PREDICTION AND CAUSALITY ANALYSIS OF CHURN USING DEEP LEARNING

Muzaffar Shah, Darshan Adiga, Shabir Bhat and Viveka Vyeth

Datoin Bangalore, India

ABSTRACT

In almost every type of business a retention stage is very important in the customer life cycle because according to market theory, it is always expensive to attract new customers than retaining existing ones. Thus, a churn prediction system that can predict accurately ahead of time, whether a customer will churn in the foreseeable future and also help the enterprises with the possible reasons which may cause a customer to churn is an extremely powerful tool for any marketing team. In this paper, we propose an approach to predict customer churn for non-subscription based business settings. We suggest a set of generic features that can be extracted from sales and payment data of almost all non-subscription based businesses and can be used in predicting customer churn. We have used the neural network-based Multilayer perceptron for prediction purposes. The proposed method achieves an F1-Score of 80% and a recall of 85%, comparable to the accuracy of churn prediction for subscription-based business settings. We also propose a system for causality analysis of churn, which will predict a set of causes which may have led to the customer churn and helps to derive customer retention strategies.

KEYWORDS

churn Analysis, Causality Analysis, Machine Learning, Business Analytics , Deep Neural Network

1. INTRODUCTION

Large enterprises in the competitive market mostly rely on their existing loyal customers as their large chunk of business is coming from these customers. As the market is becoming saturated day by day, the enterprises have realized that they need to focus more on retaining the existing customers [8]. While acquiring new customers is the backbone of the business growth but marketing experts suggest that equal importance should be given to retention policies[9]. As many marketing researchers suggest that retention rate can result in a significant impact on business[10]. But sometimes a customer may churn out from the enterprise which will be a loss for a company. In most of these cases, they give some prior signal before actually getting churned. Thus the main focus of the churn prediction system should be analyzing the customer behavior, such that they will identify those types of customers before they churn out and recommend some of the causes which may have led to churn. These things will help the marketing team of enterprises to adopt proper retention strategies and may help in increasing the average customer lifetime value. In turn, it will help in increasing the company's business value

in the market, as proposed by [12] that the market value of the company is a function of customer lifetime value.

First, we propose a set of generic features which can be used for all most all non-subscription business settings for developing churn prediction system. Then we present a system which will predict churn probability score for nonsubscription business settings. Finally, we introduce two methods for causality analysis to predict the possible causes that act as critical reasons for a customer to churn.

2. RELATED WORK

In recent years there has been a lot of work in this field and many systems have been proposed which will help in predicting customer churn. The advent of deep learning has also contributed to increased research in this direction. Surprisingly, there is a lack of substantial work in the direction of causality analysis of customer churn.

[4] has shown the effective use of Deep Learning to predict customer churn specifically in the Mobile Telecommunication Network. Their work concludes that a medium scale deep learning system is sufficient to predict customer churn in cellular network services, emphasizing more on unsupervised feature engineering involved in the process. Their proposed system targets generic feature vectors applicable to almost all subscription-based companies. (Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network by Federico Castanedo) [7] are proposing a data preparation architecture that automatically comes up with more complex features and representation for the input data in the prepaid telecommunication industry. Their motivation to investigate and consider the application of deep learning as a predictive model is to avoid time-consuming feature engineering effort and ideally to increase the predictive performance of previous models. Another Deep Learning-based customer churn prediction has been proposed by [5].

There has been extensive work on conventional machine learning algorithms as well. Wei and Chiu, 2002 a proposed churn prediction system for telecommunication businesses using decision trees [3]. Burez and Van den Poel (2006) have built a churn prediction system for European pay-TV company, using the random forest algorithm [2]. Coussement and Van den Poel (2008) have applied a support vector machine for a newspaper subscription churn prediction [1].

The above systems use both conventional ML and deep learning algorithms for the churn prediction problem and most of which provide sensible results on real-world problems. The commonality among these related work is that they model churn prediction as a classification problem. So we decided to look at churn prediction as a classification problem. However, there has been very limited work on feature engineering particularly for non-subscription based businesses and none of the related work has explored the causality analysis of customer churn.

In this paper we will focus on non-subscription based business settings and will be proposing a churn prediction system, emphasizing more on the process of feature engineering. We also propose methods for causality analysis of customer churn.

