

**ANALISIS SENTIMEN TERHADAP MASKAPAI GARUDA
INDONESIA PADA MEDIA SOSIAL TWITTER
MENGUNAKAN METODE *K-NEAREST NEIGHBOR***

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Pada
Jurusan Teknik Informatika

Oleh:

RAHMAN ADITYA

11551102857



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU**

2019

DAFTAR ISI

	Halaman
DAFTAR ISI.....	i
DAFTAR GAMBAR	iii
DAFTAR TABEL.....	iv
BAB I PENDAHULUAN.....	I-5
1.1 Latar Belakang.....	I-5
1.2 Rumusan Masalah	I-9
1.3 Batasan Masalah.....	I-9
1.4 Tujuan.....	I-9
1.5 Sistematika Penulisan.....	I-9
BAB II LANDASAN TEORI.....	II-11
2.1 Twitter	II-11
2.2 Analisis Sentimen.....	II-11
2.3 Text Preprocessing	II-12
2.3.1 Cleaning.....	II-12
2.3.2 Case Folding.....	II-13
2.3.3 Tokenizing	II-13
2.3.4 Normalisasi	II-13
2.3.5 Stemming.....	II-13
2.3.6 Negation Handling.....	II-15
2.3.7 Filtering.....	II-15
2.4 Pembobotan Kata.....	II-Error! Bookmark not defined.
2.5 Feature Selection	II-16
2.6 Klasifikasi.....	II-18
2.6.1 Algoritma <i>K-Nearest Neighbor</i>	II-19
2.6.2 Confusion Matrix.....	II-20
2.7 Penelitian Terkait.....	II-21
BAB III METODOLOGI PENELITIAN	III-1

3.1	Identifikasi Masalah	III-2
3.2	Studi Pustaka	III-2
3.3	Pengumpulan Data.....	III-2
3.4	Analisa.....	III-2
3.4.1	Kebutuhan Data	III-2
3.4.2	Pelabelan Manual Dataset.....	III-3
3.4.3	Text Pre-processing	III-3
3.4.4	Pembobotan Kata.....	III-Error! Bookmark not defined.
3.4.5	Feature Selection.....	III-4
3.4.5	Klasifikasi	III-4
3.5	Perancangan.....	III-5
3.6	Implementasi	III-5
3.7	Pengujian	III-6
3.8	Kesimpulan Dan Saran	III-6
Daftar Pustaka		v

DAFTAR GAMBAR

Gambar	Halaman
3. 1 Tahapan Metodologi Peneitian	III-1

DAFTAR TABEL

Tabel	Halaman
2. 1 Confusion Matrix	II-20
2. 2 Penelitian Terkait	II-21

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring perkembangan dunia usaha sekarang ini, pertumbuhan di bidang perekonomian baik jasa maupun industri mengalami perkembangan yang sangat pesat, contohnya pada jasa maskapai penerbangan (Wati, 2016). Jasa penerbangan adalah salah satu bentuk jasa yang melayani kebutuhan manusia untuk bermigrasi dari satu tempat ke tempat lain yang berjauhan dengan waktu yang singkat. Dilansir dari situs Tempo.co berdasarkan penyampaian oleh (kementrian perhubungan, 2017) jumlah penumpang pesawat meningkat 9,5 persen.

