

# Perbandingan Algoritma Support Vector Machine dan Naïve Bayes dalam Menganalisis Sentimen Pinjaman Online di Twitter

Yulia Ikhsani\*, Inggih Permana, Febi Nur Salisah, Mustakim, Nesdi Evrilyan Rozanda

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>\*12050320529@students.uin-suska.ac.id, <sup>2</sup>inggihpermana@uin-suska.ac.id, <sup>3</sup>febinursalisah@uin-suska.ac.id,

<sup>4</sup>mustakim@uin-suska.ac.id, <sup>5</sup>nesdi.er@uin-suska.ac.id

Email Penulis Korespondensi: 12050320529@students.uin-suska.ac.id

Submitted: 20/10/2024; Accepted: 01/12/2024; Published: 03/12/2024

**Abstrak**– Penelitian ini membandingkan performa algoritma SVM (*Support Vector Machine*) dan NBC (*Naïve Bayes Classifier*) dalam menganalisis sentimen terkait pinjaman online di *Twitter*. Data yang digunakan terdiri dari 1.200 komentar yang dikumpulkan melalui teknik *crawling*, mencakup pinjol legal dan ilegal. Data ini diproses melalui *text preprocessing* (*cleaning, tokenizing, filtering, dan stemming*) dan diberikan bobot menggunakan metode *TF-IDF*. Penelitian ini menggunakan teknik *SMOTE* untuk menyeimbangkan kelas sentimen dan membagi data menjadi 70% data latih dan 30% data uji. Hasil analisis menunjukkan bahwa algoritma SVM dengan kernel RBF memiliki kinerja terbaik pada dataset pinjol legal, dengan akurasi, presisi, dan recall masing-masing sebesar 70%, 72%, dan 70%. Sedangkan pada dataset pinjol ilegal, algoritma NBC menunjukkan kinerja yang lebih baik dengan akurasi, presisi, dan recall masing-masing sebesar 75%, 76%, dan 75%. Penelitian ini menunjukkan pentingnya penggunaan teknik *SMOTE* untuk meningkatkan performa model dalam mengatasi ketidakseimbangan data. Studi ini memberikan wawasan mengenai efektivitas kedua algoritma dalam analisis sentimen dan relevansinya terhadap pemahaman opini publik di ranah *fintech lending*.

**Kata Kunci:** Pinjaman Online; Support Vector Machine; Naive Bayes Classifier; Sentimen; Twitter

**Abstract**– This study compares the performance of the SVM (*Support Vector Machine*) algorithm and NBC (*Naïve Bayes Classifier*) in analyzing sentiment related to online loans on *Twitter*. The data used consists of 1,200 comments collected through crawling techniques, covering both legal and illegal online loans. The data was processed through text preprocessing (*cleaning, tokenizing, filtering, and stemming*) and weighted using the *TF-IDF* method. The study employs the *SMOTE* technique to balance sentiment classes and splits the data into 70% training data and 30% testing data. The analysis results show that the SVM algorithm with an RBF kernel achieved the best performance on the legal online loan dataset, with accuracy, precision, and recall of 70%, 72%, and 70%, respectively. Meanwhile, for the illegal online loan dataset, the NBC algorithm demonstrated better performance with accuracy, precision, and recall of 75%, 76%, and 75%, respectively. This research highlights the importance of using the *SMOTE* technique to improve model performance in handling data imbalance. The study provides insights into the effectiveness of both algorithms in sentiment analysis and their relevance to understanding public opinion in the realm of *fintech lending*.

**Keywords:** Online Loans; Support Vector Machine; Naïve Bayes Classifier; Sentiment; Twitter

## 1. PENDAHULUAN

Pengangguran menjadi salah satu penyebab kemiskinan di masyarakat, terbatasnya lapangan pekerjaan menyebabkan beberapa orang kesulitan untuk memenuhi kebutuhan hidup sehari-hari mereka [1]. Hal ini juga memicu tingginya permintaan pinjam meminjam di masyarakat. Dengan kemajuan teknologi banyak perusahaan *start up* keuangan yang bermunculan guna membantu masyarakat dalam hal peminjaman dana berupa aplikasi maupun *website* pinjaman online. Pinjaman online memberikan akses mudah bagi masyarakat untuk mendapatkan pinjaman secara cepat dan tanpa proses yang rumit [2]. Meskipun demikian, keberadaan pinjaman online juga memunculkan berbagai pendapat dan sentimen yang beragam, baik positif, netral maupun negatif, di kalangan masyarakat khususnya di sosial media [3].

Salah satu media sosial yang paling populer adalah *Twitter*, yang menjadi tempat bagi individu untuk menyampaikan pendapat, berbagi pengalaman, dan berinteraksi dalam diskusi publik [4]. Pentingnya analisis sentimen dalam konteks ini terletak pada kemampuan untuk memahami pandangan atau opini masyarakat terhadap suatu topik atau isu tertentu [5]. Dengan menganalisis sentimen masyarakat terhadap pinjaman online, pihak penyedia layanan maupun regulator dapat memperoleh wawasan yang berharga untuk meningkatkan kualitas layanan, mengidentifikasi potensi masalah, dan merancang kebijakan yang lebih responsif terhadap kebutuhan masyarakat.

Salah satu pendekatan yang umum digunakan untuk analisis sentimen adalah menggunakan algoritma pembelajaran mesin (*machine learning*) [6]. Di antara banyak algoritma yang tersedia, dua metode yang cukup populer adalah SVM dan NBC. Algoritma SVM dikenal dengan kemampuannya dalam klasifikasi data yang sangat baik, terutama untuk dataset yang besar dan memiliki dimensi tinggi [7]. Sementara itu, NBC lebih sederhana dan efektif dalam menangani masalah klasifikasi berbasis probabilitas, meskipun cenderung lebih sensitif terhadap asumsi independensi antar fitur [8].

Beberapa penelitian sebelumnya yang juga membahas perihal sentimen, seperti pada penelitian yang dilakukan oleh Ikhsani dan lainnya (2023), perbandingan ini dilakukan untuk mengetahui kinerja dari algoritma SVM dan *Decision Tree* yang mana diperoleh hasil akhir bahwa SVM lebih unggul dibandingkan *Decision Tree* terbukti dengan hasil akurasi yang lebih tinggi yaitu 99%[9]. penelitian oleh Arsi dan Waluyo (2021), menggunakan SVM untuk

menganalisis sentimen terkait pemindahan ibu kota Indonesia, memperoleh hasil akurasi yang tinggi sebesar 96,68% [10].

