MINING DATA TO DESIGN VALUE: A DEMONSTRATOR IN EARLY DESIGN

Bertoni, Alessandro (1); Larsson, Tobias (1); Larsson, Jonas (2); Elfsberg, Jenny (2)
1: Blekinge Institute of Technology, Sweden; 2: Volvo Construction Equipment, Sweden

Abstract

The paper presents a study run to verify the applicability of data mining algorithms as decision support in early design stages of a complex product development project. The paper describes a scenario built in two-stages providing the rationale for the application of data science in engineering design. Furthermore, it describes a demonstrator where usage data are fed back to the early design stage and used to populate value models to reduce the uncertainty in engineering design decision making. The development of a new machine for construction equipment, a wheel loader, is the subject of the demonstration and machine learning algorithms are applied on a dataset built on machine performances and contextual and environmental data. The demonstrator allows the estimation of the fuel consumption of different design concepts and the analysis of the performance variations given by a change in a contextual or environmental variable. Finally, the demonstrator allows the visualization of how much the tested performances of a new design deviate from the original designers' expectations.

Keywords: Early design phases, Case study, Data mining, Value Driven Design, Decision making

Contact:
Dr.-Ing. Alessandro Bertoni
Blekinge Institute of Technology
Mechanical Engineering
Sweden
alessandro.bertoni@bth.se

Please cite this paper as:
Surnames, Initials: *Title of paper*. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 7: Design Theory and Research Methodology, Vancouver, Canada, 21.-25.08.2017.

1 INTRODUCTION

Any development process is built upon a sequence of activities guiding the design team from the early stages of design to the product embodiment and delivery. The development of a successful product is often the result of a combination of right decisions taken by engineers in the context of high complexity and uncertainty. The human factor is still a main driver when choosing a design configuration, and engineers face the "design paradox" (Ullman, 1992) of making critical decisions in early design when information is fragmented and knowledge immature. Frameworks, methods and tools have been developed as an answer to the need of an enhanced awareness in early decision making, e.g. Quality Functional Deployment (Akao, 2004), Pugh Matrices (Pugh, 1981), Value Models (Hazelrigg, 1998; Collopy and Hollingsworth, 2011; Bertoni et al., 2016). Several authors have however also warned about the "hidden trap" given by their qualitative nature, driving decisions seldom supported by real data (Eres et al., 2014; Zhang et al., 2015).
To increase assessment reliability numerous computational models have been proposed in the recent years. However, those deal with a mix of uncertain information that is difficult to quantify numerically (Soban et al., 2011). The key challenge is that of lowering the uncertainty of the decision making by populating models with data driven information rather than experience-driven assumptions. On this line, the development of information communication technologies and of algorithms based on data mining (Anand and Buchner, 1998) and machine learning (Manyka et al., 2011) allows nowadays to create, manage, correlate and forecast a huge amount of data with relative low effort both in time and resources. The use of data mining techniques can support the tradeoff between different concepts, by allowing a dynamic indexing and retrieval of information to detect correlations and emerging trends (Braha, 2013). While the real-time analysis of product data is a practice already in use in many fields (e.g. (Pouliezos and Stavrakakis, 2013; Akhavian and Behzadan, 2013) the collection and analysis of data to derive design indications for new product development is still a poorly explored possibility. A few works have been recently published (see for instance Stockton et al., 2013; Dasari et al., 2015) however those do not focus on assessment models for early decision making. Here two main challenges need to be addressed: first the relevant subset of data representative of the product value in conceptual design needs to be identified, and second, it is unclear how to communicate the analysis result in a format immediately understandable by engineers, so to overcome what Freitas (2014) defines as "comprehensibility barrier" of data-science algorithms.
The research presented in this paper has first focused on what kind of data would be of interest, when those shall be collected, what kind of information can be generated through data-mining algorithms, and in what form the results can be best communicated to the engineers. Secondly a demonstrator has been developed in a reference scenario to test the applicability of the approach in a realistic use case. The paper has therefore the double purpose of investigating the potentiality and benefit off applying data mining in engineering design, and of promoting and disseminating such knowledge in an industrial context through the realization of a demonstrator.

2 METHODOLOGY

The research was performed through participatory action research (Whyte et al. 1989). Transcriptions and notes from interviews, internal conferences and formal and informal discussions were used to identify the industrial needs and have driven the definition of the scenario for the demonstrator. The development of the demonstrator was initiated based on an already available dataset of variables and measurements collected from machines operations in different working sites. The initial dataset encompassed machine performances, omitting variables considering contextual aspects (e.g. environmental conditions, operator expertise, conditions and topography of the site etc.) The full dataset was completed by artificially defining such "contextual" variables. The work of Cronholm (2013) was used as a reference for variables definition, and was further complemented by the variables emerged as relevant through interviews and discussions. To avoid the exposure of sensible information the data described in this paper have been artificially modified, however they conserve a reasonable degree of credibility to assure a realistic description.

3 DATA MINING AND MACHINE LEARNING IN ENGINEERING DESIGN

Data mining is defined as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large datasets (Anand and Buchner, 1998). Machine learning is instead defined as a sub-speciality of computer science concerning the design and development of algorithms that allow computers to evolve their behaviour based on empirical data (Manyka et al., 2011). The use of the two terms is often associated in literature when it comes to knowledge discovery from dataset in industrial settings. In recent years, the use of data mining in combination with the development of IT infrastructures, and of increased data collection and storage capabilities, has signed a profound shift toward more transparent, informed and autonomous decision making (Kusiak, 2006). Kusiak (2006) recognises for data mining techniques the quality of being able to fit the gap between tools used in decision-making and their linkage to data. His paper describes eight different examples of applications in manufacturing and service, spanning from process control, to production of semiconductors, to biotechnology and medical/pharmaceutical applications; although no examples of application of data mining for engineering design tasks in presented. As stated by Romanowski et al. (2006) engineering design is a multi-disciplinary, multi-dimensional, non-linear process, generating large amounts of heterogeneous data for which suitable mining methods are not readily available.

A small number of examples are available in literature concerning the application of data mining and machine learning in combination with engineering design methods. One of the first examples is the one described by Tseng and Hiao (1997) using conceptual clustering for the recognition of patterns of functional requirements. They claimed the approach to be able to integrate experts' knowledge from historical data projection and to enhance the ability to explore and utilize underlying domain knowledge more effectively. A later application by Vale and Shea (2003) focused on accelerating design synthesis by first using a "data modeler", observing and analyzing the effects of sequences of modifications on the design objectives, and then using a "modification advisor" capable of ranking potential imminent design modifications. A few researchers have focused on the use of machine learning to classify knowledge in engineering design. One of the first contribution on this topic can be identified in the work by McMahon et al. (2004) focusing on knowledge personalization and codification. Associations rules have been used in various occasions: to identify correlations between documents parts and assemblies (Woon et al. 2013), to develop person-oriented products and marketing solutions (Liao et al. 2008), to generate the recommendations on engineering changes based on an available database (Wickel and Lindemann 2015) and to create customers' segmentation based on requirements analysis (Agard and Kusiak 2004). Furthermore, historical product data have been mined in aerospace product development to extract information about process limitation and association between product dimensions (Choudhary et al. 2009). Finally, Apriori algorithms have been implemented to mine the customer knowledge to improve the development process and the customer relationship management (Liao et al. 2010), and a two-step cluster analysis have been applied to systematically analyse process interfaces based on their structural and compositional characteristics (Parraguez et al. 2016).

Among such contributions authors have shown an increased interest into investigating how to consistently integrate the use of data mining and machine learning into engineering models used in the very early stages of design. Pajo et al. (2015) have focused on the application of data mining in conceptual design by using automated classification of data from social media to identify potential lead users of new products. Quintana Amate et al. (2015) have described a case study on the design optimization of wing covers where the automated execution of machine learning methods is integrated in a knowledge based engineering implementation. Isaksson et al. (2015) have taken a step forward by proposing the integration of data mining techniques in a unique decision support system to evaluate value and sustainability of design concepts. Here machine learning techniques are seen as the enabling technology to create response surface models, so to allow engineers to evaluate, compare, and enhance design concepts by quick design space exploration and optimization. In the same work the use of machine learning is also proposed to fill the gap between qualitative expert assessment and early simulation results by extracting trends and relations from existing historical data and projection data, although no case related to such implementation is described.

Such challenge of generating more data-driven information is targeted by the work presented in this paper that also describes a demonstrator using data mining techniques for the design of construction equipment.

4 VALUE AND DATA DRIVEN DESIGN IN CONSTRUCTION EQUIPMENT: A DEMONSTRATOR

4.1 The reference scenario

The development of construction equipment (e.g. wheel loaders, articulated haulers, dump trucks and trucks) is a long and complex engineering process whose structure can be likened to a "traditional" engineering design or systems engineering process. Energy efficiency is a key performance objective for the industry. Such performance can be improved either by the more efficient usage of existing products or by developing new products capable of reducing energy consumption and recover energy (Cronholm, 2013). Both the improvement directions require good knowledge about the operating conditions of the machines. The machine behaviours, the fuel consumption, the durability of the components, and other product related performances are largely affected by the variability of the context in which the machine is operated; for such reason a machine with good fuel efficiency and performances in one type of operation does not necessary have the same behaviour in another type of operation (Cronholm, 2013).

A design team in charge of developing a machine needs to have a good knowledge about how the machine performs in different contexts and, nonetheless, the team needs to understand how a change in an engineering characteristic would impact the different usage performances.

The reference product for the development of the demonstrator was identified in a wheel loader. A wheel loader is a type of tractor provided with a bucket and arms, whose main function is to move material or dirt from a place to another lifting it, so that the material is not pushed across the ground. In terms of energy consumption, the main sources of power requested by a wheel loader are: the power needed to lift the material with the bucket, the power needed to move the loader from point A to point B and the power needed for cabin comfort and other power usage. The reference scenario is summarized in Figure 1. The design team need to trade-off two alternative design concepts, i.e. Concept A and Concept B, to decide which of the two to further develop. Engineers at this stage need to make decisions while dealing with a number of trade-offs between product performances that suffer from high variability given by the context in which the future product will be used. While traditionally experience-based assessment would be the main source of information for decision making models, the scenario integrates the use of data analysis from current

available machine data to explore relationships between operational context and machine performances. In this scenario, experienced-based assessment is not replaced, rather complemented, by the forecast derived from the real data. Both such assessments work as input into a model supporting decision making (i.e. a value model), whose role is to increase decision makers' awareness about estimated performances and system level effects.

The next section (4.2) describes the details of the definition of data and data mining approach applied in the demonstrator, while the integration of quantitative and qualitative data in the value model is not described in this paper.

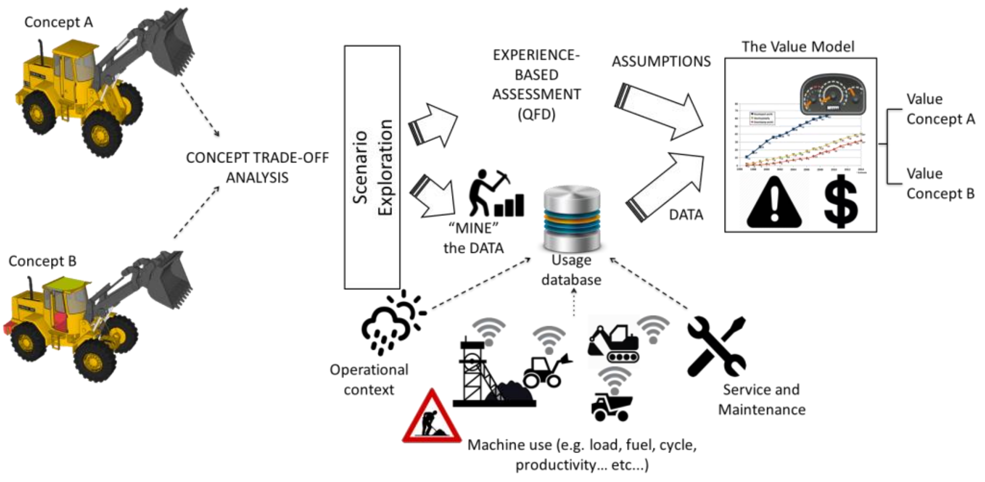

Figure 1. The reference Scenario for the demonstrator

4.2 Application of data driven approach

The demonstrator of the data driven approach has been developed to address one of the main challenges of the construction equipment industry, i.e. the energy efficiency. To address the need of both understanding how the current machines perform in different contexts, and the impact that a new design would have in comparison to a previous configuration, the demonstrator is suited for two different application stages.

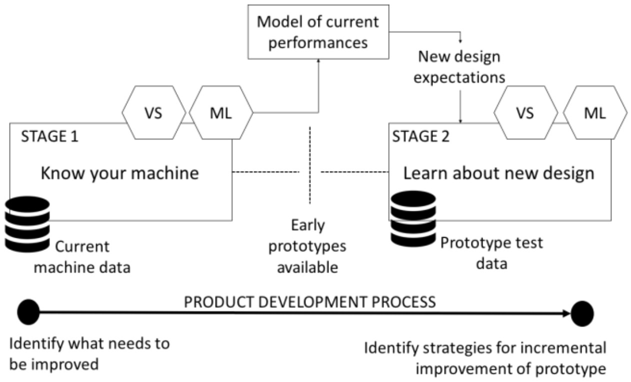

Figure 2. The two stages of the demonstrator and their use in the product development process. ML = Machine Learning; VS = visualization support

In a first step the approach is used in the stage where engineers need to increase their knowledge about how the current machine is performing in different operational context, this is done to be able to derive usage guidelines, and give customized indications to customers, but also to understand if there is room for improvement of specific machine features under defined operating conditions. Here data mining algorithms are applied on a set of data collected from real time monitoring of machines and environment, to mathematically represent machine performances under different operating variables. This first stage is defined in the demonstrator as the “Know your machine” stage. The second stage builds on the knowledge created in the first stage that is then applied to compare the performances of a new design once the first available prototypes are ready for testing. Initially engineers propose solutions to improve the current machine; the quantification of such improvement is based mostly on qualitative assessment since no real product has been created yet. It is only when the first prototypes are available that real tests data about machine performances become available and data are compared. Such data represent a key source of information for the further development and continuous improvement of the prototypes, but they need to be structured and analysed so to define the correlations and impacts that the contextual variables played on the prototype performances. Engineers then compare the results with the original qualitative expectations, and derive design indications. This second application stage is defined in the demonstrator as the “Learn about new design” stage. Figure 2 summarizes the logic of the demonstrator highlighting the two stages. A detailed description of the logic and of the algorithms applied is proposed in sections 4.2.1 and 4.2.2.

4.2.1 Stage 1: Know your machine

The dataset used for the demonstrator embedded the measurement of variables in 234 occasions (i.e. 234 instances in the database). In total 17 variables were considered as a starting point for the analysis. Part of the variables were collected during the normal machine operation, others were manually added to recreate a consistent dataset for demonstration purpose. The work by Cronholm (2013) had previously identified and ranked a list of variables with high impact on fuel consumption in relation to the performance of an articulated hauler in construction equipment. The ten most relevant variables were chosen and adapted to the context of a wheel loader. Eight out of ten variables were finally selected and complemented by 9 variables specifically identified for the case. In summary, the numerical variables considered were: cycle time, productivity, load, fuel consumption, distance, Max_dB_out, Avg_dB_out, Max_dB_in, Avg_dB_in. The discrete variables considered were: Operation (load_carry OR short_cycle OR Shot_rock), Road type (smooth OR rough OR very rough OR cross country) Ground resistance (very low OR low OR medium OR high OR very high), Driver experience (beginner OR expert OR master), Gross vehicle weight (nominal OR overload), Speed changes (low OR high OR very high), Topography (flat OR predominantly flat OR hilly OR very hilly) and curve density (low OR

medium OR high). An open source data mining and machine learning software named Weka (Hall et al. 2009) was used to run machine learning algorithms, while the input data and the output data were summarized, collected and visualized in a Microsoft Excel spreadsheet.
Given the focus on energy efficiency the fuel consumption was selected as the variable to be estimated, i.e. the "fuel consumption" was the class selected in the machine learning algorithms. The dataset was first imported into Weka and later pre-processed to exclude variables intuitively not related to fuel consumption. Since the focus of the analysis was on the assessment of the impact of contextual variables on fuel consumption, the "non-contextual" variables, i.e. productivity, load, distance and cycle time, were excluded and considered as constant. Additionally, the 4 variables related to the number of decibels produced were excluded as considered not indicative of fuel consumption variation.
To the remaining variables, a machine learning algorithm was applied for the classification of fuel consumption. Three classification algorithms were compared: multiple linear regression, K-nearest neighbour and Random Forest. A 10-folds cross-validation was applied. As shown in Table 1 the algorithm rendering the highest accuracy was the multiple linear regression that was therefore selected for the analysis of the data. Regarding the selection of the algorithm it must be highlighted that no further detailed analysis was run in the frame of the work. Since the goal of the work was that of demonstrating the applicability of the approach in an engineering design context, no further analysis of the algorithm best fitting the set of data was run more than the one presented in the table. Such choice, despite not being the optimal mathematically, was driven by the need of keeping the demonstrator to a low level of complexity to avoid a communication barrier with the engineers involved in the design.

Table 1. Machine learning algorithms tested, accuracy and errors.

Algorithm	Accuracy	Mean absolute error	Relative absolute error
Multiple linear regression	0.9792	1058	16.1%
K-nearest neighbour	0.8894	2091	31.9%
Random forest	0.9756	1334	20.6%

As a result of the multiple linear regression a function was derived estimating the fuel consumption of the machine given a combination of contextual variables. In detail: road type, ground resistance, operation, driver experience, speed changes and topography emerged as variables impacting with different extent the final fuel consumption, while the gross vehicle weight resulted to be not relevant in the calculation.

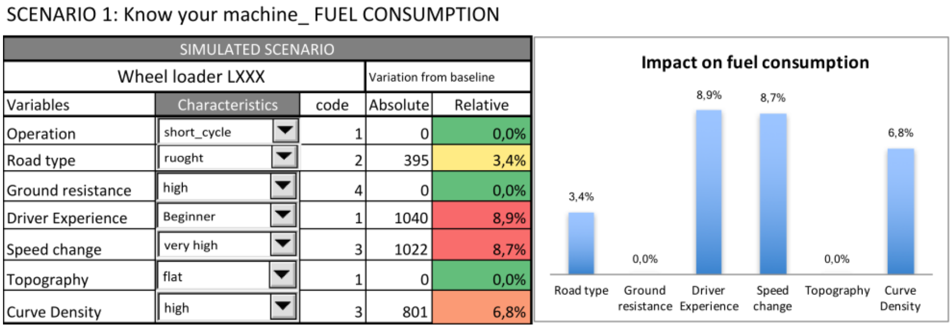

Figure 3. Screenshot from stage 1 analysis (notice: the numbers are not indicative of reality and black cells correspond to omitted values)

The obtained function was imported in an excel-based model visualizing the impact of contextual variables in respect to a baseline scenario corresponding to the ideal operating condition of the machine. Figure 3 is a screenshot of the visualization environment in which engineers could play with the available data and increase their knowledge about the behaviour of the machine in different contexts. For instance, (as from the example presented in Figure 3), the engineers can study development directions by visualizing the additional amount of fuel that is consumed on average when a beginner is driving a machine rather than an expert driver, or in case a machine is operating with continuous speed changes, or again in a scenario with high curves density.

4.2.2 Stage 2: Learn about a new design

Stage 2 of the demonstrator concerned the application of the data driven logic to the fictitious development of a new wheel loader featuring a hybrid engine. The dataset used in this stage concerned data collected from the testing activities of the early prototypes of the hybrid machine. Also in this case the same type of variables collected for 234 observations were considered, no real data were used for this stage of the demonstrator.

The goal of stage 2 was to provide indications for improving the performances of the early prototypes tested. The first step of the process was about defining what were the expected performance improvements given by the hybrid solution in respect to the machine considered during stage 1. For instance, the shift toward a hybrid engine was considered to be beneficial toward eliminating the difference in fuel consumption between expert drivers and beginners. This was because a hybrid machine was considered relatively easier to drive compared to a pure diesel machine. In the second step the expected improvements were compared with the data obtained from the application of machine learning algorithms on the prototype test data. From this a multiple linear regression model was derived following the same logic of stage 1, obtaining a function expressing the estimated fuel consumption of the hybrid machine in different operating context. Finally, the expected data from the qualitative assessment were compared with the machine learning output and the results were visualized in a common interface on a Microsoft Excel spreadsheet. Figure 4 shows the screenshot of a portion of the visualization interface, highlighting the deviation of the testing data from the expected results. For instance, as shown in Figure 4, the model highlighted that different performances in fuel consumption between beginners and expert drivers were still present, in disagreement with the initial expectation. Such kind of information is meant to be used by engineers to assess expected and unexpected benefits and drawbacks of the new product, and as a knowledge base to drive design modifications.

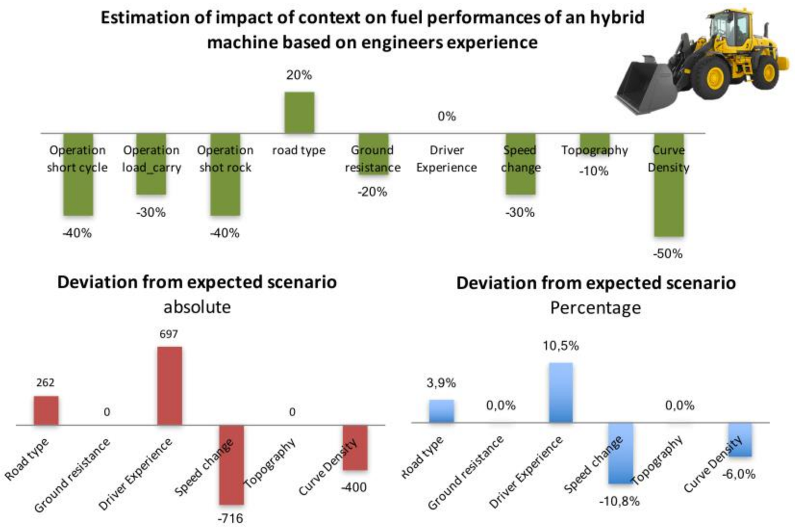

Figure 4. Screenshot from Stage 2 visualization interface: deviation of performances of the tested prototype compared to initial expectations.

5 CONCLUDING REMARKS

The use of data mining and machine learning in the early stages of engineering design is still a poorly explored field. The work presented in this paper does not claim to be an extensive example, rather it is the result of an investigative effort that had both the goal of demonstrating the potential of applying data science in an engineering design context, and of promoting and disseminating such knowledge in an industrial environment. The paper presented a scenario where machine usage data are fed back to the design stage and used as basis to populate value models for decision making. This is described as the ideal scenario providing the rationale for the application of data science in engineering design. The demonstrator presented in the paper focuses on a partial aspect of the scenario: the one related to the use of data mining technique to generate quantitative performance model. How quantitative data are merged and used in value models together with qualitative assessment has not been described in the paper. The demonstrator has been developed in collaboration with a company developing and producing construction equipment. The industry was particularly suitable due to the availability of product-related data: machine monitoring and data collection and analysis are an established reality in many business to business situation in the construction equipment sector. The establishment of a standardised approach to collect and analyse machine data in relation to the operational environment is believed to be the way toward the definition of a consistent knowledge-base to allow the creation of generic rules to be later generalised in reliable simulation models.

Running a demonstrator on a single industrial context creates an intrinsic limitation in terms of multi-context validation of the results, although the process and methods used for data analysis and visualization are widely adopted in several research contexts, and no ah-hoc solutions were developed to specifically adapt to the construction equipment sector.

The demonstrator has used basic machine learning algorithms without further exploring the possibility to define specific algorithms to increase model accuracy. The use of partially fake data has rendered results that, despite being dimensionally credible, are not representative of real machine performances. This was necessary for not revealing company private information and did not contrast with the investigative purpose of the research. From a wider research perspective, the validation of the value and data driven approach is currently limited to qualitative feedback from industrial practitioners highlighting the benefit of an increased knowledge-base on machine performances, together with the possibility of accessing data analysis in a clear and intuitive way. Further validation work is currently running to move from a demonstration stage to a pilot study to verify the applicability and benefits in a larger scale project. Further research is needed to verify the benefit in terms of knowledge creation and reduction of uncertainty, and also in studying the wider impact of the new approach in the engineering design process.

REFERENCES

Agard, B. and Kusiak, A., (2004), “Data-mining-based methodology for the design of product families”, *International Journal of Production Research*, 42(15), pp.2955-2969.

Akao, Y. (2004), *Quality function deployment – Integrating customer requirements into product design*, ISBN-13: 978-1563273131

Akhavian, R. and Behzadan, A.H., (2013), “Knowledge-based simulation modeling of construction fleet operations using multimodal-process data mining”, *Journal of Construction Engineering and Management*, 139(11), p.04013021.

Anand, S. S., & Büchner, A. G. (1998), “Decision support using data mining”, *Financial Times Management*.

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*.

Braha, D. ed., (2013), *Data mining for design and manufacturing: methods and applications (Vol. 3)*, Springer Science & Business Media.

Choudhary, A.K., Harding, J.A. and Tiwari, M.K., (2009), “Data mining in manufacturing: a review based on the kind of knowledge”, *Journal of Intelligent Manufacturing*, 20(5), pp.501-521

Collopy, P.D. and Hollingsworth, P.M., (2011), “Value-driven design”, *Journal of Aircraft*, 48(3), pp.749-759.

Cronholm, K., (2013), “Design of experiment based on VMEA (Variation Mode and Effect Analysis)”, In: *Procedia Engineering 5th International Conference on Fatigue Design, Fatigue Design 2013*, Vol. 66, pp. 369-382.

Dasari, S.K., Lavesson, N., Andersson, P. and Persson, M., (2015), “Tree-Based Response Surface Analysis”, In: *International Workshop on Machine Learning, Optimization and Big Data*, pp. 118-129, Springer International Publishing.

Eres, M.H., Bertoni, M., Kossmann, M. and Scanlan, J., (2014), “Mapping customer needs to engineering characteristics: an aerospace perspective for conceptual design”, *Journal of Engineering Design*, 25(1-3), pp.64-87.

Freitas, A.A., (2014), “Comprehensible classification models: a position paper”, *ACM SIGKDD explorations newsletter*, 15(1), pp.1-10.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H., (2009), “The WEKA data mining software: an update”, *ACM SIGKDD explorations newsletter*, 11(1), pp.10-18.

Hazelrigg, G.A., (1998), “A framework for decision-based engineering design”, *Journal of mechanical design*, 120(4), pp.653-658.

