What are my Users Looking for when preparing a Big Data Campaign

Claudio Ardagna, Paolo Ceravolo, Guido Lena Cota, Muhammad Muneeb Kiani
Department of Computer Science
Univesità degli Studi di Milano, Italy
Milan, Italy
e-mail: {claudio.ardagna, paolo.ceravolo, guido.lena, muhammad.kiani}@unimi.it

Ernesto Damiani
CINI, Consorzio Interuniversitario Nazionale per l'Informatica
Rome 00198, Italy
EBTIC, Khalifa University
Abu Dhabi, UAE
e-mail: ernsto.damiani@ kustar.ac.ae

Abstract— **This paper describes the process to elicit and classify the requirements of the TOREADOR Big Data platform. The paper provides an overview of the analysis performed on the general requirements related to project goals, models' definition, and management, as well as on the legal aspects of a Big Data Campaign. The final aim is offering a proposition on the aspects that today users perceive as innovative for a Big Data platform.**

Keywords;KANO Model, Requirement Specification, Big Data

I. INTRODUCTION

In the last few years, many companies and organizations in Europe have developed a deep awareness of the opportunities for leveraging Big Data Analytics (BDA) within their organizations. Still, many of them now realize that lack of competencies, skills, and budget may lead to unsatisfactory results and even costly failures. The TOREADOR project aims to overcome this hurdle by providing a Model-based BDA-as-a-Service framework (MBDAaaS), which will support organizations lacking in Big Data expertise in managing Big Data analytics. The TOREADOR framework includes customizable models of the entire BDA process and its artifacts [1].

As known, the first activity in the development of a new software system or technology is understanding the problem domain and what the stakeholders' needs are [2]. Classifying requirements according to their potential in eliciting stakeholders' satisfaction or dissatisfaction has been proven very useful for guiding the development phase of new systems or products [3] [4] [5]. In fact, based on this classification, system developers can prioritize the requirements to fulfill and take appropriate trade-off decisions between design alternatives. The Kano Model [6] is one of the most popular tools for understanding stakeholders' expectations and categorizing requirements accordingly. This paper reports the results achieved by TOREADOR in adopting an early involvement of users during requirement elicitation and can provide interesting insight into the priorities that drive Small and Medium-sized Enterprises (SMEs) in the set-up of a Big Data campaign. In particular, our results offer a standpoint to identify features that today users consider innovative for a platform supporting the design of a BDA campaign.

This paper is structured as follows: Section II gives an overview of the TOREADOR project and how requirement elicitation phase has been conducted. Section III describes some terminologies and processes used in requirement elicitation phase. Section IV provides an overview of Kano model, this section explains how this Kano model has been used in TOREADOR project along with a summary of requirements elicited. Also, DuMouchel methodology has been discussed in this section. Then, Section V summarizes results and provide an analysis on results achieved through the application of Kano Model and DuMouchel methodology. Finally, Section VI comprises of conclusion.

II. REQUIREMENTS CLASSIFICATION IN TOREADOR

A. TOREADOR Project Overview

The TOREADOR framework includes customizable models of the entire BDA process and its artifacts. These models direct and largely automate the configuration, provision, and execution of an architectural framework to carry out the BDA, i.e., the TOREADOR platform. Specifically, the TOREADOR platform is based on the model transformation from declarative models, specifying the goals of a Big Data campaign, to procedural models, specifying how analytics should be parallelized and executed. Such procedural, platform-independent, model is finally translated into provisioning of computational resources on actual execution platforms. The semi-automatic approach of the TOREADOR platform enables seamless transformations between these models and easy deployment of a fully operational Big Data analytics tool [1].

TOREADOR also addresses the regulatory barrier by pre-dealing with the key legal aspects related to data sharing in multi-party and outsourcing scenarios on behalf of its customers. Effectiveness and enforceability of contractual agreements are crucial for fostering a BDA-as-a-Service approach, particularly in multi-party scenarios regarding the European laws. Another major issue is the responsibility and liability of Big Data Analytics hosts/providers. TOREADOR aims to develop well-defined and internationally recognized standards, which will potentially reduce the possible controversy due to the different legal and regulatory compliance requirements existing in different EU countries.

The evaluation of the TOREADOR project involves the deployment of its MBDAaaS framework in four real in-production scenarios (also referred as pilots in the remainder of this paper), covering key areas of the EU marketplace that until now have been only marginally involved in BDA, namely, on-demand security analysis of application data, on-demand analysis of energy production and delivery, aerospace products manufacturing and maintenance, and clickstream analysis of web e-commerce applications.