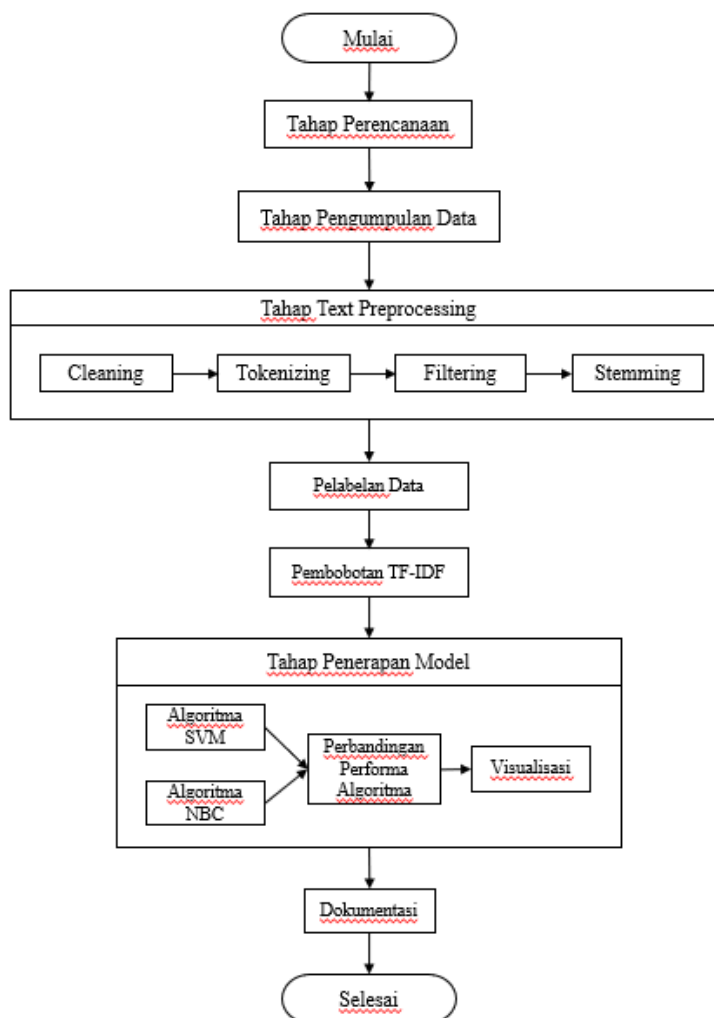
Menurut Fathonah dan Herlina (2021), penerapan *text mining* dalam menganalisis sentimen juga menggunakan algoritma NBC memperoleh hasil yang sangat baik terbukti dari nilai akurasinya yang mencapai 100% [11]. Penelitian oleh Arifin dan lainnya (2022), dalam pengujian kinerja algoritma NBC untuk mengelompokkan data penerima program keluarga harapan berhasil dilakukan. Diketahui metode NBC dikatakan baik dalam pengklasifikasian data tersebut dengan tingginya akurasi mencapai 88% [12]

Meskipun kedua algoritma ini sering digunakan pada penelitian sebelumnya, belum banyak penelitian yang membandingkan performa keduanya dalam analisis sentimen dengan menggunakan 2 dataset yaitu pinjaman online legal dan ilegal. Penelitian ini juga akan menerapkan 4 kernel pada algoritma SVM dan akan menggunakan teknik *SMOTE* untuk menyeimbangkan datanya. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja algoritma *Support Vector Machine* dan *Naive Bayes Classifier* dalam mengklasifikasikan sentimen yang terkait dengan pinjaman online di *Twitter*. Dengan melakukan perbandingan ini, diharapkan dapat ditemukan algoritma yang lebih efektif dalam memberikan hasil analisis yang akurat dan relevan untuk memahami sentimen publik terhadap pinjaman online.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Pada dasar metodologi penelitian ini, tahap awal yang dilakukan yaitu perencanaan, dilanjutkan dengan pengumpulan data melalui Twitter dengan cara *Crawling*, yang diikuti tahapan *text preprocessing*. Selanjutnya tahapan implementasi dari kedua algoritma yaitu SVM dan NBC untuk mengetahui perbandingan performa keduanya, lalu dilakukan visualisasi. Kemudian ditutup dengan tahap dokumentasi. Alur metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

## 2.2 Tahap Penerapan Model

Tahapan ini berisi proses pelabelan data, pembobotan menggunakan TF-IDF, klasifikasi menggunakan algoritma SVM dan NBC, perbandingan performa kedua algoritma dan visualisasi data komentar. Sebagai berikut :

- Pelabelan Data, data akan dibagi menjadi tiga sebagai kelas positif, negatif juga netral. Proses pelabelan dilakukan oleh pakar secara manual.
- Pembobotan dengan TF-IDF  
Tahap ini melibatkan pemberian bobot pada tiap kata memakai metode Term Frequency dan Inverse Document Frequency[13], [14].
- Klasifikasi menggunakan *Support Vector Machine* dan *Naive Bayes Classifier*  
Pengklasifikasian menggunakan algoritma SVM dan NBC akan ditampilkan dalam bentuk tabel hasil pengujian berisi nilai akurasi, presisi, dan recall dengan persamaan berikut [15].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

Singkatan yang terdapat pada formula diatas menyatakan; TP untuk *True Positif*, TN untuk *True Negative*, FP untuk *False Positif*, dan FN untuk *False Negative*.

Pada algoritma SVM juga akan menggunakan 4 kernelnya yang umum digunakan yaitu kernel Linear (4), kernel Polinomial (5), kernel RBF (6), kernel Sigmoid (7) berikut [16].

$$K = (X_i, X_j) = (X_i, X_j) \quad (4)$$

Kernel linear menghitung inner product antara dua vektor  $X_i$  dan  $X_j$ , cocok untuk data yang dapat dipisahkan secara linear.

$$K = (X_i, X_j) = (X_i, X_j + 1)^p \quad (5)$$

Kernel polinomial menambahkan inner product  $(X_i, X_j)$  dengan konstanta dan menaikannya ke pangkat  $p$ . Parameter  $p$  menentukan derajat polinomial, yang mengontrol kompleksitas dari pemisah data.

$$K = (X_i, X_j) = e^{-\gamma(X_i, X_j)^2} \quad (6)$$

Kernel Gaussian atau RBF, bergantung pada jarak kuadrat antara dua titik data. Parameter  $\gamma$  mengontrol seberapa jauh pengaruh setiap titik data, semakin besar nilai  $\gamma$  semakin sempit pengaruhnya.

$$K = (X_i, X_j) = \tanh(\eta X_i \cdot X_j + v) \quad (7)$$

Kernel sigmoid menyerupai fungsi aktivitas pada jaringan saraf. Parameter  $\eta$  dan  $v$  mengontrol bentuk kurva sigmoid, dengan  $\eta$  sebagai skala dan  $v$  sebagai bias.

Penelitian ini juga menggunakan Grid search sebagai metode tuning parameter [17], dan akan menggunakan teknik SMOTE (Synthetic Minority Oversampling Technique) guna menyeimbangkan data [18].

- Perbandingan Performa Algoritma  
Dilakukan untuk menentukan algoritma mana yang menunjukkan kinerja lebih baik dalam menganalisa sentimen, akan diukur dari nilai akurasi, presisi, dan recall. Lalu akan di dapat perolehan perbandingan dari akurasi antar algoritma pada kedua dataset yang digunakan.
- Visualisasi  
Akan menampilkan wordcloud dari keseluruhan kata dan dari sentimen positif, netral dan negatif.

## 2.3 Dokumentasi

Tahap dokumentasi adalah tahap akhir dari seluruh kegiatan penelitian, mulai dari tahap perencanaan hingga hasil akhir penelitian.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini berupa opini masyarakat terhadap pinjaman online di Twitter. Rentang waktu sejak 01 Januari 2023 sampai 30 September 2024. Teknik pengumpulan data adalah *crawling* dengan bahasa pemrograman *python* pada *google colab*. Dataset awal yang akan diolah terlihat pada Tabel 1 dan Tabel 2.

**Tabel 1.** Dataset Awal Pinjol Legal

No.	Komentar
1	Stay safe dan selalu waspada! Kalo ragu mending hati-hati daripada menyesal kemudian. #SelaluAdaBuatKamu #SelaluWaspada
2	Kalo nemuin yang kayak gitu buruan laporin ke CS resmi dan hotline AdaKami di 15000-77 supaya orang lain nggak jadi korban. #SelaluAdaBuatKamu #SelaluJagainKamu
...	...
600	@SamaPinjol Iya sih bang gue posisi sama ibu kontrakan ga bayangin malunya nih. Wah gue juga lg uatas nih setelah berapa lama bang kalau boleh tau si fc dateng

Tabel 1 merupakan data awal pinjol legal yang berjumlah sebanyak 600 data komentar, yang didapat dengan cara *crawling* di *twitter* dengan menggunakan kata kunci pinjol legal, pinjol resmi, pinjol adakami, amarnya dan adapundi.

**Tabel 2.** Dataset Awal Pinjol Ilegal

No.	Komentar
1	Ramai WNI di Luar Negeri Jadi Korban Pinjol Ilegal
2	Korban gali tutup lobang pinjol sampai kejawab 27 pinjol ilegal bunga tinggi tenor pendek Galbay aman nyaman tanpa sebar data backup cloning apk aman galbay bunda tiara siap bantu
...	...
600	@brobill123 Kak pindu ada dc lapangan ga? Terus keluar kondar ga? Soalnya kondar aku nomor yang udah ga aktif.

Tabel 2 merupakan data awal pinjol ilegal yang juga berjumlah sebanyak 600 data komentar, yang didapat dengan cara *crawling* di *twitter* dengan menggunakan kata kunci pinjol ilegal.

### 3.2 Text Preprocessing

Pada tahap ini akan ada empat proses yaitu *cleaning*, *tokenizing*, *filtering* dan juga *stemming* [19],[20]. Berikut tahapan pada *text preprocessing* data:

- a. *Cleaning*, proses ini bertujuan menghapus karakter yang bukan simbol atau huruf, seperti nama pengguna, hastag, angka, dan link. *Output* ada pada Tabel 3.

**Tabel 3.** Hasil *Cleaning*

Kategori	Sebelum	Setelah
Pinjol Legal	@SamaPinjol Iya sih bang gue posisi sama ibu kontrakan ga bayangin malunya nih. Wah gue juga lg uatas nih setelah berapa lama bang kalau boleh tau si fc dateng	iya sih bang gue posisi sama ibu kontrakan ga bayangin malunya nih wah gue juga lg uatas nih setelah berapa lama bang kalau boleh si fc dateng
Pinjol Ilegal	@brobill123 Kak pindu ada dc lapangan ga? Terus keluar kondar ga? Soalnya kondar aku nomor yang udah ga aktif	kak pindu ada dc lapangan ga terus keluar kondar ga soalnya kondar aku nomor yang udah ga aktif

Tabel 3 memperlihatkan data komentar pada pinjol legal dan pinjol ilegal sebelum dan sesudah dilakukannya proses *cleaning*. Sebelumnya pada data komentar memiliki nama pengguna (@SamaPinjol dan @brobill123) serta memiliki tanda baca (?), namun setelah dilakukan proses *cleaning* nama pengguna dan tanda baca tersebut telah dihapus. Menandakan proses *cleaning* berhasil.

- b. *Tokenizing*, adalah proses memecah kalimat menjadi bagian-bagian kata. Setiap komentar diambil minimal 3 kata untuk memahami maknanya. Proses ini terdapat pada Tabel 4.

**Tabel 4.** Hasil *Tokenizing*

Kategori	Sebelum	Setelah
Pinjol Legal	iya sih bang gue posisi sama ibu kontrakan ga bayangin malunya nih wah gue juga lg uatas nih setelah berapa lama bang kalau boleh si fc dateng	['iya', 'sih', 'bang', 'gue', 'posisi', 'sama', 'ibu', 'kontrakan', 'ga', 'bayangin', 'malunya', 'nih', 'wah', 'gue', 'juga', 'lg', 'uatas', 'nih', 'setelah', 'berapa', 'lama', 'bang', 'kalau', 'boleh', 'si', 'fc', 'dateng']

Pinjol Ilegal	kak pindu ada dc lapangan ga terus keluar kendar ga soalnya kendar aku nomor yang udah ga aktif	['kak', 'pindu', 'ada', 'dc', 'lapangan', 'ga', 'terus', 'keluar', 'kendar', 'ga', 'soalnya', 'kendar', 'aku', 'nomer', 'yang', 'udah', 'ga', 'aktif']
---------------	---	--

Setelah dilakukan proses tokenizing seperti pada Tabel 4 dimana data komentar yang awalnya masih dalam bentuk kalimat kemudian berubah menjadi satuan kata.

- c. Filtering, guna menghapus kata yang tidak bermakna secara signifikan dari token yang telah diperoleh sebelumnya. Untuk hasil lihat Tabel 5.

**Tabel 5.** Hasil *Filtering*

Kategori	Sebelum	Setelah
Pinjol Legal	['iya', 'sih', 'bang', 'gue', 'posisi', 'sama', 'ibu', 'kontrakan', 'ga', 'bayangin', 'malunya', 'nih', 'wah', 'gue', 'juga', 'lg', 'uatas', 'nih', 'setelah', 'berapa', 'lama', 'bang', 'kalau', 'boleh', 'si', 'fc', 'dateng']	['iya', 'sih', 'bang', 'gue', 'posisi', 'kontrakan', 'ga', 'bayangin', 'malunya', 'nih', 'gue', 'lg', 'uatas', 'nih', 'bang', 'si', 'fc', 'dateng']
Pinjol Ilegal	['kak', 'pindu', 'ada', 'dc', 'lapangan', 'ga', 'terus', 'keluar', 'kendar', 'ga', 'soalnya', 'kendar', 'aku', 'nomer', 'yang', 'udah', 'ga', 'aktif']	['kak', 'pindu', 'dc', 'lapangan', 'ga', 'kendar', 'ga', 'kendar', 'nomer', 'udah', 'ga', 'aktif']

Pada Tabel 5 telah diperlihatkan hasil dari sebelum dan setelah dilakukannya proses filtering, yang mana setelah dilakukan proses filtering banyak kata yang dihapus pada kalimat tersebut.

- d. Stemming, mengubah kata pada kalimat dari yang berimbuhan menjadi kata dasar, dan dilakukan pengelompokkan untuk kata-kata dengan makna dasar yang sama namun bentuknya berbeda. Ada di Tabel 6.

**Tabel 6.** Hasil *Stemming*

Kategori	Sebelum	Setelah
Pinjol Legal	['iya', 'sih', 'bang', 'gue', 'posisi', 'kontrakan', 'ga', 'bayangin', 'malunya', 'nih', 'gue', 'lg', 'uatas', 'nih', 'bang', 'si', 'fc', 'dateng']	['iya', 'sih', 'bang', 'gue', 'posisi', 'kontrak', 'ga', 'bayangin', 'malu', 'nih', 'gue', 'lg', 'uatas', 'nih', 'bang', 'si', 'fc', 'dateng']
Pinjol Ilegal	['kak', 'pindu', 'dc', 'lapangan', 'ga', 'kendar', 'ga', 'kendar', 'nomer', 'udah', 'ga', 'aktif']	['kak', 'pindu', 'dc', 'lapang', 'ga', 'kendar', 'ga', 'kendar', 'nomer', 'udah', 'ga', 'aktif']

Proses stemming pada Tabel 6 memperlihatkan perubahan kata yang memiliki imbuhan menjadi kata dasar, seperti pada kata “kontrakan” yang setelah dilakukan proses stemming berubah menjadi kata “kontrak”.

### 3.2 Pelabelan

Pelabelan data dilakukan secara manual oleh pakar, dimana kedua dataset akan terbagi menjadi masing-masing 3 label yaitu: positif, negatif, dan netral. Hasilnya dapat dilihat pada Tabel 7 dan Tabel 8.

**Tabel 7.** Hasil Pelabelan Pinjol Legal oleh Pakar

No.	Sentimen	Komentar
1	negatif	stay safe waspada ragu mending hati hati sesal selaluadabuatkamu selaluwaspada
2	negatif	nemuin kayak gitu buru laporn cs resmi hotline adakami orang nggak korban selaluadabuatkamu selalujagainkamu
...	...	...
598	negatif	iya sih bang gue posisi kontra ga bayangin malu nih gue lg uatas nih bang si fc dateng

Tabel 7 Proses pelabelan telah dilakukan terlebih dahulu oleh pakar, dimana setiap komentar diperiksa jika mengandung kata-kata yang menyatakan kepuasan, kebahagiaan, atau penghargaan maka skornya positif, jika mengandung kata-kata yang menyatakan ketidakpuasan, kemarahan, atau kritik maka skornya negatif, jika kata-katanya mengandung fakta tanpa opini maka skornya netral.

**Tabel 8.** Hasil Pelabelan Pinjol Ilegal oleh Pakar

No.	Sentimen	Komentar
1	negatif	ramai wni negeri korban pinjol ilegal

2	negatif	korban gali tutup lobang pinjol jebak pinjol ilegal bunga tenor pendek galbay aman nyaman sebar data backup cloning apk aman galbay bunda tiara bantu
...	...	...
597	netral	kak pindu dc lapang ga kondar ga kondar nomer udah ga aktif

Untuk pemberian label pada Tabel 8 juga dilakukan oleh pakar dengan menggunakan cara yang sama seperti Tabel 7.

### 3.3 Pembobotan dengan TF-IDF

Proses ini dilakukan agar nilai bobot dari suatu kata bisa didapatkan. Berikut merupakan hasil TF-IDF untuk data pinjol legal dan pinjol ilegal, lihat Tabel 9 dan Tabel 10.

**Tabel 9.** Hasil TF-IDF Pinjol Legal

No.	adakami	...	bayangin	bayar	...	zonauang
1	0	...	0	0	...	0
2	0,1721	...	0	0	...	0
3	0	...	0	0,2294	...	0
...	...	...	...	...	...	...
598	0	...	0,2429	0	...	0

Di Tabel 9 menunjukkan nilai bobot pada kata setiap komentar, kolom “aaaa” sampai “zonauang” merupakan kolom yang berisi kata-kata yang muncul dalam kumpulan data teks. Contoh pada komentar no 2 terdapat kata adakami dan memiliki nilai bobot sebanyak 0,1721 menunjukkan kepentingan kata tersebut dalam dokumen. Semakin besar nilai TF-IDF maka semakin penting kata tersebut.

**Tabel 10.** Hasil TF-IDF Pinjol Ilegal

No.	aman	...	bantu	ilegal	...	zonauang
1	0	...	0	0,0941	...	0
2	0,319076	...	0,175311	0,05238	...	0
3	0	...	0	0,11747	...	0
...	...	...	...	...	...	...
597	0	...	0	0	...	0

Tabel 10 juga menunjukkan nilai bobot pada kata setiap komentar untuk dataset pinjol ilegal, untuk konteksnya sama seperti pada Tabel 9 semakin besar nilai bobot maka semakin penting kata tersebut.

### 3.4 Pembagian Data

Pada penelitian ini menggunakan teknik pembagian data Holdout Validation dimana 70% merupakan data latih dan 30% merupakan data uji.

### 3.5 Penerapan Algoritma SVM dan NBC

Pada tahap ini data akan diimplementasikan ke dalam algoritma SVM dan NBC. Untuk statistik data dapat dilihat pada Tabel 11. Hasil pengujian dari algoritma SVM dan NBC untuk dataset pinjol legal lihat Tabel 12 dan Tabel 13 untuk dataset pinjol ilegal.

**Tabel 11.** Statistik Data Pinjol Legal dan Ilegal

Kategori	Label	Jumlah	
		Tanpa SMOTE	SMOTE
Pinjol Legal	Negatif	280	280
	Netral	153	280
	Positif	165	280
Total		598	840
Pinjol Ilegal	Negatif	300	300
	Netral	112	300
	Positif	185	300
Total		597	900

Tabel 11 menunjukkan jumlah masing-masing sentimen dari keseluruhan data komentar. Pada data pinjol legal setelah dilakukan text preprocessing jumlah keseluruhan datanya menjadi 598 komentar. Untuk sentimen negatif berjumlah 280 komentar, sentimen netral berjumlah 153 komentar, dan sentimen positif berjumlah 165 komentar. Untuk menstabilkan datanya maka digunakan teknik SMOTE, setiap sentimen negatif, netral dan positif memiliki jumlah yang sama yaitu 280 data komentar. Dengan jumlah total datanya sebanyak 840 data.

Tabel 11 juga menunjukkan jumlah keseluruhan data komentar pada dataset pinjol ilegal setelah dilakukan text preprocessing yaitu sebanyak 597 data. Dengan sentimen negatif berjumlah 300 data, sentimen netral berjumlah 112 data, dan sentimen positif berjumlah 185 data. Pada penggunaan teknik SMOTE setiap sentimen negatif, netral dan positif memiliki jumlah data yang sama yaitu 300 data. Dan jumlah total datanya sebanyak 900 data.

**Tabel 12.** Hasil Pengujian pada Dataset Pinjol Legal

Kategori	Algoritma	Kernel	Parameter	Akurasi	Presisi	Recall
Dengan SMOTE	NBC	-	-	69	69	69
	SVM	Linear	{C:10}	68	68	68
		RBF	{C:10, $\gamma$ :1}	70	72	70
		Polynomial	{C:1, d:1, $\gamma$ :10}	68	68	68
		Sigmoid	{C:10, $\gamma$ :1}	66	66	66
Tanpa SMOTE	NBC	-	-	41	41	41
	SVM	Linear	{C:1}	47	44	47
		RBF	{C:1, $\gamma$ :1}	55	57	55
		Polynomial	{C:0,1, d:1, $\gamma$ :10}	47	44	47
		Sigmoid	{C:10, $\gamma$ :0,1}	47	44	47

Tabel 12 memperlihatkan hasil pengujian menggunakan SMOTE dan tanpa SMOTE, algoritma yang dipakai yaitu NBC dan SVM beserta dengan 4 kernelnya. Parameter yang digunakan adalah C (cost) dengan nilai (0,001, 0,01, 0,1, 1, 10),  $\gamma$  (gamma) dengan nilai (0,001, 0,01, 0,1, 1, 10), dan d (degree) dengan nilai (0, 1, 5, 8, 10).

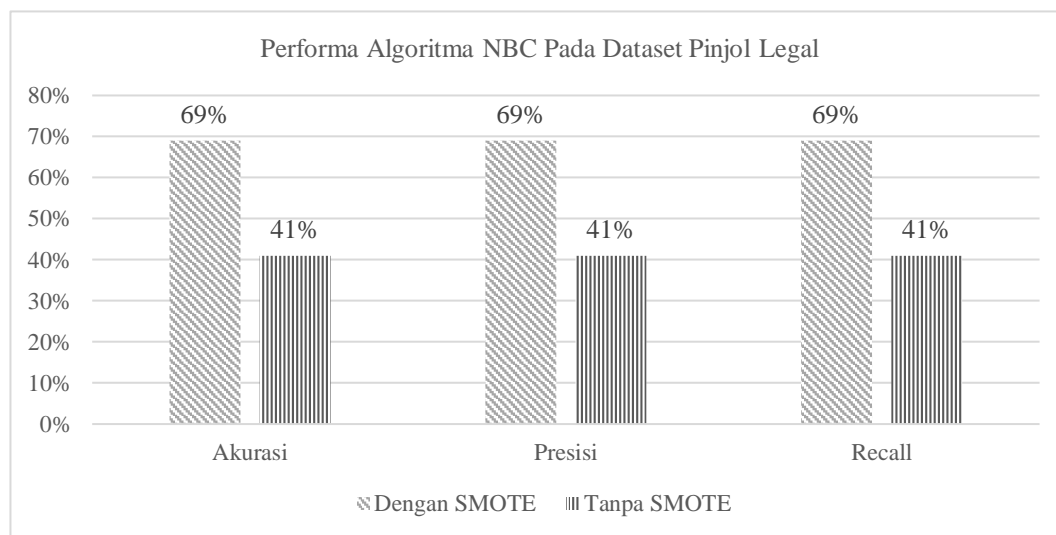
**Tabel 13.** Hasil Pengujian pada Dataset Pinjol Ilegal

Kategori	Algoritma	Kernel	Parameter	Akurasi	Presisi	Recall
Dengan SMOTE	NBC	-	-	75	76	75
	SVM	Linear	{C:10}	74	74	74
		RBF	{C:10, $\gamma$ :1}	74	77	74
		Polynomial	{C:1, d:5, $\gamma$ :1}	74	83	74
		Sigmoid	{C:10, $\gamma$ :0,1}	70	69	70
Tanpa SMOTE	NBC	-	-	42	40	42
	SVM	Linear	{C:1}	45	34	45
		RBF	{C:10, $\gamma$ :0,1}	43	36	43
		Polynomial	{C:0,1, d:1, $\gamma$ :10}	45	34	45
		Sigmoid	{C:10, $\gamma$ :0,1}	45	34	45

Tabel 13 memperlihatkan hasil pengujian menggunakan dataset pinjol ilegal, untuk keterangannya sama seperti pada Tabel 12 hanya berbeda pada nilai parameter dan nilai akurasi, presisi dan recallnya saja.

a. Algoritma Naive Bayes Classifier

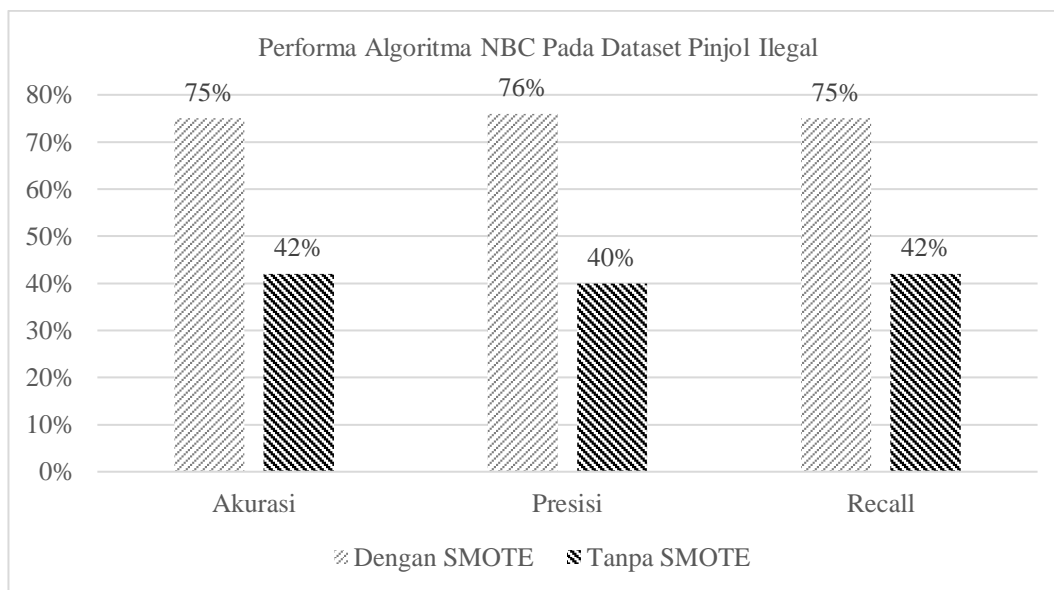
Dalam algoritma NBC diketahui bahwa hasil pengujian sentimen pinjol legal dan ilegal dengan menggunakan teknik SMOTE memiliki performa terbaik. Jelasnya lihat pada Gambar 2 dan Gambar 3.



**Gambar 2.** Performa Algoritma NBC pada Dataset Pinjol Legal

Pada Gambar 2 diperlihatkan nilai akurasi, presisi dan recall dari hasil pengujian menggunakan algoritma NBC pada dataset pinjol legal, dari hasil pengujian tanpa menggunakan teknik SMOTE kinerja algoritma NBC sangat

rendah. Dilihat dari hasil akurasi, presisi dan recallnya yang hanya mencapai 41%. Dan setelah dilakukan pengujian menggunakan teknik SMOTE kinerja dari algoritma NBC menjadi lebih baik dimana nilai akurasi, presisi dan recall yang diperoleh mencapai 69%.

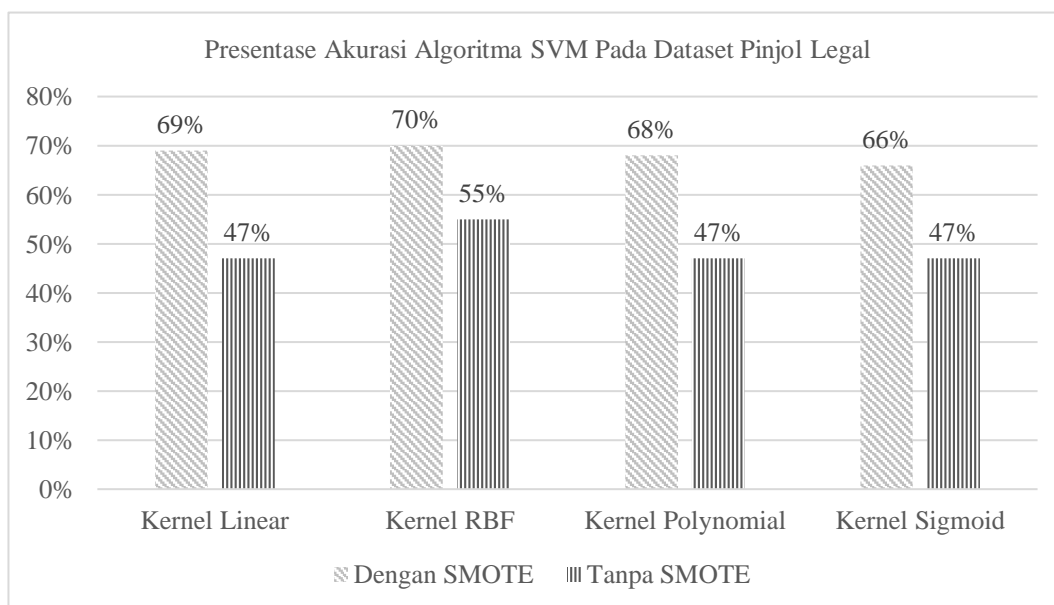


**Gambar 3.** Performa Algoritma NBC pada Dataset Pinjol Ilegal

Performa dari algoritma NBC pada dataset pinjol ilegal diperlihatkan pada Gambar 3, tanpa menggunakan teknik SMOTE nilai akurasi hanya mencapai 42%, presisi sebanyak 40%, dan recall sebanyak 42%. Setelah diterapkan teknik SMOTE pada data nilai akurasi bisa mencapai 75%, presisi mencapai 76%, dan recall mencapai 75%. Maka dapat dikatakan dengan menggunakan teknik SMOTE kinerja algoritma NBC menjadi lebih baik.

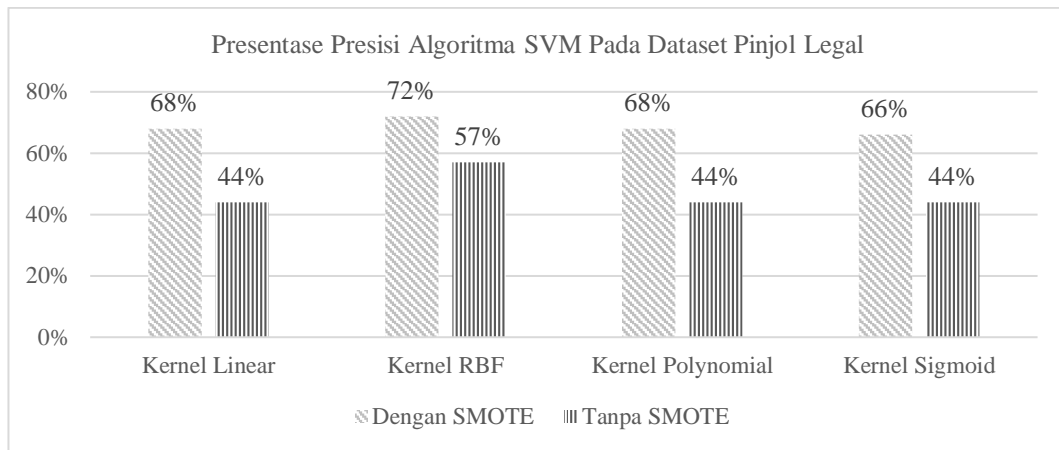
b. Algoritma Support Vector Machine

Berdasarkan Tabel 12 pada dataset pinjol legal performa algoritma SVM terhadap akurasi, presisi dan recall menjadi lebih baik jika masing-masing kernelnya menggunakan teknik SMOTE. Dan hasil performa terbaik dimiliki oleh (kernel RBF, C:10,  $\gamma$ :1) dengan akurasi sebesar 70%, presisi 72% dan recall sebesar 70%. Lebih jelasnya lihat Gambar 4, Gambar 5, dan Gambar 6.



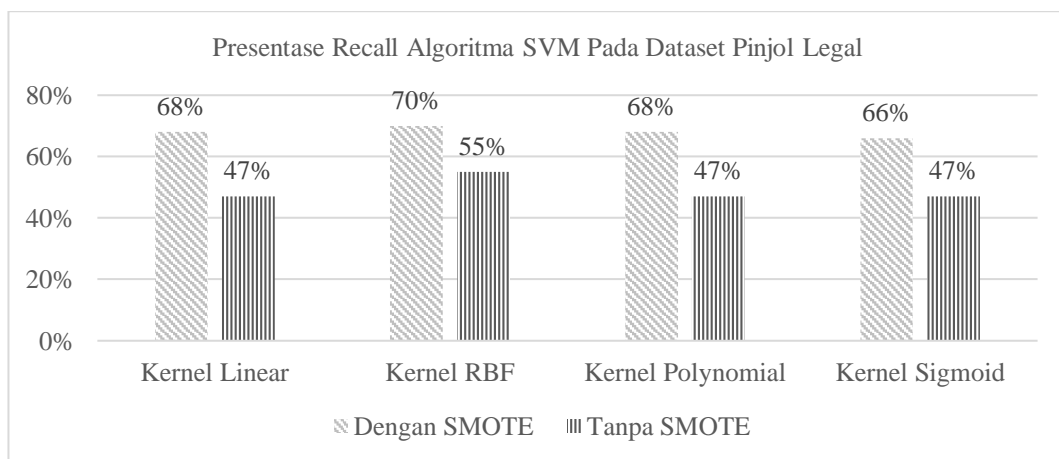
**Gambar 4.** Presentase Akurasi Algoritma SVM pada Dataset Pinjol Legal

Gambar 4 menunjukkan presentase akurasi setiap kernel algoritma SVM, sebelum menggunakan SMOTE kernel linear memiliki akurasi 47%, kernel RBF memiliki akurasi 55%, kernel polinomial akurasinya 47% dan kernel sigmoid akurasi 47%. Setelah menggunakan SMOTE akurasi masing-masing kernel mengalami peningkatan nilai akurasi yaitu, kernel linear sebesar 68%, kernel RBF sebesar 70%, kernel polinomial 68% dan sigmoid sebesar 66%.



**Gambar 5.** Presentase Presisi Algoritma SVM pada Dataset Pinjol Legal

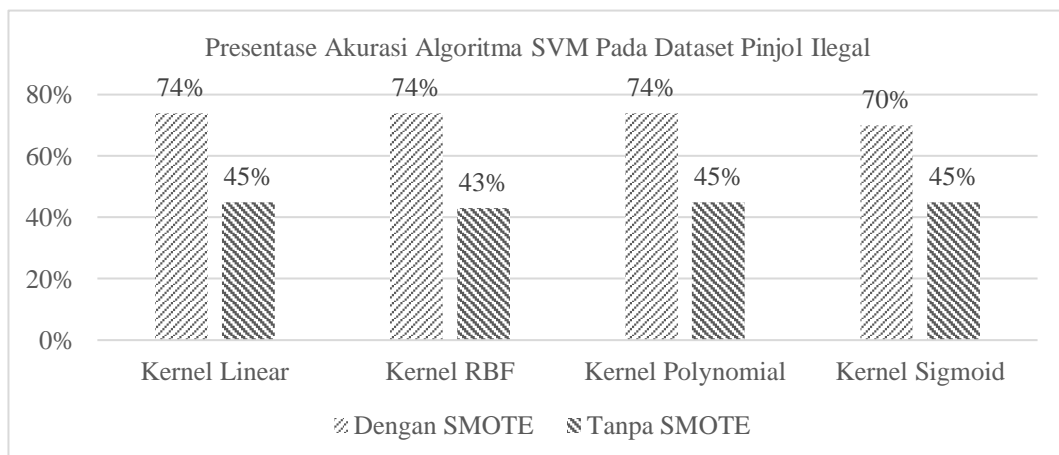
Gambar 5 memperlihatkan nilai presisi pada masing-masing kernel, tanpa SMOTE kernel linear memiliki presisi 44%, kernel RBF sebesar 57%, kernel polinomial sebesar 44% dan sigmoid sebesar 44%. Setelah menerapkan SMOTE nilai presisi setiap kernel menjadi naik, kernel linear menjadi 68%, kernel RBF 72%, kernel polinomial 68%, dan sigmoid menjadi 66%.



**Gambar 6.** Presentase Recall Algoritma SVM pada Dataset Pinjol Legal

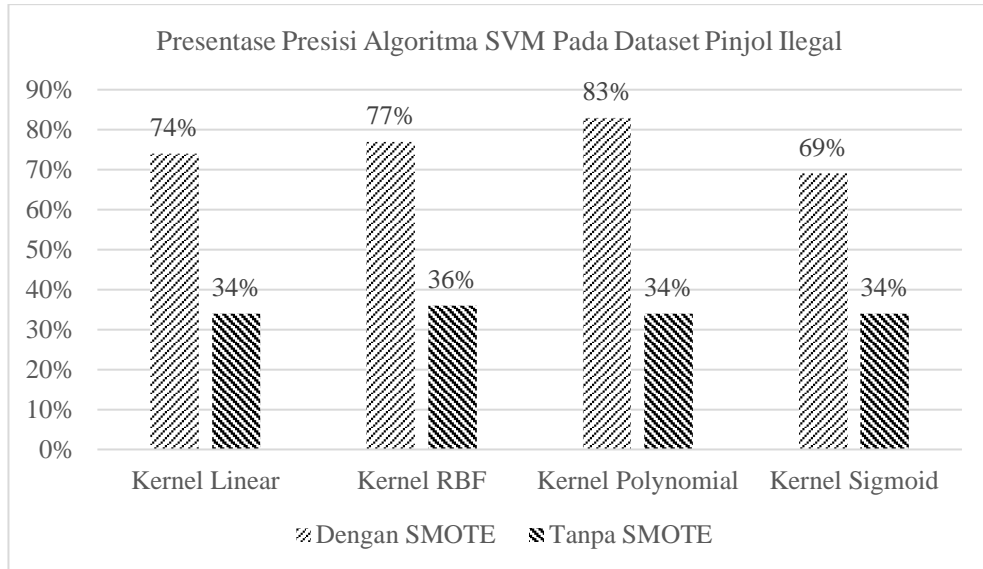
Gambar 6 memperlihatkan nilai recall setiap kernel, tanpa SMOTE kernel linear memiliki recall sebesar 47%, kernel RBF sebesar 55%, kernel polinomial 47%, dan kernel sigmoid 47%. Dengan SMOTE nilai recall kernel linear menjadi 68%, kernel RBF 70%, kernel polinomial 68%, dan kernel sigmoid 66%.

Berdasarkan Tabel 13 pada data pinjol ilegal performa algoritma SVM terhadap akurasi, presisi dan recall juga menjadi lebih baik saat menggunakan teknik SMOTE. Untuk performa SVM terbaik ada pada (kernel polinomial, C;1, d;5,  $\gamma$ :1) dengan akurasi mencapai 74%, presisi sebesar 83% dan recall sebesar 74%. Lebih jelasnya dapat dilihat pada Gambar 7, Gambar 8, dan Gambar 9.



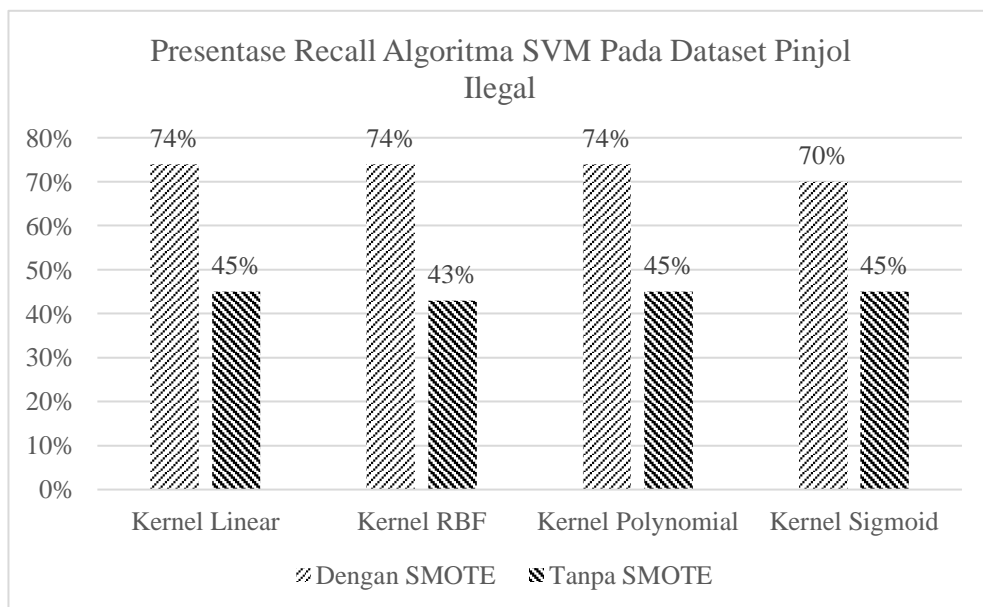
**Gambar 7.** Presentase Akurasi Algoritma SVM Pada Data Pinjol Ilegal

Gambar 7 menunjukkan presentase akurasi setiap kernel algoritma SVM, sebelum menggunakan SMOTE kernel linear memiliki akurasi 45%, kernel RBF memiliki akurasi 43%, kernel polinomial akurasinya 45% dan kernel sigmoid akurasi 45%. Setelah menggunakan SMOTE akurasi masing-masing kernel mengalami peningkatan nilai akurasi yaitu, kernel linear sebesar 74%, kernel RBF sebesar 74%, kernel polinomial 74% dan sigmoid sebesar 70%.



**Gambar 8.** Presentase Presisi Algoritma SVM Pada Data Pinjol Ilegal

Gambar 8 memperlihatkan nilai presisi pada masing-masing kernel, tanpa SMOTE kernel linear memiliki presisi 34%, kernel RBF sebesar 36%, kernel polinomial sebesar 34% dan sigmoid sebesar 34%. Setelah menerapkan SMOTE nilai presisi setiap kernel menjadi naik, kernel linear menjadi 74%, kernel RBF 77%, kernel polinomial 83%, dan sigmoid menjadi 69%

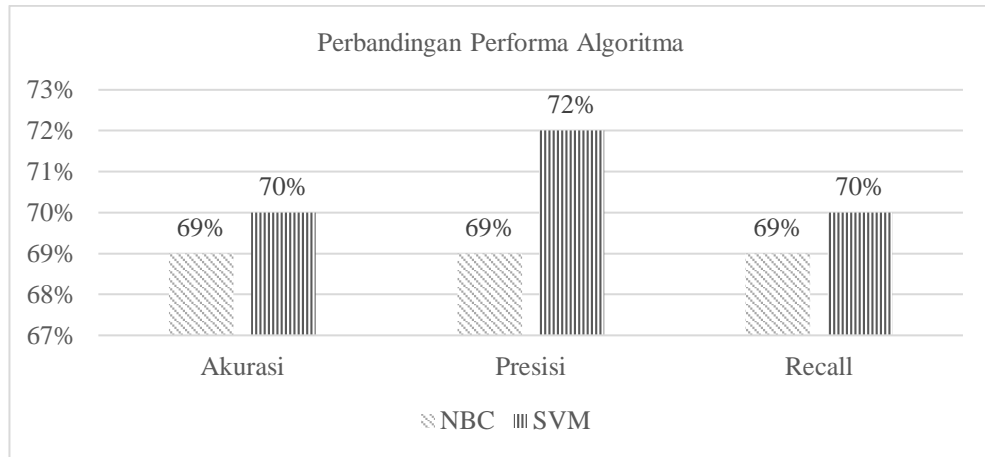


**Gambar 9.** Presentase Recall Algoritma SVM Pada Data Pinjol Ilegal

Gambar 9 memperlihatkan nilai recall setiap kernel, tanpa SMOTE kernel linear memiliki recall sebesar 45%, kernel RBF sebesar 43%, kernel polinomial 45%, dan kernel sigmoid 45%. Dengan SMOTE nilai recall kernel linear menjadi 74%, kernel RBF 74%, kernel polinomial 74%, dan kernel sigmoid 70%.

### 3.6 Perbandingan Performa Algoritma

- a. Perbandingan performa algoritma SVM dan NBC menggunakan teknik SMOTE pada data pinjol legal. Dan didapati bahwa performa algoritma SVM lebih baik dibanding dengan algoritma NBC. Hasil dapat dilihat pada Gambar 10.



Gambar 10. Perbandingan Performa Algoritma pada Data Pinjol Legal

Gambar 10 menunjukkan perbandingan kinerja kedua algoritma, dapat dilihat bahwa algoritma SVM memang lebih baik dalam mengklasifikasi data sentimen pinjol legal terbukti dengan nilai akurasi, presisi dan recall yang lebih tinggi daripada algoritma NBC.

Berikut merupakan Tabel 14 yang berisi confusion matrix dari algoritma terbaik yakni SVM (kernel RBF, C:10,  $\gamma$ :1). Dari pengujian terdapat 252 data yang dievaluasi, dengan 43 data diklasifikasikan sebagai negatif, 62 data sebagai netral dan 66 data sebagai positif.

Tabel 14. Confusion Matrix SVM

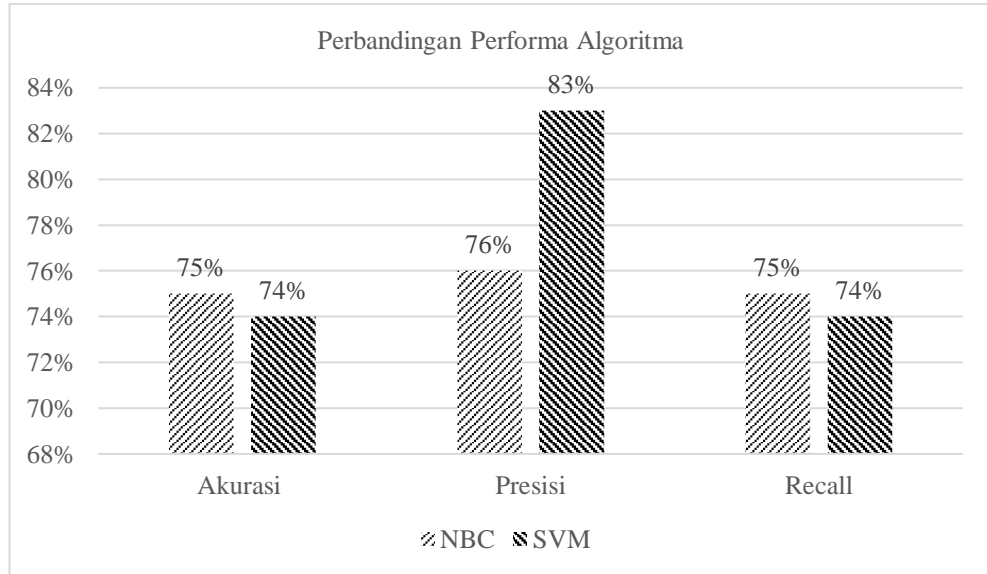
		Prediksi		
		Negatif	Netral	Positif
Aktual	Negatif	43	15	31
	Netral	17	62	7
	Positif	7	4	66

Berikut merupakan visualisasi data komentar pada dataset pinjol legal.



Gambar 11. Visualisasi Data Komentar

- b. Perbandingan performa algoritma SVM dan NBC menggunakan teknik SMOTE pada data pinjol ilegal. Didapat hasil bahwa performa algoritma NBC lebih baik dibanding dengan algoritma SVM. Sebagaimana terlihat di Gambar 12.



**Gambar 12.** Perbandingan Performa Algoritma pada Data Pinjol Ilegal

Gambar 12 menunjukkan perbandingan kinerja kedua algoritma pada sentimen pinjol ilegal, terlihat bahwa algoritma NBC memiliki nilai akurasi dan recall yang lebih baik dari SVM sebesar 75%. Sedangkan untuk nilai presisi algoritma SVM lebih unggul daripada NBC dengan nilai presisi sebesar 83%.

Berikut merupakan Tabel 12 yang berisi confusion matrix dari NBC. Terdapat 270 baris data, dengan 49 data diklasifikasikan sebagai negatif, 82 data sebagai netral dan 71 data sebagai positif.

**Tabel 15.** Confusion Matrix NBC

		Prediksi		
		Negatif	Netral	Positif
Aktual	Negatif	49	19	28
	Netral	0	82	0
	Positif	10	11	71

Berikut merupakan visualisasi data komentar pada dataset pinjol ilegal.





Gambar 13. Visualisasi Data Komentar

#### 4. KESIMPULAN

Penelitian ini berhasil membandingkan performa algoritma SVM dan NBC dalam menganalisis sentimen terhadap pinjaman online di Twitter. Pada dataset pinjol legal, SVM dengan kernel RBF unggul dengan akurasi terbaik mencapai 70%, menunjukkan kemampuannya yang kuat dalam menangani data dengan dimensi tinggi. Sebaliknya, pada dataset pinjol ilegal, NBC menunjukkan kinerja yang lebih baik dengan akurasi mencapai 75%, menegaskan efektivitasnya dalam klasifikasi probabilitas. Teknik SMOTE terbukti sangat membantu dalam meningkatkan kinerja model, khususnya pada dataset yang tidak seimbang. Sentimen yang diperoleh pada dataset pinjol legal adalah sentiment positif. Sentimen yang diperoleh pada dataset pinjol ilegal adalah negatif. Studi ini memberikan kontribusi signifikan dalam memahami sentimen publik terhadap pinjaman online dan dapat digunakan oleh penyedia layanan serta regulator untuk meningkatkan kualitas layanan dan merancang kebijakan yang lebih responsif. Rekomendasi untuk penelitian selanjutnya adalah menggabungkan pendekatan algoritma lain dan mengeksplorasi fitur-fitur tambahan yang dapat meningkatkan akurasi analisis sentimen.

#### REFERENCES

- [1] R. Imanto, M. Panorama, and R. Sumantri, "Pengaruh Pengangguran dan Kemiskinan Terhadap Pertumbuhan Ekonomi di Provinsi Sumatera Selatan," *J. Ekon. Islam*, vol. 1, no. 1, p. 59, 2020, doi: 10.32662/golder.v1i1.115.
- [2] Muhammad Ramli, Mulono Apriyanto, Ali Azhar, Feni Puspitasari, and KMS. Novyar Satriawan Fikri, "Dampak Konsumen Terhadap Pinjaman Online (PINJOL)," *PaKMas J. Pengabd. Kpd. Masy.*, vol. 3, no. 1, pp. 52–58, 2023, doi: 10.54259/pakmas.v3i1.1638.
- [3] T. P. Lestari, "Analisis Text Mining pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) dan Social Network Analysis (SNA)," *J. Inform. Ekon. Bisnis*, vol. 4, no. 3, pp. 65–71, 2022, doi: 10.37034/infeb.v4i3.146.
- [4] J. Florensus Sianipar, Y. R. Ramadhan, and I. Jaelani, "Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes," *Media Online*, vol. 4, no. 1, pp. 360–367, 2023, doi: 10.30865/klik.v4i1.1033.
- [5] E. E. Amelia and I. Yustiana, "Analisis Sentimen Pada Ulasan Produk UNIQLO dengan Algoritma Naive Bayes," *Univ. Nusa Putra, Indones.*, vol. 8, no. 1, pp. 141–148, 2024, doi: http://dx.doi.org/10.30645/j-sakti.v8i1.773.
- [6] A. Alsiaity and R. Orji, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," *Behaviour and Information Technology*, vol. 43, no. 1, pp. 139–164, 2024. doi: 10.1080/0144929X.2022.2156387.
- [7] R. W. Pratiwi, S. F. H. D. Dairoh, D. I. Af'idah, Q. R. A., and A. G. F., "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 1, pp. 40–46, 2021, doi: 10.20895/inista.v4i1.387.
- [8] I. M. A. D. Putra, I. M. G. Sunarya, and I. G. A. Gunadi, "Perbandingan Algoritma Naive Bayes Berbasis Feature Selection Gain Ratio dengan Naive Bayes Kovenasional dalam Prediksi Komplikasi Hipertensi," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 6, no. 1, pp. 37–49, 2024, doi: 10.35746/jtim.v6i1.488.
- [9] R. N. Ikhsani and F. F. Abdulloh, "Optimasi SVM dan Decision Tree Menggunakan SMOTE Untuk Mengklasifikasi Sentimen Masyarakat Mengenai Pinjaman Online," *J. Media Inform. Budidarma*, vol. 7, no. 4, pp. 1667–1677, 2023, doi: 10.30865/mib.v7i4.6809.
- [10] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindehan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [11] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naive Bayes," *J. Sains dan Inform.*, vol. 2, no. 2, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.
- [12] A. A. A. Arifin, W. Handoko, and Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan," *J-Com (Journal Comput.)*, vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.
- [13] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan TF-IDF Menggunakan Naive Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," *J. Manaj. Inform.*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.
- [14] M. Das, S. Kamalanathan, and P. Alphonse, "A Comparative Study on TF-IDF feature weighting method and its analysis using unstructured dataset," *CEUR Workshop Proc.*, vol. 2870, pp. 98–107, 2021.
- [15] G. H. Kusuma, I. Permana, F. N. Salisah, M. Afdal, M. Jazman, and A. Marsal, "Pendekatan Machine Learning: Analisis



- Sentimen Masyarakat Terhadap Kendaraan Listrik Pada Sosial Media X,” *JUSIFO (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 65–76, 2023, doi: 10.19109/jusifo.v9i2.21354.
- [16] C. M. Putri, M. Afdal, R. Novita, and M. Mustakim, “Perbandingan Evaluasi Kernel Support Vector Machine dalam Analisis Sentimen Chatbot AI pada Ulasan Google Play Store,” *J. Teknol. Sist. Inf. dan Apl.*, vol. 7, no. 3, pp. 1236–1245, 2024, doi: 10.32493/jtsi.v7i3.41354.
- [17] W. Nugraha and A. Sasongko, “Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search Hyperparameter Tuning on Classification Algorithm with Grid Search,” *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 391–401, 2022, doi: <http://sistemasi.ftik.unisi.ac.id>.
- [18] S. Maula Chamzah, M. Lestandy, N. Kasan, and A. Nugraha, “Penerapan Synthetic Minority Oversampling Technique (SMOTE) untuk Imbalance Class pada Data Text Menggunakan kNN,” *Syntax J. Inform.*, vol. 11, no. 02, pp. 56–67, 2022, doi: 10.35706/syji.v11i02.6940.
- [19] S. Ramadhani, D. Azzahra, and T. Z., “Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student’s Thesis,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 24–33, 2022, doi: 10.31849/digitalzone.v13i1.9292.
- [20] A. Mustofa and R. Novita, “Klasifikasi Sentimen Masyarakat Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Text Mining Pada Twitter,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 200–208, 2022, doi: 10.47065/bits.v4i1.1628.