Customer Churn Prediction using Predictive Analytics: Basis for the Formulation of Customer Retention Strategy in the Context of Web-based Collaboration Platform

Neil T. Awit
College of Business Administration
Lyceum of the Philippines University, Manila Philippines
neil.awit@lpu.edu.ph

Ramon M. Marticio
Higher Education Department
University of Santo Tomas- Angelicum College, Quezon City, Philippines
ramon.marticio@ustangelicum.edu.ph

Abstract

This study utilized Machine Learning (ML) models to analyze and predict customer churns in the context of web-based collaboration platform. Secondary data was used in the study, containing 109,740 observations and 12 predictors. This was divided into test dataset and out-of-time (OOT) dataset where the prior was used for model fitting and the latter is to test performance stability in unseen data. Furthermore, Synthetic Minority Over-sampling Technique (SMOTE) was performed to resolve the data imbalance, hence, preventing bias and distortion in the models' performance. These 3 ML models were assessed based on Accuracy, ROC-AUC, Precision, Recall and F1-Score. Given the business context's applicability, F1-score and Accuracy were used as bases for performance, leading to the selection of Decision Tree Classifier as the ML model in this study, with Accuracy of 92.1% and F1-Score of 63%. Furthermore, hyper-parameter tuning was performed on Decision Tree Classifier to prevent overfitting. To reinforce the model selected, Survival Analysis was implemented, specifically, Kaplan-Meier (KM) Estimator and Cox Proportional Hazard (CPH) were utilized to analyze the rate and timeframe of disengagement to the platform, revealing that beyond 72 months, it was projected to retain only 60% of its user base. Hence, these multidimensional results and insights derived from both Decision Tree Classifier and Survival Analysis were anchored in the formulation of customer retention strategy, proactively target customers who are predicted to churn.

Keywords
Customer Churn, Customer Retention Strategy, Decision Tree, Survival Analysis, Machine Learning (ML)

1. Introduction
The rapid growth of technology, as brought by the Fourth Industrial Revolution, advances the adoption of game-changing technologies such as Artificial Intelligence (AI), Internet of Things (IoT), Cloud Computing and Analytics. This also pushed the massive digital transformation across industries in the Philippines. As reported by the World Bank, 56% of the Micro, Small and Medium Enterprises (MSMEs) have adopted technology, although this was described as still "at the basic level" of adoption rate compared to other Southeast Asian nations. From the user adoption standpoint, the country's internet penetration rate at the start of 2022, has grown to 68% of the population, as reported by DataReportal. This only shows the growing traction that the digital transformation in the country has gained and is expected to sustain its momentum in the next few years.

While the massive digital transformation has stimulated the formation of new business models, especially at the height of Pandemic, it was proven that communications and collaborative platform were able to sustain various and crucial activities such as the implementation of distance education, and Telecommuting. This only proves the promising value of adopting and implementing such technology in helping to sustain various business activities.

In the Philippines, there's a wide array of collaboration platforms that users can choose from, and this is exactly one of the pain-point of organizations operating under the realm of collaboration platform - how these companies going to attract more users? Or at the least, how these companies going to maintain their existing user base? Retention of the customers has become a tall order for every business and various research papers are now looking at various methodologies to analyze customer churn. In this age of analytics, several modernistic approaches have become more available, especially when analyzing customer churn, stretching further the already extensive range of methodologies. In the examination of various literatures, the researchers saw a gap, paving way to the development of this study which revolved mainly on the particular application of Predictive Analytics. And with still few studies yet continuously flourishing under this domain, this inspired the researchers to analyze customer churn through the application of Predictive Analytics. Specifically, a study that aims to predict churns and reveal the key indicators of churn, pivotal in the formulation of customer retention strategy in the context of collaboration platform.

1.1 Objectives

This study focuses on the business communications services particularly, a web-based collaboration platform, where users can collaborate to fulfill various business/ organization-related tasks through forums, group chats, and direct messages. As this organization, which anonymized as part of ethical considerations, is continuously hounded by the customer churn, or those users who are attempting to disengaged in the use of messaging platform, brought by the present and growing competitive landscape in this space. Figure 1 shows the existing user base and the actual churn rate which motivates the formation of this study. As observed, customer base at the start of year is at 56,070, with a steady decline/ churn in number of users. At the end of the analysis, I can be seen that already 30% churned.

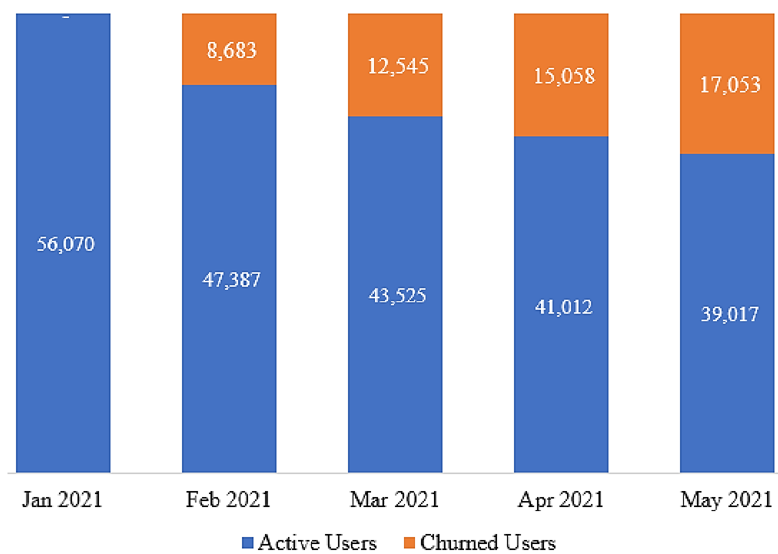

Figure 1. The User Base and the Actual Churn rate

This is deemed a costly scenario not only for the anonymous messaging company, but for other organizations who are most likely dealing with the same problem. To resolve the issue of customer churn, this paper aims to create a heuristic and contemporary approach through the application of Predictive Analytics particularly, Machine Learning (ML), in analyzing and predicting who among the active users will churn in the following month. Furthermore, this also aims to identify the predictors of churn by leveraging the anonymous organization's high volume of data. These predictions, through various ML models, will be anchored in the formulation of customer retention strategy.

