Predicting Customer Demand for Remanufactured Products: A Data-Mining Approach

Truong Van Nguyen[1], Li Zhou[2*], Alain Yee Loong Chong[3], Boying Li[3], Xiaodie Pu[3]

1: Brunel Business School, Brunel University London, Kingston Lane, Uxbridge, UB8 3PH, UK

2*: Faculty of Business, University of Greenwich, London, SE10 9LS, UK (Corresponding author: Email*: ZL14@gre.ac.uk)

3: Nottingham University Business School, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo 315100, China

Predicting Customer Demand for Remanufactured Products: A Data-Mining Approach

Abstract: Remanufacturing has received increasing attention from researchers over the last decade. While many associated operational issues have been extensively studied, research into the prediction customer demand for, and the market development of, remanufactured products is still in its infancy. The majority of the existing research into remanufactured product demand is largely based on conventional statistical models that fail to capture the non-linear behaviour of customer demand and market factors in real-world business environments, in particular e-marketplaces. Therefore, this paper aims to develop a comprehensible data-mining prediction approach, in order to achieve two objectives: (1) to provide a highly accurate and robust demand prediction model of remanufactured products; and (2) to shed light on the non-linear effect of online market factors as predictors of customer demand. Based on the real-world Amazon dataset, the results suggest that predicting remanufactured product demand is a complex, non-linear problem, and that, by using advanced machine-learning techniques, our proposed approach can predict the product demand with high accuracy. In terms of practical implications, the importance of market factors is ranked according to their predictive powers of demand, while their effects on demand are analysed through their partial dependence plots. Several insights for management are revealed by a thorough comparison of the sales impact of these market factors on remanufactured and new products.

Keywords: Data mining; Remanufactured products; Machine learning; Regression trees.

1. Introduction

1.1. Background and motivation

Remanufacturing is now a multi-billion dollar industry, with product sales increasing by 15% (approximately US$43 billion) per year. The driving forces of this growth are the enormous economic and environmental benefits gained from remanufacturing used products rather than producing new ones. These benefits include: reducing production costs by 50%; using 70% less raw materials, cutting manufacturing emissions by 80%; reducing energy consumption by a maximum of 60%; and offering lower prices to customers (Wang & Hazen, 2016). However, remanufacturing is not a panacea for the achievement of a sustainable thriving business, as its viability hinges on a high degree of uncertainty, in terms of both return and demand. The return uncertainty derives from a lack of information about the timing, quantity and quality of returned products (Zhou et al., 2016). The demand uncertainty of remanufactured products stems from unobservable remanufacturing processes, such as cleaning, disassembling, inspecting and testing, which make it difficult for customers to evaluate quality (Tereyağoğlu, 2016). Accurate predictions of demand for remanufactured products are therefore required for effective remanufacturing/closed-loop supply chain (CLSC) operations.

Compared to studies in the CLSC subject related to returned product acquisition management and the operational issues of remanufacturing, less progress is being made in demand prediction and marketing strategy for remanufactured products (Atasu, Guide, & Van Wassenhove, 2008). Atasu et al. (2008) argue that there is an urgent need to apply advanced forecasting techniques to accurately quantify the uncertain parameters in the CLSC, such as demand, price and return. They point out that the more understanding of market acceptance of remanufactured products would facilitate the development of a more sophisticated analytical model of remanufacturing/CLSC operations with high industrial relevance.

This research is motivated by these two research streams – *remanufactured product demand forecasting* and *marketing strategy* – with the aim of contributing to CLSC research from both operational and marketing perspectives. To achieve this, the research utilises a data-mining and machine-learning (ML) approach, which predicts demand for remanufactured products with high accuracy, resulting in a practicable marketing strategy for management. Related studies are reviewed below.

1.2. Market development of remanufactured products

Research into the market development of remanufactured products has attracted increasing interest over the last decade. The topic is broad and includes areas related to sales channels (Yan, Xiong, Xiong, & Guo, 2015); warranty strategies (Alqahtani & Gupta, 2017); pricing analysis (Abbey, Blackburn, & Guide, 2015); and revenue management (Ovchinnikov, 2011). Of the various sub-topics, the differences in consumer behaviour involving remanufactured products and new products is one of the most emergent subjects. A literature review on the study of consumer behaviour in relation to remanufactured products is presented below and is summarised in Table S1 in Supplementary Materials_A.

As Table S1 shows, most of the insights into consumer behaviour towards remanufactured products have been gained through the collection of primary data, in order to examine the key determinants that underpin customers' purchase intention and estimated willingness to pay (WTP) (Abbey, Meloy, Guide, & Atalay, 2015; Hamzaoui-Essoussi & Linton, 2014; Hamzaoui Essoussi & Linton, 2010; Jiménez-Parra, Rubio, & Vicente-Molina, 2014; Khor & Hazen, 2017; Wang & Hazen, 2016). The authors of these studies believe that a consumer will follow his/her purchase intention when making an actual purchase decision. However, in practice, a consumer's purchase behaviour does not always follow his/her purchase intention. For example, green awareness may be a driving force behind a consumer's intention to purchase a remanufactured product, but the actual purchase decision is based on the perceived quality and risk associated with the product, rather than on their initial intention (Khor & Hazen, 2017). To overcome this issue, another new research stream is initiated by using secondary data of real transactions to reflect real WTP, thereby exploring actual consumer purchase behaviour (Frota Neto, Bloemhof, & Corbett, 2016; Subramanian & Subramanyam, 2012; Xu, Zeng, & He, 2017).

However, a major drawback of the previous studies is that they study consumer behaviour using a linear regression (LR) model which fails to capture the non-linear relationships between various predictors. As a result, its prediction accuracy deteriorates substantially when dealing with real-world datasets characterised by highly complex and non-linear relationships between multiple market factors.

1.3. Machine-learning and regression-tree approaches

In the big-data era, ML approaches are increasingly used to overcome the drawbacks of LR and to generate data-driven decision-making. Empirical evidence suggests that companies that use data-driven decision-making can see significant improvements in both productivity and profitability (Bohanec, Kljajić Borštnar, & Robnik-Šikonja, 2017). The key difference between the statistics-based LR and ML models is that the former uses a parametric approach in which the model structure is predetermined and the input–output relationship is forced to fit certain simplified assumptions. In contrast, the latter does not start from the model structure, but uses algorithms to determine input–output relationships from the dataset. Therefore, ML can approximate the complex and unknown non-linearity of noisy, high-dimensional datasets better than statistical methods.

Demand forecasting is naturally considered to be a regression problem in ML, and aims to accurately estimate the demand level of a product based on its relationships with a given set of independent input variables (i.e. predictors). Some examples of ML algorithms for regression problems include support vector machine (SVM), classification and regression tree (CART), artificial neural network (ANN), and K-nearest neighbour (KNN) algorithms. A recent review of big-data applications in supply-chain management (SCM) by Nguyen, Zhou, Spiegler, Ieromonachou, & Lin (2018) shows that ML has been widely adopted in traditional SCM to improve demand prediction. However, in CLSCs, the application of ML is largely underexploited and merely used for product return forecasting (see, e.g. Mazhar, Kara, & Kaebernick, 2007). The use of ML for remanufactured product demand forecasting has hardly been studied.

Despite the high level of predictive performance, ML models face resistance from users, as there is a perceived lack of comprehensibility (also known as interpretability) which requires practitioners to understand the insights behind the prediction of the model. In some areas, such as credit risk analysis and medical diagnosis, comprehensibility is even more vital than prediction accuracy (Martens, Baesens, Van Gestel, & Vanthienen, 2007).

While many ML models, such as the SVM, ANN, KNN and ensemble models, are widely criticised as being ‘black box’, the regression-tree approach is a type of ML that has gained popularity for creating a good balance between predictive performance and interpretation (De Caigny, Coussement, & De Bock, 2018; Masci, Johnes, & Agasisti, 2018; Yang, Liu, Tsoka, & Papageorgiou, 2017). In single regression-tree models, such as CART and the M5 model tree, high interpretability derives from the tree-based graph in which one can quickly detect the most important variables used for node splitting and their variable interactions.

However, in more advanced regression-tree models, such as the ensemble-based random forest (RF) model, the tree visualisation is no longer possible. Alternative explanatory methods are needed to reveal the insights of such black box models. Two techniques which tackle this task effectively are variable importance ranking (VIR) and partial dependence plot (PDP). These methods are commonly used in many domains, such as education (Masci et al., 2018), ecology (Cutler et al., 2007), business risk management (De Bock, 2017), and supply-chain finance risk prediction (Zhu, Zhou, Xie, Wang, & Nguyen, 2019). To the best of our knowledge, however, they have not yet been applied in remanufacturing and CLSC management.

1.4. Research objective and contribution

The literature review makes clear that it is necessary to use more sophisticated ML techniques to improve the accuracy of remanufactured product demand prediction (RPDP), which is essential for developing a marketing strategy for remanufactured products. This paper therefore develops a data-mining approach that pursues two main objectives: (1) to obtain an accurate and robust RPDP by using the machine-learning approach; and (2) to analyse the RPDP by using VIR and PDP in order to gain an in-depth understanding of the online purchasing behaviours of consumers of remanufactured products, leading to the development of a practicable marketing strategy. A real-world database consisting of 5,693 remanufactured product listings on www.amazon.com is used to pursue the above objectives.

This research is one of the pioneering papers which apply an ML approach to the demand prediction of remanufactured products. The paper's contribution is twofold: theoretical and practical. In theoretical terms, the research (1) develops a structured business analytics approach that can balance the trade-off between prediction accuracy and comprehensibility, thereby stimulating the use of black-box ML models; (2) provides a highly accurate and robust demand prediction for remanufactured products (a research area which remains largely understudied); and (3) sheds light on the non-linear behaviours of online market factors on remanufactured product demand. In practical terms, the research provides guidelines with which managers can develop effective online marketing and selling strategies for remanufactured products, thereby increasing the viability and profitability of remanufacturing/CLSCs.

The remainder of the paper is structured as follows. Section 2 describes the methodology framework, detailing the rigorous research process adopted for this paper. Section 3 describes the data collection and the variables related to the RPDP. Section 4 details the data preparation for the prediction model. Section 5 explains how the predictive models are developed and validated. Section 6 covers the evaluation and deployment of the model, while Section 7 discusses the results and the potential insights for management. Section 8 deals with the robustness checking of the model. Finally, section 9 is the conclusion.

