Developing Machine Learning and Deep Learning Models for Customer Churn Prediction in Telecommunication Industry*

Teoh Jay Shen†, Abdul Samad Bin Shibghatullah
Institute of Computer Science & Digital Innovation, UCSI University
Cheras, Kuala Lumpur, Malaysia‡

E-mail: 1001539989@ucsiuniversity.edu.my, abdulsamad@ucsiuniversity.edu.my

Abstract

Customer churn is always a significant problem and one of the biggest concerns of telecommunication companies. The companies are attempting to create and design an approach to predict customer churn. This is why determining factors that causes the customer to churn is significant. The proposed models constructed in this work apply both the machine learning and deep learning algorithms. Those models was constructed and run under the Python environment and it used an open sources dataset that are available to everyone on www.kaggle.com. This dataset contained 7043 rows of customer's data with 21 features, and it was applied in the training and testing process of the models development. These models used four different types of machine learning and deep learning algorithms, which are the Artificial Neural Network, Self-Organizing Map, Decision Tree and a hybrid model with the combination of the Self-Organizing Map and Artificial Neural network algorithms.

Keywords: Machine Learning; Deep Learning; Churn Prediction; Telco Industry.

1. Introduction

In Telco industries or other business related field, customer churn is defined as the action of existing customers terminating their subscription of service with the company due to several reasons such as dissatisfaction of the service provided/better price offered by competitors for the same services [1] . Customers churn prediction and management is specifically intense in the mobile telecommunication industry as the market is increasingly saturated by the customers who change their subscription from one network provider to another in a very frequent manner. Furthermore, one of the main reason that customers churn are recognized as the biggest problem in any industry is due to the cost of obtaining new customers is whole lot more than keeping the existing customers [2]. Based on a research conducted by Abbas Keramati and Seyed M.S. Ardabili, it stated that the annual churn rate within telecommunication can be range from 20% to 40%, whereas the cost needed to obtain a single new customer is 5 to 10 times higher than maintaining an existing customer [3]. Therefore, to develop a churn prediction model with high accuracy is very important for a Telco service provider to be successful or bottom-line survival of a Telco service provider in this intensely saturated marketplace. On those grounds, the Telco market is ever changing and getting more and more competitive [4]. In fact, customer churn is not only the essential concern on the marketing side, but also one of the fundamental dimensions of CRM which is also known as customer relationship management [6].Churn prediction model is the model that utilized data mining techniques to transform an enormous amount of data into a meaningful insight and present it in a way that normal people could also understand.

Machine Learning carries a huge potential in the field of data mining and data analytics [6]. Machine Learning can be defined as the ability of a computer to study and discover a set of rules from the given input data or overall build a model that can be used to find correlations between the variables or a structured data set to make predictions [7]. Numerous of machine learning predictive modelling algorithms have been proposed and applied in building churn prediction models, these machine learning algorithms can be ranged from simple linear regression to more complex hybrid methodologies. These algorithms are able to work efficiently in predicting the customers who are most likely to churn. At the past few years, besides machine learning, deep learning has also become one of the most popular trends thanks to its potential in processing big data [8]. Deep Learning, also known as deep structured learning is basically a division of Machine learning that relying on algorithms that trying to model high degree abstractions on data [9). Deep Learning Algorithms (DLAs) establish a multi layered hierarchical structure of learning and indicating data. DLAs are useful in dealing with a massive amount of unsupervised data as it study and discover the correlations of data in a greedy layer structured method [9]. Thus, this project is going to develop two deep learning churn prediction models, one hybrid churn prediction model (SOM+ANN) and one machine learning churn prediction model using artificial neural network (ANN), self-organizing map (SOM), and decision tree (DT) algorithms. Lastly, CRSIP-DM methodology was applied for the development of this research.

2. Previous Work

Customer churn is one of the significant matters in the Telco industry nowadays. Numerous methods were implemented to predict customer churn in Telco companies. Majority of these methods employed machine learning, deep learning and data mining techniques. Most of the previous work done by researchers focused on implementing only one technique of data mining to retrieve knowledge, where others focused on conducting studies between several ML/DL algorithms to build a churn model. The aims and purposes of the research of such studies differ. Some of the aims are to simply evaluate if machine learning is an applicable approach for customer churn. In others, the aims may be to evaluate and select the best algorithms for prediction.

