



# Klasifikasi Komentar YouTube Menggunakan Algoritma SVM: Studi Kasus Video Mitos dan Fakta Kesehatan oleh dr. Tirta

Caswita<sup>1</sup>, Rofi Irfan Miftah<sup>2</sup>, Elkin Rilvani<sup>3</sup>

<sup>1,2,3</sup>Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa

Alamat, Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Kabupaten Bekasi, Jawa Barat 17530

E-mail : [cscaswita@gmail.com](mailto:cscaswita@gmail.com)<sup>1</sup>, [rofiirfan23@gmail.com](mailto:rofiirfan23@gmail.com)<sup>2</sup>, [elkin.rilvani@pelitabangsa.ac.id](mailto:elkin.rilvani@pelitabangsa.ac.id)<sup>3</sup>

## Article Info

### Article history:

Received July 13, 2025

Revised July 15, 2025

Accepted July 23, 2025

### Keywords:

Comment Classification, Youtube, Health Education, Support Vector Machine.

## ABSTRACT

The development of digital technology has encouraged people to access health information through social media platforms such as YouTube. Health education videos by medical personnel, one of whom is Dr. Tirta, have received high attention and generated public interaction in the form of comments. This study aims to classify the sentiment of YouTube comments on health education videos using the Support Vector Machine (SVM) algorithm. Data were collected from a YouTube video titled "NETIZEN BERTANYA LAGI, dr. TIRTA MENJAWAB KEMBALI: MITOS & FAKTA KESEHATAN" with a total of 2,414 comments. The research stages include data collection using web scraping, data pre-processing (case folding, cleaning, stopword removal, stemming, tokenization), automatic labeling of comments into positive, negative, and neutral categories, feature extraction using TF-IDF, and training a classification model using SVM with a linear kernel. The evaluation results show that the model obtained an accuracy of 93.2%, a precision of 96.2%, a recall of 65.5%, and an F1-score of 67.8%. This study concludes that the SVM algorithm is effective for classifying YouTube comments on health education content and can support automatic public opinion analysis.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Article Info

### Article history:

Received July 13, 2025

Revised July 15, 2025

Accepted July 23, 2025

### Keywords:

Klasifikasi Komentar, Youtube, Edukasi Kesehatan, Support Vector Machine

## ABSTRAK

Perkembangan teknologi digital telah mendorong masyarakat untuk mengakses informasi kesehatan melalui platform media sosial seperti YouTube. Video edukasi kesehatan oleh tenaga medis, salah satunya dr. Tirta, mendapatkan perhatian tinggi dan menghasilkan interaksi publik dalam bentuk komentar. Penelitian ini bertujuan untuk mengklasifikasikan sentimen komentar YouTube pada video edukasi kesehatan menggunakan algoritma Support Vector Machine (SVM). Data dikumpulkan dari video YouTube berjudul "NETIZEN BERTANYA LAGI, dr. TIRTA MENJAWAB KEMBALI: MITOS & FAKTA KESEHATAN" dengan total 2.414 komentar. Tahapan penelitian meliputi pengumpulan data menggunakan web scraping, pra-pemrosesan data (case folding, cleaning, stopword removal, stemming, tokenization), pelabelan otomatis komentar ke dalam kategori positif, negatif, dan netral, ekstraksi fitur menggunakan TF-IDF, serta pelatihan model klasifikasi menggunakan SVM dengan kernel linear. Hasil evaluasi menunjukkan model memperoleh akurasi sebesar 93,2%, precision 96,2%, recall 65,5%, dan F1-score 67,8%. Penelitian ini menyimpulkan bahwa algoritma SVM efektif digunakan



untuk klasifikasi komentar YouTube pada konten edukasi kesehatan dan dapat mendukung analisis opini publik secara otomatis.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



**Corresponding Author:**

Caswita

Universitas Pelita Bangsa

E-mail: [cscaswita@gmail.com](mailto:cscaswita@gmail.com)

## PENDAHULUAN

Perkembangan teknologi digital telah mengubah cara masyarakat dalam mengakses informasi, termasuk informasi seputar kesehatan. Salah satu media yang berperan penting adalah YouTube, yang kini tidak hanya digunakan sebagai sarana hiburan, tetapi juga sebagai media edukasi yang sangat populer. Menurut laporan DataReportal (2024), sebanyak 93,8% pengguna internet di Indonesia menggunakan YouTube, menjadikannya salah satu platform edukasi terbesar di era digital[1].

