

Optimizing Coverage of Churn Prediction in Telecommunication Industry

COMSATS Institute of Information
Technology, Pakistan

Sahiwal, Pakistan

Adnan Anjum¹, Adnan Zeb³, Imran
Uddin Afridi⁴, Pir Masoom Shah⁵,
Adeel Anjum⁷, Basit Raza⁸, Ahmad
Kamran Malik⁹, Saif Ur Rehman
Malik¹⁰
Department of Computer Science,

Saeeda Usman²
Department of Electrical Engineering,
COMSATS Institute of Information
Technology,

Zahid Anwar⁶
Department of Computer Science,
Bahria University, Islamabad, Pakistan

Abstract—Companies are investing more in analytics to obtain a competitive edge in the market and decision makers are required better identification among their data to be able to interpret complex patterns more easily. Alluring thousands of new customers is worthless if an equal number is leaving. Business Intelligence (BI) systems are unable to find hidden churn patterns for the huge customer base. In this paper, a decision support system has been proposed, which can predict the churning behaviour of a customer efficiently. We have proposed a procedure to develop an analytical system using data mining as well as machine learning techniques C5, CHAID, QUEST, and ANN for the churn analysis and prediction for the telecommunication industry. Prediction performance can be significantly improved by using a large volume and several features from both Business Support Systems (BSS) and Operations Support Systems (OSS). Extensive experiments are performed; marginal increases in predictive performance can be seen by using a larger volume and multiple attributes from both Telco BSS and OSS data. From the results, it is observed that using a combination of techniques can help to figure out a better and precise churn prediction model.

Keywords—Telco; Churn Prediction; Business Intelligence; Business Support Systems; Operations Support Systems; E-Churn Model (Ensembling Churn Model)

I. INTRODUCTION

Churn is dealing with the risk of a customer moving from one company to another. Churn prediction is used to recognize customers who are most probable to churn. Churn prediction and analysis can help a company to develop a sustainable strategy for customer retention programs. By getting awareness of the percentage of churners, we can easily come up with detailed analysis, causes of the churn and customer retention programs. Pakistan opened its Global System for Mobile (GSM) communications telecom services in October 1994 and there are five telecom service providers: Mobilink, Telenor, Ufone, Warid and Zong operating with over 121 million subscribers (<http://www.pta.gov.pk>, Pakistan telecommunication authority, PTA). Figure 1 shows the rate of increase in subscriber growth in the telecom sector. It shows that the largest transition is between 2004 and 2007. After 2007, linear growth can be seen.

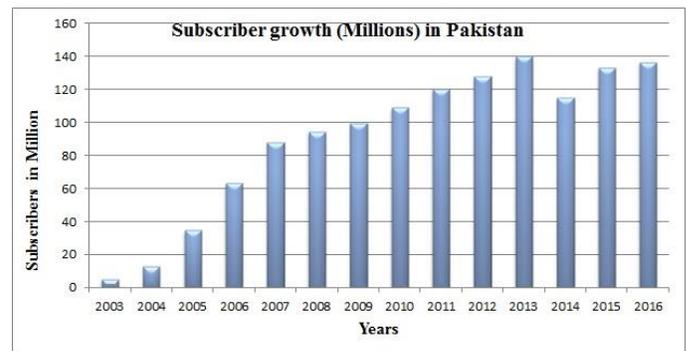


Fig. 1. Subscriber growth (millions) in Pakistan

Some facts about the telecom sector in Pakistan (PTA) are: First, subscriber growth rate has reduced to 5%, which was more than doubled each year till 2006. Second, inflation has moved up to 15% from 7% in 2007 and moved to double digit in 2013. Third, interest rates are constantly increasing and currently standing at 14%. Fourth, exchange rate against USD has gone up by more than 58% since 2006. Fifth, there is extremely low average revenue per User ARPU (< USD 2) against the world average of USD 17.

Pakistan telecom sector's main focus is on prepaid customers with very little or no legal binding. The prepaid churn rates are usually higher than postpaid churn. Over the last five years, the average lifetime of the prepaid customer has halved to only 17 months. Churn prediction is especially very difficult in Pakistan, because of many reasons such as: prepaid base with no contractual bond; limited or inaccurate subscriber information (like name, gender, age, location, etc.); extreme competition between all operators; mostly unlettered subscribers, based in rural areas; customers with very low buying power; and Mobile Number Portability (MNP) law from Pakistan Telecom Authority (PTA) removed the number binding as well. The same number can be used for other operators, IT fraud, and fake sales. Because of above-mentioned reasons, there is a dire need of special churn model with optimized coverage in developing countries like Pakistan.

Churn is still the biggest issue of the competitive telecom market of Pakistan as none of the company surveys achieved

75% coverage of the churn. In Pakistan, more than 40 million subscribers go unpredicted. The traditional methods used for churn prediction are easy to work with and to generate good results, but these are still not sufficient. There is a lack of biographical and microenvironmental variables in the existing models, which are developed on the basis of the customer revenue and usage only.

