ANALYTIC METHODS FOR PREDICTING CUSTOMER CHURN

Stephen Woods

Northwestern University, Introduction to Data Science

10/10/2021

Companies are growingly concerned with increasing the lifetime value of their customers through analytics. While companies have increasing volumes of customer data to leverage, customers also have easily accessible data on competitors. It can be 16 times more expensive to attract a new profitable customer than to keep an existing valuable customer (Domingos, Blessing, and Olawande 2021). Therefore, small reductions in customer churn rates can have large financial impacts. For example, an analytics group at Sprint generated $500 million in revenue through reducing customer churn (Davenport and Harris 2017). Management must consider various data science methods when creating a team to address customer churn.

1. Methods

Recent advances in predictive analytics allow competitive firms to better predict cohorts of customers that may be more likely to leave. This paper covers the most effective data science techniques for predicting customer churn, from more traditional machine learning techniques to deep learning and text analytics.

1.1 Traditional Machine Learning Techniques

Machine learning combines statistics and computer programming to identify correlated variables that help predict outcomes. Das Adhikary and Gupta 2020 compare the accuracy of 100 different machine learning classification models in predicting customer churn. Of these 100 classification techniques, Das Adhikary and Gupta 2020 found two random forest methods to be most accurate. The Regularized Random Forest method had the highest accuracy for classification while the Bagging Random Forests method produced the highest ROC value. Random forests are supervised machine learning algorithms that are built from many decision trees. Within random forests, each decision tree generates a prediction or class, and the average

or mode of those results it ultimately taken as the output of the random forest model (Mbaabu 2020). Other Common supervised learning techniques for churn analysis include support vector machines (SVM) and Multi-Layer Perceptron (MLP) models (Domingos, Blessing, and Olawande 2021).

1.2 Deep Learning

Deep learning is another data science method for predicting outcomes or classifying units of data into groups based on statistics and computer programming. Deep learning generally involves many levels of processing and thus more processing power than traditional machine learning techniques. Domingos, Blessing, and Olawande 2021 found that deep neural networks performed better at predicting customer churn depending on the sample size and hyperparameter tuning. Domingos, Blessing, and Olawande 2021 attributed instances of greater performance to the multiple layers of classification performed in deep learning. Deep learning has emerged as highly effective at "pattern recognition, targeted advertisements, and image processing (Domingos, Blessing, and Olawande 2021). Deep learning techniques are less studied and proven in predicting customer churn, but this promising field is likely to continue advancing to the forefront of data science.

1.3 Text Analytics

Text analytics involve formatting unstructured text into formatted text than can be analyzed to derive insights from words just like numerical data (Miller 2013). Text analytics provides another source of data for companies to analyze and produce insights from. Pustokhina 2021 incorporate text analytics powered by a metaheuristic optimization algorithm to predict customer churn in the telecommunications industry. Common in text analytics, Pustokhina 2021

converted categorical text values into numeric values, performed class labelling on that data, and executed a min-max normalization process to make data more comparable. Text analytics are powerful as they open up a whole new dimension of data for companies to analyze. Company reviews, product reviews, and textual customer service conversations are valuable data sources for text analytics.

2.0 Applications to Management

Predictive analytics can be used to reasonably asses which customers are more likely to discontinue services with a given company. Management should allocate more funds from marketing budgets to analytics focused on predicting and reducing customer churn. Management should consider which of the three general methods discussed best align with their needs and analytic capabilities. Deep learning models require advanced hyperparameter tuning and processing power to outperform more traditional machine learning techniques (Domingos, Blessing, and Olawande 2021). It may be more efficient for companies with less developed data science teams and software to use traditional machine learning techniques. With rapid advancement of predictive techniques, companies aiming to compete on analytics should invest in dynamic analytic capabilities for “adapting, integrating, and reconfiguring resources to match their environment” (Kitchens et al. 2018). An organization well versed in traditional machine learning techniques, deep learning, and text analytics will be able to address customer churn from many angles with both text and numeric data. As deep learning techniques advance and other technologies emerge, companies with systemic commitment to analytics will be poised to lead the pack in customer retention analytics.

References

Das Adhikary, Debjyoti, and Gupta, Deepak. 2020. "Applying over 100 Classifiers for Churn Prediction in Telecom Companies." *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-020-09658-z.

Davenport, Thomas H., and Jeanne B. Harris. 2017. *Competing on Analytics*: *The New Science of Winning (Updated, with a New Introduction).* Cambridge, MA: Harvard Business Review Press

Domingos, Edvaldo, Blessing Ojeme, and Olawande Daramola. 2021. "Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector." *Computation* 9, no. 3: 34–. https://doi.org/10.3390/computation9030034.

Kitchens, Brent, David Dobolyi, Jingjing Li, and Ahmed Abbasi. 2018. "Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data." *Journal of Management Information Systems* 35, no. 2: 540–74. https://doi.org/10.1080/07421222.2018.1451957.

Mbaabu, Onesmus. 2020. "Introduction to Random Forest in Machine Learning." section.io. https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/.

Miller, Thomas W. 2013. *Modeling Techniques in Predictive Analytics: Business Problems & Solutions with R.* (Links to an external site.) Upper Saddle River, NJ: Pearson.

Pustokhina, Irina V, Denis A Pustokhin, Aswathy RH, T Jayasankar, C Jeyalakshmi, Vicente García Díaz, and K Shankar. 2021. "Dynamic Customer Churn Prediction Strategy for Business Intelligence Using Text Analytics with Evolutionary Optimization Algorithms." Information Processing & Management 58, no. 6. https://doi.org/10.1016/j.ipm.2021.102706.