Data-driven Sales Operations Management

In this paper, we present a data-driven approach for the allocation of sales resources at DAW, a large, German-based manufacturer of building paint. DAW’s product portfolio mainly focuses on paint and coating solutions for the professional business and is sold via two main distribution channels: The first one comprises the wholesale business where DAW markets their products to paint wholesalers and the specialist trade. The second channel, which is the one we are focusing on in this work, comprises direct sales to large building sites. In the latter business, the company collaborates directly with potential painters and processors to win these objects and supply their products. Evidently, good customer relations are key in this business - however, the capacity of the sales team is limited and DAW needs to decide how to allocate these scarce resources. Despite the very particular setting in the paint manufacturing industry, the company faces a problem which is very common for a lot of enterprises and which can be abstracted from the concrete application: The company’s potential customers are dispersed over a large area and as a consequence, the company needs to plan which customer to visit in order to optimize the selling potential.

So far, the decision of which customer to visit is completely made by the sales agents based on their own assessment, without providing them with any information about success probabilities for specific sales opportunities. The main reason why this approach will typically lead to suboptimal decisions is that sales agents tend to contact potential customers where a successful sale is very likely. However, these are generally not the sales opportunities where an additional contact by a sales agent yields the highest increase in winning chances, which would be the better criterion.

In contrast to the hands-on approach currently in practice in many companies, we propose a new, fully data-driven method to support and improve sales operations planning. Starting with a raw data set received from DAW we obtain 1200 instances of won and lost contracts after some
preprocessing operations. This data set is then enriched with general characteristics of the project such as type of the object, location or expected total volume of the contract. Also, we consider company-related information such as the history of past projects with involved partners, preliminary work done by DAW as well as the number of visits sales agents made prior to the project decision at involved partners. We then use this augmented data set to build a predictive model that determines the probability of winning or losing a specific project. For this task, we revert to a nonparametric regression model based on boosted decision trees, xgBoost (c.f. Chen and Guestrin, 2016), resulting in an out-of-sample prediction accuracy of 90%. That is, in 90% of the cases, our prediction whether a contract is won or lost is correct.

This predictive model serves as the core building block for our prescriptive analytics solution. We find that the probability of winning such large building contracts emanates strongly from the face-to-face contacts that sales agents spent upfront with involved processors, i.e., the painters and plasterers actually processing DAW’s products. Based on these insights, we develop a data-driven decision support system with which we can identify the most promising processors that the sales team should visit: We determine the value of an additional visit by calculating the probabilities of winning the project with an assumed additional visit and multiply it with the expected contract volume. We then use this information to provide the sales agents with an ordered list of the most important customers to visit. Moreover, based on this information, we can optimize the travel routes of the sales agents in a prize-collecting travelling salesman model.

Our work is broadly related to two streams in the literature, uplift and response modeling on the one hand as well as the literature on the classical traveling salesman problem on the other hand. Many applications of response modeling in a business context can be found in the marketing discipline, e.g., to predict the success of a particular marketing action such as a telemarketing call. In that context, Cui et al. (2006) and Moro et al. (2014) show that machine learning approaches can be very valuable tools to model consumer response. However, these approaches focus on only predicting the success probability of a particular action and not on predicting the change in success probabilities following such an action. In the literature this is known as uplift modeling. Rzepakowski and Jaroszewicz (2012) for example, propose a tree-based approach for such an application. However, their approach is not transferable into the DAW setting, since it requires to actually perform the
action or treatment (in our case the visit) on a subset of the objects and compare the outcomes with a control group where there was no treatment.

On the other hand, there exist numerous papers (e.g. Bales, 1989; Bienstock et al., 1993) on finding optimal routes for sales representatives to visit their (potential) clients. However, we are not aware of approaches combining the insights of uplift models with a downstream routing problem such as the one we are considering at the example of DAW.

Our work contributes to the literature in two ways: First, we provide an uplift model based on machine learning that does not require to actually test a direct marketing action but leverages historical data about a company’s won and lost contracts. Second, we integrate these insights in a prescriptive model that optimizes the route planning of sales agents in a real-world problem.

References


