Proceedings of the ASME 2016 International Manufacturing Science and Engineering Conference
MSEC2016
June 27-July 1, 2016, Blacksburg, Virginia, USA

MSEC2016-8625

Forecasting Obsolescence Risk using Machine Learning

Connor Jennings, Dazhong Wu, Janis Terpenny

Center for e-Design
Industrial and Manufacturing Systems Engineering
Pennsylvania State University
State College, Pennsylvania, 16801, USA

ABSTRACT

With rapid innovation in the electronics industry, product obsolescence forecasting has become increasingly important. More accurate obsolescence forecasting would have cost reduction effects in product design and part procurement over a product's lifetime. Currently many obsolescence forecasting methods require manual input or perform market analysis on a part by part basis; practices that are not feasible for large bill of materials. In response, this paper introduces an obsolescence forecasting framework that is capable of being scaled to meet industry needs while remaining highly accurate. The framework utilizes machine learning to classify parts as active, in production, or obsolete and discontinued. This classification and labeling of parts can be useful in the design stage in part selection and during inventory management with evaluating the chance that suppliers might stop production. A case study utilizing the proposed framework is presented to demonstrate and validate the improved accuracy of obsolescence risk forecasting. As shown, the framework correctly identified active and obsolete products with an accuracy as high as 98.3%.

1 INTRODUCTION

As many products and systems continue to improve at an ever increasing rate, obsolescence becomes an accelerating problem by volume as well as in costs. Currently, 3% of the world's electronic components become obsolete every month [1]. The flood of electronics components into traditionally non-electronics products further exacerbates the problem of product obsolescence. Further, today's approaches to dealing with obsolescence are predominately reactionary rather than being strategically planned or implemented. Such approaches have led to an escalation of non-value added tasks and associated high costs of averting and resolving problems.

Currently proactive obsolescence forecasting methods are hindered by their inability to scale to meet larger needs of industry. This paper describes a new approach that seeks to solve this problem, leveraging the advantages of machine learning and provide estimates of part status as discontinued or actively in production. This industry friendly approach to predicting product obsolescence risk level will provide valuable insights for proactive obsolescence management.

The remainder of this paper is organized as follows: In section 2, a brief overview of how obsolescence is handled in industry is presented, including: (1) current obsolescence risk forecasting methods, and (2) difficulties experienced in industry. In section 3, a brief overview of machine learning is presented. In section 4, Obsolescence Risk Forecasting using Machine Learning (ORML) method is presented. Section 5 provides a case study of ORML that is used to predict obsolescence in the cell phone market. Section 6 provides conclusions that include a discussion of research contribution and future work

2 OBSOLESCENCE

Obsolescence can have an immensely negative effect on many industries; the ramifications of which have generated a large body of research around obsolescence related decision making and more generally studying products through the product life cycle. Obsolescence can be broken into three main categories: technical, functional, and style obsolescence. Technical obsolescence refers to parts that are rendered obsolete by technically superior products entering the market. Functional obsolescence is caused when an organization suspends support of a product and the product then is render obsolete due to the lack of functional parts. An example of this would be a manufacturer of printers discontinuing the cartridges compatible with a particular printer model. Even though that model remains functional and capable of printing, the lack of new ink cartridges renders the printer functionally obsolete. And finally, style obsolescence, occurs largely due to shifts in preference of how products look to the consumer. Examples of this would be the shift in computer case color from beige to black, updates to logos, or curvature changes in the case that may appear more sleek or modern.

Throughout the life span of a product, maintaining and managing obsolescence can be a costly undertaking. In a 2006 Department of Defense report, the cost of obsolescence and obsolescence mitigation for the U.S. government was estimated at $10 billion annually [2]. The excessive cost is largely due to limited reaction time when obsolescence occurs which leads to costlier reactionary solutions as opposed to a more proactive management system. Some examples of short term obsolescence alleviation include lifetime buy, last-time buy, aftermarket sources, identification of alternative or substitute parts, emulated parts, and salvaged parts [3, 4]. More sustainable long term solutions include redesign and design-refresh.

Obsolescence management strategies are important during the design stage and throughout the product life cycle. With 80-85% of life cycle costs resulting from decisions made during the design stage, the ability to forecast product obsolescence of components is important to reductions of cost from design through production [5]. Methods and tools that help designers and supply chain managers understand risks and make informed choices associated with design alternatives and evaluate substitutable parts are sorely needed to the proactive planning and management of obsolescence.