Salah satu konten edukasi kesehatan yang sering dijadikan referensi masyarakat adalah video edukasi dari dr. Tirta, khususnya video dengan tema “Mitos & Fakta Kesehatan” yang membahas isu kesehatan sehari-hari dengan gaya komunikasi ringan namun informatif.

Seiring meningkatnya konsumsi video edukasi kesehatan, kolom komentar di YouTube menjadi ruang interaksi yang penting untuk menilai opini publik. Komentar tersebut mencerminkan sentimen pengguna, baik dalam bentuk dukungan positif, pertanyaan, kritik, hingga penolakan informasi. Namun, mengingat volume komentar yang sangat besar, proses analisis secara manual tidak efisien dan menyulitkan analisis persepsi publik secara menyeluruh.

Dalam konteks pengolahan data teks, klasifikasi komentar merupakan solusi efektif untuk mengelompokkan opini publik secara otomatis. Salah satu metode klasifikasi teks yang banyak digunakan adalah algoritma Support Vector Machine (SVM), yang dikenal efektif dalam pengolahan data teks berdimensi tinggi. Penelitian ini bertujuan untuk mengimplementasikan algoritma SVM dalam klasifikasi sentimen komentar YouTube pada video edukasi kesehatan oleh dr. Tirta, khususnya untuk mengetahui distribusi sentimen positif, netral, dan negatif terhadap konten edukasi mitos dan fakta kesehatan.

## TINJAUAN PUSTAKA

YouTube telah berkembang menjadi salah satu platform edukasi terbesar secara global. Berdasarkan laporan DataReportal (2024), sebanyak 93,8% pengguna internet di Indonesia mengakses YouTube, yang tidak hanya menyajikan hiburan tetapi juga menyediakan berbagai konten edukasi termasuk edukasi kesehatan [1]. Salah satu contoh nyata adalah video-video edukasi dari tenaga kesehatan seperti dr. Tirta, yang secara aktif berkontribusi dalam meningkatkan literasi kesehatan masyarakat melalui media digital.



Salah satu aspek menarik dari YouTube adalah adanya kolom komentar yang menjadi ruang interaksi publik. Komentar yang ditinggalkan oleh pengguna dapat merepresentasikan berbagai macam sentimen, baik berupa apresiasi, kritik, pertanyaan, maupun ketidaksetujuan. Beberapa penelitian menyebutkan bahwa komentar-komentar ini merupakan sumber data tidak terstruktur yang dapat dimanfaatkan untuk analisis opini atau persepsi publik terhadap suatu konten [2]. Oleh karena itu, klasifikasi komentar sangat penting untuk membantu content creator maupun peneliti dalam memahami tren opini masyarakat secara cepat dan akurat.

Dalam proses analisis komentar, klasifikasi teks menjadi metode populer dalam ranah data mining. Klasifikasi teks bertujuan untuk mengelompokkan data teks ke dalam kategori tertentu seperti positif, netral, dan negatif. Penelitian sebelumnya telah membuktikan bahwa pendekatan supervised learning dapat digunakan secara efektif dalam mengidentifikasi pola sentimen dari komentar media sosial secara otomatis [3].

Salah satu algoritma supervised learning yang banyak digunakan dalam klasifikasi teks adalah Support Vector Machine (SVM). Algoritma ini memiliki kemampuan dalam menemukan hyperplane optimal untuk memisahkan data ke dalam kategori klasifikasi yang berbeda. Khomsah dan Aribowo (2022) menunjukkan bahwa penggunaan SVM dengan metode TF-IDF mampu menghasilkan akurasi sebesar 92,86% dalam klasifikasi komentar YouTube berbahasa Indonesia [4]. Selain itu, penelitian oleh Nursetyo dan Luthfi (2021) juga menunjukkan bahwa algoritma SVM mampu mencapai akurasi sebesar 93% dalam klasifikasi sentimen komentar pada video edukasi kesehatan [5].

Namun, sebagian besar penelitian klasifikasi komentar YouTube yang ada masih berfokus pada komentar umum seperti vlog, produk, atau hiburan. Penelitian spesifik mengenai klasifikasi komentar pada konten edukasi kesehatan, khususnya yang membahas mitos dan fakta kesehatan, masih sangat terbatas. Berdasarkan celah tersebut, penelitian ini dilakukan untuk mengisi gap dengan menerapkan algoritma Support Vector Machine (SVM) dalam klasifikasi sentimen komentar pada video edukasi kesehatan oleh dr. Tirta.

## **METODE PENELITIAN**

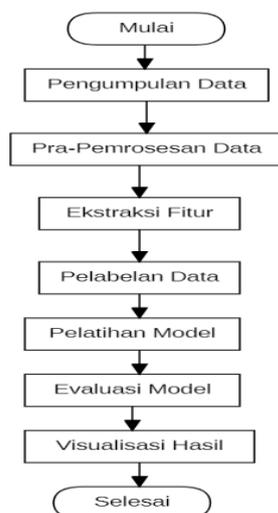
Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi teks berbasis algoritma Support Vector Machine (SVM) untuk menganalisis sentimen komentar YouTube. Studi kasus yang digunakan adalah video edukasi kesehatan oleh dr. Tirta yang berjudul “NETIZEN BERTANYA LAGI, Dr. TIRTA MENJAWAB KEMBALI: MITOS & FAKTA KESEHATAN” (<https://www.youtube.com/watch?v=YEAq10ZdbdU>). Penelitian ini mengikuti alur sistematis yang terdiri dari lima tahapan utama, yaitu: pengumpulan data, pra-pemrosesan data, ekstraksi fitur, pelatihan model klasifikasi SVM, dan evaluasi performa model.

Proses pengumpulan data dilakukan dengan teknik web scraping untuk mengambil komentar dari video YouTube yang menjadi objek studi. Setelah data dikumpulkan, dilakukan proses pra-pemrosesan seperti case folding, pembersihan teks, stopword removal, stemming, dan tokenization agar data lebih bersih dan siap digunakan. Ekstraksi fitur dilakukan menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma SVM. Selanjutnya, model dikembangkan menggunakan algoritma SVM dengan kernel linear yang sesuai dengan rekomendasi penelitian sebelumnya,



karena SVM dinilai efektif dalam menangani data teks berdimensi tinggi serta memberikan performa klasifikasi yang baik (Khomsah & Aribowo, 2022) [6].

Penelitian ini dilakukan mengikuti alur sistematis yang digambarkan melalui diagram alir penelitian pada Gambar 1. Diagram alir ini berfungsi untuk memvisualisasikan tahapan proses penelitian mulai dari pengumpulan data hingga visualisasi hasil klasifikasi komentar YouTube menggunakan algoritma Support Vector Machine (SVM), dengan tujuan akhir menghasilkan model klasifikasi sentimen yang akurat.



Gambar 1. Diagram Alur Penelitian

## HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan secara menyeluruh mengenai hasil penelitian yang meliputi proses pengumpulan data komentar YouTube, tahapan pra-pemrosesan data untuk membersihkan dan mempersiapkan data teks, proses pelabelan data ke dalam kategori sentimen positif, negatif, dan netral, pelatihan model klasifikasi menggunakan algoritma Support Vector Machine (SVM), evaluasi performa model menggunakan metrik akurasi, precision, recall, dan F1-score, serta pembahasan hasil klasifikasi komentar terhadap video edukasi kesehatan. Selain itu, bab ini juga menyajikan visualisasi hasil berupa confusion matrix dan analisis mendalam terkait efektivitas model dalam mengelompokkan sentimen komentar pada konten edukasi.

