Automating Lead Scoring with Machine Learning: An Experimental Study

Robert Nygård
Idea Development BBN
Faculty of Social Sciences, Business and Economics,
Åbo Akademi University
robenyga@abo.fi

József Mezei
School of Business Planning, Chongqing Technology and Business University
Faculty of Social Sciences, Business and Economics,
Åbo Akademi University
Jozsef.Mezei@abo.fi

Abstract

Companies often gather a tremendous amount of data, such as browsing behavior, email activities and other contact data. This data can be the source of important competitive advantage by utilizing it in estimating a contact's purchase probability using predictive analytics. The calculated purchase probability can then be used by companies to solve different business problems, such as optimizing their sales processes. The purpose of this article is to study how machine learning can be used to perform lead scoring as a special application case of purchase probabilities. Historical behavioral data is used as training data for the classification algorithm, and purchase moments are used to limit the behavioral data for the contacts that have purchased a product in the past. Different ways of aggregating time-series data are tested to ensure that limiting the activities of buyers does not result in model bias. The results suggest that it is possible to estimate the purchase probability of leads using supervised learning algorithms, such as random forest, and one can obtain novel business insights from the results using visual analytics relevant for decision makers.

1. Introduction

In the present competitive business environment, some of the most critical business decisions are related to customer acquisition. During the acquisition phase of the customer life cycle, companies try to convert leads into customers through different methods. In order to make this process as time- and cost-efficient as it is possible from the organizations point of view, various *lead scoring* methodologies [1] have been proposed and used in practice.

Lead scoring is the general procedure applied by organizations in prioritizing which customer leads to target. In the typical case, the evaluation is based on activities performed by the potential customer when interacting with the company through different channels. This may include website visits or emails. According to a basic model, each activity is assigned an importance score; the leads are ranked based on this score and the ones with the highest overall score are then pursued by sales people. This process is termed as manual lead scoring.

The main goal of this article is to understand how machine learning can assist in automating and improving the lead scoring processes in the B2C (Business-to-Consumer) context. In order to achieve this goal, real world data is utilized to illustrate the typical issues part of a general data preparation process and to build and evaluate different machine learning models as the basis of automated lead scoring. Additionally, utilizing various data visualization techniques, we try to illustrate how the lead scoring results can help in uncovering various business insights, such as the importance of different customer touch points. The research objective of the article is to understand *"how can machine learning and data analytics be used to automate lead scoring and generate business insights for the decision makers"*.

The rest of the article is structured as follows. In Section 2, a brief literature review is provided on the general topic of analytics, machine learning and automation in Customer Relationship Management. This is followed by the description of the data used in the empirical study and the data analysis methodology in Section 3. We present and discuss the results in Section 4. Finally, some conclusions are provided in Section 5.

2. Background

In present days, companies generate and collect a tremendous amount of data [2]. As a consequence of this, organizations increasingly rely on data-driven decision support [3]. Lead scoring, or marketing and customer relationship management processes of

URI: https://hdl.handle.net/10125/63916
978-0-9981331-3-3

companies are not different from these trends. A broad area of marketing, presently termed as relationship marketing, is 'the ongoing process of engaging in collaborative activities in programs with immediate and end-user customers to create or enhance mutual economic, social and psychological value, profitably' [4]. The process of relationship marketing relies largely on the availability of digital data that is increasingly relevant for organizations because of the fact that they need to have a strong digital presence in order to remain competitive [5]. Collecting this digital data allows organizations to collect data on how possible future customers and interested people, i.e. leads, have interacted with various online communication channels available.

Tracing these activities and applying various advanced business analytics tools or machine learning to the collected data can enhance customer relationship management significantly [6]. Gathering this useful information takes place via various online channels, such as e-commerce websites, software and email. In general, the overarching conclusions of numerous studies support the statement that in presence of this possibility to utilize data in marketing and customer relationship management, organizations should not have to rely on gut feeling or business intuition, but rather pursue data-driven decisions when implementing an (automated) lead scoring solution to replace or at least complement manual lead scoring [7].

