Measuring Churner Influence on Pre-paid Subscribers Using Fuzzy Logic

Louise Columelli
EURECOM,
Sophia Antipolis, France
louise.columelli@telecom-paristech.fr

Miguel Nuñez-del-Prado
Universidad del Pacifico,
Av. Salaverry 2020,
Lima-Peru
m.nunezdelpradoc@up.edu.pe

Leoncio Zarate-Gamarra
Peru IDI
Lima - Peru
leoncio.zarate@peruidi.com

***Abstract*—In the last decades, mobile phones have become the major medium for communication between humans. The site effect is the loss of subscribers. Consequently, Telecoms operators invest in developing algorithms for quantifying the risk to churn and to influence other subscribers to churn. The objective is to prioritize the retention of subscribers in their network due to the cost of obtaining a new subscriber is four times more expensive than retaining subscribers. Hence, we use Extremely Random Forest to classify churners and non-churners obtaining a Lift value at 10% of 5.5. Then, we rely on graph-based measures such as Degree of Centrality and Page rank to measure emitted and received influence in the social network of the carrier. Our methodology allows summarising churn risk score, relying on a Fuzzy Logic system, combining the churn probability and the risk of the churner to leave the network with other subscribers.**

Keywords—Churn, Data mining, Classification and Fuzzy logic

I. Introduction

Over the last two decades, we have seen mobile phones became the major medium for communication. In developed and developing countries, almost everybody has a cellphone and in some cases, some users have more than one phone with different carriers. the ITU reports 7 billions of mobile phones in the world corresponding to a penetration rate of 97% [1]. Consequently, competition between carriers to retain subscribers becomes more and more difficult as the number of carriers' grows. For example, each year about 10% of carriers subscribers churn. Since the cost of acquiring a new subscriber is in average 4 times more expensive than the cost of retaining one, mobile carriers have been investing in preventing subscribers to churn from their networks. Some works use only Call Detail Records (*CDR*) data. While other rely on Social Network Analysis (*SNA*) to classify churners and to estimate the impact of having churners among their close social neighbors. Our work focuses mainly on computing measures to quantify influence emitted as well as received for each subscriber, especially for those who will churn combining the churn probability as well as emitted and revived influence in only one score through a Fuzzy Logic system. We create features from *CDR* data, consumption statistics, and social interactions. To test the pertinence of the features and to find a suitable classification algorithm for churn prediction, we test different classification algorithms, such as *Naive Bayes* [2] classification algorithm, *Random Forest* [3], *Gradient Boosting* [4] and *Spreading Propagation* [5] algorithm. We evaluated the performance of these algorithms measuring the *Precision*, *Recall*, *F-Measure*, *Lift value* at 10%, *True Positive Rate (TPR)*, *False Positive Rate (FPR)* and *Area Under the Curve* ROC (*AUC*).

Our goal is twofold, we try to find an efficient algorithm for churn detection considering our variables and to quantify the influence of a subscriber in the social network of the Telco. Accordingly, our approach concludes a risk score computed base on three measures: (1) *the probability of churn*, (2) *the score of influence spread* and (3) *the score of influence received*.

The rest of our work is organized as follows: Section II presents related works. While Section III introduces the dataset used for experiments. Sections IV, V, and VI introduce our methodology to predict churners and to quantify the risk of churn, respectively. Finally, Section VII concludes our work and depicts futures research directions.

II. Related works

In the current section, we present related works to predict churners and to quantify the risk of the churner influence over other subscribers. Concerning churn prediction, these works use different approaches depending on the dataset. There are two approaches to detect churns: one *non-relational* (classical approach) [15], [16] and the another one *relational* (social approach) [17], [6], [18], [8], [9]. We describe these approaches in the following subsections.

a) Churn prediction: we focus on this prediction problem using the *social approach*. For instance, Phaadke *et al.* present their *Social Network Analysis-enhanced churn prediction* model [8], which is composed of three steps. The first step builds a communication graph taking subscribers as nodes (0,5 million) and edges weighted by the sum of normalized attributes. The second step, they compute the influence propagation of a given subscriber as the sum of the fractions of different influences of the subscribers' neighbors. Finally, influence attribute and other variables from the *Call Data Records* (CDR) as well as the *Customer Relationship Manager* (CRM) are introduced into a *Stochastic Gradient Boosting algorithm* [19] to predict possible churners. Authors use two months data to obtain 1.3 Lift value in the first 10% in the Lift curve. The study of Richter *et al.* [9] introduces the *Group-First Churn Prediction* approach focused on subscribers

Work	Social influence	Network partition	Learning model	Labeling	Lift
Columelli and Nunez del Prado	Centrality degree and Page rank	Louvain method	Extremely randomized decision tree	Threshold over the probability of churn	4.5-5.5
Dasgupta *et al.* [6]	Spreading Activation (SPA)	-	-	Threshold over influence given by SPA	5.5
Kim *et al.* [7]	Spreading Activation (SPA)	Louvain method	Logistic regression and Neural Networks	Threshold of the churn probability	-
Phaadke *et al.* [8]	Weighted sum of normalized variables	-	Stochastic gradient boosting	Threshold of the churn probability	1.25
Richter *et al.* [9]	Mutual information	Clusters between m and M nodes	Decision tree	Cluster marked as churner when churners $>$ a third of users	1.5-3
Verbeke *et al.* [10]	Logistic Regression	1st and 2nd order neighbors	Weighted Vote Neighbor Classification	Threshold over the mean of $Logic + WVNC$ probabilities	2.9
Zhang *et al.* [11]	Infection rule (IR) and number of iterartions (SR)	-	Neural Networks	Threshold over the probability of churn	3.5-7
Motahari *et al.* [12]	Correlation between edge variables and churner profiles		Decision Tree	Influence threshold	4
Han and Ferreira [13]	Generalized Propensity Score	-	-	Three months inactivity	
Li *et al.* [14]	Negative Inter-Subscriber Influence model	Information propagation model	Random forest	Threshold over the probability of churn	-

