DATA-DRIVEN PRODUCT RETURNS PREDICTION: A CLOUD-BASED ENSEMBLE SELECTION APPROACH

Research in Progress

Leonard Heilig, Institute of Information Systems, University of Hamburg, Hamburg, Germany, leonard.heilig@uni-hamburg.de

Julien Hofer, University of Hildesheim, Hildesheim, Germany, julien.hofer@uni-hildesheim.de

Stefan Lessmann, Humboldt-University of Berlin, Berlin, Germany, stefan.lessmann@hu-berlin.de

Stefan Voß, Institute of Information Systems, University of Hamburg, Hamburg, Germany, stefan.voss@uni-hamburg.de

Abstract

The number of product returns represents a considerable cost factor in e-commerce, especially in the apparel sector. The application of advanced information technologies and predictive analytics, enabling to capture and analyze massive amounts of user data, pave the way for a more efficient management of product returns and reverse logistics. However, we identify a lack of data-driven approaches in this area, especially regarding product returns prediction. In this paper, we present an ensemble selection approach for predicting product returns in the apparel sector. Computational experiments indicate that our approach produces satisfying results in terms of prediction quality. We further explore the correlation between sample sizes and computational times. Thereby, we demonstrate that the run-time increases exponentially when using more data records. To address heavy run-time overheads resulting from high processing and memory requirements of classifiers, we present a framework to embed ensemble selection processes into a highly scalable cloud environment. The framework explains the provisioning of cloud resources and parallelization of tasks according to ensemble selection processes. It further builds a basis for considering data streams, data splitting, and a dynamic adoption of changing customer behavior over time, which has not been considered in related work so far. The envisioned forecasting support system aids retailers in reducing product returns and increasing profit margins.

Keywords: Returns management, predictive analytics, product returns prediction, ensemble selection, cloud computing.

1 Introduction

Product returns are a major problem for e-retailers. The perceived intangibility of products in e-commerce environments increases the likelihood that customers return products after physical inspection (Mukhopadhyay and Setoputro, 2004). This implies high expenditures for reverse logistics, especially in the apparel sector where huge return quotas are common (Accenture, 2012). Compared to buying in brick-and-mortar stores, online customers face a high degree of uncertainty (Liu and Wie, 2003). False decisions on the size of shoes or fit of clothes, e.g., can easily spoil the shopping experience and lead to returns. To prevent mispurchasing, e-retailers employ a variety of approaches including, e.g., rich-media product information and customer-generated reviews. This paper focuses on one specific approach to reduce returns, namely predictive analytics. An accurate prediction of a customer's likelihood to return a product facilitates a number of preventive actions. For example, prior to purchase, a shop system could issue warnings when customers shop for items that do not match their typical buying profile. Consider a customer who commonly buys shirts of size medium. Imagine the customer is about to purchase a shirt from a specific brand A, and assume further that several other customers with the same size have returned items of brand A in the past for them being too tight. Our customer could be presented corresponding information, which might lead to a reconsideration of decisions. Alternatively, based on a prediction that the buying of the shirt is associated with a high return risk, the e-retailer could also try to promote an alternative product with lesser return likelihood. A more invasive approach would be to charge a risk premium through increasing the product price. Predictions may also facilitate post-purchase actions. For example, a customer who displays a high risk of returning could be offered a coupon conditioned by the fact that the product is not returned. The implication of the above examples is two-fold: first, predicting the likelihood of returns gives e-retailers several opportunities to proactively manage returns. Second, such predictions rely on specific attributes of individual customers and products.