According to an experimental demonstration, with the integration of both BSS and OSS data, Telco big data can considerably enhance the performance of churn prediction. BSS data covers the major components of IT, which helps the operators to run the business operations. Four different processes can be handled by BSS data, namely, product management, order management, revenue management and customer management. On the other hand, OSS can provide support in network management tasks, e.g. as network inventory, service provisioning, network configuration and fault management. Although BSS data have been utilized in churn prediction very well; even in our research, it is giving a total of 72% precision. But, it is worthwhile collecting, storing and mining OSS data, which takes around 97% size of the entire Telco data assets this could even increase the precision to 0.96%. Figure 2 presents an overview of the architecture of Telco big data platform [1].

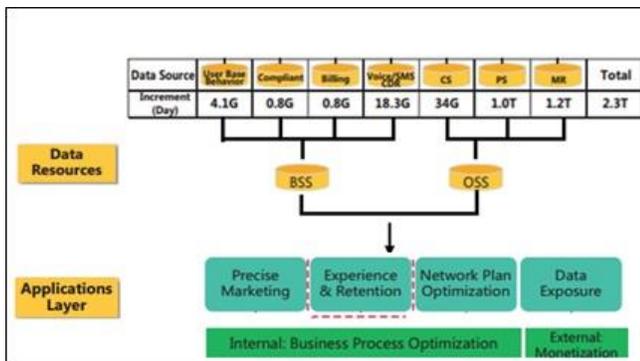


Fig. 2. The overview of platform

In this article, churn prediction mechanism is developed that can increase the recall of the prediction models and is also able to find a reason of the churn. The primary contribution is devising a new mechanism of an ensemble that can help to increase the recall factor and answer following questions: how are different algorithms cross compared to each other independent of the nature of datasets, and how well different models can be ensemble to effectively predict churn? The rest of paper is arranged as follows: Section 2 describes related work. Section 3 presents our proposed Ensembling-Churn (E-Churn) Model. Data collection and manipulation are described in Section 4. Section 5 provides formal verification of the model. Section 6 consists of results and discussion. Finally, Section 7 concludes the research and offers future directions.

II. RELATED WORK

In the telecom industry, the service provider has to realize the customer-centric business strategy. A churner joins another company in search of better rates, services, joining benefits, and very low joining fees. These trends also attract

other subscribers to switch to another company. According to the Database Marketing Institute, annual churn rates of Telecom industry varies from 10 to 67 per cent [2]. In [3], authors indicated reactive and proactive approaches, one can take to manage churn. A customer can ask the company to cancel its service relationship with him in a reactive approach. In this approach, the operator doesn't have any predictive analysis team and has negligible efficiency. While in a latter approach, the company tries to analyze the behaviour of different customers to identify the churns and proactively counter them with some lucrative services to retain them. Churn prediction is only about the proactive approach. The key here is to build as accurate churn prediction model as possible [4]. Following are the efforts made to reduce the churn figure by developing an effective churn prediction approach [5-14, 15-17]. The limiting factor in existing approaches is that it makes use of only one of the data mining techniques, i.e. classification or clustering.

Some of the studies have used more than one techniques based on cluster analysis and classification [2]. Support Vector Machine (SVM) is used to develop the churn model of a newspaper subscription. It is complex for implementation but it is a benchmark for random forecast [1, 18, 19, 20, and 21]. Both types of classifiers: single and ensemble have been used for churn dataset classification [22] and it was found that self-organizing map, Principal Component Analysis, and Heterogeneous Boosting outperform other classification methods. A study based on the text of customers for the consideration of their positive and negative influences is presented for churn analysis on a macro level but not on an individual level [23]. A churn model is also available to solve unbalanced, scatter and high dimensional problem in telecom datasets [24]. The C4.5 decision tree algorithm is applied on the dataset by achieving 80.42% precision. Rough Set Theory based on Genetic algorithms produced efficient decision rules as compared to other rule generation mechanisms named Exhaustive Algorithm, Covering Algorithm and the LEM2 algorithm for churn and non-churn classification [25]. A churn prediction model based on AUC parameter selection technique is proposed which has shown good performance in the case of noisy nonlinear business customer's dataset [26].

Some of the studies used a binomial logic regression to build the prediction model [27]. Hybrid approaches tend to be very flexible as in these approaches we can combine both classification and clustering techniques. Usually, the clustering is used to develop the model and after the model creation one can classify and predict future behaviour. There is no single hybrid approach instead multiple hybrid approaches are used to find more accurate results [28]. Available work in literature is based on a single data mining techniques; classification or clustering for the prediction of customer churn and mining of retention data of customer [9, 22], however, some studies have been conducted which apply more than one technology [2, 30].

III. E-CHURN MODEL

A new ensembling model has been proposed which can increase the churn prediction that can eventually increase the overall recall of the diverse type of data. Figure 3 explains the

abstract level process model. First, the existing ensembling techniques are used in which the top two or more accurate algorithms will be selected. Then the result of the multiple models will be compared based on the presumption that one technique could be better at predicting as compared to another. If the two models predict someone as churn it will be marked as a churning. If the model differs, propensity will be checked. If the propensity of any model is greater than 70–80%, it will be marked then as churning, else as non-churning. An algorithm

for the explanation of ensemble process is discussed below. We combined the accuracy of these models and predicted all those that were marked as True either by any of these algorithms in the modelling process. The merging models work on these logics: Create two or more models and test them. When the models agree, use that prediction. When the models don't agree, use the model prediction with the highest confidence. The best fit is opted on the basis of the highest precision and recall factor.

