Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 431-440
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Analisys of Trends And Patterns of Hololive ID Vtubers on Content Variation for Monetization Optimization Using Multiple Linear Regression Algorithm

Ananda Aulia[1*], Muhammad Iqbal[2], Khairul[3]
[1,2,3] Master of Information Technology, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: anandaaulia30@gmail.com[1], muhammadiqbal@dosen.pancabudi.ac.id[2], khairul@dosen.pancabudi.ac.id[3]

Abstract

YouTube, as a new form of mass media in the advancement of increasingly sophisticated technology, caters to a diverse audience seeking daily information. Many YouTube users leverage this platform as a medium for creativity and income generation, such as through video blogs, educational videos, gaming videos, and various other content. Not all YouTubers can express themselves directly or through livestreaming; some use virtual 2D or 3D characters created with computer software, utilizing technologies such as Face Tracking and Hand Tracking. Hololive Production is a Japanese VTuber agency owned by the Japanese technology company Cover Corporation. Hololive Indonesia (Hololive ID) was established to produce Indonesian-language VTuber content and has garnered significant attention from fans in Indonesia and worldwide. VTuber content primarily revolves around video game topics, indicating a strong interest among VTuber fans in gaming-related content. Moreover, the analysis results show that multiplayer games enabling interaction among VTubers or even with viewers are highly favored, further supporting the popularity of gaming topics among VTuber enthusiasts. The Multiple Linear Regression algorithm has been widely used for sales predictions related to buyer patterns and trends. The performance measurement is influenced by the amount of data and the number of attributes.

Keyword: *Hololive, VTuber, Super Chat*

1. INTRODUCTION

YouTube, as a new form of mass media in the develpment of increasingly advanced technology, has a diverse audience that meets its daily information needs [1]. With millions of users, YouTube ha become a vast community where users can upload and share video content with others, both as content creators and as viewers [2]. YouTubers earn a promising source of income from YouTube based on various terms set by Google as the owner of YouTube, depending on having subscribers equal to or above Google's requirements, commonly known as Google Adsense [3]. This program is an advertising initiative provided by Google [4]. If a content creator's uploaded content receives positive appreciation in the form of likes and comments, it can increase the ranking or rating of that content, which in turn affects the commision earned by the uploader [5].

Not all YouTubers express themselves directly or through live streaming, some use intermediary 2D or 3D virtual characters created with the help of computer software utilizing Face Tracking [6] and Hand Tracking technology [7]. To create a humorous impression, most virtual characters are depicted as female. The term of YouTubers who use virtual characters is known as Virtual YouTuber or VTuber for short [8].

The term "Virtual YouTuber" was first used by Kizuna Ai on her YouTube channel named "A.I Channel" [9]. This Japanese Virtual YouTuber began uploading content in November 2016 and is considered the world's first Virtual YouTuber [10]. Initially, the VTuber phenomenon mainly attracted Japanese audience (Khatami, 2023). However, over time VTuber started incorporating translations into English and other languages, including Indonesian, into their video content to make it accessible to a broader audience, not just a Japanese viewer [11].

Hololive Production is a Japanese VTuber agency owned by the Japanese technology company Cover Corporation [12]. It was founded on June 13, 2016 by Motoaki "Yagoo" Tanigo and includes various branches and subsidiaries such as Holostars, Holostars English, Hololive, Hololive English, and Hololive Indonesia [13]. Hololive Indonesia (Hololive ID) was established to produce VTuber content in the Indonesian language, attracting a large fan base both in Indonesia and worldwide [14]. In April 2020, Hololive ID debuted its first three members : Ayunda Risu, Moona Hoshinova, and Airani Iofifteen. In Desember 2020, three more members were introduced : Kureiji Ollie, Anya Melfissa, and Pavolia Reine. Later, in March 2022, Hololive ID added three additional members : Kobo Kanaeru, Vestia Zeta, and Kaela Kovalskia. With the addition of these new members, Hololive ID continued to grow, offering, a greater variety of centent for its fans to enjoy [15].

