Modeling Client Churn for Small Business-to-Business Firms

Winfred Hills Jr.
School of Data Science
University of Virginia
Charlottesville, USA
whh3rz@virginia.edu

William Daniel
School of Data Science
University of Virginia
Charlottesville, USA
wkd9th@virginia.edu

Mo Yang Lu
School of Data Science
University of Virginia
Charlottesville, USA
ml3qf@virginia.edu

Oliver Schaer
School of Data Science
University of Virginia
Charlottesville, USA
os9vu@virginia.edu

Stephen Adams
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, USA
sca2c@virginia.edu

***Abstract*—With the widespread adoption of customer relationship management (CRM) systems such as Salesforce, HubSpot and Oracle, businesses are becoming increasingly aware of their customer churn rates. Churn rates describe how many customers stop using a product or service within a certain time period and provide a sense of the businesses' long-term viability. Business-to-Business (B2B) firms place high value on the ability to predict individual customer churn, as it presents an opportunity to retain key clients in an inherently limited customer portfolio. These predictions must be both actionable and timely if a manager hopes to retain their client, since a client's churn decision occurs months before the observed churn event. This study explores the HubSpot data of a B2B organization. The objective is to determine the client characteristics that predict sustained product usage and to analyze the indicators of potential churn. Our approach was to model the predictive features of client churn, which would allow managers to directly map churn probability to business strategies. Our final models flagged a handful of management-adjustable features that were significant for predicting customer churn and survival times.**

***Index Terms*—customer churn, behavior modeling, time-series, regression, customer value, predictive modeling**

I. Introduction

Small Business-to-Business (B2B) firms typically provide specialized services to a small pool of clients. Due to this service specialization, value is derived from cultivating long-term relationships with existing clients rather than generalizing and expanding the service to a larger client pool. When a single client churns, small B2B firms risk losing a larger percentage of revenue when compared to large B2B firms with many clients or even compared to small firms that provide services directly to consumers (D2C). As an example, the data set used in this study contains only 242 unique clients over a three-year period. Given this limited client pool, managers of small B2B firms prioritize on opportunities to reduce customer churn. The key to finding these opportunities is identifying the individual clients at high-risk of churning.

Since large data are a natural fit for predictive modeling, smaller firms have been underrepresented in customer churn research. Churn research draws heavily on large D2C companies which service millions of customers. Example industries include telecommunications [1]–[3], online retail [4], and banking [5]. Yet predictive churn modeling provides unique opportunities for small B2B firms. For such firms, insights drawn from predictive modeling can lead to direct churn-risk intervention for specific clients. This is only possible through the nuanced relationship small B2B firms have with their clients. Given the sheer magnitude of customers, larger firms do not have the opportunities to target client churn at this individual level. We propose that churn modeling is just as worthwhile for smaller B2B firms as for larger D2C firms. The loss of a single customer for a small B2B firm represents a substantial loss in revenue.

Churn is traditionally treated as a classification problem, with churn as the response and customer features as the predictor variables. Considering the unique opportunity for targeted churn intervention, managerial interpretation is a high priority in modeling. Our models are designed to answer four questions:

1) What are the churn probabilities of the current client portfolio?
2) If our firm changes a specific feature of our client relationship, what is the impact on their churn probability?
3) If our firm changes a specific feature of our client relationship, what is the impact on their time to churn?
4) What features of our client relationship have the most impact on churn rates?

Our approach focuses developing on two models: a logistic regression and a time-series model. Multiple algorithms were tested for both model types, using receiver operating characteristic (ROC), area under the curve (AUC), and concordance metrics. The final models account for client features which best predict churn probability and churn time and can be directly adjusted by managers.

978-1-7281-7145-6/20/$31.00 ©2020 IEEE

This paper will begin with an overview of the business problem and the priorities of the modeling process. We follow this with a discussion on feature importance, feature selection and feature transformations, which feed into the logistic regression and survival models. We conclude by comparing model performance and discuss the business implications of our findings.

II. Related Work

The research in B2B presently is represented by work exploring non-contractual customers across different industries, basically to mirror the D2C research as closely as possible. [6]. A case study approach, evaluating customer churn of a B2B logistics company involved one of the largest in China. The company still maintained over 10000 customers [7]. B2B companies in the services industry contracts are custom. Consequently churn events are rare, leading to data imbalance in the classification problem [8]. From a qualitative perspective, contract agreements are usually long-term pacts made between people with personal relationships, although that contract affects the entire company. Employee turnover on the client and company side can adversely affect these relationships. As companies continue to grow, they have to efficiently manage these high-touch relationships to keep clients satisfied with proper attention and employees satisfied with a balanced work-life. The client business performance, unrelated to your project may affect their position to remain a client. The aforementioned factors can a lot times be out of the company's control [9]

III. Data

A. Data Overview

The data set is customer data taken from a B2B industrials company selling safety devices. Summary statistics can be seen in Table I. From their customer relationship management platform, we have a data set of 242 individual customers with 16 variables including the event or "churn variable." The data is imbalanced with only 48 of 242 data points (19.83%) of the data representing churned (1) customers (Fig. 1).

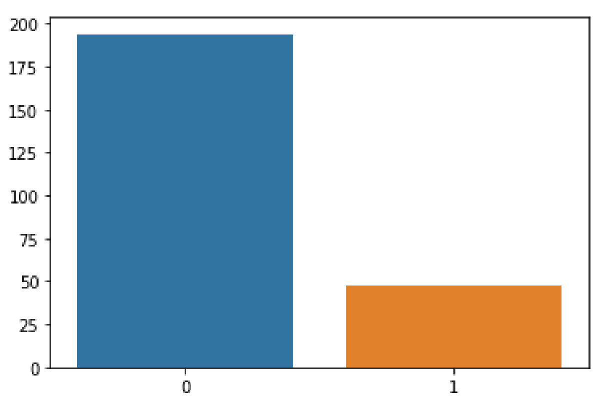

Fig. 1. Active vs Churn Clients

To clean the data, the churn column was made to be a binary denoting active (0) or inactive customers (1). The call-cycle column was made to be a continuous variable based on calls per year, instead of a categorical. Categorical variables were dummy-encoded and gauge was all but ignored, as it is a subjective measurement for renewal probability entered by sales representatives.

TABLE I
Variables in Customer Dataset

Variable Name	***Description***	***Summary Data***
churn	Binary Churn Marker	churned:0.1982, active: 0.8018
pageviews	Visits to Resource Pages	min: 0, median: 3, max: 1807
admins	Num Paid Admins	min: 0, median: 0, max: 600
employees	Employee Count	min:0, median:500, max: 10000
usecompetitors	Usage of Competitors	True: 0.2066, False: 0.7934
contractdays	Contracted Service Days	min: 0, median: 0, max: 68
callcycle_numeric	Sales Contact Cycle	min:0, median: 1, max:12
associateddeals	Num Renewal Deals	min: 0, median: 2, max: 21
timescontacted	Num Hubspot Contacts	min: 0, median: 18, max: 666
origsource	First Contact Source	string (ex: Direct Traffic, Offline Services)
sessions	Num Sessions, Marketing or Support Visits	min: 0, median: 1, max: 879
FF	Third Party Lead	True: 0.4710 False:0.5290
gauge	Manager Est. of Renewal	string (ex: green, yellow, red)
industry	Company Industry	string (ex: Construction, Energy, Insurance)
strategic	Larger Accounts	True: 0.1942, False: 0.8158
competingProducts	ComProd Type	string (ex: SAP, Microsoft, Enablon)

IV. High Level Approach

The key features of our data revolved around client characteristics, such as *industry*, *acquisition channel*, or *employee count*, and client relationship metrics, such as *time as customer*, *support site visits*, or *call cycle*. With this data, we developed a two-prong approach. The first is using a standard customer churn model, which identifies the probability of a customer churning. The second approach is to explore the time dimension of customer churn using survival modeling.

A. As a customer churn regression model

Traditionally, a customer churn model has been treated as a supervised learning problem [1] [4]. The dependent variable is a binary classifier which indicates whether a customer has already churned or not. Churn models are built from data sets which have both current customers and churned customers. Without current customers, the resulting model will be biased towards the churn event.

If model interpretability is not a concern, state-of-the-art approaches would be using a tree-based method, like random forest [10], or a boosting model, like XGBoost [11]. However, the largest downsides to these models are that they are not as interpretable as more traditional approaches, like a logistic regression model [12]. Our priority for this study is to build a model that provides managerial interpretability. For example, being able to say, "A change in variable X will lead to a decrease in risk of churning by Y percent". We found that a logistic regression model struck the best balance between accuracy, precision and ease of interpretation. However, it was important to highlight and compare the logistic regression model against baseline models suggested by the literature in the form of random forest and XGBoost.

B. As a survival model

Time is a particularly relevant feature when analyzing customer churn. In a B2B context, service contracts outline the lifetime of a client and represent repeated revenue. As clients approach their service contract renewal dates, B2B managers focus on customer retention efforts. However, shifting focus to client retention only before renewal dates is a flawed business strategy. Clients can make their renewal decision months ahead of the actual renewal date. In some circumstances, service contracts may also be cancelled before the renewal period. Given these challenges, predicting *when* a client will churn provides robust business insights. Survival models predict the timeframe for client churn, leading to opportunities such as specialized client targeting or portfolio prioritization.

Survival models are a natural fit for modeling customer churn [2], [3]. Like a regression model, we are interested in client features and the churn classifier. We are also interested in the survival time of each client, which marks how long a client lasted before the churn observation. While each observation needs a starting and ending point, these times can be independent. There is no requirement for each observation to start at the same time, which allows the model to reflect business reality of clients purchasing and churning out of the service.

Survival models are also designed to handle right-censoring, which is when the full lifetime of an observation cannot be observed [13]. In our data set, current clients are considered right-censored, since churn has not been observed before the study ends. Excluding these observations of right-censoring would bias the model towards churned clients, which would underestimate the lifespan of a client. For the response variable, we are interested in the hazard rate. It is defined as the probability of observing the event within a specific time interval [14]. Our event of interest is churn. As the time interval gets smaller, the hazard rate can be interpreted as a spot event rate for the surviving observations at time t.

We draw insights about the client portfolio using a two-stage approach. First, using the Kaplan-Meier Estimate (KME), we found the survival curve of the generalized population. This generalized model allows us to estimate the times when the churn risk increases. KME, however, is insufficient for analyzing the effect of observation features on the hazard rate, so the next step was to consider survival regression. We will account for client features by considering three commonly used survival regression models: Cox proportional hazards, Aalen's additive and Weibull accelerated failure time.

V. Models and Methodology

A. Logistic regression model

With logistic regression being our preferred model of choice for predicting customer churn, our attention turned to the following tasks: feature importance, feature selection, and feature transformation. The logistic regression model provides the interpretability that is needed for our primary objective.

In order to determine the feature importance measure of the logistic model we use three key methods: a pearson correlation matrix to determine which features are correlated with each other [15], an algorithm to determine the best K univariate models [16], and feature importance measures from our baseline models (XGBoost and random forest).

Our small data set allows us to use an exhaustive feature selection algorithm [17], which runs through every combination of features and returns the feature set with the best f1 score.

Utilizing a combination of Box-Cox transformations [18], higher-level terms, and interaction variables, we can arrive at a final model to predict customer churn.

For testing methodology, our small sample size restricts us from using a true train and test set. Thus, we use three-fold cross-validation to find precision-recall matrices, f1 score [19], ROC Curves, and AUC metrics [20]. Our best performance metric is looking at a ROC Curve, as the ROC curve gives the business a complete picture of the weight given to false positives or false negatives.

B. Survival model

To get a sense of overall client behavior, we built a KME model. KME assumes that the survival probability is dependent on time alone, so no feature selection is required for this step. The output is a survival curve, which plots client survival as a function of days as customer. KME can also account for feature variability to a limited extent. Here, we study the effect of changing one feature on the survival curve. The observations are divided into cohorts based on the condition, which allows for a direct comparison of survival rates across cohort groups. KME survival curves are a starting point for drawing churn insights and can provide a foundation for broader retention strategies. Specialized client retention strategies, which are of particular interest to B2B firms, must regress the effect of individual features on survival rates. This is beyond the scope of a KME model.

To account for the impact of individual features, our next step was building survival regression models. The first model we considered was Cox proportional hazards (CoxPH). Much like a logistic regression model, CoxPH explores the effect of multiple features. Instead of churn probability, the dependent variable is the hazard function, which describes the survival rate at time t. Each feature is assigned a regression coefficient, with negative coefficients reducing the hazard and positive coefficients increasing the hazard.

VI. Results

A. Logistic Regression

Fig. 2 is constructed using pearson correlation. Given these results, the only highly correlated features are *sessionsPerDay* and *pageviews*. To avoid multicollinearity, our model uses only one of these variables.

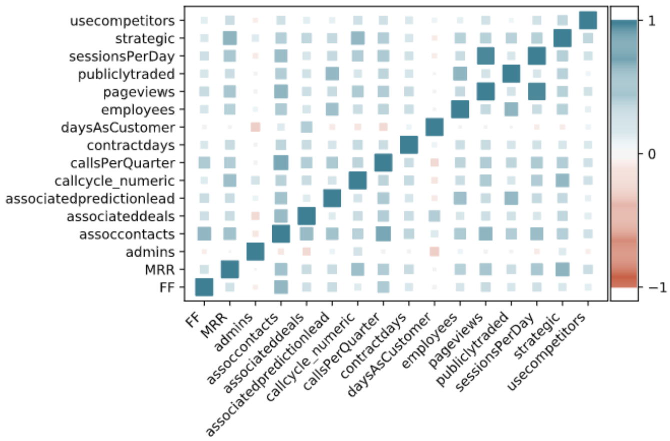

Fig. 2. Feature correlation matrix

We further trim down our feature set based on the results of our baseline models. Given random forest produced better performance results (see Fig. 3 to compare ROC curves between random forest and XGBoost), we use that base model to determine feature importance for our final logistic regression model. Using the results from Table II, we feed in the top features into our exhaustive search algorithm to determine the best subset of features for our model.

TABLE II
RANDOM FOREST FEATURE IMPORTANCE

Feature	*Relative Importance**
timescontacted	2.39
MRR	2.25
callsPerQuarter	2.19
daysAsCustomer	2.00
assoccontacts	1.82
associateddeals	1.80
sessionsPerDay	1.70
employees	1.50
pageviews	1.46
callcycle_numeric	1.00

*Relative importance calculated as a multiple of callcycle_numeric

To elevate our model, we apply the Box-Cox algorithm and high-level terms to give us the transformations listed in Table III. Given our small data set, we also take a look at every combination of interaction between features to include that into our final model as well.

TABLE III
BOX-COX RESULTS

Feature	*Box-Cox Result*	*Transform Used*
timescontacted	0.31	0.25
MRR	-0.03	logarithmic
callsPerQuarter	0.25	0.25
daysAsCustomer	-0.03	none
assoccontacts	0.17	0.20
associateddeals	0.38	0.50
sessionsPerDay	-0.09	none
employees	0.10	0.10
pageviews	0.00	none
callcycle_numeric	-1.41	square term

As can be seen in Table IV, the baseline random forest model and XGBoost model tend to do have both high precision while logistic regression has better recall.

A high precision percentage and low recall percentage suggests that random forest and XGBoost models tend to under-predict churned customers. Conversely, the logistic regression model tends to over-predict churned customers.

The precision-recall table is only a snapshot of the model, where false positives and false negatives are given equal weighting. By looking at a ROC curve, we can see what would happen to our model if we adjust that assumption.

TABLE IV
PRECISION-RECALL RESULTS

	Random Forest	**XGBoost**	**Log Regression**
precision	68%	**70%**	39%
recall	27%	44%	**75%**
f1-score	39%	**54%**	51%

Looking to the ROC Curves in Fig. 3, the logistic regression model is generally in between the random forest model and XGBoost, which can be seen numerically when looking at the AUC in Fig. 3. However, at certain thresholds of false positive rates, the logistic regression model can actually outperform the other two baseline models.

These two graphics show that the logistic regression model is likely sufficient from a performance perspective.

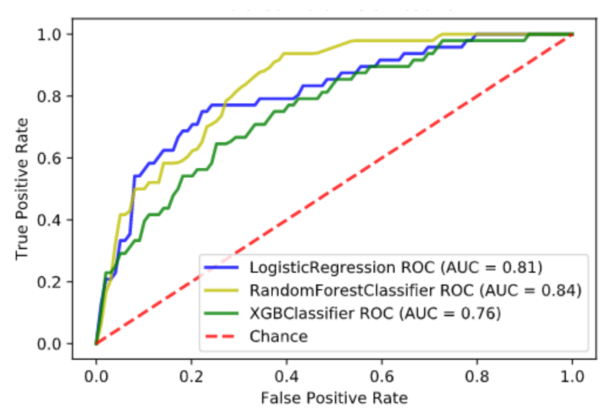

Fig. 3. ROC Curve of different predictive models

B. Results for Survival Analysis

1) Kaplan-Meier Estimates

Reviewing our overall KME, we can see in Fig. 4 The confidence interval becomes drastically larger as the days exceed 2000, due to the sparse observations of clients who survived past that point. The survival rate has substantial declines around day 1000 and day 2000.

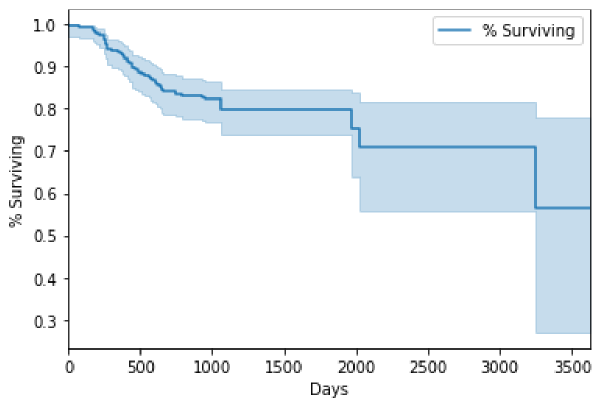

Fig. 4. Overall Client Survival

As an example of cohort analysis, we can examine the survival curves through feature differentiation. A manager can use the KME model to explore how the usage of competitive products impacts the survival rate. We see that clients that use competitive products are less likely to churn in the short term but more likely to churn after 600 days. This may point to a retention opportunity right around a client's 1.5 year mark, seen in Fig. 5. For the competing products cohort, the first decline in survival rate appears to occur before 1000 days, which is earlier than the noncompeting cohort. Later, survival rate dramatically declines for the competitive product cohort around 2000 days. From this analysis, a manager may choose to prioritize short-term retention efforts in clients that use competing products and then shift their attention to maintaining the business of clients who don't use competing products after 2000 days.

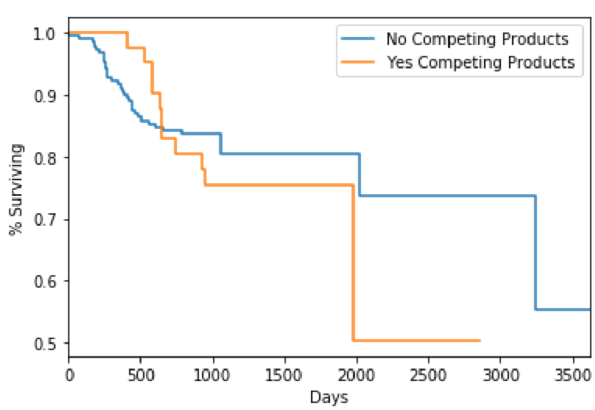

Fig. 5. KME, Competitive Products

a) Survival Regression Models

We begin the survival regression modeling by including the features flagged by Random Forest Feature Importance, seen in Table I. From this point, we individually pared down the insignificant features until we arrived at a finalized CoxPH model. Accounting for the high correlation between *sessions* and *pageviews*, the output is shown in Table V.

TABLE V
FINALIZED COXPH MODEL

Feature	***Coefficient***	***P-Value***
associateddeals	-0.32	0.01
sessions	-0.11	0.01
callcycle_numeric	0.13	0.02

This feature selection process was repeated for two additional hazard models, Aalen's additive and Weibull AFT. We compared concordance rates across all three models and selected a final model from evaluating performance and interpretability.

C. Survival Regression Results

To evaluate the performance of survival models, we compare each model's concordance index. The index is a score between 0 and 1 which measures the model's validity by comparing the order of the survival times. Better performing models have indexes close to 1. The concordance index of each model's best-performing set of features is shown in Table VI. CoxPH and Weibull ATF have a similar level of performance.

TABLE VI
SURVIVAL MODEL CONCORDANCE INDEX

Model	***Concordance Index***
CoxPH	0.78
Aalen Additive	0.76
Weibull ATF	0.78

With these models, we can explicitly explore individual client relationships, such as the survival curve of individual customers (Fig. 6) or the the changes in the survival curve when specific features are adjusted (Fig. 7).

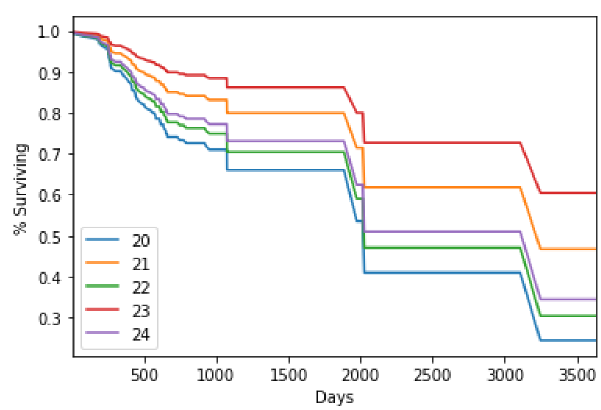

Fig. 6. Client Specific Survival Curves

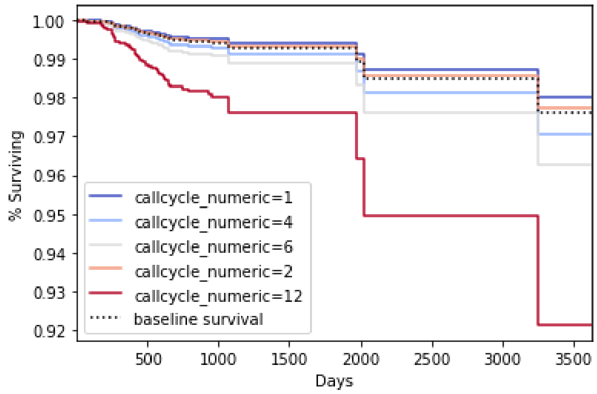

Fig. 7. Feature Adjustment, Call Cycle

A manager may use these client specific curves to explore the impact of proposed changes in a client relationship. For example, a manager may be considering changing a customer's call cycle from yearly (CC = 1) to monthly (CC = 12). The survival curve in Fig. 7 would suggest keeping the calls yearly, due to the decrease in the survival function if call cycle is changed to monthly.

VII. Discussion

A. Logistic Regression

For any classification problem, model performance is contingent upon the threshold for over-predicting or under-predicting. Especially in a business context, the cost of over-prediction and under-prediction can usually be easily quantified. In a customer churn context, we want to weigh the cost of false pursuit (i.e. giving a customer who is not going to churn extra attention) vs. cost of losing an average customer. This analysis will identify where along the ROC Curve we want our model to fall.

Our analysis shows that using a logistic regression model, which enables a business to identify and quantify key characteristics of churn, does not sacrifice overall predictability to our baseline models. This result shows that we can give a business highly accurate and interpretable results that they can use to target susceptible customers.

B. KME Survival Curve

We can draw some insights from the initial KME surival curve, which describes the entire client population without accounting for features (Fig. 4). The Hazard Rate appears to decrease in a near linear pattern from Day 0 to Day 600, where it begins to flatten. There may be a benefit in focusing client retention efforts in this first 1.5 years. The Hazard Rate stabilizes until Day 1800, which is roughly the 5 year mark. The stabilization after Day 600 may point to a transition point where a client becomes a long-term customer. It is notable that the confidence interval expands after Day 600 and balloons after Day 1800. There are fewer observations of long-term clients in the data set, so we must be cautious when drawing insights for long-term client retention.

C. Survival Regression

We see that CoxPH and Weibull ATF models have similar performance metrics. With the initial goal of interpretability in mind, our final survival model will be the CoxPH model. The feature coefficients of the final CoxPH model provide a straightforward connection to the survival rate. Third party associated deals and higher user help sessions increase churn risk, while less frequent callcycle contacts decrease the churn risk. These features can be adjusted as managers consider the client relationship. The models offer a window into the predicted customer lifetimes at the client level (Fig. 6), and can also provide a sense of the changes in that customer lifetime or survival when a specific feature of the customer relationship is adjusted (Fig. 7).

D. Closing Thoughts

Our research has shown that valuable insights about customer churn can be produced through data captured in a business' CRM system. B2B companies, who characteristically have a smaller amount of customers, need these insights to be interpretable to allow the company to act on the information. Our model has shown that interpretability need not to be sacrificed at the expense of model performance.

Further research should be developed on top of this information to determine how a B2B company can best act on this information. For example, B2B companies typically also have a set number of client relationship managers whose job it is to keep clients from churning. Given a data set like this one and with constraints around how much time can be devoted to each customer, how can a model optimize a client relationship manager's time given predicted probabilities of churn?

Alternatively, research could also focus on a generalized customer churn model across many B2B companies. What are the key attributes of customer churn across industries and/or regions of the world? Is a general model achievable?

CRM systems have brought about standardized data sets that enable businesses to understand the key drivers behind customer churn. The availability of this data will promote further research around actionable customer churn insights for years to come.

References

[1] A. Ahmed and D. M. Linen, "A review and analysis of churn prediction methods for customer retention in telecom industries," in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2017, pp. 1–7.

[2] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1659–1665, 2012.

[3] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414–1425, 2012.

[4] A. P. Patil, M. Deepshika, S. Mittal, S. Shetty, S. S. Hiremath, and Y. E. Patil, "Customer churn prediction for retail business," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. IEEE, 2017, pp. 845–851.

[5] Y. Xie, X. Li, E. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, 2009.

[6] A. T. Jahromi, S. Stakhovych, and M. Ewing, "Managing b2b customer churn, retention and profitability," *Industrial Marketing Management*, vol. 43, no. 7, pp. 1258–1268, 2014.

[7] K. Chen, Y.-H. Hu, and Y.-C. Hsieh, "Predicting customer churn from valuable b2b customers in the logistics industry: a case study," *Information Systems and e-Business Management*, vol. 13, no. 3, pp. 475–494, 2015.

[8] L. Yan, R. H. Wolniewicz, and R. Dodier, "Predicting customer behavior in telecommunications," *IEEE Intelligent Systems*, vol. 19, no. 2, pp. 50–58, 2004.

[9] C. Davidson-Pilon, "How b2b and b2c companies solve churn differently." [Online]. Available: https://www.profitwell.com/blog/how-b2b-and-b2c-companies-solve-churn-for-different-customers/

[10] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[12] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[13] P. Tobias, "Censoring." [Online]. Available: https://www.itl.nist.gov/div898/handbook/apr/section1/apr131.htm

[14] C. Davidson-Pilon, "Lifelines survival package documentation." [Online]. Available: https://lifelines.readthedocs.io/en/latest/%20Analysis%20intro.html

[15] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.

[16] E. Bisong, "More supervised machine learning techniques with scikit-learn," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Springer, 2019, pp. 287–308.

[17] S. Raschka, "Exhaustive feature selector." [Online]. Available: "https://bit.ly/3ckkflr"

[18] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–243, 1964.

[19] I. D. Melamed, R. Green, and J. Turian, "Precision and recall of machine translation," in *Companion Volume of the Proceedings of HLT-NAACL 2003-Short Papers*, 2003, pp. 61–63.

[20] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.