2. Literature Review

Customer churn has been analyzed across industries and there are various methodologies that have been applied by various researchers. As postulated by Ahmad et al. (2019), customer churn has always been a dilemma and a major concern for every business. While various statistical models have been applied to analyze churns, this paper aims to examine the customer churns through the application of Predictive Analytics. First, the application of Predictive Analytics in the context of telecommunication sector have garnered major interest among researchers. Ahmad et al. (2019), build a study in the telecommunications context through the application of XGBoost algorithm which was used in their customer churns prediction. Also, Wu et al. (2021), made a study on the churn prediction in the similar context of telecommunication where Bayesian Logistic Regression was used to conduct the factor analysis while K-means clustering was used for churns segmentation. Another study by Ullah et al. (2019), implementing the Random Forest Classifier in the prediction of customer churns in the similar context of telecommunication which also leveraged on K-means clustering for customer profiling. Similar study and context have been conducted by Geetha et al. (2020), utilizing Random Forest Classifier and Support Vector Machine (SVM) in their customer churns prediction in the telecommunication context.

On the other hand, Wu et al. (2021) analyzed the customer segmentation using K-means clustering which integrates the adaptive Particle Swarm Optimization (PSO) algorithm, showing the effectiveness and practicability of both algorithms especially for customer segmentation. Another study conducted by Ebrah and Elnasir (2019), where three ML models have been utilized to analyze and predict churn which are the Naive Bayes, SVMs and Decision Trees. In the Financial services landscape, various studies utilized Predictive Analytics. First, Karvana et al. (2019), analyzed customer churn and prediction using several data mining model in the banking sector. He et al. (2020), also utilized several ML models to predict customer churns in the insurance industry. Also, Mauritsius et al. (2020), also applied the customer churn prediction model in the similar context of insurance.

Another study by Jain et al. (2021), utilized ML in the Churn prediction and retention in the context of Banking, Telecommunication, and Information Technology (IT). In online gaming space, Predictive Analytics has also been utilized. First, Lee et al. (2019), predicted churn through the application of Survival Analysis using commercial game data. Also, Kim et al. (2020), analyzed the retention, loyalty, and churn of customers through the application of ML models in the online gaming context. Meanwhile, a growing interest in the application of Predictive analytics in customer churns have also seen in the E-commerce space. Xiahou and Harada (2022), examined the E-commerce space through the application of K-means and SVM in the churn prediction and that the results have seen significance in the customer relationship management The application of Predictive Analytics in other areas have also been examined by this paper. First, Agrawal et al. (2018), utilized the behavioral patterns analysis using deep learning to analyze the customer churn, specifically the use of Deep Learning, which was used in gauging the factors to churn. Meanwhile, Saghir et al. (2019), implemented a Neural Network-based individual and ensemble models in the churn prediction. Previous literatures discussed the used of various predictive models such as K-means, SVM, Neural Network, Random Forest, Decision Tree and XGBoost.

One predictive model that draws the interest of the researchers especially in the analysis of customer churns is the Survival Analysis. Tracing back its roots, Survival Analysis was used to be implemented in Lifesciences, though the everchanging business landscape, allowed the use of Survival Analysis outside of Lifesciences. King and Rice (2019), analyzed the churn in Mobile Telecommunications domain where it aimed to predict the timing of customer churn. Lim (2020), also examined the customer churn through Survival Analysis where it found useful in supplier convergent triple-or-quad-play services. Furthermore, Masarifoglu and Buyuklu (2019), applied Survival Analysis to predict churn in Telecommunications sector which analyzed the lifespan of customers, hence, creating proactive measures. Synthesizing the methodologies applied as well as identifying key areas where gaps found, this study focused on the application of Predictive Analytics, specifically, assessing the performance of Logistic Regression, Decision Tree Classifier and Random Forest Classifier which also supplemented by Survival Analysis in the Prediction of customer churns in the context of web-based collaboration platform.

3. Methods

In this study, the problem was modelled using ML and will be implemented using Sci-kit Learn library in Python. The researchers tested various models and subsequently selected the most appropriate technique in this study. As seen in Figure 2, the Predictive Modelling Process will come into five main phases. First, on the data collection, the data has a total of 109,740 observations, containing 13 variables. The first 54,870 observations will be used for the Test dataset where several algorithms will be tested while the other 54,870 observations will be used for the Out-of-time (OOT) validation dataset where the algorithms will test again for the actual performance. Next, for the data pre-processing, Exploratory Data Analysis (EDA) performed for descriptive analysis as well as the Correlation matrix to spot Multicollinearity as well as the treatment of missing values and outliers. The model formulation phase is where algorithms will be selected and tested. In this study, Logistic Regression, Decision Tree, and Random Forest, were chosen as the top three models based on the following: 1) applicability of the algorithms with the problem, and 2) the observed knowledge, evidence, and methodological gap. Although this can only be gauged through performance assessment which is the subsequent step in the model formulation, and finally, model selection from the 3 selected algorithms, based on the prior ML model performance assessment.