B. *Process planning*

In this section, we describe the process that supports the elicitation and classification of stakeholders' requirements for the TOREADOR project. This process occurs in the four phases shown in Fig. 1, namely, the elicitation of the requirements on the BDA-as-a-service approach supported by TOREADOR, the construction of the Kano questionnaire that enables their classification, the administration of the questionnaire to a selected group of stakeholders, and the analysis of the results of the completed questionnaires. The process involves the participation of all the industrial partners of the TOREADOR project, under the supervision of the Italian consortium for informatics CINI. In particular, due to the complex nature of the Big Data domain, the stakeholders who participated in the questionnaire were not selected from a generic audience but from the organizations that expressed a special interest in TOREADOR, so as to involve people with enough expertise and competence to understand both the methodological and technological implications of the project.

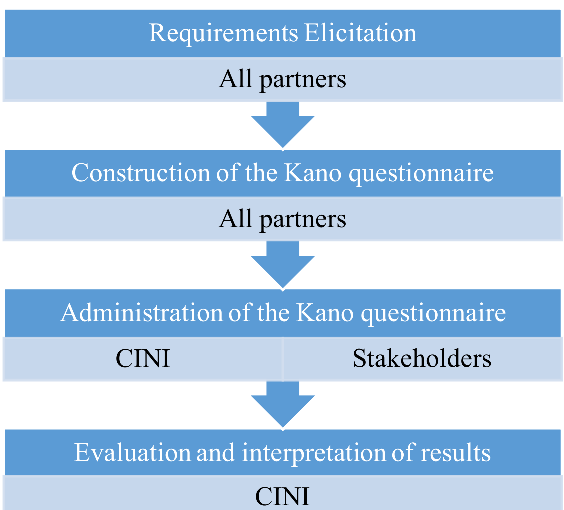

Figure 1: Phases and actors of the Requirements Elicitation and Classification Process

C. *Early involvement of users*

Usually, the requirements elicited from stakeholders are classified into functional and non-functional categories, without providing any information about the stakeholders' opinion about their importance. In [7] authors highlights benefits and challenges of users' involvement in the entire development process and suggests early user involvement is directly linked with quality of requirements and end-product. By involving users'/stakeholders early in requirement elicitation by classifying and prioritizing requirements based on their input can be a good trade off because due to the complexity and diverse nature of a Big Data platform it is a very complex task to keep users engaged during the entire process.

III. Requirements Elicitation

Requirements elicitation is the first and most critical step in the process of requirement engineering. Indeed, correct and complete requirements lead to the success of a system, whereas ambiguous and wrong requirements may result in its failure [2]. Several techniques have been proposed for requirement elicitation [8], such as, for example, interviews, focus groups, questionnaires, document analysis, domain analysis, and requirements workshops.

In the TOREADOR project, the goal of the requirements elicitation phase is to identify a strategic set of guiding requirements that the TOREADOR platform should fulfill. By guiding requirements, we refer to requirements intended to shape the project aim and means without entering on low-level details about the implementation of specific functionalities or technological infrastructures. The phase started with the collection of a draft list of requirements, based on the contribution of each project partner. Note that although no particular elicitation technique was imposed to partners, they mainly used domain analysis by capitalizing on their specific knowledge.[1] For the sake of consistency, all the partners were requested to express the elicited requirements using a uniform specification format, which includes the following information:

- *Name*: a code to identify the requirement.
- Property: a property or a function that should be possibly supported by the TOREADOR platform.
- *Rationale*: an explanation of why the requirement is needed.
- *Scope*, *Source*, and *Target*: references to the work areas, work packages, and target deliverables of the TOREADOR project that are related to this requirement.
- *Priority*: the Kano category assigned to classify the importance of this requirement (see Section IV.A).
- *Dependencies*: a list of other requirements that have some relationship with this one, such as similarities, pre-requisites, or conflicts.

The TOREADOR partners provided all the information but the Dependencies and the Priority, which are assigned by CINI at the end of the requirements elicitation and of the classification process, respectively. Once the draft list of requirements is completed, the partner CINI proceeded with their refinement (e.g., providing a homogeneous style and vocabulary) and integration (e.g., merging similar requirements, eliminating redundancies). As an example, Fig.

[1] Domain analysis can elicit requirements of an existing system by studying available documentation (e.g., business plans, market studies, research studies) and previous applications to identify relevant information.

2 presents the final specification of a requirement identified by the partner CINI.

