ISSN 2347 - 3983
Volume 7, No. 11 November 2019
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter177112019.pdf
https://doi.org/10.30534/ijeter/2019/177112019

An Exposition of Data Mining Techniques for Customer Churn in Telecom Sector

Pankaj Hooda[1] & Pooja Mittal[2]
[1,2]Department of Computer Science and Applications, Maharshi Dayanand University, Rohtak-124001, India
[1]hooda.pankaj@gmail.com, [2]mpoojamdu@gmail.com

ABSTRACT

Due to rise within the telecommunication sector and an outsized variety of competitors within the telecommunication domain, business issues face a rigorous client churn drawback. client churn could be a state of affairs that highlights a customer's intention to change from an explicit service or perhaps a service operator. Loads of client churner prediction methods square measure developed from business issues to handle client churn and to make sure client retention. However, client churn isn't wide foreseen in an exceedingly state of affairs wherever there's not enough historical information thanks to business issues, either due to a brand newly established organization or thanks to the occasional twitch of novel information or because of the shortage of that bond. With increasing number of mobile operators, user enjoys a seamless freedom to switch from one mobile operator to another if he finds the service or pricing unsatisfying. This draft deals with the enlargement of varied data processing techniques for churn analysis within the telecommunications sector..

Key words: JIT (Just In Time), CMN (Carrying Mobile Number), Genetic Programming (GP), Fast Fuzzy C-Means (FFCM), EMPC (expected maximum profit measure for customer churn).

1. INTRODUCTION

For many business issues, discovering the explanations for customer loss, activities to verify client loyalty and reclaiming the loyalty of client square measure important ideas. Business issues conduct numerous studies and campaigns to avoid the loss of their customers instead of exploiting new ones.

The telecommunication sector however receives vast amounts of knowledge; thanks to renewable information, increasing client volumes and added facilities. Uncontrolled and really fast enlargement of this scope results in accumulated losses supported fraud and technical constraints. Therefore, innovative inquiring approaches have to be compelled to be dilated.

In India, this case has been an inducement for numerous researches within the telecommunication sector that suffer from significant losses of clients. Brainstorm analysis that is commonly utilized in all domains is one of the applications of mining to get relevant knowledge to target groups accordingly. By analyzing the customers, agency would switch most service suppliers, organizations will produce specific selling campaigns aimed toward increasing client loyalty and developing selling methods for higher client retention. The aim of this study is to expedite why the telecommunications company loses its customers. Like finding a reason, defining those varieties of customers which have been lost is researched, as well.

2. RELATED LITERATURE

To detect churners in a vast consumer base, as is the case with telephone service providers, business concerns heavily rely on predictive churn models to remain competitive in a saturated market. In previous work, the expected maximum profit measure for consumer churn (EMPC) has been proposed in order to determine the most profitable churn model. However, profit concerns are not directly integrated into the model construction.

To notice churn in an exceedingly giant client base, as happens with telecom operators, business issues believe too heavily on prophetical churn models to stay competitive in an exceedingly saturated market. In previous work, the expected maximum profit measure for customer churn (EMPC) has been planned to see the foremost profitable churn model. However, profit issues aren't directly integrated into model construction. a number of these square measures given below:

Eugen Stripling et al. [1] introduced a classifier, named ProfLogit that maximizes EMPC within the training step employing a genetic algorithmic rule, wherever the inner model structure of ProFloGit resembles a Lasso-regularized provision model.

Bashar Al-Shboul and Nazeeh Ghatasheh [2] planned the churn prediction structure, that will increase the ability of predicting customers who are likely to churn. The structure supported a mixture of 2 heuristic approaches; Fast Fuzzy C-Means (FFCM) and Genetic Programming (GP).

Marcin Bienkowski et al. [3] introduced a proper model that's a remarkable generalization of the many classic on-line aggregation issues. Their main contribution is an O(w)-

competitive algorithmic rule, where w refers to the length of an IP address.

Ionut Brandusoiu [4] urged a complicated technique for predicting the churn of within the mobile telecommunications sector. He used a support vector machines algorithmic rule with four kernel functions to implement models for churner prediction. Gain measure was used to do the comparisons and

performances of the models were evaluated.

Adnan Amin, Changez Khan and Sajid Anwar [5] introduced a study that used cross-company information within the context of JIT to unravel CCP issues within the telecommunications sector, that is information from another business concern to predict churners.

