Design Principles for Explainable Sales Win-Propensity Prediction Systems

Tiemo Thiess[1], Oliver Müller[2], Lorenzo Tonelli[3]

[1] IT University of Copenhagen, Business IT, Copenhagen, Denmark; [2] Paderborn University, Data Analytics, Paderborn, Germany; [3] MAN Energy Solutions, Aftersales, Copenhagen, Denmark

tith@itu.dk, oliver.mueller@uni-paderborn.de, tonelli.lorenzo@gmail.com

Abstract: MAN Energy Solutions, one of the largest ship engine manufacturers in the world, is looking into further improving its hit rate of through-life engineering services and spare parts quotations. We help to solve this relevant field problem by building a novel machine learning based sales win-propensity prediction system that utilizes the lightGBM algorithm, SHapley Additive exPlanations, and a second layer conditional probability model of quotation age. Moreover, we build an implementation method for the broader class of such systems and extend the scientific literature on explainable machine learning by abductively developing and instantiating the design principles (DPs) of local contrastive explainability, global explainability, selective visualization, causality, confirmatory nudging, and accountability in a sales win-propensity system.

Keywords: Machine Learning, Explainability, Sales, Maritime Industry, ADR

1 Introduction

In the last years, shipbuilders and original equipment manufacturers (OEM) in the maritime industry have suffered from a significant drop in the demand for new-building of vessels and engines [1]. An ongoing oversupply of tankers and containerships in the market caused this drop. OEMs are especially challenged to rethink their traditional business models and to shift the focus in product lifecycle management from the product development phase (beginning-of-life) to the product usage phase (middle-of-life). In the approximately 15-20 years lasting usage phase of main engines, OEMs can generate earnings via spare parts sales and through-life engineering services (TES), such as maintenance, repair, and overhaul. For OEMs, the product usage phase of their installed equipment determines the aftersales market.

In this context, MAN Energy Solutions, one of the largest ship engine manufacturers in the world with high market shares in the tanker and container vessel segments and approximately 15.000 employees in over 100 destinations around the

15th International Conference on Wirtschaftsinformatik,
March 08-11, 2020, Potsdam, Germany

https://doi.org/10.30844/wi_2020_c8-thiess

world, is looking into further improving its hit rate[1] of through-life engineering services and spare parts quotations. Following the dual mission of IS, we help to solve this relevant field problem by building a novel sales quotation win-propensity[2] prediction system, while extending the scientific literature [2] on explainable machine learning by abductively developing design principles (DPs) based on a sound literature review and an authentic and concurrent evaluation of the action design research (ADR) process [3].

Win-propensity estimation is an important aspect of assessing overall sales performance [4]. Despite its importance, research on sales win-propensity estimation models is scarce [5]. In large firms such as MAN Energy Solutions, sales professionals sometimes have to deal with many open sales opportunities and quotations. To structure their work and to enable an approximate forecast of the win-propensity, sales professionals use CRM systems that enable them to assign win-propensity scores or hot-warm-cold labels manually as an outcome of an often more or less subjective judgment [6]. Such subjective judgments are prone to cognitive biases [7], such as being overly confident and thus estimating too high win-propensity scores [8]. Moreover, they can be biased due to organizational structures, politics, and socio-cultural phenomena, for example, when the management expects a positive forecast for the current sales pipeline [9]. Data-driven sales win-propensity estimation methods, on the other hand, can support resource management [10], increase efficiency, and generate explanatory insights about the sales process and its drivers [11].

Overall, in this paper, we make four scientific contributions. First, we push the state-of-the-art in sales propensity modeling by developing an approach combining ensemble machine learning techniques (esp. lightGBM) to robustly model non-linear relationships and interaction effects with a conditional probability model accounting for quotation age (Sections 4.1 and 4.2). Second, we demonstrate how methods for the human-interpretable explanation of black-box machine learning models (esp. SHapley Additive exPlanations) can be applied to improve the acceptance of predictions by users and managers, and how they help data scientists to improve model quality (Section 4.3). Third, we go beyond the pragmatic design of a single prototype and propose a method for the organizational implementation of the proposed approach in complex real-life settings (Section 4.4). Fourth, we formalize the learnings from this 1.5 years lasting action design research project as design principles for explainable aftersales win-propensity prediction systems (Section 5).

2 Explainable Machine Learning

Machine learning and data science have the ultimate goal of supporting decision making. Common sense tells us that one should only implement good decisions. But what are good decisions? Sharma et al. [12] present two characteristics of good decisions: quality and acceptance. The quality criterion is concerned with whether a

[1] At MAN hit rate is essentially calculated as orders euro / quotations euro (ex-post)
[2] Win-propensity is the hit rate expressed as a probability for a particular quotation (ex-ante)

https://doi.org/10.30844/wi_2020_c8-thiess

decision is able to reach its stated goals. The other criterion refers to whether a decision is accepted by its stakeholders, especially those responsible for successfully implementing it [13–15]. Hollander et al. [14] argue that how much stakeholders participate in the decision making process and, thus, influence the final decision, significantly impacts its acceptance and the chances of successful implementation.