Berdasarkan peningkatan penumpang pesawat tersebut, terjadi persaingan antar maskapai dalam menarik minat penumpang, termasuk maskapai penerbangan Garuda Indonesia. Maskapai Garuda Indonesia didirikan pada tanggal 1 Agustus 1947 sebagai *KLM Interinsulair Bedrijf*, dan mulai beroperasi pada tanggal 26 Januari 1949 sebagai *Indonesia Airways*. Maskapai Garuda Indonesia sempat dilarang terbang menuju Eropa pada tahun 2007 karena kejadian yang menimpa pesawat Garuda Indonesia nomor penerbangan 200. Setahun kemudian maskapai Garuda Indonesia menerima sertifikasi *IATA Operational Safety Audit* dari IATA yang menunjukkan maskapai Garuda Indonesia telah memenuhi standar keselamatan penerbangan Internasional. Pada tanggal 1 Juni 2010 maskapai Garuda Indonesia membuka kembali penerbangan ke Amsterdam menggunakan pesawat Airbus A330-200 dengan perhentian di Dubai. Pada bulan juni 2012 Garuda Indonesia mengadakan perjanjian dengan klub sepakbola Liverpool sebagai sponsor global. Pada tahun 2013 maskapai Garuda Indonesia mendapat dua penghargaan dari Skytrax yaitu “*World Best Economy Class*” dan “*World Best Economy Class Seat*”. Pada tahun 2014 Garuda Indonesia mendapatkan penghargaan “*World’s Best Cabin Crew*”.

Terlepas dari banyaknya penghargaan yang didapatkan, maskapai Garuda Indonesia juga banyak menerima pendapat atau opini mengenai baik pendapat negatif dan positif. Salah satunya tentang pemberitaan dimana maskapai Garuda Indonesia mencantumkan ucapan terimakasih kepada presiden Jokowi pada pesawat yang mengangkut jemaah haji Indonesia. Hal ini memancing banyak pendapat, diantaranya Tweet dari (Putra Melayu @ardi_riau) “itu pesawat milik publik, bukan milik Garuda, kalian hanya mengelola, kenapa di coret2 pakai bawa nama orang lain?...”. Dari (Netijen garis lurus @aku_oposisi) “Emang Jokowi yg bayar ongkos jemaah haji?? Ga ikhlas!! Mengganggu itu nama pembohong ada di badan pesawat”. Pendapat lainnya mengenai pelayanan pesawat yaitu dari (Neddy Farmanto @neddyfarmanto) “Sangat mengecewakan tidak layak lagi menyandang *5 airlines. Ini yang diberikan kepada pax di GA536 tgl 20 Mei 2019. Biasanya selalu dapat *hot meal* apalagi pada saat waktu buka puasa”. Dari (Sugiarti @Sugiart59338296) “Garuda Indonesia penerbangan yg ok servicenya bagus”. Dari banyaknya sentimen positif dan negatif terhadap maskapai Garuda Indonesia, maka untuk mengetahui kualitas jasa maskapai Garuda Indonesia dapat diketahui dengan menganalisis sentimen tersebut dari *review* online masyarakat pada media sosial. Pada penelitian ini, penulis akan menggunakan media sosial Twitter untuk pengambilan data sentimen yang akan dianalisis.

Jack Dorsey merupakan pendiri media sosial Twitter. Twitter didirikan pada tanggal 21 Maret 2006. Pengguna Twitter dapat membaca dan mengirim teks hingga 280 karakter. Twitter sudah melekat pada kehidupan manusia pada saat sekarang (Aggrawal, 2018). Menurut data yang didapat dari PT. Bakrie Telecom jumlah pengguna Twitter di dunia yaitu 500 juta dan 19,5 juta pengguna yang berasal dari Indonesia dan Indonesia peringkat kelima pengguna Twitter di dunia (Kementerian Komunikasi dan Informatika Republik Indonesia, 2013). *Postingan* pada Twitter disebut dengan *tweet* (kicauan). Kicauan pada Twitter dapat bersifat menyinggung terhadap suatu kelompok, individu ataupun perusahaan tertentu termasuk maskapai Garuda Indonesia. Untuk dapat mengambil kesimpulan mengenai jasa penerbangan Garuda Indonesia, sulit jika hanya membaca satu atau sebagian *tweet* pada media sosial

Twitter. Untuk itu maka di perlukan analisis sentimen pada maskapai penerbangan Garuda Indonesia agar kesimpulan yang diambil lebih baik.

Menurut (Liu, 2012) analisis sentimen dapat digunakan dalam berbagai kemungkinan *domain*, dari produk konsumen, jasa kesehatan, jasa keuangan peristiwa sosial dan politik pada pemilu. Menurut (Simanjuntak, 2018) tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas mempunyai arti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif, negatif atau netral.