Overall, the requirements elicitation phase of the process shown in Fig. 1 resulted in a consistent and structured dataset of 125 requirements on the MBDAaaS approach supported

NAME: CINI.REQ00
PROPERTY: Declarative models shall adopt a business−oriented perspective, letting the User focusing on her goals.
RATIONALE: Supporting a model−based approach for the set−up and management of a Big Data Analytics process.
SCOPE: TOREADOR Platform
SOURCE: WP1
TARGET DELIVERABLE: D1.1
PRIORITY: (to be assigned)
DEPENDENCIES: RELATE TO CINI.REQ01−CINI.REQ04

Figure 2: Example of a requirement elicited by the partner CINI.

by TOREADOR. These requirements were organized into the six categories listed below:

1. PRELIMINARIES: requirements on the high-level TOREADOR objectives and fundamental characteristics.
2. MODEL: requirements related to the definition, use and transformation of the models driving the Big Data process and BDA-as-a-Service in TOREADOR.
3. INFRASTRUCTURE: technical and functional requirements of the integrated TOREADOR framework.
4. SLA: requirements focusing on assurance aspects and security. Notably, this part investigates how to deal with the negotiation, monitoring, and assessment of SLAs for Big Data Analytics (e.g., privacy, timing and accuracy concerns) as well as BDA-as-a-Service assurance (e.g., Big Data opacity, diversity, security, and privacy compliance).
5. LEGAL: legal and regulatory requirements existing in different EU countries for Big Data models and process activities (e.g., privacy and data protection).
6. Pilot specific requirements: requirements related to pilots that TOREADOR is developing in the co-design.

IV. Requirements Classification

A. Overview

The output of the requirements elicitation phase is a consistent and structured dataset of stakeholders' requirements. In this section, we describe how to prioritize these requirements using the Kano model.

The stakeholders' satisfaction has become a key factor for the success of a product (e.g., a new service or software system) [5]. Satisfaction is related to the achievement of stakeholders' needs, and thereby to the presence and performance of certain stakeholders' requirements in the product. The Kano model [6] is a useful tool for understanding needs and expectations of a stakeholder based on how they affect this/her satisfaction with a given product (e.g., a new software system) to achieve stakeholders' satisfaction.

The Kano model classifies requirements based on their location along two dimensions, namely, the degree of satisfaction and the level of functionality. The degree of satisfaction goes from total satisfaction (also called delight and excitement) to total dissatisfaction (or frustration). On the other hand, the level of functionality goes from fully dysfunctional, when no functionality is provided, to fully functional, when the best possible fulfillment of the requirement is delivered. These two dimensions are the basis of the Kano Model to classify requirements depending on how stakeholders feel about the provided level of functionality. In his paper [6], Kano identifies the four categories depicted in Fig. 3 and described below.

- Must-be requirements are basic requirements of a product. If these requirements are not achieved, the stakeholder will be severely dissatisfied and not interested in the product at all.
- One-dimensional are those for which the level of functionality is proportional to the degree of satisfaction the better a requirement is achieved, the higher the stakeholder will be satisfied, and vice versa.
- Attractive requirements are usually unexpected by the stakeholders but have the greatest influence on how satisfied they will be with the achievement of a given requirement. As the level of functionality achieved by these requirement increases, the stakeholder's satisfaction increase more than proportionally. Conversely, if an attractive requirement is not met, there is no feeling of dissatisfaction. Attractive requirements are also called exciter, delighter, or added value.
- Indifferent requirements are those which their presence (or absence) does not affect the reaction of the stakeholder to the product.

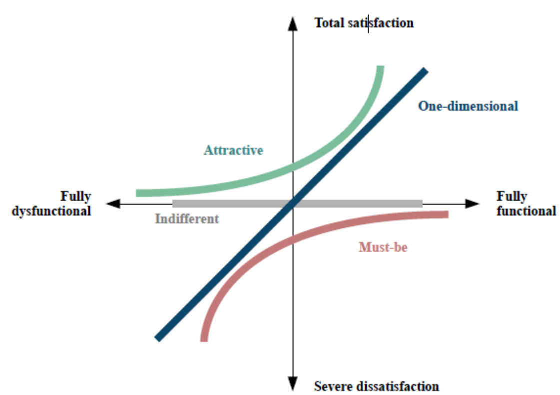

Figure 3 : The Kano model of stakeholder satisfaction

The Kano's classification of requirements can provide decision support to product design. In particular, product designers must ensure that all the must-be requirements are met. However, once a satisfactory level of functionality is reached, it is not necessary to keep investing in must-be requirements, as the effort does not translate into a substantial increase in the stakeholder satisfaction. A good performance of the one-dimensional requirements is instead essential to stay competitive with market leaders, whereas providing some attractive features allows differentiating the product from competitors. Note that, even a limited achievement of an attractive requirement will induce satisfaction. Kano's categories also provide valuable guidance in trade-off situations during the product development stage [3] [9]. For example, if two or more requirements cannot be achieved in the same product release due to technical or financial reasons, their category will indicate which one should be prioritized.

B. *Kano questionnaire*

The Kano model is constructed using a survey methodology, whereby requirements are first classified at the individual stakeholder level through a questionnaire and then aggregated. The Kano questionnaire (KQ) contains a list of question pairs for each product requirement. The question pair includes a functional question, which asks how the respondent would feel if a certain requirement is met, and a dysfunctional question, which asks how the respondent would feel if a product fails to achieve that requirement, Fig. 4 depicts a Kano question.

