

ANALISIS SENTIMEN MASYARAKAT TERHADAP KASUS JUDI ONLINE MENGGUNAKAN DATA DARI MEDIA SOSIAL X PENDEKATAN NAIVE BAYES DAN SVM

M Febrian As Shidiq¹, Debby Alita²

Sistem Informasi, Universitas Teknokrat Indonesia

Jl. ZA Pagar Alam No.9-11. Labuhan Ratu. Kec. Kedaton. Kota Bandar Lampung

e-mail: *¹m_febrían_as_shidiq@teknokrat.ac.id,
²debbyalita@teknokrat.ac.id

Abstract

Research conducted by analyzing public sentiment related to online gambling cases using datasets from *x* social media using the naïve bayes method approach and support vector machine (SVM). The analysis phase starts with data gathering or crawling, followed by data labeling, data preprocessing, and ultimately method categorization. The dataset comprises 2,866 tweets, with 1,436 classified as positive (50.12%) and 1,429 as negative (49.88%). The data before to the classification process is partitioned into training data and testing data, including 70% training data and 30% testing data. The analysis with the SVM approach yielded a classification accuracy of 83%, whereas the naïve Bayes method achieved just 79%. Upon completion of the method classification process, the subsequent phase involves visualization and assessment. During the visualization step, bar plots, word clouds, and word frequencies derived from sentiment analysis calculations are shown, alongside a visualization of words from the dataset. The investigation indicates that the SVM approach outperforms Naive Bayes in sentiment classification. The benefit of SVM resides in its capability to manage data with elevated limits and accuracy, enhancing its efficiency in discerning positive and negative thoughts. The findings of this study demonstrate that SVM is better appropriate for data exhibiting complicated distributions, whereas the Naive Bayes approach yields suboptimal results. Thus, SVM can be proposed as a more appropriate and reliable approach for similar sentiment analysis in the future.

Keyword: naive bayes, online gambling, sentiment analysis, svm, *x* social media

PENDAHULUAN

Pada Perkembangan teknologi informasi saat ini sangat cepat, termasuk di indonesia. Teknologi ini dirancang untuk memudahkan berbagai aktivitas manusia, seperti memproses dan menganalisis data guna menghasilkan informasi yang relevan, cepat, dan akurat (Rahmansyah & Darwis, 2020). Penggunaannya telah meluas di lembaga pemerintah, perusahaan swasta, dan institusi lainnya. Teknologi telah menjadi bagian tak terpisahkan dari kehidupan manusia. Kemajuan teknologi yang pesat kini meningkatkan dan menyederhanakan beberapa aspek kehidupan manusia (Sulistiani et al., 2020). Selain itu, teknologi informasi juga membuka peluang baru di masyarakat, termasuk di dunia bisnis, di mana para entrepreneur memanfaatkannya untuk mengembangkan usaha mereka (Yana Siregar et al., 2020). Internet memudahkan manusia dalam berbagai aktivitas, namun seiring perkembangan zaman, semakin banyak pihak yang menyalahgunakannya untuk kepentingan pribadi dengan merugikan orang lain. Dampak buruk dari tindakan ini sudah dirasakan banyak orang, seperti penipuan, peretasan, phishing, dan berbagai bentuk serangan lainnya (Addiyansyah et al., 2023).

Kemajuan teknologi dan komunikasi telah membawa dampak seperti munculnya perjudian online. Awalnya, teknologi hanya mendukung permainan online melalui *smartphone* yang terhubung ke internet (Tasya Jadidah et al., 2023). Namun, seiring waktu, ini berkembang menjadi perjudian online yang menawarkan keuntungan bagi pemainnya. Judi sendiri merupakan bentuk penyimpangan sosial, di mana seseorang mempertaruhkan miliknya dalam permainan dengan harapan memenangkan kembali barang tersebut, tetapi jika kalah, semua yang dipertaruhkan akan hilang (R. H. Wibowo & Ikhsan, 2020).

Aplikasi *x* menjadi wadah untuk melakukan riset penelitian ini dengan mengangkat topik mengenai judi online yang semakin hari terus membesar. *X* merupakan media sosial

dengan komentar opini masyarakat yang memberikan informasi secara *real time*, sehingga data yang diperoleh akan selalu terbaru dan update. *x* merupakan platform mikroblog yang banyak digunakan karena aksesibilitasnya yang mudah digunakan. Berbeda dengan platform media sosial lain seperti MySpace atau Facebook, *X* memungkinkan pengguna untuk saling menanggapi. Pengguna sering kali memulai dengan "tweet", pesan singkat yang dibatasi hingga 140 karakter. Selanjutnya, "RT" atau retweet menandakan tindakan memposting ulang tweet orang lain. Simbol "@" menandakan penyebutan pengguna, "D" menandakan pesan langsung untuk komunikasi pribadi, dan "#" atau tagar mengidentifikasi tema utama untuk meningkatkan kemudahan penemuan. (Rakhmah & Putri, 2019).

