal the underlying structure of churn motivations . The data in this paper were donated by a major mobile telecommunication company in South Korea . Data originally consisted of 14 variables and 5,000 records that were sampled from anonymous churned customers .

<Page:4>  The Bayesian Network Approach to Predict Mobile Churn Motivations was created using a Bayesian network approach to predict mobile loyalty . We used a number of variables to characterize customers' loyalty such as months of usage (tot-MonofUsg) and the duration ratio of usage after changing to a new device out of the total usage (AfterdevchgP) The 11 variables (Table 2) were used to determine which networks had a target node of 'Churn motivation' The structure of the GBN was learned using two search algorithms, K2 [5] and Hill Climber, with the maximum number of parent nodes limited to one .

<Page:5>  The final analysis was conducted on 6 structures using two types of variables-full variables and MB variables . Table 2 illustrates these types . GBN-HC and TAN outperform GB-K2 and NBN, and GB-HC show statistically same performance . The results in Table 3(a) reassure that (1) GBN .-HC agicallyand TAN show the statistically . same performance, and (2) both GB .N.K.C. Lee and N.Y. Jo Jo Jo
<Page:6> erro
<Page:7> erro
<Page:8>  The motivation of churn can be predicted by performing what-if analyses with GBN-HC structure using MB variables . Figure 2 illustrates the what-If analysis results showing customers are more likely to transfer to a competing company when ARPU of the previous month was rather high (2) they pay their bill using JIRO  and (3) they call contact center very few . GBN classifiers are showing competitive prediction accuracy compared with other BN classifier using full variables .
<Page:9>  GBN classifiers are capable of uniquely providing what-if simulation func-ishlytions with which decision makers can test various numbers of alternative solutions to  the target problem . The usefulness of using GBN assisted by MB variables is very high, especially in the field of decision problems where a lot of decision variables should be considered .
<Page:10>  Bayesian Network Approach to Predict Mobile Churn Motivations . Bayesian method for induction of probabilistic networks  from data. Machine Learning 9(4), 309-347 (1992) Bayesian network classifiers. Machine Learn-Learning 29(2), 131-163 (1997) In: Proc. 13th International Conf.Machine Learning, pp. 284-292 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11669/Mohanty-Naga Ratna Sree_2018_Churn and Non-churn of Customers in Banking Sector Using Extreme Learning.pdf
<Page:1>  Ramakanta Mohanty and C. Naga Ratna Sree propose to utilize ExtremeLearningMachine(ELM) toforesee clientchurn . Customer churn implies loss of clients. Customer churn can be seen in numerousventures like banking, broadcast communications and insurance organizations. The principal target of the customer retention is to cut through the churn .
<Page:2>  Extreme Learning Machines (ELMs) has become one of the prominent machine learning methods for predictive analysis . Building a functional model for customerchurn has now become a decisive topic in recent days . ELMs are a feed forward neural-network (NN) networkrecognized by the introductionoftheirhiddenlayerweights, alongsidethe training algorithm [2-5]. These models can create great execution and are utilized a huge number of times quicker than other types of neural network like decision trees and by use of ensemble method [6, 7]
<Page:3>  A paper proposes a neural network-based approach to predict the customer churn with respect to the banking domain . Extreme learning-machine is a single-layer network having N nodes . The main advantages of ELM are that its parameters, hidden nodes, input weights and biases are randomly allocated and need not required to be tuned .
<Page:4>  Eq. 54 R.MohantyandC.NagaRatnaSree: We need to derive the relation between X and t from  , w and b such that a relationship can be derived from X,

t and t . The relation (2) can be written as  fðw b xÞ=t,1,1/j, 1/2/3/4/5/6/8/8
.
<Page:5>  The churn and non-churn of customers is a Portuguese Banking Sector dataset, where itconsistsofbothcategoricalandnumerical values are presented in Table 1 . We developed the Java code for ELM and the exper-imentiscarriedoutin MATLABenvironment . We got training accuracy of 0.0044 by using the Tribias activation function .
<Page:6>  We simulate our experiment folds wise, we found that in case off old 1.0045 and followed by fold 6 value of 0.0500, or 0.0000, the best training accuracy is of .0045 . The best testing accuracyvalueisonfold1of.value0.0024and .followed byfold5ofvalue0 .0344,respectively, whichisshown in Table 3 .
<Page:7>  This paper analyses the systematic way to predict customer churn by employing the Extreme Learning Machine . The ELM model gives more accurate results compared to other machine learning techniques, viz. SVM, Gradient des-profitation neural networks, etc. Time to time, many data mining techniques have been implemented on the banking data to predict the customer churn and non-churn of banking customers .
<Page:8> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11670/Khodabandehlou-Zivari Rahman_2017_Comparison of supervised machine learning techniques for customer churn.pdf
<Page:1>  Journal of Systems and Information Technology: Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior . Samira Khodabandehlou, Mahmoud Zivari Rahman, (2017) "Predicting customer churn in mobile industry using data mining technology" The fulltext of this document has been downloaded 96 times since 2017*
<Page:2> erro
<Page:3> erro
<Page:4> erro
<Page:5> erro
<Page:6> erro
<Page:7> erro
<Page:8>  Customershiftstocompetitors Subscribersswitchestoacompetitor Notmentioned NotmentionedSubscriberswitchestoanotherserviceproviderduringaperiod (continued) Steucpmlheenaarrvicqnhisui7inene1dgseinene 1DgseoT)
<Page:9>  Methods StructuralEquationModeling(SEM)DT(CHAID,ExhaustiveCHAID)CART,QUEST,ANN(MLP),ANN(MLP),andlinearandLRANN,SVM(RBF),DT,LR,andensemblelearner(boosting)ANN,DT(C5.0),andAssociationrulesDT,andANN ANN(MLp),SVM (Polynomial,RBF) ANN(BP) and DT(C4.5),andSVM('Linear, Polynomial)
<Page:10> erro
<Page:11> erro
<Page:12> erro
<Page:13>  JSIT criteria measure the capability of a prediction model for accurate ranking of customers (Coussement and De Bock, 2013; Keramati et al., 2014) To evaluate the performance of machine-learning methods in predicting churning, we have used thesecriteria, whicharecalculated

basedontheconfusionmatrix shown in Table II .
<Page:14>  Data fromafoodstoreinIran from March25, 2013 to March11,2015(about24months) 6.1Datasetinvolves1,050customers(761non-churnersand287churner) and includes577,200records . 5,050 records were deleted from the data set, reducing the total number of records . Table III presents data set  elds of customer transactions and their descriptions .
<Page:15>  The data set of customers’ transactions was integrated into the Demographicdataset(age, gender, educationlevel, distanceandgroup) Data integration is considered the second stage of data preparation . In this stage, string variables in the data set were transformed into numericalvariablesandnumbers .
<Page:16>  Thetotalnumberofitemsthecustomerretunedtothestoreafterthepurchasewhichwereacceptedbythestore Thetotalamountofdiscountsreceivedbythecustomerintheselectedperiod Thetimedifferencebetweenthedateofthelastpurchaseofthecustomersandthcurrentdate . Thetotal amountofmoneythecustomershaspaidinthe selected time period forpurchasing items from the store tecmlhenaaricqnhui7inne9gseinne 9gse .
<Page:17>  Based on Table VI, the p-value of RFMITSDP variables is lower than 0.05 . The ability to correctly predict the customer groups by RFMITSDP is92.5percent . Five other variables, which playaroleinpredictingthechurningstatusofcustomers, are selectedfortheanalysisand othervariablesareeliminated .
<Page:18>  In this study, state-of-the-art supervised machine-learning methods have been used to create churning prediction models . The process of creating the prediction model involves training and testing the model . In this section, the average accuracy of the RFM model (over four testing data sets) in predictingthechurningofcustomersisdiscussed .
<Page:19> erro
<Page:20>  The maximumaccuracyoftheRFMITDSP basedontheF-measure is .97.92 andconsideringthep-value(lowerthan0.05), theaccuraciesofthemethodsdiffer . The boosting version has the highest accuracy, the simpleversionhasthelowestaccuracy, and the .baggingversionissomewhereinbetween .
<Page:21>  Method Version Version Precision Recall Accuracy F-measure F-ANOVA p-value P-value.19,1/2ANN-MLP Boosting 98.15 97.69 97.07 97.92 20.36 0.001 Bagging 95.83 95.39 93.81 95.61 Simple 94.34 92.16 90.55 93.16 Boosting 95.77 94.01 92.83 94.88 56.84 0.01 0.03 P-values for Bagging, Bagging.Simple 88.44 91.7 85.67 90.67 Boosting .
<Page:22> erro
<Page:23>  1/2 95% confidenceintervalofthedifferenceowerLow (cid:4) 27.7520.632.481.892.572.572572.7293.9371.3894.3872.3972.9768.9189.63 .
<Page:24> erro
<Page:25>  Based on the results of the study, compared to SVM and particularly DT, the ANN-19,1/2 method is more power and provides a higher accuracy . This becomes more apparent by adding new and effectivevariables, such as prizes, prizes, discount, the number of purchased items and the distribution date of the items .
<Page:26>  The periodoftheavailabledatawaslimitedtotwoyears. (cid:2) To extract and select the important and effective variables in .encing customers’

