Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

Available online to https://ejournal.uin-suka.ac.id
Profetik
Journal Page is available at https://ejournal.uin-suka.ac.id/isoshum/profetik/

From Pulpit to Platform: Algorithmic Mediation and the Transformation of Religious Authority on YouTube

M. Didik R. Wahyudi[1], Noorhaidi Hasan[2], Agung Fatwanto[3]
[1,2,3] Universitas Islam NegeriSunan Kalijaga Yogyakarta
**Corresponding author, e-mail: m.didik@uin-suka.ac.id*

ARTICLE INFO

Article history:
Received 3 March 2024
Reviewed 3 April 2025
Revised 7 January 2025
Revised 3 March 2025
Revised 21 November 2025
Accepted 21 November 2025
Published online: 30 December 2025

Keywords: YouTube Algorithm; Content Recommendation; Filter Bubble; Echo Chamber; Religious Authority

ABSTRACT

Social media algorithms have been widely argued to shape information exposure through personalization based on users' interaction histories, with potential implications for filter-bubble and echo-chamber dynamics. This study examined how algorithmic variables affected the relevance of YouTube content recommendations and considered broader socio-religious implications of increasingly personalized visibility. A quantitative quasi-experimental Interrupted Time Series Design (ITSD) was implemented by creating 25 test accounts with diverse demographic profiles and interest themes (informed by APJII 2024). Each account followed the same procedure across five iterations: a keyword search was conducted, the top-10 recommended videos were recorded, and three recommendations were opened using a randomized selection rule, yielding 1,250 video observations. Data was collected via the YouTube API, manually coded for recommendation relevance, transformed into numeric variables, and cleaned using the Interquartile Range (IQR) method. A logistic regression model was estimated and validated using the Hosmer–Lemeshow test, logit-linearity checks, and Variance Inflation Factor (VIF) diagnostics. Simple Exponential Smoothing (SES) and Holt's Linear Trend were applied to project recommendation patterns across iterations. Iteration emerged as the most influential predictor of recommendation relevance, whereas other variables showed small or non-significant effects. The model demonstrated acceptable fit and no problematic multicollinearity, and forecasting suggested increasing relevance across iterations. Overall, the results were consistent with the strengthening of viewing-history-based personalization, which may reduce informational diversity and may facilitate a shift of religious authority toward digital actors more adaptive to algorithmic visibility.

To cite this article with APA Style:	Wahyudi, M. D. R., Hasan, N., & Fatwanto, A. (2025). *From pulpit to platform: Algorithmic mediation and the transformation of religious authority on YouTube*. Profetik: Jurnal Komunikasi, 18(2), 352–377.

INTRODUCTION

The rapid development of digital technology and social media platforms has fundamentally transformed how societies access information, interact, and shape public opinion. Platforms such as YouTube, Facebook, Instagram, TikTok, and Twitter now function not only as communication tools but also as algorithm-driven content distribution systems (Saura, Palacios-Marqués, & Iturricha-Fernández, 2021). This transformation has created new spaces for religious expression that were previously centered in physical institutions such as mosques, pesantren, and study circles. In the digital era, YouTube operates simultaneously as a medium for religious communication and as a site for contesting religious authority. Social media algorithms play a central role in shaping content visibility by prioritizing engagement metrics, including clicks, comments, watch time, and emotional responses (Narayanan, 2023). Within this system, sensational or emotionally charged content tends to be prioritized over educational and reflective material (Rodilosso, 2024). As a result, religious popularity and authority are no longer solely determined by scholarly lineage (*sanad*) or institutional affiliation but also by algorithmic performance, how well content creators can adapt to user preferences, and platform recommendation strategies. This phenomenon has given rise to new models of religious authority that are symbolic, performative, and algorithmic in nature. In this context, algorithms generate increasingly personalized information spaces where users are more frequently exposed to content aligned with their previous behaviors and preferences. This condition contributes to the formation of filter bubbles, the confinement of information to homogeneous content (Pariser, 2011), and echo chamber, where existing views are reinforced. Unknowingly, these phenomena limit the diversity of religious perspectives accessible to users (Harambam, Helberger, & Van Hoboken, 2018). In Indonesian communication studies, the dynamics

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

of issues in digital spaces are also frequently examined through content analysis of online media coverage, for instance to map patterns of balance and framing tendencies in sensitive issues (Hindarto, 2022).

The literature indicates that the curation architecture and recommendation logic of major social media platforms differ, meaning that the variables most heavily "weighted" in ranking are not identical across platforms (Gillespie, 2014; Nieborg & Poell, 2018). YouTube emphasizes personalization based on viewing/activity history and optimizes ranking toward expected watch time (Covington, Adams, & Sargin, 2016; Davidson, Liebald, Liu, Nandy, & Van Vleet, 2010). TikTok is highly responsive to interaction signals and watch time, with recommendation factors that include user interactions, content information, and user information (including temporal context such as time zone/day) (TikTok, 2020). Meanwhile, Facebook and Twitter/X foreground relevance and social/network relations (e.g., content from connections that users frequently interact with), while also using many signals to predict the value of content for users (Bucher, 2012; Engineering, 2023; Meta, 2025). More broadly, platform incentives (engagement and the attention or advertising economy) tend to increase the visibility of high-performing and "advertiser-friendly" content, including through monetization and demonetization practices that shape a hierarchy of visibility (Bishop, 2018; Fourcade & Johns, 2020; Kingsley, Sinha, Wang, Eslami, & Hong, 2022). As a consequence, content distribution becomes non-neutral because algorithmic curation functions as a form of selection/filtering that has implications for inequalities of visibility and users' informational experience (Eslami et al., 2015; Gillespie, 2014).

To bridge technical findings and social implications, this study employs the framework of machine habitus and digital habitus. Machine habitus refers to the predisposition of recommendation systems formed through the accumulation of usage traces (viewing history and engagement signals), such that platforms tend to stabilize preferences as they are read from user data (Airoldi, 2022). Digital habitus explains how experiences mediated by personalization and imaginaries of AI shape users' dispositions in selecting and interpreting information (Romele, 2023). Within this framework, the increasing relevance of recommendations across cycles can be read as a reinforcement of machine predispositions that may narrow exposure diversity, thereby framing filter bubbles and echo chambers as gradual processes.

ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

Within the framework of algorithmic gatekeeping and reinforcing spirals, the filter-bubble phenomenon is supported by limited data transparency and algorithmic mechanisms that cannot be fully explained to the public (Narayanan, 2023), as well as by challenges in media literacy, public communication performance, and the effectiveness of regulation in mitigating the impact of misleading information on social media (Prianto, Abdillah, Syukri, Muhammad, & Yama, 2021). This can generate bias and inequality in content distribution (Ji, 2004), especially for content creators who lack adequate technical understanding of these systems (Wang, Zhang, & Yamasaki, 2019). Variables such as metadata, video duration, upload frequency, and the emotional sentiment of content are important factors in determining content visibility and recommendation on digital platforms (Wang et al., 2019). Engagement, as a core operational metric, not only shapes content distribution but also produces new forms of authority that are algorithmic and symbolic (Saura et al., 2021). Algorithms create feedback loops in which repeatedly recommending popular content can reinforce inequalities in visibility distribution (Davidson et al., 2010). The shift from scholarly lineage–based authority toward visibility-based authority thus becomes a consequence of distribution systems governed by engagement metrics and user preferences (Dujeancourt & Garz, 2023).