Sejumlah penelitian tentang analisis sentimen dengan bermacam kasus dan metode seperti penelitian (Kharisman, 2017.) tentang analisis sentimen pada *review* konsumen makapai penerbangan menggunakan kombinasi *Lexicon* berbasis *sentiwordnet* dan *supervised* model, penelitian (Wati, 2016) tentang penerapan algoritma genetika untuk seleksi fitur pada analisis sentimen *review* jasa maskapai penerbangan menggunakan *Naive Bayes*.

Pada penelitian (Kharisman, 2017) menggunakan menggunakan kombinasi *lexicon* berbasis *sentiwordnet* dan *supervised model* (Kharisman, 2017) melakukan penelitian dengan membatasi topik sentimen pada layanan Gojek variabel orientasi dalam 2 jenis yaitu positif dan negatif. Akurasi pengujian metode *lexicon* berbasis *sentiwordnet* untuk menentukan label data terhadap data uji memberikan persentase akurasi 82.00%. Pengujian metode *Machine Learning* untuk mengklasifikasikan sentimen terhadap data uji menunjukan bahwa akurasi SVM memberikan persentase yang lebih besar dengan persentase 91.00% sedangkan akurasi NBC memberikan persentase 89.50%. Pada penelitian (Wati, 2016) menggunakan metode algoritma genetika untuk seleksi fitur dan *Naive Bayes* dan menggunakan 2 variabel sentimen yaitu positif dan negatif. Setelah menguji dua model, penelitian ini menghasilkan akurasi 60.00% dan AUC sebesar 0.512 sedangkan akurasi setelah menggunakan pemilihan fitur Algoritma Genetika sebesar 89.50 dan AUC sebesar 0.919 termasuk kedalam *excellent classification* terjadi peningkatan akurasi sebesar 29.5% dan AUC sebesar 0.407.

Penelitian lainnya yaitu dari (Mentari, Fauzi, & Muflikhah, 2018) tentang analisis sentimen kurikulum 2013 pada sosial media *twitter* menggunakan metode *K-Nearest Neighbor* dan *feature selection query expansion ranking*, dari (Indriati & Ridok, 2016) tentang *sentiment analysis for review mobile applications using neighbor method weighted K-Nearest Neighbor*, dari (Nurjanah, Perdana, & Fauzi, 2017) tentang analisis sentimen terhadap tayangan televisi berdasarkan opini masyarakat pada media sosial *twitter* menggunakan metode *K-Nearest Neighbor* dan pembobotan jumlah *retweet*.

Pada penelitian ini akan dilakukan klasifikasi orientasi sentimen dalam 2 kelas yaitu positif dan negatif dengan menggunakan metode *K-Nearest Neighbor* (KNN) pada maskapai Garuda Indonesia. Berdasarkan penelitian (Utami, 2017) tentang analisis sentimen opini berita kebakaran hutan melalui komparasi algoritma *Support Vector Machine* dan *K-Nearest Neighbor* berbasis *Particle Swarm Optimazation* menghasilkan kesimpulan bahwa dengan menggunakan algoritma KNN menghasilkan akurasi lebih tinggi dari pada menggunakan SVM, yaitu tingkat akurasi KNN 85.00% sedangkan SVM sebesar 80.83%. Penelitian lainnya (Putri, Suparti, & Rahmawati, 2014) membandingkan metode KNN dengan NBC pada kasus analisis dan status kerja di Kabupaten Demak dan mendapatkan hasil metode KNN lebih baik dari pada NBC dalam mengklasifikasikan status kerja di Kabupaten Demak, yaitu tingkat akurasi NBC 94.09% sedangkan KNN sebesar 96.06%. Melihat dari penelitian-penelitian diatas, Algoritma KNN mendapatkan hasil yang baik.