3. METHODOLOGY

In the coming sections, we explain the process of feature engineering for churn prediction and we show how deep learning is used for the churn classification.

In this section, we will define important terms and definitions. Then we will give an introduction to the dataset used. After that, we will discuss the proposed feature engineering techniques, followed by a discussion on the proposed neural network used for predicting churn. In the end, we will be discussing theoretical definitions of causality analysis and how we are predicting causes of customer churn.

The churn prediction system is a data classification problem, in which we are given the historical data about customers and retention period and we have to use it to train a model which should be able to predict before a foreseeable time whether a customer will churn or not.

3.1. Type of businesses based on customer purchase behaviour

Type of business and business models has a great impact on the behavior of customers and their life cycle, which in turn affects the possibility of the customer getting churned out. Based on customer purchase behavior, there are four types of business settings:

• Contractual and discrete business:- In these types of businesses, purchase interval and sale amount both are fixed. They are also called the subscription-based business. Example:- Postpaid telecommunication, Netflix

• Contractual and continuous business:- Where purchase interval is fixed but the sale amount can vary. Example:- Credit card businesses

• Non-contractual and discrete business:- When the purchase interval can vary but the sale amount is fixed. Example:- Oil and Gas retail businesses

• Non-contractual and continuous business:- Where both sale amount and purchase interval can vary. Example:- Grocery stores, Retailer businesses

 Further, each of these business categories can be classified as business-to-business or business-to-consumer. However, the methods proposed in this work apply to both of these sub-categories of businesses. Most of the churn prediction systems until now have been proposed for Contractual and discrete and Contractual and continuous types of businesses. In this paper, we will be focusing on business type Non-contractual and continuous business but our system can be generalized to other types of businesses as well.

3.2. Payment and sales data

Different enterprises have different representations and structures of sales and payment data in their system. Also, the set of fields in the data representing the sales and payment will vary from business to business. Despite the diversity in the schema and representations of the data, at the least, every enterprise will have data about sale and payment of a dealer, per day, per month or on

some other time granularity. The proposed method makes use of this minimal set of data to generate enough features to build a basic churn prediction system.

As mentioned earlier, churn prediction is a typical classification problem in which we generate training data from historical sales and payment data, one sample per customer and assign a label to each customer whether he has churned or not. We have used sales and payment data of a business-to-business large-scale non-contractual and continuous business in all of our experiments.

The consumer of this business is called the dealer who acts as a tradesperson between the business and the end-user. In our experiments, we intend to predict the churn of these dealers based on their sales and payment data.

4. FEATURE ENGINEERING

It is often said that "data is the fuel of machine learning". This is not quite true: data is like the crude oil of machine learning which means it has to be refined into features or predictor variables, to be useful for training a model. Without relevant features, you can not train an accurate model, no matter how complex the machine learning algorithm. Algorithms are pretty naive by themselves and cannot work out of the box on raw data. Hence the need for engineering meaningful features from raw data is of utmost importance which can be understood and consumed by these algorithms.

The process of extracting features from a raw dataset is called feature engineering. Though one of the main concerns of deep neural networks is to automate the process of feature engineering (Philip Spanoudes, Thomson Nguyen [5]), i.e to shift the burden of preparing and processing of feature vectors to the underlying learning system itself. But this does not mean that the data preprocessing, feature extraction, and feature engineering are irrelevant when one uses deep learning, particularly when we are dealing with direct numeric data like sales, finance, sensory data, etc. We cannot just feed any type of sales data to the neural network and expect it to generate the churn prediction problem.

In our work, we propose a set of generic features that can be extracted from sales and payment data of almost all non-subscription based businesses. The proposed feature set in our method is based on a well-known marketing technique called RFM analysis. RFM analysis, which is an abbreviation for Recency, Frequency, Monetary, is a marketing technique used to determine quantitatively which customers are the more valuable ones, by examining how recently a customer has purchased (recency), how often they purchase (frequency), and how much the customer spends (monetary) [5]. RFM analysis is based on the marketing axiom that "80% of your business comes from 20% of your customers".