TABLE I. RELATED WORK SUMMARY

interactions. The idea is to form clusters containing a minimal m and maximal M number of subscribers. Once clusters are extracted, the authors compute individual and interaction variables in each group to identify risk groups when more than a third of subscribers in a group are churners. All these variables input to a *Decision Trees* [20] for ranking churners based on the probability to leave the carrier. Authors performed their experiments using one month data, which represents 16 millions of subscribers, obtaining between 1.5 and 3 of Lift value for the 10% in the Lift curve. Kim *et al.* [7] describe a method to predict churners based on demographic information, phone model, service satisfaction and calls data. The method builds communities using the *Louvain method* [21] to apply a *Spreading Activation* algorithm [5] in each community to measure the influence received from churners. Accordingly, the extracted variables are the input of a *Neural Network* (NN) [22], which uses a *Sigmoid function* for activation and three hidden layers. The experiments use 2.5 millions subscribers over two months observation period. To evaluate the results, authors use Hit Rate curve, which is the ratio of true positives out of the positives - in stead of Lift curve - reaching 31%. In the same spirit, Zhang *et al.* [11] relies on infection process to forecast churners using Neural Networks NN [22] as Machine Learning technique. The researchers combine traditional CDR data with network variables, such as neighbor composition, tie strength, similarity and homophily as input for the one hidden layer of the NN. The method uses an *Infection Rule Threshold* (IR) to label subscribes as churners if the probability of being a churner is greater than the IR. Once subscribers are labeled as churners, the method removes them and iterates until a given number of iterations are performed. The scientists experimented using one million subscribers dataset over a period of seven months obtaining between 3.5 and 7 of Lift score at 10% for different thresholds.

The work of Verbeke *et al.* compare different classical and social approaches to propose a strategy to combine both approaches [10]. Consequently, the algorithms considered for the former approach are: *Logistic Regression with Logit Function* [23], *Alternating Decision Tree* [24], *Random Forests* [3] *Bagging* [25] and *Bayesian Network* [26]. The techniques contemplated for the latter approach are: *Class-Distribution Relational Neighbor classifier* (CDRN) [27], *Network-only Link Based Classifier* [28], *Spreading Activation Relational Classifier* [5] and *Weighted-Vote Relational Neighbor* [27]. The study suggests three different strategies to mix classical and social techniques. (1) One sequentially strategy by inserting the churn scores computed by a social approach into the classical analyze and *vice versa*. (2) Another strategy by stacking the result of both approaches into a classifier or (3) combining them in parallel by taking the mean of the outputted probabilities. The authors found that the optimal configuration was associating the Logistic Regression and Weighted-Vote Relational Neighbor with a parallel strategy. They obtained 2.9 as Lift score value for the first 10% of the Lift curve. The experiments were performed using 673 724 subscribers, 2 414 945 edges and an observation period of five months.

b) Churn influence: we present some works in the literature on quantification of churners influence. For example, Dasgupta *et al.* [6] apply a *Spreading Activation* technique to the communication graph of the carrier subscribers over a period of four months. The idea behind this technique is to measure churners' influence over the subscriber of the communication graph. Thus, the authors initialize the spreading

Sender (subscriber of the network)	Receiver (subscriber or not)	Type of call	Duration of call in second	Number of call	Number of SMS
A	B	MOBILE	0.0	0.0	2.0
C	3030	SERVICE	0.0	0.0	1.0
D	E	MOBILE	0.0	0.0	1.0
D	A	MOBILE	60.0	1.0	1.0
E	B	MOBILE	484.0	14.0	0.0
F	C	MOBILE	102.0	2.0	0.0
F	555	SERVICE	64.0	1.0	4.0

TABLE II. EXAMPLE OF ACTIVITY DATA FROM MOBILE ORIGINATED FILE

influence process by taking the churners of the previous month as seeds. They use an energy function to transfer a portion of influence to the churners' neighbors. Once the process of influence is achieved a threshold is fixed to label new possible churners. The experiments were performed in a graph of 3.1 millions nodes and 12.3 millions edges, obtaining a Lift value of 5.5 for the 10% in the Lift curve. Motahari *et al.* [12] predict churners using a single *Decision Tree*. Authors use two month data to detect churners observing *CDR* activity, first. Then, the next two months to inspect friends of churners that have left the carrier. Based on this observation they correlate social features with churner profiles to establish features capturing the influence of churn in the social graph. Finally, they apply a single *Decision Tree* to capture influence level and to classify non churners. This method reached a Lift value of 4 at 10% using 4 months data of 270 millions of subscribers. Han and Ferreira [13] study the peer influence on churn. They use *Generalized Propensity Score* [29] with a treatment *exposure period*, where subscribers observe their friends churning and a post-treatment period, during which the researchers observe whether churners' friend leave. They used 13 months *CDR* data observing four millions of subscriber of a European carrier. Li *et al.* [14] proposed a methodology to measure the churn probability of a Chinese carrier using *Random Forest* [3] over *CDR* and interaction based features. Then, authors compute information propagation within the social graph of the carrier using *sender-centric* and *receiver-centric* models. In the former method, the sender decides the amount of influence he sends to their networks. In the latter method, it is the receiver who decide the amount of influence he receives. Finally, the use the *Negative Inter-Subscriber Influence* [14] model to quantify churner influence. Authors use for their evaluation a dataset composed of 3.5 millions of subscribers over six months. They use F-measure to evaluate the result of their method reaching between 0.254 and 0.823.

Table I summarizes the works described in the current section. More precisely, for each work we enumerate the algorithm for computing: social influence, social network partition algorithm, the classification algorithm, the method to label a churner and the Lift value at 10%, which is the most popular measure for this problem.

Different from the aforementioned works, which are summarized in Table I, our methodology outputs only one score composed of churn probability, received influence and emitted influence using a Fuzzy Logic system, which allows quantifying the risk of the churner to leave the network with other subscribers. To the best of our knowledge, there have not been other efforts using Fuzzy logic to combine churn probability and influence measure. In the next section, we present the dataset used to test our methodology.

III. DATASET DESCRIPTION

Topup file		
Subscriber	Number of charge	Price of charge
A	1.0	15.0
B	1.0	5.0
C	1.0	5.0
D	1.0	25.0
E	2.0	1000.0
F	1.0	10.0
G	1.0	10.0

TABLE III. EXAMPLE OF DATA FROM TOPUP FILE

In the present section, we describe the dataset used for the experiments. More precisely, we describe the detail of our dataset derived from pre-paid *CDR* and the features added for churners classification.

For the present study, we used daily anonymised data derived from a *CDR* of an African Telco operator. This data has 9 weeks $W = \{w_1, ..., w_9\}$ gathered from March to May 2015. This dataset is composed of: (1) Mobile Originated (*MO*), (2) Mobile Terminated (*MT*) and (3) Top-ups containing information about subscribers' consumption in terms of call, SMS, and charge, respectively.