Algoritma KNN digunakan untuk mengklasifikasi objek baru berdasarkan data latih. Menurut (Rosdiansyah, 2014) algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai dari sampel uji baru. Pada penelitian ini algoritma KNN bertindak sebagai algoritma untuk melakukan klasifikasi orientasi sentimen yaitu positif dan negatif pada kasus maskapai penerbangan Garuda Indonesia. Untuk menjaga tingkat akurasi dari data latih, labeling akan dilakukan secara manual.

Berdasarkan latar belakang di atas penulis mengajukan penelitian dengan judul ”Analisis Sentimen Terhadap Maskapai Garuda Indonesia Menggunakan Metode *K-Nearest Neighbor*”.

1.2 Rumusan Masalah

Dari latar belakang diatas maka dapat dirumuskan masalah penelitian bagaimana menerapkan algoritma *K-Nearest Neighbor* untuk melakukan analisis sentimen pada maskapai Garuda Indonesia pada komentar media sosial Twitter.

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini sebagai berikut:

1. Data opini yang diambil berasal dari media sosial Twitter berupa teks dalam bahasa Indonesia.
2. *Dataset* yang digunakan berjumlah 1000 komentar.
3. Kelas sentimen yang digunakan pada penelitian ini yaitu positif dan negatif.

1.4 Tujuan

Adapun tujuan yang ingin dicapai pada penelitian ini adalah:

1. Menerapkan algoritma KNN dalam menganalisis sentimen terhadap maskapai Garuda Indonesia Indonesia pada media sosial Twitter.
2. Mengetahui akurasi pada penelitian ini yang menggunakan algoritma KNN.

1.5 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini terdiri dari pokok-pokok permasalahan yang dibahas pada masing-masing yang diuraikan menjadi beberapa bagian:

BAB I. PENDAHULUAN

Bab ini membahas tentang gambaran umum isi tugas akhir yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan.

BAB II. LANDASAN TEORI

Bab ini membahas tentang landasan teori dan mendukung penelitian ini sebagai dasar untuk pemecahan masalah.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas langkah-langkah yang dilaksanakan dalam proses penelitian, yaitu merumuskan masalah, studi pustaka, pengumpulan data, analisa, perancangan, *preprocessing*, *feature selection*, pembobotan kata, klasifikasi, implementasi, pengujian dan kesimpulan dan saran.

BAB IV. ANALISIS DAN PERANCANGAN

Bab ini berisi pembahasan mengenai analisa pada aplikasi yang akan dibangun, meliputi analisa dan perancangan pada aplikasi sentimen analisis maskapai Garuda Indonesia.

BAB V. IMPLEMENTASI

Dalam bab ini membahas tentang implementasi aplikasi analisis sentimen maskapai Garuda Indonesia.

BAB VI. PENUTUP

Dalam bab ini akan dijelaskan mengenai beberapa kesimpulan yang didapatkan dari pembahasan pembuatan aplikasi analisis sentimen maskapai Garuda Indonesia.

BAB II

LANDASAN TEORI

2.1 Twitter

Jack Dorsey merupakan pendiri dari media sosial Twitter. Twitter didirikan pada tanggal 21 Maret 2006. Twitter merupakan sebuah *platform* yang banyak digunakan oleh tokoh publik, selebriti, bintang olahraga dan tokoh terkemuka lainnya (GOYAL, 2012). Twitter memungkinkan penggunaanya untuk membaca dan mengirim teks hingga 280 karakter. Terdapat beberapa istilah pada Twitter yaitu *Direct Message, Favorite, Follow, Follower, Following, Unfollow, Tweet, Retweet, Timeline, Hashtag, Mention, Trending Topic, Search, List, Over Heard (OH)*.