The explanation above inspires researchers to analyze trends and patterns of Hololive ID VTubers using multiple liniear regression algorithms and visualizations created with Python progaamming language. The process of identifying text with a database is known as data mining [16]. Data mining is the process of discovering anomalies, patterns, and correlations within large datasets to predict outcomes [17]. In data mining functions, appropriate algorithms and methods are required to process inputs into desired outputs. Based on this, the multiple linear regression algorithm is utilized. Multiple linear regression has a linear relationship between two or more independent variables (X1, X2, ... Xn) and dependent variable (Y). To determine the regression coeficients and constants, independent variables can be calculated using a determinant matrix [18].

Based on the aforementioned background, this study aims to analyze patterns, video trends, and factors that significanly inluence VTuber revenue. The study emoploys a multiple linear regression algorithm, with the output expected to meet the desired objectives and present the processed data in a more structured visual format [19]. The data for this research is sourced from the website kaggle.com using the keyword "VTuber". Accuracy measurement in this study is anticipated to identify trends and patterns of Hololive ID VTubers in relation to content variations that influence VTuber revenue.

2. RESEARCH METHODOLOGY

The multiple linear regression research method is a statistical analysis technique that establises a linear relationship between two or more independent variables (X1, X2, ...Xn) and a dependent variable (Y). To determine the values of the regression coefficients and constants, independent vareables can be obtained using a deteminant matrix.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 431-440
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

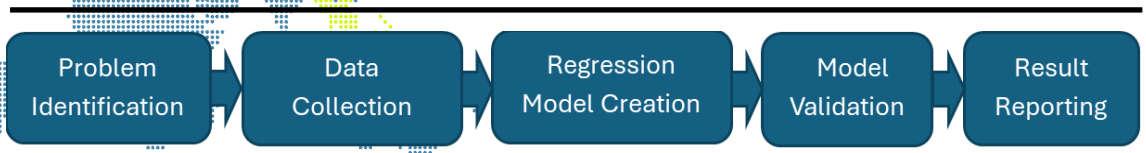

Figure 1. Multiple Linear Regression Stages

2.1. Problem Identification

a. Determine the dependent variable (Y) that will be explained or predicted.
b. Identify the independent variables (X1, X2, ..., Xn) that considered to affect the dependent variable Y.

2.2. Data Collection

a. Collect data through kaggle.com.
b. Ensure that data is not null.

2.3. Regression Model Creation

The mathematical model is given by:

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \epsilon \quad (1)$$

Where:

a. Y: Dependent variable
b. $X_1, X_2, ..., X_n$: Independent variables
c. B_0 : Intercept (constant)
d. $\beta_1, \beta_2, ..., \beta_n$: Regression coefficients
e. ϵ: : Error (residual)

2.4. Model Validation

Validate The Results Using Test Data Or Methods Such As Cross-Validation.

2.5. Result Reporting

Report The Results In The Form Of A Coefficient Table, Statistical Test, And A Brief Interpretation.

3. RESULTS AND DISCUSSION

The dataset utilized in this study comprises 13,050 VTuber records, which were partitioned into training and testing dataset using a 80:20 ratio. Consequently, the data was divided as follows:

a. The dataset used in this study comes from the YouTube platform, containing data about uploaded videos, video shorts, and livestream videos.

This dataset includes the following columns:

1. URL: The URL of the individual livestream page.
2. Video_ID: The YouTube video ID for the livestream.
3. Video_Title: The title of the livestream.
4. Creator: The Hololive ID member who is the host during the livestream.
5. Avg_viewers: The average number of viewers during the livestream.
6. Max_viewers: The maximum number of viewers during the livestream.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 431-440
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

7. SuperChat_Total: The total earnings from the livestream in USD.
8. Stream_Start_Date: The date the livestream started (GMT +0:00).
9. Stream_Start_Time: The time the livestream started (GMT +0:00).
10. Duration: The total duration of the livestream in the format (HH:MM:SS).

b. Training dataset: 2,610 records (20% of total data)

Video_Title	Creator	Avg_viewers	Max_viewers	SuperChat_total	Stream_Start_Date_UST	Stream_Start_Time_UST	duration
【Donation Reading + Free Talk】HAPPY 2023 🍞🎉【Ka...	Kaela Kovalskia	1452	1803	942.65	2023/01/01	03:10:33	04:02:02
LUVORATORRRRRY! - Kobo Kanaeru + @KureijiOllie...	Kobo Kanaeru	1734	1977	1.40	2022/12/31	16:58:09	04:21
【 New Year Countdown 】Tahun Baru Bersama holol...	Airani Iofifteen	2110	3228	176.21	2022/12/31	16:01:54	01:03:17
【New Year Countdown】CLAIM YOUR 2023 BOOSTER PA...	Kaela Kovalskia	2282	3066	898.30	2022/12/31	16:01:23	01:17:26
【NEW YEAR COUNTDOWN】 with fireworks! WOOOHOOO ...	Vestia Zeta	2175	3647	1838.14	2022/12/31	16:00:21	01:29:21
【holopro Countdown Live Watchalong】2023 IS ALM...	Anya Melfissa	1026	1554	4.24	2022/12/31	13:50:52	01:42:16
【CONCERT WATCHALONG】 #ホロライブカウントダウン: TIME TO SI...	Kureiji Ollie	1543	2301	45.70	2022/12/31	13:48:48	01:38:21

Figure 2. Training Dataset

c. Testing dataset: 10,440 records (80% of total data)

Video_Title	Creator	Avg_viewers	Max_viewers	SuperChat_total	Stream_Start_Date_UST	Stream_Start_Time_UST	duration
【UNARCHIVED KARAOKE】 Singing til the new year!...	Ceres Fauna	13542	16155	2788.90	2023/01/01	05:53:07	02:14:25
【🧘2023年🧘】書道「段」持ちがマウスで挑戦！！デジタル書き初め！！【Amane Kana...	Amane Kanata	12305	14550	2635.44	2023/01/01	05:00:56	01:58:31
【Valorant】年明け早々プラチナに行きたい！ウィスパー声縛りvalorant【Robo...	Robokosan	1684	2042	175.40	2023/01/01	04:44:19	04:27:35
NEW YEAR CHITCHAT & Rewatching my official pro...	Takanashi Kiara	2984	4309	4601.01	2023/01/01	04:32:11	03:44:19
【新年初雑談】あけましておめでとうございます✨A HAPPY NEW YEAR!! 【Lui T...	Lui Takane	4363	7513	1924.36	2023/01/01	04:01:02	03:31:35
【NEW YEARS PARTY】COUNTING DOWN 2023! I vtstats	Mori Calliope	4870	6513	4074.23	2023/01/01	04:00:40	04:34:15
【NYE】LET'S CELEBRATE! #2023 I vtstats	Watson Amelia	11736	19255	3171.46	2023/01/01	03:01:04	02:23:21

Figure 3. Testing Dataset

To ensure that the obtained dataset does not contain null data and is clean, as stated in the dataset source from kaggle.com, a check is performed on the dataset for any null values. These null values are then displayed in Google Colab.

```
data_streaming.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13049 entries, 0 to 13048
Data columns (total 10 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   URL                   13049 non-null  object
 1   Video_ID              13049 non-null  object
 2   Video_Title           13049 non-null  object
 3   Creator               13049 non-null  object
 4   Avg_viewers           13049 non-null  object
 5   Max_viewers           13049 non-null  object
 6   SuperChat_total       13049 non-null  float64
 7   Stream_Start_Date_UST 13049 non-null  object
 8   Stream_Start_Time_UST 13049 non-null  object
 9   duration              13049 non-null  object
dtypes: float64(1), object(9)
memory usage: 1019.6+ KB
```

Figure 4. Checking null on Dataset

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 431-440
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Next, the dataset that has been checked for null values will be processed by removing the index columns that are not used in the experiment. In this case, the columns that will be removed are the URL column and the video ID column, and the cleaned dataset will then be displayed in Google Colab.

	Video_Title	Creator	Avg_viewers	Max_viewers	SuperChat_total	Stream_Start_Date_
0	【UNARCHIVED KARAOKE】 Singing til the new year!...	Ceres Fauna	13542	16155	2788.90	2023/0
1	【🎍2023年🎍】書道「段」持ちがマウスで挑戦！！デジタル書き初め！！【Amane Kana...	Amane Kanata	12305	14550	2635.44	2023/0
2	【Valorant】年明け早々プラチナに行きたい！ウィスパー声縛りvalorant【Robo...	Robokosan	1684	2042	175.40	2023/0
3	NEW YEAR CHITCHAT & Rewatching my official pro...	Takanashi Kiara	2984	4309	4601.01	2023/0

Figure 5. Deleted Unused Column on Dataset

The next step is to analyze the SuperChat, Average Viewer, and Max Viewer data for VTubers from Indonesia, using a training dataset consisting of 2,610 records. This data will then be displayed in Google Colab.

