Preprints of the
16th IFAC Symposium on Information Control Problems in Manufacturing
Bergamo, Italy. June 11-13, 2018

Modeling resale value of road compaction equipment: a data mining approach

Syed Azad Chowdhery, Marco Bertoni

Blekinge Institute of Technology, Karlskrona, SE-37168
Sweden (Tel: +46-0455-385526; e-mail: syed.azad.chowdhery@bth.se, marco.bertoni@bth.se).

Abstract: Resale value is an important aspect to be addressed when companies aim to shift from one-sale models to Product Service Systems (PSS). In the initial stages of the PSS design process, it is beneficial to predict how the mechanical features of the PSS hardware will impact resale value, so to orient business strategy decisions accordingly. The objective of this work is to propose a methodology to model the resale value of road compaction equipment using data mining techniques. By scrapping and merging data sets from the machine manufacturer and from dealers of second-hand machines, the work discusses how the derived correlations may be used to populate value models for early stage decision making.

Keywords: Data-driven decision making, product service system, new product development, decision support systems, data mining.

1. INTRODUCTION

Second hand markets are an important source of revenue for manufacturers, in particular for high-end products. Hence, resale value is an important aspect to consider when companies shift from one-sale models to Product Service Systems (PSS) (Ölundh and Ritzén 2003). Intuitively, higher resale value raises the appeal of traditional 'one-sale' models because it lowers the customer Total Cost of Ownership (TCO) in the sales system, positively influencing the decision to buy a product (Oraiopoulos et al. 2012; Waldman 2003). At the same time, it reduces the appeal of more service-oriented business models. This forces PSS providers to decrease the fees of their functional offers, to make them more attractive than 'buying' (van Loon, et al. 2017).

Product-oriented services (POS) (Tukker and Tischner 2006) are a PSS type that represents an exception to this rule. POSs are an early PSS transition step for many manufacturers and are still largely associated with the sale of products to consumers. In this business model, products are sold together with additional services, such as a maintenance contract or an end of life take-back agreement. In this situation, a higher resale price is a desirable feature, and the designers of the technical hardware shall aim at maximizing its appeal for second-hand customers. A challenge is then to predict how the mechanical properties of the hardware are expected to influence the resale value dimension.

Interpreting information about resale market prices and linking it to the requirements for the PSS hardware is a difficult and time-consuming task. While several contributions discuss the role of equipment and machines in second-hand markets (e.g., Bartolomeo et al. 2003, Holmström et al. 2010), they neither attempt to link specific mechanical properties to price, nor to position the discussion from a PSS standpoint. As discussed by Harding et al. (2001) market information is not often used effectively to guide design decision making activities, the latter relying heavily on designers' intuition and gut feelings. This is a main limitation in PSS design, since research has shown (e.g., Mitchell 1997) that even very experienced professionals are able to recognize only a subset of all possible patterns linking hardware design to its value. This challenge points to the need of developing more systematic and formal approaches to support decision making in early stage design.

2. OBJECTIVES AND RESEARCH DESIGN

The objective of this paper is to propose a methodology that exploits machine learning techniques to raise the cross-functional design team awareness of how the requirements of the hardware can influence its resale value in a POS business model. Figure 1 details the process followed by the authors to identify the most value-adding engineering characteristics of a road compaction equipment in second-hand markets

These findings originate from of applied research based on a single case study conducted in collaboration with a Swedish multinational Original Equipment Manufacturer (OEM) of mobile compactors for road surfaces. The machines considered in the analysis are double drum asphalt rollers that range between 1.5 tons to 25 tons in weight, and between 1000mm to 2100 mm in drum width. The process features 10 steps. Initially, data are scrapped and restructured from the OEM website (to retrieve information about the engineering characteristics of the machines) and from the dealer websites (gathering information, such as price, model and year of manufacturing for all machines on sales). Then the two datasets are merged, filling the missing values and preprocessing the data before applying the machine learning algorithm. Eventually, results are displayed to the cross-functional team and used to support design decisions concerning new machines.

