

Analisis *Sentiment* Publik di Media Sosial X Terhadap Judi *Online* di Indonesia

Rd Muhammad Sopian Putra Pratama^{1*}, Christina Juliane²

^{1,2} Magister Sistem Informasi, Fakultas Sistem Informasi, STMIK LIKMI, Bandung, Indonesia
e-mail : sopianputrapratama77@gmail.com, christina.juliane@likmi.ac.id

*Corresponding Author

Submitted Date: April 20, 2025

Revised Date: May 30, 2025

Reviewed Date: May 15, 2025

Accepted Date: July 20, 2025

Abstract

This study analyses public perceptions of the phenomenon of online gambling through data from social media platform X (Twitter). A total of 5,846 tweets were collected using the Tweet-harvest tool, which utilises authentication tokens to efficiently obtain large amounts of data. After the pre-processing stage, 4,782 relevant tweets were analysed using the VADER (Valence Aware Dictionary and Sentiment Reasoner) method for labelling positive sentiment and negative sentiment. The labelling results showed that 4,151 tweets (86.9%) contained negative sentiment, while 631 tweets (13.2%) contained positive sentiment. Continuing with the Support Vector Machine (SVM) model, the data was divided into 70% training data and 30% test data using the linear kernel parameter with the best results at C=1.0, producing an accuracy of 97%, with negative sentiment precision of 97%, recall of 100%, and positive sentiment precision of 99% with recall of 79%. Analysis of the confusion matrix shows that the model correctly predicted 1,253 out of 1,255 negative tweets, while for positive sentiment, it correctly predicted 142 out of 180 tweets. These findings reveal the dominance of negative sentiment in public perception of online gambling, which is related to potential social impacts such as addiction and economic losses. The results of this study can serve as a basis for more responsive and data-driven policy-making in addressing the issue of online gambling in Indonesia.

Keywords: *Crawling Data, Online Gambling, Sentiment Analysis, SVM, VADER*

Abstrak

Penelitian ini menganalisis persepsi masyarakat terhadap fenomena judi *online* melalui data dari media sosial X (Twitter). Sebanyak 5.846 tweets dikumpulkan menggunakan alat *Tweet-harvest*, yang memanfaatkan *token autentikasi* untuk memperoleh data dalam jumlah besar secara efisien. Setelah tahap *pre-processing*, sebanyak 4.782 tweet yang relevan dianalisis menggunakan metode VADER (*Valence Aware Dictionary and Sentiment Reasoner*) untuk pelabelan *positive sentiment* dan *negative sentiment*. Hasil pelabelan menunjukkan bahwa 4151 tweets (86,9%) bermuatan *negative sentiment*, sedangkan 631 tweets (13,2%) bermuatan *positive sentiment*. Dilanjutkan dengan menggunakan model *Support Vector Machine* (SVM) data dibagi menjadi 70% data latih dan 30% data uji dengan menggunakan parameter *kernel linear* dengan hasil terbaik pada C=1.0 menghasilkan akurasi sebesar 97%, dengan *precision negative sentiment* sebesar 97%, *recall* 100%, dan *precision positive sentiment* sebesar 99% dengan *recall* 79%. Analisis *confusion matrix* menunjukkan model memprediksi dengan benar 1.253 dari 1.255 tweet *negative*, sementara untuk *sentimen tpositive* memprediksi benar 142 dari 180 tweet. Temuan ini mengungkapkan dominasi *negative sentiment* dalam persepsi publik terhadap judi *online*, yang berkaitan dengan potensi dampak sosial seperti kecanduan dan kerugian ekonomi. Hasil penelitian ini dapat menjadi landasan bagi pembuatan kebijakan yang lebih responsif dan berbasis data dalam menangani permasalahan judi *online* di Indonesia.

Keywords: Analisis Sentimen, *Crawling Data*, Judi *Online*, SVM, VADER



1 Pendahuluan

Fenomena judi *online* mengalami perkembangan pesat dalam sepuluh tahun terakhir, yang didorong oleh adopsi teknologi internet, digitalisasi dan penyebaran perangkat selular yang secara signifikan mempermudah akses ke platform perjudian (Fino et al., 2021). Perkembangan ini turut berkontribusi pada meningkatnya tingkat kecanduan judi *online* dibandingkan dengan judi *offline* (Abbott, 2020). Meskipun perjudian *online* dilarang oleh hukum di Indonesia (Muhammad Rizky Aditia Nugraha & Setiawan, 2023), praktik ini tetap berkembang dengan pesat (Bang Fiu, 2024), menjadikannya sebagai salah satu masalah sosial yang semakin mengkhawatirkan, mengingat dampak *negative* yang ditimbulkannya seperti gangguan psikososial, kesehatan mental yang terganggu, kerugian finansial, memicu kecanduan (Iliceto et al., 2020), meningkatkan risiko tindakan kriminal serta mengancam perlindungan privasi serta keamanan data pribadi pengguna (Ihsanudin et al., 2023).

Berdasarkan data dari PPAATK (Pusat Pelaporan dan Analisis Transaksi Keuangan), pada tahun 2023 tercatat terjadi 168 juta transaksi judi *online* dengan nilai perputaran dana mencapai Rp 327 triliun, sementara total akumulasi transaksi sejak 2017 mencapai Rp 517 triliun (Bang Fiu, 2024). Dari total transaksi tersebut, sekitar 80% dilakukan dengan nilai transaksi di bawah Rp 100 ribu (Nadia Riso, 2024). Penyebaran judi *online* ini sebagian besar dipengaruhi oleh kurangnya pemahaman masyarakat mengenai bahaya yang ditimbulkan, serta faktor-faktor psikologis seperti godaan untuk meraih keuntungan finansial dengan cepat, rasa penasaran, hiburan, pengaruh teman dan mengisi kebosanan (Fino et al., 2021), (Ihsanudin et al., 2023).

