Conceptual Modeling for Business Analytics: A Framework and Potential Benefits

Soroosh Nalchigar*, Eric Yu†
*Department of Computer Science, University of Toronto
soroosh@cs.toronto.edu
†Faculty of Information, University of Toronto
eric@cs.toronto.edu

***Abstract*—Advanced analytics solutions are becoming widespread in business organizations. While data scientists create, implement, or apply machine learning algorithms, business stakeholders need the ultimate solution to gain competitive advantage and performance improvement. How can one, systematically, elicit analytical requirements? How can one design the analytics system for addressing such requirement? How can one assure the alignment between data analytics solutions and business strategies? How can one codify and represent analytics know-how in terms of design patterns? This paper has two contributions. First, it introduces a conceptual modeling framework for addressing those challenges. Second, it assesses the potential use cases and limitations of the framework by applying it to two case studies.**

***Keywords*—Conceptual Modeling, Requirements Engineering, Business Analytics, Machine Learning, Data Analytics.**

I. Introduction

Advanced analytics solutions are becoming widespread in business organizations. Despite this ever increased interest, many businesses still struggle to identify how to use analytics to take advantage of their data [1], [2]. Requirements analysis and design of business analytics systems is proven to be a challenging task [3], [4].

While data scientists apply and implement machine learning algorithms, business stakeholders need the ultimate solution to gain competitive advantage and performance improvement. How can one, systematically, elicit analytical requirements? How can one design the analytics system for addressing such requirement? How can one ensure the alignment between analytics and business strategies? How can one codify and represent analytics know-how in terms of design patterns?

This paper has two contributions. *First,* it introduces a conceptual modeling framework for addressing such challenges. The framework includes three modeling views, namely Business View, Analytics Design View, and Data preparation View. It comes with three kinds of design catalogue that represent know-how knowledge with respect to each view. *Second,* it illustrates the potential use case of such framework through two illustrative cases. Through examples, it illustrate how the framework can be used for (1) eliciting analytics requirements, (2) clarifying analytics requirements, (3) deriving analytics solution design, (4) monitoring analytics impact on business, (5) aligning analytics solutions with business strategies, and lastly for (6) developing and deploying design patterns for analytics solutions.

This paper is organized as follows. Section II presents an overview of the case studies. Section III describes and illustrates the framework including the modeling views and design catalogues. Section IV shows different use cases of the framework in the requirements analysis and design processes of analytics systems. Section V describes findings and limitations. Section VI summarizes related works and highlights the contributions. The paper ends in Section VII with conclusions and directions for future work.

II. Illustrative Cases

This paper uses two illustrative cases to address the research objectives in previous section. The two cases were analyzed in collaboration with a participant who had work experience as a data scientist in addition to some experience in conceptual modeling and goal-oriented requirements engineering. All models in this paper are based on information from two main sources: (1) a collection of analytics case studies and white paper documents retrieved from Internet, and (2) authors' collected experience from real data mining projects in both domains. If needed, the models are supplemented with some assumptions.

Case-1: A Shopping Mobile App. The first case is about a company that offers a variety of products to its users via in-app purchases. The company aims to increase its market share and net profit by focusing on user retention and their loyalty. The stakeholders are interested in using machine learning and advanced analytics solutions to support a wide range of decisions about their marketing campaigns and reward programs. Company's data stores include users demographics, their activities within the app, and their online purchases.

Case-2: A Grocery Retailer. The second case is about a supermarket chain and food distributor. The company aims to improve its online grocery promotions, improve the physical store experience, as well as decrease logistics and operations costs. The business stakeholders are interested in applying cutting edge analytics and up-to-date datasets for achieving those objectives. The company tracks customer activities through its loyalty card system. It also has started collecting sensor data such as at store entrances in addition to external data such as metropolitan population.

