Machine learning or Statistical discrete response modelling?

Webshop conversion rate maximization at Fatboy.com

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Barbara Sophia Claudia Linssen
Student number: 4211995

To be defended in public on March, 7th 2019

Graduation committee

Chairperson	Prof. Dr ir C.G. Chorus, Transport and Logistics
First Supervisor	Dr ir S. van Cranenburgh, Transport and Logistics
Second Supervisor	Prof. Dr ir M.F.W.H.A. Janssen, Information and Communication Technology
Internal Advisor	PhD candidate ir A. Alwosheel, Transport and Logistics
External Supervisor	ir J. Hulst, SparkOptimus
External Supervisor	ir J. Joubert, SparkOptimus
External Supervisor	MSc. T. de Vos, Fatboy®

Comprehensive summary

Adaptive website content is an increasingly popular method for e-commerce platforms to personalize the content shown to its visitors in an attempt to increase the conversion rates of their platforms. However, in order to select the content relevant for a specific customer, the online platform should be able to rapidly interpret the explicit choices, as well as the implicitly sensed context, in which the choices are made. Creating a so-called context-aware e-commerce platform thus requires not only the understanding of users' content-preferences, but also of the users' context influencing this decision-making and the opportunities of the e-commerce platform to actively respond to this, in addition to analysis of customer behaviour that goes beyond traditional A/B testing used by websites and online services to quickly evaluate hypothesis with real users.

Statistical choice modelling and machine learning are both methods with the ability of doing such advanced analysis of choice behaviour. By inferring preferences and trade-offs from people's choices observed in real life, or hypothetically stated choice experiments, choice models can be estimated with which future choices can be predicted. As statistical choice models provide insight into the most important parameters that influence decision-making, conversion rates of e-commerce platforms can be maximized by optimizing those parameters. Moreover, it has long been acknowledged in discrete choice literature that the context in which a decision is made affects one's decision-making. Machine learning on the other hand, takes an algorithmic approach and bases its prediction on patterns found in the data. The characteristic of machine learning to allow for high-order interactions that are not pre-specified, could be beneficial as the context in which decisions are made, is often fuzzy and high-dimensional. However, as this is a black box method, the method will not explicitly give insight into the parameters influencing behaviour, but instead, train itself to maximize the outcome.

Although a substantial amount of research is conducted on the difference between statistical choice models and machine learning, and A/B testing and machine learning are both widely used methods for the dynamic content selection on e-commerce platforms, and thus for conversion optimization, less research however, has been done on the complementary use of the two methods within the boundaries of a simple context-aware system-experiment. Nowadays, it is often stated that machine learning has an accuracy far beyond statistical methods and that it is the recommended method for future data analytics and prediction, as machine learning enables determination of outcomes in which large number of variables with complex relationships are involved. However, should machine learning always be the recommended method for an e-commerce platform trying to maximize its conversion rate with the use of dynamic content, or is more traditional statistical choice modelling sometimes still a better solution for smaller organizations less experienced in data analytics?

The scientific objective of this study was to investigate the applicability of both statistical discrete response modelling and machine learning methods to maximize, separately or complementary, conversion rate on e-commerce platforms. By using the e-commerce platform Fatboy.com as test environment, in which both methods were required to handle the context-aware opportunities of contemporary online platforms, the potential in applicability of both methods could be examined. The objective within the case study Fatboy® was to understand how the conversion rates of e-commerce platforms can be maximized using statistical discrete response modelling, machine learning, or complementary to each other.

Firstly, the e-commerce platform, including the current interaction of the platform with its users, was examined in order to understand the context-aware systems opportunities. Secondly, it was examined which different webpage interfaces could be constructed to use as choice sets to present to the webshop visitors and which context settings of the visiting users could be tracked in order to analyse customer decision-making under different context settings and whether customer decision-making is influenced by the customers' context. After running several data collection experiments, in which the customer decision-making was analysed under different context settings, the resulting dataset, containing 2,292 observations, was analysed with both statistical discrete response models, as well as machine learning's supervised decision tree and artificial neural network classification, and unsupervised clustering to investigate the applicability of both methods. This resulted in real-case illustrations on the applicability of both statistical discrete response modelling and machine learning for conversion rate maximization, in which each method was evaluated for its applicability for conversion rate maximization,

interpretability, complexity, and predictive performance, and a reflection on the potential of the complementary use of statistical analyses and machine learning methods. Finally, practical recommendations for Fatboy® were formulated regarding the use of dynamic content for conversion rate maximization.

Table 1 provides the summary on the applicability for conversion rate maximization per method.

The statistical discrete response model, used to examine visitors' preferences for webshop content on their decision to stay longer than 30 seconds in the webshop or not, and how these preferences were influenced by their contextual circumstances, proved valuable for conversion rate maximization in its ability to estimate the relative preference for accepting specific content variations versus the preference for the alternative of leaving the platform and to estimate the influence, in direction and strength, of certain context situations on these preferences. The method is however, less capable of instantly combining various, not specifically predefined, context variables to determine which content should be shown.

To examine the applicability of a black-box method versus a more interpretable supervised machine learning method, decision tree classification and artificial neural network classification were both used to classify the webshop visitors into visitors who probably will, or will not, convert. Decision tree classification proved valuable for conversion rate maximization in its ability to indicate an order of influence of context variables on the probability of a visitor to convert, or not, and to show which combinations of context variations result in the highest probability of visitors to convert. Comparing decision tree classification with discrete response modelling, the results show how the discrete response models provided more detailed insight into the influence of a certain variable on another variable, while the added value of decision trees lies in their ability to indicate an order of the influence of context variables and to present how combinations of variables led to a certain result. Similar to decision tree classification, artificial neural network classification showed valuable applicability for conversion rate maximization in its ability to combine multiple variables simultaneously, high-dimensionally, without having to predefine all the effects beforehand.