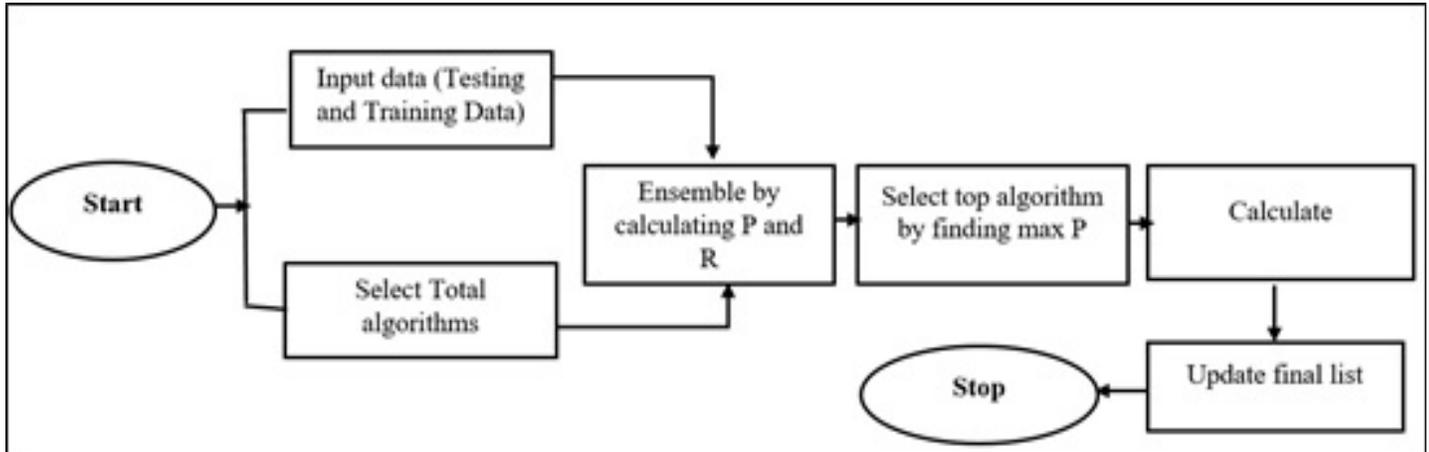


Fig. 3. Flow chart of the proposed technique

Algorithm

Combiner algorithm for ensembling Churn prediction
Require: DCHURN 1: lstChurn 2: BUC // Top n algorithms 3: Initialize: N = total number of churn prediction techniques with same base (e.g., Decision tree) 4: $D_{testing}, D_{training}$ 5: selecting based and derived variables 6: for $i=1$ to N do 7: Ensemble($i, D_{testing}, D_{training}$) 8: $BUC \leftarrow$ Calculate Precision and Recall for i_{th} churn prediction technique 9: end for 10: $BUC \leftarrow$ Select Top n techniques from BUC 11: for $i=0$ to $D_{CHURN}.count$ do 12: Calculate propensity for $D_{CHURN}[i]$ using each selected churn technique in BUC 13: if propensity > 80 % then 14: $lstChurn.add(D_{CHURN}[i])$ 15: end if 16: end for 17: Manage-Disagreement($lstChurn, D_{CHURN}[i]$) 18: for $i=1$ to $lstChurn.count$ do 19: if $lstChurn[i].ARPU > 2$ then 20: $lstChurn.delete(lstChurn[i])$ 21: end if 22: end for 23: return $lstChurn$

IV. DATA COLLECTION AND MANIPULATION

The analysis is done on the raw call detail records (CDRs) and customer demographics data of six months. The raw CDRs were parsed through massive ETL work and data was loaded in the operational data model. Data was divided into testing and training. 31 actual and 83 derived variables were

obtained from this raw data. SPSS was used as a mining tool. The list of attributes used for this experimentation is: Customer identification code, Charged SMS, Charged calls, Charged minutes, Charged revenue, Free calls, Free minutes, Free SMS, Total incoming minutes, Total outgoing minutes, Onnet calls, Onnet minutes, Onnet revenue, Recharge total load and, Revenue SMS.

A. Data Preparation

Data preparation is a significant and time-taking phase of data that covers constructing the final dataset from the initial raw data by performing data preparation tasks for several times, not in any prescribed order. Transformation and elimination of data for modelling tools as well as table, record, and attribute selection are some of these tasks. IBM Modeller uses data prepared from ETL process using the

telecommunication data warehouse. After fetching the raw data from the data warehouse, the first step is to run the data audit and see the maximum, minimum and average value of each attribute. We paid special attention to identifying if any record is having lots of null value or if the record is completely null. The data audit report is stored in an excel sheet and has the format as shown in Table 1.

