Classification of Quality Problems in Carpet Manufacturing by Using Data Mining Algorithms

Zeynep Didem Unutmaz Durmuşoğlu and Aysu Borsöken
Department of Industrial Engineering
Gaziantep University
Turkey
unutmaz@gantep.edu.tr , aysuborsoken@hotmail.com

Abstract

Nowadays, with the latest technology, the accumulation of data has increased, and it has become more important to transform the stored raw data into information. The transformation of raw data into information increases earnings and can more easily be adapted to the competitive environment. Data mining is an approach based on statistical applications and machine learning algorithms in converting raw data into information. Data mining has been an important research area to find the hidden information/knowledge inside huge amounts of data. In this paper, quality problems in carpet manufacturing are gathered in a database and the data are analyzed and classified by using data mining algorithms. The main purpose is to classify the quality problems in carpet production. The similarity of the problems resulted in the grouping of them under the same title. Thus, a deeper perspective on carpet quality problems is targeted and the number of customer complaints is expected to reduce in the long term. The data obtained from the database of a carpet producer were pre-processed and adapted to WEKA 3.9.4 software. The dataset was classified using 10-fold cross-validation and Percentage-Split with J.48, two Bayesian classifiers (NaiveBayes and BayesNet), a Nearest Neighbour algorithm (IBk), and two rule learners (OneR and JRip). The achieved conclusions present that the decision tree classifier (J48) performs best (with the highest overall accuracy), followed by the rule learner (JRip) and the BayesNet classifier. The k-NN classifiers are less accurate than the others. In consequence of the J48 classifier, the True Positive Rate is high for three of the classes – Bad, Good, and Average, while it is very low for the other two classes – Very Good and Excellent. The Precision is very high for the Bad class, high for the Good and Average classes, and low for the Very Good and Excellent classes. The acquired results are a little better for the 10-fold Cross-Validation testing option.

Keywords

Data mining (DM), Machine Learning (ML), Statistical Analysis, Classification Methods, Waikato Environment for Knowledge Analysis (WEKA) Software.

1. Introduction

Today, carpet factories are operating in a very complex and highly competitive environment. The main challenge for factories is to deeply analyze their manufacturing, to identify their uniqueness and to build a strategy for further development and future actions. The carpet industry should focus more on the creation of high-fashion, style, and performance-based products. A factory should also consider if they have all the data needed to analyze the defects of carpet at the entry point of the production or they need other data to help the managers support their decisions as how to organize the operation management and approach the promising carpet.

This paper is focused on the implementation of data mining techniques and methods for acquiring new knowledge from data collected by a carpet factory. The main goal of the research is to classify the quality problems in a carpet production factory.

The required data have been taken from one of the carpet production companies located in Gaziantep. Data belongs to the years between 2017 and 2019.

The paper is concerted in six sections. The justification and objectives for the conducted research work are presented in the introduction. A literature review work is supplied in Section 2, the research methodology is declared in Section 3, the data collection is presented in Section 4, the obtained results, the comparative analysis, and validation are imputed in Section 5. The paper concludes with a summary of the achievements.

1.1 Objectives

The specific objective of the proposed research work is to find out if there are any patterns in the available data that could be useful for predicting a carpet's quality defect. The operation management would like to know which features in the currently available data are the strongest predictors of carpet quality. They would also be interested in the data − is the collected data sufficient for making dependable predictions, is it essential to make any changes in the data collection process and how to enhance it, what other data to accumulate in order to increase the usability of the analysis results.

The main aim of this work is to describe the methodology for the implementation of the initiated data mining project at the carpet manufacturing and to present the results of a study aimed at analyzing the performance of different data mining classification algorithms on the provided dataset in order to utilize their potential usefulness for the fulfillment of the project goal and objectives. To analyze the data, the dataset is classified using 10-fold cross-validation and Percentage-Split with J.48, two Bayesian classifiers (NaiveBayes and BayesNet), The Nearest Neighbour algorithm (IBk), and two rule learners (OneR and JRip). The WEKA software is used for the study implementation since is widely used for research purposes in the data mining field.

2. Literature Review

Data mining has been comprehensively applied in different fields of manufacturing in order to classify the quality characteristics or forecasting the class of quality. The use of data mining in manufacturing begins in the 1990s (M.H. Lee, 1993).

