NOVA SCHOOL OF BUSINESS & ECONOMICS

A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

THE EXPECTED VALUE OF APPLYING MACHINE LEARNING FOR PREDICTIVE LEAD SCORING BASED ON CUSTOMER CONTACT-FORM INPUT – A CASE IN THE B2B ENERGY SECTOR

CARINA SELL

Work project carried out under the supervision of:

Dr. Rodrigo Belo (NOVA School of Business and Economics)

Dr. Rachel Marie Harris-Lethaus (Partner Company)

Fabian Fischer (Partner Company)

19/01/2023

Table of Contents

Abstract

This thesis examines the performance of machine learning to predict lead conversion probability in an early lead maturity stage based on customer contact-form data. An empirical case study was conducted developing models at two maturity stages in the lead funnel using automated machine learning and their expected value for the business was calculated. The resulting models prove the suitability of machine learning to predict lead conversion and reveal that predictions are better in a later maturity stage. Furthermore, the findings suggest including cost and benefit calculations in the development is beneficial, as not all models are profitable despite good performance.

Keywords: Predictive Lead Scoring, Lead Management, Expected Value Framework, Machine Learning, CRM

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

In times of uncertainty, intense competition, and ongoing change, the key to engagement for any company is customer-centricity and strong Customer Relationship Management (CRM) (Williams 2014, 89). Yet, one CRM area often neglected is customer acquisition (Buttle 2012, 178–80). In order to acquire new customers successfully and profitably, companies rely on a solid customer acquisition strategy, one possible option being *lead management*, pointed out as the most innovative one by several studies (for example D'Haen et al. 2016; Ohiomah, Benyoucef and Andreev 2016).

Lead management can be defined as a structured way to generate, nurture and qualify potential customers (= leads) using marketing and sales strategies (Wenger 2021, 261). Nowadays, a large portion of lead generation happens online, as potential customers inform themselves about products on the internet and reach out via different digital channels. For example, they send an email, subscribe to a newsletter, or use a contact form. However, only a tiny proportion of leads who show first interest conclude a sale, i.e., convert – depending on the industry, typical conversion rates range from 1% to 5% (Coe 2004). Especially in the B2B context, characterized by higher complexity of products, longer sales cycles, and thus low conversion rates (Wenger 2021, 253–54), identifying and prioritizing the most promising leads early on can make an enormous difference (Monat 2011, 192).

Lead scoring is a way to determine the likelihood of conversion to make a prioritization by attaching a probability value calculated from the characteristics and behaviours of a lead. Instead of randomly or chronologically pursuing leads, they can be structured based on their anticipated desire to buy. Therefore sales representatives, who previously spent approximately 20% of their time selecting prospects (Trailer 2006), can use this time more effectively, saving costs and enabling a timelier follow-up. Ohiomah, Benyoucef and Andreev (2016) argued that the chances of converting a lead into a sale are higher the earlier that contact takes place after an inquiry occurs.

In recent years the use of data mining techniques to automate the lead scoring and qualification process has emerged. Despite the practical relevance of the problem and the tremendous efficiency potential, only a few studies have dedicated their research to developing machine learning models that predict the likelihood of converting a lead. Two areas of focus in the sales funnel become apparent: *Prospect qualification models* help companies determine which prospects they should proactively contact before the customer shows a first interest utilizing behavioural data like browsing history. S*coring models* are used in a later maturity stage and utilize extensive characteristic data about the lead that first must be gathered and is often externally bought (see chapter 2.1 for a detailed literature review). However, when leads enter the CRM system, often, there is only a small amount of information available internally, and the use of machine learning (ML) in this early lead stage is currently less well researched.

Hence this thesis is focused on the use of ML for lead qualification in an early maturity stage using internal data only, on the example of leads generated through online contact forms. The main question researched is: "*Can the use of machine learning in the early entry stage of the lead management process based on customer-contact-form data add significant value to the business?".* To answer this question, an empirical case study examining the lead management process of a German energy provider in the B2B sector is conducted. Various ML models for lead scoring are developed at two maturity levels in the lead management process, and their business value is evaluated and compared using the expected value framework (Fawcett and Provost 2013, 187).

The remaining paper is structured as follows: in chapter two, a literature review outlines the theoretical background of lead management and related work using ML in this field. Then, in section three, the methodology used to develop and evaluate the model is presented, followed by an overview of the results looking at model performance and business value in chapter four. Finally, the conclusions are discussed, and limitations, as well as future research opportunities, are highlighted.

2. Literature Review

2.1 Lead Management Framework

As briefly outlined in the introduction, identifying and prioritizing leads is challenging for many businesses. The sales funnel conceptualization divides the customer acquisition process into different stages. Depending on the product or service sold, these divisions vary from study to study, and no predominant unifying lead management definition has emerged (Monat 2011, 179; Espadinha-Cruz, Fernandes, and Grilo 2021). Figure 1 gives an overview of the different stages defined in this research as leads mature in their buying journey based on the framework presented by Wenger (2021).

Suspects are all potential new customers who could hypothetically buy the company's product; in a B2B context, that would include any other company. *Prospects* are all suscepts with actual interest in the product. Both stages, called *lead generation*, comprise all marketing-related activities with the primary goal of creating awareness and turning awareness into interest. Often marketing teams track behaviour and interactions with websites and campaigns to identify marketing-qualified leads (MQL) with whom the company should initiate direct contact. At this point of initial contact or if the prospect proactively expresses

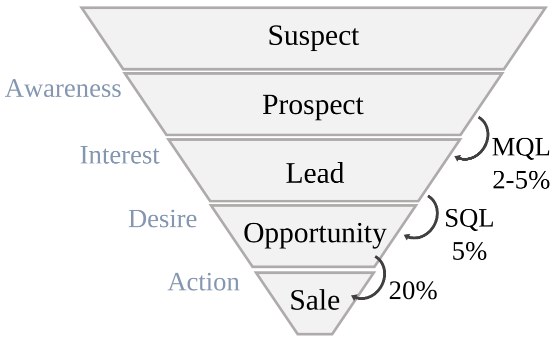


Figure 1: Sales Funnel Overview

interest through, for example, a call, e-mail, or use of a contact form, prospects become *leads* and are handed over to the lead management team.

The main objective in this maturity stage is sales *lead qualification*: more information about the lead is gathered to determine how qualified the lead is, based on a set of product-specific criteria, as well as critical aspects of the sales case (BANT = budget, authority, need, and time). When a lead is characterized as qualified by the lead management team (= sales-qualified lead; SQL) and the interest turns into desire, the lead is handed over to the sales team during the lead routing process and called an *opportunity* from now on. At this point, sales representatives take applicable action, such as product demos or client meetings, with the final goal of converting the opportunity into a *sale*.

The percentage of leads progressing to the next maturity stage varies depending on the industry. However, traditionally the conversion rate from lead to opportunity can be approximated at 5%, as shown in Figure 1 (Duncan and Elkan 2015, 1751). Looking at the total lead funnel, on average, around 1-5% of all prospects become customers (Coe 2004, 65).

To make the lead management process more efficient and increase conversion rates, a qualification is conducted within each maturity stage. The probability for the lead to become a customer is determined based on needs and information available at the respective maturity level. The key difference between lead qualification and lead scoring is that lead scoring assigns specific values to the probability of conversion (Stevens 2012, 145). Two dimensions are considered: explicit and implicit scoring (Lewis 2012, 56). Explicit lead scoring mainly uses demographic data like industry, region, and company information to define how closely a lead's profile matches the defined customer profile. Implicit scoring utilizes documented behaviour like how often and through which channels the lead engages with the company. However, qualifying and scoring leads is a complex process. The assessment is often influenced by personal judgement, and it is difficult to determine if and when a lead should be targeted (Espadinha-Cruz, Fernandes, and Grilo 2021, 3).

Considering these challenges, using data mining techniques in the lead-scoring and qualification process can benefit organizations significantly (Monat 2011, 192). Automated methods are quicker and work in real-time, instantly equipping sales personnel with comprehensive information about the leads. ML is also less prone to errors making the process more accurate and data-driven. Another significant benefit of relying more heavily on algorithms than on individuals is that a predictive lead-management system can operate at a much larger scale and be an advantage when lead numbers increase. It has the ability to carry out repetitive, laborious tasks to the same standard indefinitely. Companies who use artificial intelligence and ML algorithms in their sales process see their lead numbers rise by up to 50%, their average talk times drop by up to 60-70%, and their costs reduced by up to 40-60% (Salesforce 2018).

Given these advantages, ML in lead management has gained importance in the research environment in the last decade. The following subchapter gives a brief overview of related work in the field.

2.2 Machine Learning Models for Lead Management

An extensive literature review on the application of machine learning in CRM by Ngai, Xiu, and Chau in 2009 demonstrated that, while there are several studies around ML in customer attraction and retention, there is a considerable gap when it comes to customer acquisition. The assessment of studies since 2009 presented in this document led to the conclusion that, to date, still little research exists, similar to what Espadinha-Cruz, Fernandes, and Grilo concluded in their article in 2001. Out of all results in standard online libraries, only 15 articles were deemed relevant for the application of ML in lead management, most using cases in a B2B environment. The following paragraphs summarize these works.

More than 50% of the selected articles applied ML in the second level of the sales funnel, developing models to *automate prospect scoring*. For example, D'Haen and Van den Poel (2013) used unsupervised ML techniques combined with binary classifiers to create a profile of current customers and then ranked prospects based on their similarity to these characteristics. Jadli et al. (2022) as well as Etminan (2022), introduced models for scoring prospects based on their activities and online interactions with the company. Both tried various algorithms and found tree-based algorithms like random forest and gradient boost to be best-performing. Further, several authors presented explicit lead scoring models that use general information about the prospects available online utilizing web scraping (D'Haen, Van den Poel, and Thorleuchter 2013; D'Haen et al. 2016; Benhaddou and Leray 2017). D'Haen, Van den Poel, and Thorleuchter (2013) also compared data mining techniques combined with different data sources, finding that a bagged decision tree is the best classifier among their options. Another type of ML was used by Mezei (2020), who introduced time series forecasting based on behavioural data, achieving good model performance while trying to eliminate bias using different ways of aggregating features with a relative chronological relationship. Random Forest was concluded as the most promising classifier for lead scoring, and in addition to other studies using black box models, the authors highlighted the generation of business insights using feature importance. Bohanec, Robnik-Šikonja, and Kljajić-Borštnar (2017) also focused on business understanding and presented an explanation model that allows deeper knowledge and transparency of the opportunity prediction.

Concentrating on a later maturity stage in the sales funnel, Duncan and Elkan (2015) developed a multiclass classification model to automate conversion from *lead to opportunity* and from *opportunity to sale*. The authors used behavioural and demographic data from the internal Salesforce system, exclaiming a need for a large historical dataset and emphasizing the following features: customer company size, industry, geographical location and company market value. Behavioural data was found to be less predictive than demographic.

Similarly, Eitle and Buxmann (2019) proposed a solution creating one model to predict whether a lead will become an opportunity, a separate model for the prediction of opportunity to customer, and an overarching model from lead to customer. Using mostly static categorical variables, CatBoost was found to be an effective algorithm. However, random forest performed better in early stages due to better handling of outliers and noise. The overarching model's low Area Under the Curve (AUC) showed that it is hard to predict sales in the earlier lead phase.

Lambert (2018) and Rezazadeh (2020) focused on computing a sales-win propensity score for *opportunities* using information from the internal CRM database gathered in the prior process. The superior performance comparing the AUC of these models to prior mentioned works showed that the extensive information gathered at this point in the sales funnel enables more robust predictions. In addition, Rezazadeh was the first to develop the model within an end-to-end cloud-based workflow using Azure software, which will also be used in this study.

The only authors that tackled *the dynamic nature of the lead management process* and explored a non-static model capturing the whole process are Yan et al. (2015). The two-dimensional Hawkes dynamic clustering process allows for live assessment based on company information and variables generated from interaction with sales representatives. However, this solution relies heavily on correct and timely updates through sales representatives in the CRM system, which is often considered a problem in larger B2B firms.

While studies mentioned above purely focused on the modelling phase, Espadinha-Cruz, Fernandes, and Grilo (2021) presented an *overall framework* for the application of data mining techniques within lead management. This framework includes capturing, enriching, nourishing and tracking data for the model and generating insights for the business afterwards. The application of the framework in a telecommunications company uncovered many complications in reality: there was a general lack of quality data, an immense need for data cleaning and the wrong information was gathered. Only a few aspects mentioned in literature were covered by the discovered variables. These are conditions to remember when undertaking a ML project focused on customer acquisition.

In summation, although the last few years have seen a growth in research in the application of ML for lead management, there is still a research gap visible. Nevertheless, all work conducted provides evidence for the good performance that ML can achieve in this area. Furthermore, while a wide range of different algorithms was discussed, more complex tree-based classifiers were superior in most applications, defining the choice of classifiers for this work. Few studies demonstrated that while behavioural data is more critical for models in the prospect phase, demographic data has higher predictive power in the later opportunity phase.

Whereas the prospect and opportunity phases are documented in the literature, research has yet to discuss the role of ML in the early lead stage. Moreover, the variety of features found to be impactful leads to the conclusion that lead management processes are highly tailored to the product or service sold. Therefore, no generic model can be applied. Instead, looking closely at available information, processes, and specific goals within the respective company is essential to create a sound model that shows good performance. Besides performance, few authors have examined model interpretability to gain business insights; however, the financial value generated for the business has never been addressed to the authors' knowledge.

Based on these findings, this paper tries to identify the expected value from the use of ML within two of the less explored maturity stages: the beginning and the end of the lead phase. With this, we aim to fill a gap in the literature and contribute towards a better understanding of lead management in different businesses.

3. Methodology

To illustrate the added value that ML can offer by creating automated lead scoring, the best method to adopt for this investigation is a case study approach. Due to the lack of a standard process and the differences mentioned above in the lead management process, focusing on one company enables studying one aspect of a real-world problem in detail from different viewpoints. The present case study was pursued in a German company in the energy industry focusing on their B2B business. For the development of the model, the cross-industry standard process for data mining (CRISP-DM) was applied (Chapman et al. 2000) as it provides a cost-effective uniform framework for the project and includes business understanding as a key factor. The activities consisted of six phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment (see Appendix 1). The following subchapters will guide through each step, except deployment, which was excluded from this study.