2.1 Obsolescence Forecasting

Current methods for forecasting obsolescence fall into two broad categories: obsolescence risk and life cycle. Obsolescence risk methods calculate a risk index or probability a part will become obsolete [5–8]. Life cycle methods estimate part life span in an attempt to predict the date when the part reaches end of life [9–13]. Both the risk levels and end date estimations have unique applications and if accurate, are extremely helpful to industry.

Currently two simple models have been established in the literature for forecasting obsolescence risk level; both use high, medium and low rating of different key factors related to obsolescence risk [5, 6, 8]. Rojo conducted a survey of organizations' obsolescence manager and compiled a best practice list of factors for forecasting obsolescence risk. The key obsolescence factors are numbers of manufacturers, years to end of life, stock available vs. consumption rate, and operational impact criticality as key indicator for potential parts with high obsolescence risk [9]. Josias and Terpenny also created a risk index to measure obsolescence risk levels with the key metrics of manufacturers' market share, number of manufacturers, life cycle stage, and company's risk level [6]. Josias and Terpenny's method allows for weights to be changed between each factor to more approximately fit the model to different industries [6]. Unlike Rojo and other obsolescence risk forecasting methods, Josias and Terpenny's method uses a numerical index and not a percentage to measure the risk level. This allows only for comparison between parts on a bill of material (BOM) and not as easily between separate bills of materials. Another approach introduced by van Jaarsveld uses demand data to estimate the risk of obsolescence. The method manually groups similar parts and watches the demand over time [8]. A formula is given to measure how a drop in demand increases the risk of obsolescence [8]. However, this method cannot predict very far into the future because it does not attempt to forecast out demand, which causes the obsolescence risk to be more reactive and proactive [8].

2.2 Industry Obsolescence Forecasting Adoption

Although proactive obsolescence mitigation is a much more cost effective solution and obsolescence forecasting methods exist, companies and governments still largely do not deploy obsolescence forecasting in their obsolescence management strategies. Currently many organizations use web services that aggregate information from manufacturer's and supplier's websites to determine which products are currently in production or discontinued [15].

The reason for the lack of market adoption of obsolescence forecasting method is largely due to the insufficient ability for current methods to scale to industry needs. For a method to be scalable, the method must have the ability to adjust the capacity of predictions with minimal cost in minimal time over a large capacity range [16]. Currently, obsolescence risk forecasting methods have three key attribute that prevent scalability. First, many methods require the continual collection and updating of part's sales data. Second, systems requiring manual human inputs or a human interpretation of a market are costly due to the time and man-power required to maintain the system. The introduction of human perception into a prediction model causes immense bias in the model and predictions will vary widely depending on the operator. Third, many forecasting methods use growth over time of a single part specification, like memory, to predict when a product will become obsolete. The method works well for products that can be distinguished with one feature, like memory for the flash memory market, but with more complex products like a cellphone the method can not be implemented [10, 11]. In Table 1, current obsolescence risk forecasting methodologies from the literature are compared against the three key attribute necessary to scale to industry needs. All three methods are able to handle multi-feature products, but all methods require some form of human input either through asking a market expert for their opinions or manually manipulating the bill of material to remove parts the operator feels has a low obsolescence risk.

Table 1: Risk Level Methods and Scalability Factors

Methods	Sales Data Required	Human Inputs	Multi-Feature Capable
Josias et al. (2009)	-	✓	✓
van Jaarsveld et al. (2010)	✓	✓*	✓
Rojo et al. (2012)	-	✓*	✓

Notes: *Human bias due to manually filtering the BOM

3 MACHINE LEARNING

Machine learning is a method of pattern recognition in data analysis. It can be used to cluster instances and to predict an output. Machine learning algorithms build predictive models using data sets. These data sets can be continuously added to and will often improve the prediction model as the model "learns" from new instances. For obsolescence forecasting, the ability for a data set to be continuously updated is an important attribute for an effective predictive model. As parts become obsolete and new products are introduced, the data can be updated and the model will "learn" and adjust automatically.

