

Individual Assignment - Data and Society

What are data, datafication, and algorithms

Blaike and Priest (2019) explain that data is something that can be presented in numbers or words. They also point out that data can be analyzed. Kirk (2019) refers to raw data when data is collected and not analyzed yet. The data collected for our paper are the data of movies. This is information such as director, ratings and the movie duration. This also confirms Kirks (2019) and Blaike and Priest's (2019) suggestions of data being information in the form of numbers or words.

Raw data differs from information which refers to when data is analyzed (Blaike and Priest, 2019). Data would such be the original form of what is collected. Kirk (2019) describes data as information that may contain measurements or descriptions. The data in our analysis are information about the movies and how datafication occurs in movie information. The final analysis that was presented on our website is thus information created from analysis of the collected raw data, such as ratings and genres.

Hintz et. al. (2017) suggests datafication as "vastly enhanced possibilities to understand, predict, and control citizen activities." (p.732). A digitalization of society would occur when society is more dependent on the collection of raw data (van Dijck, 2014). A datafication would then arise when society would be more digitalized. We could see in our research paper that movies consist of data that can be used and analyzed. However, datafication would arise when there is a general interest in collecting the data. Cukier & Mayer-Scoenberger (2013) discuss the datafication as a process where objects that previously were not quantified are now converting it to data. Since digital society would consist of more data (Hintz et. al., 2017) it could make datafication and digitalization to be intertwined concepts.

Klinger & Svensson (2018) refers to algorithms as a process to solve problems. They also see algorithms as a way of control. It would then be related to datafication in how an algorithm controls what data reaches the audience. Burrell (2016) also emphasizes how algorithms are dependent on data input.

“The algorithms in question operate on data. Using this data as input, they produce an output; specifically, a classification” (Burrell, 2016, p.1). Without data, it would then not be necessary to use or possible for any algorithms to exist.

Paßmann & Boersma (2017) explains that algorithms and machines are the same things. It could become a new issue, explained as a “black box” when machines control data and what reaches the audience, and that could prevent users from understanding the underlying data.

Klinger and Svensson (2018) however, points out that algorithms are constructed based on human input. Algorithms could thus form a new tool for corporations to control users, what data they produce and make new forms of economic profit.

The Commercialization of Data

The project we presented analyzed the data and datafication of movie information through the database IMDb. IMDb was created in 1990 and collected information such as genres, duration, and actors about movies. Today, IMDb contains more than six and a half million titles with data available for anyone to access (IMDb, n.d.a).

IMDb is today owned by Amazon which is a privately owned corporation. Behind the database of IMDb is a corporation with the motive of making a profit from their business. Arsenault (2017) also claims about corporations: “Particularly in an age of corporatization, their primary focus is on delivering audiences to advertisers, not delivering content to audiences” (Arsenault, 2017, p.20). The overall goal of the corporations is not to give users a product but rather to make a profit from the product they create and make sure the data can be used for further economical purposes. The purpose of the data usage of IMDb could thus be questioned. This would create a form of commercialization of data, where there is a primary economical aim with the database and data collection. The data could be seen as a form of currency to be used by corporations.

Arsenault (2017) also emphasizes that media corporations use big data as a tool to compete in a globalized media environment. Using big data as a tool would be the datafication of society that converts the information of movies into data. Big data is not only about the large amounts of data but also focuses on the datafication of objects that were not data before (Van Dijck, 2014). Movie data is as such a form of new data that has emerged as a result of the datafication of society. Movie

reviews and ratings have now been quantified into numbers and words that can be used by the end-user of the platforms.

Today, people may rely more on reviews and scores before purchasing a product. To find reviews a user searches an internet database that collects data the user desires. This could be searching for movie ratings or reviews. IMDb holds in such regard, power over what movies a user may choose to buy or view at a cinema. IMDb and its owner Amazon would then control what information reaches the end-user, and in turn, control the user's liking of movies.

A company that owns a database could then promote their products and influence the user to buy certain products or services, which would bring economic profit for the mother company.

Algorithms, as described above, could be a tool for controlling data (Klinger & Svensson, 2018), as well as automating the data usage (Hintz, Dencik, & Wahl-Jorgensen, 2018). The algorithms of IMDb could then control what data reaches the users. Algorithms can select certain movies based on previous data inputs by the users, such as movies the user has rated positively before while excluding movies that a user has given a negative review. As the algorithms are based on human input it could also show how corporations can control the users and how to use it for the corporation's agenda.