A comparative study for different type of machine learning algorithms for predicting churn customer for prepaid services was done by Brandusoiu and Toderean in the year 2016 [10]. The machine learning algorithms used in this study includes the Neural Network, Support Vector Machine and Bayes Network. The dataset used in this study consist of 3333 rows of customer call details with 21 independent variables and 1 dependent variable which is the churn status with the values of only YES or NO. Some features also captured the data of incoming and outgoing calls and messages for each customer. In order to reduce the data dimension, the principal component analysis algorithms were used by the author. AUC which is known as the area under the curve is the technique used by the author to evaluate the performance of the ML algorithms. The AUC values were reported 99.10%, 99.70% and 99.55% respectively for Bayes Network, support vector machine and neural network [10]. A similar study was made by Yue He, Zhenglin He and Dan Zhang in 2009 [11], where they designed a new model for churn prediction on the basis of the Neural Network algorithm. This model was proposed with the intention to solve the customer churn issues in a big Chinese Telco company which consists 5 million of customers. The results of the model accuracy rate reached 91.1%, which is very a stunning result [11].

In the year 2015, 9 researchers from China conducted a research of the customer churn problem in the big data platform [12].The goal of this study was to demonstrate that big data can significantly improve the performance of predicting customer churn based on the 3Vs (variety, volume, velocity) of the data. Interpreting the data at China's largest Telco Company, a big data platform is needed to engineer the process and solve the problems.

AUC (area under the curve) was once again used to measure the model performance and Random Forest Algorithm was applied to build the model in this study (12). Other than that, at the same year 2015, a group of researchers from the United States with Maryam M Najafabadi [13] as the leader demonstrated how to use deep learning to interpret big data, extracting and discovering complex correlation hidden under a large dataset. Moreover, they classified and list out certain types of problems that can be solve by applying deep learning. These problems are problems related to prediction and classification, sematic indexing, extracting information from raw dataset and data tagging. They discussed several components of deep learning to address certain issues in big data analytics [13].

Furthermore, Niall McLaughlin and his team used the concept of a CNN, which is also known as convolutional neural network to develop an android malicious software detection system in the year 2017. It was a creative and modern application of deep learning in the domain of malware analysis. This application was proficient in learning to perform malware detection and feature extraction at the same time. Lastly the team came to a conclusion that showed the proposed model is more efficient than an n-gram based malicious software detection system. The last observation of study related to deep learning application in predicting churn was done by Federico Castanedo and his team in 2016. They introduced deep neural network as a perfect tool for customer churn prediction as it constructs numerous hidden layer and it disseminates the weights of each neurons from one layer to the next layer. Deep Learning supports automated process of extracting features with the maximum information (impact), therefore there was a remarkable melioration in the performance of the model in accuracy wise. Billions of call details from a Telco business enterprise were used as the data input to train the predictive model for customer churn prediction. 77.9% of AUC (are under the curve) values were achieved on the validation dataset. Moreover, they extended the idea into the field of fraud detection in various industries such as the banking industry, insurance industry and Telco industry.

3. Methodology and Algorithm

For this research, CRISP-DM Methodology is use to model the customer churn prediction in Telco industry. CRISP-DM contains of the following phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. In general, these phases are implemented subsequently. However in this main stream, various iterative processes can be seen due to the fact that every outcome of each phase impacts the next methodology phase. In other words, after the data understanding phase is completed, data analyst usually needs to go back to the business understanding phase to refine or reconsider the aims and objectives for the project. In the same manner, after the modelling part, mostly there is always a need for the data analyst to conduct a new data pre-processing process in order to enhance the model's performance. Moreover, the results of the evaluation phase could also potentially leads to a new start of the CRISP-DM process in the case that the models do not satisfy the aims and objectives defined previously in the research.

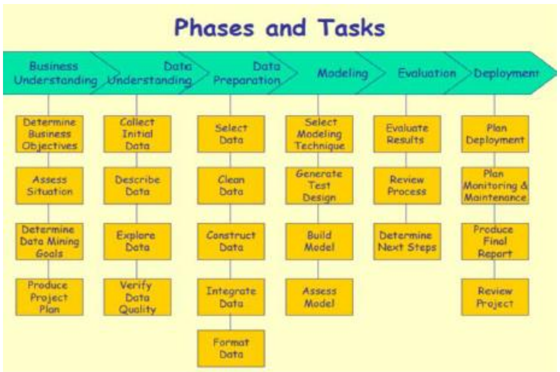

Fig.1: Phases and tasks of the CRISP-DM process

Fig. 1 shows the complete cycle of the CRISP-DM methodology and the phases within the cycle. The details are explained in the following section:

