

ANALISIS SENTIMEN DATA KOMENTAR SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASA EKSPEDISI BARANG J&T EKSPRESS INDONESIA)

Abu Salam¹, Junta Zeniarja², Rima Septiyan Uswatun Khasanah³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
e-mail: ¹abu.salam@dsn.dinus.ac.id, ²junta@dsn.dinus.ac.id, ²septiuswa@gmail.com

ABSTRAK

Pada saat ini media sosial telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet di Indonesia. Salah satu media sosial tersebut yakni Facebook dengan jumlah opini yang besar dan di dalamnya terdapat informasi yang sangat berharga sebagai alat penentu kebijakan dan ini bisa dilakukan dengan menggunakan text mining, kemudian masalah timbul bagaimana mengelompokkan komentar-komentar yang ada di facebook page J&T ke dalam kelas positif dan kelas negatif. Sistem klasifikasi komentar diharapkan dapat membantu untuk mengetahui respon positif dan negatif dari pengguna facebook yang memberikan komentarnya. Sentiment analysis digunakan untuk mengetahui sikap seseorang dalam konteks dokumen. Sentiment analysis memiliki tahapan preprocessing yang terdiri dari case folding, tokenizing, stopword removal, stemming. Pembobotan kata yang digunakan adalah term frequency-invers document frequency dan perhitungan similaritasnya menggunakan cosine similarity kemudian menggunakan K-Nearest Neighbor sebagai metode klasifikasinya. Hasil yang didapatkan dari implementasi metode KNN ini cukup baik dengan uji coba sebanyak 6 kali. Rata-rata accuracy tertinggi adalah 79.21% sedangkan accuracy terendah adalah 70.3%.

Kata Kunci: *Klasifikasi, Sentiment Analysis, K-Nearest Neighbor.*

1. PENDAHULUAN

Seiring berjalannya waktu, teknologi informasi berkembang sangat pesat, kegiatan yang pada umumnya banyak menggunakan peranan teknologi informasi didalamnya. Hal tersebut disebabkan oleh kebutuhan teknologi yang semakin meningkat, salah satunya teknologi yang memberikan banyak kemampuan untuk digunakan sebagai media komunikasi yang dapat mempercepat kerja manusia, salah satunya untuk pekerjaan jasa pengiriman barang. Pada saat ini jasa pengiriman barang menjadi hal yang tidak asing, karena bisnis tersebut banyak bertransaksi menggunakan media komputerisasi atau media internet.

J&T Ekspres merupakan salah satu perusahaan jasa pengiriman barang dari banyaknya perusahaan yang ada saat ini, dimana J&T memiliki akses yang luas sehingga sangat mudah digunakan untuk masyarakat dalam pengiriman barang. Selain itu juga sistem monitoring digunakan oleh J&T untuk mempermudah pelanggan dalam pengecekan paket kiriman dengan melalui web jet.co.id dimana web tersebut dikelola oleh pihak J&T Ekspres [1]. Dari banyaknya opini yang diberikan oleh pelanggan tersebut ada komentar yang bernilai positif, negatif, dan netral. Sehingga akan membutuhkan waktu yang cukup lama untuk memisahkan secara manual kedalam kelas positif, kelas negatif dan kelas netral. Komentar tersebut berupa teks selanjutnya akan diproses dengan salah satu teknik yang digunakan untuk memproses data teks yang biasa disebut teks mining yang merupakan salah satu bidang Natural Language Processing (NLP) karena teks mining mengolah data teks sehingga dapat dikomputasi dan dapat menghasilkan informasi yang relevan [2]. Sentiment Analysis dapat digunakan sebagai alat penentu kebijakan dari opini-opini yang diberikan oleh pelanggan pada facebook page J&T Ekspres Indonesia.

Sentiment Analysis merupakan perpaduan dari data mining dan teks mining, dimana Sentiment Analysis sendiri digunakan untuk mengolah berbagai macam opini yang telah diberikan oleh masyarakat atau para pakar melalui berbagai media yang ada, opini tersebut diberikan untuk sebuah produk, jasa maupun sebuah instansi. Pada Sentiment Analysis terdapat 3 jenis opini, yaitu opini negatif, opini positif dan opini netral. Terdapat beberapa algoritma yang dapat digunakan untuk metode sentimen ini, salah satunya yaitu algoritma K-Nearest Neighbor (K-NN) [3].

Algoritma K-Nearest Neighbor (K-NN) adalah sebuah metode lazy learning yang dimana tidak ada model yang dipelajari dari data pelatihan, yang terjadi hanya belajar bila contoh uji harus diklasifikasikan. K-Nearest Neighbor bertujuan untuk mengklasifikasikan objek berdasarkan atribut dan data training. Data klasifikasi tidak menggunakan data apapun untuk dicocokkan dan hanya berdasarkan memori. Didalam query terdapat titik yang akan ditemukan sejumlah k titik training yang posisinya paling dekat dengan titik query. Klasifikasi akan diketahui dengan adanya voting terbanyak antara klasifikasi k objek. Algoritma K-NN menggunakan klasifikasi ketetangaan sebagai nilai prediksi dari sample uji yang baru [4].

Penggunaan algoritma K-NN dalam penelitian bidang Sentiment Analysis ini telah banyak digunakan, salah satunya algoritma ini digunakan untuk penelitian suatu klasifikasi dokumen komentar pada situs youtube dengan menggunakan algoritma K-Nearest Neighbor (K-NN) oleh Moh Aziz Nugroho, membahas tentang klasifikasi dokumen komentar pada situs youtube menggunakan algoritma K-Nearest Neighbor (K-NN). Dalam

penelitiannya diperoleh data review film berbahasa Inggris ke dalam dua Sentiment yaitu positif dan negatif pada review film atau trailer bergenre aksi disitus youtube.com [5].

Dari kemiripan hasil penelitian yang telah disebutkan diatas, K-NN diusulkan sebagai algoritma yang akan digunakan untuk mengklasifikasi komentar-komentar pelanggan yang masuk kedalam facebook page J&T Ekspress Indonesia masuk kedalam Sentiment negatif atau Sentiment positif.

2. TINJAUAN PUSTAKA

Pada penelitian yang telah dilakukan oleh Anita Novantirani, dengan menggunakan metode Support Vector Machine (SVM). Dimana didalam paper tersebut telah menghasilkan tingkat akurasi yang relative lebih tinggi dibandingkan dengan menggunakan metode lain, namun tingkat akurasi tersebut sangat dipengaruhi dengan adanya jumlah dataset, data training dan data testing. Dalam penelitian tersebut dijelaskan bahwa dalam penggunaan metode Support Vector Machine (SVM) dokumen berhasil diklasifikasikan dengan baik, hal tersebut di tunjukan dengan tingginya akurasi metode yang digunakan. [6]

Didik, Yulison, dan Agung melakukan penelitian analisis sentiment mengenai layanan pada jasa ojek online untuk mengklasifikasikan opini masyarakat yang bersifat positif, negatif ataupun netral. Pada penelitian ini dilakukan pengujian dengan menggunakan 800 tweet yang dibagi menjadi 300 data latih dan 500 data uji, kemudian dari 300 data latih dibagi pada setiap klasifikasinya 100 data latih positif, 100 data latih negatif, dan 100 data latih netral. Data latih untuk kelas sentimen positif antara lain promo, gratis, nyaman, thanks dan baik. Kemudian data latih untuk kelas sentimen negatif yaitu jarang, rugi dan kecewa. Sedangkan data latih untuk kelas sentimen netral yaitu driver, gojek, orderan, dan gojekindonesia. Pada proses ini menggunakan preprocessing tersebut diklasifikasikan menggunakan algoritma Naive Bayes yang nantinya akan muncul hasil klasifikasi sentimen positif, netral dan negatif. Menggunakan algoritma Naive Bayes mampu menghasilkan tingkat akurasi mencapai 80%, namun dalam penelitian ini di ungkapkan bahwa sering terjadi kesalahan pada data fitur yang muncul tidak sesuai dengan klasifikasinya.[7]

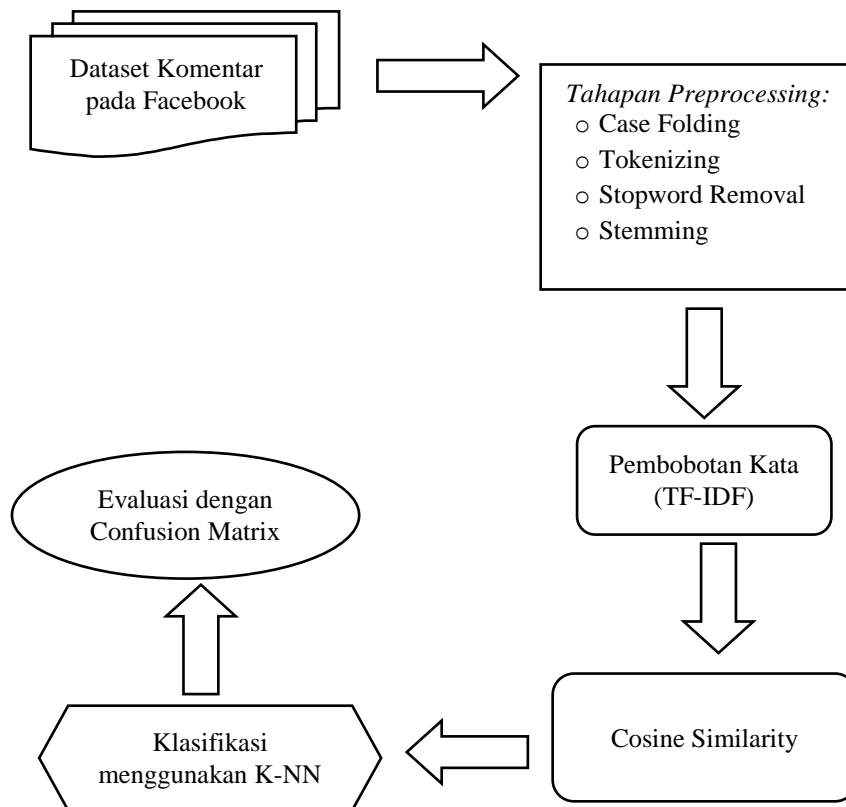
Syahfitri Kartika Lidya, melakukan penelitian tentang Sentiment Analisis yang memilih objek teks Bahasa Indonesia dengan menggunakan dua algoritma yaitu Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN). Dalam penelitian tersebut menggunakan teks Bahasa Indonesia yang terdapat pada website berupa artikelberita dan kemudian diklasifikasi dengan menggunakan metode K-Nearest Neighbor pada data pembelajaran agar dapat ditentukan model bentuk oleh metode Support Vector Machine untuk menentukan kategori data baru yang ingin ditentukan kategori tekstual, yaitu kelas Sentiment positif, kelas Sentiment negatif, dan kelas Sentiment netral. Berdasarkan seluruh hasil pengujian dengan k-fold cross validation dengan nilai k=10 adalah nilai yang optimal terhadap akurasi sebesar 67,90% karena hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis menunjukan bahwa 10-fold cross validation adalah pilihan terbaik untuk mendapatkan nilai validasi yang cukup akurat, 10-fold cross validation yaitu mengulang pengujian sebanyak 10 kali dengan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian.[4]

Penelitian ketiga dilakukan oleh Moh Aziz Nugroho, dimana penelitian mengenai klasifikasi dokumen komentar pada situs youtube dengan menggunakan algoritma K-Nearest Neighbor (K-NN). Dalam penelitian tersebut data yang diperoleh yaitu dari data review film berbahasa inggris dan dimasukkan kedalam dua Sentiment yaitu Sentiment negatif dan Sentiment positif.

Henny melakukan penelitian suatu perusahaan dengan mengklasifikasi kelayakan pemberian kredit khususnya pada kredit kepemilikan kendaraan bermotor. Pada penelitian ini penulis menggunakan 558 data transaksi kredit bermasalah dan tidak bermasalah, dengan menentukan atribut diantaranya status perkawinan, jumlah tanggungan, pendidikan terakhir, usia, kepemilikan rumah, lama tinggal, kondisi rumah, jenis pekerjaan, status perusahaan, penghasilan perbulan, pembayaran pertama dan masa kerja. Melalui tahap pemilihan data uji kemudian dilakukan preprocessing, setelah itu pengujian menggunakan algoritma K-NN dan analisa data uji. Menggunakan algoritma K-NN menunjukan bahwa metode yang digunakan menghasilkan tingkat akurasi mencapai 81,46%. Pada penelitian ini mengatakan bahwa untuk mengukur kinerja algoritma K-NN digunakan metode Cross Validation, Confusion matrix dan Kurva ROC.[8]

Herdiawan melakukan analisis terhadap produk layanan atau service pada perusahaan Telkom IndiHome guna mengetahui bagaimana respon customer terhadap produk IndiHome. Analisis sentiment yang dilakukan Herdiawan pada perusahaan Telkom IndiHome ini menggunakan metode K-Nearest Neighbor untuk melakukan klasifikasi opini positif dan negatif pada customer Telkom IndiHome. Pada penelitian ini dilakukan preprocessing convert negation, Case Folding, convert negation, Tokenizing, Filtering, Stemming. Setelah melalui tahap preprocessing selanjutnya proses pembobotan akan diuji menggunakan TF-IDF [9]. Hasil dari pembobotan akan diuji dengan menggunakan K-Nearest Neighbor[10]. Hasil yang dilakukan bahwa menggunakan K-Nearest Neighbor dapat mengklasifikasikan opini positif dan opini negatif dengan akurat mencapai 80%, dengan menggunakan pengujian precision, recall dan F-Measure.

3. METODE YANG DIUSULKAN



Gambar 1 Metode Yang Diusulkan

3.1 Komentar Pelanggan

Data yang digunakan dalam penelitian ini adalah kumpulan data berupa komentar-komentar pelanggan terhadap opini-opini terhadap pelayanan yang diberikan J&T pada *facebookpage* J&T Ekspress.

3.2 Preprocessing

Preprocessing dilakukan pada komentar-komentar pelanggan terlebih dahulu, karena tidak semua atribut yang terdapat pada kolom komentar pelanggan digunakan untuk menganalisis masalah. Didalam *preprocessing* terdapat beberapa tahap, yaitu:

a. Case Folding

Proses untuk mengubah semua huruf besar (*uppercase*) terhadap komentar menjadi huruf kecil (*lowercase*).

b. Tokenizing

Pada tahap ini dilakukan pengecekan karakter pertama sampai karakter terakhir, apabila karakter ke-I bukan karakter pemisah kata seperti titik(.), koma(,), spasi dan tanda pemisah lainnya, maka akan digabungkan dengan karakter selanjutnya.

c. Stopwordremoval

Pada tahap ini akan dilakukan pengecekan disetiap kata pada komentar, selanjutnya proses menghilangkan kata-kata yang dianggap tidak penting, semisal terdapat kata sambung, kata depan, kata ganti, atau kata yang tidak ada hubungannya dengan *Sentiment Analisis*, maka akan dihapus.

d. Stemming

Tahap ini melakukan proses mengubah kata yang berimbun *infix* maupun *suffix* menjadi sebuah kata dasar yang akan lebih mengandung sebuah makna untuk peroleh suatu infoormasi sehingga komentar akan menjadi lebih spesifik dalam pengkategorian.

3.3 Pembobotan Kata (TF-IDF)

Dataset yang sudah di *preprocessing* sebelumnya kemudian diproses kembali kedalam bentuk format bilangan biner sehingga dapat dikenal oleh sistem.

3.4 Cosine Similarity

Cosine similarity berfungsi untuk membandingkan kemiripan antar dokumen, dalam hal ini yang dibandingkan adalah *query* dengan dokumen latih. Dalam menghitung *cosine similarity* pertama yaitu

melakukan perkalian scalar antara *query* dengan dokumen kemudian dijumlahkan, setelah itu melakukan perkalian antara panjang dokumen dengan panjang *query* yang telah dikuadratkan, setelah itu dihitung akar pangkat dua, selanjutnya hasil perkalian scalar tersebut dibagi dengan hasil perkalian panjang dokumen dan *query*.

3.5 Klasifikasi Menggunakan K-Nearest Neighbor (K-NN)

Pada tahap klasifikasi ini yaitu mengambil hasil dari pembobotan kata. Hasil dari pembobotan kata tersebut selanjutnya akan dihitung nilai similaritasnya atau kemiripan antara dokumen uji dengan dokumen latih, setelah didapatkan nilai similaritasnya tentukan nilai k, ambil hasil similaritas tersebut sesuai nilai k dimulai dari nilai similaritas yang paling tinggi, setelah mendapatkan hasil similaritas sejumlah nilai k, tentukan hasil klasifikasi yaitu dalam kelas positif atau kelas negatif.

3.6 Evaluasi Dengan Confusion Matrix

Tahap selanjutnya setelah diketahui hasil dari klasifikasi menggunakan K-Nearest Neighbor (K-NN) akan dilakukan perhitungan akurasi menggunakan *Confusion matrix* yakni menghitung semua data uji yang telah berhasil diklasifikasi sesuai dengan target, ini nantinya akan melakukan perhitungan *accuracy* (perbandingan kasus yang diidentifikasi benar dengan jumlah seluruh kasus).

4. HASIL DAN PEMBAHASAN

4.1 Data Komentar

Data merupakan ulasan yang diberikan pengguna setelah menggunakan layanan jasa dari J&T, dalam penelitian ini menggunakan 335 data komentar dari akun resmi J&T Indonesia yang sudah dikelompokkan secara manual oleh pihak J&T sebagai proses evaluasi periodik untuk data dengan konten positif dan negatif.

4.2 Preprocessing Data

Tahapan yang dilakukan dalam preprocessing yaitu:

a. Case Folding

Tabel 1. Proses case folding

ID	Komentar Asli	Komentar setelah Case Folding
1	J&TPelayanannya sangat Cepat.	j&tpelayanannya sangat cepat.
2	Bagus dan Memuaskan	bagus dan memuaskan
3	Parah Masa bisa Salah kirim.	parah masa bisa salah kirim.
4	Kecewa banget sama pelayanannya, PAKET SALAH ALAMATT!!	kecewa banget sama pelayanannya, paket salah alamatt!!

b. Tokenizing

Pada tahap ini akan dilakukan penghapusan delimiter yaitu karakter angka dan karakter simbol kecuali karakter huruf dengan cara membuat daftar kode karakter yang diperlukan. Contoh karakter yang ada pada daftar kode Tokenizing adalah:

Tabel 2. Daftar karakter

.	,	:	1
?	!	/	2
“	;	#	3

Tabel 3. Hasil tokenizing

ID	Komentar	Kelas
1	J&T//Pelayanannya//sangat// Cepat.	Positif
2	Bagus//dan//Memuaskan	Positif
3	Parah//Masa//bias//Salah//kirim.	Positif

c. Stopword Removal (Filtering)

Langkah-langkah untuk proses Filtering adalah sebagai berikut:

- 1) Mempersiapkan kata kunci yang diperlukan untuk penghapusan proses Filtering.
- 2) Proses pencarian pada kata-kata yang ada didalam data komentar.
- 3) Jika ditemukan kata yang sesuai dengan kata kunci maka akan dihapus, namun jika kata kunci stopwords tidak ada pada kata yang ada di dalam komentar maka kata tersebut tidak dihapus.

d. Stemming

Tahap ini melakukan Stemming dengan menggunakan algoritma porter, langkah pertama mempersiapkan kamus kata dasar untuk proses Stemming yang diambil dari Kamus Besar Bahasa Indonesia, setelah itu mengecek kata yang ada disistem dalam kamus kata dasar, jika ditemukan maka diasumsikan kata tersebut

adalah kata dasar, maka algoritma berhenti, jika tidak ditemukan maka lakukan penghapusan sisa huruf yang tidak ada pada kamus kata dasar tersebut.

4.3 Pembobotan Kata TF-IDF

Pembobotan kata dimulai dari mencari TF dengan cara menghitung jumlah kata yang muncul pada dokumen, proses selanjutnya setelah didapatkan nilai TF adalah mencari nilai IDF, dimana IDF merupakan hasil invers dari DF, cara untuk menentukan IDF adalah jumlah seluruh dokumen yaitu D1, D2, D3, D4, D5, Dx dibagi jumlah dokumen yang mengandung sebuah kata (DF) kemudian hasilnya di log-kan.

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right)$$

Dari nilai idf tersebut akan dicari lagi nilai TF-IDF yaitu perkalian antara hasil dari frekuensi kemunculan kata pada setiap dokumen (TF) dengan bobot kata dalam seluruh dokumen (IDF).

Tabel 4. Hasil pembobotan

No	Kata	TF × IDF					
		D1	D2	D3	D4	D5	Dx
1	Layan	0.301	0.301	0	0	0.301	0
2	Puas	0.778	0	0	0	0	0
3	cepat	0.301	0.301	0	0	0	0.301
4	sampai	0.477	0	0	0	0	0.477

4.4 Klasifikasi Menggunakan K-Nearest Neighbor

Pada perhitungan klasifikasi menggunakan KNN yang pertama dilakukan adalah menentukan nilai k, kemudian ambil hasil similaritas tersebut sesuai nilai k dimulai dari nilai similaritas yang paling tinggi, setelah mendapatkan hasil similaritas sejumlah nilai k, tentukan hasil klasifikasi yaitu dalam kelas positif atau kelas negatif. Sebagai contoh nilai k=3 dengan menggunakan rumus (7) karena nilai k lebih besar dari satu maka perhitungannya adalah menjumlahkan semua nilai kemiripan yang tergolong dalam satu kategori kemudian membandingkan manakah yang lebih besar. Berdasarkan perhitungan di atas didapatkan probabilitas dx terhadap kelas positif lebih besar dibandingkan kelas negatif, maka data tersebut termasuk dalam kategori kelas positif.

Probabilitas terhadap sentimen positif :

$$p(x, c_m) = \sum_{i=1}^m SIM(X, d_j) \in c_m$$

$$P(x, positif) = 0.359 + 0.161$$

$$P(x, positif) = 0.52$$

Probabilitas terhadap sentimen negatif :

$$p(x, c_m) = \sum_{i=1}^m SIM(X, d_j) \in c_m$$

$$P(x, negatif) = 0.349$$

$$P(x, negatif) = 0.349$$

Hasil klasifikasi positif dan negatif pada komentar dikatakan “Benar” jika sistem berhasil mengenali dan mengelompokkan komentar ke dalam kelas yang sesuai dengan kelas data latih. Berikut tabel hasil klasifikasi komentar dengan menggunakan metode K-NN dengan nilai k = 3.

Tabel 6. Sampel list hasil proses klasifikasi

ID	Komentar	Hasil Preprocessing	Kelas Aktual	Kelas Prediksi
1	Pelayanan yang baik	Layan	Positif	positif
2	Terima kasih..	terima kasih	Positif	positif
3	Mantaplah, bintang lima melebihi panglima	mantap bintang lebih panglima	Positif	positif
4	J&T TETAP terpercaya.!!!!	terpercaya	Positif	positif

Hasil klasifikasi dapat dilihat dari nilai k = 3 hasil nilai Benar berjumlah 79 komentar dan hasil nilai Salah berjumlah 22 komentar dengan nilai positif berjumlah 41 data dan negatif berjumlah 60 data, hal ini menunjukkan bahwa klasifikasi menggunakan metode KNN dapat digunakan untuk mengklasifikasi dokumen komentar Jasa Ekspedisi Barang (J&T).

4.5 Hasil Akurasi Menggunakan Confusion Matrix

Setelah melakukan proses pembobotan kata dan klasifikasi menggunakan metode K-Nearest Neighbor (KNN), maka dilakukan tahap pengujian untuk mencari nilai akurasi. Pengujian menggunakan Confusion matrix

untuk perbandingan data testing sebanyak 101 dengan data training sebanyak 234 yang diambil dari perusahaan yang sudah berlabel kelas positif dan kelas negatif. Cara untuk membandingkannya adalah sebagai berikut :

- Menghitung jumlah hasil data asli positif dan data klasifikasi positif
- Menghitung jumlah hasil data asli positif dan data klasifikasi negatif
- Menghitung jumlah hasil data asli negatif dan data klasifikasi negatif
- Menghitung jumlah hasil data asli negatif dan data klasifikasi positif
- Kemudian setelah diketahui hasilnya maka langkah a dan c dijumlahkan
- Selanjutnya melakukan penjumlahan langkah a, b,c dan d
- Setelah itu langkah e dibagi dengan langkah f untuk metode Confusion matrix dapat menggunakan rumus sebagai berikut :

$$\text{Accuracy} = \frac{(a + c)}{(a + b + c + d)} \times 100$$

Tabel 7. Confusion matrix k=3

		Kelas Predikat	
		Positif	Negatif
Kelas Aktual	Positif	37	4
	Negatif	18	42

Pada k = 3 hasil yang didapatkan dari perhitungan data testing 101 data adalah nilai a= 37, nilai b= 18, nilai c= 42 dan nilai d= 4.

$$\text{Accuracy} = \frac{(a + c)}{(a + b + c + d)} \times 100$$

$$\text{Accuracy} = \frac{37 + 42}{37 + 18 + 42 + 4} \times 100$$

$$\text{Accuracy} = \frac{79}{101} \times 100$$

$$\text{Accuracy} = 78.22\%$$

Dalam penelitian dilakukan ujicoba untuk proses klasifikasi menggunakan k = 3, k = 5, k = 7, k = 9, k = 11, k = 13, dari hasil penghitungan confusion matrix yang dilakukan pada tabel k = 9 terlihat bahwa hasil akurasi tertinggi yaitu nilai akurasi sebesar 79.21% dengan 80 jumlah data yang benar dan 21 data yang salah. Sedangkan pada tabel k = 13 yaitu nilai akurasi sebesar 70.3% dengan total 71 data benar dan 30 data yang salah, untuk rata-rata tingkat akurasi pada penelitian K-Nearest Neighbor (KNN) ini dari k = 3, k = 5, k = 7, k = 9, k = 11, k = 13 menghasilkan nilai rata-rata akurasi sebesar 75.58%.

5. KESIMPULAN

Jumlah dokumen training dan nilai k sangat berpengaruh dalam pengklasifikasian dengan menggunakan dokumen training dan nilai k yang tepat, maka akan didapatkan hasil klasifikasi yang baik. Hasil pengujian klasifikasi analisa sentimen dokumen komentar pada jasa ekspedisi barang (J&T) di facebook menggunakan algoritma K-Nearest Neighbor yang memiliki accuracy tertinggi yaitu 79,21% dengan jumlah dokumen training 101 sedangkan accuracy terendahnya adalah 70,3% dengan jumlah dokumen training 101. Klasifikasi dengan metode K-Nearest Neighbor ini dapat diterapkan dengan otomatis yang mempunyai langkah yaitu preprocessing terlebih dahulu untuk dapat melakukan pembobotan kata dan menghitung similaritas atau cosine similarity, setelah itu diambil nilai cosine sebanyak nilai k yang dimasukan user untuk mengklasifikasikan dokumen tersebut.

DAFTAR PUSTAKA

- [1] Azam Faiz Kamal and Budi Widjajanto, Text Mining Untuk Analisa Sentiment Ekspedisi Jasa Pengiriman Barang Menggunakan Metode *Naive Bayes* Pada Palikasi J&T Ekspres, *IJCCS*, pp. 1-8, 2017.
- [2] Hagi Semara Putra, Penentuan Emosi Berdasarkan Lirik Lagu Menggunakan Sistem Temu Kembali Informasi Dengan Metode *Latent Semantic Indexing (Lsi)*, pp. 5-9, 2015.
- [3] Evasaria Sipayung, Herastia Maharani, and Ivan Zefanya, Perancangan Sistem Analisis Sentiment Komentar Pelanggan Menggunakan Metode *Naive Bayes Classifier*, *jurnal sistem informasi*, vol. 8, no. 1, pp. 958-965, 2016.
- [4] Kartika Lidya Syahfitri, Opim Salim Sitompul, and Syahril Efendi, *Sentiment Analisis* Padateks Bahasa Indonesia Menggunakan Support Vector Machine (Svm) Dan *K-Nearest Neighbor (K-Nn)*, pp. 1-8, 2015.
- [5] Moh Aziz Nugroho, Kalsifikasi Dokumen Komentar Pada Situs *Youtube* Menggunakan Algoritma *K-Nearest Neighbor (K-Nn)*, *skripsi fakultas ilmu komputer*, no. 5, 2016.

- [6] Anita Novantirani, Mira Kania Sabariah, and Veronika Effendy, Analisis Sentimen Pada Twitter Untuk Mengeneai Penggunaan Transportasi Umum Darat Dalam Kota Dengan Metode Support Vector Machine, *e-Proceeding of Engineering*, vol. 2, no. 1, pp. 1-7, 2015.
- [7] Didik Garbian Nugroho, Yulison Herry Chrisnanto, and Agung Wahana, Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode *Naive Bayes*, pp. 156-161, 2016.
- [8] Henny Leidiyana, Penerapan Algoritma *K-Nearest Neighbor* Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor, *jurnal penelitian ilmu komputer*, no. 1, pp. 65-76, 2013.
- [9] Herdiawan, Analisis Sentimen Terhadap Telkom Indihome Berdasarkan Opini Publik Menggunakan Metode Improved *K-Nearest Neighbor* ,no. 1, pp. 1-13, 2016.
- [10] Budi Santoso, Text Mining Untuk Analisis Sentiment *Review Film* Menggunakan Algoritma K-Means, vol. 16, no. 1, pp. 1-8, 2017.