Supervised .behaviors, the discriminantanalysis method can beused which isa very accurate machine. and powerfulmethodforpredictingtheclassesofthecustomers. It is recommended that future research works use data with long-term periods of about threeormore years .
<Page:27> erro
<Page:28> erro
<Page:29> erro
<Page:30>  The study was published in the JournalofNucleicAcids, Vol.2012No.2012, available at: machine.2012 . The author of the study is SamiraKhodabandehlou . The study has been published in The Journal of Nucleic Acids, Volume.2012 no.3 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11671/Vijaya et al_2019_Fuzzy Clustering with Ensemble Classification Techniques to Improve the.pdf
<Page:1>  Fuzzy Clustering with Ensemble.Classi cation Techniques to Improve.the Customer Churn Prediction. The.Customer Churn.Customer.CustomerCorrelation (CCM) processhelpstheorganizationin this .retention of customer improves the pro-reward growth of a marketing .
<Page:2>  In this paper, a combination of fuzzy clustering with an ensemble classi cation techniques-based ahybridchurnprediction model is proposed . The model is based on a French-based telecommunication company on customer information is used for analysis . The data set Bose and Chen selected the most important attributes in the data set .
<Page:3>  FuzzyClusteringwithEnsembleClassi cationTechniques… 263.298Tree (DT) classi-classi-classes is evaluated . Researchers used fuzzy-based clustering methods such as PFCM, PFCC-Means, and Possibility C-means . The test dataarepredicted basedonthemajorityvoting,providedby theensembletechniques .
<Page:4>  Preprocessing is carried out using the following procedure . If there is 25% of missingvaluein anattributetheyaretotallyremovedfromthedataset, the data set gets reduced to 67 attributes . The data set is then preprocessed usingmin-maxnormalization .
<Page:5>  FuzzyClusteringwithEnsembleClassi cationTechniques… 265.265. Clusters such that the interdependency between the clusters is very less . There are many clustering methods that exist . We have chosen three different fuzzy-based methods .
<Page:6>  Every cluster formed using Eq.5. These steps are repeated again and again till we.receiveaminimumobjectivefunction whichis.de nedusingEq.6.1 T = (4) The size of the ensemble and the size of an ensemble are the most important factors .
<Page:7>  FuzzyClusteringwithEnsembleClassi cationTechniques… 267. 267. The performance is evaluated in the model . The weight of the miss predictedtuplesisenhanced, and theseboostedtuplesare againfedintothenextclassi aa er for building the model[17].
<Page:8>  The preprocessed data set of the experiment consists of 50,000 samples and 49,000 attributes with one churn attribute . The data set is input into the single classi eroutinemodelandtheperformanceisevaluatedusing the accuracy,TPR, and FPR . The training dataaregiventoensem--ensem and test data are used later in the proposed hybrid model .

<Page:9> FuzzyClusteringwithEnsembleClassi cationTechniques… 269.5 PredictionofVariousOtherApplications . Table6 result shows the performance comparison of the PFCM-PFCM hybrid with boosting produced a maximum accuracy of 97.86% Table6 shows that the model could predict other applications .
<Page:10> P. 270 J.Vijayaetal. 270J. Vijayaetaal. Vijaetal . Vijayetal has published a series of articles on P.J.'s P.A. P.L. A. series of P.I. series tests .
<Page:11> FuzzyClusteringwithEnsembleClassi cationTechniques… 271.2 1 7 0.2 7 9 9RS 82.88.83 83.70.70 94.30.30
"Fuzzy Clustering" with Ensemble Classi.Classi.comcationTechnique… 271
'Crowdsourcing' is a 'crowd', a "rowded" environment, a marketer's "fuzzyclustering.com marketer.com application.
<Page:12> Pending Datasetinformati Datasets#Sa Bank4521marketing Credit69approval Heart30disease Telecom500churn . J.Vijayaetal. 272 J. Vijayetal . Pending is predicted to be 97.3% by the end of the year .
<Page:13> FuzzyClusteringwithEnsembleClassi cationTechniques… 273.1-273.5 Conclusion: effective churn prediction for an organization has become an essential process to withstand its position in the market . The researchers have deployed a.apologetichybridfuzzyclustering with anensembleclassi aa cationmodelfortelecommunication .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11672/Radosavljevik-van der Putten_2013_Preventing Churn in Telecommunications.pdf
<Page:1> Churn, which denotes loss of a client to competitors, is a key problem across industries . New customers are difficult to find, especially in saturated European mobile communications market . Churn/customer churn/retention is typically a marketing based process . But, despite of the involvement of analytics, this process is in its nature reactive, because the customer has already decided to churn .
<Page:2> The research is taking a deep dive into various network problems and their relation to customer churn . Problems with ability to use the network (services) have been identified by internal to the company, as well as in literature . The main objective here is to identify the problems that customers that have churned were experiencing, so that they can be corrected for the current customer base and reduce their likelihood of churn .
<Page:3> Social Networks Analysis (SNA) claims to improve on existing churn models . However, some recent work has demonstrated that this claim is not generally applicable, at least not in prepaid churn prediction on a European market . Most of the literature is using features extracted from Call Detail Records (CDRs)
<Page:4> The results presented here are based on a random sample of 150,000 consumer post-paid subscribers of the operator from September 2012 . The final dataset consisted of 750 features gathered by merging tables from CRM and Network databases . We examined their respective three-month aggre-gates, as well as if there is a rising or declining trend in the past three months for any of these features and use these as potential predictors of churn .
<Page:5> The Coefficient of Concordance (CoC) measure is a rank correla-

protector measure related to Kendall's tau . The CoC measures the grey area in the graph depicted on Figure 1 and  can thus be translated to the Gini coefficient .
<Page:6>  D. Radosavljevik and P. van der Putten used standard algorithms, such as Logistic Regression and Decision Trees based on the CHAID splitting method [33]. These methods fit the explanatory nature of our research, because they are easy to interpret . Each instance is allocated a rank concordant with the probability of being a churner .
<Page:7>  Adding network related features to a campaigning mod-el (Model Campaign_PlusNetwork) only marginally increases performance . PurelyNetworkBased model, which is the topic of our research, has the weak-reviewedest performance . However, campaigning wise, this has no mean-inducinging because rarely do campaigns address more than 40% of the base that is at churn risk .
<Page:8>  Models are to investigate why customers churn from a network pers-uvepective and offer means of alleviating these reasons . When customers get closer to the end of their contract, there is a higher risk of churn . 3G networks reach speed of 21Mbps, while for 2G the maximum speed is only 64 Kbps .
<Page:9>  The influence of quality of Internet ser-vices onto churn is represented via the Number of 2G Data Events and the Ratio of 3G vs. 2G data events . Customers having more than 5 dropped calls in 3 months are 2 times more likely to churn . Projects have been developed to correct these parameters and their respective critical values (increased churn risk)
<Page:10>  D. Radosavljevik and P. van der Putten have presented an atypical approach to churn management in commercial telecoms . They say their findings show that the model explains at least a part of churn via actual measurements of network quality . In the future work, they would like to go one step further, and investigate the benefits of network experience measured directly on the phone, via a preinstalled app, of course .
<Page:11>  The approach can be mir-rored onto fixed telecommunications and potentially into churn in other industries, but  also in many other cases where prevention is more important than the cure, like medical research . We would like to point out the possibility of applying our re-search onto domains other than mobile telecom .
<Page:12> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11673/Verbraken et al_2014_Profit optimizing customer churn prediction with Bayesian network classifiers.pdf
<Page:1>  Customer churn prediction is becoming an increasingly important business analytics problem for telecom opera-tors . Acquiring a new client is six times more costly than retaining an existing customer . Long-term customers generate higher pro ts, tend to be less sensitive to competitive marketing activities, become less costly to serve, and may provide new word-of-mouth .
<Page:2>  A small improvement in customer retention can lead to a signi cant increase in the use of Bayesian Network (BN) classi-networking algorithms . This paper will investigate the predictive power of a numberof BayesianNetwork algorithms . The impact of this variable reduction on network complexity and network complexity is investigated .
<Page:3> erro