Adnan Amin et al. [6] discovered an analysis of 4 rule-generation algorithms: learning from Module two version (LEM2) as an example, covering, completing, and genetic algorithms treated on public obtainable data sets..

T. Vafeiadis et al. [7] conferred a comparative study on the foremost fashionable machine learning ways applicable to the difficult drawback of predicting client churn within the telecommunications trade.

María Óskarsdóttir et al. [8] conferred a unique technique to extract statistic information from a decision network to represent off-time ever-changing client behavior with time.

Adnan Anjum et al. [9] urged a choice web, which might predict the churner behavior of a client. They came up with the method of developing an analytical system for data processing and machine learning technology.

Daskalaki, et al. [10] suggested a choice web for an outsized telecommunications company to handle client bankruptcy, that was ready to predict churn behavior by using decision trees and neural networks..

Chang's [11] research evaluates data processing methods to make a model for churn exposition. The urged approach provides excellent forecast reliability taking into account client demographics, request info, Call detail records, and service changed logs to create a churn prediction model with the help of artificial neural networks.

Turhan et al. [12] provided foreseen imperfection of modules by training models on in public obtainable NASA MDP information. In their experiments they used nearest neighbor sampling to construct static call graph based rankings to detect technique level defects.

Hilas's paper [13] deals with the identification of illicit telecommunications activity within the premises of huge organizations. The main focus is on fraud detection. The difficulty is attacked through the creation of a knowledgeable support system that consolidates the experience and apprehension of the network administrator derived from the implementation of data mining techniques on real-world data.

Wei and Chiu [14] presented design, and experimentally

conferred the planning, and through an experiment evaluated a churn-prediction technique, that churns out client written agreement info and pattern changes extracted from call detail, that could be a specific approximate time- ready to notice potential churn at contract level for the period.

3. CUSTOMER CHURN

If a customer closes a membership contract with his operator and becomes a customer of another rival concern, this customer can be defined as lost customer or Churn customer. Customer loss is correlated with customer loyalty. We cannot guarantee client satisfaction by doing providing pricing cuts only. Consequently, adding new value additions to services has been a mandatory trade norm to ensure loyalty of clients in telecommunications sector.

The main goal of lost customer research is to chalk out a client who is likely to shift and then to expedite the cost of obtaining those customers back again by arresting the churn rate. Throughout the analysis, the foremost necessary purpose is to define a churner client.In certain cases, it's terribly tough to form assumptions. A MasterCard client, for instance, will simply begin utilizing another bank's MasterCard while not canceling the MasterCard of the present bank. During this specific case, decrement in payment habits is often taken under consideration to grasp the customer's impairment. Arresting Client churn is the most tangible concern for business units that frequently lose their customers. Bank, insurance and telecommunication business units are good examples of this scenario.

For business units, it is not easy to exploit new customers as the price to do so is increasing day by day. Customers have numerous options to select among the mobile operators without any commitment and risk. Therefore, a brand new competitive era has started within the market segments. Rather than organizing campaigns to win new customers, business units focus on exploring a range of programs to arrest client dissatisfaction, increase earnings based on existing customer, and maintain high client loyalty.

The only method to achieve those goals is avoiding customer churn before it happens. Effective modeling of a customer churn can provide an important competitive advantage and a flexible workspace in this situation. A good model here unveils which consumer is on the brink of churn and which is loyal.With developments in information systems and also the variability of client behavior, there has been an unprecedented expansion in the knowledge size. This results in the extraction of unknown information and relationships amongst enormous amounts of data.Analyzing this information needs the utilization of varied ways and techniques in line with the structure of the data sets. The results of the investigation can be applied to devise targeted promotional campaigns and novel marketing strategies.

4. CHURN ANALYSIS IN TELECOM SECTOR

Telecom market, shows incredible growth in subscription with hectic competition, turns out to be at the verge of saturation This domain provides various services like quickly growing intercity and native calls; other communications services such as e-mail ,voice, fax; and other data traffic like Short Messaging Service/ Multimedia Messaging Service, Out Bound Dialers, Interactive Voice Response Systems etc.

