Software Requirements Elicitation Techniques Selection Method for the Project Scope Management

Denys Gobov[a], Inna Huchenko[b]

[a] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37, Peremohy ave., Kyiv, 03056, Ukraine*
[b] *National Aviation University, 1, Liubomyra Huzara ave., Kyiv, 03058, Ukraine*

Abstract
Project Scope Management is one of the ten knowledge areas described in PMBOK. It refers to the set of processes that ensures a project's scope is accurately defined and mapped. Elicitation is a critical part of the "Collect Requirements" process of the Scope Management that helps to derive and extract information from stakeholders or other sources. The results of elicitation are used as inputs for requirement analysis and management activities. Multiple elicitation techniques may be applied alternatively or in conjunction with other techniques to accomplish the elicitation. Business analysts can modify existing techniques or create new ones to adjust the project context. The selection of the best-suited techniques influences the business analysis approach, which is an important part of the scope management plan. This paper is intended to analyze the current practice of elicitation techniques application in the software development projects, define factors influencing technique selection based on the two-classification Machine Learning model, and predict the usage of a particular elicitation technique depending on the project attributes and business analyst background. We conducted a survey study involving 328 specialists from Ukrainian IT companies. Gathered data was used to build and evaluate the prediction models.

Keywords
Project scope management, requirements management plan, collect requirements, elicitation technique, machine learning, prototyping, business rule analysis, observation.

1. Introduction

Project Scope Management is one of the knowledge areas described in [1]. Scope Management techniques enable project managers and supervisors to allocate the right amount of work necessary to successfully complete a project – concerned primarily with controlling what is and what is not part of the project's scope. For a project manager, the scope management knowledge area is critical, and the Project Management Institute (PMI)® emphasizes this. Six processes are defined for Project Scope Management in [1], namely:
- Plan Scope Management
- Collect Requirements
- Define Scope
- Create WBS
- Validate Scope
- Control Scope

A requirements management plan, which is one of the outputs of the "Plan Scope Management" process, is a component of the project management plan that describes how project and product requirements will be defined, developed, monitored, controlled, and validated [1]. The project manager and business analyst are responsible for defining how project requirements are collected, analyzed, documented, prioritized, and traced. The success of any project is directly related to the accurate definition and documentation of stakeholder needs. Tools and techniques described in [1] for the "Collect Requirements" process intersect with the elicitation techniques list defined in [2]. However, a deeper understanding of the requirements collection could be achieved only by looking into the business analysis area.

Elicitation is defined as one of six core knowledge areas in the business analysis [3]. Requirements elicitation or gathering is a start point of defining stakeholder needs and requirements to meet objectives. The plan of elicitation activities is a part of the global project schedule. Given the fundamental role of these project activities, special attention is paid to defining the available source of requirements and selection of elicitation techniques. The preliminary decision regarding techniques is made during business analysis approach planning, but it can be changed during the first step of elicitation activities - "preparing for elicitation" [2]. The list of proven techniques and approaches are defined in international standards [4, 5] and industrial bodies of knowledge [3, 6, 7, 8]. The list of factors that influence the effectiveness and applicability of techniques is mentioned in these sources and contains such characteristics as historical information and lessons learned repositories, organization's culture, skills of the business analysis practitioner, stakeholders who are involved and their group dynamics, etc. Usually, expert judgment plays the leading part in determining the tools and techniques to be used for accomplishing the project's processes. Expert judgment is limited by the experience of a particular company's expert and can face difficulties in the case of the project context. This study was conducted to analyze the current practices of using elicitation techniques in software development projects and built prediction models based on machine learning algorithms that can support experts in selection techniques activities, and assess the accuracy of these models for selected elicitation techniques. The input data was collected during a survey of practicing specialists from Ukrainian and international companies with branches in Ukraine involved in requirement gathering for IT projects. The second section addressed the related studies regarding using machine learning in project management and business analysis, particularly for requirement elicitation techniques selection, and guidance on their use. Section 3 is devoted to the structure of the questionnaire, the machine learning model approach for technique selection, and the modeling results. Section 4 concludes the paper with a discussion of the findings of our study and future work.

