A Conceptual Modeling Framework
for Business Analytics

Soroosh Nalchigar[1(✉)], Eric Yu[1], and Rajgopal Ramani[2]

[1] Department of Computer Science, University of Toronto, Toronto, Canada
{soroosh,eric}@cs.toronto.edu
[2] Deloitte, Global Technology Services, Toronto, Canada
rramani@deloitte.ca

Abstract. Data analytics is an essential element for success in modern enterprises. Nonetheless, to effectively design and implement analytics systems is a non-trivial task. This paper proposes a modeling framework (a set of metamodels and a set of design catalogues) for requirements analysis of data analytics systems. It consists of three complementary modeling views: business view, analytics design view, and data preparation view. These views are linked together and act as a bridge between enterprise strategies, analytics algorithms, and data preparation activities. The framework comes with a set of catalogues that codify and represent an organized body of business analytics design knowledge. The framework has been applied to three real-world case studies and findings are discussed.

Keywords: Conceptual modeling · Data analytics · Machine learning · Goal-oriented requirements engineering · Business analytics

1 Introduction

The effective design and implementation of data analytics solutions has proven to be difficult. This difficulty is, in part, due to challenges such as determining the right analytics needs, utilizing the right analytics algorithms, as well as connecting them with high-level business objectives and strategies.

Requirements elicitation for data analytics systems is a complex task [12,27]. Analytics requirements are often unclear and incomplete at the early phases of projects. While business users often have a clear understanding of their strategic goals (e.g., improve marketing campaigns, reduce inventory levels), they are not clear on how analytics can help them achieve those goals. This is, to a great extent, due to a huge conceptual distance between business strategies, decision processes and organizational performance on one hand, and the implementation of analytics systems in terms of databases, preprocessing activities, and machine learning algorithms on the other hand. Previous researches report that the leading barrier to using analytics techniques is lack of understanding of how to use analytics and unlock its value to improve the business [17,19].

Moreover, *designing analytics solutions* includes making critical design decisions while taking into account softgoals and tradeoffs [18]. A large number of machine learning and data mining algorithms exists and new ones are being developed continuously. During analytics projects, one needs to make design choices such as what are potential algorithms that can address the problem at hand? What criteria should be considered to evaluate those algorithms? What/how data should be prepared to be used by algorithms? These decisions have important implications in several aspects of the eventual analytics solution, such as scalability, understandability, tolerance to noisy data and missing values.

On the other hand, *aligning analytics with business strategies* is critical for achieving value through analytics [14,17]. Lack of this alignment can result in unclear expectations of how analytics contribute to business strategies, lack of executive sponsorship, and analytics project failures. It is important for organizations to discover, justify, and establish why there is a need for the organization to allocate resources to analytics initiatives. Towards this end, discovering the business goals and translating them into analytics goals is a critical step [4,15].

This paper presents a modeling framework (i.e., a set of metamodels and a set of design catalogues) for overcoming these challenges. The framework includes three complementary modeling views: (i) The *Business View* represents an enterprise in terms of strategies, decisions, analytics questions, and required insights. This view is used to systematically elicit analytics requirements and to inform the types of analytics that the user needs. (ii) The *Analytics Design View* represents the core design of an analytics system in terms of analytical goals, (machine learning) algorithms, softgoals, and metrics. This view identifies design tradeoffs, captures the experiments (to be) performed with a range of algorithms, and supports algorithm selection. (iii) The *Data Preparation View* represents data preparation processes in terms of mechanisms, algorithms and preparation tasks. This view expresses the structure and content of data sources and the design of data preparation tasks. The three views are used together to link enterprise strategies to analytics algorithms and data preparation activities. The framework comes with three catalogues, each corresponding to a modeling view. Catalogues codify and represent reusable analytics knowledge for users.

Organization. Section 2 presents an illustration of the proposed framework in a real analytics project. Section 3 introduces primitive concepts and presents metamodels. Section 4 offers three analytics design catalogues. Section 5 discusses findings from applying the framework in three analytics projects. Section 6 reviews related work and Sect. 7 concludes the paper.

2   An Illustration

We illustrate the framework using a project aimed at developing an analytics system to predict upcoming software system outages. The company has around 300 globally accessible software applications hosted in its data centers across the world. Software system outages are costly and predicting them can enable preventive maintenance activities.
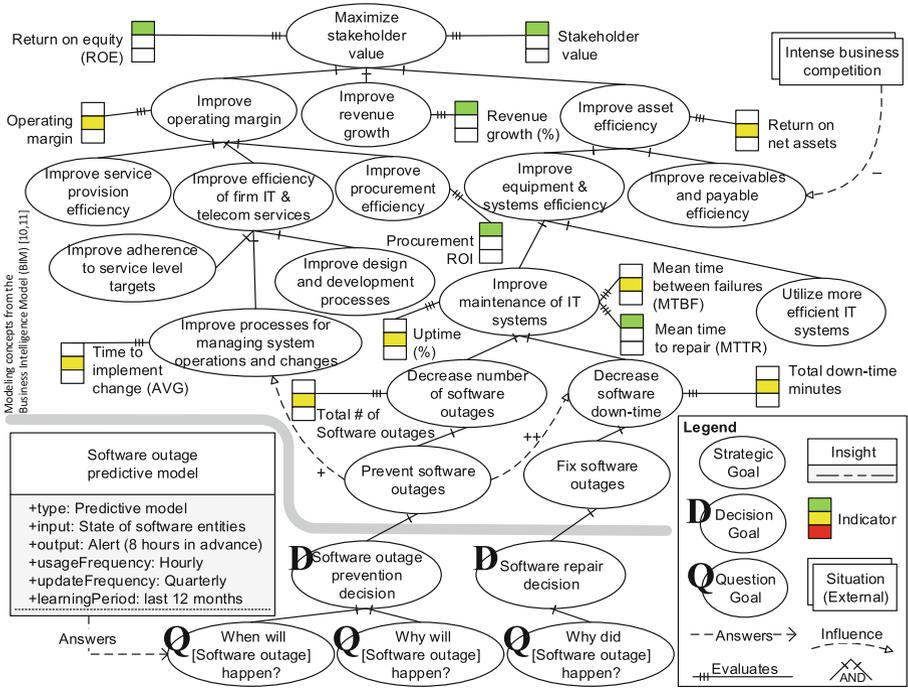
Fig. 1. Business View for the software outage prediction project (partial). This model is constructed based on interviews with domain experts, review of reporting dashboards and metrics in place, supplemented with some assumptions.

Figure 1 illustrates the *Business View* for the software outage prediction project. The purpose of this view is to represent the analytics needs of an organization and to ensure that those needs are driven by organizational decisions and strategies. This view models the business motivation for the analytics project in terms of its *strategic goals*, *indicators*, *decision goals*, *question goals*, and *insights*.

The model in Fig. 1 shows that Improve maintenance of IT systems is a strategic goal of the company. It also shows that Mean time between failures and Uptime (%) are among the indicators that the company uses to evaluate the goal. Strategic goals are decomposed into lower level strategic goals and eventually into decision goals. Software outage prevention decision is an example of a decision goal. The model indicates that in order to Prevent software outages, the corresponding actor[1] needs to decide on how to prevent a software from failing. Decision goals are further decomposed into question goals. When will [Software outage] happen? is an example of a question goal. The model depicts that for making Software outage prevention decision, the corresponding actor needs to know if a software outage will happen in the near future. Question goals are answered by insights. Software outage predictive model is an example of an insight to be generated by

---

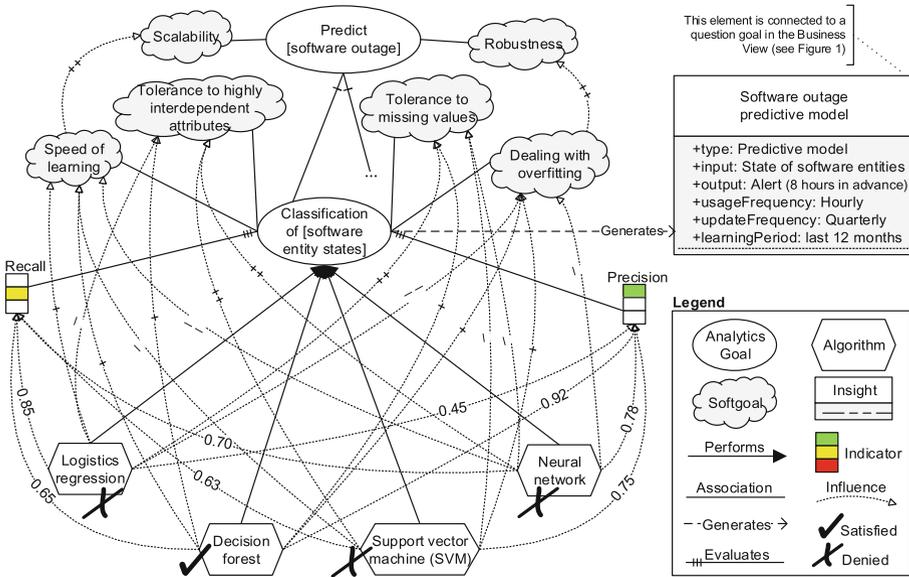[1] Actors are not shown here due to space limitations.

Fig. 2. Analytics Design View for software outage prediction project (partial).

the intended analytics system. It is a Predictive model that, in runtime, will be used Hourly to generate Alerts before an upcoming outage.

By modeling *decision goals*, this view represents the areas that need support from analytics insights. It ensures the connection between analytics, organizational decision processes, and strategic goals. This concept also facilitates linking and turning analytics-driven insights into actions, because the actions are indeed the decision outcomes. Through the *question goals*, the framework captures the business needs that the analytics work is intended to address. The catalogue of question goals (introduced in Sect. 4) can be used while performing modeling activities in this view. Eliciting the questions at the early phases of analytics will help perform the right analysis for the right user. Later in the analytics process and once the findings are generated, the questions can also facilitate the process of interpreting and framing the findings. By modeling *insights*, this view represents the knowledge that is extracted from data for answering the questions. The insight elements connect business view to analytics design view.

Figure 2 illustrates the *Analytics Design View* for the software outage prediction project. The purpose of this view is to represent the design of the analytics system, including algorithm selection. This view models an analytics system in terms of *analytics goals*, *(machine learning) algorithms*, *softgoals*, and *indicators*.

In Fig. 2, Predict software outage is an example of an analytics goal. To achieve this goal, the system needs to achieve the Classification of software entity states

goal[2]. The model shows that Neural networks and Decision forest are alternative algorithms that perform classification. Moreover, the model represents the contributions from algorithms towards indicators and softgoals. For example, the link from Decision Forest algorithm towards Precision means that during experiments, the algorithm resulted in the value of 0.92 for Precision. Also, this algorithm has a positive contribution towards the Speed of learning. By capturing these, the view supports algorithm selection while designing the analytics systems[3]. The algorithms catalogue (introduced in Sect. 4) assists users in this modeling view and supports designing analytics systems. The analytics goals connect this view to the data preparation view.
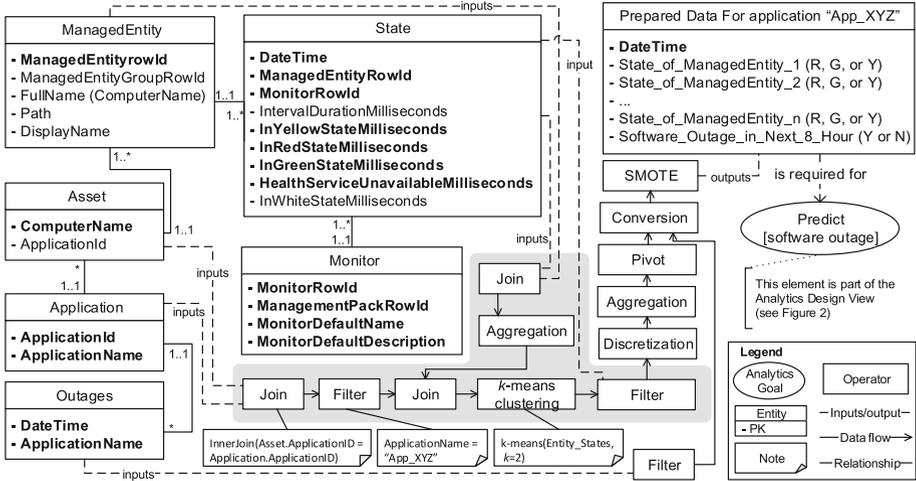

Fig. 3. Data Preparation View for software failure prediction project (partial).

Figure 3 illustrates the *Data Preparation View* for the software outage prediction project. The purpose of this view is to support the design and documentation of data preparations workflows. This view models data preparation processes in terms of *entities*, *attributes*, *mechanisms*, *algorithms*, and *preparation tasks*.

The model in Fig. 3 shows the content and structure of data sources[4]. It shows that an Application is related to many Assets and each asset in turn can have many ManagedEntities. The State data captures the status of the software

---

[2] There were several instances of classification goal that each addressed a specific prediction period, such as 8, 16, 24 h. Each of the goals is connected to a different instance of the insight element. Due to space limitations, only one pair of analytics goal and insight is illustrated here.

[3] In the first case study, the indicator Precision had highest priority which justified the choice of Decision Forest for the corresponding classification goal.

[4] The company has a cross-platform data center management system that logs computer systems operations.

entities over time. The model shows the sequence of data preparation mechanism and algorithms. Join and Filter are examples of mechanism. SMOTE is an example of an algorithm for data preparation. A set of mechanisms and algorithms together form a data preparation task. In Fig. 3 the gray shaded area shows a Data numerosity reduction task. This task is responsible for removing managed entities whose State data is not showing any meaningful relationship with software outage. The *k*-means clustering is an example of an algorithm which, in this case, performs the main part of the data reduction task. The main outcome of the workflows is the prepared dataset that is required for the analytics goal To predict [software outage]. The data preparation catalogue (see Sect. 4) assists users in this modeling view and supports designing data preparation workflows.

3   Metamodels

3.1   Business View

Figure 4 shows metamodel of the Business View in terms of a UML class diagram. Concepts in the gray shaded area are adopted from the Business Intelligence Model (BIM) [10,11]. Here we explain concepts that are added to extend BIM.
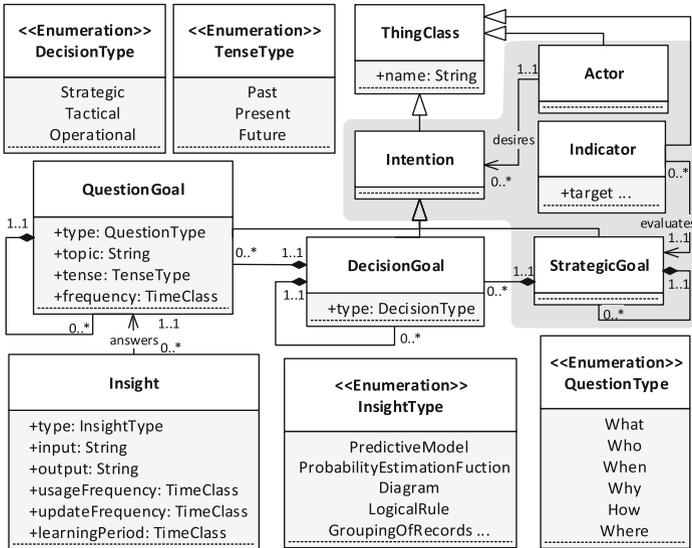

Fig. 4. Part of metamodel for the Business View.

Decision Goals. This concept represents intention of an actor for taking actions towards achieving strategic goals. Strategic goals can be decomposed into one or more decision goals.

Question Goals. This concept represents the desire of an actor for understanding or knowing something that is required for making decisions (i.e., achieving decisions goals). It captures "needs to know" of an actor. Decision goals are decomposed into one or more questions. Questions can be refined into one or more questions.

Question goals are analyzed into a *type* and *topic* as in NFR framework [5], and also *tense* (see the metamodel in Fig. 4). The question type denotes the question phrase (e.g., When in Fig. 1), while the question topic denotes the subject and focus of the (intended) analysis (e.g. [Software outage] in Fig. 1). The question tense captures the time horizon that a question goal addresses. Elicitation of question type and tense together allows specifying what kinds of analytics and machine learning algorithms are required as part of the intended system. Moreover, identification of topic allows specifying what kind of data (or what parts of database) will the intended analytics system use for mining. In addition, as shown in Fig. 4, question goals are specified in terms of their *frequency*. This attribute captures time scales and frequencies that the corresponding question is being raised. High frequency analytics question have more potential to be embedded into automated analytics systems and tools [21].

Insights. This concept represents a structured, (machine) understandable pattern (i.e., relationship among data) that is extracted from data by applying analytics algorithms. It represents a piece of information/knowledge that (partially) answers a question goal, and thereafter facilitates decision making and contributes to strategic goals. This concept has the following subtypes: *Predictive model*, *Probability Estimation Function*, *Diagrams* (e.g., trees, graphs), *Logical Rule* (e.g., association rules) and *Groupings of Records* (e.g., clusters). This concept connects to the question goals through the *answers* link. It represents the immediate output of the data analytics activities.
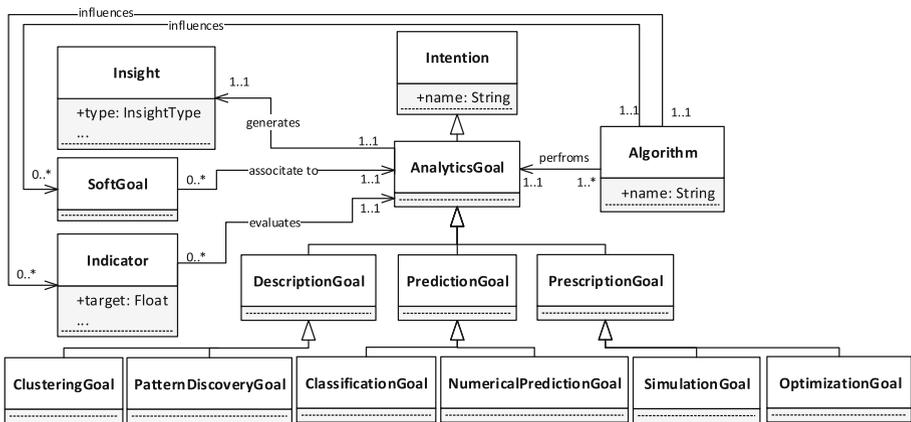

Fig. 5. Part of metamodel for the Analytics Design View.

3.2 Analytics Design View

Analytics Goals. This concept (see the metamodel in Fig. 5) represents the top-goal of the data analytics system, i.e., to extract insight from data. Analytics goals connect to insights via the link *generates*. There are three types of analytics goals. *Prediction Goal* represents an intention to predict value of a target data attribute (i.e., label attribute) by using other existing attributes in the dataset. It shows the desire to find the relationship between the target feature and other existing features in the dataset. Two subtypes of this concept are *Classification* (predicts categorical values) and *Numeric Prediction. Description Goal* represents an intention to summarize and describe the dataset and includes two subtypes: *Clustering* and *Pattern Discovery. Prescription Goal* represents an intention to find the optimal alternative among a set of potential alternatives. *Optimization* and *Simulation* are subtypes of prescription goals.

Algorithms. This concept represents a procedure that addresses an analytics goal. An algorithm is a set of steps that are necessary for an analytics goal to be achieved. It is a way through which insight is extracted from data in order to satisfy an analytics goal. This concept is connected to analytics goal through the *performs* links, representing a means-end relationship.

Indicators and Softgoals. Indicators [10] are numeric metrics that measure performance with regard to some goal (analytics goal in this modeling view). Softgoals [28] capture qualities that should sufficiently hold when performing analytics. Algorithms connect to indicators and softgoals through the *influence* links. Influence links that are directed towards an indicator, can be labeled with the corresponding numeric value.Contributions that are directed towards qualities can range from positive to negative, following i^* guidelines [28].

Analytics projects involve experimenting with different algorithms. During design time, indicators and softgoals represent criteria to be considered for evaluation/comparison of alternative algorithms that perform the analytics task at hand. They can be used to reduce the domain of experiments. During runtime they can be used for monitoring the performance of the running analytics system.

3.3 Data Preparation View

Data Preparation Tasks. This concept (see the metamodel in Fig. 6) represents the general task of preparing the data that is required for achieving some analytics goal. A data preparation task consists of one or more *Operator(s)*. It has four subtypes [9]: *Data reduction* generates a data set that is smaller in size than the input data set and yet produces the same analytical results (i.e., serves the same analytical goals). *Data numerosity reduction* (see an example in Fig. 3) and *Data dimensionality reduction* are two types of data reduction tasks. *Data cleaning* represents the tasks that remove errors from the input dataset and also treat missing values in it. *Clean missing value* and *Clean noisy attribute* are subtypes of this concept. *Data transformation* transforms the shape of data in a
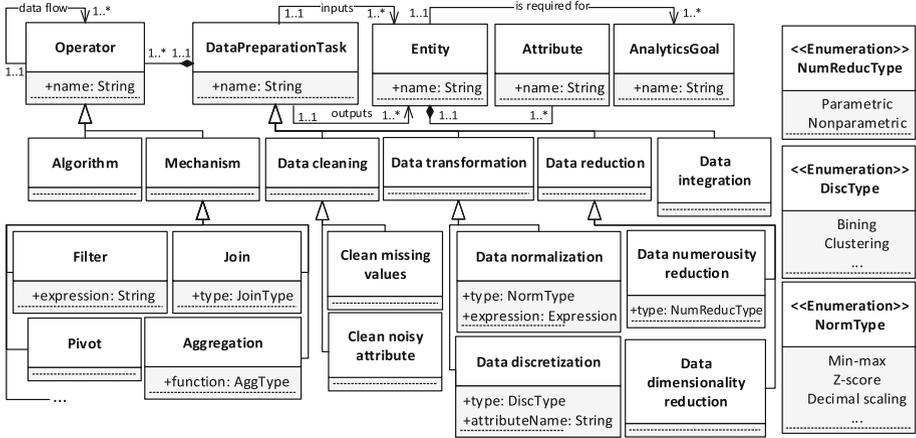
Fig. 6. Part of metamodel for the Data Preparation View.

way that is more appropriate for analytics algorithms to mine and find patterns. *Data normalization* and *Data discretization* are subtypes of this concept. *Data integration* merges data from different data sources.

Operator. It represents an atomic activity that performs (part of) a data preparation task. Operators are linked by *data flows* to represent the sequence. There are two types of operators. *Mechanism* represents fundamental data preparation operations such as *Join* and *Filter* [6,24]. *Algorithm* is identical with algorithm in the previous view. In the data preparation view, this concept captures situations where machine learning algorithms are used for preparing data, and not for performing the actual analytics task (see examples in Fig. 3).

Table 1. High-level structure of question goals catalogue. Due to space limitations, instances of each category of question goals are not provided here.

Type	Tense		
	Past	*Present*	*Future*
What	What happened?✍	What is happening?✍➻	What will happen?✱➻
Who	Who was involved in it?✍	Who is involved in it?✍	Who will be involved?✱
When	When did it happened?✍	Is it happening now?✍✱	When will it happen?✱
Where	Where did it happen?✍	Where is it happening?✍	Where will it happen?✱
Why	Why did it happen?✍	Why is it happening?✍	Why will it happen?✱➻
How	How did it happen?✍	How is it happening?✍	How will it happen?✱➻

Symbols ✍,✱, ➻ refer to analytics goal of description, prediction, and prescription respectively.

4   Cataloguing Analytics Design Knowledge

The proposed framework comes with three kinds of design catalogues. These catalogues bring relevant analytics knowledge to the attention of the project team for use and re-use during the design and development processes. They provide an organized body of analytics knowledge, accumulated from surveys (e.g., [16]), textbooks (e.g., [9]), formal ontologies (e.g., [25]), and previous experiences.

Business Questions Catalogue. This catalogue represents knowledge about the types of question goals, and their associated analytics types. It categorizes question goals based on their *type* and *tense* (see Sect. 3.1) and associates each category with relevant analytics goal(s). Table 1 presents the high level schema of the catalogue. This catalogue is populated with a wide collection of instances for each category of questions goals. For example, the question goal of Who will be [leaving the firm]? belongs to the Who will be involved in it? category in Table 1, and can be addressed by Prediction type of analytics. As another example, the question goal of When will [Software outage] happen? from Fig. 1, belongs to the When will it happen? category in Table 1. This catalogue can be used by analytcis team and stakeholders during the modeling activities of business view. It can facilitate the elicitation of analytics requirements (i.e., needs to know) by suggesting and refining question goals. It also guides users to the kinds of analytics solutions that can address their needs.

Algorithms Catalogue. This catalogue systematically organizes machine learning algorithms that are available for addressing different types of analytics goals. The catalogue provides existing metrics to be taken into account while comparing/evaluating performances of different algorithms. It also presents critical softgoals that need to be taken into account while developing analytics solutions. In addition, it encodes the knowledge on how each algorithm perform with regard to different softgoals (influence links). A portion of this catalogue is illustrated in Fig. 7. As an example, it shows that Support Vector Machine (SVM) is an algorithm that performs Numeric prediction and its performance can be evaluated using the Mean Absolute Error (MAE) metric.

The *context* semantics from [1] are used to associate context with machine learning algorithms. In this way, the catalogue represents when certain machine learning algorithms are shown to perform well based on a collection of previous evidences and experiments in the literature or relevant sources. This can guide the decision on which algorithms are more appropriate for the analytics goal and shorten the experimentation phase of the projects. In Fig. 7, context C1 shows that the Classification goal is activated when Target attribute type (the value to be predicted) is categorical. On the other hand, C2 shows that Neural network can be used for Numeric prediction, when Input dataset is scaled to a narrow range around zero. Due to space limitations, not all the contexts are given in Fig. 7.

Data Preparation Techniques Catalogue. This catalogue captures knowledge on available methods for different types of data preparation tasks. It makes use of the same modeling elements as in the algorithm catalogue. As shown in
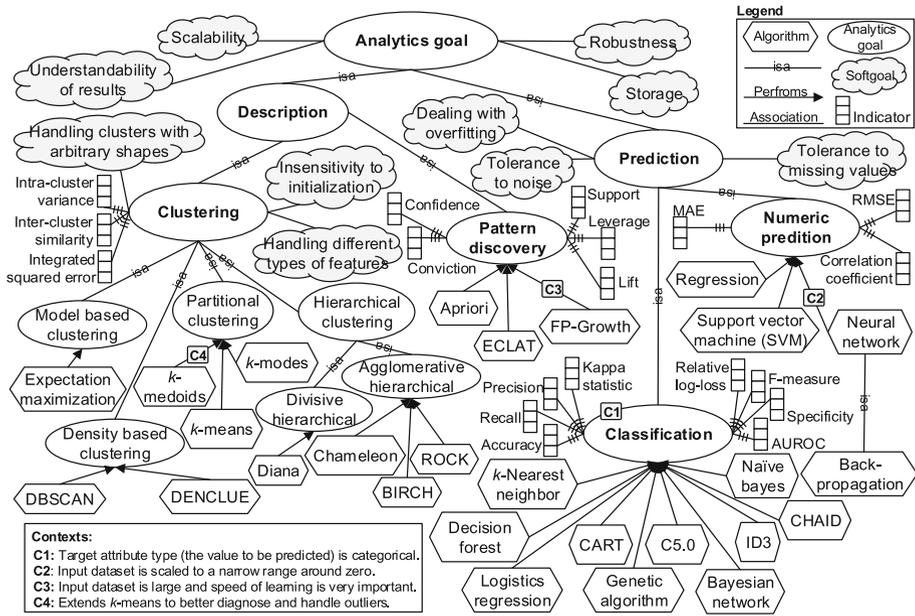
Fig. 7. A portion of algorithm catalogue. Influence links from algorithms towards soft-goals are not shown here to keep the model readable.
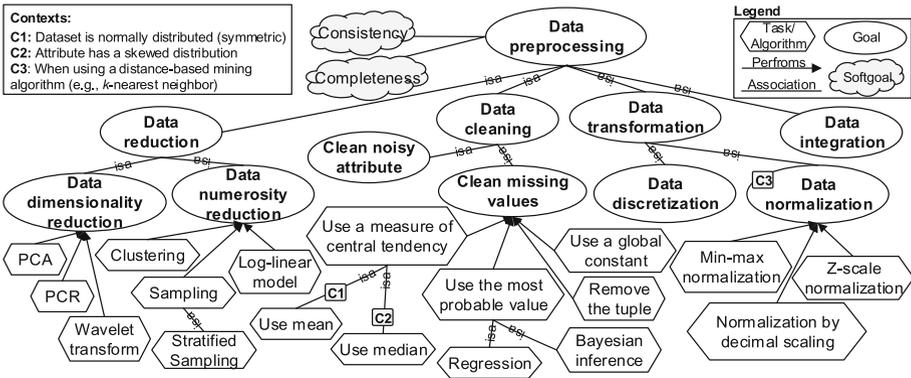

Fig. 8. A portion of data preparation catalogue. Not all the contexts are shown.

Fig. 8, Using median is a method for Cleaning missing values when the corresponding Attribute has a skewed distribution. Analytics development team can browse through this catalogue and design data preparation workflows.

5 Case Studies

The proposed framework has been applied to three analytics projects. The first two case studies were reconstructions of completed projects. The third case study was an application of the framework to an on-going analytics project. These cases together serve as an initial validation of the framework. In Sect. 2, we used the first case study for illustrating the modeling views. The *second project* focused on finance analytics. The purpose of this project was to predict an upcoming event regarding financial metrics in company's network. The *third project* focuses on search engine analytics. The purpose of this project is to use analytics to provide query suggestions to online users.

Our main observation from the first and second cases is that the modeling views together provide an adequate set of concepts for connecting strategic goals to analytics algorithms and data preparation activities. The three modeling views were instantiated for these case studies, presented to and understood by stakeholders. We observed that the framework can be used for representing analytics requirements, can show design tradeoffs and support algorithm selection, can capture data preparation activities, and can represent the alignment between analytics systems and business strategies.

Our main observation from the third case is that the framework can be useful in guiding analytics projects. A model from business view was constructed, in collaboration with stakeholders, at the requirements elicitation phase of the project. While at the beginning the focus of the project was broad and imprecise (to use analytics for improving users' search experience), the models effectively helped the team to narrow down the scope and reach an agreement about the "to-be" analytics system (to use analytics to provide query suggestions). We observed that users are able to understand the content of the model and can work with analytics team to construct and elaborate on the models. The models raised effective discussions during meetings and resulted in removing some and adding new question goals. These suggest that the framework can enhance the communication between domain experts and data scientists (who develop analytics systems). Models from data analytics design view were constructed and updated during the project, mostly by the project manager and data scientists. The softgoals (most importantly Scalability) were used for making design decisions.

6 Related Work

Conceptual Modeling for Data Warehouses. These works propose conceptual modeling approaches for requirements engineering of data warehouses. For example, the work in [20] proposes a goal-oriented, model-driven approach for development of data warehouses. Authors in [23] propose goal-decision-information model for analyzing data warehouse requirements. Reference [8] proposes a Tropos-based methodology for requirements analysis in data warehouses. While we adopt some of concepts from these works (e.g., decision goals

in [20,23]), the proposed framework supports requirements engineering for predictive and prescriptive types of analytics systems, in addition to descriptive ones.

Conceptual Modeling for ETL Processes. These works propose conceptual modelings for ETL (Extraction-Transformation-Loading) processes. The work in [26] presents a metamodel and notation for modeling ETL processes in the early stages of data warehouse projects. In [24] authors define a set of common ETL activities in terms of stereotyped classes and use UML dependencies to link them together. Reference [22] defines a model–driven architecture approach to transform ETL conceptual models to code. In [6], a BPMN-based modeling approach for ETL processes is presented. While the proposed framework reuses modeling constructs from these works (e.g., mechanism from [24]), it captures machine learning and organizational aspects of analytics solutions.

Modeling for BI. The Business Intelligence Model (BIM) [11] represents a business in terms of strategic goals, processes, performance indicators, influences, and situations. BIM supports a wide range of automated reasoning and business analyses techniques [2,10]. It is shown that the language can facilitate design and development of BI solutions [3]. BIM lacks primitive concepts for supporting design of advanced analytics solutions. This work uses and extends the modeling constructs to capture analytics work from data preparation tasks to algorithms, and thereafter to insights and question goals.

Data Mining Process Models. These models describe the sequence of tasks that should be done in order to carry out data mining projects. The work by Fayyad et al. [7] is often considered as the first reported data mining process model. The CRISP-DM model [4] is often mentioned as the most used and the *de facto* standard process model. These works do not offer a modeling language.

Data Mining Ontologies. Several efforts have been made to establish formal ontologies for supporting users during data mining processes. For example, references [13] propose ontologies for facilitating algorithm selection and designing the data mining workflows. The ontology in [25] formally represents data mining experiments to enable meta-learning. Concepts that express business and requirements aspect of analytics solutions are not included in these works.

7   Conclusion

This paper presented initial research results towards a conceptual modeling framework for business analytics. The framework has been tested in three case studies. The case studies suggest that the proposed framework can support the design and implementation of analytics solutions. It is notable that all these case studies belong to a single domain and company. In future we plan to extend the framework and evaluate it in different domains, completing other pieces of the design science research approach. We plan to conduct empirical studies with users who are not the researchers. Usage, comprehensibility and learning curve

of the modeling views can be examined for different types of roles (from business decision makers to data scientists) that are typically involved in analytics projects. These studies can lead to definition of a model-based methodology, as part of the framework, for developing analytics systems. The content of analytics catalogues can be extended, validated, and their usage can be examined in real cases. We also plan to develop tools that support the framework.

References

1. Ali, R., Dalpiaz, F., Giorgini, P.: A goal-based framework for contextual requirements modeling and analysis. Requirements Eng. **15**(4), 439–458 (2010)
2. Barone, D., Jiang, L., Amyot, D., Mylopoulos, J.: Composite indicators for business intelligence. In: Jeusfeld, M., Delcambre, L., Ling, T.-W. (eds.) ER 2011. LNCS, vol. 6998, pp. 448–458. Springer, Heidelberg (2011). doi:10.1007/978-3-642-24606-7_35
3. Barone, D., Topaloglou, T., Mylopoulos, J.: Business intelligence modeling in action: a hospital case study. In: Ralyté, J., Franch, X., Brinkkemper, S., Wrycza, S. (eds.) CAiSE 2012. LNCS, vol. 7328, pp. 502–517. Springer, Heidelberg (2012). doi:10.1007/978-3-642-31095-9_33
4. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-Step Data Mining Guide. SPSS Inc. (2000)
5. Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J.: Non-functional Requirements in Software Engineering. Springer Science & Business Media, New York (2012)
6. Akkaoui, Z., Mazón, J.-N., Vaisman, A., Zimányi, E.: BPMN-based conceptual modeling of ETL processes. In: Cuzzocrea, A., Dayal, U. (eds.) DaWaK 2012. LNCS, vol. 7448, pp. 1–14. Springer, Heidelberg (2012). doi:10.1007/978-3-642-32584-7_1
7. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Mag. **17**(3), 37–54 (1996)
8. Giorgini, P., Rizzi, S., Garzetti, M.: GRAnD: a goal-oriented approach to requirement analysis in data warehouses. Decis. Support Syst. **45**(1), 4–21 (2008)
9. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Elsevier, Waltham (2012)
10. Horkoff, J., Barone, D., Jiang, L., Eric, Y., Amyot, D., Borgida, A., Mylopoulos, J.: Strategic business modeling: representation and reasoning. Softw. Syst. Model. **13**(3), 1015–1041 (2014)
11. Jiang, L., Barone, D., Amyot, D., Mylopoulos, J.: Strategic models for business intelligence. In: Jeusfeld, M., Delcambre, L., Ling, T.-W. (eds.) ER 2011. LNCS, vol. 6998, pp. 429–439. Springer, Heidelberg (2011). doi:10.1007/978-3-642-24606-7_33
12. Kandogan, E., Balakrishnan, A., Haber, E.M., Pierce, J.S.: From data to insight: work practices of analysts in the enterprise. IEEE Comput. Graphics Appl. **34**(5), 42–50 (2014)
13. Keet, C.M., Lawrynowicz, A., dAmato, C., Hilario, M.: Modeling issues and choices in the Data Mining OPtimization Ontology. In: OWLED 2013, Montpellier, France, May 2013
14. Kohavi, R., Mason, L., Parekh, R., Zheng, Z.: Lessons and challenges from mining retail e-Commerce data. Mach. Learn. **57**, 83–113 (2004)

15. Kohavi, R., Rothleder, N.J., Simoudis, E.: Emerging trends in business analytics. Commun. ACM **45**(8), 45–48 (2002)
16. Kotsiantis, S.B.: Supervised machine learning: a review of classification techniques. Informatica **31**(3) (2007)
17. LaValle, S., Hopkins, M.S., Lesser, E., Shockley, R., Kruschwitz, N.: Analytics: the new path to value. MIT Sloan Manag. Rev. (2010)
18. Luca, M., Kleinberg, J., Mullainathan, S.: Algorithms need managers, too. Harvard Bus. Rev. **94**, 96–101 (2016)
19. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: the next frontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute (2011)
20. Mazón, J.-N., Pardillo, J., Trujillo, J.: A model-driven goal-oriented requirement engineering approach for data warehouses. In: Hainaut, J.-L., et al. (eds.) ER 2007. LNCS, vol. 4802, pp. 255–264. Springer, Heidelberg (2007). doi:10.1007/978-3-540-76292-8_31
21. Menzies, T., Zimmermann, T.: Software analytics: so what? IEEE Softw. **30**(4), 31–37 (2013)
22. Muñoz, L., Mazón, J.-N., Trujillo, J.: Automatic generation of ETL processes from conceptual models. In: DOLAP 2009, pp. 33–40 (2009)
23. Prakash, N., Gosain, A.: An approach to engineering the requirements of data warehouses. Requirements Eng. **13**(1), 49–72 (2008)
24. Trujillo, J., Luján-Mora, S.: A UML based approach for modeling ETL processes in data warehouses. In: Song, I.-Y., Liddle, S.W., Ling, T.-W., Scheuermann, P. (eds.) ER 2003. LNCS, vol. 2813, pp. 307–320. Springer, Heidelberg (2003). doi:10.1007/978-3-540-39648-2_25
25. Vanschoren, J., Blockeel, H., Pfahringer, B., Holmes, G.: Experiment databases - a new way to share, organize and learn from experiments. Mach. Learn. **87**(2), 127–158 (2012)
26. Vassiliadis, P., Simitsis, A., Skiadopoulos, S.: Conceptual modeling for ETL processes. In: DOLAP 2002, pp. 14–21 (2002)
27. Viaene, S., Van den Bunder, A.: The secrets to managing business analytics projects. MIT Sloan Manag. Rev. **53**(1), 65–69 (2011)
28. Yu, E.: Modelling strategic relationships for process reengineering. Ph.D. thesis, University of Toronto, Canada (1995)