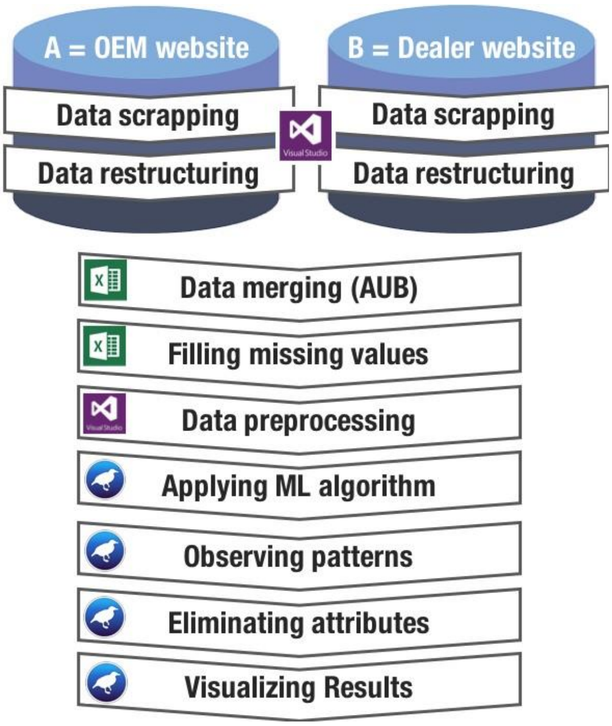


Fig. 1. Data mining methodology for resale value analysis

3. DATA MINING APPLICATIONS FOR EQUIPMENT DESIGN

Literature (e.g., Linoff and Berry 2011, Scherer et al. 2016) shows that data mining techniques can be successfully applied to determine the customer value for new products and services, identifying patterns that describes the nexus between customer satisfaction and business process, so to make more effective business decisions. Data mining (Witten et al. 2016) is the process of applying an algorithm to large datasets to identify unknown patterns that can be used as an optimization tool for industry (see also: Koh and Low 2004). Linoff and Berry (2011) describe the application of data mining in a business situation as a four-stage process: problem identification, transforming data into information, taking action and measure the outcome. The second stage ‘transform data into information’ is a vast field that falls under the domain of machine learning, and that touches upon issues from artificial intelligence research.

Data mining plays an important role in design decision making, and it is a gateway for both qualitative and quantitative models for design concept selection. Bertoni et al. (2017) recently presented a study related to the applicability of data mining algorithms as decision support in early design stages in the construction equipment sector. The authors present a scenario where machine usage data are fed back to the design stage and used as a basis to populate value models for decision making. Data mining is applied on a dataset built on machine performances and contextual and environmental data. The proposed demonstrator focuses on the estimation of the fuel consumption of alternative design concepts and estimates the performance variations given different contextual variable.

Auto Regression tree is another data mining algorithm applied by Fan et al. (2008) to predict the residual value of heavy construction equipment. The authors firstly shortlisted 50 top equipment models from different manufacturers. Then they selected different features of the machine and built a function to forecast their residual value. In this work, they treated price as a dependent attribute, while several features of the equipment were treated as a potential influencer. A research conducted by Perry et al. (1990) on farm tractors showed that manufacturing year and mileage are important decision criteria while purchasing second-hand equipment. In their work, the author showed that resale price featured a 0.75 correlation coefficient between age and hour in operation. Lee et al (2014) further applied a fuzzy logic-based algorithm to identify the value in patterned data.

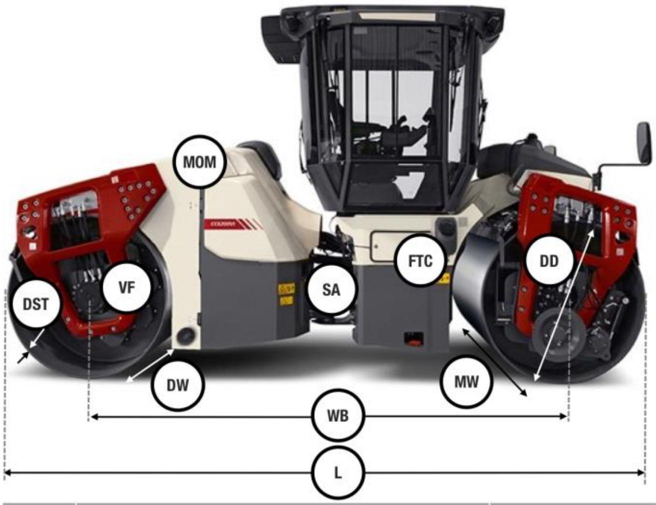
By analyzing the price of second-hand cars in Mauritius, Pudaruth (2014) provides one of the clearest examples of how to link machine features to resale value through the application of machine learning capabilities. In this work, the author identifies age as one of the most important drivers in developing a relationship between resale price and features of the vehicle. In the analysis, ‘price’ is treated as an independent variable, while four dependent attributes are ‘manufacturer’ (original equipment manufacturer OEM), ‘cylinder volume’, ‘year of manufacturing’ and ‘total mileage’ The author further applied different machine learning techniques, such as K-Nearest Neighbour (KNN), Naïve Bayes, Multiple Linear Regression Analysis and decision tree analysis, and compared their results in the resale value prediction exercise.

