

Analisis Sentimen Nasabah Pada Kotak Komentar Terhadap Pelayanan Koperasi Simpan Pinjam Bangun Mandiri Menggunakan Metode K-Nearest Neighbor

Natalia Krismonika Lingga
Teknik Informatika, Universitas Budi Darma, Indonesia
e-mail: lialingga24@gmail.com

Keywords:

Customer Sentiment;
Service;
Cooperative;
Savings and Loan;
Data Mining;
K-Nearest Neighbor.

ABSTRACT

The problem of analyzing customer sentiment that often occurs in finding out the meaning of cooperative customer comments is very difficult, especially to determine whether the comment is positive or negative because there are some languages that are not understood, sentiment needs to be known to determine the next step the cooperative does to improve the quality of cooperative services. then to analyze customer sentiment, a computer application is carried out on a computer to find much more accurate results, the steps are carried out by filtering each word comment which is then seen by the closeness of the value of each cometar to know whether it is a positive or negative comment. The computer application used is the K-Nearest Neighbor approach algorithm with data processed directly using RapidMner 9.10.8 tools. The results of the research found similarities and closeness of values to old data that are worth positive comments, so the results of research on new data are stated that the sentiment of new customers is positive.

Kata Kunci:

Sentiment Nasabah;
Pelayanan;
Koperasi;
Simpan Pinjam;
Data Mining;
K-Nearest Neighbor.

ABSTRAK

Permasalahan analisa sentiment nasabah yang sering terjadi dalam mencari tahu makna dari komentar nasabah koperasi sangat sulit, terutama untuk menentukan komentar tersebut bermakna positif atau negatif karena terdapat beberapa Bahasa yang tidak dimengerti, sentiment perlu diketahui untuk meningkatkan Kualitas dari pelayanan koperasi. selanjutnya untuk menganalisa sentiment nasabah dilakukan penerapan komputer terhadap komputer untuk menemukan hasil yang jauh lebih akurat, tahapan dilakukan dengan menyaring setiap perkata komentar yang selanjutnya dilihat kedekatan nilai masing-masing kometar untuk diketahui terhadap komentar apakah itu komentar positif atau negatif. Penerapan komputer yang digunakan adalah algoritma pendekatan K-Nearest Neighbor dengan data yang dikelola langsung menggunakan tools RapidMner 9.10.8. Hasil dari penelitian ditemukan kemiripan dan kedekatan nilai terhadap data lama yang bernilai komentar positif, sehingga hasil penelitian terhadap data baru dinyatakan sentiment nasabah baru bersifat positif.

Korespondensi Penulis *):

Natalia Krismonika Lingga
Universitas Budi Darma
Jl. Sisingamangaraja No. 338 Simpang Limun, Kota Medan

Diajukan: 19-04-2024 | Diterima: 19-04-2024 | Diterbitkan: 30-04-2024

1. PENDAHULUAN

Koperasi merupakan sebuah badan usaha milik perorangan maupun kelompok yang mirip dengan Bank yaitu memiliki cara kerja yang hampir serupa dalam urusan simpan dan peminjaman uang sesuai dengan syarat dan ketentuan yang telah diatur dan disepakati demi kemajuan dan kesejahteraan bersama dalam urusan pengolahan keuangan. Koperasi Simpan Pinjam Bangun Mandiri yang merupakan salah satu koperasi yang sangat berkembang dikalangan masyarakat kota Medan, kemudahan akses dan negosiasi dengan bunga yang rendah membuat koperasi ini sangat diminati oleh banyak kalangan masayarat dalam membantu permasalahan keuangan, baik dalam urusan

simpan uang, pinjam uang dan jaminan terhadap barang yang akan dijadikan uang. Setiap kegiatan transaksi antara pihak perusahaan dengan nasabah tentunya dikelola oleh sistem komputer.

Hubungan baik antara Koperasi Simpan Pinjam Bangun Mandiri dengan nasabah akan membuat pihak Koperasi Simpan Pinjam Bangun Mandiri menjadi lebih berkembang dan maju sesuai dengan cita-cita koperasi tersebut. Nasabah yang menaruh stigma dan komentar negatif terhadap pelayanan koperasi yang membuat tumbuh kembang dan lancarnya usaha Koperasi Simpan Pinjam Bangun Mandiri menjadi terhambat. Dampak negatif yang akan dihadapi oleh pihak Koperasi Simpan Pinjam Bangun Mandiri jika tidak memanagemen kepuasan nasabah terhadap pelayanan koperasi secara baik dapat menyebabkan pihak koperasi mengalami kehilangan kepercayaan nasabah dan tidak memiliki nasabah baru dikemudian harinya.

Hal tersebut dapat berdampak besar kepada kebangkrutan yang akan dialami oleh pihak koperasi sehingga pihak koperasi melakukan analisis komentar nasabah terhadap pelayanan koperasi menggunakan kertas note yang tersedia dalam bentuk mengetahui tingkat kepuasan terhadap pelayanan koperasi.

Kertas note yang berisi komentar nasabah tersebut menjadi masalah baru kembali dihadapi oleh pihak koperasi, dikarenakan data yang masuk masih secara manual dan sulit diketahui apakah lebih banyak komentar bernilai positif, negatif atau netral dari masing-masing nasabah. Sehingga permasalahan tersebut harus dikelola dengan sebuah teknologi yang tepat untuk mempermudah pihak koperasi mengetahui sentiment nasabah terhadap pelayanan koperasi.

Teknologi yang memproses data biasanya diterapkan ke dalam sebuah teknologi komputer dengan memanfaatkan penerapan Data Mining dalam mencapai tujuan tertentu. Data Mining merupakan teknik komputer yang membantu menggali informasi berdasarkan data yang sudah tersedia, pada penelitian ini dilakukan analisis sentiment nasabah yang ada pada kotak komentar terhadap pelayanan koperasi simpan pinjam bangun mandiri menggunakan logika Data Mining dalam bentuk algoritma K-Nearest Neighbor. Jika sudah diketahui statement yang domain dari komentar nasabah terhadap pelayanan koperasi baik itu positif, negatif atau netral maka akan memberikan arahan dan strategi baru yang akan dilakukan oleh koperasi untuk meningkatkan pelayanan koperasi kepada nasabah.

Pada penelitian terdahulu algoritma Nearest Neighbor telah digunakan oleh Rasenda pada tahun 2020 yang melakukan pengujian terhadap Analisa sentiment riba pada bank dan mendapatkan hasil penelitian yang akurat sebesar 70,59% dengan jumlah kasus sebanyak 170 dan hasil asumsi masyarakat sesuai dengan kebutuhan dan tidak berpengaruh terhadap riba[1].