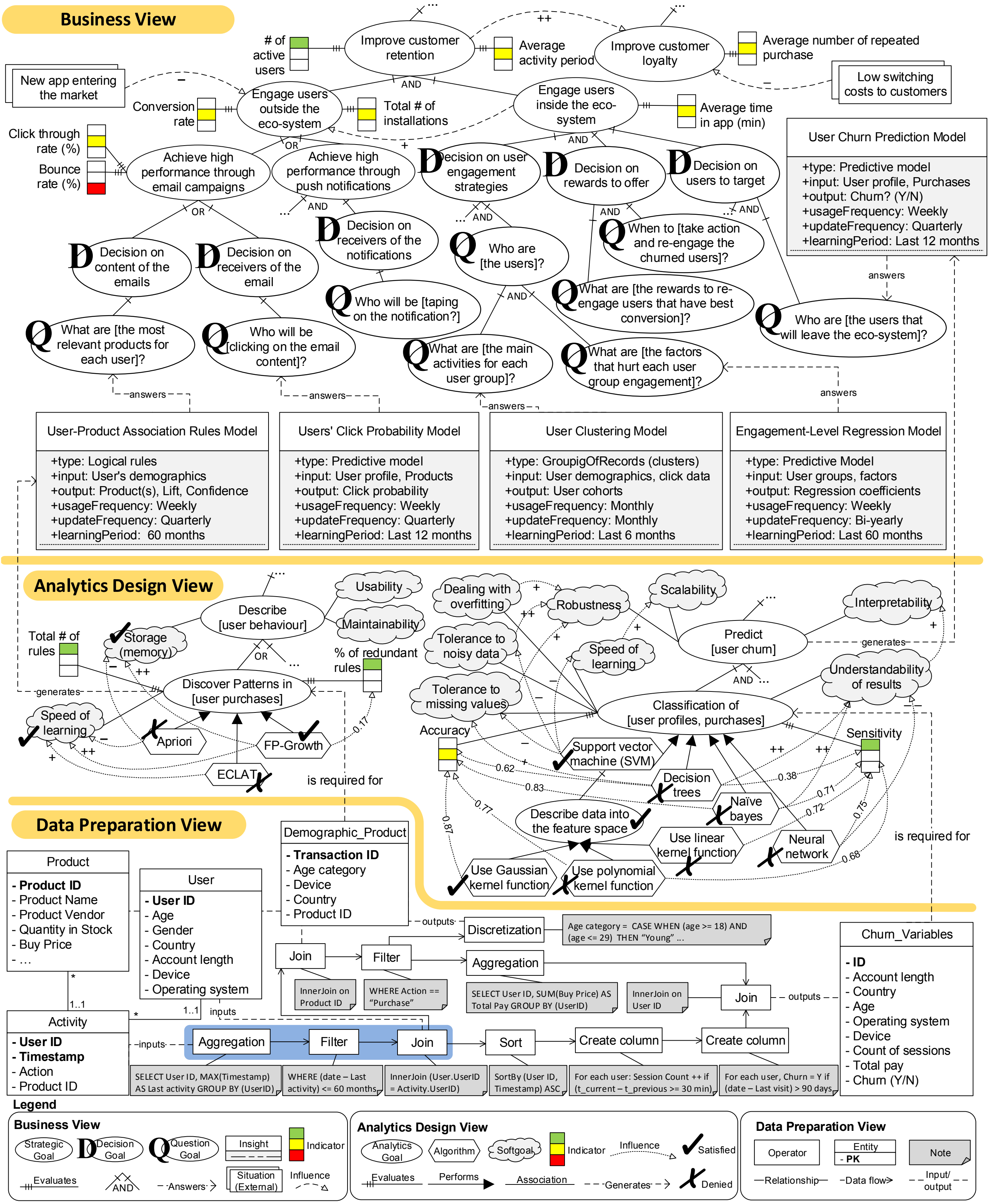

Fig. 1. Fragments of the three modeling views for Case-1. Due to space limitation, the Analytics Design View and the Data Preparation View are showing the solution for only two (out of eight) question goals in the Business View.

III. Overview of the Framework

A. Modeling Views

The proposed framework includes three complementary modeling views: *Business View, Analytics Design View,* and *Data Preparation View*. These views, while having different focuses and serving different purposes, are linked to each other and bridge the gap between strategic goals, machine learning algorithms, and data tables.

1) Business View: This view aims to (i) facilitate the elicitation and clarification of analytics requirements in business contexts, (ii) support analysis of those requirements (e.g., prioritization), and (iii) ensure the alignment of business and analytics strategies. The main modeling elements are strategic goals, decision goals, question goals, insights, indicators, influences, and situations.

Strategic Goals, adopted from the Business Intelligence Model (BIM) language [5], symbolize business objectives and strategies. In Figure 1, Improve customer retention is an example of a strategic goal. Strategic goals are refined into lower-level goals through decomposition links.

Strategic goals are decomposed into one or more *Decision Goals*. Decision goals represent the decisions that need to be made towards achieving the strategic goals. They symbolize the decisions that (will be) are supported by the (to-be) analytics system. In Figure 1, Decision on content of the emails is an example of a decision goal. It shows that in order to Achieve high performance through email campaigns, the corresponding actor needs to make the Decision on content of the emails to be sent to the target users.

A decision goal can be decomposed into one or more *Question Goals*. Question goals capture the "needs-to-know" of the stakeholders towards decisions to be made. They represent business questions that once answered (using machine learning algorithms), result in achieving decision goals and hence enable data-driven decision support towards strategic goals. Question goals are analyzed in terms of *Type, Topic, Tense,* and *Frequency*. Question type denotes the question phrase (what, who, when, where, why, how). Question topic captures the focus of analysis and reveals related parts of enterprise data stores for the problem at hand. Question tense (past, present, future) represents the temporal aspect of the focus of the analysis. In many cases, specifying the tense facilitates finding an analytics family of techniques that is most relevant to the business needs. Question frequency indicates how frequent the corresponding actors need and answer for the question goal. In Figure 1, What are the most relevant products for each user group? is an example of a question goal. It shows that in order to make the Decision on content of the emails, the corresponding actor[1] needs to know the products that are more relevant for each group/cluster of users. The Business Questions Catalogue (introduced later in Section III-B1) provides project team and/or stakeholders with a wide range of question goals and their associated analytics techniques to select from.

[1]Actors are not shown here due to space limitations.

A question goal is answered by (i.e., satisfied by) one or more *Insights*. Insight elements characterize the type of knowledge/patterns/findings that need to be extracted from datasets such that the question goal is answered. They are connected to question goals through the *answers* links. Insights are differentiated into subtypes including *Predictive Models*, *Probability Distributions*, *Grouping of Records* (e.g., clusters), *Logical Rules* (e.g., association rules), and *Diagrams* (e.g., correlation heat-maps). The type of insight suggests relevant machine learning algorithms that can be applied fro the problem at hand. In Figure 1, User-Product Association Rule Model is an example of an insight. It symbolizes a set of Logical rules (e.g., Canadian users with an age between x and y are likely to buy product z), which answer the question of What are the most relevant products for each user group?. At run-time, this insight requires User's demographics data as input, in order to generate a list of Product(s) as the answer to the question. This insight is used on a Weekly basis and the rules are mined from the dataset with a 60 months time interval. More examples of each modeling concept can be found in Figure 1.

2) Analytics Design View: This view aims to (i) support exploration of alternate approaches for the problem at hand, (ii) facilitate design of (machine learning) experiments and identifying trade-offs, and (iii) support algorithm selection and monitoring their performance over time. The main modeling elements are analytics goals, algorithms, softgoals, influences, and indicators.

Analytics Goals capture the intention of the analysis to be performed over the datasets. Three types of analytics goals are distinguished. If the analytics aims to predict the value of a data attribute (i.e., a variable or data column), it is called a *Prediction Goal*. If the analytics aims to summarize and explain the dataset, it is called a *Description Goal*. If the analytics aims to find the optimal alternative given a set of options and criteria, it is called a *Prescription Goal*. Each of these types are further refined in terms of sub-types. For example, *Numeric Prediction* and *Classification* are subtypes of the prediction goal. Also, description goals include two subtypes of *Clustering* and *Pattern Discovery*. In Figure 1, Describe user behaviour is an example of an analytics goal, representing descriptive analytics intentions. To achieve this goal, the system needs to achieve the goal Discover patterns in user purchases.

Algorithms represent machine learning algorithms that address an analytics goal. They are connected to analytics goal through the *Performs* links, showing a means-end relationship [6]. Figure 1 shows that Apriori, ECLAT, and FP-Growth as alternative algorithms for achieving the pattern discovery goal.