1. Business Understanding: This is the initial phase of CRISP-DM methodology. Three statements are produced in this phase which the statements include the statement of business objective, statement of data mining objective and statement of success criteria. This phase emphasizes on understanding the objectives and requirements of the project from a business point of view, then transforming this information into a problem

statement of the project and the draft of strategy to accomplish the objectives.
2. Data Understanding: The data understanding phase begin with the initial data collection and then follows by the data reading process with the purpose of getting familiarize with the data, identify the data quality, to determine the initial insights to the data and to discover relevance subset to create the hypothesis for hidden knowledge.
3. Data Preparation: This phase includes all the actions to develop the final dataset on the basis of the raw data. The tasks in data preparation phase are commonly been executed several times and does not follow any prescribed order. The tasks include row and feature selection, data filtering, data cleaning and transforming the dataset into the appropriate format for the modelling tools.
4. Modeling: The application of the algorithms and code developing parts are constructed under this phase. Model optimizing process is conducted in order to get the best performance form the model. Usually for one data mining problem, there are multiple algorithms and techniques available to be apply and solve the problems. However, some algorithms require specific requirements from the data. Thus, moving back to the data preparation part is often required.
5. Evaluation: In this phase, a thorough evaluation on the model and the steps of execution to construct the model is conducted in order to make sure that the business objectives are achieved by the model. The main activity of this phase is to assess and identify whether is there any critical business problems that has not been recognized adequately. In the end of this phase, a conclusion on the usage of the model results should be made.
6. Deployment. The deployment phase remains as the last phase in the CRISP-DM process. Where in this phase the data mining results are utilized as the business rule and the findings and knowledge obtained from the model needs to be visualized and presented in a way that business people can understand it.

Artificial Neural Network (ANN) is part of the model of this research. ANN was invented on the basis of a sophisticated biology research regarding neural system and human brain tissue. ANN is applied to stimulate the neural operations of knowledge processing in the human brain [14]. In ANN, the neurons which can also be called as the information processing nodes are formed in a topological structures. Therefore, the neurons disseminate their data and information in a parallel fashion. Multiple nonlinear transfer functions are combined to maps the inputs and measured output responses [15].

Self-Organizing Map (SOM) is classified as a methodology under unsupervised learning, it distribute a group of patterns into clusters or segments. Cluster analysis can be defined as the process of arranging a group of data object into different cluster. Above all, no predefined clusters are allocated. In the year 1987, Teuvo Kalevi Kohonen introduced and illustrated a brand new structure of a neural network architecture named the Self-Organizing Map (SOM). SOM turned out to become incredibly useful in the case that the input dataset are complicated and high dimensions. SOM is applied to study the correlations in a dataset and distribute the data into different cluster based on the similarity of patterns of the data where the modellers are not able to forecast the class of the classification[16]. SOM is classified as a methodology under unsupervised learning, it distribute Decision trees algorithm possess the basis of a graph structure, every decision could potentially generates a new node and eventually developed into a tree-like graph [18].

As a rule, a hybrid model is a combination of two or more machine learning or deep learning algorithms. As an example, the clustering (SOM) and classification (ANN) techniques can be combined in sequence. In other words, clustering techniques are able to be applied as a pre-processing phase to determine different pattern of clusters for later supervised learning [19]. Therefore, the result of the clustering process can either be used to determining the main clusters of a set of data given or be used as a pre-classification of unnamed features. Thus, the result of the clustering process can be included as a feature into the training set to train a prediction model. After the completion of the hybrid model, it has the capability to group or forecast new instances. In other terms, the first stage of the hybrid model is to detect the outlier of the dataset and the second stage of the hybrid model is to make prediction. Since the customer churn prediction is a supervised classification process, thus the hybrid models that are chosen in this paper are the cluster (SOM) to classifier (ANN) model.

3.1. *Data*

Only one dataset was used for this research. In order to obtain a good dataset, an online research are done by the researcher in order to find a free open source dataset that fulfilled the requirements of this research, finally a dataset that fulfils the requirements was found on www.kaggle.com. A total of 7043 rows of customer details with 21 features were provided in the dataset. The variables of the dataset used were stimulated from a Telco company from USA. Variables of the dataset are categorized into two types. Nominal types define a set of values which represents a certain meaning, while quantitative types refer to values that can be calculated and ordered, such as integer and float values. Variable "Churn" is the dependant variable, meaning it is the variable that the model is trying to predict (see Table 1, row 21).