Machine learning has grown to be a prominent tool in data analysis. The application of machine learning ranges widely from recommendations systems on Netflix and Amazon to facial recognition in pictures to cancer prediction and prognosis [16–18]. Machine learning has also been applied in the design field to help designers to more effectively gather information and feedback from previously under utilize sources. An example of this is using data mining to gather public reviews of products to better understand how consumer use and feel about individual features in products [20]. Additionally, social media can be mined to gather similar feedback about products and even predict new features consumers desire [21].

Researchers in France, applied machine learning to improve the search results for articles on a newspaper's website. The researcher used machine learning to show how manually present search grouping (tags) could become obsolete over time [22]. For example, the abbreviation CDC could stand for 'Caisse des Dépôts et Consignations' but if there is a disease outbreak, CDC as a search term on a news website could change meanings to 'Center of Disease Control' [22]. The machine learning approach was able to notify the newspapers of changing trends and identify potential tags that need updating. An overlooked contribution of the study was the first and only application of machine learning to predict obsolescence. However, the application area is rather unique and a generalizable obsolescence forecasting form was not provided.

Currently in the area of product obsolescence forecasting, creating automatic and scalable prediction models is one of the largest hindrances too large scale industry adoption. Machine learning is commonly utilized for creating nimble prediction models capable of rapidly updated with changing data too large for an individual. The following sections introduces the combination of these application areas to create a methodology to improve obsolescence forecasting models.

4 OBSOLESCENCE RISK FORECASTING WITH MACHINE LEARNING

This section will serve as an introduction to the concept of machine learning and a basic overview of how it works. Then will introduce the obsolescence risk forecasting using machine learning (ORML) methodology. Lastly, potential outputs will be discussed.

Machine learning can be broken into two main groups: unsupervised and supervised. Supervised learning develops prediction models that are capable of predicting a label. These models are created from using data with known labels. An example of this is over time, as users mark more and more emails as spam and mark other emails as important, email clients predict the label of an incoming email based on past email attributes like certain word counts. Unsupervised learning does not have a label output and instead clusters similar data points together. A common use case of unsupervised learning is clustering voters in an election. Voters with similar views and wants are grouped together then each individual group is analyzed and candidates can target their campaigns at certain groups that they wish to gain support from during the election.

For this method, supervised learning will be employed to predict the label of active or discontinued. The attributes that will be used as inputs are the technical specifications of each product.

The ORML process can be seen in Figure 1. First, components with known active or obsolete labels and their specifications are fed into the machine learning algorithm and the algorithm generates a prediction model. Components with unknown labels have their specifications inputted into the prediction model. The model then outputs the label with the highest probability for each of the unknown components.

The use of technical specifications as the input for the prediction model allows for the model to calculate how individual features indicate if a product is obsolete or is still actively in production. Because the model is created using known historical data, the model benchmarks unknown components against historical trends in the market. As the historical data grows with time, the machine learning algorithm will update the model and the relationships between certain features and the chance a component is obsolete. This aspect of a machine learning based approach to obsolescence forecasting makes ORML one of the most powerful and easily maintained methods.

5 CASE STUDY IN THE CELLPHONE MARKET

The case study will demonstrate the accuracy and scalability of ORML as a method to forecast obsolescence. The data contains specification information for 7000 unique models of cellphones and whether the phone is in production or discontinued. The specifications include weight (g.), screen size (inch.), screen resolution (pixels), talk time on one battery (min.), primary and secondary camera size (Megapixels), type of web browser, and if the phone has the following: 3.5 mm headphone jack, Bluetooth, email, push email, radio, SMS, MMS, thread text messaging, GPS, vibration alerts, or a physical keyboard.

The data was web scraped from GSM Arena, a popular online forum for comparing cellphones and can be downloaded at connor.github.io/research.html. The forum data was all user submitted so there is missing values and even misreported information. The errors and lack of information reflects some industries limitations on finding complete data sets. Even with these short falls with the data set, the machine learning algorithms can still create accurate obsolescence risk prediction models.