4. RESALE VALUE ANALYSIS FOR ROAD COMPACTION EQUIPMENT

4.1. Data scrapping and restructuring

The residual value of a road compaction equipment in second-hand markets is determined by the mix of its geometrical and topological properties. In the analysis, resale price was considered a dependent variable, while these properties were considered the independent variables. To get a better understanding of the effect of such characteristics on second-hand market price, it was decided to exclude manufacturing year and mileage from the analysis, normalizing the outcomes of the data scraping task. Figure 2 describes the geometrical and topological properties of the double drum compactor used in the analysis, with an indication of their minimum and maximum values. It also displays to which machine sub-system these properties relate to.

Machine data have been initially scrapped from the OEM website, and later imported into an MS Excel environment. A script in Visual Basic (VB) (Albright 2001) was developed by the authors to automate this task. The script was designed to fetch data about machine model, machine dimension, technical specifications, and more. As a result of data scrapping, 111 sheets, corresponding to 111 machines models, were obtained in the MS Excel workbook. In the following step, the authors created a unique master sheet to summarize the entire workbook. The same VB code was used to fetch data from dealer web pages for different machines models. The obtained workbook contains machine model name, cost, hours of operations, machine conditions, and more. In total, the scrapping activity generated more than 600 sheets in MS Excel workbook. Those machine models that were not featured in the OEM workbook (e.g., because the machine was simply too old) were eliminated from the Dealer workbook.



CODE	DESCRIPTION	RANGE
DD	Drum Diameter (the drum is the cylindrical feature in contact with the asphalt).	580–1573 mm
DW	Drum width (this is proportional to the width of the useful compaction area).	800-2130 mm
DST	Drum shell thickness (the drum it is subject to wear during operation).	11-50 mm
FTC	Fuel Tank Capacity (varies with engine and machine size).	23-320 litres
L	Machine length (total).	2040-6000 mm
MOM	Maximum Operating Mass.	1510-19000 kg
W	Machine width (total).	870-2400 mm
SA	Maximum steering angle.	30°-38°
VF	Vibration Frequency (of the drums).	29 -70 Hz
WB	Wheel base (distance between the axis of the front and the rear drum).	1350-3690 mm

Fig. 2. Independent variables for the resale value study.

Noticeably, several off-the-shelves software solutions for data scrapping were tested during the work, such as Octoparse, Mozenda, Parsehub, and more. However, hypertext in both the OEM and dealer's database was found to be rather unstructured and raw, hindering the direct implementation of data scrapping software. Due to the complexity of web structure and layout, they were not able to reliably automate the scrapping activity. Furthermore, both the OEM and Dealer workbooks featured unwanted and/or irrelevant information that needed to be eliminated, restructured and refined, which made it necessary to manually adjust the results of the scrapping activity. Also, due to the complexity of the website layout and structure (i.e., dimensions may be reported in plain text, tabular form, or figures in .png format), it was necessary to manually fill in the missing values.

4.2. Data merging

At this point, the authors started to pre-process the data, combining the OEM and dealer workbooks, so that the resulting dataset could be then processed by a data mining application.

The OEM workbook was named 'Set A' and contained data about machine model and features, i.e.

Set A (OEM workbook) = {Equipment Model, DW, DD, DST, W, L, WB, MOM, SA, FTC, VF}

The Dealer workbook was named 'Set B' and contained data about machine model and resale price.

