

A SURVEY ON CHURN ANALYSIS AND PREDICTION IN VIDEO ON DEMAND**VENKATRAMAN R, RAMESH RAGALA**

Department of CSE, School of Computing Science and Engineering, Vellore Institute of Technology Chennai Campus, Chennai, Tamil Nadu, India. Email: asha.s@vit.ac.in

Received: 23 January 2017, Revised and Accepted: 03 March 2017

ABSTRACT

Consumer loyalty is a key measure of achievement. Despondent clients will not be staying around the service. When there are unhappy clients once in a while voice their disappointment before leaving. Streaming administration, motion pictures, and television shows are gushing over the internet, not being downloaded, so we should be associated with the internet all through your watch instantly experience. Hence, to help them recognize to "disappointed clients" from the get-go in their relationship. Doing as such would permit streaming administration to find a way to enhance "client's joy" before it's excessively late. To distinguish the purposes behind clients who are producing from the spilling administration furthermore anticipating what number of clients will get stick around the gushing administration.

Keywords: Customers, Streaming, Churn.

© 2017 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2017.v10s1.19603>

INTRODUCTION

Today, web-based gushing administration showcase everywhere worldwide is confronting a serious loss of income because of furious rivalry and loss of potential clients. To keep the upper hands and to obtain whatever number clients as would be prudent, most administrators contribute a gigantic measure of income to extend their business in the absolute starting point. In this manner, it has ended up fundamental for the gushing administrations to obtain the sum contributed and to pick up no less than a base benefit inside a brief time frame. Since it is especially testing and repetitive issue to keep the clients in place for a long length because of the opposition required around here field. To make due in the market, internet gushing administrations generally offer an assortment of maintenance approaches to pull in new clients. This is the real reason for the endorsers abandoning one system and moving to another which suits their requirements. As indicated by spilling administrations, the procedure of supporters changing from keeps on happening for any web based gushing administrations; it would prompt to the considerable loss of income to the organization. In this circumstance, the main solution for defeat such business perils and to hold in the market, administrators are compelled to search for option methods for utilizing information mining strategies and measurable devices to distinguish the cause ahead of time and to take quick endeavors accordingly. This is conceivable if the previous history of the clients is examined deliberately. Luckily, spilling administrations create and keep up an extensive volume of information. They incorporate client data, see data and network information. These measurable apparatuses are to recognize the cause ahead of time and to take quick endeavors accordingly. This voluminous sum information guarantees the degree for the use of information mining procedures in media transmission database. As a lot of data are covered up in the information created by the web-based gushing administration industry, there is a great deal of degree for the analysts to break down the information in alternate points of view and to help the spilling administrations to enhance their business in different ways. The most widely recognized ranges of research in gushing administration databases are comprehensively ordered into three sorts, (i) streaming fraud detection, (ii) web-based streaming churn prediction, and (iii) network fault identification and isolation. Furthermore, not every one of the information things of the streaming administrations database is utilized by every one of the methods. Just the applicable information things which truly add to the particular investigation must be considered for any study. This study

concentrates on stir expectation, the significance of highlight extraction (trait determination) and the utilization of information mining methods in beat forecast in web-based spilling administrations information.

CHURN PREDICTION-PROBLEM STATEMENT

In a business situation, the term, client steady loss just alludes to the clients abandoning one business administration to another. Client stir or supporter beat is likewise like wearing down, which is the procedure of clients changing starting with one administration supplier then onto the next namelessly. From a machine learning point of view, beat forecast is a regulated (i.e., named) issue characterized as takes after: Given a predefined estimate skyline, the goal is to foresee the future churners over that skyline, given the information connected with every endorser in the system. The beat forecast issue spoke to here includes three stages, to be specific, (i) preparing stage, (ii) test stage, and (iii) expectation stage. The contribution for this issue incorporates the information on past memberships for every portable endorser, together with all individual and business data that is kept up by the spilling administration supplier. Furthermore, for the preparation stage, marks are given as a rundown of churners. After the model is prepared with most elevated precision, the model must have the capacity to foresee the rundown of churners from the genuine dataset which does exclude any agitate name. In the point of view of learning disclosure handle, this issue is ordered as prescient mining or prescient demonstrating. Stir prediction is a wonder which is utilized to distinguish the conceivable churners ahead of time before they leave the online video gushing administration. This anticipates endorsers who are probably going to stir in future by taking the required maintenance strategies to pull in the possible churners and to hold them. In this way, the potential loss of the organization could be stayed away from. This study uses information mining procedures to recognize the churners.