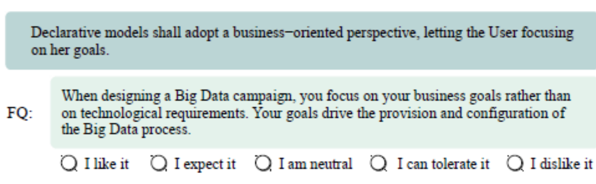

Figure 4: Kano Sample Question

The questions must be understandable by the intended respondents, and, therefore, they require an appropriate and unambiguous terminology to answer each part of the question, the respondent can choose one of five different options, which are proposed below (the different wordings are taken from [10]):

- "I like it" (or "I like it that way", "This would be helpful to me");
- "I expect it" (or "It must be that way", "This is a basic requirement to me");
- "I am neutral" (or "I do not care", "This would not affect me");
- "I can tolerate it" (or "I can live with it that way", "This would be a minor inconvenience");
- "I dislike it" (or "I dislike it that way", "This would be a major problem for me").

The answers above do not reflect a ranking of the correspondent requirement, but rather the means for its classification in terms of Kano categories. It is important to make this difference clear to the respondents, or the answers may be biased by their preferences over certain features. For instance, a respondent may intend the "I like it" answer as "give maximum priority to the fulfillment of this requirement", which will lead to misleading classifications (e.g., also an "I expect it" answer might refer to a requirement expected to be fulfilled with maximum priority).

C. *Analysis of Kano questionnaire*

The KQ is administered to a number of stakeholders, and each answer pair is aligned with the Kano Evaluation Table (KET) [10] revealing an individual stakeholder's perception of a requirement.

A requirement is questionable if there is a contradiction in the stakeholders' answers to a question pair (e.g., respondents answered) "I like it" to both the functional and dysfunctional question). If a requirement received a substantial number of questionable scores, probably the question was confusing, and it should be reformulated. On the other hand, a reversal requirement clearly indicates that stakeholders want the opposite of what it describes, as the degree of satisfaction decreases with the level of functionality provided. Of course, no reversal requirement should be included in the final product.

Table I: Table of results.

Requirement	M	O	A	I	R	Q	Total	Mode
Req.1	27.9	9.6	39.2	21.2	2.1	0.0	100	A
Req.2	3.5	2.9	89.9	2.4	0.3	1.1	100	A
Req.3	19.9	23.4	6.7	48.3	1.1	0.6	100	I
Req.4	23.3	23.6	23.1	23.6	2.8	3.6	100	O, I
M: must-be; O: one-dimension; A: attractive; I: indifferent; Q: questionable; R: Reverse								

Once having combined the answers to the functional and dysfunctional question in the KET, the classification of the individual requirements can be summarized as in Table I. The analysis and interpretation of the results in this table can use different techniques. A standard solution is to consider statistical summaries of the categories distribution, especially the mode i.e., the most frequently occurring category. The mode is simple to use and easy to understand but has some limitations in particular cases. Consider for example Req.1 and Req.2 in Table I: the Kano category assigned by the mode to both requirements is "Attractive", but it does not say anything about the fact that the consensus on the category for Req.2 more than doubles that for Req.1. It appears likely that the two requirements should be treated differently, instead, in Table I, in which despite the mode assigns the "One-dimensional" category to the requirement, there are other three Kano categories with very similar scores. To deal with tie scores, one may consider selecting the category that has the greatest impact on the overall satisfaction, which typically consists in applying these priorities:

$$M > O > A > I.$$

Another technique to interpret the Kano evaluation has been presented by Mike Timko [10] of Analog Devices, who proposed using the Satisfaction and Dissatisfaction coefficients. The coefficients indicate, in numerical terms, how strongly a requirement may influence satisfaction or, in case it is not fulfilled, stakeholders' dissatisfaction. In the original paper, Timko labels the Satisfaction and Dissatisfaction coefficients as "Better" and "Worse", respectively. By considering the total number of answers in each Kano category for a given requirement, the Better and Worse values can be calculated using the following formulas:

$$\textbf{Better} = \frac{\mathrm{A} + \mathrm{O}}{\mathrm{A} + \mathrm{O} + \mathrm{M} + \mathrm{I}}$$

$$\textbf{Worse} = \frac{\mathrm{O} + \mathrm{M}}{\mathrm{A} + \mathrm{O} + \mathrm{M} + \mathrm{I}}$$

Differently than the mode, the satisfaction and dissatisfaction coefficients provide a better discrimination among requirements. For instance, consider again Req.1 and Req.2 in Table I, which have mode "Attractive" but with different scores. In contrast, their Better and Worse values capture this difference, as Req.2 has a Better value that approaches 1 (i.e., a great influence on stakeholders' satisfaction), whereas for Req.1 is only a half (i.e., a slight increase of satisfaction).