TABLE I. DATA PREPARATION

Field	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid	Outlier	Extreme
MSISDN	Continuous	3002000992	3645916333	3145565507	55906475.59	3.453	--	2913060	61470	48183
TOTAL_CLAS S_PI	Continuous	0	21007	216.18	382.583	4.036	--	2913060	50719	18901
TOTAL_MINS PI	Continuous	0	112146.13	727.739	2073.644	6.971	--	2913060	40350	25730
TOTAL_CALL REV_PI	Continuous	0	68609.83	282.82	532.663	6.992	--	2913060	38211	18425
ONNET_CALL PI	Continuous	0	20988	131.265	302.242	5.179	--	2913060	48462	24510

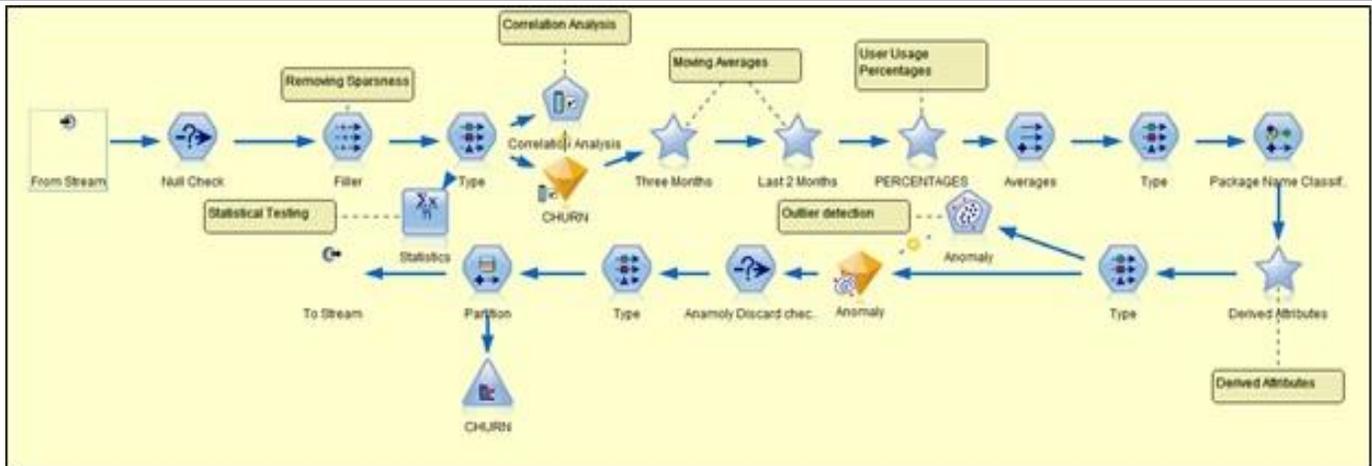


Fig. 4. Data Pre-processing

B. Data Pre-processing

Since the data preparation phase usually includes loosely controlled data and can have out-of-range values, missing values and impossible data combinations, i.e. data which has not been carefully screened. Analyzing such data can produce misleading results. Inconsistent and redundant data (due to missing values and impossible data combinations) even makes data mining phase more difficult. Data pre-processing, shown in Figure 4, involves a number of steps which can take a considerable amount of time. The data is filtered in a form which can produce more accurate results. First, correlation

analysis with target variables is conducted. Second, feature elimination then outliers detection in data is done. Smoothing is performed and in the end, sparseness is removed.

C. Correlation with Target Variables

Filtering out target variable is an important step. The correlation analysis is used to find the level of dependence of target variable over some independent attributes. The target variable is churn, with two values T or F. The system will decide the list of the important attributes to be included in the further analysis. The Pearson test was used for categorical target variable; churn.

D. Feature Elimination

A number of techniques are used to eliminate the scattered features in the data. We used standard deviation, variance and principal component analysis in this phase. First, the standard deviation is used to find out variation or dispersion from the average value of the data. Second, values are discarded with a standard deviation greater than 2. Third, the variance is used to find out how far the data spreads away from the mean value. The attribute is discarded if its variance is zero.

E. Smoothing

We have removed the short term fluctuations by moving computed averages for this purpose.

F. Sparseness

Replacing missing/null values by averages of other values removes missing/null values. It balances out the odd effect of missing/null values and has a smooth transitional pattern.

G. Outliers Detection

Identify and remove the outliers in the data, as any abnormal value can affect the model. In SPSS Modeller, a node called Anomaly Node is used to check every record and identify anomalies. The Anomaly Detection procedure

examines infrequent deviations cases from their cluster groups. The procedure is designed for explanatory data analysis step to rapidly identify unusual cases for data auditing purposes before carrying out any inferential data analysis. This node performs the operation by identifying records which are having outliers or extreme values and will affect the overall accuracy of the model. The algorithm is for generic anomaly detection. The definition of an anomalous case is not specific to any particular application.

H. Modelling

We train our models by using different algorithms as shown in Figure 5. The distribution of data for churn consists of 95.73% non-churners while only 4.27% with churning behaviour. The dataset is divided into *Training* and *Testing* data sets. The models were trained with the 70% of the dataset, which has the selected inputs and the target attribute. Later the trained model is used for testing on the other 30% of the dataset to see how much accurately the trained model can predict the Target Variable. The target variable churn has two outcomes, T and F, we used some input variables to predict it and later check the accuracy of the predicted churn variable with the actual variable.



Fig. 5. Modelling

I. Balancing Dataset

A large data file cannot be used as a sample; thus the balance node can be used to make the distribution of a categorical field more equal. Balancing is carried out by discarding records based on the conditions specified, i.e. records for which no condition holds are always passed through. In normal churn, we have 6% True while 94% false records; we reduced the true records as to balance the dataset for a fair representation.

