

## ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP APLIKASI PINJAMAN ONLINE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Aldiansyah Putra<sup>1</sup>, Retnani Latifah<sup>2,\*</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jakarta, Jl. Cempaka Putih  
Tengah 27 Jakarta Pusat, Indonesia, 10510

[retnani.latifah@umj.ac.id](mailto:retnani.latifah@umj.ac.id)

### ABSTRAK

Pinjaman online adalah fasilitas penyedia jasa keuangan yang beroperasi secara daring dengan berbasis aplikasi mobile sebagai platform. Adanya aplikasi tersebut menimbulkan banyak diskusi dimasyarakat khususnya cuitan mengenai pinjaman online pada media sosial twitter. Dimana dalam diskusi tersebut ada yang pro dan ada yang kontra, Diskusi di sosial media tersebut dapat dianalisis untuk mengetahui sentimen umum masyarakat terhadap pinjaman online. Salah satu metode yang dapat digunakan adalah metode Support Vector Machine. Berdasarkan hasil pengujian pada penelitian ini dilakukan menggunakan data tweet sebanyak 6500 data, masing-masing data training sebanyak 80% atau 5300 data tweet dan data testing 20% atau 1300 data. Pada model svm ini dilakukan skenario percobaan parameter dengan nilai  $C = 1.5$  menggunakan kernel rbf menghasilkan akurasi terbaik menggunakan evaluasi Confusion matrix yaitu 84%. Persebaran sentimen analisis terkait pinjaman online yaitu sebanyak 215 tweet positif, 538 tweet negatif dan 547 tweet netra.

**Kata kunci:** : Twitter, Data, Sentimen, Pinjaman online, Support vector machine

### ABSTRACT

*Online lending is a financial service provider facility that operates boldly based on a mobile application as a platform. The existence of this application has caused a lot of discussion in the community, especially regarding online loans on Twitter social media. These data can be analyzed to find out the general public sentiment towards online credit. One method that can be used is the Support Vector Machine. method. Based on the results of testing in this study, it was carried out using 6500 tweet data, each training data as much as 80% or 5300 tweet data and testing data 20% or 1300 data. In this svm model, a parameter experiment scenario with a value of  $C = 1.5$  using the rbf kernel was carried out to produce the best accuracy using the Confusion matrix evaluation, which was 84%. The distribution of sentiment analysis related to online credit is 215 positive tweets, 538 negative tweets and 547 neutral tweets.*

**Keywords:** *Twitter, Data, Sentimen, Online loans, Support vector machine*

## 1. PENDAHULUAN

Salah satu media yang dapat digunakan untuk menyampaikan opini adalah melalui Twitter. Aplikasi ini merupakan salah satu media sosial yang paling banyak digunakan yaitu mencapai 18.45 juta per Agustus 2022 (Rizaty, 2022). Media sosial Twitter merupakan media sosial berupa microblog yang awalnya hanya menampung 140 karakter (sekarang sudah bisa 280 karakter), dimana pesan yang ditulis tersebut biasa disebut dengan Pinjaman online merupakan fasilitas penyedia jasa keuangan yang beroperasi secara daring berupa layanan pinjam uang berbasis aplikasi (Utami & Erfina, 2021). Adanya aplikasi pinjaman *online* ini menimbulkan pro dan kontra di masyarakat *tweet* (Nurhadi, 2017).

Salah satu fitur Twitter adalah *trending* topik yang menunjukkan percakapan apa yang sedang banyak diperbincangkan pada saat tertentu. Salah satu topik yang dibahas adalah terkait aplikasi pinjaman *online* yang menjamur di Indonesia. Pinjaman *online* merupakan fasilitas penyedia jasa keuangan yang beroperasi secara daring berupa layanan pinjam uang berbasis aplikasi (Utami & Erfina, 2021). Salah satu penelitian menjelaskan bahwa adanya aplikasi pinjaman online kredivo bertujuan untuk mempermudah pengguna dalam melakukan transaksi finansial dan memberikan pelayanan keuangan yang baik kepada penggunanya (Panginan & Irwansyah, 2020).