3.1 Business Understanding

The German energy market is shaped by well-informed customers, intensified competitive activity, and increasing switching rates (Kuenzel and Lohse 2022, 365). Electricity and gas are special commodities – standardized, homogenous products without differentiation through product quality, packaging, or performance. They generate no customer benefit itself. Therefore, they can be classified more as a service than a product (Kuenzel and Lohse 2022, 366–67), and communication, customer service, and relationship with the customer are even more critical. A study from Geigenmüller and Kuhn (2022) showed that communication can be a source of differentiation in the commodity market at every stage of the CRM, underlying the need for solid lead management.

Like most energy suppliers, the company in focus generates most of its leads online. First, prospects visit the company's website, where they can use a contact form (CF) to express their interest or use an online price indicator called "pricing tool" (PT): a form that generates a price estimation for an energy or gas contract when a few key parameters are filled in.

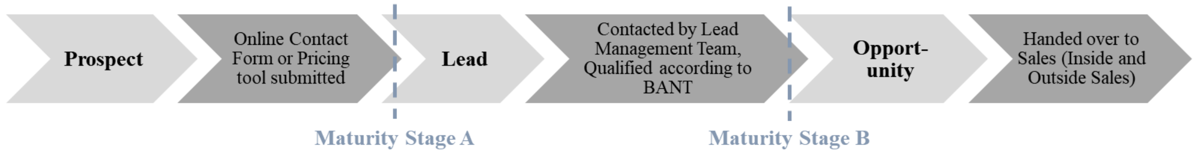


Figure 2: Company Lead Funnel

Information entered in these forms is transmitted into the company's CRM system as a new lead. This point in time is defined as *maturity stage A*. A small, dedicated lead management team then calls all new leads to gather more information about their needs and qualifies leads using BANT (see chapter 2.1). In the next step, all selected leads are routed to the respective key account manager as opportunities based on annual consumption, points of delivery and location (= *maturity stage B*).

With 100 new leads per week on average, often there is a backlog of leads, and the team reaches out to new leads based on first-come-first-serve and intuition. An automated lead scoring model would enable the team to work more efficiently, not investing time in leads with low conversion probability and ensuring that leads with high probability could be prioritized. A fast response can be seen as an indicator of good customer service, which is a crucial differentiation from competition in the commodity business at this point in the sales funnel.

For the company to be able to use the model, the following aspects should be kept in mind: the method should use internal information available at the respective maturity stage only, it should be easy to use and integrate into the current CRM system, and it should be easy to understand and interpret for the team, providing valuable business insights.

3.2 Data Understanding

A total of 7803 leads was included in the analysis for the time 01.11.2019 to 1.11.2022 based on two primary sources of data: the online contact form (n = 3412) and the online pricing tool (n = 4391). The internal CRM system captures all demographic data of the leads and activity (= behavioural) information throughout the process. When a new lead is created from the CF or PT, the system automatically saves all entries and some implicit information. In addition, every interaction between the sales representatives during the lead qualification phase is entered manually and classified by type of interaction, e.g., first call or question.

As a first step, it was identified which features are available at which maturity stage from the raw dataset. Filtering out a total of 32 relevant variables for the analysis based on completeness, quality, redundancy caused by multicollinearity, correlation with the output label, and excluding all name-based variables, the following table gives an overview of selected datatypes (detailed data dictionary in Appendix 2).

Table 1: Overview of Features

Feature Type	Source	Features
Categorical	CF	Status, Division, Postcode, Last Activity Type
	PT	Division, Postcode, Profile, Mail Domain, Last Activity Type
Binary	CF	Status, Measure Type
	PT	Measure Type, Phone, Corporate Mail, Oeko Option, Invoice Option, Optimum Option, Repricing Option
Continous	CF	Consumption, Pod, Activity Types, Activity Count, Note
	PT	Consumption, Pod, Contract Length, Time to Contract, Activity Types, Activity Count, Note
DateTime	CF	Entry Date, Firstactivity Time, Lastactivity Time
	PT	Entry Date, Contract Start, Firstactivity Time, Lastactivity Time
Text	Both	Comment

CF = Contact Form, PT = Pricing Tool, Grey Features only available for Maturity B

Next, an exploratory analysis was conducted to understand the data better. The target variable is defined using the “status” of the lead. Status “won”, indicates that the lead has made a purchase and signed a contract with the company and is defined as the positive class. The negative class equals all leads with status "lost". This status is set when the lead either actively indicates that the company is longer interested in a sale or if the lead does not reply to three consecutive contact attempts over the course of three weeks. "Open" leads still in the sales funnel were not included in the data as lead generation was stopped for the prior weeks at the data extraction date, and the online forms were deactivated. A closer look at the target illustrates the high imbalance of the dataset. Out of all CF leads, only 3.1% were labelled as "won" in their final sales status. The conversion rate for the PT is slightly better as 5.1% were "won".

3.3 Data Preparation

The most important step of the process, according to Chapman et al. (2000), is data preparation. Therefore, special attention was given to this phase, and different pre-processing methods were applied and tested with different models. The techniques used were chosen based on business context and model improvement. All pre-processing steps were applied using standard Python libraries in Jupyter Notebook run on a cloud-based computing platform: Microsoft Azure Machine Learning Service (Azure ML). This framework was chosen as Azure readily integrates into the existing CRM system, allowing for more scalability and easy deployment of the model.

After an initial cleaning of features, dealing with missing values and outliers, new features were created based on the raw data to make specific patterns more visible, reduce dimensionality and ensure anonymity in the dataset (detailed feature generation outlined in Appendix 3). The text data, namely the comment field, was processed into a form which could be consumed for model training and inference purposes using natural language processing techniques like *term frequency-inverse document frequency* (TF-IDF) (Anandarajan, Hill and Nolan 2019, 45–73). Applied text mining techniques can be seen in Appendix 4. The final dataset, therefore, consisted of 8 variables for the CF, 19 for the PT, and an additional 13 behavioural features when looking at Maturity stage B.

Lastly, several resampling methods have been tested to deal with the imbalance in the dataset. For most imbalanced classification tasks, undersampling, oversampling, or a combination of both can be applied. Due to the small amount of data, undersampling was no viable option, as it would have led to a loss of information. Random Oversampling involves simple duplication of samples from the minority class, whereas SMOTE-ENC uses K nearest neighbor to create synthetic samples. Both methods have been applied with various sampling rates to create a more balanced dataset while still controlling bias (see Appendix 5).

3.4 Modelling

To simplify the development process and maximize the use of more complex algorithms, automated machine learning (Auto ML) implemented in Azure ML was used. Auto ML dynamically combines various data mining techniques to form an easy-to-use end-to-end pipeline and therefore provides a solution for building a complex deep learning system without extensive human expertise (He, Zhao, and Chu 2021).

The Auto ML process consists of four key steps: data guardrails check, intelligent feature engineering, iterative data transformation, and iterative model building (Sawyers 2021, 38–42). Throughout the first two steps, the type and format of each column were automatically detected, and the data was transformed accordingly. Appendix 6 shows the applied feature transformation steps. Before training the model, the data was scaled using one of several scaling methods. Auto ML iterates through different combinations of Scaler and Classifier to find the best-performing one. The core of the pipeline is iterative model building. Different ML algorithms for binary classification were selected to be tested based on the results of our literature review: Logistic

Regression (LR), Decision Tree (DT), Random Forest (RF), Extra Trees (EXT), XGBoost (XGB), Naïve Bayes (NB). Appendix 7 gives an overview of these algorithms.

The data was split into a train-validation-test split of 70%:15%:15% and hyperparameter tuning was done for each of the above classifiers as part of the Auto ML process to find the best-performing classifier for each maturity level and source. To validate the results two different data splits were tested. As they have shown stable results, only the bets split is shown.

3.5 Evaluation

An appropriate evaluation metric must be applied to define "best-performing" for the presented task. The most widely used measure for comparing the performance of classification models is accuracy (Berrar 2019). *Accuracy* measures the percentage of correct predictions across all predictions made; however, if used for highly imbalanced datasets, as given in this research, the larger class is favoured making the metric invaluable for the studied problem. Therefore, a closer look was taken at the *confusion matrix* and related measures in the explicit business context.

If the model is applied in business, a positive prediction (conversion = "yes") would lead to a prioritization of the lead. In contrast, a negative prediction would lead to a depreciation of the lead or no further contact. Keeping that in mind, the worst error for the model would be predicting class 0 when the lead is class 1, as this would, lead to the loss of a customer for the company. This error can be measured with *recall/ true-positive-rate (TPR):* fraction of true positive predictions from all positive cases.

The second metric in focus is *false positive rate (FPR*): the fraction of cases precited positive that was actually negative. The lower this rate, the more money and time could be saved by not following up on leads that will likely not convert.

The model should minimize both errors, which is why the area under the *receiver operating curve (AUC ROC)* presents itself as the best metric for the task. The ROC combines the TPR and the FPR by plotting performance at various thresholds and can be aggregated across all thresholds with the area under the curve. This single score can be interpreted as the model's ability to make good predictions independently from class imbalance. Going one step further, *weighted AUC* (WAUC) includes differing weights for the values of the TPR among regions. It, therefore, focuses more on the accuracy of the minority class (Dong et al. 2012). As the presented dataset was highly imbalanced and the focus was on the minority class, WAUC was therefore proposed as the best way to measure overall performance and select the final model for each maturity stage.

On top of performance, we also wanted to measure the business's actual value add when using the lead scoring model. The *expected value framework* (Fawcett and Provost 2013, 187–213) helps estimate each model's business impact, considering the costs and benefits for each type of error. The formula to calculate the expected profit for a model is:

$$Exp.\ profit = p(1) * [p(Y|1) * v(Y, 1) + p(N|1) * v(N, 1)] + p(0) * [p(N|0) * v(N, 0) + p(Y|0) * v(Y, 0)] \quad (1)$$

where $p(x)$ represents the probability of a lead being won (=1) or not won (=0), p(Y|x) the probability of predicting class 1 (=Y) or 0 (=N) when the true value is x and $v(x, y)$ represents the value (= benefit or cost) if this outcome happens. This means that the benefit or cost of each error are multiplied with the respective error rate from the confusion matrix and factor in the class imbalance to sum the expected profit of the model.

An estimation of costs and benefits for the different model errors is needed to apply this framework. As mentioned above, the worst possible error is false negative: the lead would not be called, and the profit from the sale would be lost. As energy and gas have no fixed prices, the profit of a sale was estimated by calculating the average profit based on the contribution margin of all contracts from converted leads in the given timeframe. On the other hand, the costs that can be saved when correctly predicting a lead that will not convert were estimated using the average sales costs for the lead, and the opportunity phase in the company Appendix 8 presents the calculations as well as the resulting cost-benefit-matrix. As the costs for the lead phase have already occurred when the model is only used at maturity level B, the benefit of a true negative is smaller in this later stage. Considering the asymmetry between the costs and benefits of correct and incorrect predictions, the expected value framework can be used to find the best threshold for a model at which a maximum profit is generated. By default, all models predict a lead to be won when the calculated probability for class 1 is greater than 0.5. Adapting this threshold leads to the model predicting more or less leads as class 1. A confusion matrix was generated for each adapted threshold and multiplied with the respective cost-benefit matrix to get the expected value for each threshold. This was plotted in a profit curve, and the maximum profit of the model was derived.

Finally, the best-performing models for each maturity stage and source were further evaluated by looking at model explanation. As pointed out by Bohanec, Robnik-Šikonja, and Kljajić-Borštnar (2017) interpretation of the model is a crucial factor in generating business value with ML. As part of the Auto ML package, trained models can be explained by looking at local and global feature importance using Mimic Light GBM explainer combined with SHAP tree explainer (Microsoft 2022).

4. Results and Discussion

4.1 Results of the exploratory analysis

Several issues were identified throughout the exploratory data analysis of the company's lead data. Similar to Espadinha-Cruz, Fernandes and Grilo's (2021) study, the need for a more transparent data-driven process was found, leading to problems with consistency in data generation across sources and maturity stages of the leads. For example, some leads within the dataset were classified as won, even though no contact with the lead is noted in the CRM system, reflecting the sales representatives' lack of diligence in adding information to the system. Also, a closer look at the different activity types (see Appendix 9) showed that each sales representative seems to have a different understanding of their definitions. Often calls are registered even though they were unsuccessful, and the lead did not pick up. Thus, an activity does not necessarily equal actual contact with the lead. In addition, the data collection was adapted over the past years, but it was never designed with the data science goal in mind (see conclusions). Therefore, much uninformative information is gathered, but essential data is missing. For example, within the dataset, there is no clear indicator to distinguish at which point in the sales funnel a lead was lost. This information would enable a prediction from maturity stage to maturity stage, which could be more precise.

Another interesting finding was that opposed to the results in previous research, the time to the first contact with a lead does not have a linear relationship to the conversion. As it can be seen from the data in table 2 the conversion rate was not always higher the faster a lead was contacted. Data from both sources showed the highest conversion rates if the lead was called within less than 10 minutes or more than 20 hours.

Time to First Act.	Won Leads CF	Won Leads PT
< 10 min	2.5%	6.5%
10 - 30 min	1.1%	5.2%
30 - 60 min	1.5%	4.6%
1 – 11 hours	2.5%	3.6%
11-20 hours	2.4%	3.5%
>20 hours	3.1%	6.0%

Table 2: Conversion grouped by time to first activity

Looking at the distributions and conversion rates for all variables (see Appendix 9), especially the entry date of the lead, the profile, the consumption measure, and activity information displayed clear patterns, which were overmined by the correlation to the target. A closer look at the entry date showed that there is a sales cycle throughout the year. Based on the differences in the patterns between leads from the CF and PT, it was decided to develop separate classifiers for the two channels, to achieve greater accuracy.

