COMM5780M – DIGITAL PRACTICES

CRITICAL REFLECTION: SOCIAL MEDIA ANALYTICS

WORD COUNT: 2607

Introduction

Social media analytics is one of the key elements in the digital media landscape. This realm has been continuously changing as inventions happen daily in this field. However, the scholarships on this have shed light upon its negative aspects. This essay critically reflects on social media analytics, especially algorithms. The essay set out with a context on social media analytics and algorithms then moved on to the critical theories on them. Later, it discusses Sumpter's experiment of understanding the algorithms, followed by the author's own experience of using the method and his critical reflections on it while relating to the theories discussed earlier.

Social Media Analytics

Social media emerged as a direct electronic exchange platform. However, eventually, it shifted to one of the biggest drivers for acquiring and spreading information in different domains (Stieglitz et al., 2018, p.156). The importance of social media in daily life has been readily increasing along with their usage. It was noted that an estimated 3.6 billion users used social media, a figure projected to increase further (Statista, 2020, cited in Zachlod et al., 2022, p.1064). This growth in the use of social media resulted in a huge accumulation of data termed "Social Media Big Data" (Stieglitz et al., 2018, p.156). The analysis of social media big data, called social media analytics, opened up endless possibilities in multiple dimensions; a better understanding of society, enhanced business activities, comprehension of public opinion and influence of political results are few among them (Zachlod et al., 2022, p.1064). The interactions with these social media platforms and our online behaviour feed into the big data. However, we do not know by whom it is used, where our data is being used and in what way it has been transformed. "Everything we do online is recorded; the only questions left are to whom the data will be available and for how long" (Pasquale, 2015, p.3). Living in a society

where we are constantly monitored in a system where we can only observe the actions and results is for which Pasquale coins the term the "black box society". "Tracked ever more closely by firms and government, we have no clear idea of just how far much of this information can travel, how it is used, or its consequences" (Pasquale, 2015, p.3).

Critical Theories of Algorithm

Algorithms are one of the fundamentals of analytics. Algorithms, too, have critiques based on their ethical implications and the biases embedded in them. Few of these critiques explore how algorithms oversee and track populations and how they perpetuate discrimination in society.

Soft biopolitics

Lippold (2011) talks about how algorithms classify people into categorizations based on their online activities by which they create a "new algorithmic identity" for each one of us. "An identity formation that works through mathematical algorithms to infer categories of identity on otherwise anonymous beings" (2011, p.165). Labelling people with a new identity and categorizing them into certain groups is through a marketing logic of consumption. Lippold compares this with the consumer discrimination that happened earlier using census-laden geographies. She writes, "Mathematical algorithms allowed marketers to make sense out of these data to better understand how to more effectively target services, advertisements, and content" (2011, p.168).

Algorithms can match and classify a new person to a certain category or identity based on the existing data and behaviour models available. It is like a system where a set of identifications and features is already pre-determined, categorising people. As online users continuously interact with the web, their data gradually increases, which then adds up with their previously

collected data. However, if the algorithm sense changes to users' interaction and finds a greater match with another algorithmic identity, the users' algorithmic identity will be shifted. As Lippold (2011, p.168) writes:

> So as more data is received about a certain user's behavior online, new coded computations can be done to change who the user is believed to be and what content that user might desire (the gender of the same user might change from male to female if enough user data, such as the addition of certain web sites that user visited, are presented to statistically identify that user with a different gender).

This whole process of algorithms and data analytics, which is involved in modulating, regulating or influencing the population in a "Deleuzeian approach", resembles the notion of Biopower and Biopolitics proposed by Foucault in a subtle way, for which Lippold (2011, p. 174) terms Soft Biopower and Soft Biopolitics:

> By stepping through the concept of cybernetic categorization we can follow this important move etched by Foucault in his understanding of the social construction of categories. First, ontologies are embedded within a set of power relations. Second, the categories that are bred from those ontologies exercise a profound impact on how we as subjects encounter our world. And third, changes in our categorizing schematics are indebted to this fundamental coding of a culture, from which I find a strong parallel to the technological organization of subject identities through code and algorithm.