2.2 Analisis Sentimen

Menurut (Liu, 2012) analisis sentimen digunakan dalam beberapa kemungkinan domain seperti produk, jasa kesehatan, jasa keuangan, peristiwa sosial dan politik. Penelitian tentang analisis sentimen kecenderungan berfokus pada pendapat yang menyatakan atau menyiratkan suatu sentimen positif atau negatif. Opini atau pendapat yang dikeluarkan mewakili hampir semua aktivitas manusia, karena opini yang dikeluarkan dapat mempengaruhi perilaku seseorang. Sebagaimana setiap kita ingin membuat keputusan, kita melihat atau menayakan pendapat orang lain. Dalam aktivitas bisnis dan organisasi selalu memperhatikan opini publik tentang suatu produk atau jasa.

Menurut (Pang & Lee, 2008) analisis sentimen pada suatu kalimat menggambarkan bagian pertimbangan penilaian terhadap entitas atau kejadian tertentu. Entitas merupakan produk, layanan topik, isu, orang, organisasi atau peristiwa yang pada kalimat sentimen menjadi objek (Liu, 2012). Contoh permasalahan kalimat sentimen dalam kasus pemilihan kamera (Liu, 2012):

"(1) Saya membeli kamera Canon G12 enam bulan lalu. (2) Saya hanya menyukai itu. (3) Kualitas gambarnya menakjubkan. (4) Ketahanan pemakaian batrei juga menakjubkan. (5) Namun, istri saya berpikir itu terlalu berat baginya. "

Dari pendapat diatas dapat disimpulkan sebagai berikut (Liu, 2012):

- a. Tulisan diatas berpendapat tentang entitas yaitu kamera *Canon G12*. Pada kalimat kedua dan kalimat ketiga berpendapat positif terhadap kamera Canon G12, dimana poin positif terletak pada "*saya menyukai itu*" dan "*kualitas gambar menakjubkan*".
- b. Pada kalimat keempat juga berpendapat positif yaitu pendapat tentang ketahanan batrei.
- c. Pada kalimat kelima terdapat pendapat negatif yaitu tentang berat kamera.

Dari poin diatas dapat dilihat terdapat dua komponen yaitu (G, s), G adalah terget atau entitas dan s adalah pendapat atau opini bersifat positif, negatif, atau netral. Menurut (Liu, 2012) pengertian opini adalah (G, s, h, t), yaitu G merupakan opini atau pendapat sasaran, s merupakan opini atau pendapat tentang target, h merupakan pemegang pendapat dan t merupakan waktu kapan pendapat atau opini diungkapkan.

2.3 Text Preprocessing

Text preprocessing adalah proses awal untuk mempersiapkan teks menjadi data yang diolah lebih lanjut. Hal ini dapat dilakukan dalam beberapa tingkatan yang berbeda. Suatu dokumen dapat dipecah menjadi bab, sub-bab, paragraf, kalimat dan pada akhirnya menjadi potongan kata. Selain itu pada tahapan ini keberadaan digit angka, huruf kapital, atau karakter-karakter yang lainnya dihilangkan dan dirubah (Feldman & Sanger, 2007).

Berikut tahapan dari *text preprocessing*:

2.3.1 Cleaning

Cleaning merupakan proses pembersihan atribut-atribut yang tidak berhubungan dengan informasi yang ada pada data seperti *hashtag*, *URL*, *mention* dan *emoticon*.

2.3.2 Case Folding

Pada tahap ini mengubah semua huruf dalam kometer menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.

2.3.3 Tokenizing

Tahap *tokenizing* adalah tahapan pemotongan string input berdasarkan pada tiap kata yang menyusunnya.

2.3.4 Normalisasi

Normalisasi merupakan proses mengubah kata yang salah eja ataupun kata-kata tidak baku kedalam bahasa baku dalam kamus bahasa Indonesia.

2.3.5 Stemming

Stemming merupakan proses yang digunakan untuk menemukan kata dasar dari sebuah kata dengan menghilangkan semua imbuhan (*affixes*) baik yang merupakan awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*), dan kombinasi dari awalan serta akhiran (*cofixes*) pada kata turunan.