2. Related works

Most studies regarding using machine learning algorithms in project management are devoted to effort and duration prediction [9, 10]. The proposed models for estimation are intended to provide a decision support tool for project managers that develop or implement software systems. Separately worth noting the study on the application of machine learning for the projects' assessment and selection [11]. There are many studies conducted to analyze the practical and theoretical aspects of selection elicitation techniques and provide some guidance on their use. Hickey and Davis [12] conducted a series of in-depth interviews with some of the world's most experienced analysts to analyze how they select elicitation techniques based on a variety of situational assessments. Darwish, Mohamed and Abdelghany [13] proposed a neural network-based model for elicitation technique selection. The attributes, which are relevant to the context of the elicitation process and influence techniques selection, were analyzed using a multiple regression model to find the most important attributes that influence techniques selection and eliminate the less critical ones. A neural network based model for elicitation techniques selection was developed. The limitation of this model is that the interdependencies between elicitation techniques were not analyzed, a subset of attributes was used in the resulted model. In [14] authors proposed a framework for selecting elicitation techniques based on contextual attributes of the elicitation process and established the adequacy values of each technique for each attribute value. The set of attributes is relevant to the context of the elicitation process and influence the selection of one or other techniques were discovered. Two groups of

students were involved in the experiment, practitioners did not take part in the experiment. Wong and Mauricio [15] defined a set of factors that influenced each activity of the requirements elicitation process and, consequently, the quality: learning capacity, negotiation capacity, permanent staff, perceived utility, confidence, stress, and semi-autonomous. Tiwari, Rathore, and Gupta [16] proposed a framework that allows defining the most appropriate elicitation technique based on the project's contextual information. The limitation of this study is that the mapping function is the theoretical one.

The following elicitation techniques were selected for analysis in this study:

1. Stakeholder list, map or personas;
2. Observation;
3. Business rules analysis;
4. Prototyping.

The set of techniques was defined based on the usage frequency according to [17] and is shown in Figure 1.
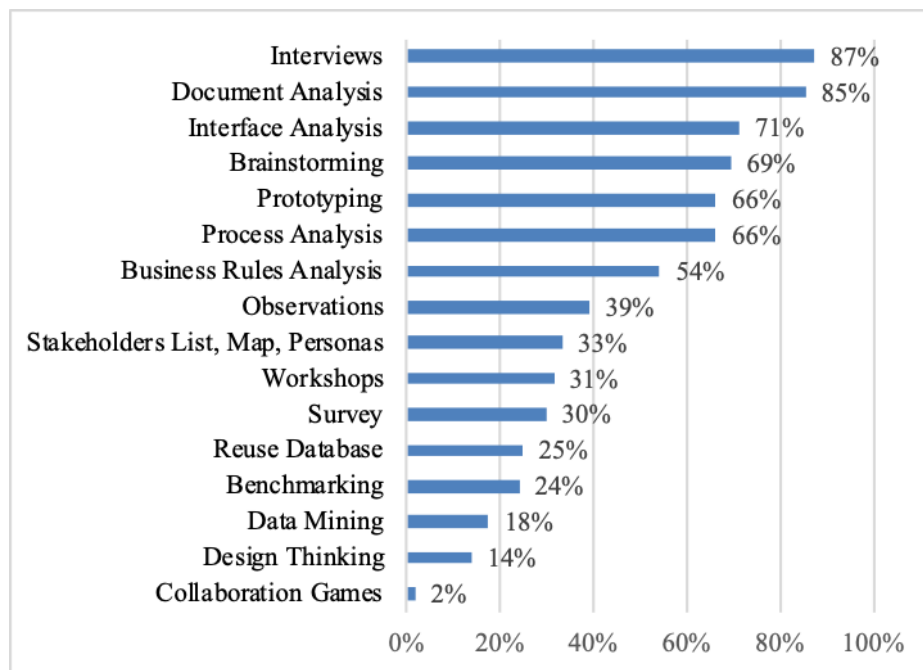
Figure 1: Elicitation techniques usage

The techniques selected for the research are neither the most popular nor the least popular. It means that they are more likely to be questioned about use. A short definition of these techniques is given below.

Business rules analysis is used to identify, express, validate, refine, and organize the rules that shape day-to-day business behavior and guide operational business decision making. It involves gathering business rules from sources, expressing them clearly, validating them with stakeholders, refining them to best align with business goals, and organizing them so they can be effectively managed and reused [2].