This nature of algorithms gains the sole control to define who we are online, and it also detaches us from understanding the meaning of this categorization, which creates our identities (2011, p.178).

Algorithmic Oppression

Algorithms and social media big data are not just controlling or influencing society; they are also reinforcing traditional oppressive social relationships. Data, statistics and algorithms are

always seen to have an objective and neutral nature. However, the algorithms are often biased since they are embedded in the socio-cultural values of the programmers who code them. Even if they tried to make it neutral, there would be some influence of their inner character bias. As argued by Pasquel (Citron, 2008, cited in Pasquale, 2015, p.38):

> Algorithms are not immune from the fundamental problem of discrimination, in which negative and baseless assumptions congeal into prejudice. They are programmed by human beings whose values are embedded into their software.

Also, studies show that programmers are often found lacking to embody good socio-cultural values. Criado Perez (Esmenger, 2010 cited in Criado Perez, 2019, p.102) writes:

> Personality profiles formalised the programmer stereotype nodded to by the computer-science teacher at the Carnegie- Mellon programme: the geeky loner with poor social and hygiene skills. A widely quoted 1967 psychological paper had identified a 'disinterest in people' and a dislike of 'activities involving close personal interaction' as a 'striking characteristic of programmers'. As a result, companies sought these people out, they became the top programmers of their generation, and the psychological profile became a self-fulfilling prophecy.

There is plenty of evidence where this human bias is evident in the outputs given by the algorithms. These new ways that digital decision reinforce oppressive relationships and enact new modes of racial profiling is termed technological redlining by Noble (2018, p.1). This bias is often against women and people of colour. Even though it is not clear how these biased outputs appear on the web because of their “black-boxed information-sorting tools” (2018, p.2). It is proven that the whole responsibility for this lies with the developers who create it, as Noble (2018, p.2) puts it:

> Some of the very people who are developing search algorithms and architecture are willing to promote sexist and racist attitudes openly at work and beyond, while we

> are supposed to believe that these same employees are developing “neutral” or “objective” decision- making tools.

Google’s photo application automatically tagged African Americans as “apes” and “animals, Google Maps search for the word “N*gger” led to the White House during Obama’s presidency, photoshopped pictures of a monkey’s face on Michelle Obama circulated through Google images, and Google search “black girls” yielded search results of black girl pornography are situations where coloured people are experiencing an “algorithmic oppression” as Noble calls it (2018, pp.1-11). Even though these instances have happened years before, they are expected to happen in the future, too, as this sort of oppressive nature is integral to algorithms. “Algorithmic oppression is not just a glitch in the system but, rather, is fundamental to the operating system of the web” (2018, p.10).

Gender Inequality

Another major issue algorithms are causing is that they widen the gender gap. It is evident when corporations hire people with artificial intelligence (AI). AI’s which have been trained using data which lacks socio-cultural values, discriminates women and shows biased behaviour:

> For example, according to Gild's data, frequenting a particular Japanese manga site is a 'solid predictor of strong coding'. Programmers who visit this site therefore receives higher scores. Which all sounds very exciting, but as O'Neil points out, awarding marks for this rings immediate alarm bells for anyone who cares about diversity. Women, who as we have seen do 75% of the world's unpaid care work, may not have the spare leisure time to spend hours chatting about manga online. O'Neil also points out that 'if, like most techdom, that manga site is dominated by males and has a sexist tone, a good number of the women in the industry will probably avoid it… Gild undoubtedly did not intend to create an algorithm that discriminated against women. They were intending to remove human

> biases. But if you aren't aware of how those biases operate, if you aren't collecting data and taking a little time to produce evidence- based processes, you will continue to blindly perpetuate old injustices (Criado Perez, 2019, p.102).

These characteristics embodied by algorithms in hiring areas will continue to contribute towards systemic biases and widen the gender gap. Gild's data provides a clear example of how seemingly innocent criteria might unintentionally bias against women in the case of AI-driven recruitment.

At this stage, one might argue that if a programmer had not intended to develop algorithms in a biased manner, is it still discrimination? The answer would be yes unless he stops ignoring the fact that the algorithms are biased. This sort of action is to be classified as what Eubanks writes about, the concept of “Rational Discrimination”. “Rational discrimination does not require class or racial hatred, or even unconscious bias, to operate. It only requires ignoring bias that already exists” (Gandy, 2009, cited in Eubanks, 2018, p. 155).

Unveiling the “Black-box”

Algorithms were only associated with secrecy and opaqueness until David Sumpter (2018), a mathematician, crafted a matrix, a method where we can understand how algorithms work. He tries to deviate the concept of algorithms from the “black box” imagery coined by Pasquale. “The black box imagery can give us a sense of helplessness, a feeling that we can’t understand what algorithms do with our data. But this feeling can be misleading. We can and should look at what goes on inside algorithms” (2018, p.23).

Sumpter took data of his Facebook friends from their last 15 posts. He then classifies it into 13 categories, assigning each post into categories and recording the number of times each friend has made that particular kind of post. His categories were family/ partner, outdoors, work,

jokes/ memes, product/ advertising, politics/ news, music/sport/ film, animals, friends, local events, thoughts/ reflections, activism and lifestyle (2018, p.24). Furtherly, he performs principal component analysis to analyse the data better by which he concludes that this classification captured genuine similarities and differences.

The Sumpter Perspective

To understand more about the Sumpter matrix and reflect on how it performs the functions of an algorithm, I tried using Sumpter's experiment to categorize my friends and reflect on the outcome. Since most of my friends have become inactive on Facebook, I used Instagram as the platform for my research. Similar to Sumpter, I also took 15 of my friends' recent posts and categorized them into the same categories Sumpter created. Since I did not plan to do principal component analysis, I only took data from 5 of my friends to reduce the data obtained for easy analysis.

Case	Friend ID	Family/Partner	Outdoors	Work	Jokes/Memes	Product/Advertising	Politics/News	Music/Sport/Film	Animals	Friends	Local Events	Thoughts/Reflections	Activism	Lifestyle
1	theghosh													1
2	theghosh									1				
3	theghosh			1										
4	theghosh									1				
5	theghosh													1
6	theghosh	1												
7	theghosh		1											
8	theghosh									1				
9	theghosh									1				
10	theghosh									1				
11	theghosh									1				
12	theghosh		1											
13	theghosh									1				
14	theghosh									1				
15	theghosh									1				
16	cecilpeter													1
17	cecilpeter								1					
18	cecilpeter		1											
19	cecilpeter		1											1
20	cecilpeter		1											
21	cecilpeter													1
22	cecilpeter													1
23	cecilpeter		1											
24	cecilpeter													
25	cecilpeter		1											
26	cecilpeter								1					