Because of enhanced regulatory standards, novel information and communication technologies, Telecommunication market is becoming competitive and is rapidly developing in many countries. Customers have numerous options to select among the mobile operators without any commitment and risk. With increasing number of mobile operators, user is entitled with unlimited freedom to switch from one mobile operator to another if he is not satisfied with service or pricing. In this situation, data processing is a mandatory requirement to know the business needs, make a proper telecommunications model as per target segment, harnessing sources in a productive manner and do the value addition in service quality as per norms. Telecommunication industry is now a nubile market and is highly sensitive to the significance of customer retention. Due to this, continuous growth is achieved on the subsequent areas:

Cross Selling and up-selling: to capitalize maximum profits from existing clients.

Retaining and up-selling: to keep profitable consumers in possession and eliminate unsuitable customers from the profile of business.

Poaching: to catch new customers from rival business units.

Maintaining existing customers is easy as compared to procuring new ones. It's for that, business unit notice that it is vital to retain existing customers and agree that churner analysis is one of the indispensable data mining application areas.

In churn analysis applications, the primary factor is access to client information. Then, factors square measure classified in line with that factors or factors influence client churn selections. when crucial that customers square measure seemingly to churn, numerous and specific selling and retention methods are often applied to focus on customers over a group period. Churn Analysis helps to find out the reasons for switching service provider. We access customer data in first place. Then, it is decided to find out which factor or factors affect consumer churn decision on the basis of classification of these factors. After determining which customers are most likely to churn, different and specific marketing and retention strategies can be applied to the target customers, in a defined time frame.

This research isn't solely applicable in marketing; this additionally applies in client service, sales and finance applications. These markets need to determine the potential outcomes of the churn prediction, the monetary impact of the business, how the sales and client services sectors get tormented by the churn. Customers could leave the organization for various reasons. Therefore, totally different definitions are often created concerning client churn. Churn customer may be defined as "customers who leave the company due to some reasons". Contract expiration, phone device changes, service quality, competition, technical and regulative changes could also be listed as reasons for patrons to lose. The churner clients are often classified in line with the party addressing the initiatory move. If the first move is started by the client, it's referred to as "voluntary churn". If the organization approaches first; this is defined as "non-voluntary churn". This is the case when, for a few reasons the corporate could attempt to discontinue its services to the client

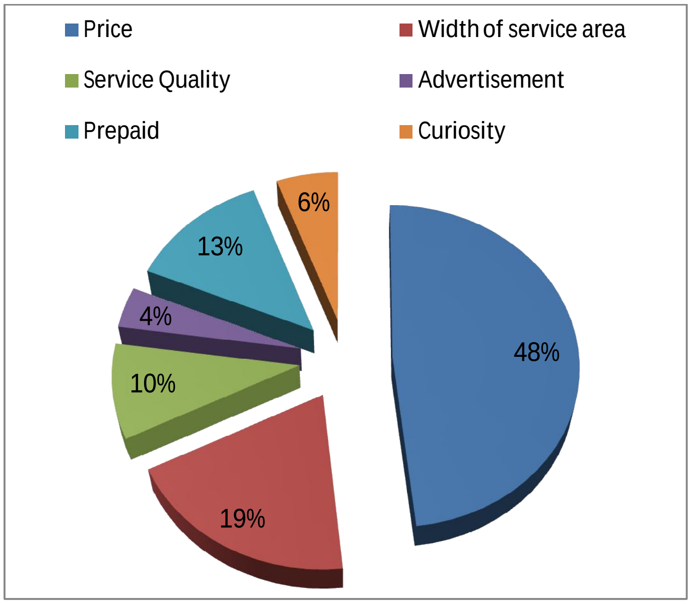

Figure 1: Possible causes for churn

Unpaid bills for several months or not to load prepaid minutes are some of the reasons for contract termination by service providers. Based on recent investigations, Carrying Mobile Number (CMN) in developed countries accounts to 27%. Pricing is the main reason for CMN. Shifting to a comparatively less expensive operator is normal. The churn rate generated by the worth is 48.3%. This alarming rate forces the service providers to launch competitive pricing for different plans to maintain market share. Operator's width of service area is the second most crucial reason for CMN and amounts to 19.4%. Discontentment of customers regarding service quality is 9.5%. Advertisements amount to 4.1% of the churn share. The proportion of the customers churning just for the sake of curiosity is 5.7%. This rate of shifting is higher than the rate of shifting due advertisements. This outcome is really attention grabbing and it is worth carrying out further investigation on this issue.

