



## Perbandingan Algoritma SVM, KNN dan NBC Terhadap Analisis Sentimen Aplikasi Loan Service

Dewi Nurmalasari\*, Teguh Iman Hermanto, Imam Ma'ruf Nugroho

Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Purwakarta, Indonesia

Email: <sup>1,\*</sup>dewinurmalasari20@wastukencana.ac.id, <sup>2</sup>teguhiman@wastukencana.ac.id, <sup>3</sup>imam.ma@wastukencana.ac.id

Email Penulis Korespondensi: dewinurmalasari20@wastukencana.ac.id

**Abstrak**—Menurut data jumlah pengguna kartu kredit di Indonesia mengalami penurunan dari tahun 2019 akhir hingga 2021 salah satu penyebabnya karena pandemi covid-19 yang melanda, karena kondisi ini banyak masyarakat yang mulai beralih ke kredit digital karena dianggap lebih meminimalisir penuluran virus dan proses nya dirasa lebih efisien dibandingkan harus menggunakan kartu kredit. Penelitian ini bertujuan untuk mengetahui perbandingan tingkat akurasi diantara algoritma ketiga algoritma, yaitu naïve bayes classifier, k-nearest neighbor dan support vector machine terhadap aplikasi kredit digital atau sering disebut loan service yaitu Kredivo, Akulaku, dan Indodana di Indonesia dengan mengklasifikasikannya kedalam dua kelas yakni positif dan negatif dengan menggunakan bantuan bahasa pemrograman python untuk menganalisa suatu sentimen dengan melewati proses text preprocessing dan pembobotan kata TF-IDF. Data diperoleh dari review pengguna aplikasi di google playstore dan total keseluruhan data yang diperoleh sebanyak 42.174 kemudian data dibagi dengan rincian 70 data latih dan 30 data uji. Hasil akurasi pada aplikasi kredivo dengan menggunakan K-NN mendapatkan nilai 84%, Naïve Bayes 88%, dan SVM mendapat 89%. Untuk aplikasi akulaku metode K-NN mendapat 79%, Naïve Bayes 86%, dan SVM 87%. Sedangkan untuk aplikasi indodana metode K-NN mendapat 81%, Naïve Bayes 88%, dan SVM 88%. Dari hasil akurasi tersebut dapat disimpulkan bahwa metode Support Vector Machine memiliki hasil akurasi yang lebih baik dibandingkan dengan metode K-Nearest Neighbor dan Naïve Bayes.

**Kata Kunci:** Analisis Sentimen; Loan Service; K-Neares Neighbor; Naïve Bayes; Support Vector Machine.

**Abstract**—According to data on the number of credit card users in Indonesia, it has decreased from late 2019 to 2021 oneThe reason is because of the Covid-19 pandemic that hit. Because of this condition, many people are starting to switch to digital credit because they are considered to minimize transmission of viruses and the process is felt to be more efficient than having to use a credit card. This study aims to compare the level of accuracy between the three algorithms, namely the naïve Bayes classifier, k-nearest neighbor and support vector machine for digital credit applications or often called loan services, namely Kredivo, Akulaku, and Indodana in Indonesia by classifying it into two classes namely positive and negative by using the help of the Python programming language to analyze a sentiment by going through text preprocessing and weighting processes said TF-IDF. The results for the accuracy of the Kredivo application using K-NN get a score of 84%, Naïve Bayes 88%, and SVM get 89%. For the application of the K-NN method, it gets 79%, Naïve Bayes 86%, and SVM 87%. As for the indodana application, the K-NN method gets 81%, Naïve Bayes 88%, and SVM 88%. From the results of this accuracy it can be concluded that the Support Vector Machine method has better accuracy results compared to the K-Nearest Neighbor and Naïve Bayes methods.

**Keywords:** Sentiment Analysis; Loan Service; K-Neares Neighbor; Naïve Bayes; Support Vector Machine.

### 1. PENDAHULUAN

Sistem pembayaran melalui kartu kredit semakin banyak digunakan oleh masyarakat, namun banyak masyarakat yang mengalami kesulitan dalam mengajukan pinjaman dari bank konvensional karena persyaratan yang ketat atau tidak memiliki jaminan yang cukup. Melihat adanya kebutuhan masyarakat akan transaksi non tunai, beberapa penyedia jasa pun menawarkan berbagai produk sebagai solusi dari transaksi non tunai[1]. Data yang didapat dari Bank Indonesia mencatat, jumlah kartu kredit yang beredar di Indonesia mencapai 16,8 juta unit pada juni 2022. Jumlah tersebut naik 0,84% dibandingkan pada juni 2021 yang sebesar 16,56 juta kartu kredit. Meski meningkat dari tahun 2021 dibandingkan dengan tahun 2022, jumlah kartu kredit yang beredar dalam negeri cenderung menurun saat tahun 2019 akhir hingga 2021 dikarenakan pandemi covid-19 melanda. Dengan adanya kondisi tersebut masyarakat lebih beralih ke kredit atau uang digital, dengan anggapan lebih aman dari penularan virus dan efisien. Dalam memilih platform kredit digital yang akan digunakan, biasanya penggunaan mempertimbangkan kenyamanan, keamanan, ketepatan transaksi, kemudahan dan banyaknya promosi[2]. Oleh karena itu, sebelum menggunakan platform kredit digital, pengguna harus melakukan pengecekan terlebih dahulu terhadap reputasi dan legalitas platform tersebut dalam ulasan di aplikasi Google Playstore. Dengan adanya review dan ratin tersebut bisa menilai apakah aplikasi tersebut aman atau tidak, karena peran opini atau pendapat seseorang disini sangat penting dalam kewaspadaan sebelum menggunakan aplikasi tersebut[3]. Hal ini menunjukkan, perlu adanya analisa terhadap sentimen aplikasi loan service untuk membantu masyarakat dalam melihat klasifikasi ulasan agar menentukan aplikasi yang bagus untuk dipilih. Contoh dari platform tersebut adalah Kredivo, Akulaku, Indodana, dll. Persaingan bisnis tersebut tentu saja bukan hanya mencari keuntungan semata, tetapi mencoba menjadi yang terbaik dan memuaskan para pelanggan. Kredivo, Indodana dan Akulaku adalah platform perbankan dan keuangan digital terkemuka di Indonesia. Kehadiran aplikasi pinjaman online hingga aplikasi kredit sebenarnya memilih konsep yang sama dengan VCC (Virtual Credit Card). Sebagai penyedia aplikasi sejenisnya, Google Play Store menyediakan berbagai fitur umpan balik dari konten yang disediakan tersebut, sehingga pengguna dapat memilih sesuai