We derive a set of features that represent the RMF properties of sales and payment data. Since almost all the businesses capture the sales and payment data to capture the RMF information, our feature engineering steps can be reproduced very easily. We use below set of fields that can be seen in sales and payment datasets, for the feature generation

Table 1. Common fields used from Sales data.

Field	Description
Sales date	The date on which sale has happened
Customer id	Unique business id of the dealer
Amount	Sale on that date

We use monthly granularity in our experiments because the monthly sale and payment of the business data we are using are sufficient for our system to generate features. Using the above common fields from Sales and Payment data we derived below features per dealer.

Table 2. Common fields used from Payment data.

Field	Description
Payment date	The date on which payment has happened
Customer id	Unique business id of the dealer
Amount	Sale on that date

Table 3. Features used from sales data

S. No.	Feature Name	Description	Importance
1	Overall sales of dealer	The overall lifetime sale of the dealer	Indicates the importance of a dealer based on his contribution to revenue. This represents the monetary part of RMF
2	The monthly average sale of dealer	The lifetime monthly average sales of the dealer	Represents the monetary part of RMP
3	Overall sales growth	Sales growth of dealer	Defines whether the business with the dealer is increasing or decreasing. This represents the monetary part of RMF
4	Average sale gap in days	Sales gap indicates how often the dealer is buying from the enterprise	Represents the frequency factor of RMF
5	Overall sales frequency growth	Defines whether the sales gap is increasing or decreasing and the quantity of changed sales frequency	Represents the frequency part of RMF
6	The recent average sale of dealer	Average sales of the dealer in the recent period	Represents recency part of RMF
7	Relative recent sales	The ratio of recent average sale and overall average sale	Represents recency part of RMF
8	Relative recent sale gap	The ratio of recent average sale gap and overall sales gap	Represents recency part of RMF
9	Recent average sale gap in days	Average sales gap of the dealer in the recent period	Indicates the recency part of RMF

| 10 | Recent sale growth of dealer | Sales growth in the recent period | Indicates the recency part of RMF |
| 11 | Recent sales frequency growth | This defines whether the sales gap is increasing or decreasing in the recent period | Indicates the recency part of RMF |

Table 4. Features used from payment data

S. No.	Feature Name	Description	Importance
1	Overall payment of the dealer	The lifetime payment of the dealer	Indicates the importance of the dealer based on revenue and it also represents the monetary part of RMF
2	The average payment per month of dealer	The lifetime monthly average payment of the dealer	Represents the monetary part of RMP
3	Average payment gap in days.	Payment gap indicates how often the dealer is paying	Indicates the frequency part of RMF
4	Overall payment gap growth.	Defines whether the payment gap is increasing or decreasing and the quantity of changed payment	Represents the frequency factor of RMF
5	Recent Payment of dealer	Payment received from the dealer in the recent period	Indicates the recency part of RMF
6	Relative recent average payment	The ratio of recent average payment and overall average payment	Represents recency part of RMF
7	Recent average payment gap in days	Payment gap in the recent period	Indicates the recency part of RMF
8	Relative recent average payment gap	The ratio of recent average payment gap and overall payment gap	Represents recency part of RMF

We calculate the growths by approximating a line on monthly data(sales, sales gap, payment, etc) of the dealer and then calculating the slope of that line. For example, if the slope of the line on the monthly sales data is positive then sales growth is positive and vice-versa. For creating training data, we consider the dealers who have not done any sales transactions from the last 120 days as churned dealers. The logic is also described in [11]. This number of days can be customized according to different use cases and particularly depends upon the average sales gap of the customer in that business.

The target field in our experiments indicates whether a customer has churned or not. To overcome the unavailability of this data we have chosen a novel solution. The data we have used contains sales and payment transactions from April 2016 to December 2018. The dealers who have not transacted from 30th September are considered as churned. One important point to note here is that we should not have to take the last transaction date of the dealer as a training feature, because we are deriving our target field from that date and also, this will not be available during inference. Also, in all the above-mentioned features, recency is considered as 6 months, it can also vary from problem to problem.