V. FORMAL VERIFICATION

For the formal verification, we have used PIPE+, a tool which supports High-Level Petri Nets (HLPN) [31]. Transition conditions are defined in terms of logic formulas [32]. For the formal verification of the combiner algorithm,

first an HLPN is developed, and then logical formulas are applied to verify it. The Table 2 explains the places that are used for the verification of the algorithm. It explains every place in detail that what will place a hold and a part of a state and the Petri net structure. It also presents the mappings of places to data types.

It provides the static semantics information that does not change throughout the system. After identifying all the places needed for the verification, the formulas are applied on the transitions. This maps the transitions to predicate logic formulas. The Figure 6 shows the formulas applied on each transition. The PIPE+ generates a Promela formula specification script as a result of model checking as shown in Figure 7.

TABLE II. PLACES AND THEIR DATA TYPES OF THE HIGH-LEVEL PETRI NET

Total Algorithms	Place holds all the algorithms for churn analysis.	$\mathbb{P}(\text{Algo_ID})$
N Algorithms	Place holds all the algorithms with the same base.	$\mathbb{P}(\text{Algo_ID})$
Data	Churn dataset i.e. Caller Detail Record (CDR).	$\mathbb{P}(\text{Cust_ID})$
Training Data	Place holds only the training data i.e. 70% of the whole dataset.	$\mathbb{P}(\text{Train_ID} \times \text{CUST_ID})$
Testing Data	Place holds only the testing data i.e. 30% of the whole dataset.	$\mathbb{P}(\text{TestingID} \times \text{CUST_ID})$
BUC	Place holds the percentages of precision and, after ensembling the algorithm, training data and testing data.	$\mathbb{P}(\text{Algo_ID} \times \text{Train_ID} \times \text{TestingID} \times \text{PxR})$
TBUC	Place holds the top algorithms which precisions are maximum.	$\mathbb{P}(\text{Algo_ID} \times \text{P} \times \text{R})$
Propensity	Place holds the propensity for the dataset of each algorithm in BUC.	$\mathbb{P}(\text{Algo_ID} \times \text{CUST_ID} \times \text{Propensity})$
List Churn	Place holds the list of all the subscribers who are most likely to churn.	$\mathbb{P}(\text{Cust_ID})$
Final Churn List	Place holds a list of subscribers who are most likely to churn and have ARPU less than 2.	$\mathbb{P}(\text{Cust_ID})$

R (Select algo with same base) = $\forall a \in A \bullet a \neq NULL \wedge \forall b \in B \bullet b := \text{SameBase}(a) \wedge B' = B \cup \{b\}$

R(separated data) = $\forall c \in C \bullet c \neq NULL \wedge \forall d \in D \bullet d \neq NULL \wedge \forall e \in E \bullet e \neq NULL \wedge d[2] := \text{TrainingData}(c) \wedge e[2] := \text{TestingData}(c) \wedge d[1] := \text{Randomly_assign_id}() \wedge e[1] := \text{Randomly_assign_id}() \wedge D' = D \cup \{d[1], d[2]\} \wedge E' = E \cup \{e[1], e[2]\}$

R(ensemble) = $\forall f \in F \bullet f \neq NULL \wedge \forall g \in G \bullet g \neq NULL \wedge \forall h \in H \bullet h \neq NULL \wedge s[2] := g[1] \wedge s[3] = h[1] \wedge s[1] := F \wedge s[2] := \text{Ensemble}(F, G, H) \wedge s[3] := \text{Ensemble}(F, G, H) \wedge S' = S \cup \{s[1], s[2], s[3]\}$

R(maximumPandR) = $\forall i \in I \bullet i \neq NULL \wedge \forall j \in J \bullet j \neq NULL \wedge i[2] := s[4] \wedge j[3] = s[5] \wedge j[1] := \text{Maximum}(i[2], i[3]) \rightarrow j[1] = i[1] \wedge J' = J \cup \{j[1]\}$

R(calculatePropensity) = $\forall k \in K \bullet k \neq NULL \wedge \forall l \in L \bullet l \neq NULL \wedge \forall m \in M \bullet m \neq NULL \wedge m[3] := \text{propensity}(K, L) \wedge m[1] := l[1] \wedge m[2] := K \wedge M' = M \cup \{m[1], m[2], m[3]\}$

R(prop > 80) = $\forall n \in N \bullet n \neq NULL \wedge \forall o \in O \bullet o \neq NULL (\text{Propensity} > 80) \rightarrow O := n[2] \wedge O' = O \cup \{o\}$

R(manageDisagreement) = $\forall p \in P \bullet p \neq NULL \wedge \forall q \in Q \bullet q \neq NULL \wedge \forall r \in R \bullet r \neq NULL (\text{ARPU} > 2) \rightarrow P \setminus \{p[1]\} \wedge P' = P \cup \{p[1]\} \wedge R' = R \cup \{r\}$