Algorithm are evaluated and compared with regard to *Indicators* and *Softgoals*. Indicators represent the numeric metrics that are used for performance evaluation and comparison of algorithms. % of redundant rules is an example of indicator (see an example of a rule in previous section). Softgoals capture quality requirements that need to be satisfied by the (to be) system. Speed of learning and Usability are examples of softgoals.

Influence Links represent the contribution and impact of each algorithm on softgoals and indicators. For example, the link from FP-Growth to the indicator % of redundant rules shows that the algorithm will result on the value of 0.17 for that indicator, found through experiments. Also, the influence link from the algorithm Apriori towards the softgoal Speed of learning shows that this algorithm will Hurt (−) achievement of that softgoal. By capturing these, the Analytics Design View supports comparison and selection of alternative algorithms.

This view is connected to the previous modeling view through the *generates* links. These links connect an analytics goal to an insight element in the Business View. The Algorithms Catalogue (introduced later in Section III-B2) provides users in this modeling view by showing what algorithm are applicable for a given analytics goal, as well as the relevant softgoals and indicators for the problem at hand. More examples of each modeling concept can be found in Figure 1.

3) Data Preparation View: This view aims to (i) support share and reuse of prepared data assets, (ii) enhance data awareness among analytics users, and (iii) ease data understanding by providing a reference for data engineers (who prepare datasets) on data preparation activities. The main modeling elements are operators, algorithms, tasks, entities, relationships, and data flows.

Entities and *Relationships* represent the conceptual structure and content of the data sources. Figure 1 shows that for each User, demographics data such as Age and Gender is being captured. *Data preparation task* represents the general task of preparing data for performing some analytics goals. *Data Cleaning, Data Transformation, Data Reduction,* and *Data Integration* are four types of preparation tasks. In Figure 1, the blue-shaded area in the Data Preparation View shows an example of a data reduction task. It shows that the system excludes those users who have not done any activity/shopping for more than five years.

A data preparation task consists of one or more *Operators*. Operators represent an atomic activity that performs (part of) a data preparation task. In figure 1, Create column and Join are examples of operators. Operators are linked by *Data Flows* to represent the sequence and dependencies. *Notes* are linked to operators to provide explanations and ease the understanding of the function being performed. For example the note For each user, Churn = Y if (date Last visit) $>$ 90 days associated with a Create column operator shows that a new data column is created and its value is Y if the the corresponding user has been inactive for more than three months. The *Input* links represent the dataflows from data stores to the operators. The *Output* links are pointed to the prepared datasets.

This view is connected to the previous modeling view through the *is required for* links. These links connect a prepared dataset to one or many analytics goals in the Analytics Design View. The Data Preparation Catalogue (briefly introduced in Section III-B3) assist users by providing methods that are available for different data preparation tasks as well as information on when to use what method. More examples of each modeling concept can be found in Figure 1.

B. Design Catalogues

An important component of the framework is a set of catalogues that support requirements analysis and design of analytics systems. The catalogues provide proven solutions to common and recurring analytics problems in business domains. The catalogues organize and represent a body of knowledge that can be used during analytics projects to speed up the development process. Three types of catalogues are distinguished in the framework.

1) Business Questions Catalogues: The focus of this catalogue is to represent a wide range of business questions that can be answered with machine learning and analytics solutions. Using this catalogue stakeholders and analytics experts are able to browse through an organized set of *Question Goals*. The catalogue categorizes question goals based on their *Type* and *Tense*. Within each category, a wide range of instances exist where each instance is mapped to an specific analytics goal. For example, the two question goals of Who will be [taping on the notification?] and Who will be [clicking on the email content]? (from Figure 1) are listed under the category of *Who* and *Future*, and both are mapped to *Prediction Goal*. In this way, the catalogue to bridge the gap between business questions and analytics techniques.

2) Algorithm Catalogue: Effective design of analytics systems requires experimentation with and selection of machine learning algorithms. This catalogues codifies the know-how knowledge on analytics techniques and algorithms. In particular, it represents different machine learning algorithms that are applicable for a given *Analytics Goal*. The catalogue also represents well-known *Indicators* (i.e., metrics) for evaluation and comparison of those algorithms. For each analytics goal, the catalogue also provide relevant *Softgoals* (i.e., quality requirements) whose lack of consideration can become major issues later in the project life-cycle. Moreover, it encodes the knowledge on how each algorithm is known to influence meeting those softgolas. For example, in this catalogue Regression and Neural Networks are among algorithms for performing Classification. Recall and Precision are among the metrics to be considered while Dealing with overfitting is represented as a quality requirement.

3) Data Preparation Catalogue: This catalogue has a similar structure to the Algorithm Catalogue, but representing the data preparation know-how knowledge. This catalogue helps developers to find existing methods for addressing data preparation tasks such as data cleaning and data value normalization. For example, in this catalogue Min-max normalization and Z-scale normalization are captured among different ways of performing Data normalization.

Due to space limitations, the metamodels and content of these catalogues are not discussed here. Readers are referred to [7] for more details.