Prototyping is used to elicit and validate stakeholder needs through an iterative process. It is a method of obtaining early feedback on requirements by providing a model of the expected solution before building it. Prototypes are also known as proof of concepts (PoC). Because prototypes are tangible, stakeholders are able to visualize and possibly experiment with a model of the product rather than discussing abstract representations of the requirements. This provides an opportunity to validate a conceptual solution against requirements to look for potential gaps [2, 9].

Observation is used to elicit information by viewing and understanding activities and their context. This technique is helpful when domain specialists are unable to spend the time needed to share their expertise or are unable to express and denote their knowledge. Observation provides a direct way of viewing individuals in their environment and how they perform their jobs or tasks and carry out

processes. It is particularly helpful for detailed processes when the people who use the product have difficulty or are reluctant to articulate their requirements [6, 7, 8].

Stakeholder lists, maps, and personas assist the business analyst in analyzing stakeholders and their characteristics. Stakeholder lists contain the key attributes of the project's stakeholders and it is central to both stakeholder analysis activities and the planning work the business analyst performs for elicitation, collaboration, and communication. Stakeholder maps are diagrams that depict the relationship of stakeholders to the solution and to one another. A persona is defined as a fictional character or archetype that exemplifies the way a typical user interacts with a product. Personas are helpful when there is a desire to understand the needs held by a group or class of users. Although the user groups are fictional, they are built to represent actual users. Research is conducted to understand the user group, and the personas are then created based upon knowledge rather than opinion [2, 6].

3. Survey study

A survey description is divided into 3 steps like questionnaire design and data gathering, machine learning-based method for elicitation techniques selection, and obtained results justification.

3.1 Questionnaire design and data gathering

The literature review has shown that many kinds of research have been conducted for identifying common patterns and problems in IT business analysis and requirements elicitation in particular. However, after studying the existing questionnaires developed for international surveys, we realized the necessity of adjusting them to Ukrainian IT companies' specifics. It was decided to take the questions' basis from NaPIRE initiative [18] and rework it concerning mentioned above sources such as [1, 2, 6, 7]. Survey items were carefully written using the business analysis vocabulary, mostly from BABOK. Types of questions used for the questionnaire are open-ended, closed-ended (multiple and single choices) and Likert scale. The total number of questions is 43. After several rounds of internal peer reviews, the questionnaire was given for validation by business analysis experts from the Ukrainian IT industry. Among the comments received as the first feedbacks, there were remarks about the time needed for answering the questions (it took too long to complete the questionnaire) and the complexity of some terms that might cause clarity problems for young professionals. After the recommended improvements were done, cognitive interviews were conducted with 10 potential respondents to determine how they interpret the terms, questions, and answer options. After that step questionnaire was ready for distribution.

Our target group of respondents was IT professionals from Ukraine, mainly business analysts but also other roles involved in business analysis or requirements engineering activities. The overall number of survey participants is 328. The questionnaire itself was created using Google forms and link to it was shared in the local Business Analysis communities, professional and social networks, and via personal contacts in TOP 10 Ukrainian IT companies. The answers were collected in one month. After that, data were merged and coded for further analysis. The following questions' categories were included in the questionnaire:
- Q1: General Information.
- Q2: Requirements Elicitation and Collaboration.
- Q3: Requirements Analysis and Design.
- Q4: Requirements Verification and Validation.
- Q5: Requirements Management.
- Q6: Attitude to the Business Analysis in the project.
- Q7: Problems, Causes, and Effects.

In the given article we focus on the Elicitation and Collaboration topic in the context of general information questions about respondents' backgrounds. The result regarding the current state of requirements elicitation techniques in different software project contexts and statistically significant relationship between project context's attributes and elicitation techniques were described in [18].

Q1: General Information. Questions in this section were intended to give the context such as:

1. Project size.
2. The main industrial sector of the current project. The set of industrial sectors was taken from [18] and reworked to domain areas within which services are offered by most of the Ukrainian IT Companies.
3. Company type: IT or non-IT. For IT companies the separation was made among Outstaff, Outsource, and Product companies.
4. Company size.
5. Class of systems or services such as business, embedded, scientific software, etc.
6. Team distribution (co-located or dispersed).
7. Role in the Project (primary and secondary).
8. Experience in business analyst (BA) or requirements engineer (RE) role.
9. Certifications.
10. Way of working in the project (adaptive vs predictive). The Likert scale with five categories was used.
11. Project category for most of the participant's projects (e.g. greenfield engineering).
12. BA/RE activities which the respondent is usually involved in.