Penelitian terdahulu digunakan algoritma oleh Akhmad Deviyanto di tahun 2018 dalam pengujian sentiment masyarakat twitter terhadap pilkada yang dilakukan pemerintahan setempat mendapatkan hasil penelitian terbagi atas dua sentiment yaitu positif dan juga negatif dengan tingkat akurasi terbesar adalah "67,2% ketika k=5, presisi tertinggi 56,94% ketika k=5, dan recall 78,24% dengan k=15"[2].

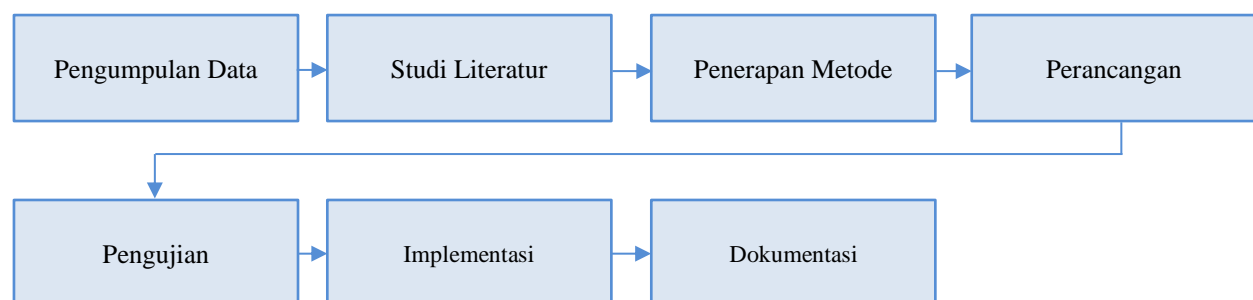
Penelitian terdahulu menjelaskan terhadap perbandingan antara metode KNN dengan Decision Tree, Naïve Bayes dalam menemukan sentiment masyarakat terhadap pelayanan penggunaan layanan BPJS dan hasil yang ditemukan dalam 1000 sampel data KNN mencapai 94%, Decision tree sebesar 96% dan naïve bayes mendapatkan hasil 89,14% [3].

Berdasarkan permasalahan dan penelitian terdahulu yang peneliti telah baca membuat peneliti tertarik menggunakan metode K-Nearest Neighbor sebagai metode yang sederhana, akurat dan banyak digunakan dalam penelitian terdahulu lainnya dalam menyelesaikan permasalahan Analisis Sentimen Nasabah Pada Kotak Komentar Terhadap Pelayanan Koperasi Simpan Pinjam Bangun Mandiri Menggunakan Metode K-Nearest Neighbor.

2. METODE PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian ini terdapat tahap-tapan penelitian yang saling berkorelasi serta menggambarkan antara variabel yang satu dengan variabel lainnya dibuat secara detail dan sistematis bertujuan untuk mencapai hasil yang maksimal dan akurat. Tahapan-tahapannya adalah sebagai berikut:



Gambar 1. Kerangka Kerja Penelitian

Kerangka kerja penelitian di atas dapat diuraikan masing-masing tahap dalam penelitian adalah sebagai berikut:

1. Studi Literatur

Tahapan studi literatur dilakukan untuk melihat penelitian terdahulu agar mengetahui penetapan tahapan dan penyelesaian permasalahan terbaik yang dapat dilakukan berdasarkan rekomendasi dan sumber-sumber terpercaya sebelumnya.

2. Identifikasi Masalah
Identifikasi masalah adalah proses mengenali masalah yang perlu diselesaikan. Proses identifikasi masalah salah satu proses mengidentifikasi masalah pada sistem keputusan pemilihan guru terbaik, dilakukan mulai dari pengamatan untuk mencari solusi dalam penyelesaian masalah hingga sampai selesai.
3. Pengumpulan Data
Pengumpulan data dilakukan dengan melihat komentar nasabah terhadap koperasi, pengumpulan data dilakukan untuk mempelajari permasalahan yang ada dalam tujuan menyelesaikan permasalahan dengan bantuan data yang tersedia.
4. Analisa
Analisa merupakan proses evaluasi setelah mendapatkan data informasi yang dibutuhkan.
5. Penerapan Metode
Tahapan ini, proses melakukan penerapan metode yang digunakan dalam penyelesaian masalah yang dihadapi pada penelitian ini. Metode yang digunakan pada masalah analisis sentiment nasabah ini menggunakan metode KNN.
6. Perancangan sistem
Pada tahapan perancangan sistem penulis menggunakan sistem *RapidMiner* untuk mempermudah menemukan hasil akurasi terhadap penelitian.
7. Pengujian
Tahapan ini, proses melakukan pengujian terhadap sistem yang telah tersedia.
8. Implementasi sistem
Implementasi sistem penulis akan melakukan penerapan terhadap sistem yang telah direncanakan sebelumnya.
9. Penulisan laporan
Penulisan laporan untuk bukti dan dokumentasi tetap bahwa penelitian telah dilaksanakan dengan baik..

2.2 Text Mining

Text mining dapat disebut juga penggalian atau pencarian data dalam bentuk teks yang berarti proses dimana dilakukannya pola ekstraksi baik dalam bentuk informasi maupun pengetahuan dan bermanfaat sebagai sumber data teks seperti dokumen yang disajikan dalam bentuk word, PDF, kutipan teks dan masih banyak contoh lainnya [4][5]. *Text mining* digunakan untuk melakukan klasifikasi, *clustering*, *information extraction* dan *information retrieval*. *Text mining* adalah ilmu pengetahuan dari data mining, yang merupakan lingkup penelitian yang komperhensif dan memiliki nilai komersial dibandingkan dengan data mining itu sendiri. Dalam *text mining*, melibatkan berbagai proses seperti melakukan kategorisasi teks dan melakukan ekstraksi baik dalam informasi maupun kata[6].