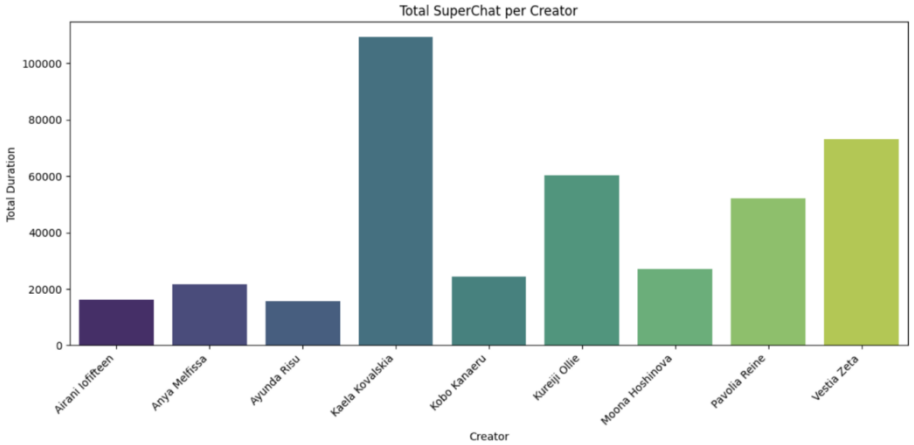

Figure 6. Total SuperChat per Creator from Training Dataset

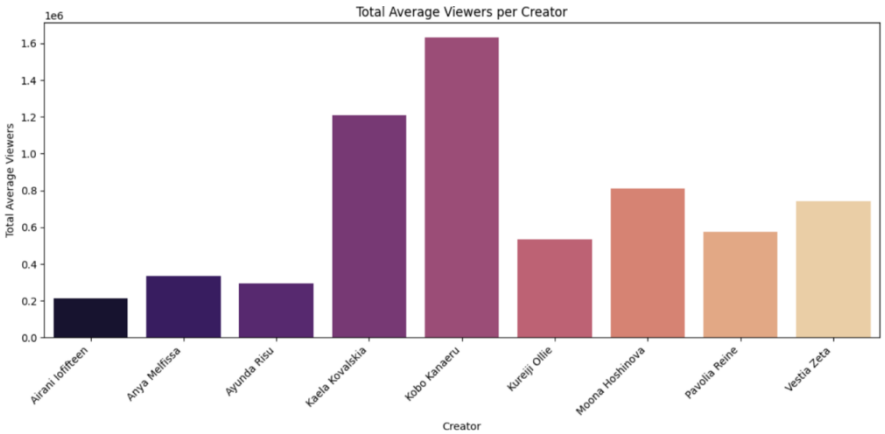

Figure 7. Total Average Viewers per Creator from Training Dataset

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 431-440
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

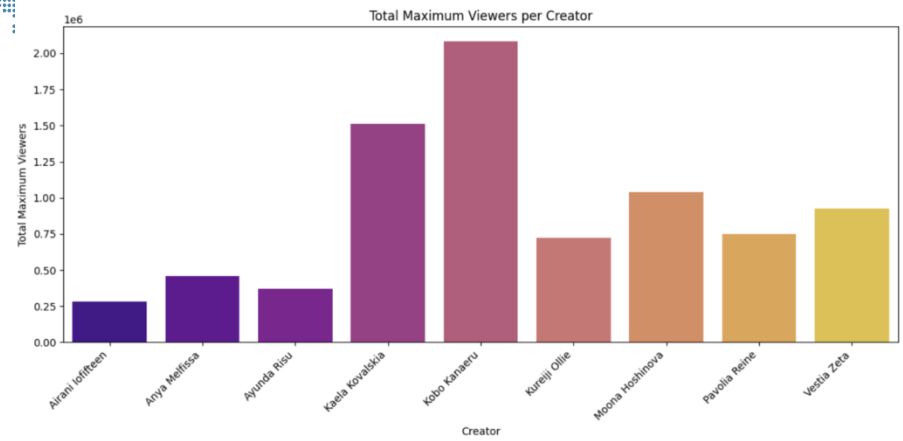

Figure 8. Total Maximum Viewers per Creator from Training Dataset

The next step is to analyze the SuperChat, Average Viewer, and Max Viewer data for VTubers except from Indonesia, using a testing dataset consisting of 10,440 records. This data will then be displayed in Google Colab.

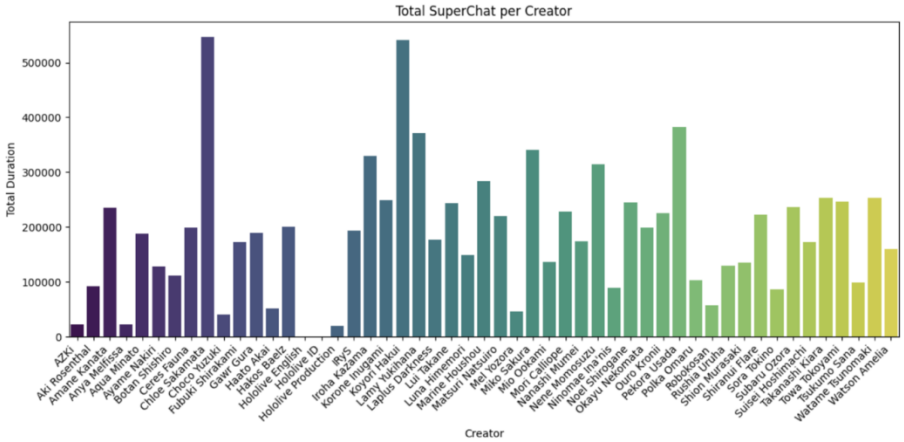

Figure 9. Total SuperChat per Creator from Testing Dataset

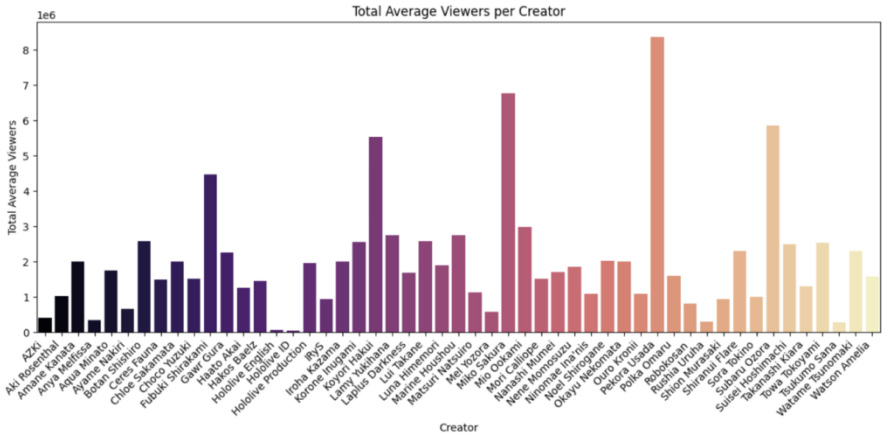

Figure 10. Total Average Viewers per Creator from Training Dataset

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 431-440
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

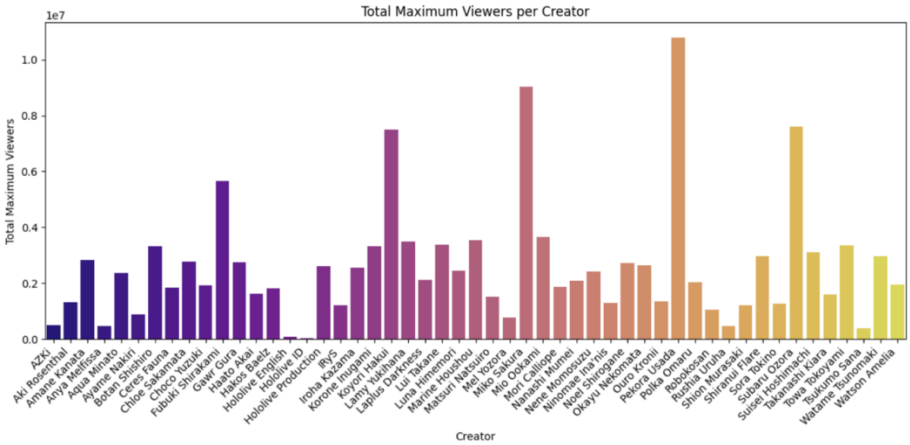

Figure 11. Total Maximum Viewers per Creator from Testing Dataset

The next step is to perform a data test using Multiple Linear Regression with the training and testing subjects, and then display the results in Google Colab.

```
Selected Data:
Model Evaluation:
Mean Squared Error: 0.0
R-squared: 1.0
Mean Absolute Error: 0.0
Mean Absolute Percentage Error: 0.0

Unselected Data:
Model Evaluation:
Mean Squared Error: 0.0
R-squared: 1.0
Mean Absolute Error: 0.0
Mean Absolute Percentage Error: 0.0
```

Figure 12. Result of Test using Multiple Linear Regression

The next step is to compare SuperChat and Max Viewers for every creator using the training and testing data by implementing the multiple linear regression algorithm.

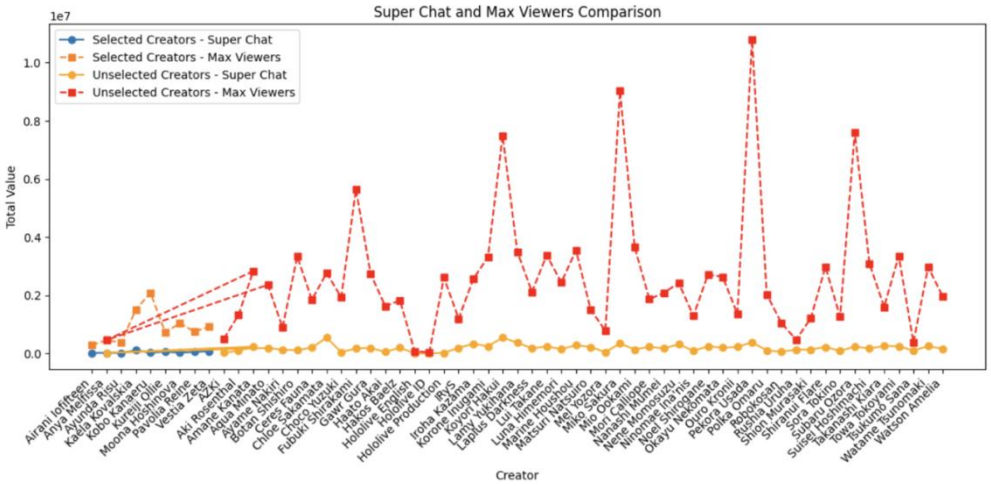

Figure 13. Result of Comparing using Multiple Linear Regression

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 431-440
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

The next step is to examine the total number of videos uploaded by each creator, and then obtain the overall results from the tests conducted.

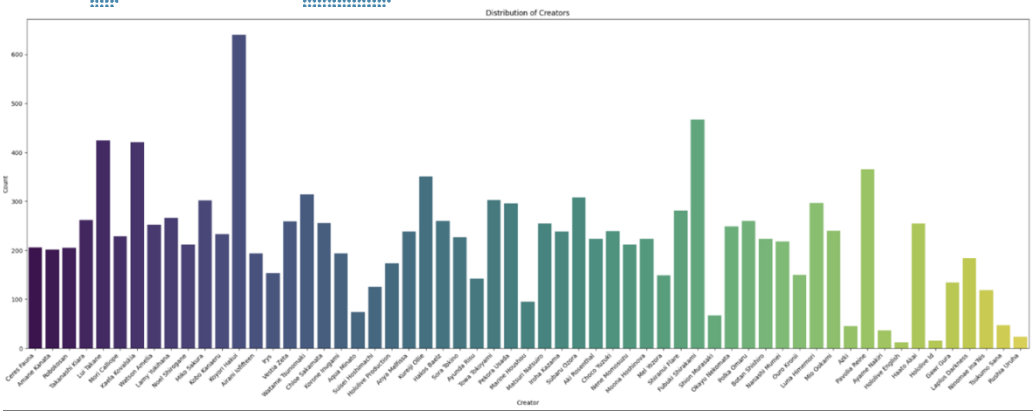

Figure 14. Result of Uploaded Video by Creators

The next step is to examine the total number of Livestreaming duration by each creator, and then obtain the overall results from the tests conducted.

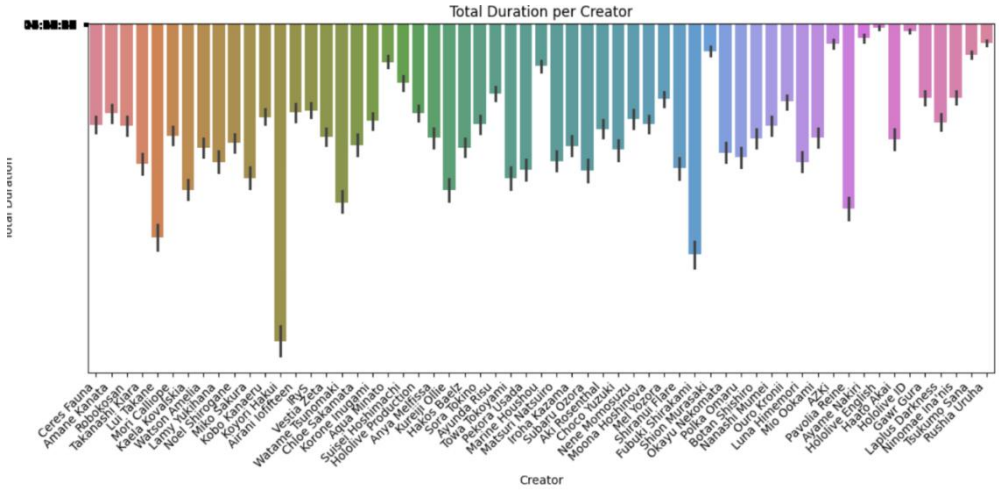

Figure 15. Result of LiveStreaming by Creators

4. CONCLUSION

From the graph obtained in the testing using Google Colab, this study successfully identified patterns in the number of videos, duration, SuperChat amount, Average Viewer, and Max Viewer. Based on Figure 4, it shows that no null data was found in the dataset used, and the data is clean, as confirmed by the source from Kaggle.com. Based on Figure 14, it was found that the VTuber with the most videos uploaded is Koyori Hakui, with 625 video uploads. Then, in Figure 15, it also shows that Koyori Hakui had the longest livestreaming duration. Figure 11 explains that the SuperChat earnings of over USD 500,000 went to Koyori Hakui. However, Figure 13 show that the Average Viewer and Max Viewer, with over 8

million (8,000,000) viewers, went to Pekora Usada. Based on the explanation above, the patterns and trends affecting monetization in VTubers were found to be influenced by the number of videos and the duration of livestreaming. VTubers who stream for longer and upload more videos tend to earn more SuperChat compared to those who rely solely on the number of viewers. Data testing was conducted using the Multiple Linear Regression Algorithm, where each test was carried out using two subjects: selected_data (Hololive Indonesia VTubers) and unselected_data (Hololive VTubers outside Indonesia). The error rate in the test results showed a Mean Squared Error of 0.0, an R-squared value of 1.0, a Mean Absolute Error of 0.0, and a Mean Absolute Percentage Error of 0.0. These test results indicate that the data tested is accurate and consistent, without any data manipulation.

REFERENCES

[1] Adzania, J. I., & Arianingsih, A. (2022). Dampak Menonton Vtuber Terhadap Penggemar Budaya Populer Jepang Di Bandung. *Mahadaya: Jurnal Bahasa, Sastra, Dan Budaya*, *2*(2). https://doi.org/10.34010/mhd.v2i2.8265.

[2] Akdeas Oktanae Widodo, Septiadi, F., & Nur Aini Rakhmawati. (2023). Analisis Tren Konten Pada Vtuber Indonesia Menggunakan Latent Dirichlet Allocation. *Jurnal Informatika Dan Rekayasa Elektronik*, *6*(1). https://doi.org/10.36595/jire.v6i1.718

[3] Asohi, Y., & Andri, A. (2020). Impelementasi Algoritma Regresi Linier Berganda Untuk Prediksi Penjualan. *Jurnal Nasional Ilmu Komputer*, *1*(3). https://doi.org/10.47747/jurnalnik.v1i3.161

[4] Aulia, A., Tanjung, M. R., Iqbal, M., & Wijaya, R. F. (2023). Implementasi Algoritma Naïve Bayes Dalam Menganalisis Jumlah Live Stream VTuber Skem. *Bulletin of Information Technology (BIT)*, *4*(4). https://doi.org/10.47065/bit.v4i4.899

[5] Bermúdez Badia, S., Kuramoto, I., & Sakuma, H. (2023). *YouTubers vs. VTubers: Persuasiveness of human and virtual presenters in promotional videos*.

[6] Carolina, A., & Inayati, I. (2021). Analisis Strategi Pemerintah dalam Upaya Meningkatkan Kepatuhan Wajib Pajak Youtuber Di Indonesia. *Syntax Literate ; Jurnal Ilmiah Indonesia*, *6*(2). https://doi.org/10.36418/syntax-literate.v6i2.2189

[7] Enoch, P. I. D., Anis, F. H., & Mamahit, C. E. M. (2023). Tinjauan Hukum Terhadap Pajak Penghasilan Profesi Youtuber Yang Memperoleh Pendapatan Dari Google Adsense. *Lex Privatum*, *7*.

[8] Fathudin &, & Muhammad Mukromin. (2021). Advertising Business Pada Google AdSense di Youtube Perspektif Ekonomi Syariah. *Jurnal Ilmiah Ekonomi Islam*, *7*(03).

[9] Khatami, F. (2023). *Pengembangan Media Vtuber Dan Implementasi Dalam Pembelajaran Bahasa Indonesia* (Vol. 1, Issue 2).

[10] Margiani Saraswati, A., Rizal Setiawan, I., & Asriyanik. (2023). *Pemanfaatan Tools Live2D terhadap Animasi menggunakan Metode Face Tracking*. *16*(1), 186–195. https://doi.org/10.51903/pixel.v16i1.1242

[11] Meng, R., & He, W. (2023). Data Mining. In *Flavoromics* (pp. 83–122). CRC Press. https://doi.org/10.1201/9781003268758-8

[12] Mohammad, W., Saifurrahman, H., Rausyan Fiqrussalam, J., Umar Sulthoni, A., & Fansyiari Austi, D. (2023). *Triwikrama: Jurnal Multidisiplin Ilmu Sosial Pengaruh Jumlah Subscribers Dan Jumlah Penayangan Terhadap Pendapatan Super Chat Pada Virtual Youtuber Kobo Kanaeru*. *01*, 30–45.

[13] Mustofa, Z. (2021). Data mining: Pengertian, proses, manfaat, dan 3 contoh penerapannya. *UNIVERSITAS STEKOM Universitas Sains & Teknologi Komputer*.

[14] Mutiah, T. (2020). Trend Youtuber di Generasi Millenial dalam Pengembangan Kreati. *Global Komunika*.

[15] Normah, Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer AMIK BSI*, *8*(2). https://doi.org/10.31294/jtk.v4i2

[16] Padilah, T. N., & Adam, R. I. (2019). *Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang*.

[17] Pratama, H. R., & Kahija, Y. F. La. (2022). Apa Artinya Menjadi Youtuber? Analisis Fenomenologis Interpretatif Tentang Pengalaman Menjadi Youtuber. *Jurnal EMPATI*, *11*(3). https://doi.org/10.14710/empati.2022.34465

[18] Puspitaningrum, D. R., & Prasetio, A. (2019). Fenomena "Virtual Youtuber" Kizuna Ai di Kalangan Penggemar Budaya Populer Jepang di Indonesia. *Mediator: Jurnal Komunikasi*, *12*(2). https://doi.org/10.29313/mediator.v12i2.4758

[19] Reyus Liman Prodi Perhotelan, M., & Yapari Bandung, S. (2023). *Pengaruh Hololive Indonesia dalam Memperkenalkan Bahasa Indonesia Kepada Penonton Internasional*. *1*(2).

[20] Saputra, D. I. S., & Setyawan, I. (2021). Virtual YouTuber (VTuber) sebagai Konten Media Pembelajaran Online. *Prosiding SISFOTEK*, *5*(1).