4.1. Churn prediction using Feed-forward neural network

Most of the related work has achieved remarkable accuracy in churn prediction [5, 6] using neural networks. We used 5 layers (MLP) feed-forward back-propagation neural network, with the input layer containing all of the input fields or features used to predict the outcome variable. The output layer contains an output field which is the target of the prediction. We have used the TensorFlow library in our neural network experiments. The architecture of the neural network used for churn prediction is mentioned in Table 5.

Table 5. Network architecture.

Hidden layers	200
Activation function	Relu in hidden layers and softmax in the output layer
Learning rate	0.01
Optimization Algorithm	Sigmoid Cross entropye

Exploring the most suitable algorithms or more complex networks for churn prediction is out the scope of our work as we are emphasizing more on feature engineering and causality analysis of churn.

5. EXPERIMENTS AND RESULTS

We have used sales data of the enterprise which has around 6000 customers, leading to 6000 training examples. In the data set, 80% of customers were still active and around 20% were already churned as per 120 days condition already mentioned in section 3.3. We have divided the data into three sets by random sampling: 5000 samples in the training set, 500 samples for evaluation set and the remaining 500 for the test set. Evaluation set is used to tune the network and algorithm to maximize the accuracy and other metrics while a test set is used to get the accuracy of the system on unseen data. Precision for churn class indicates the confidence of the model in predicting the churn. The recall represents the percentage of predicted churned customers out of the total churned customers. The F1-score is the weighted sum of the precision and recall providing a balanced score to evaluate the churn model.

We trained the network for 500 epochs and after a certain number of experiments we came up with the best configuration as mentioned in the earlier section, which resulted in an accuracy of 79.65% on the test set. Evaluation metrics of the best configuration are given in Table 6.

Table 6. Evaluation of the Churn analysis model

	Non-churn (%)	Churn class(%)
Precision	77.47	71.18
Recall	85.29	89.52
F1 Score	80.63	63.38

Generally, the number of churn events will be far less than the number of non-churn events. A representative of such a population, our training data also has fewer churn samples than the non-churn samples, making the data biased. Thus, the results are favorable towards non-churn class than the churn class. Apart from just classifying a dealer as churned or non-churned, the output of the model represents the probability of churning. This will be useful for prioritizing the dealers based on their probability of churning out, for further marketing analysis.

As per our literature study, there has been no work related to churn prediction for non-subscription based businesses and there is no benchmark dataset to compare our results with.

6. CAUSALITY ANALYSIS AND PREDICTING CAUSE OF CHURN

6.1. Introduction

Causality analysis can be defined as a process of predicting the root cause of an event. Using causality analysis we can predict the probability of a factor is a cause of certain events. Causal analysis is different from the normal regression analysis. In regression analysis, the goal is to develop a model for making predictions about the dependent variable, based on the observed values of the independent variables whereas in causal analysis, the independent variables are regarded as causes of the dependent variable.

The causal study aims to determine whether a particular independent variable affects the dependent variable and to estimate the magnitude of that effect if any. The independent variable which affects the dependent variable or may play a role in the occurrence of an event can be called a causal variable. If the variables simultaneously increase or decrease with the dependent variable but it does not cause the occurrence of the dependent variable, it may be called just a symptom of event.

In churn prediction, the causal variables are those types of features that cause customers to churn. While these causality features will vary from business to business, for the particular business setting considered in this work, the following factors can be considered as major causal factors:

1. Number of Complaints
2. Salesman or point of contact changed
3. Orders canceled due to understock
4. Returns due to defective material

6.2. How to identify the possible causes

During the data collection and preparation phase, the business representative, who has good knowledge of the business, is asked to identify the factors which may lead to churn. For instance, in this particular example, the factors identified are the ones mentioned above. Now the system

will predict the contributing score for each factor for a given customer who has a high churn risk score as per the churn prediction algorithm. The data set used in this section for experimentation is the same as used in the churn prediction system above.

6.3. Methods for predicting causality

Literature suggests different methods for generic causality analysis such as Bayesian Causal networks based model [24] [25], Structural Equation Models (SEM) [13] [14][21] [22] [23], Counterfactuals based model [13] [14] [15]. In this work, we will be proposing two models specifically for causality analysis of customer churn: Counterfactuals based causal model and predicting cause by using Bayes theorem.