2. Methodology framework

This paper follows the Cross Industry Standard Process and Data Mining (CRISP-DM) framework, one of the most popular methodologies for data analytics (Oztekin, Kizilaslan, Freund, & Iseri, 2016). In line with the CRISP-DM, the data-mining approach in this paper consists of six fundamental steps as shown in Figure 1: (1) *Business understanding* refers to the conversion of the business objective, RPDP in this paper, into a data mining problem; (2) *Data understanding* identifies the data source and obtains the variables related to the problem; (3) *Data preparation* uses several data cleaning and transforming techniques to produce a well-structured dataset prior to analysis; (4) *Predictive modelling* includes variable selection, model development, hyperparameter tuning and validation; (5) *Model evaluation* measures and compares the predictive performance of the models based on different predefined error measurements; (6) *Model deployment* generates insights to assist managerial decision-making.

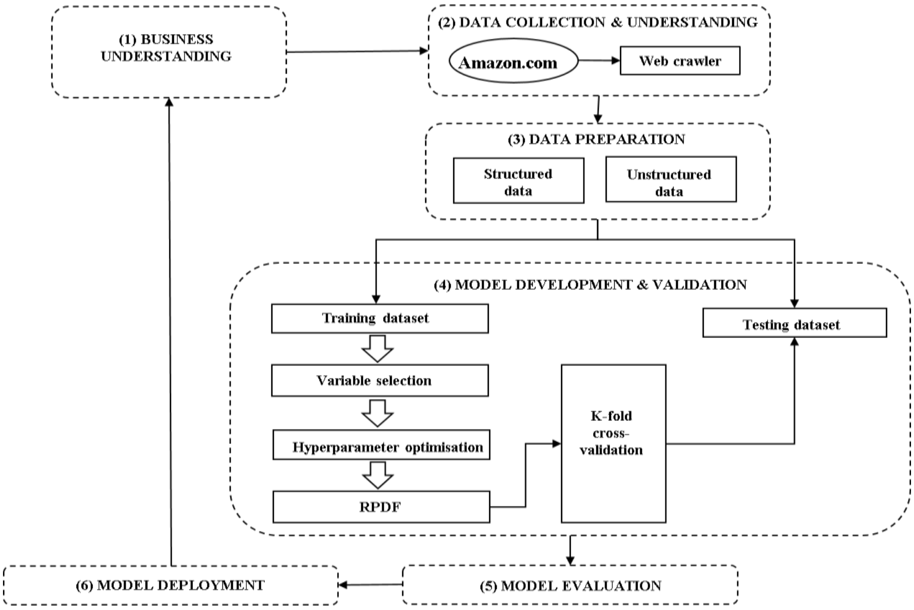

Figure 1: Methodology framework

3. Data collection and use of variables

3.1. Research context and data collection

Consistent with previous studies of the online marketplace, www.amazon.com has been chosen as the data source for this study. Amazon classifies product conditions into three major categories — 'new', 'certified refurbished' (another term for remanufactured products), and 'used'. To distinguish between remanufactured products and used products, Amazon defines remanufactured products as products that have been tested and certified by original equipment manufacturers (OEMs) or by qualified/specialised third-party remanufacturers to ensure their 'like-new' working condition. In this paper, we focus on remanufactured products, as they are among the core products in CLSCs and have a direct impact on the manufacturing processes of OEMs (Zhou & Disney, 2006).

Amazon does not publicly reveal actual sales transactions. Instead, it uses the sales ranks to indicate the sales performance, which is recognised as a well-established proxy of customer demand (Archak, Ghose, & Ipeirotis, 2011; Dekkers, 2011; Hu, Koh, & Reddy, 2014; B. Li, Ch'ng, Chong, & Bao, 2016). In general, the smaller the value of a product's sales rank, the higher its customer demand. As there are scale effects in our data, the sales rank is estimated by its natural logarithm transformation (ln) rather than its level, as suggested by Chevalier & Mayzlin (2006). Hence, the dependent variable used in this study is expressed as follows:

$$customer.demand = \frac{1}{ln\ (Salesrank\ )} \tag{1}$$

where $Salesrank$ is Amazon's sales rank of the remanufactured product listing; and $customer.demand$ is the demand level of the product and also the dependent variable in our model.

We develop a python programme to crawl the daily historical data of remanufactured products listed on www.amazon.com. For each product listing, the crawler was coded to capture a complete set of the data publicly available on the product page. It is noteworthy that Amazon sales ranks are frequently updated so that they reflect recent sales (with a maximum period of one months) (Chevalier & Mayzlin, 2006). Hence, in order to eliminate the potential simultaneity issue between independent and dependent variables, we follow the approach of Chevalier & Mayzlin (2006), which predicts the sales rank at time *t*, based on lagged explanatory variables up to one month before time *t*. In particular, sales ranks were recorded on 30 May 2018, while all the predictors were recorded during the period up to 30 April 2018 (i.e. one month before the 30 May sales rank).

Our dataset includes remanufactured products that have comparable sales ranks in the Amazon Electronics category. where remanufacturing activities are particularly important, because of the end-of-life environment effect of products (Subramanian & Subramanyam, 2012). This category includes technological products such as cell phones, computers, cameras and GPS navigation.

3.2. Theoretical background and variable description

3.2.1. Theoretical background

In the contract and economic theory, the term 'information asymmetry' refers to market interaction under conditions in which sellers have more or better information about the quality of the product and service than buyers (Boulding & Kirmani, 1993). Such an information gap would increase the buyer's quality uncertainty and the perceived risks of buying a low-quality product, making buyers willing to pay no more than the average market price. Consequently, sellers have no incentive to sell high quality products, and are likely instead to keep low quality products on hand (Frota Neto et al., 2016). This phenomenon was first described by Akerlof (1970) as the 'market of lemons'. To prevent the emergence of lemon markets, sellers often apply market signal theory, in order to address the information asymmetries of customers (Boulding & Kirmani, 1993; Wells, Valacich, & Hess, 2011). The theory provides a framework with which to understand how sellers can use extrinsic cues (i.e. signals) to convey information about their product and service quality to buyers, in order to reduce their perceived uncertainty and to facilitate trade in an environment featuring high information asymmetries (H. Li, Fang, Wang, Lim, & Liang, 2015).

As it focuses on the online marketplace of remanufactured products, the theoretical foundation of this study is primarily based on market signal theory. This is because: (1) empirical research has suggested that the burden of information asymmetry between sellers and buyers is exaggerated in the e-marketplace, as buyers have no physical interaction with the products prior to their purchase (H. Li et al., 2015); and (2) the information asymmetry of product quality seems to be more severe for remanufactured products than new products, due to the higher customer uncertainty and perceived risks associated with unobservable remanufacturing processes (Tereyağoğlu, 2016).These factors increase the need of customers for additional information about the product, which ultimately means that quality cues have a stronger influence on customers' purchasing decision-making (Frota Neto et al., 2016).

Seller-generated content (SGC) and user-generated content (UGC) can both be signalling factors. Following previous research (e.g., Frota Neto et al., 2016; Subramanian & Subramanyam, 2012; Xu et al., 2017), this study uses a number of the SGC and UGC variables which are available on Amazon to predict demand for the remanufactured product. The description and measurement of each variable are presented as follows.

3.2.2. Description and measurement of variables

- Positivity of product description

A common way for sellers to ease the problem of information asymmetry with customers is to use textual descriptions of products. The seller-provided product description is typically written in natural language with structured formats. It serves as a snapshot which summarises the key selling points that favourably differentiate their product from others. Sellers of remanufactured products often use keywords that positively describe the condition and quality of these products, such as 'like new condition', 'certified refurbishment', 'testified', 'mint', 'excellent condition' and 'no scratches'. The effect of such positive keywords on customer WTP has been previously studied, with mixed findings. For example, van Heijst, Potharst, & van Wezel (2008) find that positivity of product description is the most influential predictor of customer WTP to remanufactured products, compared to the number of product pictures and seller feedback ratings. In contrast, Frota Neto et al. (2016) find no significant statistical evidence for such a relationship. Despite the mixed result, it is therefore still very likely that customers would use such condition-related keywords as a cue about remanufactured product quality. As such, based on market signalling theory, we expect that the positivity of product description will have a significant, positive effect on customer demand for remanufactured products. The technical detail of how this variable is constructed is described in Supplementary Materials_B.1.

- Number of product pictures

In addition to textual quality cues, sellers also use product pictures as visual quality cues because some product attributes are difficult to describe in words, such as signs of wear on shoes (Frota Neto et al., 2016; van Heijst et al., 2008). Therefore, it is expected that the number of product pictures provided by the seller will have a positive effect on the remanufactured product demand.

- Warranty information disclosure

Due to the monetary and legal risks involved, the signalling role of warranties as a quality cue has gained much more attention than textual and visual cues for both new products (H. Li et al., 2015) and remanufactured products (Alqahtani & Gupta, 2017; Subramanian & Subramanyam, 2012). Although warranties may subsequently incur extra costs for sellers/manufacturers, they can bring additional profits at the point of sale by reducing customers' concerns about product reliability during their purchase decision-making. This positive effect of warranties on online sales is found to be significant for new products (H. Li et al., 2015), but has not yet been confirmed for remanufactured products (Abbey, Meloy, et al., 2015; Atasu, Guide, & Van Wassenhove, 2010; Jiménez-Parra et al., 2014; Subramanian & Subramanyam, 2012).

Instead of focusing on warranty strength (i.e. the duration of warranty), our study examines a new angle that has not been yet gained much research attention — the ways in which warranty information disclosure affects customer purchasing behaviour towards remanufactured products. According to Tereyağoğlu (2016), sellers often see the disclosure of full information about the warranty and entire remanufacturing process as one of the strategies which can reduce customers' uncertainty about the remanufactured product quality, thereby further increasing sales. This strategy is in line with transactional cost theory and the linkage principle of Milgrom & Weber (1982), suggesting that voluntary information disclosure by sellers can help reduce information asymmetries and save customers the cost of acquiring information, thereby increasing their WTP.

Some Amazon sellers of remanufactured products provide technical documents which typically specify the details of the remanufacturing process, product condition and the full terms and conditions of the product warranty. Other sellers require customers to take extra time and effort, such as by contacting the seller by email or being directed to other websites for further warranty details. To reflect the impact of warranty information disclosure, we create a binary variable, which is equal to '1' if the Amazon seller provides a technical document specifying the full warranty terms and conditions in detail, but equal to '0' if no such document is given. Following the above discussion, we would expect that remanufactured products with full warranty information disclosure are associated with high customer demand, whereas those without such information are associated with lower demand.