Table 1: Description of the Data

Index	Column	Description	Type
1	Customer ID	Customer ID	Nominal
2	Gender	Gender of the customer (Male/Female)	Nominal
3	SeniorCitizen	Determine the customer is a senior or not (1/0)	Nominal
4	Partner	Determine the customer has a partner or not (Yes/No)	Nominal
5	Dependents	Determine the customer has a dependents or not (Yes/No)	Nominal
6	Tenure	Amount of months that the customer used the company service.	Quantitative
7	PhoneService	Determine the customer has phone service or not (Yes/No)	Nominal
8	MultipleLines	Determine the customer has multiple line or not (Yes/No)	Nominal
9	InternetService	The ISP, Internet Service Provider of the customer (Fiber optic, DSL, No)	Nominal
10	OnlineSecurity	Determine the customer has online (Yes/No/No Internet Service)	Nominal
11	OnlineBackup	Determine the customer has online backup or not (Yes/No/No Internet Service)	Nominal
12	DeviceProtection	Determine the customer protection for the device or not (Yes/No/No Internet Service)	Nominal
13	TechSupport	Determine the customer has technical support or not (Yes/No/No Internet Service)	Nominal
14	StreamingTV	Determine the Customer has streaming TV or not (Yes/No)	Nominal
15	StreamingMovies	Determine the customer has streaming movies or not (Yes/No/No internet service)	Nominal
16	Contract	The contract term of the customer (Month-to-month, One year, Two year)	Nominal
17	PaperlessBilling	Determine the customer has paperless billing or not (Yes/No)	Nominal
18	PaymentMethod	The payment method of the customer (Electronic check, Mailed check, Bank transfer, Credit card)	Nominal
19	MonthlyCharges	The monthly amount charged to the customer	Quantitative
20	TotalCharges	The total amount charged to the customer	Quantitative
21	Churn	Determine the customer churned or not	Nominal

4. Results and Evaluation

The method of evaluation that will be used to evaluate the algorithm's performance in this research is the calculation of the F-measure value. The accuracy, precision, recall and the F-measures value for all of the four proposed models are calculated accurately and populated in Table 2 with its corresponding algorithm.

Table 2 Models performance comparison table.

Algorithm	Accuracy	Precision	Recall	F-measure
Artificial Neural Network (ANN)	78.90	84.03	88.24	86.08
Self-Organizing Map (SOM)	51.89	69.35	70.67	70.00
Decision Trees (DT)	73.41	85.59	80.73	83.08
Hybrid Model (SOM + ANN)	79.53	83.90	89.40	86.56

In Table 2 the performance of each model can be observed clearly. Based on the F-measure score, the hybrid model possesses an 86.56% of F-measure score which is also the highest among the proposed models, and then followed by the Artificial Neural Network model which holds almost the same F-measure score (86.08%) with the hybrid model. In addition, the Decision Tree model holds a 83.08% of F-measure score and the Self-Organizing Map model possesses the lowest F-measure score (70.00%) among the proposed models. As the Self-Organizing Map model applied the unsupervised learning method, thus it is reasonable for the SOM model to have a comparatively weaker performance compared other models that applied supervised learning method in this case. Lastly, by evaluating the results of the models we noticed that the overall performance of the model increased after implementing the Hybrid model approach.

5. Conclusion and Future Work

CRISP-DM methodology was implemented for conducting a prediction of churn customer in the Telco industry. The aim of this research is to compare the algorithms and to determine which is the most suitable and accurate to predict the results of the given problem. A uniquely modified dataset was analysed with different machine learning and deep learning models and its result were compared with an evaluation method. The algorithms used were the Artificial Neural Network

(ANN), Self-Organizing Map, Decision Tree (DT) and a hybrid algorithm which is the combination of SOM and ANN algorithm. After studying and comparing the results of each algorithms, the hybrid model with two deep learning algorithms combined together are concluded as the model that achieved the best outcomes for this research.

Moreover, the likelihood of successfully predicting the customer churn with Machine Learning and Deep Learning depends on the correct choice of data and algorithm. In order to achieve the best outcomes, choosing a suitable machine learning or deep learning algorithm for the problem or topic is crucial. However, by having the algorithm alone is unable to provide the best prediction results. Therefore, having the right variables and types of dataset is also an important factor in getting the best prediction results. Also, by looking at the scores of the feature importance is able to determine how the changes of each independent variable impact on another dependent variable. In the optimization process of the dataset, only highly statistically significance variables/predictors will be selected and grouped in an optimal dataset for independent variable. In fact, the optimal dataset does help the models to come up with better predictions. In addition, based on our literature review, implementing machine learning and deep learning model to predict customer churn is new and prevailing in the Global Telco industry. Feature engineering and data gathering process for predicting churn in the Telco industry in one of the popular and important elements of the current research Thus, it is expected that with an in-depth understanding of characteristic and behaviour of the churn customers, Telco companies can develop new strategies to address churn. These strategies need to be applied on the targeted customers who shared the similar behaviours with the churn customers group. Offering a new plan, provides better quality of service, determine the needs from different segments of customer, and provides tailored offers for different customer groups can be all included in the strategies.

