# Fitting data mining to the organization: using validity, utility and novelty in the pattern evaluation phase of data mining

Q.J.J. Meulenberg

*Faculty of Technology, Policy and Management, Delft University of Technology. Jaffalaan 5, 2625BX Delft, The Netherlands*

**Abstract**

An organization's productivity can increase by 6% if they are able to make use of existing data in their decision making process. To exploit this potential, an organization should be able to successfully apply data mining and combine it with decision making. However, in practice it is difficult to interpret data mining results and translate them to action and value for organizations. This research presents a generalizable process for pattern evaluation in organizations, that bridges the gap between data mining and decision making. It presents the quality criteria *validity, utility* and *novelty* to evaluate patterns on, and describes how these criteria should be related to each other. In order to balance objective and subjective decision making and to integrate inductive methods with deductive methods, the process decouples the evaluation phase in three steps, corresponding to the quality criteria. This process functions as a starting point for an integration of data mining and decision making. Further research should be focused on the incorporation of outranking methods for decision making and should expand the focus to the deployment of the discovered knowledge in the organization.

Keywords: data mining; decision making; process design; utility; novelty

## 1. Introduction

In current markets, data is becoming an increasingly important asset to companies. Currently, the world produces 1.7 quadrillion ($10^{15}$) bytes a minute (European Commission, 2014). Estimations of the benefits show that an organization's productivity can increase by 6% if they are able to make use of this data in their decision making process (Brynjolfsson et al., 2011). To exploit this potential, an organization should be able to successfully apply advanced data analytics in their decision making processes.

Data mining (DM) is the buzzword the applications of advanced data analytics on large datasets. It is part of the knowledge discovery process (which is currently interchangeably used with the term 'data mining'; in this report, data mining is used as the process of knowledge discovery through advanced data analytics) (Fayyad et al., 1996; Han & Kamber, 2006). It is a method to structure data to information or *pattern*s. If these patterns are interpretable, the pattern represents knowledge (Natarajan & Shekar, 2009). Because of the increasing magnitude of data and datasets in organizations, data mining is increasingly adopted to for knowledge discovery. It is the method that should enable organizations to reach the bfore mentioned increase in productivity.

The process of data mining is executed in six steps (Fayyad et al., 1996; Han & Kamber, 2006):

- Data selection
- Data preprocessing,
- Data transformation
- Data mining
- Pattern evaluation

After the pattern evaluation phase, a data scientist should have transformed data into knowledge, that is valid, previously unknown and actionable in the organization (Cabena et al., 1999).

However, in practice it is hard to interpret and implement findings of DM research (Yang & Wu, 2006). Several sources conclude that pattern evaluation is too less studied (Berkhin, 2006; Köksal et al., 2011). When applying data mining in organizations, the interpretation problems endanger the potential of successful integration of data mining in organizations.

In order to fill the gap between data mining and pattern evaluation, this paper presents a process that focusses on the interpretation of DM-results in organizations. This method takes into account the necessary requirements for a data mining pattern, as well as the interactions between decision makers and data scientists.

### *1.1. Structure*

This paper is structured as follows. Section 2 gives a short explanation of the design science methodology. In section 3, the previous literature on the interpretation of patterns is examined. that was used to design the pattern evaluation process. Section 4 describes the decoupled process for pattern evaluation. This process is validated in Section 5. Finally, section 6 presents the final conclusions and describes possibilities for future work.

## 2. Methodology

Since the goal of this research was to look into new methods to improve the interpretation of patterns, this research has been of an explorative nature. This research consisted of a literature analysis and a case study at a large bank in The Netherlands (for reasons of confidentiality, this case study is not treated in this article).

For the literature study, data mining has been considered as a socio-technical system. The interpretation of patterns is not only technically difficult, but also in the way that the stakeholders interact with each other and with data mining. Therefore, data mining is considered as a socio-technical system, that can be improved from both the technical perspective as from the process perspective .

In order to combine the insights from the case study and the literature research, the design science research of Hevner and Hevner et al.

has been adopted (2007; 2004). This methodology structures the way in which both findings have their influence on the improving the interpretation of data mining patterns.

**3. Previous literature**

One of the main problems for evaluating DM patterns is the fact that data mining delivers so many results. When so many outcomes result from an analysis, how are the good outcomes selected? This problem is defined as the rule quality problem (Choi et al., 2005).