Fenomena ini perlu ditelaah lebih lanjut, terutama dengan memperhatikan pandangan masyarakat terkait judi *online* di Indonesia. Di era digital saat ini, masyarakat kerap mengungkapkan pendapat dan opini mereka melalui berbagai platform media sosial, X (dahulu dikenal sebagai Twitter) merupakan salah satu media sosial populer yang digunakan masyarakat luas sebagai sarana berbagi dan bertukar informasi secara *real-time* (Savitri et al., 2021). Indonesia menjadi negara peringkat ke 4 dunia dengan jumlah 27 juta pengguna X pada tahun 2023[9]. Jumlah pengguna X di Indonesia menawarkan potensi yang signifikan untuk menggali aspirasi dan opini masyarakat terkait fenomena judi

online melalui analisis data yang diperoleh dari media sosial. Dengan mengklasifikasikan opini berdasarkan kategori *positive sentiment* dan *negative* (Arsi & Waluyo, 2021).

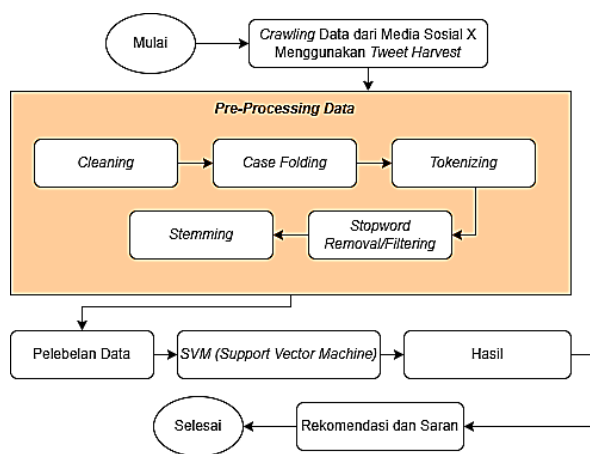
Penelitian sebelumnya telah menerapkan analisis *sentiment* pada data yang diperoleh dari media sosial X, maupun data sekunder (Damayunita et al., 2022), menggunakan pendekatan seperti Naïve Bayes, K-nearest neighbors, dan *Support Vector Machine* (SVM) (Arsi & Waluyo, 2021), (Damayunita et al., 2022), (Maharani & Effendy, 2022), (Gusdiana et al., 2023), (Ramadani et al., 2024). Berdasarkan perbandingan hasil studi, metode SVM terbukti memiliki tingkat akurasi tertinggi mencapai 92% dibandingkan metode lainnya (Damayunita et al., 2022). Keunggulan metode SVM terletak pada kemampuannya dalam mengklasifikasikan opini berdasarkan atribut penilaian yang relevan, sehingga memungkinkan pengelompokan opini secara akurat dalam kategori *positive sentiment* atau *negative* (Siti Aisah et al., 2024). Dalam penelitian ini, pelabelan data latih menggunakan metode VADER (*Valence Aware Dictionary and Sentiment Reasoner*). VADER menganalisis *sentiment* dalam teks dengan menentukan apakah teks tersebut mengandung *positive sentiment* atau *negative*, berdasarkan leksikon yang telah diberi skor *sentiment* untuk setiap kata (Hutto & Gilbert, 2014). Pendekatan ini memungkinkan pemahaman otomatis mengenai *sentiment* yang terkandung dalam teks (Anggraeni et al., 2023), yang kemudian digunakan untuk pelatihan model SVM dalam mengklasifikasikan *sentiment* untuk data sosial yang lebih luas, khususnya terkait dengan fenomena judi *online* di Indonesia.

Penelitian ini bertujuan untuk mengkaji persepsi masyarakat terhadap fenomena judi *online* melalui analisis *sentiment* ada data yang dikumpulkan dari platform media sosial X. Fokus penelitian ini adalah untuk mengklasifikasikan *sentiment* yang terkandung dalam tweet terkait judi *online*, dengan tujuan mengidentifikasi kecenderungan *sentiment negative* dan *positive*. Hasil analisis ini dapat memberikan wawasan yang berguna bagi pihak berwenang, khususnya pemerintah Indonesia, dalam merumuskan kebijakan yang lebih responsif dan berbasis data terkait isu judi *online* di Indonesia.

2 Metode Penelitian

Analisis *sentiment* ini mengklasifikasikan opini masyarakat di X ke dalam dua kategori utama yaitu *positive* dan *negative*. Proses penelitian ini

terdiri dari beberapa tahapan, yaitu: *Crawling* data untuk pengumpulan data, *Pre-processing* Data untuk mempersiapkan data yang siap dianalisis, Pelabelan Data menggunakan metode VADER untuk memberikan label *sentiment* pada dataset, serta analisis *sentiment* dengan menggunakan algoritma SVM untuk klasifikasi sentimen. Hasil dari analisis ini diikuti dengan hasil dan kesimpulan yang merangkum temuan-temuan utama dalam penelitian ini sebagaimana dijelaskan pada Gambar 1.



Gambar 1. Metode Penelitian (Terinspirasi dari (Damayunita et al., 2022))

2.1. Crawling Data

Penelitian ini mengimplementasikan metode *crawling* untuk mengumpulkan data dari platform media sosial X dengan menggunakan alat *tweet-harvest*, yang dijalankan di Google Colab menggunakan pemrograman Python. Penggunaan *tweet-harvest* mempermudah proses *Crawling* data karena hanya membutuhkan token authentication dari akun X, sehingga memungkinkan pengumpulan data dalam volume yang lebih besar dan lebih efisien, serta mengurangi kendala yang biasa ditemui pada batasan akses API konvensional (Sosial et al., 2023). Proses *crawling* ini difokuskan pada pengambilan data relevan dengan kata kunci terkait judi *online* di Indonesia, serta membatasi pencarian pada tweet berbahasa Indonesia untuk memastikan relevansi konteks sosial dan kultural, sekaligus mengurangi bias dalam analisis teks. Data yang dikumpulkan mencakup tweet, retweet, dan komentar mengenai fenomena judi *online*, baik berupa opini, protes, maupun dukungan terhadap kebijakan pemerintah.

2.2. Pre-Processing Data

Tahapan ini bertujuan untuk mengubah data mentah menjadi dataset yang siap dianalisis, yang terdiri dari dua sub-tahapan: *cleaning* data (*general preprocessing*) untuk menghapus duplikasi, dimana tahap ini merupakan tahapan tahap krusial dalam proses *text mining*, mengingat teks yang akan dianalisis seringkali mengandung noise yang dapat mengganggu kualitas dan akurasi hasil analisis (Savitri et al., 2021). Berikutnya tahapan *pre-processing* data untuk seleksi data yang relevan. Berikut adalah urutan langkah-langkah yang dilakukan dalam tahapan *pre-processing* data:

1) Cleaning (Pembersihan data *double*)

Pada tahap ini dilakukan proses pembersihan data yang mencakup penghapusan duplikasi data (*double*) yang dihasilkan dari retweet.

2) Cleaning (Data Text)

Pada tahap ini merupakan penghilangan elemen-elemen yang tidak relevan seperti Uniform Resource Locator (URL), username, tanda retweet (RT), karakter HTML, dan hashtag.

3) Case Folding

Proses ini mengkonversi seluruh teks dalam dokumen menjadi huruf kecil (*lowercase*), untuk mengurangi variasi dalam representasi kata yang dapat mempengaruhi hasil analisis. Dengan demikian, teks menjadi lebih seragam dan konsisten dalam bentuk standarnya.

4) Tokenizing

Pada tahap tokenisasi, teks dipecah menjadi unit-unit terkecil berupa kata atau token dengan menggunakan batasan tanda baca dan spasi sebagai pemisah. Tokenisasi memungkinkan pemisahan informasi dalam teks menjadi bagian-bagian yang lebih terstruktur dan dapat dianalisis lebih lanjut.

5) Stopword Removal/Filtering

Tahap ini bertujuan untuk menghilangkan kata-kata yang tidak memberikan kontribusi berarti dalam analisis, seperti kata-kata umum yang tercantum dalam daftar *stopword* (misalnya "dan", "atau", "dengan"). Penghilangan *stopword* ini bertujuan untuk meningkatkan akurasi dan efisiensi dalam analisis data teks.

6) Stemming

Proses *stemming* dilakukan untuk mengubah kata-kata yang merupakan bentuk derivatif atau turunan menjadi bentuk dasar atau akar kata. Hal ini

bertujuan untuk mengurangi variasi morfologis dalam teks, sehingga kata-kata dengan makna yang sama dapat dikelompokkan menjadi satu entitas yang lebih konsisten.

2.3. Pelabelan Data Dengan *Valence Aware Dictionary And Sentiment Reasoner (VADER)*

Valence Aware Dictionary and Sentiment Reasoner (VADER) digunakan untuk pelabelan data *sentiment* pada penelitian ini, sebuah model berbasis aturan yang dikembangkan oleh Hutto dan Gilbert (Hutto & Gilbert, 2014). VADER mengklasifikasikan teks sosial menjadi dua kategori *sentimen*, yaitu *positive* dan *negative*, berdasarkan skor *compound* yang mengukur intensitas emosional dari teks tersebut. Model ini mempertimbangkan kata-kata dengan intensitas emosional yang kuat, sehingga sangat cocok untuk analisis *sentiment* pada teks sosial yang tidak terstruktur, seperti tweet terkait judi *online*. Pendekatan ini dipilih karena efisiensinya dalam memberikan label *sentiment* secara cepat dan akurat

2.4. Support Vector Machine (SVM)

Setelah tahap pemrosesan data, langkah berikutnya adalah pembangunan model machine learning. Algoritma machine learning dilatih menggunakan data latih untuk memahami pola-pola pada setiap fitur dalam token, sehingga model dapat mengklasifikasikan *sentiment* dengan akurat. Pada penelitian ini, model yang digunakan adalah model supervised learning yang bertujuan untuk menganalisis *sentiment* terkait topik judi *online* di mana setiap token sudah memiliki label yang sesuai. Algoritma yang diterapkan adalah *Support Vector Machine (SVM)*, yang dikenal efektif dalam klasifikasi data dengan dimensi tinggi dan memiliki kinerja yang baik pada data yang memiliki margin pemisah yang jelas. SVM dipilih karena kemampuannya dalam menangani masalah klasifikasi yang kompleks dengan margin yang optimal, memberikan performa yang baik pada dataset yang terbatas atau tidak terstruktur (Savitri et al., 2021). Hasil dari algoritma SVM perlu dievaluasi secara teliti untuk menganalisis kinerja model dalam melakukan klasifikasi. Kinerja klasifikasi dapat diukur menggunakan metrik seperti *accuracy*, *sensitivity*, dan *specificity*. Selain itu, beberapa metrik lain yang diambil dari *confusion matrix*, seperti *precision*, *recall*, *F1-score*, *ROC*, *informedness*, *markedness*, dan *correlation*, juga digunakan untuk mengevaluasi performa model secara lebih komprehensif.

Confusion matrix sendiri adalah tabel yang mendefinisikan hasil klasifikasi, yang menggambarkan jumlah *True Positives (TP)*, *False Positives (FP)*, *False Negatives (FN)*, dan *True Negatives (TN)* (Hand et al., 2021), yang sangat berguna dalam mengevaluasi efektivitas model SVM dalam memisahkan kelas dengan jelas dan akurat (Savitri et al., 2021).

3. Hasil dan Pembahasan

Bagian ini membahas hasil analisis *sentiment* dan pembahasan dari proses *crawling* data, *pre-processing* data, pelabelan dan hasil analisis *sentiment* menggunakan *Support Vector Machine (SVM)* untuk mengklasifikasikan *sentimen* masyarakat terkait fenomena judi *online* dan mengevaluasi efektivitas model SVM dalam mengidentifikasi *sentiment* berdasarkan data yang diperoleh guna memberikan rekomendasi dan saran.

3.1. Crawling Data

Penelitian ini mengimplementasikan metode *crawling* untuk mengumpulkan data dari platform media sosial X dengan menggunakan alat *tweet-harvest*, yang dijalankan di Google Colab menggunakan pemrograman Python. Penggunaan *tweet-harvest* mempermudah proses *Crawling* data karena hanya membutuhkan autentikasi token dari akun X, sehingga memungkinkan pengumpulan data dalam volume yang lebih besar dan lebih efisien, serta mengurangi kendala yang biasa ditemui pada batasan akses API konvensional (Hand et al., 2021). Proses *crawling* ini difokuskan pada pengambilan data relevan dengan kata kunci terkait judi *online* di Indonesia, serta membatasi pencarian pada tweet berbahasa Indonesia untuk memastikan relevansi konteks sosial dan kultural, sekaligus mengurangi bias dalam analisis teks. Data yang dikumpulkan mencakup tweet, retweet, dan komentar mengenai fenomena judi *online*, baik berupa opini, protes, maupun dukungan terhadap kebijakan pemerintah. Penelitian ini berhasil mengumpulkan 5.846 tweet yang diposting antara 8 hingga 15 November 2024, periode yang bertepatan dengan pemberitaan terkait pernyataan Presiden Prabowo mengenai komitmennya untuk memberantas judi *online* dan komentar Kejaksaan Agung yang menyebutkan keterlibatan anggotanya dalam judi *online* sebagai tindakan iseng. Data ini digunakan untuk menganalisis persepsi publik terhadap isu judi *online* pada periode tersebut, dengan Tabel 1 menyajikan contoh tweet yang dianalisis dalam penelitian ini.

Tabel 1. Contoh Tweets Hasil *Crawling* Data

No	Tweet
1	Perjudian <i>online</i> sebagai dampak dari era digital telah menciptakan ancaman serius terhadap generasi muda dan struktur sosial. #JudiOnline #Judol #Komdigi #SitusJudi #BandarJudiA #IndonesiaEmas #LawanJudol https://t.co/ZI38SE35YV
2	Prabowo: Tak boleh ada yang lindungi judi <i>online</i> #BerantasJudiOnline https://t.co/8ipktStSpk
3	Psikiater Sebut Kecanduan Judi <i>Online</i> adalah Gangguan Kejiwaan Ini Penjelasananya https://t.co/BBdCeAD03h
.....	
.....	
5845	OLX88 Slot <i>Online</i> Mudah Menang dan Banjir Scatter Gas Langsung Daftar Sekarang di Situs OLX88 Link Daftar : https://t.co/i46bpXLFJ5 https://t.co/IZ6mAzJBN4
5846	RESULT #BANDUNGPOOLS SABTU 16 NOVEMBER 2024 4363 SHIO : HARIMAU SELAMAT KEPADA PEMENANG SALAM JP #HUKUMTOTO https://t.co/OYG91Ml8mo

3.2. Pre-Processing Data

Berikut adalah urutan langkah-langkah yang dilakukan dalam tahapan *pre-processing* data:

1) *Cleaning* (Pembersihan data *double*)

Pada tahap ini dilakukan proses pembersihan data yang mencakup penghapusan duplikasi data (*double*) yang dihasilkan dari retweet. Hasil pembersihan data *double* bisa dilihat pada Gambar 2 dibawah ini.

