

Analisis Forensik Digital Pada Komentar Youtube Live Menggunakan *Sentiment Analysis*

Uning Kristiana*¹, Alfirna Rizqi Lahitani ², Chanief Budi Setiawan ³, Nafisa Alfi Sa'diya⁴

^{1,2,3}Teknologi Informasi, FTI Unjani, Yogyakarta, Indonesia

e-mail: ¹uningkrist@gmail.com, ²alfinarizqi@gmail.com, ³chanief.b.s@gmail.com,
⁴nafisaalfi204@gmail.com

Abstract - The development of increasingly sophisticated technology can have a positive influence on various aspects of our daily lives. From the survey results of the Indonesian Internet Services Association (APJII) in the second quarter of 2019-2020, it shows that Internet users of the operator spend more time watching online videos. Youtube video content watching is open to the public and all ages can freely watch it. However, the content and comments are not necessarily suitable for audiences of all ages to read. Of course, Youtube video content can also affect behavior, especially minors. The purpose of this research is to conduct digital forensic analysis on Youtube Live Comments using sentiment analysis. The research method used applies the NIST SP 800-86 method, namely Collection, Examination, Analysis, and Reporting. Sentiment analysis resulted in 0.01 in the comments on the two videos tested, namely the PUBG and Free Fire video games. Sentiment analysis resulted in 0.01 in the comments on the two videos tested, namely the PUBG and Free Fire video games.

Keywords - Sentiment Analysis, digital forensics, TF-IDF, Cosine Similarity, Youtube

Abstrak - Perkembangan teknologi yang semakin canggih dapat memberi pengaruh dalam bentuk positif di berbagai kehidupan sehari-hari. Dari hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2019-2020 kuartal ke-2 menunjukkan pengguna internet lebih banyak menghabiskan waktu dengan menonton video online. Konten video Youtube terbuka untuk umum dan segala usia dapat menonton dengan bebas. Namun, isi konten maupun komentar yang ada belum tentu pantas di baca penonton segala umur. Tentu konten video Youtube juga dapat mempengaruhi perilaku khususnya anak dibawah umur. Adanya penelitian ini memiliki tujuan yaitu melakukan Analisis Forensik Digital pada Komentar Youtube Live menggunakan Sentiment Analysis. Metode penelitian yang digunakan menerapkan metode NIST SP 800-86 yaitu Collection, Examination, Analysis dan Reporting. Analisis sentimen menghasilkan angka 0,01 pada komentar di kedua video yang diuji yaitu video games PUBG dan Free Fire. Kesimpulan: Dari penelitian ini adalah video games PUBG dan Mobile Legend dapat ditonton semua usia namun untuk usia anak

dibawah umur lebih baik jika dalam pengawasan orang tua..

Kata kunci - Sentiment Analysis, forensik digital, TF-IDF, Cosine Similarity, Youtube

I. PENDAHULUAN

Pada era digital saat ini, penggunaan internet semakin meluas dan beragam. Salah satu platform yang sangat populer adalah *YouTube*, di mana pengguna dapat menonton dan mengunggah video secara online. Namun, keberadaan konten di *YouTube* juga dapat memengaruhi perilaku penonton, terutama di kalangan anak-anak yang masih di bawah umur. Dalam konteks ini, *cyberbullying* menjadi salah satu bentuk kekerasan yang sering terjadi di platform *YouTube*. *Cyberbullying* dapat berupa komentar berkonotasi buruk atau penyalahgunaan pembacaan dan penulisan komentar. Diharapkan bahwa analisis sentimen akan menghasilkan prediksi apakah komentar pada video bersifat positif atau negatif [1]. Tujuan dari penelitian ini adalah untuk mengetahui kelompok paling banyak penyumbang komentar positif dan negatif dalam komentar video pada platform *YouTube*, serta mengetahui hasil persentase dari penilaian komentar video yang diuji yang bersifat positif maupun negatif .

Landasan teori yang akan digunakan sebagai dasar dalam mengerjakan penelitian ini adalah berasal dari beberapa jurnal terkait dengan penelitian yang hampir sama dengan penelitian ini. Adapun topik yang relevan dibahas pada penelitian sebelumnya yaitu Penelitian yang pertama dilakukan oleh [2] hasil penelitian ini memberikan akses penuh ke semua informasi yang diterima oleh pengguna smartphone dan *WhatsApp* berbasis web. Ditampilkan juga isi pesan, kontak, dan data pengguna unik. Penelitian kedua dilakukan oleh [3] yang membahas mengenai analisis sentiment menggunakan *Support vector Machine* dengan fitur *Naïve bayes* (NBSVM) pada komentar video Youtube dengan 400 variasi data yang diuji dapat didefinisikan bahwa 251 komentar bernilai positif dan 149 komentar bernilai negative. Penelitian ketiga dilakukan oleh [4] dalam penelitian ini membahas tentang proses investigasi dan analisis forensik percakapan grup dari platform messenger diperlukan oleh penyelidik polisi dan penyelidik swasta untuk menentukan isi percakapan. Dalam studi ini, investigasi dan analisis forensik mengikuti standar NIST SP80086, dan

algoritma *Support Vector Machine* (SVM) digunakan untuk analisis dan pelaporan. Penelitian keempat dilakukan oleh [5] dengan digunakan untuk mengamati analisis sentimen kinerja Ahok. Prosesnya meliputi komentar data, preprocessing, tokenization, dan penentuan sentimen berbasis *Lexicon*. Penelitian Kelima dilakukan oleh [6] menganalisis kekerasan verbal di platform Youtube Indonesia dari perspektif kultivasi. Penelitian ini menggunakan pendekatan kualitatif dengan metode survei kepustakaan. *Youtube* telah menjadi media baru yang didedikasikan untuk menyebarkan informasi melalui konten. Oleh karena itu, bukti kekerasan verbal menjadi perhatian utama bagi semua yang terlibat. Penelitian keenam oleh [2] dengan studi kasus prostitusi online terkait Android diselidiki dengan menganalisis bukti yang diperoleh, perangkat lunak forensik *MOBILedit Express* dan *SysTools SQLite Viewer* yang disimulasikan menggunakan metode *National Justice Department* (NIJ). Penelitian ketujuh dilakukan oleh [6] dengan menganalisis bukti digital *cyberbullying* di media sosial, khususnya *Twitter*. Penelitian kedelapan dilakukan oleh [7] yang menganalisis tren akun media sosial *Twitter* menunjukkan kesamaan antara TF-IDF dan Cosine. Berdasarkan hasil pengujian yang dijalankan, sistem yang dibangun menggunakan kumpulan data otomatis yang terbentuk dari data terbaru hingga tren dari akun media sosial yang dipilih dengan akurasi menghasilkan 99,14. Penelitian kesembilan dilakukan oleh [8] yang menghasilkan Hasil akurasi klasifikasi untuk algoritma klasifikasi *Naive Bayes* pada kanal KemenPUPR menunjukkan nilai 69,23 persen dan pada kanal Kemenkeu nilai 64,10 persen. Penelitian kesepuluh dilakukan oleh [9] yang meneliti hukum normatif menggunakan pendekatan kasus analisis data mengevaluasi data yang terkumpul secara deskriptif.

Metode analisis data mengevaluasi data yang terkumpul secara deskriptif. Dimana menyajikan semua masalah yang ada secara akurat dan jelas. Kesimpulannya secara induktif dari pernyataan-pernyataan yang konkrit sehingga penyajian hasil pembahasan mudah dipahami.

II. METODE PENELITIAN

Dalam penelitian ini akan menggunakan metode penelitian seperti yang terlihat pada alur Gambar 1 :



Gambar 1. Alur Implementasi NIST SP 800-86

Berdasarkan gambar 1 alur implementasi NISTSP 800-86 [10] proses digital forensik terdiri dari empat langkah: *collection*, *examination*, *analysis*, dan *reporting*. Langkah pertama adalah *collection*, di mana data terkait kejadian yang akan diselidiki dikumpulkan dengan tanda, pencatatan, dan penyimpanan yang

akurat. Langkah kedua adalah *examination*, di mana data diselidiki menggunakan teknik forensik yang sesuai untuk menjaga keakuratan data. Observasi dapat dilakukan dengan menggunakan alat otomatis maupun proses tradisional. Langkah ketiga adalah *analysis*, di mana hasil dari *examination* dianalisis untuk mengungkap informasi yang relevan dalam menjawab pertanyaan dan tujuan penyelidikan. Langkah terakhir adalah *reporting*, di mana hasil analisis disajikan secara *visual* dan rekomendasi untuk perubahan peraturan, pedoman, standar, dan prosedur forensik ditentukan.

