



PREDIKSI JUMLAH PENONTON VIDEO YOUTUBE MENGUNAKAN MODEL DEEP NEURAL NETWORK (DNN)

**Dwin Indrawan¹, Sena Ramadona Cakrawijaya², Bagus
Dwi Wicaksono³, Erni⁴, Windu Gata⁵**

Magister Ilmu Komputer¹, Magister Ilmu Komputer², Magister Ilmu
Komputer³, Magister Ilmu Komputer⁴, Magister Ilmu Komputer⁵

Teknologi Informasi¹, Teknologi Informasi², Teknologi Informasi³,
Teknologi Informasi⁴, Teknologi Informasi⁵

Universitas Nusa Mandiri¹, Universitas Nusa Mandiri², Universitas Nusa
Mandiri³, Universitas Nusa Mandiri⁴, Universitas Nusa Mandiri⁵

14002435@nusamandiri.ac.id¹, 14002412@nusamandiri.ac.id²,
14002428@nusamandiri.ac.id³, 14002422@nusamandiri.ac.id⁴,
windu@nusamandiri.ac.id⁵

Received: May 07, 2021. **Revised:** May 18, 2021. **Accepted:** June 02, 2021.

Published: June 20, 2021. **Issue Period:** Vol.5 No.1 (2021), Pp.94-98

Abstrak: Penonton Youtube menjadi faktor utama kesuksesan seorang pembuat konten youtube. Masalahnya adalah bagaimana cara memprediksi jumlah penonton konten youtube untuk menentukan keberhasilan konten atau menjadikannya topik video yang sedang populer. Mesin atau komputer dapat memprediksi jumlah penonton yang menggunakan metode ANN. DNN adalah salah satu model ANN yang dapat memprediksi kumpulan data. Penelitian ini membandingkan tingkat keberhasilan dalam memprediksi jumlah penonton konten youtube dari dataset Youtube API. Hasil prediksi model DNN memiliki tingkat keunggulan yang lebih tinggi dibandingkan dengan penggunaan metode tradisional Linear Regression dan Naive Bayes.

Kata kunci: machine learning; dnn; ann; r2 score; youtube

Abstract: Youtube viewer becomes a major factor in a content creator's success. The problem is how can predict the number of viewers of a youtube content to determine the success of a content or make it a trending video topic. A machine or computer can predict the number of viewers using the ANN method. DNN is one of ANN models that can predict a dataset. This study compared the success rate in predicting the number of viewers of youtube content from Youtube API dataset. DNN model prediction results have a higher level of excellence compared to the use of traditional methods Linear Regression and Naive Bayes.

Keywords: machine learning; dnn; ann; prediction; r2 score; youtube

I. PENDAHULUAN

Youtuber adalah seseorang atau sekelompok orang yang memiliki akun atau channel di platform YouTube, mereka membuat konten dan mengunggah video yang mereka hasilkan ke akun atau saluran yang

DOI: 10.52362/jisicom.v5i1.463



Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

mereka miliki di platform YouTube, mereka juga mendapatkan ketenaran dan mendapatkan pengunjung setia (pelanggan) karena berbagai alasan[1]. Salah satu media sosial yang paling populer dan berkembang adalah YouTube, menurut Alexa pada tahun 2021, YouTube adalah Situs populer Kedua secara Global[2].

Saat ini, YouTube digunakan oleh sebagian orang untuk mencari penghasilan dan menciptakan profesi baru yaitu YouTuber. Semakin banyak orang termotivasi untuk menjadi YouTuber daripada memiliki menjadi karyawan, inilah alasan yang membuat banyak orang menggunakan YouTube[3].

Saat ini platform video bersaing untuk memberikan layanan yang lebih baik, Youtube sebagai platform pemimpin pasar berhasil berbagi keuntungan dengan pembuat konten mereka. Faktor penting untuk menentukan nilai konten adalah penonton konten video. Setiap pembuat konten berlomba untuk mengejar pemirsa sebanyak mungkin dalam mengejar nilai komersial dan popularitas[4].

Riwayat perilaku pengguna di Youtube secara hubungan erat sulit diprediksi karena sparsitas dan berbagai faktor eksternal yang tidak dapat diamati[5]. Hal inilah yang membuat prediksi jumlah penonton relatif sulit ditebak meskipun dari segi kualitas suatu konten memiliki nilai yang sangat baik namun belum tentu jumlah penonton berbanding lurus dengan itu.

Riwayat menonton video pengguna mungkin berisi outlier yang memengaruhi hasil rekomendasi dan sistem periklanan[6]. Penghasilan pembuat konten sangat erat kaitannya dengan layanan iklan komersial yang dilayani oleh Youtube.

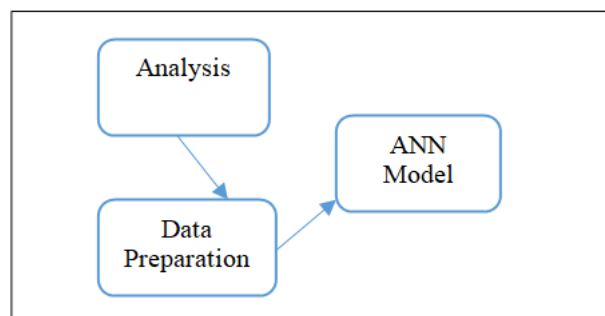
Jangkauan video YouTube melingkupi menyukai video (Like), mengomentari (Comment) di video di YouTube dan melihat video (View)[7]. Berbagi video (Share) dari situs YouTube dan Subscribing (Subscribe) Channel YouTube[8].

Pengguna YouTube Aktif diatas 2 miliar yang mengunjungi YouTube setiap bulan dan setiap hari orang menonton lebih dari satu miliar jam video dan menghasilkan miliaran penayangan[9]. Pada tulisan ini kumpulan data diperoleh dari YouTube API[10] yang dihasilkan dan dipublikasikan di kaggle.com. Kumpulan data mencakup data analitik statistik dari konten trending Youtube pada periode 2020, dengan jumlah 7600 set data.

Deep learning (juga dikenal sebagai *deep structured learning*) adalah bagian dari keluarga metode pembelajaran mesin yang lebih luas berdasarkan ANN (artificial neural networks) dengan *representation learning* [11]. Penampil prediksi Youtube menggunakan model DNN (*Deep neural network*) memiliki hasil yang lebih baik dibandingkan dengan metode tradisional seperti KNN[12] sebagai metode regretor atau Naive Bayes. Hasil DNN untuk prediksi jumlah penonton adalah 92% dari skor R2.

II. METODE DAN MATERI

Gambar. 1 menampilkan alur metode penelitian dalam makalah ini, mulai dari analisis masalah dan kumpulan data yang dipilih, kemudian memasuki tahap persiapan data sehingga set data siap diproses dan akhirnya membuat model dari ANN menggunakan DNN[13].



Gambar 1. Metode Penelitian

Penelitian ini dibagi menjadi 2 tahap, tahap pertama studi literatur yang menghasilkan makalah, tahap kedua adalah pemodelan berisi tentang Analisis Dataset (menganalisis data yang dikumpulkan), Persiapan Data, Data Spilting (pelatihan dan pengujian), model Machine Learning.

DOI: 10.52362/jisicom.v5i1.463



Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

2.1. Analisis

Pertama, Sumber data dalam penelitian ini dikumpulkan menggunakan API YouTube. Kumpulan data ini mencakup catatan harian dari video YouTube teratas yang sedang tren hingga 200 video populer per hari pada tahun 2020 khususnya untuk wilayah negara Amerika serikat. YouTube menyimpan daftar video populer teratas di platform Youtube. Menurut majalah Variety, "Untuk menentukan video yang sedang tren tahun ini, YouTube menggunakan kombinasi faktor termasuk mengukur interaksi pengguna (jumlah penayangan, berbagi, komentar, dan suka). Data mencakup judul video, judul saluran, waktu publikasi, tag, suka dan tidak suka, deskripsi, identitas saluran, Judul saluran, identitas kategori, tanggal tren, jumlah tampilan, jumlah komentar, tautan thumbnail, komentar dinonaktifkan, peringkat dinonaktifkan, deskripsi dan jumlah komentar .

Untuk informasi selengkapnya tentang kolom tertentu dalam kumpulan data, lihat metadata kolom[14] dengan jumlah data 7600 baris dan 16 kolom seperti yang terlihat di Gambar. 2.

dataYT.head()															
	video_id	title	publishedAt	channelId	channelTitle	categoryId	trending_date	tags							
0	3C66w5Z0ks	I ASKED HER TO BE MY GIRLFRIEND...	2020-08-11T19:20:14Z	UCvRTOMP2TqY5u1xNrqA2g	Brawadis	22	2020-08-12T00:00:00Z	brawadis prank basketball skits ghost funny vi...							
1	M9Pm9AB4Mo	Apex Legends Stories from the Outlands - "Th...	2020-08-11T17:00:10Z	UC0ZV6M2THA81QT9hrVWJG3A	Apex Legends	20	2020-08-12T00:00:00Z	Apex Legends Apex Legends characters new Apex ...							
2	J78aPJ3VyNs	I left youtube for a month and THIS is what ha...	2020-08-11T16:34:06Z	UCyZpXprvISY-Sf0g4vX-m6g	jacksepticeye	24	2020-08-12T00:00:00Z	jacksepticeye funny funny meme memes jacksepti...							
3	KXln3HkpjaA	XXL 2020 Freshman Class Revealed - Official An...	2020-08-11T16:38:55Z	UCbg_UmJHjg_19S2cKaKajg	XXL	10	2020-08-12T00:00:00Z	xxl freshman xxl freshmen 2020 xxl freshman 20...							
4	VIUo6yapDbc	Ultimate DIY Home Movie Theater for The LaBran...	2020-08-11T15:10:05Z	UCDVPcEbVLQgLZXOR16jo34A	Mr. Kate	26	2020-08-12T00:00:00Z	The LaBrant Family DIY Interior Design Makeove...							

Gambar 2. Dataset

2.2. Persiapan Data

Tahap persiapan kumpulan data memiliki 3 fase yaitu (1) pembersihan data, (2) transformasi data, (3) pemilihan fitur data dan (4) pemisahan data. Dari kumpulan data di atas untuk tahap awal peneliti melakukan pembersihan data dengan menghapus data nilai NAN.

Tabel I. 3 tahap berikutnya adalah transformasi data, setiap nilai set data dalam bentuk teks dikonversi menjadi angka menggunakan 'label encoder'.

Tabel I. Transformasi Data

categoryId	View_count	likes	dislikes	Comment_count	videoId_new	channelId_new
22	1514614	156908	5855	35313	104	951
20	2381688	146739	2794	16549	599	26
24	2038853	353787	2628	40221	510	579
10	496771	23251	1856	7647	1190	643
26	1123889	45802	964	2196	809	224

Tahap ketiga adalah pemilihan fitur data yang menentukan prediktor fitur sesuai dengan fitur berpengaruh. Ada 6 fitur terpilih dan 1 fitur utama untuk menentukan prediksi jumlah penonton. Langkah terakhir adalah pemisahan data dengan porsi uji data 20% dan kereta data 80% untuk menghitung hasil menggunakan metode untuk memprediksi penayangan.

2.3. Model ANN

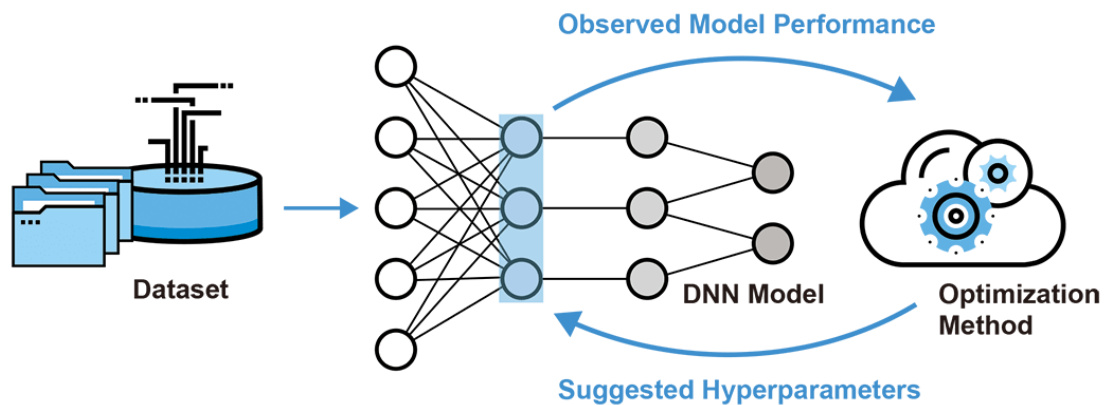
Model ANN yang kami gunakan adalah DNN dengan 3 pengoptimal sebagaimana ditunjukkan dalam Gambar. 4[15]. Untuk menghasilkan model DNN tingkat produksi, seorang peneliti harus melalui banyak langkah yang sulit dan memakan waktu, termasuk pengumpulan set data, pembersihan set data, pelabelan set

DOI: 10.52362/jisicom.v5i1.463



Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

data, augmentasi set data, konversi format set data, pemilihan model, desain model, penyetelan hiperparameter, pelatihan model, evaluasi model, dan konversi format model. Setiap langkah membutuhkan berbagai alat dan konfigurasi yang membutuhkan waktu dan upaya untuk persiapan, dan beralih di antara alat-alat ini sering membutuhkan kode tulis waktu tambahan untuk mengonversi format yang berbeda untuk digunakan dengan alat yang berbeda.



Gambar 3. DNN dengan Optimizer

III. PEMBAHASAN DAN HASIL

Setelah serangkaian uji coba dan penelitian, hasilnya telah diperoleh seperti pada tabel II di mana dari 3 tingkat pembelajaran pengoptimal, pengoptimal 'adam' adalah orang yang memberikan hasil terbaik.

Tabel II. Hasil DNN

No.	DNN				
	Optimizer	MAE	MSE	RMSE	R2
1	Adam	1061437	8122758586028	2850045	0.92
2	Adagrad	2834607	103273026078826	10162333	-0.07
3	RMSProp	2849396	104234726590183	10209540	-0.08

Sementara dibandingkan dengan metode tradisional di Tabel III yang menggunakan Linear Regression dan Naive Bayes sebagai perbandingan dalam memprediksi jumlah penonton set data yang ada.

Tabel III. Hasil Metode Tradisional

No	Traditional Method				
		MAE	MSE	RMSE	R2
1	Linear Regression	2797988	79468644494847	8914518	0.03
2	Naive Bayes	2501161	88226607552728	9392901	-0.08

DOI: 10.52362/jisicom.v5i1.463



Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Untuk prediksi penampil video Youtube kinerja terbaik adalah dengan model DNN menggunakan pengoptimal Adam karena skor R2 mendekati nilai 100%.

IV. KESIMPULAN

Dari hasil penelitian diperoleh akurasi prediksi menggunakan model Linear Regresi dengan Skor 3% R2 (0,03) yang berarti sangat jauh dari hasil yang baik dan metode Naïve Bayes dengan skor -8 % R2 (-0,08), yang berarti prediksi tren berada di arah yang berlawanan.

Dan dibandingkan dengan model ANN yang menggunakan DNN sebagai perbandingan dalam memprediksi jumlah penonton dari set data yang ada, hasil terbaik adalah oleh pengoptimal Adam dengan skor R2 92% (0,92) yang berarti kinerja terbaik. Pengoptimal lain seperti Adagrad dan RMSProp memiliki hasil negatif yang berarti tren berlawanan arah.

REFERENSI

- [1] M. Ramos-Serrano and P. Herrero-Diz, "Unboxing and brands: Youtubers phenomenon through the case study of evantubehd," *Prism. Soc.*, vol. 2016, 2016.
- [2] Alexa, "The top 500 sites on the web," 2021, 2021. <https://www.alexa.com/topsites>.
- [3] A. Gilbreath, "Trying to Understand YouTube Success," 2018, 2018. <https://longreads.com/2018/02/27/trying-to-understand-youtube-success/>.
- [4] W. Tafesse, "YouTube marketing: how marketers' video optimization practices influence video views," *Internet Res.*, vol. 30, no. 6, 2020, doi: 10.1108/INTR-10-2019-0406.
- [5] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," 2016, doi: 10.1145/2959100.2959190.
- [6] Y. J. Zhang, Z. Dong, and X. W. Meng, "Research on Personalized Advertising Recommendation Systems and Their Applications," *Jisuanji Xuebao/Chinese Journal of Computers*, vol. 44, no. 3. 2021, doi: 10.11897/SP.J.1016.2021.00531.
- [7] E. Halim, R. Anindya, and M. Hebrard, "The Impact of Motivation to Watch YouTube, Subjective Norms, Behavior Control, Information Success Model to watching YouTube Engagement," 2020 *International Conference on Information Management and Technology (ICIMTech)*. IEEE, 2020, doi: 10.1109/icimtech50083.2020.9211225.
- [8] C. Schwemmer and S. Ziewiecki, "Social Media Sellout: The Increasing Role of Product Promotion on YouTube," *Soc. Media Soc.*, vol. 4, no. 3, 2018, doi: 10.1177/2056305118786720.
- [9] YouTube, "YouTube for Press," *YouTube for Press*, 2020. .
- [10] T. Toepak, Ervin Winardo, Arwani, Issa , Afirianto, "Pembangunan Aplikasi Penyedia Informasi Lowongan Pekerjaan Menggunakan Youtube API Pada Smartphone Android," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, 2018.
- [11] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61. 2015, doi: 10.1016/j.neunet.2014.09.003.
- [12] L. M. SHavtikova, "Implementation of the kNN algorithm in the Python programming language," *TRENDS Dev. Sci. Educ.*, 2020, doi: 10.18411/ij-09-2020-05.
- [13] Manisha, S. K. Dhull, and K. K. Singh, "ECG Beat Classifiers: A Journey from ANN to DNN," in *Procedia Computer Science*, 2020, vol. 167, doi: 10.1016/j.procs.2020.03.340.
- [14] R. Sharma, "YouTube Trending Video Dataset (US Region)," *kaggle*, 2020. <https://www.kaggle.com/rsrishav/youtube-trending-video-dataset/>.
- [15] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7. 2019, doi: 10.1109/ACCESS.2019.2912200.

DOI: 10.52362/jisicom.v5i1.463



Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).