4.2 Results of the predictive analysis

Four Auto ML experiments were conducted as described in the methodology chapter: one per source and maturity stage. From here on, we refer to the models for the CF as model 1A (Maturity A) and 1B (Maturity B) and the PT as 2A and 2B. The best classifier was selected based on WAUC from each run, and key metrics were calculated. All models achieved WAUC scores greater than 0.6, showing that ML can accomplish good predictive performance with little information in the analyzed company.

Looking at the top ten classifiers and the respective hyperparameters tuned for each experiment (Appendix 10), the best-performing models were all tree-based, in line with the conclusion of the previous literature review. EXT and RF that did not use bootstrapping (=subsampling of the dataset) performed better because of the small size of the dataset. For the best models, the hyperparameter "class weight" was tuned to balance the weight, which could be one main difference to the other classifiers, as more importance is given to converted leads. Models with fewer trees (50 – 200) outperformed models with more trees. For all experiments except 2B, the XGB classifier had a low recall (< 0.25) and could not identify converted leads. RF and EXT are better at dealing with noise and less information available. As for experiment 2B, where more features are available and fewer comments, XGB outperformed the other classifiers. This was also the only experiment where the LR was in the top ten.

Any data balancing methods tested led to extreme overfitting of the models, meaning the models fit exactly against its training data and therefore did not generalize well. Though WAUC scores of the training/validation data increased compared to models without resampled data, the test scores dropped significantly. Even when creating synthetic samples using SMOTE-ENC and decreasing the sampling rate, the model was overfitting (see Appendix 11 for details). Therefore, no resampling was used within the final models.

Table 3 gives an overview of the best-performing algorithms as well as their key metrics for each experiment (detailed results for each model can be found in Appendix 12):

Table 3: Key Metrics of best models

Source	Metric	Maturity A	Maturity B
Contact Form (1)	Best Classifier	Random Forest	ExtraTree
	Recall	0.69	0.56
	FPR	0.35	0.23
	WAUC	**0.792**	**0.804**
Pricing Tool (2)	Best Classifier	ExtaTree	XGBoost
	Recall	0.41	0.81
	FPR	0.18	0.01
	WAUC	**0.612**	**0.975**

When comparing the two maturity stages, models from both sources presented better WAUC when the activity information was included (Maturity B). These results further support the idea that the more mature a lead is in its journey, the better the prediction of its probability to convert. Only a slight increase of 0.012 in WAUC was achieved for CF models, whereas the increase for the PT was 0.363. Concluding, the activity information had more predictive power for model 2 than model 1. Looking at the recall for model 1, there was a decrease in the score, meaning that the model in maturity stage B was worse in discovering converted leads. However, with the added information it was better in finding leads who will not convert. For model 2, Recall as well as FPR improved in Maturity B, leading to a very high-performance model. For the given test set, model 2B achieved a WAUC of 0.975 and an accuracy score of 0.98, only mispredicting 12 out of 659 leads.

Although the PT includes more variables and details about the leads, surprisingly, model 2A showed the lowest performance, having had 0.18 worse WAUC than model 1A. Only 41% of all converted leads were correctly identified. The ROC Curves in figure 3 on the left show the differences between the models.

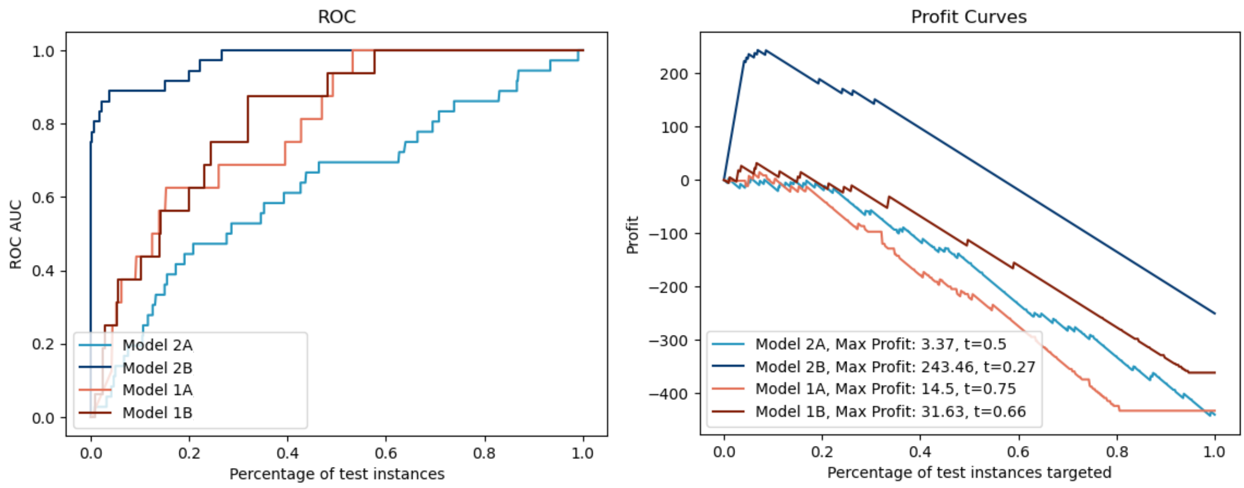


Figure 3: ROC and Profit Curves

4.3 Expected value for the models

The above results are purely based on the models' performance metrics with the default threshold of 0.5. However, if financial values are assigned to the model as described in chapter 3 the threshold and value of each model changes. The profit curves for each model resulting from the application of the Expected Value Framework are illustrated in figure 3.

As expected, the XGB model 2B had the highest value added for the business as it far exceeded the other models in accuracy for both classes. The highest profit was achieved when targeting only 6.9% of test leads, meaning they would not be contacted. Interestingly, the threshold for the model was very low at t = 0.27. In this case a lead should not be contacted whenever the likelihood of not converting, predicted by the model, is 73% or more. Deploying this model would lead to a profit of 243.46€ per lead, resulting in a total profit of 32758€ per year based on the historical lead amount per year.

As visible in the graph, all models achieved their maximum profit at 5.5 – 7.5% of the test instances targeted, which can be explained by the much higher costs of losing a potential customer compared to the sales costs associated. The difference in profit can be justified by the lower accuracy and, therefore, the higher cost of potential errors of the other models. At nearly the same percentage of test instances targeted, the threshold varied significantly between the models. For example, using model 2B, the likelihood of not converting was 73% at this point, whereas using model 1B, it was only 34%.

The maximum profits that can be achieved with the CF models are 15€ and 32€ per lead, which results in a yearly profit of 2036€ (model 2A) and 4071€ (model 2B) respectively. Model 1A, however, can only generate 361€ per year. Depending on the model's deployment costs, these models might not add a monetary value to the business with the process as-is. The only model that truly generated a significant value add is model 2B.

4.4 Feature Importance

Further analysis was performed to understand the differences among the four models checking which features drive the different results. Figure 4 gives an overview of global feature importance for all features with a value of < 0.15 for at least one model. It is apparent from this figure that the importance of the features differs a lot between the different models. Models 1A and 1B showed a similar trend that start_date, postcode and comment have high importance. However, the activity information in model B only has a negligible impact on the model, which explains that the increase in model performance with the increase in maturity is limited. In contrast, especially the activity “others”, which is often used for further phone calls and if a customer has reached out with questions, and the note length from the activity had a high feature importance for model 2B. Interestingly, no significant importance for any feature was reported for model 2A.

	1A	1B	2A	2B
start_date	**1.7620**	**1.4280**	0.0003	0.3310
postcode	0.3430	0.2580	0.0004	0.1100
consumption	0.0050	0.0170	0.0000	0.1890
comment	0.2990	0.2890	0.0030	0.0650
time_to_lastactivity		0.1520		0.0790
last_activity_type		0.1930		0.1920
act_others		0.0560		0.3640
act_questcust		0.0006		0.1550
note_length_avrg		0.0880		**0.3630**

Figure 4: Feature Importance for top features

5. Conclusion and Recommendations

Predictive lead scoring and qualification is a relevant topic for research and business. It shows tremendous potential for improving profits, especially in B2B, where customer acquisition is often lengthy and costly. This research investigated algorithms for lead scoring in the early maturity lead phase.

To answer the research question, a literature review in the area of lead management was performed to understand the sales funnel process and the benefits ML can have in this process. Automated lead scoring provides data-driven and more accurate scores faster and generates more insights. The examination of related work led to the conclusion that ML, especially more complex tree-based classifiers, has shown good performance in past studies. Especially behavioural data in the prospect phase and demographic data in the opportunity phase led to good predictions combining features from many sources. The present work aims to contribute by analyzing a model in two maturity stages within the lead phase that only uses a small amount of internally available features.

With the use of Auto ML, several classifiers were tuned and compared for leads from two different online CF from a German energy business. The results show that although preparing and preprocessing of in particular activity data is a challenge, one can obtain good classification performance in the lead phase with little information (WAUC over 0.6). The challenges faced are rooted in the lack of a clear strategy-driven process for data gathering. Additionally, the preeminence of tree-based models for lead scoring is supported by the findings.

Comparing the best models for an early lead maturity stage before any interaction, and a later maturity stage, revealed that, although more mature models are superior, the inclusion of the activity data increases predictive performance to a varying degree, depending on the available demographic features. In general, all features differed in their importance for the models, and no clear pattern appeared, pointing out that the company misses covering some essential topics in their current CRM process.

One of the more significant findings to emerge from this study is that even though all models achieve good performance in standard evaluation metrics when factoring in the benefits and costs for model errors, not all models actually add value to the business. The application of the expected value framework revealed that only the model that included more demographic information as well as the activity data and achieved an outstanding WAUC > 0.9 seems to create a significant financial value.

Taken together, the study makes two contributions to the current body of literature: it is amongst the first investigating ML in the early lead phase, and it enhances the understanding of the monetary value that can be derived from automated lead scoring going beyond standard evaluation.

The findings also suggest several courses of action for the company in focus. A key priority should be adapting the information gathered in the CRM system, simplifying how interactions are recorded and focusing on consistently tracking features important for prediction. To generate a long-term profit, ML should be used as support for lead prioritization based on the predicted probability of conversion and continuously iterated as new information is gathered. With this, the impact of changes made to the process can be directly measured. Finally, the data-driven approach presented in this work should be leveraged in combination with human assessments to achieve strong results in the current process.

6. Limitations and Further Research

While the results mentioned above give relevant insights into gaps in the literature and recommendations for practice, the study offers opportunities for further research as it is affected by some limitations. First, the presented models are bespoke to the information available from the specific contact forms for the company studied and, therefore, cannot be generalized. The commodity business provides a unique market resulting in the need for further tests of the concept in other companies with similar multifaceted sales pipelines to prove its usability and suitability in other companies and industries.

Second, the analysis was performed on a relatively small, highly imbalanced dataset. Integrating more data as new leads enter the system could increase the accuracy. As the tested resampling methods did not increase model performance but led to overfitting, other more complex resampling methods should be tried to address this issue. One approach that could be investigated is One-Class-Classification, defining the minority class as outliers.

Furthermore, although the results were stable checking on a second train-validation-test split, they could be influenced by the data split. A larger training dataset as well as cross-validation should be used to confirm the results.

The detected lack of a transparent and data-driven process to gather information in the CRM system presents another limitation and opportunity for further improvement. Using the insights from this study, the process should be adapted to gather relevant information more consistently. The models could be re-trained to check real-time improvement of performance.

Besides, the model could include more features to increase predictive performance. One opportunity would be to add external data to the model. Some studies have shown good performance using web scraping to get additional demographic information about the leads. Also, we have seen that the set-up of the contact forms greatly influences the output when comparing PT against CF. Though keeping in mind user-friendliness, forms could be adapted to get relevant information directly from the user. Further research should be conducted, including A/B tests on the best set-up of contact forms for lead management.

References

Anandarajan, Murugan, Hill, Chelsey, and Nolan, Thomas. 2019. *Practical Text Analytics: Maximizing the Value of Text Data.* 1st ed. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-95663-3

Benhaddou, Youssef, and Leray, Philippe. 2017. "Customer Relationship Management and Small Data — Application of Bayesian Network Elicitation Techniques for Building a Lead Scoring Model." *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA):* 251–55. https://doi.org/10.1109/AICCSA.2017.51

Berrar, Daniel. 2019. "Performance Measures for Binary Classification." *Encyclopedia of Bioinformatics and Computational Biology* 1: 546–60. https://doi.org/10.1016/B978-0-12-809633-8.20351-8.

Bohanec, Marko, Robnik-Šikonja, Marko, and Kljajić Borštnar, Mirjana. 2017. "Organizational Learning Supported by Machine Learning Models Coupled with General Explanation Methods: A Case of B2B Sales Forecasting." *Organizacija* 50 (3): 217–33. https://doi.org/10.1515/orga-2017-0020.

Buttle, Francis. 2012. *Customer Relationship Management: Concepts and Technologies.* 2nd ed. London, New York: Routledge.

Chapman, Peter, Clinton, Janet, Kerber, Randy, Khabaza, Tom, Reinartz, Thomas P., Shearer, Colin, and Wirth Richard. 2000. *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. S. l.: CRISP-DM consortium.

Coe, John M. 2004. "The Integration of Direct Marketing and Field Sales to Form a New B2B Sales Coverage Model." *Journal of Interactive Marketing* 18 (2): 62–74. https://doi.org/10.1002/dir.20005.

D'Haen, J., Van den Poel, D., Thorleuchter, D., and Benoit, D. F. 2016. "Integrating Expert Knowledge and Multilingual Web Crawling Data in a Lead Qualification System." *Decision Support Systems* 82:69–78. https://doi.org/10.1016/j.dss.2015.12.002.

D'Haen, Jeroen, and Van den Poel, Dirk. 2013. "Model-Supported Business-to-Business Prospect Prediction Based on an Iterative Customer Acquisition Framework." *Industrial Marketing Management* 42 (4): 544–51. https://doi.org/10.1016/j.indmarman.2013.03.006.

D'Haen, Jeroen, Van den Poel, Dirk and Thorleuchter, Dirk. 2013. "Predicting Customer Profitability During Acquisition: Finding the Optimal Combination of Data Source and Data Mining Technique." *Expert Systems with Applications* 40 (6): 2007–12. https://doi.org/10.1016/j.eswa.2012.10.023.

Dong, Y., Li, X., Li, J., and Zhao, H. 2012. "Analysis on Weighted AUC for Imbalanced Data Learning Through Isometrics." *Journal of Computational Information Systems* 8 (1): 371–78.

Duncan, Brendan Andrew, and Elkan, Charles Peter. 2015. "Probabilistic Modeling of a Sales Funnel to Prioritize Leads." *KDD '15: Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1751–58. https://doi.org/10.1145/2783258.2788578.

Eitle, Verena, and Buxmann, Peter. 2019. "Business Analytics for Sales Pipeline Management in the Software Industry: A Machine Learning Perspective." *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 1013–22. http://dx.doi.org/10.24251/HICSS.2019.125

Espadinha-Cruz, P., Fernandes, A., and Grilo, A. 2021. "Lead Management Optimization Using Data Mining: A Case in the Telecommunications Sector." *Computers & Industrial Engineering* 154 (4): 107122. https://doi.org/10.1016/j.cie.2021.107122.

Etminan, Ali. 2022. "Prediction of Lead Conversion with imbalanced data: a method based on predictive lead scoring." Master's thesis, Linköping University. http://liu.diva-portal.org/smash/record.jsf?pid=diva2%3A1566623

Fawcett, Tom, and Provost, Foster. 2013. *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking.* Beijing, O'Reilly.

Geigenmüller, Anja, and Kuhn, Elisabeth. 2022. "De-Commoditization in B2B Markets: A Communication Perspective." In *Commodity Marketing*, edited by Margit Enke, Anja Geigenmüller, and Alexander Leischnig, 413–25. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-90657-3

Han, Jiawei, Pei, Jian, and Kamber, Micheline. 2012. *Data Mining: Concepts and Techniques.* 3rd ed. Amsterdam: Morgan Kaufmann; Elsevier.

He, Xin, Zhao, Kaiyong, and Chu, Xiaowen. 2021. "AutoML: A Survey of the state-of-the-art." *Knowledge-Based Systems* 212: 106622. https://doi.org/10.1016/j.knosys.2020.106622.

Heddam, Salim. 2021. "Extremely Randomized Tree: A New Machines Learning Method for Predicting Coagulant Dosage in Drinking Water Treatment Plant." *Water Engineering Modelling and Mathematic Tools* 2021, 475–89. https://doi.org/10.1016/B978-0-12-820644-7.00013-X.

Jadli, Aissam, Hamim, Mohammed, Hain, Mustapha, and Hasbaou, Anouar. 2022. "Toward a smart lead scoring system using machine learning." *INDJCSE* 13 (2): 433–43. https://doi.org/10.21817/indjcse/2022/v13i2/221302098.

Karim, Masud, and Rahman, Rashedur. 2013. "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing." *Journal of*

Software Engineering and Applications 6 (4): 196–206. http://dx.doi.org/10.4236/jsea.2013.64025

Kuenzel, Mauela and Lohse, Lutz. 2022. “CRM in the Energy Market: Customer Relationship Management in Commodity Industries Using the Example of an Energy Service Provider.” In *Commodity Marketing*, edited by Margit Enke, Anja Geigenmüller, and Alexander Leischnig, 365-390. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-90657-3

Kuhn, Max, and Johnson, Kjell. 2016. *Applied Predictive Modeling.* 5th ed. New York: Springer.

Lambert, Marina. 2018. “Sales Forecasting: Machine Learning Solution to B2B Sales Opportunity Win-Propensity Computation.” Msc Research Project, National College of Ireland. https://norma.ncirl.ie/3437/

Lewis, David. 2012. *Manufacturing Demand: The Principles of Successful Lead Management.* Danville, CA: New Year Publishing.

Mezei, Jozsef. 2020. “Automating Lead Scoring with Machine Learning: An Experimental Study.” *Proceedings of the 53rd Hawaii International Conference on System Sciences 2020*, 1439-1448. https://hdl.handle.net/10125/63916

Microsoft. 2022. “Azure Documentation: Model Interpretability.” Azure Documentation. Accessed December 01, 2022. https://learn.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability.

Mitchell, Thomas M. 1997. *Machine Learning.* Singapore: McGraw-Hill.

Monat, Jamie P. 2011. “Industrial Sales Lead Conversion Modeling.” *Marketing Intelligence & Planning* 29 (2): 178–94. https://doi.org/10.1108/02634501111117610.

Mukherjee, Mimi, and Khushi, Matloob. 2021. "SMOTE-ENC: A Novel SMOTE-Based Method to Generate Synthetic Data for Nominal and Continuous Features." *Applied System Innovation* 4 (1): 18. https://doi.org/10.3390/asi4010018

Ngai, E.W.T., Xiu, Li, and Chau, D.C.K. 2009. "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification." *Expert Systems with Applications* 36 (2): 2592–2602. https://doi.org/10.1016/j.eswa.2008.02.021.

Ohiomah, A., Benyoucef, M., Andreev, P. 2016. "Driving Inside Sales Performance with Lead Management Systems: A Conceptual Model." *Journal of Information Systems Applied Research 9 (1):* 4–15. http://jisar.org/2016-9/

Rezazadeh, Alireza. 2020. "A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach." *Forecasting* 2 (3): 267–83. https://doi.org/10.3390/forecast2030015.

Salesforce. 2018. "2018 State of Sales Report." State of Sales. Accessed December 01, 2022. https://www.salesforce.com/content/dam/web/en_au/www/documents/research/salesforce-state-of-sales-4th-ed.pdf.

Sawyers, Dennis. 2021. *Automated Machine Learning with Microsoft Azure.* 1st ed. Packt Publishing.

Stevens, Ruth P. 2012. *Maximizing Lead Generation: The Complete Guide for B2B Marketers.* Indianapolis, Pearson Education.

Trailer, B. 2006. "Understanding What Your Sales Manager Is up Against." *Harvard Business Review* 84 (7): 48–55.

Wenger, Stephan. 2021. “Successful Lead Management: Nothing’s Gonna Stop Us Now.” In *B2B Marketing*, edited by Uwe G. Seebacher, 247–69. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-54292-4

Williams, David S. 2014. *Connected CRM: Implementing a Data-Driven, Customer-Centric Business Strategy*. 1st ed. New Jersey, John Wiley & Sons.

Yan, Junchi, Zhang, Chao, Zha, Hongyuan, Gong, Min, Sun, Changhua, Huang, Jin, Chu, Stephen, and Yang, Xiaokang. 2015. “On Machine Learning Towards Predictive Sales Pipeline Analytics.” *Twenty-Ninth AAAI Conference on Artificial Intelligence* 29 (1): 1945-1951. https://doi.org/10.1609/aaai.v29i1.9455

Appendix

List of Figures

List of Tables

List of Abbreviations

AUC	Area-under-the-curve
B2B	Business to Business
CRM	Customer Relationship Management
CF	Contact Form
ML	Machine Learning
PT	Pricing Tool
ROC	Receiver operations curve
WAUC	Weighted area under curve

Appendix 1: Applied CRISP-DM framework

The methodological approach of this research is based on the Cross Industry Standard Process for Data Mining (Chapman et al. 2000).

Figure 5: CRISP-DM Framework (based on Chapman et al. 2000)

Appendix 2: Data Dictionary

#	Feature	Type	Description	Included in Model			
				1A	1B	2A	2B
1	division	Binary	Division of lead (0 = Gas, 1 = Electricity)	Y	Y	Y	Y
2	start_date	DateTime	Date and time when lead filled in form	Y	Y	Y	Y
3	postcode	Categorical	Postcode area of lead (regions from 0-10)	Y	Y	Y	Y
4	comment_length	Numerical	Number of characters excluding whitespace that was filled in comment field	Y	Y	Y	Y
5	free_text	Text	Comment, pre-processed as described in Appendix C	Y	Y	Y	Y
6	consumption	Numerical	Excepted yearly consumption of gas/electricity	Y	Y	Y	Y
7	pod	Numerical	Expected amount of power measure points needed	Y	Y	Y	Y
8	cons_slp	Binary	Indicator of power measurement type (0 = RLM, 1= SLP)	Y	Y	Y	Y
9	profile	Categorical	"Standardlastprofil" of lead = demand profile for electricity / gas depended on business model			Y	Y
10	phone	Binary	Indicator if phone number field was filled in (=1) or not (=0)			Y	Y
11	corporate_mail	Binary	Indicator if mail domain is corporate (=1) or free domain (=0)			Y	Y
12	mail_domain_cat	Categorical	Free mail provider domain			Y	Y
13	oeko_option	Binary	Indicator if oeko option is chosen			Y	Y
14	paper_invoice	Binary	Indicator if paper invoice option is chosen			Y	Y
15	optimum	Binary	Indicator if optimum option is chosen			Y	Y
16	repricing	Binary	Indicator if repricing option is chosen			Y	Y
17	months_contract	Numerical	Length of contract requested in months			Y	Y
18	time_to_contract	Numerical	Time from start_date to contract in months			Y	Y
19	act_1call	Numerical	Number of times called first		Y		Y
20	act_2call	Numerical	Number of times called second		Y		Y
21	act_auth_fail	Numerical	Number of times authentication failed		Y		Y
22	act_followup	Numerical	Number of times a follow up interaction happened		Y		Y
23	act_others	Numerical	Number of times any other interaction happened		Y		Y
24	act_quest	Numerical	Number of times lead was called because of questions		Y		Y
25	act_questcust	Numerical	Number of times lead called company because of questions		Y		Y
26	act_update	Numerical	Number of times information was updated in CRM		Y		Y
27	activity_count	Numerical	Total sum of all activities		Y		Y
28	last_activity_type	Categorical	Type of last activity		Y		Y
29	time_to_ firstactivity	Categorical	Time Difference between Start Date and first activity		Y		Y
30	time_to_ lastactivity	Categorical	Time Difference between Start Date and last activity		Y		Y
31	note_length_avrg	Numerical	Average count of characters of notes made in CRM system		Y		Y
32	target	Binary	Conversion of Customer (0 = not converted, 1 = converted)	Y	Y	Y	Y

Table 4: Data Dictionary

Appendix 3: Detailed Feature Generation

After an initial selection of variables, ignoring features with more than 20% **missing values**, only a few missing variables were left. If these were missing, because of the nature of the process, for example, because no interaction has taken place, the respective value was set to 0. Features with low missing value ratio based on sales representative forgotten input were replaced with adequate values (mean for numeric, mode for categorical). Checking for **outliers** looking at standard deviation from the mean resulted in the deletion of 15 records in total, mainly based on outliers in the interaction data.

As a next step, new features were created based on the raw data to make certain patterns more visible, reduce dimensionality and ensure anonymity in the dataset:

- The columns **"mail"** and **"phone"** were transformed into binary features, indicating whether or not the information has been provided. In addition, the mail domain was extracted, and a binary feature was created if the domain is a corporate or a free domain, as well as a categorical feature listing the main free domains.
- The **datetime information** was used to calculate times in between certain milestones, rather than only looking at a specific date: time from start date (=when contact form has been used) to energy contract start, time from start to first interaction and time from start to last interaction. These time difference values were then aggregated into bins for interpretation purposes, for example "less than 10min", "10min to 1hour" etc.
- The **start_date** is used to derive the following features: month, day of year, day of month, week, weekday, daytime, and hour
- Following a similar procedure, the **"postcode"** variable was aggregated, resulting in a categorical variable with 10 different regions.

- **Consumption information** was summarized in one feature showing the amount of expected consumption and a binary variable indicating the power measurement method.
- The **target** is defined using the "status" of the lead. Status "won", indicates that the lead has made a purchase and signed a contract with the company and is defined as the positive class. The negative class equals all leads with status "lost". This status is set when the lead either actively indicates that the company is longer interested in a sale, or if the lead does not reply to three consecutive contact attempts over the course of three weeks. "Open" leads still in the sales funnel where not included in the data as at the date of data extraction lead generation was stopped for the prior weeks and the online forms were deactivated.
- Comment is transformed applying text mining techniques (see Appendix 4)

Lastly, a correlation matrix is generated, and features with high multicollinearity are excluded, keeping the features with higher correlation with the target.

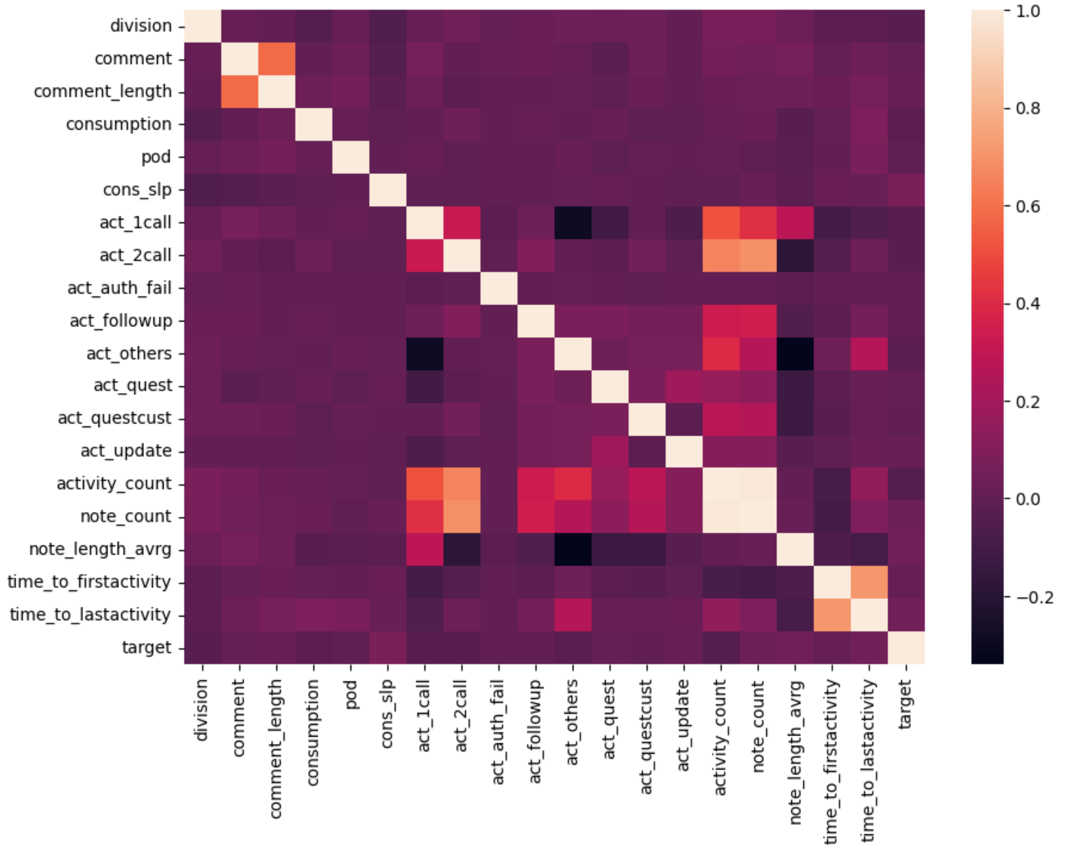


Figure 6: Correlation Matrix Contact Form

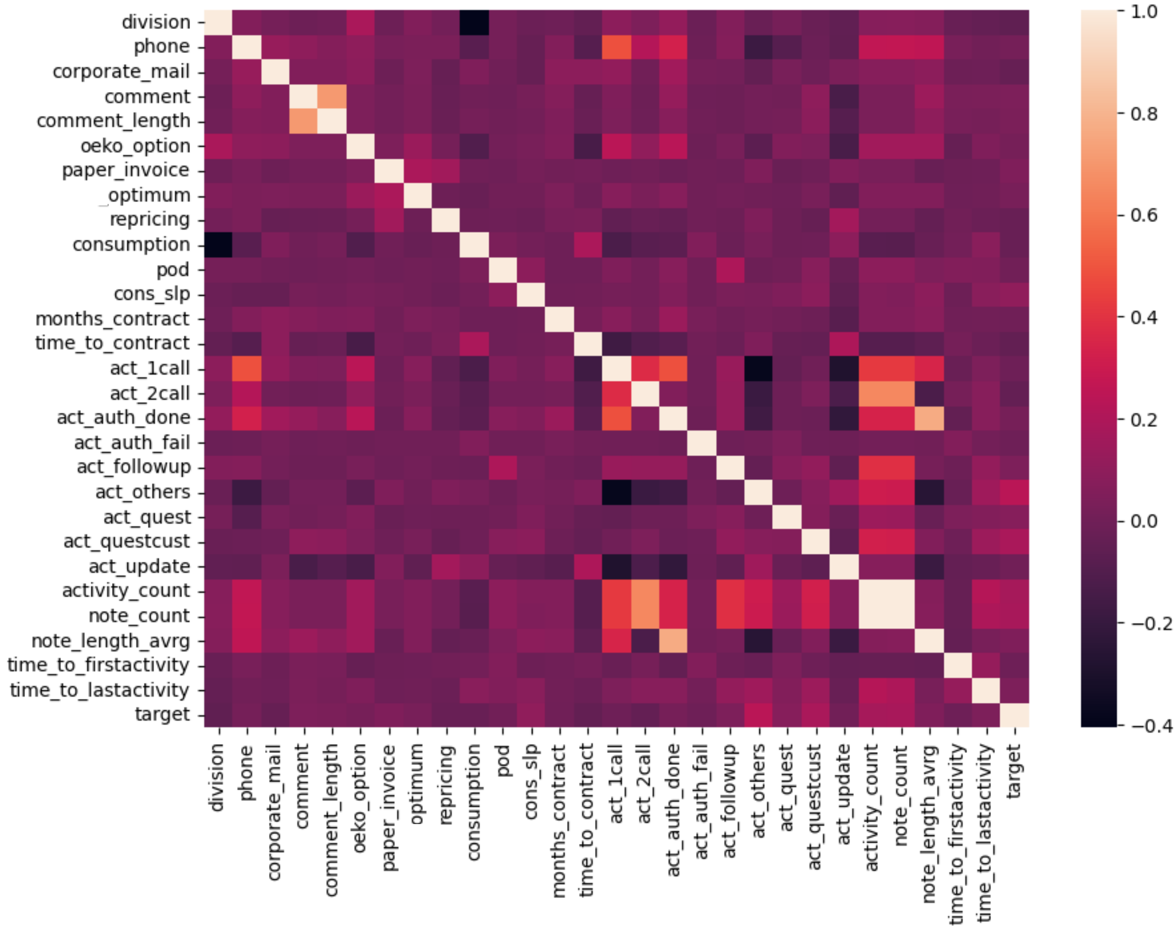


Figure 7: Correlation Matrix Pricing Tool

Appendix 4: Applied Text Mining Techniques

To be able to include the comment field in the analysis, which is a collection of unstructured data, the application of several **text mining technique** was required:

1. **Language Detection**

   Extraction of few Non-German comments using Language Modul from SpaCy (Python Software Foundation 2022)

2. **Anonymization of Comments for GDPR purposes**

   Application of own class developed using regex that anonymizes names, email addresses, website names, phone numbers and address information in text

3. **Text Cleaning**

   Replacement of symbols, dates, numbers, umlaute, punctuation, abbrevations, greetings and additional whitespace

4. **Stop Word Removal**

   Removal of stopwords with common lists

5. **Stemming/Lemmatization**

   Both approaches tested

6. **Tokenization and encoding using TFIDF**

   Application of term frequency inverse document frequency using sk-learn tfidf-vectorizer (Anandarajan 2019)

Figure 8: Text Mining Process Overview based on (Anandarajan 2019)

Appendix 5: Sampling

1. **Random Oversampling** (Etminan 2022): This technique is based on simple random duplication of samples from the minority class. The minority data points are sampled with replacement and duplicated. As the newly generated samples overlap with the original samples, it is known to cause overfitting.

2. **SMOTE-ENC** (Mukherjee and Khushi 2021): Synthetic Minority Oversampling technique uses K nearest neighbours of the minority sample to generate synthetic samples instead of simply duplicating the data points. SMOTE-ENC is specifically designed for both nominal and continuous features. Labels of a categorical variable are represented by their affinity of association with minority class target. For each minority class datapoint x, the k nearest neighbours of x are obtained using the Euclidean distance. According to the sampling rate for each x examples are randomly selected from the neighbours. Finally, for each of these neighbours this formula is used to generate new samples:

$$\$x' = x + rand(0,1) * |x - x\_k| \qquad (2)$$

$rand(0,1) =$ random number between 0 and 1, x_k = neighbor of x

Sampling Rate: For both techniques two different sampling rates were tested. The sampling rate is implemented with the parameter *sampling_strategy* in imblearn.

Sampling Rate 0,5: Original Target {0:3306, 1:106}
Resampled Target Distribution {0:3306, 1:1653}

Sampling Rate 0,2: Original Target {0:3306, 1:106}
Resampled Target Distribution {0:3306, 1:661}

Appendix 6: Feature Transformation in Azure Auto ML

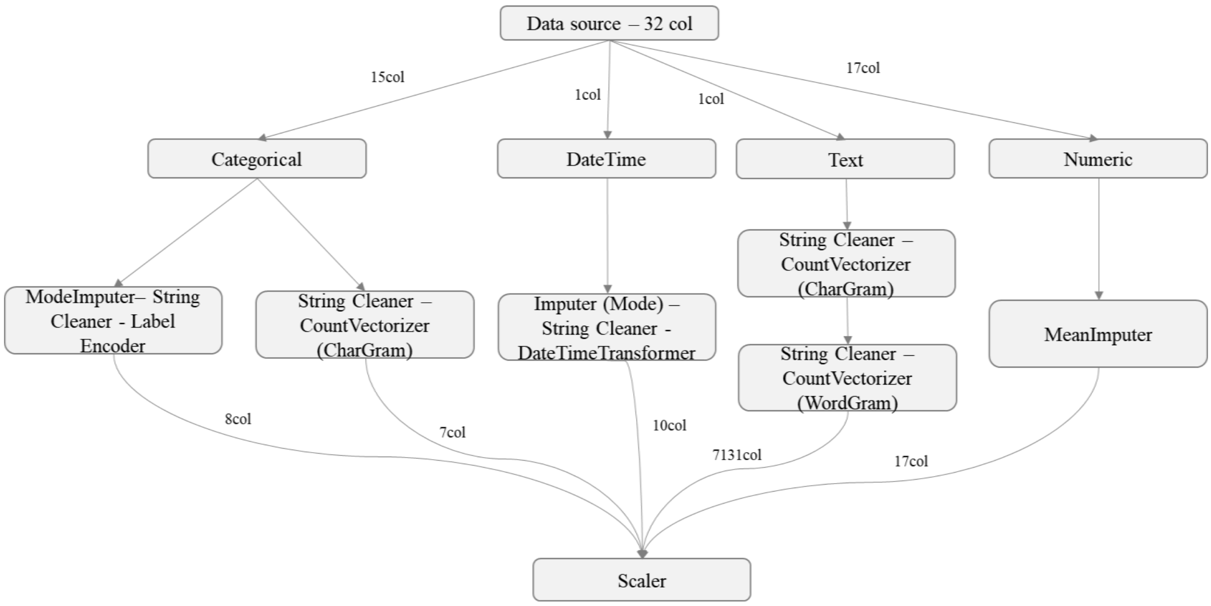


Figure 9: Feature Transformation in Azure

Appendix 7: Overview of tested algorithms

Classifier	Explanation
Logistic Regression (LR)	Applied as a baseline classification model. A class of generalized linear models which is widely used (M. Kuhn and Johnson 2016, 282–86).
Decision Tree (DT)	Data is split according to a certain parameter, like combining several if-statements. Very popular classifier as DTs are good with missing values and errors, simple yet powerful classifiers, and can be explained and visualized (Mitchell 1997, 52–62).
Random Forest (RF)	Ensemble Method based on bagging. Multiple decision trees are trained over different subsets of the training data, and the optimal split is chosen at each node. RF works well with large datasets and different data types but is complex and needs a high computational time, also the sampling of subsets may introduce bias. RF was the best performing classifier for Lead Scoring in most previous studies (M. Kuhn and Johnson 2016, 386–89).
Extra Trees (EXT)	Very similar to RandomForest however, trees are trained over the entire dataset, using different subsets of features instead of rows. Another difference is that the nodes are split randomly, making training faster and introducing less bias (Salim Heddam).
XGBoost (XGB)	Tree based ensemble method based on boosting. Included to test the performance of gradient boosting although better suited for large training sets and not ideal for Natural Language Processing problems (HAN, PEI, and KAMBER 2012, 380–82).
Naïve Bayes (NB)	A simple algorithm that utilized Bayes' rule. Chosen as NB is easy to implement, does not require much training data, and handles continuous and discrete data both well. Also, NB is highly scalable and not sensitive to irrelevant features, which is why it is often used for text classification problems. However, it assumes independence between features which is not fully given in the analyzed dataset. (M. Karim 2013).

Table 5: Explanation of algorithms

Appendix 8: Expected Value Framework

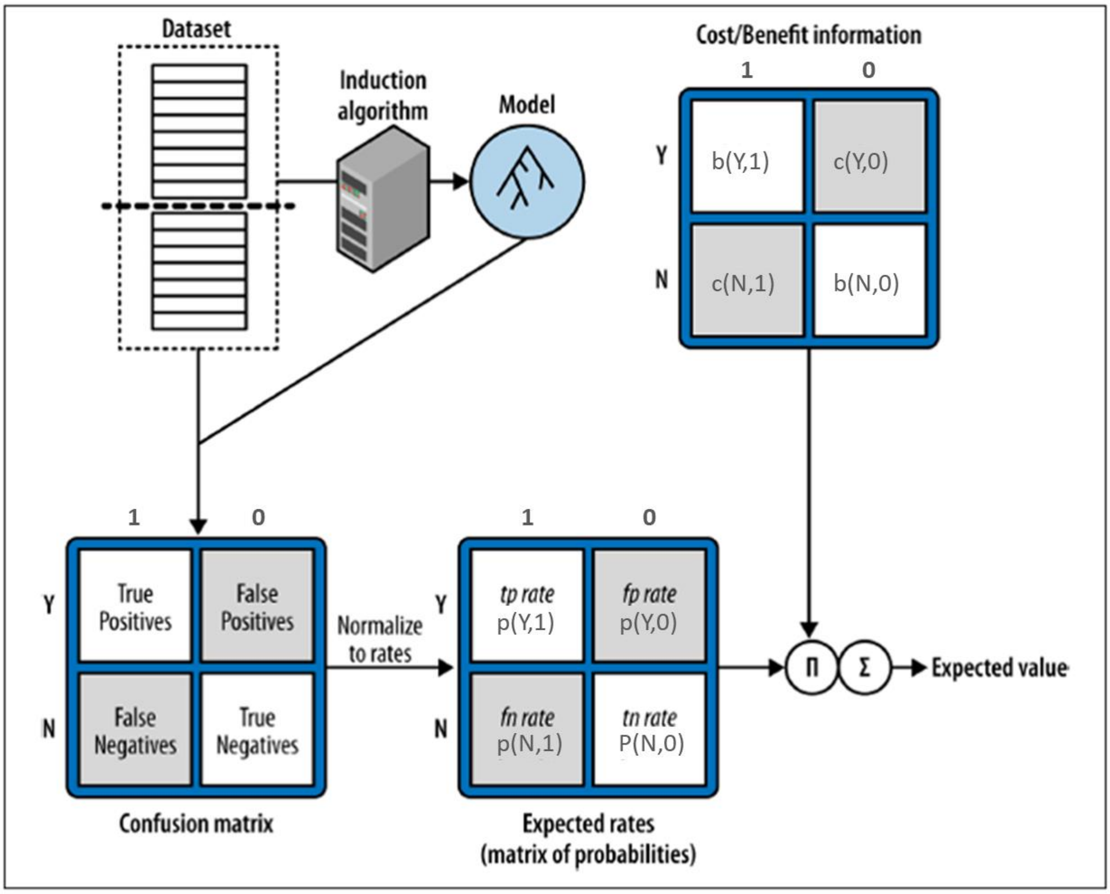


Figure 10: Expected Value Framework (own elaboration based on Fawcett and Provost 2013, 187–213)

Benefits and costs are estimations calculated based on company internal information. Due to confidentiality only the methodology how the final values were calculated can be described.

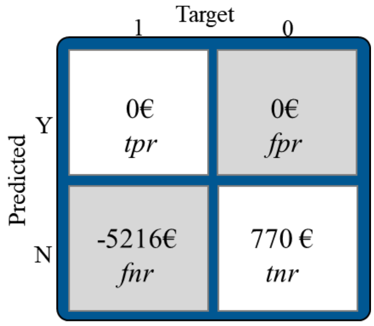
$$\begin{aligned} &\textit{Benefit} &&= \textit{Average Contribution Margin 1 of converted leads from dataset} \\ & &&= 5986€ \end{aligned} \tag{3}$$

$$\begin{aligned} &\textit{Costs Maturity A} &&= \textit{Salary for FTE in B2B-Qualityteam} * \textit{average hours invested in lead} \\ & &&= 190€ \end{aligned} \tag{4}$$

$$\begin{aligned} &\textit{Costs Maturity B} &&= \textit{Average cost to acquire weighted by sales channel} \\ & &&= 580€ \end{aligned} \tag{5}$$

The base scenario defined is that all leads are contacted, and the model is used to determine which leads do not need to be contacted. The ideal threshold is one where maximum profit can be achieved by targeting (not contacting) leads with low probability of conversion.

The cost-benefit-matrix is therefore calculated:



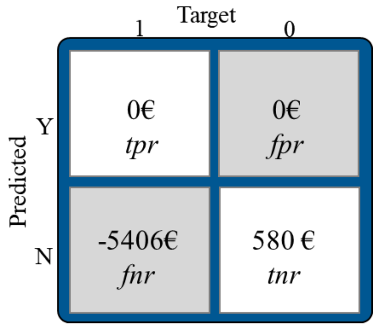
Maturity A:

b(Y,1) = 0

c(Y,0) = 0

c(N,1) = profit – costs = - 5986€ - (190€ + 580€) = - 5216€

b(N,0) = costs = - (190€ + 580€) = 770€



Maturity B:

b(Y,1) = 0

c(Y,0) = 0

c(N,1) = profit – costs = - 5986€ - (580€) = - 5406€

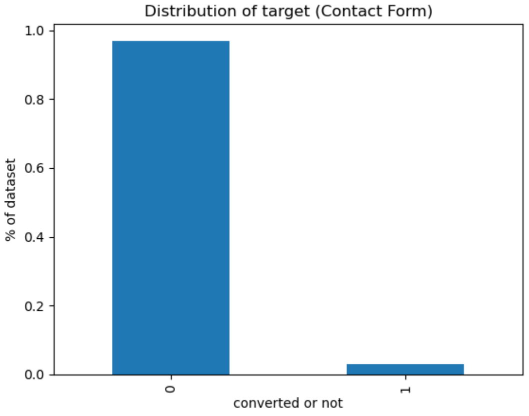
b(N,0) = costs = - 580€

Figure 11: Cost Benefit Matrices

When the model predicts that a lead will convert *Y*, the lead will be contacted. Nothing changes compared to the base scenario, so the cost and profit is 0. When the model predicts that a lead will not convert *N*, the lead will not be contacted. If this prediction is true *(N,0)*, the respective sales costs were saved and are therefore a benefit of the model usage. If the prediction is however wrong *(N,1)*, the money from the sale is lost, resulting in high costs.

Appendix 9: Results of the exploratory analysis

Figure 12: Distribution of target



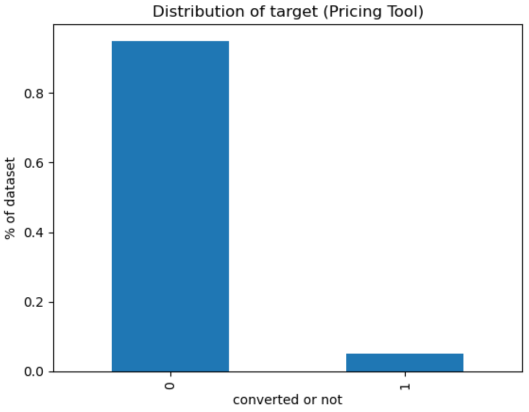


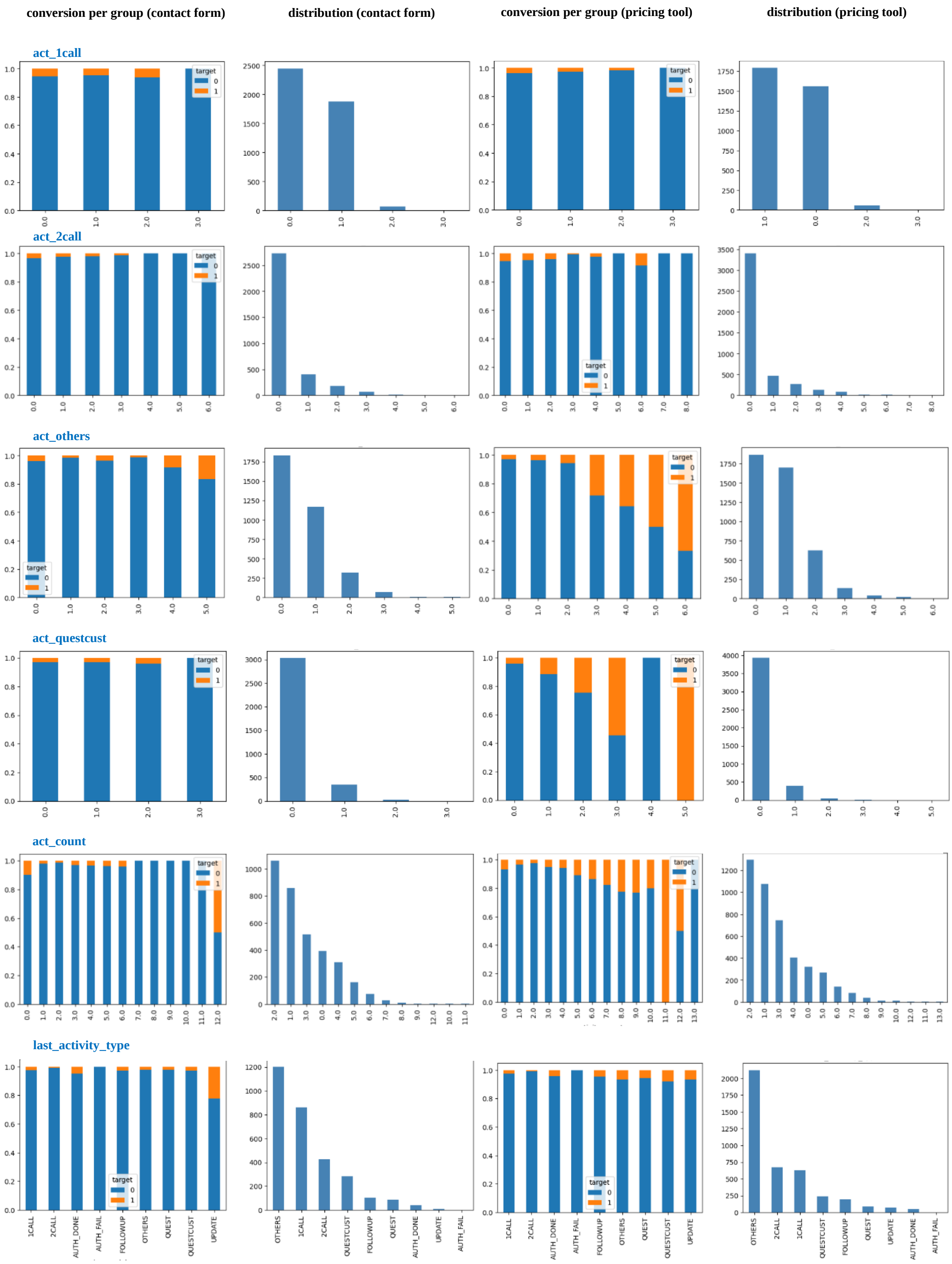
Figure 13: Descriptive statistics (Contact Form)

	consumption	pod	activity_count	note_length_avrg	comment_length
count	3412.000	3412.000	3412.000	3019.000	3412.000
mean	5786448.725	13.253	2.152	112.388	145.536
std	44552324.861	172.417	1.554	72.175	233.391
min	0.000	0.000	0.000	0.000	0.000
25%	131750.000	1.000	1.000	44.000	0.000
50%	600000.000	1.000	2.000	113.750	32.500
75%	2500000.000	2.000	3.000	174.417	215.250
max	1377000000.000	8400.000	12.000	330.000	2237.000

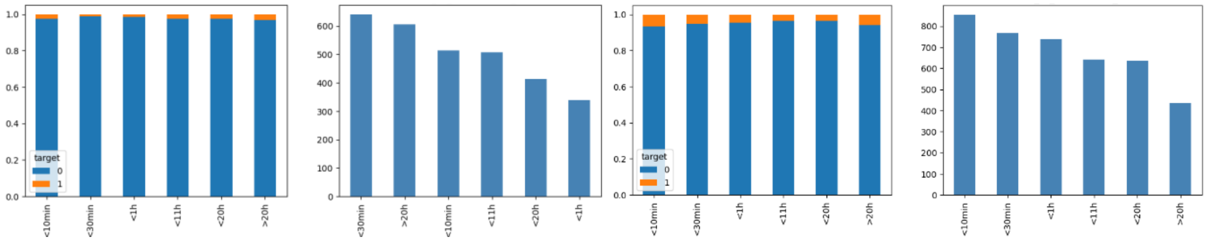
Figure 14: Descriptive statistics (Pricing Tool)

	consumption	pod	activity_count	note_length_avrg	comment_length
count	4391.000	4391.000	4391.000	4072.000	4391.000
mean	492231.660	0.322	2.466	81.993	12.615
std	637239.410	4.480	1.753	65.451	66.731
min	0.000	0.000	0.000	0.000	0.000
25%	140000.000	0.000	1.000	33.500	0.000
50%	286000.000	0.000	2.000	55.000	0.000
75%	550000.000	0.000	3.000	123.312	0.000
max	10000000.000	250.000	13.000	329.000	811.000

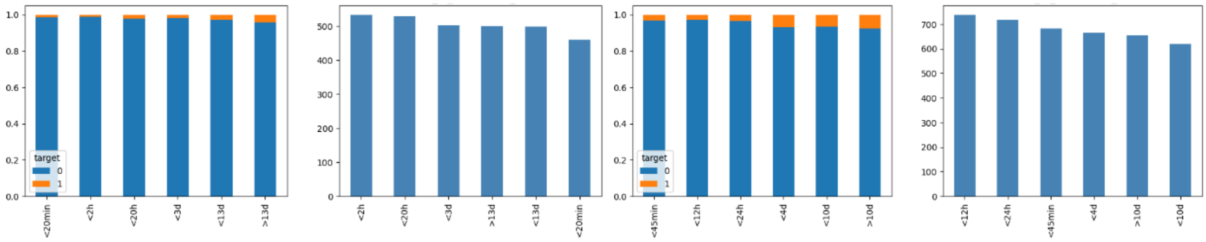
Figure 15: Conversion and Distribution tables of key features