D. *DuMouchel Analysis Methodology*

The methodologies described above to analyze the results of the KQ are based on the categories resulting from the application of the Kano Evaluation Table. However, using the KET leads to the loss of a considerable amount of information from 25 possible combinations of answers for each requirement and from each respondent to just one of six Kano categories. Moreover, in the case of the mode, the resulting categories are further aggregated into a single Kano category for each requirement. Therefore, weaker answers (e.g., "I am neutral", "I can tolerate it") get the same weight of stronger ones (e.g., "I expect it"). Consider for example the must-be category, which can result from three different combinations of answers ("I expect it" -" I dislike", "I am neutral" - "I dislike", "I can tolerate it" - "I dislike"). According to this mapping, it is impossible to distinguish must-be requirements with a majority of functional expectations from those with a majority of functional "I can tolerate it".

To overcome this limitation, William DuMouchel proposed a finer "continuous" methodology of analysing the data from the Kano questionnaires [10]. The analysis assumes the use of a Self-stated importance Questionnaire (SIQ) [3] together with the Kano Questionnaire. The SIQ makes the respondents ranking each requirement in the KQ on a rating scale of importance, in order to determine the relative importance of the individual requirement. The analysis method described by DuMouchel assigns three scores to each requirement, namely, the Functional, Dysfunctional, and Importance scores. The first two scores translate each answer to a numerical value as reported in Table II.

Table II: The scores assigned to each answer to the functional and dysfunctional

Score	Like	Expect	Neutral	Tolerate	Dislike
Functional	4	2	0	-1	-2
Dysfunctional	-2	-1	0	2	4

The Dysfunctional scale is not symmetrical to the Functional one, but it starts from -2 instead of -4. The logic of this asymmetry is based on the observation that the Kano categories resulting from answers on the negative end (i.e., reverse and questionable) are weaker than those resulting from the positive end (i.e., must-be and one-dimensional). For this reason, DuMouchel decided to give more weight to the stronger responses so as to increase their influence on the average. Finally, the Importance score typically assigns a value from 1 ("not at all important") to 9 ("extremely important"). Having this information can sharpen the distinction among requirements making clearer which are most relevant to respondents. In other words, the Importance score can be used to separate core requirements from peripheral ones and to have a better understanding of how they impact on the overall stakeholders' satisfaction.

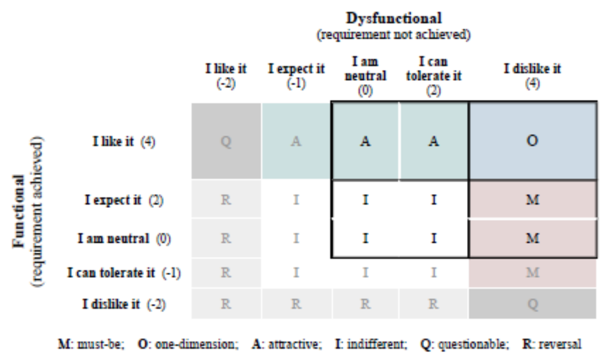

Figure 5 : The portion of the Kano evaluation table considered in the DuMouchel

In the DuMouchel method to analyze the KQ responses, the three scores lead to the categorization of the requirements within a two-dimensional coordinate system, which corresponds to the positive quadrant of the traditional KET). The rationale of focusing only on this quadrant is because it holds the strongest responses. Each requirement can be represented as a point in this coordinate system, having coordinates given by the average of the Dysfunctional and Functional scores across all respondents. More precisely, assuming Q pairs of questions, j = 1, … Q, and N respondents, i = 1, … N, the formulas to compute the average Dysfunctional (D) and Functional (F) scores of a question pair j are:

$$D[j] = \frac{\sum_i D_{ij}}{N} \quad , \quad F[j] = \frac{\sum_i F_{ij}}{N}$$

Consider for example the situation summarized in Table III, the table reports the Dysfunctional, Functional, and Importance (I) scores of four requirements, calculated based on the answers of N respondents. Fig. 5 provides a graphical representation of the four requirements in Table III into the two-dimensional categorization grid proposed by DuMouchel. The grid is divided into four quadrants, which identify the attractive (top-left), one-dimensional (top-right), must-be (bottom-right), and indifferent (bottom-left) Kano categories.