Set B (Dealer workbook) = {Equipment Model, Resale price}

The combination of set A and B was performed using VB programming, using the Boolean operation U, which represents the union of OEM workbook and Dealer workbooks (see: Chakrabarti 2003).

AUB = {Equipment model, DW, DD, DST, W, L, WB, MOM, SA, FTC, VF, Resale price}

Once the combined master-sheet was generated, the authors had to intervene again to fill in the missing values manually. It was then possible to complete the pre-processing process, by assigning attributes to the data, such as 'numerical', 'nominal' or 'string', so that the data could be analyzed. Noticeably, the machine model name was termed as 'string' (nominal), because the regression analysis was performed only for the numeric values. Hence, the machine model name was neglected for the purpose of the analysis.

5. RESULTS

The resulting dataset was imported in the Weka machine learning open source software, (see: Frank, et al. 2016) to observe hidden patterns in data. The correlation between resale price and the other machine features was displayed using scatter plots, such as the one pictured in Figure 3 and showing the relationship between resale price vs. Vibration Frequency.

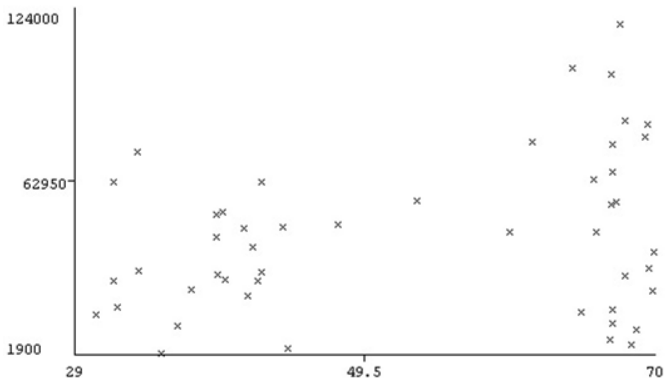


Fig. 3. Vibration Frequency (x-axis) vs. Resale price (y-axis).

The analysis shows a seemingly linear, but weak, correlation between this technical features and price. A similar weak and linear relationship with resale price was also observed for other engineering characteristics, such as Length (L), Fuel tank capacity (FTC) and Maximum operating mass (MOM). Other plots show a relationship that is even more diffuse and far less evident than the one pictured above. This is the case of the maximum steering angle (SA) dimension, as well as of drum shell thickness (DST). Interestingly, the scatter plot for DTS, (Figure 4) shows a pattern that was quite unexpected. The industrial experts during the descriptive study suggested this characteristic to likely produce a linear pattern with resale price (i.e., thicker the drum, higher the price). However, the plot reveals that this machine characteristic is not seemingly valued as foreseen by practitioners. This shall suggest decision makers that the price for second hand machines is not driven by considerations on how worn the drum shell is. At an operational level, this may suggest producing thinner drums, or to avoid using more expensive material in the construction, which is a material less subject to wearing.

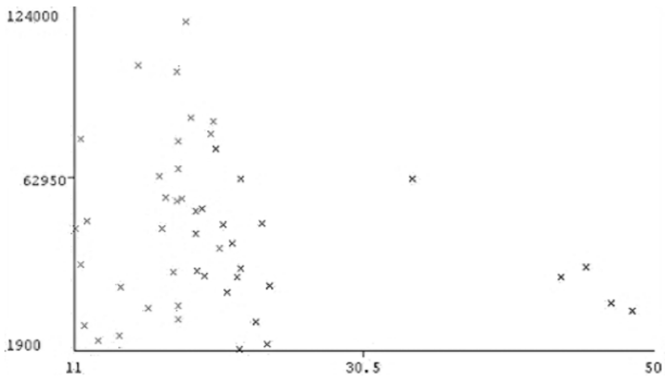


Fig. 4. Drum shell thickness (DSF, x-axis) vs. Resale price (y-axis).

These preliminary findings were followed up by a multiple regression analysis between resale price (dependent variable) and all independent variables to produce a more reliable prediction for which features are negatively correlated to the resale price. Multiple regression can be defined by (1), in which β terms are the linear parameters and ϵ is the error associated with the data-set.

$$y= \beta_{\circ}+ + \beta_2 x_2+^{\cdots}\ldots\ldots \beta_p x_p+ \epsilon \qquad (1)$$

Table 1 shows that steering angle (SA) and drum shell thickness (DST) are slightly negatively correlated with price (the latter is also evident in plot matrix in Figure 4). Wheelbase (WB), instead, features the highest positive correlation (0.55), and Drum width (DW) the second highest (0.39).

