International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

CUSTOMER STRESS PREDICTION IN TELECOM INDUSTRY

Mohammed Aaqhil.R , Mohanraj.K , Monish.B , Yogeshwaran.P
Dhaanish Ahmed College of Engineering

Abstract: Retaining the most valuable customers is a major problem companies face in this information age. Especially, the field of telecommunication faces complex challenges due to a number of vibrant competitive service providers. Therefore, it has become very difficult for them to retain existing customers. Since the cost of acquiring new customers is much higher than the cost of retaining the existing customers, it is the time for the telecom industries to take necessary steps to retain the customers to stabilize their market value. CRM uses data mining (one of the elements of CRM) techniques to interact with customers. This study investigates the use of a technique, supervised learning, for the management and analysis of customer-related data warehouse and information.

Data mining technologies extract hidden information and knowledge from large data stored in databases or data warehouses, which supports the corporate in decision making process. Several data mining techniques have been proposed in the literature for predicting the happy and stressed customer using heterogeneous customer records. Probably, the stressed customers are in the urge of moving out to competitive service providers. This project analysis the telecom customer data available in open dataset and predict the customer stress by applying supervised machine-learning algorithms mainly using Deep Neural Network , K Nearest Neighbour , Support Vector Machine and Random Forest **.**

INTRODUCTION

Enterprises in the competitive market mainly rely on the profits which come from customers. Therefore, customer relationship management (CRM) always concentrates on confirmed customers that are the most fertile source of data for decision making. This data reflects customers' actual individual behavior. This kind of behavioural data can be used to evaluate customers potential value assess the risk that they will stop paying their bills, and anticipate their future needs. Besides, because customer churning will likely to result in the loss of businesses, churn prediction has received increasing attention in the marketing and management literature over the past time. In addition, it shows that a small change in the retention rate can result in significant impact on businesses. In order to effectively manage customer churn for companies, it is important to build a more effective and accurate customer churn prediction model. In literature, statistical and data mining techniques have been used to create the prediction models. The data mining task can be used to describe (i.e. discover interesting patterns or relationships in the data), and predict (i.e.

predictor classify the behavior of the model based on available data).In other words, it is an interdisciplinary field with a general goal of predicting outcomes and employing sophisticated algorithms to discover mainly hidden patterns ,associations, anomalies, and/or structure from extensive data stored in data warehouses or other information repositories and filter necessary information from large datasets.

OBJECTIVE

Our Aim Customer Stress-ML should meet the basic requirements, i.e., predict the churning of the customer, find the accuracy of the above algorithms and find the efficient one. Efficient Churn model predicts the Customer churning and helps in relieving thecustomer from the stress and help them in retaining in the respective telecom service.

EXISTING SYSTEM

Stressed customer, tend to cease doing business with a company in a given time period which has become a significant problem for many firms. These include publishing industry, investment services, insurance, electric utilities, health care providers, credit card providers, banking, Internet service providers, telephone service providers, online services, and cable services operators.

There are numerous predictive modeling techniques for predicting customer behavior and their satisfactory level. These vary in terms of statistical technique (e.g., neural nets versus logistic regression), variable selection method (e.g., theory versus stepwise selection), and number of variables included in the model.

PROPOSED SYSTEM

In this paper, we review the existing works on customer stress prediction in three different perspectives: datasets, methods, and metrics. Firstly, we present the details about the availability of public datasets and what kinds of customer details are available for predicting customer stress, which leads to churn. Secondly, we compare and contrast the various predictive modeling methods that have been used in the literature for predicting the churners using different categories of customer records, and then quantitatively compare their performances. Finally, we summarize what kinds of performance metrics have been used to evaluate the existing churn prediction methods. Analyzing all these three perspectives is very crucial for developing a more efficient churn prediction system for telecom industries.

In a business environment, the term, customer churn simply refers to the customers leaving one business service to another, which is the process of customers switching from one service provider to another anonymously. This is because of customer stress and un-satisfaction with the business. From a machine learning perspective, customer stress (probable churn) prediction is a supervised (i.e. labeled) problem defined as follows: Given a predefined forecast horizon, the goal is to predict the future churners over that horizon, given the data associated with each subscriber in the network.