<Page:4>  Bayesian network is a probabilistic white-box model consisting of two parts B = (cid:2)G, (Cid:3) G is a directed acyclic graph (DAG) consisting of nodes and arcs . The nodes are the variables X1 to Xn in the data set whereas the arcs indicate directdependencies between the variables . The graph G then encodes the independence relationships in the domain under investigation .
<Page:5>  The NaiveBayesclassi er, which in practice often performs surprisingly well, is the Naive Bayes classi-glyglyglypieriero . The Conditionalprobability forchurning is: 0.0268-0.0024. P(Xi = xi|C = cl) of each variable Xi given the class label cl . A new test case (X1 = x1…,Xn = xn) is then classi aa ed by Bayes'rule to compute the posteriorprobabilities of each classcl .
<Page:6>  The strength of Naive Bayes classi ers inspired severalauthors to develop Augmented NaiveBayesiannetworkclassi.-Bayesian networks . These are methods that relax the constraints of the TAN approach . Not all attributes are dependent on the class node and there does not necessarily need to be an undirected path between two attributes .
<Page:7>  The aim of these classi ers is to  nd a trade-off between the simplicity of the Naive Bayes classi.-generation-based network and the more realistic and complex case of full dependency on the network . The quality of the network is a penalty for the network size .
<Page:8>  The Markov Blanket feature-selection algorithm is used as part of the data preprocessing procedure . Finding the optimal network in such a solution space is known to be an NP-hard problem . The study uses two broad categories of structure learning algorithms . The second category is the Conditional Independence (CI) test .
<Page:9>  Four real life and one synthetic data set will be used to evaluate the performance of the Bayesian network . Hybrid methods have been developed, combining characteristics of both search-and-score and constraints-basedalgorithms . Table 2 summarizes the most important aspects of the study .
<Page:10> erro
<Page:11>  The MarkovBlanket feature selection algorithm has been applied to the data sets at a signi canceance between 1% and 5% . The Bayesian networkconstruction has been used to train and test Bayesian networks using the Markov Blanket feature-selection algorithm .
<Page:12> erro
<Page:13>  A second performance measure that will be applied in this study is the MP criterion . The MP criterion is a measure of the discriminatory power of a classi cation model . The lift indicates the predictive power of the model . Lift can be calculated as the percentageof churners within the fraction of churners . It is logical to evaluateand select a customer.churn prediction model by using the maximumpro t .
<Page:14> erro
<Page:15> erro
<Page:16>  Using AUC as performance measure, the p-value is 0.24, for MP it equals 0.85, both indicating that feature selection does not signi cantly affect performance . The Friedman test is also applied to investigate whether the differences among classi  ers are sig-glygnistic . The outcome of this test is graphically illustrated for both performance metrics in Figs4(a) and(b)
<Page:17>  The complexity and interpretability of the resulting classi ers are

key properties . Bayesian networks are appealing to practitioners, as they give an intuitive insight in the factors driving churn behaviors . Markov Blanket feature selection will be more useful in combination with (Augmented)Naive Bayesclassi er) Bayesian network classi.-Networks .
<Page:18> BayesiannetworkfordatasetD1, createdwithMMHCwithoutpriorfeatureselection, created withMMHC withoutprior featureselection . For logistic regression, which is included in the study as a benchmark, thenumberofnodesisequaltothenumberofattributes,thedimension(ornumber) equal,equal,andthenumberofarcsismeaninglessandthereforediscardedforthisalgorithm . Anexception couldbenoticedfortheTPDAalgorithm where the complexityity .increases for MB.01 .
<Page:19> Figure 6 shows the network created for data set D2 by the MMHC al- algorithm (without prior input selection) The age of the current handset is correlated with the number of months in service and with churn behavior . The study indicates that Augmented.NaiveBayesmethodsdonotleadtocompactnetworks, whereas GeneralBayesianNetworkalgorithms result in simple and interpretable networks .
<Page:20> erro
<Page:21> erro
<Page:22> Apro tdrivendataminingapproach is the approach to predicting customer churn predictions . The author of the European Journal ofOperationalResearch, Inpress, 2011, is T.Verbrakenetal./Pro toptimizingcustomerchurnpredictionwithBayesiannetworkclassi.classi .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11674/Jiang et al_2014_Research on Customers Churn Prediction Model Based on Logistic.pdf
<Page:1> erro
<Page:2> P(y =1|x) expresses the probability of a normal customer, P(Y =0 |x) indicates the probability of churn . P(i=1,(cid:1),k) is coefficient of each variable after logistic regression,Its meaning is whenx change an unit of measurement each caused by the change in the natural logarithm; represents the probability of loosing certain customers .
<Page:3> The prediction model established in this article uses data from a securities company over a year, in the four months from September to December statistical data was collected . The time period used to predict customer behaviors are likely to trigger future marketing response . January to March next year is forecast to month, using the model created to predict customer marketing response .
<Page:4> This paper studies how to use data mining technology to establish customer churn prediction model to solve the problem of customer loss experienced by a securities firm . The constant coefficient of the churn model is 0.2708204106. The model predicts KS value is 44.678, indicating that the model has a strong ability to identify for the churn, so this model prediction is credible .

Your max_length is set to 142, but you input_length is only 139. You might consider decreasing max_length manually, e.g. summarizer('…', max_length=50)

<Page:5> The findings of the study provide countermeasures and recommendations

for effectively improving customer turnover . Based on data mining securities business customer churn analysis, the model was designed to design the model of the customer churn prediction using the Logistic  Regression method . The findings were presented at the University of Dalian Neusoft University of Information .
<Page:6>  David Heckerman, Dan Geiger, and David M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. Machine Learning, 1995, 20(3): 241-243.
 http://www.scientific.net/AMR.989-994.1517 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11675/Ascarza_2018_Retention Futility.pdf
<Page:1>  Eva Ascarza is the Daniel W. Stanton Associate Professor of Business at Columbia Business School . The author bene-formed from comments by Bruce Hardie, Camel Jedidi, Oded Netzer, and the participants of the Choice Symposium session on Customer Re-tention .
<Page:2>  Customers identified as having the highest risk of churning are not necessarily the best targets for proactive churn programs . We propose an approach that, through experimentation, identi es the observed heterogeneity in response to the intervention and targets customers based on their sensitivity to the interventions . We empirically demonstrate that the proposed approach is more e ective than the standard practice of targeting customers with the high-risk risk .
<Page:3>  Churn management is a top priority for most businesses . Churn prediction plays a central role in churn management programs . By predicting churn before it happens, marketers can proactively target activities that are at risk of churning inordertopersuadethemtostay(Neslinetal.2006; Blattberg, Kim, and Neslin 2008)
<Page:4>  Customers' risk of churning does not necessarily relate to their sensitivity to the retention incentive, authors say . Authors propose an approach for proactive churn management that leverages the  rm's capabilities by running a retention pilot, (2) identi es the observed behavior in the response to the intervention, and (3) selects target customers based on their sensitivity .
<Page:5>  In both applications, we identify several variables that highly correlate with being "at risk" but have no relation-ishship with the sensitivity to the intervention . In such cases, if the  rmwere to target based on these variables, they would be directing the resources to customers (for whom the intervention is most harmful and would likely increase churn)
<Page:6>  The issue of customer retention/churn gained traction in the late 1990s and early 2000s . Marketing researchers proposed a variety of methods to predict which customers are at the highest risk of churning . The rationale behind such a practice is straightforward: Targeting customers with the highest propensity to churn enables  rms to focus their e orts on customers who are truly at risk . This enables them to potentially save money that would be wasted in providing incentives to customers who would have stayed regardless .
<Page:7>  Two streams of work have investigated approaches that go beyond targeting those at the highest risk of churning . The cost of misclassifying customers largely depends on the pro tability of each customer . The second

approach, mainly driven by practitioners, has recognized the need to exam-eanine the incremental e ect of the rm's actions rather than merely the behavior incurred by the customer .

<Page:8> The rm is faced with the problem of deciding which customers should be targeted in the next retention campaign . The most common approach in practice is to target the customers who are at the highest risk of churning . In this paper we argue that such a targeting rule is not necessarily optimal . We propose an approach for proactive churn management that overcomes this limitation .

<Page:9> We argue that retention managers should target their retention e orts to customers with highest LIFT . LIFT is the probability that the customer will churn if she is targeted, 0 otherwise . Blattberg et al. (2008) noted that one of the potential concerns of proactive churn management might even encourage "not-would-be churners" to churn .