One of the first studies is, Irani et al. (1993) generalizes an ID3 (In decision tree learning, ID3 is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset.) algorithm that predicted the outcome of future experiments under various, more general conditions. The algorithm is used successfully in various semiconductor manufacturing applications in both diagnosis and process modeling.

Another application of data mining in production is, Bertino et al. (1999) reports a practical implementation of data mining techniques in the area of semiconductor fabrication. They deal with the analysis of data concerning the wafer production process with the goal of determining possible causes lots of faulty wafers.

Towards the 2000 years, the procedure of manufacturing became more and more complex and, more detailed studies are implemented. As production types increase, information management becomes difficult. Gibbons et al. (2000) describes a computer component manufacturing script that concentrates on the implementation of data mining techniques to improve information management and process improvement within a manufacturing script.

Quality control in production is a process that ensures customers receive products free from defects and meet their needs. Maki and Teranishi (2001) improves an automated data mining system designed for quality control in production.

One of the studies in the development processes of knowledge discovery is, Last and Kandel (2004) submits a new, perception-based method, for automated structure of compact and interpretable models from highly raucous data sets. Last and Kandel utilized the method on yield data of two semiconductor products and described possible instructions for the future use of automated perceptions in data mining and knowledge discovery.

One of the studies to protect product quality with decision tree method, Huang and Wu (2005) executes an analysis of product quality development in ultra-precision manufacturing industry using data mining for developing quality development strategies. Based on ultra-precision optical products, important factors effecting the product quality are identified via the decision tree method for data mining. Harding et al. (2006) examines applications of data mining in manufacturing engineering, in specially production processes, operations, defect detection, maintenance, decision support, and product quality improvement. In 2006, Wang (2006) refers the importance of using data mining techniques in manufacturing. Also, Rokach and Maimon (2006) realizes that the accessibility and abundance of knowledge made data mining a matter of importance and necessity to develop manufacturing quality.

Data mining studies are constantly continuing in different production areas. Chien et al. (2007) attempts data mining in the semiconductor fabrication process. Liu improves a data mining algorithm for designing the cellular manufacturing systems.

Yu et al. (2008) applies a knowledge-based artificial neural network model for monitoring the manufacturing process and identifying defect quality categories of the products being produced. Additively, a genetic algorithm-based rule subtraction approach is developed to explore the causal relationship between manufacturing parameters and product quality.

Another classification study in data mining is, Hsu (2009) implements a data mining structure, using anthropometric data, to improve industrial standards for adult females. By applying the proposed structure, body types can be properly classified. They can predict the proportional quantities necessary for each size, resulting in enhanced production, economic material control, and accurate production planning for specific marketplace. The proposed structure combines statistical methods and data mining techniques to explore the anthropometric data.

Charaniya et al. (2010) describes a multivariate data mining technique that combines more than one-hundred time-dependent off-line, on-line, as well as single point parameters across different production stages to forecast a key process output – the run performance. Generally, they show the power of mining process data in revealing hidden correlations between process outcome and process parameters.

3. The Research Methodology

The initiated data mining project at the factory is implemented following the **CRISP-DM (Cross-Industry Standard Process for Data Mining)** model. The CRISP-DM is selected as a research approach because it is a freely existing, and implementation-neutral standard for data mining projects, and it is widely used by researchers in the field during the last ten years.

The CRISP-DM is a cyclic approach, involving six main phases – Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. There are several internal feedback cycles between the phases, resulting from the very complex non-linear nature of the data mining process and providing the achievement of consistent and reliable results. As it can be seen in figure 1.

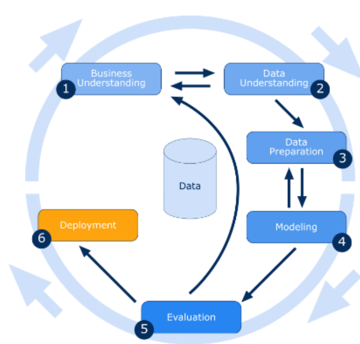

Figure 1. Data Mining Process

The software tool that is used for the project implementation is the open-source software **WEKA**, offering a wide range of classification methods for data mining (Witten et al. 2005).