Penelitian sebelumnya yang dilakukan oleh (Julianti et al., 2024) dengan judul "penerapan natural language processing pada analisis sentimen judi online di media sosial twitter" Dengan menggunakan pendekatan Naïve Bayes dan pemrosesan bahasa alami, tingkat akurasi mencapai 76%. Studi tambahan dilakukan (Oktavia et al., 2023) Penerapan sistem e-tiket di *x* yang dianalisis menggunakan pendekatan SVM menghasilkan akurasi sebesar 74%, dengan kategori netral yang berlaku. Kemudian analisis yang dilakukan (Muhammadin & Sobari, 2021) dengan judul "analisis sentimen pada ulasan aplikasi kredivo dengan algoritma SVM Dan NBC", dengan membandingkan kedua metode tersebut dengan data sebanyak 10.000 ulasan yang diambil dari *google play store* meraih nilai akurasi 83,3 untuk metode SVM, dan *naïve bayes* meraih akurasi 80,8%, jadi disimpulkan bahwa SVM lebih baik dibandingkan dengan *naïve bayes* untuk klasifikasi dan analisis sebuah data yang cukup besar.

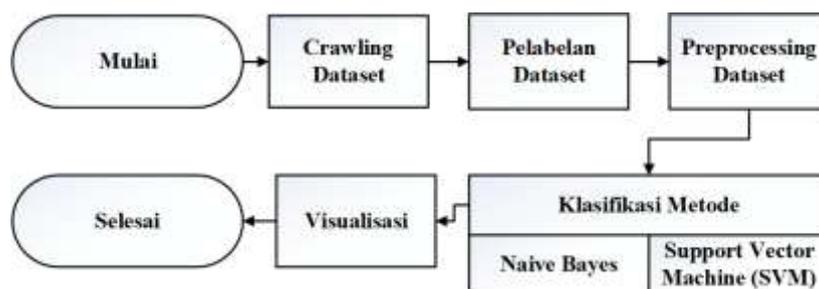
Analisis sentimen penting untuk menyaring komentar di media sosial, dengan tujuan mengidentifikasi komentar yang bersifat *negatif* maupun *positif*. Analisis sentimen adalah teknik pemrosesan bahasa alami yang mengidentifikasi perasaan dan pikiran seseorang seputar subjek tertentu (Hasiholan et al., 2022). Tahap awal melibatkan pengumpulan data atau perayapan data. *X* adalah platform media sosial yang digunakan untuk penyebaran informasi (Wongkar & Angdressey, 2019). Studi ini bertujuan untuk memastikan sentimen pengguna *x* terhadap insiden perjudian daring melalui analisis sentimen, yang mengekstrak wawasan penting dari data tak terstruktur.

Penelitian ini menggunakan perbandingan komparatif antara metodologi naïve Bayes dan support vector machine (SVM). Pendekatan klasifikasi terdiri dari beberapa langkah awal, termasuk pelabelan data dan pemrosesan data, yang memerlukan pembersihan, pelipatan kasus, tokenisasi, dan stemming. Kumpulan data selanjutnya akan dievaluasi dengan membandingkan berbagai metodologi. Naïve Bayes adalah teknik klasifikasi yang menggunakan data yang dipelajari sebelumnya (Ikasari & Widiastuti, 2021). SVM adalah metodologi pembelajaran yang memeriksa data dan mengidentifikasi pola. Tujuan SVM adalah menghitung frekuensi kemunculan kata dan mengklasifikasikan kalimat sebagai *positif* atau *negatif*. Keuntungan SVM termasuk dalam ruang input berdimensi tinggi dan penggunaan vektor dokumen (Husada & Paramita, 2021).

METODE PENELITIAN

Alur Penelitian

Alur penelitian diperlukan dalam penelitian karena tahapan proses analisis sentimen kasus judi online di *x* memiliki tahapan penting berupa, pengumpulan data (*crawling*), pelabelan data (*labelling*), pemrosesan data (*preprocessing*), klasifikasi metode, dan visualisasi (Alita, Fernando, et al., 2020). Langkah tersebut dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Dengan menggunakan *pattern recognition technology*, data mining adalah proses sorting melalui sejumlah besar data yang disimpan di *repositories* untuk menemukan menarik *patterns* dan *information* (Nabila et al., 2021). Tujuan dari data mining adalah untuk menemukan menarik *correlations*, *patterns*, dan *trends*.

Crawling Dataset

Pengumpulan data tentang judi online dilakukan melalui web *scraping* di *x* untuk mendapatkan *tweets* (Alita & Isnain, 2020). Proses ini menggunakan API *x* dengan package ‘*rtweet*’, yang dirancang khusus untuk mengumpulkan dan mengelola data dari *x* (Dongo et al., 2020).

Data yang terkumpul melalui proses *crawling* atau *web scraping* dengan API *x* menggunakan beberapa kata kunci yaitu “casino online”, “dampak judi online”, “judi online”, “judol”, “judol indo”, “slot gacor”. Dari beberapa kata kunci yang menjadi acuan terkumpulnya data opini masyarakat mengenai pengaruh dan dampak judi online terhadap remaja maupun masyarakat di *x* maka akan dianalisis menggunakan metode *naïve bayes* dan *support vector machine* (SVM) (Rachman & Pramana, 2020). Tahapan Crawling dapat dilihat pada Gambar 2.



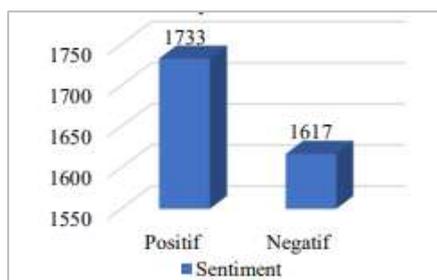
Gambar 2. Crawling Dataset

Ketika program python dijalankan dengan google collaboratory, data langsung disimpan dalam format *.csv* setelah proses *scraping* selesai. Data yang diperoleh masih mentah atau tidak terstruktur, sehingga langkah selanjutnya adalah mengkategorikan dan melakukan tahap *preprocessing* (Febriani & Sulistiani, 2021). Informasi tersebut diperoleh menggunakan teknik perayapan yang memanfaatkan paket *Harvest* dalam bahasa pemrograman Python dari situs media sosial X (Setiawan & Suryono, 2024). Tweet dari pengguna berbahasa Indonesia digunakan sebagai sumber data.

Pelabelan Dataset

Proses pelabelan data berfungsi sebagai pembagian kelas sentimen yang digunakan untuk tahap klasifikasi metode, kelas sentimen yang terbagi yaitu *negatif*, *netral*, dan *positif*. Pada proses ini data yang masih berupa bahasa indonesia akan diterjemahkan kedalam bahasa inggris yang berfungsi agar pada saat proses pelabelan data tersebut lebih akurat. Pelabelan dilakukan dengan metode *lexicon based*.

Pada penelitian terdahulu dengan tahapan yang sama dalam memproses pelabelan dataset berasal dari media sosial x. sentimen data tersebut terbagi menjadi positif, negatif, dan untuk netral data tidak terpakai karena dianggap data yang tidak bermakna atau kosong. berikut merupakan contoh hasil dari penelitian yang dilakukan oleh (Ananda & Suryono, 2024) Dengan pelabelan sentimen data sebanyak 3.350 yang terbagi menjadi data positif sebanyak 1.733 data, dan negatif sebanyak 1.617 data. Contoh hasil diagram penelitian tersebut dapat dilihat pada Gambar 3.



Gambar 3. Contoh Hasil Pelabelan

Preprocessing dataset

Preprocessing adalah langkah membersihkan dan menyiapkan teks sebelum dianalisis oleh sistem. Tahapan ini dilakukan untuk menghindari data yang hilang, berlebihan, atau tidak konsisten (Arischo & Damayanti, 2024). Kualitas data yang dihasilkan tidak secara konsisten sesuai untuk diproses (Syah & Witanti, 2022). Adapun langkah *preprocessing* yang dilakukan dalam penelitian ini adalah:

1. *Cleaning*

Pembersihan data adalah prosedur verifikasi dan peningkatan kualitas data dengan mengubah, mengamandemen, atau menghilangkan data yang berlebihan, tidak lengkap, salah, atau cacat untuk menghasilkan data berkualitas tinggi. Proses ini juga dikenal sebagai data *cleaning* atau data *scrubbing*.

2. *Case folding*

Case folding adalah proses menstandarisasi semua huruf ke dalam format yang konsisten, sering kali huruf kecil. Pada langkah ini, semua karakter dalam kalimat diubah menjadi huruf kecil.

3. *Tokenizing*

Kalimat akan dibagi menjadi beberapa kata terpisah dan diperiksa dari karakter pertama hingga karakter terakhir. Jika karakter ke-*i* bukan pemisah, seperti titik, koma, spasi, atau yang serupa, karakter tersebut akan disambung dengan karakter berikutnya.

4. *Stopword*

Filtering adalah fase yang melibatkan pemilihan kata-kata yang relevan dari hasil tokenisasi dengan menggunakan algoritma daftar henti untuk mengecualikan istilah yang kurang relevan atau daftar kata untuk mempertahankan kata-kata penting. Kata henti sering dianggap tidak penting, termasuk istilah seperti "yang mana," "dan," "dari," dan "di." Tujuan penggunaan kata henti adalah untuk mengecualikan kata kunci dengan nilai informasi yang lebih rendah, sehingga memungkinkan fokus pada kata-kata yang lebih penting.

5. *Stemming*

Pada tahap *stemming*, kata-kata dalam dokumen diubah menjadi bentuk dasarnya menggunakan aturan tertentu. Untuk bahasa Indonesia, proses ini melibatkan penghilangan *sufiks*, *prefiks*, dan *konfiks* dari kata-kata.

Klasifikasi Metode

1. *Support Vector Machine (SVM)*

Support vector machine (SVM) adalah metode yang umum digunakan dalam penelitian data dan *text mining* karena performanya yang unggul (Styawati & Mustofa, 2019). SVM beroperasi dengan menggunakan fungsi linear di dalam ruang fitur berdimensi tinggi. Meskipun demikian, teknik klasifikasi SVM terbatas pada pengelompokan data ke dalam dua kelas. (Pramudita & Musdholifah, 2020).

Prinsip utama pendekatan SVM untuk klasifikasi data adalah mengidentifikasi hiperbidang optimal yang memaksimalkan pemisahan antara dua kelas yang telah ditentukan sebelumnya (Ananda, 2024). SVM berfungsi dengan menentukan hyperplane terbaik. Hyperplane dan margin maksimum berkaitan dengan jarak antara titik data yang paling dekat dengan hyperplane, yang disebut sebagai margin. Dengan vector pendukung berada dekat dengan *hyperlane* (Tineges et al., 2020).

2. *Naïve Bayes*

Pendekatan Naïve Bayes dipilih karena kemampuan klasifikasinya, kesederhanaan rumus, dan kemudahan penerapannya. (Safira & Hasan, 2023). *Naïve bayes* dianggap lebih unggul dibandingkan metode pemisahan data terstruktur lainnya dalam hal akurasi dan efisiensi komputasi (A. Wibowo et al., 2022). Metode klasifikasi Naïve Bayes digunakan untuk membuat keputusan dengan meramalkan kasus berdasarkan hasil klasifikasi yang diperoleh. Penelitian ini menggunakan Naïve Bayes untuk menentukan sentimen topik perjudian daring dari kumpulan data X, dengan mengadopsi pendekatan yang didasarkan pada probabilitas dan statistik. (Made et al., 2020). Rumus sederhana metode *naïve bayes* dapat dilihat pada persamaan 1 berikut.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Persamaan metode tersebut menggambarkan bahwa A merepresentasikan "hipotesis data sebagai suatu kelas tertentu," B menandakan "data dengan kelas yang tidak diketahui," $P(A | B)$ menandakan "probabilitas hipotesis dengan kondisi tertentu," $P(A)$ mengindikasikan "probabilitas hipotesis," $P(B | A)$ mencerminkan "probabilitas berdasarkan kondisi hipotesis," dan $P(B)$ merepresentasikan "probabilitas B."

Visualisasi dan Evaluasi

Fase visualisasi ditujukan untuk menyajikan hasil teknik kategorisasi yang digunakan dalam penelitian ini. visualisasi yang ditampilkan berupa *histogram* dan *wordcloud* dari visualisasi kata *negatif*, *positif*, dan keseluruhan kata dalam dataset.

HASIL DAN PEMBAHASAN

A. Crawling Dataset

Proses *crawling* dataset pada topik kasus judi online dilakukan melalui media sosial *x*. Tahap tersebut diperlukan bahasa pemrograman *pyhton* dengan bantuan *tools google colab*, yang tak kalah penting adalah kode token API *x* yang mana token ini yang berfungsi sebagai kode untuk mengumpulkan data dari media sosial *x* tersebut. Berikut merupakan pada Gambar 4 hasil dari proses *crawling* yang telah mengumpulkan data mentah sebanyak 2.866 yang sudah difilter secara manual sebelum menuju tahap *preprocessing* selanjutnya.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	1.8E+18	Sac Jan 08	97	Polri mental	1.8E+18	https://jbs.twimg.com										
3	1.8E+18	Fri Jan 14	34055	Blockir dot	1.8E+18		in	Michusol	478	574	6524	https://t.c	1.7E+08	ardisatriawan		
4	1.8E+18	Mon Jan 1	19457	Merikonek	1.8E+18		in	Michusol	233	509	8017	https://t.c	1.7E+08	ardisatriawan		
5	1.8E+18	Thu Aug 2	0	lri apikan	1.8E+18	https://jbs.twimg.com		Indonesia	0	0	0	https://t.c	1.8E+08	Amansyahsyahpu3		
6	1.8E+18	Thu Jul 25	721	coba terar	1.8E+18		in	Acet: 7.3	7	21	451	https://t.c	1.8E+07	lantip		
7	1.8E+18	Sun Jul 07	0	Asak arab	1.8E+18		in		0	0	0	https://t.c	3.5E+07	cesarinae		
8	1.8E+18	Thu Jul 11	21	gimana rrv	1.8E+18	https://jbs.twimg.com		Yogyakarta	0	12	3	https://t.c	7.6E+17	jogjapokeid		
9	1.8E+18	Thu Aug 2	0	MENKOM	1.8E+18	https://jbs.twimg.com		Indonesia	0	0	0	https://t.c	1.2E+18	TotalPolitikCom		
10	1.8E+18	Thu Aug 2	1	Budi Arie J	1.8E+18	https://jbs.twimg.com		Jakarta Ca	3	0	0	https://t.c	1.1E+18	SyaMusdalifah		
11	1.8E+18	Fri Jul 17 E	0	Saran saya agar pemerintah tidak sibuk dengan judi online turunkan harga ³ sembako biaya kesehatan &amp;				Indonesia	0	0	0	https://t.c	1.8E+09	Amansyahsyahpu3		
12	1.8E+18	Thu Aug 2	0	lri kesaja	1.8E+18	https://jbs.twimg.com		Indonesia	0	0	0	https://t.c	5.1E+08	Dirragawak		
13	1.8E+18	Thu Jun 11	85	Suduh san	1.8E+18	https://jbs.twimg.com		091+09HC	4	18	39	https://t.c	1.8E+08	Dirragawak		
14	1.8E+18	Thu Aug 2	0	Merikonek	1.8E+18		in	KKM	0	0	0	https://t.c	1.8E+07	ardisatriawan		
15	1.8E+18	Thu Jul 04	0	Seorang te	1.8E+18	https://jbs.twimg.com			3	0	0	https://t.c	1E+18	slapoty		
16	1.8E+18	Sat Jan 22	25	@berakoe	1.8E+18		in	Haralots	1	15	1	https://t.c	1.5E+18	Servaisat111		
17	1.8E+18	Sun Aug 0	2	Sanku Me	1.8E+18		in		2	0	2	https://t.c	1.2E+18	Muttajansahmed		

Gambar 4. Hasil Crawling

B. Pelabelan Dataset

Tahap pelabelan dilakukan untuk mengetahui kelas sentimen dari data tersebut, yang terbagi menjadi sentimen *positif*, *negatif*, dan *netral*. Namun sentimen *netral* tidak digunakan dalam penelitian ini karena sentimen tersebut dianggap tidak bermakna atau sering disebut dengan istilah *null* atau berarti kosong. Data pada tahap ini akan diterjemahkan dalam bahasa Inggris agar hasil data yang diperoleh akan memiliki tingkat akurasi yang sesuai. Pelabelan dilakukan dengan menggunakan kamus *googletrans*. Hasil pelabelan dapat dilihat pada Gambar 5 berikut.

	id_str	username	full_text	Compound_Score	Sentiments
2	1.8E+18	DivHumas_Polri	The Nation	-0.1027	Negatif
3	1.8E+18	ardisatriawan	Block onlir	-0.3818	Negatif
4	1.8E+18	ardisatriawan	Minister o	-0.8271	Negatif
5	1.83E+18	Amansyahsyahpu3	This online	0.4939	Positif
6	1.82E+18	lantip	Try explair	-0.8504	Negatif
7	1.81E+18	cesarinae	Children a	-0.7771	Negatif
8	1.81E+18	jogjapokeid	How can y	-0.3301	Negatif
9	1.83E+18	TotalPolitikCom	MENKOM	-0.7824	Negatif
10	1.83E+18	SyaMusdalifah	Budi Arie 1	-0.6597	Negatif

Gambar 5. Hasil Pelabelan

C. Preprocessing Dataset

Dalam tahap *preprocessing* dataset memiliki beberapa tahapan penting yang harus diperhatikan, tahapan tersebut berupa *cleaning*, *case folding*, *tokenizing*, *stopword*, dan *stemming*. Langkah-langkah tersebut penting dilakukan sebelum menuju tahap klasifikasi metode yang akan digunakan. Berikut merupakan hasil dari tahapan-tahapan tersebut dapat dilihat pada Tabel 1,2,3,4,5.

1. *Cleaning*

Tabel 1. Hasil *Cleaning*

Tweet	Hasil Cleaning
The National Police through Polda Metro Jaya succeeded in uncovering criminal acts of online gambling in the Royal Domino Higgs Domino Royal Dream Boss Domino and Joker King game applications. Of the 23 suspects who were successfully arrested by the police, 5 of them were from the same family who acted as managers	The National Police through Polda Metro Jaya succeeded in uncovering criminal acts of online gambling in the Royal Domino Higgs Domino Royal Dream Boss Domino and Joker King game applications Of the suspects who were successfully arrested by the police of them were from the same family who acted as managers

2. *Case Folding*

Tabel 2. Hasil *Case Folding*

Tweet	Hasil Case Folding
The National Police through Polda Metro Jaya succeeded in uncovering criminal acts of online gambling in the Royal Domino Higgs Domino Royal Dream Boss Domino and Joker King game applications. Of the 23 suspects who were successfully arrested by the police, 5 of them were from the same family who acted as managers	the national police through polda metro jaya succeeded in uncovering criminal acts of online gambling in the royal domino higgs domino royal dream boss domino and joker king game applications of the suspects who were successfully arrested by the police of them were from the same family who acted as managers

3. *Tokenizing*

Tabel 3. Hasil *Tokenizing*

Tweet	Hasil Tokenizing
The National Police through Polda Metro Jaya succeeded in uncovering criminal acts of online gambling in the Royal Domino Higgs Domino Royal Dream Boss Domino and Joker King game applications. Of the 23 suspects who were successfully arrested by the police, 5 of them were from the same family who acted as managers	['the', 'national', 'police', 'through', 'polda', 'metro', 'jaya', 'succeeded', 'in', 'uncovering', 'criminal', 'acts', 'of', 'online', 'gambling', 'in', 'the', 'royal', 'domino', 'higgs', 'domino', 'royal', 'dream', 'boss', 'domino', 'and', 'joker', 'king', 'game', 'applications', 'of', 'the', 'suspects', 'who', 'were', 'successfully', 'arrested', 'by', 'the', 'police', 'of', 'them', 'were', 'from', 'the', 'same', 'family', 'who', 'acted', 'as', 'managers']

4. Stopword

Tabel 4. Hasil Stopword

Tweet	Hasil Stopword
The National Police through Polda Metro Jaya succeeded in uncovering criminal acts of online gambling in the Royal Domino Higgs Domino Royal Dream Boss Domino and Joker King game applications. Of the 23 suspects who were successfully arrested by the police, 5 of them were from the same family who acted as managers	['the', 'national', 'police', 'through', 'polda', 'metro', 'jaya', 'succeeded', 'in', 'uncovering', 'criminal', 'acts', 'of', 'online', 'gambling', 'in', 'the', 'royal', 'domino', 'higgs', 'domino', 'royal', 'dream', 'boss', 'domino', 'and', 'joker', 'king', 'game', 'applications', 'of', 'the', 'suspects', 'who', 'were', 'successfully', 'arrested', 'by', 'the', 'police', 'of', 'them', 'were', 'from', 'the', 'same', 'family', 'who', 'acted', 'as', 'managers']

5. Stemming

Tabel 5. Hasil Stemming

Tweet	Hasil Stemming
The National Police through Polda Metro Jaya succeeded in uncovering criminal acts of online gambling in the Royal Domino Higgs Domino Royal Dream Boss Domino and Joker King game applications. Of the 23 suspects who were successfully arrested by the police, 5 of them were from the same family who acted as managers	the national police through polda metro jaya succeeded in uncovering criminal acts of online gambling in the royal domino higgs domino royal dream boss domino and joker king game applications of the suspects who were successfully arrested by the police of them were from the same family who acted as managers

D. Klasifikasi Metode

1. Naïve Bayes

Tahap ini menampilkan hasil klasifikasi data yang telah dilakukan tahap uji yang terbagi menjadi 30% data *testing* dan 70% data *training*, dengan data sebanyak 2.866. Hasil klasifikasi metode *naïve bayes* dapat dilihat pada Gambar 6 berikut.

	precision	recall	f1-score	support
Negatif	0.76	0.82	0.79	418
Positif	0.82	0.76	0.78	442
accuracy			0.79	860
macro avg	0.79	0.79	0.79	860
weighted avg	0.79	0.79	0.79	860

Gambar 6. Klasifikasi Nive Bayes

Hasil penelitian tersebut dilakukan dengan penambahan menggunakan fungsi perhitungan *confusion matrix* yang mana menghasilkan nilai *accuracy* sebesar 79% dengan *support* data uji sebanyak 860 data. Berdasarkan keterangan gambar diatas dijelaskan hasil *precision negatif* 76%, *positif* 82%, *recall negatif* 82%, *positif* 76%, *f1-score negatif* 79%, *positif* 78%, dengan *support negatif* 418, *positif* 442. Dari hasil tersebut dapat dikatakan analisis dengan *naïve bayes* sudah cukup baik untuk jumlah data yang cukup besar.

2. Support Vector Machine (SVM)

Analisis menggunakan metode SVM dalam beberapa sumber memiliki tingkat akurasi yang sangat baik dibandingkan dengan metode lain. Data yang digunakan oleh metode SVM sama seperti metode *naïve bayes* yaitu dataset sebanyak 2.866 dengan data *testing* 30% dan data *training* 70% dengan jumlah data *testing* sebanyak 860.

Dengan adanya fungsi perhitungan *confusion matrix* pada setiap metode akan menambah nilai keakuratan perhitungan sebuah data. Berikut hasil terlihat pada Gambar 7.

	precision	recall	f1-score	support
Negatif	0.81	0.85	0.83	418
Positif	0.85	0.81	0.83	442
accuracy			0.83	860
macro avg	0.83	0.83	0.83	860
weighted avg	0.83	0.83	0.83	860

Gambar 7. Klasifikasi *Support Vector machine* (SVM)

Berdasarkan analisis yang dilakukan dengan menggunakan metode SVM pada topik kasus judi online menghasilkan nilai akurasi yang sangat baik yaitu *accuracy* 83%. Dengan nilai *precision negatif* 81%, *positif* 86%, *recall negatif* 85%, *positif* 81%, *f1-score negatif* 83%, *positif* 83% dengan *support negatif* 418 dan *positif* 442. Dari hasil tersebut nilai *support* untuk akurasi memiliki sebesar 860 data. Hasil analisis menggunakan metode SVM memiliki nilai *accuracy* yang sangat memuaskan dibandingkan dengan metode *naïve bayes* dengan perbandingan nilai *accuracy* yang berbeda.

E. Visualisasi dan Evaluasi

Tahap visualisasi ini menyajikan data atau informasi secara grafis, sehingga dapat memudahkan dalam memahami data yang sulit dilihat hanya berupa teks. Hasil yang akan ditampilkan berupa *wordcloud*, *barplot*, dan frekuensi kata. Tahap evaluasi bertujuan sebagai tahap yang berfungsi untuk menilai kualitas dan keandalan dari hasil penelitian. Evaluasi juga meliputi peninjauan metodologi, analisis data, dan kesimpulan yang diambil. Berikut ini merupakan hasil dari visualisasi dan evaluasi.

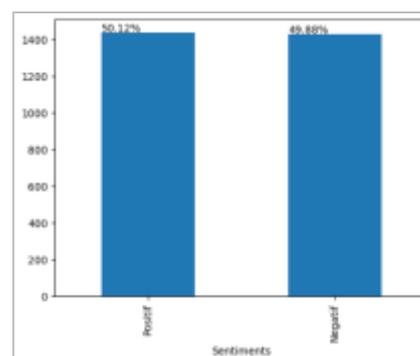
1. Visualisasi

a. Barplot

Tahap visualisasi pada barplot ini digunakan untuk menampilkan jumlah data dalam kelas sentimen dan hasil pelabelan dari dataset tersebut. Dapat dilihat hasil *barplot* pada Gambar 8 dan Gambar 9.

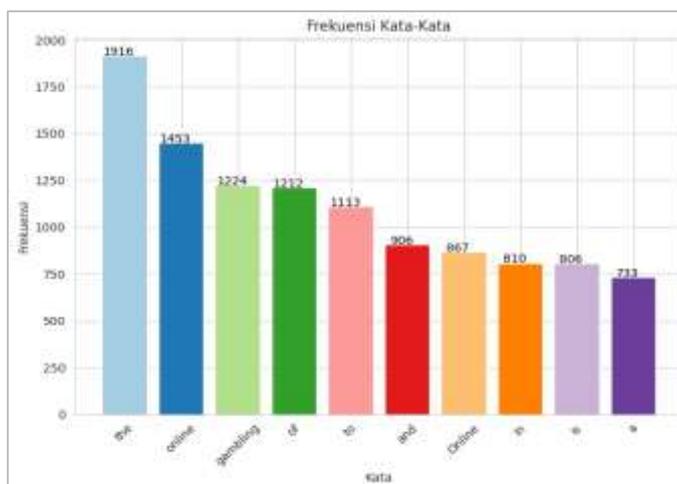


Gambar 8. Jumlah *Class Sentiment*



Gambar 9. Jumlah Sentimen Dataset

Berdasarkan hasil yang dilihat pada gambar diatas dapat dijelaskan bahwa untuk Gambar 8 menjelaskan mengenai hasil sentimen berupa jumlah banyaknya data yang terbagi menjadi 2 yaitu *negatif* sebanyak 1.429 dan *positif* 1.436 data. Sedangkan untuk Gambar 9 sama dengan gambar sebelumnya menjelaskan mengenai hasil sentimen akan tetapi berupa besaran % dengan 2 kelas sentimen yaitu *negatif* 49.88% dan *positif* 50.12%.



Gambar 12. Hasil Frekuensi Kata

Berdasarkan hasil frekuensi kata tersebut mendapatkan hasil frekuensi tertinggi yaitu kata “the” dengan jumlah 1.916, kemudian untuk kata “online” urutan ke dua dengan jumlah kata 1.453, dan pada urutan ke tiga dihasilkan pada kata “gambling (judi)” sebanyak 1.212 kata.

2. Evaluasi

Selama fase penilaian menggunakan himpunan data 2.866 entri dalam studi situasi perjudian daring, perbandingan antara pendekatan SVM dan Naïve Bayes mengungkapkan bahwa SVM mengungguli Naïve Bayes dalam hal akurasi, mencapai 83% dibandingkan dengan Naïve Bayes yang 79%. Keunggulan SVM terletak pada kemampuannya untuk menangani data dengan margin yang lebih banyak dan jelas, yang berarti bahwa metode tersebut lebih baik dalam membedakan kelas sentimen positif dan negatif dari pada naïve bayes yang cenderung lebih sederhana dan kurang optimal pada data dengan distribusi yang kompleks.

KESIMPULAN

Penelitian opini publik tentang kasus perjudian daring yang memanfaatkan data X menunjukkan bahwa teknik support vector machine (SVM) mengungguli naive bayes dalam mengkategorikan opini positif dan negatif. Dalam kumpulan data yang mencakup 2.866 contoh, yang mencakup 1.436 contoh positif (50,12%) dan 1.429 contoh negatif (49,88%), Support Vector Machine (SVM) memperoleh akurasi sebesar 83%, sedangkan Naive Bayes memperoleh akurasi sebesar 79%. Manfaat utama SVM adalah kapasitasnya untuk menangani data dengan margin dan kejelasan yang lebih besar, sehingga memfasilitasi diferensiasi yang lebih efektif antara sikap positif dan negatif. Hal ini menunjukkan bahwa SVM lebih cocok digunakan pada data dengan distribusi yang kompleks, dimana metode yang lebih sederhana seperti *naive bayes* cenderung kurang optimal. Dengan demikian, dalam konteks analisis opini publik terhadap kasus perjudian online, SVM tampaknya merupakan metode yang lebih akurat dan andal. Oleh karena itu, SVM dapat diusulkan sebagai pendekatan yang lebih efektif untuk analisis kesamaan di masa depan, terutama ketika diperlukan akurasi dalam diskriminasi sentimen.

SARAN

Seperti yang ditunjukkan dalam penelitian ini, metode SVM dapat memberikan hasil yang lebih andal dalam membedakan sentimen *positif* dan *negatif* pada dataset dengan distribusi yang hampir seimbang. Hasil analisis menunjukkan bahwa metode SVM memiliki akurasi yang lebih tinggi dibandingkan dengan *naive bayes*.

DAFTAR PUSTAKA

- Addiyansyah, W., Artikel, S., & Judi Online Remaja, K. (2023). Kecanduan Judi Online Di Kalangan Remaja Desa Cilebut Barat Kecamatan Sukaraja Kabupaten Bogor. *MANIFESTO Jurnal Gagasan Komunikasi, Politik, Dan Budaya*, 1(1), 13–22.
- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass SVM Pada Opini Publik Berbahasa Indonesia di Twitter. *Jurnal Tekno Kompak*, 14(2), 86–91. <https://doi.org/10.33365/JTK.V14I2.792>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50–58. <https://doi.org/10.23960/KOMPUTASI.V8I2.2615>
- Ananda, D., & Suryono, R. R. (2024). Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 8(2), 748–757. <https://doi.org/10.30865/MIB.V8I2.7517>
- Arischo, R. S., & Damayanti, D. (2024). Analisis Sentimen Pinjaman Online di Twitter dengan Metode Naive Bayes Classifier dan SVM. *Jurnal Media Informatika Budidarma*, 8(2), 1120–1130. <https://doi.org/10.30865/MIB.V8I2.7406>
- Dongo, I., Cadinale, Y., Aguilera, A., Martínez, F., Quintero, Y., & Barrios, S. (2020). Web Scraping versus Twitter API: A Comparison for a Credibility Analysis. *ACM International Conference Proceeding Series*, 263–273. <https://doi.org/10.1145/3428757.3429104>
- Febriani, S., & Sulistiani, H. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. *Jurnal Teknologi Dan Sistem Informasi*, 2(4), 89–95. <https://doi.org/10.33365/JTSL.V2I4.1373>
- Hasiholan, A., Cholissodin, I., & Yudistira, N. (2022). Analisis Sentimen Tweet Covid-19 Varian Omicron pada Platform Media Sosial Twitter menggunakan Metode LSTM berbasis Multi Fungsi Aktivasi dan GLOVE. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(10), 4653–4661.
- Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/TEKNIKA.V10I1.311>
- Ikasari, D., & Widiastuti, W. (2021). Sentiment Analysis Review Novel “Goodreads” Berbahasa Indonesia Menggunakan Naïve Bayes Classifier. *Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi)*, 5(1). <https://doi.org/10.30998/SEMNASRISTEK.V5I1.5040>
- Julianti, O. N., Suarna, N., & Prihartono, W. (2024). Penerapan Natural Language Processing Pada Analisis Sentimen Judi Online di Media Sosial Twitter. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2936–2941. <https://doi.org/10.36040/JATI.V8I3.9613>
- Made, N., Astari, A. J., Gede, D., Divayana, H., & Indrawan, G. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/JSI.V15I1.332>
- Muhammadin, A., & Sobari, I. A. (2021). Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma SVM dan NBC. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, 2(2), 85–91. <https://doi.org/10.31294/REPUTASI.V2I2.785>
- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108. <https://doi.org/10.33365/JTSL.V2I2.868>
- Oktavia, D., Ramadhan, Y. R., & Minarto, M. (2023). Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM). *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(1), 407–417. <https://doi.org/10.30865/KLIK.V4I1.1040>
- Pramudita, D. A., & Musdholifah, A. (2020). GSA to Obtain SVM Kernel Parameter for Thyroid Nodule Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(1), 11–22. <https://doi.org/10.22146/IJCCS.41215>
- Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Indonesian of Health*

- Information Management Journal (INOHIM)*, 8(2), 100–109.
<https://doi.org/10.47007/INOHIM.V8I2.223>
- Rahmansyah, A. I., & Darwis, D. (2020). Sistem Informasi Akuntansi Pengendalian Internal Terhadap Penjualan (Studi Kasus : CV. Anugrah PS). *Jurnal Teknologi Dan Sistem Informasi*, 1(2), 42–49. <https://doi.org/10.33365/JTSL.V1I2.388>
- Rakhmah, A. H., & Putri, T. A. (2019). Analisis Sentimen Terhadap Pasangan Calon Presiden 2019 Pada Media Sosial Twitter. *Jurnal lentera ICT*, 5(1), 1–11.
- Safira, A., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *ZONasi: Jurnal Sistem Informasi*, 5(1), 59–70. <https://doi.org/10.31849/ZN.V5I1.12856>
- Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 183–192. <https://doi.org/10.29408/EDUMATIC.V8I1.25667>
- Styawati, S., & Mustofa, K. (2019). A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3), 219–230. <https://doi.org/10.22146/IJCCS.41302>
- Sulistiani, H., Darwis, D., Shinta, D., Silaen, M., Marlyna, D., Akuntansi, S. I., & Informasi, S. (2020). Pengembangan Media Pembelajaran Akuntansi Berbasis Multimedia (Studi Kasus: SMA Bina Mulya Gading Rejo, Pringsewu). *Jurnal Komputer Dan Informatika*, 15(1), 127–136.
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59–67. <https://doi.org/10.47080/simika.v5i1.1411>
- Tasya Jadidah, I., Milyarta Lestari, U., Alea Amanah Fatiha, K., Riyani, R., Ariesty Wulandari, C., Studi Pendidikan Guru Madrasah Ibtidaiyah, P., Islam Negeri Raden Fatah Palembang, U., & H Zainal Abidin Fikri, J. K. (2023). Analisis maraknya judi online di Masyarakat. *Jurnal Ilmu Sosial Dan Budaya Indonesia*, 1(1), 20–27. <https://doi.org/10.61476/8XVGDDB22>
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, 4(3), 650–658. <https://doi.org/10.30865/MIB.V4I3.2181>
- Wibowo, A., Noor Hasan, F., Akbar Ramadhan, L., Nurhayati, R., & Arief Wibowo, dan. (2022). Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier. *Jurnal Asimetrik: Jurnal Ilmiah Rekayasa & Inovasi*, 4, 239–248. <https://doi.org/10.35814/ASIIMETRIK.V4I1.3577>
- Wibowo, R. H., & Ikhsan, M. (2020). Penegakan Hukum terhadap Penyalahgunaan Aplikasi yang Bermuatan Perjudian Online di Dunia Maya oleh Polresta Surakarta. *Prosiding University Research Colloquium*, 179–190.
- Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*. <https://doi.org/10.1109/ICIC47613.2019.8985884>
- Yana Siregar, L., Irwan Padli Nasution Prodi Manajemen, M., & Negeri Islam Sumatera Utara, U. (2020). Perkembangan Teknologi Informasi Terhadap Peningkatan Bisnis Online. *Hirarki: Jurnal Ilmiah Manajemen Dan Bisnis*, 2(1), 71–75. <https://doi.org/10.30606/hjimb>