ALGORITHM**Regression algorithm**

Regression is movement from one variable to another. The below diagram depicts the information show that the data are moved from one point to another by using the slope which is nothing but the regression line. The main algorithm used in this is logistic regression which is because when we want to analyze whether the churn has happened or not (i.e.,) if churn has happened then it is 1 else 0 that is churn has not happened (Fig. 1).

Naive Bayes

It is one among the classification technique. In addition, this algorithm contains some unidentified feature which will not be correlated to the feature which is not already present in it. Some features will be dependent on each other. The properties of these features are dependent on the probability of the class. One advantage of this is easy to build and is highly focused on larger data sets which suits for this. This is used to classify the customers based on the probability, so this is the best to outperform high classification methods among all. We will create a frequency table from that we will be finding out the likelihood table and then using those probability applying in the formula which has mentioned below in the figure based on that result, the one which is having highest posterior probability is the outcome of prediction. This classification is mainly used when there are many classes. It supports to identify the probability of various class based on different attributes. When the input variable is categorical rather than numerical variable, it will give the best results. When it is numerical variable, normal distribution is used which is called as bell curve. Finally, this algorithm is mainly used for doing prediction in real time (Fig. 2).

Decision trees

Decision tree is one kind of algorithm having the end variable which is mostly used in the case of classification. This decision tree works for both categorical and continuous input and output variables. According to this paper for creating a model to predict, who are the customers, will be churning out from the subscription video on demand. This is where decision tree helps that will segregate the customers based on the outcome of variable and identify the variable. Since we need to say whether customer will be churning out or not categorical variable decision tree is used. So now by identifying the important variable then we build a decision tree to predict the customer for the streaming video service (Fig. 3).

Artificial neural networks (ANN)

ANN is not used for studying brain but it is also used to build useful machines. ANN works on something referred as hidden states. This hidden state is referred to similar as neurons. Each of this hidden state is a transient form which has a probabilistic behavior (Fig. 4).

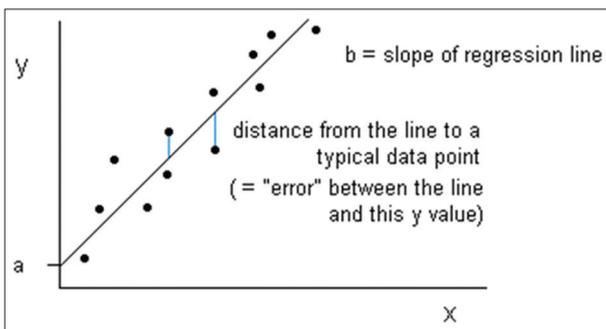


Fig. 1: Logistic regression

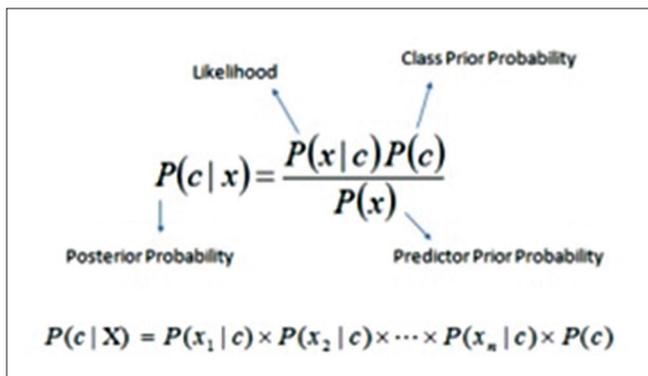


Fig. 2: Naive bayes

Support vector machines

In support vector machine, the data is plotted in the n-dimensional space. After plotting the data in the n-dimensional space, the data are splitted separating the different classes which involve method of supervised learning classification in the n-dimensional space. Based on that, we will be drawing a line which is called as hyperplane since we are drawing the line in the n-dimensional space. After drawing the hyperplane, we will see the classes which are having the highest margin. The classes which are best suited that is nearer to the hyperplane that is having more distance between nearest data point and the hyperplane (Fig. 5).