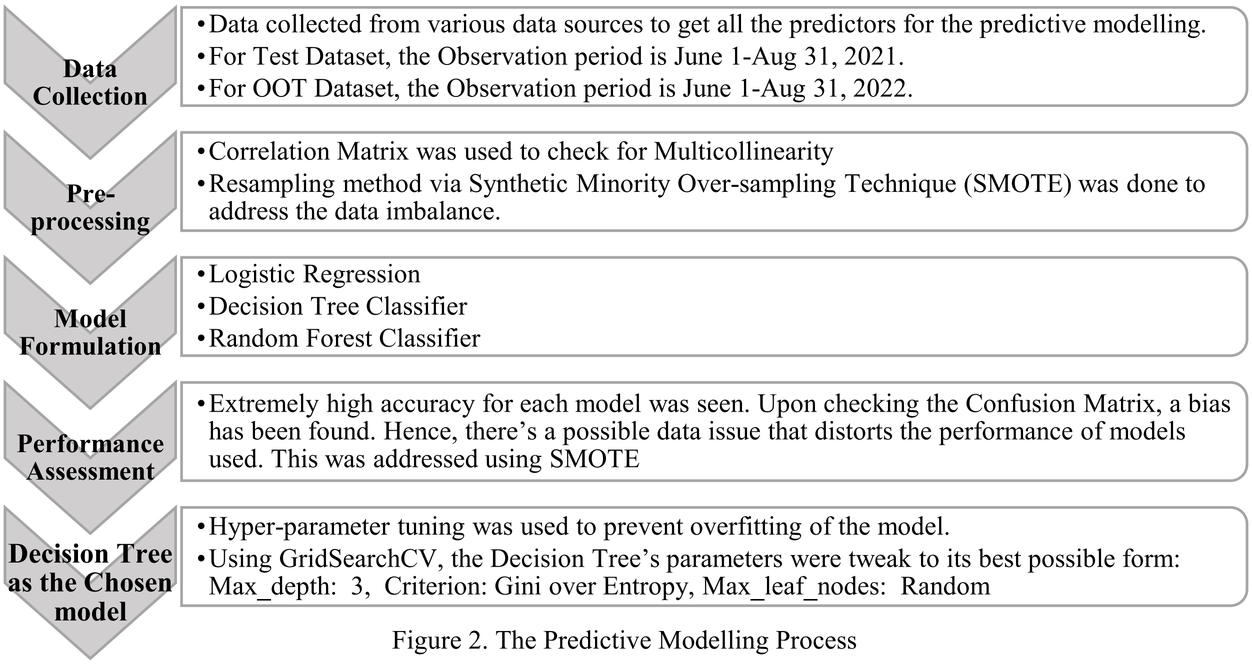

Figure 2. The Predictive Modelling Process

In the formulation of the Predictive Model, below shows the mathematical formula of the top 3 selected models in this paper. First, the Logistics Regression formula below:

$$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \qquad (1)$$

- B_0+b_1X are linear components
- b_0 is the intercept
- b_1 is the coefficient for X
- e is the error term

Both the Decision Tree Classifier and Random Forest Classifier will utilize the Information Criteria formula below:

$$Gini = 1 - \sum_{i=1}^{n} p^2(c_i)$$
$$Entropy = \sum_{i=1}^{n} -p(c_i) log_2(p(c_i)) \qquad (2)$$

For each decision tree, nodes importance is computed through the use of Gini Importance with the assumption of two child nodes:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \qquad (3)$$

- Ni_j represents the importance of the node $_j$
- w_j represents the weighted number of samples reaching node $_j$
- C_j represents the impurity value of node j
- $left_{(j)}$ represents the child node from left split on the node $_j$
- $right_{(j)}$ represents the child node from right split on the node $_j$

For the Survival Analysis, there are two statistical models that can be applied. First, Kaplan–Meier estimator will be leveraged through the following formula:

$$\widehat{S}(t) = \prod_{i:\ t_i \le t} \left(1 - \frac{d_i}{n_i}\right) \qquad (4)$$

- t_i represents the initial time when one event happened
- d_i represents the number of events at time t_i
- n_i represents the number of individuals survived until time t_i

Next, Cox Proportional Hazards (CPH) will be utilized through the following formula (Charan, 2021):

$$\begin{aligned}\lambda(t) &= \mathrm{P}(T=t|T\geq t) \\ &= \frac{\mathrm{P}(T=t)}{S(t)} \\ &\equiv \frac{f(t)}{S(t)}\end{aligned} \tag{5}$$

- S(t) represents the proportion of the respondents surviving at the indicated time t
- t_i represents the observed failure of the subjects $_i$
- T_i represents the failure time of subject $_i$

4. Data Collection

Secondary data source will be utilized in this study, which was provided by the anonymous organization as it was agreed by both researchers and data owner that the data should be treated within ethical bounds, hence, Personal Identifiable Information (PII) and as well the dataset, per se, should be treated with high degree of anonymity. The data has a total of 109,740 observations which was divided into 2 dataset- 1st for Test Dataset and the 2nd is for OOT dataset. In other words, each dataset, Test and OOT dataset, will have 54,870 observations each. This also contained a total of 13 variables, as shown in Table 1 below. The "Churned?" column serves as the label or outcome variable.

Table 1. Variables used in this Study

Variables	**Description**
Complaint Call	The number of complaints vial calls the user has done since their first use of the platform
Retention Period	The user's retention in months
Buyside Flag	Either a user is buyside or sell side users (1=Buyside, 0= Sell side)
Average Number of Active Days in a Month	Captured based on their activity within two months, this is the average number of days the user is active in a month
Total Messages Sent	The total number of messages that a user has sent for the given period
Total Chat Room Posts	The total number of posts made by the user to the chatroom features
Total Connection	The total number of unique connections by the user for the given period
Usage Decline?	1 if the message sent for the users declined in the following month, else 0
Active Days Decline?	1 if the number of active days for the users declined in the following month, else 0
Connections Decline?	1 if the number of connections for the users declined in the following month, else 0
Customer's Review	A variable showing an open-ended response of the customer's sentiment.
Customer's Review Category	Categorical data, showing the category of review
Churned?	The Status of the user whether they churned or not

5. Results and Discussion

As part of preprocessing, dataset has been examined. For the data collection, the dataset segmented into test dataset where several ML models will be tested. OOT dataset will be used for the implementation of the selected ML models. As part of Exploratory Data Analysis (EDA), Correlation Matrix was used, as seen in Figure 3, to validate the existence of multicollinearity.