Also, Kayande et al. [16] suggest that a lack of understanding of a machine learning model can lead to a refusal of acceptance and, consequently, usage by end-users, despite the fact that the model might improve decision quality. They further elaborate on this idea by proposing a three-gap framework that conceptualizes how human-interpretable explanations can be used to improve the acceptance and performance of decision support systems (DSS). In particular, they relate three different concepts, namely, the manager´s mental model, the DSS, and the true model (reality) via three distinct bi-directional gaps. The first gap, between the manager´s mental model and the DSS, can lead to reduced model acceptance when widened and improved model acceptance when narrowed. The second gap, between the DSS and the true model (reality), affects the performance of a DSS negatively when widened and positively when narrowed. The third gap, between the manager´s mental model and the true model (reality), affects the manager´s decision making performance negatively when widened and positively when narrowed.

Gregor and Benbasat [17] give arguments for why users need explanations when working with intelligent systems such as machine learning systems, namely, to solve specific problems by using the system, to learn from the system and its outputs, and to understand why anomalies have come to be. Moreover, they argue that explanations can lead to an improvement in terms of performance, learning, and the overall perception of a system. However, they also note that in order to enable such improvements, explanations should be context-specific rather than too generic and should not demand too much effort from users and, thus, if possible, be automated. Finally, they stress the importance of justificatory knowledge, which can lead to a deeper understanding by grounding, for instance, a prediction in sound causal theory.

Martens and Provost [18] have extended both the work of Kayande et al. and Gregor and Benbasat. They criticize the three-gap framework of Kayande et al. because it assumes that DSS are always superior to a manager's mental model in terms of decision quality (alignment with reality). Instead, they argue that DSS can be wrong too, for example, because of biases introduced during the model building process or overfitting a model to training data. In consequence, they extend the three-gap framework by adding a feedback loop for situations in which a manager´s mental model is closer to the true model (reality) than the DSS. The objective of this feedback loop is to improve the DSS by bringing it closer to the manager´s mental model. Moreover, they add the three different roles of developers, managers, and customers to aid understanding of how the explanatory needs of the roles differ.

Furthermore, Martens and Provost [18] extend the above-outlined arguments of Gregor and Benbasat by distinguishing between (1) explanations that lead to improved system acceptance by supporting the user in getting a causal understanding of the general real-world mechanisms that the system builds upon and (2) explanations that lead to improved acceptance by supporting the user in understanding

https://doi.org/10.30844/wi_2020_c8-thiess

how the particular system works. They further subdivide explanations of type (2) into (a) global explanations of how the overall model behaves and (b) local explanations of how it behaved in a particular instance. Such types of explanations, they argue, can lead to improved acceptance but also an improved model, which, again, can improve model acceptance but also aid in making sense of the model´s underlying causal mechanisms (reality).

3 Methodology

We followed an action design research (ADR) process inspired by Sein et al. [3], in which we started by analyzing and formulating the field problem of aftersales hit rate improvement at MAN Energy Solutions. Next, we designed initial artifacts of the class of explainable win-propensity prediction systems. Throughout many iterations of building, intervention, and evaluation, the artifacts were further shaped and refined by the design team, but also by the specific context of the maritime industry, until they reached their current state. Finally, we formalized abstracted learnings as design principles for explainable win-propensity prediction systems. During the whole process, we collected rich empirical data in the form of observation notes of our encounters at MAN Energy Solutions (see Table 1 for an overview). To collect the data, we used a form of design ethnography [19], in which one does not only study others and their behavior, but also oneself and one's artifacts, and how they interact as interventions with their environment.

Table 1: Project-related encounters at MAN Energy Solutions

Meeting Type	*Participants*	*#*	*h*	*Total (h)*
Development Meeting	Business Analyst, Junior Data Analyst, Data Management Specialist, Researcher	20	2	40
Stakeholder Presentation	Business Analyst, Department Manager, Sen. Strategy Manager, Strategy Manager, Researcher, Pricing Analyst	4	1	4
Sprints	Researcher, Business Analyst	18	4	72

4 An Explainable After-Sales Win-Propensity Prediction System

At MAN Energy Solutions, we built a system for win-propensity scoring that is integrated into the existing IT infrastructure (see Figure 1). In the spirit of ADR, this system constitutes the main practical contribution of our work. The core of the system is a lightGBM model [20] that produces base win-propensity probabilities for sales quotations and a second-level conditional probability model that accounts for the decaying effect of quotation age on the base win-propensity probabilities. Moreover,

https://doi.org/10.30844/wi_2020_c8-thiess

we train a separate explanatory SHAP (Shapley Additive exPlanations) model to open up the lightGBM black-box model by generating human-interpretable explanations of global and local (individual predictions) feature importance [21]. The model training part of the system executes over the weekend, while the prediction part of the system executes daily. Both parts work fully automated.

