TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

ACTA TECHNICA NAPOCENSIS

Series: Applied Mathematics, Mechanics, and Engineering
Vol. 63, Issue III, September, 2020

PRACTICAL DATA MINING APPLIED IN STEEL COILS MANUFACTURING

Imanol BILBAO, Javier BILBAO, Cristina FENISER, Andrei BORSA

Abstract. *Due to the big amount of data that researchers can be obtain at the present, Data Mining is an important research field that can be applied in different fields. One of these fields is the recommendation or recommender systems, which take a large number of data, values, products or characteristics to obtain some outputs as recommendations for a user that is interested in some area. Sometimes Data Mining is established as one of the stages of a more generic process called Knowledge Discovery in Databases. In this paper, a review of these processes is done and we analyze some techniques used, such as KNN, Naïve Bayes, decision trees, SVM, ANN, regression and multiclassifiers.*
Keywords: *Data Mining, Recommender Systems, Knowledge Discovery in Databases, Steel Coils*

1. INTRODUCTION

Today's organizations are gathering increasing volumes of information from all kinds of sources, including websites, business applications, social networks, mobile devices and increasingly the Internet of Things (IoT). All this large amount of data can be useful if properly managed. In this sense, data mining is the automated process of classifying large data sets to identify trends and patterns and establish relationships, to solve scientific or even business problems or generate new opportunities through the analysis of data.

There are several definitions for data mining. Fayyad et al. say that data mining is the application of specific algorithms for extracting patterns from data [1]. Liu et al. define Data Mining as a lot of, incomplete, noisy, fuzzy and random data extracted from implicit in them, and these authors continue saying that it is a set of technologies and applications, or a method for large-capacity data and data relationships between study and modeling of collections [2]. And, in our opinion, Witten et al. provide the best definition when they say that Data Mining is the process of discovering patterns in data, where this process must be automatic or (more usually) semiautomatic, and where the patterns discovered must be meaningful in that they lead to some advantage, e.g. an economic advantage [3].

On the other hand, the processing of massive volumes of data can discover valuable information that, at first glance, goes unnoticed. For example, about the purchasing behavior of products or services, or also about how to generate new products using the behaviors of customers. Here, recommender systems would come into play. A recommender system is an intelligent system that provides users with a series of personalized suggestions (recommendations) about a certain type of elements (items). Recommender systems study the characteristics of each user and through a processing of the data, find a subset of items that may be of interest to the user. In recent years, and mainly due to the overload of information that we have on the Internet, recommender systems have proliferated, which provide users, information, products, etc. that may be of interest to this user, after conducting a analysis of his profile, his taste and even the way in which the user browses the internet.

2. RECOMMENDER SYSTEMS

Recommender systems have been defined in different ways. R. Burke defines recommender systems as systems that produce personalized recommendations as output or have the effect of guiding the user in a personalized way to interesting or useful products among a large number of available products [4]. A more general definition is that Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user [5, 6, 7]. Xiao and Benbasat say that recommender systems are software agents that elicit the interests or preferences of individual consumers for products, either explicitly or implicitly, and make recommendations accordingly. They have the potential to support and improve the quality of the decisions consumers make when searching for and selecting products online [8].

Recommender systems provide the users as output a series of products, services or contents that fit their tastes or needs, which have previously been collected by the system.

The information that users have (we can call feedback) is used to know their tastes or preferences and describes what products the user liked in the past and to what extent. This information may have been obtained in different ways:

- Implicit information: the information of the users is obtained from their interaction with the system [9, 10], such as the songs they have listened to, web pages visited, if they have consulted the description of a product, time that has remained in it, actions on the products (saved, printing, etc.), etc. [11]. This type of information has the advantage that it can be obtained in a relatively simple way, but it is usually not very precise [12] and, in some cases, difficult to apply to the recommendation process.
- Explicit information: in this type of information, the user must actively provide their preferences. To do this, a system of valuation of the products is used (using a numerical scale [13], a linguistic scale) in which the user indicates, to what extent, a set of products they have liked or meet their needs (favorite products, wish list, etc.). Explicit information is a more accurate source of information than implicit information, but it is more complicated to obtain, since it requires the collaboration of users [14].

In certain systems, the use of implicit and explicit information can be considered at the same time. This can be done through a combination of the two ways [15], transforming the implicit information to explicit [9, 16] or improving the results of the recommendation of one type by refining them with those of the other type [17].

The criteria for choosing what type of information the recommender system uses depends on the domain to which it applies. As already mentioned, the implicit information is simpler to obtain and reflects the user's interaction with the system, but it is less accurate than the explicit information. Alsaleh et al. demonstrated empirically that the use of both types of information (implicit and explicit) at the same time, improves the recommendation results [18]. In addition, we must take into account what kind of information is available in the environment, the cost of it, etc.

2.1 Obtaining information

Regardless of whether the information is implicit or explicit, we will have information from each of the users about the products. We can suppose for a first approach that it is considered that the information held about users is explicit and that preference ratings are used on a numerical scale. In this way, the utility function or set of preference ratings (*pr*) can be represented as follows:

$$pr: \quad U \times P \rightarrow D \tag{1}$$

where *pr* is the set of preference scores that *U* users have given on *P* products in a domain *D*. One of the most used domains for preference ratings is D = {1, 2, 3, 4, 5}, with 1 being the lowest preference value, which indicates that a product does not suit the user or does not meet his needs, and 5 the highest preference value, which indicates that the product perfectly fits the needs of the user.

Mathematically, the problem of the recommendation can be defined as a prediction problem [19]. In this way, the set of preference ratings is taken as an incomplete function that is intended to be approximated by different techniques. Once the valuations of each product have been predicted, the product with the highest predicted preference rating (Eq. 2) is selected among the products that the user has not yet rated.

$$\forall u \in U \quad p'_u = \underset{p \in S}{\arg\max}\; pr(u, p) \quad (2)$$

where U is the set of all users, S the set of products that has not yet been valued by user u, and $pr(u, p)$ is the utility value that recommender system predicts for product p and user u. In addition, $S \in P$.

Another approach to solve the problem of the recommendation is by binary classification of the products. In this way, an attempt is made to discriminate whether a product is suitable for a user or not.

The basic tasks that a recommender system has to perform are three: the calculation of the user profile, the modeling of the products and the filtering [20].

1. User profile. Recommender systems need to have information about the user that they intend to recommend. To do this, a user profile is calculated from the information that is known about it. Different approaches have been applied to this task, such as modeling behavior, modeling interests or intentions. With this information, different actions can be made: user classifications, clustering, extract behavior patterns, make predictions about them, etc.
2. Product modeling. In order to know the products that exist in the system, a profile is also calculated. These product profiles can be used to analyze the products, perform an extraction of semantic variables, analyze the diversity of them, etc.
3. Filtered out. Once the profiles of users and products are known, a filtering is carried out based on how suitable a product is to a specific user (usefulness) or to predict the valuation that a user would give to a product. With this objective, there are different approaches, such as filtering based on the characteristics of the products, based on rules or collaborative filtering.

3. THE KDD PROCESS

Sometimes data mining is established as one of the stages of a more generic process called Knowledge Discovery in Databases (KDD), which is the database analysis process that seeks to find unexpected relationships that are of interest or value to the holder of this database [21]. In simple terms, it is to find non-trivial relationships within the data. This iterative process consists of five stages, where data mining is defined as one more phase of this procedure [22, 23]. As it is an iterative process, it is possible to return to a previous stage in case there are no satisfactory results at the end of one. These stages or phases are usually described as follows:

- Integration or Selection. In this step the variables and the sources to be considered in the complete process are chosen. That is, the creation of the data set and the study database in the process are part of this stage.
- Preprocessing. The analysis and cleaning of the data are the main lines to follow in this stage, where there is the treatment of absent values (missing) and out-of-range values (outliers). In order to implement this step, different data imputation techniques are used, ranging from a value-to-value treatment (simple imputation) to a replacement contemplating multiple variables and their values (multiple imputation).
- Transformation. Sometimes this step is underestimated, but it is a necessary stage in the process. New variables are generated from the study of the nature of the original variables; from the perspective of the scale, nominal or continuous, or from the distribution of the present values.
- Data Mining. This step of the KDD process consists in the application of data analysis to discover an ad-hoc algorithm that produces a particular enumeration of patterns from the data and that produces them considering computational capacity constraints [23]. Therefore, the model and algorithm to be

used are selected, under the assumptions that maintain the primary objectives of the study.

- Interpretation and Evaluation. This last phase involves the evaluation measures and the transposition of technical results at commercial or application levels, in such a way that the application of the procedure converges to corrective actions in the business or process that solve the studied phenomenon. Regarding the evaluation, this can be applied from two edges: technical and commercial. The first is subdivided according to the type of validation and its metrics that apply to the model, while the commercial evaluation is not standardized and surveys can generally be used to measure the practical effectiveness of the procedure. The main techniques of technical evaluation are holdout and cross-validation.

4. DATA MINING TECHNIQUES

In the data mining stage, a variety of models and perspectives can be applied. Algorithms are part of the models section and there is a big variety of types. Among them, we can describe the most used:

4.1 K-Nearest Neighbor (KNN)

The algorithm of the nearest K neighbor or KNN is one of the simplest algorithms. This algorithm does not require any parameter outside the number of neighbors to consider. In short, the algorithm can be summarized in that it gathers the nearest K neighbors and makes them vote, the class with the most neighbors wins, ..., the more neighbors we consider, the lower the error rate [24]. This closeness is usually measured based on some distance, so different results can be obtained depending on the chosen distance, since different metrics will denote different regions [25]. Its general scheme is proposed below:

4.2 Naive Bayes

In general, classification algorithms that use Bayesian learning are complex in the sense of the number of parameters. However, the Naive Bayes method converts this complexity into a feasible simplicity, because, according to Mitchel, it makes an assumption of conditional independence that reduces the number of parameters to estimate, when $P(x \mid y)$ is modeled [26]. Quantitatively, if the variable to be predicted has two values, it goes from estimating 2 (2n - 1) parameters to 2n.

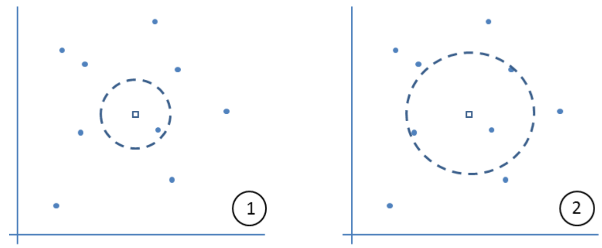

Fig. 1. Structural model of KNN. In the left side, the first iteration with only one neighbor; and in the right, the second iteration with three neighbors.

The usefulness of the Bayesian learning algorithms is that it gives a probabilistic measure of the importance of these variables in the problem, and, therefore, an explicit probability of the hypotheses that are formulated [27].

4.3 Decision Trees

Decision trees are models that are usually represented as graphs. It is a predictive model that can be used to represent both regressive models and those of classification, referring to a hierarchical model of decisions and their consequences [28]. Within a general scheme, the decision tree consists of a graph where there is a single or parental node, which contains the instances to be contemplated in the model. An example of this type of model is the LADTree, which is a type of decision tree that iterates over the ADTree that is a tree that instead of establishing criteria and dividing the sample, assign a score to the relevant categories of certain variables.

4.4 Support Vector Machines (SVM)

Unlike the previous algorithms, the Support Vector Machines use complete planes to find the best division of the instances that allows to classify them in an optimal way, whose formulation is a quadratic minimization problem with a number of variables equal to the number of training cases [29]. There are also formulations to train SVM using linear programming. These formulations are based on the consideration of the L_1 and L_∞ norms instead of the norm L_2 [30]. Therefore, for large numbers of data, a large capacity device should be used. In addition, SVM uses the optimization branch of mathematics, since the problem they address involves optimization of a convex function [31], this means that it does not contain local minimums. Another particularity that this algorithm comprises is that it does not require information about the distribution of the data set. That is, these systems search for the balance between certainty and quantity of data to be accepted. It is this combination that induces the origin of the problem of minimization of structural risk, whose solution is later translated into an optimization problem about finding the separator hyperplane between the instances.

4.5 Neural Networks

Regarding Artificial Neural Networks (ANN), all research areas say that ANN are part of them. There is no doubt that they are a powerful tool and that they can be used by different techniques or fields. Data Mining also uses ANN as a model and nowadays ANN are one of the most popular strategies for supervised learning and classification. However, due to the complexity it has, you cannot know exactly the origin of its results, which is a difficulty when explaining its operation. In a direct sense, an artificial neural network (or simply called neural network) consists of processing elements (called neurons) and the connections between them with coefficients (weights) linked to the connections, which constitute a neuronal structure, and a training and reminder algorithms attached to the structure [32]. This can be described as a pool of simple processing units that communicate by sending signals to each other over a large number of weighted connections [33]. In a purely feedforward network architecture the i layer's neurons only are connected to the i+1 layer's neurons. The input layer receives the data and transforms it in network signals. The output layer transmits its signals outside the network. The hidden layers are responsible of finding an appropriate representation of the info to achieve the desired result [34]. A general outline of this model is presented in the next figure:

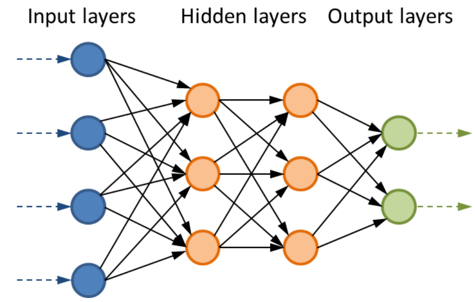

Fig. 2. Structural model of a neural network.

4.6 Regression

The regression, at present, consists of the study of the dependence of the dependent variable with respect to one or more variables (the "explanatory" variables), with the objective of estimating and / or predicting the average or average population value of the first one in terms of the known or fixed values (in repeated samples) of the last ones [35]. Therefore, this model serves to predict and classify, where its typical use in business is to predict the demand or future inventory of a company. In the case of classification, the regression that is commonly used sometimes is not very effective, since the variable to predict has a nominal connotation. However, there is a type of regression, called logistic regression, which is responsible for predicting nominal variables.

4.7 Multiclassifiers

Multiclassifiers, unlike previous models, seek to find ways to return (or combinations to return) an effective prediction or classification. For this, it is intended to explore the greatest number of accessible paths, covering all the possible information. However, when looking for this efficiency, this type of models usually falls into algorithms of great complexity.

5. INDUSTRIAL APPLICATION

There are many fields of science where these techniques can be applied. An example of application in industrial engineering is steel production. A common feature of today's industrial processes is the constant and fast growth of their capacity to store data; that is to say, there are greater volumes of historical records available every day containing information about these production processes. The information reduces uncertainty and, therefore, allows, in general, better decisions to be made. However, as the amount of data stored increases, the capacity to assimilate it decreases, making it necessary to use tools that allow useful knowledge to be extracted from large data sets. Data analysis forms the core of data mining, but the entire process also encompasses issues such as defining the business problem and developing the solution to solve it.

The fact that, in most industrial processes, the relationships among variables are non-linear, and the difficulty in obtaining explicit models that define their behavior, leads to the consideration of data-based models versus analytical models based on explicit equations. Nowadays, non-linear modelling has important techniques that have achieved great applicability thanks to the growing development of processors. These techniques include the algorithms and processes mentioned in Section 4.

Unfortunately, due to the existing conditions in the industries (electromagnetic interferences, current peaks in the starting of motors, the human factor, etc.), erroneous data, defined in the literature as outliers, are very probable among the stored data. Thus, many authors have studied how to work with them: Zhu et al. [36], Knorr [37], Mishra et al. [38], Hampel et al. [39]. For example, Hampel et al. [39] state that routine data contain between 1 and 10% of serious errors and that even the best quality records cannot be guaranteed to be error-free. The presence of outliers in the data set causes a worse, sometimes far from optimal, fit of the model obtained, so it is very important to treat them in some way so that they do not harm the models sought. Normally, one of these two strategies is used: use of outliers diagnostic techniques with which erroneous data are detected and eliminated before proceeding with the construction of the model ([40], [41], [42]); or avoid this type of data pre-processing by using modeling techniques that are robust to outliers ([43], [44], [45]).

The chemical composition of steel, the heat treatment to which it is subjected, as well as the manufacturing process used, define its mechanical properties. Although the basic element that constitutes steel is iron, the addition of small quantities of other elements has a marked effect on the type and properties of steel. In addition, it should be borne in mind that when heat treatments are applied to the material, including cooling at a certain rate from a particular temperature peak, these elements produce different responses. It should not be forgotten either, that the production process uses combinations of heat treatments and mechanical work, which are of critical importance. In this way, and using the manufacture of galvanized steel coils as an example of application, around fifteen variables of the chemical composition of the casting (carbon, manganese, silicon content, etc.), and no less than three process variables from the galvanizing line (among them: the temperature and average speed of the strip) can be included in the study of the process to predict the elastic limit, resistance to breakage and elongation of the coils.

It is important to note that usually not a single Data Mining technique is used in the studies. Thus, in the example of steel manufacturing, we can differentiate phases of analysis and preparation of the data set, data classification, and pattern search, in which, in addition to "traditional" statistical tools, classifiers such as Support Vector Machines or modeling through ANN can be used.

5.1 Case Study in Steel Coils

Authors such as Wang [46] and Sebzalli and Wang [47] used linear projectors such as Principal Component Analysis (PCA). The purpose of Principal Component Analysis is, explained very briefly, to transform an X matrix of n variables into a Y matrix of uncorrelated virtual variables ordered from highest to lowest variance. The selection of factorial axes is carried out in order of "relevance" in terms of information contribution (eigenvalues), so that each axis determined must contribute less and less information. This method makes it possible to reduce the data considerably, replacing the observed variables with a small group of derived variables, but it has some drawbacks, such as the fact that it makes no specific mention of data measurement errors or that the projection on the axes is linear, so that other more complex relationships cannot be inferred.

Projectors and clustering tools are useful for classifying patterns. The problem of modelling in steel manufacturing that explains the operation of the system must be addressed separately for each type of coil. This is due to the fact that the behavior of each coil in front of the furnace depends on its dimensions and on the type of steel of the coil. Therefore, the first point to be studied is whether the types of steel can be grouped into larger groups according to their behavior. To do this, we use different projectors and Data Mining techniques. The points of operation of the galvanizing process are very dependent on the type of hardness of the steel, so that, this could be used as classifier of coils and to continue making the study grouping the coils according to their similarities of metallurgical composition.

By using different projectors and clustering techniques, we try to obtain some objective criteria for classifying coils according to the metallurgical composition of the steels in them. Table I shows an example of a table with the data of only three coils and only the chemical composition of the casting.

Table I

Example of data for chemical composition (‰).

	C	Mn	Si	P	S	Cu	Al	Ni	B	Ti	N	V	Nb	Cr	Ceq
1	2.5	128	6.8	10.1	8.2	13.9	25.7	26.4	0.1	63.8	3.8	2.0	0.9	16	22.5
2	3.2	141	8.1	8.5	7.5	20.1	25.9	17.5	0.1	71.0	3.8	2.3	1.1	21	26.0
3	4.1	131	7.5	8.8	9.8	18.1	31.0	18.9	0.1	67.0	3.4	1.7	0.8	25	23.9

By first using a PCA projector, it is obtained that the intrinsic structure of the data is sufficiently defined with two axes, since the first two axes of the PCA projection cover a high percentage of the variance (more than 95%). The previous way of implementation is quite useful although it is convenient to verify it with clustering algorithms, since it can be used to determine the validity or not of the previous projections. Using for example the algorithm of the K-Nearest Neighbor or KNN, Figure 3 is obtained, where five different groups of coils can be observed.

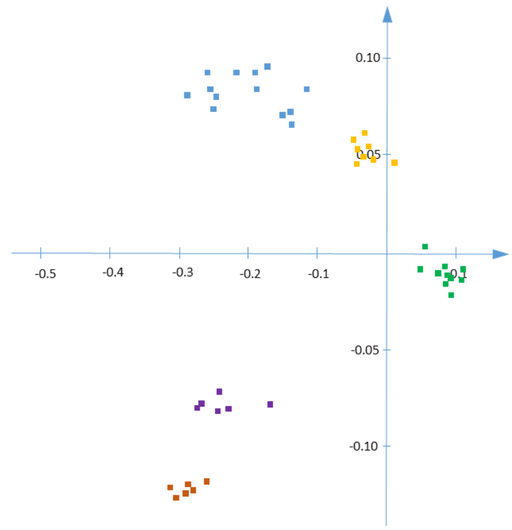

Fig. 3. Distribution of the coils with five different groups

6. CONCLUSIONS

Volume of information is increasing in the last years in an exponential way. Websites, business applications, social networks, mobile devices and increasingly the Internet of Things (IoT) are using for companies, governments and all citizens. All this large amount of data can be useful if properly managed. In this sense, data mining is the automated process of classifying large data sets to identify trends and patterns and establish relationships, to solve scientific or even business problems or generate new opportunities through the analysis of data.

Recommender systems are taking more importance for companies that want to buy their products and also for users that want to find their desires as soon as possible and as easy as possible. These systems provide the users as output a series of products, services or contents that fit their tastes or needs, which have previously been collected by the system. We have analyze briefly some techniques used for this purpose: KNN, SVM, ANN, etc. Among them, regression technique can be the easiest technique to be applied for new researchers, but SVM and ANN are more powerful.

7. REFERENCES

[1] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: *From Data Mining to Knowledge Discovery in Databases*. AI Magazine Volume 17 Number 3, 37–54 (1996).

[2] Liu, X., Zhang, X., Li, X., Sun, Z.: *Research on data mining clustering algorithm in cloud computing environments*. BTAIJ, 10(17), 9562–9566 (2014).

[3] Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: *Data Mining. Practical Machine Learning Tools and Techniques*. 4th edn. Elsevier (2017).

[4] Burke, R.: *Hybrid recommender systems: Survey and experiments. User Modelling and User-Adapted Interaction* 12(4), 331–370 (2002).

[5] Resnick, P., Varian, H.R.: *Recommender systems. Communications of the ACM* 40 (3), 56–58 (1997).

[6] Burke, R.: *Hybrid web recommender systems. In: The Adaptive Web*, pp. 377–408. Springer, Berlin / Heidelberg (2007).

[7] Mahmood, T., Ricci, F.: *Improving recommender systems with adaptive conversational strategies*. In: C. Cattuto, G. Ruffo, F. Menczer (eds.) Hypertext, pp. 73–82. ACM (2009).

[8] Xiao, B., Benbasat, I.: *E-commerce product recommendation agents: use, characteristics, and impact*. MIS Quarterly 31(1), 137–209 (2007).

[9] Lee, T.Q., Park, Y., Park, Y.T.: *A time-based approach to effective recommender systems using implicit feedback.* Expert Systems with Applications, 34(4), 3055–3062 (2008).

[10] Nuñez-Valdez, E.R., Cueva Lovelle, J.M., Sanjuan Martinez, O., Garcia-Diaz, V., Ordoñez de Pablos, P., Montenegro Marin, C.E.: *Implicit feedback techniques on recommender systems applied to electronic books*. Computers in Human Behavior 28(4), 1186–1193 (2012).

[11] Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J.: *Grouplens: applying collaborative filtering to usenet news*. Communications of the ACM 40(3), 77–87 (1997).

[12] Nichols, D.: *Implicit rating and filtering*. In Proceedings of 5th DELOS Workshop on Filtering and Collaborative Filtering, ERCIM, pp. 31–36. (1998).

[13] Desrosiers, C., Karypis, G.: *A comprehensive survey of neighborhood based recommendation methods*. In Recommender Systems Handbook, chapter 4, pp. 107–144. Springer US (2011).

[14] Zanker, M., Jessenitschnig, M.: *Case-studies on exploiting explicit customer requirements in recommender systems*. User Modelling and User-Adapted Interaction 19(1-2), 133–166, (2009).

[15] Jeong, B., Lee, J., Cho, H.: *An iterative semi-explicit rating method for building collaborative recommender systems*. Expert Systems with Applications 36(3), 6181–6186 (2009).

[16] Lee, S.K., Cho, Y.H., Kim, S.H.: *Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations*. Information Sciences 180(11), 2142–2155 (2010).

[17] Koren, Y., Bell, R.: *Advances in collaborative filtering*. In Recommender Systems Handbook, chapter 5, pp. 145–186. Springer US (2011).

[18] Alsaleh, S., Nayak, R., Xu, Y., Chen, L.: *Improving matching process in social network using implicit and explicit user information*. In: Springer-Verlag, Proceedings of the 13th Asia-Pacific web conference on Web technologies and applications, pp. 313–320, Berlin, Heidelberg (2011).

[19] Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A.: *Incorporating contextual information in recommender systems using a multidimensional approach*. ACM Transactions on Information Systems 23(1), 103–145 (2005).

[20] Gao, M., Liu, K., Wu, Z.: *Personalization in web computing and informatics: Theories, techniques, applications, and future research*. Information Systems Frontiers 12(5), 607–629, (2010).
[21] Burke, R.: *Integrating Knowledge-based and Collaborative filtering Recommender Systems*. In Workshop on AI and Electronic Commerce (1999).
[22] Barranco, M.J., Martinez, L.: *A method for weighting multi-valued features in content-based filtering*. In: Springer-Verlag, Proceedings of the 23rd international conference on Industrial engineering and other applications of applied intelligent systems – Volume Part III, IEA/AIE'10, pp. 409–418, Berlin, Heidelberg (2010).
[23] Blanco-Fernandez, Y., Lopez-Nores, M., Gil-Solla, A., Ramos-Cabrer, M., Pazos-Arias, J.J.: *Exploring synergies between content-based filtering and Spreading Activation techniques in knowledge-based recommender systems*. Information Sciences 181(21), 4823–4846 (2011).
[24] Burke, R.: *Knowledge-based recommender systems*. Encyclopedia of Library and Information Systems 69(32), 175–186 (2000).
[25] Burke, R., Hammond, K.J., Young, B.C.: *The findme approach to assisted browsing*. IEEE Expert: Intelligent Systems and Their Applications 12(4), 32–40 (1997).
[26] De Campos, L.M., Fernandez-Luna, J.M., Huete, J.F., Rueda-Morales, M.A.: *Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks*. International Journal of Approximate Reasoning 51(7), 785–799 (2010).
[27] Cover, T.M., Thomas, J.A.: *Elements of information theory*. Wiley-Interscience, New York, NY, USA (1991).
[28] Dantas Bezerra, B.L., Tenorio de Carvalho, F.A.: *A symbolic approach for content-based information filtering*. Information Processing Letters 92(1), 45–52 (2004).
[29] De Campos, L.M., Fernandez-Luna, J.M., Huete, J.F.: *A collaborative recommender system based on probabilistic inference from fuzzy observations*. Fuzzy Sets and Systems 159(12), 1554–1576 (2008).
[30] Adomavicius, G., Tuzhilin, A.: *Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions*. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005).
[31] Franke, N., Schreier, M., Kaiser, U.: *The "I designed it myself" effect in mass customization*. Management Sciences 56(1), 125–140 (2010).
[32] Burke, R.: *Interactive critiquing for catalog navigation in E-commerce*. Artificial Intelligence Review 18(3-4), 245–267 (2002).
[33] Burke, R.: *Hybrid web recommender systems*. In The adaptive web, pp. 377–408 (2007).
[34] Bilbao, J., Bilbao, I., Feniser, C.: *Generalized Delta Rule with Entropy Error Function*. Acta Technica Napocensis - Applied Mathematics, Mechanics and Engineering 60(2), 165–170 (2017).
[35] Bridge, D., Ferguson, A.: *An expressive query language for product recommender systems*. Artificial Intelligence Review 18(3-4), 269–307 (2002).
[36] Zhu, J., Ge, Z., Song, Z., Gao, F.: *Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data*. Annual Reviews in Control 46, 107–133 (2018).
[37] Knorr, E.M.: *Outliers and data mining: finding exceptions in data*. University of British Columbia, (2002). doi: 10.14288/1.0051497
[38] Mishra, S., Chawla, M.: *A comparative study of local outlier factor algorithms for outliers detection in data streams*. In: Abraham, A., Dutta, P., Mandal, J., Bhattacharya, A., Dutta, S. (eds) *Emerging Technologies in Data Mining and Information Security*. Advances in Intelligent Systems and Computing, vol 813. Springer, Singapore (2019).