Table 1. Correlations between variables in the study

	Price	*WB*	*W*	*DD*	*L*	*DW*	*FTC*	*SA*	*MOM*	*VF*	*DST*
Price	1,00										
WB	0,55	1,00									
W	0,37	0,82	1,00								
DD	0,33	0,75	0,93	1,00							
L	0,33	0,79	0,94	0,98	1,00						
DW	0,39	0,84	1,00	0,93	0,93	1,00					
FTC	0,29	0,79	0,93	0,95	0,98	0,92	1,00				
SA	-0,30	-0,16	0,22	0,34	0,39	0,19	0,42	1,00			
MOM	0,34	0,81	0,95	0,92	0,94	0,95	0,94	0,27	1,00		
VF	0,22	0,21	-0,30	-0,46	-0,41	-0,27	-0,35	-0,71	-0,28	1,00	
DST	-0,09	0,22	0,62	0,70	0,67	0,59	0,65	0,57	0,71	-0,73	1,00

These analytic results (plot matrix, multiple regression analysis, and correlation coefficient table) were followed by the creation of a regression model with a reduced number of independent variables. Table 2 shows the results of such model, which further confirm that DW is the more valuable characteristic for second-hand market customers This is quite intuitive, considering that an equipment with a larger drum width (DW) has a larger surface working area, hence reducing operational time and associated cost.

Table 2. Regression analysis (only selected independent variables).

	Coefficients
Intercept	-13653,40
WB	16,50
W	-392,36
DD	52,70
L	-14,17
DW	438,68
MOM	0,07

6. DISCUSSION: USING DATA MINING RESULTS IN EARLY STAGE PSS DESIGN

Both the qualitative and quantitative approaches for value modeling require human decision makers to interpret and establish relationships between design properties and value. The results from the application of the proposed methodology are of great interest to populate both qualitative (i.e., the House of Quality, part of Quality Function Deployment) and quantitative (monetary) value models for early stage design concept selection.

6.1. Populating the House of Quality

Quality Function Deployment (QFD) is a well-established and widely adopted technique to identify customer needs and translates these into technical requirements. The House of Quality (HoQ) is the first as well as one of the most fundamental steps in QFD: here customer needs are listed, and then transformed into product features and functions, or design requirement. The numerical weights featured at each intersection of the matrix are often the results of a gut-feeling process, where patterns are recognized based on previous experience of the practitioners (Weiss 2013). A major issue with the HoQ is that professionals must often predict patterns of future relationships for radical new concepts, which are difficult to substantiate because historical data are lacking.

In the HoQ, the machine learning results can be used to validate the assumptions used to generate correlations between customer desires and the engineering characteristics which may be relevant to those desires. Also, they can help set targets for the engineering characteristics in the matrix, which are often set by design engineers, based on their knowledge and experience of the product type (Harding et al. 2001). In the presented example, while it is intuitive to relate the year of manufacturing to resale price, it is less intuitive to assess how the capacity of the fuel tank might affect this criterion. The results of the machine learning analysis can support decision-makers in associating the appropriate numeric weights to each intersection of the matrix, supporting the assumption that resale value features strong, moderate or weak correlation with the geometrical or functional properties.

6.2 Determining the value function for Multi-Criteria Decision Analysis

PSS design literature further shows that the relationship between customer satisfaction and machine properties features a nonlinear behavior. The 'Receiver State Parameters' approach proposed by Kimita et al. (2009) (Figure 5), for instance, is based on the use of an exponential function, similar to the one proposed by Xing et al. (2013) to measure the functional fitness of a PSS hardware in a time perspective. The CODA model (Eres et al. 2014) and the EVOKE model (Bertoni et al. 2018) are two QFD extensions use non-linear merit functions in PSS design to map the engineering characteristics of a product to its expected value contribution.