Fig1.1. Partial screenshot of the data collection

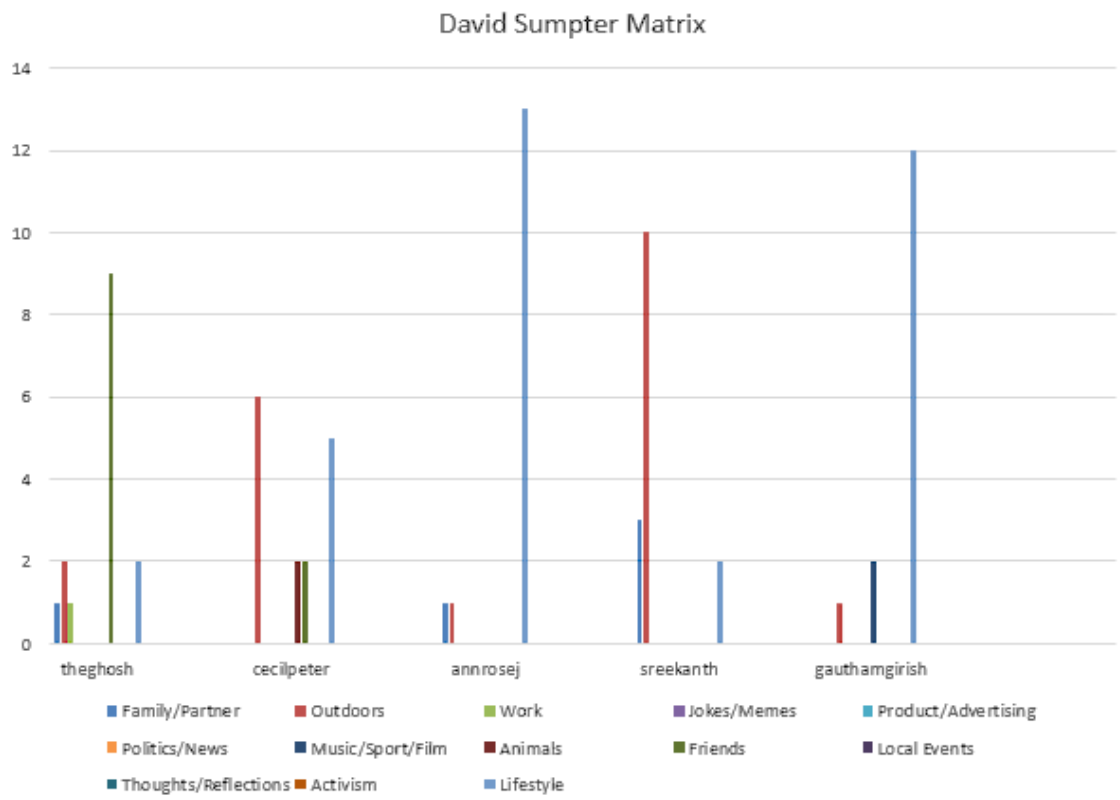

Fig 1.2. Graphical representation of the data obtained.

During the data collection, I first noticed that something was lacking in the categories. I could not classify some posts because I could not find categories for them, such as there was no category for “emotion”. Even though I did classify some of them as thoughts, emotions would be a more apt category. Also, I felt Sumpter created these categories with an image of his friends in mind. Because none of my friends had posted about activism, jokes/memes, product/advertising, and local events, another thought rushed into me was: how can I categorize a post if it was meant to be sarcasm; how will I understand the true intention behind each post?

Method Iteration & Reflections

The next step was to provide an iteration to the method to fit more into our context and include some categories to classify the data I collected. Even though I had an idea of what to do, I felt it almost impossible to write new categories. In this process, I had a feeling that I was

considering my real-life friends in a few columns of categories. It felt like quantifying a person's characteristic/ behaviour is limiting the understanding we have of a person, and to think this is how an algorithm considers a human felt similar to what Lippold (2011) says about how algorithm models determine identity: "Ultimately, it moves the practice of identification into an entirely digital, and thus measurable, plane" (2011, 165).

According to the study conducted by Lippold (2011), algorithms create a "new algorithmic identity". If so, would each friend I took data from have an algorithmic identity? In the dataset, two of my friends are females, and the rest are males; one of my female and male friends are very similar as they have similar interests and characteristics. How can algorithms classify them? How will algorithms identify their gender if their posts fall into the same categories?

Apart from that, from analysing their online behaviours, many differences existed with their real personalities. For instance, my friend with the username "theghosh" is lonely. I understood this during our conversations, but most of her posts were in the "friends" category. When I asked her about this, she agreed that she feels primarily lonely but does not want to show it to others. This observation here relates to the self-discrepancy theory for virtual settings. As Hu et al. argue:

> Moreover, the cyberspace provides people the chance to reconstruct their online identity based on their own discretions. People can hide and/or fabricate their personal information online, such as hiding their educational backgrounds or even faking their gender (2017, p.2).

