**Project definition:**

* Data: submissions and comments from the subreddit **Change my view**. About 35K submissions and 2.5M comments.
  + Filter: only comments with more than 150 characters which were posted in the week after the submission was posted – about 0.5M comments
* Causal question: what is the causal effect of a quote of the submitter on the delta, i.e whether a comment that quotes the submitter caused him to give delta to this comment.
* Units: each unit will be a comment and its submission.
* Treatment: whether the comment quotes the submitter: either the submission itself or its parent that was written by the submitter.
* Outcome: binary, whether the submitter gave delta to the comment.
* Features:

1. **Length of the submission**
2. **Length of the title**
3. **Length of the comment**
4. **Sentiment of the submission (http://text-processing.com/docs/sentiment.html)**
5. **Sentiment of the submission’s title**
6. **Sentiment of the comment**
7. **Cosine similarity between the Sentiment of the submission and the Sentiment of the comment**
8. **Number of submissions that the submitter wrote**
9. **Number of submissions that the comment author wrote**
10. **Number of comments that the submitter wrote**
11. **Number of comments that the comment author wrote**
12. **Time between the submission and the comment**
13. **Similarity between the submission and the comment – the idf from all the data – maybe 2 features**
14. **% of the Adjective in the comment- kind of רגש in these POS**
15. Topic modeling – give any word its topic – give the sentence the topic with the highest dist. Of its words.
16. **If this is the first comment of this user to this submission**
17. **Number of comments of this user in this submission until the unit**
18. **Time between the submission and the first comment - Time between the submission and the comment (this unit)**
19. **Number of times the submitter responded to the comment user in this submission**
20. **Number of times the submitter responded in total**
21. **Number of times the submitter responded to the comment user in this submission/number of comments to the submission**
22. **Number of times the submitter responded to the comment user in this submission/ Number of times the submitter responded in total**
23. **Number of times the submitter responded in total/ number of comments to the submission**
24. **The seniority of the commenter in change my view (how many days passed since his first comment in the data)**
25. **The seniority of the submitter in change my view (how many days passed since his first comment in the data)**
26. דני יוגטמה עשה עבודה על טקסט קלאסיפיקיישן- להסתכל מה הפיצ'רים שהשתמש בהם
27. noah smith -NLP for social science