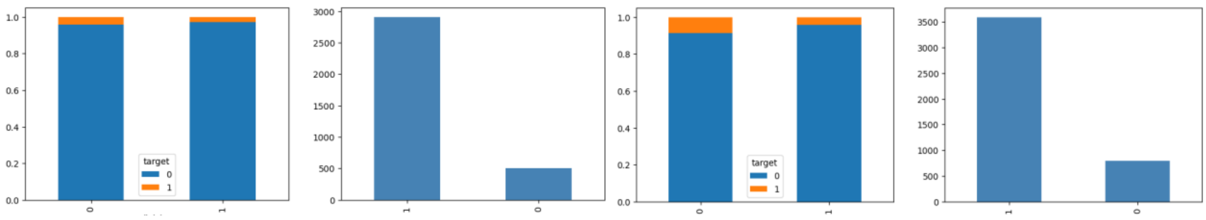
time_to_firstactivity



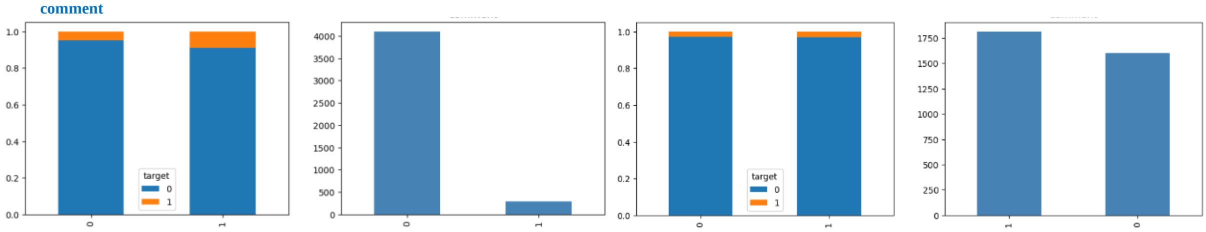
time_to_lastactivity



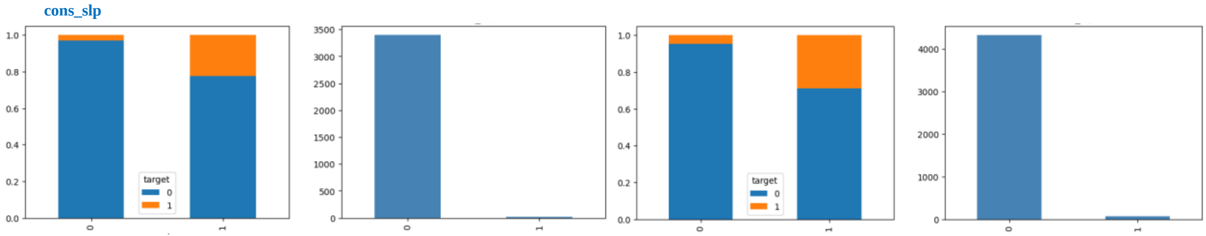
division



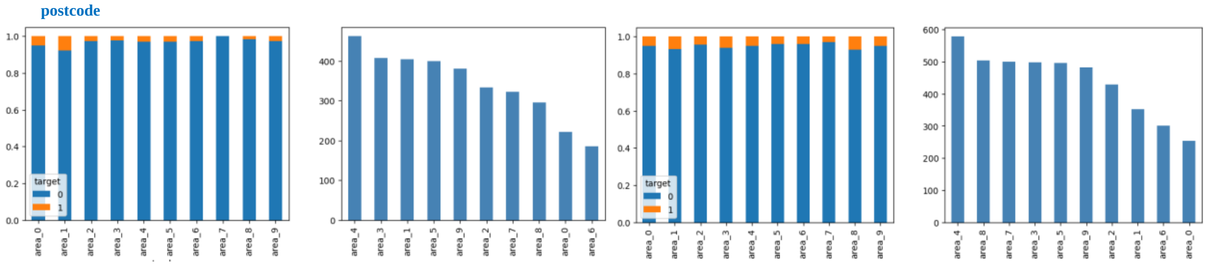
comment



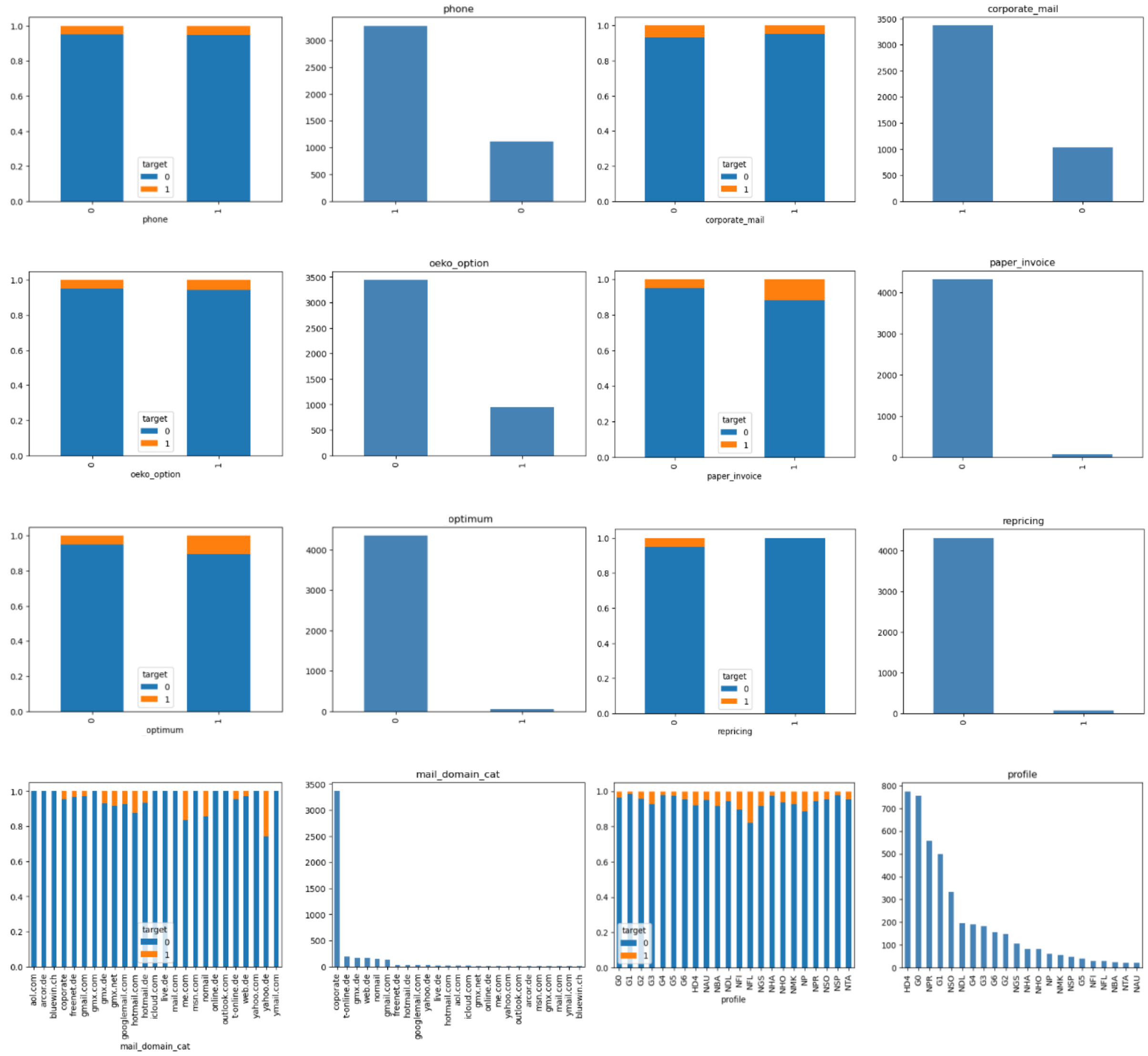
cons_slp



postcode



Additional features from Pricing tool; conversion and distribution:



Appendix 10: Top models and hyperparameters trained for each experiment

Table 6: Top 10 Classifiers CF Maturity A (M 1A)

	Classifier	Scaler	Val WAUC	Train Time	Hyperparameters
1	Random Forest	Standard Scaler	0.78618	55s	bootstrap: false, class_weight : balanced, criterion : entropy, max_features : 0.6, min_samples_leaf : 0.036, min_samples_split: 0.197, n_estimators: 50
2	Extra Trees	Standard Scaler	0.78400	45s	bootstrap: false, class_weight : balanced, criterion : gini, max_features : null, min_samples_leaf : 0.01, min_samples_split: 0.057, n_estimators: 200
3	Extra Trees	Standard Scaler	0.77361	52s	bootstrap: false, class_weight : balanced, criterion : gini, max_features : 0.9, min_samples_leaf : 0.01, min_samples_split: 0.198, n_estimators: 10
4	Extra Trees	Standard Scaler	0.77279	54s	bootstrap: false, class_weight : balanced, criterion : gini, max_features : 0.7, min_samples_leaf : 0.01, min_samples_split: 0.198, n_estimators: 200
5	Random Forest	Standard Scaler	0.75465	58s	bootstrap: false, class_weight : null, criterion : entropy, max_features : 0.7, min_samples_leaf : 0.036, min_samples_split: 0.244, n_estimators: 400
6	XGBoost Classifier	Standard Scaler	0.75155	37s	booster : gbtree, colsample_bytree : 0.7, eta : 0.3, gamma: 0, max_depth: 5, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 1.563, subsample: 0.7, tree_method: auto
7	Extra Trees	Standard Scaler	0.75046	54s	bootstrap: true, class_weight : balanced, criterion : gini, max_features : 0.9, min_samples_leaf : 0.01, min_samples_split: 0.057, n_estimators: 50
8	XGBoost Classifier	Standard Scaler	0.74626	1m 3s	booster : gbtree, colsample_bytree : 0.9, eta : 0.5, max_depth: 4, max_leaves: 0, n_estimators: 400, objective: reg logistic, reg_alpha: 1.875, reg_lambda: 1.875, subsample: 0.9, tree_method: auto
9	Extra Trees	MaxAbs Scaler	0.73214	53s	bootstrap: true, class_weight : null, criterion : entropy, max_features : null, min_samples_leaf : 0.01, min_samples_split: 0.01, n_estimators: 10
10	Random Forest	MaxAbs Scaler	0.72986	1m 9s	bootstrap: false, class_weight : null, criterion : entropy, max_features : 0.5, min_samples_leaf : 0.01, min_samples_split: 0.01, n_estimators: 400

Table 7: Top 10 Classifiers CF Maturity B (M 1B)

	Classifier	Scaler	Val WAUC	Train Time	Hyperparameters
1	ExtraTrees	StandardScaler	0.78181	36s	bootstrap: false, class_weight : balanced, criterion : gini, max_features : null, min_samples_leaf : 0.01, min_samples_split: 0.057, n_estimators: 200
2	XGBoost Classifier	StandardScaler	0.77543	53s	booster : gbtree, colsample_bytree : 0.6, eta : 0.1, gamma: 1, growpolicy: lossguide, max_bin: 1023, max_depth: 9, max_leaves: 7, n_estimators: 100, objective: reg logistic, reg_alpha: 0, reg_lambda: 2.188, subsample: 0.9, tree_method: hist
3	XGBoost Classifier	StandardScaler	0.77215	38s	booster : gbtree, colsample_bytree : 0.7, eta : 0.3, gamma: 0, max_depth: 5, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 1.562, reg_lambda: 2.188, subsample: 0.7, tree_method: auto
4	XGBoost Classifier	MaxAbsScaler	0.77178	33s	tree_method: auto

5	XGBoost Classifier	StandardScaler	0.76613	51s	booster : gbtree, colsample_bytree : 1, eta : 0.3, gamma: 0, max_depth: 8, max_leaves: 7, n_estimators: 100, objective: reg logistic, reg_alpha: 0.208, reg_lambda: 1.250, subsample: 0.9, tree_method: auto
6	XGBoost Classifier	StandardScaler	0.75374	49s	booster : gbtree, colsample_bytree : 0.7, eta : 0.3, gamma: 0, max_depth: 9, max_leaves: 3, n_estimators: 400, objective: reg logistic, reg_alpha: 1.250, reg_lambda: 1.459, subsample: 1, tree_method: auto
7	ExtraTrees	StandardScaler	0.74043	43s	bootstrap: true, class_weight : balanced, criterion : gini, max_features : 0.9, min_samples_leaf : 0.01, min_samples_split: 0.057, n_estimators: 50
8	XGBoost Classifier	StandardScaler	0.74034	44s	booster : gbtree, colsample_bytree : 0.5, eta : 0.3, gamma: 5, max_depth: 6, max_leaves: 0, n_estimators: 50, objective: reg logistic, reg_alpha: 0.834, reg_lambda: 0.834, subsample: 0.7, tree_method: auto
9	XGBoost Classifier	StandardScaler	0.73642	52s	booster : gbtree, colsample_bytree : 0.9, eta : 0.3, gamma: 0, max_depth: 10, max_leaves: 255, n_estimators: 200, objective: reg logistic, reg_alpha: 2.5, reg_lambda: 1.354, subsample: 0.9, tree_method: auto
10	XGBoost Classifier	SparseNormalizer	0.73223	54s	booster : gbtree, colsample_bytree : 1, colsample_bylevel : 1, eta : 0.5, gamma: 0, growpolicy: lossguide, max_bin: 255, max_depth: 6, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 0, reg_lambda: 2.083, subsample: 1, tree_method: hist

Table 8: Top 10 Classifiers PT Maturity A (M 2A)

	Classifier	**Scaler**	**Val WAUC**	**Train Time**	**Hyperparameters**
1	Extra Trees	MaxAbs Scaler	0.70260	35s	bootstrap: false, class_weight : balanced, criterion : entropy, max_features : log2, min_samples_leaf : 0.01, min_samples_split: 0.01, n_estimators: 400
2	Extra Trees	Standard Scaler	0.70207	35s	bootstrap: false, class_weight : balanced, criterion: gini, max_features : null, min_samples_leaf : 0.01, min_samples_split: 0.057, n_estimators: 200
3	XGBoost Classifier	Standard Scaler	0.68079	44s	booster : gbtree, colsample_bytree : 0.7, eta : 0.5, gamma: 10, max_depth: 9, max_leaves: 3, n_estimators: 10, objective: reg logistic, reg_alpha: 0, reg_lambda: 0.208, subsample: 0.6, tree_method: auto
4	XGBoost Classifier	Truncated SVDWrapper	0.67887	54s	booster : gbtree, colsample_bytree : 0.9, eta : 0.001, gamma: 0.1, max_depth: 6, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 1.354, reg_lambda: 0.104, subsample: 1, tree_method: auto
5	Extra Trees	Standard Scaler	0.67488	38s	bootstrap: false, class_weight : null, criterion: entropy, max_features : 0.05, min_samples_leaf : 0.01, min_samples_split: 0.01, n_estimators: 25
6	XGBoost Classifier	Truncated SVDWrapper	0.67488	55s	booster : gbtree, colsample_bytree : 0.8, eta : 0.001, gamma: 0, max_depth: 8, max_leaves: 7, n_estimators: 50, objective: reg logistic, reg_alpha: 0, reg_lambda: 0.83, subsample: 0.6, tree_method: auto
7	Random Forest	MaxAbs Scaler	0.66931	40s	bootstrap: true, class_weight : balanced, criterion : gini, max_features : 0.2, min_samples_leaf : 0.036, min_samples_split: 0.057, n_estimators: 10
8	XGBoost Classifier	Standard Scaler	0.66702	33s	booster : gbtree, colsample_bytree : 0.9, eta : 0.1, gamma: 0, max_depth: 6, max_leaves: 3, n_estimators: 25, objective: reg logistic, reg_alpha: 0, reg_lambda: 0.729, subsample: 0.5, tree_method: auto

9	Extra Trees	MaxAbs Scaler	0.66380	39s	bootstrap: false, class_weight : null, criterion: entropy, max_features : 0.05, min_samples_leaf : 0.01, min_samples_split: 0.01, n_estimators: 100
10	XGBoost Classifier	Sparse Normalizer	0.66330	31s	booster : gbtree, colsample_bytree : 0.8, eta : 0.3, gamma: 0, max_depth: 6, max_leaves: 0, n_estimators: 10, objective: reg logistic, reg_alpha: 0, reg_lambda: 0.625, subsample: 0.8, tree_method: auto

Table 9: Top 10 Classifiers PT Maturity B (M 2B)

	Classifier	**Scaler**	**Val WAUC**	**Train Time**	**Hyperparameters**
1	XGBoost Classifier	Standard Scaler	0.86386	36s	booster : gbtree, colsample_bytree : 0.7, eta : 0.3, gamma: 0, max_depth: 5, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 1.563, reg_lambda: 2.188, subsample: 0.7, tree_method: auto
2	XGBoost Classifier	Standard Scaler	0.85110	49s	booster : gbtree, colsample_bytree : 0.9, colsample_bylevel : 1, eta : 0.05, gamma: 0, max_depth: 10, max_leaves: 1023, n_estimators: 800, objective: reg logistic, reg_alpha: 0.625, reg_lambda: 1.56, subsample: 1, tree_method: auto
3	XGBoost Classifier	Standard Scaler	0.84291	46s	booster : gbtree, colsample_bytree : 0.9, eta : 0.1, gamma: 0, max_depth: 6, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 2.083, reg_lambda: 0.313, subsample: 0.8, tree_method: auto
4	XGBoost Classifier	Standard Scaler	0.84016	46s	booster : gbtree, colsample_bytree : 1, colsample_bylevel: 0.8, eta : 0.5, gamma: 5, max_depth: 9, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 0.104, reg_lambda: 2.396, subsample: 1, tree_method: auto
5	LogisticRegression	Standard Scaler	0.82955	48s	C: 1526.42, class_weight: null, multi_class: multinomial, penalty: l2, solver: newton-cg
6	XGBoost Classifier	Standard Scaler	0.82841	42s	booster : gbtree, colsample_bytree : 0.7, eta : 0.2, gamma: 1, max_depth: 9, max_leaves: 0, n_estimators: 200, objective: reg logistic, reg_alpha: 1.25, reg_lambda: 1.458, subsample: 0.6, tree_method: auto
7	XGBoost Classifier	MaxAbs Scaler	0.82485	29s	tree_method: auto
8	XGBoost Classifier	Standard Scaler	0.82364	46s	booster : gbtree, colsample_bytree : 0.7, eta : 0.3, gamma: 0, max_depth: 7, max_leaves: 0, n_estimators: 100, objective: reg logistic, reg_alpha: 1.77, reg_lambda: 0, subsample: 0.6, tree_method: auto
9	XGBoost Classifier	Standard Scaler	0.81948	46s	booster : gbtree, colsample_bytree : 0.8, eta : 0.05, gamma: 1, max_depth: 8, max_leaves: 255, n_estimators: 100, objective: reg logistic, reg_alpha: 2.188, reg_lambda: 1.25, subsample: 1, tree_method: auto
10	XGBoost Classifier	Truncated SVDWrapper	0.81579	49s	booster : gbtree, colsample_bytree : 0.5, eta : 0.2, max_depth: 3, max_leaves: 0, n_estimators: 400, objective: reg logistic, reg_alpha: 0, reg_lambda: 0.833, subsample: 0.7, tree_method: auto

Appendix 11: Sampling Results

Below Table shows the resulting WAUC for Validation as well as Test Set on the example of the model for the Pricing Tool Maturity A. It can be seen that the sampling led to overfitting of the model for all tested methods as test results are lower than validation results. Especially the Recall is very high for all models in the validation dataset and close to 0 in the test set. Therefore, resampling was not used in the final model. Similar results were observed for the other experiments.

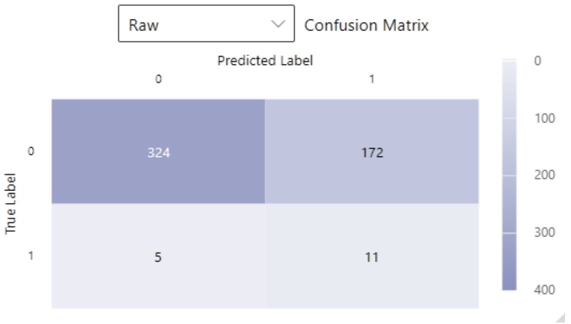
Sampling	Metric	Validation Result	Test Result
Oversampling (Rate 0.5)	Recall	1.0	0.083
	FPR	0.085	0.003
	WAUC	**1.0**	**0.512**
Oversampling (Rate 0.2)	Recall	0.897	0.103
	FPR	0.089	0.056
	WAUC	**0.887**	**0.512**
SMOTE-ENC (Rate 0.5)	Recall	0.851	0.083
	FPR	0.002	0.005
	WAUC	**0.978**	**0.573**
SMOTE-ENC (Rate 0.2)	Recall	0.734	0.143
	FPR	0.04	0.01
	WAUC	**0.876**	**0.601**

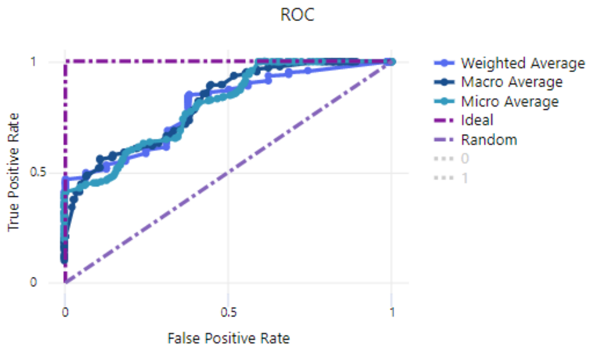
Table 10: Sampling results model 2A

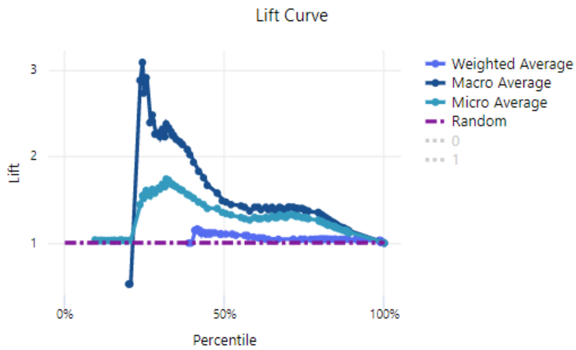
Appendix 12: Model Results

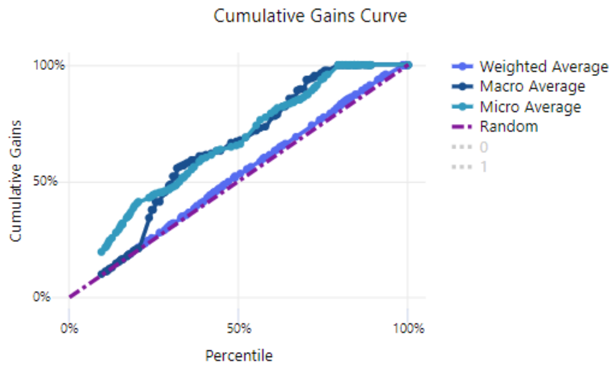
Figure 16: Results of Contact Form Maturity A

Metric	Validation Result	Test Result
Accuracy	0.655	0.654
WAUC	0.786	0.792
Precision	0.531	0.523
F1 Score	0.457	0.448
Recall Score	0.748	0.670









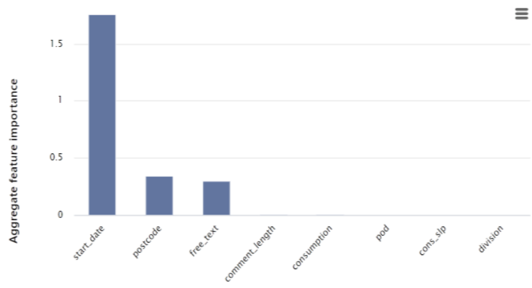
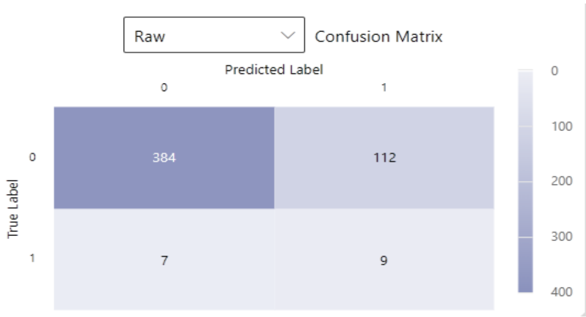


Figure 17: Results of Contact Form Maturity B

Metric	Validation Result	Test Result
Accuracy	0.782	0.768
WAUC	0.782	0.804
Precision	0.516	0.528
F1 Score	0.486	0.499
Recall Score	0.589	0.668



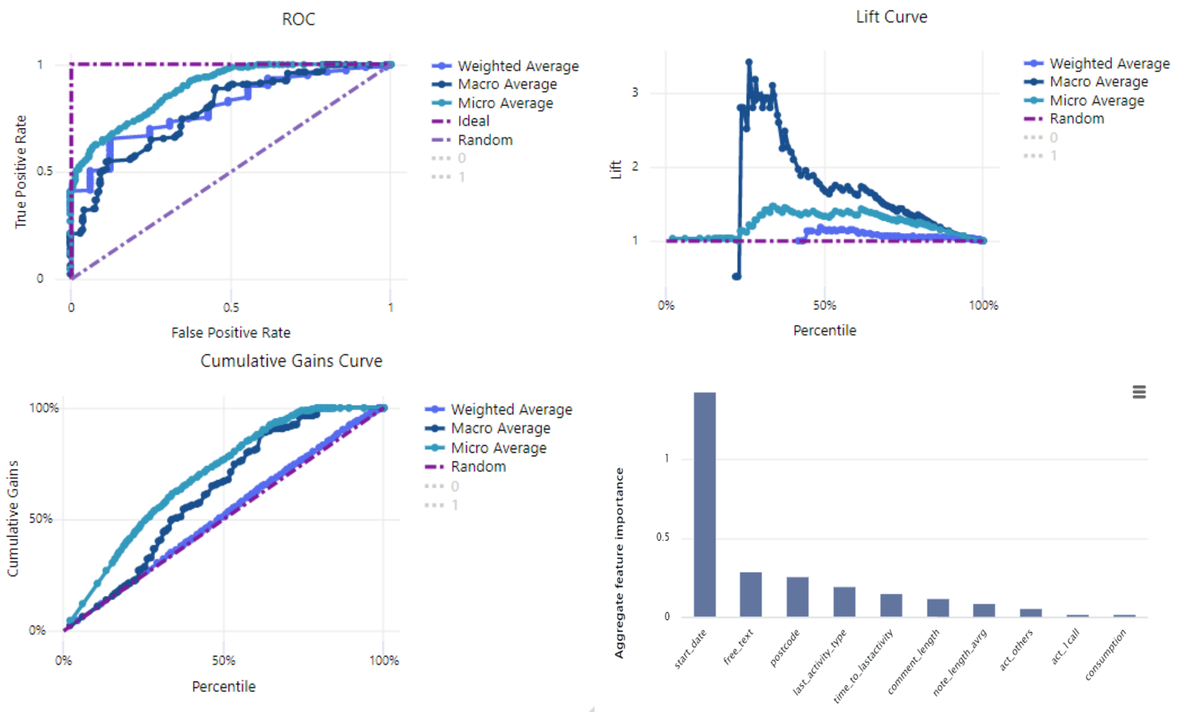
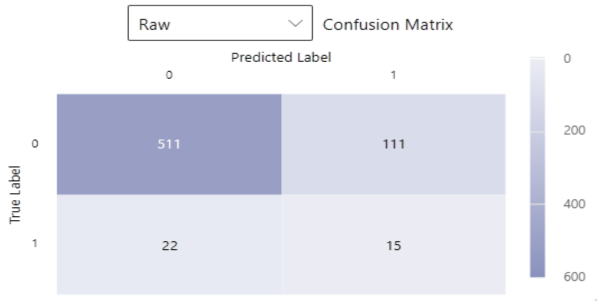
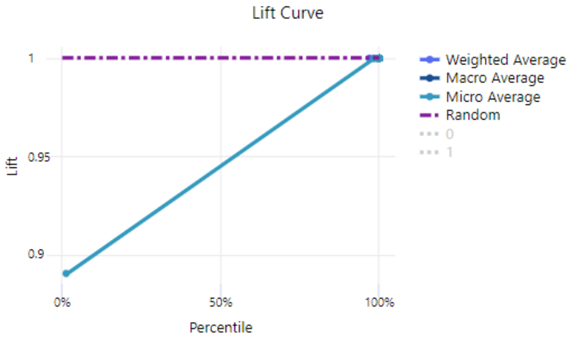
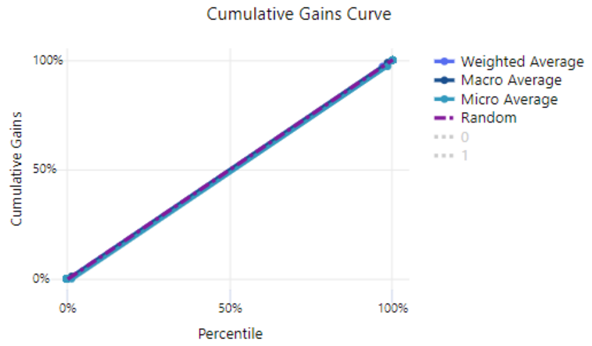


Figure 18: Results of Pricing Tool Maturity A

Metric	Validation Result	Test Result
Accuracy	0.827	0.798
WAUC	0.703	0.612
Precision	0.557	0.539
F1 Score	0.563	0.534
Recall Score	0.672	0.613







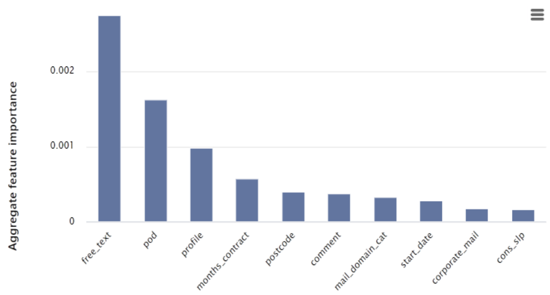


Figure 19: Results of Pricing Tool Maturity B

Metric	Validation Result	Test Result
Accuracy	0.943	0.982
WAUC	0.864	0.975
Precision	0.692	0.923
F1 Score	0.677	0.912
Recall Score	0.665	0.901