Clustering was used as an unsupervised learning technique to find homogenous subgroups among Fatboy®'s webshop visitors: whether associations could be found within the context situations of customers who did prefer similar content. This thesis shows how it is possible to define customer subgroups within Fatboy®'s webshop visitors and what the context-similarities are per customer subgroup. However, no insights into the weights of context variables, e.g. the importance of those contextual circumstances relative to each other, is provided by the resulted clusters.

Table 1: Summary of the applicability of the investigated methods to maximize conversion rate

	Statistics	*Machine Learning*		
	Discrete response modelling	Decision tree classification	Artificial neural network classification	K-modes clustering
Applicability for conversion rate maximization	▪ Ability to estimate the relative preference for accepting specific content variations versus the preference for the alternative of leaving the platform. ▪ Ability to estimate the influence, in direction and strength, of certain context situations on these preferences.	▪ Ability to indicate an order of influence of context variables on the probability of a visitor to convert or not. ▪ Ability to show which combinations of context variations result in the highest probability of visitors to convert for a specific content.	▪ Ability to combine multiple variables simultaneously, while predicting whether a visitor will convert or not.	▪ Ability to find homogenous subgroups among webshop visitors: whether associations could be found within the context situations of customers who did prefer similar content.

The strength of discrete response modelling lies in the interpretability of the results, providing detailed insight into the influence of a certain variable on another variable. However, a disadvantage is that the method is less capable of combining variables without being specifically predefined for, which, regarding the applicability of the method for the use of context-aware webshop content for conversion rate maximization, is precisely the capability to look for: desirably, the model should be able to instantly combine all the available context factors in order to predict the best content leading to conversion rate maximization.

The strength of artificial neural network classification lies in its predictive performance, resulting in perfectly fitting models by combining multiple variables simultaneously without having to predefine all the effects beforehand. However, the main disadvantages are that the method is a black box, and thus less interpretable than discrete response models and decision trees, and that the method is not suitable for small datasets due to overfitting.

On all three criteria, decision tree classification can be positioned between statistical discrete response modelling and neural network classification. Decision trees are interpretable, providing detailed insight into how combinations of variables led to a certain result, and are able to combine multiple variables simultaneously, high-dimensionally, without having to predefine all the effects in advance. However, the predictive performance of decision trees is lower than the artificial neural network, and more similar to statistical discrete response modelling. Moreover, in terms of complexity, decision trees are less capable of handling unbalanced datasets: poor distribution of labelled target variables substantially impacts the classification accuracy of a decision tree.

Table 2 provides a brief summary of the comparison of the different methods applied (the complete analysis on the methodological differences is included in the conclusion of this thesis).

Table 2: Summary of the investigated methods to maximize conversion rate

	Statistics	*Machine Learning*		
	Discrete response modelling	Decision tree classification	Artificial neural network classification	K-modes clustering
Interpretability	▪ Highly interpretable, detailed insight into the influence of a certain variable on another variable	▪ Interpretable, detailed insight into how combinations of variables led to a certain result	▪ Less interpretable. Black box	▪ Interpretable, provides insight into how variables are associated with each other
Complexity	▪ Relationships between variables are required to be predefined beforehand. ▪ Better capable of handling smaller amounts of data	▪ Able to combine multiple variables simultaneously, high-dimensionally, without having to predefine all the effects beforehand. ▪ Less capable in handling an unbalanced dataset, and less capable in handling a smaller dataset	▪ Able to combine multiple variables simultaneously, high-dimensionally, without having to predefine all the effects beforehand. ▪ Not suitable for small datasets due to overfitting, and less capable in handling an unbalanced dataset	▪ Able to identify associated variables without having to predefine expected associations beforehand
Predictive performance	▪ Good. Accuracy rate of 77.8%	▪ Good. Accuracy rate of 80.8%. ▪ Unbalanced dataset substantially impacts the classification accuracy of a decision tree while it is better capable than a neural network in handling a smaller dataset	▪ In principle, excellent: accuracy rate of 100.0% on training data. However, overfitting issue, network failed to hold a similar performance on test data: accuracy of 74.1% on test data. Method is not suitable for handling small datasets	

Logically, the potential use of the examined methods lies in the complementary use of both methods to benefit from the strength of both: statistical discrete response modelling provides the required insights to understand the behaviour of platform the visitors, while machine learning offers the ability to predict which content should be shown to which visitor, based on the visitor's high-dimensional context situation. Prediction without understanding is not desired as no insights, for instance for the production of new webshop content, would then be achieved, while vice versa, understanding without prediction is not desired as this is the basis behind the use of dynamic content for conversion rate maximization. Clustering can be used to identify homogenous groups within the platform visitors, whereas statistical discrete response modelling can be used to identify the specific influences of context and content preference on a visitor's probability to convert. Combining the two, for example by making discrete response models per visitor cluster, should lead to more specified content recommendations for each specific customer segment. Similarly, it is shown how statistical discrete response modelling can result in detailed insights into the importance of a context factor and a visitor's content preference on his or her preference for an alternative, while supervised classification can result in accurate prediction of the conversion probabilities of visitors. Combined, the insights achieved from statistical discrete response modelling on the most important context variables influencing behaviour can be used to optimize the classification model. And similarly, the insights achieved from statistical discrete response modelling, on for example content preferences, can be used for the production of new webshop content, to be in turn included in the predictive classification model.