Understanding four empirically tested algorithmic variables (engagement, personalization, content relevance, and transparency) is essential for designing ethical and effective strategies for producing and distributing content, particularly religious content, within an ecosystem governed by algorithmic logic (Larsson, 2020). In this study, engagement was operationalized through observable interaction signals available in the dataset (e.g., channel- and video-level engagement indicators), personalization was captured by user-profiled account attributes and repeated exposure across iterations, content relevance was measured via manual relevance coding of recommended videos to the search intent, and transparency was approached indirectly through observable recommendation outputs rather than internal platform explanations. While social network dynamics remains analytically important for understanding how recommendations circulate and gain visibility, it was treated here as part of the conceptual framework rather than an empirical predictor because it was not directly measured in the dataset. Based on this framework, the main proposition tested was that repeated

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

interactions across iterations strengthen personalization signals, thereby increasing the probability that recommended videos are coded as relevant.

In response to this logic, content creators develop various optimization strategies to increase the acceptance and spread of their content on platforms, including paid promotions or amplification through popular accounts to expand distribution reach (Wulandari & Nuraniwati, 2023). Content optimization cannot be separated from a deep understanding of algorithmic variables (Delmonaco et al., 2024). By knowing how algorithms work, content creators can craft more adaptive and effective strategies to reach wider audiences while avoiding pitfalls such as filter bubbles and purely sensational content. The identification of these five algorithmic variables serves as a conceptual foundation for analyzing religious content production strategies (Griffiths et al., 2024). This opens up opportunities for religious actors to formulate distribution strategies for *dakwah* that are not only adaptive to algorithmic systems but also preserve the integrity of authentic religious messages (Harambam et al., 2018).

METHODOLOGY

This study adopted a quantitative approach using a quasi-experimental design, operationalized as a repeated-measures Interrupted Time Series Design (ITSD), in which recommendation outcomes were observed across sequential iterations (Creswell, 2014; Shadish, Cook, & Campbell, 2002). The design was used to examine whether changes in recommendation relevance across iterations were consistent with increasingly history-dependent personalization and with patterns commonly discussed in the filter-bubble and echo-chamber literature.

A total of 25 YouTube accounts were created with varied biodata (age, gender, domicile, topics of interest, and search keywords) based on the 2024 APJII data on internet user profiles in Indonesia (APJII, 2024). Each account was used to search for videos, record the top 10 recommendations, and then randomly open 3 of those videos. This process was repeated for 5 iterations, resulting in a total sample of 1250 videos.

The dataset was collected via the YouTube API and then processed in several stages. First, manual coding was performed to assess the relevance of the videos to the initial search theme (Krippendorff, 2019). Second, data transformation was conducted from text to numerical format (e.g., converting duration to seconds, age to months, and encoding categorical variables) (Allison, 2012). Third, the dataset underwent a data cleaning phase using the Interquartile Range

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

(IQR) method to identify and remove outliers (Tukey, 1977).

Data analysis was carried out using logistic regression to test the significance of the independent variables (age, gender, domicile, subscribers, duration, comments, views, likes, dislikes, iteration) on the dependent variable (relevance). Validation was performed through the Hosmer-Lemeshow (goodness-of-fit) test, linearity of the logit, and Variance Inflation Factor (VIF) (Hair, Black, Babin, & Anderson, 2010; Hosmer Jr., Lemeshow, & Sturdivant, 2013). To project recommendation trends in subsequent iterations, a forecasting model with Simple Exponential Smoothing (SES) and Holt's Linear Trend Model (HLTM) was used (R.J. Hyndman & Athanasopoulos, 2018). The results of this quantitative analysis will then be used as a basis for interpreting the socio-religious implications related to the algorithm's potential to drive shifts in religious authority, content distribution bias, and the tendency to promote specific content.

RESULTS AND DISCUSSIONS

Experiment on Algorithmic Variables on YouTube

A quasi-experiment is a quantitative research design approach used to test causal relationships, particularly in contexts where researchers do not have full control over the random assignment of participants into treatment and control groups (Cook, 1979). Unlike a true experiment, an experiment still involves the application of a treatment and the measurement of its impact, but the groups being compared are often naturally or administratively formed before the study begins. This makes experimental designs commonly used in real-world settings such as schools, hospitals, or social communities, where randomization is often unethical or impractical (Shadish et al., 2002).

The experimental design employed in this study is the Interrupted Time Series Design (ITSD), which involves a series of measurements taken on the same group at multiple time points before and after an intervention(Creswell, 2014). The ITSD framework for this research includes the following components:

1. Observing the initial state of YouTube video recommendations generated by the algorithm and the state of recommendations after a viewing history has been established by the YouTube viewer.
2. Preparing 25 Google accounts that will be used to access YouTube. Each of the 25 accounts will be configured with different profile attributes based on date of birth, age

ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

group, gender, theme, location, and keywords. This profiling is based on the results of the 2024 survey by the Indonesian Internet Service Users Association (APJII), particularly the section on the most accessed social media platforms in Indonesia and users' favorite content by age group. Each YouTube account will be assigned a specific theme and keyword for video searches. Table 1 below presents the scenario and demographic profiles of the 25 Google users that will be used in this experiment.

3. In accessing YouTube, each user will use the Google Chrome browser in incognito mode to ensure independence and avoid bias in content recommendations between YouTube accounts. Each YouTube account will log in and search for videos based on predetermined keywords, and then the top 10 videos recommended by YouTube will be taken. From the first 10 recommended videos, 3 videos will be randomly selected to be opened/played and will also be a reference for YouTube in recommending the next video. From these YouTube recommended videos, the top 10 videos will be taken again, and then 3 videos will be taken to be opened. This step will be repeated up to 5 iterations. So that each iteration will collect 250 videos from 25 YouTube accounts and from five iterations a total of 1250 YouTube sample videos will be collected.

Table 1: Scenarios and demographics of 25 Google users

No	IDYT	DATE OF BIRTH	GENDER	THEME	ADDRESS	KEYWORD
1	ID02	04/05/1952	L	Politik	Jawa	Ijazah jokowi roy suryo
2	ID05	17/03/1960	L	Politik	Jawa	Makan bergizi gratis
3	ID07	26/02/1962	L	Politik	Kalimantan	Perekonomian Indonesia
4	ID09	08/07/1954	P	Infotainment	Kalimantan	Bisnis kuliner artis
5	ID10	10/12/1958	P	Infotainment	Jawa	Ivan gunawan dan ruben onsu
6	ID12	11/12/1963	P	Infotainment	Sulawesi	Karir politik artis Indonesia
7	ID14	11/06/1964	P	Infotainment	Jawa	Artis hijrah

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

No	IDYT	DATE OF BIRTH	GENDER	THEME	ADDRESS	KEYWORD
8	ID15	04/11/1965	L	Olahraga	Papua	Hamilton pindah ferrari
9	ID16	23/05/1967	L	Olahraga	Jawa	Liga champion
10	ID18	09/10/1970	L	Olahraga	Kalimantan	Formula 1
11	ID21	28/01/1973	L	Olahraga	Jawa	UFC
12	ID22	06/07/1977	P	Kesehatan	Sumatera	Gejala stroke
13	ID23	19/03/1972	P	Kesehatan	Jawa	Bersahabat dengan diabetes
14	ID24	21/09/1979	P	Kesehatan	Sulawesi	Chiropractic
15	ID26	28/04/1975	P	Kesehatan	Sumatera	Menjaga kesehatan jantung
16	ID29	24/01/1981	L	Ekonomi	Jawa	Efisinsi anggaran belanja negara
17	ID32	18/09/1984	L	Ekonomi	Jawa	Pengangguran dan kemiskinan di Indonesia
18	ID33	27/07/1995	L	Ekonomi	Bali	UMKM
19	ID35	16/08/1990	P	Iptek	Sulawesi	Teknologi hybrid mobil listrik
20	ID36	28/05/1989	P	Iptek	Jawa	Nvidia dan AI
21	ID37	12/02/1994	P	Iptek	Sumatera	Pertanian dengan IOT
22	ID40	23/03/2001	L	Budaya/Wisata	Jawa	Glamping dan kamping
23	ID41	04/05/2000	L	Budaya/Wisata	Sumatera	Wisata kuliner
24	ID43	16/04/1998	P	Politik	Sumatera	DPR
25	ID45	15/10/2004	P	Mancanegara	Jawa	Konflik india pakistan

Dataset Compilation

Video sample collection is done with the help of Python code with the help of YouTube API, where the input required is YouTube Video ID. From this process, each sample taken will be categorized by theme to determine whether the video recommended by YouTube based on viewing history is relevant to the specified theme or not. This categorization process was carried out using a manual coding approach based on observations of the video titles, thumbnails, and descriptions, which were placed in the RELEVAN_TM column with category values of "RELEVANT" and "NOT RELEVANT" (Krippendorff, 2019). Figure 1 shows

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

an example of the dataset layout, organized with the help of a Python script.

Figure 1 : Example of experimental results of YouTube video recommendations

IDYT	Usia	Gender	Domisili	Kata Kunci	TEMA	ELEVAN_TE	TEMA_VID	IDV	RELEVAN_K	NmChanel	JmlSubsc	JudulVid	TglUpload	DurasiVid	JmlKomen	JmlView	JmlLike	JmlDisLike	[illegible]
ID01	02-Mar-61	L	Sumatera	mutasi ketu	Politik	TIDAK	Infotainme	[illegible]	TIDAK	trendsmesi	2310000	WANITA SEO	2024-06-2	00:08:30	585	2298321	9990	182	1
ID01	02-Mar-61	L	Sumatera	mutasi ketu	Politik	RELEVAN	Politik	[illegible]	RELEVAN	Hersubeno	989000	CERITA SERU	2025-05-0	00:19:34	2118	311878	7415	23	1
ID02	04-May-52	L	Jawa	ijazah jokos	Politik	RELEVAN	Politik	[illegible]	RELEVAN	Refly Harun	3570000	BREAKING	2025-05-0	00:44:28	1421	76584	2425	29	1
ID13	27-Sep-82	F	Jawa	koleksi mo	Infotainme	RELEVAN	Infotainme	[illegible]	TIDAK	tanboy kun	20000000	MUKBANG N	2024-09-1	00:23:44	2699	5892146	59151	1072	2
ID13	27-Sep-82	F	Jawa	koleksi mo	Infotainme	TIDAK	Musik	[illegible]	TIDAK	Melody Mu	2150	NIKI - You'll	2025-05-0	00:29:55	28	414389	1241	12	2
ID14	11-Jun-64	F	Jawa	artis hijrah	Infotainme	RELEVAN	Infotainme	[illegible]	RELEVAN	kisah islam	149000	KISAH MUAL	2025-05-1	00:56:08	13	5333	112	4	2
ID14	11-Jun-64	F	Jawa	artis hijrah	Infotainme	RELEVAN	Infotainme	[illegible]	RELEVAN	Dedeyuli	475000	HAJAB!! Mua	2025-05-1	00:22:30	259	43460	588	3	2
ID33	27-Jul-95	L	Bali	umkm	Ekonomi	RELEVAN	Ekonomi	[illegible]	TIDAK	SUARA BER	252000	Punya AMBIS	2024-12-1	00:58:48	180	241907	5071	46	3
ID33	27-Jul-95	L	Bali	umkm	Ekonomi	RELEVAN	Ekonomi	[illegible]	RELEVAN	ZAINI M	1300	Bisnis sukses	2025-05-0	00:08:25	3	273	2	0	3
ID33	27-Jul-95	L	Bali	umkm	Ekonomi	RELEVAN	Ekonomi	[illegible]	TIDAK	Leon Harto	994000	4 PASIF INCO	2025-04-1	00:18:03	679	370015	5721	106	3
ID14	11-Jun-64	F	Jawa	artis hijrah	Infotainme	RELEVAN	Infotainme	[illegible]	TIDAK	MUALAF	735000	BERGETAR M	2025-05-1	00:51:22	128	4625	314	1	3
ID14	11-Jun-64	F	Jawa	artis hijrah	Infotainme	RELEVAN	Infotainme	[illegible]	RELEVAN	VERMIA	122000	MURTAD	2025-05-1	00:31:53	72	2606	186	4	3
ID15	04-Nov-65	L	Papua	hamilton	Olahraga	RELEVAN	Olahraga	[illegible]	TIDAK	Overtake	12400	Emilia Roma	2025-05-1	00:09:44	39	14760	154	3	3
ID15	04-Nov-65	L	Papua	hamilton	Olahraga	TIDAK	Ekonomi	[illegible]	TIDAK	Raymond	2920000	Project Nega	2025-05-0	00:13:07	5268	990846	29573	61	3
ID34	10-Sep-93	F	Jawa	masa depa	Iptek	RELEVAN	Iptek	[illegible]	TIDAK	Kodex Arca	33600	Saat Kamu M	2025-04-2	00:50:34	1024	477436	13205	93	3
ID34	10-Sep-93	F	Jawa	masa depa	Iptek	RELEVAN	Iptek	[illegible]	RELEVAN	Economy	34000	How AI Engin	2025-04-1	00:08:44	265	102335	2671	168	3
ID35	16-Aug-90	F	Sulawesi	mobil listri	Iptek	RELEVAN	Iptek	[illegible]	RELEVAN	Auto secret	15100	New BYD Yan	2025-03-3	00:10:45	2545	2551258	27302	814	3
ID22	06-Jul-77	F	Sumatera	gejala strok	Kesehatan	RELEVAN	Kesehatan	[illegible]	TIDAK	SAHABAT	15600	Tubuhmu Od	2025-05-1	01:57:59	3	511	31	0	3
ID22	06-Jul-77	F	Sumatera	gejala strok	Kesehatan	RELEVAN	Kesehatan	[illegible]	TIDAK	Gaya Hidup	74100	Posisi tidur P	2025-04-1	00:19:19	243	321455	3484	98	3
ID23	19-Mar-72	F	Jawa	bersahabat	Kesehatan	RELEVAN	Kesehatan	[illegible]	RELEVAN	[illegible]	2290000	Makan Enak	2024-04-1	00:09:13	78	24568	400	2	3
ID23	19-Mar-72	F	Jawa	bersahabat	Kesehatan	RELEVAN	Kesehatan	[illegible]	TIDAK	Nurse Guid	1220	Cholesterol E	2025-05-1	00:23:58	5	429	20	1	3
ID23	19-Mar-72	F	Jawa	bersahabat	Kesehatan	RELEVAN	Kesehatan	[illegible]	RELEVAN	Sehat Dala	2350	3 Vitamin Ter	2025-05-1	00:24:07	5	3833	33	0	3
ID41	04-May-00	L	Sumatera	wisata kul	Budaya/Wi	RELEVAN	Budaya/Wi	[illegible]	RELEVAN	usa street	25100	What You Ca	2025-04-1	00:52:21	58	58959	242	2	3
ID42	07-Sep-87	L	Jawa	adat istiadi	Budaya/Wi	TIDAK	Politik	[illegible]	TIDAK	Nessie Jud	11300000	PRESIDEN KE	2025-05-0	00:25:33	2100	303035	12066	195	3
ID42	07-Sep-87	L	Jawa	adat istiadi	Budaya/Wi	RELEVAN	Budaya/Wi	[illegible]	TIDAK	Dunia Alam	230000	Harta Karun	2025-04-1	00:20:01	1325	543065	14173	161	3

