

# Klasifikasi Opini Masyarakat Pada Facebook Terhadap Operator Seluler Menggunakan Metode K-Nearest Neighbor Dan Support Vector Machine

M. Bagus Abrianto<sup>1</sup>, Hary Junaidi Hatman<sup>2</sup>, mirfan<sup>3</sup>, Michael Oktavianus<sup>4</sup>

<sup>1</sup>S1 sistem informasi, Universitas DIPA Makassar, Indonesia

<sup>2</sup>S1 sistem informasi, Universitas DIPA Makassar, Indonesia

<sup>1</sup>aditjunaidi440@gmail.com, <sup>2</sup>bagusabrianto30@gmail.com, <sup>3</sup>mirfan@undipa.ac.id

<sup>4</sup>michaeloktavianusdipa@gmail.com

## ABSTRAK

*Dalam beberapa tahun terakhir jumlah pelanggan yang dilayani oleh operator seluler meningkat secara signifikan pada operator seluler Telkomsel, Indosat dan XL. Oleh karena itu media sosial menjadi media pilihan banyak orang untuk menyampaikan pesannya dan operator seluler telkomsel, indosat dan xl memilih Facebook dengan menggunakan metode klasifikasi K-Nearest Neighbors (KNN) dan metode Support Vector Machine (SVM). Hasil analisis menunjukkan bahwa implementasi metode SVM memiliki akurasi yang lebih baik dalam mengklasifikasikan sentimen pengguna terhadap operator seluler. Analisis sentimen mengungkapkan bahwa sentimen negatif mendominasi dibandingkan dengan sentimen positif pada setiap operator. Hal ini mengindikasikan adanya kecenderungan masyarakat yang masih merasa ragu terhadap kualitas layanan operator seluler, Telkomsel, Indosat, dan XL, mungkin akibat dari berbagai faktor risiko dan kurangnya kepercayaan.*

**Kata kunci:** klasifikasi opini masyarakat, facebook, K-Nearest Neighbor, Support Vector Machine

## I. PENDAHULUAN

Operator seluler sangat penting dalam kehidupan sehari-hari karena mereka menyediakan layanan data seperti internet. Dalam beberapa tahun terakhir, jumlah pelanggan yang dilayani oleh operator seluler meningkat secara signifikan, yaitu operator seluler Telkomsel, Indosat dan XL.

Tiga operator seluler tersebut merupakan perusahaan operator terbesar di Indonesia dan menyediakan layanan komunikasi nirkabel melalui jaringan seluler, menawarkan berbagai layanan seperti panggilan suara dan pesanteks, serta layanan data seperti Internet dan video streaming. Oleh karena itu, media sosial menjadi media pilihan banyak orang untuk menyampaikan pesannya dan operator seluler telkomsel, indosat dan xl memilih Facebook karena setiap pengguna Facebook bebas mengunggah postingan atau komentar tanpa ada batasan.

Namun informasi yang didapat dari komentar Facebook belum dapat mengidentifikasi opini publik baik atau buruk, sehingga data jumlah komentar di halaman Facebook Telkomsel, Indosat, dan XL perlu dikumpulkan,

Data dapat dikumpulkan melalui Facebook Dengan menggunakan Teknik crawling teks data tersebut akan ditarik dengan menggunakan ekspor komen yang ada di website, jumlah target data yang akan ditarik minimal 1000 data, Jika data sudah ditarik maka akan dikumpulkan ke format excel menjadi dataset.

Setelah data sudah di kumpulkan maka data itu selanjutnya akan dicross validation memakai rapid miner untuk mengevaluasi mana data yang akan dijadikan data training dan data testing setelah itu data tersebut akan dievaluasi dengan metode K-Nearest Neighbor (KNN) dan support vector machine (SVM). Pada tahap akhir, data tersebut akan dievaluasi dengan menerapkan perintah *confusion matrix* pada Rapid Miner untuk mengecek tingkat keakuratan data.

Diharapkan penelitian dengan menggunakan metode K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM) guna mendapatkan pemahaman yang lebih baik tentang pandangan dan sentimen masyarakat terhadap layanan operator seluler agar masyarakat dapat memilih operator mana yang lebih baik dan dapat bertahan sampai beberapa tahun kedepan.