Fig. 6. Formulas applied on the transitions

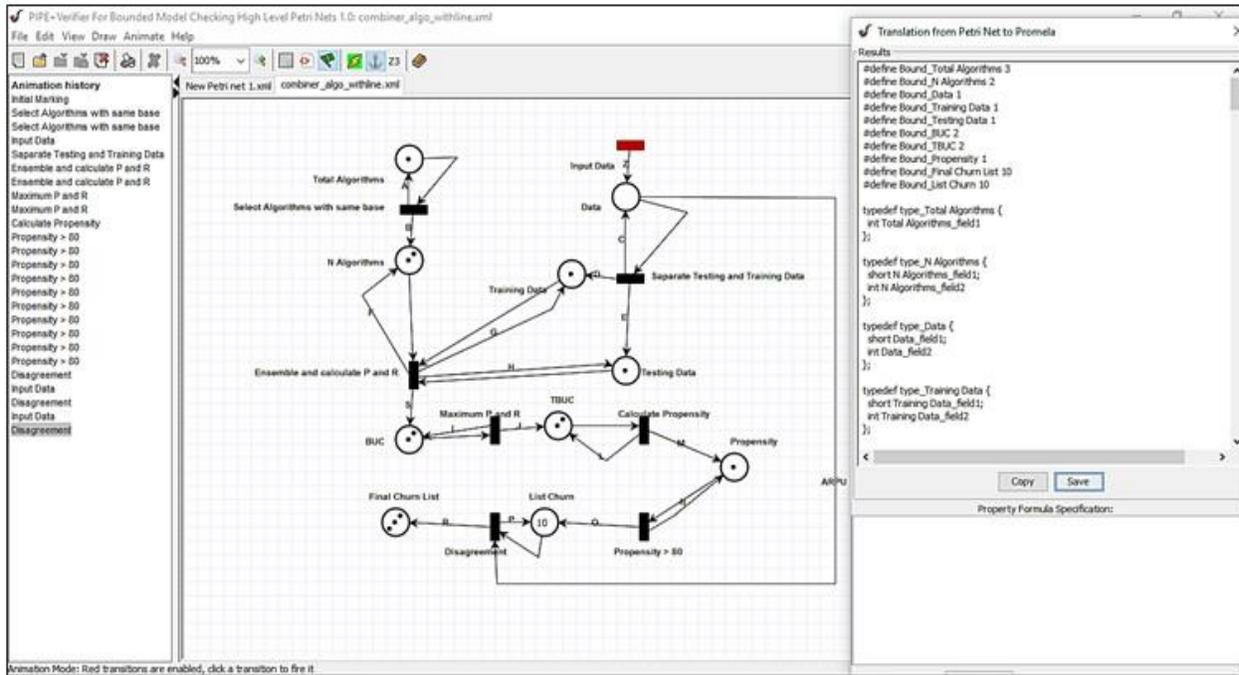


Fig. 7. Modelling checking in PIPE+

VI. RESULTS AND DISCUSSION

Data selection, experimentation, ensembling, and final results are step by step processes. We proposed a 360-degree view on the problem that will cover dimensional model, data cleansing, data preparation and churn prediction with different prediction algorithms, ensembling results of different algorithms. Deeply analyzed the accuracy of each modelling algorithm and studied how their accuracies may be improved. Later these algorithms are compared and the best algorithm is declared for prepaid subscriber base. We used decision tree based algorithms for prediction due to their rule-based nature which makes them easy to understand and implement. They provide “reasoning”, which branch is causing churn based on their proven results in other Telco data sets. Algorithms used are C5, Logistics Regression, Decision List, C & R-tree, QUEST and, CHAID.

A. Results before Ensembling

The results for C5 Model are explained here and shown in Figure 8. For True cases approach, the model was able to correctly predict 81% of the churners, which means it, predict that these subscribers will churn out and testing data confirmed these figures.

Model incorrectly predicted 19% of churners, who were not actually churners, but it marked them as churners. For

False cases approach, the model was able to correctly predict 71% of the non-churners correctly, which means Model predicts that these subscribers will not churn out and testing data confirmed these figures. Model incorrectly predicted 29% of non-churners, who were actually churners, but model marked them as non-churners. Overall model accuracy is determined to be 72.19%, which is quite good, especially in telecom. The results for CHAID as shown in Figure 9 are: For True cases approach, the model was able to correctly predict 77% of the churners correctly. Model incorrectly predicted 23% of churners. Similarly, for False cases approach, the model was able to correctly predict 69% of the non-churners correctly. Model incorrectly predicted 31% of non-churners. Overall model accuracy is determined to be 69.71%. The CRT results in Figure 10 shows that for True cases approach, the model was able to correctly predict 82% of the churners correctly and incorrectly predicted 18% of churners. Whereas, for False cases approach, the model was able to predict 60% of the non-churners correctly and incorrectly predicted 40% of non-churners. Overall model accuracy is determined to be 60.99%. The QUEST results as shown in Figure 11 are: For True cases approach; the model predicted 84% of the churners correctly and incorrectly predicted 16% of churners. For False cases approach, the model predicted 55% of the non-churners correctly and incorrectly predicted 45% of non-churners. Overall model accuracy is determined to be 53.99%.