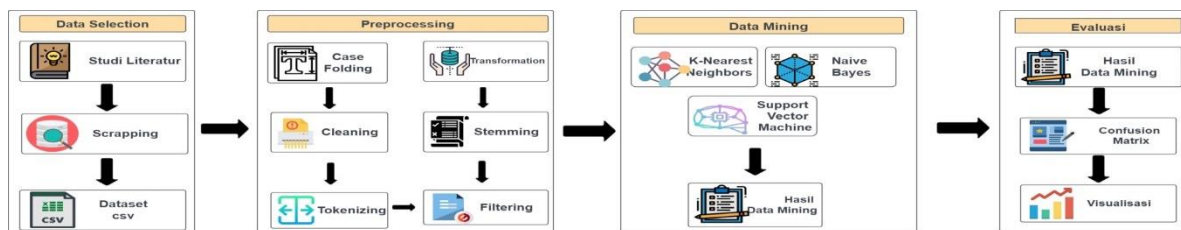


kemauan atau rating yang telah ditampilkan berdasarkan ulasan pengguna sebelumnya yang telah ada[4]. Analisis sentimen adalah proses menentukan sentimen dan mengategorikan polaritas teks dalam sebuah dokumen atau frasa sehingga pembagiannya dapat digambarkan sebagai sentimen positif, negatif, atau netral. Saat ini, para akademisi umumnya menggunakan analisis emosi sebagai cabang penelitian ilmu komputer[5]. Melalui proses analisis sentimen, maka akan terlihat bagaimana kecenderungan opini seseorang terhadap suatu topik atau permasalahan yaitu dengan menetapkan klasifikasi sentimen ke dalam dua atau lebih kelas [6]. Maka dari itu, analisis sentimen pada aplikasi loan service tersebut dapat berguna bagi masyarakat yang ingin memilih aplikasi mana yang bagus untuk digunakan atau mempunyai ulasan positif terbanyak dibandingkan ulasan negatif dari para penggunanya. Penelitian sebelumnya yang dilakukan oleh Rahmat dan Kahir yang berjudul Comparison of Naïve Bayes and SVM Algorithm based on Sentimen Analysis Using Review Dataset menghasilkan akurasi yang di dapatkan pada algoritma SVM mendapatkan nilai 82,48% dan pada algoritma Naïve Bayes bernilai 76,56%. Nilai presisi dan recall pada algoritma SVM adalah 90,33% dan 81,79%. Pada algoritma Naïve Bayes bernilai 89% dan 83,75% untuk nilai presisi dan recall[7]. Kedua algoritma dapat memprediksi nilai secara bagus, namun algoritma SVM lebih bagus dalam hal tersebut. Kemudian penelitian yang dilakukan oleh Fatmanisa, dkk. Dengan judul Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor, hasilnya menunjukkan bahwa algoritma k-NN terbukti menunjukkan hasil terbaik. Dengan rata-rata accuracy OVO sebesar 94.25%, LinkAja 89.63%, Gopay 88.23%, DANA 88.10% dan Shopeepay sebesar 87.03% [8]. Berdasarkan penelitian sebelumnya, pada penelitian ini peneliti akan menggunakan metode Naïve Bayes Classifier, K-Nearest Neighbor dan Support Vector Machine dalam melakukan analisis terhadap aplikasi loan service di Google Play Store. Naïve Bayes Classifier merupakan model yang digunakan untuk pengelompokan berdasarkan probabilitas pada setiap kelas berdasarkan pembagian kata dalam suatu dokumen [9]. Algoritma K-Nearest Neighbor yaitu menentukan dan mencari jarak terdekat dengan nilai k neighbor terdekat dalam data training dengan data yang akan diuji [10]. Sedangkan Support Vector Machine (SVM) merupakan sebuah algoritma yang bekerja dengan cara mendapatkan hyperplane terbaik sebagai pemisah antara dua buah kelas [11]. Adapun tujuan dari penelitian ini yaitu untuk mengetahui perbandingan tingkat akurasi diantara algoritma ketiga algoritma, yaitu Naïve Bayes Classifier, K-Nearest Neighbor dan Support Vector Machine terhadap aplikasi loan service di Google Play Store.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Berikut merupakan tahapan dalam penelitian ini, tahapannya antara lain Data selection, Preprocessing, data mining, dan evaluasi. Diagram alir tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

### 2.2 Data Selection

Pada pengumpulan data ini terdapat 3 tahapan antara lain Studi literatur , crawling data , lalu data yang di dapat akan berupa dataset dengan format csv. Selanjutnya tahapan studi literatur akan dilakukan penelitian mengenai teori-teori yang berkaitan dengan topik yang akan diteliti. Studi literatur ini dilakukan untuk mencari dan membaca mengenai referensi penelitian-penelitian sebelumnya yang telah dilakukan bersumber dari buku-buku, artikel ilmiah ataupun jurnal penelitian yang berkaitan dengan analisis sentimen menggunakan algoritma K-Nearest Neighbors, Naïve Bayes dan SVM. Pada penelitian ini, dataset yang diambil berasal dari Google Play Store dan objek yang digunakan yaitu 3 jenis aplikasi loan service yaitu Kredivo, Akulaku dan Indodana. Total data yang digunakan adalah 42.119 ulasan. ini data yang telah diperoleh melalui scrapping dari data Google Play Store meliputi aplikasi Kredivo, Akulaku dan Indodana menggunakan Google Colab.