Adanya aplikasi pinjaman *online* ini menimbulkan pro dan kontra di masyarakat. Di paruh kedua 2021, aplikasi pinjaman *online* menjadi salah satu topik yang diperbincangkan di Twitter karena banyaknya korban aplikasi peminjaman *online*. Beberapa diantaranya adalah dengan hashtag #IndonesiaDaruratPinjol (Asria, 2021), #JokowiBrantasPinjolIllegal (redaksi, 2021), #JokowiStopPinjolBaru (Aco, 2021) #AyoPolisiSikatPinjol (Rmolnetwork, 2021) dan lain-lain.

Dalam trending dengan hashtag-hashtag tersebut, banyak masyarakat yang menggunakan kata “pinjol” (yang merupakan singkatan dari pinjaman

online) untuk menyatakan pendapat mereka terkait tentang aplikasi pinjaman *online* tersebut. Pendapat ataupun opini yang disampaikan terkait pinjaman online tersebut beragam meskipun sepiantas terlihat kebanyakan opini yang diberikan adalah negatif terutama terkait pinjaman online, namun terdapat juga *tweet* positif terkait dukungan terhadap upaya pemerintah. Untuk dapat menganalisis lebih jauh mengenai sentiment masyarakat terhadap keberadaan pinjaman online, diperlukan suatu analisis pada teks-teks data *tweet* yang ada. Namun, jika hal tersebut dilakukan secara manual akan memakan waktu yang cukup lama.

Beberapa teknik pada *machine learning* dapat digunakan untuk membantu melakukan analisis sentimen secara otomatis. Sentimen analisis atau *opinion mining* adalah riset komputasi atas opini, sentimen, dan emosi yang diungkapkan secara tekstual yang dilakukan untuk melihat opini tentang suatu masalah, atau untuk mengidentifikasi tren dalam berbagai hal di pasar. Saat ini opini publik menjadi sumber penting dalam pengambilan keputusan tentang suatu produk (Ipmawati, Kusriani, & Luthfi, 2017).

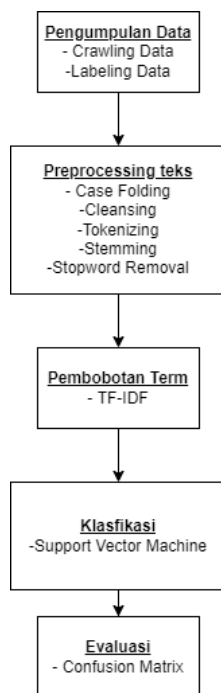
Pada penelitian sebelumnya, peneliti telah melakukan analisis sentiment terhadap kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) di tahun 2021 dengan menggunakan metode Support Vector Machine (SVM) dengan akurasi mencapai 60% (Putra, Haeirudin, Khairunnisa, & Latifah, 2021). Nilai yang diperoleh tidak terlalu bagus, Namun pada penelitian lain terkait analisis sentiment menggunakan SVM dan lexicon-based pada layanan telekomunikasi dengan akurasi 84% (Nurjanah, Perdana, & Fauzi, 2017). Selain itu, pada analisis sentimen vaksin COVID-19 di Indonesia menggunakan berbagai metode klasifikasi juga ditemukan bahwa evaluasi terbaik adalah menggunakan metode SVM dengan akurasi 84% (Latifah, Baddalwan, Meilina, Saputra, & Adharani, 2021).

Publikasi terkait analisis sentimen pinjaman online sudah dilakukan pada tahun 2021 dengan menggunakan SVM dengan akurasi 62%, tetapi tidak

disebutkan waktu pengambilan datanya (Utami & Erfina, 2021). Seperti disebutkan sebelumnya, isu pinjaman online menaik di paruh kedua tahun 2021, hal tersebut mendasari dilakukannya penelitian ini.

## 2. METODE PENELITIAN

Tahapan Penelitian dapat dijelaskan dengan alur diagram berikut ini.



**Gambar 1.** Tahapan Penelitian

Pengumpulan data dilakukan dengan menggunakan API untuk melakukan crawling pada Twitter. *Crawling* data merupakan tahap dalam penelitian yang bertujuan untuk mengumpulkan atau mengunduh data dari suatu database (Sembodo, Setiawan, & Baizal, 2016). Tujuan awal dibentuknya twitter API ini adalah untuk mengetahui relasi dan interaksi antar pengguna, namun sebaliknya twitter API banyak digunakan untuk menggali informasi komunitas tertentu menurut pandangannya terhadap suatu topik (Muis & Affandes, 2015).

Data diambil selama kurun waktu 11 – 15 Oktober 2021 dengan kata kunci “pinjol”. Dari hasil diperoleh sebanyak 15.000 data tweet dengan kata kunci tersebut. Data kemudian difilter untuk menghilangkan duplikasi sehingga

terdapat 6500 tweet yang dilakukan pelabelan otomatis menggunakan *sentistrength* bahasa Indonesia dimana jika skor sentiment >0 maka termasuk positif, jika skor <0 maka termasuk negative dan jika skor = 0 maka netral (Santoso & Nugroho, 2019). Data yang sudah diberi label secara otomatis tersebut kemudian divalidasi secara manual.

Sebelum dilakukan pemodelan, dilakukan pra proses data teks dengan tahapan yang sama seperti penelitian sebelumnya di (Putra, Haeirudin, Khairunnisa, & Latifah, 2021) dan (Latifah, Baddalwan, Meilina, Saputra, & Adharani, 2021) yaitu melakukan *case folding / lower casing* atau mengubah huruf besar menjadi huruf kecil, *cleansing* atau menghilangkan *noise* seperti angka, simbol dan kata-kata stopword, serta melakukan tokenisasi dan stemming menggunakan library sastrawi. *Stopword* adalah sebuah kata yang dieliminasi pada saat melakukan *Preprocessing* pada *text mining* yang merupakan kata-kata yang paling sering muncul dalam suatu dokumen yang memiliki sedikit nilai dan sedikit membantu ketika memilih dokumen yang dibutuhkan pengguna. Sedangkan stemming adalah proses mendapatkan kata dasar dengan menghilangkan imbuhan pada kata (Susilowati, Sabariah, & Gozali, 2015). Contoh hasil pra proses data tweet dapat dilihat pada tabel 1.

Selanjutnya adalah melakukan pembobotan data dengan menggunakan metode TF-IDF. Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah salah satu proses dari teknik ekstraksi ciri dengan proses pemberian nilai pada setiap kata pada tweet data latih. Adapun rumusnya adalah mengalikan nilai frekuensi term atau kata (TF) dengan invers dari frekuensi dokumennya (IDF) (Pravina & Adikara, 2019).

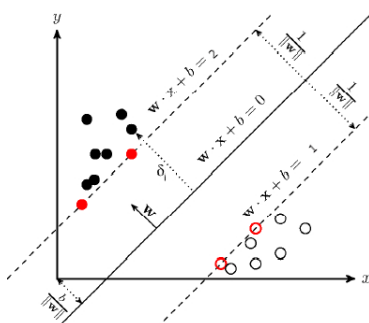
$$W_t = tf_t * idf_t$$
$$idf_t = \log_{10}(N/df_t)$$

dimana  $tf$  adalah frekuensi kemunculankata pada setiap tweet,  $N$  adalah jumlah tweet dan  $df_t$  adalah jumlah tweet yang memiliki kata  $t$ .

**Tabel 1.** Contoh Hasil Pra Proses

Input Tweet	Output Case folding dan cleansing	Output stemming, stopword dan tokenisasi
@Vivacoid Penagihan pinjol online berada di bawah pengawasan OJK pun tidak jauh beda dengan ilegal. <a href="https://t.co/ztN6SCD">https://t.co/ztN6SCD</a> NgR	pinjol online berada di bawah pegawasn ojk pun tidak jauh beda dengan ilegal.	tagih,pinjol,online, bawah,awas, ojk, beda,illegal

Pemodelan data yang sudah direpresentasikan dalam bobot TFIDF, dilakukan dengan menggunakan Support Vector Machine (SVM) Kelebihan pada SVM ini adalah efektif pada *high dimensional space*, efektif dalam kasus dengan jumlah dimensi yang lebih banyak daripada jumlah sampelnya, dengan menggunakan subset titik pelatihan sehingga lebih memori efisien (Adrian, Putra, Rafialdy, & Rakhmawati, 2021). Atas dasar tersebut, SVM cocok digunakan untuk pengolahan data teks yang memiliki dimensi yang sangat tinggi. Gambar 1 menunjukkan ilustrasi dari algoritma SVM.



**Gambar 2** Visualisasi SVM (Susilowati, Sabariah, & Gozali, 2015)

Pada gambar di atas terlihat bahwa terdapat dua kelas data yang terpisah secara linier yaitu kelas bulatan hitam dan kelas bulatan putih. *Hyperplane* memisahkan kedua kelas dengan persamaan

$$w \cdot x + b = 0$$

dimana  $w$  adalah normal bidang dan  $b$  adalah bias atau posisi bidang relatif terhadap pusat koordinat. Support vector merupakan titik data terdekat dengan hyperplane dari masing-masing kelas.

Persamaan yang terbentuk dari garis pembatas dari kedua kelas adalah :

$$\text{untuk } y = +1, \text{ maka } w \cdot x + b = 1$$

$$\text{untuk } y = -1, \text{ maka } w \cdot c + b = -1$$

sedangkan untuk melakukan klasifikasi dari data pengujian  $x$  dapat ditentukan dengan rumus :

$$f(x) = \text{sgn}[\sum_{i=1}^n a_i y_i x_i \cdot x - b]$$

### 3. Hasil dan Pembahasan

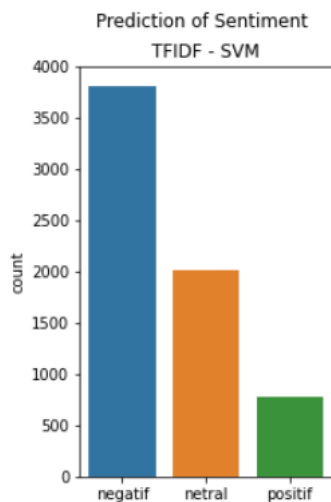
Algoritma SVM diuji dengan menggunakan precision, recall dan F1 score. Digunakan 6500 data *tweet* yang sudah divalidasi secara manual dan dibagi menjadi data training dan data testing. Dari hasil percobaan menggunakan berbagai penalty nilai  $C$  menggunakan library sklearn diketahui bahwa dari  $C = 1.5$  menunjukkan nilai f1 score di atas 80% dengan masing-masing kategori sentimen memiliki precision, recall dan f1 score di atas 60% seperti ditunjukkan pada tabel 2. Dari hasil tersebut menunjukkan bahwa algoritma SVM dengan kernel RBF dan penalty  $C=1,5$  memiliki evaluasi yang cukup baik untuk melakukan otomatisasi pelabelan sentimen secara otomatis.

**Tabel 2.** Hasil Evaluasi Algoritma SVM dengan berbagai nilai  $C$

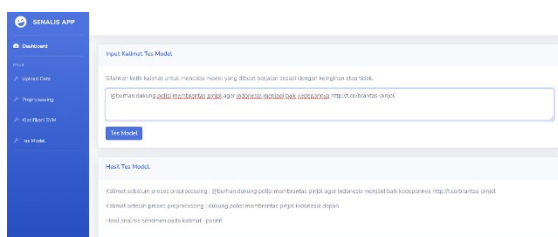
Kategori	C=0,5			C=1		
	Precision	recall	f1 score	Precision	recall	f1 score
negatif	0,71	0,95	0,81	0,8	0,93	0,86
netral	0,97	0,97	0,77	0,95	0,73	0,83
positif	0,34	0,34	0,51	0,57	0,98	0,72
macro average	0,67	0,86	0,7	0,78	0,88	0,8
weighted average	0,85	0,76	0,77	0,86	0,83	0,83

Kategori	C=1,5			C=5		
	Precision	recall	f1 score	Precision	recall	f1 score
negatif	0,83	0,92	0,87	0,83	0,92	0,87
netral	0,93	0,76	0,83	0,93	0,76	0,83
positif	0,63	0,94	0,75	0,63	0,94	0,75
macro average	0,79	0,87	0,82	0,79	0,87	0,82
weighted average	0,86	0,84	0,84	0,86	0,84	0,84

Model tersebut kemudian digunakan untuk memprediksi 6500 *tweet* lain yang belum memiliki label. Hasilnya diperoleh 3800 *tweet* memiliki sentiment negatif, 2000 memiliki sentiment netral dan 700 memiliki sentiment positif seperti terlihat pada gambar 3. Hasil ini sesuai dengan hasil penelitian mengenai analisis sentiment pinjaman *online* yang dilakukan pada saat yang hampir bersamaan dimana diketahui salah satu puncak tren ada pada 14 Oktober 2021 dimana penangkapan pertama dilakukan (Utami, Purnama, & Hidayanto, 2022). Selain itu, model juga digunakan untuk membangun sistem berbasis web yang dapat digunakan untuk melakukan pelabelan secara otomatis seperti terlihat pada gambar 4.



**Gambar 3.** Visualisasi klasifikasi non label



**Gambar 4.** Tampilan Sistem Prediksi Sentimen Otomatis

Dari hasil analisis *word cloud* pada sentiment negatif di gambar 5 terlihat bahwa kata-kata yang sering muncul adalah illegal, bantai, dan polri serta hashtag #JokowiBrantasPinjolIllegal. Hal ini menunjukkan bahwa pinjaman *online* banyak yang illegal dan perlu ditangani oleh pihak berwajib. Sedangkan pada sentiment positif menunjukkan kata-kata yang berhubungan dengan polisi yang merujuk pada dukungan masyarakat terhadap penindakan pinjaman *online* illegal oleh kepolisian.



**Gambar 5.** *Word cloud* data non label untuk sentiment negatif dan positif



#### 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan diketahui bahwa algoritma SVM kernel RBF dengan penalti  $C=1.5$  dapat digunakan untuk melakukan prediksi sentimen data *tweet* terkait topik pinjaman online dengan kinerja *f1 score* mencapai 84%. Selain itu, lebih dari 50% *tweet* menunjukkan respon negative terhadap adanya pinjaman *online* dan perlunya Tindakan dari pemerintah untuk menyelesaikan polemik tersebut. Masyarakat juga mendukung adanya partisipasi aktif dari instansi pemerintah termasuk dari kepolisian terlihat dari *word cloud tweet* bersentimen positif.

#### DAFTAR PUSTAKA

- Aco, H. (2021, October 16). *Pinjol Ilegal Marak dan Makan Korban hingga Bunuh Diri, #JokowiStopPinjolBaru Menggema di Twitter*. Retrieved from Tribunnews: <https://www.tribunnews.com/nasional/2021/10/16/pinjol-ilegal-makin-marak-dan-makan-korban-jokowistoppinjolbaru-menggema-di-twitter>
- Adrian, M. R., Putra, M. P., Rafialdy, M. H., & Rakhmawati, N. A. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB. *Jurnal Informatika Upgris*.
- Asria, D. (2021, June 18). *Trending di Twitter, Indonesia Darurat Pinjaman Online, Warganet: Hati-hati! Data Pribadi Tersebar*. Retrieved from Kabar Banten Pikiran Rakyat: <https://kabarbanten.pikiran-rakyat.com/nasional/pr-592077461/trending-di-twitter-indonesia-darurat-pinjaman-online-warganet-hati-hati-data-pribadi-tersebar>
- Fikri, M., & Sarno, R. (2019). A comparative study of sentiment analysis using SVM and SentiWordNet. *Indonesian Journal of Electrical Engineering and Computer Science* 13(3).
- Ipmawati, J., Kusriani, & Luthfi, E. T. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security* 6 (1).
- Latifah, R., Baddalwan, R., Meilina, P., Saputra, A. D., & Adharani, Y. (2021). Sentiment Analysis of COVID-19 Vaccines from Indonesian Tweets using Various Machine Learning Techniques. *2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. Jakarta.
- Muis, I. A., & Affandes, M. (2015). Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet. *SiTekIn : Jurnal Sains, Teknologi dan Industri*.
- Nurhadi, Z. F. (2017). Model Komunikasi Sosial Remaja Melalui Media Twitter. *Jurnal ASPIKOM*.
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Sentiment Analysis of Television Shows Based on Public Opinion on Twitter Social Media using the K-Nearest Neighbor Method and Weighting the Number of Retweets. *JPTIHK : Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- Panginan, E. K., & Irwansyah. (2020). Fenomena Aplikasi Kredit dan Pinjaman. *Jurnal Komunikasi Dan Kajian Media Karya*.
- Pravina, A. M., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi informasi dan Ilmu Komputer*, (pp. 2789-2797). Malang.
- Putra, A., Haeirudin, D., Khairunnisa, H., & Latifah, R. (2021). Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma SVM. *SEMNASSTEK 2021*. Jakarta.

- redaksi. (2021, October 15). *Viral Kantor Pinjol Ilegal Digrebek, #JokowiBrantasPinjolIllegal Trending di Twitter*. Retrieved from [indonesiatech.id: https://www.indonesiatech.id/2021/10/15/viral-kantor-pinjol-ilegal-digrebek-jokowibrantaspinjolilegal-trending-di-twitter/](https://www.indonesiatech.id/2021/10/15/viral-kantor-pinjol-ilegal-digrebek-jokowibrantaspinjolilegal-trending-di-twitter/)
- Rizaty, M. A. (2022, Agustus 10). *Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022*. Retrieved from [dataindonesia.id: https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022](https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022)
- Rmolnetwork. (2021, October 23). *Tagar #AyoPolisiSikatPinjol Trending di Twitter*. Retrieved from [Rmoljatim.id: https://www.rmoljatim.id/2021/10/23/tagar-ayopolisisikatpinjol-trending-di-twitter](https://www.rmoljatim.id/2021/10/23/tagar-ayopolisisikatpinjol-trending-di-twitter)
- Santoso, E. B., & Nugroho, A. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik di Facebook. *Jurnal Eksplora Informatika*.
- Sembodo, J. E., Setiawan, E. B., & Baizal, A. (2016). Data Crawling Otomatis pada Twitter. *Conference: Indonesia Symposium on Computing*. Bandung.
- Susilowati, E., Sabariah, M., & Gozali, A. A. (2015). Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter. *E-Proceeding of Engineering*, (pp. 1478-1484).
- Utami, D. S., & Erfina, A. (2021). ANALISIS SENTIMEN PINJAMAN ONLINE DI TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM). *SISMATIK : Seminar Nasional Sistem Informasi dan Manajemen Informatika*, (pp. 299-305).
- Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). Fintech Lending in Indonesia: A Sentiment Analysis, Topic Modelling, and Social Network Analysis using Twitter Data. *International Journal of Applied Engineering & Technology* 4(1).