Table III: Examples of Dysfunctional (D), Functional (F), and Importance (I) scores

Requirement	D	F	I
Req1	1.09	2.57	8.6
Req2	2.44	3.24	6.7
Req3	1.74	1.15	2.1
Req4	3.64	0.58	4.1

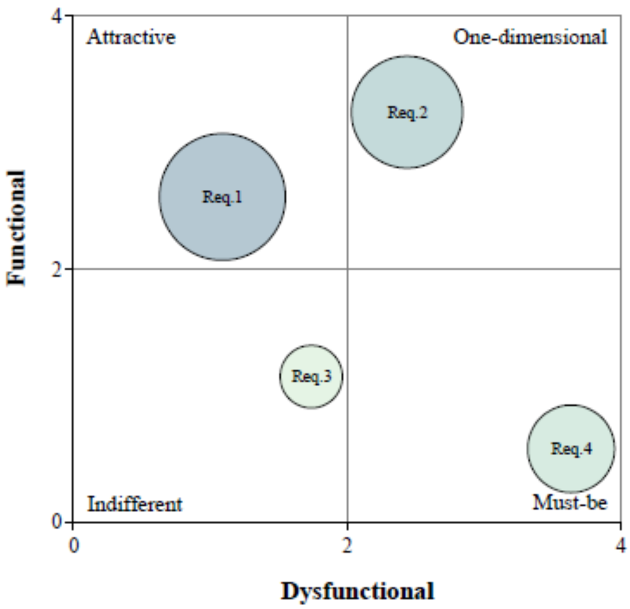

Figure 6: Graphical representation of the Dysfunctional (D), Functional (F), and Importance (I) scores assigned to the four requirements

The category of each requirement is delineated by the quadrant into which its point with coordinates (D; F) falls. For instance, a requirement such as Req.1 in Fig. 6 should be considered an attractive requirement, whereas Req.4 should be viewed as a must-be requirement. The closer a point falls to one of the corners, the more unanimous the answers associated with that requirement must have been. For example, Fig. 6 indicates that the respondents of the KQ show a higher agreement over their view of requirement Req.4 rather than for Req.3. Finally, the Importance score can be visualized in several ways, such as using different colors or different sizes proportional to I (as illustrated in Fig. 6). There are many other statistical and conceptual techniques for analyzing and interpreting the results summarized in the KET. Mikulic and Prebezac [11] have provided an extensive review of these techniques, along with a clear definition of their advantages and disadvantages. However, rather than considering them as alternative means to classify requirements in the Kano model, they should be considered possible extensions.

Fig. 6 provides a graphical representation of the Dysfunctional (D), Functional (F), and Importance (I) scores assigned to the four requirements listed in Table III referred as complementary options. The benefit of adopting such inclusive approach has been already been explained in earlier that the mode and the customer satisfaction coefficient together provide clearer and more precise information on the considered requirements.

V. Analysis and Results

In this section, we discuss the execution of the last three phases of the requirements elicitation and classification process illustrated in Fig. 1.

A. Construction of the Kano Questionnaire

To prioritize the dataset of elicited requirements on the MBDAaaS approach supported by TOREADOR, we created a Kano Questionnaire comprising 91 pairs of (functional and dysfunctional) questions. To reduce the length of the questionnaire, similar requirements were tested by the same question pair. This improves the participation level of the respondents as well as their available attention [10]. We grouped the questions of the KQ into the categories presented in Section III. For reasons of space, in this paper, we restrict our focus to the PRELIMINARIES and MODEL categories, which include 14 pairs of questions testing 26 requirements. More precisely, the KQ includes five PRELIMINARIES question pairs (labeled as P1, … , P5) and nine MODEL question pairs (M1, …, M9).

The requirements tested by P1 are related to the ease of use and abstraction of the TOREADOR platform, which enables the users of this platform to focus on overall goals without indulging into technical specifications. Also requirements in P2 focus on usability and flexibility for the users, who can easily express the goals of their BDA campaign through simple declarative models. Requirements tested by P3 concern the assisted specification of user goals supported by TOREADOR. The questions pair P4 is related to legal aspects as well as the capability of TOREADOR to advise users on the right implementation of legal constraints about Intellectual Property Rights. Lastly, the requirements in P5 focus on assurance aspects and security.

Concerning the MODEL questions, M1 provides details on the interconnections between specifications addressing Data Analytics and Data Processing aspects. More specifically, the expectation is that the TOREADOR platform should support the user in identifying the Data Processing model based on the desired Data Analytics, or vice versa. The requirements tested by M2 specify that the configuration of the deployment model should also be pre-setup by the system based on the general goals defined by the user. Questions pairs M3 and M9 are related to the capability of the TOREADOR platform to support the specification, monitoring, and verification of Service Level Agreements (SLAs) on the status of the BDAaaS process. Finally, the requirements tested by the remaining MODEL question pairs specify detail about data preparation, presentation, and visualization aspects that the TOREADOR platform should address in accordance with the general goals selected by the user in designing a BDA campaign. Among the others, particular relevance has been assigned to the interrelationships between data anonymization and analytics, as users are often unable to fully capture the effects that data anonymization compels to the results returned by analytics.