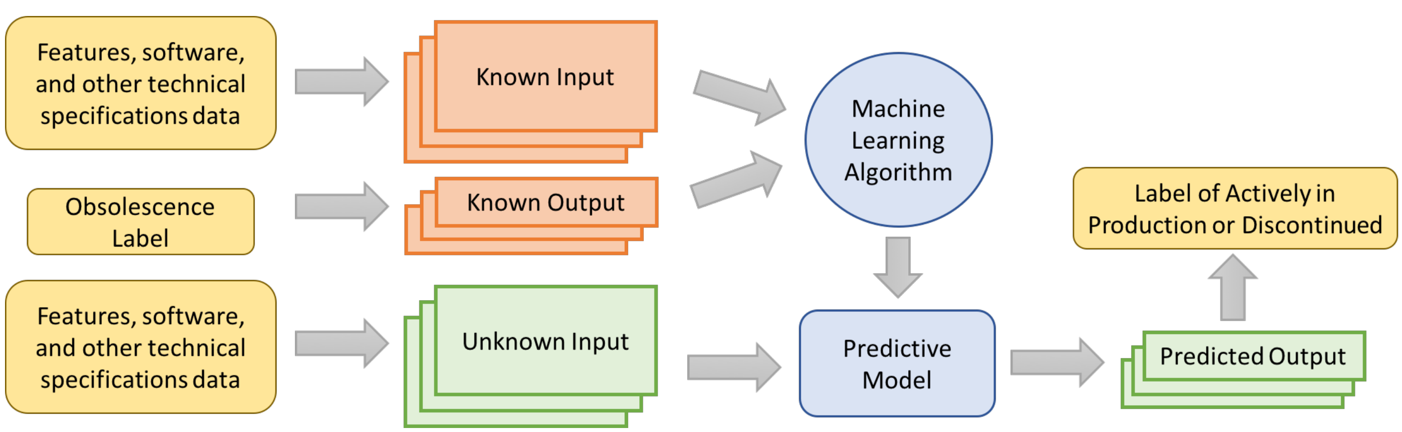

Figure 1: Obsolescence Risk Supervised Learning Process

The case study was conducted by splitting the data set into two random groups. The first group contains 2/3 of the data set and is called the training set because the model will be trained using this data. The second data set represents the other 1/3 of the data and is called the test set because the accuracy of the prediction model, created from the training set, will be tested using the test set. The practice of splitting the data into a training and test set is a common method for model validation [23]. However, currently in most obsolescence forecasting model literature the same data is used in model creation and model testing [5, 7–12]. Because of this, the model's accuracy will tend to be artificially higher in other literature than the true model accuracy. The training and test sets are for an initial analysis of the accuracy of the prediction model using confusion matrixes. A more in-depth analysis was performed to investigate how the change in the proportion of train set size to test size effects the model accuracy (Table 4).

After splitting the data, the training sets were run through a machine learning algorithm to generate a prediction model. Machine learning has many algorithms and each algorithm has sub-variations that can be implemented to increase the accuracy. In this case study, three algorithms were used to create prediction models, artificial neural networks (ANN), support vector machines (SVM), and random forest (RF) [23–25]. ANN was selected because of it's use for predicting obsolescence of descriptive tags in the newspaper industry that was discussed in Section 3. In the list, "Top 10 Algorithms in Data Mining", RF and SVM, were ranked the top three algorithms in machine learning and data mining. The other algorithm named in the top three was K-mean. The ORML method requires supervised algorithms and since K-means is an unsupervised method, it will not be used in this case study.

Then the algorithms were tested on four key areas identified in Zhang & Bivens 2007, accuracy, evaluation speed, interpretability, and maintainability/flexibility [27]. Accuracy is accessed by the percent of cellphones classified correctly. The evaluation speed was calculated by taking the average time for creating 10 prediction models for each of the three algorithms tested. Interpretability was a non-performance based characteristic used to measure the ability for obsolescence managers to glean information from the prediction model. Maintainability/flexibility, another non-performance based characteristic, measures the ability of the algorithm to adapt and scale to industries needs.

The first step was processing the data. All missing numeric values were replaced with the median of the variable and all missing categorical variables were replaced with the most common category for each variable. This was done by using the *na.roughfix* function from the randomForest package [28]. For the ANN and SVM, all categorical variables were converted to numeric variables. If the variable category did not have an obvious order, each category was split into a binary variable that was assigned 1 if the cellphone belonged to that category and 0 if not.

The first algorithm tested was neural networks. The *nnet* function from the nnet R package was utilized to create the prediction models. All the ANN in this study were constructed with 2 hidden layers. The confusion matrix is shown in Table 2. The confusion matrix is a visual representation of how the model classified products. The columns represent the status of the product as predicted by the model and the rows represent the true status of the product. 1295 phones were classified as available and were available while 860 phones were classified as discontinued and were discontinued. However, 67 models were classified as discontinued when they were available and 129 phones were predicted available when actually discontinued. All of these 2351 instances were classified correctly giving the algorithm an accuracy of 91.66%.

Table 1: Neural Networks confusion matrix

	Prediction			
		Available	Discontinued	Total
Actual	Available	1295	67	1362 (95.08%
	Discontinued	129	860	989 (86.96%)
	Total	1424 (90.94%)	927 (92.77%)	2351 (91.66%

Next, the accuracy of support vector machines was tested. Support vector machines creates cuts between the groups of available parts and discontinued parts [26]. The *svm* function from the e1071 R package was used to create the model and a radial basis kernel was selected [29]. The resulting prediction model was able to achieve an accuracy of 92.41% and increase of .75% over neural networks (Table 3).

Table 2: Support Vector Machine confusion matrix

	Prediction			
		Available	Discontinued	Total
Actual	Available	1218	76	1294 (94.13%)
	Discontinued	92	827	919 (89.99%)
	Total	1310 (92.98%)	903 (91.58%)	2213 (92.41%)

The last algorithm tested was random forest. Random forest creates decision trees used to classify parts based on a series of conditional statements based on the product's specifications [25]. The R package randomForest function *randomForest* was used to create the prediction model. In each model 500 trees were generated for each forest [28]. The prediction model created by the random forest algorithm was able to obtain an accuracy of 92.56%, an accuracy higher than both neural networks and support vector machines (Table 4).

Table 3: Random Forest confusion matrix

	Prediction			
		Available	Discontinued	Total
Actual	Available	1243	72	1315 (94.52%)
	Discontinued	98	873	971 (89.91%)
	Total	1341 (92.69%)	945 (92.38%)	2286 (92.56%)

To better understand the relationship between training set size and model accuracy, a sensitivity analysis was conducted to investigate the relationship. Table 4, shows the results of that analysis. The training set was set at 50% of the data set then 60% all the way to 100%; this change can be seen in the left most column.

The prediction models were fitted using the training set, then the models predicted the labels for the training set and testing set and the overall data set (both the training and test combined). To assess the accuracies, each training set size was done ten times randomly sampling a new training set of the same size, then the average accuracy was taken for the training, testing and overall.

Although random forest and support vector machine have relatively close results in the confusion matrix analysis, a study of the change of training set size gives a greater sense of random forest's superiority.

The statistical difference between the accuracies of the algorithms at each training size level was tested using a standard hypothesis test. For training set sizes 80% and 90%, the difference in training set accuracy between random forest compared to ANN and random forest compared to SVM was statistically significant. All other comparisons showed no significant difference. This means random forest is more accuracy than neural networks and support vector machines in larger training sizes, but there is no difference between neural networks and support vector machines.

The next metric evaluated in the study was the speed which a prediction model can be constructed. Ten models were created for each training using the same varying training set size steps as above in Table 5, the creation times were recorded and then averaged for each algorithm for each training set step. The results of this were plotted in Figure 2. Neural Networks were the slowest model but saw a drop as the training set grew in size. However, random forest and support vector machines grow at a constant rate as the training set grew.

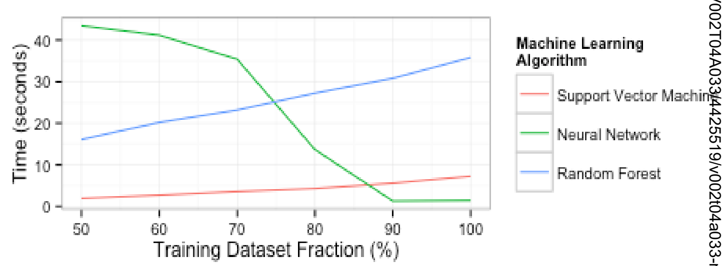

Figure 2: Overall average evaluation speed by training dataset fraction

Table 4: Average Accuracy of Predictions by Training Size

Training Size (%)	Random Forest			Neural Network			Support Vector Machine		
	Training (%)	Testing (%)	Overall (%)	Training (%)	Testing (%)	Overall (%)	Training (%)	Testing (%)	Overall (%)
50	98.8	92.2	95.5	91.8	91.2	91.5	90.9	91.7	91.3
60	98.5	92.5	96.1	91.4	91.7	91.5	91.0	92.2	91.4
70	98.5	92.9	96.8	91.5	91.9	91.6	91.3	92.3	91.6
80	98.2	93.3	97.2	91.7	91.1	91.6	91.6	91.7	91.6
90	98.2	94.3	97.8	91.7	91.2	91.6	91.7	91.2	91.6
100	-	-	98.3	-	-	91.1	-	-	91.6

