Gergo Gyori (gegy@itu.dk), BSc Data Science
Katalin Literati-Dobos (klit@itu.dk), BSc Data Science
Ludek Cizinsky (luci@itu.dk), BSc Data Science
Lukas Rasocha (lukr@itu.dk), BSc Data Science

Reflections on Data Science 2023

Social Study: Herding effect on Reddit

1   Introduction

Posts that achieve popularity on social media platforms, like Reddit, often outperform average posts significantly, which indicates a qualitative difference between popular posts and the rest. With this phenomenon an interesting question emerges: is the success of a post attributable to its inherent quality, or is it influenced by the herd effect? The social psychological behaviour that leads individuals to perceive an action as the appropriate course simply because it's what "everyone else" seems to be doing. This effect has already been documented in various studies such as [1] [2]. In this study we aim to investigate the herd effect on the Reddit platform, more concretely, we want to find out to what extent does an initial upvote to a post influence its future score?

2   Experiment implementation

Similar to Muchnik et al. [1], to test the effect of a social influence on upvoting reddit posts we conducted a randomised control trial. For the continous period of seven days we collected and monitored recently published posts with no upvotes nor comments from the subreddit *r/all*, and randomly assigned them to two categories: *treatment* and *control*. The treatment group received an initial upvote, while the control group did not. The experiment was scheduled to run daily to ensure that the collected posts had at least 24 hour period between the score loggings.

After the experimentation period ended, we aimed to test the following null hypothesis, with a significance threshold α set to 0.05: *A treatment of a post does not influences its score (number of upvotes minus downvotes) in the future.* Before the testing, some pre-processing steps were necessary. First, we removed the posts which were not continously monitored for a week and secondly, to account for our initial upvote, we subtracted it from the posts' last day scores. This resulted in a total number of 1578 posts, i.e. 789 treated posts and 789 control posts. Finally, since we did not want to make assumptions about the underlying distribution of the data, we used non-parametric tests, namely Empirical bootstrap and Kolmogorov-Smirnov test, to confirm or reject the studied null hypothesis.

3   Results and discussion

Discuss your findings in light of the theories and concepts covered in the course (e.g., causality, experimental designs, extraneous variables, social influence, reproducibility, etc.)

Since we learned a bit too late about the vote fuzzing our experiment only recorded the overall score of the posts, while not looking at changing number of comments which might have been a better indication (no comment count fuzzing).

4   Conclusion

References

[1] Lev Muchnik, Sinan Aral, and Sean J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

[2] Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.

[3] C.T. Bergstrom and J.D. West. *Calling Bullshit: The Art of Skepticism in a Data-Driven World*. Random House Publishing Group, 2020.