### 2.3 Preprocessing

Data preprocessing digunakan agar gangguan terhadap data bisa berkurang. Hal ini penting dilakukan karena data yang diambil dari google play store tentunya tidak selalu menggunakan bahasa yang baku. Pada proses data preprocessing ini juga termasuk ke dalam salah satu tahapan untuk menghilangkan masalah yang dapat mengganggu hasil proses data. Teknik yang dilakukan dalam tahap preprocessing ini adalah case folding,



cleaning, tokenization, stemming dan stopword removal. Tujuan dalam tahapan ini adalah untuk mendapatkan atau menemukan pola yang berguna untuk dijadikan informasi dengan tujuan tertentu [12]. Tahapan pertama yaitu case folding ialah proses penyamaan case dalam sebuah dokumen. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital[13]. Tahapan ke dua proses cleaning ialah membersihkan atau menghilangkan karakter yang tidak diperlukan pada data[14]. Tahapan yang ke tiga ada Tokenizing merupakan proses mengambil kata-kata penting dari hasil token. Pada proses ini akan dilakukan pembersihan dari spesial karakter, URL link, username, serta emoticon[15]. Pada Tahapan yang ke empat yaitu Filtering ialah untuk mengambil kata-kata penting dari hasil token [16]. Selanjutnya ditahapan ke lima yaitu Stemming merupakan tahap yang juga diperlukan untuk mengurangi jumlah indeks yang berbeda dari sebuah data sehingga sebuah kata yang memiliki akhiran dan awalan akan kembali ke bentuk dasarnya [17] Tahapan terakhir yaitu transformation , Pada tahapan ini juga bisa dilakukan dengan memberikan suatu inialisasi terhadap data yang memiliki nilai atau tipe data nominal menjadi numerik, kondisi ini dilakukan ketika proses modelling memerlukan pemrosesan dalam hal numerik[18].

**2.3 Data Mining**

Berdasarkan kategori tertentu dokumen dapat diklasifikasikan dengan kata-kata yang terkandung dalam dokumen tersebut. . Data mining juga didefinisikan sebagai proses otomatis terhadap data yang berukuran besar dan bertujuan untuk menemukan hubungan antar data tersebut[19]. Berbagai macam kumpulan kata pada dokumen dapat digunakan untuk menentukan kategori yang memiliki nilai atau makna tertentu. Beberapa algoritma yang digunakan untuk klasifikasi teks pada data mining yaitu K-Nearest Neighbor, Naïve Bayes dan Support Vector Machine (SVM).

**2.4 Evaluasi**

Informasi yang didapatkan pada tahapan data mining akan sulit kita lihat atau mengerti dalam bahasa manusia, maka dari itu perlu adanya evaluasi agar hasil pola yang didapatkan menjadi lebih mudah dimengerti oleh pihak yang akan membaca atau menggunakan[20]. Sebagai evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Confusion matrix digunakan untuk membandingkan hasil klasifikasi dengan data asli dan prediksi sebagai pembandingnya. Pada tahapan evaluasi ini dilakukan pengujian kinerja algoritma pada data mining yaitu K-Nearest Neighbor, Naïve Bayes dan Support Vector Machine (SVM).

**3. HASIL DAN PEMBAHASAN**

**3.1 Scraping Data**

Di dalam tahapan ini dilakukan proses scrapping dengan menggunakan library google play scrapper dan data yang diambil berdasarkan relevansi ulasan pada setiap aplikasi. Dataset yang dihasilkan dari hasil scrapping adalah Kredivo sebanyak 14.471 ulasan, Akulaku sebanyak 14.142 ulasan dan Indodana sebanyak 13.606 ulasan. Maka, total data yang digunakan adalah 42.119 ulasan / dataset.

**3.2 Pelabelan**

Pada tahap ini data diberikan label. Pelabelan dilakukan sesuai dengan rating yang diberikan pengguna. rating 1 dan 2 akan masuk kelas negatif, rating 4 dan 5 masuk kelas positif sedangkan rating 3 dihapus karena terlalu ambigu. Berikut kode program untuk pelabelan dataset terlampir pada Gambar 2.

Kredivo				Akulaku				Indodana			
content	score	sentiment		content	score	sentiment		content	score	sentiment	
Aplikasi pinjaman online terbaik pelayanan nya...	5	positive		Sangat lengkap sekali fitur akulaku mau pinjam...	5	positive		platform yang sangat andal, dengan begitu bany...	5	positive	
Sangat beruntung sekali menemukan aplikasi ini...	5	positive		Untuk aplikasi ini sih Oke sekali, karna bema...	5	positive		Tidak pernah ada tunggakan tidak pernah telat ...	1	negative	
Membantu sekali dengan ada nya aplikasi ini pr...	5	positive		Dari dalam banyak aplikasi. Saya lebih suka me...	5	positive		aplikasi ini sangat menyenangkan dan sangat me...	5	positive	
Setelah update aplikasi tidak bisa di buka. S...	1	negative		Berlangganan paylater di akulaku sangat lah me...	5	positive		Aplikasi ini sangat mengecewakan sekali, untuk...	1	negative	
Bulan kedua pemakaian, aplikasi lagi lagi ga b...	1	negative		Suka banget sama aplikasi akulaku ini. Sangat ...	5	positive		Dunia perbelanjaan saya selalu memakai paylate...	4	positive	

Gambar 2. Pelabelan Dataset

**3.3 Preprocessing Data**

Dalam tahap preprocessing data ini dibagi lagi menjadi beberapa proses mulai dari case folding, cleansing, tokenizing, filtering, stemming dan transformation dengan hasil sebagai berikut :

1. Case folding, Tahapan ini mengubah data huruf semua teks menjadi lowercase atau huruf kecil. Terlihat pada Tabel 1 proses perubahan data menjadi huruf kecil.

Tabel 1. Proses Case folding

Input	Output
aplikasi payah. padahal bayar selalu tepat tidak pernah telat dan akun sudah menjadi premium, tapi	aplikasi payah padahal bayar selalu tepat tidak pernah telat dan akun sudah menjadi premium tapi



Input	Output
tiba2 limit yg besar menjadi nol seketika hilang tidak bisa digunakan. aplikasi yang mengecewakan butuh waktu lama untuk menjadi akun dari basic ke premium, setelah jadi premium malah limit hilang.	tiba limit yg besar menjadi nol seketika hilang tidak bisa digunakan aplikasi yang mengecewakan butuh waktu lama untuk menjadi akun dari basic ke premium setelah jadi premium malah limit hilang

2. Cleaning, Terlihat pada Tabel 2 proses pada tahapan cleaning, dilakukan penghapusan tanda baca yang tidak digunakan pada saat melakukan data mining nantinya. Tahapan tersebut antara lain, seperti penghapusan tanda baca titik, koma dan tanda baca lainnya,

**Tabel 2.** Proses Cleaning

Input	Output
aplikasi payah. padahal bayar selalu tepat tidak pernah telat dan akun sudah menjadi premium, tapi tiba2 limit yg besar menjadi nol seketika hilang tidak bisa digunakan. aplikasi yang mengecewakan butuh waktu lama untuk menjadi akun dari basic ke premium, setelah jadi premium malah limit hilang.	aplikasi payah padahal bayar selalu tepat tidak pernah telat dan akun sudah menjadi premium tapi tiba limit yg besar menjadi nol seketika hilang tidak bisa digunakan aplikasi yang mengecewakan butuh waktu lama untuk menjadi akun dari basic ke premium setelah jadi premium malah limit hilang

3. Tokenizing, pada Tabel 3 di dalam proses tokenizing, akan dilakukan sebuah proses untuk memecah atau memotong kalimat menjadi urutan kata atau per-string. Terlihat pada Tabel 3 dan Tabel 4 proses tokenizing.

**Tabel 3.** Proses Input Tokenizing

Input
aplikasi payah. padahal bayar selalu tepat tidak pernah telat dan akun sudah menjadi premium, tapi tiba2 limit yg besar menjadi nol seketika hilang tidak bisa digunakan. aplikasi yang mengecewakan butuh waktu lama untuk menjadi akun dari basic ke premium, setelah jadi premium malah limit hilang.

**Tabel 4.** Proses Output Tokenizing

Output
aplikasi payah bayar telat akun premium limit yg nol hilang aplikasi mengecewakan butuh akun basic premium premium limit hilang

4. Filtering, pada tahapan proses ini, digunakan tahapan stopword removal atau membuang kata penghubung yang tidak berpengaruh dalam sebuah kalimat. Terlihat pada Tabel 5 dan Tabel 6 untuk proses tahapan sebelum di filtering dan sesudah difiltering.

**Tabel 5.** Proses Input Filtering

Input
aplikasi payah padahal bayar selalu tepat tidak pernah telat dan akun sudah menjadi premium tapi tiba limit yg besar menjadi nol seketika hilang tidak bisa digunakan aplikasi yang mengecewakan butuh waktu lama untuk menjadi akun dari basic ke premium setelah jadi premium malah limit hilang

**Tabel 6.** Proses Output Filtering

Output
aplikasi payah bayar telat akun premium limit yg nol hilang aplikasi mengecewakan butuh akun basic premium premium limit hilang

5. Steaming, tahapan ini adalah proses perubahan kata dasar atau menghilangkan imbuhan atau mengubah kata jamak menjadi kata dasar dalam suatu teks dipaparkan pada Tabel 7 untuk prosesnya sebelum dan sesudah steaming.



**Tabel 7.** Proses Input Steaming

Input				
aplikasi	payah	bayar	telat	akun
premium	limit	yg	nol	hilang
aplikasi	mengecewakan	butuh	akun	basic
premiun	premiun	limit	hilang	

**Tabel 8.** Proses Output Steaming

Output
aplikasi payah bayar telat akun premium limit yg nol hilang aplikasi kecewa butuh akun basic premiun premiun limit hilang

6. Transformation

Pada transformation, akan dilakukan pembobotan kata menggunakan TF-IDF. Pembobotan kata menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) bekerja dengan menghitung bobot pada setiap kata dan menyajikan skor dari frekuensi kata penting atau tingkat frekuensi kemunculan yang tinggi dari sebuah kata. Berikut tahapan pembobotan kata menggunakan TF-IDF :

- a) Mengambil sampel dokumen yang telah dilakukan text preprocessing untuk dilakukan proses TF-IDF, terlampir pada Tabel 9 proses sampel dokumen TF-IDF.

**Tabel 9.** Proses Pengambilan sampel dokumen TF-IDF

D1	terbaik cepat proses
D2	cepat proses percaya rahasia data
D3	pengajuan mudah limit besar

- b) Setelah mengambil sampel pada Tabel 9, selanjutnya melakukan frekuensi kemunculan (tf) suatu kata dalam suatu dokumen, dijelaskan pada Tabel 10 proses pencarian frekuensi tf.

**Tabel 10.** Proses Mencari frekuensi TF

Term	TF		
	D1	D2	D3
terbaik	1	0	0
cepat	1	1	0
proses	1	1	0
percaya	0	1	0
rahasia	0	1	0
data	0	1	0
pengajuan	0	0	1
mudah	0	0	1
limit	0	0	1
besar	0	0	1

- c) Data yang sudah dicari pada Tabel 10, lalu Tahapan selanjutnya Menghitung jumlah frekuensi kata terpilih yang sudah diketahui kemudian menghitung jumlah dokumen (idf), terlihat pada Tabel 11. Proses Data IDF yang sudah dihitung.

**Tabel 11.** Proses Menghitung Jumlah IDF

Term	DF	TD/DF	IDF
terbaik	1	3	0,477
cepat	2	1,5	0,176
proses	2	1,5	0,176
percaya	1	3	0,477
rahasia	1	3	0,477
data	1	3	0,477
pengajuan	1	3	0,477
mudah	1	3	0,477
limit	1	3	0,477
besar	1	3	0,477

- d) Setelah hasil data perhitungan didapat pada Tabel 11, selanjutnya tahapan ini mencari nilai TF-IDF dengan cara mengalikan tf dan idf yang dijelaskan prosesnya di Tabel 12.

**Tabel 12.** Proses Mencari Nilai TF dan IDF



TF*IDF		
D1	D2	D3
0,477	0	0
0,176	0,176	0
0,176	0,176	0
0	0,477	0
0	0,477	0
0	0,477	0
0	0	0,477
0	0	0,477
0	0	0,477
0	0	0,477

e) Terakhir jumlahkan semua nilai TF-IDF yang ada pada document, hasil akhir yang didapat menghasilkan nilai akhir yang dijelaskan prosesnya pada Tabel 13.

**Tabel 13.** Proses Hasil Akhir Nilai TF-IDF

D1	D2	D3
0,829	1,783	1,908

**3.4 Data mining**

Setelah proses TF-IDF selesai kemudian masuk ke tahap proses data mining untuk memproses datanya. Proses data mining dalam penelitian ini menggunakan model K-NN, Naïve Bayes dan SVM. Berikut gambar model klasifikasi

```

K-NN
from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier()
knn.fit(X_train, y_train)

y_pred_knn = knn.predict(X_test)
y_pred_knn_prob = knn.predict_proba(X_test)

print('=====')
print(confusion_matrix(y_test, y_pred_knn, labels = knn.classes_))
print('=====')
print('=====')
print(classification_report(y_test, y_pred_knn))
    
```

**Gambar 3.** Metode K-NN

```

Naïve Bayes
from sklearn.naive_bayes import MultinomialNB

nb = MultinomialNB().fit(X_train, y_train)
y_pred_nb = nb.predict(X_test)

print('=====')
print(confusion_matrix(y_test, y_pred_nb, labels = nb.classes_))
print('=====')
print('=====')
print(classification_report(y_test, y_pred_nb))
    
```

**Gambar 4.** Metode Naïve Bayes

```

Support Vector Machine
from sklearn import svm
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.svm import SVC

svm = SVC(kernel = 'rbf', probability = True)
svm.fit(X_train, y_train)

y_pred_svm = svm.predict(X_test)
y_pred_svm_prob = svm.predict_proba(X_test)

print('=====')
print(confusion_matrix(y_test, y_pred_svm, labels = svm.classes_))
print('=====')
print('=====')
print(classification_report(y_test, y_pred_svm))
    
```

**Gambar 5.** Metode SVM

**3.5 Evaluasi**

Hasil dari proses klasifikasi model selanjutnya dilakukan tahap evaluasi untuk melihat performance dari ketiga algoritma yang digunakan terhadap ke 3 aplikasi loan service.

1. Hasil Data Mining

a). Aplikasi Loan Service Kredivo

Pada tahap evaluasi confusion matrix di google collab dengan algoritma K-NN , Naïve bayes dan SVM mengambil sampel data kredivo menghasilkan nilai dengan algoritma K-NN akurasi 84%, presisi 79% , recal 73%, dengan menggunakan Naïve Bayes menghasilkan nilai akurasi 88% , presisi 83% dan recal 83% dan F1-Score 83% , dan menggunakan SVM menghasilkan nilai akurasi 89% , presisi 81% . recal 89% dan F1-Score 85%. Pada Tabel 14. dapat disimpulkan bahwa SVM mempunyai nilai algoritma tertinggi di aplikasi kredivo.

**Tabel 14.** Hasil Pengujian Data Kredivo

	K-NN	Naïve Bayes	SVM
Akurasi	84%	88%	89%
Presisi	79%	83%	81%
Recal	73%	83%	89%
F1-Score	76%	83%	85%

b). Aplikasi Loan Service Akulaku

Pada tahap evaluasi confusion matrix di google collab dengan algoritma K-NN , Naïve bayes dan SVM mengambil sampel data akulaku menghasilkan nilai dengan algoritma K-NN akurasi 79%, presisi 78% , recal 70%, dan F1-Score 74% , dengan menggunakan Naïve Bayes menghasilkan nilai akurasi 86% , presisi 81% dan recal 88% dan F1-Score 84% , dan menggunakan SVM menghasilkan nilai akurasi 87% , presisi 81% . recal 91% dan F1-Score 86%.



**Tabel 15.** Hasil Pengujian Data Akulaku

	K-NN	Naïve Bayes	SVM
Akurasi	79%	86%	87%
Presisi	78%	81%	81%
Recal	70%	88%	91%
F1-Score	74%	84%	86%

c). Aplikasi Loan Service Indodana

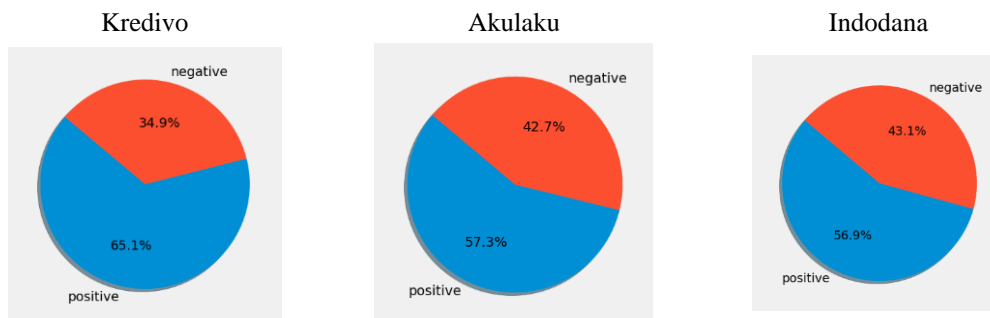
Pada tahap evaluasi confusion matrix di google collab dengan algoritma K-NN, Naïve bayes dan SVM mengambil sampel data akulaku menghasilkan nilai dengan algoritma K-NN akurasi 81%, presisi 85% , recal 69%, dan F1-Score 76% , dengan menggunakan Naïve Bayes menghasilkan nilai akurasi 88% , presisi 86% dan recal 86% dan F1-Score 86% , dan menggunakan SVM menghasilkan nilai akurasi 88% , presisi 85% . recal 89% dan F1-Score 87%.

**Tabel 16.** Hasil Pengujian Data Indodana

	K-NN	Naïve Bayes	SVM
Akurasi	81%	88%	88%
Presisi	85%	86%	85%
Recal	69%	86%	89%
F1-Score	76%	86%	87%

**3.6 Visualisasi**

Pada Gambar 6. dapat disimpulkan bahwa perbandingan sentimen pada aplikasi kredivo dominan memiliki sentimen yang positif. Sentimen positif sebanyak 65.1% atau 9.351 ulasan bersifat positif. Pada Gambar 7. dapat disimpulkan bahwa perbandingan sentimen pada aplikasi Akulaku dominan memiliki sentimen yang positif. Sentimen positif sebanyak 57.3% atau 8.092 ulasan bersifat positif. Walaupun masih dominan positif, ulasan yang bersifat negatif juga terbilang masih banyak dari total data sebanyak 14.123 ulasan. Pada Gambar 8. Dapat disimpulkan bahwa perbandingan sentimen pada aplikasi Indodana dominan memiliki sentimen yang positif. Sentimen positif sebanyak 56.9% atau 7.784 ulasan bersifat positif. Sama halnya dengan perbandingan sentimen pada data Akulaku, sentimen pada data Indodana juga terbilang masih tidak berbeda jauh secara signifikan.



**Gambar 6.** Sentimen Kredivo    **Gambar 7.** Sentimen Akulaku    **Gambar 8.** Sentimen Indodana

Hasil pengolahan data direpresentasikan dengan wordcloud. Setiap aplikasi mengimplementasikan visualisasi wordcloud menggunakan data pelatihan yang telah mengalami pengolahan data. seperti gambar 9 , 10 dan 11 pada gambar dibawah ini.



**Gambar 9.** WordCloud Kredivo



**Gambar 10.** WordCloud Akulaku



**Gambar 11.** WourCloud Indodana

**3.7 Confusion Matrix Kredivo**

Pada Gambar 12, 13 dan 14 menunjukkan confusion matrix berukuran 2 x 2 yang merepresentasikan setiap kelas klasifikasi positif dan negatif. Hasil Dari confusion matrix tersebut dapat dihitung secara manual untuk melihat nilai akurasi seperti dijelaskan pada tabel 17, 18 dan 19. Hasil Nilai akurasi tertinggi terdapat pada algoritma SVM 90,3%.

**Tabel 17.** Nilai Akurasi K-NN Kredivo

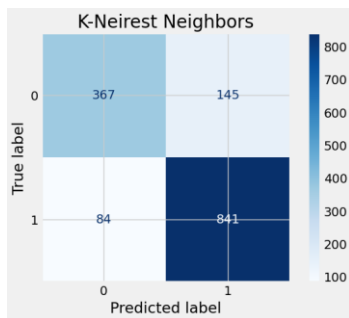
$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{841 + 367}{841 + 145 + 367 + 84} \times 100\% \\
 &= \frac{1.208}{1.437} \times 100\% \\
 &= \mathbf{84,06\%}
 \end{aligned}$$

**Tabel 18.** Nilai Akurasi Naïve Bayes Kredivo

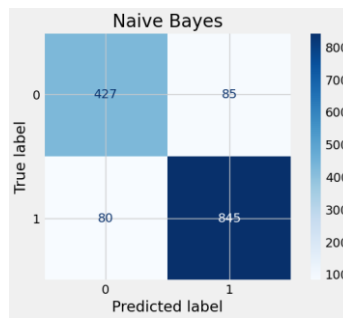
$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{845 + 427}{841 + 85 + 427 + 80} \times 100\% \\
 &= \frac{1.272}{1.437} \times 100\% \\
 &= \mathbf{88,51\%}
 \end{aligned}$$

**Tabel 19.** Nilai Akurasi SVM Kredivo

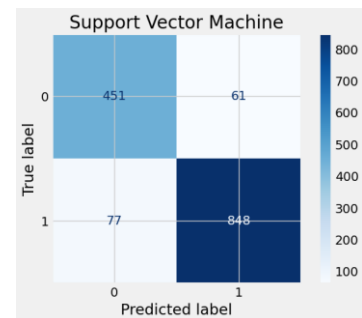
$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{848 + 451}{848 + 61 + 451 + 77} \times 100\% \\
 &= \frac{1.299}{1.437} \times 100\% \\
 &= \mathbf{90,3\%}
 \end{aligned}$$



**Gambar 12.** Confusion Matrix K-NN



**Gambar 13.** Confusion Matrix Naïve Bayes



**Gambar 14.** Confusion Matrix SVM

**3.8 Confusion Matrix Akulaku**

Selanjutnya Aplikasi Akulaku, Pada Gambar 15, 16 17 menunjukkan confusion matrix berukuran 2 x 2 yang merepresentasikan setiap kelas klasifikasi positif dan negatif. Hasil Dari confusion matrix tersebut dapat dihitung secara manual untuk melihat nilai akurasi seperti dijelaskan pada tabel 17, 18 dan 19. Hasil Nilai akurasi tertinggi terdapat pada algoritma SVM 87,4%.

