

DOI: 10.1515/orga-2017-0020



Organizational Learning Supported by Machine Learning Models Coupled with General Explanation Methods: A Case of B2B Sales Forecasting

Marko BOHANEC¹, Marko ROBNIK-ŠIKONJA², Mirjana KLJAJIĆ BORŠTNAR³

¹ Salvirt, Ltd, Dunajska 136, 1000 Ljubljana, Slovenia
marko.bohanec@salvirt.com

² University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, 1000 Ljubljana, Slovenia
marko.robnik@fri.uni-lj.si

³ University of Maribor, Faculty of Organizational Sciences, Kidričeva 55a, 4000 Kranj, Slovenia
mirjana.kljajic@fov.uni-mb.si

Background and Purpose: The process of business to business (B2B) sales forecasting is a complex decision-making process. There are many approaches to support this process, but mainly it is still based on the subjective judgment of a decision-maker. The problem of B2B sales forecasting can be modeled as a classification problem. However, top performing machine learning (ML) models are black boxes and do not support transparent reasoning. The purpose of this research is to develop an organizational model using ML model coupled with general explanation methods. The goal is to support the decision-maker in the process of B2B sales forecasting.

Design/Methodology/Approach: Participatory approach of action design research was used to promote acceptance of the model among users. ML model was built following CRISP-DM methodology and utilizes R software environment.

Results: ML model was developed in several design cycles involving users. It was evaluated in the company for several months. Results suggest that based on the explanations of the ML model predictions the users' forecasts improved. Furthermore, when the users embrace the proposed ML model and its explanations, they change their initial beliefs, make more accurate B2B sales predictions and detect other features of the process, not included in the ML model.

Conclusions: The proposed model promotes understanding, foster debate and validation of existing beliefs, and thus contributes to single and double-loop learning. Active participation of the users in the process of development, validation, and implementation has shown to be beneficial in creating trust and promotes acceptance in practice.

Keywords: *decision support; organizational learning; machine learning; explanations; B2B sales forecasting*

1 Introduction

Business-to-Business (B2B) sales forecasting is a complex (inter)organizational process that is tightly related to decision making. The dynamic environment (economic and

political), multi-stage sales processes, multiple participants with possibly conflicting interests (sellers, buyers), and multiple interrelated attributes all contribute to the complexity of the process. B2B sales forecasts serve as the basis for managerial decisions that result in resource

allocation. Implications of incorrect forecasting can lead to non-optimal decisions and consequently to a waste of resources.

Forecasting of the sales outcomes is a well-researched subject, especially in the context of time series. Although there is a vast body of literature and technological advancement on the topic of forecasting (Fildes, Goodwin and Lawrence, 2006; McCarthy, Davis Golcic and Mentzer, 2006; Armstrong, Green and Graefe, 2015), there is a weak evidence on successful business implementations. Decision-makers remain skeptical about recommendations offered by forecasting support systems (FSS) and rather rely on applying their own mental models (Goodwin, Fildes, Lawrence and Stephens, 2011) the resulting forecasts are often 'sub-optimal' because many judgmental adjustments are made when they are not required. An experiment was used to investigate whether restrictiveness or guidance in a support system leads to more effective use of judgment. Users received statistical forecasts of the demand for products that were subject to promotions. In the restrictiveness mode small judgmental adjustments to these forecasts were prohibited (research indicates that these waste effort and may damage accuracy. Mental models are reflected as deeply rooted assumptions and generalizations that influence the way individuals act and are often unconsciously reflected in behavior that limits the organization's development capabilities. If an organization wants to improve its efficiency (i.e. decreasing the gap between the forecasts and realization), it needs to reflect upon those anchored mental models.

Organizational learning thus represents a constant effort to create organizational knowledge, which according to Senge (1990) consists of team learning, personal mastery, mental models, building a common vision, and systemic reflection as an all-inclusive, fifth discipline. Furthermore, it refers to organizations' ability to adapt effectively to changes in its environment. In an analogy to individual learning, it can be described as an alteration of the behavior based on an individual/group experience (Škraba et al., 2007) or a process of detection and correction of errors (Argyris and Schön, 1996). In contrast to individual learning, the organizational learning is more complex since it is not only a sum of individual learning but an exchange of individual models, beliefs, and behaviors. When it is based on a feedback, and individuals and groups change their mental models (beliefs and behaviors), we speak about double-loop learning. It can occur on an individual level, but it is rarely observed outside an organizational setting (Kljajić Borštnar et al., 2011).

The two types of learning were defined by Argyris and Schön (1996). A problem solving oriented single-loop learning is often efficiently supported by a "black-box" models ("know-how"). Double-loop learning assumes critical reflection that leads to an understanding of a system ("know-why"). Understanding why complex (inter)

organizational systems operate in a certain way, helps in identifying changes in the environment and reacting to them, effectively supporting organizational learning. Transparent models thus create the basis for double loop learning and promote organizational learning (Größler, 2000; Flesichmann and Wallace, 2005). However, many models, including top performing machine learning (ML) models, are black boxes and cannot support such learning.

The learning process should provide sufficient knowledge for effective decision-making (Škraba et al., 2007), while feedback is the key part of the learning process (Kuchinke, 2000). Simon (1960) proposed the model of the decision process, which consists of three basic phases: intelligence, design, and choice among options. The first phase deals with the identification of the problem and the collection of data they describe. The design phase deals with the preparation and design of different decision options, which are then evaluated in accordance with the set criteria in the third phase with the goal of choosing the best option.

In this paper, we focus on all three Simons' (1960) decision-making phases, especially focusing on organizational aspects and people-related evaluation. The novel use of general explanation methodology applied to business (B2B) sales forecasting process was introduced in Bohanec et al. (2017a) in the context of a decision-making framework with double-loop learning (Bohanec et al., 2017b). Analysis of the proposed organizational model and its implementation as a part of research design is discussed in this paper. Our hypothesis is that it is possible to use interpretable machine-learning models to support both single and double-loop learning, and at the same time foster acceptance by involving users. As a consequence sellers and companies make fewer mistakes in sales forecasting.

The proposed model will, through a combination of machine learning methods, knowledge and practice of experts, surpass the shortcomings of partial approaches and make the decision-making process transparent and consequently comprehensible to the participants. By making forecasting support systems models transparent, the users are encouraged to reflect not only on the outcomes but also on the reasons for the specific outcome. In this way individuals' and organizations' mental models are tested and the underlying model can improve (Grossler, Maier and Milling, 2000; Senge and Sterman, 1992), thus the gap between the forecasts and actual outcomes can be reduced. Furthermore, the machine learning model, built on a small set of features and supported by visualizations to support the reflection of a decision-makers, addresses important limitations of human decision-makers (Simon, 1991; Sterman, 1994) and states the requirements for successful learning. The feedback loop (FL; Kljajić Borštnar, Kljajić, Škraba, Kofjač and Rajkovič, 2011) and helps building trust (Fleischmann and Wallace, 2005).

The rest of the paper is organized as follows. Section

2 gives an overview of related work. Section 3 describes research approach, which is interconnected with companies' organizational process. In Section 4 we introduce an explanatory toy example. Analysis of an implementation in a company is presented in Section 5. Conclusions are put forward in Section 6.

2 Related work

Many techniques and solutions support forecasting of sales result - both for the business to consumer (B2C) and B2B segments. They can be grouped as quantitative (relying on data collected over longer period of time) and qualitative, based on judgment, intuition and informed opinions (and inherently subjective) (Davis and Mentzer, 2007; Kerkkänen and Huiskonen, 2007; Ingram, LaForge, Avila, Schwepker and Williams, 2012; Armstrong and Green, 2014). If a company has a large number of stored transactions, it is possible to use probability estimation techniques based on the development of opportunity, i.e. Sales funnel (Lodato, 2006; Duran, 2008; Söhnchen and Albers, 2010). Such an approach less applicable where there are fewer sales opportunities. In this case, the importance of sales forecasts for a company is all the more important as it could get either no sale closed or get all sales closed (even the unexpected cases). Additionally, the size of the opportunity also matters, since the company needs to allocate its resources (Duran, 2008).

A survey of the leading companies in various industries has shown that companies relying on data-driven decision-making (DDDM) achieve better results (Provost and Fawcett, 2013). On average, the top one-third DDDM companies from their industry are on average 5% more productive and 6% more profitable compared to their competition (Brynjolfsson, Hitt and Kim, 2011; McAfee and Brynjolfsson, 2012).