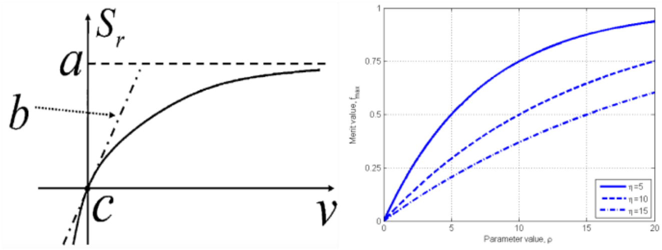


Fig. 5. Exponential value functions in Kimita et al. (2009) (left) and Bertoni et al. (2018) (right)

The main challenge when working with these more sophisticated mappings is that the nature and type of value functions is difficult to determine with precisions. Literature (e.g.: Gervásio and da Silva 2012) highlights the existence of several preference functions for multi-attribute decision making, from impulse to step functions, from linear to parabolic ones. Determining the nature and shapes of these curves is often an iterative task (Alarcon et al. 2010), which can be supported by the proposed data mining methodology, as shown in the resale value example.

Machine learning can be exploited to detail type and shape of the function that characterizes the relationship between customer value and a value-creating resource (such as a geometrical feature of the hardware). In a first iteration, machine learning algorithms can support engineers in evaluating tendencies (increase or decrease) in the value function. In the example described in Figure 3, the value function describes a (weak) increase relationship between vibration frequency (VF) and resale price. This information can be used, for instance, in the EVOKE model to determine the relationship function (i.e., maximization) at the intersection of these two dimensions.

In a second step, the retrieved information can be used to define the points controlling the overall shape of the value function, such as those corresponding to minimum, average or maximum satisfaction. Again, using EVOKE as an example, the plot shown in Figure 3 suggest setting the neutral point for the maximization function (i.e., the level to which a 50% satisfaction is obtained), at approximatively 60 Hz. In a third iteration, the results of the Weka analysis can help in defining the exact shape of value function (e.g., linear, concave, convex, S-shaped). Eventually, decision makers can use the identified patterns to deliberate with more confidence on the precise mathematical expression of the value function.

7. CONCLUSIONS

PSS design can be described as a game of making thoughtful trade-offs between technical aspects (costs, performances), business dimensions and consumer experience. A major gap in PSS design research is how to build models that, since an early design stage, can support decision-makers in predicting the future impact of hardware configurations on a number of value-related aspects. The paper shows that capabilities exist to exploit data-driven design approaches to support early stage design decisions in a PSS context. The presented case study illustrates that a data mining approach can be applied to support decision makers in gathering facts about how resale value is affected by changes in the characteristics of the product. By raising awareness on tacit and hidden patterns, engineers and designers can build arguments in favour (or against) certain technical solutions, and can deliberate on design trade-offs with more confidence.

Future work in the road compaction equipment industry will aim at increasing the number of dealers to get a larger data-set, to identify the patterns more accurately and precisely. Also, the authors will expand the number of geometrical and mechanical properties included in the model, to include other relevant characteristics, such as the size of the cabin, volume of the water tank and more. Furthermore, running a demonstrator on a single industrial context creates an intrinsic limitation in terms of multi-context validation of the results. Future work will also aim at growing lessons learned from the application of the proposed approach in heterogeneous case studies and industrial sectors.

ACKNOWLEDGMENTS

The research leading to these results has received financial support by the Swedish Knowledge and Competence Development Foundation (Stiftelsen för kunskaps- och kompetensutveckling) through the Model Driven Development and Decision Support research profile at Blekinge Institute of Technology.

REFERENCES

Alarcon, B., Aguado, A., Manga, R., and Josa, A. (2010). A value function for assessing sustainability: application to industrial buildings. Sustainability, 3(1), 35-50.

Albright, S.C. (2001). VBA for Modelers: Developing Decision Support Systems with Microsoft Excel. 4th edition.

Bartolomeo, M., Dal Maso, D., De Jong, P., Eder, P., Groenewegen, P., Hopkinson, P., James, P., Nijhus, L., Örniunge, M., Scholl, G., Slob, A. and Zaring, O. (2003). Eco-efficient producer services—what are they, how do they benefit customers and the environment and how likely are they to develop and be extensively utilised?. Journal of Cleaner Production, 11(8), 829-837.

Bertoni, M., Bertoni, A., and Isaksson, O. (2018). Evoke: A value-driven concept selection method for early system design. Journal of Systems Science and Systems Engineering, 27(1), 46-77.