K-nearest neighbor (KNN)

In example acknowledgment field, KNN is a standout among the most vital non-parameter calculations and it is an administered learning calculation. The arrangement standards are produced by the preparation tests themselves with no extra information. The KNN order calculation predicts the test's class as indicated by the K preparing tests which are the closest neighbors to the test, and judge it to that classification which has the biggest class likelihood.

LITERATURE SURVEY

Data mining process

Data mining has risen over the pate years as a to a great degree intense way to deal with extricating important data from huge databases and data distribution centers. Since the expanded computerization of business exchanges, changes away and preparing limits of personal computers (PCs), and in addition huge advances in data discovery calculations, those all have added to the evolution of the data mining. The method of information mining sees the exposure of information from a database as a four-organize get ready. To begin with, the business issue must be perceived. After the issue is portrayed and related information is assembled, the accompanying step is to handle the accumulated information by information change, information cleaning, thus on for the later mining process. The third step is to apply some specific mining algorithm (s) over the took care of information. In this paper, estimate/portrayal figurines can be used. Finally, the mining result is evaluated to take a gander at whether the finding is useful for the business issue.

Data pre-processing

The data in this present reality are constantly fragmented, boisterous, and conflicting in light of not relevant, human or PC mistake at data section, blunders in data transmission, or from various data sources, and so on. In this manner, the real undertakings in data pre-processing incorporates data cleaning, data reconciliation, data change, data lessening, and data discretization.

Data cleaning is one of the three most serious issues in data warehousing (Kimball, 1996). In data cleaning process, a few errands might be to fill in missing qualities, recognize anomalies, smooth out loud data, rectify conflicting data, and resolve repetition brought about by data combination. Absent and boisterous data are settled by utilizing credit intend to fill as a part of, or utilizing a relapse capacity to locate a fitted esteem by and large.

Data incorporation is to consolidate data from different sources into a cognizant store. In incorporation preparing, the excess data issue dependably happen, since a similar true element, quality qualities from various sources have diverse names, or one trait might be an inferred property in another table. In this manner, analysts ought to precisely distinguish certifiable substances from different data sources by utilizing connection investigation. Something else, watchful combination of the data from various sources may diminish/keep away from redundancies and irregularities and enhance mining velocity and quality.

Since complex data mining may set aside a long opportunity to keep running on the entire data set, data lessening is normally utilized in