However, research on a development of sales forecasting (McCarthy, Davis, Golicic and Mentzer, 2006) showed that the knowledge of forecasting techniques from both categories, quantitative and qualitative, is declining. Similarly, the review of sales forecasting (Armstrong, Green and Graefe, 2015) showed that forecasting practice had seen little improvement, despite major advances in forecasting methods and development of sophisticated statistical procedures. Other researchers note a negligible positive effect of forecasting techniques, which is a result of decision makers' doubts about their reliability and comprehensibility (Lawrence, Goodwin, O'Connor and Önkäl, 2006). In practice, managements easily and quickly decide qualitatively, in its own discretion, based on experience, knowledge and mental models of individuals. The reasons for weak user acceptance are generally low trust in technology, doubts about the quality of data and the doubts about the benefit of such recommendation systems. This is especially true in domains where process data cannot be

easily measured, like for example in B2B business sales.

In contrast to the B2B domain, the B2C domain receives more attention in the academic literature (Lilien, 2016). In B2C, large amounts of data are generated from user behavior. In contrast, in B2B sales process data are interpreted and collected by the sales experts. Subjective interpretations of the process' features decrease confidence in data quality. On the other hand, it enables organizations to capture »soft« features of the sales process (i.e. expectations of the client about the offered solution), which is a good starting point for describing preferences and sales processes with qualitative attributes that describe the state of opportunity (Söhnchen and Albers, 2010; Monat, 2011).

B2B sales acquisition applications of ML have been discussed in (Yan et al., 2015; D'Haen and Van der Poel, 2013). Yan et al. (2015) explicated that ML methods outperform subjective judgment. Furthermore, when sellers were provided with scorings for their resource allocation decision, their results improved, indicating the regenerative effect between prediction and action. An iterative three-phased automated ML model for identifying promising clients (sales opportunities) in a B2B environment was proposed by D'Haen and Van der Poel (2013). They emphasize the importance of documenting the decisions made, steps taken, etc., to incrementally improve client acquisition. We address feedback issues with explanations of predictions. A comprehensive review of the literature on B2B sales leads explicates that there is little research, lack of rigor (theoretical grounding or validation) and no corroborative data (Monat, 2011) on this subject.

A review of 52 articles addressing the application of ML in Decision Support Systems (DSS) between 1993 and 2013, suggests that ML usefulness depends on the task, the phase of decision-making and applied technologies (Merkert, Mueller and Hubl, 2015). Furthermore, these researchers found that ML methods (i.e. support vector machines and neural networks) are used mostly in the first two phases of the decision-making process, intelligence and design, as described by Simon (1960), while the third phase, choice, is less supported. We address the identified gap by using ADR approach, which includes all three phases in an organizational context.

In many classification problems, users are concerned with more than predictive performance, and in decision support, the interpretability of prediction models is of great importance. In order to apply prediction models, users have to trust them first and models' transparency is a crucial step in ensuring the trust. As many comparative studies show, complex models, like random forests, boosting, and support vector machine, achieve significantly better predictive performance than simple interpretable models such as decision trees, Naive Bayes, or decision rules (Caruana and Niculescu-Mizil, 2006). Unfortunately, complex models are also difficult to interpret. This can be alleviated either by sacrificing some prediction perfor-

mance and selection of transparent model or by using an explanation method that improves the interpretability of complex models, like the general explanation methodology that can be applied to any classification or regression model (Robnik-Šikonja and Kononenko, 2008; Štrumbelj, Kononenko and Robnik-Šikonja, 2009).

The present paper builds on our previous work, where we proposed a novel use of general explanation methodology inside an intelligent system in a real-world case of B2B sales forecasting (Bohanec et al., 2017a). We first assembled a set of attributes from academic literature (Bohanec, Kljajić Borštnar and Robnik-Šikonja, 2015a), developed an optimization process to define the minimum sized data set (Bohanec, Kljajić Borštnar and Robnik-Šikonja, 2016b) and built machine learning model (Bohanec et al., 2017a). The ML model, enhanced with the general explanation methods, was applied in a real-world B2B process so that users could validate their assumptions with the presented explanations and test their hypotheses using the presented what-if parallel graph representation. The results demonstrate effectiveness and usability of the methodology. A significant advantage of the presented method is the possibility to evaluate seller's actions and to outline general recommendations for sales strategy. The results on the use of the framework were discussed in (Bohanec et al., 2017b). The results suggest that the provided ML model explanations efficiently support business decision makers, reduce forecasting error for new sales opportunities, and facilitate discussion about the context of opportunities in the sales team. In this paper we focus on the organizational context. We analyze the evidences of single and double-loop learning occurring in the process of model building and use.

3 Methodology

Our research idea is grounded in the Design Science Research paradigm (Hevner, March, Park and Ram, 2004), which deals with the development of IT artifacts. Due to weak user acceptance of the developed IT artifacts, the participatory research design was employed. Action Design Research (ADR) is a participatory design, which combines action research (Avison and Fitzgerald, 2006) with design science research (Hevner et al., 2004). Here users and researchers cooperate in the development of the solution in an organizational context/setting (Sein, Henfridsson, Purao, Rossi and Lindgren, 2011). By involving users, the acceptance of the developed IT artifact is addressed at the same time.

The ADR methodology has four stages, each supporting certain key principles as shown in Figure 1. In our context, ADR results in an organizational artifact, represented by the comprehensible explanations of top-performing black-box ML models supporting decision-making in B2B sales forecasting. The artifact is bound by the context of

the organization. Different organizations require re-conceptualization of learning from the specific solution (as presented in this paper) into knowledge needed to create other instances of solutions (i.e. B2B sales forecasting decision-making process in another organization).

The four ADR stages with their application as defined by Sein et al. (2011) are as follows:

1. Problem formulation, which is triggered by a problem perceived in practice or anticipated by researchers.
2. Building, intervention, and evaluation builds upon problem framing and theoretical premises from Stage 1.
3. Reflection and learning enable the move from the conceptual solution for a particular instance to a more general solution. This stage runs in parallel with Stages 1 and 2, recognizing that the research process involves more than problem-solving.
4. Formalization of learning formalizes the learning from the ADR project into general solution concepts for a class of field problems.

3.1 CRISP DM

Cross industry standard process for data mining (CRISP-DM) phases are presented in Figure 1. We shortly describe the methodology through its use in the context of our ADR based approach.

In the first phase, Business understanding, the team identified a problem, set up goals and expectations. In the second phase, data were exported from the existing customer relationship management system (CRM), followed by preliminary analysis and data-driven problem understanding. Since standard CRM attributes showed to have little predictive value, the ADR team selected 23 attributes describing the specific context of the participating company from the list of attributes identified in the literature review. In the Data preparation phase, the selected attributes were added to the companies' CRM. The Modeling phase is comprised of the model building based on the collected data (selected attributes) from sales history. Prediction models were built and discussed with the ADR team in order to identify outliers, check data quality (Bohanec et al., 2015b), and foster critical reflection on the predictions and thus sales forecasting process (Figure 2). This fostered user acceptance of the models and the whole process. We used several software packages to build and present ML models, e.g., Orange, WEKA, and R. Once the consensus on data quality and presentation format was achieved among ADR team, we compared different ML methods (random forest, naive Bayes classificatory, decision trees, artificial neural nets, and support vector machines). Besides the classification accuracy (CA) measure, we observed also the ROC curve (AUC measure). ML models explanations were generated by IME and EXPLAIN methods (Bohanec et al., 2017a).

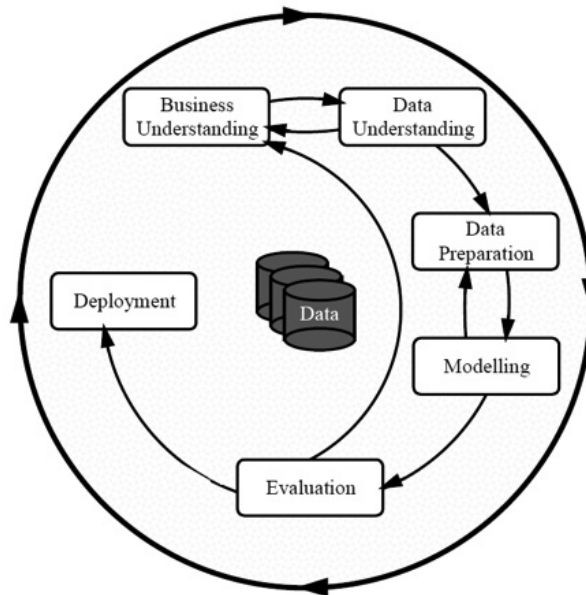


Figure 1: CRISP-DM methodology (Wirth and Hipp, 2000)

In the Evaluation phase, the final artifact was presented to the larger group of users (in the participating company) in a form of a workshop. The ML model predictions were interpreted on new sales opportunities. Users identified some erroneous data and the cycle was rerun.