During the **Business Understanding Phase**, a prevalent literature review is performed in order to study the existing problems at manufacturing that have been untied by the implementation of data mining techniques and methods in previous research projects. Formal interviews with representatives of the business management at factories are also conducted, for finding out the specific problems at the factory which have not yet been solved but are considered very important for the improvement of the manufacturing performance and more effective and efficient management. Some insights are gathered from informal talks with a product manager, business engagement director, operators, and representatives of the administrative staff (IT experts and managers). Based on the results of the performed research, the project aims and objectives, and the basic research questions are formulated.

The basis project goal is to find out the high potential of data mining applications for the manufacturing, referring to the optimal usage of data mining methods and techniques to deeply analyze the gathered historical data. The project-specific purpose, to classify the quality problems in a carpet production factory according to the manufacturing performance results, in data mining terms is considered a classification problem to be solved by using the available quality defect data. This is a task for manufacturing process because the classification models are constructed from data where the target (or response) variable is known.

During the **Data Understanding Phase**, the application process for classify quality at the factory is studied, including the formal procedures and application documents, in order to identify the types of data collected from stored in the factory databases in electronic format. The rules and procedures for collecting and storing data about the manufacturing process of the carpets are also reviewed. Discussions with representatives of the administrative staff responsible for the factory data collection, storage are also carried out. Production data is basically stored in two databases. All the data related to the product sales campaigns are stored in the carpets production database, including personal data of customer orders (names, addresses, quality of product, preferred products, designs, etc.), data about the organization and performance of the quality, information achieved by the most defective products, data related to the final classification of high and low quality carpets, etc. All the data classify the quality problems is stored in the carpet factory database, including weaving, yarn quality, operating process, and administrative data, the information achieved at the defects on the different department, etc.

During the **Data Preprocessing Phase**, quality defect data from the databases are extracted and organized in a new flat file. Data Preprocessing Phase is explained in detail in the data collection section.

During the **Modeling Phase**, the methods for building a model that would classify the qualities into the five classes (categories), depending on their quality defect, are considered and selected. Several different classification algorithms are implemented during the performed research work, selected because they have the potential to yield good results. Popular WEKA classifiers (with their default settings unless specified otherwise) are used in the experimental study, including a common decision tree algorithm C4.5 (J48), two Bayesian classifiers (NaiveBayes and BayesNet), The Nearest Neighbour algorithm (IBk), and two rule learners (OneR and JRip). The achieved research results are submitted in the next paper section.

4. Data Collection

Quality defect data from the databases were used as the data collection instrument. The preliminary research sample is provided by the software technical staff responsible for the data collection and includes data about 24.008 quality defects data, described by 65 parameters, including abrage (weaving), abrage (sourced from the supplier of yarn), design defect, vapor stain, warp yarn, overlock defect, etc. The provided data is subjected to many transformations. Some of the parameters were removed, e.g., mold, tear, stained on the yarn, and curve cutting, because the data did not concern scarcely any affect quality defect. Some of the variables containing important data for the research are fields where weaving knowledge is being entered at the data collection stage. Therefore, these variables are processed and turned into numerical variables with a limited number of distinct values. The data is also being studied for missing values, which are very few and could not affect the results, and for obvious mistakes, which are corrected.

Basically, the challenge in the presented data mining project is to predict the carpet quality for classification based on the collection of attributes providing information about quality defects of the carpets. The selected target variable in this case, or the concept to be learned by the data mining algorithm, is the “quality defect class”. A categorical target variable is constructed based on the original numeric parameter product defect. It has five distinct values (categories):

"excellent", "very good", "good", "average" and "bad". The five categories (classes) of the target (class) variable are determined from the total quality defect achieved by the data. "Excellent" qualities are considered those who have a total quality defects in the range between 0 and 100 pieces, "very good" – in the range between 100 and 300 pieces, "good" – in the range between 300 and 750 pieces, "average" – in the range between 300 and 400 pieces, and "bad" – in the range above 750 pieces.

The final dataset used for the project implementation contains 24020 instances (1068 in the "excellent" category, 2167 in the "very good" category, 3784 in the "good" category, 1091 in the "average" category, and 15910 in the "bad" category), each described with 14 attributes (1 output and 13 input variables), nominal and numeric. The study is limited to quality defect of carpets data for quality classification in carpet production (for the time year between 2017 and 2019).

5. The Achieved Results

The primary objective of the study is to find out if it is possible to predict the class (output) variable using the explanatory (input) variables which are maintained in the model. Several different algorithms are implemented for building the classification model, each of them using different classification techniques. The WEKA Explorer application is used at this phase. Each classifier is implemented for two testing options–cross-validation (using 10 folds and applying the algorithm 10 times – each time 9 of the folds are used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing).

5.1 Decision Tree Classifier

Decision trees are strong and popular tools for classification. A decision tree is a tree-like construction, which begins from root attributes and ends with leaf nodes. Generally, a decision tree has several branches consisting of different attributes, the leaf node on each branch presenting a class or a kind of class distribution. Decision tree algorithms define the relationship among attributes, and the connection importance of attributes. The advantages of decision trees are that they represent rules which could easily be understood and commented on by users, do not necessitate complex data preparation, and perform well for numerical and categorical variables.

The WEKA J48 classification is implemented on the dataset during the experimental study. It is based on the C4.5 decision tree algorithm, structuring decision trees from a set of training data using the notion of information entropy. The J48 classifier classifies accurately about 2/3 of the instances (73.93 % for the 10-fold cross-validation testing and 71.76 % for the percentage split testing), produces a classification tree with a size of 4205 nodes and 2103 leaves. As it can be seen in table 1 that classification of the instances is into the five classes.

Table 1. Results for the decision tree algorithm (J48)

Class	J48-10-fold Cross Validation		J48-Percentage-Split	
	TP Rate	Precision	TP Rate	Precision
Bad	0.667	0.571	0.624	0.651
Average	0.692	0.756	0.631	0.685
Good	0.772	0.823	0.735	0.739
Very Good	0.485	0.541	0.475	0.493
Excellent	0.327	0.332	0.321	0.350
Weighted Average	0.588	0.604	0.557	0.583

The conclusions for the detailed accuracy by class, involving the True Positive (TP) rate (the proportion of examples which were classified as class x, among all examples which truly have class x) and the Precision (the proportion of the examples which truly have class x among all those which were classified as class x), are submitted in Table 1.

The results expose that the True Positive Rate is high for three of the classes – Bad (62-67 %), Good (73-77 %), and Average (63-69 %), while it is very low for the other two classes – Very Good (47 %) and Excellent (32 %). The Precision is very high for the Bad class (57-65 %), high for the Good (73-82 %) and Average (68-75 %) classes, and low for the Very Good (49-54 %) and Excellent (33-35 %) classes. The acquired results are a little better for the 10-fold Cross- Validation testing option.

5.2 Bayesian Classifiers

Bayesian classifiers are statistical classifiers that estimate class membership by probabilities, such as the probability that a given sample pertains to a particular class. Several Bayes' algorithms have been improved, among which Bayesian networks and Naive Bayes are the two basic methods. Naive Bayes algorithms suppose that the effect that an attribute on a given class is independent of the values of other attributes. However, in operation, dependencies often exist among attributes; hence Bayesian networks are graphical models, which can identify partner conditional probability distributions. Bayesian classifiers are popular classification algorithms due to their simplicity, calculation efficiency, and performance for real-world problems. Another important advantage is also that the Bayesian models are fast to evaluate, and have a very high accuracy in many areas.

The two WEKA classification filters implemented on the dataset are the NaiveBayes and the BayesNet. The NaiveBayes and the BayesNet are tested for 10-fold cross-validation and percentage split options. The achieved results are shown in Table 2.

Table 2. Results for the Bayesian Classifiers.

Class	NaiveBayes				BayesNet			
	10-fold Cross Validation		Percentage Split		10-fold Cross Validation		Percentage Split	
	TP Rate	Precision	TP Rate	Precision	TP Rate	Precision	TP Rate	Precision
Bad	0.334	0.286	0.320	0.292	0.538	0.551	0.522	0.530
Average	0.328	0.513	0.323	0.397	0.725	0.643	0.623	0.619
Good	0.491	0.487	0.480	0.473	0.730	0.677	0.687	0.665
Very Good	0.153	0.257	0.151	0.241	0.475	0.462	0.463	0.437
Excellent	0.236	0.171	0.149	0.128	0.317	0.295	0.227	0.279
Weighted Average	0.308	0.342	0.284	0.306	0.557	0.525	0.504	0.506

The general average of the Bayesian classifies is about 40 % which is not high, and it is worse compared to the performance of the decision tree classifier (55-60 %). The detailed accuracy results for the Bayesian classifiers find out that the True Positive Rate is not so high for the Bad class (32-53 %), low for the Very Good (15-47 %), high for Good (48-73 %) classes, high for the Average class (32-72 %), and so low for the Excellent class (15-31 %). The Precision is low for the Bad class (28-55 %), not so high for the Good (47-67 %) and Very Good (24-46 %) classes, not so high for the Average (39-64 %) and so low for Excellent (12-29 %) classes.

The Naive Bayes algorithm classifies the defects taking into account the independent effect of each attribute to the classification, and the final accuracy is specified based on the results procured for all the attributes. The BayesNet classifier creates a simple graph, containing all input attributes at the first level.

5.3 The k-Nearest Neighbour Classifier

The k-Nearest Neighbor algorithm (k-NN) is a method for classifying objects according to closest quality defects examples in the feature space. The k-NN algorithm is amongst the easy of all machine learning algorithms: an object is classified by a generality vote of its neighbors, with the object being appointed to the class most common amongst its k nearest neighbors (k is a positive integer). The best selection of k depends upon the data; in general, larger values of k decrease the impact of noise on the classification but make differences between classes less distinct. The accuracy of the k-NN algorithm can be severely reduced by the presence of noisy or irrelevant features.

The WEKA IBk classification choosing is implemented to the dataset, which is a k-NN classifier. The algorithm is executed for two values of the parameter k (50 and 500), and the two testing options – 10-fold cross-validation and percentage split. The results are submitted in Table 3.

Table 3. Results for the k-NN Classifier

Class	k-NN Classifier K=500							
	K=50				K=500			
	10-fold Cross Validation		Percentage Split		10-fold Cross Validation		Percentage Split	
	TP Rate	Precision	TP Rate	Precision	TP Rate	Precision	TP Rate	Precision
Bad	0.437	0.384	0.408	0.337	0.311	0.244	0.298	0.299
Average	0.363	0.350	0.337	0.311	0.057	0.116	0.031	0.108
Good	0.424	0.482	0.384	0.418	0.098	0.162	0.056	0.147
Very Good	0.044	0.020	0.034	0.024	0	0	0	0
Excellent	0.049	0.105	0.041	0.056	0	0	0	0
Weighted Average	0.263	0.268	0.240	0.229	0.093	0.104	0.077	0.110

The k-NN classifier accuracy is about 30 % and varies by the selected value for k. The conclusions are slightly better for k = 50 if compared to k =500. This classifier works with higher accuracies for the bad (32-43 %), with low accuracy for the Average (1-30%) class and good (1-42 %), and performs very badly for the Very good (0 %) and Excellent (0 %) classes. The Precision is good for the Bad class (35 %), but not so high for the Average (1-35 %) and Good (10-48 %) classes.

5.4 Rule learners

Two algorithms for composing classification rules are considered. The OneR classifier composes a one-level decision tree stated in the form of a set of rules that all test one particular attribute. It is a simple method that usually produces effective rules with high accuracy for characterizing the structure in data. The JRip classifier applies the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm. Classes are viewed in increasing size and an initial set of rules for the class is composed using incremental reduced-error pruning. The conclusions are presented in Table 4.

Table 4. Results for the rule learners

Class	OneR				JRip			
	10-fold Cross Validation		Percentage Split		10-fold Cross Validation		Percentage Split	
	TP Rate	Precision	TP Rate	Precision	TP Rate	Precision	TP Rate	Precision

Bad	0.361	0.367	0.380	0.354	0.487	0.634	0.472	0.604
Average	0.357	0.435	0.283	0.257	0.631	0.771	0.611	0.801
Good	0.444	0.354	0.425	0.328	0.673	0.780	0.659	0.761
Very Good	0.101	0.247	0.108	0.231	0.347	0.579	0.321	0.524
Excellent	0.115	0.113	0.088	0.045	0.291	0.380	0.279	0.348
Weighted Average	0.275	0.303	0.256	0.243	0.485	0.628	0.468	0.607

The achieved conclusions show that, as expected, the JRip classifier performs better than the OneR classifier. The overall accuracy of the JRip classifier is 60 about 50 %, and for the OneR classifier, it is about 30 %. Both rule quality defects perform not so bad for the Good class, the JRip classifier submitting slightly better results than the OneR classifier. Besides, The OneR quality defect uses the minimum-error attribute for prediction and in this case, this is a good conclusion for carpet.

5.5 Results

The conclusions for the performance of the chosen classification algorithms (TP rate, percentage split test option) are summarized and submitted in Figure 2.

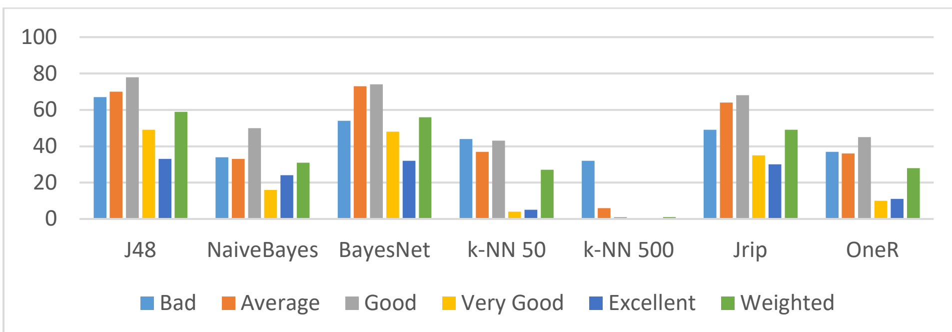

Figure 2. Classification Algorithms Performance Comparison

The achieved conclusions present that the decision tree classifier (J48) performs best (with the highest overall accuracy), followed by the rule learner (JRip) and the BayesNet classifier. The k-NN classifiers are less accurate than the others. However, all tested classifiers are showing an overall accuracy below 60 % which means that the error rate is not low and the predictions are not very reliable.

When the detailed accuracy for the different classes is concerned, it is visible that the predictions are worst for the Excellent and Very good class, the k-NN=500 classifiers being certainly unable to predict them. The predictions for the Bad, Average, and Good classes are more exact than for the other classes, and all classifiers perform with accuracies around 50-70 %. The decision tree classifier (J48) and the BayesNet are most reliable because they perform with the highest accuracy for all classes. The k-NN classifier is not able to predict the classes which are less represented in the dataset.

5.6 Validation

To validate the model, a 10-fold cross-validation method and percentage split are applied. In the 10-fold cross-validation, firstly, the data is divided into ten equal parts, then the network is trained and tested ten times. For example, for the first training of the network, the first nine part of the data is considered as the training and the last part as the test. For the second run, the first eight part and the 10th part of the data is considered as the training set and the ninth part for the test and so on. In addition, the Percentage split has 2/3 of the dataset used for training and 1/3 – for testing.

All tested methods are showing an overall accuracy below 60 % which means that the error rate is not low and the predictions are not very reliable.

6. Conclusion

The conclusions achieved by implementing selected data mining algorithms for classification on quality problems in a carpet production factory sample data exhibit that the prediction rates are not remarkable (vary between 50-60 %). Furthermore, the classifiers show differently for the five classes.

The conclusions from the performed study are actually the initial steps in the realization of an applied data mining project at a carpet production factory. The conclusions made from the conducted research will be used for defining the further steps and directions for quality defects in carpet manufacturing data mining project application, including possible transformations of the dataset, tuning the classification algorithms' parameters, etc., in order to achieve more accurate results and to extract more important knowledge from the available data. Recommendations will also be provided to the business management, concerning the sufficiency and availability of quality defects data, and related to the improvement of the data collection process.

References

M.H. Lee, Knowledge based factory, *Artificial Intelligence in Engineering*, Available: https://doi.org/10.1016/0954-1810(93)90021-7, pp. 109-125, 1993.

K.B. Irani, J. Cheng, U.M. Fayyad, et al, Applying machine learning to semiconductor manufacturing, Proc. SPIE 1293, *Applications of Artificial Intelligence VIII*, Available: https://doi.org/10.1117/12.21147 , pp. 41-47, 1993.

E. Bertino, B. Catania, E. Caglio (Applying data mining techniques to wafer manufacturing), *in: J.M. Zytkow, J. Rauch (Eds.), PKDD'99, LNAI*, vol. 1704, Springer-Verlag, Berlin, pp. 41-50, 1999.