B. Administration of the Kano Questionnaire

The TOREADOR partners have administered the KQ to 27 respondents, selected according to the following criteria: (i) covering both the public and private sectors, (ii) involving stakeholders from SMEs, and, mainly, (iii) involving Domain Experts, Data Scientists, and Data Engineers. The last criterion allows building a rather homogeneous group of respondents with significant knowledge in the TOREADOR area of interest. Overall, almost three-quarters of the respondents are from the private sector, and nearly a half of them represents SMEs. The distribution of the respondent profiles ranges over the three categories, with 13 respondents who qualified themselves as Domain Expert, 11 as Data Engineers and 3 as Data Scientists.

C. Evaluation and interpretation of results

We analyzed the results of the 27 completed Kano questionnaires using the DuMouchel methodology presented in Section IV.D. The distribution of the answers received for the requirements related to the 14 question pairs in the PRELIMINARIES and MODEL categories indicates that the majority of respondents have always assigned a positive ranking to functional questions. Particularly, almost three-quarters of the most frequent answers were "I like it", and the remaining were "I expect it". It is, thus, reasonable to conclude that the performance of the requirements under consideration will positively affect the stakeholders' satisfaction. By considering the dysfunctional questions, it appears that only a quarter of the answers were on the positive side of the scale; the remaining 75% of answers are evenly distributed among the neutral, can tolerate, and dislike response types. In particular, the most frequent answer to dysfunctional questions are "I am neutral" and "I dislike it". In other words, the majority of respondents showed to be either indifferent or strongly dissatisfied if the basic requirements related to work area 1 are not met.

According to the DuMouchel methodology, we are first required to calculate the Functional and Dysfunctional scores of every set of requirements, by applying the mapping to the respondents' answers. Second, the Importance score is computed by averaging the responses given to the Self-Stated Importance questionnaire administered together with the KQ. Tables IV and V report the scores calculated for the questions pair to carry out the DuMouchel analysis, along with the resulting Kano categories. Fig. 8 and Fig. 9 provides a graphical representation of the results of the DuMouchel analysis, where "P" and "M" refer to the PRELIMINARIES and MODEL questions pairs, respectively.

Table IV: Scores calculated for the DuMouchel analysis of the PRELIMINARIES questions, and results

ID Questions	Dysfunctional	Functional	Importance	Category
P1	0.82	2.50	4.46	A
P2	0.24	2.82	4.42	A
P3	0.30	2.60	4.15	A
P4	1.67	2.75	4.27	A
P5	1.44	2.48	3.60	A

Table V: Scores calculated for the DuMouchel analysis of the MODEL questions, and results

ID Questions	Dysfunctional	Functional	Importance	Category
M1	1.50	3.00	3.67	A
M2	-0.50	2.25	4.00	-
M3	1.57	2.29	3.65	A
M4	2.44	2.44	3.89	O
M5	1.89	2.44	3.33	A
M6	1.67	2.67	4.11	A
M7	1.78	1.67	3.44	I
M8	2.33	2.67	4.00	O
M9	1.89	2.44	3.80	A

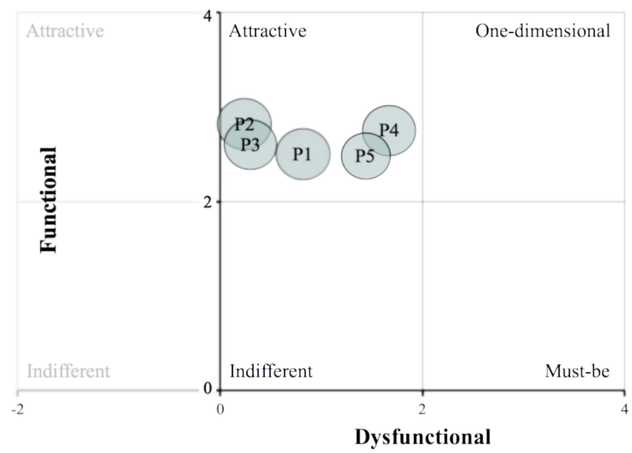

Figure 7: Result classification of the PRELIMINARIES requirements using the DuMouchel analysis methodology