Once the ADR team was satisfied with the model, the deployment phase followed. The model was used in the monthly forecasting process for several consecutive months. Every month the users and external consultant, using the ML model coupled with explanations, produced the sales outcome predictions. The users were presented with the results of ML model predictions and predictions of the consultant (using the ML model and explanations). This resulted in revised forecast of users. At the end of each month, the predictions and actual outcomes were analyzed.

Figure 2 presents the proposed research framework, which is grounded in ADR methodology and consists of several design cycles (following CRISP-DM methodology), in which business users together with researchers define the problem, build an ML model, use the model in an organizational decision-making process, and use new insights to update the model. Figure 2 depicts single loop learning (supported by the ML model), and double-loop learning (supported by the ML model, enhanced with explanations).

4 Introduction to general explanation methodology with examples

We use two general explanation methods IME and EXPLAIN (Robnik-Šikonja and Kononenko, 2008; Štrumbelj et al., 2009). These two methods explain model's predictions as contributions of individual attributes. The explanations are based on the structure of the model and visualize the context of individual opportunities.

In general, prediction explanations can be divided into two levels - domain level and model level explanations. The domain level explanations would show a true causality between dependent and independent variables, and can only be achieved for artificially constructed problems, where relations between variables and distribution probabilities of outcomes are known. In real world problems, these relations are not known and only the built model can be used to explain the causalities (model level explanations). The model level explanation transparently presents the prediction process with a particular ML model, which is trained from examples described by the attributes. The research on artificial problems shows that models with better predictive capacity allow better explanation (Štrumbelj et al., 2009).

From the model, we can get explanations of individual cases or the explanation of the whole model. The whole model explanations average the explanations of training set examples. They display the impact of the attributes as a whole as well the influence of individual values of attributes in the model. Since individual values of attributes affect different outcomes differently, each outcome

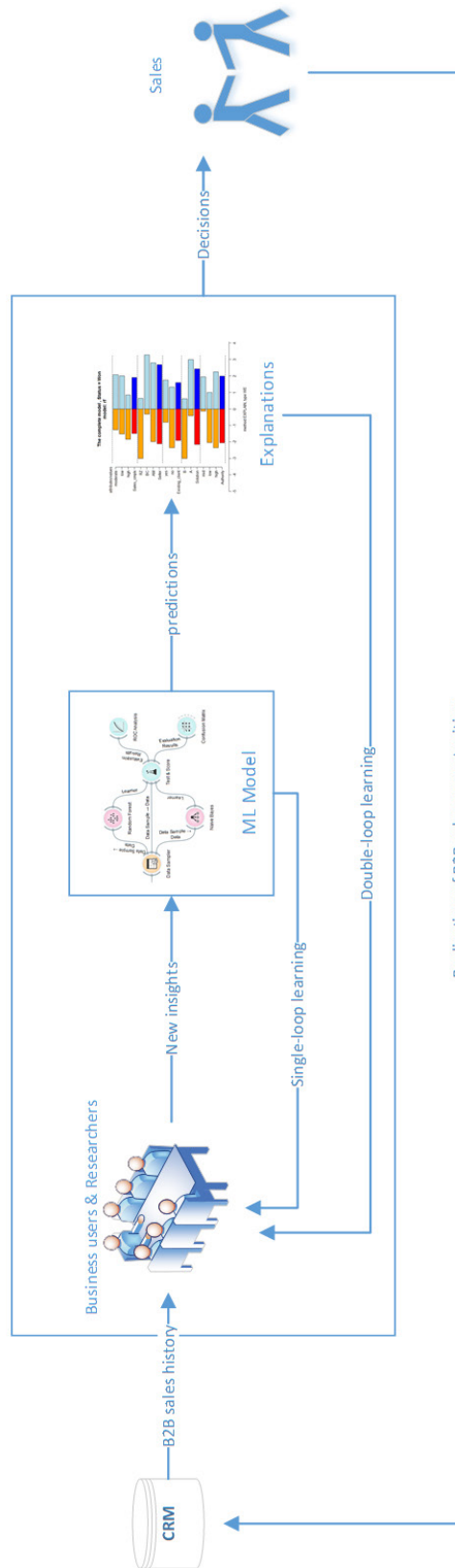


Figure 2: Proposed organizational learning based on ML model, enhanced with explanation methods

(i.e. class in ML terminology) has to be taken into account separately. The EXPLAIN method observes changes of one variable at a time. In case there are strong redundancies between the input variables (i.e. disjunctive relationship with the class), we can get unrealistic assessments of contributions. The IME method samples contributions of groups of input variables and thereby avoids the problem of redundancy but is more time-consuming. In practice, we test the performance of both methods and if they produce similar results, we use the faster EXPLAIN method, otherwise, we have to use the IME method.

We present a simple, toy example to introduce the use of the proposed explanation methodology applied to B2B sales problem. The data set contains a basic description of B2B sales events of a fictional company. The practice shows that simpler visualizations achieve better acceptance and foster trust at the beginning of learning.

We assume that our fictional company offers two complex solutions, A and B, on a B2B market. Their key customers are business managers and a certain level of sales complexity is expected (e.g., several units with competing priorities at client side). The company has grown on the success of their initial Solution A, however recently Solution B was added to the sales portfolio to open new sales opportunities. Preferably, the company offers the Solution B to existing clients, a practice called cross-selling. The sales personnel tries to pursue sales deals in which they can offer complex solutions together with the company's deployment consultants. Their previous experience shows that for a successful sale, the sales team should attempt to engage senior business leaders at prospects, with the authority to secure the budget and participate in the definition of requirements. The data for the fictional company is described in Table 1.

For example, the attribute *Authority* represents the authority level of a key contact at the prospect. It has three values with the following meaning: "high" (e.g., a person can secure the funding), "moderate" (e.g., a person influences the project specification, but lacks budget), and

"low" (e.g., a person just collects information and has no power to make important decisions). The toy example fictional data set consists of 100 instances. We randomly take 80 percent of the instances as a training set and the remaining 20 percent as a testing set. To build a classifier, we use the ensemble learning method Random Forest (RF) (Breiman, 2001). The RF model is passed as the input to the EXPLAIN or IME explanation methods. Figure 3 introduces an example of an explanation for a specific case (the sales opportunity named instance 14), where the sale was discussed with the high-level manager at a new prospect, the *Seller* AM offered the *Solution A* and was experiencing high complexity in the sales effort.

The left-hand side of Figure 3 outlines the attributes, with the specific values for the selected instance on the right-hand side. For this instance, the probability returned by the model for the outcome *Status* = "Won" is 0.74, and "Won" is the true outcome of this instance. The impact of attributes on the outcome is expressed as the weight of evidence (WE) and shown as horizontal bars. The length of the bars corresponds to the impact of the attribute values on the outcome predicted by the model. Right-hand bars show positive impacts on the selected outcome (*Status* = "Won" in this case, see the header of the figure), and left-hand sidebars correspond to negative impacts. The thinner bars (light gray bars) above the explanation bars (dark gray) indicate average impacts (obtained from training instances) for particular attribute values.