W.M. Gibbons, M. Ranta, T.M. Scott, et al., Information management and process improvement using data mining techniques, in: R. Loganantharaj et al. (Eds.), *IEA/AIE 2000, LNAI, 1821*, Springer-Verlag, Berlin, pp. 93–98, 2000.

H. Maki, Y. Teranishi, Development of automated data mining system for quality control in manufacturing, in: Y. Kambayashi, W. Winiwarter, M. Arikawa (Eds.), *DaWak 2001, LNCS 2114*, Springer-Verlag, Berlin, pp. 93–100, 2001.

M. Last, A. Kandel, Discovering useful and understandable patterns in manufacturing data, *Robotics and Autonomous Systems 49*, Available: doi: 10.1016/j.robot.2004.09.002, pp. 137–152, 2004.

H. Huang, D. Wu, Product quality improvement analysis using data mining: a case study in ultra-precision manufacturing industry, in: L. Wang, Y. Jin (Eds.), *FSKD 2005, LNAI, 3614*, Springer-Verlag, Berlin, pp. 577–580, 2005.

J.A. Harding, M. Shahbaz, S. Srinivas, A. Kusiak, Data mining in manufacturing: a review, *Journal of Manufacturing Science and Engineering 128*, pp. 969–976, 2006.

K. Wang, Data mining in manufacturing: the nature and implications, in: K.Wang, G. Kovacs, M. Wozny, et al. (Eds.), *International Federation for Information Processing (IFIP) Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management*, vol. 207, Springer-Verlag, Boston, pp. 1–10 , 2006.

L. Rokach, O. Maimon, Data mining for improving the quality of manufacturing: a feature set decomposition approach, *Journal of Intelligent Manufacturing 17 (3)*, Available: doi:10.1007/s10845-005-0005-x, pp. 285–299, 2006.

C. Chien, W. Wang, J. Cheng, Data mining for yield enhancement in semiconductor manufacturing and an empirical study, *Expert Systems with Applications 33,* Available: doi:10.1016/j.eswa.2006.04.014 , pp. 192–198, 2007.

C. Liu, A data mining algorithm for designing the conventional cellular manufacturing systems, in: M.A. Orgun, *J. Thornton (Eds.), AI 2007, LNAI, 4830,* Springer-Verlag, Berlin, pp. 715–720, 2007.

J. Yu, L. Xi, X. Zhou, Intelligent monitoring and diagnosis of manufacturing processes using an integrated approach of KBANN and GA, *Computers in Industry 59*, Available: https://doi.org/10.1016/j.compind.2007.12.005, pp. 489–501, 2008.

C. Hsu, Data mining to improve industrial standards and enhance production and marketing: an empirical study in apparel industry, *Expert Systems with Applications 36 (3)*, Available: doi:10.1016/j.eswa.2008.04.0092009, pp. 4185–4191, 2009.

Salim Charaniya, Huong Le, Huzefa Rangwala, Keri Mills, Kevin Johnson, George Karypis, Wei-Shou Hu, Mining manufacturing data for discovery of high productivity process characteristics, *Journal of Biotechnology*, Available: doi:10.1016/j.jbiotec.2010.04.005, 2008.

Witten, I., E. Frank. Data Mining: Practical Machine Learning Tools and Techniques. *Morgan Kaufmann Publishers, Elsevier Inc*., 2005.

Biographies

Zeynep D. U. Durmuşoğlu received her BSc, MSc and PhD. at Industrial Engineering of Gaziantep University. She worked as a research assistant at the same department for six years. She is currently associate professor at Industrial Engineering Department of Gaziantep University, Turkey. She has research interest in: business analytics, medical modelling, technology analysis, simulation optimization, heuristic optimization and negotiation.

Aysu Borsöken received her BSc in Industrial Engineering from Gaziantep University, in 2017. She is a student MSc in industrial engineering from Gaziantep University between 2018 and 2021. She works as an independent researcher in a wide variety of fields. Besides, she worked in industrial companies as a production planning engineer, and R&D engineer. She works International Sales Specialist at a carpet company. Her research interests include data mining, the textile sector, customer relationship management, statistical method, productivity, business model, optimization, and machine learning.