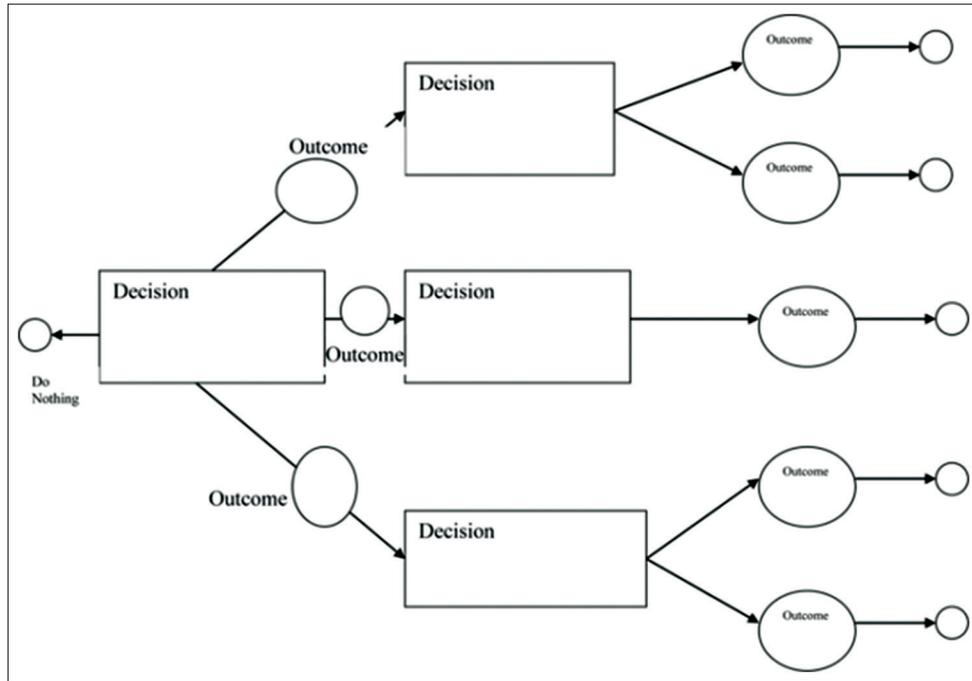


Fig. 3: Decision tree

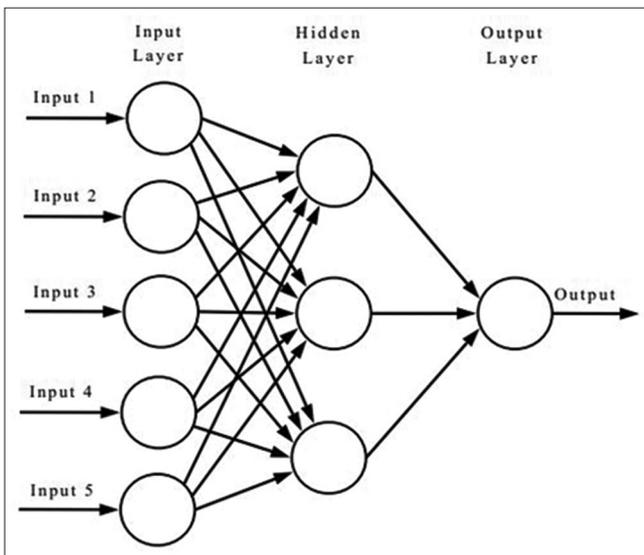


Fig. 4: Artificial neural networks

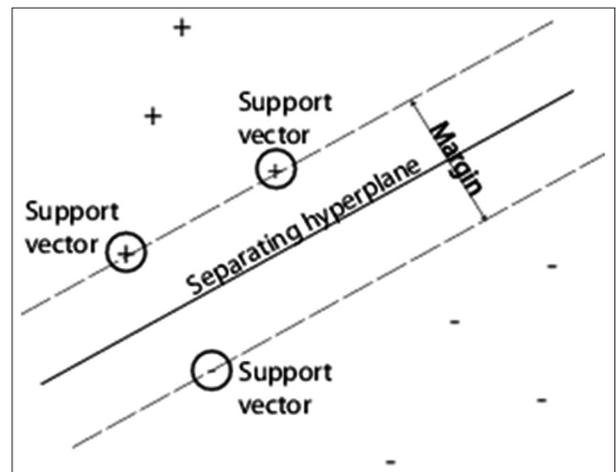


Fig. 5: Support vector machines

the data pre-processing stage. Data decrease errands incorporate data shape accumulation, measurements lessening (e.g., evacuating irrelevant properties), data pressure, and to diminish data volume by picking elective, littler types of data representation.

At long last, data discretization isolates the scope of a constant property into interims since some characterization calculations just acknowledge straight out traits. After data pre-processing, data investigation and mining can continue proficiently and successfully.

PROPOSED SOLUTION

This paper is about churn analysis in subscription in video on demand. Lots of customer takes a subscription for trial period for the video streaming service but after that trial period whether the customers will be staying around the streaming service or not we do not know. Either customer will be satisfied from the trial period of the subscription service and moving to paid subscription of the streaming service.