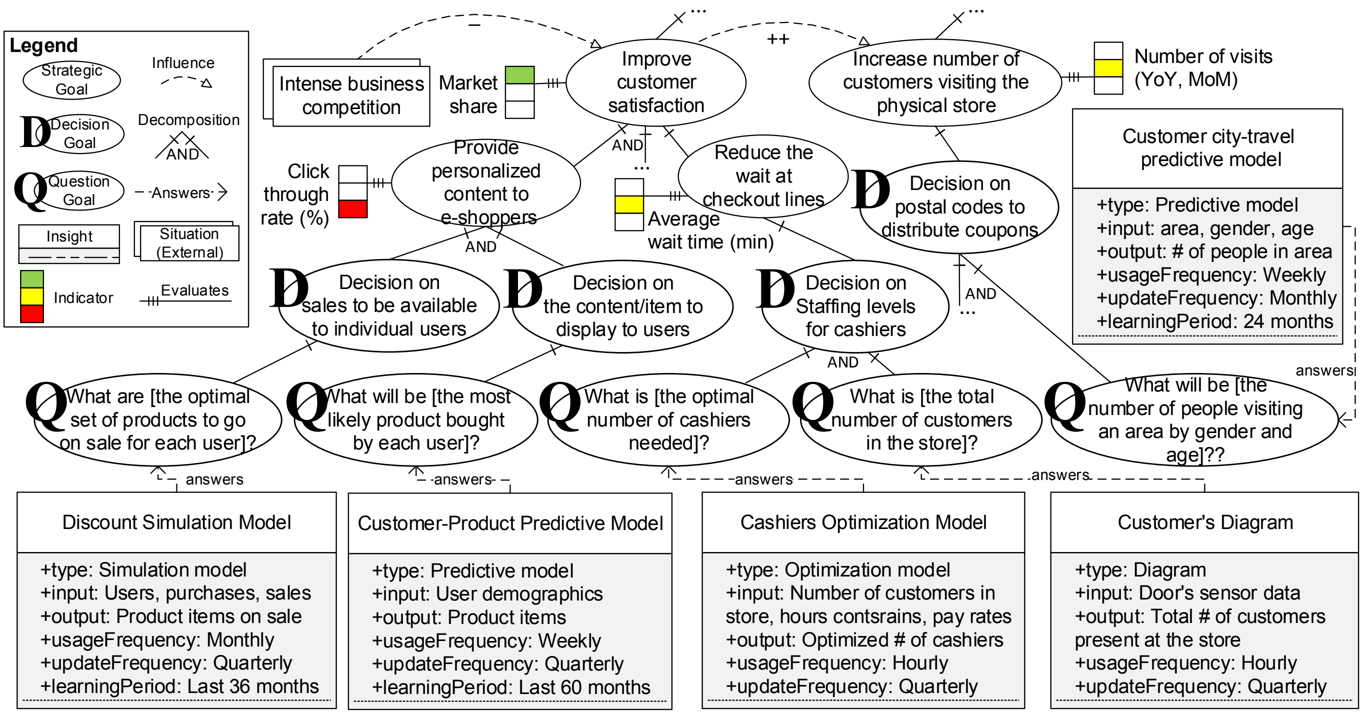

Fig. 2. Partial Business View for the grocery retailer (Case-2).

IV. Why Conceptual Modeling for Business Analytics?

In this section, we present a number of ways in which the conceptual modeling framework can help in the data analytics requirements analysis and design processes. These can be viewed as use cases for the framework. We describe them using examples from the case studies.

A. Eliciting Analytics Requirements

1) The Challenge: Requirements elicitation for the advanced analytics systems is a challenging task [3]. This is to a great extent due to the huge conceptual gap between business stakeholders and analytics experts. The continuous and rapid growth of the machine learning and analytics algorithms, technologies, and applications intensifies the mentioned gap. Studies show that the lack of understanding on how to use business analytics techniques is a leading barrier to effective design and implementation of these systems [2]. While in many real-world business contexts, stakeholders admit the importance and necessity of analytics systems, they lack a clear understanding of what kinds of analytics capabilities are required and where they are located (within their business area/function).

2) How Modeling Helps: Figure 2 shows a fragment of the Business View for Case-2. It shows that the corresponding retail company aims to Improve customer satisfaction as one of its strategic goals. Towards that end, the company aims to Reduce the wait at checkout lines and to Provide personalized content to e-shoppers. Moreover, the company uses a set of performance indicators to monitor how well it is doing with respect to those goals. Click through rate (%) and Average wait time (min) are indicators that are associated with the aforementioned goals.

Achieving strategic goals requires business stakeholders to make critical decisions. For example, in order to Reduce the wait at checkout lines, the store manager needs to make the Decision on Staffing levels for cashiers, (i.e., to decide how many of the store staff should act as cashier at a given time).

In order to make decisions, business stakeholders need to know the answer(s) to some questions. For example, in order to make the Decision on Staffing levels for cashiers, the store manager needs to know What is the total number of customers in the store?, as well as What is the optimal number of cashiers needed? (See Figure 2). In order to answer business questions, stakeholders need to rely on data-driven insights and findings to be generated by analytics techniques and solutions. For example, in order to answer the question of What is [the optimal number of cashiers needed?], the store manager needs an Optimization model, that receives Number of customers in store as well as Cashiers' pay rates as input and generates Optimized # of cashiers as output. This optimization model, during runtime, is used on an Hourly basis, and its parameters are updated Quarterly.

Characterizing the business in terms of strategies, decisions, analytical questions and insights is a critical step towards effective design and implementation of analytics systems. Understanding business strategies helps stakeholders and project team to justify why they are performing the analytics work. In the framework, this is represented as *Strategic Goals*, such as

Improve customer satisfaction. Without taking strategy into account, the project team and stakeholders would not know the *why* behind analytics initiatives. Understanding business decisions results in discovering areas that need support from analytics solutions and data-driven initiatives. In the framework this is captured in terms of *Decision Goals*, such as Decision on Staffing levels for cashiers in Figure 2. Figure 1 also includes examples of decision goals. This modeling element ensures the connection between analytics solution and organizational decision processes. Moreover, it facilitates linking analytics-driven insights into actions and leveraging the analytics findings in business operations and decisions.

Eliciting business questions results in discovering the focus of analytics project and the issues that it is intended to inform. In the framework, this is represented in terms of *Question Goals*, such as What is the optimal number of cashiers needed?. By modeling question goals, one is indeed eliciting the needs-to-know of stakeholders towards their decisions, which will result in performing the right analysis for the right user. Moreover, confirming the question goals with stakeholders support the process of understanding and communicating analytics findings, once they are generated.

Understanding analytical insights help characterizing the type of findings that are required for answering the business questions. In the framework, this is represented in terms of *Insights*, such as Cashiers Optimization Model. This allows specification of the actual outcome of the machine learning algorithms. By modeling the desired outcome, indeed the project team reveals the (group of) analytics techniques to be used for the problem at hand. During the process of modeling, by refining strategic goals into sub-goals and thereafter into decision goals and question goals, one can elicit analytics requirements of the stakeholders. In summary, the Business View model provides a systematic way of revealing advanced analytical requirements by representing "who" needs to know "what", and "why".

B. Clarifying Analytics Requirements

1) The Challenge: Analytics requirements often need to be clarified for both stakeholders and analytics team. Lack of congruency between the business problems perceived as critical by the stakeholders and the problems actually addressed by the analytics system/team is a key cause of failure in analytics projects [8]. Data science projects include asking and experimenting with a series of (initially wrong) questions in order to improve, modify, refine, and eventually get to better questions, insights, and valuable decisions [9].

Depending on how business questions are formulated, the analytics work (including the choice of algorithms, techniques, and design of data preparation workflows) varies considerably. The study in [10] reports that during analytics projects, stakeholders tend to raise unstructured questions which usually include ambiguities in the definitions of key variables. Clarifying these ambiguities and formulating the right business questions is a critical step in this process, which needs tremendous amount of work and interactions with business stakeholders [3]. Lack of those clarifications can result in misinterpretation of the outputs/findings that emanate from the analytics work, and eventually loss of time and resources [11].

2) How Modeling Helps: The Business View model in Figure 1 shows that in order to make the Decision on user engagement strategies, the corresponding actor needs to know Who are the users? (a broad question that includes ambiguities). Towards answering that question, the actor needs to know What are the main online activities of each user groups? and also What are the factors that hurt each user groups engagement?. The model shows that by having a User Clustering Model one can answer the former question. This insight receives User demographics and their click data as input and generates User cohorts, which answers the question of What are the main online activities of each user group?.

By refining business questions into sub-questions, one can discuss and resolve early ambiguities that are raised by business stakeholders. In the framework, this is represented in terms of *Decomposition Links* that break a question goal into sub-goals. For example, in Figure 1, the question goal of Who are the users? is refined into sub-questions. In addition, question goals are analyzed in terms of *Type*, *Topic*, *Tense*, and *Frequency*. Specifying these attributes for each question goal assists in arriving at a set of clear and accurate requirements in addition to enhancing the communication and understanding between developers (usually referred to as data scientists) and stakeholders.

Insight elements characterize findings/outputs of the (machine learning) solution in terms of *Type*, *Input*, *Output*, *Usage Frequency*, *Update Frequency* and *Learning Period*. Figure 2 includes several examples of such element. It shows that insight elements clarify the type of knowledge that is required for answering the question goals. During the process of modeling, by refining question goals into sub-questions and thereafter specifying the insights, one can clarify the analytics requirements, reduce ambiguities, while having the stakeholders involved in the process.

C. Deriving Analytics Solution Design

1) The Challenge: Analytics requirements, once discovered, must eventually lead into analytics design, experimentations with machine learning algorithms, and implementation. A large number of algorithms exist and more are being developed. For a given analytics goal (e.g., numerical prediction), usually several alternative algorithms exist (e.g., linear regression, neural networks, support vector machine). Algorithm selection is a critical design decision that influences several aspects of the eventual analytics solution, such as understandability of results, scalability, memory, tolerance to noisy data, and missing values.

Meeting these quality requirements can be crucial to the success of the system [12]. Moreover, the algorithm selection task requires taking into account different (sometimes competing) numerical metrics. To trade-off and find the most suitable technique is a challenging task.

2) How Modeling Helps: The middle section of Figure 1 shows part of an Analytics Design View model for the Case-1. On the right side, the model shows the analytics goal of Predict user churn. Towards that goal, the analytics solution needs to achieve the Classification of user profiles and purchases. The model shows that there are several alternative algorithms that can perform the classification goal, such as Support Vector Machine (SVM), Decision Trees, Nave Bayes, and Neural Networks. These algorithms are evaluated with regard to some numeric metrics such as Accuracy and Sensitivity. The model also shows that softgoals such as Tolerance to missing values, and Tolerance to noisy data are considered while designing the system. The model also represents how each algorithm would influence the metrics (numeric labels) and the softgoals (qualitative labels). For example, use of Neural Network would result in the value of for 0.75 for Sensitivity while it would Break ($--$) the softgoal Understandability of results. The model shows that the selected algorithm is Support Vector Machine (SVM) with the Use Gaussian kernel function[2].

At design time, by knowing the desired types of outputs, one can find the kinds of analytics techniques that needs to be performed. In the framework, this is captured through *Insight* elements, their *Type*, *Analytics Goal*, and *Generates* links. The insight type specifies what kinds of machine learning output would be required for the business question at hand. The type of insight, once clarified, reveals the category of machine learning algorithms that can be used for the requirements at hand. For example, in Figure 1, the insight User Churn Prediction Model with the Predictive Model type, suggests the need for predictive analytics (i.e., prediction goal). In Figure 1, this is represented in terms of the prediction goal of Predict [user churn].

The type of analytics goal, once revealed, suggests a relevant set of alternative algorithms for the problem at hand. The Algorithm Catalogue (see Section III-B2) presents existing algorithms, metrics, and soft-goals for various types of analytics goals. The project team can browse through the catalogue to derive the design of the analytics system. In Figure 1 the prediction goal is decomposed into the Classification of user profiles and purchases which can be performed by alternative algorithms [3,4].

Designing analytics system include making decisions on algorithms with respect to criteria. In the framework, those criteria are modeled in terms of *Softgoals* and *Indicators*. The goal-oriented reasoning techniques [13] can be used to reason about alternative algorithms for performing analytics goals. Soft-goals, their influence, analytics indicators along with their priorities will be used during those reasoning and analysis. Lack of these considerations can result in an implementation where critical soft-goals are not satisfied.

[2]Assuming that the Accuracy metric has the highest priority among the metrics and softgoals

[3]In Algorithm Catalogues, Classification Goal is modeled as a type of a Prediction Goal to be used in situations where the target variable to be predicted is categorical.

[4]Due to space limitations, the model in Figure 1 is showing only one of the classification goals. There can be several classification models for predicting user churn each with a different prediction period and time interval.

D. Monitoring Analytics Impact on Business

1) The Challenge: It is essential for an enterprise to define and agree on a set of metrics that can be used to measure and monitor the impact of analytics on the business [14][15]. Such metrics can be used to justify the need for analytics, obtain executive sponsorship, and to assure analytics-driven value creation over time. To systematically discover and use those metrics is a difficult task. Lack of such measures could result in evaluating the right analytics system based on a wrong set of metrics and business success criteria. On the other hand, early definition of these metrics is reported to be critical to the success of the business analytics initiative [16].

2) How Modeling Helps: Figure 2 shows a fragment of the Business View model for Case-2. The model shows that the retailer aims to Reduce the wait at checkout lines. The model shows that the company is tracking Average wait time (min) attached to that goal as an indicator. It also captures decomposition of such goal to the Decision on staffing levels for cashiers. The model shows the use of Cashiers Optimization Model and Total Customers Diagram insights as analytics-driven results to support such decision. By agreeing on and monitoring Average wait time (min) over time, the project team can understand the impact and the business values derived from those analytics insights.

Understading the impact of analytics on enterprise requires taking into account the relationship between analytics work, decision processes and organizational performance [17]. In the framework, these relationships are captured mainly through *Strategic Goals*, their associated *Indicators*, *Decomposition Links*, *Decision Goals*, *Question Goals* and *Insights*. Indicators, represent numeric metrics that show how well an organization is doing with regard to some strategic goal. The strategic goals are decomposed into decision goals, which are (eventually) linked to analytics insights through the question goals. By capturing these connections, the framework indeed creates links from performance indicators to analytics systems and findings.

At design time, by elaborating on and refining strategic goals and identifying relevant business indicators, the stakeholders along with the project team arrive at a set of metrics that can be monitored for analyzing the impact of analytics solution on business. During the modeling process, these measures can be identified and attached to strategic goals that are at the higher level of decision goals. At run-time, the target and current values of the indicator can be compared over time to analyze the changes before and after introducing the analytics solutions.

E. Aligning Analytics Solutions with Business Strategies

1) The Challenge: Aligning analytics systems and techniques with enterprise strategies is critical for eventual success of the analytics initiatives [2][18]. Such alignment results in an ongoing understanding of enterprise objectives by the analytics team while securing continuous business support and executive sponsorship. Without a strategic perspective, the stakeholders and analytics team would not know what it is that they are trying to achieve through analytics work, how to allocate analytics resources, or what data to focus on [10].

2) How Modeling Helps: The model in Figure 2 shows that the retail stakeholders desire an answer to (i.e., need to know) What will be the most likely product bought by each user?. Knowing that, is required for making the Decision on the content/item to display to users. Such decision would be part of accomplishing the strategic goal of Provide personalized content to e-shoppers and thereafter Improve customer satisfaction. The model also shows that achieving such strategic goal, has a strong positive (++) influence the other strategic goal Increase number of customers visiting the physical store. By capturing these, the framework indeed ensures that the analytics effort and findings are informing a relevant business question and decision towards achieving enterprise strategies.

Aligning analytics and business includes an understanding of business objectives, identification of decision processes and issues, and clarifying how analytics system would contribute to them. In the framework, these are captured through *Strategic Goals*, *Decision Goals*, *Question Goals* and connections among these elements in terms of *Decomposition Links* and *Influence Links*. From a top-down point of view, while building the models, business stakeholders can assure that the analytics solution is supporting business strategies and enables data-driven decisions. This can also help justifying resources for performing the analytics projects. From a bottom-up perspective, and while developing a solution, the machine learning and data science team can assure that they are generating insights for valid business questions, supporting critical decision processes, and hence driving values from analytics initiatives.

F. Developing and Deploying Design Patterns for Analytics Solutions

1) The Challenge: Machine learning and advanced analytics applications are new capabilities for many organizations. A shortage of talent with deep expertise in statistics and machine learning is reported to be an obstacle towards effective use of analytics [19]. Rapid growth and advances in the machine learning domain adds to sush challenges, making the design of such system more difficult. Moreover, in order to extract value from analytics, business managers and stakeholders need to know about machine learning algorithms and their potential applications [20].

2) How Modeling Helps: The model in Figure 3 shows a fragment of Algorithms Catalogue. The model formally expresses know-how knowledge on how to perform Classification. It shows that *k*-Nearest neighbor and Random Forest are among algorithms that can perform Classification, which itself is-a type of Prediction Goal. It represents Perceptron and Back-propagation as different types of Neural networks. The model also shows Recall and Precision as indicators that can be used for measuring the performance and evaluation of those algorithms. In addition, the model express the softgoals such as Speed of learning that need to be considered while using those algorithms. The model, through influence links from algorithms to softgoals, captures knowledge on how the algorithms is commonly known to perform with regard to those qualities. For example, it shows that Logistics regression is known to be a fast algorithm. Such catalogue is used for constructing (part of) the Analytics View model, such as the Classification goal in Figure 1.

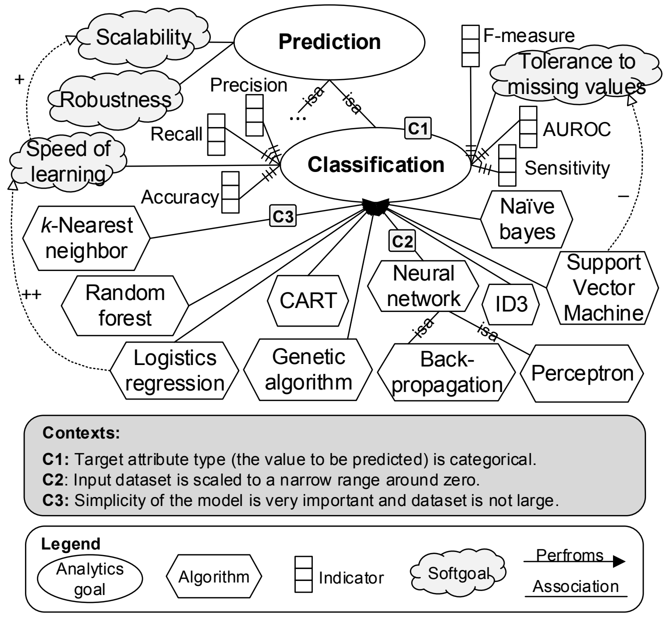

Fig. 3. A fragment of Algorithm Catalogue. To keep the model readable, not all Influence links and Contexts are shown here.

A design pattern is a three-part relation between a context, a problem, and a solution [21]. It provides a description of well-proven solutions to recurring design problems. In the framework, such patterns are captured within three kinds of design catalogues (introduced in Section III-B), mainly in terms of *Question Goals*, *Analytics Goals*, *Tasks/Algorithms*, *Indicators*, *Softgoals*, and *Means-End Links*. An important part of the catalogues is the representation of the knowledge on when to use what algorithms. In the framework, this is captured through the *Context* elements. For example, in Figure 3, the context C3 shows that *k*-Nearest neighbor is suitable to be used when simplicity of the model is very important and dataset is not large. On one hand, the catalogues represent formalized design solutions and best practices for commonly known business analytics problems. On the other hand, the framework provides a formal language to expert community to express their knowledge and collective experience and make it available to others.

V. Discussions

We demonstrated different ways in which the modeling framework can be used in two cases. Such examples serve as a *preliminary validation* of its expressiveness. We illustrated instances of models in three modeling views and described some of the analyses that they can enable. Such illustrations suggest that the framework can have a positive impact in the requirements analysis and design of analytics solutions.

Aside from the potential use cases of the framework, the case studies and involvement of the participant helped us to receive some feedback and learn about some limitations and discuss potential improvements of the framework:

- From a meta-model design perspective, currently the Business View Model only captures the decomposition type of link among decision goals. In reality, decisions can have other kinds of relationships such as followed by, triggers, and influences. Whether or not to extend the current metamodel is a design decision that needs further research and considerations from the organizational decision theory. We also need to investigate what kinds of new analysis would those potential additions add to the framework.
- In addition, within the Analytics Design View, all the indicators that are attached to the same analytics goal are treated equally. Through the case studies, we observed examples where the analytics metrics can have different degrees of importance and also can be conflicting (i.e., an increase in one would typically decrease the other one). This requires the framework to capture importance and priorities of the indicators and softgoals. Such extensions would enhance the expressiveness, ease the algorithm selection, and support justification on why a given algorithms was perceived better than others. On the other hand, the models may become more complex and harder to learn and use.
- The framework suggests strategic goals to be decomposed into one or more decision goals, and thereafter the decisions to be refined into one or more question goals. We observed that the participant (here in the role or analyst) draw some models where such sequence of decomposition is not followed. For example, we observed instances where a strategic goal was decomposed into some decision goals and thereafter into strategic goals. While each strategic goal and decision goal on its own was modeled correctly, the use of decomposition links was observed to be misinterpreted with sequence links. We also found that analysts might mix goals with meta-goals (goals about goals) all in the same diagram. Clear guidelines are needed in he framework to avoid such problems.
- In the course of the case studies, we identified that each goal (e.g., to increase x) is naturally paired with an implicit decisions (e.g., decision on how to increase x). This can be included in guidelines to help modelers to construct enriched models during the requirements elicitation activities.
- We observed that the analyst can encounter difficulties in labeling the question goals. Question goals symbolize the needs-to-know of actors towards decisions to be made. Labeling them correctly is essential for arriving at a set of accurate and precise analytical requirements; since they reveal the type of required analytics (predictive, descriptive, or prescriptive). Several guidelines were created to encounter such difficulties. For example, at the leaf of the model, each question goal label should start with one of the six types (what, who, when, where, why, and how) and cannot start with phrases such as "is it". We faced similar issues with naming of the decision goals which resulted in additional labeling guidelines.
- The participant, having data science experience, were asked to develop new instances of the Algorithm Catalogue for some specific analytics areas such as clustering. In the initial modeling attempt, some difficulties were experienced because clustering as a kind of descriptive analytics, can be performed with various objectives. We observed that catalogues need to be separated (initially by the top goal) and can be categorized based on expertise of the eventual user. We found that the formal semantics of catalogues need to be further developed and guidelines for creating catalogues should also be prepared.

Several factors can impact the validity of the findings and limit the generalizability of observations in this paper. *First,* while the testing of the framework was conducted initially by a participant who was not involved in the development of the framework, the authors subsequently assisted the participant in revising the models during several weekly meetings. The modeling was performed by the participant as part of an individual studies course supervised by one of the authors. The content of models were modified and syntactical issues were resolved during those meetings and after. *Second,* the case studies in this paper did not involve any real business stakeholder(s) of those cases. As a result, the findings in this paper are mostly reported in the form of *potentials* which need further validations. *Third,* the benefits and limitations that were discussed are by no means comprehensive. The study involved only one participant and the findings in the paper mostly relate to only two (out of three) modeling views.

VI. Related Work

Modeling for Data Warehouses. Some works focus on developing modeling the requirements for data warehouses. Authors in [22] propose the GDI (Goal-Decision-Information) model for analyzing data warehouse requirements. They develop a decision requirements metamodel [23] and use informational scenarios [24] to elicit data warehouse requirements. The work in [25] proposes a goal-oriented approach to requirement analysis of data warehouses, based on the Tropos methodology. The framework in this paper is different in the sense that it focuses on requirements analysis and design of advanced analytics and machine learning solutions.

Modeling for Business Intelligence (BI). These works propose modeling approaches for developing BI solutions. The Business Intelligence Model (BIM) language represents enterprise in term of strategies, processes, indicators adn more to bridge the gap between business and data [5]. Authors in [26] extend BIM metamodel to support modeling and reasoning on business plans. The work in [27] extends BIM to enable stress testing of business strategies. The framework in this paper extends the BIM language by introducing new concepts (such as question goals, decision goals, insights, algorithms, and operators) and design catalogues to support requirements analysis and design of advanced analytics solutions.

Data Mining Ontologies. Some works propose formal ontologies to support users during data mining projects. For example, the work in [28] for supporting users at various choice points of the data mining process. Such ontologies do not capture concepts relevant to business requirement such as actors, goals, softgoals, and influences.

Information Systems Research on Analytics. Data analytics has increasingly attracted the interest of information systems (IS) research community [29]. An important part of this body of literature focuses on the usage and impact of analytics on the organization and society. For example, [17] provides a research agenda for understanding the relationship between business analytics, decision making processes, and organizational performance. These contributions are in terms of a set of general managerial principles and guidelines, towards theories. There is a lack of enterprise models that allow for analysis and design of data analytics solutions.

Existing Tools. A number of (commercial) software and platforms exist for performing analytics, including IBM Watson Analytics, Microsoft Azure ML, SAS, etc. While these tools automate and facilitate data preparation and experimentation with (machine learning) algorithms, they do not support business and requirements aspect of analytics solutions.

A version of the modeling framework was presented in [7]. In this paper, we provided a detailed illustration of the usage of the framework in two case studies and the potential benefits, as well as limitations and shortcomings in preliminary testing.