Hintz, Dencik, & Wahl-Jorgensen (2018) also argues for how digital citizenship generates data that can be used by both governments and corporations for their specific agenda. Who accesses the data is important to consider from an ethical perspective. "it is still necessary to ask critical questions about what all this data means, who gets access to what data, how data analysis is deployed, and to what ends. "Boyd and Crawford, p.664)

Van Dijck (2014) also discusses the concept of dataveillance to describe how the monitoring data differs from surveillance. Dataveillance would be a continuation to use data for undefined purposes, which is an ethical issue for the participants or platform users who do not know what their data is used for. Van Dijck (2014) points out that society is more dependent on the data, and our study showed that as well movie data can be analyzed for the understanding of the new emerging movie genres.

IMDb offers a Pro feature, which allows the users on the IMDb community, additional functions on the platforms. One of these functions is as explained in the group project, the ability to rate for some movies that get highlighted, and also add featured images for movies (IMDb, n.d.b). This

function suggests IMDb encouraging users to pay to unlock more functions and tools for users. This creates a form of hierarchy and classifying the users of the platform where the user who has the money gets the ability to control the data from others.

Since IMDb is based on user-generated data, the IMDb Pro users would then have an even larger power over what data the platforms include.

A movie database as IMDb with these forms of functions could then be a form of capitalism. This could create one of the data divides Andrejevic (2014) mentions. Some users may be excluded from finding and using the data they desire while the richer users control what data is used on the platform and in turn controlling algorithms.

What our project never discusses, is if the IMDb as a database is biased to promoting American movies since it is an American database. Boyd and Crawford (2012) point out the importance of as well as understanding biases behind platforms. Perhaps IMDb could promote American movies to increase the profit for American based corporations or companies that cooperate with Amazon. In the same manner, IMDb can also exclude data about movies that are rivaling corporations. With the power IMDb has over the datafication of movies, it becomes important for the platform to use the data as was its original purpose, to provide data about movies. Misuse of the platform's data can, for instance, generate issues were people or researchers are being mistaken for actors. Misuse of movie genres can as well be manipulated to make fiction look like a documentary and through movies and a database create new forms of disinformation or fake news on the internet.

References

- Andrejevic, M. (2014). Big data, big questions| the big data divide. *International Journal of Communication*, 8(17).
- Arsenault, A. H. (2017). The datafication of media: Big data and the media industries. *International Journal of Media & Cultural Politics*, 13(1-2), pp. 7-24. DOI: 10.1386/macp.13.1-2.7_1
- Blaike, N. & Priest, J. (2019). *Designing social research: The logic of anticipation*. Cambridge: Polity press.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), pp. 1-12. DOI: 10.1177/2053951715622512
- Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of big data: How it's changing the way we think about the world. *Foreign Aff.*, 92, 28.
- Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance| digital citizenship and surveillance society - introduction. *International Journal of Communication*, 11(9). pp. 731–739.
- IMDb (n.d.a). *IMDb Statistics*. Retrieved from: https://www.imdb.com/pressroom/stats/?ref=helpms_ih_gi_siteindex
- IMDb (n.d.b). *What is included with an IMDbPro Membership*. Retrieved from: https://help.imdb.com/article/imdbpro/membership-benefits/what-s-included-with-an-imdbpro-membership/G6EFQ2AYEG3NKTM2?ref=helpms_helpart_inline#
- Kirk, A. (2019). *Data visualisation: a handbook for data driven design*. Los Angeles: Sage Publications.
- Klinger, U., & Svensson, J. (2018). The end of media logics? On algorithms and agency. *New Media & Society*, 20(12), pp. 4653-4670. DOI: 10.1177/1461444818779750
- Paßmann, J. & Boersma, A. (2017). Unknowing Algorithms: On Transparency of Unopenable Black Boxes. In M. T. Schäfer. & K. van Es (Ed.), *The Datafied Society: Studying Culture through Data*. (pp. 139-146). Amsterdam: Amsterdam University Press.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), pp. 197-208. DOI: 10.24908/ss.v12i2.477