Q2: Requirements Elicitation and Collaboration. Within the given questions category we were interested in elicitation sources, techniques, and project roles that have primary responsibility for the solution requirements (functional, non-functional requirements) elicitation on the respondent's ongoing project.
The following types of elicitation sources were considered:
1. Collaborative (relies on stakeholders' expertise and judgments)
2. Experiments, e.g. observational studies, proofs of concept, and prototypes.
3. Research, i.e. information from materials or sources that are not directly known by stakeholder.
4. 16 elicitation techniques were proposed as answer options with the ability to select as many as needed for reflecting the full range used by respondents. Typical cases were taken as the base for the requirements elicitation responsibility topic and resulted in the following options:
5. Business Analyst/Requirements Engineer.
6. Product Owner/Business Analyst.
7. Product Owner/Product Manager.
8. Project Lead/Project Manager
9. Solution Architect.
Also, we considered the case when in fact nobody has the primary responsibility.

3.2 Machine learning-based method for elicitation techniques selection

After questionnaire data were received and cleaned, it was decided to apply machine learning (ML) for getting the recommendations on elicitation techniques using depending on the specific combinations of factors. MS Azure Machine Learning Studio (classic) [18] was used as the tool for that purpose. Descriptive statistics based on obtained answers are out of the scope of the given article. The visual representation of the elicitation techniques selection algorithm is shown in Figure 2. The steps are as follows:
1. Form dataset based on survey results.
2. Remove irrelevant records. If respondent does not take part in elicitation activities, his/her answers are removed from dataset.
3. Select features corresponding to Q1 and Q2 described in Section 3.1.
4. Transform answers about elicitation technique usage so that for every participant ID usage of the particular elicitation technique is set to:
   - "1" if the technique was selected
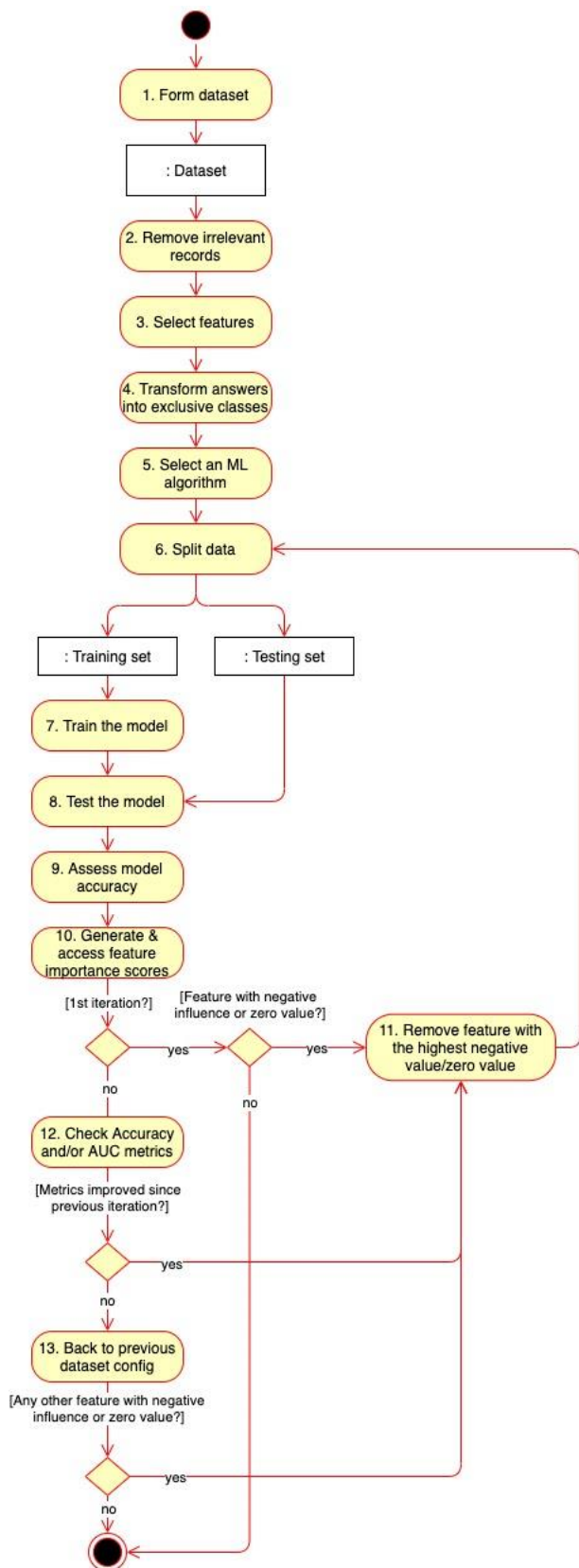   - "0" if was not selected

Figure 2: Machine learning-based algorithm for elicitation techniques selection

That allowed us to use further a two-class classification model as the answers were divided into two mutually exclusive classes.

5. Select an algorithm. Among different algorithms existing for the given model, the Decision Jungle Tree (DJT) was empirically selected as the most efficient from the Accuracy and Area Under Curve (AUC) metrics perspective.
6. Split data. For model training, the initial set of data was split randomly into 2 parts (for training set and testing set) using the 0,75 coefficient.
7. Train the model based on the training set.
8. Test the model based on the testing set.
9. Assess model accuracy. The predicted by ML model influencing factors were analyzed considering the Accuracy and AUC metrics.
10. Generate and assess feature importance scores. Generated scores were analyzed from the positive/negative scores perspective. If it is a first iteration of algorithm, then go to the step 11. Otherwise go to the step 12.
11. If there is a feature with the negative influence or zero value:
    - remove the factor with the highest negative influence or zero value from the dataset. E.g. in the Table 1 the subset of positive/negative scores for the Business Rule Analysis technique are shown. In this case the factor "Data mining" should be removed.
    - go to the step 6.
    Otherwise, **End** the algorithm.

Table 1
Example of feature importance for Business Rule Analysis (before filtering)

Feature	Score
Company Size	0.074074
Company type	0.037037
Team Distribution	0.024691
Survey/Questionnaire	0.024691
System Class	0.012346
Experience	0
Project Size	0
Project Category	-0.012346
Prototyping	-0.024691
Data mining	-0.037037

12. If Accuracy and/or AUC metrics have improved in comparison to the results of the previous iteration, then go to the step 11. Otherwise, go to the step 13.
13. Back to the previous dataset configuration. If there is another factor with negative influence or zero value that has not been tested for the current dataset, then remove the factor with the highest negative influence or zero value from the dataset and go to step 6.
    Otherwise, **End** the algorithm.

3.3   Machine learning model applying

Four elicitation techniques were studied with the help of the Machine Learning two-classification model "Decision Jungle Tree": Stakeholder list, map or personas, Prototyping, Business rules analysis, and Observation. The following metrics were considered as main for model accuracy assessment:

1. Accuracy – a metric that measures the goodness of the classification model as the proportion of true results to total cases.
2. AUC – metric that measures area under curve plotted with true positives on the y-axis and false positives on the x-axis. It provides a single number that lets us compare the models of different types.

3. Precision – a metric that measures the proportion of true results overall positive results.
4. Recall – a metric measuring the fraction of all correct results returned by the model.

The most accurate results were obtained while the DJT algorithm was used listed in Table 2.

Table 2

Accuracy metrics' values

Techniques	Metric for Two Class Decision Jungle algorithm				
	Accuracy	AUC	Precision	Recall	Threshold
Stakeholder list, map or personas	0.738	0.734	0.714	0.435	0.5
Prototyping	0.778	0.802	0.731	1	0.52
Business Rule Analysis	0.741	0.764	0.829	0.708	0.57
Observation	0.716	0.76	0.605	0.742	0.44

The comparison of accuracies before and after filtering is shown in Figure 3 for the Business Rule Analysis elicitation technique. The bold curves designate the results after features filtering based on negatively or zero value scored features exclusion.
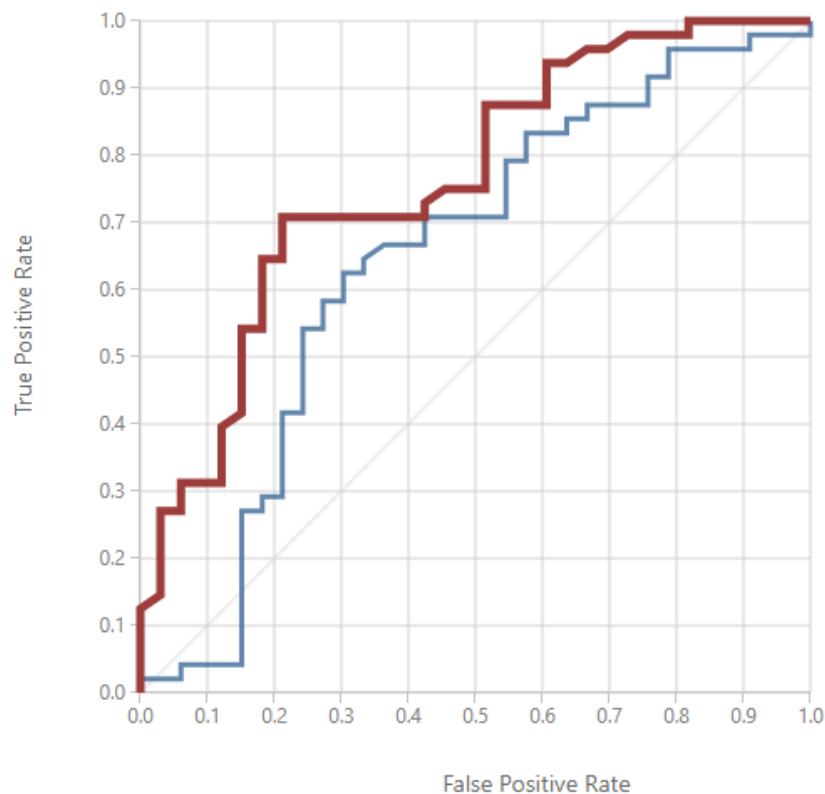
Figure 3: ROC curve before and after features filtering for Business Rule Analysis elicitation technique

The permutation feature importance (PFI) algorithm results are shown in Table 3 below. These results are part of the scoring step of the applied ML model. The most important features are defined per each elicitation techniques independently. Important features are usually more sensitive to the shuffling process and thus result in higher importance scores [19]. According to the received result, we can notice that "Project category" is defined as an important feature for all analyzed elicitation techniques except Business Rule Analysis. "Benchmarking and Market Analysis", "Data Mining", "Observation", "Primary Project Role", System Class" and "Workshop and Focus Groups" are defined as an important feature for two of four elicitation techniques.

Table 3
Positive feature for models with positive values

Stakeholder list, map or personas	Prototyping	Business Rule Analysis	Observation
Benchmarking and Market Analysis	Document analysis	Experience	Benchmarking and Market Analysis
Company size	Elicitation responsibility	Process analysis	Data mining
Data mining	Experience	Project category	Design thinking
Industry sector	Multiple project roles		Document analysis
Observations	Observations		Multiple project roles
Primary project role	Stakeholder list, map or personas		Primary project role
Project category	System class		Process analysis
Reuse database	Ways of working		Project category
Secondary project role	Workshops and Focus groups		Survey/Questionnaire
System class			Workshops and Focus groups

4. Conclusion

"Collect Requirements" is one of the processes in the Project Scope Management. This process is tightly related to the requirements elicitation activity as part of the project's business analysis. The survey study dedicated to the analysis of the current state of requirements elicitation techniques in different software project contexts has been conducted. The survey structure was built based on the worldwide known industrial standards. The survey was held among practitioners from the Ukrainian IT and non-IT companies. 328 specialists, including project managers, business analysts and product owners, took part in the survey. Participants' background factors and combinations of selected elicitation techniques were analyzed using the machine learning-based method involving the Decision Jungle Tree algorithm. Influence of the project factors on the particular requirements elicitation techniques selection was detected. The results showed that it is possible to predict/recommend, with a high level of accuracy, the usage of the following elicitation techniques depending on the project context and other selected techniques: Stakeholders list, map or personas, Prototyping, Business rules analysis, and Observation.

Our study had several limitations. The list of techniques included in the survey is not exhaustive. Elicitation techniques may be applied alternatively or in conjunction with other techniques. Due to specific project context business analysts are encouraged to modify techniques or invent new ones. Considering that the survey was limited to one country only, its results cannot be extrapolated to the worldwide software industry (even though the IT industry in Ukraine is integrated into international environments, especially outsourcing and outstaffing companies, whose employees were the majority of respondents (65%). Several directions for future research can be considered. Other classification models such as Two Class Boosted Decision Tree, Two Class Decision Forest, Two Class Neural Network, etc. can be used for building prediction rules. More sophisticated algorithm can be applied for analysis the space of feature importance scores, for example the slump-vector method, simulated annealing or ant algorithm. The proposed approach can be used for building prediction model for selection techniques used in Define Scope, Validate Scope, and Control Scope processes of the Project Scope Management knowledge area.

5. References

1.   Project Management Institute, A Guide to the Project Management Body of Knowledge (PMBOK® Guide), ver. 6., Project Management Institute, 2017.
2.   International Institute of Business Analysis, A guide to the business analysis body of knowledge (BABOK Guide), ver. 3., IIBA, 2015.
3.   D. Gobov, et al., Approaches for the Concept "Business Analysis" Definition in IT Projects and Frameworks, in: Proceedings of the 9th International Conference "Information Control Systems & Technologies" (ICST 2020), Odesa, Ukraine, 2020, pp. 321–332, CEUR Workshop Proceedings.
4.   International Institute of Business Analysis, A Core Standard A Companion to A Guide to the Business Analysis Body of Knowledge (BABOK® Guide), ver. 3, IIBA, 2017.
5.   ISO/IEC/IEEE, Systems and software engineering – Life cycle processes – Requirements engineering. ISO/IEC/IEE, Standard 29148–2011, 2011.
6.   K. Pohl, Requirements engineering: fundamentals, principles, and techniques, Springer Publishing Company, 2010.
7.   Project Management Institute, The PMI Guide to BUSINESS ANALYSIS, PMI, Newtown Square, Pennsylvania, 2017.
8.   D. Paul, et al., Business analysis, 3rd ed., BCS, The Chartered Institute for IT, 2014.
9.   P. Pospieszny, B. Czarnacka-Chrobot, A. Kobylinski, An effective approach for software project effort and duration estimation with machine learning algorithms, Journal of Systems and Software 137 (2018) 184–196. doi.org/10.1016/j.jss.2017.11.066.
10.   A. Ali, C. Gravino, A systematic literature review of software effort prediction using machine learning methods, Journal of Software: Evolution and Process 31(10) (2019) e2211. doi.org/10.1002/smr.2211.
11.   F. Costantino, G. Di Gravio, F. Nonino, Project selection in project portfolio management: An artificial neural network model based on critical success factors, International Journal of Project Management 33(8) (2015) 1744–1754. doi.org/10.1016/j.ijproman.2015.07.003
12.   A. M. Hickey, A.M. Davis, Elicitation technique selection: how do experts do it?, in: Proceedings of the 11th IEEE International Requirements Engineering Conference, 2003, pp. 169–178. doi.org/10.1109/icre.2003.1232748
13.   N. Darwish, A. Mohamed, A. Abdelghany, A hybrid machine learning model for selecting suitable requirements elicitation techniques, International Journal of Computer Science and Information Security 14(6) (2016) 1–12.
14.   O. Dieste, N. Juristo, Systematic review and aggregation of empirical studies on elicitation techniques, IEEE Transactions on Software Engineering 37(2) (2011) 283–304. doi.org/10.1109/tse.2010.33.
15.   Wong, L., Mauricio, D.: New Factors That Affect the Activities of the Requirements Elicitation Process, Journal of Engineering Science and Technology13(7), (2018) 1992–2015.
16.   S. Tiwari, S. Singh Rathore, A. Gupta, Selecting Requirement Elicitation Techniques for Software Projects, in: Proceedings of the CSI 6th International Conference on Software Engineering, IEEE, 2012, pp. 1–10.
17.   D. Gobov, I. Huchenko, Requirement Elicitation Techniques for Software Projects in Ukrainian IT: An Exploratory Study, in: Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, FedCSIS 2020, IEEE, Sofia, Bulgaria, 2020, pp. 673–681. doi.org/10.15439/2020f16.
18.   D. Fernandez, S. Wagner, Naming the pain in requirements engineering: A design for a global family of surveys and first results from Germany, Information and Software Technology 57 (2015) 616–643. doi.org/10.1016/j.infsof.2014.05.008.
19.   M. Bihis, S. Roychowdhury, A generalized flow for multi-class and binary classification tasks: An Azure ML approach, in: Proceedings of the IEEE International Conference on Big Data (Big Data), IEEE, 2015, pp. 1728–1737. doi.org/10.1109/bigdata.2015.7363944
20.   X. Zhang, et al, Machine Learning Studio (classic) Module Reference, 2019. URL: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference.