

Perbandingan Metode Naïve Bayes dan K-Nearest Neighbor Terhadap Sentimen Analisis Pinjaman Online

Thomas Boris Asalodan Tokan^{1*}, Donatus Joseph Manehat², Yovinia Carmeneja Hoar Siki³, Emerensiana Ngaga⁴, Sisilia Daeng Bakka Mau⁵

^{1,2,3}Ilmu Komputer Universitas Katolik Widya Mandira

Email: ¹tokan.boris@gmail.com, ²donatusjmanehat@unwira.ac.id, ^{3*}yoviniacarmeneja@unwira.ac.id,

⁴emerensianangaga@unwira.ac.id, ⁵sisiliamaui@unwira.ac.id

Email Penulis Korespondensi: yoviniacarmeneja@unwira.ac.id

Abstrak– Penelitian ini bertujuan untuk mengetahui opini masyarakat indonesia terhadap keberadaan pinjaman online. Pinjaman online merupakan salah layanan perbankan melalui teknologi yang dikenal dengan Industri *Financial Technology (FinTech)*. Menurut laporan Otoritas Jasa Keuangan, pada bulan Agustus 2023 lebih dari 13,37 juta akun menggunakan jasa ini. Layanan pinjaman online dianggap memberikan kemudahan dan kenyamanan bagi konsumen. Namun muncul banyak kasus merugikan, seperti naiknya suku bunga yang sangat tinggi dan praktik penagihan yang agresif sehingga meresahkan masyarakat. Untuk itu dilakukan analisis sentiment untuk mengetahui opini masyarakat yang dapat dijadikan referensi bagi penyedia layanan pinjol maupun pelaku pinjol. Dengan menggunakan data *Tweet* sejumlah 11.288 data berbahasa Indonesia opini masyarakat akan pinjol diketahui. Penelitian menggunakan dua metode analisis sentimen, yaitu *Naïve Bayes* dan *K-Nearest Neighbor*. Hasil dari penelitian menunjukkan sentimen terhadap pinjaman online yang bernilai negatif 62.29%, netral 33.19% dan positif 5.52%. Hasil dari penelitian ini juga menunjukkan metode *Naive Bayes* memiliki akurasi yang sedikit lebih tinggi (67%) dibandingkan dengan metode *K-Nearest Neighbor* (63%). Diharapkan sentiment ini memberikan dampak positif bagi penyedia layanan pinjol maupun pelaku pinjol.

Kata Kunci: Sentimen Analisis; Klasifikasi; *Naïve Bayes*; *K – Nearest Neighbors*; *Tweet*; Perbandingan

Abstract– This study aims to understand the public opinion in Indonesia regarding the existence of online loans. Online loans are a type of banking service through technology known as the *Financial Technology (FinTech)* Industry. According to a report by the Financial Services Authority, in August 2023, more than 13.37 million accounts were using this service. Online loan services are considered to provide convenience and comfort for consumers. However, many harmful cases have emerged, such as extremely high interest rates and aggressive debt collection practices that have caused public concern. Therefore, a sentiment analysis is conducted to understand public opinions, which can serve as a reference for online loan service providers and operators. By using 11,288 Indonesian-language tweets, the public opinion on online loans is analyzed. The study employs two sentiment analysis methods: *Naïve Bayes* and *K-Nearest Neighbor*. The results of the study show that the sentiment toward online loans is 62.29% negative, 33.19% neutral, and 5.52% positive. The results also indicate that the *Naïve Bayes* method has slightly higher accuracy (67%) compared to the *K-Nearest Neighbor* method (63%). It is hoped that this sentiment will have a positive impact on online loan service providers and operators.

Keywords: *sentiment Analysis; Classification; Naïve Bayes; K - Nearest Neighbors; Tweet; Comparison*

1. PENDAHULUAN

Perkembangan teknologi yang sangat pesat saat ini mendominasi gaya hidup masyarakat yang mempermudah masyarakat dalam mengakses berbagai hal melalui *smartphone*. *Financial Technology / FinTech* merupakan salah satu bentuk perkembangan teknologi menggunakan *smartphone*. Bank Indonesia, menyatakan bahwa *FinTech* merupakan perpaduan antara layanan keuangan dan teknologi yang mengtransformasi model bisnis dari tradisional menjadi lebih moderen [1]. Otoritas Jasa Keuangan (OJK) melaporkan sampai dengan Maret 2023, terdapat 102 perusahaan *FinTech Lending* yang telah memperoleh izin [2]. Salah satu jenis penyelenggaraan *FinTech* adalah layanan pinjaman online (pinjol). Pinjol adalah layanan keuangan yang menyediakan *platform* untuk memfasilitasi perjanjian pinjaman langsung antara pemberi pinjaman dan penerima pinjaman melalui sistem elektronik [3]. Regulasi hukum yang mengatur penerapan *FinTech* pada skema pembayaran di Indonesia terdiri dari [4]:

1. Peraturan Bank Indonesia No. 18/40/PBI/2016 mengatur administrasi pencatatan pembayaran;
2. Surat edaran Bank Indonesia No. 18/22/DKSP tentang administrasi layanan finansial digital;
3. Peraturan Bank Indonesia No. 18/17/PBI/2016 tentang uang elektronik.
4. OJK mengatur pinjaman online melalui Peraturan 2 No. 77/POJK.01/2016 yang membahas layanan peminjaman berbasis teknologi.

OJK melaporkan nilai penyaluran pinjol di Indonesia telah mencapai angka Rp.20,54 triliun pada bulan Agustus 2023 [5]. Layanan pinjol menyediakan akses cepat dan mudah ke peminjam. Layanan ini memiliki daya tarik tersendiri bagi individu dan usaha kecil yang membutuhkan dana tambahan dalam waktu singkat. Faktor – faktor seperti proses pengajuan yang minimalis, persyaratan yang mudah dipenuhi, dan kenyamanan dalam mengajukan pinjaman melalui aplikasi seluler telah membuat pinjol menjadi alternatif yang populer untuk memenuhi kebutuhan keuangan yang mendesak [6].

Meskipun layanan pinjol memberikan kemudahan, namun banyak kasus merugikan muncul seperti suku bunga tinggi, praktik penagihan agresif, dan pelecehan terhadap peminjam. Kesulitan keuangan menyebabkan beberapa

peminjam terjebak dalam hutang yang berkepanjangan. Menurut laporan OJK, total utang pinjol warga Indonesia mencapai Rp51,46 triliun pada Mei 2023 dengan sebagian peminjam mengalami kemacetan hingga Rp1,72 triliun pada bulan yang sama [7]. Kasus – kasus buruk ini disebabkan oleh transaksi dengan Perusahaan pinjol ilegal. OJK mencatat bahwa hingga 8 Juli 2023, ada 429 perusahaan pinjol ilegal terdaftar. Berbagai persepsi masyarakat muncul, ada yang setuju dengan pinjol dan ada yang tidak. Analisis sentiment diperlukan untuk mengetahui kasus pinjol ilegal yang menimbulkan kekhawatiran di masyarakat. Analisis sentiment biasanya merujuk pada perasaan, emosi, sikap ataupun pendapat. Platform media social seperti X (sebelumnya Twitter) sering digunakan oleh masyarakat Indonesia sebagai tempat membagikan pendapat dan pandangan mereka tentang pinjol [8].

Pada penelitian ini data yang digunakan sebanyak 11.288 data diambil dari Tweet aplikasi X. Data akan dikategorikan ke 3 kelas sentimen yaitu positif, negatif, dan netral menggunakan metode Naïve Bayes. Sebagai pembandingan digunakan metode K-Nearest Neighbor dengan melihat hasil akurasi dari kedua metode. Penggunaan kedua metode tersebut untuk mengetahui seberapa besar sentimen masyarakat terhadap pinjol. Beberapa penelitian tentang pinjaman online menggunakan Algoritma Naïve Bayes menggunakan 650 data menghasilkan akurasi sebesar 75% [9]. Penelitian lain yang membandingkan metode Naïve Bayes dan K-Nearest Neighbor terhadap sentimen pengguna E-Wallet menggunakan 1292 data tweet menghasilkan algoritma K – Nearest Neighbor mendapat hasil *accuracy* tertinggi yaitu 89.44%[10].

Penelitian ini bertujuan untuk mengetahui opini masyarakat Indonesia terhadap keberadaan pinjaman online. Pinjaman online merupakan salah layanan perbankan melalui teknologi yang dikenal dengan Industri *Financial Technology (FinTech)*. Menurut laporan Otoritas Jasa Keuangan, pada bulan Agustus 2023 lebih dari 13,37 juta akun menggunakan jasa ini. Layanan pinjaman online dianggap memberikan kemudahan dan kenyamanan bagi konsumen. Namun muncul banyak kasus merugikan, seperti naiknya suku bunga yang sangat tinggi dan praktik penagihan yang agresif sehingga meresahkan masyarakat. Untuk itu dilakukan analisis sentiment untuk mengetahui opini masyarakat yang dapat dijadikan referensi bagi penyedia layanan pinjol maupun pelaku pinjol.

2. METODOLOGI PENELITIAN

Penelitian dilakukan melalui beberapa tahap mulai dari tahap pengumpulan data, pelabelan data, *preprocessing*, klasifikasi, hingga perhitungan akurasi. Gambar 1 berikut menunjukkan tahapan penelitian.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data dikumpulkan melalui proses *crawling* tweet aplikasi X menggunakan tools “*tweet-harvest@latest*”. Proses *crawling* data dilakukan secara otomatis menggunakan “*auth_token*” yang didapat pada halaman web X pada bagian “inspeksi”. Data *tweet* berbahasa Indonesia, dikumpulkan sejak tanggal 1 September 2023 sampai dengan 30 September 2023 menggunakan kata kunci “pinjaman online dan pinjol”. Total data dikumpulkan sebanyak 11.288 dataset.

2.2. Pelabelan Data

Pelabelan data merupakan proses pemberian label atau keterangan tertentu untuk setiap data tweet. Tujuannya adalah untuk memberi label sentiment kepada setiap teks. Pelabelan data dilakukan secara manual dengan membaca setiap teks dan memberi label sentimen yang sesuai.

2.3. Cleaning Awal

Cleaning awal adalah proses teks mentah dibersihkan dan dipersiapkan sebelum dilakukan analisis lebih lanjut. Langkah-langkah dalam *cleaning* awal mencakup menghapus karakter khusus, yang tidak relevan seperti *URL*, *tag HTML*, dan *emoticon*.

2.4. Preprocessing

Preprocessing bertujuan membersihkan data yang telah dikumpulkan sebelum dianalisis. Beberapa tahap yang dilakukan pada tahap *preprocessing* adalah sebagai berikut.

- Case Folding* : merubah semua huruf menjadi huruf kecil untuk menjaga konsistensi dalam analisis. Tujuannya adalah menghapus perbedaan antara huruf kapital dan huruf kecil, sehingga mempermudah analisis dan pencarian teks.
- Remove Punctuation* : menghapus semua tanda baca dari teks karena tanda baca sering kali tidak memberikan informasi.
- Stopword Removing*: mengurangi kata-kata umum seperti “dan”, “atau”, dan “yang” yang sering muncul pada teks.
- Tokenizing* : memisahkan teks menjadi item-item berupa kata atau bagian lain dari teks. Ini adalah langkah penting karena memungkinkan kita untuk menganalisis teks secara lebih terperinci.
- Stemming* : mengubah kata menjadi bentuk dasar atau akar kata. Tujuannya adalah untuk mengurangi perbedaan variasi kata yang memiliki akar yang sama, sehingga analisis teks menjadi lebih mudah.

2.5. Visualisasi

Visualisasi data merupakan langkah krusial dalam analisis sentimen yang membantu memahami pola atau *trend* dalam data teks. Teknik visualisasi yang digunakan dalam penelitian adalah diagram lingkaran dan *word clouds*.

- Diagram Lingkaran : berguna untuk menunjukkan proporsi relatif dari beberapa kategori atau bagian dari keseluruhan data.
- Word cloud* : merupakan representasi visual dari kata-kata yang paling sering muncul dari data teks. Kata-kata yang sering muncul akan ditampilkan dengan ukuran yang lebih besar dalam *word cloud*.

2.6. Klasifikasi

Pada tahap ini, data yang telah melalui proses *preprocessing* dianalisis dan diklasifikasi ke dalam 3 kategori yaitu sentiment positif, negatif dan netral. Klasifikasi dilakukan melalui tahap – tahap berikut.

a. Splitting Data

Pada *splitting* data, dataset dibagi menjadi 2 *subset* berbeda menjadi set pelatihan dan set pengujian. Set pelatihan berfungsi untuk melatih model klasifikasi, sedangkan set pengujian berfungsi menguji kinerja model yang dilatih. Penelitian ini menggunakan 80 % data latih dan 20 % data uji. Proporsi 80 : 20 umum digunakan dan sering dianggap kompromi yang baik antara penggunaan data untuk *training* dan memastikan evaluasi yang dapat dipercaya terhadap model yang dikembangkan [11]. Data latih 80% dapat mempelajari pola yang kompleks dari data yang tersediasehingga memungkinkan model untuk menjadi lebih akurat dalam memprediksi data baru.

b. Pembobotan Data

Merupakan proses pemberian nilai numerik kepada kata - kata dalam teks berdasarkan frekuensi kemunculan kata dalam teks. Teknik Pembobotan kata menggunakan teknik *TF-IDF*.

c. Naïve Bayes

Dengan menggunakan *Teorema Bayes*, algoritma ini menghitung probabilitas kelas tertentu berdasarkan fitur-fitur yang dalam data. Algoritma *Naïve Bayes* mengasumsikan bahwa setiap variabel adalah independen dan tidak memiliki keterkaitan dengan yang lainnya [12]. Dengan demikian, sebuah teks dianggap sebagai sekumpulan kata yang membentuknya. proses ini menghasilkan kategori sentiment teks menjadi kelas sentimen positif, negatif, atau netral sesuai dengan fitur - fitur atau kata - kata yang terkandung dalam teks tersebut. Berikut diberikan beberapa persamaan *Naïve Bayes Classifier* yang digunakan.

$$P(C_i) = \frac{fD(C_i)}{|D|} \quad (1)$$

$P(C_i)$ = Probabilitas dari suatu kategori dokumen; $fD(C_i)$ = Frekuensi dokumen yang memiliki kategori C_i ; $|D|$ = Jumlah seluruh dokumen latih.

$$P(W_{kj} | C_i) = \frac{f(C_i) + 1}{f(C_i) + |W|} \quad (2)$$

$P(W_{kj} | C_i)$ = Probabilitas kemunculan kata W_{kj} pada suatu dokumen dalam kategori kelas C_i ; W_{kj} = frekuensi kata ke- k pada setiap kategori; W = Jumlah kata pada dokumen *test*; $f(C_i)$ = frekuensi dokumen berkategori kelas C_i .

d. K-Nearest Neighbors

Algoritma *K-NearestNeighbors* (K-NN) mengklasifikasi secara sederhana dan intuitif. Metode klasifikasi ini bekerja dengan cara mencari nilai k data *latih* terdekat. Untuk mengukur tingkat kemiripan digunakan *Cosine Similarity*. Berikut diberikan persamaan *Cosine Similarity* yang bekerja pada K-NN [13].

$$\cos Sim(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{(\sum_{i=1}^n A_i)^2} \sqrt{(\sum_{i=1}^n B_i)^2}} \quad (3)$$

A = Data uji; B = Data latih; A_i = Bobot *term* yang terdapat pada data uji; B_i = Bobot *term* yang terdapat pada data latih.

2.7. Hasil

Hasil merupakan tahap akhir dalam penelitian ini. Pada tahap ini dilakukan perhitungan akurasi dan perbandingan akurasi untuk mengevaluasi akurasi dan membandingkan performa algoritma *Naïve Bayes* dan *K-Nearest Neighbors* yang diterapkan pada 11.288 data.

Akurasi masing – masing metode diperoleh melalui perhitungan *accuracy*, *precision*, *recall* dan *f1-score* yang dihitung menggunakan rumus :

$$Accuracy = \frac{TP+TN+TN}{TP+TN+TN+FP+FNT+FN} * \quad (4)$$

$$100\% Precision = \frac{TP}{FP+TP} * 100\% \quad (5)$$

$$Recall = \frac{TP}{FN+FNT+TP} * 100\% \quad (6)$$

$$F1-Score = \frac{2 * Recall * Precision}{(Recall + Precision)} \quad (7)$$

True Positive (TP); True Netral (TNt); False Positive (FP); False Negative (FN); FalseNetral (FNT); TrueNegative (TN). Dari hasil *accuracy*, *precision*, *recall* dan *f1-score* diperoleh perbandingan kinerja dua model klasifikasi *NaïveBayes* dan *K-Nearest Neighbors* dalam memprediksi sentimen analisis terhadap pinjaman *online*.

3. HASIL DAN PEMBAHASAN

3.1 Crawling Data

Dengan menggunakan tools “*tweet-harvest@latest*” crawling data dilakukan pada *Google Colab* menggunakan bahasa pemrograman Python. *Tweet* yang di kumpulan menggunakan dua kata kunci yaitu “pinjaman online dan pinjol”. Data dikumpulkan sejak tanggal 1 September 2023 hingga 30 September 2023. Total data yang berhasil dikumpulkan sebanyak 11.288 dataset berbahasa Indonesia. Tahap *crawling* dapat dilihat pada gambar 2.



Gambar 2. Crawling data menggunakan Google Colab

3.2 Pelabelan Data

Pelabelan terdiri dari 3 label yaitu label 1 untuk sentiment positif, 2 untuk sentimen netral, dan 3 untuk sentimen negatif. Proses pelabelan dilakukan secara manual dengan membaca data *tweet* satu per satu. Tahap pelebelan dapat dilihat pada gambar 3.



Gambar 3. Pelabelan data dalam excel

3.3 Cleaning Awal

Proses pembersihan data (*cleaning*) dilakukan untuk memastikan data bebas dari noise dan dapat diinterpretasikan dengan benar. Pada tahap ini, dilakukan penghapusan elemen yang tidak relevan seperti nama pengguna *Twitter*, *hashtag*, *retweet*, tautan, angka di awal teks, baris duplikat, serta kolom kosong. Hasil *cleaning* dapat dilihat pada gambar 4.

	Full_text	Isi_text
0	revisi orang gila bisa semua with bangkai... 0.0	
1	... 0.0	
2	... 0.0	
3	... 0.0	
4	... 0.0	
...		
11284	... 0.0	
11285	... 0.0	
11286	... 0.0	
11287	... 0.0	
11288	... 0.0	

Gambar 4. Hasil *Cleaning Awal*

3.4 Preprocessing

a. Case Folding

Pada tahap ini semua teks diubah menjadi huruf kecil. Hasil *case folding* ditunjukkan pada gambar 5.

	Full_text	Case_Folding	Isi_text
0	revisi orang gila bisa semua with bangkai... 0.0	revisi orang gila bisa semua with bangkai... 0.0	
1	... 0.0	... 0.0	
2	... 0.0	... 0.0	
3	... 0.0	... 0.0	
4	... 0.0	... 0.0	
...			
11284	... 0.0	... 0.0	
11285	... 0.0	... 0.0	
11286	... 0.0	... 0.0	
11287	... 0.0	... 0.0	
11288	... 0.0	... 0.0	

Gambar 5. Hasil *Case Folding*

b. Remove Punctuation

Tahap ini semua teks dibersihkan dari semua karakter tanda baca seperti titik, koma, tanda seru, tanda tanya, dan simbol lainnya. Hasil *remove punctuation* ditunjukkan pada gambar 6.

	Full_text	Remove_punct	Isi_text
0	revisi orang gila bisa semua with bangkai... 0.0	revisi orang gila bisa semua with bangkai... 0.0	
1	... 0.0	... 0.0	
2	... 0.0	... 0.0	
3	... 0.0	... 0.0	
4	... 0.0	... 0.0	
...			
11284	... 0.0	... 0.0	
11285	... 0.0	... 0.0	
11286	... 0.0	... 0.0	
11287	... 0.0	... 0.0	
11288	... 0.0	... 0.0	

Gambar 6. Hasil *Remove Punctuation*

c. Stopword Removing

Pada tahap ini dilakukan penghapusan kata – kata umum yang sering muncul yaitu “yang”, “dan”, “atau”. Hasil *stopword removing* dapat dilihat pada gambar 7.

id	review_positif	review_negatif	sentimen
0	bet orang gila sama semua waktu buat...	bet orang gila sama semua waktu buat...	0.0
1	gak menter yg baru udah mulai semua...	gak menter yang baru udah mulai semua...	0.0
2	semoga aish azzah via jala menandakan...	semoga aish azzah via jala menandakan...	0.0
3	darurat yang harus diwujudkan peng...	darurat yang harus diwujudkan peng...	0.0
4	rekrutasi buat peng...	rekrutasi buat peng...	0.0
11204	di indonesia halangan itu menjadi kata...	di indonesia halangan itu menjadi kata...	0.0
11205	leth paling benar ecomen gausah ada...	leth paling benar ecomen gausah ada...	0.0
11206	kekasabih quaternya dari animal commu...	kekasabih quaternya dari animal commu...	0.0
11207	besar untuk agunan waji atas nama caka...	besar untuk agunan waji atas nama caka...	0.0
11208	juatin pernah tua gaj gende kaka terus...	juatin pernah tua gaj gende kaka terus...	0.0

Gambar 7 Hasil Stopword Removing

d. Tokenzing

Pada tahap *tokenzing* teks hasil *stopword removing* kemudian dipisah menjadi item-item berupa kata-kata untuk dianalisis secara lebih terperinci. Hasil *tokenzing* pada gambar 8.

id	tokenzing_positif	tokenzing_negatif	sentimen
0	bet orang gila sama semua waktu buat...	bet orang gila sama semua waktu buat...	0.0
1	gak menter baru mulai semua...	gak menter baru mulai semua...	0.0
2	semoga aish azzah via jala menandakan...	semoga aish azzah via jala menandakan...	0.0
3	darurat harus diwujudkan peng...	darurat harus diwujudkan peng...	0.0
4	rekrutasi buat peng...	rekrutasi buat peng...	0.0
11204	indonesia halangan itu menjadi kata...	indonesia halangan itu menjadi kata...	0.0
11205	leth paling benar ecomen gausah ada...	leth paling benar ecomen gausah ada...	0.0
11206	kekasabih quaternya dari animal commu...	kekasabih quaternya dari animal commu...	0.0
11207	besar untuk agunan waji atas nama caka...	besar untuk agunan waji atas nama caka...	0.0
11208	juatin pernah tua gaj gende kaka terus...	juatin pernah tua gaj gende kaka terus...	0.0

Gambar 8 Hasil Tokenzing

e. Stemming

Stemming merupakan proses menghapus imbuhan dari kata untuk mendapatkan bentuk kata dasar yang disebut "stem". Hasil Stemming pada gambar 9.

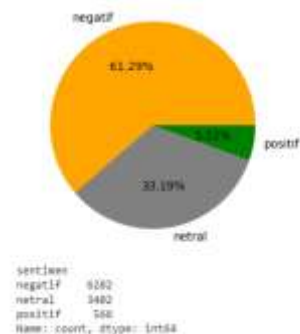
id	tokenzing	stemming	sentimen
0	[bet, 'orang', 'gila', 'sama', 'waktu', 'buat', 'nonton', '...]	bet orang gila sama waktu buat nonton...	0.0
1	[gak, 'menter', 'baru', 'mulai', 'semua', 'nangsa', 'peng', '...]	gak menter baru mulai semua nangsa peng...	0.0
2	[semoga, 'aish', 'azzah', 'via', 'jala', 'men', '...]	mega aish azzah via jala men...	0.0
3	[darurat, 'harus', 'diwujudkan', 'peng', '...]	darurat harus diwujudkan peng...	0.0
4	[rekrutasi, 'buat', 'peng', '...]	rekrutasi buat peng...	0.0
11204	[indonesia, 'halangan', 'itu', 'menjadi', 'kata', 'negatif', '...]	indonesia halangan itu kata negatif...	0.0
11205	[leth, 'paling', 'benar', 'ecomen', 'gausah', 'ada', 'peng', '...]	leth paling benar ecomen gausah ada peng...	0.0
11206	[kekasabih, 'quaternya', 'dari', 'animal', 'commu', '...]	kekasabih quaternya dari animal commu...	0.0
11207	[besar, 'untuk', 'agunan', 'waji', 'atas', 'nama', 'caka', '...]	besar agun waji atas nama caka...	0.0
11208	[juatin, 'pernah', 'tua', 'gaj', 'gende', 'kaka', 'terus', '...]	juatin pernah tua gaj gende kaka terus...	0.0

Gambar 9. Hasil Stemming

3.5 Visualisasi Data

a. Diagram Lingkaran

Diagram lingkaran berguna untuk menunjukkan proporsi relatif dari beberapa kategori atau bagian dari keseluruhan data. Hasil diagram lingkaran dilihat pada gambar 10.



Gambar 10. Hasil visualisasi diagram lingkaran

Hasil visualisasi dari diagram lingkaran pada gambar 10 menunjukkan dataset yang bernilai negatif 62.29% atau sebanyak 6282 dataset, netral 33.19% atau sebanyak 3402 dataset dan positif 5.52% atau sebanyak 566 dataset.

b. Word Clouds

Word clouds merepresentasi visual dari kata - kata yang paling sering muncul. Kata-kata ini akan ditampilkan dengan ukuran yang lebih besar dalam *word clouds*. Hasil *Word Clouds* menunjukkan kata-kata kunci yang paling banyak muncul. Hasil *word clouds* dapat dilihat pada gambar 11.a.,11.b dan 11.c.

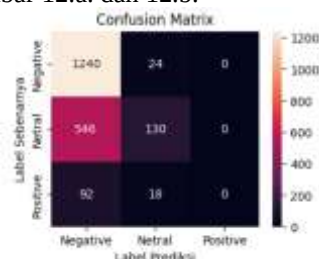
Gambar 11.a. Hasil *Word Clouds* PositifGambar 11.b. Hasil *Word Clouds* NetralGambar 11.c. Hasil *Word Clouds* Negatif

Hasil dari *Word Clouds positif* menunjukkan kata yang paling banyak muncul adalah kata “pinjam, uang, mudah, pinjol”. Hasil dari *Word Clouds Netral* menunjukkan kata yang paling banyak muncul adalah kata “pinjol, pinjam, mulai, buat”. Sedangkan hasil dari *Word Clouds Negatif* menunjukkan kata yang paling banyak muncul adalah kata “pinjam, harus, pinjol, slot”.

3.6 Hasil Klasifikasi

Data yang telah dibersihkan kemudian diberi bobot yakni data training sebanyak 80% dari total data dan data testing sebanyak 20%. Pembangunan model dilakukan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbors* dengan memanfaatkan data *training*. Kemudian dilakukan *testing* menggunakan data *testing* sebagai data pengujian model dalam melakukan analisis sentimen.

Hasil klasifikasi metode *naïve bayes* mendapat nilai *accuracy* sebesar 0.6682 atau 66.82% yang Dapat dilihat pada gambar 12.a. dan 12.b.

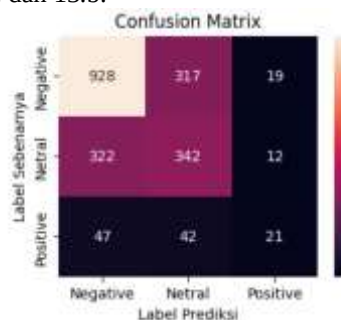
Gambar 12.a. Hasil analisis *Naïve Bayes*

	precision	recall	f1-score	support
-1	0.66	0.98	0.79	1264
0	0.70	0.19	0.31	676
1	0.00	0.00	0.00	118
accuracy	0.67			2058
macro avg	0.47	0.39	0.37	2058
weighted avg	0.66	0.67	0.50	2058

Hasil Akurasi Metode Naive Bayes : 0.6682926829268293

Gambar 12.b. Hasil evaluasi kinerja metode *Naïve Bayes*

Hasil klasifikasi metode *K-Nearest Neighbors* mendapat nilai *accuracy* sebesar 0.6297 atau 62,97% ditunjukkan pada gambar 13.a. dan 13.b.

Gambar 13.a. Hasil analisis *K-Nearest Neighbors*

	precision	recall	f1-score	support
-1	0.72	0.73	0.72	1264
0	0.49	0.51	0.50	676
1	0.40	0.19	0.26	118
accuracy	0.63			2058
macro avg	0.54	0.48	0.49	2058
weighted avg	0.62	0.63	0.62	2058

akurasi metode KNN : 0.6297560975609756

Gambar 13.b. Hasil evaluasi kinerja metode *K-Nearest Neighbors*

3.7 Perbandingan

Merujuk pada hasil klasifikasi menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* dalam analisis sentimen terhadap pinjaman online diperoleh perbandingan kinerja yang disajikan pada tabel 1 berikut.

Table 1. Perbandingan metode *Naïve Bayes* dan *K-Nearest Neighbor*

<i>Naïve Bayes</i>	<i>K-Nearest Neighbor</i>
Accuracy = 67%	Accuracy = 63%
Precision Positif = 66%	Precision Positif = 72%
Precision Netral = 76%	Precision Netral = 49 %
Precision Negatif= 0%	Precision Negatif=40 %
Recall Positif = 98%	Recall Positif = 73%
Recall Netral = 19 %	Recall Netral = 51%
Recall Negatif = 0%	Recall Negatif = 19%
F1-Score Positif = 79%	F1-Score Positif = 72%
F1-Score Netral = 31%	F1-Score Netral = 50%
F1-Score Negatif = 0%	F1-Score Negatif = 26%

Berdasarkan perbandingan, model *Naive Bayes* memiliki akurasi sedikit lebih tinggi (67%) dibandingkan *K-Nearest Neighbor* (63%). Untuk kelas -1, recall *Naive Bayes* unggul (98% vs. 73%), sedangkan untuk kelas 0, *K-Nearest Neighbor* lebih baik (51% vs. 19%). Dari segi precision, *Naive Bayes* lebih akurat pada kelas 0 (76% vs. 49%), sementara *K-Nearest Neighbor* lebih baik pada kelas 1 (40% vs. 0%). Ini menunjukkan bahwa kinerja kedua model bervariasi tergantung kelas, sehingga pemilihan model bergantung pada tujuan dan konteks spesifik.

4. KESIMPULAN

Berdasarkan hasil klasifikasi dan visualisasi data didapatkan persentase masing-masing kategori dari keseluruhan data menunjukkan sentimen terhadap pinjaman *online* yang bernilai negatif 62.29% atau sebanyak 6282 dataset, netral 33.19% atau sebanyak 3402 dataset dan positif 5.52% atau sebanyak 566 dataset. Hasil yang didapat juga menunjukkan kata *positif* yang paling banyak muncul adalah kata ‘pinjam, uang, mudah, pinjol’, yang diartikan sebagai pinjaman *online* sangat mudah dalam melakukan pinjaman uang kepada masyarakat. Sedangkan hasil yang menunjukkan kata *Negatif* yang paling banyak muncul adalah kata ‘pinjam, harus, pinjol, slot’. Yang diartikan bahwa pinjaman *online* harus ditangani karena hanya digunakan untuk melakukan judi slot.

Untuk akurasi hasil perbandingan didapatkan metode *Naive Bayes* mendapatkan akurasi lebih tinggi (67%) dari pada metode *K-Nearest Neighbor* (63%). *Naive Bayes* menunjukkan *recall* lebih tinggi pada kategori negatif (98% vs. 73%), tetapi metode *K-Nearest Neighbor* menunjukkan *recall* lebih tinggi untuk kelas netral (51% vs. 19%) metode *K-Nearest Neighbor* juga menunjukkan precision lebih tinggi untuk kelas positif (40% vs. 0%). Diharapkan sentiment ini memberikan dampak positif bagi penyedia layanan pinjol maupun pelaku pinjol.

REFERENCES

- [1] Maulidya, G. P, Afifah, N. 2021. Perbankan Dalam Era Baru Digital : Menuju Bank 4.0. Prosiding Bisnis Seri V. Pascasarjana FE Universitas Tanjungpura, Jl. Imam Bonjol.
- [2] Otoritas Jasa Keuangan. 2023. Penyelenggara Fintech Lending Berijin di OJK per 9 Oktober 2023. <https://ojk.go.id/id/kanal/iknb/financial-technology/Pages/Penyelenggara-Fintech-Lending-Berizin-di-OJK-per-9-Oktober-2023.aspx>
- [3] Yovie Bramantyo Adji. Y. B, Muhamad, W. A, Akrabi, A. N. L, Noerlina. 2023. Perkembangan Inovasi Fintech di Indonesia. Jurnal BECCOS : Business Economic, Communication, and Social Sciences, Vol. 5 No. 1, January 2023, pp : 47 – 58. DOI : 10.21512/becossjournal.v5i1.8675
- [4] Otoritas Jasa Keuangan. 2016. POJK Nomor 77/POJK.01/2016 tentang Layanan Pinjam Meminjam Uang Berbasis Teknologi Informasi
- [5] Otoritas Jasa Keuangan. 2023. Siaran Pers Sektor Jasa Keuangan Tetap Terjaga Stabil Ditopang Permodalan Yang Kuat Dan Likuiditas Memadai.
- [6] Latifa, F. A. 2023. Analisis Penggunaan *Financial Technology* pada UMKM di Yogyakarta Pasca Covid -19. Academic Collection, Universitas Islam Indonesia. <https://dspace.uui.ac.id/handle/123456789/45984>
- [7] Otoritas Jasa Keuangan. 2023. Statistik Lending Periode Juli 2023.
- [8] S. Surohman, S. Aji, R. Rousyati, and F. F. Wati, “Analisa Sentimen Terhadap Review Fintech Dengan Metode *Naive Bayes Classifier* Dan *K- Nearest Neighbor*,” *Evolusi*, vol. 8, no. 1, Mar. 2020, doi: 10.31294/evolusi.v8i1.7535.
- [9] T. Ramadhan, D. Wahiddin, and E. Awal, “Klasifikasi Sentimen Terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma *Naive Bayes*,” *Scientific Student Journal for Information, Technology and Science*, vol. 4, no. 1, Art. no. 1, Jan. 2023.
- [10] F. M. D. Maharani, A. L. Hananto, S. S. Hilabi, F. N. Apriani, A. Hananto, and B. Huda, “Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma *Naive Bayes* dan *K-Nearest Neighbor*,” *METIK JURNAL*, vol. 6, no. 2, Art. no. 2, Dec. 2022, doi: 10.47002/metik.v6i2.372.
- [11] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, “Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode *Naive Bayes*,” *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 3, pp. 279–285, Dec. 2019, doi: 10.26418/jp.v5i3.34368.



- [12] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "TEXT MINING DALAM ANALISIS SENTIMEN ASURANSI MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER," *Prosiding Seminar Sains Nasional dan Teknologi*, vol. 1, no. 1, Art. no. 1, Sep. 2016, doi: 10.36499/psnst.v1i1.1506.
- [13] L. R. Dharmawan, I. Arwani, and D. E. Ratnawati, "Analisis Sentimen pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 3, Art. no. 3, Jun. 2020
- [14] T. A. Aliah, "Penerapan Metode Naive Bayes dalam Analisis Sentimen pada Twitter (Studi Kasus: Hasil Debat Calon Presiden 2019)," *SISTEM INFORMASI, FAKULTAS ILMU KOMPUTER UNIVERSITAS JEMBER*, 2019, Accessed: Nov. 01, 2023. [Online]. Available: <https://repository.unej.ac.id/xmlui/handle/123456789/98626>
- [15] H. Nobertus Krisandi, "ALGORITMA k-NEAREST NEIGHBOR DALAM KLASIFIKASI DATA HASIL PRODUKSI KELAPA SAWIT PADA PT. MINAMAS KECAMATAN PARINDU," *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 2, no. 1, Mar. 2013, doi: 10.26418/bbimst.v2i1.1540.