5. APPLICATIONS OF CHURN ANALYSIS

We can appraise the problem and obtain the perfect solution from the set of available alternatives with the help of a rigorous churn analysis. It comprises of the following steps:

5.1. Business Understanding

This initial section focuses on understanding the objectives of the project; needs specification as per industry requirements; then using this output to define a data mining problem; and then designing an initial set up so as to attain the objectives.

5.2. Data Understanding

This section begins with gathering preliminary data and carries out activities to become familiar with the data. Predicting information quality issues, finding raw information and discovering attention grabbing subsets to make hypotheses from concealed information are the activities carried out in this step.

5.3. Data Preparation

This section includes all activities for the creation of the ultimate dataset from the initial information. Tasks include table, record, and attribute selection as well as transformation and elimination of data for modeling. Data preparation work is performed multiple times, and not in any specific order.

5.4. Modeling

In this step, numerous modeling techniques are chosen and applied, and their parameters are calibrated to optimal values. Usually, there are many techniques which can be used for the same data mining problem type. Since some techniques require specific form of data, going back to the data preparation phase is frequently required. We do this activity till we find the most suitable data model as per our requirements.

5.5. Evaluation

At this stage, you have built a model (or models) that seem to have high quality, from a data analysis perspective. Before deploying this model, it is vital to get the model evaluated more rigorously, analyze the execution of steps to construct the model, and to be certain that the business objectives are properly achieved by it. The foremost objective is to check out whether there is some vital business issue that has been missed out. At the completion of this phase, we should have reached to a decision on the use of the data mining outcomes.

5.6. Deployment

This phase does not imply the end of the project. Even though the model aims to extend knowledge of the data, the procured knowledge will need to be organized and conferred in a way usable to the customer. As per the requirement, the deployment phase can be either as complex as executing a repetitive data mining process or as straightforward as generating a report. In numerous cases it will not be the data analyst, but the customer, who will carry out the deployment steps. So, it becomes mandatory for the customer to fully understand the actions which will be needed to carry out in order to make use of the models in a true spirit.

6. CONCLUSION

In this study, the data of business concerns which are operating in telecommunication domain is analyzed with the help of various data mining techniques. Our aim was to demonstrate models to foresee churner customer behavior, improve relationship with the customer, and plan various marketing campaigns and strategies to arrest customer churn and enhance loyalty. After removing non-related data and preparing stages, various methods are applied to determine the reasons for customer switching to other operators. This data can be used by the business concerns for creative visualizations. Useful outcomes can be derived to create a more powerful and holistic representation of a single user’s multiple transactions from calls to mobile data usage with the help of call detail record. This work can be extended to other industries like Insurance, Retail or Banking, and can be used as a basis for developing a more conclusive picture of consumer behavior because payment transactions are more frequently done with the help of mobile phones.

The table below shows the future scope of varied works done, techniques used, findings and analysis work done by numerous authors within the telecommunication sector. It provides a helpful comparison of various data mining methods and steps to make sure client retention with the assistance of novel and existing tools and techniques to focus on specific market areas..

Author's Names	Techniques used	Findings	Future Scope
Arno De Caigny *et al [15]*	Decision Trees, Logistic Regression, Random Forest, Logistic Model tree, Logit Leaf Model	This method provides a conceptually easy, but efficient and accurate model. It provides the average model training time for the different classifiers in which the LLM scores well.	It is possible to further improve the model by imposing more complicated rules for the number of leaves and the leaf sizes. Also there are lots of opportunities in model variations.
María Óskarsdóttir *et al [17]*	Similarity forests method, Multivariate time series extraction	This approach performs well when predicting further into the future, but only for certain types of churn. The combination of our early churn detection technique using dynamic behavior and traditional classification techniques based on consumer data , would result in a holistic approach to the churn prediction problem that captures both short-and long-term churn for different reasons.	In a future study, our approach could be applied to identify the source of this social contagion ,e.g. by extracting different types of features that represent the structure of the social network and the individual's position within it
Adnan Amin *et al [6]*	Classifier decision based on distance factor, Naïve Bayes classification algorithm. Tool used is MATLAB.	The proposed model predicts level of certainty that leads to expected level of accuracy. This can be used to select good cases for training the classifier efficiently and more accurately. This can also be used to predict outliers in training data that can have negative effect on the classification.	Empirical results can be provided on the balanced dataset with multiple base classifiers.
Adnan Amin *et al [5]*	It uses the capability of the SVM as base classifier, homogeneous and heterogeneous ensemble methods inthe proposed JIT-CCP model	The JIT-CCP model without any ensembles methods achieved the best performance (i.e., accuracy: 55.3±7.13, f-measure:47.26 ± 11.12, psep: 0.117±0.16 and kappa: 0.106), but with the application of homogeneous ensemble method which has improved the JIT-CCP model's performance by 3.94% in accuracy, 3.94% in misclassification error rate and 9.47% in f-measure. Finally, the heterogeneous ensemble method is applied, which further improves the performance of proposed JIT-CCP model to 18.03% in accuracy, 14.69% inf-measure and also reduce the misclassification error rate down to 18.03 %.	future study would be incorporating and testing of additional empirical studies with multiple cross-business concerns' datasets and multiple baselinepredictive classifiers to further support the outcomes of this work
Muhammad Azeem *et al [18]*	Fuzzy classifiers namely FuzzyNN, VQNN, OWANN and FuzzyRoughNN Tool used is WEKA	Dimensionality reduction and variable selection techniques form another major development in the machine learning field, and may be useful for retention research and practice, which face an overflow of potential defection predictors.	Future work is intended towards using the Fuzzy based feature selection methods i.e. fuzzy rough subset.
María Óskarsdóttir *et al[8]*	Relational Classifiers, collective inference methods, Social network Analysis	Since the authors used only one classifier, they are unable to verify whether there is correlation between network construction and classifier when it comes to performance.	Extension work includes determining the optimal way of constructing networks, with regard to edges and weights, and to determine the actual timespan of CDR data.
Eva Ascarza *et al [16]*	Dimensionality reduction and variable selection techniques, Deep learning	Dimensionality reduction and variable selection techniques form another major development in the machine learning field, and may be useful for retention research and practice, which face an overflow of potential defection predictors.	Management attention and academic research efforts need to be broadened beyond identifying consumers who are most likely to churn. There are several other areas very fertile for future work.
Eugen Stripling *et al[1]*	ProfLogit classication technique is introduced which is based on the logistic model structure but optimizes the regression parameters according to the EMPC, real-coded genetic algorithm	Contrasting the EMPC estimates reveals that using the RGA optimizer attains the highest average 629 and median EMPC performance in 9 out of 9 data sets (Figure 8). Hence, we can conclude that the 630 application of RGA is preferable over PSO and DE.	Concerning future research, we intend to develop a similar profit maximizing classification algorithm that substitutes the logistic model structure with a decision tree induction algorithm.
Adnan Anjum *et a [9]l*	Algorithms used are C5, Logistics Regression, Decision List, C & R-tree,QUEST and CHAID	Overall model accuracy is determined to be 72.19%, which is quite good, especially in telecom.	This work can be used as a basis to create a more conclusive picture of consumer behaviour that can be extended to other industries like Retail or Banking, due to increase in payments transactions via mobile phones
Marcin Bienkowski *et al[3]*	Prefix Aggregation, Competitive Analysis, Software Defined Networking	The nature of the problem is more related to online ski rental and technically different: achieving a constant competitive ratio is simple, but what the optimal constant is remains an open question. The authors present a 3.603-competitive solution	A rigorous study of the aspects like performance, efficiency and granularity at which FIB updates can be flushed to the hardware constitutes another interesting subject for future research.

REFERENCES

[1] E. Stripling, S.V. Broucke, K. Antonio, B. Baesens, M. Snoeck, **Profit maximizing logistic model for consumer churn prediction using genetic algorithms**, Swarm and Evolutionary Computation BASE DATA (2018)
https://doi.org/10.1016/j.swevo.2017.10.010

[2] B. Al-Shboul , H. Faris , N. Ghatasheh, **Initializing Genetic Programming using Fuzzy Clustering and its Application in Churn Prediction in the Telecom Industry**, 213-22, Malaysian Journal of Computer Science., Vol. 28(3), 2015
https://doi.org/10.22452/mjcs.vol28no3.3

[3] M. Bienkowski, N. Sarrar, S. Schmid, S. Uhli, **Online Aggregation of the Forwarding Information Base: Accounting for Locality and Churn**, *IEEE/ACM Transactions On Networking,* (1-14), Vol. X, No. Y, February 2017
https://doi.org/10.1109/TNET.2017.2787419

[4] I. Brandusoiu, G. Toderean, **Churn Prediction In the Telecommunications Sector Using Support Vector Machines**, Annals Of The Oradea University, Fascicle of Management and Technological Engineering, Issue #1,(19-22), May 2013
https://doi.org/10.15660/AUOFMTE.2013-1.2772

[5] A. Amin, F. Al-Obeidat, B. Shah, M. Al Tae, C. Khan, H.U.R. Durrani, S.Anwar, **Just-in-time consumer churn prediction in the telecommunication sector**, Springer, LLC 2017
https://doi.org/10.1007/s11227-017-2149-9

[6] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, A. Hussain, **Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Consumer Churn Prediction Case Study**, *IEEE Access*, 7940-7957, Volume 4, 2016

[7] T. Vafeiadis, , K. I. Diamantaras, G. Sarigiannidis, K. Ch. Chatzisavvas, **A comparison of machine learning techniques for consumer churn prediction,** Simulation Modelling Practice and Theory, June 2015
https://doi.org/10.1016/j.simpat.2015.03.003

[8] M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, J. Vanthienen, **Social Network Analytics for Churn Prediction in Telco: Model Building, Evaluation and Network Architecture**, Expert Systems With Applications (2017)
https://doi.org/10.1016/j.eswa.2017.05.028

[9] A. Anjum, A. Zeb, I.U. Afridi, P. M. Shah, A. Anjum, B. Raza, A.K. Malik, S.U.R. Malik, S. Usman and Z. Anwar, **Optimizing Coverage of Churn Prediction in Telecommunication Industry**, International Journal of Advanced Computer Science and Applications, (179-188), Vol. 8, No. 5, 2017
https://doi.org/10.14569/IJACSA.2017.080523

[10] S. Daskalaki, et al., **Data Mining for Decision Support on Consumer Insolvency in Telecommunications Business**, European Journal of Operational Research, 145, 239-255 (2003).

[11] Y.T. Chang, **Applying Data Mining to Telecom Churn Management**, International Journal of Reviews in Computing, 69-77, (2009).

[12] B.Turhan, G. Koçak, A. Bener, **Data Mining Source Code for Locating Software Bugs: A Case Study in Telecommunication Industry**, Expert Systems with Applications, 36, 9986- 9990 (2009).
https://doi.org/10.1016/j.eswa.2008.12.028

[13] C.S. Hilas, **Designing an Expert System for Fraud Detection in Private Telecommunications Networks**, Expert Systems with Applications, 36, 11559-69 (2009).

[14] C.P. Wei, I.T. Chiu, **Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach,** Expert Systems with Applications. 23, 103-112, (2002).
https://doi.org/10.1016/S0957-4174(02)00030-1

[15] A. D. Caigny , K. Coussement , K.W.D. Bock , **A New Hybrid Classification Algorithm for Consumer Churn Prediction Based on Logistic Regression and Decision Trees**, European Journal of Operational Research (2018)
https://doi.org/10.1016/j.ejor.2018.02.009

[16] E. Ascarza, S.A. Neslin , O. Netzer, Z. Anderson, P.S. Fader, S. Gupta, B.G.S. Hardie, A. Lemmens, B. Libai, D. Neal, F. Provost, R. Schrif, **In Pursuit of Enhanced Consumer Retention Management: Review, Key Issues, and Future Directions**, workshop on “Consumer Retention”, 10th Triennial Invitational Choice Symposium, University of Alberta (2016)

[17] M. Óskarsdóttir, T. V. Calster, B. Baesens, W. Lemahieu, J. Vanthienen, **Time series for early churn detection: Using similarity based Classification for dynamic networks**, Expert Systems With Applications, 106, 55-65, (2018)
https://doi.org/10.1016/j.eswa.2018.04.003

[18] M. Azeem, M. Usman, A.C.M. Fong, **A churn prediction model for prepaid consumers in telecom using fuzzy classifiers**, Telecommunication Systems, 66, 603-614, (2017)
https://doi.org/10.1007/s11235-017-0310-7