VERSION 0.5
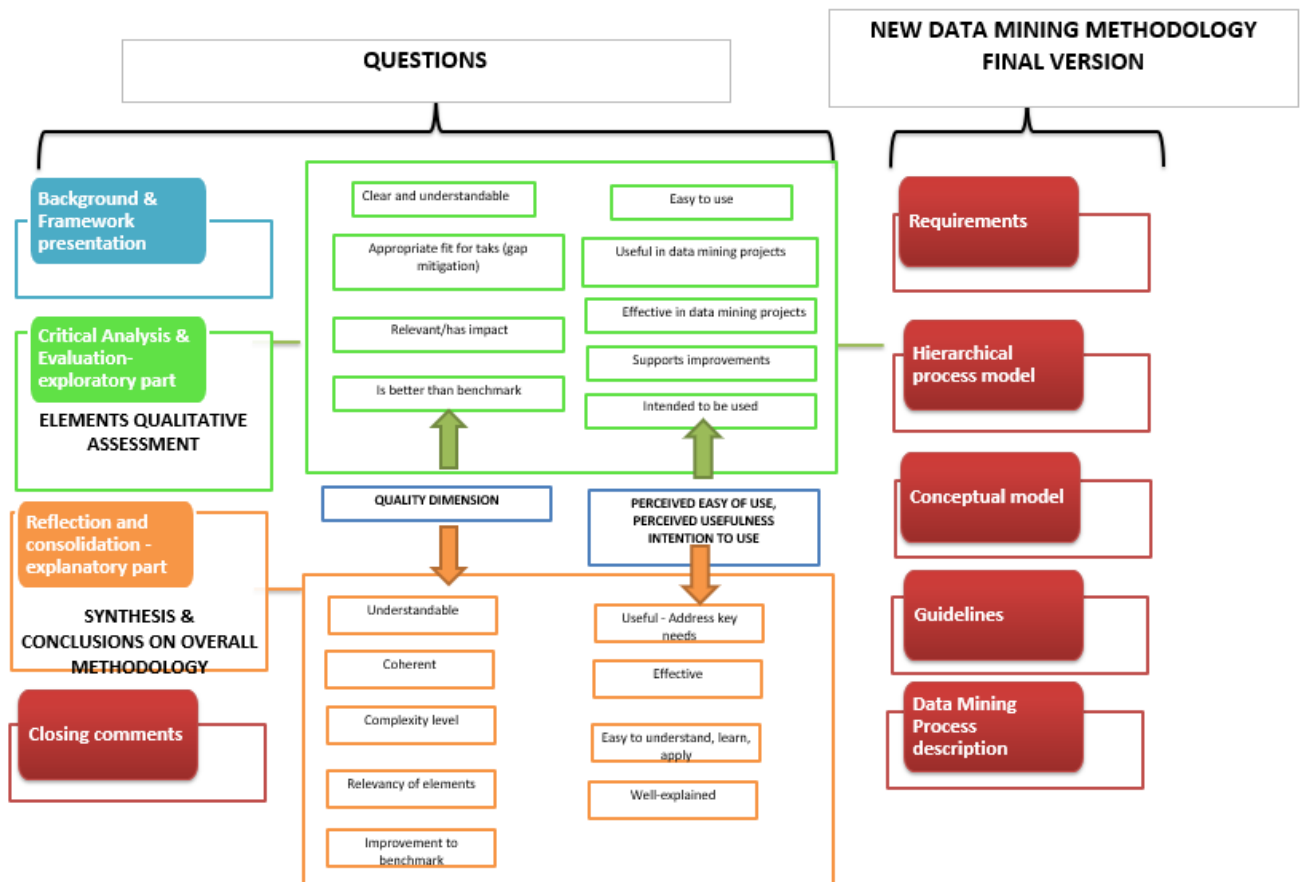
Dated: 30th July 2020
Note: Draft interview guide

Interview Guide Structure

Phase 1 Background (10-15 min) + Framework presentation (20 min)

There are two objectives of this interview phase:
1) ***Objective 1*** *is to introduce this research project and get understanding and short background of interviewee.*

   NB! Part is rather short for interviewees who have been met in previous research project, and is more in-depth for new interviewees who contribute the first time

2) ***Objective 2*** *aims to present to interviewees the adapted data mining framework. Initially, the gaps identified in the benchmark (CRISP-DM) data mining methodology are presented. Then, the adapted framework is demonstrated focusing on adjustments introduced to address the gaps.*

1. Questions

1.1. Let me provide You with explanation about the research project – read out from Consent Form (provided in **Appendix 1**)

1.2. Based on the Consent Form presented to You, do You have any questions on the research project and study background? Is everything clear?

1.3. As mentioned, let me introduce You the prototype of the new framework. And as good benchmark example, let's relate it to the use case(s) we discussed with You in the first interview rounds. Please also feel free to ask if You have any questions on the go

NB! Detailed Gaps table is demonstrated during the talk (Worskheet 1 from Excel doc)

Initially, we have analyzed publications concerning data mining methodologies (both in broad set of domains and in banking domain). We have identified 8 types of gaps, which are the following:

- **Interdependencies gaps** relate to the lack of iterations between different phases of CRISP-DM
- **Requirements gaps** relate to lack of lack of tasks for validation and modification of existing requirements, and the elicitation of new ones
- **Universality gaps** concern a lack of support for various analytical outcomes, unsupervised and specialized techniques, as well as deployment formats
- **Validation gap and Actionability gap** concern the Evaluation and Deployment phases respectively. These gaps refer to a lack of support for piloting models in real-life settings
- **Privacy and regulatory compliance gaps** relate to lack of tasks to address regulatory compliance (in particular, GDPR)
- **Process gaps** concern entire data mining process life-cycle. These gaps encompass data mining process controls, quality assurance, and critical process enablers (data, code, tools, infrastructure and organisational factors, are not taken into consideration) required for the effective execution of data mining projects
- **Knowledge management gap** relates to lack of formalization and mechanisms to share and disseminate various specialized business context knowledge across different data mining projects

1.4. Further, we have proposed the solution how to tackle these gaps. Let me in detail present You the solutions, and please ask clarifying questions on the go.

So, to summarize we propose three types of mitigations introduced into benchmark methodology CRISP-DM:
- New tasks
- New phases
- And elements of most established frameworks from IT delivery domain (ITIL) and IT governance (COBIT).

NB! Detailed new data mining framework prototype is demonstrated during the interview part described above as well as additional context and explanations around each new task, phases and other adjustments are provided verbally (Worksheet 3, 4 and 5 in Excel doc)

Regarding new tasks:
- We propose **Universality tasks** in Business Understanding, Modelling and Evaluation phases
- We propose **Business Validation Tasks** in Business Understanding and Evaluation phases

Concerning new phases:
- We propose **Post-deployment and Life-cycle management phase** which immediately follows Deployment
- We propose **Requirements phase** which will be ongoing throughout the whole data mining project cycle
- We propose **Compliance phase**, which similarly to Requirements phase will be encompassing all the data mining project cycle

To tackle **Process gaps** we propose to incorporate certain artifacts from other well-known IT management and governance frameworks.

In relation to that, are You familiar with ITIL and/or COBIT frameworks?

NB! Focus is on elements acquired from ITIL and COBIT frameworks and incorporated into proposed framework (Worksheet 2 and 3 in Excel doc).
More extensive discussion of ITIL and COBIT with experts of these frameworks (part of the interviewees' cohort).

Internal validity checkpoint: Let me summarize our discussion so far…..[conclusions, findings are summarized and reconfirmed with participant]

Phase 2 Exploratory part – critical analysis and evaluation of each proposed elements of the presented framework (ca 40 min)

The objective of this phase is to discuss and elicit opinion from interviewee on each adaptation artifact of the proposed framework based on the pre-determined criteria. They concern **quality dimension of the proposed artifact** *and investigate if proposed artifact:*

- o *Is appropriate for the task*
- o *Produces relevant result(s)/has impact*
- o *Is formulated and presented in the form that participant can relate and understand how to apply and use it*
- o *Is better than available guidance (in our case benchmark CRISP-DM)*

We also investigate completeness, we discuss with participant if there are some other artifacts and elements which are not captured, but would be relevant to include to address potential issues.

There is also other set of criteria which concerns three complementary dimensions of the proposed artifact and is based on TAM (Technology Acceptance Model) and TTF (Task Technology Fit) paradigms. These dimensions refer to **perceived usefulness/ease of use, satisfaction, and intention to use.** *These criteria help to determine if proposed artifact is perceived by user(s) as:*

- o *Easy to understand, learn to use and apply*
- o *Is useful in data mining projects*
- o *Is efficient and effective for data mining projects execution*
- o *Is intended to be used by interviewee if the new proposed framework is available*

These criteria are covered in the complementary questionnaire to be distributed and filled in by participants after the interview.

2. Questions:

2.1. In relation to **Universality tasks**, in Your opinion (please comment and motivate):

- How would You have applied it in the context of the use case/data mining project we discussed earlier?
- Does adaptation artifact serves its purpose and is adequate for adaptation task?
- Would it have been useful for Your data mining project execution (as part of overall framework)? Why?

2.2. In relation to **Business Validation tasks**, in Your opinion (please comment and motivate):….(repeat questions above)

2.3. In relation to **Post-Deployment Phase**, in Your opinion (please comment and motivate):….(repeat questions above)

2.4. In relation to **Requirements Phase**, in Your opinion (please comment and motivate):….(repeat questions above)

2.5. In relation to **Compliance Phase**, in Your opinion (please comment and motivate):….(repeat questions above)

2.6. In relation to selection of **ITIL management practices**, in Your opinion (please comment and motivate):….(repeat questions above)

2.7. In relation to selection of **ITIL xxxx**, in Your opinion (please comment and motivate):….(repeat questions above)

2.8. In relation to selection of **COBIT Internal Controls and Quality Assurance**, in Your opinion (please comment and motivate):….(repeat questions above)

Internal validity checkpoint: Let me summarize our discussion so far…..[conclusions, findings are summarized and reconfirmed with participant]

Phase 3 Reflective and Consolidation Part (20-30 min)

The objective of this phase is to obtain interviewee reflections, retrospect opinion/view as regards proposed data mining methodology holistically.

3. Questions:

3.1. Has that been easy for You to understand proposed methodology? Yes/Partially/ No. Why? Which could be design elements which need improvement?

3.2. In relation to previous, how You perceive and evaluate the framework presentation, in particular, what is You feedback about:
- Glossary section?
- Elements definitions?

For example, are they concrete? Are they clear? Do You understand how to act, what are the steps to undertake to execute on pre-scribed elements?

3.3. Is the level of complexity adequate or in contrast: (1) it could be simplified or (2) it could be enhanced more?

3.4. Which elements might overlap or address the same gap/issues?

3.5. Do You perceive if the proposed framework covers all (or most) of potential issues? Or in contrast, there are other artefacts, elements missing? Please motivate

3.6. Does the new methodology address Your needs as Data Scientists (or Project Manager, or Operational leader)? Yes/Partially/No. Why? Which elements do not meet Your needs? Please motivate

3.7. Would it in Your opinion address the needs of other potential framework users – Your colleagues in the bank and outside it, operating in other financial institutions? Yes/Partially/No. Why? Please motivate

3.8. Will the new methodology be an improvement to current CRISP-DM process in terms of addressing its gaps? Why? Please motivate Your answer

Internal validity checkpoint: Let me summarize our discussion so far…..[conclusions, findings are summarized and reconfirmed with participant].

Phase 4 Closing comments

Expressing gratitude for interviewee, informing on the transcript sign-off routine, informing how the results of the study will be summarized and shared within research community.

APPENDICES

Appendix 1 Consent Form

1. **Names of researchers and contact information**

PhD researcher Veronika Plotnikova, University of Tartu, Institute of Computer Science

2. **Purpose of the research project and case study.**

The researcher in earlier works has gathered extensive feedback and information both from expert interviews and meta-analysis of substantial publications corpus existing in the domain. Based on the given information and feedback, the gaps in the existing benchmark methodology CRISP-DM were identified. The research has constructed adapted data mining methodologies to address the gaps. The aim of this case study interviews is to:
- ✓ validate the proposed data mining methodology with the experts based on the number of criteria

3. **Procedures used in the study, that is a short description of what the participant should do during**

Participants participate in semi-structured interviews where they are presented with the proposed prototype of novel data mining methodology adapted to financial services domain. After introduction of the prototype each adaptation element is discussed with interviewee and detailed feedback is registered. The prototype framework is adjusted to reflect experts' opinion to which extent it meets the criteria set.

4. **The study and what steps the researcher will carry out during these activities**

Semi-structured interview and documentation of results. Further, consolidation of findings and publishing.

5. **A text clearly stating that the participation is voluntary, and that collected data will be anonymous.**

Participation is voluntarily and interviewing is fully anonymous.

6. **A list of known risks.**

No any risk to participants as results and interviewing is anonymous, and consolidated findings are reflected in publication.

7. **A description of how confidentiality will be assured. This includes a description of how collected material will be coded and identified in the study.**

Interview will be recorded based on consent of participant or alternatively documented via notes taken in electronic format. During interview participant will remain anonymous and will not called upon or associated with personal details which would allow for participants identification.

None of primary information is to be shared. The findings will be consolidated on higher level of abstraction ensuring further anonymization. Any transcribes of interview will be sign-offed by participant.

The conductor of the study (Veronika Plotnikova) is the only person on the project directly handling data collection. Supervisors of the projects do not participate in the data collection, processing and transcribing. Supervisors direct the PhD student in applying scientific methods and interpretation of the documented findings, results analysis and synthesis. Further, supervisors have access only to final publications draft. Supervisors do not access or review any raw sourced data (interviews, and their recordings in any format as well as non-consolidated, intermediate analysis of collected evidence).

8. **Information about approvals from participating organization.**

PhD researcher has received approvals from direct manager in organization.