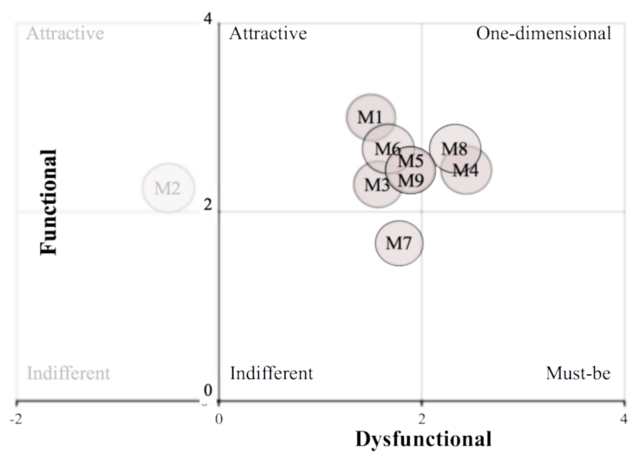

Figure 8: Result classification of the MODEL requirements using the DuMouchel analysis methodology

The analysis of the requirements included in the PRELIMINARIES section indicates that all the basic features and objectives of TOREADOR have been considered attractive by the majority of the respondents (Fig. 8). This is a clear and positive acknowledgment of the relevance of the problem addressed by the TOREADOR project, as well as of the soundness and potentials of the proposed solution.

As illustrated in Fig. 9, the classification of the requirements related to the MODEL section is more varied. The DuMouchel analysis has in fact identified three categories of requirements, namely, attractive, one-dimensional, and indifferent. Also, the set of requirements related to the availability in TOREADOR of platform independent models would have been classified attractive if using the original KET. However, this classification has been discarded by the DuMouchel analytical model because the high number of positive answers (i.e., "I expect it") to the dysfunctional question weaken its reliability. This is most certainly due to ambiguities in the phrasing of the dysfunctional question, which failed to highlight its contrasts with the complementing functional question. Same kind of contrasting results has been produced for other sections to after application of DuMouchel analysis.

VI. Conclusion

The TOREADOR project implemented early involvement of users to better understand project goals. Some of the results achieved can be generalized to any BDA campaign.

The result of our study indicates that all basic functions related requirements have been considered attractive by the majority of the respondents. With the help of DuMouchel methodology, it is also revealed that there are various other requirements which are not deemed very important initially but are highly important to achieve to customer satisfaction so should be assigned high priority. Some of the requirement received very low priority, one reason for that is they are highly technical in nature. Overall this entire exercise helps in improving requirement elicitation phase in the development of a Big Data platform. In particular, we identified three areas that TOREADOR users and stakeholders considered particular innovative for a platform supporting the design of a BDA campaign:

- **Model-driven design**: specifications start by abstract goals and the system guide the user in selecting compatible specifications.
- **Data Preparation support:** the system supports the user in identifying the effects that data preparation, in particular data anonymization, has on analytics.
- **Legal aspects:** the system guides the user in checking whatever data preparation and processing is consistent with legal aspects such as intellectual property and personal data protection.

The results of this work manifest a relevant overlapping with indications coming from our previous propositions [12].

References

[1] M. Leida, C. Ruiz and P. Ceravolo, "Facing big data variety in a model driven approach," in IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2016.

[2] B. Nuseibeh and S. Easterbrook, "Requirements engineering: a roadmap," in Proceedings of the Conference on the Future of Software Engineering. ACM, 2000, 2000.

[3] K. Matzler and H. Hinterhuber, "How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment," Technovation, vol. 18, no. 1, pp. 25-38, 1998.

[4] F. R. Frederick and E. S. W., "Zero Defections: Quality Comes to Services," Harvard Business Review, vol. 68, no. 5, pp. 105-111, 1990.

[5] Y. Akao, "Quality function deployment," Productivity press, 2004.

[6] N. Kano, N. Seraku, F. Takahashi and S. Tsuji, "Attractive quality and must-be quality," JOURNAL OF THE JAPANESE SOCIETY FOR QUALITY CONTROL, 1984.

[7] S. Kujala, "User involvement: a review of the benefits and challenges," Behaviour and information technology, vol. 22, no. 1, pp. 1-16, 2003.

[8] IIBA, A guide to the business analysis body of knowledge, International Institute of Business Analysis (IIBA), 2009.

[9] M. Conklin, K. Powaga and S. Lipovetsky, "Customer satisfaction analysis: Identification of key drivers," European journal of operational research, vol. 154, no. 3, pp. 819-827, 2004.

[10] C. Berger, R. Blauth, D. Boger, C. Bolster, G. Burchill, W. DuMouchel, F. Pouliot, R. Richter, A. Rubinoff, D. Shen and others, "Kanos methods for understanding customer-defined quality," Center for quality management journal, vol. 2, no. 4, pp. 3-35, 1993.

[11] J. Mikulic and D. Prebevzac, "A critical review of techniques for classifying quality attributes in the Kano model," Managing Service Quality: An International Journal, vol. 21, no. 1, pp. 46--66.

[12] C. A. Ardagna, P. Ceravolo and E. Damiani, "Big data analytics as-a-service: Issues and challenges," in IEEE International Conference on Big Data(Big Data), 2016.