ALGORITHMS

The machine learning algorithms that has to be applied to the training data to build the models are

(1) Deep Neural Network

Logistic regression is suitable for predicting a binary dependent variable, such as positive/negative; deceased/alive; or in this study, admit/not admit. The technique uses a logistic link function to enable the calculation of the odds of an outcome occurring. The second algorithm that was used was a random forest, specifically recursive partitioning from the RPART package.

The RPART package is an implementation based on the model presented by Breimanand colleagues . This algorithm splits the data at each node based on the variable that best separates the data until either an optimal model is identified or a minimum number of observations exists in the final (terminal) nodes .The resulting tree can then be pruned to prevent overfitting and to obtain the most accurate model for prediction. The third algorithm was a SVM which creates two vectors and tries to classify the data based on it . The last algorithm is KNN where the N defines the number of nearest neighbours and the value changes as and when the n changes.

Thus all the algorithms can increase the prediction and accuracy of the Churn prediction in Telecom Industry

SYSTEM REQUIREMENTS

The system requirement is a main part in the analyzing phase of the project. The analyzer of the project has to properly analyze the hardware and the software requirements, otherwise in future the project designer will face more trouble with the hardware and software required. Below specified are the project hardware and software requirements.

HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by the software engineers as the starting point for the system design. It shows what the system does and not how it should be implemented.

- SYSTEM:PENTIUM DUAL CORE
- HARD DISK :120 GB
- MONITOR : 15’’LCD
- INPUTDEVICE:KEYBOARD,MOUSE
- RAM :8 GB

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

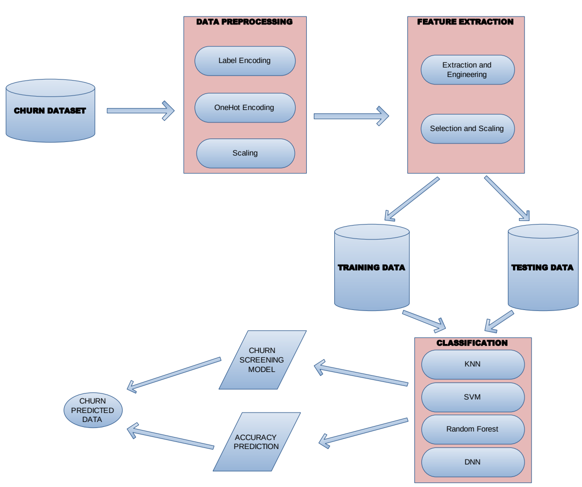
SOFTWARE REQUIREMENTS

The software requirements document is the software specification of the system. It should include both a definition and a specification of a requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the team's progress throughout the development activity.

- OPERATING SYSTEM :WINDOWS 10
- PROGRAMMINGLANGUAGE : PYTHON
- TOOLS : ANACONDA, SPYDER, PYTHON

DATABASE : KERAS LIBRARY

ARCHITECTURE DIAGRAM



MODULES

Data Collection

Data preprocessing

Data visualization and descriptive

statistics

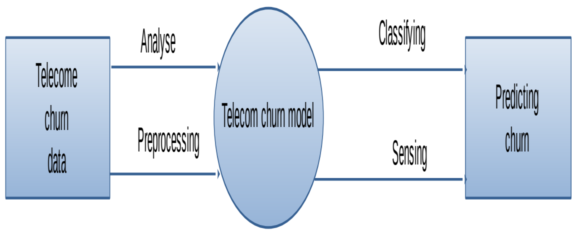
Data splitting

Model tuning using the training set

Predicting Stress

DATA COLLECTION

We have collected the Social opinion dataset of customer telecom dataProcessing .We are using this data info to predict stress for the current yearandmake appropriate preparations.Importing the Pandas gives massive functionality to work on data .Along with Pandas importing all the required libraries are even imported DATA COLLECTION DATA PRE PROCESSINGThe data has to Cleaned before loading into the Neural Classifier .