A. Bahan Dan Alat Penelitian

Penelitian sentiment analysis menggunakan data teks yang diperoleh melalui teknik crawling dari platform Youtube, khususnya komentar pada video live. Jumlah data testing yang digunakan berkisar antara 200 hingga 1000. Parameter komentar mencakup video games dari channel *Youtube FF Esport ID*, *PUBG MOBILE Indonesia*, dan *Mobile Legend: Bang Bang Official*, termasuk konten lama dan konten terbaru dari tahun 2021-2022. Penelitian ini membutuhkan laptop dengan spesifikasi yang memadai, perangkat lunak yang disesuaikan, koneksi internet, serta sistem operasi dan program aplikasi yang relevan. Adapun sistem operasi dan program aplikasi yang diperlukan untuk memenuhi jalannya penelitian ini sebagai berikut :

1. Sistem operasi : Windows 10
2. Aplikasi : *Microsoft Excel*, *Quickhash*, dan *Google Colaboratory*
3. Bahasa pemrograman: *Python*

B. Perancangan Sistem Analisis Sentimen

Penelitian ini adalah penelitian yang menganalisis peristiwa yang sudah terjadi dalam proses penggunaan teknologi yang mempunyai indikasi adanya kejahatan di dunia maya khususnya *platform Youtube*. Penelitian memiliki bahan utama analisis adalah data komentar pada video Youtube live. Penelitian ini melakukan implementasi NIST SP 800-86 yang terdiri dari 4 langkah, sebagi berikut :

1. Collection

Proses *collection* ini meliputi pengumpulan data yang akan diselidiki, pemberian tanda, pencatatan dan pengambilan data sekaligus penyimpanan data agar keutuhan dan keakuratan data masih terjaga. Tahap *collection* memiliki 2 proses, sebagai berikut :

1. Identifikasi masalah

Proses melakukan analisis terhadap peristiwa yang terjadi di internet dimana kejadian tersebut termasuk dalam salah satu bentuk kejahatan internet pada penggunaan platform Youtube.

2. Pengumpulan Data

Pengumpulan data komentar video dari platform YouTube menggunakan pendekatan forensik digital melibatkan proses *crawling*. *Crawling*

adalah proses pengambilan data dari sekumpulan data yang kemudian akan diproses sesuai dengan kebutuhan penyelidikan. Tujuan dari crawling adalah mengambil data dari halaman internet melalui aplikasi script program dan melakukan scanning untuk mengambil data yang diperlukan. Dalam penelitian ini, data komentar diambil dari beberapa jenis video game di YouTube, seperti Game *FreeFire*, *Player Unknown's Battlegrounds* (PUBG), dan *Mobile Legend*. Klasifikasi dilakukan dengan menggunakan data training dan data testing. Algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan dalam penelitian ini. Data yang dikumpulkan meliputi 1650 komentar, dengan 550 komentar digunakan sebagai data training dan 1100 komentar sebagai data testing. Jumlah komentar yang diuji adalah 550 per video game untuk memperoleh variasi yang lebih baik dalam penelitian [11].

2. Examination

Tahapan *examination* merupakan proses untuk melakukan pemeriksaan terhadap data yang telah dikumpulkan dari tahap *collection* dengan menggunakan kombinasi metode otomatis dan manual. Sehingga dapat menilai dan melakukan ekstraksi data dengan tetap menjaga integritas data itu sendiri yang sudah diamankan[12].

3. Analysis

Tahap *analysis* meliputi proses penguraian hasil dari examination dengan tujuan mengetahui informasi yang dapat membantu dalam mengungkap teka-teki yang dijadikan sebuah alasan dilakukannya pengumpulan serta pemeriksaan, meliputi beberapa tahap yaitu :

1. Tahap Pengolahan Data

Proses pengolahan data setelah diambil dan disimpan melibatkan tahap pre-processing. Pre-processing adalah proses persiapan data teks agar dapat diproses lebih lanjut. *Text pre-processing* memiliki pengaruh yang signifikan terhadap hasil akhir dari teknik text mining yang digunakan. Pre-processing melibatkan beberapa tahap penting. Tahap-tahap tersebut meliputi mengubah variabel teks menjadi karakter, mengubah huruf kapital menjadi huruf kecil (*case folding*), memecah komentar menjadi kumpulan data (*tokenizing*), dan menghapus kata-kata yang tidak penting.

Urutan pre-processing sebagai berikut ini :

- Case folding* (Mengganti semua data yang akan dianalisis dari huruf aslinya kecil maupun kapital menjadi huruf kecil semua(lowercase)).

- Tokenizing* (Memisahkan dokumen teks menjadi satuan token, apabila telah melalui proses penghilangan karakter unik)
- Filtering* (penyeleksian kata-kata dalam data yang memiliki kata yang tidak penting maupun tidak bermakna atau sering dikenal dengan sebutan stopword akan dihapus)

2. Tahap Uji Coba

Tahap uji coba merupakan proses pengujian klasifikasi sentimen dengan data yang telah melalui beberapa pengolahan sebelumnya

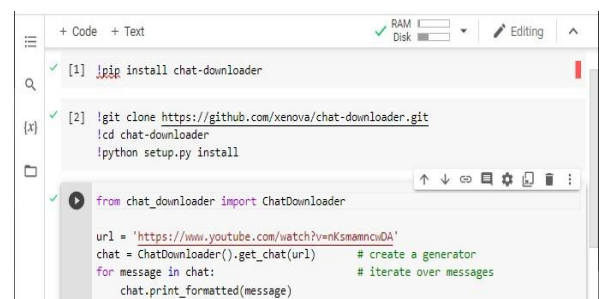
III. HASIL DAN PEMBAHASAN

A. Identifikasi Masalah

Konteks *cyberbullying* dalam penelitian ini melibatkan bentuk-bentuk kekerasan tidak langsung, seperti penggunaan bahasa berkonotasi buruk atau penyalahgunaan komentar dan penulisan. Khususnya dalam platform video *YouTube* yang banyak ditonton oleh remaja dan anak-anak. Riset dari APJII menunjukkan bahwa penggunaan internet, terutama menonton video secara online, telah meningkat dalam kurun waktu kuartal ke-2 tahun 2019 hingga 2020. Oleh karena itu, konten yang ada di *YouTube* memiliki dampak signifikan terhadap perilaku penonton, terutama pada anak-anak di bawah umur.

B. Pengumpulan Data

Tahap *crawling* data menggunakan bahasa pemrograman Python dan google colaboratory sebagai coding environment bahasa pemrograman *Python* dengan format *notebook.Tool* sederhana yang dapat digunakan mengambil pesan obrolan dari live streaming, video maupun siaran yang telah ditayangkan. Proses crawling data komentar terlihat seperti Gambar 2.



```

+ Code + Text
RAM
Disk
Editing
[1] !pip install chat-downloader

[2] !git clone https://github.com/xenova/chat-downloader.git
    !cd chat-downloader
    !python setup.py install

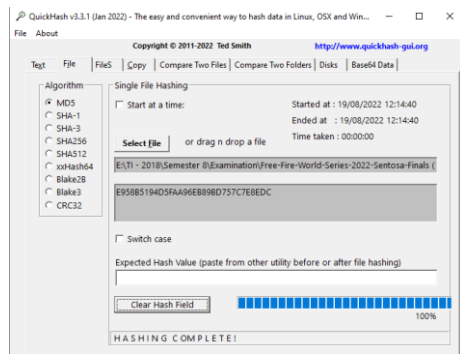
from chat_downloader import ChatDownloader

url = 'https://www.youtube.com/watch?v=NksmancvDA'
chat = ChatDownloader().get_chat(url) # create a generator
for message in chat:                 # iterate over messages
    chat.print_formatted(message)
  
```

Gambar 2. Crawling Data Komentar Live Youtube

C. Tahap Pengolahan Data

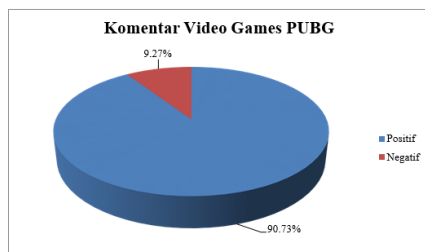
Tahap ini merupakan tahap pemrosesan data yang telah dikumpulkan dari tahap collection. File yang dimiliki merupakan data komentar video games *Free Fire World Series 2022 Sentosa Finals* diolah melalui proses *hash*, dapat dilihat pada Gambar 3 :



Gambar 3. Hasil Hash Barang Bukti Free Fire

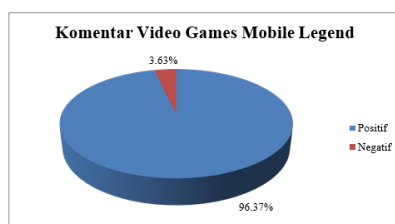
D. Pembahasan

Hasil *sentiment analysis* dari komentar video games PUBG yaitu 9,27% dari 550 komentar bersifat negatif. Sedangkan komentar positif yaitu 90,73%, hal ini memperlihatkan bahwa video games PUBG masih bisa ditonton disegala usia namun khusus anak dibawah umur sebaiknya masih dalam bimbingan maupun pengawasan orangtua atau orang dewasa. Hasil analisis dalam bentuk diagram lingkaran terlihat pada Gambar 4 :



Gambar 4. Diagram Persentase Komentar Video Games PUBG

Hasil *sentiment analysis* dari komentar video games Mobile Legend yaitu 3,63% dari 550 komentar bersifat negatif. Sedangkan komentar positif yaitu 96,37%, hal ini memperlihatkan bahwa video games Mobile Legend masih bisa ditonton disemua usia meski nilai persentase komentar negatif termasuk kecil namun khusus anak dibawah umur sebaiknya masih dalam bimbingan maupun pengawasan orangtua atau orang dewasa ketika menonton video games tersebut. Hasil analisis dalam bentuk diagram lingkaran terlihat pada Gambar 5 :



Gambar 5. Diagram Persentase Komentar Video Games Mobile Legend

Secara garis besar berikut adalah perbandingan banyaknya komentar buruk yang terdapat di dua video yang diuji dan video pembandingnya, terlihat pada Tabel 1.

Tabel 1. Perbandingan Komentar di Video Free Fire, PUBG & Mobile Legend

No.	Video	Jenis Data	Jumlah Komentar dengan term kata buruk
1.	Free Fire	Pembanding (q)	87
2.	PUBG	Uji	51
3.	Mobile Legend	Uji	20

IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan pada dua video game, yaitu *31st SEA Games | PUBG MOBILE Mode Squad Hari Ke-1 dan M3 Playoffs Hari 7 MLBB Championship 2021 Singapura* | Bahasa Indonesia, ditemukan bahwa kedua video game tersebut memiliki hasil rata-rata cosine similarity sebesar 0,1. Hasil ini juga dibandingkan dengan data komentar video games Free Fire World Series 2022 Sentosa Finals. Dari hasil tersebut, terlihat bahwa jumlah komentar buruk pada video PUBG dan Mobile Legend masih lebih sedikit dibandingkan dengan video Free Fire. Oleh karena itu, penulis menyarankan untuk melakukan pengembangan lebih lanjut pada penelitian-penelitian yang memiliki pembahasan serupa agar hasil analisis menjadi lebih efisien dan akurat.

DAFTAR PUSTAKA

- [1] M. A. Ahdiyati, "Kekerasan Verbal di Konten Youtube Indonesia Dalam Perspektif Kultivasi Cultivation Perspective," vol. 5 No.2, 2021.
- [2] N. Anwar, "Analisis Investigasi Forensik WhatsApp Messenger Smartphone Terhadap WhatsApp Berbasis Web," *Jurnal Ilmu Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 3, pp. 1–10, 2017.
- [3] A. N. Muhammad, "Analisis Sentimen Positif dan Negatif Komentar Video Youtube Menggunakan Metode Naïve Bayes – Support Vector Machine (NBSVM) Classifier'.,", 2019.
- [4] D. Hariyadi, M. Wahyu Indriyanto, and Muhammad Habibi, "Analisis Barang Bukti Digital Aplikasi Paziim Pada Ponsel Paziim Digital Evidence Analysis Application On Android," vol. 2, pp. 52–56, 2019.
- [5] F. I. Tanesab, I. Sembiring, and H. D. Purnomo, "Sentiment Analysis Model Based On Youtube Comment Using Support Vector

- Machine,” *International Journal of Computer Science and Software Engineering (IJCSSE)*, vol. 6 No.8, pp. 180–185, 2017.
- [6] Hariani, “Analisis Bukti Digital Cyberbullying Pada Jejaring Sosial Menggunakan Naïve Bayes Classifier (Nbc),” 2017.
- [7] K. N. Adi and D. Sebastian, “Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 4, pp. 2229–2443, 2018.
- [8] P. Y. Saputra, “Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan,” *Polinema, Jurnal Informatika*, pp. 209–213, 2019.
- [9] M. Muliadin, “Kebijakan Kriminal dalam Menanggulangi Kejahatan Kesusilaan Melalui Internet Pendahuluan masyarakat . Hal ini telah menyebabkan perubahan cara berpikir dan bertindak (Electronic Funds Transfer System atau sistem transfer dana elektronik),” vol. 8 No.2, pp. 167–181, 2020.
- [10] K. Kent, “Penerapan Metode Composite Logic Untuk Perancangan Framework Pengumpulan Bukti Digital Pada Media Sosial,” *ILKOM Jurnal Ilmiah*, pp. 1–121, 2006.
- [11] A. Indriani, “Analisa Perbandingan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor Terhadap Klasifikasi Data,” *Sebatik*, vol. 24, no. 1, pp. 1–7, Jun. 2020, doi: 10.46984/sebatik.v24i1.909.
- [12] M. N. Al Jum’ah, “Penerapan Metode Composite Logic Untuk Perancangan Framework Pengumpulan Bukti Digital Pada Media Sosial,” *ILKOM Jurnal Ilmiah*, vol. 11 No.2, 2019.