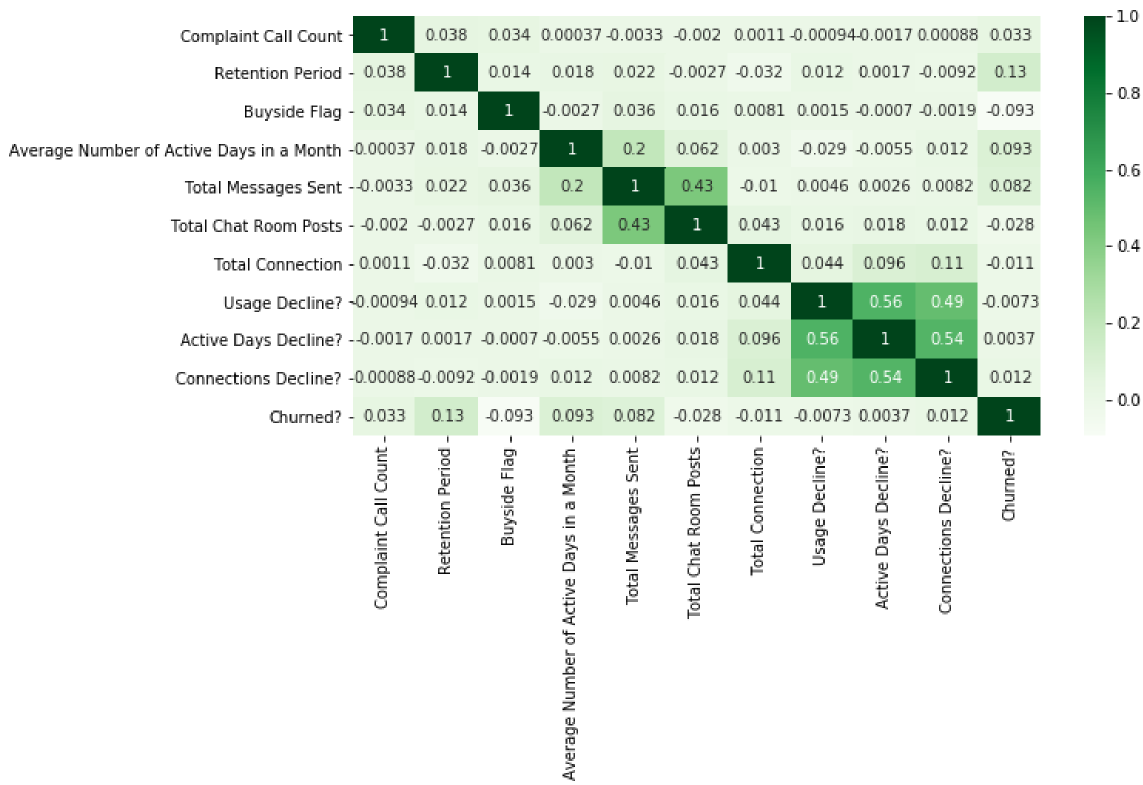

Figure 3. Correlation Matrix Heat Map

In the formulation of the model, the top 3 algorithms selected were Logistic Regression, Decision Tree Classifier, and Random Forest Classifier, based on the applicability on the problem that this paper aims to solve. When the performance of these algorithms assessed, the researchers saw high accuracy on the 3 models selected, hence, this was validated through Confusion Matrix and have seen biases on the data. This will potentially distort the performance of the models, hence, resampling method via SMOTE was done to address the data imbalance. Table 2 shows the heavy data imbalance seen in the current data which causes bias and unrealistic inflation of the accuracy of the tested models.

Table 2. Current Data Label vs. SMOTE Application

	Current Data Labels	SMOTE Applied in Data
Churned Users	6,069	35,239
Not Churned Users	48,801	35,239

Logistic Regression was first tested in this study and shows an accuracy of 67%. Next, Decision Tree Classifier was tested, showing an accuracy of 92.1%. For the Decision Tree, a hyper-parameter tuning was done to get the optimal parameter for the Decision Tree. In this regard, the Grid Search CV was used and gave the result: Criterion: Gini, Max_depth: 3. Take note that the researchers opted to set the Max_depth to 3 as it balanced the optimum between the Accuracy and Recall. Lastly, Random Forest Classifier was used with accuracy of 84%. For the Random Forest, the GridSearchCV was also performed and gave the result: Number of Estimators: 30 Decision Trees, Criterion: Entropy, Max_depth: 3. Take note that the researchers opted to set the Max_depth to 3 as it balanced the optimum between the Accuracy and Recall.

The business is keen in determining who among their users will churn. In the list of metrics provided, the business would want to balance the Precision and Recall. This means that while they wanted to get the predicted churned as much as possible, they also wanted most of it as correctly classified. This led to the selection of Decision Tree among the other models tested. Aside from getting highest accuracy, precision, and F1-Score, its ROC-AUC score is impressive enough that it can identify two classes (Churned or Non-churned). Figure 4 shows the Model performance

Comparison and Decision Tree algorithm is the most consistent when looking at the test and OOT dataset across the three metrics shown, hence, further justifies the model selection.

