A GENERALIZED FLOW FOR B2B SALES PREDICTIVE MODELING

SAATHWIK CHANDAN NUNE
IIM KOZHIKODE

PURPOSE:
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

AIM:
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.

OBJECTIVE
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 —a company recently acquired by Salesforce.com that focuses on data automation and predictive modeling—and Insight Squared, which sells software that includes a capability to forecast sales outcomes.

ABSTRACT: Predicting the outcome of sales opportunities is a core part of successful business management. Conventionally, making this prediction has relied mostly on subjective human evaluations in the process of sales decision making. In this paper, we addressed the problem of forecasting the outcome of business to business (B2B) sales by proposing a thorough data-driven Machine Learning (ML) workflow on a cloud-based computing platform: Microsoft Azure Machine Learning Service (Azure ML). This workflow consists of two pipelines: (1) An ML pipeline to train probabilistic predictive models on the historical sales opportunities data. In this pipeline, data is enriched with an extensive feature enhancement step and then used to train an ensemble of ML classification models in parallel. (2) A prediction pipeline to utilize the trained ML model and infer the likelihood of winning new sales opportunities along with calculating optimal decision boundaries. The effectiveness of the proposed workflow was evaluated on a real sales dataset of a major global B2B consulting firm. Our results implied that decision-making based on the ML predictions is more accurate and brings a higher monetary value.

Keywords: Costumer Relation Management; Business to Business Sales Prediction; Machine Learning; Predictive Modeling;

INTRODUCTION In the Business to Business (B2B) commerce, companies compete to win high-valued sales opportunities to maximize their profitability. In this regard, a key factor for maintaining a successful B2B enterprise is the task of forecasting the outcome of sales opportunities. B2B sales process typically demands significant costs and resources and, hence, requires careful evaluations in the very early steps. Quantifying the likelihood of winning new sales opportunities is an important basis for appropriate resource allocation to avoid wasting resources and sustain the company's financial objectives [1–4]. Conventionally, forecasting the outcome of sales opportunities is carried out mostly relying on subjective human rating [5–8]. Most of the Customer Relationship Management (CRM) systems allow salespersons to manually assign a probability of winning for new sales opportunities [9]. This probability is then used at various stages of the sales pipeline, i.e., calculating a weighted revenue of the sales records [10,11]. Often each salesperson develops a non-systematic intuition to forecast the likelihood of winning a sales opportunity with little to no quantitative rationale, neglecting the complexity of the business dynamics [9]. Besides, as often as not, sales opportunities are intentionally underrated to avoid any internal competition with other salespersons or overrated to circumvent the pressure from sales management to maintain a higher performance [12]. Even though the abundance of data and improvements in statistical and machine learning (ML) techniques have led to significant enhancements in data-driven decision-making, the literature is scarce in the subject of B2B sales outcome forecasting. Yan et al. [12] explored predicting win-propensity of sales opportunities using a two-dimensional Hawkes dynamic clustering technique. Their approach allowedfor live assessment of active sales although relied heavily on regular updates and inputs from salespersons in the CRM system. This solution is hard to maintain in larger B2B firms considering each salesperson often handles multiple opportunities in parallel and would put less effort into making frequent interaction with each sales record [13]. Tang et al. [9] built a sales forecast engine consist of multiple models trained on snapshots of historical data. Although their paradigm is focused on revenue forecasting, they demonstrated the effectiveness of hybrid models for sales predictive modeling. Bohane et al. [5] explored the idea of single and double-loop learning in B2B forecasting using ML models coupled with general explanation methods. Their main goal was actively involving users in the process of model development and testing. Built on their earlier work on effective feature selection [14] they concluded random forest models were the most promising for B2B sales forecasting. Here, we proposed an end-to-end cloud-based workflow to forecast the outcome of B2B sales opportunities by reframing this problem into a binary classification framework. First, an ML pipeline extracts sales data and improves them through a comprehensive feature enhancement step. The ML pipeline optimally parameterizes a hybrid of probabilistic ML classification models trained on the enhanced sales data and eventually outputs a voting ensemble classifier. Second, a prediction pipeline makes use of the optimal ML model to forecast the likelihood of winning new sales opportunities. Importantly, the prediction pipeline also performs thorough statistical analysis on the historical sales data and specifies appropriate decision boundaries based on sales monetary value and industry segment. This helps to maximize the reliability of predictions by binding the interpretation of model results to the actual data. The proposed workflow was implemented and deployed to a global B2B consulting firm's sales pipeline using Microsoft Azure Machine Learning Service (Azure ML). Such a cloud-based solution readily integrates into the existing CRM systems within each enterprise and allows for more scalability. Finally, we compared the performance of the proposed solution with salespersons' predictions using standard statistical metrics (e.g. accuracy, AUC, etc.). To make the comparison more concrete, we also looked into the financial aspect of implementing this solution and compared the monetary value of our ML solution with salespersons' predictions. Overall, we have found that the proposed ML solution results in a superior prediction both in terms of statistical and financial evaluations; therefore, it would be a constructive complement to the predictions made by salespersons

