International Journal of Internet, Broadcasting and Communication Vol.13 No.2 218-223 (2021)
http://dx.doi.org/10.7236/IJIBC.2021.13.2.218

IJIBC 21-2-27

A Study on Veracity of Raw Data based on Value Creation -Focused on YouTube Monetization

[1]Seoyeon CHOI , [2] Seung-Jung SHIN

[1]Doctor Candidate, Department of IT Convergence, Hansei University, Korea
1okpro@naver.com
[2] Professor, Department of ICT, Hansei University, Korea
expersin@gmail.com(corresponding author)

Abstract

The five elements of big data are said to be Volume, Variety, Velocity, Veracity, and Value. Among them, data lacking the Veracity of the data or fake data not only makes an error in decision making, but also hinders the creation of value. This study analyzed YouTube's revenue structure to focus the effect of data integrity on data valuation among these five factors. YouTube is one of the OTT service platforms, and due to COVID-19 in 2020, YouTube creators have emerged as a new profession. Among the revenue-generating models provided by YouTube, the process of generating advertising revenue based on click-based playback was analyzed. And, analyzed the process of subtracting the profits generated from invalid activities that not the clicks due to viewers' pure interests, then paying the final revenue. The invalid activity in YouTube's revenue structure is Raw Data, not pure viewing activity of viewers, and it was confirmed a direct impact on revenue generation. Through the analysis of this process, the new Data Value Chain was proposed.

Keywords: *Value of Data, Veracity of Data, Data Value Chain, Big Data, Raw Data*

1. Introduction

In January 2021, the chatbot service “Iruda” stopped after three weeks of launch. “Iruda” is a 100% AI (artificial intelligence) chatbot that was ambitiously launched in December 2020 but raised the problem of AI that learned data of prejudice by making a statement mixed with discrimination such as the disabled, homosexuality, and women [1]. This is the similar case of the 'Tay', an AI chatbot created by Microsoft. It was launched in 2016 about five years ago, but it was shut down after 16 hours like “Iruda”. Damage caused by false news and information is also increasing rapidly due to fake data. Knowledge and information are likely to be distorted in the process of production and distribution, and "Distortion Phenomenon" is taking place throughout society [2]. This has a great implication as an opportunity to consider not only the amount of data

Manuscript Received: April. 11, 2021 / Revised: April. 14, 2021 / Accepted: April 16, 2021
Corresponding Author: expersin@gmail.com
Tel: +82-31-4505274, Fax: +82-31-450-5137
Professor, Department of ICT, Hansei University, Korea

to be learned by the technology of artificial intelligence, but also the quality, that is, the veracity of the raw data. The technology of rapidly collecting and storing large amounts of various data has led to the development of artificial intelligence. However, data lacking veracity leads to contradictory decisions in value creation. For this reason, the importance of qualitative data rather than quantitative data is emerging. Compared to studies on the accuracy and reliability of the data collection and storage process, the researched in terms of production of raw data is insufficient. In this study, Data Value Chain was proposed by proving the truthfulness of the time when raw data is produced and the direct connection to Value Creation. Based on the YouTube advertising revenue model, we analyzed the actual application process and valuation process of unveracity Raw Data. However, the analysis of how to adjust the final revenue that the creators receive is covered in this study, but details on how to calculate revenue of Youtube is excluded.

2. Veracity of Data

2.1. Raw Data

Data is defined as having undergone a process of processing analysis and valuation so that it can be used for the purpose of creating added value [3]. The Veracity of Data refers to the level of accuracy in the collected data [4]. The data collection is the first step in the Data Value Chain, refers to collecting among the data that has already been produced that aligned with objectives. The dictionary meaning of raw data is defined as “Data generated or produced by an individual or organizations" and is also including generated by devices as smart phones, IOT, and variety of sensors. The comparison of Data and Raw Data is as shown in Table 1.

Table 1. Comparison of Data and Raw Data

Data	What has gone through the value processing analysis to be used for the purpose of creating added value.
Raw Data	Data generated or produced by individuals, organizations or other devices before collection, storage, process and analysis of data.

All data begins from production before collection, and Raw Data is included in the broader data. Big Data can bring new value creation effects [5].

2.2. Big Data 5V

The Veracity of data is also expressed in terms of accuracy and reliability and includes not only the quality of the data itself, but also the reliability of the data source, type and processing [6]. The quality of the data itself and the Authenticity in terms of data production are defined as the veracity of the data as shown in figure 1.

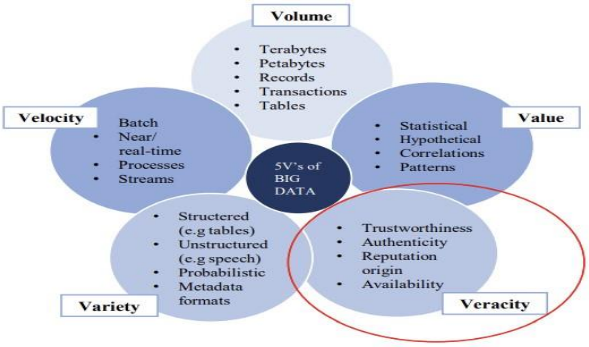

Figure 1. The 5 V’s of Big Data [Reorganization]

2.2.1. Volume

Big data refers to data on a massive scale of more than terabytes. It is worthless because a single data is unable to find commonality or pattern. However, the larger the scale, the easier it is to find common patterns and the reliability of the results increases **[7]**.

2.2.2. Variety

Data is classified in a wide variety of forms and structures. It varies from structured, semi-structured and unstructured data to heterogeneous data. In the unstructured data-oriented is a variety of images, pictures, comments, reactions, etc. that are produced by IT devices such as smartphones, GPS technology on social networks **[7]**.

2.2.3. Velocity

It is to be able to utilize data as fast as possible and as close to real-time. Data that has passed the required time is likely to become useless data **[7]**.

2.2.4. Veracity

If the truthfulness of the data is not given, no matter how quickly various data is collected, stored, and analyzed, it can lead to an error in decision making **[7]**.

2.2.5. Value

Data should be collected, stored, and analyzed to create value through utilization [3]. Regardless of the final goal, data collected or not meeting the goal is not related to value creation.**[7]**.

3. The revenue model of Youtube

Due to Corona 19, YouTube has gained a favorable response from all generations, including the “Stay at home”, MZ generation, and seniors in their 50s and above, and YouTube has established itself as one of the best OTT. According to the “2020 Elementary and Middle School Career Status Survey” by the Ministry of Education and the Korea Vocational Competency Development Institute, Youtube creator ranked fourth as the career path of elementary school students [8]. Youtube channel is used not only as an individual creator, but also as a corporate advertisement and marketing, and has established itself as a tool for creating revenue for individuals and companies. YouTube shall thoroughly analyze the revenue generated by the invalid activities and deduct the amount from the final revenue. We selected a revenue model according to the number of video views among YouTube's revenue models and analyze the process of generating AdSense advertising revenue among them.

4. How to monetize YouTube: AdSense ads

4.1. Revenue model based on number of Video Views.

AdSense ad revenue generation is a method in which Google (YouTube) contracts with advertiser to place advertisements on Creator's video. The creator firstly becomes a YouTube partner and uploads the video. YouTube exposes advertisements from advertisers that have already been contracted to videos in various ways such as pre-roll, mid-roll, and post-roll. Both advertisement placement and revenue distribution are performed through Google AdSense as shown in figure 2 [9].

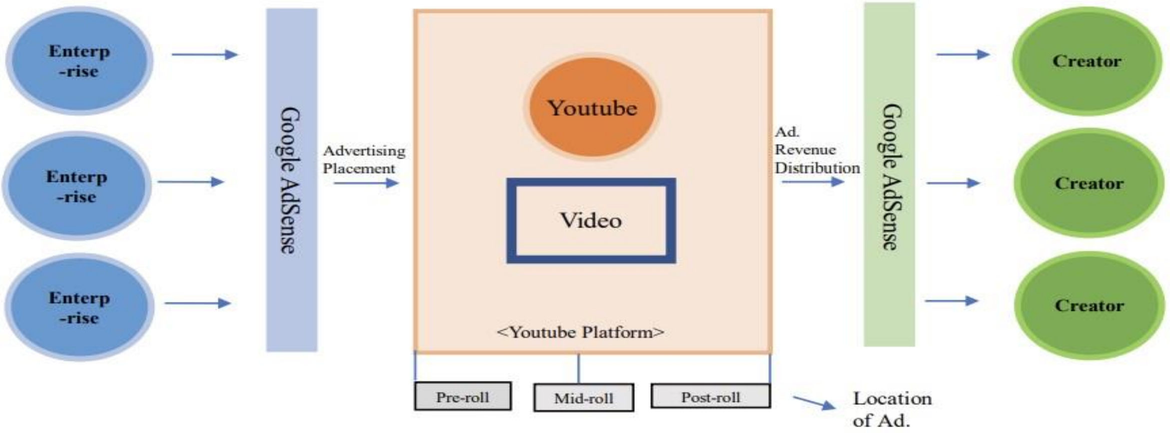

Figure 2. YouTube AdSense ad Revenue Model Conceptual Diagram [Reorganization]

4.2. Definition the amount of advertising revenue

Ad revenue refers to the revenue generated by clicking on an advertisement placed on a video according to the viewer's pure interest. In other words, the veracity of the viewing is judged by the viewer's watching and clicking on the advertisement. The clicks generate revenue for the creator, and at the same time, the advertiser incurs advertising costs.

5. YouTuber revenue adjustment

5.1. What is invalid activity?

Also known as invalid traffic, these are clicks or impressions that are intentionally lacking in veracity to increase advertisers' costs or publishers' income. Google stipulates that clicks from YouTube ads must be clicks on ads that are driven by viewers' interest. In other words, all actions that manipulate or artificially generate clicks or exposures of advertisements are defined as invalid activities in accordance with AdSense's policy. The invalidation activities specified by Google are considered data lacking truthfulness and have a direct impact on revenue generation and are strictly prohibited. Invalid traffic designated by Google specifies both intentional fraudulent traffic and unintentional clicks [10].

5.2. Correction of invalid activity

Invalid traffic, or invalid activity, refers to click activity that lacks veracity. If the level is high, Google may suspend or disable the publisher's account. It also protects advertisers and users by limiting or stopping ad serving when the quality of traffic cannot be verified. Youtube has a final revenue adjustment policy at the end of each month for the revenue generated from inaccurate data. Currently, YouTube is not sharing details about the algorithm for detecting invalid activity [10].

5.3. Conceptual diagram of revenue payment after adjustment

The raw data in YouTube advertising revenue model by play-based advertising are clicks. YouTube adjusts between expected and final earnings by determining the authenticity of the click, when the creator creates and uploads a video, YouTube and the advertiser post a pre-contracted advertisement, and the viewer clicks the advertisement while watching the video [10]. Expected revenue are not paid as they are, and revenue arising from invalid traffic at the end of each month are deducted and pays as shown in Figure 3.

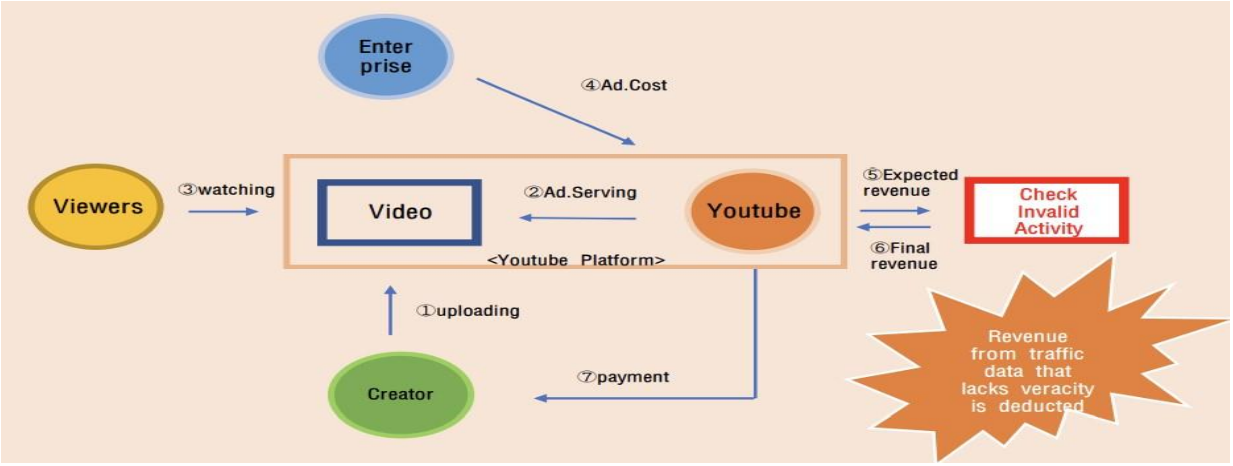

Figure 3. Conceptual Diagram of Revenue Payment

6. How to monetize YouTube

The process of data collection, storage, processing, analysis, and utilization is called the Data Value Chain [11]. Before data collection and storage, the stage that must be considered is the production stage. Since the integrity of the raw data produced has a direct impact on the value creation, the value chain of data must begin with the production of data. As the first step in the data value chain, the data collection, is a step that collects data from various sources, either directly or indirectly. It was confirmed that the integrity of raw data directly affects the change in data value. The data value chain is completed only when data production is added to the stage before the data collection stage as shown in figure 4.

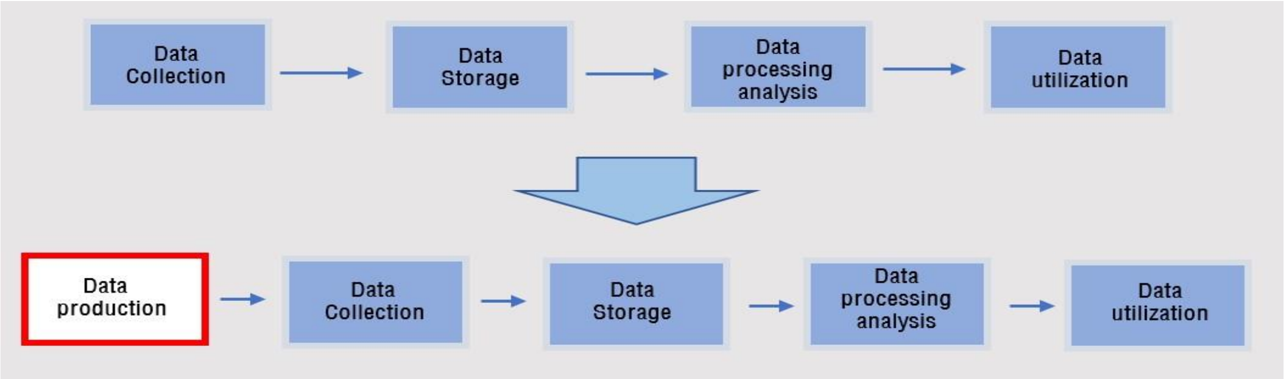

Figure 4. New Data Value Chain

7. Conclusion

By analyzing the process of creating and paying YouTube's revenue, it was confirmed that the revenue generated from raw data that lacked truth was reduced and the final revenue was deducted. The greater the veracity of the raw data, the greater the value creation, and the less the lack of the truth, the less the value creation, so the relationship between the raw data and the value creation is in direct proportion was confirmed. Therefore, a new Data Value Chain was proposed by adding data production from the existing data value chain of data collection, storage, processing, analysis, and utilization. This study analyzed the value chain that the veracity of raw data has on value creation through the YouTube revenue payment process, but a detailed study on the contribution to value creation of raw data in the future should be further conducted. It will be a better study if the process of the value chain of raw data is traced by incorporating blockchain technology in the future research.

References

[1] News Post, [Artificial Intelligence]② Controversy over "AI chatbot Iruda" created by human hatred and prejudice, http://www.newspost.kr/news/articleView.html?idxno=92145

[2] Seung-Jung Shin, “A Study on Data Reliability Used on the Internet”, Korea Information Processing Society, pp. 9 57 - 958, 2009, ISSN 2005-0011, Nov 2009

[3] Seoeyon CHOI, “A Study on Expansion Proposal of Data Dividend Qualification Based on the Contribution of Platform Workers.”, The Journal of The Institute of Internet, Broadcasting and Communication (JIBC), Vol. 12. No. 6. pp 267-272. Apr 2021.
DOI: https://doi.org/10.7236/JIIBC.2021.21.2.187.

[4] Hye-Suk Kim, “Analysis of Popular YouTube Video Content using Data Mining”, Journal of Digital Contents Society, Vol. 21, No. 4, pp. 673-681. 2020.
DOI: https://doi.org/10.9728/dcs.2020.21.4.673

[5] Seung-Jung Shin, “SNS using Big Data Utilization Research.”, The Journal of The Institute of Internet, Broadcasting and Communication (JIBC), Vol. 12. No.6. pp 267-272. Dec 2012.
DOI: http://dx.doi.org/10.7236/JIWIT.2012.12.6.267

[6] GetCheck, “Veracity: The Most Important “V” of Big Data”, USA, 2019 *https://www.gutcheckit.com/blog/veracity-big-data-v*

[7] Okan Şen, “Big data usage in electrical distribution systems: A review”, International Journal of Applied Business and Management Studies, Vol. 4, No.2; 2019 ISSN 2548-0448, OCT 2019

[8] Department of Education and the Korea Vocational Competency Development Institute, “2020 Elementary and Secondary Career Education Status Survey Results”, 2021

[9] Jae-uk OH, “A study on the revenue model of art creators on YouTube - Focused on singer-turned-YouTube music creators”, Master Thesis. The Catholic University of Korea, Seoul, South Korea 2021

[10] AdSense Google Support, *https://support.google.com/adsense/?hl=en#topic=3373519*

[11] LI, Wendy C.Y. NIREI, Makoto, YAMANA, Kazufumi, “Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy”, RIETI Discussion Paper Series 19-E-022, *JAPAN* 2019