If people are in self- discrepancy, the algorithms which create new algorithmic identities will further deviate far from their real-life personalities.

Moreover, with references to Noble's (2018) argument, the oppressive nature of algorithms yields racially discriminated results (2018, pp.1-11). My friends also might have encountered or viewed this knowingly or unknowingly. Even though we cannot use the data obtained using

Sumpter's Matrix to determine the above claim, it is understood that until algorithms become neutral (if they will), they will influence shaping the ideals of the young generation.

Conclusion

In conclusion, social media analytics and big data are huge realms that can influence the world by far. It is essential in this scenario to analyse and deeply inspect to correct the fundamentals and ethics. As discussed in this essay, the theories of soft biopolitics, algorithmic oppression, and gender inequality are related to our daily lives on the internet. It is essential to know about these critiques, even for people who use social media and the internet for personal use. Reflecting and examining our daily use of social media can make one realize the logic offered by the critical scholars of this field.

Bibliography

Cheney-Lippold, J. 2011. A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, culture & society*. **28**(6), pp.164–181.

Criado-Perez, C. 2019. *Invisible women: Data Bias in a World Designed for Men*. Abrams Press.

Eubanks, V. 2019. *Automating inequality: How high-tech tools profile, police, and punish*. New York, NY, USA: St Martin's Press.

Ghosh, V S. 2023. _theghosh. [Instagram]. [Accessed 9 December 2023]. Available from: https://www.instagram.com/_theghosh/?utm_source=ig_web_button_share_sheet&igshid=OGQ5ZDc2ODk2ZA==

Girish, G. 2023. gautham._.girish [Instagram]. [Accessed 9 December 2023]. Available from: https://www.instagram.com/gautham._.girish/?utm_source=ig_web_button_share_sheet&igshid=OGQ5ZDc2ODk2ZA==

Hearn, A. 2020. Verified: Self-presentation, identity management, and selfhood in the age of big data *In*: *Self-(re)presentation now*. Routledge, pp.9–24.

Hu, C., Kumar, S., Huang, J. and Ratnavelu, K. 2017. Disinhibition of negative true self for identity reconstructions in cyberspace: Advancing self-discrepancy theory for virtual setting. *PLoS ONE*. **12**(4).

Jayram, S. 2023. Storiesbysreekanth. [Instagram]. [Accessed 9 December 2023]. Available from: https://www.instagram.com/storiesbysreekanth/?utm_source=ig_web_button_share_sheet&igshid=OGQ5ZDc2ODk2ZA==

Joseph, A R. 2023. _ann.rose.j_. [Instagram]. [Accessed 9 December 2023]. Available from: https://www.instagram.com/_ann.rose.j_/?utm_source=ig_web_button_share_sheet&igshid=OGQ5ZDc2ODk2ZA==

Noble, S.U. 2018. *Algorithms of oppression: How search engines reinforce racism*. New York, NY, USA: New York University Press.

Pasquale, F. 2016. *The black box society: The secret algorithms that control money and information*. London: Harvard University Press.

Peter, C. 2023. cecilpeter_ [Instagram]. [Accessed 9 December 2023]. Available from: https://www.instagram.com/cecilpeter_/?utm_source=ig_web_button_share_sheet&igshid=OGQ5ZDc2ODk2ZA==

Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C. 2018. Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International journal of information management*. **39**, pp.156–168.

Sumpter, D. 2018. *Outnumbered: From Facebook and Google to fake news and filter-bubbles - the algorithms that control our lives*. London: Bloomsbury Sigma.

Zachlod, C., Samuel, O., Ochsner, A. and Werthmüller, S. 2022. Analytics of social media data – State of characteristics and application. *Journal of business research*. **144**, pp.1064–1076.