So for that, we are having customer information data, subscription information data and finally track services data. Here in this paper am using technology which is used to store the information. The customer information is stored completely in the Gigya, subscription video information is stored in the Youbora and finally the tracking information is stored in enterprise service bus (ESB). Gigya contains all the customer information like customer login information and information which is anything related to customer is stored in Gigya. Youbora contains all video subscription information like what are the videos are watched and how many video series are yet to be watched like all those information are stored in that. Now when it comes to ESB which contains all the tracking services like all the payment information and subscription information is stored like when is starting and ending timestamp of the trial period of the customer and about the customer information when the customer has started his trial period. Also whether the customer has taken the monthly subscription for the streaming service or he taken a yearly subscription (Fig. 6).

Now from that exploratory data analysis is done. After that, we start data preparation upon that building a machine learning model which

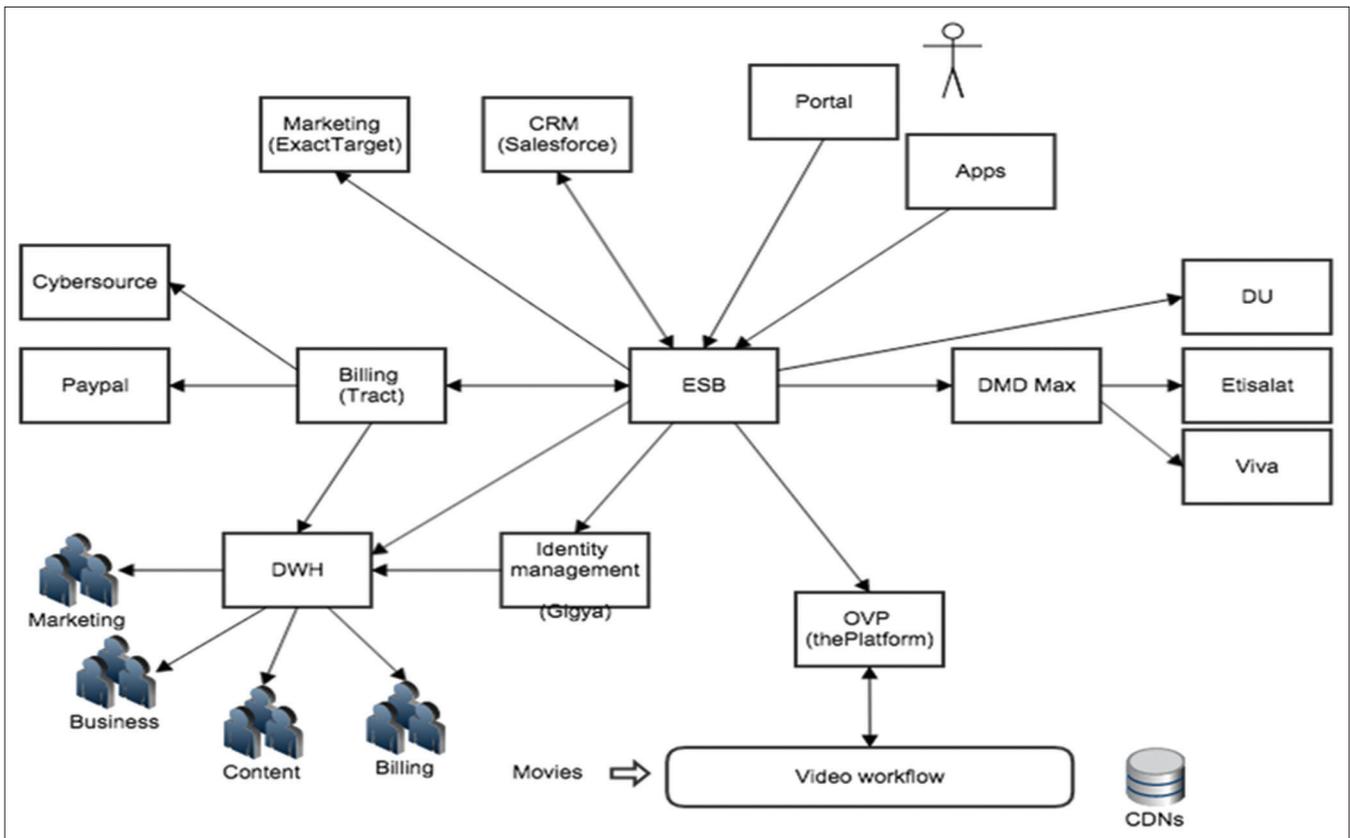


Fig. 6: Work flow

suits the best. So based on that model, we can identify the customers who are churning out from the video subscription service. Furthermore, we can identify who are the customers are staying around the video subscription for a long time so that we can group among them. Later after building the model, we are going to visualize the model using advanced data visualization like D3.

CONCLUSION

Maintenance of potentially agitating clients' has developed to be as vital for spilling administrations as the procurement of new clients. High beat rates and considerable income misfortune because of agitating have turned right stir expectation and aversion to an indispensable business handle. In spite of the fact that beat is unavoidable, it can be overseen and kept in adequate level. There are various methods for beat expectation and new systems keep on emerging. Great expectation models must be always created and a mix of the proposed systems must be utilized. Profitable clients must be recognized, therefore prompting to a mix of agitate forecast strategies with client lifetime esteem procedures.