There are some constraints in this research that needs to be recognized and improvement needs to be done against them. Firstly, because of the privacy issues of the Telco companies, real time dataset are not allowed to be used in this research. This research focused on study and compares the performance of different models, thus the future work will focus on the feature selection and the research of the customer needs by applying different data mining techniques. Moreover, future research can be improved by implement others machine learning and deep learning methods such as: K-Nearest Neighbours, Logistic regression, Learning Vector Quantization, Support Vector Machines and Random Forest. They can be used to have a better understanding of each machine learning algorithm results provides a statistical results to aids the algorithm selection process.

Acknowledgements

The authors wish to thank ICSDI, and CERVIE UCSI University for supporting this project.

References

1. Mishra, A., & Reddy, U. S. (2017). A Novel Approach for Churn Prediction Using Deep Learning. 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC).
2. Wei, C., & Chiu, I. (2002). Turning telecommunications call details to churn prediction: A data mining approach. Expert Systems with Applications, 23(2), 103-112.
3. Keramati, Abbas & M.S. Ardabili, Seyed. (2011). Churn analysis for an Iranian mobile operator. Telecommunications Policy. 35. 344-356. 10.1016/j.telpol.2011.02.009.
4. Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. IEEE Transactions on Neural Networks, 11(3), 690-696.
5. Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification Expert Systems with Applications, 36(2), 2592-2602.
6. Rathor, A., & Gyanchandani, M. (2017). A review at Machine learning algorithms targeting big data challenges. 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT).
7. Kotsiantis. (2007). Supervised Machine Learning: A Review of classification Techniques, Informatica. Informatica,31, 249-268.
8. Sejnowski, T. J. (2018). The deep learning revolution. Cambridge, MA: The MIT Press
9. Hordri, Nur & Yuhaniz, Siti & Shamsuddin, Siti Mariyam. (2016). Deep Learning and Its Applications: A Review.

10. Brandusoiu, Ionut & Toderean, G & Beleiu, Horia. (2016). Methods for Churn Prediction in the Pre-paid Mobile Telecommunications Industry. 97-100. 10.1109/ICComm.2016.7528311.
11. He, Y., He, Z., & Zhang, D. (2009). A Study on Prediction of Customer Churn in Fixed Communication Network Based on Data Mining. 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery.
12. Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B.,Zeng, J. (2015). Telco Churn Prediction with Big Data. Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data - SIGMOD 15.
13. Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1).
14. Markopoulos, A. P., Manolakos, D. E., & Vaxevanidis, N. M. (2008). Artificial neural network models for the prediction of surface roughness in electrical discharge machining. Journal of Intelligent Manufacturing, 19(3), 283-292.
15. Sharma, A., & Panigrahi, P. K. (2011). A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. International Journal of Computer Applications, 27(11), 26-31.
16. Kohonen, T. (1987). Adaptive, associative, and self-organizing functions in neural computing. Applied Optics, 26(23), 4910.
17. Tsai, C., & Lu, Y. (2009). Customer churn prediction by hybrid neural networks. Expert Systems with Applications,36(10), 12547-12553.
18. Quinlan, J. (1987). Simplifying decision trees. International Journal of Man-Machine Studies, 27(3), 221-234. official website, https://www.robocup.org/

Authors Introduction

Mr Teoh Jay Shen

He received his Diploma of Information Technology from UCSI University in 2018 and continue his Bachelor of Science in Computing at the same university and completed in 2020. His research interest are in machine learning and artificial intelligent.

Dr. Abdul Samad Bin Shibghatullah

He received the bachelor accounting degree from Universiti Kebangsaan Malaysia, Bangi, Malaysia in 1999, the M.Sc. degree in computer science from the Universiti Teknologi Malaysia, Skudai, Malaysia in 2002, and the Ph.D. degree in computer science from the Brunel University of Uxbridge, United Kingdom. He is currently Associate Professor at Institute of Computer Science & Digital Innovation, UCSI University, Kuala Lumpur, Malaysia. His current research interests include optimization, modelling and scheduling.