https://doi.org/10.1007/978-981-13-1498-8_31
[39] Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W.: *Robust statistics, the approach based of influence functions*. Wiley. (1986).
[40] Gupta, M., Gao, J., Aggarwal, C.C., Han, J.: *Outlier Detection for Temporal Data: A Survey*. IEEE Transactions on Knowledge and Data Engineering, 26(9), 2250–2267, (2014). doi: 10.1109/TKDE.2013.184
[41] Castejón Limas, M., Ordieres Meré, J.B., Martínez de Pisón Ascacibar, F.J., Vergara González, E.P.: *Outlier detection and data cleaning in multivariate non-normal samples*. The PAELLA algorithm. Data Mining and Knowledge Discovery, 9, 171–187, (2004).
[42] Rousseeuw, P.J., Leroy, A.: *Robust Regression and Outlier Detection: Diagnostic Regression Analysis*. New York: John Wiley and Sons (1987).
[43] Ahmed, M., Farag, A.: *A neural approach to zoom-lens camera calibration from data with outliers*. Image and Vision Computing, 20, 619–630 (2002).
[44] Chuang, C., Su, S., Hsiao, C.: *The annealing robust backpropagation (ARBP) learning algorithm*. IEEE Transactions on Neural Networks, 11(5), 1067–1077 (2000).
[45] Wang, B., Mao, Z.: *Outlier detection based on a dynamic ensemble model: Applied to process monitoring*. Information Fusion, 51, 244–258, (2019). doi: 10.1016/j.inffus.2019.02.006
[46] Wang, X.Z.: *Data Mining and Knowledge Discovery For Process Monitoring and Control. Advances in Industrial Control*. Ed. Springer. London, (1999).
[47] Sebzalli, Y.M., Wang, X.Z.: *Knowledge discovery from process operational data using PCA and fuzzy clustering*. Engineering Applications of Artificial Intelligence, 14 (2001).

DATA MINING APLICATA PRACTIC IN FABRICAREA BOBINELOR DE OTEL

Abstract. Din cauza cantitatii mari de date care pot fi obținute in present, Data Mining este un important câmp de cercetare care poate fi aplicat in mai multe domenii.Unul dintre aceste domenii sunt sistemele de recomandare care iau un numar mare de date, valori, produse sau caracteristici pentru a obține niște rezultate ca recomandări pentru un utilizator care este interesat intr-un domeniu.Cateodata, Data Mining este stabilită ca una dintre nivelele unui proces mult mai generic numit Knowledge Discovery in Databases. in aceasta lucrare este realizata o analiza a acestor procese, cum sunt realizate, și analiza unor tehnici folosite ca și KNN, Naive Bayes, SVM, ANN, Regresie și Multiclasificare.

Imanol BILBAO, PhD.Student, University of the Basque Country (UPV|EHU), Applied Mathematics Department, Engineering School, imanoles@gmail.com, +34946014151, Pl. Ing. Torres Quevedo,1, 48010, Bilbao, Spain
Javier BILBAO, PhD.Eng., Professor, University of the Basque Country (UPV|EHU), Applied Mathematics Department, Engineering School, javier.bilbao@ehu.eus, +34946014151, Pl. Ing. Torres Quevedo,1, 48010, Bilbao, Spain
Cristina FENISER, PhD.Eng.Ec., Associate Professor, Technical University of Cluj Napoca, Management and Economical Engineering Department, cristina.feniser@mis.utcluj.ro, +40752105451, Augustin Bena St.,14, 510120, Alba Iulia, Romania
Andrei BORSA, PhD.Eng., Assistant Professor, University of Agricultural Sciences and Veterinary Medicine, Food Technology Department, andrei.borsa@usamvcluj.ro, +40740166493,Calea Dorobantilor St.,.99-101, ap.1, Cluj Napoca, Romania