To reformulate these observations in our specific context, we can state that automated marketing is the process of utilizing data from tracking online actions of potential leads to learn about behavioral patterns of these potential buyers that can aid in identifying the ones who are more likely to turn into actual customers [8]. While the tools to support these processes in an automated way are readily available, there are very few studies attempting to understand and develop new models on how companies can utilize 'these tools to guide potential buyers engaged in different stages of the B2C sales process' [8]. Based on this brief discussion, we present a brief literature review on lead scoring and machine learning applications in automated customer relationship management.

2.1 Manual lead scoring

Before discussing the main components of automated lead scoring, it is important to discuss the dominant approach used in practice as identified in the introduction section: manual lead scoring.

According to Marion [9], there are several problematic issues with manual lead scoring. Most importantly, manual lead scoring fails to base the recommendations on statistical support. Additionally, as typically, manual lead scoring relies on a wide set of demographic, behavioral or firmographic data, lack of some specific information for some leads with high assigned scoring weight can significantly distort the results. Finally, as the manual lead scoring process is based on a lead scoring matrix, if companies aim to keep up with the constantly changing business environment, they have to manually reevaluate and update this scoring matrix continuously.

Activity	Points
Form/Landing Page Submission	+ 5
Submitted "Contact Me" Form	+25
Received an Email	0
Email Open	+1
Email Clickthrough	+3
Registered for Webinar	+3
Attended Webinar	+10
Downloaded a Document	+5
Visited a Landing Page	+2
Unsubscribed from Newsletter	-2
Watched a Demo	+8
Contact is a CXO	+5
Visited Trade Show Booth	+3
Visited Pricing Page	+10

Figure 1: Example manual lead scoring matrix [9]

An example of a scoring matrix adopted from [9] can be observed in Figure 1. In the study, the authors conducted an experiment of 800 leads scored according to manual lead scoring. They found no statistical difference between being able to convert scored leads that were determined "ready for sales" and randomly choosing leads that were not scored at all. Marion [9] asserts that there is absolutely no way that someone without experience in statistics could score or weigh these activities properly. Furthermore, it is a very time-consuming process to always keep adjusting the scores and that the time used could be spent more effectively elsewhere. Bohlin [10] also claims that manual lead scoring is not a recommended approach, even if rules and weights developed through assumptions are used together.

2.2 Components of lead scoring

Lead scoring can be seen as a subtask of customer relationship management (CRM). The process of lead scoring attempts to assign a numeric value (lead score) to potential customers of an organization [11].

A higher lead score implies that the contact, or lead, is more likely to engage with the company; consequently, it allows companies to prioritize their sales. According to [12], high priority leads should be passed on to sales and low priority leads should be engaged in lead nurturing campaigns.

The most crucial task that largely influences the quality of the lead scoring system's output is the selection of variables included in the lead scoring models. One can divide collected data into two main classes [12]: implicit data (obtained from collecting data on the actions of potential leads) and explicit data (obtained directly from the customer's own input). The best performing companies usually included three or more implicit variable attributes in their lead scoring model, while the highest performing companies tend to have more complex scoring models than their competitors [12].

From a methodological perspective, lead scoring is part of the general domain of predictive analytics: we try to estimate the likelihood of a lead turning into an actual customer: predicting future purchasing behavior of leads. According to [13], predictive analytics is *'an umbrella term that covers a variety of mathematical and statistical techniques to recognize patterns in data or make predictions about the future'*. In the case of lead scoring, mathematical and statistical techniques and machine learning are typically used to find patterns in the data to estimate the likelihood of a lead turning into a purchase.

When predictive analytics is applied to the purpose of scoring leads, it is part of predictive marketing [13], *'a customer-centric marketing approach that aims to enrich the customer's experience throughout the customer life cycle'*. This experience is made possible due to the availability of technology that captures data previously inaccessible to the everyday marketer. Another factor that contributes to the success of predictive marketing is the dramatic decrease in computing costs.