For the given instance 14 in Figure 3, we can observe that *Solution* = "A", *Authority* = "high" and *Sales_cplx* = "high" are in favor of closing the deal. The attribute *Existing_client* = "no" is not supportive of a positive outcome. For the attribute *Seller* with a value AM, there is no bar present, exposing the role of AM as completely neutral in the context of this instance. The thinner bars show that on average both positive and negative impacts of these values are observed, with different intensities. The average value for the attribute *Solution* with value A is the most biased towards a positive outcome. This is in-line with our toy

Table 1: Data for the fictional company (Bohanec et al., 2017a)

Attribute	Description	t
Authority	Authority level at a client side.	low (24), mid (37), high (39)
Solution	Which solution was offered?	A(51), B(49)
Existing_client	Selling to existing client?	no(47), yes(53)
Seller	Seller name (abbreviation).	RZ(35), BC(29), AM(36)
Sales_complexity	Complexity of sales process.	low(31), moderate(53), high(16)
Status	An outcome of sales opportunity.	lost(45), won(55)

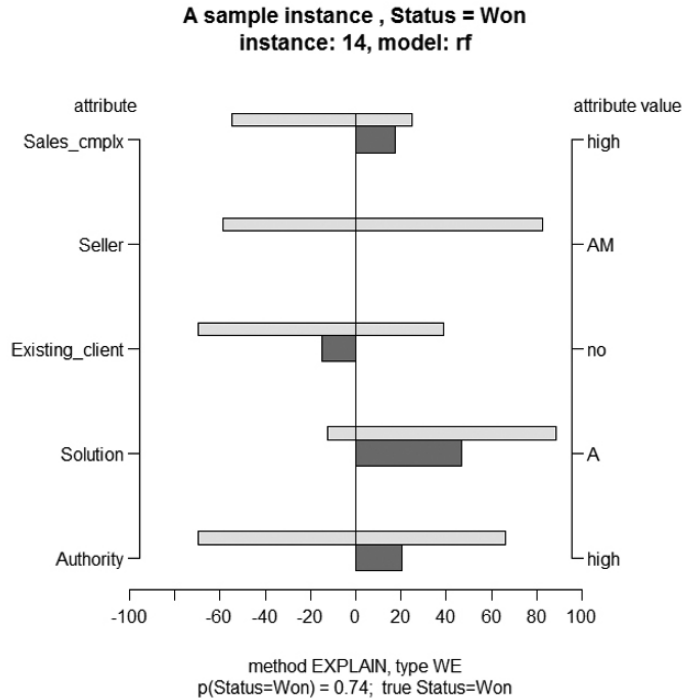


Figure 3: Explanation for a specific case (the sales opportunity named instance 14)

scenario where the Solution A is a flagship product selling very well.

To understand the problem on the level of the model, all explanations for the training data are combined. Visualization of the complete model showing all attributes and their values (separated by dashed lines) is shown in Figure 4.

From Figure 4 we can observe that the impact indicators for attributes (dark gray bars) spread across the horizontal axis, which indicates that both positive and negative impacts are possible for each attribute. Dark gray bars representing attributes are weighted averages of the impact of their values that are shown above them. For each attribute value (light gray bar), an average negative and positive impact are presented. Specific attribute values often contain more focused information than the whole attribute. For example, moderate sales complexity or dealing with mid-level managers indicate a stronger tendency towards positive outcome than towards negative outcome. The value “yes” for the attribute *Existing_client* has a prevailing positive impact on the positive outcome, but the value “no” can also have a positive impact. Note the scale of the horizontal axis in Figure 3 and 4. While on Figure 4 original values of WE are shown, we normalized the sum of contributions to 100 in Figure 3. Such normalization can be useful if we compare several decision models or if we want to assess the impact of attributes in terms of percentages.

As indicated in (Robnik-Šikonja and Kononenko,

2008), the method EXPLAIN is fast but does not capture the disjunctive and redundant interactions among attributes. To indicate these effects, Figure 5a utilizes the method IME (calculation takes somewhat longer) for a visualization of the whole model. A comparison with Figure 4 shows certain differences, therefore, we will subsequently use the IME method for the toy example. When users want to focus their analysis on a particular subset of attributes, they can single them out. For example, Figure 5b only presents attributes *Sales_cmplx* and *Authority* in the model view.

The explanation follows the underlying model; therefore if the model is wrong for a particular testing instance, the visualizations will reflect that. For example, Figure 6 shows the instance 9, where the RF model estimates that the probability of successful closure is 0.41 (which is less than the threshold 0.50, indicating the outcome “Lost”). However, this instance was actually “Won”.

In practice, sellers are interested in explanations of forecasts for new (open) cases, for which the outcome is still unknown. Figure 7a visualizes an explanation of such a case. The initially predicted probability of successful sale is 0.49.

The explanation reveals a positive influence of the fact that the sale is discussed with an existing client, also the impacts of attributes *Sales_complexity* and *Authority* are positive. The *Seller* RZ (thin bar indicates his low sales performance) has a negative impact. The fact that the Solu-

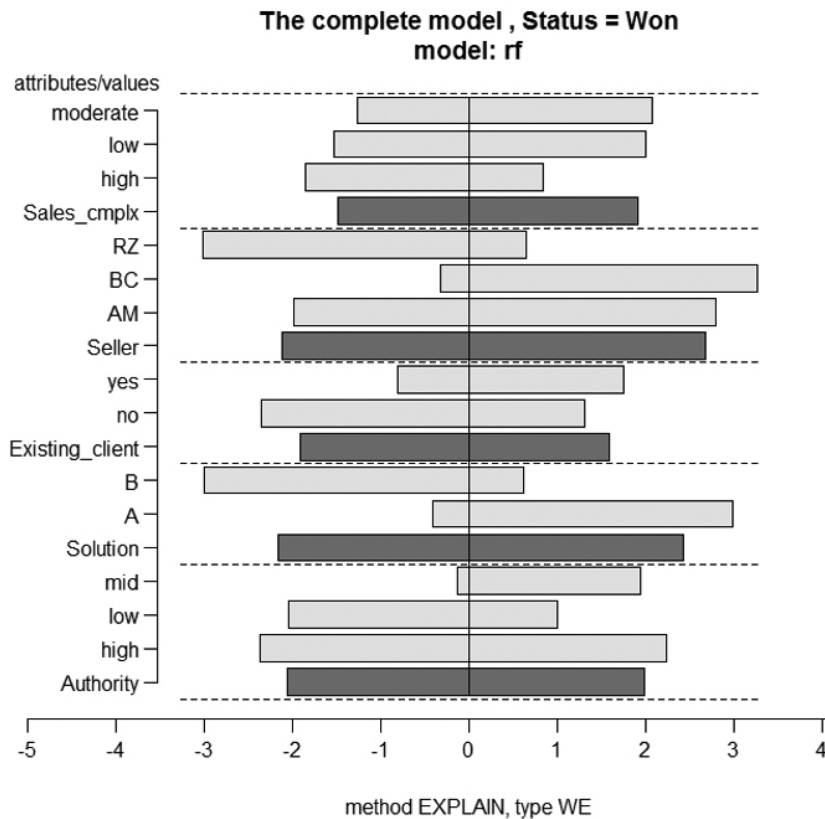


Figure 4: Visualization of the complete model, method EXPLAIN

tion B is offered shows a clear tendency toward the negative outcome.

As this sales opportunity is critical for the company, increasing the chances for a positive outcome is paramount. The low predicted probability triggers a discussion about the actions needed to enhance the likelihood of winning the contract. Unfortunately, not a lot of attributes can be influenced by the company as they are controlled by the prospect (e.g., *Authority* = "mid" cannot be changed). They consider the effect on the outcome a change of seller from "RZ" to "AM" would cause, with all other attribute values left the same. Figure 7b shows the implications of this change. The likelihood of winning the deal rises to 0.85. The explanation bars indicate strong positive influences of all attribute values but *Solution*, which follows our intuition given that, *Solution B* is a new offering.

By introduction of explanation methodology in a decision-making process, users are supported in transparent reasoning. This provides evidence for informed decisions and challenges prior assumptions.

5 Results and discussion

The process of model building and some examples of the model use on a real company data are described in this section. In contrast to toy example, described in Section 4, we wanted to explicate the complexity of the real life application. Further, we analyze the effects of implementation of the model in the company over the observed 16 months.

5.1 Features

A sample of the final set of attributes which is an input to the ML scheme for a real-world case study is presented in Table 1. The complete list is described in Bohanec et al (2017a, 2017b). A detailed description of attributes identification, analysis of attributes importance, and final attribute selection can be found in (Bohanec et al., 2015a, 2015b; Bohanec et al., 2016a, 2016b).

The data set was developed in the course of several months (Bohanec, 2016). It consists of 448 instances described by 22 attributes and one class attribute (the outcome of the sales opportunity: won, lost). There are 51% instances of the class »won«, and 49% with the class »Lost«. Data set is publicly available (Bohanec, 2016) to

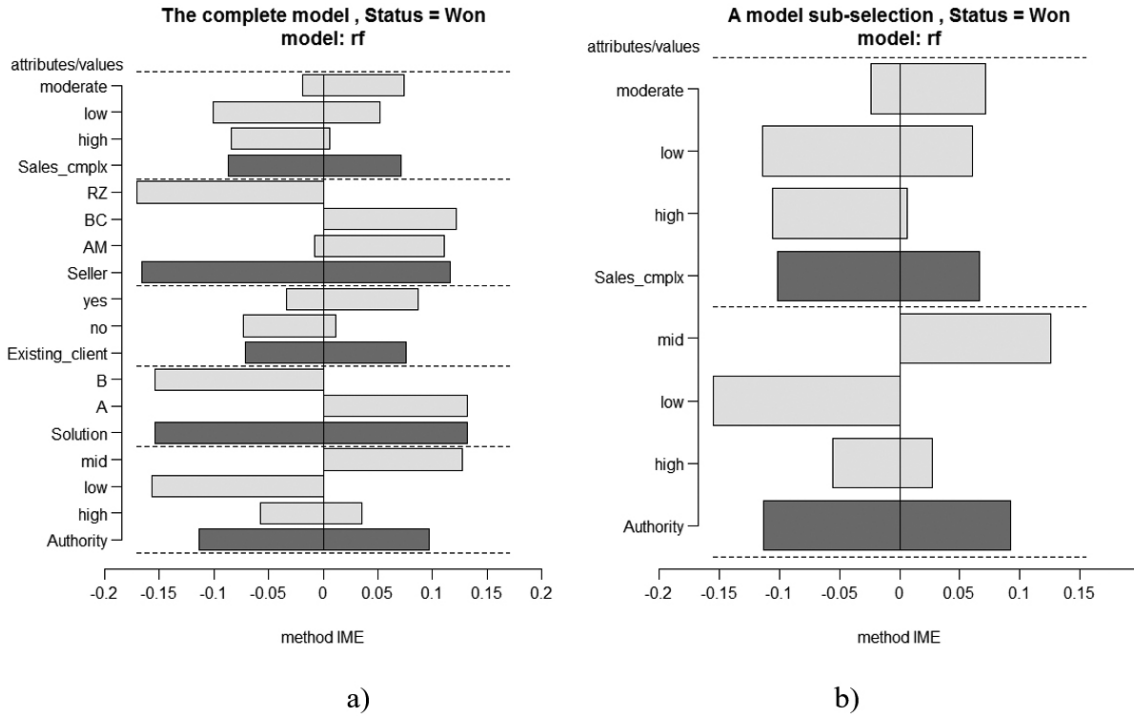


Figure 5: a) Method IME model explanations, b) drilling into attributes to visualize attributes of interest

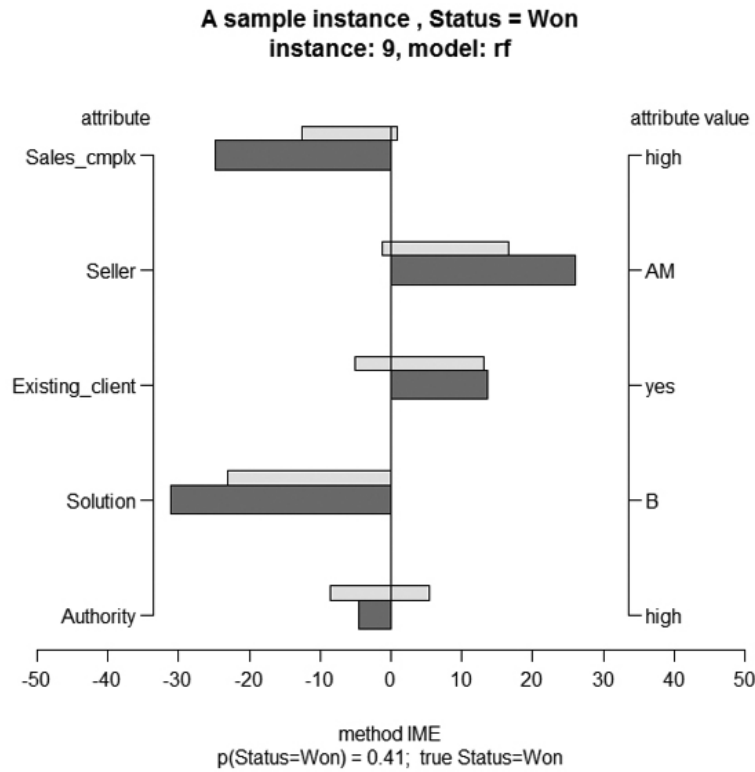


Figure 6: An example of the wrong prediction

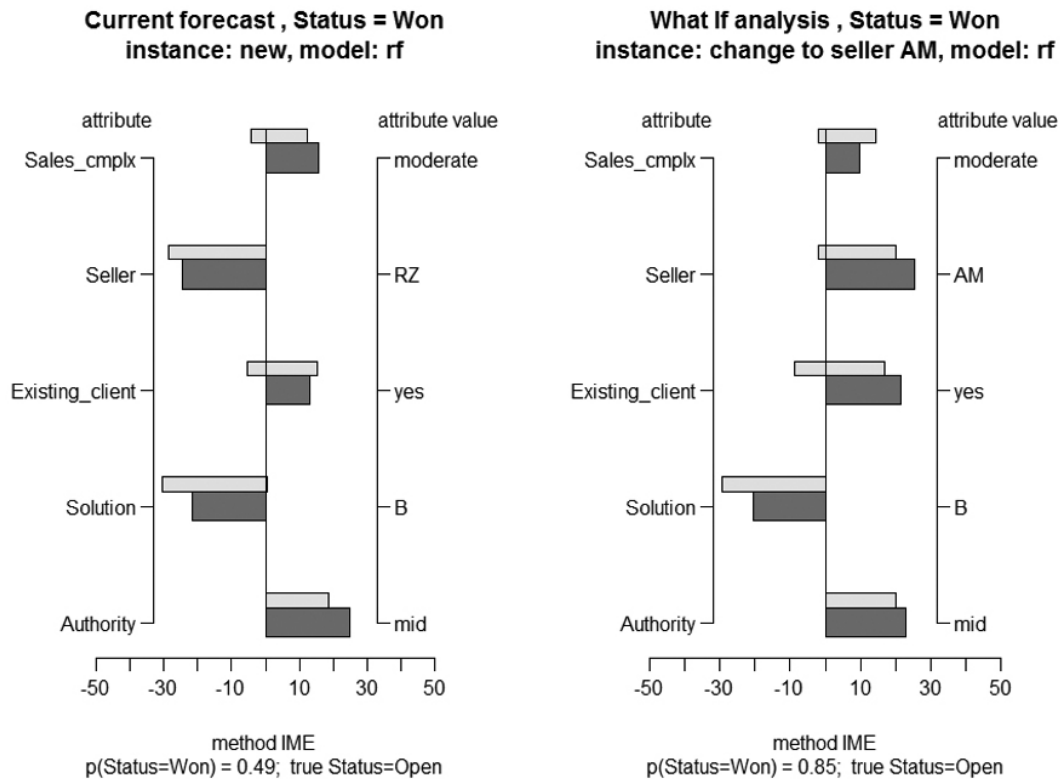


Figure 7: a) explanation of a prediction a new sales opportunity, and b) “what-if” analysis for the new sales opportunity

reduce the gap in the field of B2B research data (Lilien, 2016).

5.2 Building an ML model

The process of ML model building, presented in Figure 1, is in itself a circular process, where users (B2B sales experts) contribute input data for classification model building, examine the results of the model with explanation methods and evaluate the proposed decisions. In this process, new insights are generated and fed back into the system (the group uses new insights in the second cycle of model building, examines the updated model etc.). The process of model building and usage is contributing to better understanding of the model and B2B sales forecasting process, and thus to acceptance of the model.

Table 2 contains several classification algorithms and their performances, measured with the classification accuracy and ROC. The data set was divided into a training set (80% of cases used for training) and testing set (20% of cases used for evaluation). The process of classifier training is repeated 30 times to increase stability and reliability of the performance estimation. Average results of this experiment, along with standard deviation values, are presented in Table 2. We evaluated the performance of ran-

dom forest (RF), Naive Bayes (NB), decision tree (DT), neural nets (NN) and support machine vectors (SVM) method. The RF algorithm performed best in two different experiment settings (Bohanec et al., 2017a, 2017b).

5.3 Model use in the company

The presented approach enables explanation of predictions and other analyses, for example, »what-if« analysis and exploration of how individual attributes influence the possibilities of closing a deal with a new client (presented in Section 4 with the toy example).

An example of visualization of the model built on the data from a real world company is presented in Figure 8. Figure 8a shows all attributes with their values for a new sales opportunity. For more clear visualization, it is preferable to concentrate only on the most impactful attributes. Users can limit the threshold of attribute impact shown on the graph and in the same time test the implication of changes for a particular attribute(s) of interest. Figure 8b shows the predicted impact of getting confirmation about the budget (*Budget* = “Yes”) and formal purchase process (*Purchase_dept* = “Yes”). In Figure 8b, only the values higher than the threshold for $WE=1$ are shown. This improves graph readability and supports discussion focused

Table 1: Sample of the final list of attributes describing our B2B sales process

Attribute	Description	Values
Authority	Authority level at a client side.	low, mid, high
Product	Offered product.	e.g. A, B, C, etc.
Seller	Seller's name.	Seller's name.
Competitors	Do we have competitors?	no, yes, unknown
Client	Type of a client.	new, current, past
...
Attention to client	Attention to a client.	first deal, normal, etc.
Status	An outcome of sales opportunity.	lost, won

Table 2: Average results of 30 repetitions of classification models training (Bohanec et al., 2017a)

Classification method	CA average	CA std.dev.	AUC average	AUC std.dev.
RF	0.782	0.045	0.853	0.034
NB	0.777	0.036	0.835	0.040
DT	0.742	0.040	0.764	0.039
SVM	0.567	0.259	0.589	0.325
NN	0.702	0.051	0.702	0.051

on the most impactful attributes.

The participating company was interested in improving the efficiency of acquisition of new customers. In order to adapt the presented methodology, we adjusted the data set to contain only instances relevant to the context of the question. This means that only instances with the value of attribute *Client* = "New" were extracted from the database (in our case, there were 158 matching records). As shown in Figure 9, the company shall focus on attracting prospects to attend sales events and on opportunities to collaborate with other companies. The performance of different products varies significantly.

5.4 Analysis of the implementation of the model in the company

For several months, the participating company used the proposed solution in the forecasting process of B2B sales prediction. The process started at the beginning of each month after the management together with the sellers (users) predicted which opportunities would be successfully completed by the end of the month. The forecasts were recorded in the CRM system, thus creating baseline sales forecasts. At that moment, the data were forwarded to an external consultant (researcher) who applied ML model and generated predictions together with their explanations. Based on that, the external consultant prepared his sales

predictions. The predictions of "consultant + ML prediction" were passed back to the company together with explanations for each sales opportunity. The sellers were encouraged to reconsider their initial forecast. Challenged by nonmatching outcomes or large differences in predicted probabilities they sometimes changed their initial forecasts, which resulted in revised forecast.

The users took some time to embrace the proposed model into their regular monthly forecasting process. Figure 9 shows that in the last few months the users changed their initial forecasts based on the ML model predictions and surpassed the performance of "consultant + ML model predictions". This supports our initial hypothesis that users (domain experts) can use the proposed model (ML model predictions with explanations) to reflect upon the B2B sales forecasting process and learn from it on an individual, as well as on the group level. Double loop learning is explicated by the revision of the users' beliefs and mental models. In this way, the process of model building and usage is contributing to improved understanding of the model and the B2B sales forecasting process and thus acceptance of the model. Furthermore, they can identify slippage of a sales opportunity, which is an opportunity that will not be closed within that month but will slip into the following months. The current ML model cannot predict these slippages.

We calculated the accuracy of the forecasts. Figure 9

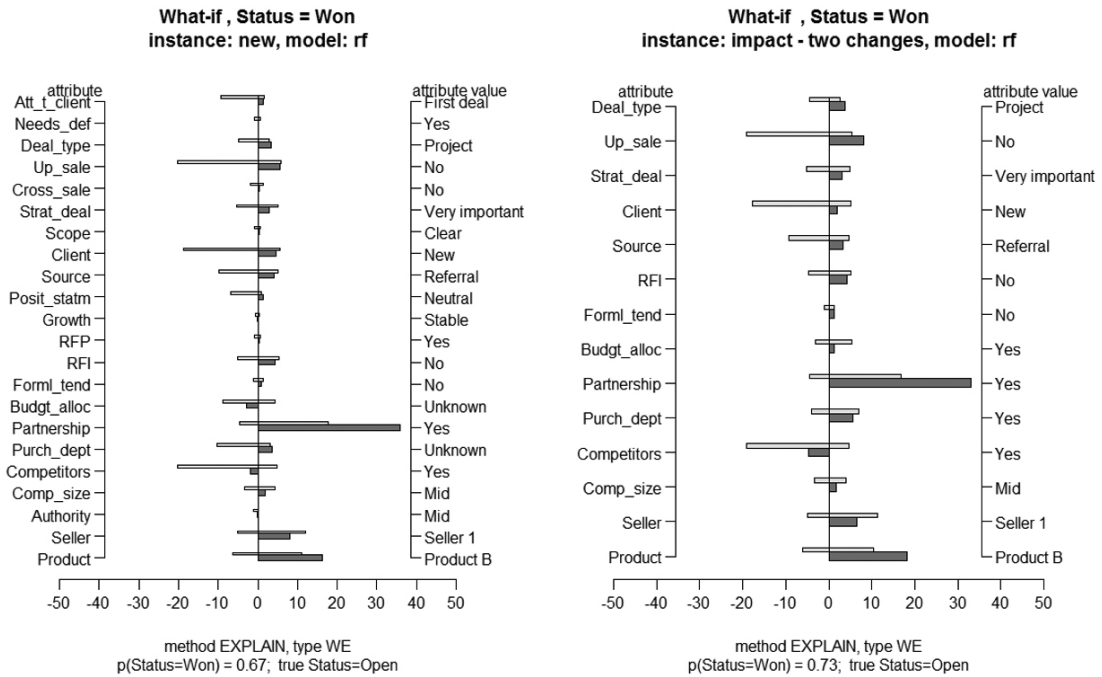


Figure 8: a) All attributes with their values for a new sales opportunity, b) analyzing impact after changing certain values of attributes (using the threshold of 1 to reduce complexity of visualization)

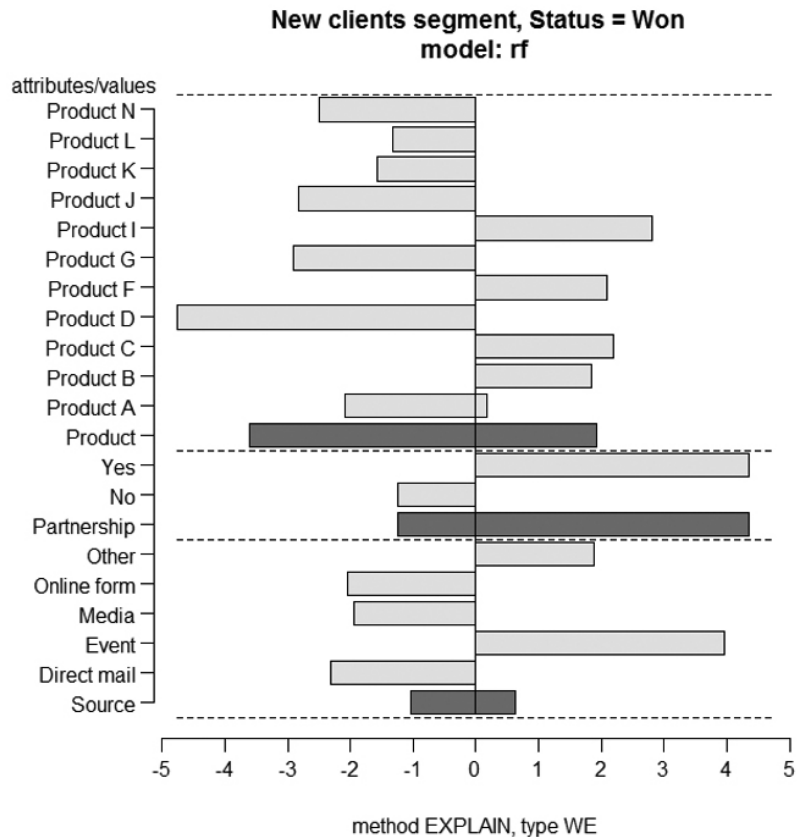


Fig 9: Drill into analysis for a specific business question

shows the results for a period between March 2016 and June 2017. The lines in Figure 9 represent the accuracy for the following types of forecasts:

- The initial forecast (double line) - the accuracy of the sellers at the beginning of the month.
- Consultant + ML model prediction (dot with dashed line) - The accuracy of the external consultant using the model and its explanations.
- An updated forecast (dashed line) - the performance of company using feedback with explanations of individual forecasts.
- Time lag (dotted line) – the percentage of slippages, (opportunities that were not decided in the current month) i.e. the opportunity has shifted.

As an example, we take the month March 2017. Only 17% of initial predictions were correct at the end of the month. An external consultant who took into account the ML model predictions along with the explanations achieved far better precision (43%). After the company reviewed the forecasts of the external consultant and explanation according to the developed ML model, their prediction accuracy rose to 65% for the observed month.

The implementation of the model allows for higher accuracy of the forecast compared to the company's baseline forecasts. By comparing the initial predictions with the updated ones (based on the proposed model), the users can recognize overly optimistic predictions, which are not supported by the data, and review their understanding. In all months, a large percentage of delayed opportunities exist (shown by the dotted line), reflecting too optimistic initial predictions of the users at the beginning of a month. In the future, this problem could be addressed by including additional attributes that will reflect too optimistic forecasting

opportunities.

For the most of the observed months in Figure 9, the revised company forecasts (dashed line) are outperforming the external consultants' forecasts (dot with dashed line). This is in line with the intuition that an internal team in the company can better evaluate the context of opportunities than the external consultant. We notice a significant change in behavior as a consequence of individual/group exposure to a specific experience (Kljajić Borštnar et al., 2011). The results show that double-loop learning helps to establish new mental models (which are reflected in revised forecasts) and repeal existing ones (changes in initial forecasts). This is confirmed by improvements in prediction accuracies of the revised forecasts compared to the initial ones.

6 Conclusions

In this paper, we addressed the problem of weak performance in judgmental B2B forecasting. We chose the action design research approach to develop an ML model, coupled by general explanation methods (IME, EXPLAIN), and introduce it into the organizational process. In this way, we involved users in all stages of model development, testing, and use.

In the proposed process, we first identified the problem and described it with a minimum set of features (attributes). Since the existing data from the company CRM were of little value, this phase consumed a lot of time and effort. When we agreed upon the attributes, the company started to collect the data and we built the data set and used it to train the ML prediction model. Based on the CA and AUC performance measures, we selected Random forest as the most appropriate method. Users reported their B2B sales

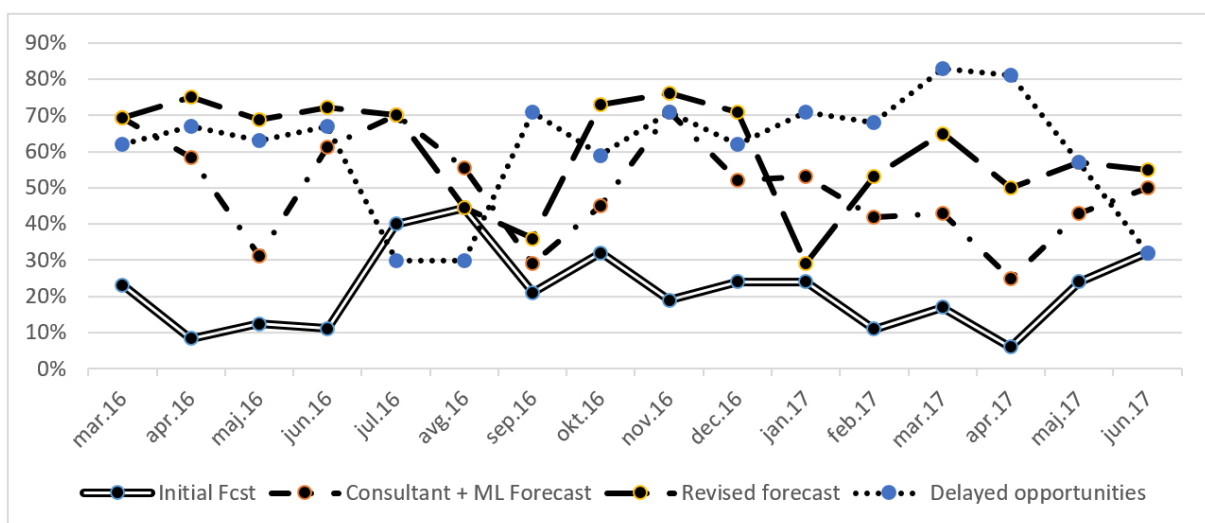


Figure 10: Comparison of prediction accuracy of the users, ML model + consultant and users + ML model

forecasts monthly and compared them to the ML model forecasts and the forecasts made by the external consultant, using ML model with explanations for every sales opportunity. We evaluated the forecasts to actual outcomes the next month and repeated the activities.

In the observed period, it was evident that the users were too optimistic in predicting outcomes of sales opportunities. Interestingly, the external consultant, using the ML model predictions with explanations, frequently achieved better results compared to users' initial forecasts. In most months, the revised forecasts of the users outperformed the forecasts of the ML model and of the external consultant, reflecting their better understanding of the total context of their business. This is in line with the intuition that internal teams in a company can better evaluate the context of opportunities compared to external consultants. The double loop learning is explicated by the revision of users' beliefs and mental models. We recognize a significant change in behavior due to individual/group exposure to a specific experience (Kljajić Borštnar et al., 2011). The results show that double-loop learning helps to establish new mental models (which are reflected in revised forecasts) and repeal existing ones (change in initial forecast). This is also corroborated by the data since the accuracy of the revised forecasts is better than the initial ones (Figure 9).

There are several limitations to this approach. First, the ML model is built upon data, collected by the users, and it reflects users' misperceptions (as evident from Figure 9). We addressed this problem to some extent by standardizing understanding of each attribute and its values. Furthermore, it is evident that the users are over-optimistic in predicting sales outcomes, resulting in several slippages. When confronted with ML model predictions, coupled with explanations, the users can identify slippage of a sales opportunity (the opportunity that will not be closed within a current month but will slip into the following months). The ML model in the current state cannot predict slippages.

Business environments are changing fast, which affects changing of modeled concepts (known as concept drift). It is important for the users to continuously reflect upon the predictions and their explanations in order to detect those changes and identify the need for additional attributes to be taken into account.

Finally, we recommend to actively support users in the phases of selection of attributes, the definition of their values, and implementation of data collection in the organizations' information system. It is important to regularly re-evaluate the values describing open sales opportunities. This can contribute to reduced noise in the data, improved accuracy of models, and builds trust in the model. An important lesson of this research is that neither ML models nor human decision-makers alone can successfully address the problem of B2B sales predictions. However, human

decision-makers supported by the ML models enhanced by explanations can surpass the limitations of human rationality.

Acknowledgement

We are grateful to the company Salvirt, Ltd., for funding the research and development presented in this paper. Mirjana Kljajić Borštnar and Marko Robnik-Šikonja were supported by the Slovenian Research Agency, ARRS, through research programs P5-0018 and P2-0209, respectively.

Literature

- Argyris, C. & Schön, D. (1996). *Organizational Learning II: Theory, Method and Practice*. Addison Wesley.
- Armstrong, J. S., Green, K. C. & Graefe, A. (2015). Golden Rule of Forecasting: Be conservative. *Journal of Business Research*, 68 (8), 1717-1731, <http://dx.doi.org/10.1016/j.jbusres.2015.03.031>
- Avison, D., & Fitzgerald, G. (2006). Methodologies for Developing Information Systems : A Historical Perspective. *The Past and Future of Information Systems*, 27–38, https://doi.org/10.1007/978-0-387-34732-5_3
- Bohanec, M. (2016). *Anonimized data set for B2B sales history*. Retrieved 15.07.2017 from <http://www.salvirt.com/research/b2bdataset>
- Bohanec, M., Kljajić Borštnar, M. & Robnik-Šikonja, M. (2015a). Feature subset selection for B2B sales forecasting. In: *13th International Symposium on Operational Research*, Bled, Slovenia, 285-290.
- Bohanec, M., Kljajić Borštnar, M. & Robnik-Šikonja, M. (2015b). Machine learning data set analysis with visual simulation. In: *InterSymp 2015*, Baden-Baden, Germany, 16-20.
- Bohanec, M., Kljajić Borštnar, M. & Robnik-Šikonja, M. (2016a). Nabor atributov za opisovanje medorganizacijske prodaje [A collection of attributes describing business to business (B2B) sales]. *Uporabna informatika*, XXIV (2), 74-80.
- Bohanec, M., Kljajić Borštnar, M. & Robnik-Šikonja, M. (2016b). Sample size for identification of important attributes in B2B sales. In: *16th International Conference on Operational Research*, Osijek, Croatia, 133.
- Bohanec, M., Kljajić Borštnar, M. & Robnik-Šikonja, M. (2017a). Explaining Machine Learning Predictions. *Expert Systems with Applications*, 71, 416-428.
- Bohanec, M., Robnik-Šikonja, M., & Kljajić Borštnar, M. (2017). Decision-making framework with double-loop learning through interpretable black-box machine learning models. *Industrial Management & Data Systems*, 117(7), in print (July, 2017b), <http://dx.doi.org/10.1108/IMDS-09-2016-0409>
- Breiman L. (2001), Random Forests, *Machine Learning*

Journal, 45, 5–32.

- Brynjolfsson, E., Hitt, L. M. & Kim, H. H. (2011). *Strength in numbers: How does data-driven decision-making affect firm performance?* Retrieved 11 May 2017 from http://ebusiness.mit.edu/research/papers/2011.12_Brynjolfsson_Hitt_Kim_Strength%20in%20Numbers_302.pdf
- Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 161–168). New York, NY, USA: ACM.
- Davis, D. F. & Mentzer, J. T. (2007). Organizational factors in sales forecasting management. *International Journal of Forecasting*, 23 (3), 475–495, <http://dx.doi.org/10.1016/j.ijforecast.2007.02.005>
- D’Haen, J., & Van der Poel, D. (2013), Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework, *Industrial Marketing Management* 42, 544–551, <http://dx.doi.org/10.1016/j.indmarman.2013.03.006>
- Duran, R. E. (2008). Probabilistic sales forecasting for small and medium-size business operations. In B. Prasad (Ed.), *Soft Computing Applications in Business*, 129–146, Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-540-79005-1_8
- Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, 42(1), 351–361, <https://doi.org/10.1016/j.dss.2005.01.003>
- Fleischmann, K. R., & Wallace, W. A. (2005). A covenant with transparency. *Communications of the ACM*, 48(5), 93–97, <https://doi.org/10.1145/1060710.1060715>
- Goodwin, P., Fildes, R., Lawrence, M., & Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega*, 39(3), 242–253, <https://doi.org/10.1016/j.omega.2010.07.001>
- Grossler, A., Maier, F. H., & Milling, P. M. (2000). Enhancing Learning Capabilities by Providing Transparency in Business Simulators. *Simulation & Gaming*, 31(2), 257–278, <https://doi.org/10.1177/104687810003100209>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Huber, G.P. (1991). Organizational Learning: The Contributing Processes and the Literatures. *Organization Science*, 2(1), Special Issue: Organizational Learning: Papers in Honor of (and by) James G. March, 88–115, <https://doi.org/10.1287/orsc.2.1.88>
- Ingram, T. N., LaForge, R. W., Avila, R. A., Schwepker Jr, C. H. & Williams, M. R. (2012). *Sales Management: Analysis and Decision Making*. ME Sharpe.
- Kerkkänen, A. & Huiskonen, J. (2007). Analysing inaccurate judgmental sales forecasts. *European J. Industrial Engineering*, 1 (4), 355–369, <https://doi.org/10.1504/EJIE.2007.015387>
- Kljajić, M., & Farr, J. V. (2010). The Role of Systems Engineering in the Development of Information Systems, In M. Hunter (ed.), *Strategic Information Systems: Concepts, methodologies, tools, and applications*, Hershey, PA, Information Science Reference. 369–381.
- Kljajić Borštnar, M., Kljajić, M., Škraba, A., Kofjač, D., & Rajković, V. (2011). The relevance of facilitation in group decision making supported by a simulation model. *System Dynamics Review*, 27(3), 270–293, <https://doi.org/10.1002/sdr.460>
- Kuchinke, K. P. (2000). The role of feedback in management training settings. *Human Resource Development Quarterly*, 11 (4), 381–401, [http://dx.doi.org/10.1002/1532-1096\(200024\)11:4%3C381::AID-HRD-Q5%3E3.0.CO;2-3](http://dx.doi.org/10.1002/1532-1096(200024)11:4%3C381::AID-HRD-Q5%3E3.0.CO;2-3)
- Lawrence, M., Goodwin, P., O’Connor, M. & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22 (3), 493–518, <http://dx.doi.org/10.1016/j.ijforecast.2006.03.007>
- Lilien, G. L. (2016). The B2B knowledge gap. *International Journal of Research in Marketing*, 33, 543–556, <http://dx.doi.org/10.1016/j.ijresmar.2016.01.003>
- Lodato, M. W. (2006). *Integrated sales process management: A methodology for improving sales effectiveness in the 21st century*. AuthorHouse.
- McAfee, A. & Brynjolfsson, E. (2012). Big data. The management revolution. *Harvard Business Review*, 90 (10), 61–67.
- McCarthy, T. M., Davis, D. F., Golicic, S. L. & Mentzer, J. T. (2006). The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, 25 (5), 303–324, <http://dx.doi.org/10.1002/for.989>
- Merkert, J., Mueller, M., & Hubl, M. (2015), *A Survey of the Application of Machine Learning in Decision Support Systems*, 23rd European Conference on Information Systems (ECIS) 2015 Completed Research Papers, Münster, Germany, 26.-29.05.2015.
- Monat, J. (2011). Industrial sales lead conversion modeling. *Marketing Intelligence and Planning*, 29, 178–194, <http://dx.doi.org/10.1108/02634501111117610>
- Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1 (1), 51–59, <http://dx.doi.org/10.1089/big.2013.1508>
- Robnik-Šikonja, M. & Kononenko, I. (2008). Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering*, 20 (5), 589–600, <http://dx.doi.org/10.1109/TKDE.2007.190734>
- Sein, M., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *Manage-*

- ment Information Systems Quarterly*, 35(1). Retrieved 12.01.2017 from <http://aisel.aisnet.org/misq/vol35/iss1/5>
- Senge P.M & Sterman J.D (1992). Systems thinking and organizational learning—acting locally and thinking globally in the organization of the future. *European Journal of Operational Research*, 59(1), 137–150, [http://dx.doi.org/10.1016/0377-2217\(92\)90011-W](http://dx.doi.org/10.1016/0377-2217(92)90011-W)
- Simon, H. (1960). *The new science of management decision*. Prentice-Hall.
- Simon, H. A. (1991). Organizations and Markets. *Journal of Economic Perspectives*, 5(2), 25–44. Retrieved 31.7.2017 from <http://people.ds.cam.ac.uk/mb65/mst-ir/documents/simon-1991.pdf>
- Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, 10(2–3 (Special Issue: Systems thinkers, systems thinking)), 291–330. <https://doi.org/10.1002/sdr.4260100214>
- Söhnchen, F. & Albers, S. (2010). Pipeline management for the acquisition of industrial projects. *Industrial Marketing Management*, 39 (8), 1356-1364.
- Škraba A, Kljajić M, & Kljajić Borštnar M. (2007). The role of information feedback in the management group decision-making process applying system dynamics models. *Group Decision and Negotiation*, 16, 77–95.
- Štrumbelj, E., Kononenko, I. & Robnik-Šikonja, M. (2009). Explaining instance classifications with interactions of subsets of feature values. *Data & Knowledge Engineering*, 68(10), 886–904, <https://doi.org/10.1016/j.datak.2009.01.004>
- Verikas, A., Gelzinis, A., & Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests, *Pattern Recognition*, 44(2), 330–349, <http://dx.doi.org/10.1016/j.patcog.2010.08.011>
- Wirth, R., & Hipp, J. (n.d.). CRISP-DM: Towards a Standard Process Model for Data Mining. Retrieved 12.2.2017 from <https://www.scribd.com/document/76859286/CRISP-DM-Towards-a-Standard-Process-Model-for-Data>
- Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C., Huang, J., Chu, S., & Yang, X. (2015), “On machine learning towards predictive sales pipeline analytics”, In: *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 1945–1951.
-
- Marko Bohanec** received his Ph.D. in Management Information Systems from the University of Maribor. He received his MSc from Faculty of Economy and BSc from Faculty of Computer and Information Science, both University of Ljubljana, Slovenia. Professionally, he supports companies to minimize risks in their sales performance and staff development. His research interests relate to improvements in business management by utilizing advancements in the field of machine learning.
-
- Mirjana Kljajić Borštnar** received her Ph.D. in Management Information Systems from the University of Maribor. She is an Assistant Professor at the Faculty of Organizational Sciences, University of Maribor and is a member of Laboratory for Decision Processes and Knowledge-Based Systems. Her research work covers decision support systems, multi-criteria decision-making, system dynamics, data mining, and organizational learning. She is the author and co-author of several scientific articles published in recognized international journals and conferences, including *Group Decision and Negotiation* and *System Dynamics Review*.
-
- Marko Robnik-Šikonja** received his Ph.D. in computer science and informatics in 2001 from the University of Ljubljana. He is an Associate Professor at the University of Ljubljana, Faculty of Computer and Information Science. His research interests include machine learning, data and text mining, knowledge discovery, cognitive modeling, and their practical applications. He is a (co)author of more than 90 publications in scientific journals and international conferences and three open-source analytic tools.