

Pendekatan TF-IDF, SMOTE, dan SVM dalam Klasifikasi Sentimen Masyarakat terhadap Pemblokiran Judi Online

Robert Antonius, Achmad Rizky Zulkarnain, Hafiz Irsyad*

Fakultas Ilmu Komputer dan Rekayasa, Informatika, Universitas Multi Data Palembang, Palembang, Indonesia

Email: ¹robert.antonius@mhs.mdp.ac.id, ²rizkyzulkarnain@mhs.mdp.ac.id, ^{3,*}hafizirsyad@mdp.ac.id

Email Penulis Korespondensi: hafizirsyad@mdp.ac.id

Abstrak—Judi online adalah topik hangat di kalangan masyarakat. Salah satu pembicaraan terkait judi online yang muncul adalah apakah seharusnya pemerintah melakukan pemblokiran situs judi online. Ada beberapa sisi dari pembahasan ini, seperti apakah pemblokiran itu benar akan membantu mencegah adiksi judi online dan apakah justru seharusnya pemerintah melegalkan judi online. Untuk membantu dalam menavigasi wacana hangat ini, dibangunlah sebuah sistem yang dapat mendeteksi dua sisi sentimen terhadap pemblokiran judi online. Model dilatih dengan dataset yang diseimbangkan dengan SMOTE karena tidak meratanya kelas klasifikasi, lalu diboboti dengan TF-IDF untuk dapat fokus pada kata-kata berbobot tinggi. Model klasifikasi yang dibangun dengan *Support Vector Machine* mencapai tingkat akurasi 61.54% dengan tolak ukur evaluasi confusion matrix.

Kata Kunci: SMOTE, support vector machine, TF-IDF

Abstract—Online gambling is currently a hot topic among internet netizens. One of the talking points in the discourse was how should the government handle blocking online gambling sites. There is multiple sides to the discourse, such as does blocking the sites actually help in preventing gambling addiction or would legalizing it be the right policy instead. To help navigate this controversial topic, a system was built to differentiate the two sides of the argument towards blocking online gambling sites. The model is trained on a dataset that is first balanced with SMOTE, then weighted with TF-IDF to give focus to vocal tokens of the discourse. The classification model was built with Support Vector Machine and reached an accuracy level of 61.54% when evaluated with a confusion matrix.

Keywords: SMOTE, support vector machine, TF-IDF

1. PENDAHULUAN

Judi *online* adalah salah satu topik yang sedang memanas di media sosial, berkat tersebar luasnya budaya perjudian di Indonesia melalui togel, *lottery*, taruhan permainan sepak bola, atau taruhan permainan mahjong. Walau ada banyak masyarakat yang mendukung pemblokiran judi *online*, juga banyak pula masyarakat yang merasa bahwa judi *online* tidak perlu diblokir [1]. Opini masyarakat kontra pemblokiran umumnya berasal dari berbagai pandangan, seperti karena menurut mereka dampak buruk judi adalah tanggung jawab diri sendiri, karena menurut mereka pelegalan judi *online* seperti Amerika Serikat cara terbaik untuk mencegah kecanduan judi, atau karena mereka sendiri sudah kecanduan judi [1].

Opini masyarakat pro pemblokiran judi *online* umumnya berasal dari bahayanya kecanduan judi yang dapat mempengaruhi semua orang bahkan anak-anak kecil [2]. Opini ini sendiri disetujui oleh banyak rakyat kontra pemblokiran judi *online*, bahwa perjudian termasuk perjudian *online* dapat berbahaya karena dapat menyebabkan seseorang kecanduan [2]. Dua sisi argumen sama-sama memiliki dukungan kuat dari berbagai kalangan masyarakat, dan sulit menentukan bagaimana tindakan terbaik terkait pemblokiran situs judi *online* karena kedua sisi sama-sama setuju pada poin bahayanya judi *online* [1].

Untuk itu, pembuatan klasifikasi sentimen terhadap judi *online* dapat berguna untuk memisahkan argumen dari kedua belah pihak dan memahami argumen masing-masing pihak dengan lebih seksama. Selain itu, mengklasifikasi sentimen juga dapat membantu mengidentifikasi orang yang kecanduan judi *online* yang umumnya memiliki sentimen negatif terhadap pemblokirannya [3]. Sentimen analisis telah dilakukan pada penelitian serupa seperti untuk analisa sentimen pelajaran daring [4].

Term Frequency - Inverse Document Frequency (TF-IDF) adalah metode pembobotan token pada korpus dalam melakukan *information retrieval* [5]. TF-IDF berguna untuk membantu menurunkan bobot kata yang umum dipakai dan memberi bobot lebih pada kata yang lebih unik dalam korpus [6]. Pada topik judi *online*, ada beberapa kata seperti judi dan pemerintahan yang sering muncul pada kedua kelas sentimen, sehingga informasi yang dapat diperoleh dari kedua kata tersebut terbatas [2]. TF-IDF dapat membantu memboboti kata yang dapat menjadi fitur yang lebih baik [7]. TF-IDF telah digunakan untuk penelitian serupa seperti klasifikasi sentimen analisis terhadap kinerja POLRI dengan hasil evaluasi 66.07% sentimen positif dan 33.79% sentimen negatif [8].

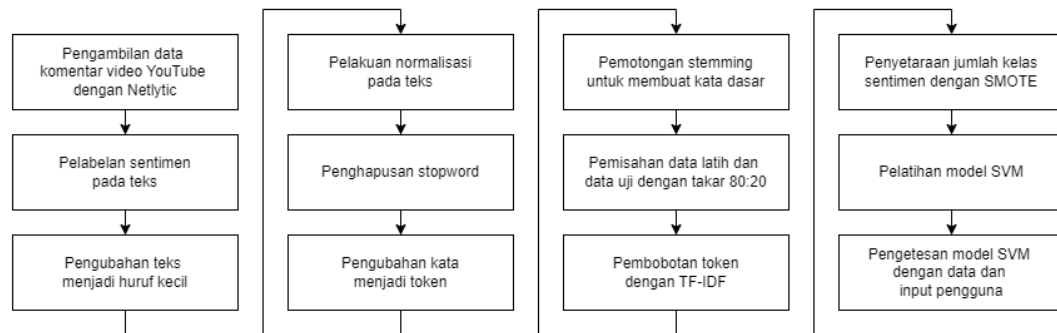
Synthetic Minority Oversampling Technique (SMOTE) adalah metode penyesuaian sampel data untuk menyeimbangi jumlah data pada tiap kelas klasifikasi [9]. SMOTE berguna untuk membantu menyetarakan jumlah data per kelas pada *dataset* yang memiliki data tidak seimbang [10]. Pada topik perjudian *online*, ada ketidakseimbangan sentimen yang mendukung pemblokiran judi *online*, sehingga penerapan SMOTE akan berguna untuk menutupi ketidakseimbangan dan membuat klasifikasi memiliki lebih sedikit bias [2]. SMOTE digunakan pada penelitian serupa seperti dalam klasifikasi sentimen pengguna terhadap aplikasi Gojek dengan akurasi tertinggi 81.09% sedangkan Grab dengan akurasi tertinggi 73.20% dengan teknik SMOTE [11].

Support Vector Machine (SVM) adalah metode klasifikasi data berkelas dengan menggunakan fungsi hyperspace untuk menentukan kelas dari fitur data [12]. SVM berguna untuk mengklasifikasi data dengan fitur numerik, seperti data bobot hasil TF-IDF [13]. Pada topik klasifikasi sentimen judi *online*, SVM membantu memisahkan antara data sentimen positif dan data sentimen negatif yang melebur antara satu sama lain karena topik judi *online* selain adalah sebuah topik kompleks yang masih diperdebatkan solusinya sampai saat ini [14]. SVM telah digunakan untuk kasus serupa seperti mengklasifikasikan tanggapan rakyat terhadap prediksi resesi 2023 dengan akurasi 98.67% [15] dan mengklasifikasikan sentimen masyarakat terhadap aplikasi MyPertamina di Google Play Store dengan akurasi 95%, Twitter dengan akurasi 76%, dan Youtube dengan akurasi 99% [16]. Penelitian ini dilakukan untuk menerapkan dan mengevaluasi penggunaan TF-IDF, SMOTE, dan SVM untuk mengklasifikasi sentimen masyarakat terhadap pemblokiran judi *online*.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

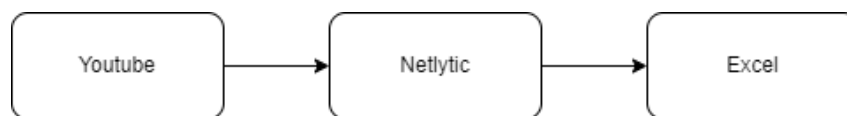
Penelitian dimulai dengan pengambilan data dengan *data scraping*. Data lalu dianalisa guna proses *preprocessing* dan *labeling* sentimen. Data dipisahkan menjadi data latih dan data uji dengan takar 80:20. Data diberi bobot dengan TF-IDF, lalu diratakan kelas antar sentimennya dengan SMOTE. Data dibagi menjadi data latih dan data uji, dan data latih digunakan untuk melatih model SVM. Terakhir, model SVM diuji pada data uji untuk mengukur akurasi dan *F1-score* model.



Gambar 1. Flowchart proses penelitian

2.2 Pengambilan Data

Data didapatkan dari platform *media social* YouTube melalui *website data scraping* Netlytic. Data yang diambil berupa teks komentar penonton dari tiga video yang membahas tentang judi *online*.



Gambar 2. Alur proses pengambilan data

Sumber data komentar yang digunakan berasal dari 3 video Youtube terkait topik pemblokiran judi *online*, yaitu *podcast* yang dilakukan oleh Ferry Irwandi, wawancara yang dilakukan oleh Aiman Witjaksono, dan *podcast* yang dilakukan oleh Denny Sumargo. Informasi terkait sumber data ditampilkan lebih lengkap pada Tabel 1.

Tabel 1. Sumber data komentar

Index	Judul Video	Tautan Video	Jumlah Komentar
1	ADA ARTIS JUDI ONLINE JADI CALEG!!TAU KALIAN? FAKTA GELAP TIKTOK DAN GACOR!!- Ferry Irwandi Podcast	https://www.youtube.com/watch?v=brA4tbGkNQ	433
2	● EKSKLUSIF! AIMAN WITJAKSONO BONGKAR SINDIKAT JUDI ONLINE, TERNYATA BANDARNYA “ORANG PENTING”	https://www.youtube.com/watch?v=1F_C4ESr3q0	382
3	SELAMA INI KALIAN SEMUA DI TIPU, MENANG JUDI ITU CUMAN SETTINGAN !! (DENNIS LIM) -Curhat Bang	https://www.youtube.com/watch?v=MsuHPBsiPok	500

2.3 Pelabelan Sentimen

Dilakukan proses pelabelan sentimen pada data komentar yang telah didapat. Pelabelan sentimen dilakukan secara manual dengan bantuan KBBI dengan label positif untuk komentar yang mendukung pemblokiran situs judi *online*, negatif untuk yang tidak mendukung pemblokiran situs judi *online*, dan netral untuk komentar yang tidak berkaitan dengan pemblokiran situs judi *online*.

Tabel 2. Pelabelan sentimen komentar

Index Video	Jumlah Sentimen		
	Positif	Netral	Negatif
1	120	230	83
2	131	171	80
3	58	394	47

2.4 Preprocessing

Data yang diperoleh berupa komentar yang ditulis oleh pengguna media sosial YouTube terhadap video yang ditayangkan, sehingga ada banyak kesalahan penulisan dan kata tidak baku yang dipergunakan. Untuk membuat data lebih mudah diolah oleh model, maka data dilakukan *preprocessing* terlebih dahulu.

2.4.1 Casefolding

Casefolding adalah proses pengubahan seluruh teks menjadi huruf kapital atau huruf kecil guna menyelaraskan kapitalisasi teks. *Casefolding* dilakukan karena komputer menyimpan huruf kapital dan huruf kecil secara terpisah, sehingga melakukan *casefolding* memungkinkan program untuk mengelolah huruf dengan sama seperti layaknya manusia.

Tabel 3. Penerapan *casefolding*

Komentar	Hasil <i>Preprocessing</i>
Masagus @Zulkifli menyala bgs amat performa kemaren nggak salah dtg rela lari dri rumah sakit !!!	masagus @zulkifli menyala bgs amat performa kemaren nggak salah dtg rela lari dri rumah sakit !!!

2.4.2 Text Cleaning

Text Cleaning adalah proses penghapusan bagian tulisan yang maknanya sulit dikuantifikasi untuk pembuatan model seperti angka, emotikon, dan simbol-simbol.

Tabel 4. Penerapan *text cleaning*

Komentar	Hasil <i>Preprocessing</i>
Masagus @Zulkifli menyala bgs amat performa kemaren nggak salah dtg rela lari dri rumah sakit !!!	masagus zulkifli menyala bgs amat performa kemaren nggak salah dtg rela lari dri rumah sakit

2.4.3 Normalisasi Teks

Normalisasi teks dilakukan untuk memudahkan model menerima teks dengan arti yang sama walau dengan pengejaan berbeda. Proses ini dilakukan dengan mengubah ejaan tidak baku dengan ejaan baku dengan makna kata yang sama.

Tabel 5. Penerapan Normalisasi text

Komentar	Hasil <i>Preprocessing</i>
masagus zulkifli menyala bgs amat performa kemaren nggak salah dtg rela lari dri rumah sakit	masagus zulkifli menyala bagus amat performa kemarin tidak salah datang rela lari dari rumah sakit

2.4.4 Penghapusan Stopword

Tahap *stopword removal* merupakan tahapan yang dilakukan untuk menghilangkan kata-kata yang tidak deskriptif. Kata-kata deskriptif berupa menurut konteks, namun umumnya adalah kata-kata seperti “dia”, “adalah”, “semua”, “diriku” yang merupakan kata-kata diskrit yang digunakan untuk melancarkan tata bahasa kalimat namun tidak memberi arti signifikan pada kalimat itu sendiri.

Tabel 6. Penerapan *Stopword Removal*

Komentar	Hasil <i>Preprocessing</i>
masagus zulkifli menyala bagus amat performa kemarin tidak salah datang rela lari rumah sakit	masagus zulkifli menyala bagus amat performa salah datang rela lari rumah sakit

2.4.5 Tokenisasi

Tokenisasi adalah proses memecah teks menjadi daftar token-token. Token mencakup satu atau lebih kata dan digunakan untuk menyimpan makna yang ada pada kata dasar yang memiliki lebih dari satu kata, seperti rumah bersalin.

Tabel 7. Penerapan Tokenisasi

Komentar	Hasil <i>Preprocessing</i>
masagus zulkifli menyala bagus amat	“masagus”, “zulkifli”, “menyala”, “bagus”, “amat”, “performa”,
performa salah datang rela lari rumah sakit	“salah”, “datang”, “rela”, “lari”, “rumah sakit”

2.4.6 Stemming

Stemming adalah proses pemenggalan imbuhan pada kata untuk mendapat kata dasarnya. Imbuhan kata dihilangkan untuk mencegah model menganggap bentuk-bentuk kata sebagai kata terpisah, walau kata-kata tersebut memiliki makna yang sama.

Tabel 8. Penerapan *Stemming*

Komentar	Hasil <i>Preprocessing</i>
“masagus”, “zulkifli”, “menyala”, “bagus”, “amat”, “performa”, “salah”, “datang”, “rela”, “lari”, “rumah sakit”	“masagus”, “zulkifli”, “nyala”, “bagus”, “amat”, “performa”, “salah”, “datang”, “rela”, “lari”, “rumah sakit”

2.5 Penyeimbangan SMOTE

SMOTE adalah turunan dari metode penyeimbang kelas melalui *oversampling*. SMOTE dilakukan dengan membuat sampel baru yang menyerupai data-data pada kelas minoritas untuk menyeimbangkan rasio antar kelas dataset.

2.6 Pembobotan TF-IDF

TF-IDF dilakukan untuk memberi bobot lebih kuat pada token-token yang sering muncul pada korpus namun muncul pada sedikit dokumen. TF-IDF digunakan untuk memperkuat fitur frekuensi teks sehingga memperlancar proses klasifikasi. Hasil dari proses TF-IDF adalah sebuah model bobot dari setiap token berdasarkan token-token yang digunakan untuk data latih.

$$TF(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

$$IDF(t, D) = \log \frac{N}{|\{d: d \in D \text{ and } t \in d\}|} \quad (2)$$

$$TFIDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

Rumus (1), (2), dan (3) digunakan untuk menghitung bobot TF-IDF setiap token. *t* merepresentasikan satuan token yang diboboti, *d* merepresentasikan satuan dokumen yang diboboti, dan *D* merepresentasikan keseluruhan dokumen pada korpus. Rumus (1) digunakan untuk menghitung bobot TF token, rumus (2) untuk bobot IDF token, dan rumus (3) digunakan untuk menghitung bobot TF-IDF token berdasarkan bobot TF dan bobot IDF token.

2.7 Pelatihan Model SVM

Model SVM dilatih dengan data latih yang telah disiapkan. Data latih adalah 80% data yang dipisahkan dari dataset untuk melatih model klasifikasi, dan 20% sisanya digunakan sebagai data uji untuk mengukur dan mengevaluasi model klasifikasi.

Model SVM yang dilatih menggunakan kernel *Radial Basis Function* (RBF). Kernel RBF didapat bahwa pada kedua dokumen teks, ada beberapa kata yang sama yang mendapat bobot TF-IDF besar, seperti kata “menang”, dan “slot”, dan “judol”. Kernel RBF memiliki fleksibilitas kuat dalam membagi fitur data menjadi hyperspace, sehingga lebih baik dalam poin data yang melebur seperti pada kasus pemblokiran judi *online*.

2.8 Evaluasi Model SVM

Model SVM diuji dengan confusion matrix terhadap data uji yang telah disiapkan. *Confusion matrix* digunakan untuk menghitung jumlah prediksi *true positive* (prediksi tepat), *false positive* (prediksi salah), *false negative* (gagal prediksi), dan *true negative* (bukan prediksi). Hasil dari *confusion matrix* lalu digunakan untuk menghitung metrik performa seperti akurasi model atau *F-measure* yang dicapai model.

$$ACC = \frac{TP+TN}{P+N} \quad (4)$$

$$F_1 score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

Pengujian model klasifikasi dilakukan dengan memasukkan fitur data uji sebagai masukan model dan membandingkan keluaran model dengan kelas data uji. Tolak ukur yang digunakan untuk mengevaluasi performa model adalah akurasi dan *F1-score*, yang mengukur ketepatan klasifikasi model. Rumus (4) dan (5) merupakan rumus yang digunakan untuk menghitung nilai akurasi dan nilai *F1-score* dari model klasifikasi. Nilai akurasi dihitung dengan membagi prediksi model yang akurat dengan jumlah prediksi yang dibuat model seperti pada rumus (4). Nilai *F1-score* dihitung dengan membagi 2 kali jumlah prediksi *true positive* dengan total 2 kali jumlah prediksi *true negative*, jumlah prediksi *false positive*, dan jumlah prediksi *false negative*.

3. HASIL DAN PEMBAHASAN

Kode untuk klasifikasi sentimen teks komentar dibuat dalam bahasa Python dengan bantuan Google Colab untuk memudahkan penggunaan banyak module berbeda yang diperlukan pada penelitian. Kode mengikuti proses yang tertera pada Gambar 1. Proses pembuatan model dapat dibagi menjadi beberapa tahap, antara lain tahap *preprocessing* data, persiapan data, dan pembuatan model.

3.1 Preprocessing Data

Pada tahap *preprocessing* data, data teks diberlakukan proses pembersihan data sesuai tata bahasa Indonesia. Proses *preprocessing* antara lain pengecilan kapitalisasi huruf, normalisasi kata tidak baku, penghapusan *stopword*, tokenisasi penggalan kata menjadi token, dan mengubah kata-kata pada token menjadi kata dasar. Data kemudian disimpan pada file terpisah supaya proses *preprocessing* tidak perlu diulang berkali-kali. Hasil proses *preprocessing* dapat dilihat pada Gambar 5.

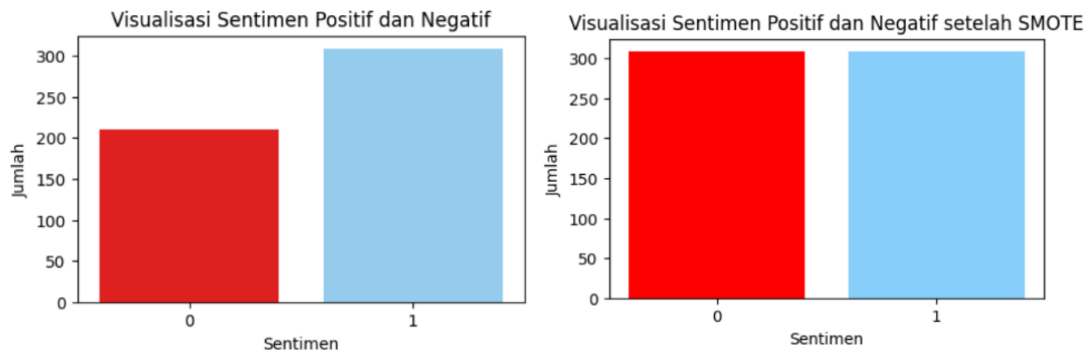


Gambar 5. Visualisasi hasil *preprocessing*

Penerapan preprocessing menghapus kata-kata stopwords dan tidak baku, serta menormalisasi penulisan tidak baku menjadi bentuk kata dasar yang baku, seperti yang ditampilkan pada Gambar 5. Penerapan preprocessing ditampilkan dalam bentuk word cloud yang menampilkan token berdasarkan frekuensi token pada sumber data awal seperti pada kata “judi” yang merupakan token yang paling sering muncul pada kedua sentimen.

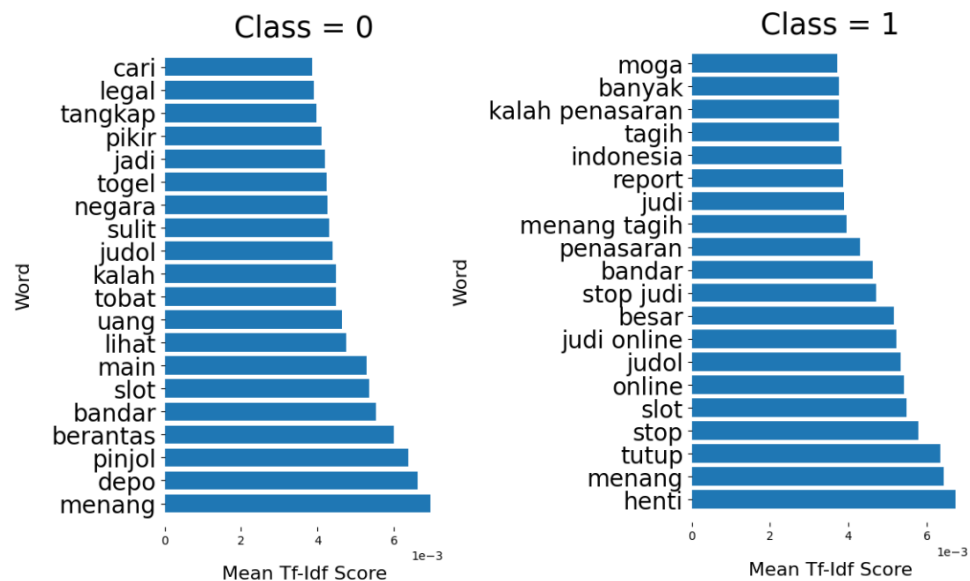
3.2 Penerapan SMOTE dan TF-IDF

Pada tahap persiapan data, data yang telah dibersihkan dipisahkan menjadi data latih dan data uji, diboboti dengan TF-IDF, dan disetarakan dengan SMOTE. Setelah dilakukan pembobotan dan penyetaraan, data divisualisasi untuk memastikan bahwa proses berhasil seperti pada Gambar 6 dan Gambar 7.



Gambar 6. Sebelum dan setelah penyeimbangan SMOTE

Pembobotan SMOTE dilakukan untuk menyetarakan takar sentimen positif dan sentimen negatif. Dataset yang digunakan memiliki lebih banyak data dengan sentimen positif daripada sentimen negatif dengan rasio sekitar 2:3. Setelah dilakukan pembobotan SMOTE, rasio data menjadi seimbang 1:1.



Gambar 7. Hasil dari pembobotan TF-IDF

Pembobotan TF-IDF dilakukan untuk memberikan bobot pada kata tertentu yang dapat lebih menentukan klasifikasi sentimen terhadap topik pemblokiran judi *online*. Pembobotan TF-IDF memberi bobot lebih besar pada kata seperti “henti” dan “menang”, seperti yang terlihat pada Gambar 7. Model pembobotan TF-IDF disimpan untuk memboboti kalimat baru.

3.3 Pembuatan Model

Pada tahap pemodelan sentimen masyarakat, digunakan metode SVM untuk membuat model klasifikasi. Model SVM dilatih dengan data latih yang telah disiapkan dengan kernel RBF. Setelah model selesai dilatih, model diuji dengan data uji untuk mendapat evaluasi nilai akurasi dan *F1-score*-nya. Disiapkan pula aplikasi sederhana untuk mengklasifikasikan kalimat masukan pengguna berdasarkan model sentimen pemblokiran judi *online* seperti pada Tabel 2.

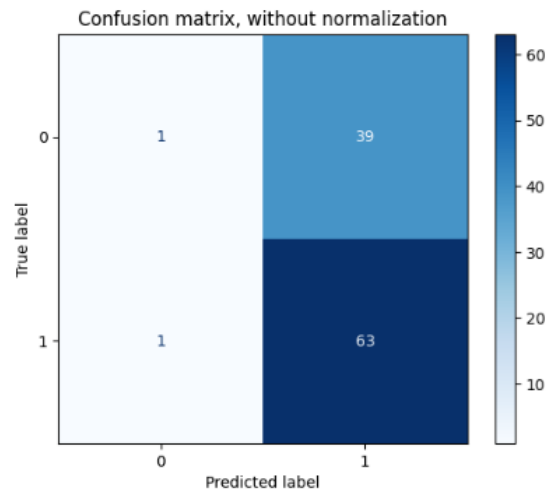
Tabel 9. Hasil klasifikasi model

Output	Target	Confusion Matrix
Positif	Positif	True Positive
Positif	Negatif	False Positive
Negatif	Positif	False Negative
Negatif	Negatif	True Negative

3.4 Evaluasi Model

Model klasifikasi sentimen yang telah dilatih dievaluasi dengan data uji menggunakan *confusion matrix*. Hasil klasifikasi model dipetakan menjadi 4 kategori, yaitu *true positive* jika model berhasil memprediksi kalimat sentimen positif sebagai sentimen positif, *false positive* jika model memprediksi kalimat sentimen negatif sebagai sentimen

positif, *false negative* jika model mengklasifikasikan kalimat sentimen positif sebagai sentimen negatif, dan *true negative* jika model mengklasifikasikan kalimat sentimen negatif sebagai sentimen negatif.



Gambar 8. Hasil *confusion matrix*

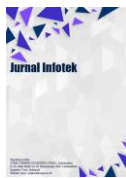
Hasil dari evaluasi menunjukkan bahwa ada 1 data negatif yang diklasifikasikan sebagai sentimen negatif, 39 data negatif yang diklasifikasikan sebagai sentimen positif, 1 data positif yang diklasifikasikan sebagai sentimen negatif, dan 63 data positif yang diklasifikasikan sebagai sentimen positif. Dari hasil *confusion matrix* yang ditampilkan pada Gambar 8, didapat bahwa akurasi dari model adalah 61.84%, dengan nilai *F1-score* model adalah 0.7590.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa klasifikasi sentimen masyarakat terhadap pemblokiran situs judi *online* dapat dilakukan dengan menggunakan metode SVM dengan pembobotan TF-IDF dan penyetaraan data SMOTE. Klasifikasi sentimen dari teks komentar YouTube mencapai nilai akurasi sebesar 61.84% dan mencapai nilai *F1-score* 0.7590.

REFERENCES

- [1] S. M. Gainsbury, "Online Gambling Addiction: the Relationship Between Internet Gambling and Disordered Gambling," *Curr Addict Rep*, vol. 2, no. 2, pp. 185–193, 2015, doi: 10.1007/s40429-015-0057-8.
- [2] M. Chóliz, "The Challenge of Online Gambling: The Effect of Legalization on the Increase in Online Gambling Addiction," *J Gambl Stud*, vol. 32, no. 2, pp. 749–756, 2016, doi: 10.1007/s10899-015-9558-6.
- [3] D. Fitriya *et al.*, "MENELAHAH FENOMENA JUDI ONLINE (SLOT) DI KALANGAN MAHASISWA DALAM PERSPEKTIF HUKUM ISLAM DI INDONESIA," *Jurnal Kajian Agama dan Dakwah*, vol. 2, 2024, doi: 10.333/Tashdiq.v1i1.571.
- [4] J. Homepage *et al.*, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," vol. 1, pp. 24–33, 2021.
- [5] S. Qaiser, U. Utara, M. Sintok, M. Kedah, A. Ramsha, and T. Analytics, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents Text Mining," 2018.
- [6] M. Liang and T. Niu, "Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs," *Procedia Comput Sci*, vol. 208, pp. 460–470, 2022, doi: <https://doi.org/10.1016/j.procs.2022.10.064>.
- [7] S.-W. Kim and J.-M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, p. 30, 2019, doi: 10.1186/s13673-019-0192-7.
- [8] R. Kristianto Hondro, "Jurnal Pendidikan Teknologi Informasi Dan Komputer Analisis Penerapan Text Mining dan TF-IDF dalam Mengetahui Sentimen Masyarakat Terhadap Kinerja POLRI," 2023, [Online]. Available: <https://journal.grahamitra.id/index.php/petik>
- [9] A. J. Mohammed, "Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 3161–3172, Jun. 2020, doi: 10.30534/ijatcse/2020/104932020.
- [10] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," 2018.



- [11] Hermanto, A. Y. Kuntoro, T. Asra, E. B. Pratama, L. Effendi, and R. Ocanitra, "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012102.
- [12] C. Cortes, "Support-Vector Networks," 1995.
- [13] A. Baita and N. Cahyono, "ANALISIS SENTIMEN MENGENAI VAKSIN SINO VAC MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN)."
- [14] D. S. Utami and A. Erfina, "ANALISIS SENTIMEN PINJAMAN ONLINE DI TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," 2021.
- [15] D. Iskandar Mulyana, N. Lutfianti, S. Tinggi Ilmu Komputer Cipta Karya Informatika, J. Radin Inten No, and J. Timur, "Analisis Sentimen Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023 Analysis Sentiment Using the SVM Algorithm in Netizen Responses to News of the 2023 Recession," vol. 13, no. 1, 2023, doi: 10.30700/jst.v13i1.1339.
- [16] A. Nursalim and R. Novita, "SENTIMENT ANALYSIS OF COMMENTS ON GOOGLE PLAY STORE, TWITTER AND YOUTUBE TO THE MYPERTAMINA APPLICATION WITH SUPPORT VECTOR MACHINE," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 6, pp. 1305–1312, 2023, doi: 10.52436/1.jutif.2023.4.6.1059.