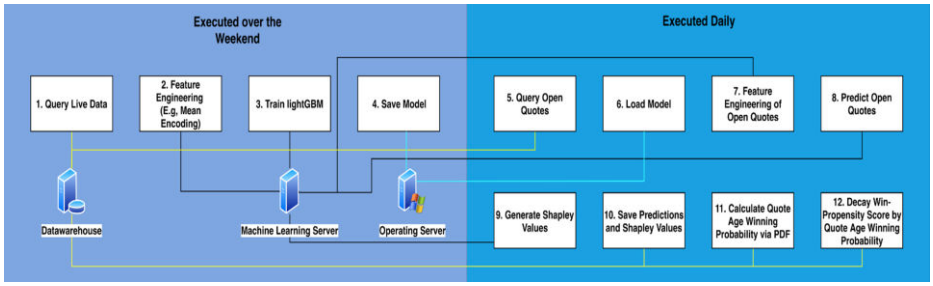

Figure 1. Implemented back-end process

4.1 lightGBM-based Win-Propensity Prediction Model

lightGBM [20] is an advanced implementation of the boosting algorithm [22] that utilizes gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). GOSS excludes instances with small gradients (residuals) from the data and focusses on instances with larger gradients to compute information gain. By this, lightGBM is faster than other implementations of boosting. With EFB, lightGBM can further improve performance by reducing the number of features via bundling variables that are mutually exclusive together. We chose lightGBM mostly due to its performance properties. In our case, the lightGBM algorithm was by far the most efficient tree ensemble method when trained on our data of up to 3 million records of quote positions and 15 carefully selected features (see Table 2). Through cross-validation, we get an average accuracy of 76% and an AUC (area under the receiver operator curve) of 0.74; meaning that there is a chance of 74% that the model can successfully distinguish between a randomly selected won quotation and a randomly selected lost quotation. Compared to a model with no separability power (AUC of 0.5), our model provides a lift of 24%. Furthermore, as the model calculates win-propensity probabilities with an average Brier score of 0.20 and not just binary labels, its outputs can be used by sales professionals directly to evaluate and prioritize sales quotations.

Table 2: Example of features used in the model

Feature	Equip. in Plant (id)	Material (id)	Discount (percent)	List Price (euro)	Processing Time (days)
Encoding	Mean	Mean	Numeric	Numeric	Numeric

https://doi.org/10.30844/wi_2020_c8-thiess

4.2 Second-level Conditional Probability Model for Quotation Age

During the implementation of the system, we faced the challenge of incorporating the time-dependent decay of win-propensity probabilities into the lightGBM model (in other words: the older a quotation gets, the less likely it is that it will be transformed into an order). The technical problem was that for all non-hit training records (i.e., rejected quotations that were never transformed into orders), we lack the reference (order date) to calculate the difference in days between quotation creation and order date.

To overcome this challenge, we developed a two-layer modeling approach. The first step in this approach is to estimate the probability density function (PDF) of the win-propensity score for quotation age. For this, we only use the subset of won-quotations that have an order date. We first calculate the difference between quotation creation and order date (quotation age) and then calculate the frequency of won-quotations and group them by quotation age. As we have access to a large amount of data and quotation age is a continuous random variable, we chose to estimate the PDF via a histogram, which as a non-parametric estimation method is suitable in this case [23]. From the PDF, we can draw probabilities for each quotation age. Eventually, we calculate the time-decayed win-propensity as a conditional probability and integrate it into the user interface (Equation 1 and Figure 2):

$$P(Propensity \mid Age) = \frac{P(Propensity\ and\ Age)}{P(Propensity)} \quad \textbf{(1)}$$

Figure 2. Quotation view of win-propensity prediction (HR pred) and time-adjusted prediction (blurred for confidentiality reasons)

4.3 SHAP Model

SHAP [21] is grounded in the game-theoretical concept of Shapley values. If one imagines that players are collaborating in a team (coalition) to win a game, then Shapley values are the marginal contribution of a player´s performance to the overall success of the team. Based on Shapley values, all players could be paid fairly by their clubs according to their contribution to winning the game.

Machine learning researchers adapted this idea and developed algorithms for local-level machine learning model explanations (per prediction) [21, 24–27]. The idea here is that the prediction, in our case predicting win-propensity, is the game, and the feature values are the players. Thus, if we can calculate the marginal contribution of each feature value, we have a consistent method of feature importance that is superior

https://doi.org/10.30844/wi_2020_c8-thiess

to standard feature importance methods such as gain (in terms of Gini index) or splitcount. Moreover, compared to the local interpretable model-agnostic explanations algorithm (LIME) [28, 29], SHAP is more interpretable, since its explanation values add up to the model output. Also, SHAP-based global feature importance allows visualizing non-monotonic relationships (bi-directional; see Figures 3 and 4).