- *Activity data* is composed of call and SMS data, as illustrated in Table II, containing: id caller, id callee, type (mobile line, land line, international line or service line), duration, the number of exchanged calls, the number of SMS exchanged. Once duplicates were removed, we aggregated the daily *MO and MT* data per week.
- *Top-up* correspond to the charge of mobile phones per day as presented in Table III. This dataset is composed of : subscriber id, the number of transactions, amount of transactions. As the aforementioned dataset, *Top-up* are aggregated per week.
- *Derived data* from *CDR*. Based on the above-mentioned dataset, we compute statistics such as: (1) the number and duration of incoming and outcoming calls; (2) the number and duration of calls as well as number of SMS to complaint, service, top-up and international numbers; (3) the entropy of the number of contacts, the mean, standard deviation, minimal, maximal of call durations and number of SMS, respectively; and (4) the ratio between the value of the week n and the week $n-1$ of all features. Table VIII presents the different added features.
- *Social network data*, extracted from activity data (*c.f.,* Table II), is a graph where users and connexion are represented by nodes and edges, respectively. Using

this social graph, we build features about the degree of centrality, connection, churn score in a community and Page rank score of each community.

Table IV summarizes the different datasets from both pre-paid and post-paid CDRs used in related works presented in Section II. More precisely, in the first part, Table IV shows the number of observed subscribers, the number of links in millions and the data gathering period. The second part of the table illustrates the proportion of the class churner and the set of tools used in different works of the literature. These tools measure the effectiveness of the classification process.

In the next section, we introduce our methodology to classify churners (*c.f.,* Section IV), to compute influence (*c.f.,* Section V) and to combine both churn prediction and influence score in a global risk score (*c.f.,* Section VI).

IV. Churner Classification

In the present effort, we compare different classifiers, such as: Naive Bayes Classifier [2], Extremely Randomized Trees [30], Extreme Gradient Boosting [4] and Spreading Activation Algorithm [6] to discriminate churners from Telco subscribers. We use for training phase weeks from w_1 to w_6 and testing from w_4 and w_9. Therefore, we make the hypotheses that a churner is a subscriber who does not have any activity for two weeks.

Concerning the *fine tuning* of the parameters for the aforementioned algorithms, we have tested different configurations varying the proportion of churners oversampling the churner class from 20% to 50% and the threshold of the probability to consider a subscriber as churner from 40% to 80% both with steps of 10%. For each algorithm, we obtain the maximal Lift value with the following configuration: Naive Bayes Classifier with 50% class proportion and 80% threshold; Extremely Randomized trees with 40% and 50%; and Gradient Boosting with 40% for both class and threshold.

Table V reports the best parameters, which optimize the Lift value at 10% for our dataset. For each tested algorithm, it includes the proportion of churners, the probability threshold to label a subscriber as churner, whether it includes social features and the value of the Lift curve. We observe that *Naive Bayes* algorithm performs the worst, while the *Extremely Random Forest* gives the best Lift value.

Once optimal parameters to maximize Lift value are found and summarized in Table V, we compare them to establish the best classification algorithm for our dataset. Accordingly, we present the *ROC* (*c.f.,* Figure 1) and *Lift* (*c.f.,* Figure 2) curves for the classifiers with the best area under the ROC curve (AUC) and the best Lift value at 10% for each classification algorithm.

In both, figures 1 and 2, we observe *Extremely Randomized Tree* and *Gradient Boosting* with and without social features performing almost the same and better than *Naive Bayes* and *SPA*. Accordingly, we compare the classification algorithm in terms of *precision, recall, training time, F-measure, Lift at 10%*, True Positive Rate (*TPR*), False Positive Rate (*FPR*) and Area Under the curve ROC (*AUC*) metrics. We measure the aforementioned metrics for *Naive Bayes*, *Extremely Randomized* and *Gradient Boosting* classifiers using the parameters

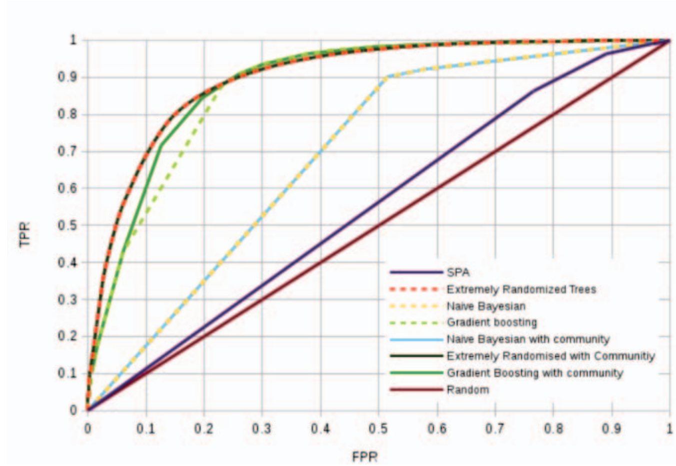

Fig. 1. ROC curves of classifiers trained with and without community features and the propagation algorithm results

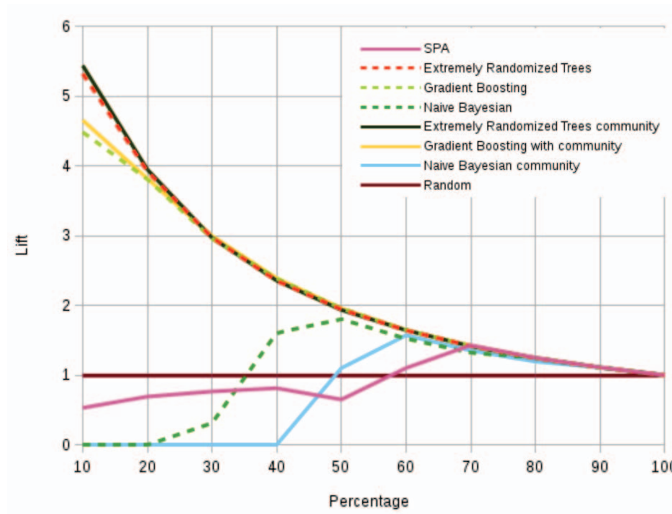

Fig. 2. Lift curves of classifiers trained with and without community features and the propagation algorithm results

setting, which optimizes the Lift value. The best classifier will output high values for all metrics except for *training time* and *FPR*. Figure 3 depicts the outperformance of the *Extremely Randomized tree* in terms of F-measure, *Lift value* at 10% and *AUC*. This classification algorithm presents low values for *training time* and *FPR* and high values for all other variables . Consequently, we use this classification algorithm to detect churners. However, we have 223 features as the input of the classification algorithm. Since the time of training is important in churn detection, we tried to find the minimum number of features to have similar results in terms of *Lift* curve and *AUC*. To meet this goal, we rely on the feature ranking establish by the *Extremely Randomized tree*. Thus, Figure 4 shows the similar performance of the classifier composed of 15 to 80 features, we appreciate classifiers with less than 15 features perform poorly in terms of *Lift* and *ROC* curve. Concerning the time function in Figure 4, we note that increasing the number of features augments the computation time.