REFERENCES

1. Au WH, Chan KC, Yao X. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Trans Evol Comput* 2003;7(6):532-45.
2. Den Poel DV, Larivière B. Customer attrition analysis for financial services using proportional hazard models. *Eur J Oper Res* 2004;157(1):196-217.
3. Domingos P. The role of Occam's razor in knowledge discovery. *Data Min Knowl Discov* 1999;3(4):409-25.
4. Hadden J, Tiwari A, Roy R, Ruta D. Churn prediction: Does technology matter. *Int J Intell Technol* 2006;1(1):104-10.
5. Iwata T, Saito K, Yamada T. Recommendation method for extending subscription periods. In *KDD'06: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM; 2006. p. 574-9.
6. Mitchell T. *Machine Learning*. : WCB/McGraw Hill; 1997.
7. Quinlan JR. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1993.
8. Kaur M, Singh K, Sharma N. Data mining as a tool to predict the churn behaviour among Indian bank customers. *Int J Recent Innov Trends Comput Commun* 2013;1(9):720-5.
9. Miguéis VL, van den Poel D, Camanho AS, Falcão e Cunha J. Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Syst Appl* 2012;39(12):11250-6.
10. Balle B, Casas B, Catarineu A, Gavalda R, Manzano-Macho D. the architecture of a churn prediction system based on stream mining. In: *Artificial Intelligence Research and Development. Proceedings of the 16th International Conference of the Catalan Association for Artificial Intelligence*. Vol. 256. 2013. p. 157.
11. Berson A, Smith S, Thearling K. *Building Data Mining Applications for CRM*. New York: McGraw-Hill; 1999.
12. Kotler P. *Marketing Management: Analysis, Planning, Implementation, and Control*. New Jersey: Prentice-Hal; 1994.
13. Reichheld F, Sasser WE. Zero defections: Quality comes to services. *Harv Bus Rev* 1990;68(5):105-11.
14. Hadden J, Tiwari A, Roy R, Ruta D. Computer assisted customer churn management: State-of-the-art and future trends. *Comput Oper Res* 2007;34(10):2902-17.
15. Kasiran Z, Ibrahim Z, Ribuan MS. Mobile phone customers churn prediction using Elman and Jordan recurrent neural network. In: *Computing and Convergence Technology (ICCT)*. 7th International Conference on; 2012. p. 673-8.
16. Jin S, Meng Y, Fan C, Peng F, Chen Q. The research on applying data mining to telecom churn management. *2nd International Conference Materials Mechatronics Automation Lecture Notes Information Technology*. Vol. 15. 2012. p. 39-45.