Results for output field CHURN				
Comparing SC-CHURN with CHURN				
'Partition'	1 Training		2 Testing	
Correct	122,341	77.33%	547,038	72.19%
Wrong	35,870	22.67%	221,185	27.81%
Total	158,211		795,223	

a) C5 model analysis for testing and training models

SC-CHURN			
CHURN		F	T
F	Count	604529	235764
	Row%	71.944	28.056
T	Count	21309	91850
	Row%	18.831	81.169

b) C5 model result breakdown

Fig. 8. C5 Results

Results for output field CHURN				
Comparing SC-CHURN with CHURN				
'Partition'	1 Training		2 Testing	
Correct	116,684	73.44%	554,342	69.71%
Wrong	42,207	26.56%	240,881	30.29%
Total	158,891		795,223	

a) CHAID model analysis for testing and training models

SC-CHURN			
CHURN		F	T
F	Count	583132	257832
	Row%	69.341	30.659
T	Count	25256	87903
	Row%	22.319	77.681

b) CHAID model result breakdown

Fig. 9. CHAID Results

Results for output field CHURN				
Comparing SC-CHURN with CHURN				
'Partition'	1 Training		2 Testing	
Correct	112,227	70.73%	484,973	60.99%
Wrong	46,445	29.27%	310,250	39.01%
Total	158,572		795,223	

a) CRT model analysis for testing and training models

SC-CHURN			
CHURN		F	T
F	Count	5048860	335876
	Row%	60.050	39.950
T	Count	20819	92340
	Row%	81.602	18.398

b) CRT model result breakdown

Fig. 10. CRT Results

Results for output field CHURN				
Comparing SC-CHURN with CHURN				
'Partition'	1 Training		2 Testing	
Correct	109,215	69.08%	444,858	55.94%
Wrong	48,891	30.92%	350,365	44.06%
Total	158,106		795,223	

a) QUEST model analysis for testing and training models

SC-CHURN			
CHURN		F	T
F	Count	459406	380764
	Row%	54.680	45.320
T	Count	18492	94667
	Row%	16.342	83.658

b) QUEST model result breakdown

Fig. 11. QUEST Results

B. Ensembling

We selected the top two algorithms C5 and QUEST. The C5 returned an accuracy of 81% and QUEST returned an accuracy of 84% for the True case. During the modelling process, the accuracy of these models along with all those predicted customers who were marked as True was combined by either of the two algorithms. These two algorithms can predict up to the accuracy of 93% for Churn TT (Actual and Predicting). Therefore, we can use a C5 and QUEST for scoring churn in future. Below is the interpretation of ensembling different algorithm to increase the True cases.

By Ensembling C5 and CHAID, for the True case, the model predicted 80% of the churners correctly and for False cases approach, the model was able to correctly predict 72% of the non-churners. Through Ensembling three algorithms

C5, QUEST, and CHAID for the True case, the model correctly predicted 81% of the churners whereas for the False case, the model predicted 67% of the non-churners correctly. By Ensembling C5, CHAID, QUEST, and CRT, for the True case, correctly prediction accuracy is 82% of the churners. However, for False cases, the accuracy of correctly predicted non-churners is 68%. Through Ensembling C5 with QUEST, for the True case, the model was able to correctly predict 93.4% of the churners. Whereas for the False case, the model was able to correctly predict 47% of the non-churners. Other Ensembling combinations can be seen in Table 3 similarly, other Ensembling combination can be created. By using the combination of C5 and QUEST we can cover nearly full of the churner base, as it can predict up to 93.4% churners correctly. Normally, algorithms accuracy can go up to 80%, which means that 20% of the churners always remained unattended

by the telecom companies. Overall accuracy dramatically increases by ensembling the output of different algorithms as shown in Table 3.

TABLE III. COMBINE MODELS

Models	Overall Accuracy (Testing %)	Recall %	Precision %
C5 + CHAID	72.93	80.623	72.693
C5 + CHAID+ CRT	71.51	72.433	71.969
C5 + CHAID+QUEST+CRT	68.84	81.378	68.321
C5+QUEST+CRT	70.29	81.686	69.924
C5+QUEST	52.5	93.319	47.106
C5 + CHAID+QUEST	68.51	84.433	67.969

Since it is a combination of different models, one need not tune the model every three months; apparently, the results will remain effective for many months. The algorithm accuracy for true cases can be further improved once we have the contact history data from campaigns against these churners. This data will eventually increase the TT (*Actual & Predicting*) results because the campaign will tend to increase the subscriber usage, wrongly predicted churners will always have a different kind of behaviour to the campaigns as compared to actual churners.

VII. CONCLUSION AND FUTURE WORK

This research makes use of multiple churn prediction models to find the suitable way to predict all the churners and also identify the most probable reason of churn, by using many different algorithms; we can save the model development and training time and effort so they can be targeted effectively to reduce the churn rate. Through research, it is observed that by combining C5 and QUEST algorithms, we can cover nearly full of the churner base, as it can predict up to 93.4% churners correctly of the BSS data. Further by using a combination of OSS and BSS data the prediction of churners was increased (0.96 precision) and higher than the previous churn prediction system deployed in which uses only BSS data (0.68 precision). Moreover, Telco companies can use their data for useful visualizations. With the help of CDRs, useful measures can be derived to create a more powerful and holistic representation of a single user's multiple transactions from calls to mobile data usage. 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.

REFERENCES

[1] Hung, S.Y., Yen, D.C., Wang, H.Y. (2006) Applying data mining to telecom churn management. *Expert Syst Appl*, 31, 515–524.

[2] Hung, S.Y., Yen, D.C., Wang, H.Y. (2006) Applying data mining to telecom churn management. *Expert Syst Appl*, 31, 515–524.

[3] Burez, J., Poel, V.D. (2009) Handling class imbalance in customer churn prediction. *Expert Syst Appl*, 36, 4626–4636.

[4] Tsai, C.F., Lu, Y.H. (2009) Customer churn prediction by hybrid neural networks. *Expert Syst Appl*, 36, 12547–12553.