Text preprocessing merupakan tahap awal dalam sebuah *text mining*. Proses yang dilakukan dalam text preprocessing adalah pemilihan data, klasifikasi dan ekstraksi dengan tujuan untuk mengubah teks yang tidak terstruktur menjadi lebih terstruktur [7]. Ada beberapa tahapan dalam text preprocessing diantaranya [4]:

1. *Case Folding*
Case folding merupakan proses yang sederhana namun tetap efektif dan seringkali diabaikan. Tujuan dari proses *case folding* adalah mengubah semua teks menjadi *lower case*, menghapus angka, menghapus tanda baca, dan menghapus karakter kosong (*white space*).
2. *Tokenizing*
Tokenizing merupakan uraian kalimat menjadi kata-kata. Yang termasuk token ialah kata, angka, symbol, tanda baca dan entitas lainnya.
3. *Stopword Removal*
Stopword removal merupakan bagian dari *filtering* yang dimana tahap ini hanya mengambil kata-kata penting dan bermakna dari proses *tokenizing* dengan menggunakan algoritma *stoplist* untuk membuang ataupun menghilangkan kata yang kurang penting dan tidak memiliki makna dan menyimpan kata-kata yang penting.
4. *Term Frequency Inverse Document Frequency (TF – IDF)*
TF-IDF merupakan pembobotan yang dilakukan untuk mengekstraksi data pada teks dengan tujuan menentukan jumlah dari kata yang sudah diketahui dan setelah itu dikalikan dengan banyaknya *tweet* yang muncul. Berikut adalah rumus untuk menentukan pembobotan menggunakan TF-IDF.

$$Idf = \log \frac{N}{df_i} \dots\dots\dots (1)$$

2.3 Analisis Sentimen

Sentiment Analysis (SA) merupakan proses memahami dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimatatau teks yang berupa opini. Tujuan dilakukan sentiment analysis untuk melihat pandangan atau pendapat teks yang berkaitan terhadap sebuah masalah atau objek, apakah cenderung berpandangan *positif* atau *negatif*. *Sentiment Analysis* terdiri dari pemrosesan bahasa alami, analisis teks dan komputasi linguistik untuk mengidentifikasi sentimen dari suatu dokumen .

Sentiment Analysis dapat dibedakan berdasarkan sumber datanya, beberapa level yang sering digunakan dalam penelitian *Sentiment Analysis* adalah *Sentiment Analysis* pada level dokumen dan *Sentiment Analysis* pada level kalimat[8]. Berdasarkan level sumberdatanya *Sentiment Analysis* terbagi menjadi 2 kelompok besar yaitu :

1. *Coarse-grained Sentiment Analysis*
 Pada *Sentiment Analysis* jenis ini, *Sentiment Analysis* yang dilakukan adalah pada level dokumen. Secara garis besar fokus utama dari *Sentiment Analysis* jenis ini adalah menganggap seluruh isi dokumen sebagai sebuah *sentiment positif* atau *sentiment negatif*.
2. *Fined-grained Sentiment Analysis*
Fined-grained Sentiment Analysis adalah *Sentiment Analysis* pada level kalimat. Fokus utama *Fined-grained Sentiment Analysis* adalah menentukan sentiment pada setiap kalimat pada suatu dokumen, dimana kemungkinan yang terjadi adalah terdapat sentiment pada level kalimat yang berbeda pada suatu dokumen.

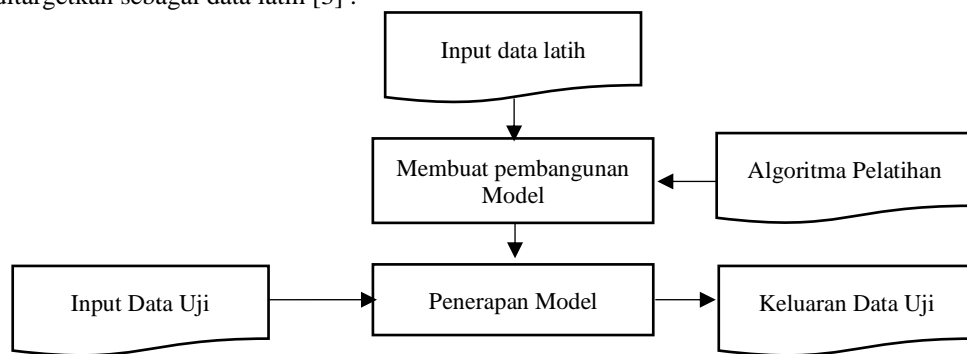
2.4 Data Mining dan Teknik Klasifikasi

Menurut Eko Prasetyo pada buku *Data Mining Konsep dan Aplikasi Menggunakan Matlab* mengatakan “Teknik data mining adalah bagaimana mencari data yang tersedia untuk menciptakan sebuah model, lalu memanfaatkan model tersebut untuk mengenali pola data lain yang tidak tersedia didalam basis data yang tersimpan [2]. Dalam data mining pengelompokan data dapat juga dilakukan untuk mengetahui pola secara universal dari data yang tersedia agar dilakukan langkah tindak lanjut lainnya yang berguna sebagai pendukung kegiatan dan tujuan akhir tertentu”

Dari pengertian diatas maka dapat diambil kesimpulan bahwasanya data mining merupakan sebuah proses penemuan serta pengenalan pola dan juga pengelompokkan data-data yang besar serta bertujuan untuk mendapatkan hubungan atau pola yang memberikan manfaat.

Teknik klasifikasi merupakan sebuah tata cara untuk mengelompokkan data menjadi tiap kelas-kelas dan merupakan salah satu teknik terpenting dari data mining dimana teknik ini melakukan pekerjaan untuk melakukan pembangunan terhadap data acak dalam sebuah model sebagai *prototype* untuk dijadikan *database*, disimpan sebagai memori lalu membangun sebuah model pada suatu objek data untuk mengetahui objek data tersebut bagian objek mana saja.[2].

Klasifikasi dapat dinyatakan sebagai teknik untuk memilih dan membaca data dengan cara mempelajari sekumpulan data yang menghasilkan data-data terbaru dimana data tersebut sebelumnya tidak pernah dipelajari, setelah itu data digali untuk menentukan kategori maupun kelas terhadap objek data. Membangun dan mengembangkan model klasifikasi biasanya didirikan berdasarkan pengetahuan para pakar/ahli, mengingat sangat banyak jumlah data yang akan dilakukan proses klasifikasi sehingga model ini juga diterapkan dalam bidang *machine learning*. Hal ini juga dikatakan sebagai proses belajar secara otomatis terhadap data atau suatu himpunan data yang telah ditargetkan sebagai data latihan [3] .



Gambar 2. Proses Kerja Klasifikasi

Berdasarkan kinerja dan cara pengukuran hasil kerja klasifikasi digunakan di antaranya : Matriks Konfusi.

Matriks Konfusi adalah matrik atau tabel yang mencatat hasil kerja klasifikasi, seperti tabel II-1 merupakan contoh matriks konfusi yang melakukan klasifikasi masalah biner hanya dua kelas, yaitu kelas 0 dan kelas 1. Setiap sel f_{ij} dalam matriks menyatakan jumlah record/data dari kelas i yang hasil prediksinya masuk ke kelas j .