Recommendations for future research are: (1) to examine the impact of changing to a context-aware system, as more in-depth research should be done into both the technical consequences of such a change, including the IT infrastructure requirements, as well as the business-related consequences of changing to a context-aware system, including the impact on a company's business model and a company's organizational structure; (2) to elaborate on the complementary use of statistical discrete response modelling and machine learning in order to result in a real-case illustration on the complementary, or even mathematically combined, use of both methods for conversion rate maximization, for example by making discrete response models per visitor cluster in order to result in more specified content recommendations per customer segment; (3) to elaborate on the difference in applicability of statistical discrete response modelling and machine learning for conversion rate maximization, as a study with more observed visitors, more observed contextual factors, and more elaborated models could result in more in-depth insights on the applicability of both methods, as well as a more fair comparison between both method; (4) to validate the results in order to substantiate the positive effects of the recommended content changes per customer segment on the client's conversion rate; (5) to investigate whether the threshold of staying more than 30 seconds on the platform can truly be regarded as the best proxy for conversion, or whether another proxy leads to a more accurate representation of conversion; (6) to investigate the optimization of full online customer journeys, as conversion rate maximization cannot be achieved by only optimizing the homepage; and (7) to examine whether the focus on conversion rate maximization over platform engagement is proportionate, or whether a more balanced focus between conversion rate maximization and platform engagement is desirable. Each recommendation for future research is extensively described in the recommendation chapter of this thesis.

Table of content

2.

Literature research, knowledge gap and research questions

An extensive amount of research is written on both statistical choice modelling and machine learning for predictive usage (Arentze, Dellaert, & Chorus, 2015; Edwards, New, & Parker, 2012; Ganapathi et al., 2009; Oppewal, Tojib, & Louvieris, 2013). Applied in all kinds of research domains, these studies mostly discuss specific case studies, in which one of both methods is applied (Abu-Nimeh, Nappa, Wang, & Nair, 2007; De Bekker-Grob & Chorus, 2013; Hansen et al., 2013; Molin, Meeuwisse, Pieters, & Chorus, 2018). The studies focussing on comparison of statistical choice modelling versus machine learning prediction often describe the difference in characteristics of statistical models versus machine learning (Ebner, 2016; Harrell, 2018; Levy, 2018), or focus on the comparison of one specific purpose, e.g. for classification (Lim, Loh, Shih, & Cohen, 2000). For the literature search, these search strategies were combined, as separate search strings yielded far too many hits.

2.1 Literature search methodology

Multiple resources were used to gather scientific papers relevant for this study. The search on Google Scholar, for papers in English or Dutch, was performed up to and including 09-09-2018. The keywords “machine learning” AND “choice modelling” AND “conversion rate” resulted in eight papers, of which two were selected for discussion. The keywords “machine learning” and “A/B testing” AND “adaptability” resulted in 89 hits (excluding patents), of which one paper was selected for discussion. Information was only regarded as relevant when explicitly addressing differences in characteristics of the two methods, differences in predictive performance, differences in usage, or addressing complementary usage of both methods. Knowing that Sander van Cranenburgh and Ahmad Alwosheel studied the combined use of artificial neural networks (ANN) and statistics, the query “modelling" AND "machine learning" AND “Cranenburgh” AND “Alwosheel” resulted in two papers, which were both included in this research. Additionally, papers on context-aware systems were provided by Prof. Dr ir Marijn Janssen (TU Delft); a paper on the predictive performance of machine learning versus statistical models was provided by Prof. Dr ir Caspar Chorus (TU Delft); papers on the difference between explanatory and predictive modelling and on the data modelling and the algorithmic modelling culture was provided by PhD candidate Ahmad Alwosheel (TU Delft); and a paper on the combined use of machine learning’s decision trees and discrete choice was provided by Dr ir Sander van Cranenburgh (TU Delft). After inclusion of all papers, cross-referencing was performed to complete the search. Blogs were also used to achieve information, so-called ‘grey literature’, as blogs are nowadays often used by scientists to spread their ideas and knowledge.

Figure 1 presents an overview of the literature selection process. An overview of the included studies for the literature research to examine existing knowledge gaps in included in Appendix I.

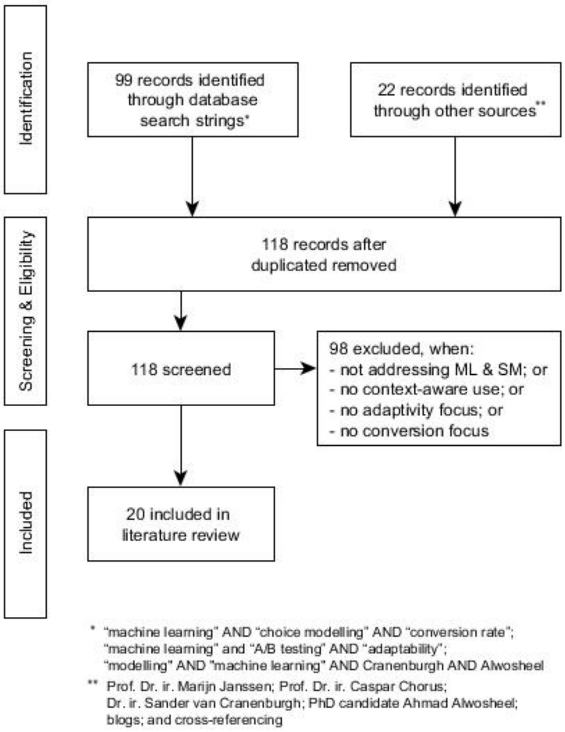

Figure 1: Literature selection overview

2.2 Context-aware systems

As described previously, context and context-aware systems in this study are defined as stated by Dey & Abowd (1999): context-aware system being "a system that "uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task" (p. 6); context being "any information, either explicitly or implicitly indicated by the user, that can be used to characterize the situation of a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves" (pp. 3-4). Van Engelenburg, Janssen, & Klievink (2018) make a split between context and its context variables, defining a context variable as "an attribute of an object that is relevant" (p. 101), and context as "the set of context variables" (p.101). They explicitly describe the criteria for an attribute of an object to be a context variable and provide a method for determining whether the criteria are met and the variable belongs to the context (Van Engelenburg et al., 2018).

