THE IMPACT OF DATA SCIENCE IN DIGITAL JOURNALISM

Noreen Aleem[1]
Ali Raza Qureshi[2]
Fazli Hussain[3]

Abstrac*t: Social media is a robust platform and it has a lot of beneficial implications. In an era of technological advancements, mankind is surrounded by electronic means of communication and networking or communication has become easier than ever. It is very easy for a person sitting in an Asian city to interact with someone in Northern America. Likewise, people are now more informed about daily events and news taking place all over the world.These social networks have become a substantial tool for unstructured data from different domains which include business, government organizations and individuals. Many data analytical firms are extracting this data in a systemic pattern which refers to the term "Data Science". However, the purpose of this aligned data is to target defined and scattered audience to control their mindsets or picture in the head. The prime focus of this study is to empericaly calculate the impact of data science in Digital Journalism as well as how alarming this can be in terms of authenticity, reliability or being stated as a word of mouth.The study depicts how technology-dependent man has become and the way technology has deterred the process of thinking and evaluation*

Keyword: *Data Science, Big Data, Social Media, Digital Journalism, Analytics*

[1]Author is Assistant Professor in Department of Media & Communication Studies at Sindh Madresstul Islam University, Karachi

[2]Author is Education Quality Auditor at Lahore Garrison University, dy.director.qec@lgu.edu.pk

[3]Author is Assistant Professor in Department of Media Sciences at IlMA University Karachi

Introduction

The 20th Century was the turning point in the world of science due to the advent of computer. It was followed by the 21st Century often referred to as the era of Information technology due to the advancement in large scale networking and Internet. The unison of computers and the Internet gave birth to a new and unique concept, "social media" which is the biggest source of data mining & creating insights.

But it's not a commonly apprehended phenomena that Buzzword, Data Science is a process to create insight which in a true sense is based on extracting, analyzing, visualizing, managing and storing data that help organizations and companies' as a mighty data-driven decisions (Chiara, Ramon & Jeanson, 2019). We have a number of platforms that knowing or unknowingly collect and provide data of more significance for all the convetional, electronic and digital medias.

Social media is relatively a new medium of communication as compared to mainstream print and electronic media which distinguishes the informer from the ones informed (Hutchison, at el. 2016) It is very clear that this is the unique forum, where people interact allowing equal participation from both ends and where there is no proper differentiation between the informer

and the informed. In other words, social media is a true reflection of Mashal Mc.Luhan's Global Village. In this sense, networking and interaction have always proved to be beneficial and people with strong contacts are often deemed successful in not only the professional domain but also in other sheres of life. Moreover, social media has enabled many political figures to disseminate their narrative as well as businesses to market their services and products. Similarly online learning through discussion forums and communities is also a noteworthy aspect of social media (Mayfield, 2008).

Respectively, the benefits of social media are innumerable and that is why professionals in walks of life tend to attain social media visibility. The benefits of social media owe their entity to the fact that almost 67% of internet users in the USA are registered on the popular social networking websites. (Duggan & Brenner, 2013). Here are some popular socoal media networking websites that are in services to provide a empowered platform for data science.

Types of Social media: A generally accepted idea of social media is social networking websites such as the Facebook, Twitter, LinkedIn, Instagram, YouTube, Google+, etc, but, in the broader sense, social media goes beyond that as it's also inclusive of blogging, wikis, podcasts, forums, content communities as well as micro-blogging.

Blogs: Blogs are online journals characterized by their tone and the subject they cover. A blog could be generalyl, ranging from stories based on different topics or that revolves around only a particular subject such as science and technology, art and literature or blogs related to simple medical remedies. But in reality blogs contain links to other informative websites which can add further weight to the point made in it. Although comments are welcomed in blogs by bloggers after approval but stil this comment section serves as a discussion or debate forum.

Resultantly, social media's biggest advantage is connectivity and as blogs are a component of social media, one can connect to blogs by subscribing to it, also. Email notifications are sent to all subscribers of the blog if a new entry is published by the blogger.

Social Networking Websites: A large fraction of public accesses social media through the platform of social networking websites. Popular social networking websites are Facebook, that allows to share status updates, photos, videos, links, check-ins and also offers live chatting, Twitter, allows sharing responses within a 140 characters limit, shows what is

trending all over the world or in a particular region, also allows sharing of links, photos and videos, Instagram allows users to share photos, videos and tells what is trending all over the world or in a particular region, LinkedIn helps users connect to professional counterparts and enables building a professional network.

Podcasts: Podcasts are also noteworthy type of social media. It refers to audios that are uploaded on the World Wide Web to facilitate users or simply for entertainment purposes. People can subscribe to regular podcasts that build an audience. Podcasts have blurred the conventional lines between media territories and also given control to the audience to consume a media presentation of their own choice and at their own designated time.

Wikis: Wikis are perhaps the most popular form of social media among literary circles. People can make their contributions or amendments to the content on wikis. Wikis are not just restricted to reference articles for term papers but also cater for citizen journalism during natural catastrophes or political unrest in a region.

Forums: Forums are perhaps the oldest of all types of social media. A forum is a discussion platform where internet users can initiate a thread based on a specific subject and participate in the on-going online discussion (Stone, 2014).. It is different from blogs due to the modus operandi of both. Blogs are owned by one person and discussion leads are sketched by the blog owner whereas forums allow all participants to initiate any thread which is worthy of a discussion or debate.

Content Communities: Content communities tend to work similarly as that of social networking websites. What distinguishes content communities from social networking websites is that they revolve around only a particular type of content (Hussain, 2020). A notable example is YouTube, which enables users to share videos. The videos can vary from informative ones to those solely for entertainment purposes but the mode is similar.

In addition to posting content in a content community, users can also connect with other users and build a network by subscribing to their content.

Each one of the types is engaging people in terms of content which include text, images, videos, interactive applications and creating two-way communication. Social media's growing popularity has made it an integral part of internet users which leaves a notable impact on human minds especially on that of adolescents (Hussain, 2018). The impact of data science through social media and how it affects Digital Journalism is the main subject of this research paper and will be discussed throughout. This study covers the advantages and disadvantages along with the connection of data sciences in digital journalism. The conclusion depicts how technology-dependent man has become and the way technology has deterred the process of thinking and evaluation.

Pro &Con of Data Science

The field of Data Science is monstrous and an ongoing process through re-analyzing and evaluating data on user's engagement. The following are the core advantages and disadvantages of data science.

Advantages

Highly Paid Career: As an economically driven subject, data scientists are one of the highest-paid employees. According to research conducted by Glassdoor, data scientists make an average of $116,100 per year.

Affluence Opportunity: A certain skill set is needed to become a complete data scientist, who makes it less saturated in different sectors. The field of Data Science is high popular however low in supply.

Refined Data: Based on user analysis through data science with advance computing, data is improved and enriched for companies and organizations in better reach and insights target marketing.

Social Change & Enhancements: Artificial intelligence is used with mined data by different industries to create the best user experience

which includes user navigation & human behavior.

Automate Redundant Tasks: AIl Machines are trained with historical data to perform repetitive tasks which were done by human beings before.

Disadvantages

Unexpected Results: A Data Scientist analyzes the data and makes careful predictions to facilitate the decision-making process. But many times, the data provided is arbitrary and does not yield expected results. This can also fail due to weak management and poor utilization of resources.

Data Privacy: For many industries, data is the primary tool, which helps companies make data-driven decisions. However, the data advance in the process may breach the privacy of customers, viewers or users. The personal data of clients are visible to the parent company and may at times cause data leaks due to lapse in security. The ethical issues regarding the preservation of data-privacy and its usage have been a concern for many industries.

Blurry Term: It is very hard to define the exact term of Data Science due to its vast spreading in different cores. The method of use to generate insight by using data in any particular sector can conceive as Data Science.

New Approach: As the world is rapidly evolving through technology advancements day by day, companies and organizations need to update their tools and methods accordingly along with new and accurate driven data for a better target hit (David, 2003)..

Possible Failure: Some companies and organizations use it as a play tool in their application and structure, ignoring real-time data can create huge and destructive problem in their businesses and surroundings as well.

Data Science & Digital Journalism

When we state Human beings as social animals, we are pointing to the curious part "to look in another window or to stay communicated". Today in the modern era, mass communication is turned out to be a wide and more effective medium through internet and technology with rapid growth. The world seems to be like a small village with lots of information about daily events, updates, and connectivity with each other.

So the journalism shapes up with the new trend & advancement which is known as Digital Journalism or Online Journalism. Nowadays we are just a click away by feeding our mind about what's happening in the surroundings (Hussain, 2019). Despite the fact, after social media, it's very easy to get insights & trending of the viewers as compared to traditional media. The complete report of the viewers including their emotional and psychological behaviors or emotions is now very easy to extract as data. This. Secondly, different business entities and political bodies are setting their agendas in creating word of mouth among people .

Literature Review

Alzahrani (2016) stated about the individuals, inspired by the content which include data, news & thoughts of different clients on social networks (Facebook, Twitter, Linkedin, Google Plus) growing rapidly. These social networks are providing them a huge amount of data & information from different sources at various scales, which helps them get insights for their businesses to plan to action marketing strategies.

Schroeder (2014) highlights the ethical & social issues in Big Data extraction from Facebook. Emotional Behaviors to change the human-made environment and its implication to improve web traffic as per their commercial requirements in a competitive market. The data draws a complete picture of online and offline activity demography, once analyzed through different tools and rationality.

Nehali (2018) in the context of The Hamilton Spectato – a local newsroom in Canada states the use of traffic based matrix for a more informed & engaged audience. In which they monitor and track the page viewing and click ratio through the user experience and generate content accordingly. This helps them creating

enhanced and provoking stories with the placement on the platform mostly viewed.

Stone (2014) stated about the growing amount of user's data extracted from various devices like Computers and Mobile Communication to create its application for editorial and business uses by implementing strategy and techniques as well as Audience interaction based on their emotional and psychological reaction for the particular targeted content

Significance of the Study

People today, are more and more engaged in the world of the internet, all their infotainment, entertainment or source of knowledge is just a click away. They are unaware of the fact that every information, navigation they are creating is recorded in a systematic data pattern to use for further purposes which can be dangerous for them if uses negatively

Theoretical Framework

Keeping in mind, Cambridge Analytical and Facebook influence in the US's Elections 2018 in which Cambridge harvested data of millions of people from their Facebook profile without their consent which result in an outreach output in Trump's victory. This, later on, encourages more data mining companies and corporate entities to hire data scientist's and use relevant tools

Research Methodology

To identify the exact and accurate impact of Data Science, the quantitative method will be used. The selected group for sampling will be youngsters between 18 to 35 years old. The sampling tool will be google forms with the close-ended questionnaire for accuracy.

The objectives of the Study

To find out internet users' average time in a day. User's social networking ratio and other surfing analytics.

Most influential social network in terms of data searching.

Current affairs & news content ratio and sources on all devices especially the phone.
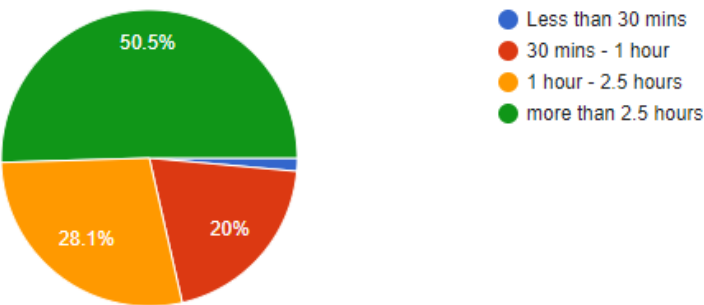
Newsfeeds or posts as a word of mouth.

Findings and Discussion

Table 1

On average how much time do you spend surfing the internet in a day?	
More than 2.5 hours	50%
2 hours	28%
30 minutes	20%
Less than 30 minutes	2%
N = 200	

On the average, how much time do you spend surfing the internet in a day?

210 responses

- ● Less than 30 mins
- ● 30 mins - 1 hour
- ● 1 hour - 2.5 hours
- ● more than 2.5 hours

Interpretation: Table 1 clearly illustrates that the majority of 50% of respondents spend more than 2.5 hours of a day surfing on the internet. The other lesser majority of 28% and 20% of the respondents spend 2 hours and 30 minutes surfing the internet respectively. The remaining 2% of the minority spend less than 30 minutes on the internet.

Table 2

What part of this time do you spend on social networking websites?	
25% of the time	32%
50% of the time	35%
75% of the time	27%
100% of the time	6%
N = 200	

What part of this time, do you spend on social networking websites?

209 responses

Legend:
- 25%
- 50%
- 75%
- 100%

27.8%
35.9%
32.1%
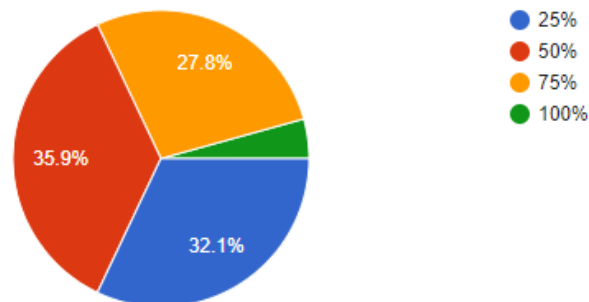
Interpretation: The pie chart in table 2 states that the majority of 36% of the respondents opined that they spend 50% of their internet time while surfing the social networking websites. A large number of 32 of the respondents tend to spend 25% of their internet time on social websites while the other 27% of the respondents said that they devote 75% of their internet time to social networking websites.

Table 3

On which social network you spend most of your time	
Facebook	47%
Twitter	4%
linkedIn	2%
other	46%
N = 200	

## Which social network, you spend most of your time on

209 responses
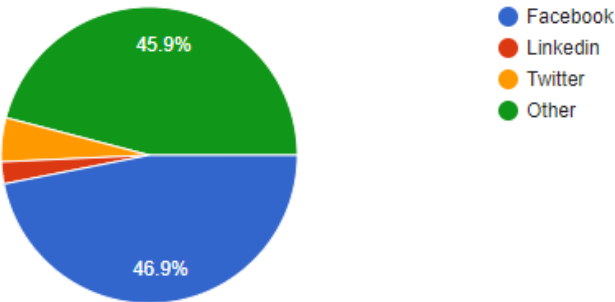
Interpretation: The pie chart in fig 3 exemplifies that a majority of 47% of the respondents use prefer to use Facebook for their social networking purposes while the minorities of 4% and 2% of users use Twitter and Linked In respectively. The remaining 46% of respondents mentioned that they use other websites for social networking while surfing the internet.

Table 4

On average, how much of the current affairs and news related content do you find on social media?	
25% of the content	47%
50% of the content	32%
75% of the content	17%
100% of the content	3%
N= 200	

On the average, how much content you find related to Current Affairs & News on that particular social network you use?

209 responses
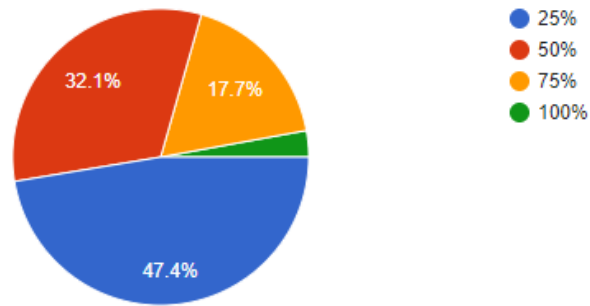
Interpretation: The numbers in table 4 expresses that a majority of 47% of the respondents were 25% successful to find out the current affairs related content on social media while a large number of 32% of the respondents found 50% of the material related to news and current affairs on their social networking website. The remaining 17% and 3% of the respondents found 75% and 100% news and current affairs data on social media.

Table 5

On average how often do you scroll these feeds on your phone?	
Always	19%
Mostly	47%
Sometimes	31%
Never	3%
N = 200	

On the average, how often you scroll these feeds through your phone?

209 responses

- ● Always
- ● Mostly
- ● Sometimes
- ● Never

31.1%

47.4%

19.1%
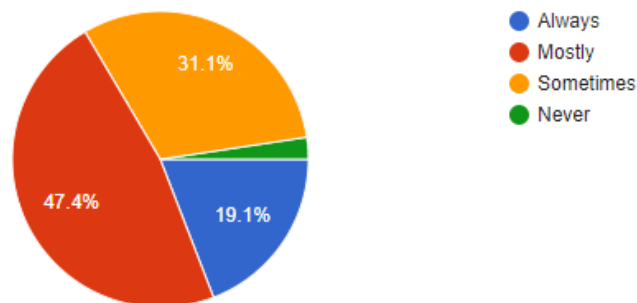
Interpretation: The table 5 indicates that a majority of 47% of the responders mostly scroll through the news feed on their phone while a large number of 31% responders sometimes take a look at the newsfeed on their phone. A minority of 19% of respondents always view the newsfeed on their mobile phones.

Table 6

What is the source of the news content in your feed?	
direct news organizations' page	45%
Shared by friends	36%
Other pages	20%
N = 200	

From which source, you often get that Current Affairs & News content in your Feed?

209 responses

- ● Direct News Sources Pages
- ● Shared by a Friend / Colleague
- ● Other Pages

35.9%

20.6%

43.5%
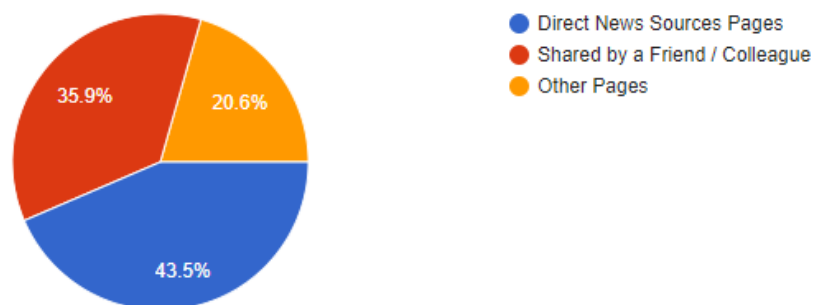
Interpretation: The figures in table 6 illustrates that the 45% majority of the respondents obtain the current affairs and news content via direct news organizations' page while a lesser majority of 36% have got their friends who share the news items on their timeline. The remaining 20% have obtained the data from other pages.

Table 7

Do you think that you or your friends become the word of mouth for that posts in surrounding	
Yes	20%
No	21%
Sometimes	59%
N = 200	

Do you think you or your known, being a word of mouth for that posts in surrounding?

209 responses

- Yes
- No
- Sometimes

59.3%

20.6%

20.1%
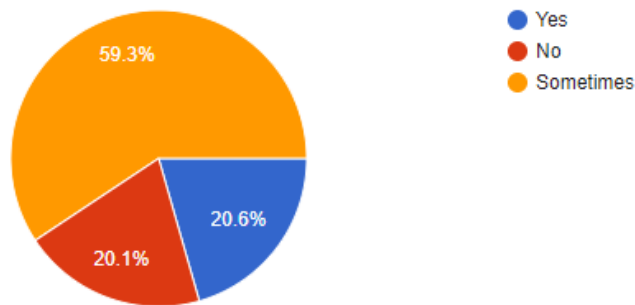
Interpretation: the pie chart illustration in figure 7 indicates that a majority of 59% responders think that sometimes their friends or themselves become a word of mouth for the news posts in the surrounding. While 21% of the respondents think they or their friends do not become the word of mouth for that post in the surroundings. On the other hand, the remaining 20% of the respondents think that they or their friends sometimes spread the news around.

Conclusion

As a result of the survey, people are engaged in the current affairs and news contents with Facebook and other platforms mostly and considered as heavy users which are bound or weave on the web of the internet. They are unaware of the knowledge that how data is breaching their other shared content on the internet and showing them what they are supposed to look for changing their picture in the head/mindset.

Mostly users are scrolling feeds with their phones which is a more dangerous object in collecting data as we are giving rights to use our information whenever we install any application. They share information as a word of mouth in their surroundings.

This disaster of information could collapse mind thinking-process among youngsters and could create rapid negative thinking among them causing harsh behaviors & emotional dilemmas although hybrid newsrooms can be a controlled solution for that when it comes to the authenticity of information and positive data-driven tool.

Limitation

Defining the clear term of data science in the required context.

Reaching the scattered audience for sampling which looped through social media and networks.

Easy questions and terms for technology slob.

Quantify breaching through google forms for perfect results.

Recommendation

PEMRA (Pakistan Electronic Media Regulatory Authority) should issue an advisory for all private news channels for building hybrid newsrooms for authentic news & current updates.

Gatekeeper or blockers should be assigned on social platforms as an artificial intelligence before sharing the content with the perspective in the head.

Improved privacy policy and terms of use will be implemented by PEMRA (Pakistan Electronic Media Regulatory Authority) for both users and news bodies.

Reference

Alzahrani, H. (2016). Social Media Analytics using Data Mining. *Global Journals of Computer Science and Technology:* C Software & Data Engineering.

Carlson, David (2003). "History of Online Journalism". In Kawamoto, Kevin (ed.). *Digital Journalism: Emerging Media and the Changing Horizons of Journalism. Lanham, Md*.: Rowman & Littlefield.

Chiara V, Ramon Portugal F, Jeanson R (2019) Social intolerance is a consequence, not a cause, of dispersal in spiders. *PLoS Biol* 17(7): e3000319. https://doi.org/10.1371/journal.pbio.3000319

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. Educational Technology & Society."The State of the News Media" Pew Project for Excellence in Journalism. Stateofthemedia.org, *Harvard Business Review*. 25 January 2018, "Are You Setting Your Data Scientists Up to Fail?".. Retrieved 26 May 2018.

Hussain, F., & Kamal, A.- e. (2020). THREATS TO JOURNALISTS IN SINDH: EVENTS AND PERCEPTIONS (2000-2017). *Journal of Social Sciences and Humanities*, *57*(2), 193-209. https://doi.org/10.46568/jssh.v57i2.6

Hussain, Fazli. (2019). Press freedom and Threats to Journalists in Pakistan. (unpublished PhD thesis) Federal Urdu University of Arts, Science and Technologies Karachi.

Hussain,F.(2019). Gender discrimination in media as work place a study media environment in sindh. *Pakistan Journal of Gender Studies*, DO - 10.46568/pjgs.v18i1.23

Hutchison, M. L., Schiano, S., & Whitten-Woodring, J. (2016). When the fourth estate becomes a fifth column: The effect of media freedom and social intolerance on civil conflict. *The International Journal of Press/Politics*, *21*(2), 165-187. https://doi.org/10.1177/1940161216632362

Jeff Leek (12 December 2013). "The keyword in "Data Science" is not Data, it is Science"

Mayfield, A. (2008, January 1). What is Social Media? An e-book by Antony Mayfield. Retrieved from http://www.iCrossing.com/ebooks

Neheli, N. B. (2018). News by Numbers - The evolution of analytics in journalism. *Taylor Francis Online*.

Schroeder, R. (2014). Big Data and the brave new world of social media research. *SAGE Journals.*

Stone, M. L. (2014). Big Data for Media. University of Oxford - *Reuters Institute for the study of Journalism.*