Tabel 1. Matriks Konfusi Untuk Dua Kelas

f_{ij}		Kelas Hasil Prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas Asli (i)	Kelas = 1	F_{11}	F_{10}
	Kelas = 0	F_{01}	F_{00}

Mengetahui jumlah data yang diklasifikasi dengan, dapat mengetahui akurasi hasil klasifikasi, dengan mengetahui jumlah data yang diklasifikasikan secara salah, digunakan rumus formula akurasi sebagai berikut ini:

$$Akurasi = \frac{Jumlah\ Data\ Yang\ Diprediksi\ Benar}{Jumlah\ Prediksi\ Yang\ Dilakukan} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}} \dots\dots\dots (1)$$

Mengetahui laju erorr digunakan formula :

$$Akurasi = \frac{Jumlah\ Data\ Yang\ Diprediksi\ Salah}{Jumlah\ Prediksi\ Yang\ Dilakukan} = \frac{f_{10}+f_{01}}{f_{11}+f_{10}+f_{01}+f_{00}} \dots\dots\dots (2)$$

Semua algoritma klasifikasi berusaha membentuk model yang mempunyai akurasi tinggi dan laju erorr yang rendah. Umumnya, model yang dibangun dapat memprediksi dengan benar semua data yang menjadi data latihnya.

2.5 K-Nearest Neighbor (KNN)

K-Nearest Neighbor(KNN) merupakan salah satu dari jenis algoritma data mining yang tergolong dalam klasifikasi *supervised learning* dimana data yang sudah tersedia dilakukan proses pengenalan kedekatan antara kasus baru dengan kasus lama berdasarkan pada pencocokan bobot yang dimiliki oleh atribut[4].

Algoritma KNN menggunakan tahapan sebagai berikut ini[4]:

1. Menentukan parameter K, parameter ini nantinya akan menjadi bahan utama dalam batasan yang akan dilakukan pencarian dengan menggunakan metode ini, agar penelitian ini memiliki pondasi awal dari gambaran hal yang ingin ditemukan.
2. Menghitung kuadrat jarak *euclidian* masing-masing obyek terhadap data sampel yang diberikan.
3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *euclid* terkecil.
4. Mengumpulkan kategori Y.
5. Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung.

$$Dij = \sqrt{(x_{ii} - x_{ji})^2 + x_{2i}^2 - x_{2j}^2 + \dots (x_{ii} - x_{ji})^2} \dots\dots\dots (3)$$

Keterangan :

- Dij = Jarak dari data ke i ke pusat cluster i
- Xxi = Data dari data ke- i keattribute data data ke-K
- Xkj = Data dari data ke- j keattribute data data ke-K

3. HASIL DAN ANALISIS

3.1 Analisis dan Penerapan Metode

Pada analisa menjelaskan secara keseluruhan sistem yang ada dalam mengetahui sentiment nasabah pada kotak komentar, sebelum melakukan proses analisa sentiment data yang telah ada dilakukan *preprocessing data* yaitu dengan tahapan penentuan *query* dengan melakukan normalisasi, setelah dilakukan normalisasi terhadap 9 data yang ada dibentuk dengan memilah dokumen sehingga membentuk angka yang dapat diproses dalam menggunakan metode *K-Nearest Neighbor*. Pada penelitian ini metode *K-Nearest Neighbor* digunakan pada penerapan aplikasi *RapidMiner* untuk mendapatkan nilai akurasi dari pengujian sentiment nasabah terhadap pelayanan koperasi.

Text Preprocessing terdapat beberapa tahapan yang merubah data yang tidak resmi menjadi data dan kosa kata yang resmi dan mudah dimengerti, pada penenggunaan *Text Preprocessing* digunakan agar mempermudah dalam membaca data untuk mengelompokan data dalam hasil negatif maupun positif. Adapun proses penerapannya sebagai berikut ini

Tabel 2. Data Nasabah

No	Komentar	Sentiment
1	#Staff Pelayanan sombong!!!!!!	Negatif
2	pelayanan Sangat nyaman x dan mudah dimengerti	Positif
3	Pelayanan yg sabar :) dan jujur	Positif
4	**koperasi yg buruk & memeras pelanggan **	Negatif
5	Pelayanan sangat lama kali dan tidak professional !!	Negatif
6	Staf ramah dengan orang tua dan memberi pengarahan 😊	Positif
7	Staf koperasi sangat sportif tapi sangat arogan	Negatif
8	Staf yg sangat ramah dan sabar	Positif
9	Koperasi memiliki fasilitas yg bagus & menarik	Positif
10	Koperasi terbaik tapi sangaat lama x pelayanannya	???

Sumber: Koperasi Simpan Pinjam Bangun Mandiri,2023

Data yang masih merupakan keterangan ada pada kriteria tabel Jaminan sehingga harus dirubah menjadi angka, Adapun nilai yang dimiliki masing-masing dibentuk dengan menggunakan tahapan *Query*. Adapun tahapan tersebut dikenal sebagai tahapan “pembobotan kata” sebagai berikut ini:

1. Mengumpulkan data yang tersedia dan memiliki label apakah positif atau negatif.
2. *Case Folding*
 Tahapan ini merubah semua data yang memiliki tulisan huruf besar atau kapital menjadi huruf kecil semuanya. Adapun data pada *Case folding* sebagai berikut ini

Tabel 3. Hasil tahapan *Case Floding*

No	Komentar	Sentiment
1	#Staff Pelayanan Sombong!!!!!!	negatif
2	Pelayanan Sangat Nyaman X Dan Mudah Dimengerti	positif
3	Pelayanan Yg Sabar :) Dan Jujur	positif
4	**Koperasi Yg Buruk & Memeras Pelanggan **	negatif
5	Pelayanan Sangat Lama Kali Dan Tidak Professional !!	negatif
6	Staf Ramah Dengan Orang Tua Dan Memberi Pengarahan ☺	positif
7	Staf Koperasi Sangat Sportif Tapi Sangat Arogan	negatif
8	Staf Yg Sangat Ramah Dan Sabar	positif
9	Koperasi Memiliki Fasilitas Yg Bagus & Menarik	positif
10	Koperasi Terbaik Tapi Sangaat Lama X Pelayanannya	???

3. *linguistic preprocessing: stopword dan stemming*
 Pada tahapan ini dilakukan penghapusan kata penghubung, kata bersifat pemberhentian, penghubung, imbuhan dan lain-lainnya seperti *symbol emoticon*, titik, koma, dan, yang, untuk, sehingga meninggalkan kalimat utama yang baik dan benar sesuai kamus. Adapun tahapan ini menghasilkan data sebagai berikut ini:

Tabel 4. hasil tahapan *linguistic preprocessing: stopword dan stemming*

No	Komentar	Sentiment
1	staff pelayanan sombong	negatif
2	pelayanan sangat nyaman mudah mengerti	positif
3	pelayanan sabar jujur	positif
4	koperasi buruk memeras pelanggan	negatif
5	pelayanan sangat lama tidak profesional	negatif
6	staf ramah orang tua memberikan pengarahannya	positif
7	staf koperasi sangat sportif sangat arogan	negatif
8	staf sangat ramah sabar	positif
9	koperasi memiliki fasilitas bagus menarik	positif
10	koperasi terbaik tetapi sangat lama pelayanan	???