- Stock information

Stock information is another binary variable which is considered to be a potential predictor of remanufactured demand. This variable is equal to '1' if the seller specifies a limited quantity of stock left; otherwise, it is equal to '0'. We adopt this variable from the marketing perspective. By leveraging the psychological effect of the scarcity principle, sellers use the limited availability of a product to increase its perceived quality and boost sales (Cialdini, 2009). The existing literature contains abundant empirical evidence on the importance of the scarcity effect for purchasing decisions (Swami & Khairnar, 2006). However, it is not known whether this effect also exists for remanufactured product. In this study, we expect that the scarcity principle still applies. As such, remanufactured products with limited availability will be associated with a higher demand level than those with abundant stocks.

● Price difference between remanufactured and corresponding new products

As Abbey, Blackburn, et al. (2015) state, managers often experience fear and uncertainty when making pricing decisions about remanufactured products — fear of the risks of cannibalisation by new products and uncertainty about customers' WTP for remanufactured products. Following the law of demand in microeconomic theory, which holds that a lower price will lead to a higher demand, many firms have been setting the prices of their remanufactured products 10% to 80% lower than those of the corresponding new products (Ovchinnikov, 2011). In this way, the seller uses price as a marketing tool with a discounting effect. This practice is supported by a small number of studies which found that price discounts should have a positive, linear effect on the perceived attractiveness and sales of the remanufactured product (Abbey, Blackburn, et al., 2015). However, the existing literature has also found some evidence for the non-linearity or even negative effects of price discounts on the attractiveness of, and demand for, remanufactured products (Ovchinnikov, 2011). This study therefore aims to determine whether a price differential between a remanufactured product and the equivalent new product will increase demand for the remanufactured one.

The variable representing the price difference is expressed in Eq (2):

$$Price\_difference = \frac{Average\ selling\ price\ of\ new - selling\ price\ of\ remanufactured}{Average\ selling\ price\ of\ new} \times 100\% \qquad (2)$$

The method used to obtain the price of corresponding new products from Amazon is described in Supplementary Materials_B.2.

● Product promotion rate

As well as selling remanufactured products at a lower price than the equivalent new products, sellers also offer some promotional discounts, aiming for an immediate short-term increase in sales. According to transactional utility theory, higher discount rates increase customers' utility, and therefore lead to higher sales (Dekkers, 2011). In this paper, we expect there to be a positive link between a promotional discount rate and remanufactured product demand. The discount rate is collected directly from each Amazon product page.

● Overall product rating

Like most e-commerce sites, Amazon provides an overall product rating which is the average of the ratings from previous customers. This numerical index ranges from the lowest rating of 1 to the highest of 5. It indicates the valence dimension of the seller's reputation and the overall attractiveness to consumers based on their past sales performances and future prospects, which provides signals that reduce the burden of information asymmetry and build customer trust in the sellers (H. Li et al., 2015). Previous studies have repeatedly found that the overall product rating has a positive and linear effect on sales of new products (Archak et al., 2011; Chevalier & Mayzlin, 2006; B. Li et al., 2016; H.

Li et al., 2015). However, the effect of the product rating for remanufactured products has not been sufficiently studied, and this study hopes to address that gap. We expect the overall product rating to have a positive effect on remanufactured product sales.

- Number of service failures and number of service successes

Previous studies have found that the effect of negative customer feedback on sales of new products can be more pronounced than that of positive feedback, as the damage to the seller's reputation significantly increases the risk perceived by prospective customers (Chevalier & Mayzlin, 2006; Cui, Lui, & Guo, 2012; De Maeyer, 2012). Subramanian & Subramanyam (2012) confirm the finding for remanufactured products. However, their finding is based on the linear model, which motivates us to re-examine whether such negative bias remains significant in the non-linear model. As a result, this study aims to determine whether the number of service failures and the number of service successes are important predictors of remanufactured product demand, and whether the effect of service failures on sales is more significant than that of service successes. Our method of obtaining these two variables from Amazon is explained in Supplementary Materials_B.3.

- Number of helpfulness votes

For each product, we take the sum of helpful votes received from all customer reviews. This metric can serve as a quality indicator of the reviews, according to other consumers. A large number of helpful votes indicates that the product reviews are of high quality, meaning that the reviews contain a large amount of helpful information and can influence the purchasing decisions of other buyers. The positive effect linking the helpfulness of reviews and the online sales of new products have been found in previous literature (e.g. B. Li et al., 2016). In particular, this effect can become more powerful for less popular products (De Maeyer, 2012). This study aims to test the validity of these findings in the case of remanufactured products.

- Number of answered questions

As well as indirect interaction between users through helpfulness votes, Amazon has initiated 'Customer questions and answers', which allows customers to directly ask and respond to each other's questions. According to trust transference theory, a large number of answered questions can stimulate social interactions which further boost online trust for higher purchasing intention (Ng, 2013). Previous studies have found a significant and positive effect of this variable on the sales predictions for new products (Dekkers, 2011; B. Li et al., 2016). Hence, it is expected that the variable will have a strong predictive power on demand for remanufactured products.

- Average sentiment of customer reviews

In addition to the numerical ratings of the product, Amazon also allows a customer to post a textual review that provides a context-specific explanation of his/her shopping experience with varying degrees of polarity sentiment (e.g. strongly or moderately negative, positive or neutral). Such polarity sentiments provide rich information to the reader who goes beyond numerical ratings (Hu et al., 2014). Although people often consider the sentiment of a textual review to be consistent with the numeric product rating, previous research indicates that these two are not always aligned (Hu et al., 2014). Archak et al. (2011) posit that only using numerical ratings may not fully capture the effect of product reviews on customer purchasing behaviour, unless the reader of a review has exactly the same preferences as its writer. According to trust transference theory, when potential customers read online reviews, their emotions may be affected by the sentiments expressed in the product review, and their emotional status may influence their evaluation of the product and direct their purchase decisions. Previous studies have found that review sentiments play a significant role in the prediction of new product demand (Archak et al., 2011; Dekkers, 2011; Hu et al., 2014). In this study, we examine whether such strong predictive power is also true of remanufactured products. We expect that remanufactured products indicating positive review sentiment would be associated with high customer demand. The technical detail of our construction of the variable representing the review sentiment of each remanufactured product is explained in Supplementary Materials_B.4.

- Brand equity

Brand equity can be used to reduce customers' perceived risks and uncertainties about product quality and increase their perceived level of trust. When a product has high brand equity, the brand is considered to have high value, and the product is associated with high levels of quality, reliability and awareness. Hence, there is likely to be high demand for it. Previous research has identified brand equity to be important for remanufactured products (Abbey, Meloy, et al., 2015). Therefore, in this study, brand equity is included in the model as a control variable. More specifically, we classify brand equity as high or low, based on the customer voting on www.ranker.com, where top brands are listed. This website provides the most comprehensive brand rank lists for a wide range of niche markets and product types. These rank lists are voted on by millions of visitors every month, and therefore accurately reveal the brand conception of the crowd. If a product brand is listed on www.ranker.com, it is marked as being of high brand equity; if not, it is marked as being of low brand equity.

4. Data preparation

There are both structured (e.g. numerical data) and unstructured (e.g. textual data) types of variables in the studied Amazon dataset that require different pre-processing techniques before they can be analysed.

4.1. Structured data preparation

The main task here is handling the missing data in the dataset. Previous studies (e.g. De Caigny et al., 2018; Verbeke, Dejaeger, Martens, Hur, & Baesens, 2012) have suggested that for continuous variables with more than 5% of missing data, the missing values should be treated using appropriate imputation procedures. However, for those with less than 5% of missing values, the missing values should be removed to limit the effect of imputations (Little, 1988). In addition, all the categorical variables should be transformed into dummy variables (1 or 0), which are then treated as numerical variables.

Outlier treatment is also carried out, to ensure data consistency. Outliers are out-of-ordinary values that are typically defined as being more than three standard deviations (SD) away from the mean. Univariate outliers among continuous variables are visually detected using a box-and-whisker plot, and winsorised into acceptable values that are within 1.5*IQR (Interquartile ranges), using R programming.

To address the scale effect between continuous independent variables, their values are normalised into the range between 0 and 1.

4.2. Unstructured textual data preparation

Customers' textual reviews and sellers' textual product descriptions can be highly unstructured and extremely noisy. Specific text pre-processing techniques are therefore required to transform such data into a usable, structured format. The techniques used in this paper include: (1) tokenisation: breaking the text sentences into word vectors; (2) stopwords filtering: removing unnecessary stopwords for grammar rules (e.g. a, an, the); and (3) part-of-speech (POS) tagging: referring to a category of words having similar grammatical properties such as noun, verb, adjective, etc. The POS tagging is very helpful as it filters out all the words that do not convey sentiments (i.e. relevant or meaningful information). More details of these pre-processing steps can be found in Alaei, Becken, & Stantic (2017). As a result, each of the customer reviews and the seller's product descriptions are transformed into vector space models, in which each vector dimension corresponds to a separate term (or keyword) that contains sentiments. The term vectors were used to construct two of the predictors used in our proposed remanufactured product demand model, namely the positivity of product descriptions and average review sentiments, as described in Section 3.2.2.

5. Model development and validation

As Figure 1 shows, we employ three specific procedures that are important for a fair and valid comparison of the performance between different ML models. These are variable selection, hyperparameter optimisation and k-fold cross-validation (CV).

5.1 Variable selection

After the well-structured dataset is prepared and aggregated, the variables are selected as predictors in the RPDP model. Variable selection is a critical step in ML applications, as using excessively large datasets with many irrelevant/noisy variables often slows the algorithm, consumes more resources and can even damage predictive performance. Based on the concept of variable relevance, Nilsson, Peña, Björkegren, & Tegnér (2007) classified variable selection into two categories of problem: (1) *all-relevant problems* — finding all strongly and weakly relevant variables; and (2) *minimal–optimal problems* — finding the subset of strongly relevant variables and removing the subset of weakly relevant variables that contains only redundant information. Variable relevance here is a qualitative measure and is independent of classifier types, which is distinct from the variable importance in the VIR analysis, as discussed later (Rudnicki, Wrzesień, & Paja, 2015).

To increase the applicability of the research to future work, we tackle both of the feature selection problems. The Boruta algorithm is adopted for the all-relevant set, while the recursive feature elimination (RFE) algorithm is adopted to select the minimal–optimal set of predictors. All the predictive algorithms for RPDP discussed in Section 5.2 are run with these two input sets. The procedures of both algorithms are detailed in Supplementary Materials_C.1 and C.2.

5.2 Regression tree model descriptions

Regarding the predictive algorithms for RPDP, three regression tree models (CART, M5 and RF) are employed.

- **CART model**

Like most decision tree algorithms, CART adopts a 'divide and conquer' strategy in order to construct the tree-based model. It aims to identify a predictor and its breaking-point value as the tree node for the binary splitting of the training space into the most homogeneous (or purest) subsets. The formal description of the CART model is presented in tSupplementary Materials_C.3.

Despite its high interpretability and efficiency, there are some practical issues affecting both 'divide and conquer' steps in the CART algorithm: (1) using the greedy search to grow the tree is an efficient approach, but it only returns local optimality, which makes the tree very sensitive to even a small change in a dataset, and bias can occur if the dataset is imbalanced; (2) the high variance of the tree leads to the instability of predicted outcomes, because the prediction rule based on the mean value is too simple (Witten, Frank, & Hall, 2011). As a result, there have been a few attempts to improve these shortcomings of CART, such as the M5 model tree and RF.

- **M5 model tree**

Introduced by Quinlan (1992), the M5 model tree attempts to improve the CART prediction rule by applying an LR model to predict the value in each leaf node, rather than taking the mean value. The construction of the model is also based on the 'divide and conquer' approach. At the first stage, the tree is grown with the same principle as CART. The only difference is that CART uses standard deviation of the output as the loss function, which therefore maximises the expected standard deviation reduction (SDR). The details of the M5 model are presented in Supplementary Materials_C.4**.**

- **RF model**

Random forest (RF) adopts the bagging ensemble method which combines multiple base classifiers with equally distributed weights, in order to increase the predictive performance of a single decision tree (Witten et al., 2011). The detailed process of bagging in RF is provided in Supplementary Materials_C.5.

By injecting randomness into both subsampling and input selection, RF has several distinct advantages, including (1) high prediction accuracy; (2) relative robustness to data outliers and noise; (3) higher speed, compared to other bagging and boosting models; (4) unbiased (low variance) internal error estimations, predictive strength, correlation and variable importance; and (5) easy hyperparameter tuning (Breiman, 2001). Therefore, RF has been commonly used for complex forecasting problems in diverse areas, such as finance (Krauss, Do, & Huck, 2017), ecology (Cutler et al., 2007) and many others. Using the CART tree as the base classifier, it is, however, difficult to interpret the RF results when numerous random single trees are added to the forest model. Therefore, like most other powerful ML prediction models (e.g. ANN, SVN), RF shares the main disadvantage of being a 'black box' (Witten et al., 2011).

5.3 Hyperparameter tuning

It is essential to make a comparison between different ML models based on the best performance of each one, obtained by running with their optimal parameter settings (De Caigny et al., 2018). Therefore, we perform the sensitive analysis of each model with regard to its own tuning parameters and using both the Boruta and RFE sets of inputs.

5.4 K-fold cross validation (CV)

When comparing the prediction accuracy of multiple models, it is common to use k-fold CV to avoid the bias issue relating to data sampling (Oztekin et al., 2016). Given the sample size used in this paper, we run all models with ten-fold CV, because the empirical research shows that k=10 is the optimal number of folds that optimises the computational time while also minimising bias and variance issues during the validation process (Kohavi, 1995). To totally remove the resampling bias when dividing *k* fold, we further repeat ten-fold CV three times.

6. Model evaluation and deployment

6.1 Model evaluation

Three commonly used statistical measurements — mean absolute error (MAE), root mean square error (RMSE) and coefficients of determination (R^2) — are calculated in order to evaluate the predictive performance of different ML models. The higher the value of R^2, the better the predictive performance achieved by the model. In contrast, the lower the value of MAE and RMSE, the better the predictive performance of the model.

6.2 Model deployment

We perform two analyses, VIR and PDP, in order to interpret the predictive power and marginal effect of the predictors of RPDP.

6.2.1 Variable importance ranking (VIR)

A simple way to extract insights from ML-based prediction is to rank the importance of the explanatory variable, based on its predictive strength to the response variable. The method of measuring variable importance will vary, depending on the nature of the problem (classification or regression) and the learning algorithms used, as described by Grömping (2015).

For all three regression tree models in this study (CART, M5 and RF), the importance of a variable can be ranked according to its relative influence, based on whether it is chosen for node splitting when growing trees and the extent to which it improves the average loss function across all trees (Krauss et al., 2017). All variable importance measures are then scaled up to a range between 0 and 100 and ranked in such a way that the most important variable has a maximum index value of 100. To avoid bias issues, we also incorporate the VIR analysis into the k-fold CV during the model training process.

However, the major limitation of VIR analysis is that it can only tell managers which predictors are important and which are not important, but cannot explain why. Therefore, despite its great popularity in practice, it is commonly criticised for being too theoretical and unable to provide a deep understanding of the subject matter (Grömping, 2015). To overcome this limitation, we employ the PDP method to extract more insights from RPDP, as discussed in the following section.

6.2.2 Partial dependence plot (PDP)

In practice, managers will often want to know not only which predictors are most important, but also how they affect the response variable. For ML models, especially black-box ones such as RF, a graphical visualisation such as a PDP is one of the most effective ways of determining this. PDP can visualise the non-linear, complex marginal effect of single and multiple (usually two) predictors on the response variable, while also taking into account the average effects of other predictors. Two points worth noting about using a PDP are that: (1) a PDP can only partially provide the marginal effect of a variable over a certain range of its values rather than its complete behaviour; and (2) a PDP can be misrepresented in the presence of a high-order interaction or a strong correlation between predictors (Goldstein, Kapelner, Bleich, & Pitkin, 2015). A brief description of Friedman's PDP is provided in Supplementary Materials_C.6.

7. Results and discussions

Following the data collection described in Section 3, we obtain a database with 5,693 remanufactured products listed on www.amazon.com. These include 169 remanufactured products which have no corresponding new products for price comparison. Since the missing data comprises less than 5% of the dataset, we remove these products, as explained in section 4.1. In addition, there are also 3,957 remanufactured products that have no customer review data. Given the fact that five out of the thirteen independent variables in our model are consumer-generated data, any imputation approach used to handle the missing data of customer reviews may mislead the research findings. We therefore exclude these products from the experiment. As a result of the pre-processing step, we have a database of 1,567 remanufactured products in total, along with 5,512 product pictures and 201,514 customer reviews. The dataset is split into two subsets: a training set (60% of the total data) and a testing set (40% of the total data). Table 1 presents the distribution summary and Variation Inflation Factor (VIF) of the variables in both datasets. The VIF is used to check for multicollinearity issues between variables. The maximum VIF is 3.89 in the training set and 3.14 in the testing set (far below the suggested threshold value of 10), indicating no significant multicollinearity issue within our datasets (Xu et al., 2017).

Table 1: Statistic description of data

	Training dataset[1] (N=943 observations)					Testing dataset[1] (N=624 observations)				
	Min	**Max**	**Mean**	**SD**	**VIF**	**Min**	**Max**	**Mean**	**SD**	**VIF**
Sales rank (in the electronics category)	4	434,734	117,177	143,940	1.35	26	434,734	118,179	146,179	1.32
Price difference (in fraction)	0	1	0.26	0.24	1.03	0	0.95	0.26	0.24	1.04
Product promotion rate (in fraction)	0	0.55	0.01	0.05	1.02	0	0.5	0.01	0.04	1.03
Positivity of product description	0	13	4.69	3.8	1.10	0	12	4.53	3.76	1.17
Number of product pictures	1	7	3.49	2.29	1.05	1	7	3.56	2.34	1.02
Overall product rating	1	5	3.78	0.68	1.43	1	5	3.78	0.7	1.41
Number of service successes	0	158	45.71	56.20	3.89	0	158	46.54	56.29	3.14
Number of service failures	0	80	21.53	27.07	3.48	0	80	22	27.79	2.58
Number of questions answered	0	135	31.09	38.5	2.30	0	135	30	38.29	2.81
Total number of helpful votes	0	54	11.00	13.87	1.05	0	54	10.81	14.32	1.05
Average sentiment of customer reviews	-3	3	0.65	0.84	1.26	-3	3	0.65	0.86	1.28
Stock information	Yes (332); No (611)					Yes (45); No (579)				
Warranty information disclosure	Yes (80); No (883)					Yes (65); No (559)				
Brand equity	High (635) ; Low (318)					High (426) ; Low (198)				

[1] *Sales ranks were recorded on 30 May 2018, while all predictors were recorded in the period up to 30 April 2018 (i.e. one month before the May sales rank).*

In this paper, all the algorithms are fitted into the training dataset and then validated with the testing set. A ten-fold CV repeated three times is applied throughout the modelling process in order to remove the resampling bias. The result of Boruta and RFE are shown in the Supplementary Materials_D.

7.1. Model predictive performance evaluation

To avoid bias issues, this paper employs five algorithms to train the prediction model of remanufactured products: a linear model (i.e., linear regression, LR); two non-linear, single tree-based ML models (i.e., CART and M5); a non-linear, tree-based, advanced ML model (i.e., RF); and a non-linear, non-tree-based, advanced ML model (i.e., the artificial neural network, ANN). In LR, the input-output relationship is modelled using the linear predictor functions. ANN, inspired by the human brain, typically consists of input, hidden and output layers connected by processing units, so-called neurons. Neurons between layers are connected by the synaptic weights, and the ANN learning algorithm updates their weights to map the relationship between the predictors and the target variable. The sum of the weighted predictors is applied to the activation function in order to generate the prediction value (Witten et al., 2011). Due to the limited space, the description of the LR and ANN model can be referred to Witten et al. (2011).

The model performance is assessed according to three measurements: MAE, RMSE and R^2. For a fair comparison, each model is run with its optimal hyperparameter settings and for both the RFE and the Boruta predictor sets, as shown in Table S4 in Supplementary Materials_E.

The prediction results in the model training and testing phase are shown in Figure 2. In this figure, the boxplots represent the statistical distribution (min, mean and max) of each performance metric which results from training the data with the ten-fold, three-time repeated CV, while the red point in each boxplot is the model performance based on the testing dataset in the model validation phase. In both the training and validation phase, it can be seen that RF outperforms the other predictive algorithms with the lowest values for errors and the highest R^2. It can also be seen in Figure 2 that all the values of MAE, RMSE and R^2 in the testing phase (the red points) fall within their corresponding distributions in the training phase (the boxplots). This indicates that there is no significant under-/overfitting problem. This validates our proposed approach to demand prediction.