*3.1. Pattern interestingness*

In order to cope with the rule quality problem, previous literature focuses on measuring the interestingness of patterns (Bong et al., 2014; De Bie & Spyropoulou, 2013; Freitas, 1999; Silberschatz, 1995). As so many patterns result from an analysis, not all patterns are interesting. Therefore, previous efforts have focuses on designing evaluation measures for objective-, subjective- and semantic interestingness. These measures can be applied as a filtering or ranking mechanism (Geng & Hamilton, 2006a).

Objective measures are solely based on data and rely on statistics, probability theory or information theory (Han & Kamber, 2006; Natarajan & Shekar, 2005; Tan et al., 2004).

Subjective measures take do not take into account the data that is used, but the user's beliefs and problem context that examines the data. These measures focus on the current knowledge of the user and try to prevent that already existing knowledge is rediscovered (Bie, 2013; Geng & Hamilton, 2006b; Silberschatz, 1995).

Semantic measures represent the ability to translate the patterns into value for the user or the client.

However, in the current knowledge base, almost no previous research conducted did incorporate all three types of interesting measures. Subjective and semantic interestingness are still just theoretical concepts, since it is unknown how to incorporate them with objective interestingness (Bie, 2013). Perhaps the term interestingness is slightly misleading, as it implies that an overall interestingness can be measured, while in practice the three types of interestingness do not measure the same thing. Therefore, it is almost impossible to jointly evaluate patterns on these requirements.

*3.2. Quality requirements for patterns*

In current literature, the quality requirements for data mining patterns have been well described. Several sets of requirements exists for patterns to be classified as a DM-pattern. In their research for data mining in business, Sharma & Osei-Bryson describe a data mining pattern as ‘valid, novel, potentially useful and ultimately understandable’ (2009). For data mining in a business context, Cabena et al. combine the usefulness and understandability to the actionability of a pattern (Cabena et al., 1999). The problem with these requirements is that they are used as broad terms, but that there operationalization is not taken into account. For instance, the understandability of a pattern is a hard requirement to measure.

In order to redefine the requirements for data mining patterns, they must be related to the key elements of what data mining consists of. Since data mining consists of advanced statistical analysis, patterns must be *statistically valid*. Data mining is an inductive methodology, so it focusses on finding *new information*. Finally, since organizations want to create value, the data mining patterns should be *valuable*. These 3 elements are translated to the following three requirements for a pattern:

- Validity;
- Utility;
- Novelty

These terms seem more applicable than the term interestingness, as it clearly indicates that each quality requirement measures something totally different.

However, since they are so clearly different, these measures create some conflict within the pattern evaluation, namely (see Figure 1) :

- Objectivity vs subjectivity
- Deduction vs induction

*3.3. Objectivity vs subjectivity*

With the introduction of subjective measures, objective nature of data mining and the subjective nature of decision making occurs. In data mining, there is some hesitance to incorporate the user’s preference, as this

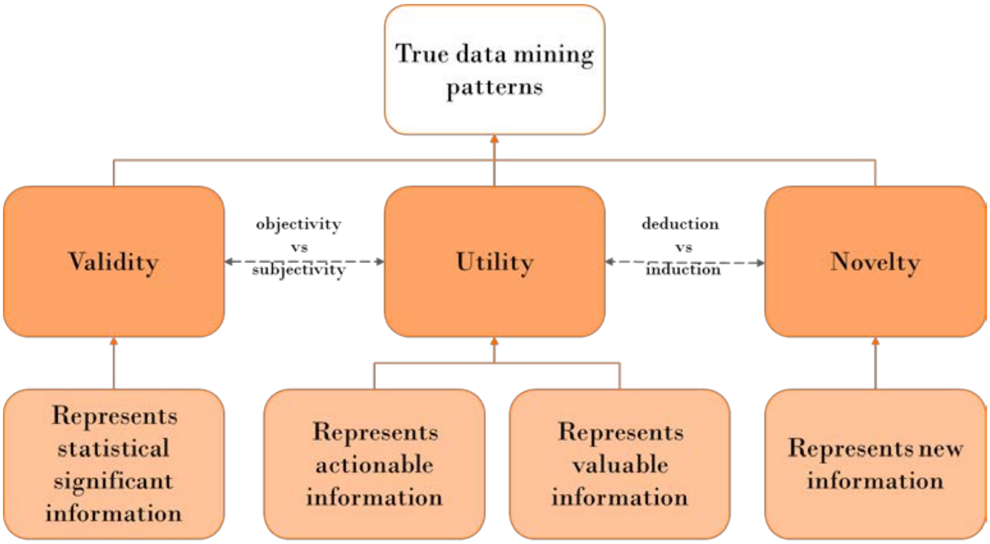

Figure 1: three criteria for data mining patterns: validity, utility and novelty

subjective evaluation may violate the objectivity of data mining.

