PREDICTING AND DEFINING B2B SALES SUCCESS USING MACHINE LEARNING

1. U.Naresh,assistant professor,CSE,Sri Indu Institute of Engineering&Technology(SIIET), Sheriguda, Ibrahimpatnam,Hydarabad
2. Yaramala Anusha,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
3.Pandiri Naveen,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
4.Veera Shaiva Priyanka,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
5.Surepalli Sai Vamshi Madhur,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
6.T Nikhil Kyumar,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

ABSTRACT:
The objectives of this project are two-fold: 1) to use statistical modeling techniques to help a Fortune 500 paper and packaging company codify what drives sales success and 2) to develop a model that can predict sales success with a reasonable degree of accuracy. The desired long-run result is to enable the company to improve both top-line revenue and bottom-line profits by increasing sales close rates, shortening sales cycles, and decreasing the cost of sales. The research team generated several models to predict win propensities for individual sales opportunities, choosing the model with the greatest predictive power and ability to generate insights to use as the backbone for a client tool. To accomplish this, the team leveraged structured and unstructured data from the company's Salesforce.com customer relationship management system. The team experimented with several techniques including binomial logit and various decision tree methods, including boosting with gradient boost and random forest. Individual attributes of customers, opportunities, and internal documentation methods that have the greatest influence on sales success were identified. The best model predicted win propensity with an accuracy of 80%, with precision and recall of 86% and 77%, respectively, which proved to be an improvement over current sales forecast accuracy.

INTRODUCTION:
The paper and packaging company that provided the data for this research has a long history of sales expertise. This expertise is captured predominantly in the intuition of sales representatives, many of whom have worked in the industry for 20 years or more. Intuition is not easy to record and disseminate across an entire sales force, however, and thus one of the company's most valuable resources is inaccessible to the broader organization. As a result, the company tasked this team with extracting the most important factors in driving sales success and modeling win propensities using data from their customer relationship management (CRM) system.

Most prior work in this space has been performed by private companies, both those that have developed proprietary technologies for internal use and those that sell B2B services related to predictive sales modeling. As a result, research in the field is typically unavailable to the public. Some examples include Implisit [1]—a company recently acquired by Salesforce.com that focuses on data automation and predictive modeling—and InsightSquared [2], which sells software that includes a capability to forecast sales outcomes.

The academic work that does exist either is related to forecasting aggregate sales instead of scoring opportunitylevel propensity, or is based on custom algorithms that fall outside the standard tools used by data scientists in industry. The earliest relevant publication dates only to 2015, in which a joint team from Chinese and US universities employed a two-dimensional Hawkes Process model on seller-lead interactions to score win propensity [3]. Other relevant research has centered around applying highly accurate machine learning algorithms based on sales pipeline data to integrate the insights they produce into an organization's practices, and explaining the output of black-box machine learning models .

Considering the lack of visibility into work predicting sales outcome propensity, this research serves to create an initial baseline of understanding on the subject. This project applies and compares several well-known methods for classifying and scoring propensities, a majority of which fall into the category of decision tree modeling.

MODELS AND METHODOLOGY

The research team employed several well-known classification models to extract important features from the data, in addition to calculating the win/loss propensity for each opportunity record. With the goal of modeling probability, the team chose different supervised machine learning algorithms that fit these criteria: Logistic Regression, Decision Tree, Random Forest, and XGBoost. In each of these supervised algorithms, the classifier was pre-defined with an iterative variable selection process. A classification model was then built with a training set split from the master table and used to predict win propensities

Prior to this date, portions of the company used the system, but it had not been rolled out companywide. examined by the actual win or loss of the opportunities in the testing set built from the remainder of observations.