Predictive analytics [14] can be characterized as a set of techniques used to generate insights from data, in the form of statistical models or machine learning algorithms. In general, machine learning algorithms can be classified into three main groups: supervised, unsupervised and reinforcement learning. The main goal of lead scoring is to obtain a numeric value that predicts the likelihood of a customer lead turning into a sale. This is a typical problem that can be classified as supervised learning: 'supervised' by historical data of previous leads including their characteristics and the observed outcome of the lead (i.e. whether it actually turned out to be a customer or not), we try to build a model that can predict the outcome for future leads.

While the number of contributions utilizing machine learning techniques is not extensive, one can identify a handful of articles. We not here that in contrast, a seach in a patent database reveals a large number of related patent applications, highlighting the relevance and timeliness of the topic.

In [11], a lead scoring model is constructed utilizing Bayesian networks. This approach allows combining expert knowledge and historical data in a straightforward manner requiring a small amount of data. In [15], the author analyzes the impact of utilizing modern information technologies such as machine learning to improve the efficiency of managing the customer journey, including how to effectively shorten the customer journey and related sales cycle in business-to-business firms using new technologies.

2.3 Machine learning examples from customer relationship management

In the following, we present some relevant applications of machine learning in customer relationship management to illustrate the potential insights we can gain with these models. In [16], a collection of literature is discussed regarding the application of machine learning in customer relationship management. Based on different stages of the customer journey, the authors identify seven different types of machine learning methods used in the literature. According to their literature review, the most widely used machine learning models in customer relationship management include association rules mining, classification, clustering, forecasting, regression, sequence discovery and visualization. The most common machine learning algorithms used include association rule, decision tree, genetic algorithm, neural networks, K-nearest neighbor and linear as well as logistic regression [16]. This finding was one of the main reasons for the selection of algorithms tested in the empirical study presented in the main part of this article.

In [17], a decision support tool is constructed that aids in predicting customer loyalty in a non-contractual setting using random forest, logistic regression and neural networks. Logistic regression was included as a comparison point for the more advanced models. The random forest algorithm is used in lieu of a decision tree algorithm due to their robustness and superior performance. The model is evaluated using accuracy and AUC. The model was successful in detecting future partial defection and there were no noticeable differences in the models created by the three algorithms.

In [18], genetic algorithm and an artificial neural network are applied to maximize expected profit from direct mailing. The genetic algorithm is used to select different subsets of variables to pass on to the neural network, the results are evaluated, and the best subset is

then chosen for the final analysis. This is done to minimize the number of variables to increase the interpretability of the neural network model, which potentially allows marketers to extract key drivers of consumer response. However, reducing the number of variables could lead to a decrease in accuracy. The method produced a model that considers campaign costs and profit per additional customer, maximizing the expected profit and having a higher interpretability due to using a smaller set of features.

3. Methodology

In the study, the general recommended process from [19] for predictive analytics in information systems research is applied. With the focus of the research being on the construction and evaluation of possible predictive machine learning models for automated lead scoring, data understanding focuses on examining the data and identifying and correcting potential problems present in it. In the data preparation process, the data is transformed in order to deal with missing values and outliers, and to create a variable structure utilizing feature extraction, filtering and feature selection that is appropriate for further machine learning model building. In the next steps, several models are built and evaluated using machine learning algorithms. After the optimal model is identified, the main results are interpreted utilizing visualization tools.

3.1 Data description and preprocessing

As specified above, the main goal in this article is to illustrate the usefulness and added value that machine learning can offer by creating automated lead scoring models. In order to do so, we conducted an experiment using real life data from an international company, focusing on its potential leads in Finland.

The general lead processing of the company relies on obtaining information about potential leads through mainly online and sometimes offline data. The collected information is sent to local contractors for further processing. Finally, they make the decision on whether initiating a contact with the lead for further inquiries or not. While the company has both B2B (Business-to-Business) and B2C (Business-to-Consumer) lines, in this analysis we focus on the data available for B2C leads. Data is included in the analysis for the time period 18.2.2018- 16.11.2018. In the analysis, two main sources of data are utilized:

- contact-level data from the company's internal systems (data on customer name, country, location, the source of the lead and whether the lead has made a purchase)
- activity data (website visits, email sends, email opens, email click throughs, form submits, etc.)

In the analysis, each step of data preprocessing, model building and evaluation was performed using RapidMiner software [20]. A summary capturing the most important steps of data preprocessing is shown in Figure 2.

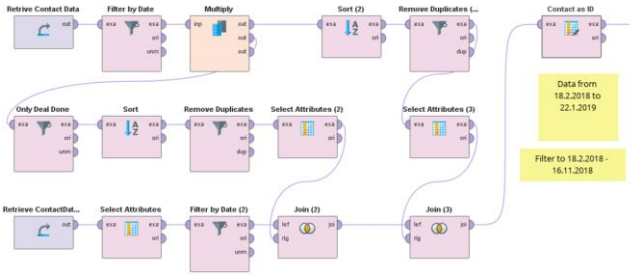

Figure 2: Main steps of data pre-processing and filtering in RapidMiner

By working with the raw datasets, it was possible to extract more than 200 variables. However, many of the extracted variables were found not to be useful for the purpose of our data analysis. Variable for the analysis were filtered out based on the following criteria:

- correlation with the output label
- number of unique values (in categorical variables)
- number of missing values

From the main dataset, the following datatypes were selected:

- Identifier: links the same contacts in different data sets
- Location: specifies the region of the lead
- Marketing unit: specifies the location of marketing unit
- Date created and modified: timestamp for events related to the lead
- Email address domain: specifies the email domain
- Contact status and time: identifies buyers and their purchase moment

The activity data was converted into a long-table format with three columns: (i) contact; (ii) activity type, and (iii) the time of activity. As we are dealing with censored data, it is important to realize the uncertainty related to the outcome of leads that has no associated purchase yet. In case of converted leads, i.e. leads that

turn into actual purchase, the output clearly can be assigned the value of 1, while for leads with no actual purchase in the dataset we only know that they were not converted into actual customers until the time of the analysis.

The activity data is used in the models after aggregation, e.g. a count for different event types for each lead is calculated. As the basis of the aggregation, the end date for the lead conversion process needs to be determined. For converted leads, a natural choice for the end date is the time of the first purchase; this implies that data collected about the customer after the first purchase is not used in the analysis. For non-customers, it is not straightforward to specify an end date, and for this reason, different aggregations based on various possible end dates will be evaluated in the experiment in order to minimize the bias present in the modelling process. Different ways to specify the end date include the following:

- the end of the time period considered in the data, which is 16.11.2018;
  the lead's last activity;
- a random date between the lead's first and last activity;
- the date of the last activity before a randomly chosen end date between the first and last activity of the lead.

To offer an overview of the underlying data, we present in Table 1 some descriptive information after performing the aggregation using the lead's last activity (as we will discuss later, this is chosen as the most unbiased way to aggregation):

Table 1: Descriptive data of some activity measurements

Activity	Mean	Standard deviation
EmailSends	1.09	1.38
Bouncebacks	0.01	0.09
EmailOpens	1.99	5.34
EmailClickthroughs	0.14	0.79
Subscribe	0.73	0.44
Unsubscribe	0.09	0.28
PageView	4.12	27.21
WebVisit	1.75	5.59
FormSubmit	1.63	1.59

Based on the above considerations, the following list of final variables was included in the model from the activity dataset:

- Contact: lead identifier
- daysToEnd.max: days between the first activity and the end date
- daysToEnd.avg: average number of days between all activities and the end date
- Sum: total number of activities
- 1daySum: number of activities within 1 day of the end date
- 3daySum: number of activities within 3 days of the end date
- 1weekSum: number of activities within 1 week of the end date
- 2weekSum: number of activities within 2 weeks of the end date
- 4weekSum: number of activities within 4 weeks of the end date
- 10percentSum: number of activities within 10 percent of the total time prior to the end date
- 40percentSum: number of activities within 40 percent of the total time prior to the end date
- 80percentSum: number of activities within 80 percent of the total time prior to the end date

4. Results

Based on the final dataset described in the previous section, four different machine learning algorithms were selected to be tested motivated by the findings in our literature review on the most widely used algorithms in customer relationship management:

- Logistic regression (LR) [14]: a widely used class of generalized linear models used in binary classification tasks
- Decision trees (DT) [21]: a family of tree-based models that result in a set of nested if-then statements derived from the variables found in the data set. An important advantage of tree-based models in practice that they offer intuitive explanations on how the predicted class is arrived at
- Random forests (RF) [22]: another family of tree-based models that attempt to alleviate the decision tree algorithm's instability problems by simultaneously creating several de-correlated decision tree models and calculating their average as the basis of predicting the output
- Neural networks (NN) [23]: non-linear algorithms and models with the most common algorithms utilizing back-propagation and a small number of hidden layers. In recent years, thanks to advances in deep learning, neural

networks because the number one choice in most supervised (and unsupervised) learning applications.

In order to evaluate the performance of the constructed machine learning models, as it is common in practice, different evaluation metrics based on the confusion matrix are used [14]. By differentiating between correct and incorrect classifications on the two possible output classes, we can count true positive (TP), true negative (TN), false positive (FP) and false negative predictions (FN). In this paper, a positive case refers to a converted lead and negative case refers to leads with no actual purchase. Additionally to the basic accuracy measure, in order to account for the different types of errors, we can calculate metrics such as precision, recall, sensitivity and specificity. The final evaluation measure utilized in this paper is the Area under the Curve (AUC) that can be obtained by calculating the area under the Receiver Operating Characteristic (ROC) curve. ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) across different probability thresholds.

Finally, regarding the model building process, as the sample dataset was largely unbalanced, SMOTE up-sampling was used to tackle this issue. Additionally, 10-fold cross-validation was used for resampling to obtain a fair estimate of the different models' performance. In the following, we start with discussing a preliminary investigation of different possible data aggregation procedures, model performance for them and associated estimated bias. Based on assessing the involved bias, one final aggregation procedure is selected and a more detailed analysis is performed for that case.

4.1 Evaluating different data aggregation strategies

In this section, we will look at five possible strategies to activity data aggregation. The aggregation methods were selected to demonstrate the importance of correctly handling the classes to reduce the amount of bias. Based on discussions with experts and the experience of the participating data analyst, these approaches cover the most important views that are normally considered when evaluating the value of specific events based on the time when the lead performed it.

In Aggregation 1, the end date for non-customers was set as the end of the time period considered in the data, while for converted leads it was set to be the same as their first purchase date. The results for this case can be seen in Table 2. In this case, non-customers have very different aggregated values depending on when they were active, resulting in a high bias.

In Aggregation 2, the end date for non-customers was set as the date of their last activity, while for converted leads it was set to be the end of the time period considered in the dataset. According to the results in Table 2, the models become very good at predicting buyers. This is a bias since the effectiveness mostly stems from the fact that the aggregations are calculated in slightly different ways for both classes.

Table 2: Evaluation for different aggregation strategies

Aggregation method	Models /evaluation metric	Precision for positive class	Precision for negative class	Recall for positive class	Recall for negative class
A1	LR	0.35	0.99	0.90	0.88
	DT	0.37	0.99	0.88	0.87
	RF	0.39	0.99	0.91	0.89
	NN	0.60	0.97	0.64	0.97
A2	LR	0.26	0.99	0.88	0.80
	DT	0.91	0.99	0.94	0.99
	RF	0.98	0.99	0.93	0.99
	NN	0.98	0.99	0.88	0.99
A3	LR	0.13	0.97	0.77	0.68
	DT	0.15	0.96	0.66	0.69
	RF	0.15	0.97	0.69	0.69
	NN	0.23	0.95	0.36	0.90
A4	LR	0.15	0.98	0.83	0.64
	DT	0.21	0.97	0.69	0.79
	RF	0.21	0.98	0.83	0.64
	NN	0.33	0.95	0.36	0.94
A5	LR	0.28	0.99	0.87	0.82
	DT	0.32	0.98	0.82	0.86
	RF	0.32	0.99	0.90	0.85
	NN	0.48	0.97	0.57	0.95

In Aggregation 3, the end date for non-customers was set as date of their last activity, while for converted leads it was set to be the date of the last activity before their purchase. As can be seen in Table 2, while this method fixes the bias that occurred in Aggregation strategies 1 and 2, the recall and precision values have dropped. However, this seems to be the fairest, most unbiased method of aggregating the activity data.

In Aggregation 4, the end date for non-customers was set as the date of a randomly chosen date between their first and last activity, while for converted leads it was set to be the date of the last activity before their purchase. The results for this case can be seen in Table 2. Selecting a random end date between the first and last activity for non-customers is meant to simulate them in different stages of the customer life cycle, which may altogether be a fairer way to teach the models. However, this approach implies that the last of the non-buyers activities will always be left out.

In Aggregation 5, the end date for non-customers was set as the date of a randomly chosen date between their first and last activity, while for converted leads it was set to be the date of their purchase. The results in Table 2 can also be seen as biased to some extent as for customers, a predetermined end date used without considering the time between the last activity and their last preceding action, while for non-buyers a randomly generated one inside of their activity timeline is used. Additionally, the last of the non-buyers activities will always be left out as in the previous approach.

The summary of our observations on the associated bias across methods together with the best performing model based on AUC in each case is presented in Table 3. As we can observe, based on this widely used metric, independently of the aggregation strategy, random forest is always the best performing model. Additionally, based on this evaluation Aggregation strategy 3 is selected for more detailed analysis as it is the one with the least possible bias.

Table 3: Comparison of aggregation strategies

Aggregation	Bias	Best model AUC
1	High	Random forest: 0.955
2	High	Random forest: 0.991
3	None	Random forest: 0.761
4	Low	Random forest: 0.843
5	Medium	Random forest: 0.935

4.2 Model evaluation for the chosen data aggregation strategy

A comparison of the performance of different models is presented in Table 4. As expected, the decision tree model is not as effective as the random forest model. The created decision tree has a maximum depth of 10 after pruning, which would make it challenging to use it in practice to derive specific explanations for predictions, which would be the main benefit of using this model.

The random forest model was created using 100 decision trees and has the best overall score. Based on this model, it is possible to produce the attribute importances, which will be presented in the following section.

Logistic regression was mainly included in the procedure to obtain a benchmark for what a linear classification algorithm could achieve compared to the more complex, non-linear machine learning algorithms. The model achieved the highest sensitivity, albeit the

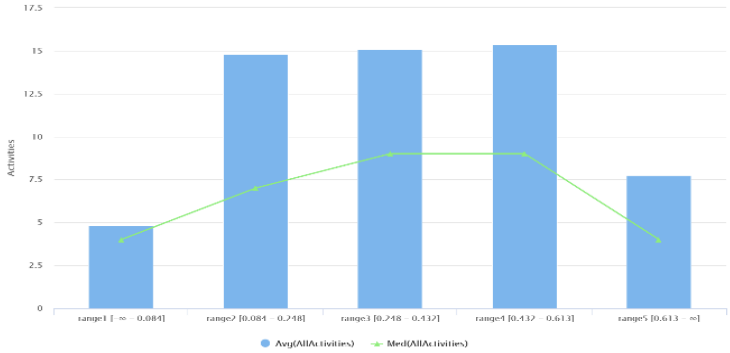

Figure 3: Average and median activity amount per purchase probability group

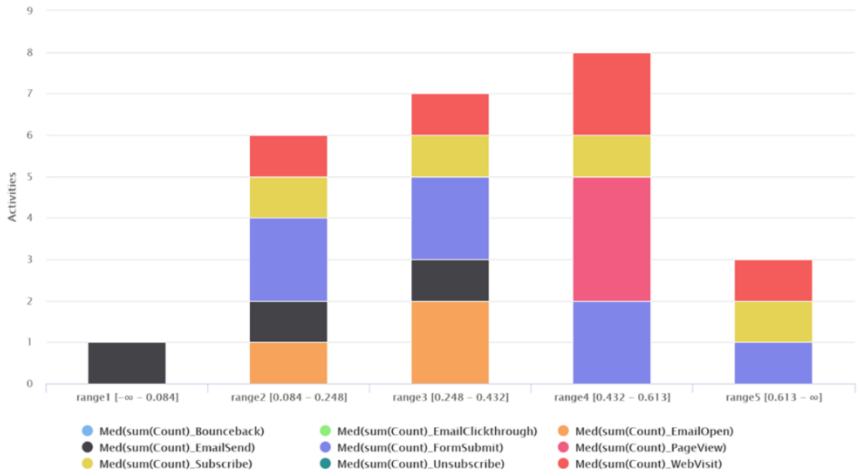

Figure 4: Median activity amount per purchase probability group

lowest specificity, which means it is better at identifying the positive class, but at the cost of being worse at identifying the negative class.

Despite having the highest accuracy and only slightly lower AUC, the neural network model has treated the classes very differently. This can be observed by looking at the sensitivity and specificity values of the model. It seems to have correctly guessed 90.22 % of the negative class, but only 36.27 % of the positive class. In lead scoring or marketing in general, one could argue that it is more important to be able to detect the positive than the negative class.

In conclusion, the random forest model is selected as the best performing model. This decision is based on the model having the highest overall performance score and the possibility to interpret the model through attribute importances. However, if one were to assign financial values such as the cost of losing a potential lead versus the cost of contacting a lead, the value of each model could change. For example, the logistic regression model may be better than the other models if sensitivity were to have a higher value than specificity.

Table 4: Model performance comparison for the chosen aggregation strategy

Model	Accuracy	AUC	Sensitivity	Specificity
LR	0.59	0.70	0.77	0.58
DT	0.69	0.72	0.66	0.69
RF	0.69	0.76	0.69	0.69
NN	0.86	0.75	0.36	0.90

We performed further analysis to understand the differences among groups of data points partitioned based on the estimated purchase probability obtained in the random forest model. Five groups were constructed with keeping the number of data points in the groups approximately equal, with the probability threshold values between groups set as [0.084, 0.248, 0.432, 0.613]. Figure 3 presents the average and median number of activities corresponding to different purchase groups. As we observe from the figure, leads with the lowest and highest estimated probability tend to have a fewer number of activities in contrast to the other three groups that behave similarly to each other. A possible reason for the fewer number of activities can be that they correspond to leads that already know with certainty that they will purchase and know what they are looking for, consequently require less interaction with the company to make their final purchase decision.

Finally, we also looked at the different activity types performed by the leads in different purchase probability groups. Figure 4 can help sales employees to further understand customer groups and improve sales processes. For example, leads in the second highest average purchase probability group have a high median value for Page Views on the company website that is not present in any of the other groups. Company employees, to further understand the reason for this distinct difference, can look at information of this kind.

5. Conclusions

In this article, an empirical study is presented to evaluate the feasibility and performance of utilizing various machine learning model for automating lead

scoring as an alternative to the still widely used manual lead scoring process. As we identified in the literature review, this problem is not sufficiently well represented in the academic literature as much as the practical relevance of the problem would presumably require. In this article, we tested the most widely used machine learning approaches from the literature. Additionally, as a second contribution, we identified several feasible aggregation strategies to identify relevant actions for leads that have not resulted in an actual purchase in the considered timeframe of the data analysis, and evaluated these approaches from the perspective of classification performance and bias introduced in the modelling process. We found that, while there is a significant challenge in preparing and preprocessing in particular activity data on potential leads, one can obtain good classification performance even when controlling for the bias involved in model building. Additionally, we found that the random forest algorithm had the best overall performance out of all the different models. However, there is still room for improving the models through extensive parameter optimization, in particular in case of the neural network model. Since there are countless algorithms and other data manipulation procedures that are not included in this thesis, it is impossible to say that the random forest algorithm is the best among them.

There were no comparisons with lead scoring using machine learning and manual lead scoring, so it is not possible to say with complete certainty, which one is better. However, we have shown that machine learning-based lead scoring models offer a viable alternative.

Some areas of possible future research would be to add customer lifetime value to lead scoring, resulting in a monetary value which may seem more tangible than a simple purchase probability. For example, one could just multiply customer lifetime value with the purchase probability. Another example would be to use regression instead of classification to estimate the customer lifetime value of leads. In addition, identifying different lead types would be beneficial for companies. That way, they could treat the different types of leads with different types of marketing material, for example through nurturing campaigns. This could be done using unsupervised learning, since it is unknown how many different types of leads there are. Finally, different steps in the machine learning-model building could be further optimized. For example, a more thorough feature selection process, e.g. with forward selection, backwards elimination or Lasso approaches, could potentially improve the final classification performance.

6. References

[1] Syam, N. and Sharma, A., 2018. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management*, *69*, pp.135-146.

[2] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.

[3] Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133-39.

[4] Sheth J. N., Parvatiyar A., Sinha M., (2015). The conceptual foundations of relationship marketing: Review and synthesis. *Journal of economic sociology*, *16*(2), 119-149.

[5] Leeflang, P. S., Verhoef, P. C., Dahlström, P., & Freundt, T. (2014). Challenges and solutions for marketing in a digital era. *European management journal*, *32*(1), 1-12.

[6] Chorianopoulos, A. (2016). *Effective CRM using predictive analytics*. John Wiley & Sons.

[7] Duncan, B. A., & Elkan, C. P. (2015, August). Probabilistic modeling of a sales funnel to prioritize leads. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1751-1758). ACM.

[8] Järvinen, J., & Taiminen, H. (2016). Harnessing marketing automation for B2B content marketing. *Industrial Marketing Management*, *54*, 164-175.

[9] Marion, G. 2016. *Lead Scoring is Broken. Here's What to Do Instead.* URL: https://medium.com/marketing-on-autopilot/lead-scoring-is-broken-here-s-what-to-do-instead-194a0696b8a3 (Retrieved 24.09.2018)

[10] Bohlin, E. (2017). Sorting Through the Scoring Mess. URL: https://www.siriusdecisions.com/blog/sorting-through-the-scoring-mess (Retrieved 24.09.2018)

[11] Benhaddou, Y., & Leray, P. (2017, October). *Customer Relationship Management and Small Data—Application of Bayesian Network Elicitation Techniques for Building a Lead Scoring Model*. In Computer Systems and Applications (AICCSA), 2017 IEEE/ACS 14th International Conference on (pp. 251-255). IEEE.

[12] Michiels, I. (2008). *Lead Prioritization and Scoring: The Path to Higher Conversion.* Aberdeen Group.

[13] Artun, O., & Levin, D. (2015). *Predictive marketing: Easy ways every marketer can use customer analytics and big data*. John Wiley & Sons.

[14] Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York: Springer.

[15] Adam, M.B. (2018). *Improving complex sale cycles and performance by using machine learning and predictive*

analytics to understand the customer journey (Doctoral dissertation, Massachusetts Institute of Technology).

[16] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, *36*(2), 2592-2602.

[17] Wouter, B., & Van den Poel, D. (2005). Customer base analysis: Partial defection of behaviorally-loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*(1), 252-268.

[18] Kim, Y., & Street, W. N. (2004). An intelligent system for customer targeting: a data mining approach. *Decision Support Systems*, *37*(2), 215-228.

[19] Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS quarterly*, 553-572.

[20] Mierswa, I., & Klinkenberg, R. (2018). RapidMiner Studio (9.1) [Data science, machine learning, predictive analytics]. Retrieved from https://rapidminer.com/

[21] Karim, M., & Rahman, R. M. (2013). Decision tree and naive bayes algorithm for classification and generation of actionable knowledge for direct marketing. *Journal of Software Engineering and Applications*, *6*(04), 196.

[22] Larivière, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, *29*(2), 472-484.

[23] Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics.