Complex Systems Informatics and Modeling Quarterly (CSIMQ)
eISSN: 2255-9922
Published online by RTU Press, https://csimq-journals.rtu.lv
Article 215, Issue 39, June/July 2024, Pages 65–85
https://doi.org/10.7250/csimq.2024-39.04

Requirements Template for Analytics Projects

Aritha Kumarasinghe and Marite Kirikova*

Institute of Applied Computer Systems, Riga Technical University,
6A Kipsalas Street, Riga, LV–1048, Latvia

balasuriyage-aritha-dewnith.kumarasinghe@rtu.lv, marite.kirikova@rtu.lv

Abstract. Data analytics projects have become a common accomplishment in many enterprises. However, establishing a data analytics project requires consideration of many factors that are not always recognized at the very beginning of the project. This study seeks to identify what generic requirements must be defined for data analytics projects and what analytics project attributes need to be addressed by these requirements. It proposes a requirements template for the generic requirements of the analytics projects. The template is intended to be used to reduce the complexity of starting the analytics projects by providing a checklist of requirements to be considered at the beginning of the project. The template is derived from analyzing 16 data analytics project reports for descriptive, diagnostic, predictive, and prescriptive analytics tasks. The template is then validated by analyzing its compliance with 20 analytics projects within the real estate domain using the corresponding research articles.
Keywords: Requirements, Data Analytics, Requirements Template, Real Estate Analytics.

1 Introduction

Data analytics is the field of study that systematically analyzes real-world systems using mathematical and statistical techniques on data that represents the systems in question [1]. There are several classifications used to differentiate data analytics projects [1], [2], [3]. In this work, we have applied one of the most common classifications [1] with the following four types of analytics:

1. Descriptive Analytics: Utilizes data to depict a system's or object's past states.
2. Diagnostic Analytics: Engages with near real-time data to explore a system's or object's current states.
3. Predictive Analytics: Aims to forecast a system's or object's future states by predicting potential events.
4. Prescriptive Analytics: Based on anticipated future events, seeks to determine future courses of action.

* Corresponding author

Reference: A. Kumarasinghe and M. Kirikova, "Requirements Template for Analytics Projects," Complex Systems Informatics and Modeling Quarterly, CSIMQ, no. 39, pp. 65–85, 2024. Available: https://doi.org/10.7250/csimq.2024-39.04

Additional information. Author ORCID iD: A. Kumarasinghe – https://orcid.org/0009-0000-5915-8919, M. Kirikova – https://orcid.org/0000-0002-1678-9523. PII S225599222400215X. Article received: 7 February 2024. Accepted: 27 May 2024. Available online: 31 July 2024.

All four analytics types are considered to ensure that the created generic requirements template would apply to most analytics projects.

When seeking the generic requirements template, in this research, the following analytics project phases were considered to structure the requirements template:

1. Initiation: The project's outset involves stakeholders defining goals, project plans, and initial requirements.
2. Acquisition: Involves gathering data from diverse sources and transforming it into an analyzable format.
3. Analysis: Application of data analysis techniques to glean insights from the acquired data.
4. Presentation: Conveying insights generated during analysis to project stakeholders.

The requirements defined in the initiation phase influence the rest of the data analytics project. Therefore, it is essential to have a full set of requirements to ensure that the phases that follow are as smooth as possible. Establishing such requirements from scratch may require dealing with high complexity due to the diverse issues to be considered and the ambiguity in stakeholder expectations regarding the results [4]. To lower the level of complexity, this article aims to establish a generic requirements template for data analytics projects. The template is elicited by analyzing already accomplished data analytics projects so that their commonalities can be transferred into a generic requirements template applicable in the initial phases of data analytics projects to aid with the definition of project requirements by providing a checklist of the minimum list of requirements that need to be defined to complete a data analytics project. Using the template would allow the analytics project team to gain a running start in the requirements engineering process in the data analytics project.

The ISO/IEC/IEEE International Standard – Systems and software engineering – Life cycle processes – Requirements engineering, in ISO/IEC/IEEE 29148:2018(E) [5] defines a requirement as a statement that translates or expresses a need and its associated constraints and conditions. In data analytics projects, a special emphasis should be placed on defining the constraints that relate to a data analytics project, such as what data sources, datasets, software, tools, and visualization means can be used within a specific analytics project [2]. Given this unique nature of constraints, it is important to respect their impact on requirements engineering for data analytics projects. Therefore, in this article, the requirements are considered in relation to different analytics project attributes that were amalgamated based on research articles (e.g., the data sources used in the analytics projects and the algorithms used in the project for data analysis). This relation to the analytics project attributes ensures that the requirements template aids users when defining requirements by considering as many aspects of the analytics project as possible in the initiation stage of the analytics project. Building the requirements template based on the analytics project attributes also enables the standardization of key information relating to analytics projects, allowing for easier knowledge reuse.

This article extends a previously published work by the authors [6], where only two types of analytics were considered. The research method is discussed and demonstrated in Section 2. The proposed templates for each type of analytics project separately and as a common template are presented in Section 3. Section 4 discusses the validity of the template through the analysis of real estate analytics project attributes of 20 real estate data analytics projects. Section 5 briefly concludes the article.

2 Research Method

To identify generic requirements for analytics projects, we employed a bottom-up approach, consisting of two main steps:

1. Analysis of Analytics Project Cases: Initially, we scrutinized research articles documenting various analytics projects to delineate project-specific requirements.

2. Derivation of Generic Requirements: Based on the project requirements identified in the first step, we extrapolated generic requirements to encompass commonalities across all projects.

Each type of analytics (descriptive, diagnostic, predictive, and prescriptive) was represented by the analysis of four projects, with a fifth project serving to validate the findings.

The analysis of data analytics cases involved a meticulous review of research articles detailing these projects. We identified sentences or phrases within these articles that defined specific processes or products within the analytics projects. These product or process definitions served as the basis for inferring potential project requirements. Each such project requirement was denoted as "Pro requirement" followed by a specific numbering convention. For instance, Pro requirement 4.2 indicates the second project requirement derived from case number four. We ensured traceability by including text from the relevant research articles within these project requirements.

Subsequently, we derived generic requirements from the defined project requirements. These generic requirements served as overarching guidelines applicable to all analytics projects. We validated these generic requirements by ensuring that project requirements satisfying the defined generic requirements could be identified in subsequent analytics project cases of the same type. Additionally, a final fifth project was used to validate all previously defined generic requirements.

To maintain traceability throughout the methodology, each generic requirement was assigned a specific number. A generic requirement retained the same number as the first project from which the requirement was derived.

The generic requirements were validated based on analytics project attributes (Table 1) that represent various aspects of analytics projects, such as data sets used, data sources, etc., ensuring that generic requirements define necessary constraints relating to projects. The attributes were acquired from a total of 21 research articles that stated (as is the case with [1]) or allowed to derive the attributes (as is the case with [7]). The generic requirements defining the values for analytics project attributes also aided in standardizing the analytics project information format allowing for easier reuse of knowledge which was discovered in other research works that utilized the analytics project attributes.

Table 1. Analytics Project Attributes

Analytics Project Attribute	Potential (example) Values of Attributes	Source	Comments on the Attribute
Project Context/ Field	Business	[7]	This attribute refers to the problem domain within which the data analytics project is being carried out.
	Biomedical		
Object/ System/ Measure of Interest	Health care	[8]	This attribute defines what object(s) or system(s) the analytics project is concerned with.
	Geothermal Energy	[9]	
Analytics Type	Descriptive	[1]	Defining the type of analytics that is being carried out within the project is very important, especially for the analysis phase of the project, since the goal of the project relates to the type of analytics.
	Diagnostic		
	Prescriptive		
	Predictive		
Project Documenta-tion	Data Dictionary	[10]	The documentation that is provided can change the amount of detail an analyst provides on things that relate to company data, available resources, and expected results.
	Data Flow Diagram	[11]	
Data Types Used	Numeric Data	[12]	The types of data that are used with any analytics project can be placed into one of these categories, regardless of the format. The type of data used within a project can affect factors such as the types of analytics methods that will be used as well as the possible visualization methods.
	Textual Data	[13]	