[5] Almana, A.M., Aksoy, M.S., Alzahrani, R. (2014) A survey of data mining techniques in customer churns analysis for the telecom industry. *Journal of Engineering Research and Applications*, 4, 165–171.

[6] Oseman, K.B., Shukor, S.B.M., Haris, N.A., Bakar, F.B.A. (2010) Data mining in churn analysis model for the telecommunication industry. *Journal of Statistical Modeling and Analytics*, 1, 19–27.

[7] Kim, H.S., Yoon, C.H. (2004) Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommun Policy*, 28, 751–765.

[8] Hadden, J., Tiwari, A., Roy, R., Ruta, D. (2007) Computer-assisted customer churn management: State-of-the-art and future trends. *Comput Oper Res*, 34, 2902–2917.

[9] Karnstedt, M., Rowe, M., Chan, J., Alani, H., Hayes, C. (2011) The effect of user features on churn in social networks. *Proceedings of WebSci11*, 14–17 June, Koblenz, Germany, pp. 14–17, ACM, New York, USA.

[10] Kim, N., Jung, K.H., Kim, Y.S., Lee, J. (2012) Uniformly subsampled ensemble (use) for churn management: Theory and implementation. *Expert Syst Appl*, 39, 11839–11845.

[11] Lemmens, A., Croux, C. (2006) Bagging and boosting classification trees to predict churn. *J Marketing Res*, 43, 276–286.

[12] Lima, E., Mues, C., Baesens, B. (2009) Domain knowledge integration in data mining using decision tables: Case studies in churn prediction. *J Oper Res Soc.*, 60, 1096–1106.

[13] Lu, N., Lin, H., Lu, J., Zhang, G. (2014) A customer churn prediction model in telecom industry using boosting. *IEEE Trans. on Industrial Informatics*, 10, 1659–1665.

[14] Neslin, S., Gupta, S., Kamakura, W., Lu, J., Mason, C. (2006) Detection defection: Measuring and understanding the predictive accuracy of customer churn models. *J Marketing Res*, 43, 204–211.

[15] Pushpa, Shobha, D.G. (2012) An efficient method of building the telecom social network for churn prediction. *IJDKP*, 2, 31–39.

[16] Radosavljevik, D., Putten, P., Larsen, K.K. (2010) The impact of the experimental setup in prepaid churn prediction for mobile telecommunications: What to predict, for whom and does the customer experience matter? *Transactions on Machine Learning and Data Mining*, 3, 80–99.

[17] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B. (2012) New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *Eur J Oper Res*, 218, 211–229.

[18] Farquard, M.A.H., Ravi, V., Raju, S.B. (2014) Churn prediction using comprehensible support vector machine: An analytical CRM application. *Appl Soft Comput*, 19, 31–40.

[19] Coussement, K., Poel, D.V.D. (2008) Churn prediction in subscription services: An application of support vector machines while comparing two parameter selection techniques. *Expert Syst Appl*, 34, 313–327.

[20] Tang, Z., MacLennan, J. (2005) *Data-mining with SQL Server*. John Wiley & Sons, U.S.

[21] Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., Abbasi, U. (2014) Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, 24, 994–1012.

[22] Fathian, M., Hoseinpoor, Y., Minaei-Bidgoli, B. (2016) Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. *Kybernetes*, 45, 732 – 743.

[23] Lee, E.B., Kim, J., Lee, S.G. (2017) Predicting customer churn in mobile industry using data mining technology. *Ind Manage Data Syst.*, 117, 90 – 109.

- [24] Li, H., Yang, D., Yang, L., Lu, Y., Lin, X. (2016) Supervised Massive Data Analysis for Telecommunication Customer Churn Prediction. Proceedings of BDCloud-SocialCom-SustainCom16, Atlanta, GA, USA, 8-10 October, pp. 163-169, IEEE, USA.
- [25] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., Huang, K. Customer Churn Prediction in Telecommunication Sector using Rough Set Approach. Advance Access published December 3, 2016, 10.1016/j.neucom.2016.12.009.
- [26] Gordini, N., Veglio, V. Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter selection technique in B2B e-commerce industry. Advance Access published August 16, 2016, 10.1016/j.indmarman.2016.08.003.
- [27] Poel, D.V.D., Lariviere, B. (2004) Customer attrition analysis for financial services using proportional hazard models. *Eur J Oper Res*, 157, 196-217.
- [28] Bisoi, R., Dash, P.K. (2014) A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter. *Appl Soft Comput*, 19, 41-56.
- [29] D17.2. (2002) Churn analysis case study. Telecom Italia Lab, Torino, Italy.
- [30] Verbeke, W., Martens, D., Baesens, B. (2014) Social network analysis for customer churn prediction. *Appl Soft Comput*, 14, 431-446.
- [31] Liu S, He X. (2015) PIPE+ Verifier: A Tool for Analyzing High-Level Petri Net. In: *International Conference on Software Engineering and Knowledge Engineering*; Pittsburgh, USA.
- [32] Liu, S., Zeng, R., Sun, Z., He, X. (2014) Bounded model checking high-level Petri nets in pipe+ verifier. Proceedings of ICFEM14, Luxembourg, Luxembourg, 3-5 November, pp. 348-363, Springer International Publishing, Switzerland.