Label Encoding

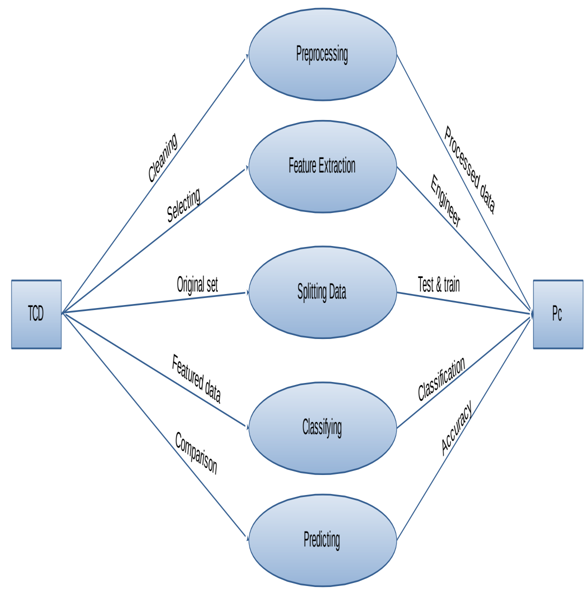
Before running a model, we need to make this data ready for the model.To convert any kind of categorical textdata into model-understandable numerical data, we use the Label Encoder class, all we have to do is to labelencode the first column. We import the LabelEncoder class from the sklearn library, fit and transform the firstcolumn of the data, and then replace the existing text data with the new encoded data.

One hot encoding

One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction. This is why we use one hot encoder to perform "binarization" of the category and include it as a feature to train the model.

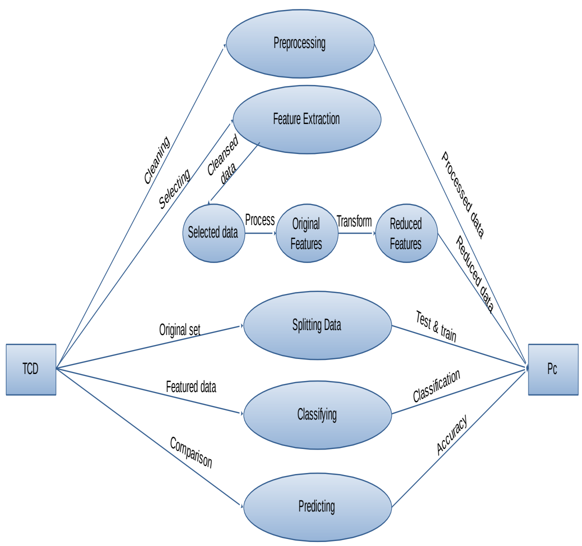
DATA PREPROCESSING

The extracted features are usually plotted against the output to check its relation to the output .



DATA SPLITTING

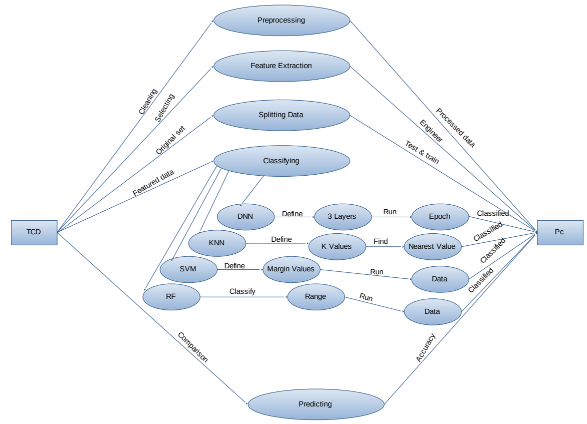
In statistics and machine learning we usually split our data into two subsets:training data and testing data (and sometimes to three: train, validate and test),and fit our model on the train data, in order to make predictions on the test data.When we do that, one of two thing might happen: we overfit our model or we unde rfit our model. We don't want any of these things to happen, because they affect the predictability of our model we might be using a model that has loweraccuracy and/or is ungeneralized (meaning you can't generalize your predictions onother data).



DATA SPLITTING

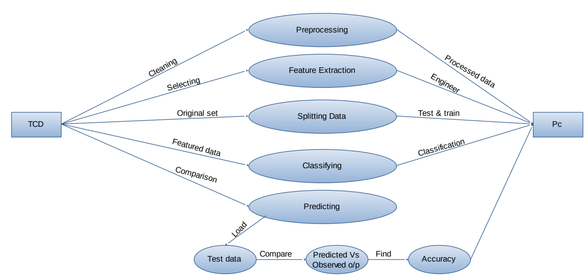
MODEL TUNING USING THE TRAINING SET

A deep neural network is a neural network with a certain level of complexity, a neural network with more than two layers. Deep neural networks use sophisticated mathematical modeling to process data in complex ways. A neural network, in general, is a technology built to simulate the activity of the human brain – specifically, pattern recognition and the passage of input through various layers of simulated neural connections.Many experts define deep neural networks as networks that have an input layer, an output layer and at least one hidden layer in between.