4. Pemisahan kata *tokenization*
 Pada tahapan pemisahan kata *tokenization* yang bersifat memisahkan setiap potongan kalimat menjadi satu bagian. Sehingga hasil dari data sampel yang dilakukan *tokenization* menjadi tabel berikut ini:

Tabel 5. Hasil Tahapan *Tokenization*

Dok 1	Dok2	Dok 3	Dok 4	Dok 5	Dok 6	Dok 7	Dok 8	Dok 9	Dok 10
staf	pelayanan	pelayanan	koperasi	pelayanan	staf	staf	staf	koperasi	koperasi
pelayanan	sangat	sabar	buruk	sangat	ramah	koperasi	sangat	memiliki	terbaik
sombong	nyaman	jujur	memeras	lama	orang	sangat	ramah	fasilitas	tetapi
	mudah		pelanggan	tidak	tua	sportif	sabar	bagus	sangat
	mengerti			profesional	memberikan	sangat		mearik	lama
					pengarahan	arogan			pelayanan

5. Normalisasi
 Pada tahapan ini dilakukan pembuangan kata yang berlebihan digantikan dengan kata yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI):

Tabel 6. Normalisasi

Dok 1	Dok2	Dok 3	Dok 4	Dok 5	Dok 6	Dok 7	Dok 8	Dok 9	Dok 10
staf	layan	layan	koperasi	layan	staf	staf	staf	koperasi	koperasi
layan	nyaman	sabar	buruk	lama	ramah	koperasi	ramah	milik	baik

sombong	mudah	jujur	peras	tidak	orang	jujur	sabar	fasilitas	tetapi
	erti		langgan	profesional	tua	sombong		bagus	lama
					beri			tarik	layan
					arah				

6. Proses *Indexing*

Tahapan *indexing* yaitu mendeteksi berapa kali kata yang sama muncul dalam dokumen, tahapan ini membuat kata yang muncul berulang kali dimasing-masing dokumen dibentuk menjadi satu kata yang muncul serta dihitung jumlah kemunculan kata berapa kali. Adapun tabel *Indexing* dapat dilihat pada tabel berikut ini:

Tabel 7. Proses *Indexing*

Kosa kata	(Dokumen : Frekuensi)					
staf	1:1	6:1	7:1	8:1		
layan	1:1	2:1	3:1	5:1	10:1	
sombong	1:1	7:1				
nyaman	2:1					
mudah	2:1					
erti	2:1					
sabar	3:1	8:1				
jujur	3:1	7:1				
koperasi	4:1	7:1	9:1	10:1		
buruk	4:1					
peras	4:1					
langgan	4:1					
lama	5:1	10:1				
tidak	5:1					
professional	5:1					
ramah	6:1	8:1				
orang	6:1					
tua	6:1					
beri	6:1					
arah	6:1					
milik	9:1					
fasukitas	9:1					
bagus	9:1					
tarik	10:1					
tetapi	10:1					

Pada data di atas dibentuk berdasarkan kata dalam dokumen berapa muncul.

7. Pembobotan Kata

Pembobotan kata dilihat berdasarkan *indexing* tabel sehingga hasil pembobotan sebagai berikut ini:

Tabel 8. Pembobotan Kata

Kosa Kata	Dok 1	Dok 2	Dok 3	Dok 4	Dok 5	Dok 6	Dok 7	Dok 8	Dok 9	Dok 10	DF
staf	1	0	0	0	0	1	1	1	0	0	3
layan	1	1	1	0	1	0	0	0	0	1	5
sombong	1	0	0	0	0	0	1	0	0	0	2
nyaman	0	1	0	0	0	0	0	0	0	0	1
mudah	0	1	0	0	0	0	0	0	0	0	1
erti	0	1	0	0	0	0	0	0	0	0	1
sabar	0	0	1	0	0	0	0	1	0	0	2
jujur	0	0	1	0	0	0	1	0	0	0	2
koperasi	0	0	0	1	0	0	1	0	1	1	4

buruk	0	0	0	1	0	0	0	0	0	0	1
peras	0	0	0	1	0	0	0	0	0	0	1
langgan	0	0	0	1	0	0	0	0	0	0	1
lama	0	0	0	0	1	0	0	0	0	1	2
tidak	0	0	0	0	1	0	0	0	0	0	1
professional	0	0	0	0	1	0	0	0	0	0	1
ramah	0	0	0	0	0	1	0	1	0	0	2
orang	0	0	0	0	0	1	0	0	0	0	1
tua	0	0	0	0	0	1	0	0	0	0	1
beri	0	0	0	0	0	1	0	0	0	0	1
arah	0	0	0	0	0	1	0	0	0	0	1
milik	0	0	0	0	0	0	0	0	1	0	1
fasukitas	0	0	0	0	0	0	0	0	1	0	1
bagus	0	0	0	0	0	0	0	0	1	0	1
tarik	0	0	0	0	0	0	0	0	0	1	1
tetapi	0	0	0	0	0	0	0	0	0	1	1

Tahapan selanjutnya dicari nilai kesamaan tetangga atau *K-nearest neighbor* dengan rumus *Cosine Similarity*, pada tahapan ini diketahui total dari kasus lama 9 dan kasus baru yang akan dicari tahu hasil keputusannya 1 sehingga total $n = 10$, rumus dari *Cosine Similarity* dilakukan dengan rumus $\text{Log}(n/df)$ sehingga data pada tabel 4.6 yang telah diterapkan menggunakan rumus $\text{log}(n/df)$ dapat dilihat pada tabel 9 berikut ini:

Tabel 9. Hasil Nilai IDF

Kosa Kata	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	DF	IDF
staf	1	0	0	0	0	1	1	1	0	0	3	0.522879
layan	1	1	1	0	1	0	0	0	0	1	5	0.30103
sombong	1	0	0	0	0	0	1	0	0	0	2	0.69897
nyaman	0	1	0	0	0	0	0	0	0	0	1	1
mudah	0	1	0	0	0	0	0	0	0	0	1	1
erti	0	1	0	0	0	0	0	0	0	0	1	1
sabar	0	0	1	0	0	0	0	1	0	0	2	0.69897
jujur	0	0	1	0	0	0	1	0	0	0	2	0.69897
koperasi	0	0	0	1	0	0	1	0	1	1	4	0.39794
buruk	0	0	0	1	0	0	0	0	0	0	1	1
peras	0	0	0	1	0	0	0	0	0	0	1	1
langgan	0	0	0	1	0	0	0	0	0	0	1	1
lama	0	0	0	0	1	0	0	0	0	1	2	0.69897
tidak	0	0	0	0	1	0	0	0	0	0	1	1
professional	0	0	0	0	1	0	0	0	0	0	1	1
ramah	0	0	0	0	0	1	0	1	0	0	2	0.69897
orang	0	0	0	0	0	1	0	0	0	0	1	1
tua	0	0	0	0	0	1	0	0	0	0	1	1
beri	0	0	0	0	0	1	0	0	0	0	1	1
arah	0	0	0	0	0	1	0	0	0	0	1	1
milik	0	0	0	0	0	0	0	0	1	0	1	1
fasukitas	0	0	0	0	0	0	0	0	1	0	1	1
bagus	0	0	0	0	0	0	0	0	1	0	1	1
tarik	0	0	0	0	0	0	0	0	0	1	1	1
tetapi	0	0	0	0	0	0	0	0	0	1	1	1
Total	3	4	3	4	4	6	4	3	4	5	39	-
Q	0	0.301	0	0.39	0	0	0.39	0	0.39	0		

Setelah diketahui nilai IDF selanjutnya masing-masing nilai baris atau nilai “*tf*” dikalikan dengan nilai IDF yang diblok dengan warna biru, setelah didapatkan nilai Wdt atau nilai bobot masing-masing kosa kata tersebut selanjutnya dan hasil akhir di jumlahkan secara kolom. Adapun hasil dapat dilihat pada tabel berikut ini:

Tabel 10. Hasil Nilai Wdt

Kosa kata	Wdt= <i>tf</i> * <i>idf</i>									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
staf	0.5229	0	0	0	0	0.5229	0.5229	0.5229	0	0
layan	0.301	0.301	0.301	0	0.301	0	0	0	0	0.301
sombong	0.699	0	0	0	0	0	0.699	0	0	0
nyaman	0	1	0	0	0	0	0	0	0	0
mudah	0	1	0	0	0	0	0	0	0	0
erti	0	1	0	0	0	0	0	0	0	0
sabar	0	0	0.699	0	0	0	0	0.699	0	0
jujur	0	0	0	0	0	0	0	0	0	0
koperasi	0	0	0	0.3979	0	0	0.3979	0	0.3979	0.3979
buruk	0	0	0	1	0	0	0	0	0	0
peras	0	0	0	1	0	0	0	0	0	0
langgan	0	0	0	1	0	0	0	0	0	0
lama	0	0	0	0	0.699	0	0	0	0	0.699
tidak	0	0	0	0	1	0	0	0	0	0
professional	0	0	0	0	1	0	0	0	0	0
ramah	0	0	0	0	0	0.699	0	0.699	0	0
orang	0	0	0	0	0	1	0	0	0	0
tua	0	0	0	0	0	1	0	0	0	0
beri	0	0	0	0	0	1	0	0	0	0
arah	0	0	0	0	0	1	0	0	0	0
milik	0	0	0	0	0	0	0	0	1	0
fasukitas	0	0	0	0	0	0	0	0	1	0
bagus	0	0	0	0	0	0	0	0	1	0
tarik	0	0	0	0	0	0	0	0	0	1
tetapi	0	0	0	0	0	0	0	0	0	1
Total	1.5229	3.301	1	3.3979	3	5.2219	1.6198	1.9209	3.3979	3.3979

Nilai data yang akan uji ada pada data dokumen ke 10, Adapun kata yang dipakai dalam dokumen ke 10 untuk melihat apakah data komentar bersifat positif atau negatif, kata yang ada pada dokumen ke 10 yaitu kalimat “koperasi baik tetapi lama layan” sehingga nilai yang diperoleh untuk dihitung ke tahap melihat kedekatan nilai ada pada tabel sebagai berikut ini:

Tabel 11. Nilai Kesimpulan Data

Data Komentar	IDF	Q	WDT
D1	3	0	1.5229
D2	4	0.301	3.301
D3	3	0	1
D4	4	0.39	3.3979
D5	4	0	3
D6	6	0	5.2219
D7	4	0.39	1.6198
D8	3	0	1.9209
D9	4	0.39	3.3979
D10	5	0	3.3979

Pada tabel 11 ditemukan berdasarkan total masing-masing nilai IDF dan nilai Q yang merupakan nilai dari data dokumen 10 ditemukan dari data komentar yang memiliki kesamaan kemunculan kata pada dokumen lain

terdapat pada tabel 10, dan nilai WDT terdapat pada total nilai pada tabel 4.8, selanjutnya diterapkan rumus *Cosine Similarity* untuk menghitung kemiripan nilai dokumen ke-10 dengan dokumen D1,D2,D3,D4,D5,D6,D7,D8,D9,

Tahapn rumus ini dilakukan dengan cara nilai $D_n = Q / (IDF * WDT)$ sebagai berikut ini:

1. $D1 = 0 / (3 \times 1,5229) = 0$
2. $D2 = 0,301 / (4 \times 3,301) = 0,022796$
3. $D3 = 0 / (4 \times 1) = 0$
4. $D4 = 0,39 / (4 \times 1) = 0,028694$
5. $D5 = 0 / (6 \times 3,3979) = 0$
6. $D6 = 0 / (4 \times 3) = 0$
7. $D7 = 0,39 / (3 \times 5,2219) = 0,060193$
8. $D8 = 0 / (4 \times 1,6198) = 0$
9. $D9 = 0,39 / (3 \times 3,3979) = 0,028694$
10. $D10 = 0 / (5 \times 3,3979) = 0$

Sehingga hasil nilai dari perhitungan ditemukan kemiripan dan kedekatan nilai sebagai berikut ini:

Tabel 12. Hasil Kedekatan Nilai

No	Komentar	Sentimen	Nilai Kedekatan
D1	Staff Pelayanan sombong	Negatif	0
D2	Pelayanan sangat nyaman dan mudah dimengerti	Positif	0.022796
D3	Pelayanan yang sabar dan jujur	Positif	0
D4	Koperasi yang buruk dan memeras pelanggan	Negatif	0.028694
D5	Pelayanan sangat lama dan tidak profesional	Negatif	0
D6	Staf ramah dengan orang tua dan memberi pengarahan	Positif	0
D7	Staf koperasi sangat sportif tetapi sangat arogan	Negatif	0.060193
D8	Staf yang sangat ramah dan sabar	Positif	0
D9	Koperasi memiliki fasilitas yang bagus dan menarik	Positif	0.028694
D10	Koperasi terbaik tetapi sangat lama pelayanannya	???	