Hasil pembersihan data *double* bisa dilihat pada Gambar 2 Pada baris pertama, kode ini digunakan untuk menghapus duplikasi dalam dataset berdasarkan kolom "full_text", dengan hanya mempertahankan entri pertama dari setiap teks yang sama. Hal ini dilakukan untuk memastikan bahwa hanya data teks yang unik yang digunakan dalam analisis, menghindari distorsi yang dapat disebabkan oleh data yang terduplikasi. Dengan menggunakan parameter *inplace=True*, perubahan ini diterapkan langsung pada dataset tanpa perlu membuat salinan baru. Setelah itu, pada

baris berikutnya, fungsi digunakan untuk menampilkan informasi umum tentang struktur dataset, termasuk jumlah baris, kolom, tipe data, serta jumlah nilai yang tidak kosong pada setiap kolom. Langkah ini penting untuk memverifikasi apakah pembersihan data berhasil dan untuk memastikan integritas data yang akan dianalisis lebih lanjut.

```

Menghapus data double

data.drop_duplicates(subset = "full_text", keep = 'first', inplace = True)
data.info()

<class 'pandas.core.frame.DataFrame'>
Index: 4782 entries, 0 to 5845
Data columns (total 15 columns):
 #   Column              Non-Null Count  Dtype  
---  -
 0   conversation_id_str  4782 non-null  float64
 1   created_at          4782 non-null  object 
 2   favorite_count      4782 non-null  int64  
 3   full_text           4782 non-null  object 
 4   id_str              4782 non-null  float64
 5   image_url           913 non-null   object 
 6   in_reply_to_screen_name 2475 non-null object 
 7   lang                4782 non-null  object 
 8   location            2374 non-null object 
 9   quote_count         4782 non-null  int64  
10  reply_count         4782 non-null  int64  
11  retweet_count        4782 non-null  int64  
12  tweet_url           4782 non-null  object 
13  user_id_str         4782 non-null  float64
14  username            4782 non-null  object 
dtypes: float64(3), int64(4), object(8)
memory usage: 597.8+ KB

```

Gambar 2. Proses Penghapusan data *double*

2) *Cleaning* (data *text*)

Pada tahap ini merupakan penghilangan elemen-elemen yang tidak relevan seperti Uniform Resource Locator (URL), username, tanda retweet (RT), karakter HyperText Markup Language (HTML), dan hashtag. Tabel 2 merupakan perbedaan tweets sebelum dan sesudah di *cleaning*.

Tabel 2. Contoh Tweets Sebelum dan Sesudah *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
Hampir 100 Orang	Hampir Orang
Dirawat di RSCM	Dirawat di RSCM
karena Kecanduan Judi	karena Kecanduan
<i>Online</i>	Judi <i>Online</i>
https://t.co/uSELetTk7m	

Setelah dilakukan pembersihan, hasilnya adalah teks yang lebih bersih dan lebih relevan untuk analisis, seperti yang terlihat pada kalimat "Hampir 100 Orang Dirawat di RSCM karena Kecanduan Judi *Online*", yang sudah bebas dari URL dan tanda retweet. Dengan pembersihan ini, analisis *sentiment* atau pemahaman teks menjadi lebih akurat, karena hanya berfokus pada informasi inti yang disampaikan dalam kalimat tersebut.

3) Case Folding

Proses ini mengkonversi seluruh teks dalam dokumen menjadi huruf kecil (*lowercase*), untuk mengurangi variasi dalam representasi kata yang dapat mempengaruhi hasil analisis. Dengan demikian, teks menjadi lebih seragam dan konsisten dalam bentuk standarnya. Tabel 3 merupakan perbedaan tweets sebelum dan sesudah *Case Folding*.

Tabel 3. Contoh Tweets Sebelum dan Sesudah *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
Hampir Orang	hampir orang
Dirawat di RSCM	dirawat di rscm
karena Kecanduan	karena kecanduan
Judi Online	judi online

Setelah *Case Folding*, perbedaan kapitalisasi pada teks dihilangkan dengan mengubah semua huruf menjadi huruf kecil, sehingga teks menjadi lebih konsisten dan seragam.

4) Tokenizing

Pada tahap tokenisasi, teks dipecah menjadi unit-unit terkecil berupa kata atau token dengan menggunakan batasan tanda baca dan spasi sebagai pemisah. Tokenisasi memungkinkan pemisahan informasi dalam teks menjadi bagian-bagian yang lebih terstruktur dan dapat dianalisis lebih lanjut. Tabel 4 merupakan perbedaan tweets sebelum dan sesudah *tokenizing*.

Tabel 4. Contoh Tweets Sebelum dan Sesudah *Tokenizing*

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
hampir orang dirawat di rscm karena kecanduan judi online	['hampir', 'orang', 'dirawat', 'di', 'rscm', 'karena', 'kecanduan', 'judi', 'online']

Sebelum *tokenizing*, teks merupakan satu kesatuan kalimat utuh, sedangkan setelah *tokenizing*, teks dipisahkan menjadi unit-unit kata yang lebih kecil atau tokens, seperti dalam contoh yang terlihat, di mana setiap kata dalam kalimat dipisahkan dan diubah menjadi elemen dalam sebuah daftar. Hal ini memungkinkan analisis lebih lanjut pada tingkat kata, seperti analisis frekuensi atau klasifikasi kata.

5) Stopword Removal/Filtering

Tahap ini bertujuan untuk menghilangkan kata-kata yang tidak memberikan kontribusi berarti dalam analisis, seperti kata-kata umum yang tercantum dalam daftar stopwords (misalnya "dan", "atau", "dengan"). Penghilangan stopwords ini bertujuan untuk meningkatkan akurasi dan efisiensi dalam analisis data teks. Tabel 5 merupakan perbedaan tweets sebelum dan sesudah *Stopword Removal/Filtering*.

Tabel 5. Contoh Tweets Sebelum dan Sesudah *Stopword Removal/Filtering*

Sebelum <i>Filtering/Stopword Removal</i>	Sesudah <i>Filtering/Stopword Removal</i>
['hampir', 'orang', 'dirawat', 'di', 'rscm', 'karena', 'kecanduan', 'judi', 'online']	['orang', 'dirawat', 'rscm', 'kecanduan', 'judi', 'online']

Setelah *filtering/stopword removal*, kata-kata yang dianggap tidak penting atau umum (seperti preposisi dan konjungsi) dihapus, sehingga hanya kata-kata yang lebih relevan untuk analisis yang tetap ada. Hal ini meningkatkan efisiensi dan akurasi dalam pemrosesan teks lebih lanjut.

6) Stemming

Proses *stemming* dilakukan untuk mengubah kata-kata yang merupakan bentuk derivatif atau turunan menjadi bentuk dasar atau akar kata. Hal ini bertujuan untuk mengurangi variasi morfologis dalam teks, sehingga kata-kata dengan makna yang sama dapat dikelompokkan menjadi satu entitas yang lebih konsisten. Tabel 6 merupakan perbedaan tweets sebelum dan sesudah *stemming*.

Tabel 6. Contoh Tweets Sebelum dan Sesudah *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
['hampir', 'orang', 'dirawat', 'di', 'rscm', 'karena', 'kecanduan', 'judi', 'online']	orang rawat rscm candu judi online

3.3. Pelabelan Data Dengan VADER

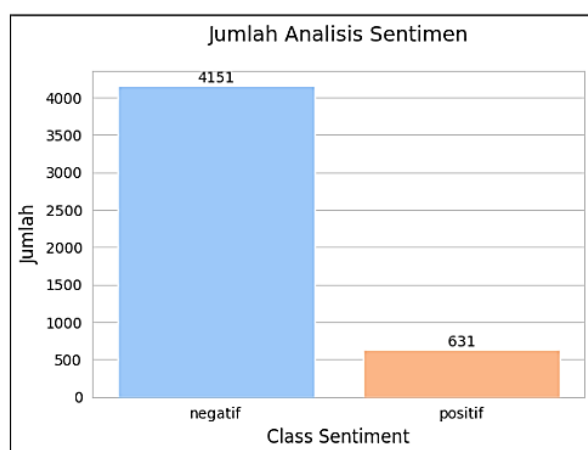
Pada tahap ini data yang sudah melalui tahap *pre-processing* diberikan pelabelan data *sentiment*

menggunakan VADER, Tabel 7 merupakan beberapa hasil pelabelan dengan VADER.

Tabel 7. Contoh Tweets yang Sudah Dilabeli VADER

Tweets	Sentimen
orang rawat rscm candu judi <i>online</i>	Negatif
daftar amp main slot asli gacor sapud situs slot <i>online</i> amp taruh live lengkap aman amp percaya jackpot bayar lunas buru main bos link akses sapud	Positif
slot - link slot <i>online</i> deposit amp situs gacor gampang jp slot link situs slot gacor <i>online</i> minimal deposit murah k via dana dapat bonus depo menang jp maxwin	Positif
gerak cepat anggota tni hukum libat skandal judi <i>online</i> sanksi	Negatif
basmi tumpas judi <i>online</i> judolbencananasional	Negatif

Hasil dari tahap pelabelan didapatkan 4151 tweets diklasifikasikan sebagai *negative* dan 631 diklasifikasikan sebagai *positive*. Data inilah yang akan digunakan untuk melatih model pembelajaran mesin menggunakan SVM. Gambar 3 merupakan diagram tweets *positive* dan *negative*.



Gambar 3. Class *Sentiment* dengan VADER

3.4. Data Latih Dan Data Uji

Sebagai langkah pertama, data yang telah disiapkan dibagi menjadi dua subset, yaitu data

latihan data uji, untuk memastikan model dapat dilatih secara optimal dengan data latih dan dievaluasi kinerjanya menggunakan data uji. Dalam penelitian ini, proporsi yang digunakan adalah 70% untuk data latih dan 30% untuk data uji dengan jumlah data latih: 3347 dan jumlah data uji: 1435.

3.5. Pelatihan Dan Pengujian Model

Tahap kedua menggunakan kernel linear untuk menganalisis *sentiment* terkait judi *online* di Twitter yang dijelaskan pada tabel 8.

Tabel 8. Parameter *Kernel Linear*

Parameter Model SVM:			
kernel:linear			
c: 1.0			
Support Vector:	(0, 530)	0.25742272357777507	

Pelatihan data dilakukan menggunakan kernel linier dengan parameter terbaik, yaitu pada nilai $C = 1.0$.

Dalam evaluasi model yang dilakukan berhasil mencapai akurasi sebesar 97%. Berdasarkan *classification report*, model menunjukkan kinerja yang sangat baik dalam mengklasifikasikan *sentiment negative* dengan presisi tinggi sebesar 97% dan *recall* hampir sempurna, yakni 100%. Sementara itu, untuk *sentiment positive*, meskipun memiliki presisi yang sangat baik sebesar 99%, *recall*-nya sedikit lebih rendah, yakni 79%. Temuan ini mengindikasikan bahwa model ini sangat efektif dalam mendeteksi *sentiment negative*, namun masih memiliki ruang untuk perbaikan dalam mengidentifikasi *positive sentiment* secara lebih konsisten.

Tabel 9. Hasil Evaluasi Model

Akurasi: 0.9721254355400697				
Classification Report:				
	precision	recall	f1-score	support
negatif	0.97	1.00	0.98	1255
positif	0.99	0.79	0.88	180
accuracy			0.97	1435
macro avg	0.98	0.89	0.93	1435
weighted avg	0.97	0.97	0.97	1435
ConfusionMatrix:				
[[1253 2] [38 142]]				

Dalam visualisasi data, digunakan word cloud untuk menggambarkan kata-kata yang paling sering muncul dalam tweet dengan *positive sentiment* dan *negative*. *Word cloud* ini memudahkan identifikasi cepat terhadap kata-kata dominan yang muncul dalam setiap *sentimen*, memberikan gambaran jelas tentang topik atau istilah yang sering dibahas oleh pengguna. Dengan demikian, visualisasi ini memperkaya pemahaman terhadap distribusi *sentiment* dalam data, serta memberikan wawasan tambahan mengenai aspek-aspek yang paling banyak diperbincangkan dalam tweet *positive* maupun *negative*.

Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	1253	2
Actual Positive	38	142

Dilakukan analisis secara manual dengan cara membaca untuk mengetahui isi tweets yang terlabeli *positive* mengenai judi *online*. Dapat

disimpulkan bahwa semua tweets berlabel *positive* teridentifikasi iklan atau ajakan untuk bermain judi. Iklan-iklan ini biasanya disertai dengan kata-kata yang sedang ramai diperbincangkan pada waktu tweet itu diposting. Hal ini menunjukkan bahwa para pelaku iklan berusaha memanfaatkan tren atau topik yang sedang populer untuk menarik pengguna X. Tabel 10 menunjukkan contoh tweets yang berlabel *positive*.

Tabel 10. Contoh Tweets Berlabel *Positive*

Tweets Berlabel Positif
main timah udah konek mudah jp nya amanah link nya semuaorang slotonline slotgacor slotgacorhariini slotpgwallet beritaterkini pik beritaterkini viral menangtiadahenti timahcuanabis link daftar
gak bosen bosennya nih main wild bandito gas main pt rasa cuan alir batas link login klik pt slothoki rtpertinggi slotgacor slotonline spin autocuan cuanterus autocuan cuanterus maxwin
result bandungpools sabtu november shio harimau selamat menang salam jp hukumtoto
slot gacor <i>online</i> baik percaya situs mudah main member langsung hoki daftar bukti hoki juga slot slotgame bijakd viral fyp
muraipoker login situs slot gacor jam <i>online</i> baik milik bonus limpah tunggu main rasa sensasi pecah tiada henti nya link daftar slotonline slotgacorindonesia malamjumatcrot slotgacorhariini

Sebagai langkah selanjutnya, disarankan untuk lebih memfokuskan pada identifikasi dan pengawasan tweet yang mengandung promosi ilegal melalui penerapan algoritma deteksi otomatis yang mampu mengenali pola bahasa dalam iklan judi. Untuk mendukung hal ini, penting juga melakukan kerja sama dengan platform media sosial untuk memblokir atau menurunkan visibilitas tweet yang mempromosikan judi *online*. Selain itu, kolaborasi dengan *influencer* dan konten kreator dalam kampanye anti-judi *online* dapat memperluas jangkauan kesadaran masyarakat, sementara penggunaan teknologi machine learning yang lebih canggih dapat meningkatkan akurasi dalam mendeteksi pola bahasa promosi judi. Pengembangan kebijakan sosial yang lebih ketat dari platform media sosial terkait konten judi juga diperlukan untuk mempersempit ruang bagi promosi ilegal. Disisi lain, penting untuk menyediakan alat pelaporan yang mudah diakses oleh pengguna untuk melaporkan tweet yang

berpotensi merugikan. Tak kalah pentingnya, penelitian lebih lanjut mengenai dampak sosial media terhadap perilaku pengguna serta pembentukan komunitas digital yang mendukung pemberantasan judi *online* akan memperkuat upaya menciptakan kesadaran yang lebih besar tentang bahaya perjudian *online*.

4. Kesimpulan

Penelitian ini menganalisis *sentiment* terkait judi *online* di Indonesia melalui media sosial. Diawali dengan melakukan *Crawling* data menggunakan tweets harvest yang berhasil memperoleh 5.846 tweets terkait judi *online*. Dilanjutkan ke tahap *Pre-processing* data yang menghasilkan 4.782 tweet yang lebih bersih yang digunakan untuk pelabelan menggunakan VADER yang menghasilkan 4151 tweet *negative* dan 631 tweet *positive* yang digunakan untuk melatih model *machine learning* dengan SVM. Pada tahap SVM data sebelumnya dibagi menjadi 70% data latih dan 30% data uji dengan menggunakan parameter kernel linear dengan hasil terbaik pada $C=1.0$. Berikutnya dilakukan evaluasi model yang menghasilkan akurasi 97% serta untuk classification report, *sentimentnegative* memiliki presisi tinggi sebesar 97% dan *recall* sempurna 100%, sebaliknya *positive sentiment* sedikit lebih rendah dengan presisi 99% dengan *recall* 79%. Tahap terakhir untuk *CONFUSION MATRIX* model mampu memprediksi 1253 dari 1255 tweet *negative* dengan benar, namun terdapat kesalahan pada 2 tweets, sedangkan untuk *positive sentiment* 142 prediksi benar dari total 180. *Sentimentnegative* yang dominan mengindikasikan ketidaksetujuan publik terhadap keberadaan judi *online*, yang berkaitan dengan potensi dampak sosial seperti kecanduan, kesehatan mental dan kerugian ekonomi. Selain itu, ketidakpercayaan terhadap efektivitas penegakan hukum juga tercermin dalam temuan ini. Berdasarkan hasil ini, disarankan agar pemerintah memperkuat penegakan hukum terhadap judi *online* dengan peningkatan pengawasan, tindakan tegas dan adil terhadap situs judi *online* serta orang-orang yang terlibat dalam mempromosikan judi *online* dan pihak berwajib bisa berkolaborasi dengan penyedia layanan internet untuk pemblokiran akses. Edukasi publik mengenai risiko judi *online* juga harus diperkuat guna mengurangi dampak *negative* yang ditimbulkan

References

- Abbott, M. W. (2020). The changing epidemiology of gambling disorder and gambling-related harm: public health implications. *Public Health*, 184(May), 41–45. <https://doi.org/10.1016/j.puhe.2020.04.003>
- Anggraeni, W. A., Fahru Roji, F., & Alkautsar, M. (2023). Analisis *Sentiment* Publik terhadap Kebijakan Insentif Perpajakan Dengan Pendekatan VADER (*Valence Aware Dictionary and Sentiment Reasoner*). *Jurnal Proaksi*, 10(4), 465–477. <https://doi.org/10.32534/jpk.v10i4.4732>
- Arsi, P., & Waluyo, R. (2021). Analisis *Sentiment* Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine* (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Bang Fiu. (2024). Gawat! Jumlah Fantastis Usia Anak Main Judi *Online*. <https://www.ppatk.go.id/News/Read/1373/Gawat-Jumlah-Fantastis-Usia-Anak-Main-Judi-Online.Html>. <https://www.ppatk.go.id/news/read/1373/gawat-jumlah-fantastis-usia-anak-main-judi-online.html>
- Damayunita, A., Fuadi, R. S., & Juliane, C. (2022). Comparative Analysis of Naive Bayes, K-Nearest Neighbors (KNN), and *Support Vector Machine* (SVM) Algorithms for Classification of Heart Disease Patients. *Jurnal Online Informatika*, 7(2), 219–225. <https://doi.org/10.15575/join.v7i2.919>
- Fino, E., Hanna-Khalil, B., & Griffiths, M. D. (2021). Exploring the public's perception of gambling addiction on Twitter during the COVID-19 pandemic: Topic modelling and *sentiment* analysis. *Journal of Addictive Diseases*, 39(4), 489–503. <https://doi.org/10.1080/10550887.2021.1897064>
- Gusdiana, R., Alfian, I., & Juliane, C. (2023). Implementation of *Text Processing* for *Sentiment* Analysis of Tax Payment Interest After the “Rubicon” Phenomenon. *Jurnal Teknik Informatika (Jutif)*, 4(5), 1157–1164. <https://doi.org/10.52436/1.jutif.2023.4.5.1014>
- Hand, D. J., Christen, P., & Kirielle, N. (2021). F*: an interpretable transformation of the F-measure. *Machine Learning*, 110(3), 451–456. <https://doi.org/10.1007/s10994-021-05964-1>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for. Eighth International AAAI Conference on Weblogs and Social Media, 216–225. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
- Ihsanudin, R., Dewi, D., & Adriansyah, M. (2023). Maraknya Judi *Online* Di Kalangan Remaja Kelurahan Derwati Kecamatan Rancasari Kota Bandung. *Jurnal Cerdik: Jurnal Pendidikan Dan Pengajaran*, 3(1), 73–87. <https://doi.org/10.21776/ub.jccerdik.2023.003.01.08>
- Iliceto, P., Fino, E., Schiavella, M., & Candilera, G. (2020). Individual differences in interpersonal security predict suicidal ideation and problem gambling. *Personality and Individual Differences*, 162(January), 110031. <https://doi.org/10.1016/j.paid.2020.110031>
- Maharani, W., & Effendy, V. (2022). Big five personality prediction based in Indonesian tweets using machine learning methods. *International Journal of Electrical and Computer Engineering*, 12(2), 1973–1981. <https://doi.org/10.11591/ijece.v12i2.pp1973-1981>
- Muhammad Rizky Aditia Nugraha, & Setiawan, D. A. (2023). Penegakan Hukum Tindak Pidana Perjudian *Online* di Kabupaten Tasikmalaya Dihubungkan dengan Undang-Undang Nomor 19 Tahun 2016 Tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik. *Bandung Conference Series: Law Studies*, 3(1), 81–86. <https://doi.org/10.29313/bcsls.v3i1.4916>
- Nadia Riso, A. R. (2024). Menko Polhukam: 3,2 Juta Orang Main Judi *Online*, Transaksi di Bawah Rp 100 Ribu. *KumparanNEWS*. <https://kumparan.com/kumparannews/menko-polhukam-3-2-juta-orang-main-judi-online-transaksi-di-bawah-rp-100-ribu-22bWXfkWAg>
- Ramadani, N. C., Tahyudin, I., & Shouni Barkah, A. (2024). Perbandingan Algoritma *Support Vector Machine*, Decision Tree, dan Logistic Regresion Pada Analisis *Sentiment* Ulasan Aplikasi Netflix. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 10(2), 110–117. <https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/2746>
- Savitri, N. L. P. C., Rahman, R. A., Venyutzky, R., & Rakhmawati, N. A. (2021). Analisis Klasifikasi *Sentiment* Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1), 47–58. <https://doi.org/10.28932/jutisi.v7i1.3216>
- Siti Aisah, I., Irawan, B., & Suprpti, T. (2024). Algoritma *Support Vector Machine* (Svm) Untuk Analisis *Sentiment* Ulasan Aplikasi Al Qur'an Digital. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3759–3765. <https://doi.org/10.36040/jati.v7i6.8263>
- Sosial, J., Di, B., & Pertanahan, K. (2023). *Jurnal widya bhumi*. 3(2), 113–136.