After the performance characteristics are assessed, the non performance characteristics, interpretability and maintainability/flexibility were evaluated. Random forest ranked highest in both of these categories due to the ease of use and simplicity of decision trees and also the ability to handle numeric and categorical variables while SVM and ANN must convert all categorical variables into numeric.

Table 5: Summary of model preference ranking

	RF	ANN	SVM
Performance based characteristics			
Accuracy	1st	2nd	2nd
Evaluation Speed	2nd	3rd	1st
Non-performance based characteristics			
Interpretability	1st	3rd	2nd
Maintainability/flexibility	1st	2nd	3rd

The last step in the case study analysis was to tally all four key metrics for the three algorithms (Table 5). Random forest was able to capture an impressive three 1^{st} place ranks and one 2^{nd}. This means random forest is the best algorithm to employ for predicting obsolescence risk in the cellphone market. Support vector machine was second overall and neural networks came in last place with two 2^{nd} and two 3^{rd} places.

6 CONCLUSION

In this paper, a framework for using supervised machine learning to predict product obsolescence risk was introduced. The method, ORML, was then applied to classify products as available or discontinued using 7000 unique cell phone models. The case study demonstrated the power of Obsolescence Risk forecasting using Machine Learning (ORML) by correctly identifying available and discontinued cell phone with an accuracy as high as 98.3%. Of the three algorithms tested, random forest was selected as the best candidate for creating obsolescence risk prediction models in the cell phone data. Random forest was selected because it was ranked highest in accuracy, interpretability, and maintainability/flexibility and second highest in creation speed.

The machine learning based method was able to predict obsolescence risk level with a high degree of accuracy and speed. The high accuracy of these prediction models validates machine learning as an appropriate approach to forecasting product obsolescence. Another benefit of this result is to show how fast ORML can accurately predict obsolescence. The slowest ORML models were all still under a minute; compared to many of the current obsolescence forecasting methods that require manual manipulations and inputs which would take days to predict 7000 cellphones.

With obsolescence effecting almost all industries, reducing the cost of impact would save millions of dollars annually. The easiest way to reduce the impact is by involving obsolescence mitigation planning in earlier phases of design and supply chain management. This shift from a reactionary approach to a proactive approach would only be possible through more accurate obsolescence forecasting that can scale to industries' needs. This research establishes machine learning as a capable technique to meet industries' large scale needs while maintaining an extremely high accuracy for predicting obsolescence.

The successful application of ORML in the cell phone market case study demonstrates that the ORML methodology can be utilized by industry, the next step is to create more value added tools to take advantage of the ORML framework. In future works, additional industry case studies would demonstrate the robustness of this model between industries and markets. In this paper, the obsolescence status of only one product type (cell phones) was predicted, the ORML framework could be applied to many products, these many product predictions could be applied to a bill of materials. A system could be design for designers to submit differing bills of materials to a ORML enabled framework and obsolescence risk levels of every component is returned. The individual risk levels could be combined to make a composite score to compare between differing designs to assess which design has the highest risk of obsolescence.