VII. Conclusions and Future Work

This paper introduced a conceptual modeling framework for business analytics and illustrated some of its potential benefits in two cases. The cases were used as a preliminary validation of framework's expressiveness and as a means to show potential use cases and to uncover limitations of the approach. We are currently involved in two collaborations with industrial partners to validate the framework and improve it. Such collaborations would also allow us to understand who would use what modeling view(s), how, and when. Those findings would lead to development of a methodology for using such a framework. Future work includes investigating and improving the usability and learnability of the notation and method. Practical applicability of the framework may require special training on the syntax and semantics of modeling views which needs to be investigated in future. We are also interested in developing tools for supporting different aspects of the framework.

References

[1] S. Ransbotham, D. Kiron, and P. K. Prentice, "Beyond the hype: the hard work behind analytics success," *MIT Sloan Management Review*, vol. 57, no. 3, 2016.

[2] S. LaValle, M. S. Hopkins, E. Lesser, R. Shockley, and N. Kruschwitz, "Analytics: The new path to value," *MIT Sloan Management Review*, vol. 52, no. 1, pp. 1–25, 2010.

[3] E. Kandogan, A. Balakrishnan, E. M. Haber, and J. S. Pierce, "From data to insight: work practices of analysts in the enterprise," *IEEE computer graphics and applications*, vol. 34, no. 5, pp. 42–50, 2014.

[4] S. Viaene and A. Van den Bunder, "The secrets to managing business analytics projects," *MIT Sloan Management Review*, vol. 53, p. 65, 2011.

[5] J. Horkoff, D. Barone, L. Jiang, E. Yu, D. Amyot, A. Borgida, and J. Mylopoulos, "Strategic business modeling: representation and reasoning," *Software & Systems Modeling*, vol. 13, no. 3, pp. 1015–1041, 2014.

[6] E. Yu, "Modelling strategic relationships for process reengineering," *Social Modeling for Requirements Engineering*, vol. 11, p. 2011, 2011.

[7] S. Nalchigar, E. Yu, and R. Ramani, "A conceptual modeling framework for business analytics," in *Conceptual Modeling: 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings 35*. Springer, 2016, pp. 35–49.

[8] I. Kolyshkina and S. Simoff, "Customer analytics projects: addressing existing problems with a process that leads to success," in *Proceedings of the sixth Australasian conference on Data mining and analytics-Volume 70*. Australian Computer Society, Inc., 2007, pp. 13–19.

[9] J. Sullivan, "Get the right data scientists asking the wrong questions," March 2014, Harvard Business Review.

[10] T. H. Davenport, J. G. Harris, W. David, and A. L. Jacobson, "Data to knowledge to results: building an analytic capability," *California Management Review*, vol. 43, no. 2, pp. 117–138, 2001.

[11] L. Fahey, "Exploring analytics to make better decisions–the questions executives need to ask," *Strategy & Leadership*, vol. 37, no. 5, pp. 12–18, 2009.

[12] M. Luca, J. Kleinberg, and S. Mullainathan, "Algorithms need managers, too," *Harvard business review*, vol. 94, no. 1, p. 20, 2016.

[13] J. Horkoff and E. Yu, "Comparison and evaluation of goal-oriented satisfaction analysis techniques," *Requirements Engineering*, vol. 18, no. 3, pp. 199–222, 2013.

[14] N. Chandler, B. Hostmann, N. Rayner, and G. Herschel, "Gartners business analytics framework," 2011.

[15] T. H. Davenport, B. E. Bensoussan, and C. S. Fleisher, *The Complete Guide to Business Analytics (Collection)*. FT Press, 2012.

[16] G. G. Shanks, N. Bekmamedova, and L. P. Willcocks, "Business analytics: Enabling strategic alignment and organisational transformation," in *Proceeding of 20th European Conference on Information Systems*, 2012.

[17] R. Sharma, S. Mithas, and A. Kankanhalli, "Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations," *European Journal of Information Systems*, vol. 23, no. 4, pp. 433–441, 2014.

[18] R. Kohavi, L. Mason, R. Parekh, and Z. Zheng, "Lessons and challenges from mining retail e-commerce data," *Machine Learning*, vol. 57, no. 1-2, pp. 83–113, 2004.

[19] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers, "Big data: the next frontier for innovation, competition, and productivity," *The McKinsey Global Institute*, 2011.

[20] M. Yeomans, "What every manager should know about machine learning," *Harvard Business Review*, vol. 93, no. 7, 2015.

[21] F. Buschmann, K. Henney, and D. Schimdt, *Pattern-oriented Software Architecture: On Patterns and Pattern Language*. John Wiley & Sons, 2007, vol. 5.

[22] N. Prakash and A. Gosain, "An approach to engineering the requirements of data warehouses," *Requirements Engineering*, vol. 13, no. 1, pp. 49–72, 2008.

[23] N. Prakash, D. Prakash, and D. Gupta, "Decisions and Decision Requirements for Data Warehouse Systems," in *CAiSE Forum*, ser. LNBIP, 2010, vol. 72, pp. 92–107.

[24] N. Prakash, Y. Singh, and A. Gosain, "Informational Scenarios for Data Warehouse Requirements Elicitation," in *ER 2004*, ser. LNCS, 2004, vol. 3288, pp. 205–216.

[25] P. Giorgini, S. Rizzi, and M. Garzetti, "Goal-oriented requirement analysis for data warehouse design," in *Proceedings of the 8th ACM international workshop on Data warehousing and OLAP*. ACM, 2005, pp. 47–56.

[26] F. Francesconi, F. Dalpiaz, and J. Mylopoulos, "TBIM: A Language for Modeling and Reasoning about Business Plans," in *ER 2013*, ser. LNCS, 2013, vol. 8217, pp. 33–46.

[27] A. Maté, J. Trujillo, and J. Mylopoulos, "Stress Testing Strategic Goals with SWOT Analysis," in *ER 2015*, ser. LNCS, 2015, vol. 9381, pp. 65–78.

[28] C. M. Keet, A. Ławrynowicz, C. dAmato, A. Kalousis, P. Nguyen, R. Palma, R. Stevens, and M. Hilario, "The data mining optimization ontology," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 32, pp. 43–53, 2015.

[29] R. Agarwal and V. Dhar, "Editorial-big data, data science, and analytics: The opportunity and challenge for is research," *Information Systems Research*, vol. 25, no. 3, pp. 443–448, 2014.