SCOPE OF PROJECT

The earliest relevant publication dates only to 2015, inwhich a joint team from Chinese and US universitiesemployed a two-dimensional Hawkes Process model onseller-lead interactions to score win propensity. Otherrelevant research has centered around applying highlyaccurate machine learning algorithms based on sales pipelinedata to integrate the insights they produce into anorganization's practices , and explaining the output ofblack-box machine learning modelsConsidering the lack of visibility into work predictingsales outcome propensity, this research serves to create aninitial baseline of understanding on the subject. This projectapplies and compares several well-known methods forclassifying and scoring propensities, a majority of which fallinto the category of decision tree modeling.

LITERATURE REVIEW

On Machine Learning towards Predictive Sales Pipeline Analytics

AUTHORS: Junchi Yan123, Chao Zhang2 , Hongyuan Zha14, Min Gong2 , Changhua Sun2 , Jin Huang2 , Stephen Chu2 , Xiaokang Yang3

Sales pipeline win-propensity prediction is fundamental to effective sales management. In contrast to using subjective human rating, we propose a modern machine learning paradigm to estimate the winpropensity of sales leads over time. A profile-specific two-dimensional Hawkes processes model is developed to capture the influence from seller's activities on their leads to the win outcome, coupled with lead's personalized profiles. It is motivated by two observations: i) sellers tend to frequently focus their selling activities and efforts on a few leads during a relatively short time. This is evidenced and reflected by their concentrated interactions with the pipeline, including login, browsing and updating the sales leads which are logged by the system; ii) the pending opportunity is prone to reach its win outcome shortly after such temporally concentrated interactions. Our model is deployed and in continual use to a large, global, B2B multinational technology enterprize (Fortune 500) with a case study. Due to the generality and flexibility of the model, it also enjoys the potential applicability to other real-world problems.

Integration of machine learning insights into organizational learning: A case of B2B sales forecasting

AUTHORS: M. Bohaneca, M.K. Borstnarb, M. Robnik-Sikonja

Business-to-business (b2b) sales forecasting can be described as a decision-making process, which is based on past data (internal and external), formalized rules, subjective judgment, and tacit organizational knowledge. Its consequences are measured in profit and loss. The research focus of this paper is aimed to narrow the gap between planned and realized performance, introducing a novel approach based on machine learning techniques. Preliminary results of machine learning model performance are presented, with focus on distilled visualizations that create powerful, yet human comprehensible and actionable insights, enabling positive climate for reflection and contributing to continuous organizational learning.

Explaining machine learning models in sales predictions

AUTHORS: M. Bohaneca, M.K. Borstnarb, M Robnik-Sikonja

ISSN:0975-3583,0976-2833 VOL12,ISSUE05,2021

A complexity of business dynamics often forces decision-makers to make decisions based on subjective mental models, reflecting their experience. However, research has shown that companies perform better when they apply data-driven decision-making. This creates anincentive to introduce intelligent, data-based decision models, which are comprehensive and support the interactive evaluation of decision options necessary for the business environment. Recently, a new general explanation methodology has been proposed, which supports the explanation of state-of-the-art black-box prediction models. Uniform explanations are generated on the level of model/individual instance and support what-if analysis . We present a novel use of this methodology inside an intelligent system in a real-world case of business-to-business (B2B) sales forecasting, a complex task frequently done judgmentally. Users canvalidate their assumptions with the presented explanations and test their hypotheses usingthe presented what-if parallel graph representation. The results demonstrate effectiveness andusability of the methodology. A significant advantage of the presented method is the possibility to evaluate seller's actions and to outline general recommendations in sales strategy. This flexibility of the approach and easy-to-follow explanations are suitable for many different applications. Our well-documented real-world case shows how to solve a decision supportproblem, namely that the best performing black-box models are inaccessible to human in-teraction and analysis. This could extend the use of the intelligent systems to areas where they were so far neglected due to their insistence on comprehensible models. A separation of the machine learning model selection from model explanation is another significant benefit for expert and intelligent systems. Explanations unconnected to a particular prediction model positively influence acceptance of new and complex models in the business environment through their easy assessment and switching

EXISTING SYSTEM:

The 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. 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.

DISADVANTAGES OF EXISTING SYSTEM:

- Most prior work in this space has been performed by private companies.
- The research in the field is typically unavailable to the public.
- 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.

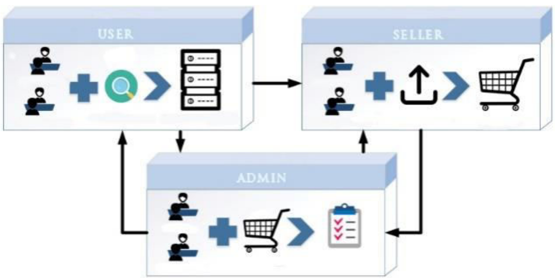
PROPOSED SYSTEM:

To clean the data and cut out inessential information prior to modeling, the team first filtered out all entries when the system was formally launched for the company. Variables with a high percentage of null values were then excluded to ensure a sufficient sample size. The remaining variables were further screened based on potential importance determined by conversations between the team and key company stakeholders. Additionally, data exploration resulted in several opportunities for feature engineering and custom variables to capture potential influence not captured in the default fields.

ADVANTAGES OF PROPOSED SYSTEM:

- Count of the number of fields completed in one record.
- Count of the number of tasks for the customer account associated with an opportunity.
- After a number of iterations between modeling and feature engineering, the final master table used in this analysis included 15 variables and was built on the opportunity-level.

ARCHITECTURE



MODULES:

- user
- seller

- admin
- Machine learning

USER

customer relationship management system (SFDC). SFDC is a software-as-aservice application that allows sales teams to record details about customer relationships and sales opportunities as they move through the sales pipeline. The data included a static snapshot of details on sales employees, customer accounts and account histories, individual customer opportunities, sales representative activities, and contact information. Some inputs in the system were automatically generated and easily readable by machine.

SELLER

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. 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.

ADMIN

admin gives the activation permission to users and seller ,admin can verify the ordered details .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. The data included a static snapshot of details on sales employees, customer accounts and account histories, individual customer opportunities, sales representative activities, and contact information. Some inputs in the system were automatically generated and easily readable by machine

MACHINE LEARNING

Machine learning refers to the computer's acquisition of a kind of ability to make predictive judgments and make the best decisions by analyzing and learning a large number of existing data. The representation algorithms include deep learning, artificial neural network, decision tree, enhancement algorithm and so on. The key way for computers to acquire artificial intelligence is machine learning. Nowadays, machine learning plays an important role in various fields of artificial intelligence. Whether in aspects of internet search, biometric identification, auto driving, Mars robot, or in American presidential election, military decision assistants and so on, basically, as long as there is a need for data analysis, machine learning can be used to play a role.

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].

6. Gephart, M.A., Marsick, V.J., Mark, E., VanBuren, M.E., Spiro, M.S.: Learning organizations come alive. Training Development 50(12), 36–41 (1996)

7. Nonaka, I., Takeuchi, H.: The Knowledge Creating Organization. Oxford University Press,New York (1995)

8. Senge, P.: The Fifth discipline: The Art & Practice of the Learning Organization. Doubleday Currency, New York (1990)

9. Bohanec, M., Kljajić Borštnar, M., Robnik-Šikonja, M.: Modeling attributes for forecasting B2B opportunities acquisition. In: Proceedings of 34th Conference of Organizational science development, Portorož, Slovenia (2015)

10. Ngai, E.W.T., Xiu, L., Chau, D.C.K.: Application of data mining techniques in CRM: a literature review and classification. Expert Syst. Appl. 36, 2592–2602 (2009)

11. Monat, J. P.: Industrial sales lead conversion modeling. Market. Intell. Plan. 29(2), 178–194(2011)

12. Rieg, R.: Do forecast improve over time? Int. J. Account. Inform. Manage. 18(3) (2010)

13. Alvarado-Valencia, J.A., Barrero, L.H.: Reliance, trust and heuristics in judgmental forecasting. Comput. Hum. Behav. 36, 102–113 (2014)

14. Maaß, D., Spruit, M., Waal, P.D.: Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques, pp. 1–17. Decision Analytics, Springer(2014)

15. Witten, I.H., Eibe, F., Hall, M.A.: Data mining—practical machine learning tools and techniques, 3rd edn. Elsevier (2011)