### 1. Pengumpulan Data

Data komentar diambil dari video YouTube berjudul “NETIZEN BERTANYA LAGI, dr. TIRTA MENJAWAB KEMBALI: MITOS & FAKTA KESEHATAN” menggunakan teknik web scraping dengan bantuan bahasa pemrograman Python. Tools yang digunakan adalah youtube-comment-downloader untuk mengekstrak seluruh komentar publik dari video yang ditentukan. Total komentar yang berhasil dikumpulkan sebanyak 2.414 komentar dalam satu sesi scraping untuk menjaga konsistensi data. Data kemudian disimpan dalam format CSV dengan tiga atribut utama, yaitu **Username**, **Comment**, dan **Likes**. Pengumpulan data dilakukan pada tanggal [isi tanggal



pengambilan data], dan file hasil pengambilan data dinamai **data\_komentar\_dr\_tirta.csv**. Contoh isi data yang dikumpulkan dapat dilihat pada Tabel 3.1. Data inilah yang kemudian digunakan untuk proses analisis lebih lanjut melalui tahapan pra-pemrosesan, pelabelan, dan klasifikasi sentimen.

username	comment	likes
@nicohereid	dok, kringat dingin itu karna apa ya	1102
@nicohereid	di tunggu jawabannya dok 🙄	75
@damaihati7151	Keringat dingin dapat terjadi karena berbagai faktor, baik yang bersifat fisik maupun psikologis. Penyebab umum meliputi stres, cemas, rasa takut, serta kondisi medis seperti infeksi, syok, dan masalah jantung. Selain itu, kondisi seperti gula darah rendah, kekurangan oksigen, dan efek samping obat-obatan juga bisa memicu keringat dingin.	100
@tiangjemuran7580	Keringat dingin itu disebabkan oleh dep kolektor	54
@Aznayafamily	Cicilan bang 🙄	38
@ken2dw	Bang, coba pas keringetan dikipasin. Itu udah keringat dingin.	7
@electricgodyan3963	keringat lu kedinginan	2
@elhubby	@nicohereid keringat kan normalnya keluar bersama suhu tubuh, kalau keringat dingin berarti berkeringat tp tanpa tujuan melepaskan suhu. Tp jawaban yg benarnya akan disampaikan oleh Suhunya 🙏	5
@kurniawandwizulianto7734	Laper bang	5
@christiangerrardd	@nicohereid nonton emyu main bang 🙄	4

Tabel 3. 1 Data Komentar YouTube yang Diperoleh

File lengkap komentar disimpan dalam file `data_komentar_dr_tirta.csv` dan dilampirkan dalam lampiran penelitian.

## 2. Pra-Pemrosesan Data

Pra-pemrosesan dilakukan untuk membersihkan data komentar sehingga siap untuk proses klasifikasi. Tahapan pra-pemrosesan dilakukan secara sistematis untuk memastikan data yang digunakan bersih dan terstruktur dengan baik. Adapun hasil dari setiap tahapan pra-pemrosesan adalah sebagai berikut:



comment	processed_comment
dok, kringat dingin itu karna apa ya	dok kringat dingin karna apa
di tunggu jawabannya dok 🙏	tunggu jawab dok
Keringat dingin dapat terjadi karena berbagai faktor, baik yang bersifat fisik maupun psikologis. Penyebab umum meliputi stres, cemas, rasa takut, serta kondisi medis seperti infeksi, syok, dan masalah jantung. Selain itu, kondisi seperti gula darah rendah, kekurangan oksigen, dan efek samping obat-obatan juga bisa memicu keringat dingin.	keringat dingin jadi bagai faktor baik sifat fisik maupun psikologis sebab umum liput stres cemas rasa takut kondisi medis infeksi syok masalah jantung kondisi gula darah rendah kurang oksigen efek samping obatobatan picu keringat dingin
Keringat dingin itu disebabkan oleh dep kolektor	keringat dingin sebab dep kolektor
Cicilan bang 🙏	cicil bang
Bang, coba pas keringetan dikipasin. Itu udah keringat dingin.	bang coba pas keringetan dikipasin udah keringat dingin
keringat lu kedinginan	keringat lu dingin
@nicohereid keringat kan normalnya keluar bersama suhu tubuh, kalau keringat dingin berarti berkeringat tp tanpa tujuan melepaskan suhu. Tp jawaban yg benarnya akan disampaikan oleh Suhunya 🙏	nicohereid keringat kan normal keluar sama suhu tubuh kalau keringat dingin arti keringat tp tuju lepas suhu tp jawab yg benarnya sampai suhu
Laper bang	laper bang
@nicohereid nonton emyu main bang 🙏	nicohereid nonton emyu main bang

Tabel 3. 2 Komentar Sebelum dan Sesudah Pra-Pemrosesan

### 3. Pelabelan Data

Sebanyak 2.414 komentar dilakukan pelabelan otomatis untuk kebutuhan supervised learning dengan tiga kelas sentimen, yaitu:

- Positif: komentar yang mengandung apresiasi, pujian, atau dukungan terhadap isi video maupun narasumber.
- Negatif: komentar yang menunjukkan kritik, keluhan, atau ketidaksetujuan terhadap isi video, narasumber, maupun topik pembahasan.
- Netral: komentar berupa pertanyaan, opini netral, atau pernyataan informatif tanpa kecenderungan sentimen tertentu.

Proses pelabelan otomatis dilakukan menggunakan metode berbasis pencocokan kata kunci dan rule-based classification, dengan daftar kata kunci positif, negatif, dan netral yang telah ditentukan sebelumnya. Hasil pelabelan kemudian diperiksa secara acak untuk memastikan akurasi pelabelan awal sebelum dilanjutkan ke tahap pelatihan model klasifikasi.



username	comment	likes	processed_comments	label
@nicohereid	dok, kringat dingin itu karna apa ya	1102	dok kringat dingin karna apa	Netra 1
@nicohereid	di tunggu jawabannya dok 🙏	75	tunggu jawab dok	Netra 1
@damaihati7151	Keringat dingin dapat terjadi karena berbagai faktor, baik yang bersifat fisik maupun psikologis. Penyebab umum meliputi stres, cemas, rasa takut, serta kondisi medis seperti infeksi, syok, dan masalah jantung. Selain itu, kondisi seperti gula darah rendah, kekurangan oksigen, dan efek samping obat-obatan juga bisa memicu keringat dingin.	100	keringat dingin jadi bagai faktor baik sifat fisik maupun psikologis sebab umum liput stres cemas rasa takut kondisi medis infeksi syok masalah jantung kondisi gula darah rendah kurang oksigen efek samping obatobatan picu keringat dingin	Netra 1
@tiangjemuran7580	Keringat dingin itu disebabkan oleh dep kolektor	54	keringat dingin sebab dep kolektor	Netra 1
@Aznayafamily	Cicilan bang 🙏	38	cicil bang	Netra 1
@ken2dw	Bang, coba pas keringetan dikipasin. Itu udah keringat dingin.	7	bang coba pas keringetan dikipasin udah keringat dingin	Netra 1
@electricgodyan3963	keringat lu kedinginan	2	keringat lu dingin	Netra 1
@elhubby	@nicohereid keringat kan normalnya keluar bersama suhu tubuh, kalau keringat dingin berarti berkeringat tp tanpa tujuan melepaskan suhu. Tp jawaban yg benarnya akan disampaikan oleh Suhunya 🙏	5	nicohereid keringat kan normal keluar sama suhu tubuh kalau keringat dingin arti keringat tp tuju lepas suhu tp jawab yg benarnya sampai suhu	Netra 1
@kurniawandwizulianto7734	Laper bang	5	laper bang	Netra 1
@christiangerrardd	@nicohereid nonton emyu main bang 🙏	4	nicohereid nonton emyu main bang	Netra 1



Tabel 3. 3 Hasil Pelabelan Data

4. Pelatihan Model

Model klasifikasi dilatih menggunakan algoritma **Support Vector Machine (SVM)** dengan kernel linear. Dataset dibagi menjadi 80% data latih dan 20% data uji secara acak menggunakan fungsi `train_test_split`. Proses ekstraksi fitur menggunakan **TF-IDF Vectorizer** dengan konfigurasi sebagai berikut:

- **Maximum Features** = 1000
- **N-Gram Range** = (1,1) (unigram)
- **Stopword** = bahasa Indonesia

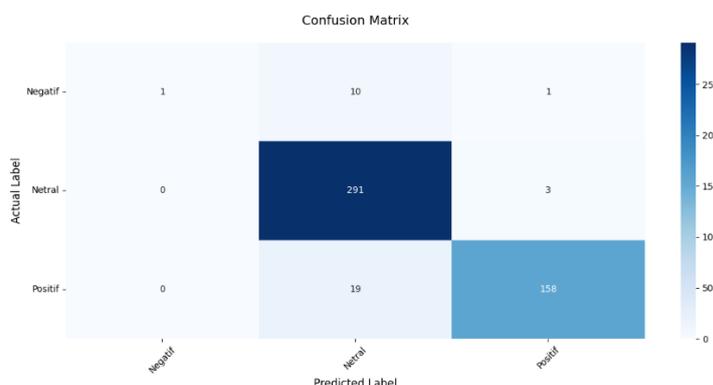
Model kemudian dilatih menggunakan data terlabel untuk menghasilkan model klasifikasi teks berbasis SVM.

5. Evaluasi Model

Evaluasi model dilakukan menggunakan metrik standard klasifikasi, yaitu akurasi, precision, recall, dan F1-score, serta confusion matrix. Hasil evaluasi model dapat dilihat pada Tabel 3.4 berikut:

HASIL EVALUASI MODEL SVM	
Metrik	Nilai (%)
Akurasi	93.2
Precision	96.2
Recall	65.5
F1-Score	67.8

Tabel 3. 4 Hasil Evaluasi Model SVM



Gambar 3. 1 Matrix Confusion



## KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Support Vector Machine (SVM) dengan kernel linear untuk klasifikasi sentimen komentar YouTube pada video edukasi kesehatan. Proses pra-pemrosesan data seperti case folding, cleaning, stopword removal, stemming, dan tokenization terbukti efektif meningkatkan kualitas data. Hasil evaluasi menunjukkan model SVM mampu mencapai akurasi 93,2%, precision 96,2%, recall 65,5%, dan F1-score 67,8%, dengan dominasi komentar netral pada dataset. Secara keseluruhan, model yang dikembangkan dapat diandalkan untuk klasifikasi sentimen komentar secara otomatis dan dapat dijadikan acuan dalam analisis opini publik terhadap konten edukasi kesehatan.

## DAFTAR RUJUKAN

- DataReportal, "Data Reportal," DataReportal, 2024, [Online]. Available: <https://datareportal.com/reports/digital-2024-indonesia>
- B. D. S. Nashrullah Nashrullah, "Klasifikasi Emosi pada Komentar YouTube menggunakan Algoritme Support Vector Machine," J. Pengemb. Teknol. Inf. Dan Ilmu Komput., vol. 7, no. 9, 2023, [Online]. Available: [https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12783?utm\\_source=chatgpt.com](https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12783?utm_source=chatgpt.com)
- S. Sudianto, J. A. A. Masheli, N. Nugroho, R. W. Ananda Rumpoko, and Z. Akhmad, "Comparison of Support Vector Machines and K-Nearest Neighbor Algorithm Analysis of Spam Comments on Youtube Covid Omicron," J. Tek. Inform., vol. 15, no. 2, pp. 110–118, 2022, doi: 10.15408/jti.v15i2.24996.
- Muhammad Hilmy Riwanto and Pramudhitya Ardhiansyah, "Analisis Sentimen Komentar YouTube Terkait Penerapan Makan Bergizi Gratis Menggunakan Model Algoritma SVM," Kohesi J. Multidisiplin Saintek, vol. 6, no. 12, pp. 1–14, 2025.
- C. S. Salsabila et al., "max 18 words ( In English ) Bold , Align Text Left , One Single Space , 14pt ) Poverty Level Clustering in Districts / Cities Using the K-Medoids Method Based on Population Data," vol. 2, no. 2023, pp. 1–7, 2024.
- S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, S. Khomsah, and A. S. Aribowo, "Terakreditasi SINTA Peringkat 2 Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," Masa Berlaku Mulai, vol. 1, no. 3, pp. 648–654, 2017.

## LAMPIRAN PENELITIAN



data\_komentar\_dr\_tir  
ta.csv

<https://drive.google.com/file/d/1svL03m8m8KL05CcB7DdvVE8TSmrJSOzX/view?usp=sharing>