Variable selection was a critical component of this project. As previously stated, variables came directly from the SFDC system and went through a series of data processing steps. The main purpose of this research was to interpret features that gave the most useful information in terms of win propensity prediction accuracy. Both the quality and quantity of variables significantly affected the accuracy and efficiency of all algorithms. An important consideration about the current data was the widely varying quality of variable inputs. This issue created constraints on the algorithm-generated selection results. Therefore, the variable selection process also involved constant communication and validation between the team and company.

The four algorithms used in this research are briefly described below:

- Multiple Logistic Regression — a generalized linear model (GLM) that describes the relationship between a binary dependent variable and more than one predictor.
- Decision Tree — a non-parametric algorithm that makes sequential, hierarchical decisions about the outcomes based on the predictors.
- Random Forest — an ensemble algorithm that constructs a multitude of decision trees and outputs the mode of the classes, correcting the overfitting habit of decision trees.
- XGBoost — an implementation of gradient boosted decision trees that minimize the loss when producing an ensemble of weak decision trees.

CONCLUSION

This research served as a first step in the development of a broader initiative for a Fortune 500 paper and packaging company to operationalize predictive modeling on sales success. As such, the challenges with any large company often include requiring the building of deep local knowledge of the data, in addition to corralling a large organization to assist with accurate data collection. Despite initial inconsistencies in the data, overall accuracy appeared promising and indicated further improvements could be made with better data quality and quantity, more featurerelated investigation and tuning, or perhaps different methods such as neural nets.

The analysis also uncovered new insights into what is important regarding sales success. But new insights are often accompanied by new questions: For instance, what kinds of data need to be captured to improve the model's predictive capabilities? How does the culture need to change to improve data capture? This cascade is to be expected, as the broader project lends itself to being a heavily iterative process.

There may appear to be a seemingly infinite pool of potential next steps to take in this case. With this in mind, there are a few the team would recommend as the most prudent to consider. Currently, the company could feasibly use the non-meta-variable model to attempt prediction on opportunities in progress for those divisions where accuracy is adequate. To better achieve the objective of predicting open opportunities, it would be prudent to capture and model how opportunity fields change over time, perhaps via periodic snapshots. This way, the company would be able to make predictions at different stages in the opportunity lifecycle.

Another important application of these kinds of prediction models is to assist in determining where to invest sales time and resources for business planning optimization. Predictions from accurate models are also worth rolling up into aggregate sales forecasts and adjusting existing "bottom-up" methods.

Before these applications would be addressed however, data ops resources would be required to perform a number of critical tasks: continue building and tuning the model for better accuracy, establish a cadence around maintaining the models and incorporating new kinds of information, and connecting with the other business units to understand strategic priorities for operationalization.

REFERENCES

[1] Implisit (Sales Cloud by Salesforce.com). [Online]. Available: https://www.salesforce.com/blog/2014/08/infographic-7-powerfulpredictors-closed-won-opportunity-gp.html

[2] Insight Squared. [Online]. Available: https://www.insightsquared.com/features/sales-forecasting/

[3] J. Yan, C. Zhang, H. Zha, et all, "On Machine Learning towards Predictive Sales Pipeline Analytics." Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 1945-1951, 2015. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/download/9 444/9488 [Accessed: Mon. 24 Sept. 2018].

[4] M. Bohaneca, M.K. Borstnarb, M. Robnik-Sikonja, "Integration of machine learning insights into organizational learning: A case of B2B sales forecasting." 28th Bled eConference, June 7-10, 2015. [Online]. Available: https://domino.fov.unimb.si/proceedings.nsf/Proceedings/B12ECF238 1AB59EEC1257E5B004B39B7/$File/2_Bohanec.pdf [Accessed: Tue. 25 Sept. 2018]. [5] M. Bohaneca, M.K. Borstnarb, M Robnik-Sikonja, "Explaining machine learning models in sales predictions." Expert Systems with Applications, no. 71, pp. 416-428, 2017. [Online]. Available: http://lkm.fri.uni-lj.si/rmarko/papers/Bohanec17-ESwA-preprint.pdf [Accessed: Tue. 25 Sept. 2018].