Table 1. Continued

Analytics Project Attribute	Potential (example) Values of Attributes	Source	Comments on the Attribute
Datasets Used	Training Dataset	[14]	Whether or not the datasets used in the analytics project exist already or need to be created by collecting data from different sources changes the work required within the data acquisition phase. The dataset used can have various features that impact how the analytics will be carried out. In the context of machine learning models, the train and test data sets need to be defined.
	Testing Dataset		
Data Source(s)	Corporate Websites	[15]	The source from which data is collected affects the data extraction methods that are required, as well as how external factors, such as data privacy laws, affect the analytics project.
	Social Networking Sites		
	Search Engines		
Data Extraction Technique	Web Wrappers	[16]	The technique used to extract data is dependent on the form of data that is extracted and who or what data is acquired from. This makes the data extraction process a large part of the data analytics project.
Data Integration	Record Linkage	[17]	If there are multiple sources of data, then data integration might be necessary to combine them into a single dataset.
Analysis Technique/ Method/ Methodo-logy	Correlation Analysis	[7]	The technique of analysis being performed impacts most aspects of the analytics project and vice versa. Therefore, the selection of the analysis techniques used in the project and the definition of the chosen technique are essential for the success of the project.
	Regression Analysis		
	Text Analysis	[13]	
Analysis Tools/ Software	Pandas (library) in Python 3 (Programming Language)	[18]	The tools that are used to carry out the analysis change the nature of the analytics project similar to the selected analysis technique. A good demonstration of this is how the software will affect "Data types used" because different tools have different inbuilt data types, as well as rules regarding how those data types can be manipulated.
	Hadoop	[19]	
Model(s) Used for Analysis	WaterGap Model	[20]	This attribute relates to preexisting mathematical or statistical models that will be used to carry out the analytics. The models represent real-world entities/events and can be used to analyze these entities/events.
Algorithm(s) Used for Analysis	SVM	[21]	Many algorithms can be utilized in analytics such as K-means and decision trees. These algorithms can be used to analyze or to create analytics models.
	Random Forests		
Report Specification	Results Formatting	[22]	Most if not all analytics projects produce a report that is given to the clients as a project artifact. Therefore, formatting this report most acceptably is vital to client satisfaction.
Graphical Results	Box Plot	[23]	The visualizations used within a data analytics project are not the most vital characteristic, but they determine how well the findings are presented, therefore it is the culmination of the project. The visualization is dependent on the characteristics within the acquisition and analysis phase.
	Bar Chart		
Interactive Results	Tableau Dashboard	[24]	Various interactive software tools, such as business intelligence dashboards, can be applied to produce a result of the analytics, one of the tools that are used for this is Tableau.
Created Model(s)	Predictive Analytics Model for the Power Consumption in Manufacturing	[25]	Some analytics projects result in the creation of models that have been trained using data. The model(s) that have to be created as a result of the analytics project must be defined.
Method Used for Verification/ Validation of Results	k-Fold Cross Validation	[26]	Once the result of the analytics project is created there can be a need to test it. The method used to verify the results must be defined.

3 The Analysis of Analytics Cases

This section demonstrates the analysis of analytics project cases and introduces generic requirements for descriptive, diagnostic, predictive, and prescriptive analytics projects. For each analytics type, four projects (referred to as cases) were analyzed to gather the generic requirements and one case was used to test their applicability. The analysis was carried out by analyzing published research articles that relate to the analytics project cases (associated with search terms 'Diagnostic Analytics,' 'Descriptive Analytics,' 'Predictive Analytics,' and 'Prescriptive Analytics') that were present in bibliographic databases IEEE Xplore and ScienceDirect. The process of analysis is illustrated for one project of descriptive analytics (in Subsection 3.1). The description of the analysis of the other projects (cases), which was done in the same way, is available on GitHub [27]. In the analysis, all requirements are organized according to the analytics project phases so that all phases can be respected from the very beginning of the project. Subsections 3.1–3.4 present the sets of generic requirements that can be used as requirements acquisition templates for the respective types of analytics.

3.1 Descriptive Analytics Project Cases

The form of analysis is demonstrated here in one of the descriptive analytics projects (Case 1) *"Students' perceptions of a community health advocacy skills building activity: A descriptive analysis"* [28]. The analysis produced the following results [6].

Initiation phase:

Pro requirement 1.1: This analytics project must "explore students' perceptions of the benefits of a discussion activity about a controversial health issue, and to describe the impact of the opportunities to form valid arguments using empirical evidence on students' perceptions of their ability to be advocates".

Gen requirement 1.1: An analytics project must have a clearly defined goal.

Pro requirement 1.2: The methods used in this project will consist of "students were invited to provide feedback on their perceptions of activity benefits. Descriptive analyses were conducted".

Gen requirement 1.2: The analytics project must have a clearly defined strategy that will be used to achieve the mentioned goal. The main point of emphasis is what type of data analytics will be required to achieve the goal.

Acquisition phase:

Pro requirement 1.3: This project will use "post assignment survey" and "included questions asking how much the activity helped the student learn the following advocacy skills: (1) form a valid argument using scientific evidence; (2) use credible sources when forming opinions; and (3) begin to see themselves as advocates for improving the health of individuals and communities."

Gen requirement 1.3: The data analytics project must have a source(s) of data and how it will be collected.

Analysis phase:

Pro requirement 1.4: The project will carry out descriptive analysis by using "Descriptive statistics".

Gen requirement 1.4: The specific method(s) that will be used to carry out the analysis must be selected.

Pro requirement 1.5: The project will use IBM's "SPSS" software to conduct descriptive statistics.

Gen requirement 1.5: The software or tools that will be used to carry out the analysis must be explicitly mentioned.

Presentation phase:

Pro requirement 1.6: The insights provided by the data analytics project will be presented in the form of a bar chart showing the "frequency distribution" for each of the responses by each category of students (graduate or undergraduate).

Gen requirement 1.6: If and how the findings of the data analysis must be visualized should be defined.

Four other cases analyzed are listed below:

- Alopecia areata: Descriptive analysis in a Brazilian sample [29] (Descriptive Analytics Case 2).
- Restaurant closures during the COVID-19 pandemic: A descriptive analysis [30] (Descriptive Analytics Case 3).
- Descriptive analytics using visualization for local government income in Indonesia [31] (Descriptive Analytics Case 4).
- Factors contributing to coronavirus disease 2019 vaccine hesitancy among healthcare workers in Iran: A descriptive-analytical study [32]. (Descriptive Analytics Case 5; this case was used for testing the applicability of the generic requirements).

The results of the analysis of all descriptive analytics cases are available in the file "Descriptive Analytics Case Study Analyses.pdf" on GitHub [27].

Through the analysis of Cases 1–4, the following generic requirements for descriptive analytics projects were defined, with Gen(des) denoting that the generic requirements relate to descriptive analytics projects:

Initiation phase

1. Gen(des) requirement 1.1: An analytics project must have a clearly defined goal.
2. Gen(des) requirement 1.2: The analytics project must have a clearly defined strategy that will be used to achieve the defined goal.
3. Gen(des) requirement 3.1: The level of detail and technicality required when describing the analytics methods used within the project must be defined based on the knowledge level of the client(s).

Acquisition phase

4. Gen(des) requirement 1.3: The analytics project must have a source(s) of data and an approach to how it will be collected.
5. Gen(des) requirement 2.1: The permissions regarding the use of incomplete data sets must be defined within the context of the analytics project.
6. Gen(des) requirement 3.2: How data was collected from the data source must be defined for an analytics project.
7. Gen(des) requirement 4.1: The data contained within the data source must be defined as well as which of that data will be used for the data analytics.
8. Gen(des) requirement 4.2: Specification regarding the ETL (Extract Transform Load) must be defined for analytics projects.
9. Gen(des) requirement 4.3: How data warehousing is carried out in the analytics project, and then the specifications regarding the data warehouse must be specified.

Analysis phase

10. Gen(des) requirement 1.4: The specific method(s) that will be used to carry out the analysis should be selected.
11. Gen(des) requirement 1.5: The software or tools that will be used to carry out the analysis must be explicitly defined.
12. Gen (des) requirement 3.3: To be used practices that relate to the reliability of analysis results must be defined for the analytics project.

Presentation phase

13. Gen(des) requirement 1.6: How the findings of the analysis must be visualized should be defined.
14. Gen(des) requirement 4.4: How the findings of the analysis can be used must be defined.

These generic requirements can form a requirements template that can be used to define values for attributes of data analytics projects. The relationships between the analytics project attributes

and the generic requirements for descriptive analytics are represented using a requirement-attribute relationship graph (Figure 1). The colored arrows in this model represent which attributes can be defined by each descriptive analytics' generic requirement, based on whether the project (Pro) requirement defined to satisfy a generic (Gen) requirement will be able to indicate the value for a specific attribute. For instance, in the analytics case 1 discussed at the beginning of this subsection, in the case of Pro requirement 1.1, it can be concluded that the project context is 'health care education' and the Object of interest is students' perceptions of the benefits of a discussion activity about a controversial health issue.

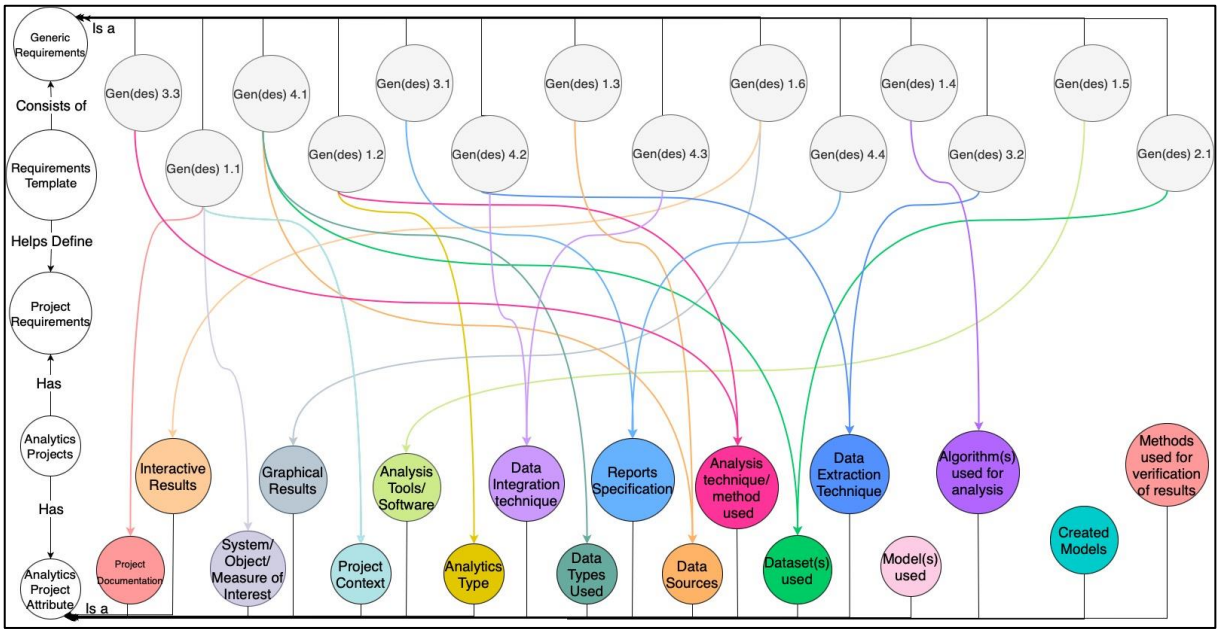

Figure 1. Relationships between generic requirements for descriptive analytics and the analytics project attributes

3.2 Diagnostics Analytics Project Cases

Similarly, as for descriptive analytics, five diagnostic analytics projects were analyzed:

- Diagnostic Analysis for outlier detection in big data analysis [33] (Diagnostic Analytics Case 1).
- Diagnostic analysis of regional ozone pollution in Yangtze River Delta, China: A case study in Summer 2020 [34] (Diagnostic Analytics Case 2).
- Mixed logit model-based diagnostic analysis of bicycle-vehicle crashed at daytime and nighttime [35] (Diagnostic Analytics Case 3).
- Diagnostic analysis of distributed input and parameter datasets in Mediterranean basin streamflow modeling [36] (Diagnostic Analytics Case 4).
- Diagnostic analysis of a single-cell proton exchange membrane unitized regenerative fuel cell using numerical simulation [37] (Diagnostic Analytics Case 5; this case was used for testing the applicability of the derived generic requirements).

The results of the case analysis are available in the file "Diagnostic Analytics Case Study Analyses.pdf" on GitHub [27]. Through the analysis of the diagnostic analytics cases (projects), the following generic requirements for diagnostic analytics projects Gen(dia) were defined:

Initiation phase:

1. Gen(dia) requirement 1.1: The analytics project must have a clearly defined goal.
2. Gen(dia) requirement 1.2: The analyst must be aware of the level of expertise of the client and define the key terminology within the context of the data analytics project accordingly.

3. Gen(dia) requirement 1.3: The analytics project must have a clearly defined system/object on which the analysis is carried out.
4. Gen(dia) requirement 1.4: The analytics project must have a quantitative metric(s) that is used to evaluate the system.
5. Gen(dia) requirement 4.1: The analytics project must define how the results of the data analysis will be validated or verified.

Acquisition phase:

6. Gen(dia) requirement 1.5: The analytics project must have defined which data sets will be used, and where and how these data sets will be acquired.
7. Gen(dia) requirement 1.6: The properties of the dataset that is used within the analytics project must be defined.
8. Gen(dia) requirement 2.1: The model(s) used within the analytics project, along with what said models are used for, must be defined for a diagnostic analytics project.
9. Gen(dia) requirement 4.2: The derived data used in the analytics project must be defined.

Analysis phase:

10. Gen(dia) requirement 1.7: The method that will be employed to carry out the analysis must be defined.
11. Gen(dia) requirement 1.8: An analytics project must have an in-depth definition of the method that will be used to carry out the analytics. This includes specific equations that will be used and the variables within said equations.
12. Gen(dia) requirement 2.2: The analytics project must have clearly defined tools that are going to be used and what those tools will be used for.

Presentation phase:

13. Gen(dia) requirement 1.9: The graphical representations that are required when presenting the analytic results must be defined.
14. Gen(dia) requirement 1.10: The format in which the results of an analytics project are textually presented must be defined.
15. Gen(dia) requirement 3.1: The format by which the different causes of the issue must be categorized, must clearly be stated when presenting the results of the analytics project.

Figure 2 shows the relationship between the generic requirements for diagnostic analytics projects and the attributes of the analytics project.

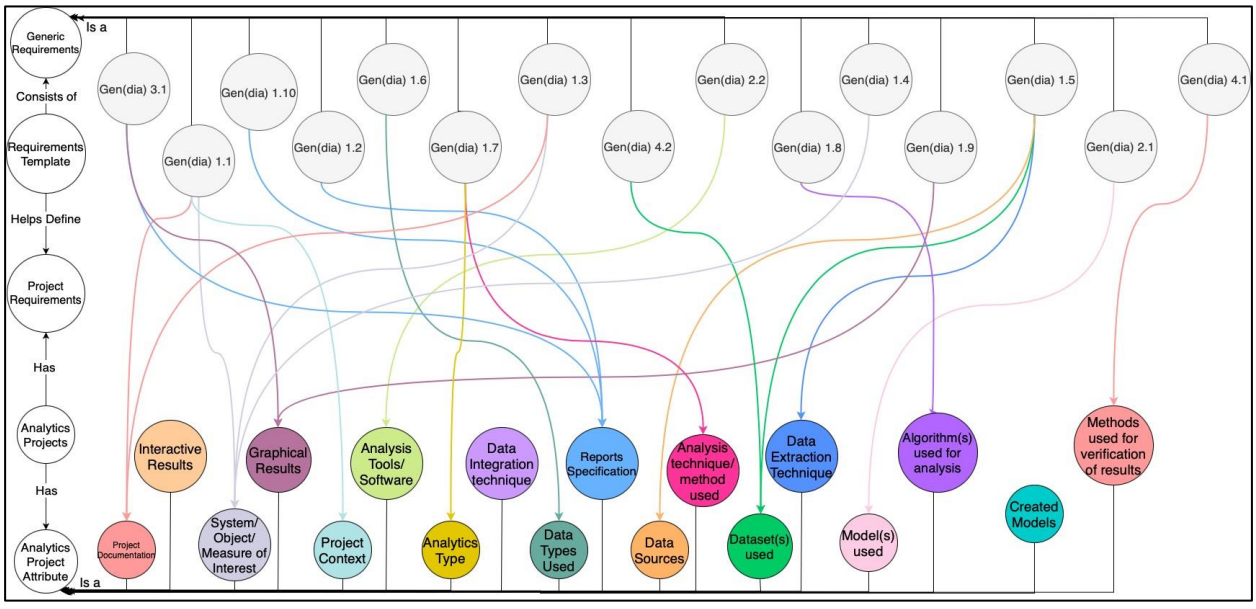

Figure 2. Relationships between generic requirements for diagnostic analytics and the analytics project attributes

As in Figure 1, the colored arrows represent which attributes can be defined by each diagnostic analytics generic requirement. The model in Figure 2 shows that attributes relating to the 'Creation of models' and 'Interactive results' are not defined by the generic requirements. This indicates that diagnostic analytics projects rarely result in the creation of models or interactive results. Additionally, there is an emphasis placed on the selection of a system(s)/object(s) as well as the appropriate measures to represent the object(s)/system(s) that the diagnostic analytics is being carried out on, which is shown by the three generic requirements that are used to define it. This means that in diagnostic analytics, more specific than in descriptive analytics, information is required to properly define the attribute relating to the selection of an object or a system and to define the appropriate measure(s) used to evaluate the system.

3.3 Predictive Analytics Project Cases

Also, for predictive analytics, five predictive analytics projects were analyzed:

- Predictive analytics using big data for increased customer loyalty: Syriatel Telecom company case study [38] (Predictive Analytics Case 1).
- Operating data-driven predictive analytics for tele-diagnosis of refrigeration systems: A Case Study [39] (Predictive Analytics Case 2).
- A Case Study of Sri Lanka oil price fluctuations and its influencing factors using predictive analytics [40] (Predictive Analytics Case 3).
- Predictive analytics of donors in crowdfunding platforms: A case study on Donorschoose.org [41] (Predictive Analytics Case 4).
- Automation in financial reporting by using predictive analytics in SAP Analytics Cloud for gold mining industry: a Case study [42] (Predictive Analytics Case 5; this case was used for testing the applicability of the derived generic requirements).

The results of the case analysis are available in the file "Predictive Analytics Case Study Analyses.pdf" on GitHub [27]. Through the analysis of the predictive analytics cases (projects), the following generic requirements for predictive analytics projects Gen(prd) were defined:

Initiation phase:

1. Gen(prd) requirement 1.1: The goal of the analytics project must be defined.
2. Gen(prd) requirement 1.2: The object to which the analytics is directed must be defined.

Acquisition phase:

3. Gen(prd) requirement 1.3: The sources of data that will be utilized in the analytics project must be defined.
4. Gen(prd) requirement 1.4: The features that will be used to train the model must be defined.
5. Gen(prd) requirement 1.5: The feature selection methods utilized in the analytics project must be defined.
6. Gen(prd) requirement 1.6: The tools and software used for acquiring and storing the data must be defined along with their utility within the project.
7. Gen(prd) requirement 4.1: The analytics project must define all preprocessing done to the data before being it inputted into the predictive model.

Analysis phase:

8. Gen(prd) requirement 1.7: The model(s)/algorithm(s) that must be used within the analytics project must be defined.
9. Gen(prd) requirement 1.8: The tools and software to be used to develop the model must be defined along with their utility within the project.
10. Gen(prd) requirement 1.9: The performance measures used to evaluate any created models must be defined alongside the acceptable ranges for the values of these measures.

Presentation phase:

11. Gen(prd) requirement 1.10: The output(s) of the model must be defined alongside the excepted utility of the models' outputs.
12. Gen(prd) requirement 2.1: The analytics project must specify the medium (mobile application, analytics dashboard, or report) by which the created analytics model is used.

Figure 3 shows the relationship between the generic requirements for predictive analytics projects and the attributes of the analytics projects. The colored arrows in this model represent which attributes can be defined by each predictive analytics generic requirement. All the attributes can be defined using the generic requirements especially those relating to usage and creation of models. Additionally, the generic requirement relating to the definition of the goal of the analytics project can be applied to define the type of analytics to be used (the case study analysis revealed that the definition of the project goal always uses to word 'predict').

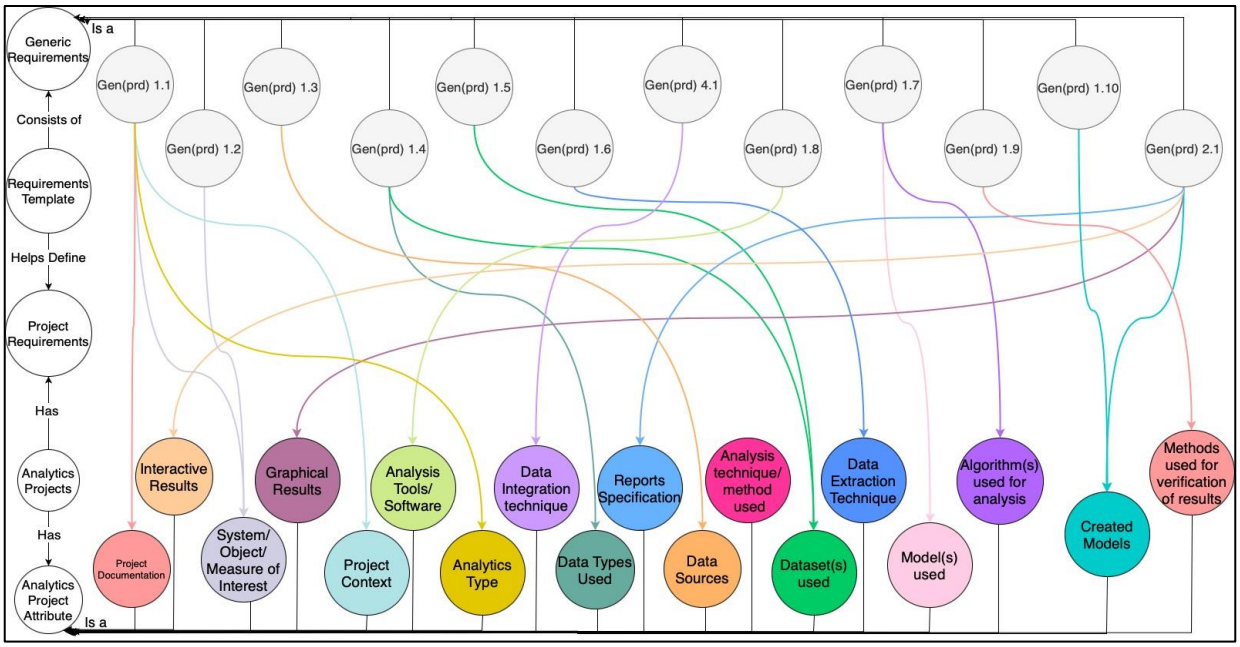

Figure 3. Relationships between generic requirements for predictive analytics and the analytics project attributes

3.4 Prescriptive Analytics Project Cases

Similarly to other types of analytics, five prescriptive analytics projects were analyzed:

- Case studies of clinical decision-making through prescriptive models based on machine learning [43] (Prescriptive Analytics Case 1).
- Study on facility location of air taxi skyports using a prescriptive analytics approach [44] (Prescriptive Analytics Case 2).
- Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework [45] (Prescriptive Analytics Case 3).
- Prescriptive analytics in public-sector decision-making: A framework and insights from charging infrastructure planning [46] (Prescriptive Analytics Case 4).
- Design and evaluation of a process-aware recommender system based on prescriptive analytics [47] (Prescriptive Analytics Case 5; this case was used for testing the applicability of the derived generic requirements).

The results of the case analysis can be found in the file "Prescriptive Analytics Case Study Analyses.pdf" on GitHub [27]. Through the analysis of the prescriptive analytics case (projects), the following generic requirements for prescriptive analytics projects Gen(prs) denoting generic requirements for prescriptive analytics projects were defined (it should be noted that prescriptive

analytics utilizes predictive models, and therefore some generic (Gen) requirements were defined relating to predictive models):

Initiation phase:

1. Gen(prs) requirement 1.1: The goal of the analytics project must be defined.
2. Gen(prs) requirement 1.2: The object to which the analytics is directed must be defined.

Acquisition phase:

3. Gen(prs) requirement 1.3: The sources of data that will be utilized in the analytics project must be defined.
4. Gen(prs) requirement 1.4: The features that will be used to train the model must be defined.

Analysis phase:

5. Gen(prs) requirement 1.5: The predictive model(s)/algorithm(s) that must be used within the analytics project must be defined.
6. Gen(prs) requirement 1.6: The expected output of the predictive analytics model must be defined for the analytics project.
7. Gen(prs) requirement 1.7: The performance measures for the created predictive model/algorithm must be defined in the project.
8. Gen(prs) requirement 1.8: The method/model(s) used to create the prescriptive analytics model or algorithm must be defined.
9. Gen(prs) requirement 1.9: The performance measures for the created prescriptive model/algorithm must be defined for the analytics project.
10. Gen(prs) requirement 1.10: The method used to define hyperparameters for the predictive analytics model(s)/algorithm(s) must be defined.
11. Gen(prs) requirement 2.1: The tools/software used within the project must be defined.

Presentation phase:

12. Gen(prs) requirement 1.11: The output of the model must be defined alongside the excepted utility of the models' outputs.

Figure 4 shows the relationships between the generic requirements for prescriptive analytics projects and the attributes of the analytics project.

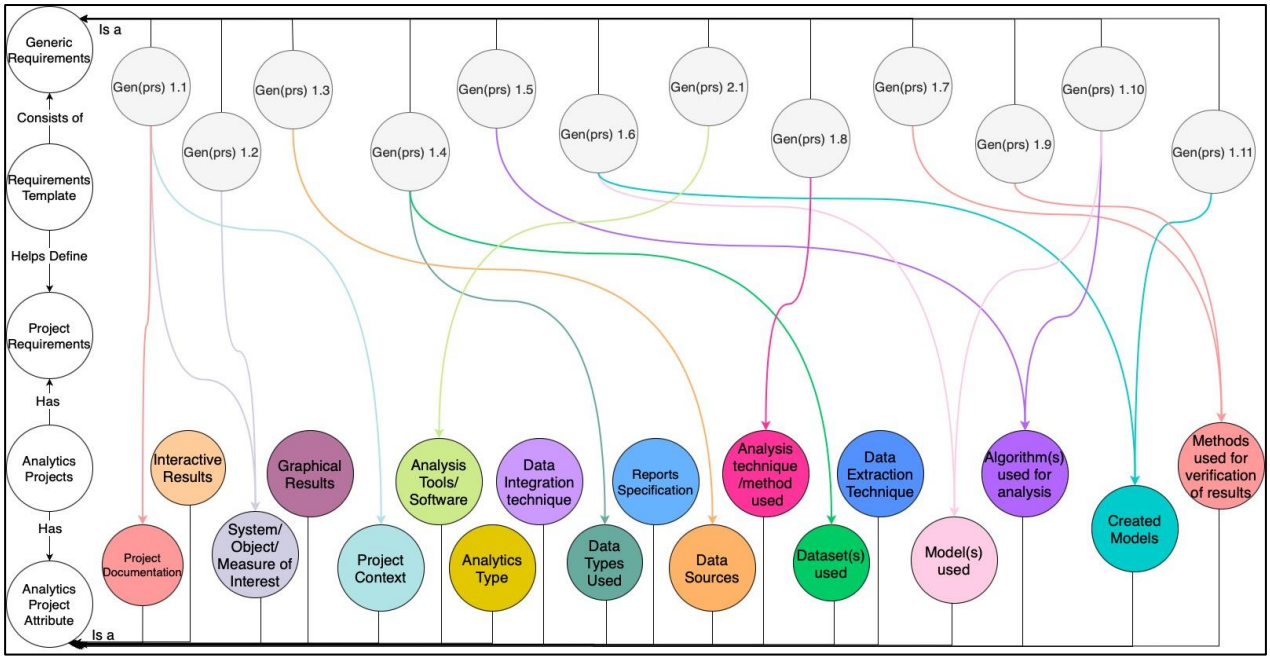

Figure 4. Relationships between generic requirements for prescriptive analytics and the analytics project attributes

The colored arrows in Figure 4 depict which attributes can be defined by each generic requirement. The graph shows that attributes relating to graphical results, interactive results, analytics type, report specification, and data extraction are not defined by the generic requirements. This indicates that prescriptive analytics projects might not have complex data extraction processes and they might not focus on the visual or textual results they create. There is also a higher emphasis on the creation, use, and validation of models evidenced by a higher number (relative to predictive analytics) of the generic requirements defining the attributes 'Models used' and 'Methods used for verification of results'. This is likely due to prescriptive analytics projects utilizing/creating more than one model within their projects, where at least one model is used to predict future outcomes and at least one other model to prescribe the best course of action based on the predictions.

3.5 Definition of Requirements Template Based on Analytics Project Attributes

The generic requirements presented in Section 3 can be put into two groups. One group consists of requirements common to all analytics types; another group consists of requirements that refer to only one or some analytics types but not all. To better understand the impact of requirements on analytics project attributes by analytics type, Table 2 was created based on the analytics requirements-attribute graphs (Figures 1–4). Table 2 shows which generic requirements in the requirements templates defined which analytics project attributes (Table 1). An example of how Table 2 was created is that Gen(des) 1.1, Gen(dia) 1.1, Gen(prd) 1.1, and Gen(prs) 1.1, which define such attributes as the project context/field and the project documentation, all are represented in the same rows of the table (Project Context Field and Project Documentation).

Table 2. Relationship between generic requirements and analytics project attributes

	Generic requirements that define the attribute			
Analytics Project Attribute	Descriptive (Gen(des))	Diagnostic (Gen(dia))	Predictive (Gen(prd))	Prescriptive (Gen(prs))
Project Context/Field	1.1	1.1	1.1	1.1
Object/System/Measure of interest	1.1	1.1, 1.3, 1.4	1.1, 1.2	1.1, 1.2
Analytics type	1.2	1.7	1.1	N/A
Project Documentation	1.1	1.1, 1.3	1.1	1.1
Data types used	4.1	1.6	1.4	1.4
Dataset(s) used	4.1	1.5, 4.2	1.4, 1.5	1.4
Data source(s)	2.1	1.5	1.3	1.3
Data extraction technique	3.2, 4.2	1.5	1.6	N/A
Data Integration	4.2, 4.3	N/A	4.1	N/A
Analysis technique/method used	1.2, 3.3	1.7	N/A	1.8
Model(s) used for analysis	N/A	2.1	1.7	1.6, 1.10
Algorithm(s) used for analysis	1.4	1.8	1.7	1.5, 1.10
Analysis tools/software	1.5	2.2	1.8	2.1
Report Specification	3.1, 4.4	1.2, 1.10, 3.1	2.1	N/A
Graphical results	1.6	1.9, 3.1	2.1	N/A
Interactive results	1.6	N/A	2.1	N/A
Created Model(s)	N/A	N/A	2.1, 1.10	1.6, 1.11
Method(s) used for verification of results	N/A	4.1	1.9	1.7, 1.9

Using Table 2 it is possible to create a common requirement template by organizing generic requirements from individual requirements templates presented in Section 3 for the different types of analytics based on the attributes they define. The common template shows in one group the

requirements that are the same for all types of analytics and in separate groups the requirements that are specific to the types of analytics.

Thus, based on coverage of the attributes, it was possible to organize the requirements in a common generic requirements template where, for each project phase, the requirements that are present for all types of analytics are put separately from the analytics type-specific requirements. Table 3 shows a small excerpt of the common requirements template for the initiation phase of the project and two types of analytics (descriptive and diagnostic). The full common requirements template is available in the appendix.

Table 3. A fragment of a template for generic requirements for analytics projects

Requirement Types	**Analytics Project Phase**
	Initiation
Common (for all types of analytics)	The goal of the analytics project must be defined, specifying the field/context of the project.
	The system(s)/object(s) on which the analysis is being carried out must be defined.
	If the results of the analytics project must be verified, specify the strategy that must be used to verify the results of the analysis.
Descriptive Analytics Specific	Gen (Des) 3.1: The level of detail and technicality required when describing the analytics methods used within the project must be defined based on the knowledge level of the client(s).
Diagnostic Analytics Specific	Gen (Dia) 1.4: The quantitative metric(s) that must be used to evaluate the object(s)/system(s) of interest must be defined.
	Gen (Dia) 1.2: The analyst must be aware of the level of expertise of the client and define the key terminology within the context of the data analytics project accordingly.

The developed common generic requirements template is intended to be used as the basis for developing a tool supporting requirements definition for analytics projects.

4 Validation Based on Real Estate Analytics

In the authors' previous work [6] the initial version of the generic requirements template (consisting of descriptive and diagnostic analytics) was applied in the pre-investigation phase of a multi-contextual real estate analytics project. With the help of the template, 13 new requirements were defined through an interview and a questionnaire that required a total investment of 70 minutes by one of the stakeholders. Although the stakeholder commented that the defined requirements needed to be more specific, the generic requirements template aided as a starting point for the requirements definition process.

Further, the proposed analytics project attributes and the generic requirements were tested using the related work consisting of reports on 20 real estate analytics research project cases. The attributes for these real estate analytics cases were defined by applying the requirements template to the research articles that described these cases. The defined attributes for each real estate analytics case are available in the file "Real_Estate_Analytics_Attributes.xlsx" on GitHub [27].

To achieve better visibility, the real estate analytics attribute models were created for some groups of attributes. These models represent the values for different analytics project attributes relating to real estate analytics cases (projects). One model demonstrates the values of multiple attributes. Creating a model for each attribute would result in the creation of 18 attribute models.

In these models, the real estate analytics projects are represented in groups based on the type of analytics being conducted, given that the generic requirements templates are organized based on the types of analytics being conducted. The analysis of these 20 real estate analytics projects showed that these projects can be related to the requirements template and the attributes represented in Table 1. The created models showed additional utility as they gave a useful overview of the analytics projects in the domain of interest. For instance, the model in Figure 5

that represents the system/object/measure of interest within the analytics project can help identify which research projects/articles are related to the actual analytics project for which the requirements are defined. Thus, for the multi-contextual real estate analytics project which concerned the return on investment and the context of energy in the real estate domain, for instance, the papers 'Return on Investment (ROI) of property rentals' [48] and 'Geographic information in relation to household energy' [49] were identified. On the other hand, Figure 6 can be utilized to see what data sources were utilized for real estate analytics projects. Figure 7 shows what analysis methods/techniques and algorithms can be used in real estate analytics. Figure 8 can be used to see what models and interactive and graphical results can be created in analytics projects. The full-size models are available in the file 'Real Estate Analytics Projects Models' on GitHub [27].

The models like the ones that were used to validate the template can be developed, maintained, and used in parallel with the use of a generic requirements template to explore the related works in the domain of interest for achieving high-quality requirements.

Additionally, the real estate analytics attributes were used to create semi-automated knowledge graphs that allow for the reuse of knowledge gained in research relating to data analytics [50], [51].

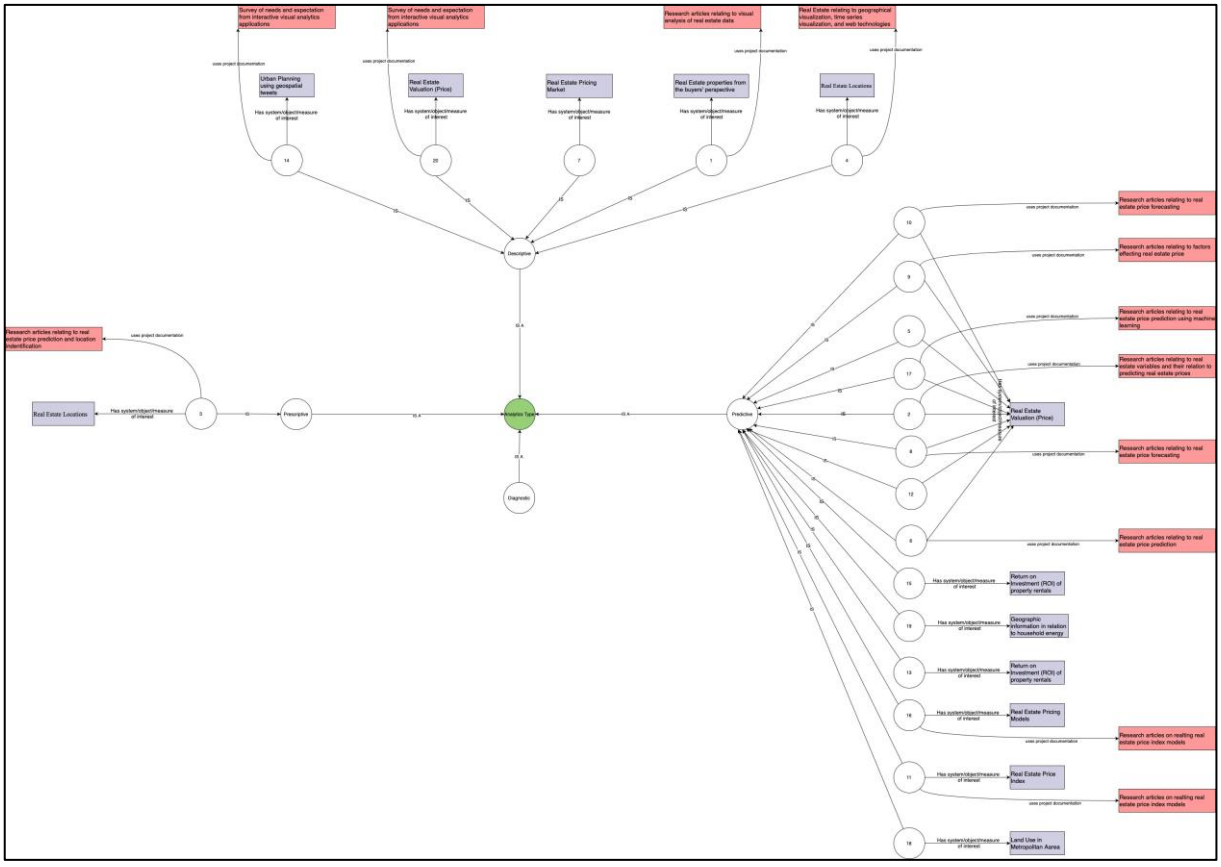

Figure 5. Real estate analytics attributes model for Object/System/Measure of Interest (purple nodes) and Project documentation (pink nodes)

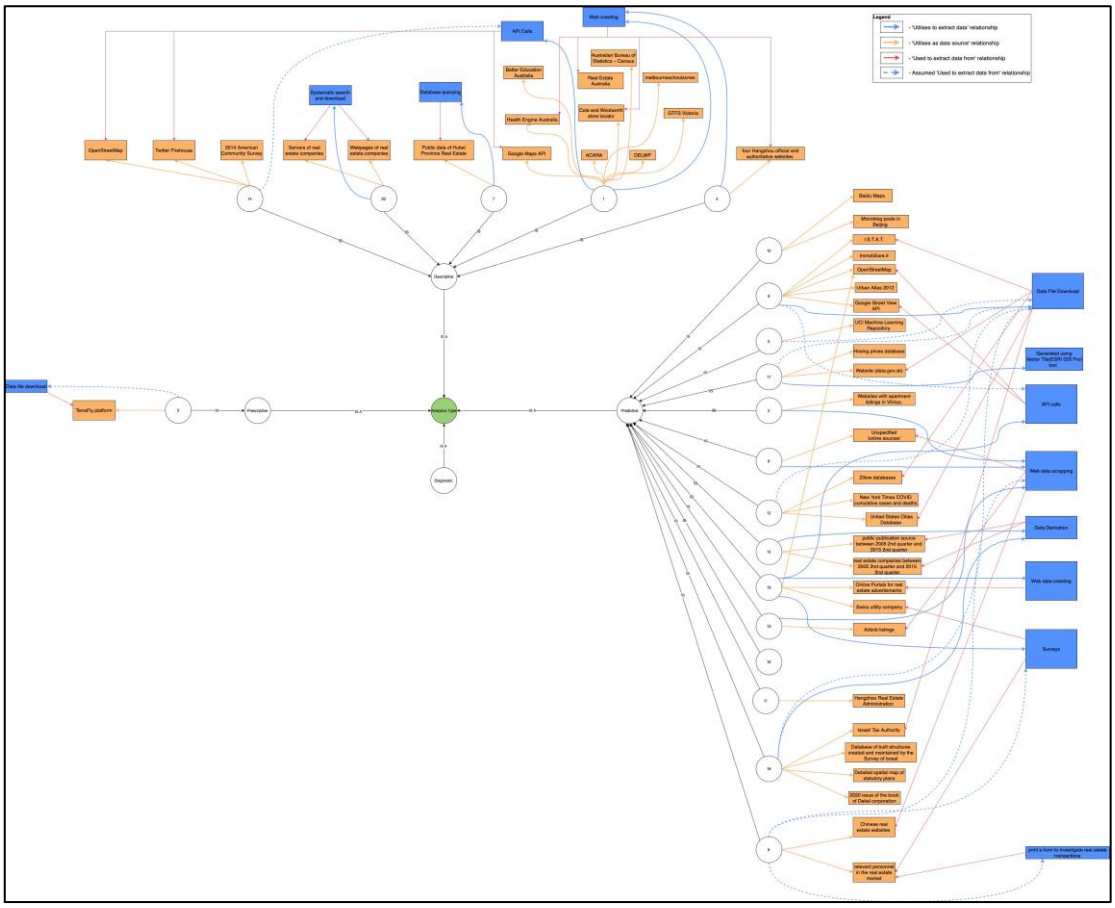

Figure 6. Real estate analytics attributes model for data sources (yellow nodes) and extraction techniques (blue nodes) used

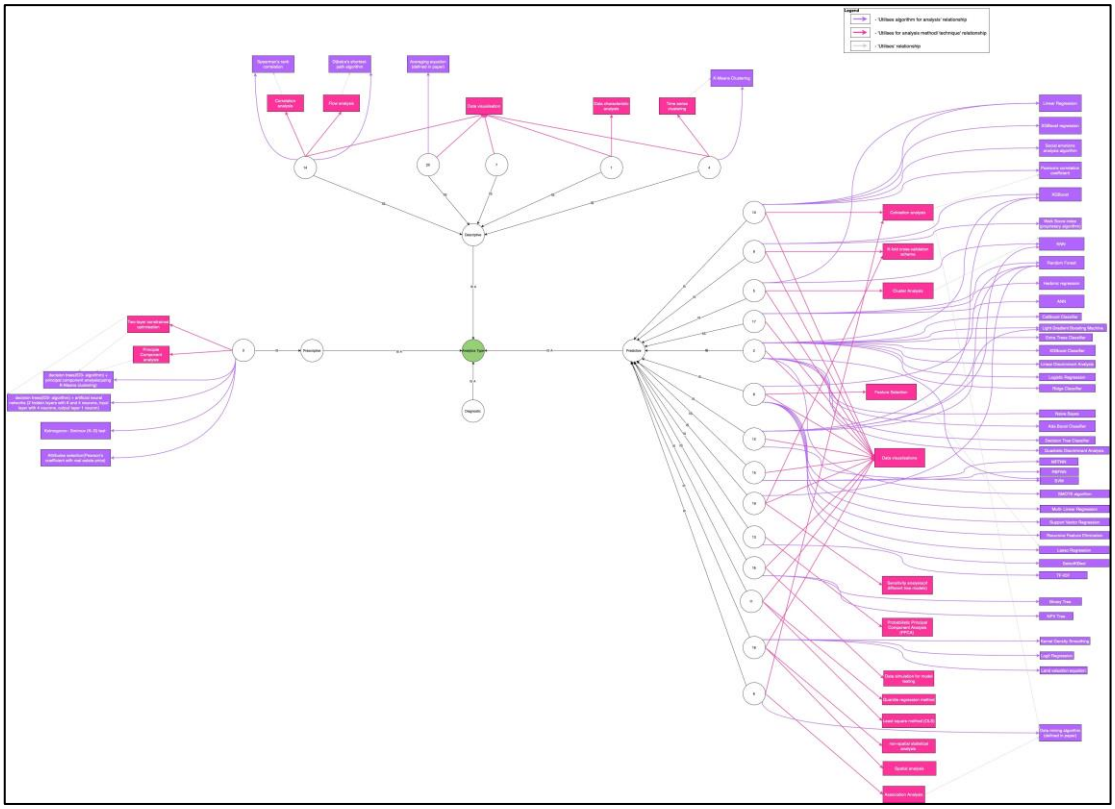

Figure 7. Real estate analytics attributes model for analytics techniques/methods (pink nodes) and algorithms (purple nodes) used

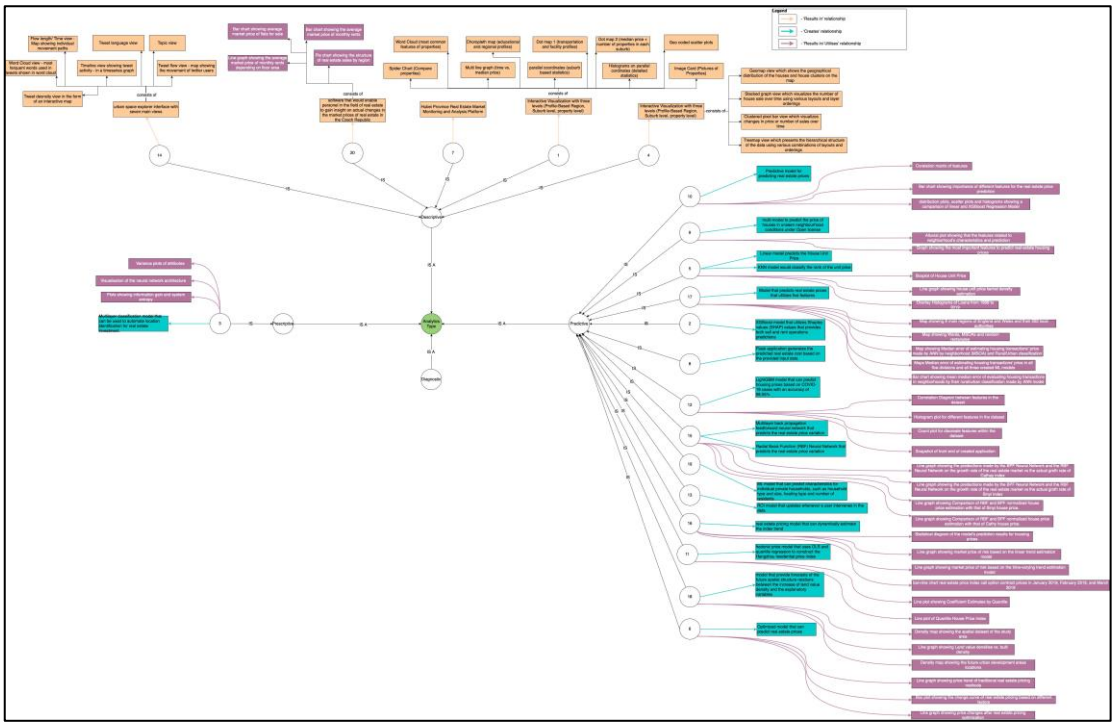

Figure 8. Real estate analytics attributes model for the graphical (purple nodes), interactive(yellow), and models (green) results created by the analytics projects

5 Conclusion

This work seeks to extend previous research [6] that was done by the authors with the goal being the analysis of analytics project cases to define a generic requirements template for analytics projects. The previous work only considered descriptive and diagnostic analytics, while this work considered two other types of analytics (namely, predictive and prescriptive) to create a common generic requirements template that would apply to the majority of analytics projects. This common generic requirements template was built in relation to analytics project attributes which represent the different aspects of the analytics projects (e.g., data sources, data types, analytics algorithms, models, and visualization methods). The intended use of the common generic requirements template is to help in the definition of requirements within the initial(planning) stage of an analytics project while considering all aspects of an analytics project (represented by the analytics project attributes) by providing a checklist for the requirements that need to be defined.

The created generic requirements template was then used to identify analytics project attribute values for 20 real estate analytics cases. The fact that the values for attributes could be defined for all projects demonstrates the requirements templates utility within analytics projects to define requirements and constraints relating to analytics project attributes which concern different aspects of analytics projects. Additionally, these attributes were then used to create attribute models that visually represent and help to analyze various aspects of the completed analytics projects such as object(s)/system(s)/measure(s) of interest, dataset and sources used, and results (models, interactive, graphical) created by the analytics project. These attribute models gave an insight into the state of the art of analytics in the particular domain, which is useful knowledge when defining new analytics projects.

The results achieved so far contribute to the opportunity to develop a new requirements engineering tool that combines analytics attributes and an accompanying common generic requirements template and helps define requirements and attributes related to new data analytics projects.

A “side effect” of this work is a created knowledge repository of real estate analytics projects in the form of attribute models that was possible thanks to the generic requirements template-based

way in which information relating to previously completed analytics projects was represented. Further research can leverage this effect to create new models for different domains of data analytics, such as, e.g., healthcare, that represent the preexisting works in the domain. There is also the possibility to generate more complex models in the form of related knowledge graphs representing different data analytics domains.

Acknowledgment

The choice of the real estate domain for testing the generic requirements template was inspired by the IT Competence Center project id. nr. 5.1.1.2.i.0/1/22/A/CFLA/008.

References

[1] F. Balali, J. Nouri, A. Nasiri, and T. Zhao, "Data Analytics," in *Data Intensive Industrial Asset Management*, Springer, 2020, pp. 105–113. Available: https://doi.org/10.1007/978-3-030-35930-0_7

[2] T. Fawcett and F. Provost, *Data Science for Business*. O'Reilly, 2013.

[3] V. N. Gudivada, M. T. Irfan, E. Fathi, and D. L. Rao, "Cognitive Analytics: Going Beyond Big Data Analytics and Machine Learning," *Handbook of Statistics*, vol. 35, pp. 169–205, 2016. Available: https://doi.org/10.1016/bs.host.2016.07.010

[4] A. Hussain, E. O. Mkpojiogu, and F. M. Kamal, "The Role of Requirements in the Success or Failure of Software Projects," *IRMM*, vol. 6, no. 7, pp. 306–311, 2016.

[5] "ISO/IEC/IEEE International Standard – Systems and software engineering – Life cycle processes – Requirements engineering," in *ISO/IEC/IEEE 29148:2018(E)*, pp. 1–104, 2018. Available: https://doi.org/10.1109/IEEESTD.2018.8559686

[6] A. Kumarasinghe and M. Kirikova, "Generic requirements template for data analytics," *Joint Proceedings of the BIR 2023 Workshops and Doctoral Consortium, 22nd International Conference on Preceptive in Business Informatics Research, CEUR Workshop Proceedings*, CEUR-WS.org, pp. 84–93, 2023. Available: https://ceur-ws.org/Vol-3514/paper72.pdf

[7] T. A. Runkler, *Data Analytics*. Springer, 2020. Available: https://doi.org/10.1007/978-3-658-29779-4

[8] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," *International Journal of Medical Informatics*, vol. 114, pp. 57–65, 2018. Available: https://doi.org/10.1016/j.ijmedinf.2018.03.013

[9] P. M. B. Abrasaldo, S. J. Zarrouk, and A. W. Kempa-Liehr, "A systematic review of data analytics applications in above-ground geothermal energy operations," *Renewable and Sustainable Energy Reviews*, vol. 189, p. 113998, 2024. Available: https://doi.org/10.1016/j.rser.2023.113998

[10] S. M. Rashid et al., "The Semantic Data Dictionary – An Approach for Describing and Annotating Data," *Data Intelligence*, vol. 2, no. 4, pp. 443–486, 2020, Available: https://doi.org/10.1162/dint_a_00058

[11] N. Olayan, V. Patu, Y. Matsuno, and S. Yamamoto, "A Dependability Assurance Method Based on Data Flow Diagram (DFD)," *2013 European Modelling Symposium*, 2013, pp. 113–118. Available: https://doi.org/10.1109/EMS.2013.20

[12] M. Gidea and Y. Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 820–834, 2018. Available: https://doi.org/10.1016/j.physa.2017.09.028

[13] V. Pichiyan, S. Muthulingam, S. G, S. Nalajala, A. Ch, and M. N. Das, "Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis," *Procedia Computer Science*, vol. 230, pp. 193–202, 2023. Available: https://doi.org/10.1016/j.procs.2023.12.074

[14] F. Huang, Z. Teng, Z. Guo, F. Catani, and J. Huang, "Uncertainties of landslide susceptibility prediction: Influences of different spatial resolutions, machine learning models and proportions of training and testing dataset," *Rock Mechanics Bulletin*, vol. 2, no. 1, p. 100028, 2023. Available: https://doi.org/10.1016/j.rockmb.2023.100028

[15] D. Blazquez and J. Domenech, “Big Data sources and methods for social and economic analyses,” *Technological Forecasting and Social Change*, vol. 130, pp. 99–113, 2018. Available: https://doi.org/10.1016/j.techfore.2017.07.027

[16] E. Ferrara, P. de Meo, G. Fiumara, and R. Baumgartner, “Web data extraction, applications and techniques: A survey,” *Knowledge-Based Systems*, vol. 70, pp. 301–323, 2014. Available: https://doi.org/10.1016/j.knosys.2014.07.007

[17] X. L. Dong and D. Srivastava, “Big data integration,” *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, 2013, pp. 1245–1248. Available: https://doi.org/10.1109/ICDE.2013.6544914

[18] J. Unpingco, *Python Programming for Data Analysis*. Springer, 2021. Available: https://doi.org/10.1007/978-3-030-68952-0

[19] M. D. A. Praveena and B. Bharathi, “A survey paper on big data analytics,” *2017 International Conference on Information Communication and Embedded Systems (ICICES)*, 2017, pp. 1–9. Available: https://doi.org/10.1109/ICICES.2017.8070723

[20] B. Lehner, G. Czisch, and S. Vassolo, “The impact of global change on the hydropower potential of Europe: a model-based analysis,” *Energy Policy*, vol. 33, no. 7, pp. 839–855, 2005. Available: https://doi.org/10.1016/j.enpol.2003.10.018

[21] A. Adadi, “A survey on data-efficient algorithms in big data era,” *J Big Data*, vol. 8, article 24, 2021. Available: https://doi.org/10.1186/s40537-021-00419-9

[22] IBM documentation, Available: https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=guide-formatting-reports

[23] M. Streit and N. Gehlenborg, “Bar charts and box plots,” *Nature Methods*, vol. 11, no. 2, pp. 117–117, 2014. Available: https://doi.org/10.1038/nmeth.2807

[24] J. Hoelscher and A. Mortimer, “Using Tableau to visualize data and drive decision-making,” *Journal of Accounting Education*, vol. 44, pp. 49–59, 2018. Available: https://doi.org/10.1016/j.jaccedu.2018.05.002

[25] S. J. Shin, J. Woo, and S. Rachuri, “Predictive Analytics Model for Power Consumption in Manufacturing,” *Procedia CIRP*, vol. 15, pp. 153–158, 2014. Available: https://doi.org/10.1016/j.procir.2014.06.036

[26] J. D. Rodriguez, A. Perez, and J. A. Lozano, “Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569–575, 2010. Available: https://doi.org/10.1109/TPAMI.2009.187

[27] A. Kumarasinghe, “Supplementary-Material-for-Research-Regarding-Requirements-Template,” GitHub, 2023. Available: https://github.com/ArithaRTU/Supplementary-Material-for-Research-Regarding-Requirements-Template.git

[28] F. Hardin-Fanning, K. R. Hartson, L. Galloway, N. Kern, and R. Gesler, “Students’ perceptions of a community health advocacy skills building activity: A descriptive analysis,” *Nurse Education Today*, vol. 120, p. 105627, 2023. Available: https://doi.org/10.1016/j.nedt.2022.105627

[29] A. S. de A. Lopes, L. D. N. Santos, M. de C. Razé, and R. Lazzarini, “Alopecia areata: descriptive analysis in a Brazilian sample,” *Anais Brasileiros de Dermatologia*, vol. 97, no. 5, pp. 654–656, 2022. Available: https://doi.org/10.1016/j.abd.2021.04.016

[30] D. Sedov, “Restaurant closures during the COVID-19 pandemic: A descriptive analysis,” *Economics Letters*, vol. 213, p. 110380, 2022. Available: https://doi.org/10.1016/j.econlet.2022.110380

[31] N. Irzavika and S. H. Supangkat, “Descriptive Analytics Using Visualization for Local Government Income in Indonesia,” *2018 International Conference on ICT for Smart Society (ICISS)*, 2018, pp. 1–4, Available: https://doi.org/10.1109/ICTSS.2018.8550006

[32] M. Dinmohammadi, S. Mohammadi, M. Taherkhani, and M. Ali Vadegary, “Factors contributing to coronavirus disease 2019 vaccine hesitancy among healthcare workers in Iran: A descriptive-analytical study,” *Clinical Epidemiology and Global Health*, vol. 18, p. 101182, 2022. Available: https://doi.org/10.1016/j.cegh.2022.101182

[33] F. Ridzuan and W. M. N. Wan Zainon, “Diagnostic analysis for outlier detection in big data analytics,” *Procedia Computer Science*, vol. 197, pp. 685–692, 2022. Available: https://doi.org/10.1016/j.procs.2021.12.189

[34] L. Li et al., “Diagnostic analysis of regional ozone pollution in Yangtze River Delta, China: A case study in summer 2020,” *Science of The Total Environment*, vol. 812, p. 151511, 2022. Available: https://doi.org/10.1016/j.scitotenv.2021.151511

[35] S. Liu, Y. Li, and W. (David) Fan, “Mixed logit model based diagnostic analysis of bicycle-vehicle crashes at daytime and nighttime,” *International Journal of Transportation Science and Technology*, vol. 11, no. 4, pp. 738–751, 2022. Available: https://doi.org/10.1016/j.ijtst.2021.10.001

[36] P. Milella, T. Bisantino, F. Gentile, V. Iacobellis, and G. Trisorio Liuzzi, “Diagnostic analysis of distributed input and parameter datasets in Mediterranean basin streamflow modeling,” *Journal of Hydrology*, vol. 472–473, pp. 262–276, 2012. Available: https://doi.org/10.1016/j.jhydrol.2012.09.039

[37] M. Arif, S. C. P. Cheung, and J. Andrews, “Diagnostic analysis of a single-cell Proton Exchange Membrane unitised regenerative fuel cell using numerical simulation,” *International Journal of Hydrogen Energy*, vol. 46, no. 57, pp. 29488–29500, 2021. Available: https://doi.org/10.1016/j.ijhydene.2020.11.165

[38] W. N. Wassouf, R. Alkhatib, K. Salloum, and S. Balloul, “Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study,” *J Big Data*, vol. 7, article 29, 2020. Available: https://doi.org/10.1186/s40537-020-00290-0

[39] T. Lu, J. Du and R. J. Jiao, “Operating Data-driven Predictive Analytics for Tele-diagnosis of Refrigeration Systems: A Case Study,” *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2018, pp. 1647–1651. Available: https://doi.org/10.1109/IEEM.2018.8607436

[40] D. S. A. Kandawala, M. A. M. Mansoor, R. T. Ramanayake, D. M. Wanniarachchi, K. G. L. Bogahawatte and P. P. G. D. Asanka, “A case study of Sri Lanka oil price fluctuations and its influencing factors using predictive analytics,” *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, 2016, pp. 1–6. Available: https://doi.org/10.1109/ICIAFS.2016.7946518

[41] J. V. Dirisam, D. Bein and A. Verma, “Predictive Analytics of Donors in crowdfunding platforms: A case study on Donorschoose.org,” *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, 2021, pp. 0834–0838, Available: https://doi.org/10.1109/CCWC51732.2021.9376094

[42] S. Nararya, M. Saputra and W. Puspitasari, “Automation in Financial Reporting by using Predictive Analytics in SAP Analytics Cloud for Gold Mining Industry: a Case Study,” *2021 International Conference on ICT for Smart Society (ICISS)*, 2021, pp. 1–6, Available: https://doi.org/10.1109/ICISS53185.2021.9532504

[43] W. Hoyos, J. Aguilar, M. Raciny, and M. Toro, “Case studies of clinical decision-making through prescriptive models based on machine learning,” *Computer Methods and Programs in Biomedicine*, vol. 242, p. 107829, 2023. Available: https://doi.org/10.1016/j.cmpb.2023.107829

[44] A. Amitanand Sinha and S. Rajendran, “Study on facility location of air taxi skyports using a prescriptive analytics approach,” *Transportation Research Interdisciplinary Perspectives*, vol. 18, p. 100761, 2023. Available: https://doi.org/10.1016/j.trip.2023.100761

[45] S. Srinivas and A. R. Ravindran, “Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework,” *Expert Systems with Applications*, vol. 102, pp. 245–261, 2018. Available: https://doi.org/10.1016/j.eswa.2018.02.022

[46] T. Brandt, S. Wagner, and D. Neumann, “Prescriptive analytics in public-sector decision-making: A framework and insights from charging infrastructure planning,” *European Journal of Operational Research*, vol. 291, no. 1, pp. 379–393, 2021. Available: https://doi.org/10.1016/j.ejor.2020.09.034

[47] M. d. Leoni, M. Dees and L. Reulink, “Design and Evaluation of a Process-aware Recommender System based on Prescriptive Analytics,” *2020 2nd International Conference on Process Mining (ICPM)*, 2020, pp. 9–16. Available: https://doi.org/10.1109/ICPM49681.2020.00013

[48] R. Rivera-Castro and E. Burnaev, “CAUSALYSIS: Causal Machine Learning for Real-Estate Investment Decisions,” *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, 2021, pp. 1–3. Available: https://doi.org/10.1109/DSAA53316.2021.9564210

[49] K. Hopf, “Energy Informatics Mining volunteered geographic information for predictive energy data analytics,” *Energy Inform*, vol. 1, article 4, 2018. Available: https://doi.org/10.1186/s42162-018-0009-3

[50] A. Kumarasinghe and M. Kirikova, “Knowledge Graph for Reusing Research Knowledge on Related Works in Data Analytics,” *Advanced Information Systems Engineering Workshops. CAiSE 2024. Lecture Notes in Business Information Processing*, vol. 521, Springer, pp. 186–199, 2024. Available: https://doi.org/10.1007/978-3-031-61003-5_17

[51] A. Kumarasinghe, Knowledge Graph for Reusing Research Knowledge on Related Works in Data Analytics (Version 2.0.4) [Computer software], 2023. Available: https://github.com/ArithaRTU/Knowledge-Graph-for-Reusing-Research-Knowledge-on-Related-Works-in-Data-Analytics.git

Appendix: Common generic requirements template for data analytics

Requirement Types	Analytics Project Phase			
	Initiation	***Acquisition***	***Analysis***	***Presentation***
Common (for all analytics types) requirements	The goal of the analytics project must be defined, specifying the field/context of the project.	The analytics project must have defined dataset(s) that must be used to carry out the analysis specifying properties (data types, data columns/attributes) of a dataset.	The algorithm(s) used for analysis within the analytics project must be defined, including specific equations and variables.	The medium (written report, interactive dashboard, application) by which the results of the analysis (findings, model(s)) must be presented must be defined.
	The system(s)/object(s) on which the analysis is being carried out must be defined.	The sources of data that can be used within the analytics project must be defined.	If a model(s) is being created as a result of the analysis, the expected output(s) of the model(s) must be defined for the analytics project alongside the performance measures that will be used to evaluate the models' performance	If a written report(s) must be formulated as a result of the analysis, the contents and the formatting of the written report(s) must be defined.
	If the results of the analytics project must be verified, specify the strategy that must be used to verify the results of the analysis.	What data acquisition method(s)/technique(s), as well as the tool(s)/software(s), must be used to acquire the data from the data source, including what and how data is to be derived, must be specified.	The tool(s)/software that must be used for analysis must be specified.	The expected utility of the result(s) of the analytics project must be defined
		If data is from multiple sources, the data integration technique(s) that must be used to create the dataset(s) must be defined.		
		If Extract, Transfer, Load (ETL) is needed, specifications regarding the process must be defined.		
		If data warehousing is needed, specifications for data warehousing must be defined.		
		Permissions regarding the use of incomplete data must be explicitly stated.		

Requirement Types	**Analytics Project Phase**			
	Initiation	***Acquisition***	***Analysis***	***Presentation***
		If preexisting mathematical/machine learning models must be used within the analytics project, what the model(s) are, the source(s) of the model(s), and how it must be utilized should be defined.		
Descriptive Analytics Specific	The level of detail and technicality used when reports are being written to the stakeholders of the project must be defined, specifying terminology/jargon that will be used within the context of the analysis.		The specific method(s) that will be used to carry out the analysis should be selected.	Gen (Des) 1.6: How the findings of the analysis must be visualized should be defined.
			To be used practices that relate to the reliability of analysis results must be defined for the analytics project.	
Diagnostic Analytics Specific	The analytics project must have a quantitative metric(s) that is used to evaluate the system.		Gen (Dia) 1.7: The method that will be employed to carry out the analysis must be defined.	Gen (Dia) 1.9: The graphical representations that are required when presenting the analytic results must be defined.
	The analyst must be aware of the level of expertise of a client and define the key terminology within the context of the data analytics project accordingly.			Gen (Dia) 3.1: The format by which the different causes of the issue must be categorized must clearly be stated when presenting the results of the analytics project.
Predictive Analytics Specific		The features that will be used to train the model must be defined.		
Prescriptive Specific Requirements		The features that will be used to train the model must be defined.	The method/model(s) used to create the prescriptive analytics model or algorithm must be defined.	