**Tabel 20.** Nilai Akurasi K-NN Akulaku

$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{692 + 421}{692 + 185 + 421 + 115} \times 100\% \\
 &= \frac{1.113}{1.413} \times 100\% \\
 &= \mathbf{78,76\%}
 \end{aligned}$$

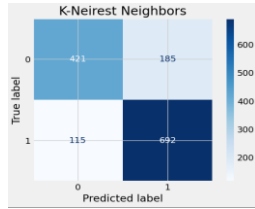
**Tabel 21.** Nilai Akurasi Naïve Bayes Indodana

$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{641 + 561}{641 + 69 + 561 + 141} \times 100\% \\
 &= \frac{1.202}{1.412} \times 100\% \\
 &= \mathbf{85,12\%}
 \end{aligned}$$

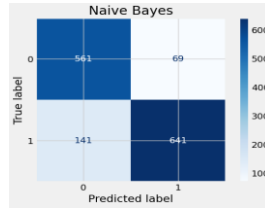
**Tabel 22.** Nilai Akurasi SVM Indodana

$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{691 + 544}{691 + 62 + 544 + 116} \times 100\% \\
 &= \frac{1.235}{1.413} \times 100\% \\
 &= \mathbf{87,4\%}
 \end{aligned}$$

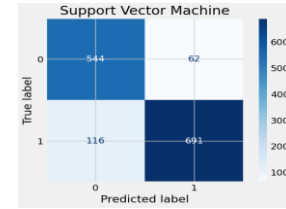




**Gambar 15.** Confusion Matrix K-NN



**Gambar 16.** Confusion Matrix Naïve Bayes



**Gambar 17.** Confusion Matrix SVM

### 3.9 Confusion Matrix Indodana

Pada Gambar 18, 19 dan 20 menunjukkan confusion matrix berukuran 2 x 2 yang merepresentasikan setiap kelas klasifikasi positif dan negatif. Hasil Dari confusion matrix tersebut dapat dihitung secara manual untuk melihat nilai akurasi seperti dijelaskan pada tabel 23, 24 dan 25. Hasil Nilai akurasi tertinggi terdapat pada algoritma SVM 87,42%.

**Tabel 23.** Nilai Akurasi K-NN Indodana

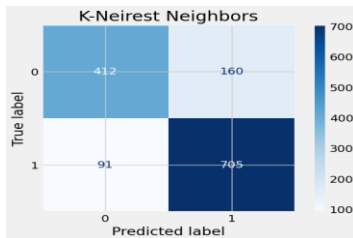
$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{705 + 412}{705 + 160 + 412 + 91} \times 100\% \\
 &= \frac{1.117}{1.368} \times 100\% \\
 &= \mathbf{81,65\%}
 \end{aligned}$$

**Tabel 24.** Nilai Akurasi Naïve Bayes Indodana

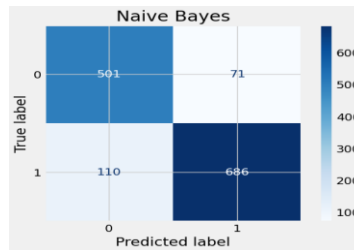
$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{686 + 501}{686 + 71 + 501 + 110} \times 100\% \\
 &= \frac{1.187}{1.368} \times 100\% \\
 &= \mathbf{86,77\%}
 \end{aligned}$$

**Tabel 25.** Nilai Akurasi SVM Indodana

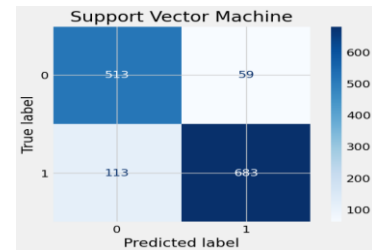
$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\
 &= \frac{683 + 513}{683 + 59 + 513 + 113} \times 100\% \\
 &= \frac{1.196}{1.368} \times 100\% \\
 &= \mathbf{87,42\%}
 \end{aligned}$$



**Gambar 18.** Confusion Matrix K-NN

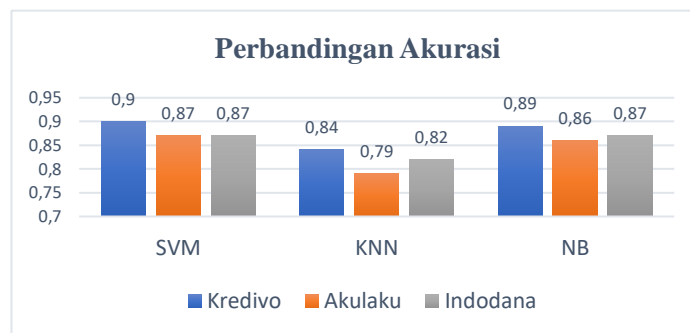


**Gambar 19.** Confusion Matrix Naïve Bayes



**Gambar 20.** Confusion Matrix SVM

Pada Gambar 21, dapat disimpulkan bahwa algoritma Support Vector Machine memiliki rata-rata nilai terbaik diantara ketiga algoritma selain KNN dan Naïve Bayes pada ketiga data. Algoritma Support Vector Machine dapat dibuktikan bahwa cocok digunakan untuk analisis sentimen pada ulasan aplikasi loan service di platform google play store.