The objective of this research is to devise a forecasting support system (FSS) that aids e-retailers in reducing returns. In the last decade, the application of data-driven prediction methods has contributed to increasing sales and improving customer experience in e-commerce (Kohavi, 2001). Surprisingly, research related to predicting returns is scarce. Hess and Mayhew (1997) present a simple adjusted hazard model to describe the time-to-return and compare it with a regression model. The probability of returns is calculated on the basis of the return history of customers without considering other customer- and product-based characteristics. Larger datasets and the inclusion of additional variables would add new insights into the phenomenon of returns. In a recent review of the product returns management literature, Walsh et al. (2014) identify three main research streams. Studies in the first stream emphasize pricing issues, including pricing strategies (Ketzenberg and Zuidwijk, 2009), price-sensitivity factors (Yao et al., 2008), and the effect of product design on pricing and return policies (Mukhopadhyay and Setaputra, 2007). The studies in the other stream focus on handling returns by considering the relationship between manufacturers and retailers (Gurnani et al., 2010) and the structure of reverse channels (Yalabik et al., 2005). Finally, behavioral aspects of online customers (Mollenkopf et al., 2007) or the adoption of virtual try-on for apparel shopping (Kim and Forsythe, 2008) are examined in the third stream. In general, this shows that the prediction of returns is essential and can be linked to activities for returns management and reverse logistics. Moreover, the review evidences the lack of data-driven approaches leveraging customer- and product-based data for predicting and managing product returns.

This paper lays the foundation for the envisioned FSS to aid returns management. More specifically, given the sparsity of empirical insights related to returns, a first question to be answered is that of a suitable modeling methodology. We argue that the prediction engine represents a key component of the FSS. Considering the previous examples how e-retailers can employ predictions of return probabilities, actions to prevent returns – before or after purchase – depend directly on the recommendations (i.e., predictions) of the FSS. This suggests that predictive accuracy will be an important determinant of the effectiveness of the overall FSS. Therefore, the choice of a suitable modeling methodology to generate predictions is a pivotal decision in the design of the FSS. To identify such methodology, we review related literature and perform an empirical study related to predicting returns in the apparel industry. A large body of literature (see, e.g., Shmueli and Koppius, 2010) as well as the popularity of large-scale forecasting competitions (e.g., Kaggle[1]) show that empirical experimentation is a well-established approach to test the relative effectiveness of alternative prediction methods. An empirical comparison of alternative prediction methods is also consistent with Hevner's et al. (2004) design science paradigm in the sense that it facilitates a comprehensive assessment of an IT artifact (i.e., prediction model) in a realistic environment (i.e., using real-world e-commerce data).

To select candidate prediction methods and to frame the overall experiment, we deduce three requirements from the application context of returns management. First, the prediction model should forecast with high *accuracy*. This requirement follows directly from the responsibility of the model to guide preemptive actions for prohibiting returns. Consequently, the economic consequences of such actions can be traced back to model predictions. For example, well-targeted product recommendations can reduce returns and thus the costs of reverse logistics, whereas complicating return processes for customers, incorrectly predicted as likely returners, can reduce sales. Second, the prediction model should display high *scalability*. The justification of this requirement is twofold. First, e-commerce settings require the model to generate predictions with minimal latency for each visitor. For example, observing that a customer is about to purchase a product with high return risk, an e-retailer may wish to dynamically reconfigure the checkout process requiring predictions to be calculated prior to displaying the website. Second, previous literature suggests that accurate predictions require fine-grained data on customer and product characteristics (Walsh et al., 2014). Processing such detailed data is associated with high computational costs, in particular in real-time e-commerce systems generating a tremendous amount of data (e.g., from clickstreams, social media streams) at unprecedented rates. Further, a fundamental problem is that generated models need to efficiently generalize the behavior of the customer, which is subject to changes over time. Consequently, we have to ensure that the used training data is consistent and updated thus requiring a constant generation of new prediction models. Last, *adaptability* is an important requirement the prediction model has to fulfil.

Returns prevention is a managerial decision problem. Hence, costs and benefits are the main measures to judge the success of the envisioned FSS. In particular, its ability to reduce logistics costs for handling returns will depend on two factors: i) the effectiveness of the countermeasures to prevent returns and ii) the accuracy with which the FSS recommends actions for different customers. In general, correctly identifying customers, who would – without treatment – return a product, gives the e-retailer an opportunity to prevent undesirable customer behavior. Consequently, a correct prediction has some economic value. In the same way, incorrect predictions carry a cost (e.g., when losing a sale or customer loyalty due to increasing prices in response to a high return likelihood). Quantifying these costs and benefits is challenging and requires in-depth knowledge of the cost structure and customer base of an e-retailer. However, the prediction methodology underneath the envisioned FSS should be able to account for the costs and benefits of wrong and correct predictions to comply with the business context.

Previous research suggests that the ensemble selection (ES) framework of Caruana et al. (2004) is a suitable approach for returns predictions in e-commerce. More specifically, it has been shown that ES leads to an improved accuracy compared to a single classifier (Tsoumakas et al., 2008), can learn from multiple physically distributed data sets (Tsoumakas et al., 2008), and is able to accommodate the requirements of specific application settings (Baumann et al., 2015). However, the degree to which our second requirement, scalability, can be fulfilled has not been investigated. As we explain in detail below, ES combines many different prediction models whose creation requires considerable computing resources. On the other hand, we conjecture that an appropriate system framework could mitigate computational issues and facilitate the use of ES in e-commerce applications and for returns prediction in particular. This suggests the following research agenda for developing an ES-based FSS. First, we need to confirm the ability of ES to predict returns with high accuracy (i.e., better than challenging benchmark

[1] http://www.kaggle.com.

models). Second, it is essential to devise, implement, and test a highly scalable system framework that facilitates predicting returns in real-time for data streams by parallelizing computations on different levels. Third, examining specific e-retailers, we need to develop a cost function capturing the relative merits and demerits of correct/incorrect returns predictions and test the ability of our ES-based approach to generate predictions in such a way that an actual cost function is minimized. In this paper, we address the first and partially address the second point of our research agenda. We contribute to the literature by examining the effectiveness of ES to predict returns in the apparel industry. A second contribution comes from the fact that our study is, to the best of our knowledge, the first to systematically investigate the resource requirements of ES. This provides important insights related to computational bottlenecks and more generally the design of a scalable FSS for returns management. Last, drawing on the principles of cloud computing, we propose a distributed system framework for the envisioned FSS. Given the scalability of the envisioned framework, we aim to further explore dynamic ES by taking into account the heterogeneity of return reasons (e.g., intended returns, returns due to unsatisfactory products) as well as changing customer behavioral patterns over time. The test of this framework as well as the ability of ES to minimize logistics costs, however, is left to future research.

The remainder of this paper is organized as follows. Section 2 summarizes the ES principle and describes the used experimental setup. In Section 3, we describe the used dataset before presenting the results of the conducted experiments in Section 4. Moreover, we discuss the impact of computational times and elaborate on our cloud-based framework. Finally, we define a path for subsequent research and conclude the paper with a discussion of main findings and implications.

2 Ensemble Selection

The prediction of returns is part of predictive analytics (PA). Previous studies in other domains have evidenced the efficacy of ensemble models (e.g., Bhattacharyya et al., 2011; Lessmann and Voß, 2010; Tsoumakas et al., 2008) and, in particular, of the ES paradigm of Caruana et al. (2004). For setting up our experiments, we define the generated model library of canidate models, performance metrics, and methods for selecting candidate models in the following.

The success of ensemble building strategies relies on the diversity of candidate models (Kuncheva, 2004). Given literature recommendations (e.g., Caruana et al., 2004; Partalas et al., 2010), we create a model library of well-known single classifiers and ensemble learners (see Table 1).

Classification Method	**# Models**	**Meta-Parameter**	**Ensemble Setting**
Classification and Regression Trees (CART)	6	Min. size of nonterminal nodes	{10, 20, 50, 100, 200, 500}
Support Vector Machine with Linear Kernel (SVM LibLin)	11	Regularization factor	$2^{\{-5, \dots, 5\}}$
Logistic Regression (LogR)	1	-	-
Multilayer Perceptron (MLP)	60	Number of layers Hidden nodes Regularization weights	2 1-10 {0.001, 0.003, 0.01, 0.003, 0.1, 0.3}
Random Forest (RF)	6	No. of member classifiers No. of covariates selected for splitting	{25, 50, 100, 250, 500, 1000} 10
AdaBoost (AdaB)	3	No. of iterations	{10, 20, 30}

Table 1. Classification methods and meta-parameters of our model library.

To evaluate predictive accuracy, we consider two widely used performance measures (Sokolova et al., 2006), namely *classification accuracy* and *Area Under the ROC Curve* (*AUC*).

To generate ensembles, we implement a directed hill-climbing strategy (Caruana et al., 2004) and a *bagging* strategy (Breiman, 1996). In the following, these strategies are briefly explained.

- *Directed hill-climbing*: Starting with the *N* best performing candidate models, we add an additional candidate model and assess the resulting performance of the new ensemble. The candidate model remains in the ensemble if it improves the predictive accuracy. The incremental growing of the ensemble continues until adding additional candidate models stops to increase performance.
- *Bagging*: Directed hill-climbing is a greedy search strategy, which may be trapped in a local optima. It has been shown that applying hill-climbing to a randomly selected subset of candidate models and repeating this process multiple times (i.e., using bagging) can further increase performance (Caruana et al., 2006). The explanation is twofold: First, bagging generally decreases the variance of prediction models and, second, pooling multiple sub ensembles increases diversity in the final ensemble.

3 Data

To evaluate the application of ensemble selection, we conduct an empirical study related to the e-commerce apparel sector. For this purpose, we use the DMC 2014 competition dataset, which includes a training sample of 14 variables (see Table 2) and 481.092 observations from an online apparel retailer. These observations are related to sales of over 3.000 product items, which were either returned (*return delivery*) or kept by the customer. To the best of our knowledge, this is the only available public dataset related to returns management. We apply common modeling steps reported in Shmueli and Koppius (2011) to build empirical prediction models, including an extensive preprocessing of the data.

Order ID	Integer	Item price	Double
Order date	Date	Salutation	String
Delivery date	Date	Date of birth	Date
Item ID	Integer	State of the customer	String
Item size	String	Customer account creation date	Date
Item color	String	Return delivery (yes/no)	Binary
Manufacturer ID	Integer		

Table 2. Variables of the product returns data set.

4 Results

We examine the performance of our ES approach both in terms of prediction quality and computational time according to our research questions. The proposed framework has been implemented in MATLAB 2014b and executed on a server equipped with an Intel Xeon 3.3 Ghz and 128 GB of RAM.

Metric	Sample (%)	CART	SVM LibLin	LogR	MLP	RF	AdaB	ES	Best
AUC	100	.5935	.6030	.6769	.6826	.7613	.7311	**.7629**	ES
	80	.5810	.6371	.6773	.6974	.7597	.7316	**.7609**	ES
	60	.5748	.6457	.6777	.6943	.7552	.7350	**.7568**	ES
	40	.5735	.5072	.6765	.6835	.7471	.7296	**.7483**	ES
	20	.5919	.6600	.6735	.6753	.7387	.7259	**.7407**	ES
	10	.6101	.6556	.6777	.6944	.7386	.7293	**.7411**	ES
Accu-racy	100	.5819	.6075	.6286	.6373	.6879	.6672	**.6887**	ES
	80	.5303	.6065	.6275	.6407	.6863	.6657	**.6866**	ES
	60	.5166	.5708	.6290	.6319	**.6830**	.6676	.6828	RF
	40	.5615	.5626	.6264	.6307	.6757	.6640	**.6758**	ES
	20	.5127	.5712	.6253	.6305	.6684	.6618	**.6712**	ES
	10	.5540	.6157	.6297	.6398	.6745	.6648	**.6749**	ES

Table 3. Prediction performance evaluation for different sample sizes.

4.1 Performance Assessment

We assess the performance of an ensemble selection approach in terms of returns prediction quality (see Table 3). For this, we calculate the performance indicators for all single classifiers (CART, SVM LibLin, LogR, MLP), ensemble learners (AdaB, RF), and for the ES approach. For the comparison, we select the best performing model in each class of models. To assess the impact of the dataset size and to examine the trajectory of runtimes, we incrementally increase the sample size from 10% to 100% of the available data records (divided into 60% training data and 40% test data). The results show that ES outperforms all single classifiers and, except two cases, all ensemble learners. Moreover, the results demonstrate that more data leads to an improvement of prediction quality in terms of both accuracy and AUC. Given the full amount of sample data, we see that ES provides high quality prediction results.

As ES is an advanced modeling paradigm and can capitalize on a large library of candidate models when producing the ensemble, superior performance compared to single classifiers may seem trivial. Compared to standard ensemble methods, though, the results demonstrate that those advanced classifiers like RF can be improved by forming ensembles. Nonetheless, the results also confirm the competitiveness of off-the-shelf ensemble methods. In general, the empirical results provide strong evidence for the effectiveness of ES for predicting returns.

Next, we analyze the computational time of the two strongest classifiers, namely RF and ES. As depicted in Figure 1, we observe that the improvement of prediction quality leads to an exponential growth of computational time. Naturally, the computational time is also impacted by the size of the model library. While the computational time for the ES process increases linearly, the time for generating all models of the model library (*ES overall*) results in exceptionally long computational times up to 39 hours. From a practical point of view, such computational times cannot be accepted as the prediction model needs to be updated (i.e., rebuilt) frequently to learn from concept-drifting data streams (see, e.g., Tsoumakas et al., 2008). This is especially important in modern e-commerce platforms that take into account different sets of data to reflect customer- and product-specific characteristics (e.g., clickstreams, social media streams). Further, predictions need to be calculated in real-time allowing the e-retailer to react in a timely manner. This effect is intensified in a big data era; thus, it is essential to minimize the run-time overhead by establishing a scalable cloud architecture (Tsoumakas et al., 2008).

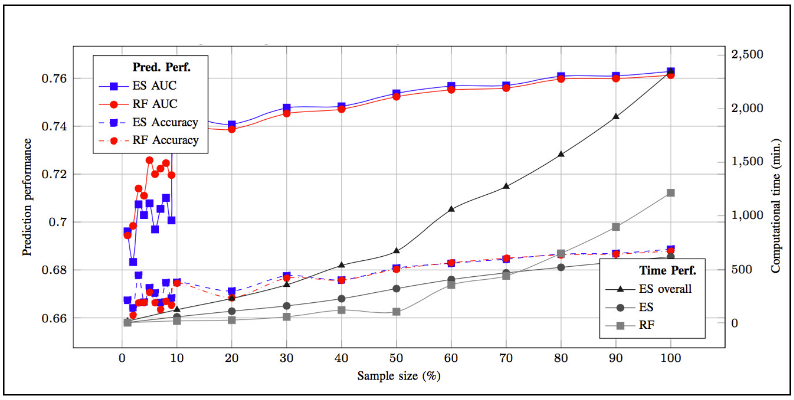

Figure 1. Performance and run-time correlation using a single server.

4.2 Cloud Framework for Ensemble Selection Processes

As computations are exceedingly resource-intensive and time-consuming, especially for large datasets, we present an approach to embed the envisioned FSS in a highly scalable cloud environment (for an overview on cloud computing the reader is referred to Heilig and Voß, 2014).

The generic cloud framework, depicted in Figure 2, allows to deploy an appropriate number of well-provisioned computing resources (1) allowing to parallelize the parallel generation of classifier-based candidate models (2). While the creation of decision tree models naturally implying huge memory requirements, other models (e.g., lazy learning methods) may have a considerable processing requirements (Tsoumakas et al., 2008). An approach to select the right number and provisioning of cloud computing resources that considers cost and computational performance requirements in real-time has been proposed in Heilig et al. (2016). The envisioned framework allows both a parameter-based (3) and model-based parallelization (4), if it is supported by the respective classifier. Consequently, several virtual machines (VMs) are horizontally scaled to generate a set of candidate models for each classifier in parallel. Vertical scaling mechanisms adjust computing resources corresponding to the classifiers' requirements. After all candidate models have been collected in a model library (5), the combination of models, using ensemble selection techniques, can be started (6). The resulting ensemble can be used to predict returns for customer- and product specific datasets (7) and provides the results for the envisioned FSS.

The framework shall further support learning from stream mining. That is, data streams characterizing the behavior of customers shall not only be used to predict product returns, but also to update data sets (8) as soon as the result and the reason of the customer's behavior becomes evident (e.g., after describing the reason of a product return in a dedicated return form). The centralized stream processing allows data gathering from multiple e-commerce platforms (in practice, large e-retailers often operate several e-commerce platforms). This requires a dedicated storage strategy of stream data as well as a dynamic ES approach to update the respective ensemble(s) over time. With the implementation of this framework, we aim to build a foundation for conducting related experiments, also indicating the assessment of available big data technologies (e.g., provided by the Apache Hadoop ecosystem), such as related to stream processing (see, e.g., Apache Spark, https://spark.apache.org). Respective design decisions for implementing the proposed framework will be explained in future research.

Another, not yet considered approach, is the splitting of the overall data into data chunks (i.e., subsamples) for which a respective model library is generated in parallel. As seen in Figure 1, the computational time for processing smaller data chunks is drastically lower. Based on these model libraries, an ensemble is created in the subsequent phase. This structure is similar to the MapReduce programming paradigm, known as a de-facto standard for addressing big data problems, where large data sets are split to data chunks and processed in parallel in a distributed computing cluster, before the results are aggregated.

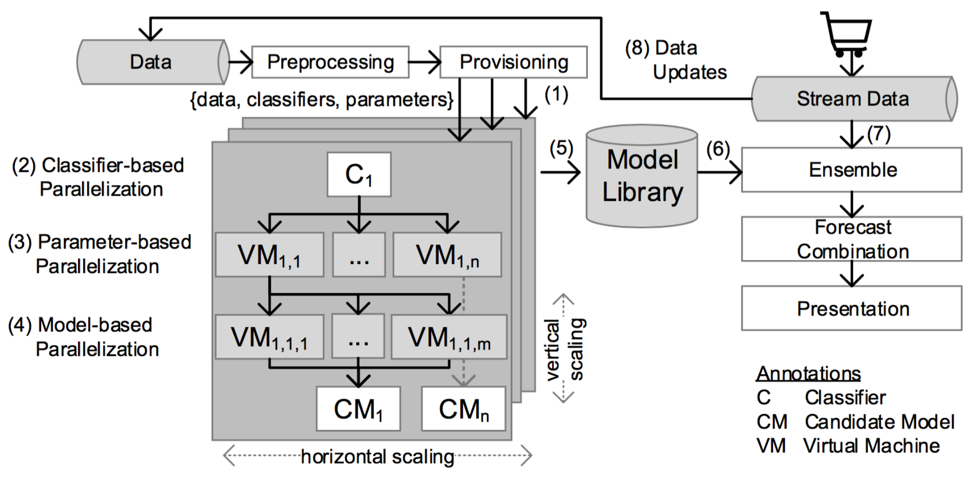

Figure 2. Generic cloud-based system framework approach.

5 Conclusion and Future Research

An accurate prediction of a customer's likelihood to return a product facilitates a number of preventive actions aiding e-retailers in reducing returns and increasing profit margins. Given the lack of data-driven approaches, we aim to develop a cloud-based FSS considering essential requirements in terms of prediction accuracy, scalability, and adaptability.

In this paper, we have presented two important contributions. First, our study lays the foundation of the envisioned FSS by examining the effectiveness of ES to predict returns in the apparel industry. The empirical results clearly demonstrate the dominance of our ES approach compared to advanced single classifiers and sophisticated ensemble learners in terms of widely applied prediction performance indicators and thus evidences its applicability as a suitable modeling methodology. In this context, we also observe a positive trend between dataset size and predictive accuracy, indicating that more data leads to a higher accuracy. Second, we systematically investigate the resource requirements of our ES approach. The results indicate an exponential growth of computational times when using more data for the prediction. To address the run-time overheads, we propose a generic cloud framework focusing on the parallelization of modeling activities on different abstraction levels. The generic framework builds a foundation for implementing a highly scalable cloud platform able to vertically and horizontally adjust its computing resources according to the user- and resource requirements. As such, it may build a foundation for exploring the design opportunities and performance impact of ES approaches for big data applications. Moreover, we present the first empirical data-driven study in the area of product returns management providing valid measures for tackling the problem of predicting product returns. These measures could be applied for improving product returns management and reverse logistics processes and may therefore stimulate research and development aimed at utilizing product returns predictions. In this regard, we outlined some ideas for preventing and responding to product returns. Moreover, the benchmark study builds the basis to further explore and assess the application of predictive analytics.

For extending this research, we aim to address the remaining points of our research agenda. To evaluate the contribution of a cloud-based FSS for high-speed model generation and real-time predictions, we aim to implement an ES-based prediction engine in a cloud environment using the proposed framework as a blueprint. Given the cloud-based platform, we aim to explore the impact of a MapReduce-like ES approach as well as the implementation of a dynamic ES approach, considering real-time data stream updates. Concerning adaptability, we aim to link returns forecasts to economic considerations by developing a cost function that builds upon quantifying the costs and economic value of false and correct predictions, respectively, to comply with the business context. The research activities will result in the envisioned FSS that need to be integrated and tested in a real environment for a large e-retailer handling several e-commerce shops.

References

Accenture (2012). *European cross-border e-commerce: The challenge of achieving profitable growth*. European Retail Round Table (ERRT). URL: http://www.ecommerce-europe.eu/stream/accenture-2012-european-cross-border-e-commerce.-the-challenge-of-achieving-profitable-growth.

Bhattacharyya, S., S. Jha, K. Tharakunnel, and J. C. Westland (2011). "Data mining for credit card fraud: A comparative study." *Decision Support Systems* 50 (3), 602–613.

Breiman, L. (1996). "Bagging predictors." *Machine Learning* 24, 123–140.

Baumann, A., Lessmann, S., Coussement, K., and De Bock, K. W. (2015). "Maximize what matters: Predicting customer churn with decision-centric ensemble selection." In: *Proceedings of the 23rd European Conference on Information Systems (ECIS)*. Münster, Germany: AIS, Paper 15.

Caruana, R., A. Niculescu-Mizil, G. Crew, and A. Ksikes (2004). "Ensemble selection from libraries of models." In: *Proceedings of the 21st International Conference on Machine Learning (ICML),* Ed. by R. Greiner and D. Schuurmans. Banff, Alberta, Canada, pp. 18–25.

Caruana, R., Munson, A., and Niculescu-Mizil, A. (2006) "Getting the most out of ensemble selection." In: *Proceedings of the 6th International Conference on Data Mining (ICDM)*. Hong Kong, China, pp. 828-833.

Gurnani, H., A. Sharma, and D. Grewal (2010). "Optimal returns policy under demand uncertainty." *Journal of Retailing* 86 (2), 137–147.