Dataset Codification

The dataset was generated from a collection of video samples obtained in the previous stage, still containing a mix of text and numerical data. Therefore, the next step involves transforming the data, particularly the non-numeric columns, by reclassifying them into numeric or categorical variables to fit the input structure required by the logistic model(Allison, 2012). Logistic regression is a binary outcome approach, as it estimates the probability of an event occurring based on a combination of numeric and categorical variables(Hosmer Jr. et al., 2013). Additionally, logistic regression provides insights into the direction of relationships, the strength of influence through the odds ratio (OR) values, and the statistical significance of each predictor variable. Table 2 presents the dataset conversion rules used in this process, and will produce a new dataset format as presented in Table 3 below.

Table 2: Dataset Conversion Rules

OLD COLUMN NAME	OLD CONTENT	NEW CONTENT	NEW COLUMN NAME
DURASIVID	Time Format	Convert to seconds	PDurasi
DOMISILI	Jawa	1	IX_Domisili
	Sumatera	2	
	Kalimantan	3	
	Sulawesi	4	

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

OLD COLUMN NAME	OLD CONTENT	NEW CONTENT	NEW COLUMN NAME
	Bali	5	
	Papua	6	
RELEVAN_TM	RELEVAN	1	IX_Relevant
	TIDAK	0	
GENDER	L	1	IX_Sex
	P	0	
USIA	Date Format	Convert to months	BUsia

Table 3: Example dataset after data transformation

BUsia	IX_Sex	IX_Domisili	Jml Subsc	PDurasi	Jml Komen	Jml View	Jml Like	Jml DisLike	Iterasi Ke	IX_Relevant
9458	1	5	23800	19	93	101265	787	12	2	1
10897	1	5	4360	60	74	781676	2922	4	1	0
10897	1	5	78100	106	1157	6551537	14982	421	1	0
10897	1	5	207000	206	901	4658439	16110	137	1	0
9458	1	5	1240000	36	218	115046	1991	52	2	1
10897	1	5	91900	191	262	1416087	8053	115	1	0
9458	1	5	627000	19	504	1031949	5689	450	5	1
9458	1	5	30900	33	655	398866	8306	182	4	1
9458	1	5	1730000	37	970	1325954	28376	9	2	1
9458	1	5	30300	58	163	40180	702	6	2	1
10897	1	5	249000	51	616	341939	9767	7	2	1
10897	1	5	592000	59	1529	1021413	27535	512	1	1
10897	1	5	229000	26	597	478270	10445	47	4	1
9458	1	5	1190000	29	1336	1898040	14398	288	4	1

Determining Dependent and Independent Variables

Dependent variables are variables used to predict or explain categorical dependent variables such as binary, which can be continuous, discrete, ordinal, or nominal data, and play a role in forming a model that estimates the log odds (logit) of the outcome(Hosmer Jr. et al., 2013). From table 3, the independent variables are the columns: 'BUsia', 'IX_Sex', 'IX_Domisili', 'JmlSubsc', 'PDurasi', 'JmlKomen', 'JmlView', 'JmlLike', 'JmlDisLike', 'IterasiKe'. At the same time, the dependent variable is the 'IX_RELEVANT' column.

Outlier Dataset Analysis with Interquartile Range (IQR)

Before analyzing the dataset with logistic regression, outliers are first identified in the dataset. Outliers are observations of random samples from a population that are at an abnormal distance from other values(Barnett & Lewis, 1994). Once outliers are identified, data that falls

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

into the category will be removed from the dataset. This aims to improve the accuracy of the model resulting from dataset processing. This process uses the Interquartile Range (IQR) method. IQR is one of the most commonly used statistical techniques to identify outliers in numeric data(Tukey, 1977). The way IQR works is by calculating the difference between the third quartile (Q3) and the first quartile (Q1) which describes the middle 50% range of the data distribution. Mathematically, the standard IQR formula is written as :

$$\mathbf{IQR = Q3 - Q1}$$

Data values are considered outliers if they are outside the range determined by:

$$\mathbf{Lower\ Limit = Q1 - 1{,}5 \times IQR}$$

$$\mathbf{Upper\ Limit = Q3 + 1{,}5 \times IQR}$$

Thus, any value smaller than the lower limit or larger than the upper limit is classified as an outlier. This approach is non-parametric, so it does not rely on the assumption of a normal distribution and is more resistant to distortion due to extreme values. The process of identifying and removing outlier data is carried out using the Python programming language. From this process, the results obtained from the initial dataset of 1250 data, 628 records were identified as outliers which were then removed from the dataset, leaving 622 records in the dataset.

Logistic Regression Analytic

Analysis of 622 records of the dataset that has been cleaned from outliers using logistic regression with the help of Python programming language coding. This process produces complete model parameter estimates and provides complete statistical inference results, including regression coefficients, p-values, and odds ratios. This model aims to identify the factors that are statistically significant in influencing content relevance (IX_RELEVAN) and to provide stable estimates, especially for dependent variables that have a strong relationship with the target variable.

From Figure 2, which presents the results of logistic regression processing using Python programming, it is evident that the logistic regression modeling yields statistically significant results. This is indicated by the very low p-value (2.657e-05), well below the 0.05 threshold. The dataset demonstrates a statistical relationship between the independent and dependent variables. Based on the odds ratios and confidence intervals, it can be observed that each additional subscriber slightly reduces the odds of content being relevant. Similarly, an increase in dislikes also slightly decreases relevance. However, the variable IterasiKe (iteration count) shows a potential to increase content relevance by 10% to 45%, with an

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

odds ratio of 22%. Therefore, IterasiKe is identified as the most influential variable affecting the relevance of recommended content themes.

Figure 2: Results of executing the logistic regression Python code

```
                           Logit Regression Results
==============================================================================
Dep. Variable:              IX_RELEVAN   No. Observations:                  622
Model:                           Logit   Df Residuals:                      613
Method:                            MLE   Df Model:                            8
Date:                 Fri, 26 Dec 2025   Pseudo R-squ.:                 0.07005
Time:                         10:44:21   Log-Likelihood:                -232.41
converged:                        True   LL-Null:                       -249.92
Covariance Type:             nonrobust   LLR p-value:                 2.657e-05
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.6545      0.517      1.266      0.206      -0.359       1.668
BUsia       4.348e-05    2.1e-05      2.066      0.039    2.24e-06    8.47e-05
JmlSubsc   -2.238e-07   7.59e-08     -2.948      0.003   -3.73e-07    -7.5e-08
PDurasi        0.0003      0.000      1.618      0.106   -5.43e-05       0.001
JmlKomen       0.0002      0.000      0.530      0.596      -0.001       0.001
JmlView    -1.665e-06   8.77e-07     -1.899      0.058   -3.38e-06    5.37e-08
JmlLike     8.785e-05      0.000      0.836      0.403      -0.000       0.000
JmlDisLike    -0.0158      0.011     -1.463      0.144      -0.037       0.005
IterasiKe      0.1993      0.088      2.265      0.024       0.027       0.372
==============================================================================


Odds Ratios dan Interval Kepercayaan:
                  OR  Lower CI  Upper CI
const       1.924206  0.698257  5.302587
BUsia       1.000043  1.000002  1.000085
JmlSubsc    1.000000  1.000000  1.000000
PDurasi     1.000257  0.999946  1.000568
JmlKomen    1.000199  0.999462  1.000937
JmlView     0.999998  0.999997  1.000000
JmlLike     1.000088  0.999882  1.000294
JmlDisLike  0.984362  0.963789  1.005373
IterasiKe   1.220549  1.027168  1.450337
```

Model Testing : Hosmer-Lemeshow Test

After the analysis is done with logistic regression, the next stage is to test the suitability of the model. This is an important step to assess the extent to which the model built can represent the observed data. The testing method that will be used in this study is the Hosmer-Lemeshow Test. This method is designed to test whether the probability prediction distribution generated by the logistic model is consistent with the distribution of actual results from the data. The Hosmer-Lemeshow test is able to provide validation of whether the logistic regression model used is able to accurately describe the relationship between the independent and dependent variables. This test method provides protection against overfitting and helps ensure that the selected model truly reflects the patterns in the data(Hosmer Jr. et al., 2013). This test will compare the

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

predicted probabilities generated by the model and the actual events in the data, by dividing observations into several groups based on the predicted values. The basic formula is:

$$\chi^2 = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g(1 - \hat{p}_g)}$$

where :

- Og = number of actual observations in group g
- Eg = number of observations predicted by the model in group g
- p^g = average predicted probability in group g
- G = number of groups (usually 10)

The Hosmer-Lemeshow test process is carried out with the help of a Python program code whose execution results will produce output as in Figure 3. The test results shown in Figure 3 indicate that the Chi-square statistic is 2.636 with 8 degrees of freedom, resulting in a p-value of 0.955. Since this p-value is much greater than the significance threshold of 0.05, it can be concluded that there is no significant difference between the values predicted by the model and the actual values observed in the data. Therefore, the null hypothesis (H_0) (that the model fits the data), cannot be rejected. Additionally, based on the group table, the differences between the observed and expected values for each decile are relatively small and consistent. This indicates that the model has good capability in mapping the probability of events in accordance with the actual distribution of the data. Thus, the Hosmer-Lemeshow test results confirm that the logistic regression model built in this study has a good goodness-of-fit, meaning that the model is able to represent the data pattern accurately and does not exhibit symptoms of overfitting or underfitting.

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

Figure 3: Results of executing the hosmer-lemeshow test python code

```
=== Hosmer-Lemeshow Test ===
Chi-square Statistic: 2.6361
Degrees of Freedom: 8
P-value: 0.9551

Grouped Table:
                 observed  total   expected  non_observed  non_expected
decile
(0.471, 0.754]         13     19  12.920771             6      6.079229
(0.754, 0.806]         16     19  14.889451             3      4.110549
(0.806, 0.842]         14     18  14.873537             4      3.126463
(0.842, 0.868]         17     19  16.279422             2      2.720578
(0.868, 0.885]         16     19  16.651458             3      2.348542
(0.885, 0.897]         17     18  16.033971             1      1.966029
(0.897, 0.908]         16     19  17.142911             3      1.857089
(0.908, 0.919]         17     19  17.356076             2      1.643924
(0.919, 0.93]          17     18  16.651794             1      1.348206
(0.93, 0.943]          18     19  17.807023             1      1.192977
```

Logit Linearity Testing

Logit Linearity Testing aims to ensure that the relationship between continuous predictor (independent) variables and the logit of the probability of an event is linear. In logistic regression, the relationship that occurs is linear between continuous independent variables and the logit of the probability of an outcome (Hosmer Jr. et al., 2013). Logit is a logarithmic transformation of the odds, which is formulated as:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

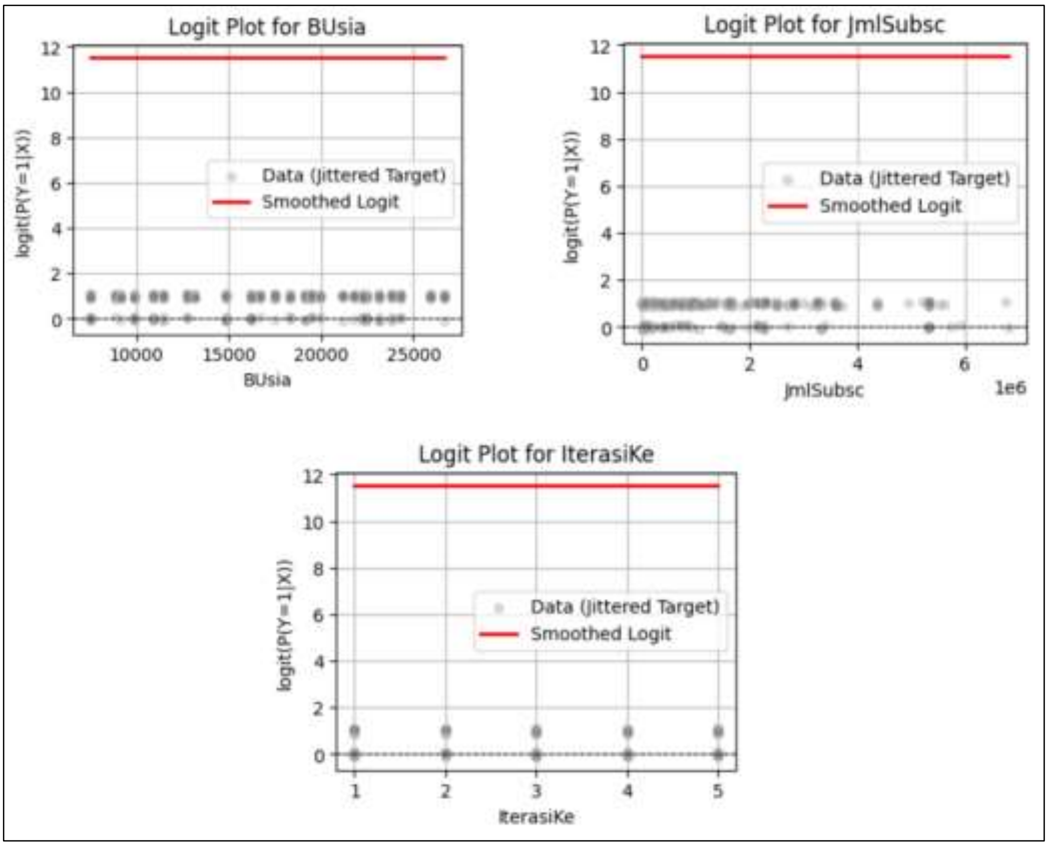
wnere:

- p is the probability of an event,
- p/(1 – p) is called the odds or odds ratio.
- log is the natural logarithm (base e).

To produce valid coefficient estimates, the logistic model assumes that each continuous predictor variable has a linear relationship to logit(p), not to p itself. If this assumption is violated (e.g., a non-linear relationship), the model may produce biased estimates, incorrect interpretations, or poor fit.

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

Figure 4: Results of executing the Python code for testing Logit Linearity



Based on Figure 4, which presents the logit plots from the logit-linearity test, the smoothed logit curves do not show a strong curvilinear pattern for the variables BUsia, JmlSubsc, and IterasiKe. Visually, this indicates that the relationship between the predictors and the log-odds of the response probability can be regarded as adequately linear, meaning that the logit-linearity assumption for the logistic regression model is satisfied for these variables. With this assumption met, the estimated logit coefficients can be interpreted appropriately without strong indications of distortion due to nonlinearity.

Variance Inflation Factor (VIF) Testing

The next stage is to carry out multicollinearity testing. Multicollinearity is a condition in regression analysis where two or more independent variables exhibit a very strong linear relationship with each other. When this occurs, the regression model struggles to estimate the individual influence of each variable on the dependent variable because the information carried by these variables becomes

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

redundant or overlapping. Multicollinearity can be detected through a correlation-based approach among independent variables, and its presence can seriously affect the reliability of regression models used for data-driven decision-making(Farrar & Glauber, 1967). This study applies the Variance Inflation Factor (VIF) method to test for multicollinearity using the Python programming language. VIF measures the extent to which the variance of a regression coefficient increases due to correlation with other variables in the model(Gujarati, 1978). The results of the Python program execution on the Multicollinearity test displayed in Figure 5 show that the VIF value is less than 5, this indicates that there is no significant multicollinearity. If the VIF value is between 5 and 10, it means moderate multicollinearity, and a VIF value above 10 means high multicollinearity(Hair et al., 2010). It can be concluded that the variables JmlLike, JmlView, JmlDisLike, JmlKomen, JmlSubsc, PDurasi, BUsia, and IterasiKe which have VIF values < 5 do not indicate significant multicollinearity so that the independent variables in this model are proven to be mutually independent, not highly correlated with each other.

Figure 5: Test linearity with VIF result from the Python code

```
   Variable        VIF
      const  15.791667
   JmlSubsc   1.003820
 IterasiKe   1.002793
      BUsia   1.001058
```

Forecasting Model

From the test conducted, the results show that the relationship between the independent variable IterasiKe and the dependent variable IX_RELEVAN is very strong. Therefore, the next stage is to create a mathematical model to forecast the trend that occurs if the IterasiKe process is continued. For this reason, the dataset will be used as an aggregate table model with the following IterasiKe and IX_RELEVAN variables :

Table 4: Aggregate dataset table

ITERATION TO	RELEVANT	NOT	TOTAL
1	67	27	94
2	110	14	124
3	109	15	124
4	124	10	134
5	126	20	146
TOTAL	**536**	**86**	**622**

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

To see the trend of content relevance, table 4 is presented in the form of a bar chart as shown in Figure 6 below.

Figure 6: Content relevance trend based on iterations

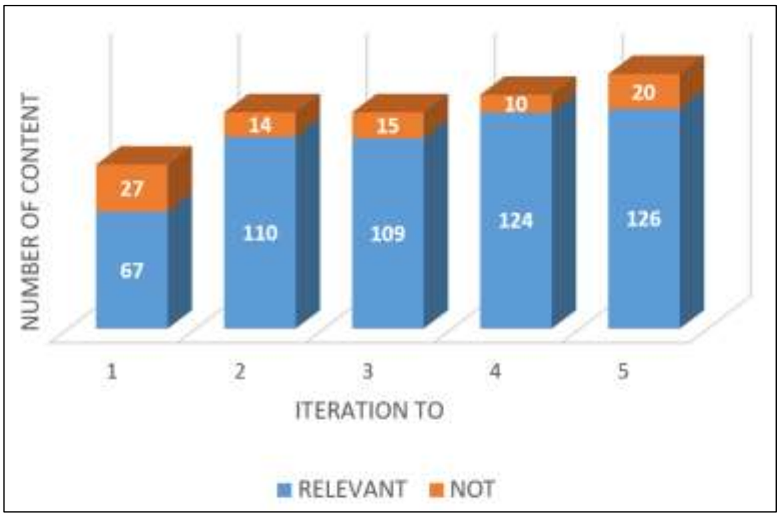


Table 6 summarizes the number of relevant and non-relevant contents obtained from five iterations of the search process. Figure 6 further shows an increase in the number of YouTube contents that are relevant to users' viewing histories as the iterations progress. Each iteration represents one observation point in an ordered time series; thus, the five iterations can be treated as five sequential observations within a single time series. To estimate the tendency in subsequent iterations (after the fifth iteration) and to capture short-term patterns in the iterative process, this study applies mathematical modelling. Statistically, having more observations than parameters is always preferred; however, when random variation is relatively low, basic estimation can be performed with a minimal number of observations. Theoretically, a regression line can be estimated with only three observations (Rob J. Hyndman & Kostenko, 2007). With five chronologically ordered observations, these data provide an adequate basis for analysing changes in the number of relevant contents across iterations and for projecting values in the next iteration as an indication of a short-term trend.

Based on Table 6 and Figure 6, the proportion of relevant recommendations increases from iteration 1 to iteration 2, drops by one percentage point in iteration 3, and

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

then rises again in iterations 4 and 5. This pattern indicates a strengthening of personalization alongside the accumulation of viewing history and engagement signals, such that the system increasingly stabilizes preferences as they are read from user data. Within the machine habitus framework, this increase can be interpreted as a reinforcement of the recommendation system's predispositions that "settle" from usage traces, while at the user level it has the potential to shape a digital habitus through repeated exposures that guide how users choose and interpret information. This fluctuation suggests that algorithmic curation is not entirely linear or monotonic, which may be due to random variation in a limited dataset or to the presence of exploration/diversification mechanisms in the recommendation system. Despite the decline in iteration 3, the level of relevance remains substantially higher than in the initial iteration, so overall the findings are consistent with explanations of filter bubbles and echo chambers as gradual processes formed through reinforced personalization. To examine the short-term tendency of this dynamic, one estimation technique used is Exponential Smoothing (ES), a time-series forecasting method that smooths historical data using exponentially decreasing weights, giving greater weight to more recent observations. This study applies two ES variants, Simple Exponential Smoothing (SES) and Holt's Linear Trend Model (HLTM), implemented using Python code. SES is appropriate when the data show no clear trend or seasonality and is therefore often applied to stationary or stable series (R.J. Hyndman & Athanasopoulos, 2018). Mathematically, the basic SES formula is expressed as:

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

where :

- Ft+1 is the forecast value for time t + 1,
- Yt is the actual value at time t,
- Ft is the previous forecast value at time t,
- α is the smoothing constant, 0< α <10.

Holt's Linear Trend Model often referred to as Double Exponential Smoothing, is a development of the SES method. This model is used to predict time series data that has a linear trend pattern, either a gradual increase or a decrease. By considering two main components, namely level, and trend, this model is able to provide more accurate predictions than SES when the data shows a consistent pattern of changes in direction. The basic formula of Holt's Linear Trend Model is as follows :

$$l_t = \alpha \cdot y_t + (1-\alpha) \cdot (l_{t-1} + b_{t-1})$$

$$b_t = \beta \cdot (l_t - l_{t-1}) + (1-\beta) \cdot b_{t-1}$$

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

where:

- *lt* : Estimated level in the period *t*
- *bt* : Estimated trend in the period *t*
- *yt* : Actual value in the period *t*
- *α* : Smoothing level parameters (0 ≤ α ≤ 1)
- *β* : Smoothing trend parameters (0 ≤ β ≤ 1)

With the combination of these two parameters, the model can make predictions taking into account the current level plus an estimate of the continuing trend.:

$$\hat{y}_{t+h} = l_t + h \cdot b_t$$

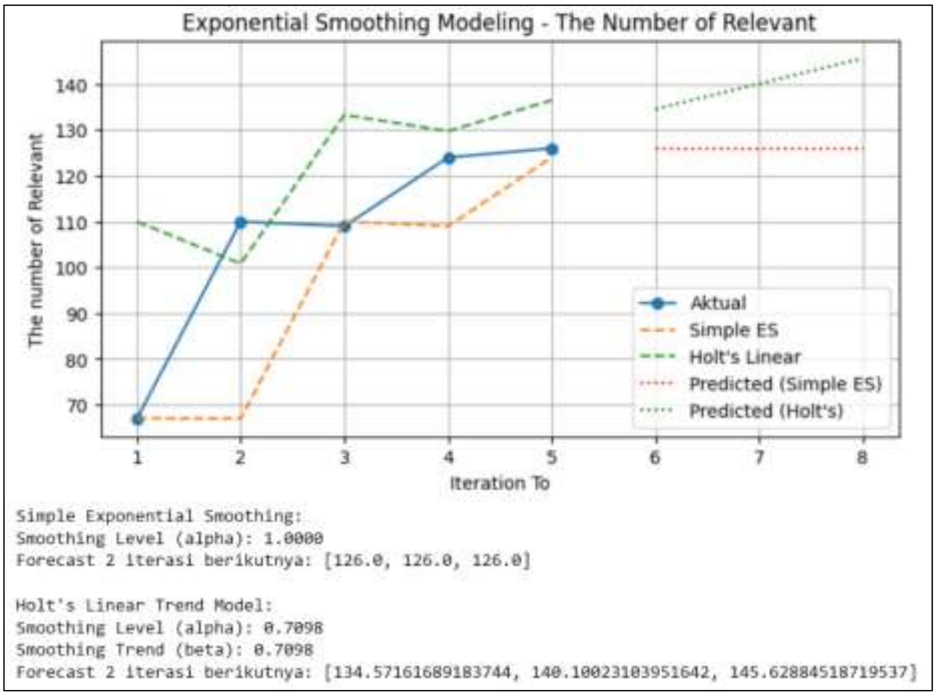
where :

- *h* : The number of periods into the future you want to predict
- *lt* and *bt* : Last estimated levels and trends

The results of executing the Python program code in creating a forecasting model using the Simple Exponential Smoothing method and Holt's Linear Trend Model displayed in Figure 7 show that in the initial stage of analysis using the SES method which uses an alpha value of 0.5 to model time series data, it gives the same weight between current data and historical data. This indicates that the SES model is able to follow the pattern of the actual data more closely; however, SES has limitations because it only accounts for the level component and does not explicitly consider trends. The SES forecast line tends to lag behind the actual data. This is due to the fact that SES is not designed to capture clear trends in the data. The SES forecast remains constant, around 126.0**,** even though the actual data shows a significant upward trend. This demonstrates that SES is less suitable for making realistic projections in the context of data with an obvious trend.

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

Figure 7: Results of executing the Exponential Smoothing Python



An alternative model used in this study is Holt's Linear Trend Model (HLTM), which is an extension of SES. HLTM is capable of modeling data with stable upward or downward trends. In this implementation, the alpha and beta parameters are both set to 0.7098. The results of HLTM are considerably more significant compared to SES, as shown by the following points:

1. The HLTM forecast line (illustrated in green) successfully captures the upward pattern in the actual data. This model is closer to the actual data than SES because HLTM explicitly incorporates trends into its estimation.
2. The forecasted values from HLTM also display a logical upward trend, with gradually increasing projections: 134.571, 140.1, and 145.629 for iterations 6, 7, and 8, respectively. This reflects the model's ability to project future trends more accurately.

The results of the logistic regression and forecasting model show that the IterationNumber variable is the dominant factor in determining the relevance of content recommendations, while other variables such as subscribers and dislikes only have a marginal influence. This finding confirms the initial hypothesis that the longer the user's interaction history, the greater the likelihood of YouTube displaying content deemed relevant. This iterative process aligns with the

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

concept of an echo chamber, where the algorithm narrows the recommendation space and repeats similar themes, and the algorithm functions as a digital gatekeeper that determines what is seen and what is hidden (Gillespie, 2014). The trend of increasing relevance shown by Holt's Linear Trend model indicates that the algorithm is not neutral but actively optimizes user preferences, with the consequence of reducing informational diversity.

The implications of these findings affirm that algorithmic logic has significant socio-religious impacts. The filter bubble is seen to limit the audience's access to diverse religious views, and the results of this study reinforce this in the context of YouTube (Harambam et al., 2018). In the Indonesian context, Islamic practices are increasingly mediated by digital media (Slama, 2018), while visibility on platforms like Instagram and YouTube shapes new patterns of influence and authority for digital preachers (*dai*) (Nisa, 2018; Solahudin & Fakhruroji, 2019). This condition confirms that figures who are adaptive to the algorithm are more advantaged than traditional authorities based on a scholarly lineage (*sanad*). Thus, the algorithm is not merely a technical mechanism for content distribution but also a cultural actor that plays a role in the formation of religious authority and the knowledge structure of the community in the digital space.

CONCLUSION

The logistic regression results show that iteration is the most influential variable affecting content relevance. Accordingly, the mechanism of exposure narrowing can be traced more systematically. Using an experimental design based on 25 YouTube accounts and 1,250 sampled videos, the study models recommendation dynamics across search cycles and quantitatively tests the determinants of recommendation relevance through logistic regression, supported by diagnostic validation tests (Hosmer–Lemeshow, logit-linearity, and multicollinearity) and forecasting models.

The results show that the Iteration variable is the most dominant factor influencing recommendation relevance, while other variables such as subscriber count and dislikes exert only marginal effects. The logistic regression model is validated, and Holt's Linear Trend modelling indicates an increasing tendency in content relevance in subsequent iterations. These findings confirm that YouTube's algorithm operates iteratively by narrowing the recommendation space in line with viewing history, thereby reinforcing filter-bubble and echo-chamber dynamics.

ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

The implications are significant for the socio-religious context: (1) religious authority may shift from scholarly lineage and traditional institutions toward figures who can optimize algorithmic performance; (2) the distribution of da'wah content is not neutral because the algorithm tends to privilege popular and emotionally charged content over reflective-educational content; and (3) recommendation homogeneity reduces the diversity of Islamic perspectives accessible to users. Thus, the algorithm functions not only as a technical mechanism for content distribution but also as a cultural actor shaping religious authority and the religious landscape in digital spaces.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

M. Didik R. Wahyudi : Writing Conceptual Draft, Methodology, Data curation, Supervision, Reviewing and Editing. **Noorhaidi Hasan** : Supervisions, Reviewing. **Agung Fatwanto** : Methodology, Supervisions, Reviewing.

DATASET

The dataset is available for download **here**

DECLARATION OF COMPETING INTEREST

We certify that there is no conflict of interest with any financial, personal, or other relationships with other people or organizations related to the material discussed in the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank all parties who were involved in the research. Many thanks are also addressed to the reviewers and editor of the Profetik: Jurnal Komunikasi.

REFERENCES

Airoldi, M. (2022). *Machine habitus: Toward a sociology of algorithms*. Cambridge Medford (Mass.): Polity press.

Allison, P. D. (2012). *Logistic Regression Using SAS: Theory and Application* (2nd ed.). SAS Press.

APJII. (2024). *Survei Penetrasi Internet Indonesia 2024*.

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*.

Bishop, S. (2018). Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. *Convergence*, *24*(1), 69–84. https://doi.org/10.1177/1354856517736978

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on

Facebook. *New Media & Society*, *14*(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Cook, T. D. ; C., D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin.

Covington, P., Adams, J., & Sargin, E. (2016, September). *Deep neural networks for youtube recommendations*. 191–198. Association for Computing Machinery, Inc. https://doi.org/10.1145/2959100.2959190

Creswell, J., W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE.

Davidson, J., Liebald, B., Liu, J., Nandy, P., & Van Vleet, T. (2010). *The YouTube video recommendation system*. 293–296. https://doi.org/10.1145/1864708.1864770

Delmonaco, D., Mayworm, S., Thach, H., Guberman, J., Augusta, A., & Haimson, O. L. (2024). “What are you doing, TikTok?”: How Marginalized Social Media Users Perceive, Theorize, and “Prove” Shadowbanning. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1). https://doi.org/10.1145/3637431

Dujeancourt, E., & Garz, M. (2023). The effects of algorithmic content selection on user engagement with news on Twitter. *Information Society*, *39*(5), 263–281. https://doi.org/10.1080/01972243.2023.2230471

Engineering, T. (2023, March 31). Twitter’s recommendation algorithm. Retrieved from X Engineering Blog website: https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., … Sandvig, C. (2015). “I always assumed that I wasn’t really that close to [her]”: Reasoning about Invisible Algorithms in News Feeds. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 153–162. Seoul Republic of Korea: ACM. https://doi.org/10.1145/2702123.2702556

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*, *49*(1), 92–107.

ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

https://doi.org/10.2307/1937887

Fourcade, M., & Johns, F. (2020). Loops, ladders and links: The recursivity of social and machine learning. *Theory and Society*, *49*(5–6), 803–832. https://doi.org/10.1007/s11186-020-09409-x

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press. https://doi.org/10.7551/mitpress/9780262525374.001.0001

Griffiths, S., Harris, E. A., Whitehead, G., Angelopoulos, F., Stone, B., Grey, W., & Dennis, S. (2024). Does TikTok contribute to eating disorders? A comparison of the TikTok algorithms belonging to individuals with eating disorders versus healthy controls. *Body Image*, *51*, 101807–101807. https://doi.org/10.1016/j.bodyim.2024.101807

Gujarati, D. N. (1978). *Basic econometrics*. McGraw-Hill.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson Prentice Hall.

Harambam, J., Helberger, N., & Van Hoboken, J. (2018). Democratizing algorithmic news recommendations: How to materialize voice in a technologically saturated media ecosystem. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133). https://doi.org/10.1098/rsta.2018.0088

Hindarto, I. H. (2022). INVESTIGATING HOW THE NATIONAL ONLINE MEDIA REPORTED THE LGBT COMMUNITY DURING THE COVID-19 PANDEMIC. *Profetik: Jurnal Komunikasi*, *15*(2), 208–227. https://doi.org/10.14421/pjk.v15i2.2577

Hosmer Jr., D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.

Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts.

Hyndman, Rob J., & Kostenko, A. V. (2007). Minimum Sample Size Requirements for Seasonal Forecasting Models. *FORESIGHT*, (6), 12–15.

Kingsley, S., Sinha, P., Wang, C., Eslami, M., & Hong, J. I.

ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

(2022). "Give Everybody [..] a Little Bit More Equity": Content Creator Perspectives and Responses to the Algorithmic Demonetization of Content Associated with Disadvantaged Groups. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2). https://doi.org/10.1145/3555149

Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology* (Fourth). Los Angeles: SAGE Publications Ltd.

Larsson, A. O. (2020). Winning and losing on social media: Comparing viral political posts across platforms. *Convergence*, *26*(3), 639–657. https://doi.org/10.1177/1354856518821589

Meta. (2025, June 11). Our approach to Facebook Feed ranking. Meta Transparency Center. Retrieved from https://transparency.meta.com/features/facebook-feed-ranking/

Narayanan, A. (2023). *Understanding Social Media Recommendation Algorithms*.

Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, *20*(11), 4275–4292. https://doi.org/10.1177/1461444818769694

Nisa, E. F. (2018). Social media and the birth of an Islamic social movement: ODOJ (One Day One Juz) in contemporary Indonesia. *Indonesia and the Malay World*, *46*(134), 24–43. https://doi.org/10.1080/13639811.2017.1416758

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press.

Prianto, A. L., Abdillah, A., Syukri, S., Muhammad, F., & Yama, A. (2021). COMBATING INFODEMIC COVID-19: GOVERNMENT RESPONSE AGAINST FAKE NEWS ON SOCIAL MEDIA. *Profetik: Jurnal Komunikasi*, *14*(2), 255–275. https://doi.org/10.14421/pjk.v14i2.2386

Rodilosso, E. (2024). Filter Bubbles and the Unfeeling: How AI for Social Media Can Foster Extremism and Polarization. *Philosophy and Technology*, *37*(2). https://doi.org/10.1007/s13347-024-00758-4

Romele, A. (2023). *Digital Habitus: A Critique of the Imaginaries of Artificial Intelligence* (1st ed.). New York: Routledge.

Profetik Jurnal Komunikasi
ISSN: 1979-2522 (Print), ISSN:2549-0168 (Online)
DOI: https://doi.org/10.14421/pjk.v18i2.3622

https://doi.org/10.4324/9781003400479

Saura, J. R., Palacios-Marqués, D., & Iturricha-Fernández, A. (2021). Ethical design in social media: Assessing the main performance measurements of user online behavior modification. *Journal of Business Research*, *129*, 271–281. https://doi.org/10.1016/j.jbusres.2021.03.001

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (Nachdr.). Belmont, CA: Wadsworth Cengage Learning.

Slama, M. (b). (2018). Practising Islam Through Social Media in Indonesia. *Indonesia and the Malay World*, *46*(134), 1–4. https://doi.org/10.1080/13639811.2018.1416798

Solahudin, D., & Fakhruroji, M. (2019). Internet and Islamic Learning Practices in Indonesia: Social Media, Religious Populism, and Religious Authority. *Religions*, *11*(1), 19. https://doi.org/10.3390/rel11010019

TikTok. (2020, June 18). How TikTok recommends videos #ForYou. Retrieved from https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

Wang, X., Zhang, Y., & Yamasaki, T. (2019, October). *User-aware folk popularity rank: User-popularity-based tag recommendation that can enhance social popularity*. 1970–1978. Association for Computing Machinery, Inc. https://doi.org/10.1145/3343031.3350920

Wulandari, T. D., & Nuraniwati, T. (2023). THE INVISIBLE HAND PROGRAM BY PT PERTAMINA MOR IV TO EDUCATE AND INCREASE PUBLIC AWARENESS ON SUBSIDIZED 3KG LPG. *Profetik: Jurnal Komunikasi*, *16*(1), 123–138. https://doi.org/10.14421/pjk.v16i1.2277