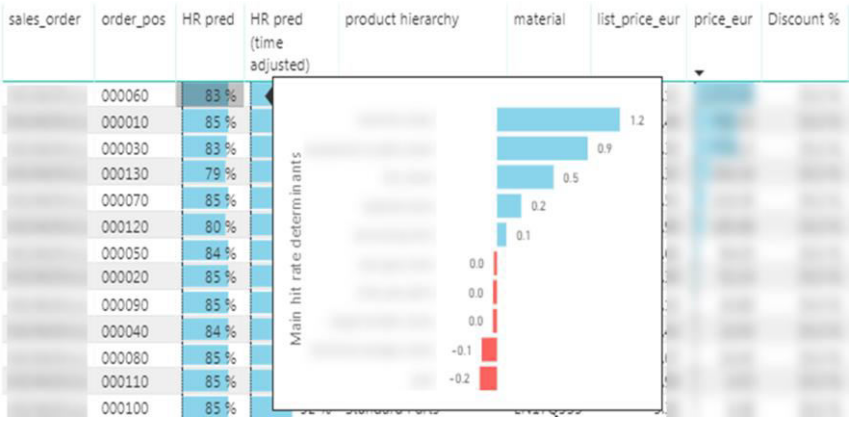

Figure 3. Local instance-level SHAP explanation (blurred for confidentiality reasons)

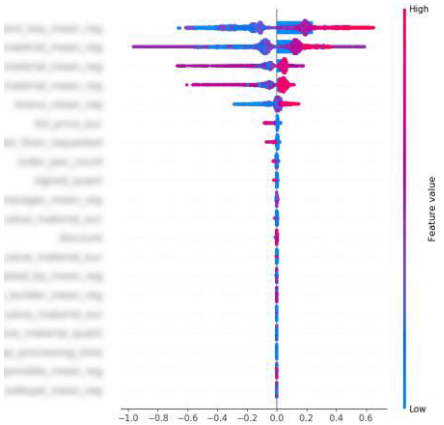

Figure 4. Explanations showing both magnitudes (X-axis) and non-monotonic value impact (color code; the figure is blurred for confidentiality reasons)

4.4 Implementation Method

Next to the core system consisting of back-end and user interface, we designed an implementation method for our system that was, despite its initial theoretical grounding, developed abductively based on the learnings of the different ADR building, intervention, and evaluation cycles (see Figure 4). One notable highlight of the method is the utilization of domain knowledge to develop hypotheses of drivers of

https://doi.org/10.30844/wi_2020_c8-thiess

hit rate that, in the following, are tested via SHAP in an explanatory analysis. Furthermore, we add steps of regular stakeholder presentations as well as the utilization of UI mockups for those presentations. Also, we distinguish between model building and evaluation in experimental lab situations and in a more naturalistic production environment, which, in our experience from the case, can give different results and, therefore, valuable insights into the modeling and implementation process. Moreover, we added steps of deployment in production UI, change management, and live monitoring to account for the fact that the data generating processes that our models rely on may change.

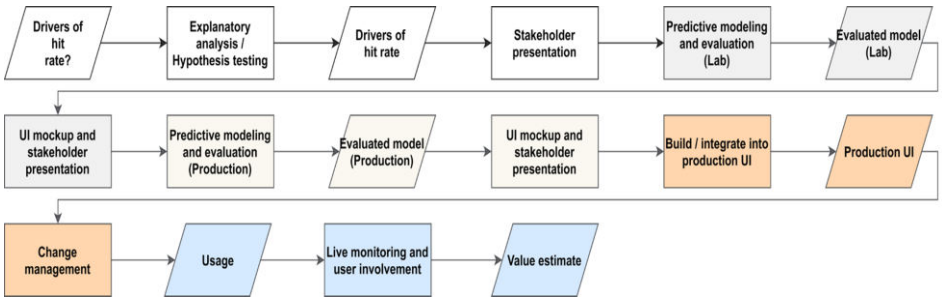

Figure 5. Implementation method

5 Design Principles

The following design principles abstract from the concrete artifacts described in Section 4 and capture general prescriptive knowledge that should enable others to build explainable machine learning systems, in particular, win-propensity prediction systems. In the spirit of ADR, these design principles constitute the main scientific contribution of our work.

5.1 Local-contrastive Explainability: Present model explanations to users on an instance-level to support contrastive explanation processes

Due to the complexity of the aftersales market in the shipbuilding industry and its industry-specific challenges, such as intransparent owner structures, bulk orders, and product heterogeneity, sales professionals at MAN rely on implicit domain knowledge and collaborate closely with other domain experts (e.g., engineers). Not surprisingly, we found that these sales professionals often do not trust the predictions of a machine learning model. While the sales professionals are not necessarily interested in fully understanding how the machine learning model has generated a score, they still want to know why the model predicts a specific win-propensity score for a specific quotation. They may ask: “Why has quotation X a win-propensity score of 0.9 and not 0.2?” According to Lipton [30] and Miller [31], answering such why-questions requires the explainee to contrast the observed event (score of 0.9) with an imagined counterfactual event (score of 0.2) to abductively infer the most plausible explanation

https://doi.org/10.30844/wi_2020_c8-thiess

of the observed event (score of 0.9) [32]. We support this cognitive process of abductive reasoning by displaying Shapley values alongside main variables such as discount % or material in our user interface (Figure 3). In the abduction process, the Shapley and variable values function as candidate hypotheses of causes for the observed event (score of 0.9), which users can compare with their mental model to assess their plausibility and eventually answer the contrastive question: "Why has quotation X a win-propensity score of 0.9 and not 0.2?"

5.2 Selective Visualization: For local explanations, visualize only the top contributing features to reduce explanation complexity

Our front-end shows a report of open quotations along with the predicted win-propensity scores and a time-decayed version of it that accounts for quotation age. This report already contains much information, and processing it puts a high cognitive load on users. Hence, we decided to not increase the information processing load further with our explanations. Instead, we wanted to limit the complexity of our instance-level explanations by providing on-demand visualizations of only the top-5 most important positive and negative features.

This empirically motivated design decision can be backed up with psychological theory. Psychological research suggests that human short-term memory can only recall 4-7 chunks of information at a time [33, 34]. A chunk is the largest unit of information that human memory can represent. How the human brain creates these chunks depends on its prior knowledge. When confronted with familiar concepts, our brain can create larger chunks, and therefore, recall more information. Visualization supports this cognitive chunking process by grouping (or pre-chunking) information into symbolic representations so that one can display much larger amounts of information that otherwise could not have been recalled simultaneously [35].

5.3 Accountability: Schedule regular management presentations to increase data scientists´ need for justification

In our implementation method (Figure 5), we propose repeated stakeholder presentations to create and sustain organizational support. Committing to those presentations comes with the side effect of having to justify one´s approach and progress to the stakeholders. As a result, one can be made accountable for what one has done between the meetings.

Research from the field of psychology suggests that accountability, the need for justification of one´s viewpoints towards other people, lets decision-makers judge in more complex ways, rely less on prior beliefs, and be more evidence-based (and supposedly more data-driven). By this, accountability affects decision making in a debiasing way [36–38]. For a developer (data scientist) that follows our method, this self-created accountability increases the need to understand how its machine learning model works. Thus, it motivates developers to align the gap between their mental model and the machine learning model [18], which eventually can lead to improved

https://doi.org/10.30844/wi_2020_c8-thiess

model quality. Moreover, having a solid understanding of a machine learning model is a pre-requisite for explaining it to others in a simple, but not simplistic, way.

5.4 Global Explainability: Explain the machine learning model to managers on a global level to increase acceptance, enable process accountability, and share outcome accountability

In our implementation method, we included a step of explanatory analysis/hypothesis testing, based on our experience that managers became much more engaged, contributed with domain knowledge, and seemed to be more positive towards the project, once we presented the results of our explanatory analysis. We presented not only findings concerning the drivers of hit rate but also how the different feature values on average affect the prediction of the model (global explanations).

While there is, as mentioned above, research that indicates a positive impact of accountability on decision making, there is also research suggesting negative forms of impact for some types of accountability [38, 39]. In particular, Simonson and Staw [38] suggest that outcome accountability triggers a mechanism by which accountable persons perceive an increased need to self-justify past behavior and decisions, which, in turn, leads to an escalation of commitment to such behaviors. Process accountability, on the other hand, leads to a more thorough alternative evaluation in decision making, but also a decrease of the need to self-justify past behavior, since one can justify behavior via a thoroughly evaluated and transparently reported process instead of exploiting or defending an eventual outcome only.

Based on our experience from the case, we argue that in machine learning projects, it is difficult for managers to comprehend the complex processes and mechanisms that underly a system. In reaction to this, managers may tend to make developers (data scientists) outcome accountable. However, when faced with a task such as implementing a novel machine learning system, where the outcome uncertainty is high, outcome accountability increases the stress-level of data scientists (see [40]). The reason for this is that in high outcome-uncertainty situations, it is particularly challenging for evaluators to assess the effort-outcome relation so that even when data science teams deliver high-quality work, the project can fail due to factors that are out of their control. In such situations, process accountability may be preferable to relieve some of the stress related to the low effort-outcome reliability [41] and its negative consequences [40].

Nevertheless, it is hard to evaluate the quality of a process if it is not explainable. We experienced that presenting global explainability methods such as average SHAP feature importance (Figure 4) to managers, helps them to align their mental model with the machine learning model and the mental model of the data scientist. It allows evaluating whether a course of action (process) chosen by the developer makes sense or not, which eventually enables managers to make data scientists process accountable and, consequently, share the outcome accountability of machine learning projects and systems.

https://doi.org/10.30844/wi_2020_c8-thiess

5.5 Confirmatory Nudging: Use language and representation devices that align well with users´ and managers´ mental models to increase acceptance of the machine learning model

In our system, we made sure that we use a vocabulary (esp. feature names) that is familiar to the stakeholders from the maritime industry, such as motor manager (customer), or equipment_in_plant (engine installed on a vessel) instead of non-speaking feature names such as x1, x2, x3 or features names that are uncommon in the given company and industry. Using such names when explaining the machine learning model helps to narrow the gap between a stakeholder´s mental model and the machine learning model, which, in turn, should increase its acceptance. Moreover, we made sure to present a working prototype early on (Figures 2, 3, and 4) by integrating the machine learning scrips and models into the existing infrastructure, which enabled us to demonstrate the model in an already familiar user interface.

Confirmation bias [42] describes a tendency to favor information that aligns well with one´s prior beliefs (mental model). By adding domain-specific traits to the system, we exploit this cognitive bias to influence the behavior of users and managers in a predictable positive way (nudging) [43].

5.6 Causality: Choose the machine learning model that aligns best with reality and design it as if it was an explanatory rather than a predictive model to increase model acceptance by users, managers, and developers

Shmueli [44, p. 293] discusses the differences between explanatory and predictive modeling, amongst others, based on the following two characteristics. (1) Causation-association: "In explanatory modeling *f* represents an underlying causal function, and *X* is assumed to cause *Y*. In predictive modeling *f* captures the association between *X* and *Y*". (2) Theory-data: "In explanatory modeling, *f* is carefully constructed based on *F* in a fashion that supports interpreting the estimated relationship between *X* and *Y* and testing the causal hypotheses. In predictive modeling, *f* is often constructed from the data". In our implementation method, we incorporated a step of reaching out to the business in order to develop hypotheses (low-level theory) of how the features relate to the target. In the next step, explanatory analysis, we are testing those hypotheses with accessible data. So instead of looking for associations only, a typical approach when the objective is mostly predictive, we start with developing a causal theory of how the features relate to the target variable, which is common when the objective is explanatory.

A more technical distinction between explanatory and predictive modeling objectives is the treatment of multicollinearity [43, p. 288]: "Multicollinearity is not a problem unless either (i) the individual regression coefficients are of interest, or (ii) attempts are made to isolate the contribution of one explanatory variable to Y, without the influence of the other explanatory variables. Multicollinearity will not affect the ability of the model to predict." Also, SHAP (Figures 3 and 4) assumes independent features [21]. A violation of this assumption can bias the Shapley values for dependent (multi-collinear) variables since the algorithm cannot attribute the

https://doi.org/10.30844/wi_2020_c8-thiess

distinctive contribution of each feature to the prediction. In reaction to that, we treat multicollinearity like one would do when having purely explanatory objectives. First, we identify collinearity via correlation matrices and multicollinearity via variance inflation factor (VIF) analysis. Based on this and the developed causal model, we remove the collinear variables or merge them.

To summarize, we are designing the model to achieve both predictive and explanatory objectives. In our case, this comes with the benefit of increased explainability, while keeping the loss in predictive power neglectable.

6 Discussion and Conclusions

In this paper, we presented an explainable two-level win-propensity prediction system that utilizes the lightGBM algorithm (4.1), a conditional probability model for quotation age (4.2), SHAP explanations on both local and global levels (4.3), an interactive user interface (4.3), an implementation method (4.4), and abductively developed design principles (5).

To the best of our knowledge, our work is the only one that provides an implementation method and derives design principles based on learnings from designing and implementing a novel sales win-propensity prediction system in a real-world environment. Also, there is no other sales win-propensity approach for predicting the probability of converting a sales quotation into a sales order (hitting). Moreover, there is no specific machine learning approach for sales predictions in the maritime manufacturing industry.

Nevertheless, there are some approaches for predicting the win-propensity of sales leads or opportunities, which is a comparable sales conversion process that, however, happens earlier in the sales funnel. In this domain, researchers from IBM developed with OnTARGET a logistic regression model that predicts the propensity of customers to buy IBM´s products [46]. Zan et al. [47] applied a neural network-based approach. Yan et al. propose a win-propensity approach based on modeling the interaction of users with the sales support system as Hawkes Processes [9]. Duncan and Elkan propose a pure probabilistic model [48]. Compared with these approaches, our approach is theoretically either superior in terms of predictive or explanatory power, and always superior in balancing predictive and explanatory power.

The approach by Bohanec et al. [8, 11, 49] and Eitle and Buxmann [50] are the only other approaches that come close in terms of theoretical predictive and theoretical explanatory power. They utilize with random forest and gradient boosting machines some of the empirically proven best-performing prediction algorithms, that, however, are more resource-intensive when compared to lightGBM. Furthermore, they do not address the issue of time-decay in win-propensity scores that we approach with our second layer conditional probability model. Also, the explanation methods IME and EXPLAIN [55, 56] used by Bohanec et al. and LIME [28] used by Eitle and Buxmann do not fulfill the criterion of explanation accuracy that SHAP fulfills [21]. None of the approaches explicitly deals with multi-collinearity, which potentially makes their approaches less aligned with reality, and due to this less suitable to align

https://doi.org/10.30844/wi_2020_c8-thiess

well with the mental models of users, managers, and also their own mental models, which in turn can lead to decreased trust, low acceptance, and low model quality (see [16, 18]).

We build our approach during a 1.5 years lasting ADR project at MAN Energy Solutions. It means that the final shape of the system, the implementation method, and the design principles are not necessarily generalizable to other environments. However, they should be transferable to similar environments that are concerned with similar problems. While our system (Section 4) deals with challenges that should be transferable to many other B2B environments, our design principles are even further abstracted to the class of explainable machine learning, which should be transferable even to B2C environments.

In the future, we want to study further how the system, with its explanatory capabilities, affects the acceptance by both users and managers. Moreover, we want to compare the accuracy of the win-propensity predictions generated by the system with those generated by users. Furthermore, we plan to integrate the aggregated win-propensity predictions for a current sales pipeline as an operations-level forecast into a more general strategic forecasting algorithm.

References

1. Danish Ship Finance: Shipping Market Review. (2018).
2. Benbasat, I., Zmud, R.W.: Empirical Research in Information Systems: The Practice of Relevance. MIS Q. 23, 3 (2006).
3. Sein, Henfridsson, Purao, Rossi, Lindgren: Action Design Research. MIS Q. (2011).
4. Monat, J.P.: Industrial sales lead conversion modeling. Mark. Intell. Plan. 29, 178–194 (2011).
5. Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C., Huang, J., Chu, S., Yang, X.: On machine learning towards predictive sales pipeline analytics. In: Twenty-ninth AAAI conference on artificial intelligence (2015).
6. Xu, X., Tang, L., Rangan, V.: Hitting your number or not? A robust & intelligent sales forecast system. In: 2017 IEEE International Conference on Big Data (Big Data). pp. 3613–3622. IEEE (2017).
7. Tversky, A., Kahneman, D.: Judgment under uncertainty: Heuristics and biases. Science (80-.). 185, 1124–1131 (1974).
8. Bohanec, M., Robnik-Šikonja, M., Kljajić Borštnar, M.: Organizational Learning Supported by Machine Learning Models Coupled with General Explanation Methods: A Case of B2B Sales Forecasting. Organizacija. 50, 217–233 (2017).
9. Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C., Huang, J., Chu, S., Yang, X.: On machine learning towards predictive sales pipeline analytics. (2015).
10. D'Haen, J., Van den Poel, D.: Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework. Ind. Mark. Manag. 42, 544–551 (2013).
11. Bohanec, M., Kljajić Borštnar, M., Robnik-Šikonja, M.: Explaining machine learning models in sales predictions. Expert Syst. Appl. 71, 416–428 (2017).