Bertoni, A., Bertoni, M., Panarotto, M., Johansson, C., and Larsson, T.C. (2016). Value-driven product service systems development: Methods and industrial applications. CIRP Journal of Manufacturing Science and Technology, 15, 42-55.

Bertoni, A., Larsson, T., and Elfsberg, J. (2017). Mining data to design value: A demonstrator in early design. In ICED17, International Conference on Engineering Design, 64, 306-311.

Chakrabarti, S. (2003). Mining the web: Discovering knowledge from hypertext data. Morgan Kaufmann Publishers.

Eres, M.H., Bertoni, M., Kossmann, M., and Scanlan, J. (2014). Mapping customer needs to engineering characteristics: an aerospace perspective for conceptual design.7 Journal of Engineering Design, Vol. 25 No. 1-3, pp. 64-87.

Fan, H., AbouRizk, S., Kim., and Zaïane, O. (2008). Assessing Residual Value of Heavy Construction Equipment. Journal of Computing in Civil Engineering. ASCE, 22 (3), 181-191.

Frank, E., Hall, M.A., and Witten I.H (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques". Morgan Kaufmann, 4th Edition.

Gervásio, H., and da Silva, L.S. (2012). A probabilistic decision-making approach for the sustainable assessment of infrastructures. Expert Systems with Applications, 39(8), 7121-7131.

Harding, J.A., Popplewell, K., Fung, R.Y., and Omar, A.R. (2001). An intelligent information framework relating customer requirements and product characteristics. Computers in Industry, 44(1), 51-65.

Holmström, J., Ala-Risku, T., Auramo, J., Collin, J., Eloranta, E., and Salminen, A. (2010). Demand-supply chain representation: a tool for economic organizing of industrial services. Journal of Manufacturing Technology Management, 21(3), 376-387.

Kimita, K., Shimomura, Y., and Arai, T. (2009). A customer value model for sustainable service design. CIRP Journal of Manufacturing Science and Technology, 1(4), 254-261.

Koh, H.C. and Low, C.K. (2004). Going concern prediction using data mining techniques. Managerial Auditing Journal, 19, 462–476.

Lee, J., Kao, H., and Yang, S. (2014). Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. Part of the special issue: Product Services Systems and Value Creation. Procedia CIRP, 16, 3-8.

Linoff, G.S., and Berry, M.J.A. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley Publishing, Inc. Third edition.

Mitchell, T.M., (1997). Machine learning. Burr Ridge, IL: McGraw Hill.

Oraiopoulos, N., Ferguson, M.E. and Toktay, L.B. (2012). “Relicensing as a Secondary Market Strategy.” Management Science 58 (5): 1022–1037.

Perry, G. M., Bayaner, A., and Nixon, C.J. (1990). The effect of usage and size on tractor depreciation. American Journal of Agricultural Economics, 72(2), 317–325.

Pudaruth, S. (2014). Predicting the Price of Used Cars using Machine Learning Techniques. International Journal of Information and Computation Technology. International Research Publications House, 4, (7), 753-764.

Scherer, J.O., Kloeckner, A.P., Ribeiro, J.L.D., Pezzotta, G., and Pirola, F. (2016). Product-Service System (PSS) design: using Design Thinking and Business Analytics to improve PSS design. Procedia CIRP, 47, 341 – 346.

Tukker, A., and Tischner, U. (2006). Product-services as a research field: past, present and future. Reflections from a decade of research. Journal of cleaner production, 14(17), 1552-1556.

Van Loon, P., Delagarde, C., and Van Wassenhove, L.N. (2017). The role of second-hand markets in circular business: a simple model for leasing versus selling consumer products. International Journal of Production Research, 1-14.

Waldman, M. 2003. “Durable Goods Theory for Real World Markets.” Journal of Economic Perspectives 17 (1): 131–154.

Weiss, S.I. (2013). Product and systems development: a value approach. John Wiley and Sons.

Witten, I.H., Frank, E., Hall, M.A., and Pal, C.J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Xing, K., and L. Luong. 2009. ‘Modelling and Evaluation of Product Fitness for Service Life Extension.” Journal of Engineering Design 20(3): 243–263, (0954–4828).

Ölundh, G., and Ritzén, S. (2003). How do functional sales affect product development and environmental performance? In DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm.