

Task 2:

- Describe what info we need to answer our research question(s)
- Decide many possible data collection method(s) + choose one + justify choice
- Discuss variants of the chosen method + justify
- Use method literature as references to justify choices

In order to start this research, we need this kind of information

- 1) Information about what are exactly cyberbullying and online harassment.
- 2) Information about what are the boundary definitions about cyberbullying and online harassment.
- 3) Determine the comments which are the toxic and which are not. Also, is there any level of toxicity. If there is, what are the level of toxicity.
- 4) Find out what kind of text classification techniques used in the past, which technique(s) better than others and what is the performance criteria about identifying toxic comments.
- 5) Steps of reaching the state-of-art level.
- 6) Effectively and properly usage of Conversation AI's open dataset.
- 7) Thinking about how will it affect this research to future research.

Data collection method

The aim of this project is to explore state-of-the-art methods in text classification and build a high performing model on the task of identifying toxic comments in data from online communities. In accordance with this purpose, words are key factor in this research so, qualitative data method is selected. The research needs big and variety kind of data to make very efficient and designed model. Also, we need actual comments data from users for analyzing whence it is not possible to collect with pure data at questionnaires, interviews and surveys. Therefore, selection of documents and reports will help to collect useful data for our purpose[1]. Conversation AI's publicly available dataset of online comments, data of research provides from it.

Method literature references

[1] Johannesson, P., & Perjons, E., (2012, p.30). A design Science Primer. CreateSpace Independent Publishing Platform.

Açıklamalı [NL1]: what do you mean here?

Açıklamalı [HU2R1]: I mean some words or sentences might not be cyberbullying but some are them might be. So, i think we should identify some boundries about which sentences are cyberbullying. For example, are every racism sentences cyberbullying or some of them can be just jokes?

Açıklamalı [NL3R1]: I see, I think this can be part of the definition of cyber bullying. Do you think we need to have it a separate point?

Açıklamalı [HU4R1]: I think if we don't define and make some boundries then how can we say this is cyberbullying and this is not?

Açıklamalı [NL5R1]: I think the boundaries are implied

Açıklamalı [HU6R1]: okay

Açıklamalı [HU7]: I have combined three statements in

Açıklamalı [NL8R7]: Alright! I changed it to 'overview'

Açıklamalı [HU9R7]: Okay, maybe we can narrow some

Açıklamalı [NL10]: what do you mean here? what's an

Açıklamalı [HU11R10]: I mean we can learn informatio

Açıklamalı [NL12R10]: hmm we use the dataset for

Açıklamalı [HU13R10]: Okay it can be remain about

Açıklamalı [NL14R10]: I think removing it is best

Açıklamalı [HU15R10]: okay

Açıklamalı [NL16]: do you mean the model here?

Açıklamalı [HU17R16]: I thought we get this informati

Açıklamalı [NL18R16]: what is this information?

Açıklamalı [HU19R16]: I mean what we did extra or

Açıklamalı [NL20R16]: Oh I see! I don't think this

Açıklamalı [HU21R16]: okay,great

Açıklamalı [NL22]: since we're analyzing said words, I

Açıklamalı [HU23R22]: He said if you analyze words,

Açıklamalı [HU24R22]: What do you think about this ?

Açıklamalı [NL25R22]: I think the words in our case are

Açıklamalı [NL26]: our data is online textual comments

Açıklamalı [HU27R26]: I thought if we have a lot of

Açıklamalı [NL28R26]: oh okay, thanks for clarifying!

Açıklamalı [HU29]: I arranged sentence about

Açıklamalı [NL30R29]: Cool! I think 'observation studie

Açıklamalı [HU31]: New sentence about data collection.

Açıklamalı [NL32R31]: Great! included it in the merge