## II. TINJAUAN PUSTAKA

### A. Landasanteori

Salah satu tugas yang dapat dilakukan dengan data mining adalah pengklasifikasian. Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasi suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (ataudikenaldengannama Carolus Linnaeus) yang pertama kali mengklasifikasikan spesies berdasarkan karakteristik fisik. Selanjutnyadiadikenal sebagai bapak klasifikasi[1]. Algoritma atau metode K-NN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang diuji. Pengklasifikasian kerabat terdekat (*Nearest Neighbor*) didasarkan pada pembelajaran dengan analogi, yaitu dengan membandingkan data yang dites dengan kumpulan data pelatihan yang mirip dengan itu[2]. Sumber data yang digunakan pada *Text mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Tujuan dari *Text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen[3]. Telekomunikasi telah menjadi kebutuhan yang tidak dapat dipisahkan bagi masyarakat luas saat ini. Perkembangan teknologi yang begitu pesat ini menyebabkan pergeseran kebutuhan terhadap media telekomunikasi dari sekunder atau bahkan tersier menjadi kebutuhan primer. Jika dulu telepon seluler menjadi barang mewah konsumsi kelas menengah keatas, sekarang hampir seluruh elemen kelas masyarakat telah memiliki telepon seluler sebagai bagian dari kebutuhan dan gaya hidup[4]. *Facebook* adalah media sosial yang populer saat ini. Sebagai media sosial, *Facebook* menjadi media yang digunakan untuk berkomunikasi. Media sosial tersebut memberikan berbagai kemudahan dalam melakukan komunikasi. *Facebook* memberikan manfaat dalam menjalin komunikasi dengan berbagai kalangan. Fitur yang disediakan *Facebook* untuk berkomunikasi adalah perbaruan status[5]. *Rapidminer* adalah salah satu *Software* untuk pengolahan data mining. Pekerjaan yang dilakukan oleh *Rapidminer Text* mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari dataset yang besar dan

mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database[6].

### B. Penelitian Terkait

Pada penelitian ini dilakukan penelitian mengenai analisis sentimen terhadap asuransi berdasarkan data ulasan di *Facebook* menggunakan metode *K-Nearest Neighbor Classifier* untuk membangun sistem yang dapat memberi informasi terhadap penilaian sentimen yang mengarah ke sentimen positif dan sentimen negative[7]. Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking[8]. Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter[9]. Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode K-Nearest Neighbor[10],[11].

## III. METODE PENELITIAN

Penelitian ini akan di kerjakan pada bulan Juni 2023 hingga September 2023. Berlokasi di tempat tinggal yang beralamatkan Jalan Perintis Kemerdekaan VII Pondok Nirwana baru. Dalam menyelesaikan skripsi ini, jenis penelitian yang dilakukan adalah Jenis penelitian mengacu pada klasifikasi atau kategori berbeda dari pendekatan yang digunakan dalam proses penelitian berdasarkan tujuan, metode, dan sifatnya. Adapun beberapa jenis penelitian, pada penelitian ini menggunakan dua jenis penelitian yaitu penelitian (eksperimental) dan penelitian lapangan (*field research*). Penelitian Eksperimental adalah jenis penelitian yang mencari data serta mengkonfirmasi data pada kolom komentar halaman page Telkomsel di media social *Facebook* yang di klasifikasikan dengan metode *K-Nearest Neighbor* dan *Support Vector Machine* sedangkan, penelitian lapangan (*field research*) Penelitian Lapangan adalah jenis penelitian yang dilakukan dengan cara mengamati secara langsung objek penelitian, dalam hal ini penulis meneliti pada kolom komentar halaman page Telkomsel, Indosat, dan XL di *Facebook*.

Adapun metode penelitian yang digunakan dalam pengumpulan data adalah penulis menggunakan teknik pengumpulan data

observasi. Teknik observasi sendiri merupakan teknik yang diaplikasikan guna memperoleh data dengan cara menilik data secara sistematis berdasarkan dari objek penelitian. Dalam penelitian ini observasi dilakukan dihalaman page *Facebook*.