6.3.1. Counterfactuals based causal model

In this model, we backtrack the event and try to undo the effects of a possible cause and then re-run the experiment [16] [17]. The goal is to find if undoing the effect will change the magnitude of the event or will not stop the event from happening [18] [19] [20]. In our example the model will work as:

• Take all the customers for whom our churn prediction system have proposed a high churn probability

• Now for every customer take all the possible causal factors and try to undo their effect one by one in the data (data used for generating features for churn)

• Now after the change in data, regenerate the features for that customer and re-calculate the churn probability of that customer

• If the recalculated churn probability is more or equal to actual probability, then we can say that this is not the major cause of churn

• But if this is less than previous churn probability, then we can say that this is one of the causes of churn magnitude of the cause can be calculated from the difference of churn probability

Example of Counterfactual method Let's consider a bunch of customers who have high churn probability as predicted by our churn prediction system. Now let us calculate the effect of the causal factor 'orders getting blocked' on the possible churning of these customers, so as a first step we have to try to undo the effect of blocked orders. Thus we will assume that the orders were not blocked, so we will add the blocked order amount to his sales data. With the change in sales data, payment data is also going to get affected, so we will try to add the same amount periodically in payment data as well. Thus the data features (used for predicting churn) have changed. So, we will apply our churn prediction on the manipulated data (feature sets). If the probability of churn gets increased or remains unchanged we can say the factor 'orders blocked' was not the causal factor, but if the probability decreases we can claim that the blocking of orders was one of the causes for customer to get churned and the magnitude of the probability of the cause can be calculated by difference of churn probabilities. One important thing to mention here is that the possibility of going through and undoing the changes is not possible in every situation with high accuracy. So, for the cases where we found it difficult to undo the effect of the cause,

like the ` change in point of contact or salesperson', some other causal model like Bayes theorem may come to the rescue.

6.3.2. Using Bayes theorem to find the probability of cause

Let us assume that y is an event that a particular cause has led to the customer churn and let x represent that customer c has been churned with churn score Pc. Thus p(y | x=Pc) represents that a particular factor being the possible cause of churn for the customer c with churn score Pc. So, the problem remains to calculate p(y | x=Pc) According to the Bayes theorem [26] , the posterior probability of an event given the predictor can be calculated as:

$$p(y|x) = (p(x| y) * p(y))/p(x)$$

Where

• p(y) is the prior probability of an event.

• p(x | y) is the likelihood which is the probability of predictor given event.

• p(x) is the prior probability of predictor.

• We can further apply a Gaussian distribution [27] calculation to calculate the probability distribution of a continuous variable as shown below.

$$PDF(x, mean, sd) = (1 / (sqrt(2 * PI) * sd)) * exp(-((x-mean^2)/(2*sd^2)))$$

Example of the Bayes method In this example, let us consider one of the above-mentioned causes like "blocked orders due to insufficient stock" and let us assume that the blocked orders due to insufficient stock will be considered as the cause, if the block per month is 25% of order for a dealer for more than 6 months. Let y be the event that blocked orders is the cause of churn (with above-mentioned conditions) for a customer c and x be predictor indicating that the customer c has been churned by churn probability Pc. Therefore p(y | x=Pc) can be defined as the probability of blocking of the orders being a cause of churn for a given customer whose churn probability score is Pc.

To calculate the above probability by Bayes theorem, we need to calculate p(x=Pc | y) from the data, which means we have to calculate the probability of a customer having a churn score of Pc given his orders have been blocked (with the above condition). Since Pc is a continuous value we need to calculate the probability distribution of using the Gaussian distribution formula. From the data, we have to calculate the mean and standard deviation of churn score having 25% of blocked orders for more than 6 months.

Given a churn score Pc of a customer c, we will use that in equation (2) with mean and standard deviation calculated from the data and we will get p(x=Pc | y). From the data, we also have to calculate p(y), which indicates the probability that a customer's order is getting blocked. We also have to calculate p(x) which is the posterior probability of the predictor. The p(x) indicates the probability of customers having a given churn score Pc. Since this is also a continuous variable, so we will use the Gaussian distribution formula to calculate the value. For this, we have to calculate the mean and standard deviation of the churn score of customers in data. Given the

churn score Pc of the customer, we can calculate the p(x). Using the above-calculated values in equation (2), we can get the probability of event say 'block being a cause' given a churn score Pc. If the probability is greater than 0.5 we can say it is a possible cause.

Unlike the counterfactual method, the Bayes method can be applied to almost all the scenarios. Bayes method of finding the cause is more applicable when we are interested in the magnitude of change in probability score.

6.4. Result of Causality analysis

We have experimented with both the proposed methods using our data on two different causes of customer churn. For order blockage (order blockage being the cause) we have used counterfactuals and found that if we unblock these orders (as proposed above) it will change the payment and sales data of these customers which in turn changes the features used for churn prediction.

When we used these new feature vectors to predict the churn, around 45% of customers, whose orders were blocked and had predicted as churned before unblocking, have been predicted as non-churned after unblocking. We also found that there has been an average reduction of 0.3 in the overall churn score of customers using counterfactuals on blocked orders.

For the cause of 'returns due to defect', we used the Bayes method as proposed above and found that about 64% of customers who returned their orders due to defect and have churned have 'returns due to defect' as a possible cause of churning. The causality analysis of our dataset shows the below results.

Table 7. Results of causality

Cause	Method	Average Change in Churn Score	Percentage of Customers churned due to the cause
Order Blockage	Counterfactual	0.3	45%
Order returned due to defect	Bayes Theorem	-	64%

The above results are specific to a particular dataset and business and can not be generalized. Also, the magnitude of the numbers does not depict the betterness of method or algorithm, instead they act as strong significance of various causes leading to churn in that particular dataset. As per our knowledge, the causality analysis of customer churn has not been explored untill now and there is a lack of baseline experiments to compare against.

7. CONCLUSION

Churn prediction and customer retention is an important problem in the subscription as well as non-subscription based businesses and can have a considerable impact on the business if ignored. In this paper, we have investigated and developed a generic representation of feature engineering steps that could be applied to any non-subscription based business for predicting churn. We have also proposed an MLP network that is used to predict the churn probability of customers using the

data prepared though the above feature engineering process. Furthermore, based on the prediction results, it has been proved that we have achieved considerable good results targeted specifically for non-subscription based business.

The novel feature engineering process proposed by this paper has shown significant results in the churn prediction. In the second part of this paper, we have proposed two different methods to predict the possible causes of customer churn which would help the businesses to improve their customer relationship by knowing the possible causes of customer churn.

Extending the feature engineering techniques described and the use of business-specific datasets other than just payment and sales data to improve results could be a possible future work. Other causality analysis techniques, particularly for churn analysis, need to be explored further. The proposed system of causality analysis can be applied to different domains like machine failure detection, classification and diagnosis of patients, etc and there is a huge scope for experimenting with different causality analysis algorithms like Bayes Casual Network, SCM, etc.

REFRENCES

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

[2]   Burez, J., Van den Poel, D. (2006) CRM at a Pay-TV Company: Using analytical models to reduce customer attrition by targeted marketing for subscription services, Expert Systems with Applications, 32(2), 277-288.

[3]   Wei, C. P., Chiu, I. T. (2002). Turning telecommunications call details to churn prediction: A data mining approach, Expert Systems with Applications, 23, 103-112.

[4]   Anuj Sharma and Dr Prabin Kumar A Neural network-based Approach for Predicting Customer Churn in Cellular Network Service, International Journal of Computer Applications (0975 - 8887)

[5]   APhilip Spanoudes and Thomson Nguyen Deep Learning in Customer Churn Prediction: Unsupervised Feature Learning on Abstract Company Independent Feature Vectors by Philip Spanoudes and Thomson Nguyen , https://arxiv.org/pdf/1703.03869.pdf

[6]   Anuj Sharma and Dr Prabin Kumar A Neural network-based Approach for Predicting Customer Churn in Cellular Network Service, International Journal of Computer Applications (0975 - 8887)

[7]   AYen-Liang, Mi-Hao, Shin-Yi and Kwei Tang Discovering recency, frequency, and monetary (RFM) sequential patterns from customers purchasing data , Electronic Commerce Research and Applications journal (https://fardapaper.ir/mohavaha/uploads/2017/12/FardapaperDiscovering-recency-frequency-and-monetary-RFM.pdf)

[8]   AYen-Liang, Mi-Hao, Shin-Yi and Kwei Tang Using Deep learning to predict customer churn in Mobile telecommunication network , http://wiseathena.com/

[9]   Hadden, J., Tiwari, A., Roy, R., Ruta, D. (2005) German mobile cellular telecommunications market. Telecommunications Policy, 25, 249-269. , Computer assisted customer churn management: State-of-the-art and future trends. Computers Operations Research, 34, 2902- 2917.

[10]   Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E., Kaushansky, H. (2000) Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry , . IEEE Transactions and Neural Networks, 11, 690- 696.

[11] M. Rowe, Mining User Development Signals for Online Community Churner , ACM Transactions on Embedded Computing Systems, vol. 5, no. 1, 2015.

[12] Gupta, S., Lehmann, D.R., Stuart, J.A., 2004 Valuing customers. , Journal of Marketing Research 41 (1), 7-18.

[13] Peter Spirtes, 2010 Introduction to Causal Inference by Peter Spirtes. , Journal of Machine Learning Research 11 (2010) 1643-1662 Submitted 2/10; Published 5/10

[14] Finnian Lattimore and Cheng Soon Ong 2018 A Primer on Causal Analysis, arXiv:1806.01488v1 [cs.LG] 5 Jun 2018

[15] JOHN R.ANDERSON Causal Analysis and Inductive Learning , Proceedings of the Fourth International Workshop on MACHINE LEARNING June 22-25, 1987 University of California, Irvine 1987, Pages 288-299

[16] PR Rosenbaum and DB Rubin. The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1):41-55, 1983. URL http://biomet.oxfordjournals.org/content/70/1/41.short.

[17] DB Rubin C Estimating causal effects of treatments in randomized and nonrandomized studies. , Journal of educational Psychology, 1974. URL http://psycnet.apa.org/journals/edu/66/5/688/.

[18] DB Rubin. Bayesian inference for causal effects: The role of randomization , The Annals of Statistics, 1978. URL http://www.jstor.org/stable/2958688.

[19] DB Rubin. Causal Inference Using Potential Outcomes. , ournal of the American Statistical Association, 100(469):322-331, mar 2005. ISSN 0162-1459. doi: 10.1198/016214504000001880. URL http://www.tandfonline.com/doi/abs/10.1198/016214504000001880.

[20] DB Rubin. For objective causal inference, design trumps analysis. , The Annals of Applied Statistics, 2(3):808-840,sep 2008. ISSN 1932-6157. doi:10.1214/08-AOAS187. URL http://projecteuclid.org/euclid.aoas/1223908042.

[21] S Wright. Correlation and causation. , PJournal of agricultural research, 1921. URL http://www.ssc.wisc.edu/soc/class/soc952/Wright/Wright팴표푯푯팺푈푦푈푯푯푈푇푯푈푇푯팴푦푤푇 푟푦푇푯푯표푈.푷푈푯.

[22] T Haavelmo. The statistical implications of a system of simultaneous equations. , Econometrica, Journal of the Econometric Society, 11(1):1- 12, 1943. URL http://www.jstor.org/stable/1905714.

[23] Judea Pearl. Causality: models, reasoning and inference. , MIT Press, Cambridge, 2000.

[24] David Heckerman and Dan Geiger. Learning Bayesian networks: a unification for discrete and Gaussian domains. In Philippe Besnard and Steve Hanks, editors, , Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, pages 274-282. Morgan Kaufman, 1995.

[25] Yimin Huang and Marco Valtorta Identifiability in causal Bayesian networks: A sound and complete algorithm. , In Proceedings of the Twenty-First National Conference on Artificial Intelligence, pages 1149-1154, Edinboro, Scotland, 2006. AAAI Press

[26] Lawrence M. Rudner and Tahung Liang Automated Essay Scoring Using Bayes' Theorem , THE JOURNAL OF TECHNOLOGY, LEARNING AND ASSESSMENT. VOL 1 NO 2 (2002)

[27] G. Moser ; J. Zerubia ; S.B. Serpico. SAR amplitude probability density function estimation based on a generalized Gaussian model , PIEEE Transactions on Image Processing Volume 15 Issue 6 âĂć June-2006.