### *3.4. Deduction vs induction*

When applied in organizations, data mining is combined with regular decision making regimes. There is a discrepancy between these two worlds.

The utility of a pattern consists of its possible value of an outcome to the organization that is should be applied in. In this part, data mining is used for decision making. In order to know the potential value, the pattern must be related to the current knowledge base, otherwise, patterns are not accepted (Boisot, 2004; Potes Ruiz et al., 2014). When incorporating the available knowledge for determining utility of a pattern, the outcomes are automatically converged to the existing knowledge base of the environment.

On the other hand, when limiting yourself to what you already know, the chance to find new knowledge is decreased.

One of the key advantages of DM is that it is an inductive method that does not take into account any pre-existing knowledge. In order to find new insights, the analysis should not be limited to the current knowledge base.

So, if the patterns should be evaluated on utility and novelty, a balance between the objectivity and the subjectivity and between induction and deduction needs to be found. The roles of the different evaluation measures need to be fitted to each other, in order to create valid solution to the rule quality problem.

## 4. Case study

The case study was performed at a Dutch bank, from here on ‘ Bank’. In Bank, there was a goal to sophisticatedly reduce the call flow at the call center. At Bank, it was assumed that clients also made calls as a consequence of their preceding activity on the web site. So, by improving the web site, the call flow can be reduced sophisticatedly. Therefore, online activity of customers that called and have been online in the 24 preceding hours is mined.

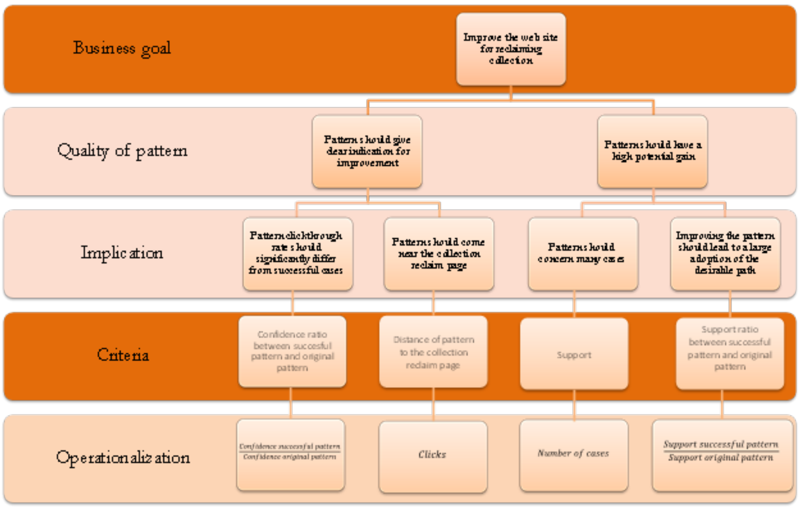


**Figure 2: the deduction of utility criteria**

### 4.1. *Setup*

This case consisted of a input dataset of 10.000 cases, of which full patterns and subpatterns could be mined. All the input data concerned clients that called to retransfer a direct debit. While this is a case of mining longitudinal association rules, the sequence mining method has been applied using the 'TraMineR' package (Gabadinho et al., 2011, 2015)

### 4.2. *Validity evaluation*

In this case study, the three evaluation criteria had been taken into account each in a separate evaluation step. First, the validity evaluation has been performed. After a first scan of the data, it was recognized that the data was scattered around 1448 unique pages. Besides, nearly two-third of the input patterns were unique. In this way, the validity critierion had to be set relatively low. Patterns that occurred more than 30 times were taken into account.

### 4.3. *Utility evaluation*

With respect to the utility evaluation, the value of a pattern was defined by its support in the dataset. The actionability of a pattern was defined by how clear the pattern differed from a dataset were people successfully retransferred a direct debit. In this way, the business could conclude that the type of behavior was significantly different from the desired behavior. Figure 3 shows the breakdown of the utility criteria.

### *4.4. Novelty evaluation*

The novelty evaluation is targeted to find latent constructs that may have influence on the utility value of a pattern. To this end, the patterns are divided in a high-utility group and a low utility group. When a latent construct in the high-utility group does not exist in the low-utility group (or the other way around), a possible explanation for the utility value of a pattern is found.

These latent constructs are approached from a totally different perspective, in order to broaden the view of the utility evaluation. In the case study, Bank did not approach the novelty evaluation from the process perspective, but from the client perspective. In this way, a theory could be formed on what types of customers' could benefit from Bank's web site improvement.

The setup of the novelty evaluation is an exploratory factor analysis (EFA) with difference between groups (Costello & Osborne, 1994). Exploratory factor analysis is a variable reduction method that identifies the number of latent constructs and the underlying factors of the input variables. The methods does so by taking account of the common variance of the variables in the data set (Suhr, 2006). In more practical terms, EFA groups variables that together seem to measure the same object, that is not specified in the dataset.

### *4.5. Results*

With the large amount of pages that have been visited at Bank's web site, the theoretical amount of patterns was extremely high: $1.43 \cdot 10^{27}$.* After the validity evaluation, 59.063 patterns remained. This is a reduction of more than 99% of the theoretical amount possible. Although the theoretical maximum is just used as an approximate of all the subpatterns in the dataset, it gives an indication on the need for validity evaluation.

The utility evaluation was used to prune patterns on their existence in the 'successful'-dataset. Moreover, the patterns were ranked on their overall utility score. The pruning had a large effect on the size of the set of resulting patterns. Nearly 96% of the valid patterns were exempted from the utility evaluation as a result of the pruning step.

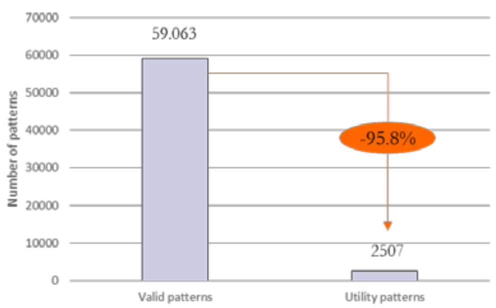


**Figure 3: the effect of utility pruning**

| Variable | High-utility factors | | | | Low-utility factors | |
|---|---|---|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor A | Factor B |
| Time spent in closed payments sector | + | | | | + | |
| Time spent in closed environment | | | + | | | |
| Internet visits of the past 3 months | + | | | | + | |
| Internet visits of the past month | + | | | | + | |
| Average login/month | | | - | | | |
| App use /month | | | | + | | |
| Logins of the past 12 months | + | | | | + | |
| Number of direct debits | | + | | | | + |
| Amount of bank transfers | | | | + | | |
| Age | | | + | | | |
| Value of planned transactions | | + | | | | + |

**Figure 4: the factor solutions of novelty evaluation**

In the novelty evaluation, the exploratory factor analysis proved a distinction on two factors between the high-utility pattern set and low-utility pattern set: the *app-use* and the *inexperience with the online environment.* This gave the team a preliminary insight that changes in the app may be needed as well.

**5. Design: The decoupled process for pattern evaluation**

In order to tackle these problems and to cope with the rule quality problem, a new process has been designed that deliberately takes validity, utility and novelty into account. This process enables data scientist to incorporate other evaluation measures in a simple and valid way. In this way, a balance between objective and subjective evaluation is found, as well as between induction and deduction.

In order to secure the adoption of other measures than statistical measures, the evaluation steps are as much decoupled as possible. They still have dependencies, but each evaluation step has a clearly different goal.

* This is calculated by the $\sum_{k=1}^{4} \binom{1448}{k}$ combinations of pages (following the nCr principle).

### *5.1. Validity evaluation*

Validity is the label of the traditional methods to evaluate DM outcomes. Validity measures are often based on frequency values. By incorporating validity measures, data mining delivers patterns that are significant, so that the researcher can make conclusions based on the patterns. Data mining outcomes are then scientifically acceptable.

From both the data mining perspective as the business perspective, validity evaluation is a logical step. Validity does not include user- or problem specific dynamics in the evaluation, so the results of the evaluation phase are objective observations. Validity evaluation is used as the first filter of patterns and is treated as a constraint. If a pattern does not satisfy the validity constraint, it is not taken into account for the consequent phases. This seems logical, since if a pattern is not statistically true, it should represent no value to the decision maker.

### *5.2. Utility evaluation*

The utility evaluation is decoupled from the validity evaluation, in order to separate the objective steps from the subjective steps. In this process, the constraint that merely statistic measures should be used is released, resulting in more effective ways to define utility (Choi et al., 2005). This exact definition is however user dependent, so lower level prescriptions for this evaluation approach are not possible.

Patterns that have been judged as valid, serve as input for the utility evaluation. In the utility evaluation phase, patterns are judged on their potential contribution to attaining a certain business goal. Since it is decoupled from the validity evaluation, utility does not have to be defined statistically (Jagannath, 2003) (Braynova & Pendharkar, 2005). This offers decision makers more freedom to specify what is relevant to them.

This freedom is needed, because there is no generalizable utility function that can be applied on every data mining project. Since the utility of a pattern is dependent of the user, the user needs to specify a utility function. This utility function needs to serve to subcriteria: the actionability of a pattern and the value of a pattern to the organization (see Figure 1).

In order to maintain the objective determination of pattern utilty, the subjective evaluation should be carried out in an objective process. To this end, the Analytical Hierarchy Process, as developed by Saaty (1990), is applied on the utility evaluation. In this process, decision makers have to specify criteria and their weights before the data mining takes place. In this way, the evaluation process is done objectively, but with subjective information. By establishing criteria, decision makers are forced to make their user preferences explicit, leading to a transparent and objective process. Besides, the AHP partly balances induction and deduction, since the alternatives are acquired through induction, but tested with deduction.

At the end of the utility evaluation, every valid pattern now also represents a certain utility value. Since the lower-utility patterns are needed for the novelty evaluation, it is advisable to use utility evaluation as a ranking technique instead of pruning.

*5.3. Novelty evaluation*

Although AHP provides a way to incorporate induction with deduction, an extra inductive step is added to explicitly look for new insights: novelty evaluation. In this research, three types of novelty have been indicated:

1. Patterns that not have identified before and may contradict existing hypotheses or form new hypotheses;
2. New knowledge that can be obtained by reflecting on the insights, gained from validity and utility evaluation, from a different perspective;
3. Patterns that can be obtained by reflecting on a large set of high-utility patterns (meta-novelty).

With the first type of novelty, a data scientist looks in the initial dataset for patterns that may contradict the knowledge of the decision maker. These patterns provide new knowledge on the same level of abstractness as the high-utility patterns do. With this type of novelty, the friction between utility and novelty is not adequately dealt with.

The second type of novelty takes the lessons learned from the previous validation steps as starting point. The starting point then is a set of valid patterns, all with a utility score (step 1 in Figure 5). The data scientist can review the two sets of patterns from a totally different perspective (step 2 in Figure 5), in order to see if there is a general rule that can help the decision makers build new hypotheses (step 3 in Figure 5).

Meta-novelty is the search for higher level knowledge. When data mining is performed a number of times on a number of processes, two large sets of patterns can be identified: a set of high-utility patterns and a set of low utility patterns. When sufficient patterns are acquired, it becomes interesting to reflect on the patterns. To do this, the initial discovered patterns (that result from the utility evaluation) serve as input for a higher-level

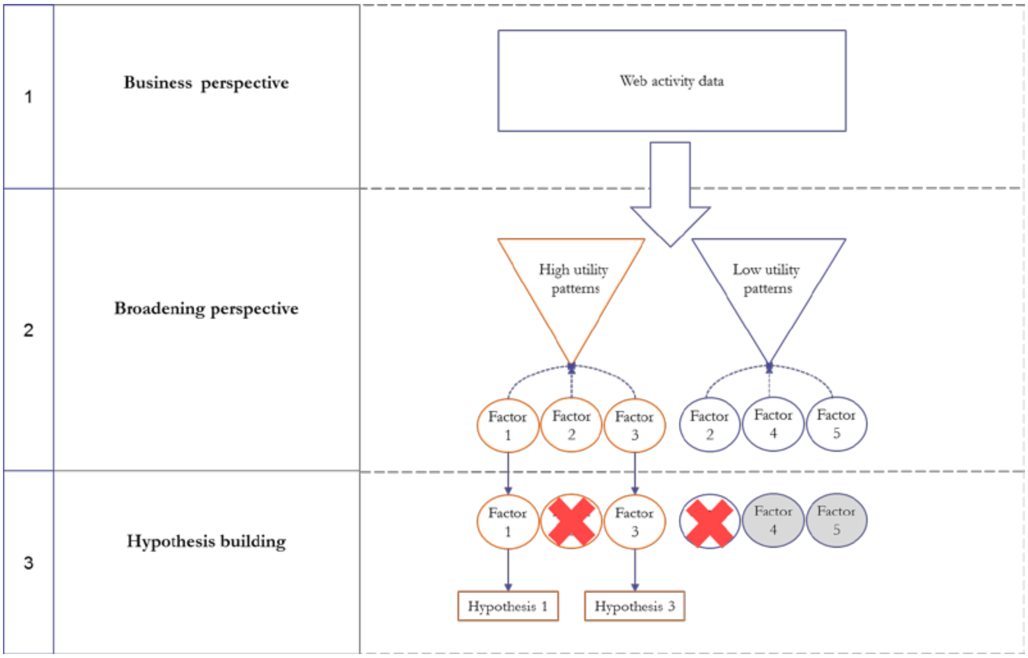


**Figure 5: novelty evaluation as an additional learning step after utility evaluation**

data mining project. Novelty evaluation is then a new exercise to discover a pattern within the patterns.

However, this search for new insights should still be useful, as a novel insight should still be useful to be deployed in the organization. Therefore, the novelty evaluation is targeted at finding an additive explanations for patterns having a high utility value. In principle, why some patterns represent value and some do not, is because to score different on the predefined criteria. However, when looking inductively at these observations, a general rule may possibly be a root cause for a pattern being valuable or relevant.

*5.4. Decoupling of the process*

Where validity evaluation aims at delivering statistically valid patterns, utility evaluation focusses on action and novelty evaluation focusses knowledge discovery. Every type of evaluation measure, measures something completely differen.t Moreover, validity and novelty evaluation is dominated by the DM-perspective, while utility evaluation is dominated by the business perspective. This, and the fact that no such algorithm exists that combines the three evaluation steps, makes it hard to integrate the three evaluation steps in one single evaluation.

Firstly, validity of a pattern should in all cases be treated a hard constraint. Non-valid patterns should be excluded from further analysis, so only patterns are selected that are sufficiently supported by the data. If validity of a pattern was not a hard constraint, a decision maker can design improvements while not supported by the data. Validity evaluation is a form of inductive reasoning, as it aims to find regularities based on observations. In this case, the outcomes are not yet influenced by the decision maker's preferences.

Utility evaluation is used as a way to reduce the rule quantity and –quality problem, by focusing on the patterns that are valuable for the business. The patterns resulting from the validity evaluation are pruned or ranked according to the predetermined criteria of business value. This leaves a set of high-utility patterns and a set of low-utility patterns. By using criteria, a decision maker is able to deduct patterns that are relevant for him. Since the criteria are predetermined, the user preference does not violate the objective nature of pattern evaluation.

Novelty evaluation should take into account both the results from the validity and utility evaluation. In organizations, a valid and novel insight is still useless if it has a low utility value. In that case, the analysis for new information needs to be more focused than just looking for new patterns in the total database. Therefore, the goal of novelty evaluation is that it should be an additive value to the outcomes of the previous analyses.

**6. Discussion**

In order to evaluate the design, the improvements are tested along the non-functional criteria of a data mining framework, as indicated by Meulenberg (2015) :

*6.1. Interactivity*

Since the utility evaluation is carried out through the Analytical Hierarchy Process, the evaluation criteria have to be thought of beforehand. Since the decision makers are most influential in the utility evaluation, they have to commit earlier on in the process and discuss with data scientists. Moreover, the decoupling of the evaluation steps serves as a turn based evaluation of the one on the other’s work.

*6.2. Simplicity*

As described in the above, the different type of interestingness measures were unable to deploy at once on the dataset because that was too complex. By decoupling the evaluation steps, the process becomes simpler. In this way, the evaluation every time serves one goal, instead of serving multiple goals at a time.

On the other hand, a very preliminary academic idea is that the data scientist should develop storytelling skills (Cunningham, 2015). In this way, the total process could be carried out by the data scientist. This whil in every case lead to simpler interactions, and possibly it will provide a simpler pattern evaluation. However, the concept of storytelling skills is brand new, so its potential and implications needs to be scientifically defined.

*6.3. Generalizability*

Since DM consists of a broad variety of methods and techniques, it is unlikely that a single evaluation measure for every type of DM exists. Therefore, the validity, utility and novelty evaluation cannot be further specified without losing on generalizability. However, the criteria are more detailed than in previous research. However, in the case study validation and the expert interview, the generalizability of these concepts have been confirmed (Cunningham, 2015).

*6.4. Scalability*

Since the evaluation becomes more time consuming than before, the scalability of this method is lower than for traditional data mining evaluation methods. However, combined with clustering or with an outranking mechanism, the evaluation phase may become more scalable.

*6.5. Objectivity*

The objectivity of this process is confirmed, although subjective information is used in the evaluation. Since this information is decoupled from the statistic information, valid DM patterns are still specified objectively. On the other hand, utility evaluation is per definition not objective, since decision makers are biased and have their own goals and interests. Since utility criteria are determined beforehand, the evaluation for something as subjective as utility is still done in an objective way.

However, the utility evaluation is also subjective because of the incorporation of weight factors. Although the utility criteria are pairwise compared to a establish the weight factor, the outcomes of this phase are sensitive to changes in the weight factors.

## 7. Conclusions and recommendations

In this report, a new pattern evaluation process is presented. This process takes into account the requirements for a data mining pattern, as well as the different desires of the data scientists and decision makers. This process effectively copes with the rule quality problem, while balancing the objectivity and subjectivity and the inductive nature of data mining with the deductive nature of decision making. The process can function as a starting point to fully integrate data mining in organizational decision making processes.

Two elements that have been derived from this research are proposed as future research. Firstly, the notion of a data scientist as a storyteller has been introduced as a possible new solution (Cunningham, 2015). However, this concept has not yet been scientifically researched. Further research can be conducted to develop a theory on this, for the notion to be regarded as a fully-fledged alternative for bridging the gap of pattern evaluation. Secondly, this research presents a new way of looking at the patterns: from a validity, utility and novelty perspective. Several types of novelty have been indicated in this research. Although well-founded, this is a first discovery of novelty in the evaluation phase of data mining. Further research on the novelty of patterns can really position the use of novelty evaluation in the data mining process.

## References


Berkhin, P. (2006). A Survey of Clustering Data Mining. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.), *Grouping Multidimensional data* (pp. 25–71). Berlin: Springer Berlin Heidelberg. doi:10.1007/3-540-28349-8_2

Bie, T. De. (2013). Subjective Interestingness in Exploratory Data Mining, 19–31.

Boisot, M. (2004). Data , information and knowledge : have we got it right ?, (February).

Bong, K. O. K. K., Joest, M., Quix, C., Anwar, T., & Manickam, S. (2014). Selection and aggregation of interestingness measures: a review. *Journal of Theoretical and Applied Information Technology*, *59*(1), 146–166.

Braynova, E., & Pendharkar, H. (2005). Frequent Pattern Mining with Preferences – Utility Functions Approach. In M.-S. Hacid, Z. W. Ras, & S. Tsumoto (Eds.), *Foundations of Intelligent Systems: 15th International Symposium Ismis 2005* (pp. 364–372). Saratoga Springs, New York, USA: Springer.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? SSRN Electronic Journal*. Cambridge, Massachusetts. doi:10.2139/ssrn.1819486

Cabena, P., Choi, H. H., Kim, I. S., Otsuka, S., Reinschmidt, J., & Saarenvirta, G. (1999). *Intelligent Miner for Data Applications Guide*. San Jose, California.

Choi, D. H., Ahn, B. S., & Kim, S. H. (2005). Prioritization of association rules in data mining: Multiple criteria decision approach. *Expert Systems with Applications*, *29*(4), 867–878. doi:10.1016/j.eswa.2005.06.006

Costello, A. B., & Osborne, J. W. (1994). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 27–29. doi:1531-7714

Cunningham, S. W. (2015). Validation interview. Delft.

De Bie, T., & Spyropoulou, E. (2013). Subjective Interestingness in Exploratory Data Mining: recent insights and challenges ahead. In *An Information Theoretic Framework for Data Mining Categories and Subject Descriptors*.

European Commission. (2014). European Commission and data industry launch € 2 . 5 billion partnership to master big data Organisation of the PPP. Retrieved from http://europa.eu/rapid/press-release_IP-14-1129_en.htm

Fayyad, U., Piatetsky-shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37–54.

Freitas, A. A. (1999). On rule interestingness measures. *Knowledge-Based Systems*, *12*(5-6), 309–315. doi:10.1016/S0950-7051(99)00019-2

Gabadinho, A., Ritschard, G., Studer, M., & Nicolas, S. M. (2011). *Mining sequence data in R with the TraMineR package :* Geneva.

Gabadinho, A., Studer, M., Muller, N., Buergin, R., & Ritschard, G. (2015). *Package " TraMineR ."*

Geng, L., & Hamilton, H. J. (2006a). Interestingness measures for data mining. *ACM Computing Surveys*, *38*(3), 9–es. doi:10.1145/1132960.1132963

Geng, L., & Hamilton, H. J. (2006b). Interestingness Measures for Data Mining: A Survey. *ACM Computing Surveys*, *38*(3), 9–es. doi:10.1145/1132960.1132963

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (second). San Franciso: Elsevier. doi:10.1016/B978-0-12-381479-1.00001-0

Hevner, A. R. (2007). A Three Cycle View of Design Science Research A Three Cycle View of Design Science Research. *Scandanavian Journal of Information Systems*, *19*(2), 87–92.

Hevner, A. R., March, S. T., & Park, J. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. doi:10.2307/25148625

Jagannath, S. (2003). *Utility Guided Pattern Mining*. Raleigh, North Carolina, USA.

Köksal, G., Batmaz, İ., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications*, *38*(10), 13448–13467. doi:10.1016/j.eswa.2011.04.063

Meulenberg, Q. J. J. (2015). *Towards successful data mining in organization: Developing a framework to successfully apply data mining in a business context*. Delft.

Natarajan, R., & Shekar, B. (2005). Interestingness of association rules in data mining : Issues relevant to e-commerce. *Sãdhanã*, *30*(June), 291–309.

Natarajan, R., & Shekar, B. (2009). Interesting Knowledge Patterns in Databases. In J. Erickson (Ed.), *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1652–1666). Hershey, PA: IGI Global. Retrieved from https://books.google.nl/books?id=Nn3jBBZSk9cC&pg=PA1653&lpg=PA1653&dq=solution+to+%22rule+quantity%22+problem&source=bl&ots=lbVs4FZNAU&sig=3vg0c0Ka4VSeoQJtt6NwnFaw1Pw&hl=nl&sa=X&ei=H78vVbnmOo35aJTxgWA&ved=0CCEQ6AEwAA#v=onepage&q=solution to %22rule quanti

Potes Ruiz, P. A., Kamsu-Foguem, B., & Grabot, B. (2014). An Interactive Approach for the Post-processing in a KDD Process. In D. K. Bernard Grabot, Bruno Vallespir, Gomes Samuel, Abdelaziz Bouras (Ed.), *Advances in Production Management Systems: Innovative and Knowledge-Based Production Management in a Global-Local World:* (pp. 92–100). Ajaccio, France: Springer. Retrieved from https://books.google.nl/books?id=6FxbBAAAQBAJ&dq=data+mining+comprehend+decision+maker&hl=nl&source=gbs_navlinks_s

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*, 9–26. doi:10.1016/0377-2217(90)90057-I

Sharma, S., & Osei-Bryson, K.-M. (2009). Framework for formal implementation of the business understanding phase of data mining projects. *Expert Systems with Applications*, *36*(2), 4114–4124. doi:10.1016/j.eswa.2008.03.021

Silberschatz, A. (1995). On Subjective Measures of Interestingness Discovery in Knowledge Bell Laboratories Measures. In *KDD-95 proceeedings* (pp. 275–281). AAAI.

Suhr, D. (2006). Exploratory or confirmatory factor analysis? *Statistics and Data Analysis*, 1–17. doi:10.1002/da.20406

Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, *29*(4), 293–313. doi:10.1016/S0306-4379(03)00072-3

Yang, Q., & Wu, X. (2006). 10 Challenging Problems in Data Mining Research. *International Journal of Information Technology & Decision Making*, *05*(04), 597–604. doi:10.1142/S0219622006002258