Setelah diketahui nilai jarak atau kedekatan dari masing-masing sentimen selanjutnya berdasarkan metode KNN memiliki parameter K-3 sebagai pembatas kedekatan nilai tetangga dengan *query* terdapat nilai yang berdekatan pada komentar D2 nilai “0.022796” dengan sentimen “Positif”, D4 nilai “0.028694” dengan sentiment “Negatif” dan D9 nilai “0.028694” dengan sentimen “Positif”.

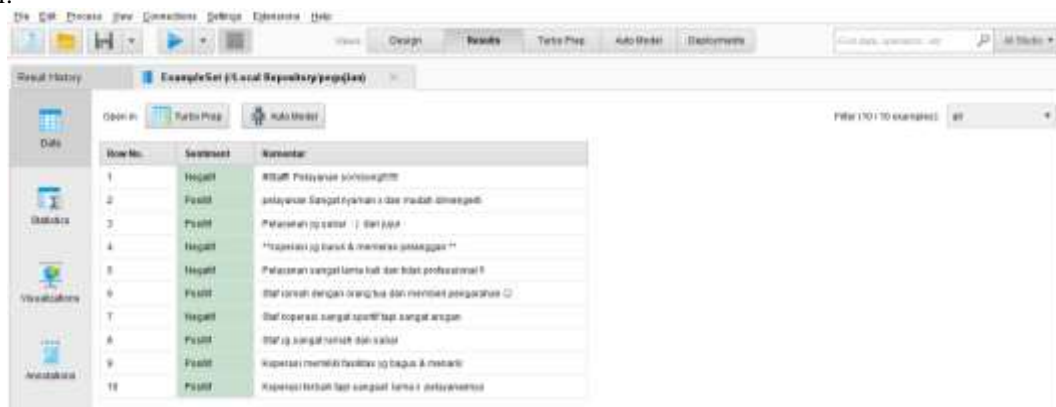
Pada hasil sentiment dengan parameter K-3 yaitu positif =2 dan negatif =1 , karena terdapat lebih banyak pernyataan postifi pada kedekatan nilai sehingga data D10 dinyatakan sentimen berpotensi “**positif**”

Tabel 13. Hasil Sentimen

No	Komentar	Sentimen
D10	Koperasi terbaik tetapi sangat lama pelayanannya	Positif

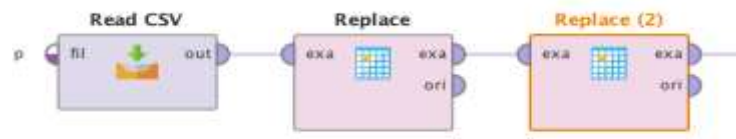
3.2 Implementasi

Penelitian ini data diolah menggunakan *RapidMiner versi 9.10.8*. Pengolahan data *K-Nearest Neighbor* pada *RapidMiner* menggunakan data pada kriteria sementara keputusan berupa data keterangan sehingga data sebagai berikut ini:



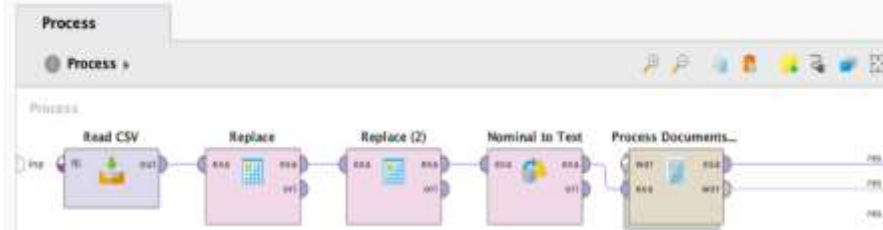
Gambar 2. Data Yang Analisis

Text Preprocessing terdapat beberapa proses salah satunya ialah proses *replace*. Disini terdapat dua kali proses *replace* yaitu menghilangkan tulisan simbol. *Replace* selanjutnya menghilangkan data linl. Berikut ini adalah *tools replace* yang digunakan pada *RapidMiner*:



Gambar 3. Replace dan Replace (2)

Berikut ini adalah proses tahapan pemodelan *preprocessing* pada *tools RapidMiner* yang digunakan :



Gambar 4. Pre-processing

Pada gambar diatas merupakan alur *preprocessing* yang berisikan *process document to data* yang terdapat sub bagian yang memiliki beberapa proses lain didalamnya untuk mendapatkan hasil pembobotan data.

a. *Tokenize*

Tokenize merupakan suatu proses untuk memisahkan antar kata komentar. Proses *tokenize* menggunakan karakter spasi sebagai pemisah katanya.



Gambar 5. Tokenize

b. *Transform Cases*

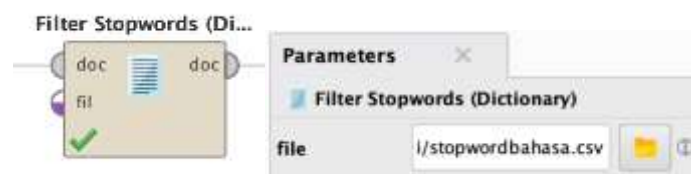
Transform cases adalah proses mengubah semua huruf pada teks sesuai dengan pilihan pengguna yaitu menjadi huruf kecil ataupun kapital. Proses *transform cases* pada penelitian semuanya diubah menjadi huruf kecil karena berisikan opini.



Gambar 6. Transform Cases & Parameters

c. *Filter Stopwords (Dictionary)*

Filter stopwords berfungsi untuk menghilangkan teks yang tidak memiliki hubungan dengan penelitian analisis sentimen sehingga dimensi dari teks tersebut berkurang tanpa mengurangi isi teks tersebut. File *stopwordbahasa.csv* diperoleh dari <https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist> merupakan dataset dari Kaggle dengan judul Indonesia Stoplist.



Gambar 7. Filter Stopwords dan Parameters

d. *Filter Tokens (by Length)*

Filter tokens (by length) adalah proses memfilter token berdasarkan panjang teks dan karakter yang dikandung didalamnya. Tujuan dari proses ini adalah untuk membuat pemilihan token dengan minimal ukuran panjang 4 karakter dan maksimal ukuran panjang 25 karakter.



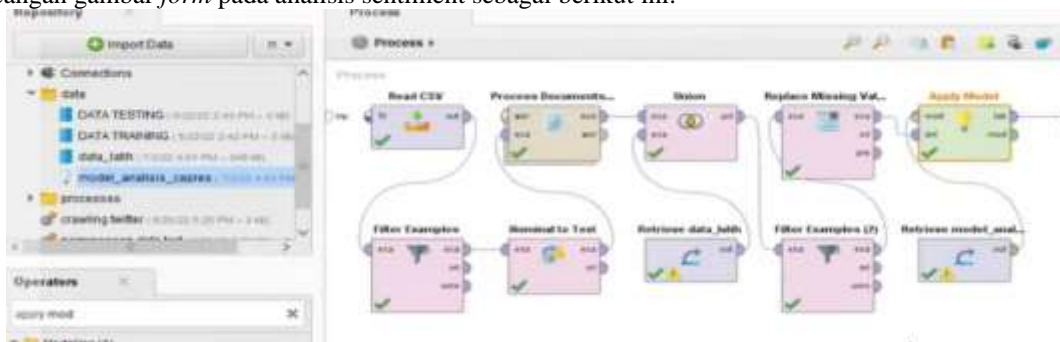
Gambar 8. Filter Tokens dan Parameters

e. Transformation

Setelah menyelesaikan proses *pre-processing* maka langkah selanjutnya adalah *transformation*. Transformation sendiri adalah sebuah proses perubahan dokumen kedalam bentuk numerik dengan memberikan pembobotan pada setiap kata dalam data. Metode pembobotan yang digunakan adalah metode TF-IDF yang dilakukan secara otomatis pada *RapidMiner*.

3.3 Pengujian Aplikasi Rapid Miner

Pada tahapan perancangan penggunaan metode K-NN dalam analisis data menerapkan tahapan yang cukup Panjang pada tahapan *preprocessing* data, pada tahapan tersebut dilakukan proses pengubahan data keterangan komentar menjadi data angka, setelah semua data berubah menjadi nominal dilanjutkan dengan menghubungkan algoritma k-NN dan Apply model dan dilakukan proses *Run (Play)* untuk melihat hasil prediksi. Adapun tampilan rancangan gambar *form* pada analisis sentiment sebagai berikut ini:



Gambar 9. Form Pengujian sentiment menggunakan K-NN

3.4 Hasil Akurasi Metode

Hasil akhir merupakan hasil yang tampil setelah dilakukan *design proses* dan digunakan tombol *play* dalam proses hasilnya. Pada pengujian ini yang menggunakan metode perbandingan dilakukan dua tahap design proses pada ikon penerapan metode dan memiliki dua hasil juga, Adapun hasil pada penelitian sebagai berikut ini:

	True Negatif	True Positif	class predisi
pred. Negatif	1	0	100.00%
pred. Positif	3	8	95.00%
class recall	20.00%	100.00%	

Gambar 10. Hasil Akurasi Nearest Neighbor Pada RapidMine

Row No.	Nombre	predicted...	confidence...	confidence...	Message
1	Negatif	Positif	0.999	0.999	Model Prevision: negatif
2	Positif	Positif	0.999	0.999	Model Prevision: positif
3	Positif	Positif	0.975	0.975	Model Prevision: positif
4	Negatif	Positif	0.999	0.977	Model Prevision: positif
5	Negatif	Positif	0.977	0.999	Model Prevision: positif
6	Positif	Positif	0.999	0.999	Model Prevision: positif
7	Negatif	Positif	0.999	0.999	Model Prevision: positif
8	Positif	Positif	0.999	0.999	Model Prevision: positif
9	Positif	Positif	0.999	0.977	Model Prevision: positif
10	Positif	Positif	0.999	0.977	Model Prevision: positif

Gambar 11. Hasil Apply Model Nearest Neighbor Pada RapidMine

Hasil dari kedua penelitian yang diterapkan dalam rapidminer terdapat tingkat akurasi 100% dari kedua metode, data training data dan data uji terhadap komentar yang dilontarkan menghasilkan sentimen yang baru. Hasil prediksi menunjukkan akurasi penelitian sebanyak 70% akurat sehingga penggunaan metode ini dalam menganalisa sentimen menghasilkan akurasi yang cukup tinggi.

Hasil menunjukkan sentimen nasabah terhadap koperasi sebanyak 25% negatif dimana pernyataan komentar terdapat kata positif sebanyak 3 dokumen tetapi menghasilkan sedikit komentar negatif dan sebanyak 100% komentar positif dengan data positif semua benar-benar tidak terselip kata negatif, dan terdapat data komentar positif sebanyak 66.67%

4. KESIMPULAN

Hasil dari memprediksi potensi nasabah kredit Tidak maka dapat disimpulkan bahwa:

1. Prosedur dalam menerapkan penujian sentiment nasabah terhadap pelayanan koperasi yaitu semua data komentar diterapkan tahapan *text preprocessing* dengan tujuan menemukan data dan kata yang tepat dan benar sesuai Kamus Besar Bahasa Indonesia (KBBI) sehingga mempermudah menemukan kalimat sentimen positif dan negatif
2. Penerapan metode *K-Nearest Neighbor* digunakan dalam pengujian *RapidMiner* sebagai operator yang mengolah data dan memperhitungkan setiap kemungkinan prediksi terhadap komentar nasabah.
3. Penerapan *Rapidminer versi 9.10.8* mendapatkan hasil akurasi sebesar 70% persen secara keseluruhan dengan peringkat data negatif yang memiliki unsur komentar positif tetapi hasil negative sebanyak 25% dan data komentar positif terhadap pelayanan koperasi *pure* positif sebanyak 100% dengan data komentar positif sebanyak 66.67%.

REFERENSI

- [1] R. Rasenda, H. Lubis, and R. Ridwan, "Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 4, no. 2, p. 369, Apr. 2020. [[Available](#)]
- [2] A. Deviyanto and M. D. R. Wahyudi, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," JISKA (Jurnal Informatika Sunan Kalijaga), vol. 3, no. 1, p. 1, Dec. 2018. [[Available](#)]
- [3] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," Jurnal Informatika Universitas Pamulang, vol. 5, no. 4, p. 646, Dec. 2021. [[Available](#)]
- [4] E. Buulolo, *Buku Data Mining Untuk Perguruan Tinggi*, I. 2020.
- [5] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, and J. Li, "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis," *Energy and Built Environment*, vol. 1, no. 2, pp. 149–164, Apr. 2020. [[Available](#)]
- [6] H. Yan, N. Yang, Y. Peng, and Y. Ren, "Data mining in the construction industry: Present status, opportunities, and future trends," *Automation in Construction*, vol. 119. Elsevier B.V., p. 103331, Nov. 01, 2020. [[Available](#)]
- [7] Z. Wang, Y. Li, D. Li, Z. Zhu, and W. Du, "Entropy and gravitation based dynamic radius nearest neighbor classification for imbalanced problem," *Knowl Based Syst*, vol. 193, no. xxxx, p. 105474, 2020. [[Available](#)]
- [8] M. I. Maulana and A. A. Soebroto, "Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi- square," vol. 3, no. 7, pp. 6662–6669, 2019.