**Gambar 21.** Perbandingan Nilai Akurasi

## 4. KESIMPULAN

Berdasarkan hasil penelitian dengan menggunakan data yang diperoleh melalui review pada google play store terhadap aplikasi loan service yaitu kredivo, akulaku dan indodana dengan total data kredivo 14368, akulaku



14134, dan indodana 13672. Kemudian data diberikan label yang menghasilkan 65,1% data positif 34,9% data negative kredivo, 57,4% data positif 42,6% data negatif untuk akulaku, dan 56,7% data positif 43,3% data negatif indodana. Setelah data diberikan label selanjutnya data diolah melalui proses text preprocessing dan dilanjutkan dengan tahap pembobotan kata menggunakan TF-IDF kemudian baru masuk ke tahap klasifikasi dengan menggunakan metode K-NN, Naïve Bayes dan SVM. Hasilnya akurasi pada aplikasi kredivo dengan menggunakan K-NN mendapatkan nilai 84%, Naïve Bayes 88%, dan SVM mendapat 89%. Untuk aplikasi akulaku metode K-NN mendapat 79%, Naïve Bayes 86%, dan SVM 87%. Sedangkan untuk aplikasi indodana metode K-NN mendapat 81%, Naïve Bayes 88%, dan SVM 88%. Dari hasil akurasi tersebut dapat disimpulkan bahwa metode Support Vector Machine memiliki hasil akurasi yang lebih baik dibandingkan dengan metode K-Nearest Neighbor dan Naïve Bayes.

## REFERENCES

- [1] E. K. Panginan et al., "FENOMENA APLIKASI KREDIT PINJAMAN ONLINE KREDIVO DI INDONESIA," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 4, pp. 247–253, 2022, doi: 10.55123/insologi.v1i4.770.
- [2] 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 J. Sains dan Manaj.*, vol. 8, no. 1, pp. 93–105, 2020, doi: 10.31294/evolusi.v8i1.7535.
- [3] A. Muhammadiyah and I. A. Sobari, "Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma Svm Dan Nbc," *Reputasi J. Rekayasa Perangkat Lunak*, vol. 2, no. 2, pp. 85–91, 2021, doi: 10.31294/reputasi.v2i2.785.
- [4] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.31539/intecomsv5i1.3708.
- [5] Pristiyono et al., "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, 2021, doi: 10.1088/1757-899X/1088/1/012045.
- [6] Junadhi, Agustin, M. Rifqi, and M. K. Anam, "Sentiment Analysis of Online Lectures using K-Nearest Neighbors based on Feature Selection," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 216–225, 2022, doi: 10.23887/janapati.v11i3.51531.
- [7] A. M. Rahat, A. Kahir, A. Kaisar, and M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," 2019.
- [8] F. M. Delta Maharani, A. Lia Hananto, S. Shofia Hilabi, F. Nur Apriani, A. Hananto, and B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor," *Metik J.*, vol. 6, no. 2, pp. 97–103, 2022, doi: 10.47002/metik.v6i2.372.
- [9] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
- [10] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN , Naïve Bayes Dan Decision Tree," vol. 9, no. 2, pp. 1109–1119, 2022.
- [11] F. E. Kavabilla, T. Widiharih, and B. Warsito, "Analisis Sentimen Pada Ulasan Aplikasi Investasi Online Ajaib Pada Google Play Menggunakan Metode Support Vector Machine Dan Maximum Entropy," *J. Gaussian*, vol. 11, no. 4, pp. 542–553, 2023, doi: 10.14710/j.gauss.11.4.542-553.
- [12] Imamah and F. H. Rachman, "Twitter sentiment analysis of Covid-19 using term weighting TF-IDF and logistic regresion," *Proceeding - 6th Inf. Technol. Int. Semin. ITIS 2020*, pp. 238–242, 2020, doi: 10.1109/ITIS50118.2020.9320958.
- [13] D. Alita and A. Rahman, "Pendeteksan Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *J. Komputasi*, vol. 8, no. 2, pp. 50–58, 2020, doi: 10.23960/komputasi.v8i2.2615.
- [14] A. B. Putra and S. Nita, "Perancangan dan Pembangunan Sistem Informasi E-Learning Berbasis Web ( Studi Kasus Pada Madrasah Aliyah Kare Madiun )," *Semin. Nas. Teknol. Inf. dan Komun.* 2019, vol. 1, no. 1, pp. 81–85, 2019.
- [15] I. Susianti, S. S. Ningsih, M. Al Haris, and T. W. Utami, "Analisis Sentimen Pada Twitter Terkait New Normal Dengan Metode Naive Bayes Classifier," *Pros. Semin. Edusainstech FMIPA UNIMUS*, pp. 354–363, 2020, [Online]. Available: <https://prosiding.unimus.ac.id/index.php/edusaintek/article/view/576/578>
- [16] P. R. Agatha and A. M. Polina, "Analisis Sentimen Terhadap Penggunaan Marketplace di Indonesia Menggunakan Metode Support Vector Machine dengan Seleksi Fitur Chi Square," *SINARINT (Seminar Nas. Ris. dan Inov. Teknol.*, vol. 1, pp. 314–323, 2022.
- [17] M. Ikhsan, N. Adhiatma, and M. Sauri, "APLIKASI PELAPORAN DATA GAIN AND LOSSES SPBU 23.361.03 KOTA JAMBI BERBASIS WEB," *LP2M STMIK NURDIN HAMZAH JAMB*, pp. 39–42, 2019.
- [18] Ivan Dwi Nugraha and Y. Azhar, "Deteksi Depresi Pengguna Twitter Indonesia Menggunakan LSTM-RNN," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 320–329, 2022, doi: 10.23887/janapati.v11i3.50674.
- [19] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, "Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization," 2020 5th Int. Conf. Informatics Comput. ICIC 2020, vol. 2, 2020, doi: 10.1109/ICIC50835.2020.9288655.
- [20] A. Sofiyana and A. Azkiya, "PENERAPAN METODE ROUGH SET MENGANALISIS PENYAKIT YANG SERING DIKELUHKAN PASIEN ( STUDI KASUS PUSKESMAS JAYA MUKTI DUMAI )," *Inform. (Jurnal Inform. Manaj. dan Komput.*, vol. 14, no. 1, pp. 31–40, 2022.