Pada penelitian ini, metode pengujian yang akan digunakan untuk menguji tingkat akurasi pada analisis ini adalah *confusion matrix*. *Confusion matrix* (matriks kebingungan) adalah alat evaluasi yang umum digunakan dalam klasifikasi untuk mengukur performa model berdasarkan prediksi dan label sebenarnya pada data uji. Ini membantu menggambarkan sejauh mana model mampu mengklasifikasikan data dengan benar dan mengidentifikasi jenis kesalahan yang dilakukan meliputi:

1. **True Positive (TP):** Data yang sebenarnya positif dan diprediksi positif dengan benar.
2. **True Negative (TN):** Data yang sebenarnya negatif dan diprediksi negatif dengan benar.
3. **False Positive (FP):** Data yang sebenarnya negatif tetapi diprediksi positif (kesalahan tipe I).
4. **False Negative (FN):** Data yang sebenarnya positif tetapi diprediksi negatif (kesalahan tipe II).

Berdasarkan empat komponen tersebut, beberapa metric evaluasi umum dapat dihitung pada penelitian ini, untuk menguji model dengan metode *confusion matrix* dalam Rapid Miner menghitung metric evaluasi dengan menggunakan nilai dari *confusion matrix* Akurasi, precision, dan recall sebagai berikut:

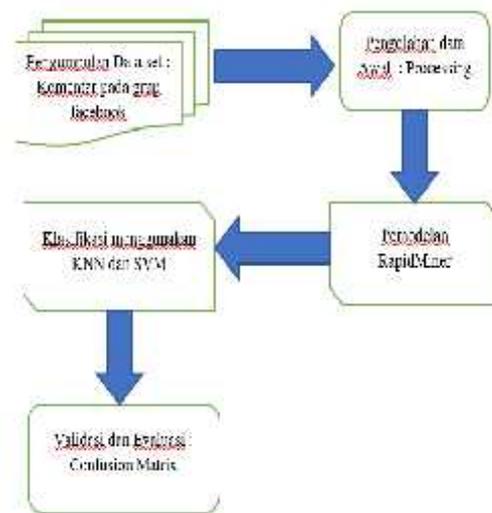
1. **Akurasi (Accuracy):** Persentase total prediksi yang benar dari keseluruhan dataset.  $Akurasi = (TP + TN) / (TP + TN + FP + FN) * 100\%$
2. **Precision:** menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model.  $Precision = (TP) / (TP + FP)$
3. **Recall:** menggambarkan keberhasilan model dalam menemukan kembali

sebuah informasi.  $Recall = TP / (TP + FN)$ .

#### IV. HASIL DAN PEMBAHASAN

##### 1. Perancangan Solusi

Pada gambar 4.1 menunjukkan gambaran arsitektur umum yang digunakan dalam penelitian ini. Dataset yang telah dikumpulkan dari facebook sebelumnya akan melalui praproses data, yang terdiri dari *cleaning text*, *case folding*, *filtering* dan *tokenizing*. Kemudian data yang sudah melalui tahap preproce singakan dimasukan pemodelan rapid miner untuk bekerja sebagai validasi data pada operator seluler telkomsel, indosat dan xl dengan dilanjutkan klasifikasi menggunakan metode K-nearest Neighbor dan support vector Machine Setelah pembuatan model, maka selanjutnya ialah melakukan evaluasi dengan menggunakan *confusion matrix*. Berdasarkan hasil *confusion matrix* dapat dilakukan perhitungan akurasi, presisi, dan recall yang digunakan untuk mengukur tingkat keberhasilan dari model yang telah dibuat.



Gambar 4.1 Arsitektur Umum

##### 2. Data pada komentar facebook

Data yang diperoleh dari ulasan atau komentar setelah menggunakan layanan operator selulertelkomsel,indosat,dan xl melalui form diskusidarifacebookdalam penelitian ini jumlah datanya sebanyak 1.061 yang masing-masingnya dari telkomsel sebanyak 309 komentar,indosat 382 komentar dan xl 370 komentar dan sudah di kelompokkan sebagai proses evaluasi untuk data testing pada table 4.1

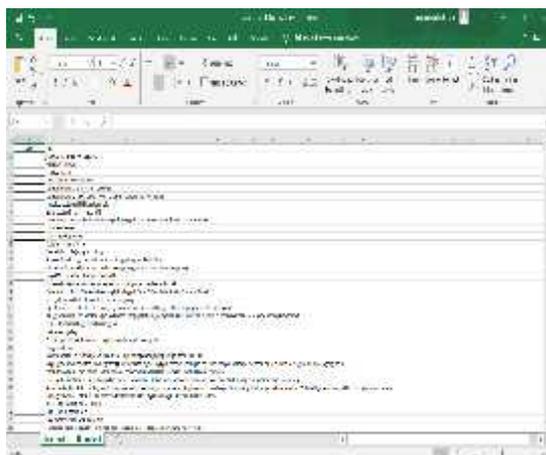
Tabel 4.1 Data komentar yang sudah dikelompokkan

No.	Data Set
1 Telkomsel	309
2 Indosat	382
3 Xl	370
Jumlah	1061

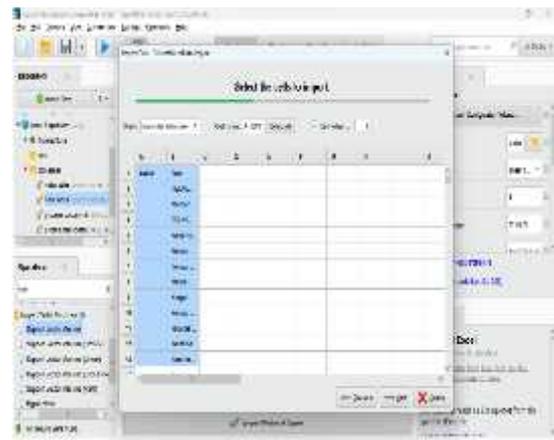
Jumlah komentar yang sudah terbagi tiga operator seluler pada table 4.1 diatas merupakan data training yang akan diolah melalui dengan pendekatan preprocessing dan penerapan klasifikasi op ini dengan melakukan tahap normalisasi data.

3. Proses Menentukan Label Positif atau Negatif di RapidMiner

Pada gambar 4.2 di bawah merupakan hasil dari penarikan data komentar operator telkomsel menggunakan export comment yang disimpan file excel.

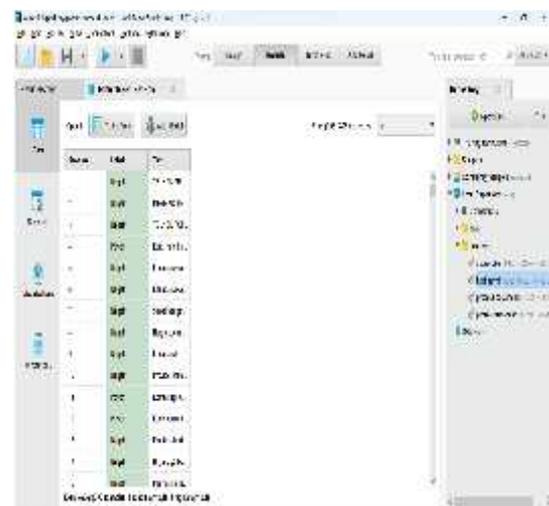


Gambar 4.2 Dataset komentar Telkomsel



Data set gambar 4.2 diatas ini merupakan data komentar operator telkomsel yang di ambil dari facebook menggunakan Export Comment dengan jumlah dataset sebanyak 309 komentar yang akan nantinya diberi label positif atau negative otomatis pada rapidminer.

Gambar 4.3 proses impor ke rapid miner



Pada Gambar 4.3 diatas merupakan proses import data komentar yang berjumlah 309 dari excel ke Rapid miner untuk menentukan mana data komentar yang bersifat positif atau negative secara otomatis

Gambar 4.4 Hasil dari proses data penentuan label positif atau negatif

Pada Gambar 4.4 diatas merupakan hasil dari proses penentuan label positif atau negative untuk data komentar operator telkomsel.

4. Perhitungan manual menggunakan K-Nearest Neighbor

Berikut ini adalah perhitungan manual penyelesaian menggunakan metode *K-Nearest Neighbor* dalam mengklasifikasi Opini Masyarakat Terhadap Operator seluler telkomsel.

No.	Komentar
1.	Sinyal sangat buruk tidak ada perubahan
2.	kenapa daily Check sekarang sudah tidak dapat poin setelah kita check in
3.	sekarang telkomsel bagus buat maen game
4.	Telkomsel buat nonton film bagus
5.	Harga kuota elit jaringan sulit
6.	Hanya orang bodoh yang pake telkomsel bos tinggalin sayang duit kalian
7.	Jaringan tambah hari tambah buruk
8.	Kenapa sekarang jaringan parah didaerah lemot banget
9.	stop berlangganan cara nya gimana
10.	telkomsel pelit tidakada kuota gratisnya

Tabel 4.17 Data uji telkomsel

Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*)

Tabel 4.18 Pembobotan Tf-Idf

No.	Term	Id dokumen	Jumlah
1	Sangat	1	1
2	Buruk	1,7	2
3	Bagus	3,4	2
4	Lemot	8	1
5	Stop	9	1
6	Pelit	10	1
7	Parah	1	0
8	Bodoh	6	1
9	Sayang	6	1

10	Elit	5	1
11	Tidak	1,2,10	3
12	Dapat	2	1
13	Poin	2	1
14	Gratis	10	1
15	Kuota	10	1
16	Nonton	4	1
17	Tambah	7	2
18	Bos	6	1
19	Duit	6	1
20	Banget	8	1

Hasil dari term frekuensi data training berupa token kata yang selanjutnya dilakukan tahapan klasifikasi, akan tetapi lebih dahulu dilakukan pelabelan kelas pada setiap opini dengan tujuan tools yang digunakan dapat mengenali dokumendari operator telkomsel.

Tabel 4.19 Data set Pembagian Kelas

No.	Komentar	Kelas
1.	Sinyal sangat buruk tidak ada perubahan	Negatif
2.	kenapa daily check sekarang sudah tidak dapat poin setelah kita check in	Negatif
3.	sekarang telkomsel bagus buat maen game	Positif
4.	Telkomsel buat nonton film bagus	Positif
5.	Harga kuota elit jaringan sulit	Negatif
6.	Hanya orang bodoh yang pake telkomsel bos tinggalin sayang duit kalian	Negatif
7.	Jaringan tambah hari tambah buruk	Negatif
8.	Kenapa sekarang jaringan parah didaerah lemot banget	Negatif
9.	stop berlangganan cara nya gimana	Negatif
10.	telkomsel pelit tidak ada kuota gratisnya	Negatif

Tabel 4.20 Data uji

NO	Komentar	kelas
11	Paket tambah murah baik sangat telkomsel	?

Pada tabel 4.19 labelisasi data training komentar pada operator seluler telkomsel yang akan selanjutnya akan diuji dengan algoritma K-Nearest Neighbor. Sebelum melakukan perhitungan jarak, perlu menentukan nilai k tetangga terdekat yang akan digunakan. Pada permasalahan ini, nilai parameter k yang akan digunakan yaitu k = 1 sampai k = 10. Tahap berikutnya yaitu menghitung jarak berdasarkan rumus Euclidean Distance yang ada pada persamaan 1.

Proses klasifikasi data latih(training):

d1=

$$Jarak = \sqrt{(X_{uji1} - X_{pelatihan1})^2 + (X_{uji2} - X_{pelatihan2})^2 + \dots + (X_{ujin} - X_{pelatihann})^2}$$

d1=

$$\sqrt{\sum_{i=1}^p (1 - 1)^2 + (1 - 1)^2 + (0,427 - 0,333)^2 + (0 - 0,427)^2 + (1 - 1)^2 + (0,333 - 1)^2}$$

$$d1 = \sqrt{\sum_{i=1}^p 0 + 0 + 0,8836 + 0,182329 + 0 + 0,444889}$$

$$d1 = \sqrt{\sum_{i=1}^p 1,510818}$$

$$d1 = \sqrt{\sum_{i=1}^p 0,755409}$$

Keterangan :

D1 = id dokumen

x1= nilai data training

x2= nilai data testing

n = dimensi data

i= variabel data

Diatas merupakan cara manual untuk menghitung jarak menggunakan rumus Euclidean Distance pada data latih 1, selanjutnya akan dihitung menggunakan Microsoft excel dengan rumus = SQRT((Data Latih-Data Uji)^2). Rumus SQRT adalah rumus untuk mencari akar kuadrat positif.

Tabel 4.21 Perolehan Jarak Data Uji Terhadap Data Latih

Data Uji	DataSample	Jarak
	Iddokumen=1	0,755409
	Iddokumen=2	0,762130
	Iddokumen=3	1.102391

Iddokumen=x	Iddokumen=4	0.810230
	Iddokumen=5	1.073201
	Iddokumen=6	0.490144
	Iddokumen=7	1.092312
	Iddokumen=8	1.313201
	Iddokumen=9	0.967309
	Iddokumen=10	0.865411

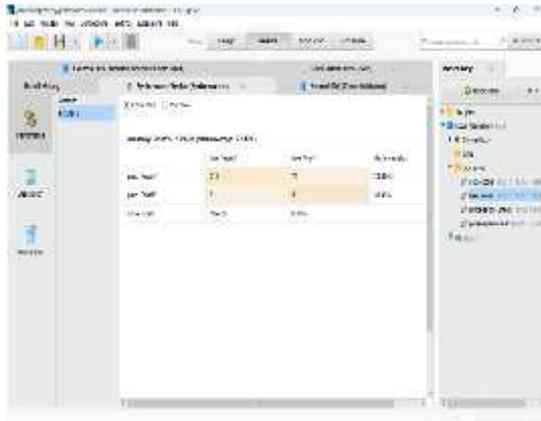
Tabel 4.21 menunjukkan perolehan jarak data uji ke setiap data latih, tahap selanjutnya adalah mengurutkan nilai jarak tersebut dari yang terkecil (terdekat) ke jarak terbesar. Kemudian dilakukan pemeriksaan ke setiap kelas berdasarkan inputan nilai k (tetangga terdekat). Karena pada penelitian ini nilai parameter k yang digunakan yaitu k = 1 sampai k = 10, maka yang diambil 10 terkecil.

Tabel 4.22 Hasil Urutan Jarak

Data Uji	DataSample	Jarak
Iddokumen=x	Iddokumen=6	0.490144
	Iddokumen=1	0,755409
	Iddokumen=2	0,762130
	Iddokumen=4	0.810230
	Iddokumen=10	0.865411
	Iddokumen=9	0.967309
	Iddokumen=5	1.073201
	Iddokumen=7	1.092312
	Iddokumen=3	1.102391
	Id dokumen = 8	1.313201

Pada tabel 4.22 menunjukkan perolehan jarak data uji ke setiap data latih, Hasil tersebut menunjukkan id dokumen = 6 merupakan nilai jarak yang terdekat dan id dokumen 8 merupakan nilai jarak yang terbesar dari parameter hitung nilai K.

### 5. Confusion Matrix data telkomsel dengan metode KNN



Gambar 4.14 Confusion matrix telkomsel knn

Nilai *accuracy* yang didapatkan algoritma *K-Nearest Neighbor*. Confusion Matrix Data pada telkomsel adalah 72.41% dengan *margin* +/- 2.65% dengan nilai rata-rata *mikro* sebesar 72.40%. Jadi, dari proses *crossvalidation* yaitu *training* dan *testing*, algoritma *K-Nearest Neighbor*, tentang Operator Seluler Telkomsel menghasilkan nilai *accuracy* 72.41% dengan rumus sebagai berikut:

$$\text{Precision} = \frac{23}{23+7} = \frac{23}{30} = 0.7667 = 76.67\%$$

$$\text{Recall} = \frac{23}{23+1} = \frac{23}{24} = 0.9583 = 95.83\%$$

$$\text{Accuracy} = \frac{23+285}{23+7+285+1} = \frac{308}{316} = 0.9747 = 97.47\%$$

**V. KESIMPULAN**

Berdasarkan Analisis dan pengujian yang dilakukan, maka kesimpulan yang dapat diambil adalah hasil dari implementasi perbandingan algoritma Support Vector Machine (SVM) dengan algoritma K-Nearest Neighbor (KNN), pada pengujian algoritma Support Vector Machine menunjukkan bahwa tingkat keakurasian untuk sentimen operator telkomsel yaitu 79.61%, operator XL yaitu 90.88% dan operator indosat yaitu 99.21%. Sedangkan pada pengujian algoritma K-Nearest Neighbor menunjukkan bahwa tingkat keakurasian untuk sentimen operator telkomsel yaitu 72.41%, operator XL yaitu 90.25% dan operator indosat yaitu 70.88%. Hasil akurasi pada penelitian ini dengan operator xl dinilai cukup baik karena termasuk Excellent Classification pada AUC yang dapat dilihat pada panduan keakuratan nilai AUC seperti 0.90-1.00 =

*Excellent classification*, 0.80-0.90 = *Good Classification*, 0.70-80 = *Fair Classification*, 0.60-70 = *Poor classification*, 0.50-0.60 = *Failure*. Analisis sentimen dengan algoritma Support Vector Machine (SVM) dan algoritma K-Nearest Neighbor. Berhasil mengklasifikasikan sentiment masyarakat di Facebook tentang operator telkomsel, xl dan indosat. Hasil klasifikasi menunjukkan sentiment negative lebih mendominasi dari sentiment positif pada setiap operator, untuk pengujian algoritma Support Vector Machine (SVM) terdapat 100% ulasan Negatif pada operator telkomsel, 100% ulasan negatif pada operator xl dan terdapat 99.21% ulasan negatif pada operator indosat. Sedangkan 71% ulasan Positif pada operator telkomsel, 84% ulasan positif pada operator xl dan 0% untuk ulasan positif pada operator indosat. Sedangkan untuk pengujian algoritma K-Nearest Neighbor terdapat 100% ulasan negatif pada operator telkomsel, 79% ulasan negatif pada operator xl dan 99% pada operator indosat. Sedangkan 50% untuk ulasan positif pada operator telkomsel, 79% untuk ulasan positif pada operator xl dan 0% untuk ulasan positif pada operator indosat. Dapat disimpulkan masih banyak orang yang ragu terhadap operator telkomsel, xl dan indosat dikarenakan resikonya yang cukup tinggi. Dan banyaknya korban serta kurangnya kepercayaan masyarakat.

**REFERENSI**

[1] Yuli Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD). Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.

[2] A. Salam *et al.*, "ANALISIS SENTIMEN DATA KOMENTAR SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASA)," pp. 480–486, 2018.

[3] S. Gusriani, K. Diah, K. Wardhani, and M. I. Zul, "Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)." *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.

[4] D. T. T. Santoso and E. Purwanti,

- “Pengaruh Faktor Budaya, Faktor Sosial, Faktor Pribadi, Dan Faktor Psikologis Terhadap Keputusan Pembelian Konsumen Dalam Memilih Produk Operator Seluler Indosat-M3 Di Kecamatan Pringapus Kab. Semarang,” vol. 6, no. 12, pp. 112–129, 2019.
- [5] K. E. P. Setiawan and W. Zyuliantina, “Analisis Kesalahan Berbahasa Indonesia Pada Status Dan Komentar Di Facebook,” *Tabasa J. Bahasa, Sastra Indones. dan Pengajarannya*, vol. 1, no. 1, pp. 96–109, 2020, doi: 10.22515/tabasa.v1i1.2605.
- [6] M. Faid, M. Jasri, and T. Rahmawati, “Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi,” *Teknika*, vol. 8, no. 1, pp. 11–16, 2019, doi: 10.34148/teknika.v8i1.95.
- [7] L. Oktasari *et al.*, *TEXT MINING DALAM ANALISIS SENTIMEN ASURANSI MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER*.
- [8] F. Rahmat, Z. Ngadengon, and N. S. M. Zawawi, “Perbandingan Alat Pengekstrakan Data Teks Janaan Pengguna,” *e-BANGI J. Sains Sos. dan Kemanus.*, vol. 17, no. 3, pp. 57–70, 2020.
- [9] L. Indah Prahartiwi, S. Informasi, S. Nusa Mandiri, J. Damai No, and W. Jati Barat Jakarta Selatan DKI Jakarta, “Pencarian Frequent Itemset pada Analisis Keranjang Belanja Menggunakan Algoritma FP-Growth,” *Inf. Syst. Educ. Prof.*, vol. 2, no. 1, pp. 1–10, 2017.
- [10] E. W. Sholeha, S. Yunita, R. Hammad, V. C. Hardita, and K. Kaharuddin, “Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 203–208, 2022, doi: 10.35746/jtim.v3i4.178.
- [11] H. Irsyad, A. Farisi, and M. R. Pribadi, “Klasifikasi Opini Masyarakat Terhadap Jasa ISP MyRepublic dengan Naïve Bayes,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 1, p. 30, 2019, doi: 10.22146/jnteti.v8i1.487.