Furthermore, we also calculate the p-value in the t-test to check the significance of the difference in the model performance between RF and the other algorithms. Based on Tables S5.1, S5.2 and S5.3 in

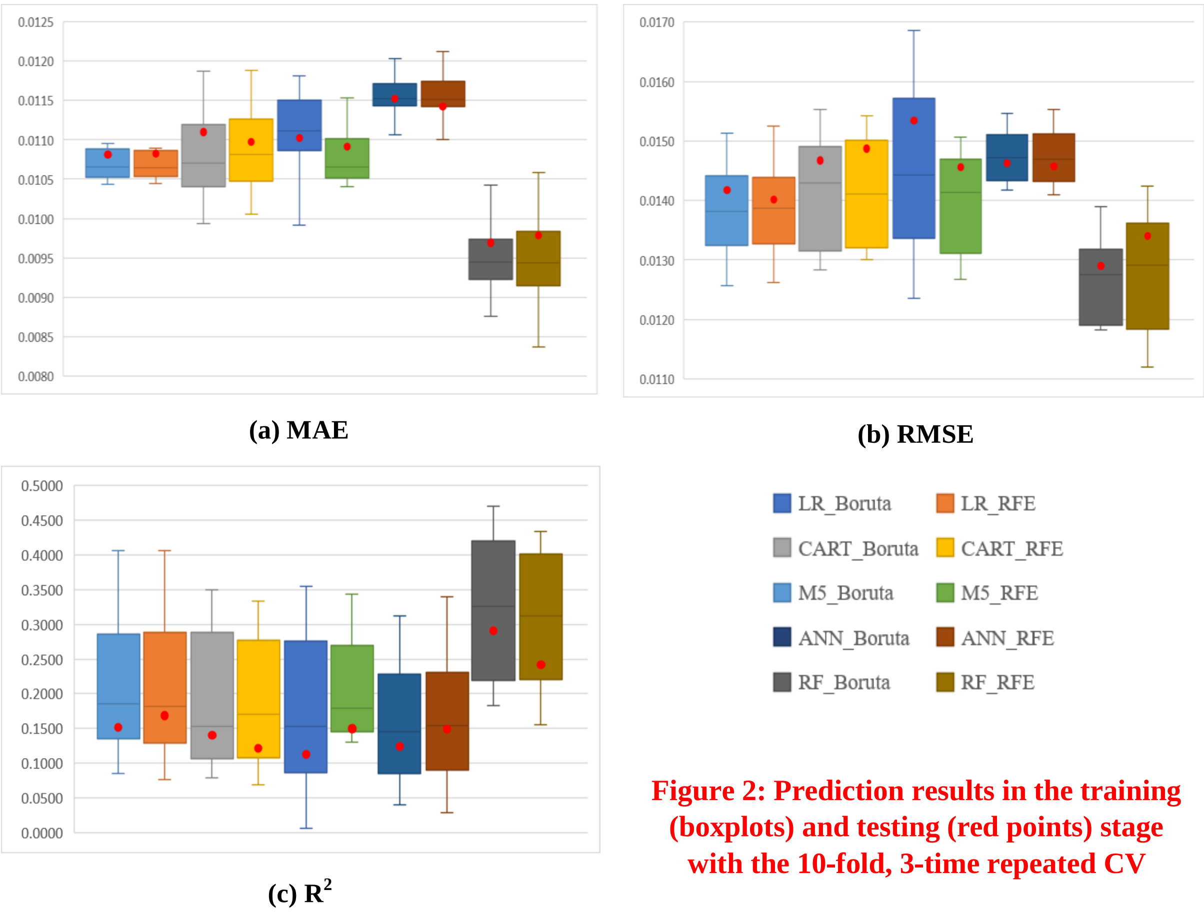

Figure 2: Prediction results in the training (boxplots) and testing (red points) stage with the 10-fold, 3-time repeated CV

the Supplementary Materials_F, the results of the p-values confirm that, compared to the other models, RF significantly improves the prediction accuracy of remanufactured product demand. In addition, there is an insignificant performance difference between the RF when using the Boruta predictor set and when using the RFE predictor set.

RF outperforms the LR model, indicating that remanufactured product demand is highly non-linear, a finding which requires the use of a non-parametric approach. Another interesting finding is that there is no significant improvement in model performance when using the Boruta predictor sets, compared to the RFE predictor sets. This implies that the predictive power of price difference, the number of product pictures and stock information on customer demand of remanufactured products are weak.

The RF model outperforms the other models in the prediction of remanufactured product demand, so the discussion in the following section is based on the results of this model.

7.2. Model deployment – interpretation of results and managerial insights

As Kunc & O'Brien (2018) suggest, this paper adopts a multi-methodological approach that combines different predictive and descriptive analytics tools to support the development of a marketing strategy for remanufactured products. Based on the RF prediction model in Section 7.1, we further employ two descriptive analytics tools in order to gain data-driven marketing insights into remanufactured products. In particular, the VIR tool is used to identify the most influential market factors based on their predictive powers in the RF model. The PDP tool is then applied, in order to demonstrate the non-linear effect of these factors on customer purchasing behaviours. Such marketing insights can help marketers to understand the different ways in which customers of remanufactured products respond to marketing tools, compared to customers of new products, thereby developing an effective marketing strategy that can maximise the sales of remanufactured products.

7.2.1 Results and discussion of VIR analysis

In order to gain management insights from the RF-based prediction model, we first conduct the VIR analysis for both the REF (Figure 3a) and Boruta predictor sets (Figure 3b). For the comparison, the VIR chart normalises the predictive power of each independent variable in the range between 0 (weakest) and 100 (strongest). In both the RF_RFE and RF_Boruta models, the number of questions answered, the number of service failures and brand equity (especially low brand equity) are perceived as the most important predictors of remanufactured product demand. The VIR also confirms that price difference, the number of product pictures and stock information are weakly important predictors, as explained in the previous section. The remaining predictors have moderately predictive powers on customer demand of remanufactured products. These include the number of service successes, the number of helpful votes, the overall product rating, the positivity of the product description, and the average sentiment of customer reviews.

However, rank the importance of variables is not enough on its own to provide managers with practicable insights. Therefore, a complementary analysis of PDP is necessary to examine the

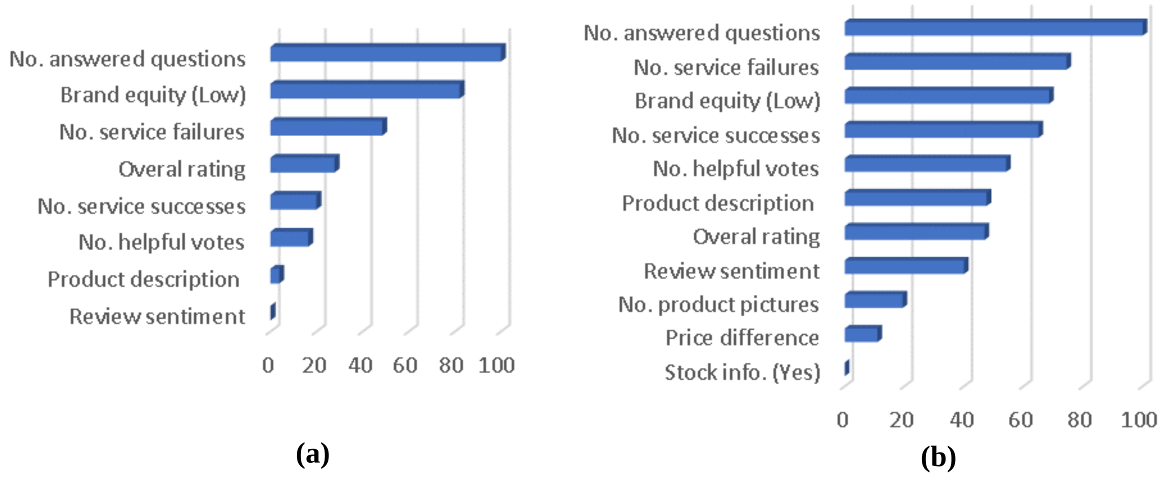

Figure 3: VIR based on the prediction result of RF_RFE (a) and RF_Boruta (b)

direction and magnitude of the impact of predictors on customer demand.

7.2.2 Results and discussion of PDP analysis

To extract the practicable insights from the RPDP model, we analyse the PDP plots that illustrate the marginal effect of eight strongly and moderately important predictors of remanufactured product demand, as previously suggested by the VIR. It is notable that the PDP plots are drawn from partial dependence functions that are each populated with only a single predictor at any one time. As such, they show the nature of the relationship of a single predictor to customer demand after taking into account the average effects of all the other predictors in the model. While these plots do not fully represent the effect of each variable, they can serve as a useful basis for interpretation (Goldstein et al., 2015). In order to enhance the insights for management, we also analyse the PDP contour plot representing the joint effect between two predictors on the response variable.

- Sales effect of the number of service failures and service successes

As with new products, the VIR analysis suggests that the number of service failures is more important than the number of service successes. The result is also supported by the PDP in Figure 4, which shows that the sales impact of service successes is more complex and non-linear than that of service failures. As a result, our predictive algorithm (RF) captures less statistical (linear) correlation between it and customer demand. This can be explained by negativity bias: in other words, the psychological tendency for people to pay more attention, and give greater diagnostic weight, to negative information than neutral and positive information.

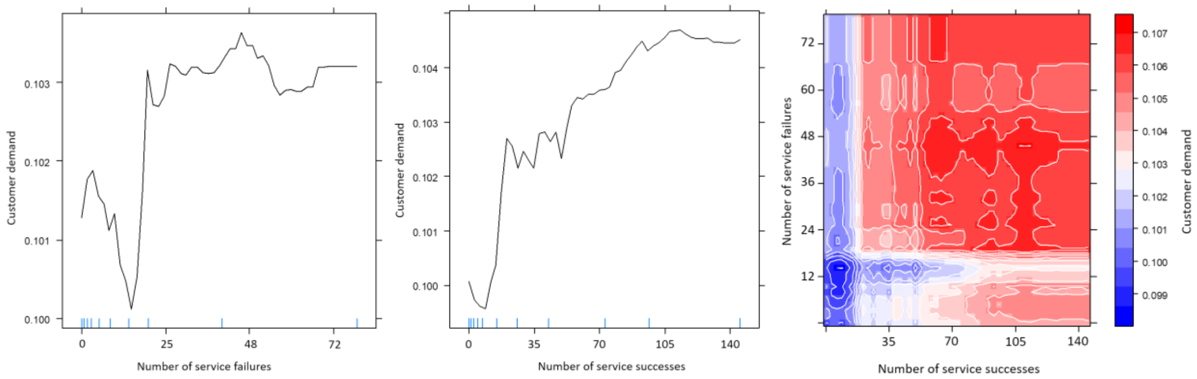

Figure 4: The PDP of number of service failures and service successes

Additionally, according to Park & Lee (2008), the number of service failures can play two roles in terms of social impact: an informant role, indicating the informativeness of customer feedback, and a recommender role, indicating product popularity. It is interesting at this point to examine how these two roles of the variable can influence customer purchasing behaviour related to remanufactured products. Interestingly, the PDP shows that when only a small amount of negative customer feedback is available (between 4 and 15 cases of negative feedback), the variable has a significant and negative association with customer demand. However, the effect becomes fairly positive and more non-linear when more than 15 customers give negative feedback.

This finding can be explained using the elaboration likelihood model (ELM), which is a theory of social psychology that describes the ways in which the information process of online customer reviews changes customer purchasing behaviour (Ho & Bodoff, 2014). According to the ELM, when only a limited amount of negative feedback is available, customers are more motivated to engage in a thoughtful and effortful information process, meaning that they read through the review content to look for quality cues of the product before making a purchasing decision. As customers examine the negative feedback carefully to gain better product knowledge, the informativeness effect of this variable outweighs its popularity effect. This means that an increasing number of service failures will significantly reduce customer demand. We also observe an interesting finding which indicates a positive effect of negative customer feedback on sales, as in the case of remanufactured products with a very low number of service failures (fewer than 4 cases of negative feedback). As the ELM suggests, customers are very likely to carefully evaluate the product attributes from different reviews, leading to enhanced product knowledge and higher confidence, which may accordingly translate into higher sales. In contrast, when a large amount of negative feedback (more than 15 cases) is available, the ELM suggests that customers tend to adopt a low-depth information process which uses non-content, peripheral cues, such as a numerical index of product popularity, in order to make their buying decisions. Therefore, in such cases, when the number of service failures increases, indicating higher product popularity, the sales of remanufactured products also increase. However, the PDP suggests that this positive effect is much less significant and more non-linear, compared to the case outlined above of a low number of service failures.

Likewise, the marginal effect of the number of service successes is also in line with the ELM, and can therefore be explained in a similar way. On average, its effect on remanufactured products with low popularity (less than 15 cases of positive feedback) is significantly positive, whereas the effect on those with high popularity is less significant because of the highly complex and non-linear variances.

The joint effect of the number of service failures and the number of service successes on sales of remanufactured products also deserves attention. The contour plot suggests that there are strong interactions between these two variables. More particularly, remanufactured products with a high number of service failures (more than 15 cases of negative feedback) are associated with high customer demand only if there is also a high number of service successes (more than 15 cases of positive feedback), and vice versa. This finding highlights the recommender role of these two variables, showing that they are important indicators of product popularity. By increasing the amount of customer feedback, the seller can signal a strong awareness effect about the existence of the product, thereby placing it in the choice set of potential customers.

- Sales effect of overall ratings and customer review sentiments

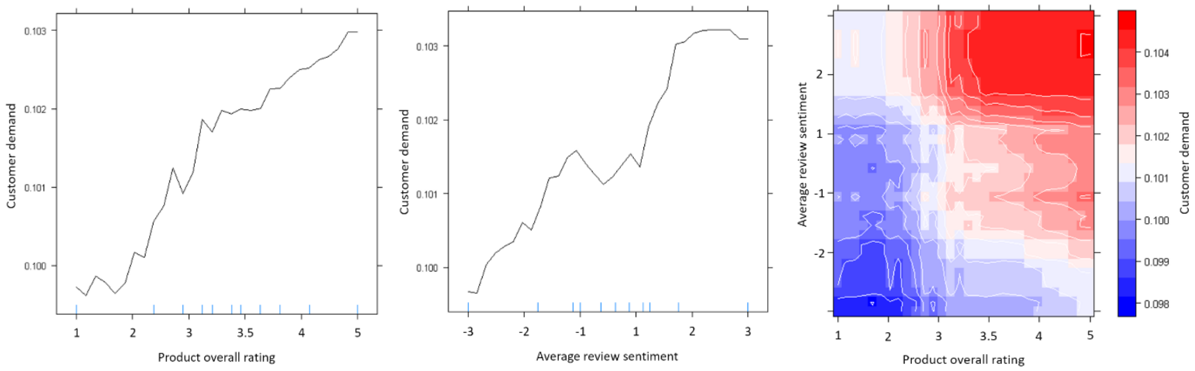

Figure 5: The PDP of overall ratings and customer review sentiments

Overall ratings and review sentiments are both representations of crowd intelligence. They are therefore perceived as trustworthy sources of information and are closely related to customer perception. The VIR result shows that the sales impact of overall ratings is more influential than that of review sentiments. This contradicts previous findings for new products which suggest that review sentiment has a more direct and substantial effect on demand than overall ratings (Hu et al., 2014). Our result is supported by rationality boundary theory, which suggests that customers often seek to reduce their cognitive efforts by making product evaluations and purchasing decisions based on information that takes less effort to process and align, such as numerical ratings, rather than more effortful strategies such as reading textual customer reviews (Shah & Oppenheimer, 2008).

As Figure 5 demonstrates, the PDP shows that the overall product rating has a monotonic positive association with remanufactured product demand. This means that there tends to be higher demand for remanufactured products with higher overall ratings, which indicate positive seller reputation and goodwill accumulated over a long period. However, the magnitude of this impact varies greatly over the rating value range. This effect is moderately significant for remanufactured products with overall ratings above 3.5 stars; very significant for those rated between 1 and 3 stars; and mostly levelled off for those with less than 2 stars. Like the number of service failures, this finding can also be explained using negativity bias: i.e. unfavourable product ratings are likely to have a greater effect on purchase intention than favourable ones.

In contrast, the effect of review sentiments on remanufactured product demand is, on average, more non-linear in a heterogeneous way. In particular, the impact is relatively linear and significantly positive for remanufactured products with customer reviews which indicate negative polarity. For products with customer reviews indicating a positive polarity, the sales impact of the variable is significantly positive. However, such an effect is levelled off if the customer reviews indicate a strongly positive polarity. Finally, customer demand for remanufactured products with reviews indicating neutral polarity becomes very complex and non-linear. This is because prospective customers often need more time and cognitive effort to process and judge product reviews which have neutral sentiments (i.e. neither positive nor negative). Following rationality boundary theory, they therefore tend to be more risk-averse to remanufactured products which trigger great debate and require greater cognitive effort to evaluate.

The joint effect of overall ratings and review sentiments also provides some interesting insights for management. The contour plot shows that these two variables interact strongly in a heterogeneous way to affect customer demand for remanufactured products. It further suggests that remanufactured products with good reputations indicated by high overall ratings (3 to 5 stars) do not necessarily produce high sales, unless the average sentiment based on customer textual reviews also indicates a positive polarity. This is because only using numerical ratings cannot fully capture the information embedded in the product reviews. Indeed, by compressing complex, text-based customer reviews to a single number, the product is implicitly assumed to be one-dimensional, even though the economic theory for product differentiations posits that products are compounded with multiple attributes and each attribute has a different perceived importance to consumers, based on their individual preferences. Therefore, prospective customers do not simply base their decisions on the numerical overall ratings in order to reduce their cognitive effort, but also read textual reviews to gain more detail-rich information about the experiences, feelings and emotions of past customers.

- Sales effect of product description positivity

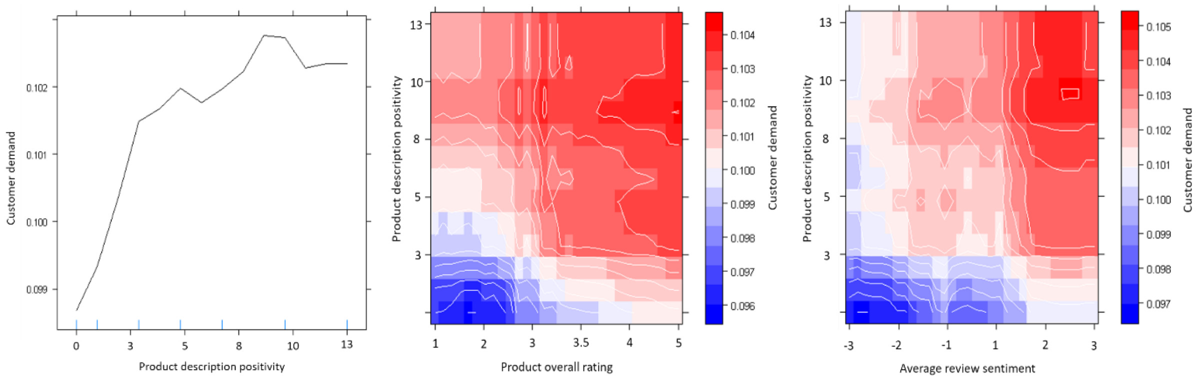

Figure 6: The PDP of number of product description positivity

Of the three quality cues designed by online sellers (textual product descriptions, product pictures and warranty information disclosure), our VIR suggests that the textual product description is the most effective signalling tool with which to convey the quality and condition of remanufactured products. In contrast to new products, this study suggests that the product description is one of the more important quality cues for customers of remanufactured products. As a result, it is ranked as a moderately important predictor of demand in the VIR analysis. This may be because seller-generated product descriptions of new products are static and do not generally include intangible features, such as product quality, robustness, performance and so on. They may therefore be unimportant determinants of customers' purchasing decisions. However, such intangible product features are often contained in the sellers' descriptions of remanufactured products – with descriptors such as 'like-new', 'certified' and 'no scratches' – in order to directly signal the condition of the product to customers. In this way, sellers can counteract the asymmetric information of the buyers and their perceived uncertainty, leading to higher sales, as suggested by market signalling theory.

Regarding the marginal effect of this variable, the PDP in Figure 6 shows that, on average, the positivity of product descriptions has a positive association with sales. This effect is significant and linear for remanufactured products with descriptions indicating low positivity (i.e. those that contain fewer than three positive quality-related keywords). However, such an effect becomes moderate and relatively non-linear for RPs with strongly positive descriptions (containing between 4 and 9 such keywords). Nevertheless, as product descriptions are seller-generated rather than consumer-generated, the excessive use of this marketing tool may increase customer scepticism. As a consequence, our PDP shows that the variable has no effect or even a negative effect on sales when the product descriptions contain too many positive keywords (more than 9).

For practical uses, we use the PDP contour plots to examine how the positivity of product descriptions interacts with the other strongly and moderately important predictors of remanufactured product demand (excluding brand equity, as it is a categorical variable). The results suggest that the interactions between product description positivity and most of the consumer-generated variables are weak, and that increasing the positivity of product descriptions does not increase sales. See Figure 6 (right) for an example of its weak interaction with average review sentiment. However, this variable is found to have a strong interaction with overall product ratings, especially in the case of negative product ratings (lower than 3 stars). As Figure 6 (middle) suggests, when remanufactured products suffer low demand because of low product ratings which indicate a negative reputation, managers can raise sales to the average level by using product descriptions (ideally containing between 4 and 9 positive quality-related keywords) with strong implications about the high quality of the product.

- Sales effect of number of answered questions, number of helpful votes, and brand equity

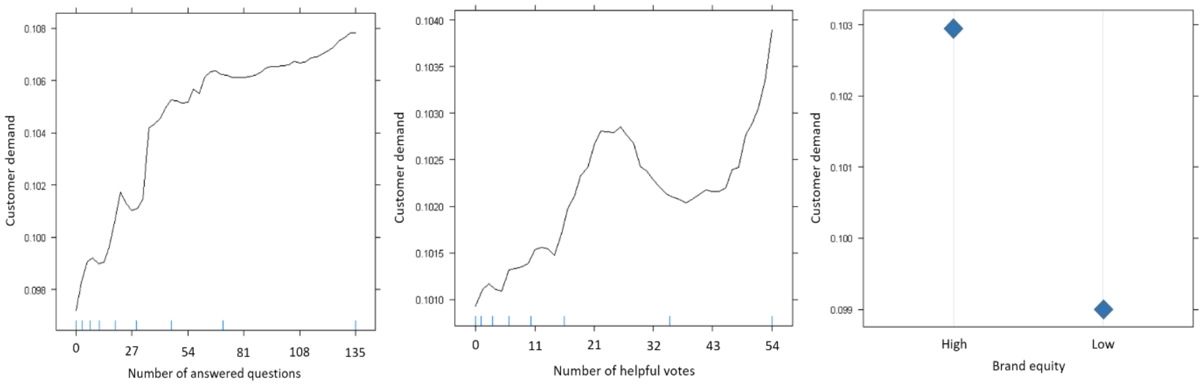

Figure 7: The PDP of number of answered questions, number of helpful votes

The VIR results suggest that the number of answered questions plays an important role in predicting customer demand for remanufactured products. More specifically, the PDP in Figure 7 shows that, on average, this variable has a monotonic positive influence on customer demand. Such an impact is significant when the number of questions ranges from 0 to 40. According to trust transference theory, when customers post more questions about a product, they stimulate social interactions and information exchange in the online community, which can help to ease the burden of information asymmetry, increase online trust and ultimately lead to higher demand. However, the positive effect of this variable becomes less significant when its value is high (more than forty questions answered). According to bounded rationality theory, customers' rationality is restrained by their cognitive limitations. They are therefore more likely to avoid information that requires a large amount of cognitive effort, such as when too many answered questions are available.

According to make signal theory, the number of helpful votes indicates the credibility of the review and this has a positive link with customers' online trust and purchase intentions. Previous research has therefore found a significant and positive effect connecting this variable with sales of new products (De Maeyer, 2012). For remanufactured products, our results show that its predictive power of demand is moderately high and that the impact on sales is generally positive but in a non-linear fashion. More specifically, based on the PDP in Figure 7, the impact on sales is significantly positive when the perceived credibility of product reviews is either low (fewer than 21 helpful votes) or high (more than 37 helpful votes). However, when review credibility is at an average level (between 21 and 37 helpful votes), the impact on sales becomes negative. In signalling theory, over such a range of values, the variable sends out mixed signals (neither high nor low) about review credibility, which often requires prospective customers to make a greater cognitive effort in order to arrive at a judgement. However, since the cognitive resource of customers is limited, as rationality boundary theory suggests, they tend to become more risk-averse and less attracted to products which require a greater effort to evaluate. This accordingly has negative effects on sales. Our PDP result also supports a positive link between brand equity and sales of remanufactured products, so that products with high/low brand equity are associated with high/low sales. In addition, the VIR suggests that brand equity is a strongly important predictor of remanufactured product demand, especially for those with low brand equity. Following market signalling theory, the brand equity of the OEM can be perceived as an initial cue signalling the reliability and quality of the product, justifying its perceived attractiveness and WTP.

8. Robustness check

8.1. Test 1: Using the number of customer reviews

We replace separate measures of product popularity and seller reputation (i.e. the number of service failures and the number of service successes) with a single measure: the total number of customer reviews. We do not find any significant change in the prediction performance. As seen in Supplementary Materials_G, the VIR shows that the new variable replaces the number of service failures as one of the three most important predictors. The importance ranking and PDP of the other variables remains the same. Unlike the number of service failures and the number of service successes, the number of customer reviews only indicates the popularity of the product (performing the recommender role), so its marginal effect is less complex than theirs. In line with signalling theory and the ELM, its sales impact is significant and positive for low popular remanufactured products, and becomes nonlinear and insignificant for high popular remanufactured products.

8.2. Test 2: Using the new dataset in different time periods

In the previous sections, we use the data of 1,567 remanufactured products listed on Amazon between April and May 2018. For a robustness check, we use the tracking crawler to collect new data for these products over the following six months. In particular, the sales rank was recorded up to 15 November 2018, while the lagged market factors were recorded up to 15 October 2018. There are 1,200 among 1,567 products available over this period. We find results that are consistent with the April-May data with regard to the predictive performance, VIR and PDP.

8.3. Test 3: Using different types of remanufactured products

To check whether our proposed approach for RPDP is valid across different product categories, we collect a new April-May dataset that includes remanufactured products from the Home & Kitchen category in Amazon. This is the second largest category for remanufactured products after Electronics. It includes a wide range of small household appliances and kitchen equipment, such as food processors, coffee makers, mixers and vacuum cleaners. Previous CLSC literature distinguishes these remanufactured household products from the remanufactured technological products using the product repulsive level. Product repulsion refers to a customer's irrational and ingrained belief that pre-owned products are permanently tainted and therefore repulsive (Abbey, Meloy, et al., 2015). This means that technological products such as remanufactured laptops are "around-you" products which are associated with low repulsion. Household products such as remanufactured food processors are "on-you" products with medium to high repulsion. Personal care products such as remanufactured toothbrushes are "in-you" products which are associated with high repulsion. A more detailed classification of remanufactured products based on product repulsion can be found in Abbey, Meloy, et al. (2015).

We find 967 remanufactured household products over the period between April and May 2018. The RPDP approach described above is used to predict the May log sales rank based on the lagged market factors in April. Regarding predictive performance, RF still outperforms the other model with the highest prediction. The results show that eleven of the thirteen independent variables (excluding brand equity and product description positivity) have a consistent impact, as in the case of technological products. This means that the product category and product repulsion on customer purchasing behaviour of remanufactured products did not have pronounced effects.

However, there are still some interesting results that deserve attention. As the VIR figures show (see Supplementary Materials_H), household products contrast with technological products, as brand equity and product description positivity have the lowest predictive powers, indicating that customers do not perceive them as cues signalling quality. This is because when customers hold a high repulsive perception, as in the case of household products, the perceived risks and uncertainty about the quality of the remanufacturing process become more severe. This means that they are more reliant on consumer-generated variables reflecting crowd intelligence than on seller-generated variables as trustworthy sources of information. This is why all consumer-generated variables for remanufactured household products have moderate to significant predictive power of customer demand, as shown in the VIR.

A summary of our main findings based on the VIR and PDP analysis is presented in Table 3.

Table 3: Summary of main findings in this paper

Predictor	Signal Implication	Predictive power of RP demand	Management implications of the findings on remanufactured product (RP) demand	Previous findings on new products
Number of questions answered	Social interaction	Strong	(1) The effect is monotonic positive. (2) Such an effect is more significant when the number of answered questions ranges from 0 to 40, and less significant when there are more than 40 questions answered.	Significantly positive
Brand equity	Quality cue	Strong (technological RP)	For technological products featuring a low customer repulsion level, the brand equity of the original manufacturer has a significant and positive effect on RP demand. This means that RP demand decreases when moving from high to low brand equity. The effect of low brand equity is also more pronounced than that of high brand equity.	Moderately positive
		Limited (Household RP)	For household products featuring medium to high repulsion level, the brand equity of the original manufacturer does not have significant effect on RP demand. Customers are more reliant on information generated by peer customers, such as customer reviews and ratings to make a decision.	
Number of Service failures	Seller reputation, Product popularity	Strong	(1) For low popular products, the effect is significantly negative when there is only a small amount of negative feedback available (between 4 and 15), and becomes significantly positive when there are fewer than 4 cases of negative feedback. (2) For high popular products, the effect becomes somewhat positive and more non-linear when there is a high number of service failures (more than 15 negative feedback). (3) The impact of service failures is more important than service successes.	Significantly negative
Number of Service successes	Seller reputation, Product popularity	Moderate	(1) For low popular products, the effect, on average, is significantly positive when there are less than 15 cases of positive feedback. (3) For high popular products, the effect is highly complex and very non-linear, and hence statistically unimportant. (3) Products with the highest customer demand are those which have	Moderately positive

			both a high number of negative feedback (more than 11 cases of feedback) and a high number of positive feedback (more than 15 cases of feedback), which indicate the maximum product popularity and customer awareness.	
Overall product rating	Seller reputation	Moderate	(1) On average, the overall rating has a monotonic positive effect on sales, but the impact magnitude varies greatly. (2) The positive effect is moderately significant for products with positive ratings (more than 3.5 stars), very significant for those with neutral and negative ratings (between 2 and 3 stars), and mostly levelled off for those with very negative ratings (fewer than 2 stars) (3) However, its positive effect is largely dependent on the textual sentiments of customer reviews. For example, products with good reputation indicated by high overall ratings (3 to 5 stars) does not necessarily lead to high sales, unless the average sentiment based on customer textual reviews also indicates a positive polarity.	Moderately positive
Customer review sentiment	Seller reputation	Moderate	(1) This variable is less important than the overall rating because, on average, the effect of this variable is more non-linear than that of the overall rating in a heterogeneous way. (3) The effect is significantly positive for products with customer reviews which indicate negative polarity. (2) The effect is significantly positive for products with customer reviews which indicate positive polarity, but such an effect is levelled off for products with strongly positive polarity. (3) The effect is very complex and non-linear for products with customer reviews which indicate neutral polarity (neither negative nor positive polarity).	Significantly positive
Number of helpful votes	Review credibility	Moderate	(1) On average, the effect is significantly positive, except for products which indicate neither very low nor very high review credibility (i.e. 21 to 37 helpful votes) for which the effect becomes negative.	Significantly positive
Product description	Textual quality cue	Moderate (technological RP)	(1) The effect is significant and linear when the product descriptions indicate low positivity (containing fewer than three positive quality-related keywords); and becomes moderate and relatively non-linear with strongly positive product descriptions (containing between 4 and 9 such keywords. The effect is levelled off or becomes negative if the product descriptions contain too many positive keywords (more than 9). (2) The variable moderates the sales effect of the overall product rating. Therefore, when products have low demand because of low product ratings (fewer than 3-stars), managers can raise sales to the average level by using product descriptions (which ideally contain between 4 and 9 positive quality-related keywords).	Insignificant
		Limited (Household RP)	For remanufactured versions of household products, customers do not perceive the product descriptions provided by the sellers as reliable quality cues.	
Product pictures	Visual quality cues	Limited	Customers do not see the product pictures as important cues signalling the quality and condition of the RP.	Positive
Warranty information disclosure	Quality cue	None	Since few Amazon sellers voluntarily provide the full terms and conditions of warranties for the RPs, the effect of this variable on RP demand is unconfirmed in this study, but is worth investigating in future studies.	Positive
Price difference	Price discount	Limited	Using price incentives is not an effective way of increasing sales. Sellers should instead focus on improving their reputation and goodwill by obtaining a higher number of positive ratings from past customers and by becoming more actively involved in question-and-answer activities.	Positive
Promotion rate	Price discount	None		
Stock information	Limited availability	Limited	The psychological effect of the scarcity principle does not apply to RP customers; therefore, the selling strategy of using limited stock availability does not help boost sales.	Positive

9. Conclusion

This paper develops a comprehensible data-mining approach in order to predict remanufactured product demand and to develop a marketing strategy. Based on the real dataset on www.amazon.com, the results show that the ensemble regression tree model, RF, can provide the most accurate and robust prediction result. A VIR analysis is then applied to determine the most influential market factors based on this prediction model. The effect of these factors on RP demand are examined using PDP analysis. As a result, a number of data-driven marketing insights are revealed, providing guidelines to help managers design an effective marketing strategy specific to remanufactured products, as summarised in Table 3 above.

We believe that this research can be grouped with other pioneering studies which provide a structured framework showing how business analytics can be applied to support data-driven demand forecasting and marketing strategy development in the remanufacturing/CLSC literature. Future research should focus on three aspects. (1) The findings of this paper rely on data from only one online marketplace, www.amazon.com. In practice, however, customer behaviour could change when comparing product deals from multiple online sales channels. Therefore, it may be beneficial to aggregate data from various sources. (2) ML models are often seen as black boxes, which limits their applications in industry. Future research should make a greater effort to develop new methodologies that can effectively explain the results of these apparently incomprehensible models. (3) The application of (big-)data analytics in reverse logistics and CLSCs is yet to be fully appreciated, and there is currently little research into these areas. This research is expected to serve as an example and stimulate further investigation.

Reference:

Abbey, J. D., Blackburn, J. D., & Guide, V. D. R. (2015). Optimal pricing for new and remanufactured products. *Journal of Operations Management*, *36*, 130–146.

Abbey, J. D., Meloy, M. G., Guide, V. D. R., & Atalay, S. (2015). Remanufactured Products in Closed-Loop Supply Chains for Consumer Goods. *Production and Operations Management*, *24*(3), 488–503.

Akerlof, G. A. (1970). The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488.

Alaei, A. R., Becken, S., & Stantic, B. (2017). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, *58*(2), 175–191.

Alqahtani, A. Y., & Gupta, S. M. (2017). Evaluating two-dimensional warranty policies for remanufactured products. *Journal of Remanufacturing*, *7*(1), 19–47.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, *57*(8), 1485–1509.

Atasu, A., Guide, V. D. R., & Van Wassenhove, L. N. (2008). Product Reuse Economics in Closed-Loop Supply Chain Research. *Production and Operations Management*, *17*(5), 483–496.

Atasu, A., Guide, V. D. R., & Van Wassenhove, L. N. (2010). So What If Remanufacturing Cannibalizes My New Product Sales? *California Management Review*, *52*(2), 56–76.

Bohanec, M., Kljajić Borštnar, M., & Robnik-Šikonja, M. (2017). Explaining machine learning models in sales predictions. *Expert Systems with Applications*, *71*, 416–428.

Boulding, W., & Kirmani, A. (1993). A Consumer-Side Experimental Examination of Signaling Theory: Do Consumers Perceive Warranties as Signals of Quality? *Journal of Consumer Research*, *20*, 111–123.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book

Reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Cialdini, R. B. (2009). *Influence: Science and Practice* (5th ed.). Pearson.

Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, *17*(1), 39–58.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792.

De Bock, K. W. (2017). The best of two worlds: Balancing model strength and comprehensibility in business failure prediction using spline-rule ensembles. *Expert Systems with Applications*, *90*, 23–39.

De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*(2), 760–772.

De Maeyer, P. (2012). Impact of online consumer reviews on sales and price strategies: a review and directions for future research. *Journal of Product & Brand Management*, *21*(2), 132–139.

Dekkers, R. (2011). Impact of strategic decision making for outsourcing on managing manufacturing. *International Journal of Operations & Production Management*, *31*(9), 935–965.

Frota Neto, J. Q., Bloemhof, J., & Corbett, C. (2016). Market prices of remanufactured, used and new items: Evidence from eBay. *International Journal of Production Economics*, *171*, 371–380.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65.

Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, *7*(2), 137–152.

Hamzaoui-Essoussi, L., & Linton, J. D. (2014). Offering branded remanufactured/recycled products: at what price? *Journal of Remanufacturing*, *4*(1), 9.

Hamzaoui Essoussi, L., & Linton, J. D. (2010). New or recycled products: how much are consumers willing to pay? *Journal of Consumer Marketing*, *27*(5), 458–468.

Ho, S. Y., & Bodoff, D. (2014). The Effects of Web Personalization on User Attitude and Behavior: An Integration of the Elaboration Likelihood Model and Consumer Search Theory. *MIS Quarterly*, *38*(2), 497–520.

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems*, *57*, 42–53.

Jiménez-Parra, B., Rubio, S., & Vicente-Molina, M.-A. (2014). Key drivers in the behavior of potential consumers of remanufactured products: a study on laptops in Spain. *Journal of Cleaner Production*, *85*, 488–496.

Khor, K. S., & Hazen, B. T. (2017). Remanufactured products purchase intentions and behaviour: Evidence from Malaysia. *International Journal of Production Research*, *55*(8), 2149–2162.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137–1145).

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689–702.

Kunc, M., & O'Brien, F. A. (2018). The role of business analytics in supporting strategy processes: Opportunities and limitations. *Journal of the Operational Research Society*, *70*(6), 974–985.

Li, B., Ch'ng, E., Chong, A. Y.-L., & Bao, H. (2016). Predicting online e-marketplace sales performances: A big data approach. *Computers & Industrial Engineering*, *101*, 565–571.

Li, H., Fang, Y., Wang, Y., Lim, K. H., & Liang, L. (2015). Are all signals equal? Investigating the differential effects of online signals on the sales performance of e-marketplace sellers. *Information Technology & People*, *28*(3), 699–723.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198.

Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research*, *183*(3), 1466–1476.

Masci, C., Johnes, G., & Agasisti, T. (2018). Student and school performance across countries: A machine learning approach. *European Journal of Operational Research*, *269*(3), 1072–1085.

Mazhar, M. I., Kara, S., & Kaebernick, H. (2007). Remaining life estimation of used components in consumer products: Life cycle data analysis by Weibull and artificial neural networks. *Journal of Operations Management*, *25*(6), 1184–1193.

Milgrom, P. R., & Weber, R. J. (1982). A Theory of Auctions and Competitive Bidding. *Econometrica*, *50*(5), 1089.

Ng, C. S.-P. (2013). Intention to purchase on social commerce websites across cultures: A cross-regional study. *Information & Management*, *50*(8), 609–620.

Nguyen, T., ZHOU, L., Spiegler, V., Ieromonachou, P., & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, *98*, 254–264.

Nilsson, R., Peña, J. M., Björkegren, J., & Tegnér, J. (2007). Consistent Feature Selection for Pattern Recognition in Polynomial Time. *The Journal of Machine Learning Research*, *8*, 589–612.

Ovchinnikov, A. (2011). Revenue and Cost Management for Remanufactured Products. *Production and Operations Management*, *20*(6), 824–840.

Oztekin, A., Kizilaslan, R., Freund, S., & Iseri, A. (2016). A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*, *253*(3), 697–710.

Park, D.-H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, *7*(4), 386–398.

Quinlan, J. R. (1992). Learning with continuous classes. *Machine Learning*, *92*, 343–348.

Rudnicki, W. R., Wrzesień, M., & Paja, W. (2015). All Relevant Feature Selection Methods and Applications. In *Feature Selection for Data and Pattern Recognition* (pp. 11–28). Berlin: Springer.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Subramanian, R., & Subramanyam, R. (2012). Key Factors in the Market for Remanufactured Products. *Manufacturing & Service Operations Management*, *14*(2), 315–326.

Swami, S., & Khairnar, P. J. (2006). Optimal normative policies for marketing of products with limited availability. *Annals of Operations Research*, *143*(1), 107–121.

Tereyağoğlu, N. (2016). Market Behavior Towards Remanufactured Products. In *Environmentally Responsible Supply Chains* (pp. 19–28). Springer, Cham.

van Heijst, D., Potharst, R., & van Wezel, M. (2008). A support system for predicting eBay end prices. *Decision Support Systems*, *44*(4), 970–982.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, *218*(1), 211–229.

Wang, Y., & Hazen, B. T. (2016). Consumer product knowledge and intention to purchase remanufactured products. *International Journal of Production Economics*, *181*, 460–469.

Wells, J. D., Valacich, J. S., & Hess, T. J. (2011). What Signal Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions. *MIS Quarterly*, *35*(2), 373.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining : practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.

Xu, X., Zeng, S., & He, Y. (2017). The influence of e-services on customer online purchasing behavior toward remanufactured products. *International Journal of Production Economics*, *187*, 113–125.

Yan, W., Xiong, Y., Xiong, Z., & Guo, N. (2015). Bricks vs. clicks: Which is better for marketing remanufactured products? *European Journal of Operational Research*, *242*(2), 434–444.

Yang, L., Liu, S., Tsoka, S., & Papageorgiou, L. G. (2017). A regression tree approach using mathematical programming. *Expert Systems with Applications*, *78*, 347–357.

Zhou, L., & Disney, S. M. (2006). Bullwhip and inventory variance in a closed loop supply chain. *OR Spectrum*, *28*(1), 127–149.

Zhou, L., Xie, J., Gu, X., Lin, Y., Ieromonachou, P., & Zhang, X. (2016). Forecasting return of used products for remanufacturing using Graphical Evaluation and Review Technique (GERT). *International Journal of Production Economics*, *181*, 315–324.

Zhu, Y., Zhou, L., Xie, C., Wang, G.-J., & Nguyen, T. V. (2019). Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *International Journal of Production Economics*, *211*, 22–33.