7 ACKNOWLEDGEMENTS

This work was funded by the National Science Foundation through Grant 1238335. Any opinions, findings, and conclusions or recommendations presented in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] QTEC, 2006.
[2] E. Payne, "DoD DMSMS Conference," Charlotte, N.C., 2006.
[3] R. Rai and J. Terpenny, "Principles for Managing Technological Product Obsolescence," *IEEE Trans. Compon. Packag. Technol.*, vol. 31, no. 4, pp. 880–889.
[4] M. Pecht and D. Humphrey, "Uprating of electronic parts to address obsolescence.," *Microelectron. Int.*, vol. 23, pp. 32–36.
[5] J. Terpenny, "MIE 754: Manufacturing & Engineering Economics," Marston Hall at UMass, 1998.
[6] C. Josias and J. Terpenny, "Component Obsolescence Risk Assessment," presented at the Industrial Engineering Research Conference, 2004.
[7] C. Josias, "Hedging Future Uncertainty: A Framework for Obsolescence Prediction, Proactive Mitigation and Management," University of Massachusetts - Amherst, ScholarWorks, 2009.
[8] W. van Jaarsveld and R. Dekker, "Estimating Obsolescence Risk From Demand Data - A Case Study," *Int. J. Prod. Econ.*, vol. 133, pp. 423–431, 2010.
[9] F. J. R. Rojo, R. Roy, and S. Kelly, "Obsolescence Risk Assessment Process Best Practice," presented at the Journal of Physics Conference, 2012, p. 365.
[10] R. Solomon, P. Sandborn, and M. Pecht, "Electronic Part Life Cycle Concepts and Obsolescence Forecasting," *IEEE Trans Compon. Packag. Technol.*, vol. 23, no. 1, pp. 190–193, Mar. 2000.
[11] P. Sandborn, F. Mauro, and R. Knox, "A Data Mining Based Approach to Electronic Part Obsolescence Forecasting," presented at the DMSMS Conference, 2005.
[12] P. Sandborn, "A Data Mining Based Approach to Electronic Part Obsolescence Forecasting," *IEEE Trans Compon. Packag. Technol.*, vol. 30, no. 3, pp. 397–401, 2007.
[13] P. Sandborn, V. Prabhakar, and O. Ahmad, "Forecasting electronic part procurement lifetimes to enable the management of DMSMS obsolescence," *51*, pp. 392–399, 2011.
[14] L. Zheng, R. Nelson, J. Terpenny, and P. Sandborn, "Ontology-Based Knowledge Representation for Obsolescence Forecasting," *J. Comput. Inf. Sci. Eng.*, vol. 13, no. 1, 2012.
[15] IHS, "IHS Electronics & Media Parts Management Solutions Information Resources and Tools for the life of Your Products." 2015.
[16] P. Spicer, Y. Koren, M. Shpitalni, and D. Yip-Hoi, "Design Principles for Machining System Configurations," *CIRP Ann. - Manuf. Technol.*, vol. 51, no. 1, pp. 275–280, 2002.
[17] J. Bennett and S. Lanning, "The Netflix Prize," 2009.
[18] J. A. Cruz and D. S. Wishart, "Applications of Machine Learning in Cancer Prediction and Prognosis," presented at the Cancer informatics, 2007.
[19] J. Wright, "Sparse Representation for Computer Vision and Pattern Recognition," presented at the Proceedings of the IEEE, vol. 98, pp. 1031 – 1044.
[20] C. Tucker and K. M. Harrison, "PREDICTING EMERGING PRODUCT DESIGN TREND BY MINING PUBLICLY AVAILABLE CUSTOMER REVIEW DATA," *Int. Conf. Eng. Des.*, Aug. 2011.
[21] S. Tuarob and C. Tucker, "FAD OR HERE TO STAY: PREDICTING PRODUCT MARKET ADOPTION AND LONGEVITY USING LARGE SCALE, SOCIAL MEDIA DATA," *ASME 2013 Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, Aug. 2013.
[22] F. Wolinski, F. Vichot, and M. Stricker, "Using Learning-based Filters to Detect Rule-based Filtering Obsolescence," presented at the Recherche d'Information Assistée par Ordinateur, RIAO 2000, 2000.
[23] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.*, 2nd ed. 2009.
[24] W. McCulloch and W. Pitts, "A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY," 1943.
[25] L. Breiman, "Random forest," *Mach. Learn.*, vol. 45.1, pp. 5–32.
[26] V. Vapnik and A. Chervonenkis, "Support Vector Network," 1963.
[27] R. Zhang and A. Bivens, "Comparing the use of Bayesian networks and neural networks in response time modeling for service-oriented systems," in *In Proceedings of the workshop on service-oriented computing performance*, Monterey, 2007, pp. 67–74.
[28] D. Meyer, "Package 'e1071,'" 05-Aug-2015.
[29] A. Liaw, "Package 'randomForest,'" 20-Feb-2015.