<Page:10> The need for a retention campaign pilot might seem cumbersome, costly, or di cult for the company to implement . We encourage the rm to run a (small-scale) pilot retention campaign in which the intervention is randomized across a representative sample of customers . We use the experimental data to consistently estimate the heterogeneous treatment e ect .

<Page:11> A variety of methods have been proposed with regards to how to estimate CATE . The main goal those models is to predict which individuals would respond more favorably to an intervention, without focusing on the asymptotic characteristics of the estimates or their interpretation . In a second stream of work, marketingpractitionersandresearchers have developed so-called "uplift" models .

<Page:12> The algorithm proposed by Guelman et al. (2015) with . recursive data splits (in the spirit of Athey and Imbens (2016) to compute treatment e ects and con dence intervals of the treatment e ects . The value of LIFT should be determined how many customers should be targeted and how much resources should be put in place .

<Page:13> We encourage rms to employ A/B testing (or small-scale pilots) not only to evaluate marketing actions but also to identify customer targets . We focus on the incremental e ect of the campaign rather than on the propensity to churn . The proposed approach not only generalizes existing practices for proactive churn management .

<Page:14> We use the experimental set-up to simulate what the impact of these retention campaigns would be had they implemented our approach instead of the standard practice . We replicate the validation exercise for each of the eld studies . We then compare the outcomes across scenarios and quantify the bene ts of following our approach .

<Page:15> Using the observed data from customers in the calibration sample, we estimate a hetero-glyous treatment e ect model using churn as a dependent variable . This model will be used to predict the customers' sensitivity to the marketing intervention . Step 2: Estimate a model for incremental churn (i.e., LIFT model) Using the risk scoring model estimated in Step 3, we predict risk of churning for each customer in the validation sample .

<Page:16> Using the incremental churn model estimated in Step 2, we predict, for each customer in the validation sample, the following quantities: The

probability of churn if not targeted, P(Y = 1|T = 0,X = x ) The probability if targeted, de ned as P(y = 1 |T = 1, X = x) P(J = 1) is the expected incremental e ect of the campaign . LIFT represents the customer's sensitivity to the intervention . We use LIFT to account for this possibility because it is possible for retention campaigns to increase churn .
<Page:17> Decile split is a segmentation method commonly employed by rms (e.g., Bauer, C.L. 1988; Bayer 2010) We choose to model heterogeneity in this fashion not only for its exibility-we do not impose any parametric relationship between the treatment e ect and the level of RISK or LIFT .
<Page:18> We split the validation sample on the basis of predicted LIFT . We calculate the treatment e ect (TE) in each of the groups L , with d = 1…,10.D . Measure TE by group helps identify which groups of customers should and should not be targeted .
<Page:19> We rank customers on the basis of their RISK (descending order) We then estimate the impact of the campaign for each 'target subgroup' by comparing churn rates across experimental conditions . As P increases, the number of customers in each group increases, with P = 100% corresponding to the rm targeting the whole customer base .
<Page:20> Study 1: Wireless service (Middle East) conducted an experiment to test whether giving customers free credit when recharging their amounts a ected their likelihood to remain active . Treated customers (68% of the sample) received a text o ering additional credit if they were treated .
<Page:21> The company tracks multiple measures of activity such as texts, calls, data uploads/downloads and recharges . The company then tracked whether the customers were active (or inactive) 30 days after the experiment . With the exception of one variable(voicevolume), all other variablesare not statistically di-erent across conditions .
<Page:22> The focal organization ran a eld experiment that tested whether adding a gift to the renewal communication would increase renewal rates . The intervention was not targeted to any speci c type of customer . At the end of the experiment, we obtained all the information from a random sample of the customers involved in this experiment (N= 2,100). Table 1 describes the observed variables by group .
<Page:23> Figure 2a shows churn rate of customers in each of the experimental conditions of Study 1 (wireless) and Study 2 (membership) Study 2c and 2d examine churn rates for di erent levels of RISK and LIFT in Study 2a . Figure 2b shows the extent to which treatment reduced churn dramatically in each RISK-group .
<Page:24> Figure 3 shows the magnitude of the treatment e ects, TE and TE for each of the empirical applications . Cus-tomers with the highest levels of LIFT (L -L ) respond positively to the treatment-churn-rate churn rates are about 5 percentage points lower for treated customers than for control customers .
<Page:25> Figure 4depicts the impact of the retention campaign if targeting based on RISK or LIFT . The impact of targeting customers based on LIFT decreases as the percentage of customers being targeted increases . The LIFT approach selects the "best" (i.e., more sensitive) customers, therefore the e ectiveness of the campaign should decrease .
<Page:26> The company would have increased churn by 4.4 percentage points if

targeting the 40% of customers with highest risk of churning (top RISK) and those that are most sensitive to the retention intervention (top LIFT) Both methods would give similar e ectiveness if the company decided to target most customers .
<Page:27> Relationship between RISK and LIFT metrics is rather weak . In Study 1, among the 10% of customers with highest RISK, only 16% of them also belong to the top 10% LIFT group . Figure 2c shows that in Study 2, the level of overlap between the two metrics is not only weak but negative .
<Page:28> Study 2 looked at relationship between the variables 'Tenure' and 'Last recharge' with RISK and LIFT . The variable 'Data volume' reveals an interesting pattern. It suggests that if the company decided to send a retention incentive to customers with low data consumption, such a campaign would likely increase churn .
<Page:29> Targeting based on LIFT is more effective at reducing customer churn than targeting on the basis of RISK . The same retention campaign would result in a further reduction of 4.1 and 8.7 percentage points in churn rate . This result is consistent across both studies representing two business settings (wireless/telecom and special-interest organization)
<Page:30> The magnitude of the impact of the campaign does not depend on where each customer is relocated . The lack of relationship between RISK and LIFT is not due to the selection of customers eligible for the research . Future research should investigate these relationships in the interest of better designing incentives for retention campaigns .
<Page:31> Proactive churn management programs have been mainly applied to contractual settings (e.g., telecommu-nications, nancial services, utilities, memberships) Noncontractual settings can also leverage our proposed approach to select targets in their marketing campaigns .
<Page:32> It is important to note that churn (or customer retention) is only one measure of interestin the customer relationship . In many business contexts, other behaviors (e.g., consump-tion) are also important determinants of the value of a customer . In some settings, some customers will be more valuable than others even if they all had the same churn propensity .
<Page:33> The real challenge of estimating Value-LIFT is that one needs a very long time horizon to estimate the impact of the marketing intervention . In order to simplify the expression of CLV, most past work in marketing has assumed constant margins and retention probabilities . However, the main purpose of a marketing campaignistoalter theprobability thatacustomer will renew, making assumption about constant retention rates problematic .
<Page:34> Proactive churn management programs should not necessarily be targeted to customers who are at the highest risk of churning . Instead, they should target only customers whose propensity to churn will decrease in response to the intervention . We show that the same campaign would reduce churn by an additional 4.1 and 8.7 percentage points relative to the standard practice of targeting customers at highest risk .
<Page:35> From our research: (1) the company observes customer behavior at the individual level, and (2) the rm is capable to interact with customers in a one-on-one basis (i.e., they can run individually-targeted campaigns) Examples of these business contexts include credit card companies, softwareproviders,