To better understand the optimal number of features, we test the performance in terms of *Lift* value at 10% and *AUC*

Work	Subscribers (millions)	Links (millions)	Data collection (months)	Lift curve	F-measure	AUC	Cost curve	Hit Rate	Profit customer	CDR data	Network data	Class distribution
Columelli *et al.*	8	40	1.1	✓	✓	✓	✓	✗	✗	✓	✓	8.75%
Dasgupta *et al.* [6]	3.1	12.3	4	✓	✓	✗	✗	✗	✗	✗	✓	-
Kim *et al.* [7]	0.5	-	2	✓	✗	✗	✗	✗	✗	✓	✓	-
Phaadke *et al.* [8]	16	-	1	✓	✗	✗	✗	✗	✗	✓	✓	-
Richter *et al.* [9]	1	-	7	✓	✓	✓	✗	✗	✗	✓	✓	-
Verbeke *et al.* [10]	0̃.67	2̃.5	5	✗	✗	✗	✗	✗	✓	✓	✓	0.52%
Zhang *et al.* [11]	2.4	-	2	✗	✗	✗	✓	✓	✗	✓	✓	-
Motahari *et al.* [12]	270	22 120	4	✓	✗	✗	✗	✗	✗	✓	✓	-
Han and Ferreira [13]	4	-	10	✗	✗	✗	✗	✗	✗	✓	✓	-
Li *et al.* [14]	3.5	-	4	✗	✓	✗	✗	✗	✗	✓	✓	-

TABLE IV. RELATED WORK DATASETS SUMMARY

Classifiers selected for Lift curves comparison							
Parameters	SPA	Naive Bayesian with social features	Extremely Randomised Trees with social features	Gradient Boosting with social features	Naive Bayesian without social features	Extremely Randomised Trees without social features	Gradient Boosting without social features
Proportion	-	50	40	40	50	40	40
Threshold	0.4	0.8	0.5	0.4	0.8	0.5	0.4
Social Features	NO	YES	YES	YES	NO	NO	NO
Lift	0.53	0.00004	5.44	4.65	0.00004	5.38	4.6

TABLE V. TABLE SUMMARIZING THE PARAMETERS OF THE CLASSIFIERS WITH THE OPTIMIZED LIFT VALUE

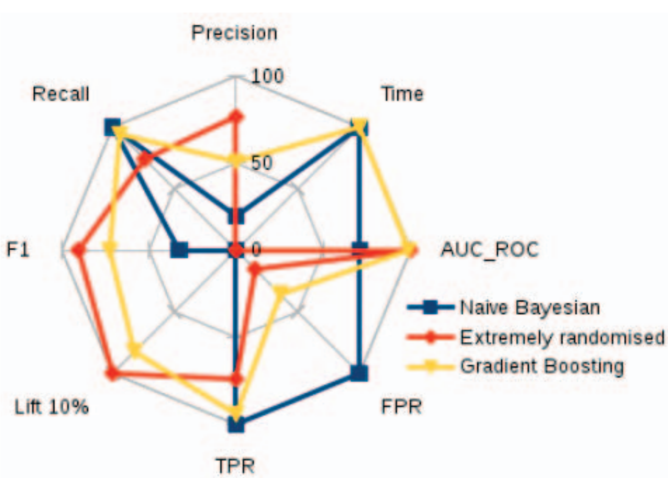

Fig. 3. Spider graph of the three classifiers with the best Lift value.

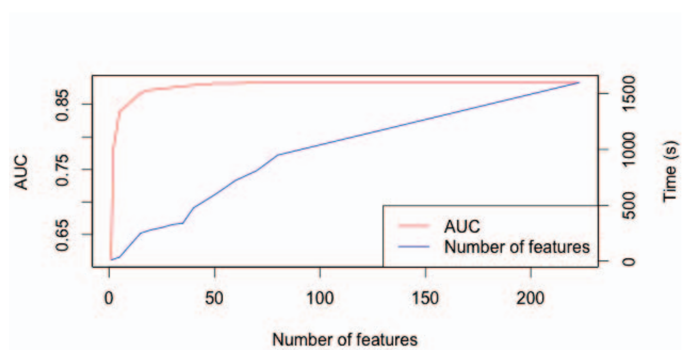

Fig. 4. Time and Area Under de ROC Curve (AUC) with respect to the number of features used in the extremely randomized algorithm for classification.

increasing the number of features. From Figure 5, we can see that the more features we have, the higher the Lift at 10% is. Hence, we find a maximum of Lift value when the classifier is trained using the 70 most significant features and then it falls down to match the score of the classifier trained with all the features as illustrated in Figure 5. Figure 6 confirms that the number of optimal features due to highest AUC value is obtained with 70 features. This phenomenon is due to the increasing information also adds noise. Moreover, correlated features do not increase discriminant information. In terms of computation time, training the classifier with 70 features takes the half of the time that training with all features. Finally, we conclude from figures 4, 5 and 6 that the 70 most significant

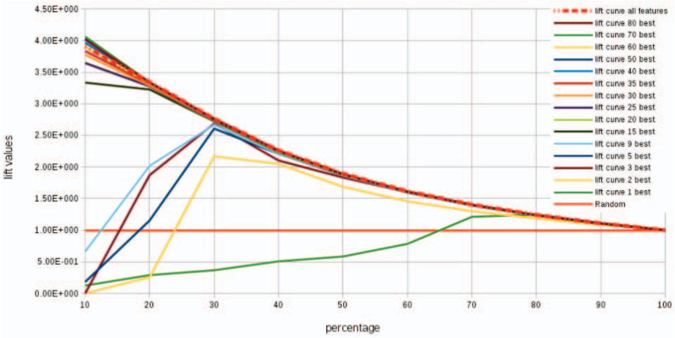

Fig. 5. Lift curves of of Extremely Randomized Forest classifiers trained with a subset of the best features