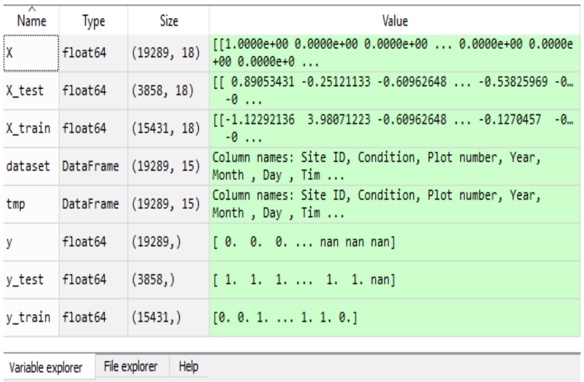


MODEL TUNING USING THE TRAINING SET PREDICTING STRESS

Predicting the test set result. The prediction result will give you probability of the People voting opinion for the given period .We can also allocate the resources assigned to them . We will convert that probability into binary 0 and predicting stress



SCREENSHOTS



Name	Type	Size	Value
X	float64	(19289, 18)	[[1.0000e+00 0.0000e+00 0.0000e+00 ... 0.0000e+00 0.0000e+00 0.0000e+0 ...
X_test	float64	(3858, 18)	[[0.89053431 -0.25121133 -0.60962648 ... -0.53825969 -0… -0 ...
X_train	float64	(15431, 18)	[[-1.12292136 3.98071223 -0.60962648 ... -0.1270457 -0… -0 ...
dataset	DataFrame	(19289, 15)	Column names: Site ID, Condition, Plot number, Year, Month , Day , Tim ...
tmp	DataFrame	(19289, 15)	Column names: Site ID, Condition, Plot number, Year, Month , Day , Tim ...
y	float64	(19289,)	[0. 0. 0. ... nan nan nan]
y_test	float64	(3858,)	[1. 1. 1. ... 1. 1. nan]
y_train	float64	(15431,)	[0. 0. 1. ... 1. 1. 0.]

REFERENCES

[1] M. Rejc and M. Panto, "Short-term transmission-loss forecast for the Slovenian transmission power system based on a fuzzy-logic decision approach," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1511–1521, Mar. 2011.

[2] F. Sulla, M Koivisto, and J. Seppanen, "Statistical analysis and forecasting of damping in the nordic power system," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 306–315, Jan. 2015.

[3] Z. Li *et al.*, "Short-term wind power prediction based on extreme learning machine with error correction," *Protection Control Modern Power Syst.*, vol. 1, no. 1, pp. 1–8, 2016.

[4] M. Kezunovic, L. Xie, and S. Grijalva, "The role of big data in improving power system operation and protection," in *Proc. IEEEIREP Symp.Rethymnon Bulk Power Syst. Dyn. Control-IX Optim. Security ControlEmerging Power Grid*, 2013, pp. 1–9.

[5] A. Azadeh, M. Saberi, S. F. Ghaderi, and V. Ebrahimipour, "Improved estimation of electricity demand function by integration of fuzzy system and data mining approach," *Energy Convers.Manage.*, vol. 49, no. 8, pp. 2165–2177, 2008.

[6] J. Liu,W. Fang, X. Zhang, and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," *IEEE Trans.Sustainable Energy*, vol. 6, no. 2, pp. 434–442, Apr. 2015.

[7] W. C. Hong, "Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model," *Energy Convers.Manage.*, vol. 50, no. 1, pp. 105–117, 2009.

[8] W. Sun and Y. Liang, "Least-squares support vector machine based on improved imperialist competitive algorithm in a short-term load forecasting model," *J. Energy Eng.*, vol. 141, no. 4, pp. 04014037-1–04014037-8,2015.

[9] D. X. Niu, H. F. Shi, and D. D. Wu, "Short-term load forecasting using Bayesian neural networks learned by hybrid Monte Carlo algorithm,"*Appl. Soft Comput.*, vol. 12, no. 6, pp. 1822–1827, 2012.

[10] Y.Wang, Q. Xia, and C. Kang, "Secondary forecasting based on deviation analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 500–507, May 2011.

[11] K. B. Song *et al.*, "Hybrid load forecasting method with analysis of temperature sensitivities," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 869–876,