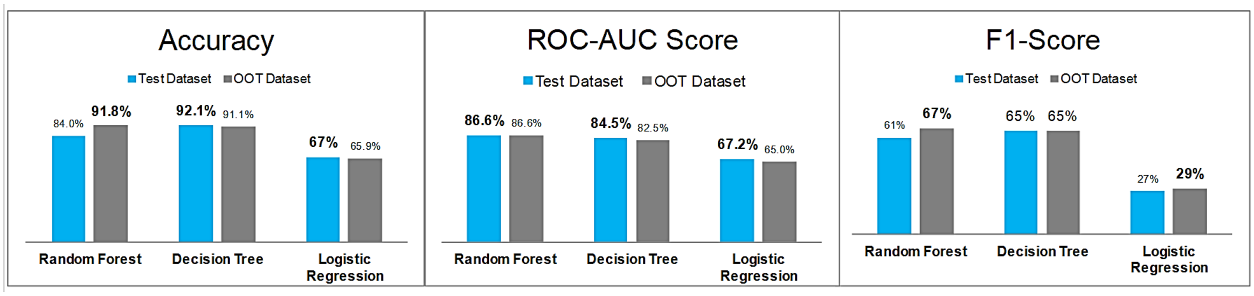

Figure 4. Model Performance Comparison

Figure 5 shows the Decision Tree output. As seen from the visualization, various probabilities of customer churn can be seen. Take note that this is the result from OOT dataset which depicts the true performance of the Decision Tree algorithm. The application, methodology and results aligned and agreed with the previous studies (Xiahou & Harada 2022; Wu et al. 2021; Jain et al. 2021; Geetha et al. 2020; He et al., 2020; Kim et al. 2020; Mauritsius et al. 2020; Ullah et al. 2019, Ebrah & Elnasir 2019).

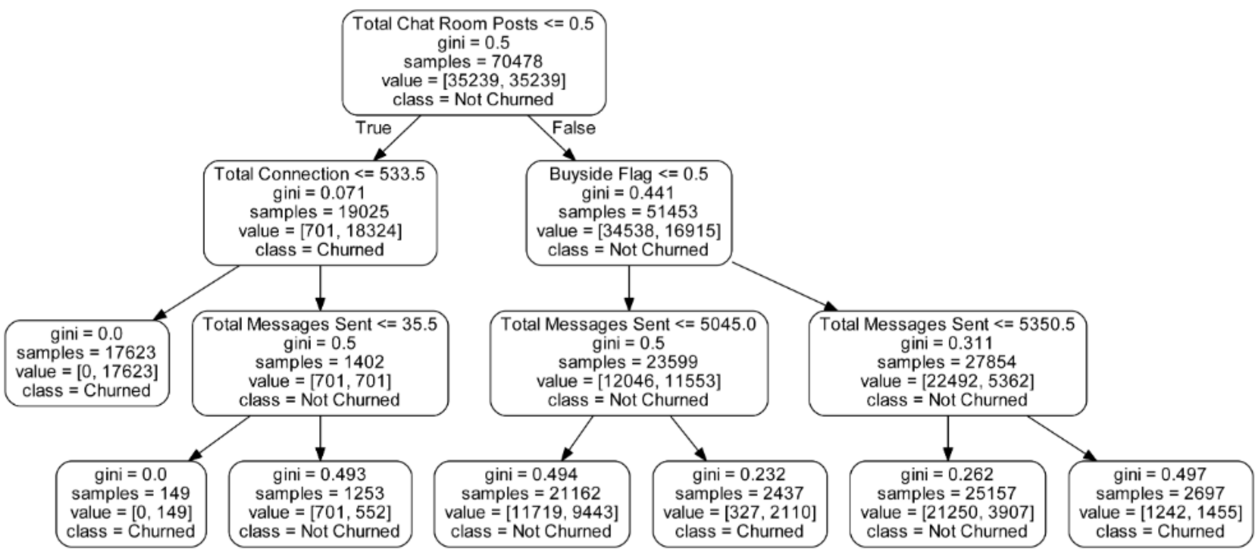

Figure 5. Decision Tree Output

As a supplementary, customer churn has been studied using Kaplan-Meier Estimator and Cox Proportional Hazard. While the business was able to determine the potential users who will disengaged to in the messaging platform, the company is also determined in the potential indicators that a user will churn. One basis for the selection of this type of analysis is the nature of the study, meaning churns were analyzed based on the cohort of users and observed their longevity on the specific period. In the implementation of Survival Analysis, the OOT dataset was used, although it was tested as well with the test dataset. Although this type of analysis doesn't entail the train-test-split process.

The survival rate shows the rate of disengagement of the users can be visualized through the Kaplan-Meier survival curve as shown in Figure 6. In this study, churn is relatively low based on the given duration which means that after 72 months, the business was expected to retain roughly around 60% of its users. Interestingly, the survival rate from the onset up to 50 months has seen marginal decrease while it took a sharp decline beyond that up to 72 months.

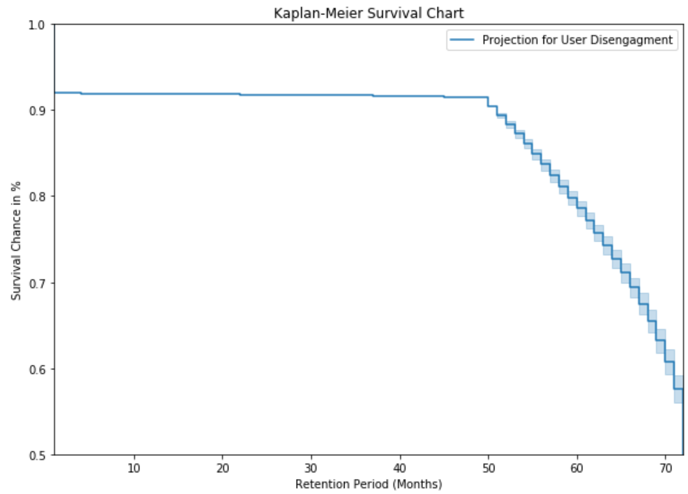

Figure 6. Kaplan-Meier Survival Chart- Retention Period

Taking a different perspective, it can be seen the survival rate of the users based on the number of complaints. Figure 7 also shows an interesting story, indicating that the survival chance of the users can be anchored on their satisfaction to our product. This means that once a user escalated a complaint/s in through service desk of the company, their survival chance will instantly drop to 70%. This will further decline as the number of their complaints increases. The business can leverage on this by analyzing and doing a series of smoke test to identify and subsequently improve several key functionalities in our platform, ensuring the positive user-experience.

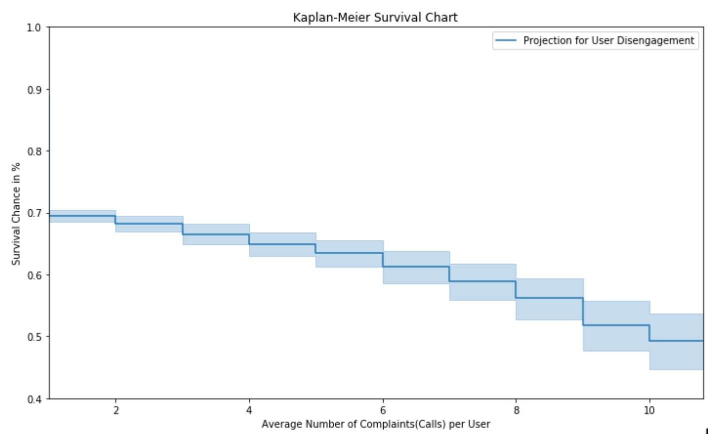

Figure 7. Kaplan-Meier Survival Chart- Average Number of Complaints

This study aims to examine the impact of the features in the dataset, this is possible through the Cox Proportional Hazards Model AKA Survival Regression model. Through Cox Proportional Hazards, this paper examined the cohort and were able to identify the indicators of churn. That means if the Connections decline in the following month, as shown in Figure 8, user will likely disengage. Same with the decline in Days active and the number of Complaints (Call) the company gets from the users. The application, methodology and results aligned and agreed with the previous studies (King & Rice 2019; Masarifoglu & Buyuklu 2019; Lim 2020).

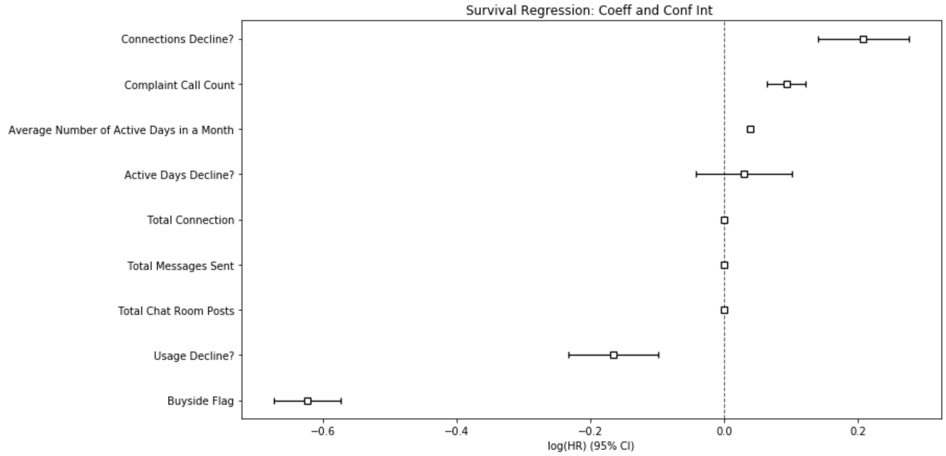

Figure 8. Survival Regression

6. Conclusion

In this study, the ML algorithm wanted to solve two main issues. First, identify the users that potentially will churn and identify the key churn indicators. As the models that mainly deals with the customer churns analysis, this will be impactful as it added new dimensions to the way the organization analyzed the customer churns especially the insights that both the Decision Tree Classifier and Survival Analysis were able to generate. Hence, the business can gain a multi-dimensional perspective on how to analyze and manage the churn and ultimately, improve customer retention. Identifying key areas on where and what causes the disengagement of the users, as well as the proactive measures that the organization can implement, given that decreasing rate of customers affects the revenue stream of the organization. The results from both Decision Tree Classifier and Survival Analysis have shown alignment with the previous studies (The application, methodology and results aligned and agreed with the previous studies (Xiahou & Harada 2022; Wu et al. 2021; Jain et al. 2021; Geetha et al. 2020; He et al. 2020; Kim et al. 2020; Mauritsius et al. 2020; Ullah et al., 2019, Ebrah & Elnasir 2019; King & Rice 2019; Masarifoglu & Buyuklu 2019; Lim 2020). Hence, these combined algorithms were synthesized and applied in the formulation of retention strategy to proactively target customers who are predicted to churn.

As part of the Customer Retention Strategy Framework, Table 3 below shows the retention strategies that specifically been selected in this study as viability was deemed. This will be used in the subsequent presentation of the Customer Retention Strategy framework.

Table 3. Customer Retention Strategy Description

Strategies	**Description**
Customer Feedback Loop implementation	The organization will leverage on survey method such as Net Promoter Score, to understand the sentiment of the users. This will also help the organization to quickly adjust to meet the changing preference of the users.
Personalized Customer Experience	The organization will leverage on the personalized customer experience which they have not been implement in the past. The organization will delegate an account manager help the users in customizing the features and setup the platform based on their preference.
Intensify Customer engagement	To supplement the regular delivery of newsletters, a support team will be deployed to reach out the predicted churning users with the main goal of getting direct sentiments from these users regarding on their user experience, and immediately address whenever possible.
Engagement through Newsletter	Communicate with users/ customers through newsletters to keep them informed on the latest updates on the platform and to continually maintain engagement. These updates normally contained the recent upgrade on the feature. This will also serve as the marketing channel for the organization.
A/B Testing	The company will perform A/B testing or split testing on the existing user base to improve the user experience on the platform as well as to better understand the

	preference of their existing users. This will also be an avenue for the company to perform "smoke test" or ascertain if the new features introduced are working as expected

Figure 9 shows the process workflow of the Customer Retention Strategy framework, built through Lucidchart.com. First, data should be collected and provided by the data management team. Next, the Analytics team subsequently will initiate the Predictive Analytics Algorithm. Any users that will be predicted by the algorithm to churn will be targeted in a proactive manner which will help in their retention before their actual disconnection from the platform. Key strategies for the predicted churning users will focus on intensified engagement as well as improvement on the user experience. Meanwhile, standard protocol will be applied for the rest of the user base, which apparently will focus on the maintained levels f engagement through newsletters and other regular ads posting, and also continuous user experience (UX) improvement through A/B Testing. As seen from the flow chart below, various scenarios can be captured for users that will be predicted to churn and that each of these scenarios will be addressed by the organization through the combination of the customer retention strategies, as realistically, one strategy might not suffice to address the churning users.

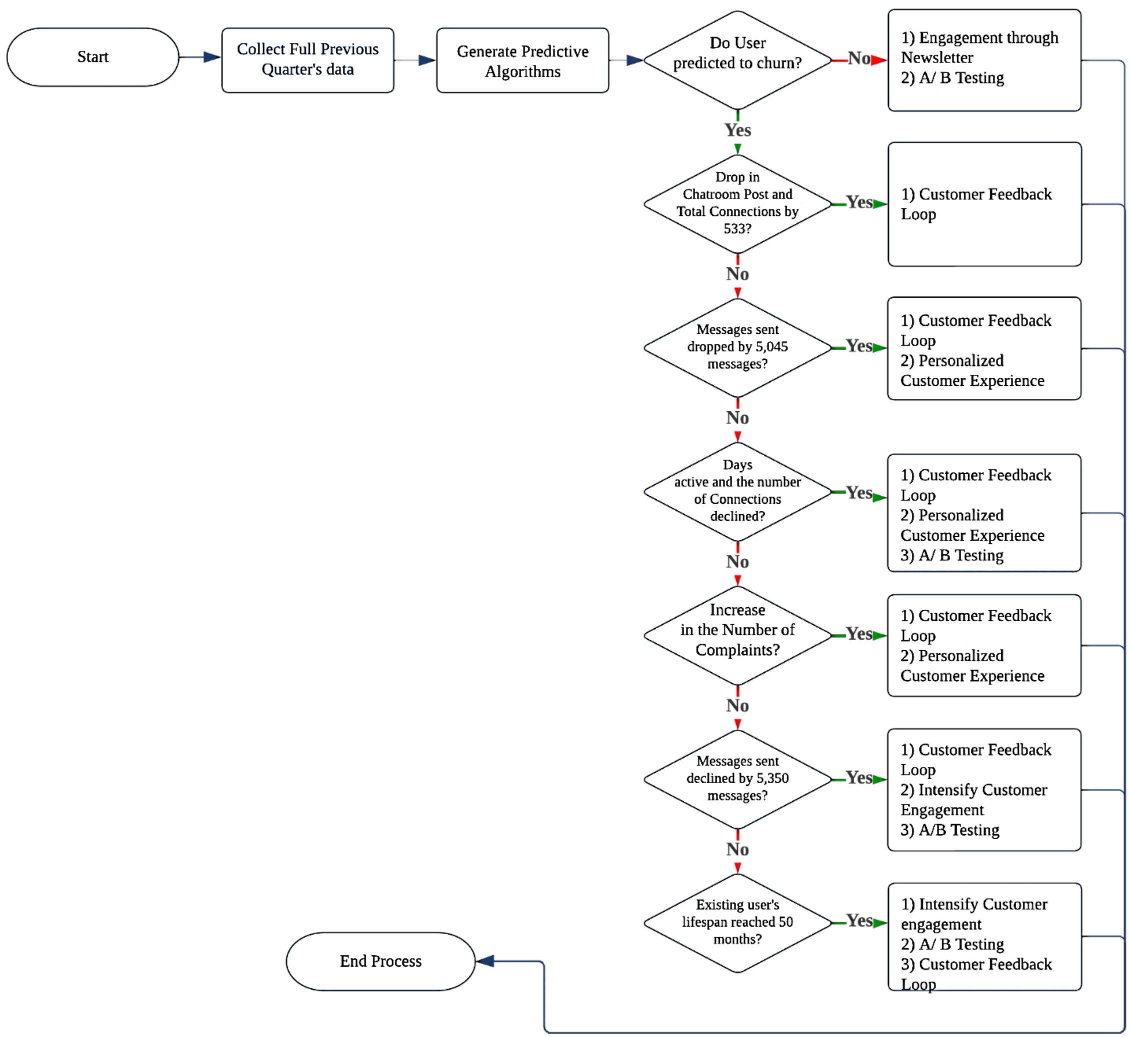

Figure 9. Customer Retention Strategy Workflow

In terms of scheduling, this proposed customer retention strategy framework should be carried out within the first month of the following quarter, since it is required to capture the entire data from the previous quarter, which will be fed into the algorithm. In other words, the predictive algorithm-based strategy should be done four times a year, in order to ensure the efficacy of the proposed customer retention strategy framework.

As for the direction of this paper, the low dimensionality of the dataset used in this study might be an opportunity for future researchers to delve further into similar study. This paper also paved way to the application of Predictive Analytics, not only in the context of web-based collaboration platform or software, but also, to other domains. At present, there were only few Predictive Analytics research papers on the realm of web-based collaboration platform. Furthermore, future researchers can drill further into this domain using other methodologies. Given the recommendations and directions for future researchers, this only shows the great potential of Predictive Analytics in solving real world problems.

References

Agrawal, S., Das, A., Gaikwad, A., and Dhage, S., Customer churn prediction modelling based on behavioral patterns analysis using deep learning, *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, pp. 1-6, 2018.

Ahmad, A.K., Jafar, A., and Aljoumaa, K., Customer churn prediction in telecom using machine learning in big data platform, *Journal of Big Data*, vol. 6, no. 28, 2019.

Charan, R., The cox proportional hazards model, available: https://towardsdatascience.com/the-cox-proportional-hazards-model-35e60e554d8f, July 24, 2020.

Ebrah, K. and Elnasir, S., Churn prediction using machine learning and recommendations plans for telecoms, *Journal of Computer and Communications*, vol. 7, no. 11, pp. 33-53, 2019.

Feng, X., Wang, C., Liu, Y., Yang, Y., and An, H. G., Research on customer churn prediction based on comment emotional tendency and neural network, *Journal of China Academy of Electronics Science*, vol. 13, no. 3, pp. 340-345, 2018.

Geetha, V., Punitha, A., Nandhini, A., Nandhini, T., Shakila, S., and Sushmitha, R., Customer churn prediction in telecommunication industry using random forest classifier. *2020 International Conference on System, Computation, Automation and Networking (ICSCAN),* pp. 1-5, 2020.

He, Y., Xiong, Y., and Tsai, Y., Machine learning based approaches to predict customer churn for an insurance company, *2020 Systems and Information Engineering Design Symposium (SIEDS), pp.* 1–6, 2020.

Jain, H., Yadav, G., and Manoov, R., Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques, *Advances in Machine Learning and Computational Intelligence,* pp. 137–156, 2022.

Karvana, K., Yazid, S., Syalim, A., and Mursanto, P., Customer churn analysis and prediction using data mining models in banking industry, *2019 International Workshop on Big Data and Information Security (IWBIS),* pp. 33–38, 2019.

King, B. and Rice, J., Analysis of churn in mobile telecommunications: Predicting the timing of customer churn, *AIMS International Journal of Management,* vol. 13, pp. 127, 2019.

Lee, E., Jang,Y., Yoon, D.-M., Jeon, J., Yang, S.I., and Lee, S.K., Game data mining competition on churn prediction and survival analysis using commercial game log data, *IEEE Transactions on Games,* vol. 11, no. 3, pp. 215-226, 2019.

Lee,E., Kim,B.,Kang, S., Kang, B., Jang, Y., and Kim, H.K., Profit optimizing churn prediction for long-term loyal customers in online games. *IEEE Transactions on Games*, vol. 12, no. 1, pp 41-53, 2020.

Lim, T., Applying survival analysis for customer retention: A U.S. regional mobile service operator, *2020 Zooming Innovation in Consumer Technologies Conference (ZINC), pp.* 338-342, 2020.

Masarifoglu, M., and Buyuklu, A., Applying survival analysis to telecom churn data, *American Journal of Theoretical and Applied Statistics*, Vol. 8, no. 6, pp. 261-275, 2019.

Mauritsius, T.; Kristianto, Sayoga, R.Y., Alamas, N., Anggraeni, M., and Binsar, F., Customer churn prediction models for PT. XYZ insurance, *2020 8th International Conference on Orange Technology (ICOT),* pp. 1–6, 2020.

Saghir, M., Bibi, Z., Bashir, S., and Khan, F. H., Churn prediction using neural network-based individual and ensemble models, *2019 16th International Bhurban Conference on Applied Sciences and Technology* (IBCAST), pp. 634-639, 2019.

Ullah, I., Raza, B., Malik, A.K., Imran, M., Islam, S.U. and Kim, S.W., A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector, in *IEEE Access,* vol. 7, pp. 60134- 60149, 2019.

Wu, S., Yau, W. C., Ong, T. S., and Chong, S. C., Integrated churn prediction and customer segmentation framework for telco business, *in IEEE Access,* vol. 9, pp. 62118-62136, 2021.

Wu, J., Shi, L., Yang, L., Niu, X., Li, Y., Cui, X., Tsai, S., and Zhang, Y., User value identification based on improved RFM model and k-means++ algorithm for complex data analysis, *Wireless Communications and Mobile Computing*, pp. 1–8, 2021.

Xiahou, X., and Harada, Y., B2C E-commerce customer churn prediction based on k-means and SVM, *Journal of Theoretical and Applied Electronic Commerce Research. 2022,* vol. 17, no. 2, pp. 458–475, 2022.

Zhuang, Y., Research on e-commerce customer churn prediction based on improved value model and XG-boost algorithm. *Management Science and Engineering*, vol. 12, no. 3, pp. 51-56, 2018.

Biographies

Neil T. Awit is currently a Senior Business Analyst for Sales Analytics and Reporting at RingCentral Corp. He has more than 10 years of experience in Analytics where he started his career with Thomson Reuters in 2012. He also serves as industry lecturer at the Corporate Management department of Lyceum of the Philippines University- Manila. He completed his bachelor's degree in Business Economics and Master of Business Administration from Colegio de San Juan de Letran- Manila. He is also a graduate of Master in Business Analytics at Mapua University.

Ramon M. Marticio is currently a Professor of Graduate School at Emilio Aguinaldo College. He is also professorial lecturer at College of Business Management and Accountancy at Trinity University of Asia and UST Angelicum College. He became the Program Chair /Head of Business Administration in the College of Business Administration and Accountancy at Colegio de San Juan de Letran from 2002-2006 and 2013-2015. Also, became the Associate Professor of Graduate School at the same institution from 2015-2020. He was conferred with Doctor of Business Administration at Colegio De San Juan de Letran in 2017 and Master of Business Administration at Lyceum of the Philippines University in 2001. He also has a Diploma in Economic Development & Crisis Management-from Korea Development Institute (KDI) in 2004.