Heilig, L. and S. Voß (2014). "A scientometric analysis of cloud computing literature." *IEEE Transactions on Cloud Computing* 2 (3), 266–278.

Heilig, L., E. Lalla-Ruiz, and S. Voß (2016). "A cloud brokerage approach for solving the resource management problem in multi-cloud environments." *Computers and Industrial Engineering* 95, 16–26.

Hess, J. D. and G. E. Mayhew (1997). "Modeling merchandise returns in direct marketing." *Journal of Direct Marketing* 11 (2), 20–35.

Hevner, A. R., S. T. March, J. Park, and S. Ram (2004). "Design science in information systems research." *MIS Quarterly* 28, 75–105.

Ketzenberg, M. E. and R. A. Zuidwijk (2009). "Optimal pricing, ordering, and return policies for consumer goods." *Production and Operations Management* 18 (3), 344–360.

Kim, J. and S. Forsythe (2008). "Adoption of virtual try-on technology for online apparel shopping." *Journal of Interactive Marketing* 22 (2), 45–59.

Kohavi, R. (2001). "Mining e-commerce data: the good, the bad, and the ugly." In: *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA, USA: ACM, pp. 8–13.

Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley.

Lessmann, S. and S. Voß (2010). "Customer-centric decision support: A benchmarking study of novel versus established classification models." *Business & Information Systems Engineering* 2 (2), 79–93.

Liu, X. and K. K. Wie (2003). "An empirical study of product differences in consumers' e-commerce adoption behaviour." *Electronic Commerce Research and Applications* 2 (3), 229–239.

Mollenkopf, D. A., E. Rabinovich, T. M. Laseter, and K. K. Boyer (2007). "Managing internet product returns: a focus on effective service operations," *Decision Sciences* 38 (2), 215–250.

Mukhopadhyay, S. K. and R. Setaputra (2007). "A dynamic model for optimal design quality and return policies." *European Journal of Operational Research* 180 (3), 1144–1154.

Mukhopadhyay, S. K. and R. Setoputro (2004). "Reverse logistics in ebusiness: Optimal price and return policy." *International Journal of Physical Distribution & Logistics Management* 34 (1), 70–89.

Partalas, I., G. Tsoumakas, and I. Vlahavas (2010). "An ensemble uncertainty aware measure for directed hill climbing ensemble pruning." *Machine Learning* 81 (3), 257–282.

Reed, R. D. and R. J. Marks (1998). *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks.* Cambridge, MA, USA: MIT Press.

Shmueli, G. and O. R. Koppius (2011). "Predictive analytics in information systems research." *MIS Quarterly* 35, 553–572.

Sokolova, M., N. Japkowicz, and S. Szpakowicz (2006). "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation." In: *Advances in Artificial Intelligence (AI)*. Ed. by L. Lamontagne and M. Marchand. Quebec City, Canada: Springer, pp. 1015–1021.

Tsoumakas, G., I. Partalas, and I. Vlahavas (2008). "A taxonomy and short review of ensemble selection." In: *18th European Conference on Artificial Intelligence (ECAI), Workshop on Supervised and Unsupervised Ensemble Methods and their Applications. Patrac, Greece,* pp. 41–46.

Walsh, G., M. Möhring, C. Koot, and M. Schaarschmidt (2014). "Preventive product returns management systems – a review and model." In: *Proceedings of the 22nd European Conference on Information Systems (ECIS)*. Tel Aviv, Israel: AIS, pp. 1–12.

Yalabik, B., N. C. Petruzzi, and D. Chhajed (2005). "An integrated product returns model with logistics and marketing coordination." *European Journal of Operational Research* 161 (1), 162–182.

Yao, Z., S. C. H. Leung, and K. K. Lai (2008). "Analysis of the impact of price-sensitivity factors on the returns policy in coordinating supply chain." *European Journal of Operational Research* 187 (1), 275–282.