https://doi.org/10.30844/wi_2020_c8-thiess

12. Sharma, R., Mithas, S., Kankanhalli, A.: Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. Taylor & Francis (2014).
13. Drucker, P.F.: The Effective Decision, (1967).
14. Hollander, E.P., Vroom, V.H., Yetton, P.W.: Leadership and Decision-Making. Adm. Sci. Q. (1973).
15. Sutanto, J., Kankanhalli, A., Tay, J., Raman, K.S., Tan, B.C.Y.: Change Management in Interorganizational Systems for the Public. J. Manag. Inf. Syst. 25, 133–176 (2009).
16. Kayande, U., De Bruyn, A., Lilien, G.L., Rangaswamy, A., van Bruggen, G.H.: How incorporating feedback mechanisms in a DSS affects DSS evaluations. Inf. Syst. Res. (2009).
17. Gregor, S., Benbasat, I.: Explanations from intelligent systems: Theoretical foundations and implications for practice. MIS Q. (1999).
18. Martens, D., Provost, F.: Explaining data-driven document classifications. (2013).
19. Baskerville, R.L., Myers, M.D.: Design ethnography in information systems. Inf. Syst. J. 25, 23–46 (2015).
20. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y.: LightGBM : A Highly Efficient Gradient Boosting Decision Tree. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). pp. 3149–3157 (2017).
21. Lundberg, S.M., Erion, G.G., Lee, S.-I.: Consistent Individualized Feature Attribution for Tree Ensembles. (2018).
22. Freund, Y., Schapire, R.E.: Experiments with a New Boosting Algorithm. (1996).
23. Izenman, A.J.: Recent Developments in Nonparametric Density Estimation. J. Am. Stat. Assoc. 86, 205 (1991).
24. Štrumbelj, E., Kononenko, I.: A General Method for Visualizing and Explaining Black-Box Regression Models. Presented at the (2011).
25. Štrumbelj, E., Kononenko, I.: Explaining prediction models and individual predictions with feature contributions. Knowl. Inf. Syst. 41, 647–665 (2014).
26. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions, (2017).
27. Lundberg, S.M., Lee, S.-I.: Consistent feature attribution for tree ensembles. (2017).
28. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why Should I Trust You?" Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '16. 1135–1144 (2016).
29. Ribeiro, M.T., Singh, S., Guestrin, C.: Model-Agnostic Interpretability of Machine Learning. (2016).
30. Lipton, P.: Contrastive Explanation. R. Inst. Philos. Suppl. 27, 247–266 (1990).
31. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences, (2019).
32. Harman, G.H.: The Inference to the Best Explanation. Philos. Rev. 74, 88 (1965).
33. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev. 63, 81 (1956).
34. Cowan, N.: The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behav. Brain Sci. 24, 87–114 (2001).
35. Larkin, J.H., Simon, H.A.: Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cogn. Sci. 11, 65–100 (1987).

https://doi.org/10.30844/wi_2020_c8-thiess

36. Tetlock, P.E.: Accountability: The neglected social context of judgment and choice. Res. Organ. Behav. 7, 297–332 (1985).
37. Tetlock, P.E., Skitka, L., Boettger, R.: Social and Cognitive Strategies for Coping With Accountability: Conformity, Complexity, and Bolstering. (1989).
38. Simonson, I., Staw, B.M., Haas, W.A.: Deescalation Strategies: A Comparison of Techniques for Reducing Commitment to Losing Courses of Action. (1992).
39. Lerner, J.S., Tetlock, P.E.: Accounting for the effects of accountability. Psychol. Bull. 125, 255 (1999).
40. Siegel-Jacobs, K., Yates, J.F.: Effects of procedural and outcome accountability on judgment quality. Organ. Behav. Hum. Decis. Process. 65, 1–17 (1996).
41. Wiseman, R.M., Gomez-Mejia, L.R.: A Behavioral Agency Model of Managerial Risk Taking. Acad. Manag. Rev. 23, 133–153 (1998).
42. Nickerson, R.S.: Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Rev. Gen. Psychol. 2, 175–220 (1998).
43. Thaler, R.H., Sunstein, C.R.: Nudge: Improving decisions about health, wealth, and happiness. Penguin (2009).
44. Shmueli, G., Galit Shmueli: To explain or to predict? Stat. Sci. 25, 289–310 (2010).
45. Wheelwright, S., Makridakis, S., Hyndman, R.J.: Forecasting: methods and applications. John Wiley & Sons (1998).
46. Lawrence, R., Perlich, C., Rosset, S., Khabibrakhmanov, I., Mahatma, S., Weiss, S., Callahan, M., Collins, M., Ershov, A., Kumar, S.: Operations research improves sales force productivity at IBM. Interfaces (Providence). 40, 33–46 (2010).
47. Zhang, C., Li, X., Yan, J., Qui, S., Wang, Y., Tian, C., Zhao, Y.: Sufficient Statistics Feature Mapping over Deep Boltzmann Machine for Detection. In: 2014 22nd International Conference on Pattern Recognition. pp. 827–832. IEEE (2014).
48. Duncan, B., Elkan, C.: Probabilistic modeling of a sales funnel to prioritize leads. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2015-Augus, 1751–1758 (2015).
49. Bohanec, M., Robnik-Šikonja, M., Kljajić Borštnar, M.: Decision-making framework with double-loop learning through interpretable black-box machine learning models. Ind. Manag. Data Syst. 117, 1389–1406 (2017).
50. Eitle, V., Buxmann, P.: Business Analytics for Sales Pipeline Management in the Software Industry: A Machine Learning Perspective. (2019).
51. Štrumbelj, E., Kononenko, I., Robnik Šikonja, M.: Explaining instance classifications with interactions of subsets of feature values. Data Knowl. Eng. 68, 886–904 (2009).
52. Robnik-Šikonja, M., Kononenko, I.: Explaining classifications for individual instances. IEEE Trans. Knowl. Data Eng. 20, 589–600 (2008).

https://doi.org/10.30844/wi_2020_c8-thiess