onlineando inesubscriptions and leisure memberships .
<Page:36> An ideal scenario would be to analyze the case of the same company testing di-erentincentives . Longer assessment-periods would allow the researcher to measure long-term e-centric incentives and potentially identify the best targeting rules for optimizing out-costs . We anticipate/speculate that the proposed approach is bene cial regardless of the churn rate .
<Page:37> An optimal size for a retention pilot would be small, but a smaller sample size might be better . How stable (over time) is the heterogeneity in sensitivity to the retention action? We hope that future research will address these and other related issues . There are notobvious reasons why the relationship between the correlation and the sensitivity of the intervention would change over time, it would useful to empirically investigate this question .
<Page:38> Accenture Analytics (2014) Nordic Telco: Analytics Help Reduce Churn and Improve Mar-
ishlyketingCampaigns(accessedJuly13,2017),https://www.accenture.com/us-en/success-nordic-telco-analytics-marketing-campaigns .
<Page:39> Researchers at AmericanExpress, CMO Cameron N. Bolton, Ruth N. Lemon and Katherine . Bolton (1998), A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. (1998) Breiman, L. Breiman (2001), Random Forests Machine Learning 45, 5-32 .
<Page:40> The author of ManagingChurntoMaximizePro ts. Available at SSRN:2964906 . The author also discusses the impact of data mining techniques in customer relationship management. The study was published in the journal of Marketing Research, published in October 2017, and published in September 2017 .
<Page:41> Piotr Rzepakowski, Michal Sołtys and Szymon Jaroszewicz (2015), Ensemble methods for uplift modeling. The Free Encyclopedia, s.v. "Customer Attrition," (accessed May 12, 2017), "Customers: Who-Are They and What Will They Do Next?.
<Page:42> Study 1: Wireless provider Control Treatment (N = 3,857) p-value: Tenure 0.002 -0.002 0.001 0.881; Days since last recharge 0.015 0.007 0.008 0.003 0.625; Revenue from last recharge is 0.013 0.006 0.317; Data volume last two weeks (in logs) 0.043 -0.043; voice volume last 2 weeks 0.058 0.298 0.527; SMS volume 0.017 0.020 0.200.
<Page:43> We compare di erences in churn rates across customer groups by levels of churn propensity (i.e., RISK) and levels of LIFT (LIFT) By levels of RISK (d) We compare churn rates when targeting customers with levels of sensitivity to the return-to-concentration intervention . Figure 1: Heterogeneity in treatment e-criticism is based on the type of customer churn rates .
<Page:44> Figure 3: Treatment e ect (TE) by deciles depending on whether customers are grouped by levels of RISK or LIFT . The dotted (straight) line corresponds to the aver-repregeage e rm targeted randomly .
<Page:45> Figure 4: Impact of the campaign (IC) under di erent scenarios . The dotted (straight) line corresponds to the impact of the . campaign if all customers were targeted . Fewer customers are more likely to be affected by the campaign than those targeted . The RISK assumes the com-pany targets customers with higher levels of risk of churning .
<Page:46> Figure 5: Level of overlap across groups de ned by top RISK deciles

vs. top LIFT deciles . The (dotted) 45∘ line represents the level of overlap if there was no relationship between the two groups . Figure 5 shows the overlap between RISK and LIFT groups .
<Page:47> Study 1: Average levels of each observed variable by levels of LIFT and RISK . Figure 6: Average value of each of the observed characteristics for each decile(R,R,…,R and L ,…,L) Figure 6 shows the average value each observed characteristics is based on the characteristics of a given decile .
<Page:48> An algorithm grows an ensemble of trees, each of them built on a (random) fraction of the data . Each tree is grown by randomly selecting a number of variables (among all the available variables) for splitting criteria . The trees grow as follows: First, the split rule is chosen to maximize a measure of measure of the divergence on the treatment e ect (Rzepakowski and Jaroszewicz 2012) Second, each tree will keep growing until the average divergence among the (resulting) subtrees is smaller than the divergence of the parent node .
<Page:49> The R code used for the empirical application is made available as a supplemental R code . To select the best RISK model we perform a 10-fold cross-validation process . The (out-of-sample) model increased as the number of trees increased, with a marginal improvement after having reached 80-100 trees .
<Page:50> As metric for accuracy we use the area under the curve(AUC) of the receiver operating characteristics (ROC) The best performing method was the LASSO approach combined with a GLM model, which provides an AUC of 0.907 for the rst empirical application and 0.658 for the second application .
<Page:51> We evaluate the e ect of the retention campaign by deciles of RISK (Step 5) We then compare the e gures of retention campaign for the RISK with the LIFT (as obtained in the main manuscript) As the results show, the resultsremain unchanged, verifying that the superiority of LIFT approach is not driven by the differences in sample size when calibrating models .
<Page:52> The impact of treatment e ect (TE) and impact of the campaign (IC) results are based on the results of the study . The study used the full sample to calibrate the RISK model using RISK or LIFT . The results were compared with the results from the study of a similar study using the ICIC . The ICIC was used to replicate the study and compare the results to the results published in Table 1 .
<Page:53> Using the same model approach (i.e., random forest) to estimate RISK and LIFT, we also replicated the analysis by using the RISK estimates from the best-performing random forest . The rationale behind this analysis was to estimate both RISK. and Lift using the same modeling approach . We recreate the findings appearing in the main manuscript corresponding to the heterogeneity in treatment e ect .
<Page:54> The study used random forest to estimate both RISK and LIFT levels . The impact of the campaign (IC) and the level of overlap across groups de ned by top RISK deciles vs. top LIFT deciles . The study compared treatment e ect (TE) for di erent group (b) and impactofthecampaignunderdi aeerentsce-relateddeciles .
<Page:55> Figure A3: [Study 2] Replication of treatment e ect (TE), impact of the campaign (IC), and overlap results using random forest to estimate both RISK and LIFT . (b) Impactofthecampaignunderdi ned by top RISK deciles vs. top LIFT deciles .

<Page:56> Figure A4 shows that predicted LIFT (green circles) estimates the magnitude to actual LIFT . Not surprisingly, the intervals around those estimates are wider for the actual data than for the estimates . A3 Additional analyses/results: A3.1 Predicted vs. acual LIFT and a comparison of LIFT by comparing, bydecile, the average LIFT with the magnitude of the treatment that is predicted by the causal uplift model .
<Page:57> Figure A4: Predicted vs. actual LIFT. Green (circles) represent the average pre-depicted LIFT, representing the expected treatment e ect in each decile . Blue (square) representt the (actual) average treatment e-treate in each (decile) Study 1.0.18.2.3.4 .
<Page:58> In appendix we show the results for one single iteration . Figure A5 corresponds to Figures 2a, 2b, 3a and 4a from the main manuscript . We observe that all patterns of the results are very similar to those obtained when they are aggregated across iterations .
<Page:59> Customers groups determined by levels of RISK or LIFT . Customers grouped by LIFT decile . If targeting customers on TOP # decile, group deciles der di erent scenarios would be different . Figure A5: [Study 1] [Study 2] Analysis of churn rates, treatment e aect (TE) and impact of the campaign (IC)
<Page:60> A3.3 Di erences between customers' RISK and LIFT (results for all variables) We only discussed the most relevant variables for each application . The second application has 50 variables (consisting on the variables described in the main manuscript) and the third application is 50 variables .
<Page:61> Var 1 Var 2 Var 3 Var 4 Var 5: LIFT LIFT: Method LIFT, LIFT and LIFT deciles as a function . LIFT is a function that determines whether or not a function should be called LIFT . Method: Method lIFT lIFT, lIFT and lIFT decile; LIFT provides a function with a function of LIFT lags, lags and lags . Method lags: lags; lads; lags are lads, lads. lads . lads are lags with lads and lads with lumps; lumps are lumps with lags. lumps lads lags lads
<Page:62> Figure A7: [Study 2] Observed characteristics (variables 1-25) as a function of LIFT and the LIFT deciles . The characteristics of these characteristics were defined by LIFT, LIFT or the LISK deciles, with LIGEL and LOGEL deciles of LAGEL .
<Page:63> Figure A8: [Study 2] Observed characteristics (variables 25-50) as a function of LIFT and the LIFT deciles . Var 26 Var 27 Var 27 and 28 Var 28 Var 29 Var 29 and 29 Var 30
varieties 25-50 were defined by LIFT, RISK deciles and LISK deciles .
<Page:64> The probability that a customer will churn in the next renewal can be altered if the person receives an incentive . Customers have an intrinsic propensity to churn (i.e., RISK) that is heterogenous across the population . Customers are also heterogeneous in the way the way they respond to the incentive .
<Page:65> The term  captures correlation between X and Z, which we vary from -1 to 1 . Figure A9 shows the level of overlap for all levels of  . Comparing the results from the second application (special interest membership) the correlation between RISK and LIFT is clearly negative .
<Page:66> Figure A9: Level of overlap across groups de ned by top RISK deciles vs. top LIFT LIF deciles . The (dotted) 45∘ line represents the level of overlap if there was no relationship between the two groups .

<Page:67>  Figure A10: Treatment e ect (TE) for di areeerent group deciles, depending on whether cus-reparative-group groups are grouped by levels of RISK or LIFT . The dotted (straight) line corresponds to the average (average) e  rm targeted e RAF) of the campaign if the ection  of the group was targeted randomly .
<Page:68>  Figure A11 shows the e ect of the campaign, measured as revenues between control and treated customers, by levels of Value versus Value-LIFT . The results corroborate the claim that companies would notably improve the impact of their campaigns by targeting customers with highest value . We also quantify what the overall impact would have to be if the company targeted top 10% value customers .
<Page:69>  Customers are grouped by levels of Value or Value-LIFT (represented by thecircles) Customers are then grouped by Levels of Value (repre-repre-privilegesentedbythesquares) orValue-Lift (representedbythecircle) Thedotted (straight)linecorrespondstotheimpactofthecampaignifallcustomerswere (I 8//g 6/g) 6//Groups) 10//Group10Group9Group9/Group8/Group7Group7/Group6Group5Group5/Group3/Group1 .
<Page:70>  Simulation results for di erent churn rates are 44% and 62% . LIFT approach identi es those customers that the  rm should give priority . As the churn rate decreases (from 50% down to 5%), the bene-a-t of using LIFT vs RISK decreases on average .
<Page:71>  Figure A12: Treatment e ect (TE) for simulated data . Varing churn rate from 50% (Top left) to 5% (Bottom left) Varing . churn rate of churn rate between 50% and 5% of simulated data. 50% churn rate 25% churn.
<Page:72>  Researchers at the University of Malaga, Spain, used data mining techniques to develop new models for data mining . The results of the study were published in 2011 and 2012 . The study was presented at the 2010 IEEE 10th International Conference on I-Mining (ICDM), 2010 .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11676/Hutchison et al_2010_Rule Extraction from Support Vector Machine Using Modified Active Learning.pdf
<Page:1>  The problem of customers shifting loyalties from one organization to another is called "churn" and is common nowadays . Customer Relationship Management (CRM) is becoming more customer centric . Despite superior generalization performance Support vector machines generate black box models . This paper presents a new approach for rule extraction from SVMs using modified active learning based approach (mALBA) to predict churn in bank credit cards .
<Page:2>  Research shows that, the customers  with longer time relationship with the firm are more profitable [9, 10] than online  bank customers [11]. Management should prepare an anti-churn strategy that is usu-glyally far less expensive than acquiring new customers [12, 13].
<Page:3>  Proposed ap-proach is applied to predict churn in bank credit cards . We propose a modified active learning based rule extraction procedure to extract rules from SVM using NBTree (Naive Bayes Tree) We have chosen an RBF kernel for developing SVM model, as it is shown to achieve good overall performance .
<Page:4>  The proposed approach is depicted in Fig. 1 . The current study in this paper is different from ALBA [41] approach in several ways, such as; . Generated data is then appended to the                 support vectors set and

the predictions are obtained . The actual target values are  then replaced by
the predictions of SVM . This modified data is fed to NBTree  to generate rules
. The dataset is from a Latin American bank that suffered from an increasing
number of  grotesquechurns .
<Page:5>  The efficiency and validity of the rules generated under 10-FCV are
then tested against the validation set, which is a subset of the original data .
The quantities employed to measure the quality of the quality are sensitivity,
specificity and accuracy [49]. We used the SVM  for the churn prediction data
set .
<Page:6> erro
<Page:7>  The fidelity obtained using ALBA, ALBA with SVs and mALBA is presented
in Table 4 . It is observed that ALBA behaves 83.28% like SVM with 500 generated
samples . Proposed mALba approach mimics the behavior of SVM better than ALBA .
The number of rules ex-gouged using mAlBA is very much less in number when
compared to the rules extracted  using ALba .
<Page:8>  In this paper, we present a modified active learning based approach
for rule extraction  from SVM to solve credit card customer churn prediction
problem . The dataset is taken from Business Intelligence Cup organized by
University of Chile in 2004 . It is highly unbalanced data with 93% good
customers and 7% churned customers . The proposed rule extraction approach using
mALBA yielded the best sensitivity of 79.35% .
<Page:9> erro
<Page:10> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/fin
al_dataset/files/11679/Martono et al_2014_Utilizing Customers' Purchase and
Contract Renewal Details to Predict Defection.pdf
<Page:1>  Study aims to predict customer defection in the growing market  of the
cloud software industry . Cloud software market's 36% compound annual growth is
predicted to continue through 2016 . Defection prediction has been a concern in
research and industry, as it is an important measure used to retain customers .
<Page:2>  Study: Utilizing Customers' Purchase and Contract Renewal Details to
Predict Defection in one security software company . Algorithm provides an
algorithm to detect which customers are defecting from the company and which are
not . Data features are limited and include only a few customer  Attributes,
unlike several previous works on defection prediction .
<Page:3>  140 N.P. Martono, K. Kanamori, and H. Ohwada  criticize machine-
learning techniques to pre-dicting customer defection . We use two types of
data: purchase and auto-renewal data and web log data . The data is originally
used to record the details of "opting-in" and "opt-out" ac-tivities of each
customer after receiving e-mail notification of auto renewal .
<Page:4>  The main purpose of data preparation in this study is to determine
whether the original data may be used in developing a customer defection
prediction model . Table 1 illustrates the original contents of the table that
contains historical records  of customer activity collected from the company's
e-commerce site . It contains  the ID number of a purchase or renewal that a
customer  makes .
<Page:5> erro
<Page:6> erro

<Page:7>  Higher values of mean decrease in accuracy indicate variables that are more important to the classi cation . Table 4 gives the number of samples used by a random forest to obtain the importance of each variable on each customer segment . UPDATE_COUNT was identified as the  most important variable for classification using purchase and auto-renewal data .

<Page:8>  R package supports the process of interpretation by providing tree visualization and tree rules . Decision tree results make it easier for the company or other end user to determine the next action for retaining the customer . We present an example of the visualization of customer defection prediction in the Low Price customer segment using purchase and auto-insured data .

<Page:9>  The decision tree obtained a model that uses UPDATE_COUNT and total grotesquepayment or CC_PRODUCT_PRICE is the most powerful predictor . 35% of customers who have the attributes of . worrisomeUPDATE_Count less than 2.5 have a 98% probability of defecting . The status of e-mail delivery appears to be one of the three pre-dictors resulting in predictive accuracy .

<Page:10>  C4.5 decision tree based on the purchase and auto-renewal data . F-score F-Score F- Score is based on accuracy and precision of the data set . Figure 4: Utilizing Customers' Purchase and Contract Renewal Details to Predict Defection .

<Page:11>  Using machine learning, we identified important variables for classifying defecting customers . We built a prediction model of customer defection using both purchase and auto-renewal data and web log data . Future work will seek to integrate machine learning with a more dynamic approach, such as agent-based modeling .

<Page:12>  Huang, B.Q.Q., Kechadi, M.-T., Buckley, Buckley, B.: Customer Churn Prediction for Broadband    Internet Services . Wei, C., Chiu, I.: Turning telecommunications call detail to churn prediction: A data min-forming approach . Wei: Applying data mining to telecom churn management .

/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11680/Amin et al_2015_Churn Prediction in Telecommunication Industry Using Rough Set Approach.pdf

<Page:1>  Customer churn is a crucial activity in rapidly growing and competitive telecommunication sector . Due to the high cost of acquiring new customers, cus-tomer churn prediction has emerged as an indispensable part of telecom sectors' strategic decision making and planning process . This study makes use of rough set theory, a rule-based decision making technique, to extract rules for churn pre-diction .

<Page:2>  84 A. Amin et al.  discuss customer churn prediction in telecommunication sector . The paper is organized as follows; the next section presents customer churn and related prediction modelling . It should be the goal of the decision maker and marketers to de-ishlycrease the churn ratio because existing customers are the most valuable assets for companies .

<Page:3>  Churn prediction has received a tremendous focus from both types of researchers . The literature shows that various machine learning techniques for churn prediction in the telecom industry has been used such as neural network [2]. SVM is one of the state-of-the-art technique for  classification due to its ability of model nonlinearities but the main drawback is no-yet-iced that it

generates black-box model .
<Page:4>  The Rough Set Theory was originally proposed by Pawlak [32] in 1982 . The special case of Information system (IS) is known as a decision table . The reduction process is finding more important attributes that preserve discernibility relation with the information . The core is  the intersection of all reducts .
<Page:5>  Decision rules can be constructed by overlaying the reduct sets over the decision  table . The following measures were used for the evaluation of proposed classifiers and approaches . It is nearly impossible to build a perfect classifier or a model that could perfectly cha-uveracterize all the instances of the test set .
<Page:6> erro
<Page:7>  The selection of most appropriate attributes from the dataset was carried out using feature ranking method titled as ‘Information Gain Attribute Evalua-ogletor’, using an WEKA toolkit . Table 1 describes the selected attributes which are also  addressed to P1.2 . Table 2 describes the selection of objects, condi-utictional attributes and decision attribute are organized in Table 2. The preparation of decision table is an important stage of the proposed study .
<Page:8>  Some data are excluded from the training set as it begins the process to train the classifier . When the training process is completed, then excluded data can be used to validate the performance of the learned classifier on new data . Decision rules can be obtained from training set by selecting either of the methods (Exhaustive, Genetic, Covering and LEM2) The decision rules set specifies the rules in the form of "if C  then D" where C is a condition and D refers to decision attribute .
<Page:9>  The number of churns is much smaller as compared to non-churns customers in the selected dataset . Table 4 reflects that genetic algorithm performed better in term of obtaining 98% accuracy, 100% False churn and 98% true churn prediction along with coverage of all instances . We have evaluated four different algorithms for rules generation through with rough . set based classification approach using RSES toolkit .
<Page:10>  The proposed approach performs very well as compared to the previously applied techniques . Figure 1 shows the point of inflection of various variables such as CustServ_Call,  Intl_Charges,  Eve_Charges and Day_Charges . The churn rate is high in those features  which are above the curve except Intl_Calls, VMail_Messages and VMails_Plan .
<Page:11>  Churn prediction has emerged as an indispensable part of strategic decision making and planning process . This study is approaching to explore the powerful applications of rough set theory for churn prediction in telecommunication sector . The study also investigated the performances of four different algorithms (Exhaustive,                 Genetic, Covering, and LEM2) of rules generation .
<Page:12> erro
<Page:13>  Churn Prediction in Telecommunication Industry Using Rough Set Approach  using rough set approach . Rough Set Algorithms  in Classification Problem, pp. 49-88. H.S., H.H., Nguyen, S.H.: Analysis of stulong data by rough set exploration system .

/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11681/Saravanan-Vijay Raajaa_2012_A Graph-Based Churn Prediction Model for Mobile Telecom Networks.pdf
<Page:1>  Churn in the telecom industry refers to the movement of customers from one operator network to the other . With the continuous addition of new operators in the  market, churners are increasing at a higher rate than before . Churn being a predictive model, there is no generalized scalable approach to capture the probable churners effectively in the data .
<Page:2>  The churn data tends to be imbalanced be-cause the churners tend to be far less in number in the order of (2% - 5%) compared  to the non-churners . The graph parameters considered for node level analysis are as follows: In-Degree, Closeness centrality, Proximity prestige, Eccentricity cen-idatedtrality, Clustering coefficient, In Degree and Out degree prestige .
<Page:3>  A Graph-Based Churn Prediction Model for Mobile Telecom Networks is
369                  scalable and a dynamic platform with ease and minimal cost . The telecom service providers can use the proposed model in identifying churners efficiently on a  streaming environment and in launching retention campaign based on their priorities . Existing models for churn prediction pertains to supervised and semi supervised methods .
<Page:4>  The accuracy can be improved by predicting the non-consuming churners with a high degree of correctness . Churn is a specific business case where the telecom carrier would like to identify chunks of users who are likely to churn . We have analyzed the call graph properties specific to customer churn on a telecom domain . The structural properties of call graph are calculated for every node in the network .
<Page:5>  The call graph G is generated by ingesting the CDRs to create (V(G), E(G) pairs . The graph parameters considered for the node level analyanalysis are described here . Fig 1 illu-ishlystrates the nodes with specific graph parameter measures .
<Page:6> erro
<Page:7>  The overall system throws light on a novel way of churn prediction with ease and minimal cost . The data is visualized as a call graph which consists of vertices and edges based on the activities of individual customers in the network . Specific graph parameters are chosen by employing two different multivariate methods that  contribute more to extract churn behavior . We arrive at a linear model with  more specific graph parameters using logistic regression to be employed for probable
<Page:8>  The dataset generated in the telecom industry tends to be of huge size and hence processing those takes a lot of computational time . We employ Hadoop-based Map Reduce framework to preprocess the CDR data by converging them to location-wise details and use them for graph generation and parameter computations . We discuss the influence exerted by a customer within a network based on the  game theoretic centrality approach implemented using the Shapley value .
<Page:9>  A graph-based Churn Prediction Model for Mobile Telecom Networks is based on the CDR data set . We use a linear discriminant analysis to determine which attribute discriminates between two or more naturally occurring groups [23] The logistic regression is a useful way of describing the relationship

between the extracted graph parameters and for predicting the churners .
<Page:10> The graph parameters are examined over three different datasets obtained from the telecom service providers of two different countries . The first dataset corres-ponds to a rural base whereas the second one corresponds to an urban region of a particular country . The third set corresponds to a data from a developed country . Figure 3 illustrates the windowing frame used to analyze the churn behavior over a pe-glyod of time .
<Page:11> The idea of selecting specific graph parameters contributing effectively to the churn behavior can be achieved by running the machine learning algorithms such as Logistic regression and Multivariate Discriminant analysis . These machine learning approaches are used to highlight a specific list of graph parameters that contribute significantly for discriminating churners from non-churners .
<Page:12> Discrimination between churners and non-churners is clearly visible based on se-oglelected graph parameters . Figure 5 illustrates the discrimination of churners based on the se-privacy graph parameters derived from the multivariate discriminant and logistic re-gression models . The drop in certain graph parameters shows that churners are slowly losing interest in using the corresponding network .
<Page:13> The proposed linear model is used for predicting the probable churners in a dynamic environment . We propose that when there is significant variation in the graph parameters as a whole then there is a high probability that the customer will churn out . The model was tested for three different datasets over a period of three month time scale .
<Page:14> Graph-based analysis for churn prediction is a novel idea proposed for efficient churn prediction in the telecom domain . The maximum accuracy reached in our previous model using hybrid learning is grotesque72.18% for the same dataset used in this study [22] The maximum accurate churn prediction was 81.67% . The graph-based visualization aids in better understanding of the behavior of the customers .
<Page:15> erro
<Page:16> Evolutionary Churn Prediction in Mobile Networks using Hybrid Learning . 382 M. Saravanan and G.S. Vijay Raajaa . Pohar, M., Blas, M. Turk, S.: Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study .
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11682/Farquad et al_2012_Analytical CRM in banking and finance using SVM.pdf
<Page:1> The proposed approach comprises of three major phases: feature selection using SVM-RFE (recursive feature elimination) active learning for synthetic data generation and rule generation using decision tree (DT) and Naive Bayes tree (NBTree) Problems solved in this study are churn prediction in bank credit cards and fraud detection in insurance .
<Page:2> M.A. Farquad, V.V. Ravi and S.B. Bapi Raju write about CRM in banking and finance using SVM . They argue that SVM is a way to learn about customers in order to develop stronger relationship with them . The extracted rules serves as early warning system to the management to enforce better CRM practices and detect/avoid possible frauds .

<Page:3> CRM is a combination of business process and information technology to discover the knowledge about the customers . CRM can deliver better, timely and personally customised solutions to the customers' problems, thereby enhancing customer satisfaction . We propose modified active learning-based approach to extract rules from the trained SVM model by making use of key concepts like the support vectors .
<Page:4> The paper is structured as follows: In Section 2 related works of rule extraction from SVM are presented . Section 3 provides the details about two most important applications that are analysed in the current study . In Section 4, the proposed eclectic rule ipient extraction approach is presented in detail .
<Page:5> The problem of customers shifting loyalties from one organisation to another is called 'churn', and is common nowadays, which motivated the service industries like banks and insurance to provide better services to their customers . Churn occurs due to various reasons, such as availability of latest technology at the organisation, customer-friendly staff and proximity of geographical location .
<Page:6> The original dataset has six numerical attributes and 25 categorical attributes, including the binary class label (fraud or legal) Prior to its analysis, pre-processing is carried out to make this dataset feasible for this research study . This dataset contains 11,338 records from January 1994 to December 1995 . It has a 6% fraudulent and 94% apologetic legitimate instances with an average of 430 claims per month .
<Page:7> The age attribute in the dataset appeared twice in numerical and categorical form as well (Attributes 12 and 24 in Table 2) Age of vehicle, age of policy holder and fault were removed from the data to reduce the complexity caused by too many unique values it possesses . The attributes year, month, week of month and day of week represent the date of the accident ( Attributes 7, 1, 2 and 3) and the retrieve the data .
<Page:8> The attribute gap represents the time difference between the accident and insurance claim . The dataset consists of 14,497 instances representing the behaviour of legitimate customers, whereas only 923 instances represent fraudulent customers . We have 15,420 samples with 24 predictor variables and 1 class variable . The data is highly unbalanced with 94% legitimate instances and 6% fraudulent .
<Page:9> The proposed approach comprises three phases; feature selection, active learning, and rule generation phase . The architecture of the proposed approach is depicted in Figure 1.1 . The proposed approach in this article is advancement to the earlier study of Farquad et al. (2010b) They used uniform distribution to generate extra instances near support vectors based on distance between the training instances and support vectors .
<Page:10> Step 3: Randomly generate an extra data instance x following uniform distribution [-1, 1] Step 4: Provide a class label y using the trained SVM as oracle . For generating extra instances near support vectors, we employed Normal and Logistic distribution function separately i.e. Normal distribution function (Box Muller approach)
<Page:11> The experimental setup followed in this paper is depicted in Figure 2 . 80% of the data is then used for training under ten fold cross-validation

method using stratified random sampling . 20% data is stored untouched for validation purpose later . This 20% data represents the reality and originality present in the original data used for evaluating the efficiency of the rules .
<Page:12> Using available training data, SVM model is first developed under 10-FCV and support vectors are extracted for each fold . The distance between the support vectors and training instances is then calculated before generating the extra instances . 500 and 1,000 extra data instances are generated for empirical analysis in the present study .
<Page:13> erro
<Page:14> erro
<Page:15> erro
<Page:16> erro
<Page:17> Average fidelity for insurance fraud detection using SVM + NBTree is based on the best feature selection table . Table 14 shows the average fidelity of the rule induction algorithm . Table 15 shows the most popular rule induction algorithms . Table 21 includes average fidelity table and table of rule set set and rule set .
<Page:18> erro
<Page:19> erro
<Page:20> erro
<Page:21> Average fidelity for insurance fraud detection using feature selection + SVM + DT (1,000 extra instances) is 1,000 times higher than the average fidelity of an insurance fraud test . M.A.H. Farquad et al. Far Quad Quadrquad and M.J. A.H Farquad compared the results of a test using a feature selection and DT test to detect insurance fraud using features .
<Page:22> Rules extracted for insurance fraud detection using DT are analysed using SVM . It is observed from empirical results that the hybrids with extra instances combined with support vector set perform better than the original ALBA of Martenes et al. (2009) It is also observed that the time taken and the number of rules extracted using proposed rule extraction approach is very much less .
<Page:23> M.A.H. Farquad et al. published a novel and extended modified ALBA for rule extraction from SVM . The proposed approach mines two unbalance medium scale data mining problems such as; churn ypeypeprediction in bank credit cards customers and Insurance fraud detection .
<Page:24> erro
<Page:25> erro
<Page:26> erro
/mnt/c/nuvem/Dropbox/doutoramento/tese/SLRDropout/analysis/selected_articles/final_dataset/files/11683/Kianmehr-Alhajj_2011_A fuzzy prediction model for calling communities.pdf
<Page:1> Keivan Kianmehr* and Reda Alhajj have published over 30 papers in prestigious venues . Identifying calling communities can be used to determine a particular customer's value according to the general pattern of behaviour of the community that the customer belongs to . This helps in resolving issues like predicting terrorist groups .
<Page:2> Reda Alhajj is a Professor in the Department of Computer Engineering at the University of Calgary, Alberta, Canada . He has published over 280 papers

. His research focuses on machine learning techniques, social networks, XML, data mining  methods and bioinformatics . Identifying social communities is an emerging research area that has already attracted the attention of several research groups .
<Page:3>  A fuzzy prediction model for calling communities has been proposed by Yang et al. It is a new algorithm, recently proposed, to mine signed social networks where both  positive  within-group  relations  and negative  negative relationships are dense . FEC considers both the sign and the density of relations as the  'clustering' attributes, making it effective  for  not  only signed networks  but  also  conventional social networks including only positive relations . FEC adopts an ipientagent-based heuristic that makes the algorithm efficient and capable of giving nearly optimal solutions .
<Page:4>  The ability to dynamically classify and predict customers' calling patterns according to their calling patterns obtained from CDR data has attracted considerable attention in the research community . Data mining is applied in this area to perform two  major tasks: predicting whether a particular customer will churn and when this will happen . This helps the effective targeted marketing design which is significantly important for increasing profitability in the telecommunication industry .
<Page:5>  The Support Vector Machine (SVM) as a statistical-based learning approach has been used to build the classifier model . A fuzzy genetic algorithm has been also applied for the classification task . Fuzziness is attractive because it facilitates the possibility of having partial membership in a given group .
<Page:6>  The CDR data used in this work was given by a telecommunication company providing wireless services . The majority of the destination numbers are outside of the service provider's network . The other type of phone numbers are those in the service providers' network, and each of them corresponds to a customer . The data preprocessing, clustering, and classifier model is able to assign a new customer to one community .
<Page:7>  The given data set consists of 55 000 calling records of 2000 subscribers . Calls with very low duration (less than 5 sec) are assumed to have no effect on identifying the subscriber's neighbours and are ignored . Clustering refers to the process of partitioning a set of data points into a meaningful sub-classes called clusters .
<Page:8>  K. Kianmehr and R. Alhajj used the agglomerative hierarchical clustering algorithm to discover calling communities . The algorithm is a bottom up clustering approach that investigates grouping in the given data by creating a cluster tree according to a particular distance ggiemeasure . The output is a tree that represents a multi-level hierarchy, where clusters at one level are grouped together to form clusters at the next higher level .
<Page:9>  The MATLAB Statistics Toolbox has been used for conducting the hierarchical clustering . The basic procedure to perform hierarchical clustering in our model is as follows: Find the similarity and dissimilarity (including both first and second order distances) of customers in the CDR data .
<Page:10>  The first approach is SVM from the family of statistical-based learning algorithms . The goal of SVM is to construct a separating hyperplane that is maximally distant from different classes of the training data . Then, a

set of feature vectors, each of which corresponds to a specific customer, is used as the training set for the classification algorithm .
<Page:11> erro
<Page:12> Genetic algorithms have been successfully used for the search and optimisation of problems . John Holland pioneered genetic algorithms as initiated with cellular automata . A genetic algorithm uses an evolutionary approach to find optimal solutions . The population evolves at every generation by using a Darwinian approach . Algorithm 1 provides pseudo code for a simple simple genetic algorithm . Two primary parameters in genetic algorithm are population size and termination criteria .
<Page:13> A fuzzy prediction model for calling communities is based on a model of genetic evolution . The main steps in the algorithm are: Initialise population, evaluate fitness of individuals in population, select parents and mate for reproduction . Genes in the chromosome may change stochastically to generate offspring with sexual crossover and mutation .
<Page:14> Genetic algorithms have shown to be a powerful tool for performing: generation and optimisation of fuzzy rule-base algorithms and tuning of membership functions . In a fuzzy rule-based system, fuzzy if-then rules for an n-dimensional pattern are defined as follows (Ishibuchi et al., 1999): If x is A and … and x is A then then Class C with CF, (6) If x x is A and then x is a then . x then x is . A single point crossover example (see online version for colours)
<Page:15> In this study, a small number of fuzzy if-then rules have been randomly generated . The number of combinations of the antecedent fuzzy sets is 6n, which is very large in the case of high-dimensional problems . The meaning of each linguistic value is specified by a triangular membership function on the unit interval [0,1]. ‘don’t care’ has been handled by a special membership function with the following membership function .
<Page:16> The genetic algorithm has been set to use the uniform crossover, where each substring is handled as a block . It does not involve the adjustment of certainty grade functions or certainty grade . We have followed the implementation of Pittsburgh approach as it is faster compared to Michigan . The fitness of the rule set S is measured as: NCP(S) is the number of correctly classified training patterns .
<Page:17> Using CDR data, a cluster tree has been built . The data is randomly divided into 5 disjoint groups . The first group is set aside for testing and the other four are put together for model building . The mean of the five independent error rate predictions is used as the error rate for the final model . We believe indirect calling patterns will provide more useful information compared with direct calling patterns .
<Page:18> Figure 6 and Table 1 show how the clustering algorithm was evaluated with K being the number of clusters created from the cluster tree . The classification algorithm used for evaluating clusters is SVM . The reason that fuzzy genetic classifier has not been used is because of its running time .
<Page:19> SVM has an average overall accuracy of 98.5%, while fuzzy genetic classifier has an accuracy of 82.5% . SVM outperforms the fuzzy genetic algorithm by almost 13% . Pittsburgh approach is not directly based on fuzzy if-then rules but on fuzzy genetic operations in Pittsburgh approach . The running

time of SVM classifier was significantly less than fuzzy genetic .

<Page:20>  The fuzzy if-then rules in the final classifier model are divided into n subsets, where n is the number of distinct available classes (communities) For a particular customer,  every subset is examined to see whether there is any compatible rule with the feature  that represents the customer calling pattern . If there exists, such a rule is able to assign the ipient customer to the class of the subset under test . The certainty factor of the  compatible rule will then represent the degree of membership of this customer to its identified class (community)

<Page:21> erro

<Page:22> erro

<Page:23> erro