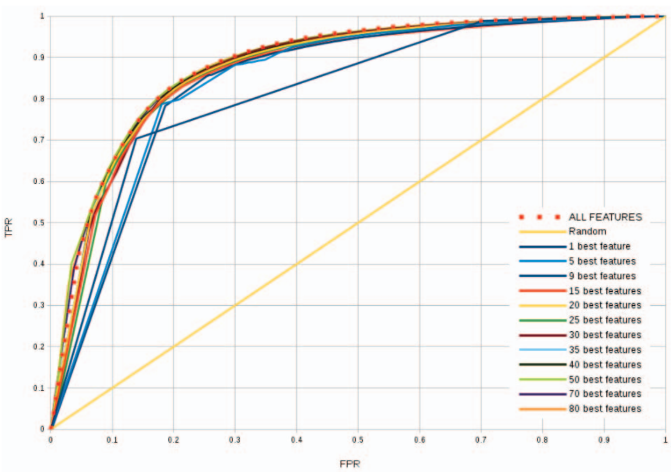

Fig. 6. ROC curves of Extremely Randomized Forest classifiers trained with a subset of the best features

features are enough for our dataset. Based on these optimal parameters and the minimal number of features to classify a churner, we obtain a Lift value at 10% of 5.5. We compare this result with the works found in the state of the art in Figure 7. The combined algorithm used in [11] performs the best in comparison to single methods. Our approach, which is a single method performs the best among the single methods. Nevertheless, we are conscious that these results are no directly comparable due to the datasets differences.

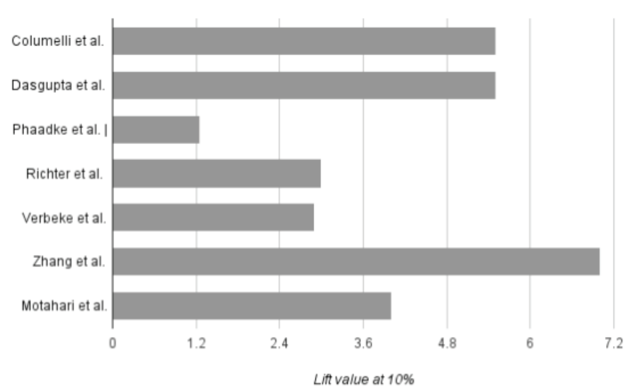

Fig. 7. Comparaison of churn classification algorithms.

Since we are able to classify churners. We need to quantify the emitted and received influences of subscribers to measure the risk of having high influence churners in communities with receptive subscribers. In the next section, we explain how to capture these influences.

V. Subscribers influence

Since we are able to detect churners in a Telco communication graph built from calls and SMS exchanges, we test different social based metrics, such as: (1) *Degree of Connection to a Churner*, (2) *Score of Churn in Communities*, (3) *Degree of Centrality* and (4) *Page rank score*. We observed that the *Degree of Connection to Churner* and *Score of Churn in Communities* quantify how close a subscriber is from a churner and the probability to churn in the community, respectively. Nonetheless, these variables do not really quantify influence. Hence, we explore the *Degree of Centrality* and *Page rank score* metrics. The former quantifies for a node v in the network, the fraction of nodes v is connected to within the community (*i.e.,* popularity). The latter computes a ranking of the nodes in the graph G based on the structure of the incoming links. It was originally designed as an algorithm to rank web pages. The *Page rank* algorithm gives each node a ranking based on its importance. A node v is considered more "important" based on the number of external nodes connected to v. Thus, *Page rank* models how influenceable a node v is.

Consequently, our methodology returns a vector composed of the *probability of churn* as well as the *score of the spread of influence* and *score of received influence* for each subscriber in order to quantify the value and/or the risk of the users. The Degree of Centrality captures how influent a subscriber is. For example two churners having a high probability to churn but one with high and the another one with low emitted influence. Therefore, it is more convenient to retain the subscriber with the highest *Degree of Centrality* or emitted influence due the potential economic impact of losing him with his followers. For instance, when subscribers have a high *Page rank score*, they are more susceptible to churn influence from other subscribers of the network. If a subscriber has a low *probability of churn* but a high score of *Page rank*, it can be interesting to look at the churn probability of their close neighbors to see if there is a higher risk that this subscriber becomes a churner under the influence of his neighbors.

Figure 8 depicts the relation between *emitted* and *received* influence of a subscriber, which is inversely proportional *i.e.,* the more influence someone spreads, the less influence he receives within his community. Thus, these scores can be used by the marketing department of the Telco company to understand further the real value of losing a subscriber. For the sake of explanation. Table VII presents the churn probability as well as the emitted and received influence. The objective of this table is to have the churn risk and the risk of influence from churners. Therefore, a churner with high influence is more dangerous and then the retention of that subscriber is more important due to the potential economic cost of losing influential subscribers. Accordingly, we need to combine these three measures (namely, Churn probability, emitted and revived influence) in only one global risk metric. Thus, in the next section, we present the fuzzy system, which transforms our three measures in a single metric.

Churn		**Low**			**Medium**			**High**		
Centrality		**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Page rank	**Low**	Low	Low	Medium	Low	Medium	Medium	Medium	Medium	Medium
	Medium	Low	Medium	Medium	Medium	Medium	Medium	Medium	Medium	High
	High	Medium	Medium	Medium	Medium	Medium	High	Medium	High	High

TABLE VI. FUZZY RULES SUMMARY TO COMBINE CHURN PROBABILITY, DEGREE OF CENTRALITY AND PAGE RANK.

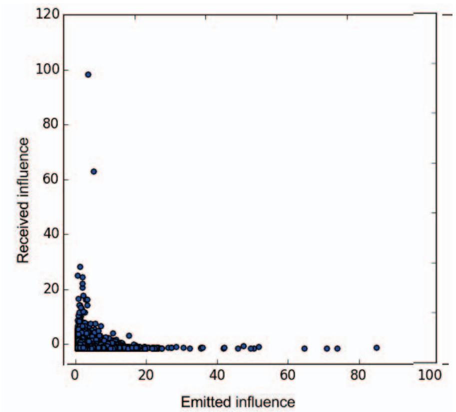

Fig. 8. Representation of the repartition of the influence received (Page rank) and emitted (degree of Centrality in the community)

VI. FUZZY SYSTEM

In the present section, we define and describe the implementation of our own fuzzy system to combine *Churn probability*, *Emitted influence* and *Received influence* in one global metric. Thus, based on the output of the classification algorithm as well as the Degree of Centrality and the Page rank, we obtain for each subscriber the *Churn probability*, *Emitted influence* and *Received influence*. Nonetheless, there is still the question about how to conclude using only one global score taking into account the three aforementioned variables.