Isaksson, O., Bertoni, M., Hallstedt, S. and Lavesson, N., (2015), “Model Based Decision Support for Value and Sustainability in Product Development”, In: *DS 80-1 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 1: Design for Life*, Milan, Italy, 27-30.07.15

Kusiak, A., (2006), “Data mining: manufacturing and service applications”, *International Journal of Production Research*, 44(18-19), pp.4175-4191.
Liao, S.H., Chen, Y.J. and Deng, M.Y., (2010), “Mining customer knowledge for tourism new product development and customer relationship management”, *Expert Systems with Applications*, 37(6), pp.4212-4223.
Liao, S.H., Hsieh, C.L. and Huang, S.P., (2008), “Mining product maps for new product development”, *Expert Systems with Applications*, 34(1), pp.50-62.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H., (2011) *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute
McMahon, C., Lowe, A. and Culley, S., (2004), “Knowledge management in engineering design: personalization and codification”, *Journal of Engineering Design*, 15(4), pp.307-325.
Pajo, S., Vandevenne, D. and Duflou, J.R., (2015), “Systematic Online Lead User Identification-Case Study For Electrical Installations”, In: *DS 80-10 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 10: Design Information and Knowledge Management Milan*, Italy, 27-30.07. 15.
Parraguez, P., Eppinger, S. and Maier, A., (2016), “Characterizing Design Process Interfaces as Organization Networks: Insights for Engineering Systems Management”, *Systems Engineering*, 19(2), pp.158-173.
Pouliezos, A. and Stavrakakis, G.S., (2013), *Real time fault monitoring of industrial processes (Vol. 12)*, Springer Science & Business Media.
Pugh, S., (1981), “Concept Selection-a method that works”, *Proceedings of the International Conference on Engineering Design*, Rome, p. 497-506
Quintana-Amate, S., Bermell-Garcia, P., Balcazar, L. and Tiwari, A., (2015), “A new knowledge sourcing framework to support KBE development”, In: *DS 80-8 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 8: Innovation and Creativity*, Milan, Italy, 27-30.07. 15.
Romanowski, C.J., Nagi, R. and Sudit, M., (2006), “Data mining in an engineering design environment: OR applications from graph matching”, *Computers & operations research*, 33(11), pp.3150-3160.
Soban, D.S., Price, M.A. and Hollingsworth, P., (2012), “Defining a research agenda in Value Driven Design: Questions that need to be asked”, *Journal of Aerospace Operations*, 1(4), pp.329-342.
Stockton, D.J., Khalil, R.A. and Mukhongo, M.L., (2013), “Cost model development using virtual manufacturing and data mining: part I—methodology development”, *The International Journal of Advanced Manufacturing Technology*, 66(5-8), pp.741-749.
Tseng, M.M. and Jiao, J., (1997), “A variant approach to product definition by recognizing functional requirement patterns”, *Journal of Engineering Design*, 8(4), pp.329-340.
Ullman, D.G., (1992), *The mechanical design process (Vol. 2)*, New York: McGraw-Hill.
Vale, C.A. and Shea, K., (2003), “A machine learning-based approach to accelerating computational design synthesis”, In: *DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design*, Stockholm.
Whyte, W.F., Greenwood, D.J. and Lazes, P., (1989), “Participatory Action Research: Through Practice to Science in Social Research". *The American behavioural scientist*, 32(5), p.513.
Wickel, M.C. and Lindemann, U., (2015), “How to integrate information about past engineering changes in new change process?”, In: *DS 80-3 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 3: Organisation and Management*, Milan, Italy, 27-30.07. 15.
Woon, Y.K., Ng, W.K., Li, X. and Lu, W.F., (2003), “Efficient web log mining for product development”, In: *Cyberworlds*, 2003. pp. 294-301.
Zhang, X., Tong, S., Eres, H., Wang, K. and Kossmann, M., (2015), “Towards avoiding the hidden traps in QFD during requirements establishment”, *Journal of Systems Science and Systems Engineering*, 24(3), pp.316-336.

ACKNOWLEDGMENTS

The research leading to these results has received financial support by the Swedish Knowledge and Competence Development Foundation through the Model Driven Development and Decision Support research profile at Blekinge Institute of Technology.