Khedo (2006) describes that the way context-aware applications make use of context can be categorized into three classes: (1) presentation of context information to the user or use context to propose actions to the user; (2) automatic execution of a service on behalf of the user according to context changes; and (3) attachment of context information for later retrieval. According to Khedo (2006), most context-aware systems fall into the first category. The challenge of context-aware systems, according to Khedo (2006), is, that the systems have to gather context information and process it in such a way that it is meaningful to the context-aware application, as the context data is often noisy and ambiguous. As Khedo (2006) explains: "Incorrect sensing of context, or inappropriate reactions to context, can be as great a problem as insensitivity to context" (p. 3).

Benou & Vassilakis (2010) focused their research specifically on context-aware mobile commerce applications. They define mobile commerce, or m-commerce, as "any activity related to a commercial transaction (or a potential one) – a transaction that includes a monetary value – and is conducted via wireless and mobile communication networks and uses wireless and mobile devices as user interface" (p. 140). Benou & Vassilakis (2010) explain that context-aware mobile commerce applications should be able to adapt their interface, services, and content towards a certain context, but that the design of such applications is therefore difficult, as a mobile context is far more diverse than that of a stable desktop. Benou & Vassilakis (2010) make a distinction between context and context information. Context is seen as "the set of all possible conditions and states that surround an electronic commerce operation" (p. 143), whereas context information is defined as "the set of data elements comprising the operation context; every piece of information which may be used to characterize a state of an entity, which can be considered to be relevant to the interaction of the user with the particular application" (pp. 143-144).

According to Zimmerman, Lorenz, & Oppermann (2007), context information can be classified into five categories: (1) individuality context, which can be observed about an entity, such as an entity's state, or individual tastes and preferences for language or modality of interaction; (2) time context, like the entity's time zone or current time of the day; (3) location context, such as an entity's country or work environment; (4) activity context, which includes the activities the entity is currently, or in future, involved in; and (5) relations context, which covers the relations of an entity to other entities. In other words, these five categories of context information describe the dimensions of how context can be of influence on customers' decision-making. Wang (2004) describes the context of a content-aware mobile learning (CAML) system by six dimensions: identification of learner, spatio-temporal dimension (including time and location), facility dimension (the mobile devices that is used for learning), learning activities (such as status of assignments), learner dimension (such as her of his emotional state), and community dimension (interaction with others). As a CAML should detect the learning contexts of a student and should adapt to changes in the student's learning environment, Wang (2004) describes four modes of interaction through context adaption: spatio-temporal dependent interface (adapt interface according to time and location contexts), contextual event notification (send notification learner as reminder to finish activities), context-aware communication (allow communication to interrupt learner only when learning context allows it), and navigation and retrieval of learning materials (deliver accurate learning materials to learner). Wang (2004) concludes however, by emphasizing that various challenges still need to be overcome before CAML is achieved, as for instance reliably sensing the context of learners via sensors.

2.3 Statistical choice modelling and machine learning: the differences

Harrell (2018) emphasizes three main differences between statistical models and machine learning: 1) "statistical models take uncertainty into account by specifying a probabilistic model for the data" (para. 1); 2) "statistical models typically start by assuming additivity of predictor effects when specifying the model" (para. 1); and 3) "machine learning is more empirical, including allowance for high-order interactions that are not pre-specified, whereas statistical models have identified parameters of special interest" (para. 1). Harrell (2018) lists several considerations to consider when choosing one of the two models. He recommends choosing a statistical model when uncertainty is inherent, isolation of the effects of variables is desired, or if the entire model should be interpretable (Harrell, 2018). He recommends machine learning when overall prediction is desired over understanding the impact of specific variables, and if the model being a black box is not a problem.

This difference between statistical models' 'white box' modelling and machine learning's 'black box' modelling is clearly explained by Breiman (2001b): statistical models assume that the data are generated by a given stochastic data model: "data are generated by independent draws from response variables = f (predictor variables, random noise, parameters); the values of the parameters are estimated from the data and the model then used for information and/or prediction" (p. 199) (fig. 2).

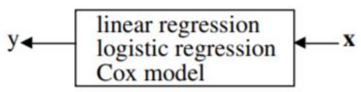

Figure 2: Example data model of data modelling culture (Breiman, 2001b, p. 199).

Machine Learning models, on the contrary, consider the inside of the box complex and unknown: "their approach is to find a function f(x) – an algorithm that operates on x to predict the responses y" (Breiman, 2001b, p.199) (fig. 3).

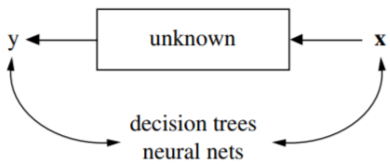

Figure 3: Example data model of algorithmic modelling culture (Breiman, 2001b, p. 199).

The difference between inference ("the creation of a mathematical model of the data-generation process to formalize understanding or test a hypothesis about how the system behaves" (p. 233)) and prediction ("forecasting unobserved outcomes or future behaviour, which does not require understanding of the underlying mechanisms" (p. 233)) is explained by Bzdok, Altman, & Krzywinski (2018), and they state that statistics as well as machine learning can be used for both inference as well as prediction. The main difference is that statistical models are often project-specific, "allowing to compute a quantitative measure of confidence that a discovered relationship describes a true' effect that is unlikely to result from noise" (p. 233), while machine learning uses "general-purpose learning algorithms to find patterns in often rich and unwieldy data" (p. 233). As Bzdok et al. (2018) explain: "statistics requires us to choose a model that incorporates our knowledge of the system, and ML requires us to choose a predictive algorithm by relying on its empirical capabilities" (p. 234). Bzdok et al. (2018) argue therefore, that statistical inferences become less precise as the included number of input variables and possible associations among them increase, and machine learning in that case could be the desired method.

Shmueli (2010) also emphasizes the difference between exploratory modelling and predictive modelling, explaining that both play a different role in generating and testing theories, and thus require different scientific usage. Shmueli (2010) defines explanatory modelling as "the application of statistical models to data for testing causal hypotheses about theoretical constructs" (p. 291) and predictive modelling as "the process of applying a statistical model or data mining algorithm to data for the purpose of predicting new or future observations" (p. 291).

Makridakis, Spiliotis, & Assimakopoulos (2018) studied the differences in predictive performance between machine learning and statistical models, evaluating in their research the performance of several machine learning methods with statistical methods across multiple forecasting horizons, using a large subset of 1045 monthly time series. As Makridakis et al. (2018) describe: "the objective of machine learning methods is the same as that of statistical ones. They both aim at improving forecasting accuracy by minimizing some loss function, typically the sum of squared errors. Their difference lies in how such a minimization is done with machine learning methods, utilizing non-linear algorithms to do so, while statistical ones use linear processes. Machine learning methods are computationally more demanding than statistical ones, requiring greater dependence on computer science to be implemented, placing them at the intersection of statistics and computer science" (p. 2). In their research, Makridakis et al. (2018) compared both methods on four factors: symmetric Mean Absolute Percentage Error (sMAPE), Mean Absolute Scaled Error (MASE), Computational Complexity (CC), and Model fit (MF). Their result showed that machine learning outperformed the statistical methods in accuracy for all examined horizons and that the computational requirements of machine learning are greater than those of statistical methods.

Mottini & Acuna-Agost (2018) focused their research on modelling air passenger choices of flight travel plans. They start with explaining the advantages and the shortcomings of multinomial logit (MNL) models, traditionally used for this purpose. The mentioned advantages of MNL models are: (1) simplicity; (2) general good performance; and (3) and ease of interpretation (Mottini & Acuna-Agost, 2018). The mentioned shortcomings are: (1) the linear combination of the input features considered by the model; (2) its Independence of Irrelevant Alternatives (IIA) property; and (3) its inability to take the order of alternatives into account (Mottini & Acuna-Agost, 2018). They present therefore a deep choice model using pointer networks to predict the alternative that is going to be selected when presented a user a sequence of alternatives. After evaluating the model on a dataset of matched airline bookings and online search logs, they concluded that the proposed model outperforms the traditional MNL model (Mottini & Acuna-Agost, 2018).

Van Cranenburgh (2018) discussed in his comparison between statistical, theory-driven, models, and machine learning, data-driven, models the strengths and weaknesses of both modelling paradigms. The strengths of the statistical models, including its good interpretation (due to the model parameters, confidence intervals, and statistical measures for model fit comparison) and transparency, are contra versa the weaknesses of the machine learning models (Van Cranenburgh, 2018). Similarly, the weaknesses of statistical choice models, including being restrictive on data, mostly a lower prediction performance, and the various assumptions made to generate the model, are the contra versa the strengths of machine learning models (Van Cranenburgh, 2018). As described by Van Cranenburgh (2018), the assumptions made when generating a statistical model are various, ranging from assumed decision rules, the involved attributes, interaction-effects, and error term distributions to random sample data collection,, or for example the assumption than an individual has well-defined preferences (Van Cranenburgh, 2018).

2.4 Statistical choice modelling and machine learning: the similarities

Ebner (2016) argues that machine learning and statistics are actually quite similar, as both focus on the question how to learn from data, and that both only tend to emphasize different things: machine learning focusses more on computers and systems, using terminology as 'weights' and often software Python or Matlab, while statistics, being more a mathematical discipline, uses terminology as 'parameters', and often software R. Ebner (2016) also argues that, despite these differences, both methods are eventually fundamentally similar as they have the same purpose. Levy (2018) on the contrary, argues that the issue of a false dichotomy is moot, as statistical models and machine learning are shown to be different, but that a better question would be how to combine the two methods, the conditions for combining and for which purposes.

2.5 Statistical choice modelling or machine learning: complementary or combined approach?

Less research, however, is written on the combined use of the two methods within the boundaries of a simple context-aware system-experiment. This is interesting as A/B testing and machine learning are both widely used methods for the dynamic content selection on e-commerce platforms, and thus for conversion optimization (Krishnan, 2016); (Serdyukov, 2017); (Urban, Sreenivasan, & Kannan, 2016); (Moatti, US Patent No. Pub No: US 2007/0124192 A1, 2007). Nowadays, it is often said that machine learning has an accuracy far beyond statistical methods and that it is the recommended method for future data analytics and prediction (Fernández-Delgado, Cernadas, Barro, Amorim, & Amorim Fernández-Delgado, 2014; Wainer, 2016). The inventors Miikkulainen & Iscoe (U.S. Patentnr. US2017/0193367A1, 2017) explain that conversion optimization includes testing multiple variations of webpages at the same time, which often lead to complex combinations that all need to be analysed in order to determine the most user-engaging combination of webpage elements. According to Miikkulainen & Iscoe (U.S. Patentnr. US2017/0193367A1, 2017), machine learning systems are therefore useful as they enable running tests to determine outcomes in which large number of variables with complex relationships between them are involved. In (U.S. Patentnr. US2017/0193367A1, 2017), artificial neural networks are therefore used to identify the most successful webpage designs. However, should machine learning for an e-commerce platform, trying to maximize its conversion rate with the use of dynamic content, always be the recommended method, or is more traditional statistical choice modelling sometimes still a better solution for smaller companies less experienced in data analysis?

Van Cranenburgh & Alwosheel (2017) combined machine learning and statistical modelling. By training pattern recognition ANNs, they were able to detect patterns in sequences of choices that are more likely to be associated with certain decision rules than others. This enabled them to distinguish RUM decision-makers (Random Utility Maximization decision-makers, whose choices are driven by the wish to maximize the expected utility (Chorus, 2012)) from RRM decision-makers (Random Regret Minimization decision-makers, whose choices are driven by the wish to minimize the anticipated regret (Chorus, 2012)) based on an observed sequence of choices. Alwosheel, Van Cranenburgh, & Chorus's (2018) research into the sample size requirements showed that a minimum sample size of fifty times the number of weights in the ANN is recommended when using artificial neural networks for discrete choice analysis.

Brathwaite et al. (2017) emphasizes that machine learning methods for modelling discrete choices have never been used in combination with economic theories of human decision-making, and they hypothesize that this is because machine learning methods are considered to be 'black-boxes, that lack a theoretical basis for interpreting and understanding human behaviour' (p. 2). Their paper aims to contribute to literature by (1) trying to connect machine learning's decision trees to economic theory; (2) combining discrete choice models with decision trees; and (3) applying both their decision tree model as their statistical model on a case study (Brathwaite et al., 2017). Brathwaite et al. (2017) show that machine learning's decision trees can make probabilistic predictions and deal with heterogenous non-compensatory rules, estimation uncertainty, context-dependent preference heterogeneity, and monotonicity, but that there are no decision trees that can account for all these considerations simultaneously. The Bayesian model tree proposed in their paper accounts for estimation uncertainty and context-dependent preference heterogeneity (Brathwaite et al., 2017). The outcomes of their research is that (1) their Bayesian model tree was far more accurate than the multinomial logit (MNL) model; (2) forecasting by the Bayesian model tree outperformed forecasting by the MNL model; and (3) qualitatively different insights were provided by the Bayesian model tree versus the MNL model (Brathwaite et al., 2017).

Cottrell, Girard, Girard, Mangeas, & Muller (1995) combine a statistical stepwise method for weight elimination with neural modelling for time series, by comparing the use of a statistical stepwise method for weight elimination with other pruning techniques and applying it to artificial series (Cottrell et al., 1995). They found that the statistical stepwise method is well capable for eliminating nonsignificant weights.

Cruz-Benito et al. (2018)'s study focusses on the goal of getting user's complete large questionnaires by making the web forms adapt to user preferences and behaviour to show the user the version of the questionnaire that most fits the user's profile. With the use of A/B testing, user preferences for web forms were detected. With the use of machine learning, users with similar characteristics were clustered and their performance in completing webforms was investigated. With this information, Cruz-Benito et al. (2018) generated guidelines to lead users to the most adequate version of a questionnaire fitting their user profiles. Cruz-Benito et al. (2018) explains that they used a Random Forest classifier algorithm, as their variable to predict was of categorical level, as thus classification was required. The results were promising that adapting web forms to user profiles for questionnaires can increase the completing of large questionnaires.

2.6 Knowledge gap and research objectives

The most compelling and clear research gap lies in the complementary – or combined – use of machine learning and statistical choice models. In order to investigate the possibilities of such a complementary use, this study has examined both methods for conversion rate maximization of the e-commerce platform Fatboy.com. This was an interesting test environment as it required both methods to handle the context-aware opportunities of nowadays online platforms.

The scientific objective of this study was to investigate the applicability of both statistical choice behaviour analysis and machine learning methods, separately or complementary, to maximize conversion rate on e-commerce platforms.

The objective within the case study Fatboy® was to understand how the conversion rates of e-commerce platforms can be maximized using statistical choice modelling, machine learning, or complementary to each other.

The study resulted in real-case illustrations on the applicability of both statistical choice modelling and machine learning for conversion rate maximization, in which each method was evaluated for its applicability for conversion rate maximization, interpretability, complexity, and predictive performance, and a reflection on the potential of the complementary use of statistical analyses and machine learning methods. Finally, practical recommendations for Fatboy® were formulated regarding the use of dynamic content for conversion rate maximization.

2.7 Sub questions and research approach

In order to answer the main research question, five sub questions were formulated. First, full understanding of the e-commerce platform, including the current interaction of the platform with its users, was necessary to understand the context-aware systems opportunities.

1. How does the e-commerce platform currently interact with its visiting users, and what are the context-aware systems opportunities?

Secondly, customer decision-making must be analysed under different context settings and different system behaviour to be able to examine whether customer decision-making is influenced by the customer's context. Therefore, different webpage interfaces must be constructed to use as choice sets to present to the webshop visitors and which context settings of the visiting users could be tracked.

2. Which different webpage interfaces should be constructed to use as choice sets in data collection experiments and under what context variables should they be tested?

After running the data collection experiments, in which customer decision-making was analysed under different context settings, the data must be analysed with both statistical choice models as well as machine learning methods to investigate the applicability of both methods for conversion rate maximization of e-commerce platforms with the use of context-aware webshop content.

3. What is the applicability of the statistical choice model and the machine learning models for conversion rate maximization, including their predictive performance and potential for further use?

4. What is the potential of the complementary use of statistical analyses and machine learning methods?

This resulted in real-case illustrations on the applicability of both statistical choice modelling and machine learning for conversion rate maximization, in which each method was evaluated for its applicability for conversion rate maximization, interpretability, complexity, and predictive performance, and a reflection on the potential of the complementary use of statistical analyses and machine learning methods.

Final steps were to compare the applicability of both methods and to evaluate the use of dynamic content for conversion rate maximization in practice.

5. How do the results of the statistical choice model and the machine learning methods differ and what are the key learnings and recommendations for e-commerce platforms on how to use dynamic content for conversion rate maximization in practice?

The e-commerce platform of Fatboy® has been used as a case study to explore in depth the possibilities of both methods. The case study approach resulting in recommendations on how to optimize Fatboy®'s e-commerce platform was combined with a modelling approach in which machine learning and statistical choice modelling both were applied. As this study has combined both qualitative research as well as quantitative research, it can be seen as mixed-methods research.

Qualitative research has been used to (1) understand the current status of Fatboy®'s e-commerce platform, the interaction with its visitors, and its environment; (2) understand the current status of Fatboy®'s conversion optimization process; and explore (3) the possibilities of dynamic content; and (4) the possibilities of context-aware systems.
Quantitative research has been be used to (1) investigate the applicability of statistical choice modelling to maximize conversion rate, (2) investigate the applicability of machine learning to maximize conversion rate, and (3) investigate the effect of complementary use of both methods. The advantage of this mixed-methods research approach is that the quantitative research has been executed after profound understanding of the contextual setting in which the quantitative research is applied (Creswell & Clark, 2011a). This was important to eventually provide useful answers to the research questions and recommendations for Fatboy®, which would not have been possible using qualitative or quantitative research alone (Creswell & Clark, 2011a). The challenge was to keep the study within the set time frame, as mixed-methods research requires time reserved for both the qualitative, as for the quantitative, research (Creswell & Clark, 2011a). The combined use of qualitative and quantitative research is visualized in the Research Flow Diagram in Appendix II. The mixed-methods research design can be best classified

as an exploratory design, as the qualitative results were used to make decisions about the quantitative research (Creswell & Clark, 2011b). Final notes where has to be aware of during the taken approach were that: (1) at the end of the research the specific case study results had to be used to develop a more general proposition or theory about the use of statistical choice modelling and machine learning for conversion rate maximization; (2) during the study, the case study-specific factors influencing a general theory had to be noted in order to take them into account at the end of the research; (3) a model will always be a simplification of the real-world, taking a limited number of variables into account.

2.8 Reading guide

This thesis is structured as follows:

Chapter 3 – System and contextual understanding, addresses how the system, Fatboy®'s e-commerce platform, currently interacts with its visiting users and what the context-aware systems opportunities are for Fatboy.com. Section 3.1 gives insight into the structure of a webshop. Section 3.2 summarizes Fatboy®'s current state of conversion optimization and user interaction. Section 3.3 describes Fatboy®'s e-commerce platform as possible context-aware system, including how context variables could influence a customer's decision-making, the opportunities of using dynamic content, and how the context data can be used for dynamic content use, resulting in a context-aware system.
Chapter 4 – Experimental setup, describes the test experiment set up together with Fatboy®, to analyse the choice behaviour of their webshop visitors when different content was presented, similarly collecting the context setting in which decision-making was made. Section 4.1 describes the context variables that were measured for each platform visitor and section 4.2 the constructed webpages used as choice sets in the experiment to analyse different customer preferences and behaviour. Section 4.3 describes the data collection process.
Chapter 5 – Methodology, describes the conducted data investigation in section 5.1, as well as the chosen statistical choice modelling and machine learning approach in section 5.2. Section 5.2.1 describes the methodology followed for the discrete response model, including theoretical background on discrete response modelling, the applicability of the method within this study, the model generation and estimation process, and the chosen data preparation steps. Section 5.2.2 describes the methodology followed for the machine learnings models, including theoretical background on machine learning, the applicability of machine learning within this study. The model generation and estimation process and the taken data preparation steps are discussed per machine learning model: 5.2.2.1 focusses on decision tree generation, 5.2.2.2 on artificial neural network generation, and 5.2.2.3 on cluster generation. Section 5.2.3 describes the method evaluation criteria.
Chapter 6 – Results, describes the results. Section 6.1 describes the results related to the estimated discrete response models, section 6.2 the results related to the estimated decision tree classification models, section 6.3 the results related to the estimated artificial neural network classification models, and section 6.4 the results related to the estimated clusters.
Chapter 7 comprises the conclusion, Chapter 8 the discussion and Chapter 9 the recommendations for future research. The literature list is included in Chapter 10. The appendices follow thereafter.

10. References

Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). A comparison of machine learning techniques for phishing detection. *European Journal of Marketing*, 60–69. https://doi.org/10.1145/1299015.1299021

Arentze, T. A., Dellaert, B. G. C., & Chorus, C. G. (2015). Incorporating Mental Representations in Discrete Choice Models of Travel Behaviour: Modelling Approach and Empirical Application. *Transportation Science Publication*, *49*(3), 577–590. https://doi.org/10.1287/trsc.2013.0513

Benou, P., & Vassilakis, C. (2010). The conceptual model of context for mobile commerce applications. *Electronic Commerce Research*, *10*(2), 139–165. https://doi.org/10.1007/s10660-010-9050-4

Brathwaite, T., Vij, A., & Walker, J. L. (2017). Machine Learning Meets Microeconomics: The Case of Decision Trees and Discrete Choice. *Preprint Submitted to Elsevier*, 1–39. Retrieved from http://arxiv.org/abs/1711.04826

Breiman, L. (2001b). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–231. https://doi.org/10.2307/2676681

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of Significance: Statistics versus machine learning. *Nature Methods*, *15*(4), 233–234. https://doi.org/10.1038/nmeth.4642

Cottrell, M., Girard, B., Girard, Y., Mangeas, M., & Muller, C. (1995). Neural Modeling for Time Series: A Statistical Stepwise Method for Weight Elimination. *IEEE Transactions on Neural Networks*, *6*(6), 1355–1364. https://doi.org/10.1109/72.471372

Cruz-Benito, J., Vazquez-Ingelmo, A., Sanchez-Prieto, J. C., Theron, R., Garcia-Penalvo, F. J., & Martin-Gonzalez, M. (2018). Enabling Adaptability in Web Forms Based on User Characteristics Detection Through A/B Testing and Machine Learning. *IEEE Access*, *6*, 2251–2265. https://doi.org/10.1109/ACCESS.2017.2782678

Cumming, G. (2013). *Understanding The New Statistics. Effect sizes, confidence intervals, and meta-analysis*. New York: Taylor & Francis Group, LLC.

De Bekker-Grob, E. W., & Chorus, C. G. (2013). Random regret-based discrete-choice modelling: An application to healthcare. *PharmacoEconomics*, *31*(7), 623–634. https://doi.org/10.1007/s40273-013-0059-0

Dey, A. K., & Abowd, G. D. (1999). Towards a Better Understanding of Context and Context-Awareness. *Computing Systems*, *40*(3), 304–307. https://doi.org/10.1007/3-540-48157-5_29

Ebner, J. (2016, May 09). *What's the difference between machine learning, statistics, and data mining?* Retrieved from Sharp Sight: https://www.sharpsightlabs.com/blog/difference-machine-learning-statistics-data-mining/

Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, *49*, 591–603. https://doi.org/10.1016/j.enbuild.2012.03.010

Ganapathi, A., Kuno, H., Dayal, U., Wiener, J. L., Fox, A., Jordan, M., & Patterson, D. (2009). Predicting multiple metrics for queries: Better decisions enabled by machine learning. *IEEE International Conference on Data Engineering Predicting*, 592–603. https://doi.org/10.1109/ICDE.2009.130

Hansen, K., Montavon, G., Biegler, F., Fazli, S., Rupp, M., Scheffler, M., … Müller, K. R. (2013). Assessment and validation of machine learning methods for predicting molecular atomization energies. *Journal of Chemical Theory and Computation*, *9*(8), 3404–3419. https://doi.org/10.1021/ct400195d

Harrell, F. (2018, June 21). *Road Map for Choosing Between Statistical Modeling and Machine Learning*. Retrieved

Levy, D. (2018, May 14). *Navigating Statistical Modeling and Machine Learning*. Retrieved from Statistical Thinking: http://www.fharrell.com/post/stat-ml2/

Lim, T.-S., Loh, W.-Y., Shih, Y.-S., & Cohen, W. W. (2000). A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-three Old and New Classification Algorithms. *Machine Learning*, *40*(1992), 203–229. https://doi.org/10.1023/A:1007608224229

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods : Concerns and ways forward. *PLoS ONE*, *13*(3), 1–26. https://doi.org/10.1371/journal.pone.0194889

Miikkulainen, R., & Iscoe, N. (2017). *U.S. Patentnr. US2017/0193367A1*.

Molin, E., Meeuwisse, K., Pieters, W., & Chorus, C. (2018). Secure or usable computers? Revealing employees' perceptions and trade-offs by means of a discrete choice experiment. *Computers and Security*, *77*, 65–78.

https://doi.org/10.1016/j.cose.2018.03.003

Mottini, A., & Acuna-Agost, R. (2018). Deep Choice Model Using Pointer Networks for Airline Itinerary Prediction. *KDD '17 Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1575–1583. https://doi.org/10.1145/3097983.3098005

Oppewal, H., Tojib, D. R., & Louvieris, P. (2013). Experimental analysis of consumer channel-mix use. *Journal of Business Research*, *66*, 2226–2233. https://doi.org/10.1016/j.jbusres.2012.02.002

Shmueli, G. (2010). To Explain or To Predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.2139/ssrn.1351252

Van Cranenburgh, S. (2018, 10 08). TRAIL 2018. Discrete Choice Analysis: micro-econometrics and machine learning approaches. Artificial neural networks for Discrete choice analysis. Delft.

Van Cranenburgh, S., & Alwosheel, A. (2017). Using artificial neural networks to investigate decision- rule heterogeneity, 1–2.

Van Engelenburg, S., Janssen, M., & Klievink, B. (2018). What Belongs to Context? A De finition, a Criterion and a Method for Deciding on What Context-Aware Systems Should Sense and Adapt to. *Springer International Publishing AG 2018, A. Cerone and M. Roveri (Eds.): SEFM 2017 Workshops, LNCS 10729*, 101–116.

Zimmerman, A., Lorenz, A., & Oppermann, R. (2007). An operational definition of context. In B. Kokinov, D. Richardson, T. Roth-Berghofer, & L. Vieu, *CONTEXT 2007. LNCS (LNAI), vol. 4635* (pp. 558-571). Heidelberg: Springer.