More precisely, for each subscriber, we have computed a value, which is obtained by nonconventional techniques relying on *Fuzzy sets* as proposed by Zadeh [31]. Accordingly, we are able to attain a qualitative metric using quantitative values obtained from the *Extremely Randomized* classification algorithm, Degree of Centrality and Page rank. We use qualitative measurement scales such as: *low*, *medium* and *high* for each aforementioned variable. Hence, we built 27 *fuzzy implications* (fuzzy rules) that are summarized in Table VI

These rules were constructed based on triangular and trapezoidal fuzzy sets, which are founded on qualitative measurement scale as depicted in Figure 10. We use these fuzzy sets due tos crisp values range from 0 to 1 and the fuzzy sets are built from subintervals. Therefore, it is necessary to have a function, which models the membership with a gradual slope. Otherwise, the membership degrees are almost the same for all values in the non-fuzzy axis. For instance, the membership functions for the *Churn probability*, denoted by $\mu(x)$, indicates the *membership degree* of the variable value with a given fuzzy set. More precisely, $\mu_M(x_0)$ indicates the correspondence of the membership degree value of the Churn probability x_0 to the fuzzy set "Medium' [32]. The structure of the fuzzy rules: *medium, low, medium* and *medium* for *Churn probability*, *Degree of Centrality* (emitted influence), *Page rank* (received influence) and the result, respectively, are extracted from Table VI and depicted in Figure 9.

The values of *Churn probability* x_0, *Degree of Centrality* y_0 and *Page rank* z_0 are associated to the fuzzy set through their membership degrees (*i.e.,* fuzzification process). Then, the associated fuzzy sets correspond to the fuzzy rules in Table VI, which activates or triggers some rules based on the minimum mechanism of Mamdami [33]. This technique relies on outputted fuzzy sets union from the fuzzification process in accordance to the minimum of the membership degrees as illustrated in Figure 9. Finally, we obtain as the result a fuzzy set from which, quantitative values could be extracted from the defuzzification process [34]. There are different strategies for this process, in our case the defuzzification method used was the one proposed by Tsukamoto [35], which is illustrated in the scheme depicted in Figure 11.

For the normalized input values, from Table VII, we are able to rank the combined in a churn index, which is the output and result of the fuzzy system. Finally, using the described fuzzy system, we are able to rank subscribers base on the global churn metric, which is the output of the Fuzzy system.

VII. CONCLUSION

In the present work, we have presented a Fuzzy system to conclude a global risk index taking into account the Churn probability, the Degree of Centrality (emitted influence) and the Page rank (receiver influence). We observe that this new rank of subscriber takes into account the risk of high influence while churning. For example, in Table VII subscriber R has the highest Churn probability. Thus, regarding only the Churn probability, the Telco operator will try to retain R. However, subscriber C has a high Churn score as well as a high influence over his neighbors. Then, our methodology gives more importance to C than R due to his influence. Regarding subscriber A, even if he has medium Churn probability and emitted influence, he is the second more risky subscriber. The aforementioned facts are depicted in Figure 12, which gives a visual description of the new ranking of the risk of the subscribers. Consequently, our global risk metric captures potential damage of high emitted influence subscriber and penalizes those with a high received influence.

We have also studied different classification algorithms to identify potential churn, on pre-paid Telco service. As we have seen different ways of computing churn with several variables that significantly change the results. We have benchmarked the different algorithm in terms of performances showing that *Extremely Random Forest* classification fits the best for classifying churners reaching a Lift value of 5.5. It is worth noting that we were not able to use datasets described in the related works because they are not public.

On summary the present study objective is twofold. On one hand, it highlights and compares various techniques for analyzing the churn in the Telco industry, which includes

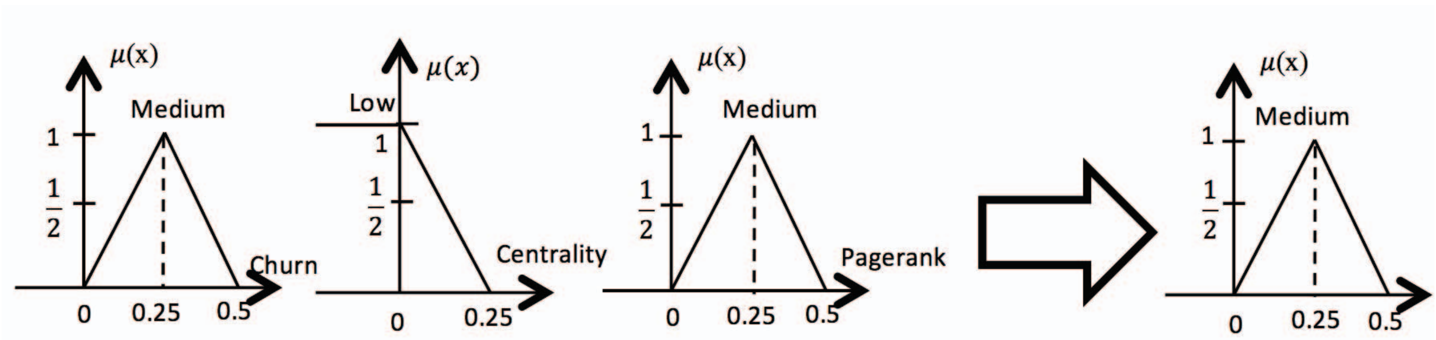

Fig. 9. The structure of the fuzzy rule: Medium $\wedge$ Low $\wedge$ Medium $\rightarrow$ Medium.

Subscriber churn	Churn probability	Influence spread (Degree of Centrality	Received influence (Page rank)	Fuzzy system output
A	0.4	0.5	0.2415558980725615	0.262816
C	0.6	1	0.5884190043917654	0.45
R	0.7	0	1	0.15
T	0.4	0	0.6966726471636972	0.13
D	0.3	1	0.0946519898487528	0.118640

TABLE VII. EXAMPLE OF INPUT AND OUTPUT VALUES OF THE FUZZY SYSTEM.

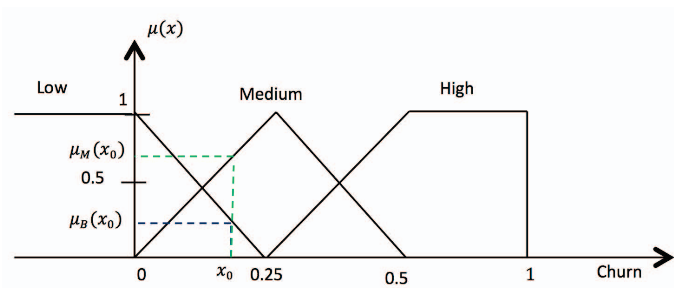

Fig. 10. Fuzzy set for Churn variable

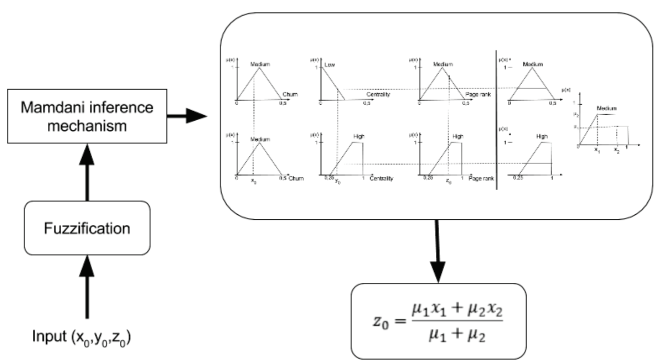

Fig. 11. Expert system based on fuzzy rules.

several variables for analysis, providing new insights into this phenomenon, which can be useful for managers to address this problem. On the other hand, it opens up new opportunities in the field of analysis of unstructured data, as well as a methodology to improve churn prediction and influence quantification through Fuzzy systems.

In the future, we plan to take different research avenues. Concerning features, we plan to build new features such as total number of churn neighbors, total number of non churn neighbors ratio of churn neighbors, average of call counts to/from churn neighbors, average of call counts to/from non churn neighbors, similarity to churn neighbors and similarity to non-churn neighbors that will help to discriminate churners.

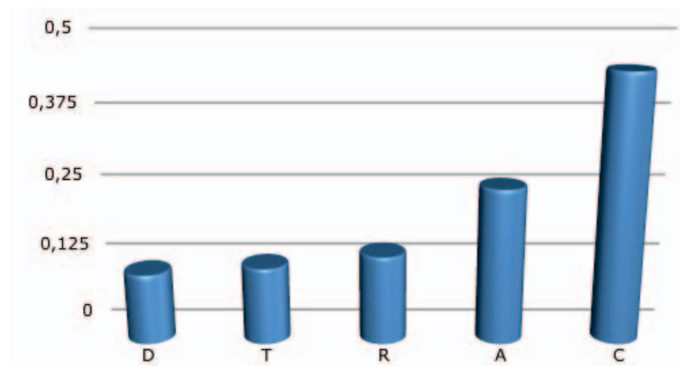

Fig. 12. Results obtained by the fuzzy system

Regarding the algorithms to detect churners, we will study in detail the Spreading Activation algorithm by personalizing weights between users by creating an influence function that takes into account the frequency of calls, the number of SMS exchanged, the proportion of friends in common and other consumption data. We would like to compare the Extremely Randomized Forest to Deep Learning based algorithms for predicting churners.

In the matter of the communication graph, we will add non-relational attributes to compute more precisely influence impact. Finally, about evaluation metric, we plan to use a function to measure economic loss to compare the traditional churn rank using Lift curves and our approach, which rank by the global risk metric.

REFERENCES

[1] B. Sanou, "Ict facts & figures. the world in 2015 - itu," 2015.

[2] K. P. Murphy, "Naive bayes classifiers," *University of British Columbia*, 2006.

[3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: http://dx.doi.org/10.1023/

[4] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[5] D. Aswath, S. T. Ahmed, J. D'cunha, and H. Davulcu, "Boosting item keyword search with spreading activation," in *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*. IEEE, 2005, pp. 704–707.

[6] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, "Social ties and their relevance to churn in mobile telecom networks," in *International Conference on Extending Database Technology: Advances in Database Technology*, 2008, pp. 668–677.

[7] K. Kim, C.-H. Jun, and J. Lee, "Improved churn prediction in telecommunication industry by analyzing a large network," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6575–6584, 2014.

[8] C. Phadke, H. Uzunalioglu, V. B. Mendiratta, D. Kushnir, and D. Doran, "Prediction of subscriber churn using social network analysis," *Bell Labs Technical Journal*, vol. 17, no. 4, pp. 63–75, 2013.

[9] Y. Richter, E. Yom-Tov, and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups." in *SDM*, vol. 2010, 2010, pp. 732–741.

[10] W. Verbeke, D. Martens, and B. Baesens, "Social network analysis for customer churn prediction," *Applied Soft Computing*, vol. 14, pp. 431–446, 2014.

[11] X. Zhang, J. Zhu, S. Xu, and Y. Wan, "Predicting customer churn through interpersonal influence," *Knowledge-Based Systems*, vol. 28, pp. 97–104, 2012.

[12] S. Motahari, T. Jung, H. Zang, K. Janakiraman, X.-Y. Li, and K. S. Hoo, "Predicting the influencers on wireless subscriber churn," in *Wireless Communications and Networking Conference*, 2014, pp. 3402–3407.

[13] Q. Han and P. Ferreira, "The role of peer influence in churn in wireless networks," in *International Conference on Social Computing*, 2014, p. 11.

[14] H. Li, D. Wu, G.-X. Li, Y.-H. Ke, W.-J. Liu, Y.-H. Zheng, and X.-L. Lin, "Enhancing telco service quality with big data enabled churn analysis: Infrastructure, model, and deployment," *Journal of Computer Science and Technology*, vol. 30, no. 6, pp. 1201–1214, 2015.

[15] F. Castanedo, G. Valverde, J. Zaratiegui, and A. Vazquez, "Using deep learning to predict customer churn in a mobile telecommunication network."

[16] E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, "A proposed churn prediction model," *International Journal of Engineering Research and Applications*, vol. 2, pp. 693–697, 2012.

[17] M. N. Abd-Allah, A. Salah, and S. R. El-Beltagy, "Enhanced customer churn prediction using social network analysis," in *Proceedings of the 3rd Workshop on Data-Driven User Behavioral Modeling and Mining from Social Media*. ACM, 2014, pp. 11–12.

[18] A. A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, and A. Joshi, "On the structural properties of massive telecom call graphs: findings and implications," in *International conference on Information and knowledge management*, 2006, pp. 435–444.

[19] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[20] J. R. Quinlan, "Simplifying decision trees," *International journal of man-machine studies*, vol. 27, no. 3, pp. 221–234, 1987.

[21] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[22] L. Personnaz and I. Rivals, *Réseaux de neurones formels pour la modélisation, la commande et la classification*. CNRS éd., 2003.

[23] P. Macdonald, "The logit transformation: With special reference to its uses in bio-assay," *Journal of the Operational Research Society*, vol. 25, no. 1, pp. 201–202, 1974.

[24] Y. Freund and L. Mason, "The alternating decision tree learning algorithm," in *International Conference on Machine Learning*, vol. 99, 1999, pp. 124–133.

[25] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996. [Online]. Available: http://dx.doi.org/10.1007/BF00058655

[26] F. V. Jensen, *An introduction to Bayesian networks*. UCL press London, 1996, vol. 210.

[27] S. Macskassy and F. Provost, "Classification in networked data: A toolkit and a univariate case study," *The Journal of Machine Learning Research*, vol. 8, pp. 935–983, 2007.

[28] Q. Lu and L. Getoor, "Link-based classification," in *International Conference on Machine Learning*, vol. 3, 2003, pp. 496–503.

[29] M. Bia, A. Mattei *et al.*, "A stata package for the estimation of the dose–response function through adjustment for the generalized propensity score," *The Stata Journal*, vol. 8, no. 3, pp. 354–373, 2008.

[30] P. Geurts and G. Louppe, "Learning to rank with extremely randomized trees," in *nternational Conference of Machine Learning*, vol. 14, 2011, pp. 49–61.

[31] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," *Information and control*, vol. 20, no. 4, pp. 301–312, 1972.

[32] C. C. Lee, "Fuzzy logic in control systems: fuzzy logic controller. i," *Systems, Man and Cybernetics*, vol. 20, no. 2, p. 404418, 1990.

[33] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," *Computers, IEEE Transactions on*, vol. 100, no. 12, pp. 1182–1191, 1977.

[34] C. C. Lee, "Fuzzy logic in control systems: fuzzy logic controller. ii," *Systems, Man and Cybernetics*, vol. 20, no. 2, pp. 419–435, 1990.

[35] Y. Tsukamoto, *An approach to fuzzy reasoning method*. Amsterdam: NorthHolland, 1979.

70 Best Features found by Extremely Randomised Trees algorithm	
Features	Score of importance in the algorithm between 0 and 100
Number of contact as sender week 3	85.6842405603
Mean of call as sender week 3	72.1977136435
Max of call as a sender week 3	67.6883892538
Std of call as a sender week 3	65.8078816395
Different service number week 3	62.594591353
Mean of call as receiver week 3	56.7376715045
Number of contact as receiver week 3	38.1658769021
Max of call as a receiver week 3	35.6447452349
Min of call as a receiver week 3	34.3498342095
Std of call as a receiver week 3	29.9006362374
Number of contact as sender week 2	26.47257198
Mean of call as sender week 2	25.0617360922
Max of call as sender week 2	25.0134134171
Different complaint number week 3	23.5019044418
Mean of call as receiver week 2	19.6876325386
Std of call as a sender week 2	18.5124327113
Second degree connection to churner as receiver of sender	18.1764491978
Variation of duration of international call between weeks 2 and 3	16.588543703
Number of contact as receiver week 2	15.5987021206
Min of call duration as a sender week 3	14.5551961457
Different service number week 2	13.2803020136
Std of call duration as a sender week 3	12.9105401967
Std of call duration as a receiver week 3	12.8109016229
Mean of call duration as a sender week 3	12.7969274608
Min of call as a receiver week 2	11.8617726378
Max of call duration as sender week 3	11.4444924704
Variation of mean of SMS sent between weeks 2 and 3	9.9188579774
Std of call as a sender week 1	9.8089648616
Variation of duration of call to service number between week 2 and 3	9.6308167903
Std of call as receiver week 2	9.598031029
Max of SMS as sender week 3	9.4720119872
Max of call duration as receiver week 2	8.9687411232
Variation of call duration with international number between weeks 1 and 2	8.929425153
Mean of call duration as receiver week 3	8.6227041776
Max of call as receiver week 2	8.3332668123
Number of different complaint number week 2	7.4620020266
Std SMS as sender week 3	7.3628276947
Number of contact as receiver week 1	7.2037561236
Variation of mean SMS sent as receiver between weeks 1 and 2	6.9991481872
Variation of min duration of call as receiver between weeks 2 and 3	6.7490675154
Mean SMS sent week 3	6.6695614819
Number of contact as a sender week 1	6.6413824665
Variation of call duration to solde between weeks 2 and 3	6.1692854106
Max call as receiver week 1	5.830458951
Std call duration as sender week 2	5.8198061423
Mean call as receiver week 1	5.6953587667
Variation of mean sms sent between weeks 1 and 2	5.6409139681
Mean call duration of receiver week 2	5.3481486856
Variation of number of complaint number used between weeks 2 and 3	5.1077404387
Max call duration as receiver week 3	4.9571658325
Min sms sent week 3	4.6726663789
Different service number used week 1	4.5485166731
Std call as receiver week 1	4.5276886419
Variation of mean of sms received between weeks 2 and 3	4.3343463049
Std of call duration as receiver week 2	4.1140551017
Max of call duration as sender week 2	4.0812076735
Variation of min of call received between weeks 2 and 3	3.3482673875
Min of call sent week 2	3.2169180963
Std of call duration as sender week 1	2.9471368301
Variation of number of contact as sender between weeks 2 and 3	2.7962831718
Variation of max sms received between weeks 2 and 3	2.6469065629
Min of call duration as receiver week 3	2.4483871724
Second degree connection to churner as receiver of receiver	2.3907089592
Variation of max call as sender between weeks 2 and 3	2.3649956267
Variation of complaint call duration between weeks 2 and 3	2.3303259297
Mean of call duration as sender week 2	2.3282528857
Mean of call duration as sender week 1	2.2918336008
Variation of min of call duration as receiver between week 1 and 2	2.2608068387
Variation of number of topup bought between weeks 1 and 2	2.1579493352
Variation of number of different solde number used between 2 and 3	1.9228006512
Variation of price of charge in topup between weeks 2 and 3	1.8503215638

TABLE VIII. LIST OF 70 BEST FEATURES SELECTED FOR THE CHURN PREDICTION