Algoritma *stemming* kata pada bahasa Indonesia dengan performa yang paling baik saat ini adalah algoritma *Enhanced Confix Stripping* (ECS) (Mahendra & Kerta, 2008). Algoritma ini diperkenalkan oleh Jelita Asian sebagai pengembangan dari algoritma Nazief dan Adriani pada tahun 1996. Algoritma ECS merupakan algoritma perbaikan dari algoritma *stemming confix stripping* (CS). Perbaikan yang dilakukan oleh *stemming* ECS adalah perbaikan beberapa aturan pada tabel acuan pemenggalan imbuhan. Selain itu, algoritma *stemming* ECS juga menambahkan langkah pengambilan akhiran jika terjadi penghilangan akhiran yang seharusnya tidak dilakukan (Tahitoe & Purwitasari, 2010).

$$[DP + [DP + [DP]]s]Kata\ Dasar\ [[+DS]s[+PP][+P]]$$

Keterangan :

DP (Derivational Prefixes): kumpulan awalan yang diberikan langsung pada kata dasar.

DS (Derivational Suffixes): kumpulan akhiran yang ditambahkan langsung pada kata dasar.

PP (Possessive Pronoun): Kata ganti kepemilikan.

Particle : Partikel.

Algoritma *Enhanced Confix Stripping* mempunyai tahapan sebagai berikut (B. Prasetyo & Susana, 2012):

1. Kata yang hendak di *stemming* dicari terlebih dahulu pada kamus. Jika ditemukan, berarti kata tersebut adalah kata dasar, jika tidak maka lakukan langkah 2.
2. Cek *rule precedence*. Apabila suatu kata memiliki pasangan awalan- akhiran “be-lah”, “be-an”, “me-i”, “di-i”, “pe-i”, atau “te-i” maka langkah *stemming* selanjutnya adalah (5,3, 4, 6). Apabila kata tidak memiliki pasangan awalan– akhiran tersebut, langkah *stemming* berjalan normal (3, 4, 5, 6).
3. Hilangkan *inflectional particle* P (“-lah”, “-kah”, “-tah”, “-pun”) dan kata ganti kepemilikan atau *possessive pronoun* PP (“-ku”, “-mu”, “-nya”).
4. Hilangkan *derivational suffixes* DS (“-i”, “-kan”, atau “-an”).
5. Hilangkan *Derivational Prefixes* DP (“di-”, “ke-”, “se-”, “me-”, “be-”, “pe-”, “te-”).
 - a. Identifikasi tipe awalan dan hilangkan. Awalan ada dua tipe:
 1. Standar: “di-”, “ke-”, “se-”, yang dapat langsung dihilangkan dari kata.
 2. Kompleks: “me-”, “be-”, “pe-”, “te-” adalah tipe-tipe awalan yang dapat bermorfologi sesuai kata dasar yang mengikutinya.
 - b. Cari kata yang telah dihilangkan awalnya ini di dalam kamus. Apabila tidak ditemukan, maka langkah 5 diulangi kembali. Apabila ditemukan, maka keseluruhan proses dihentikan.

6. Jika semua gagal, maka masukan kata yang diuji pada algoritma ini dianggap sebagai kata dasar.

2.3.6 Negation Handling

Untuk meningkatkan akurasi dalam klasifikasi juga dapat menerapkan *negation handling* dalam prosesnya, dimana apabila terdapat kata yang bersifat negasi seperti tidak, bukan, nggak dan lainnya, ini akan membuat makna kata berikutnya menjadi berlawanan dengan makna sesungguhnya. Berikut langkah-langkah *negation handling* menurut (Narayanan, Arora, & Bhatia, 2013).

1. Inisialisasi negasi = *False*
2. *For each* kata in dok

If negasi = *True*

Ubah kata menjadi ="kata_negasi"+"kata"

If kata = "tidak" or "bukan" or "nggak" or "belum"

Negasi = *True*

Else negasi = *False*

Proses *negation handling* ini yang dilakukan adalah mencari kata yang bersifat negasi, jika ditemukan kata negasi maka kata negasi tersebut digabungkan dengan kata selanjutnya dan selanjutnya digantikan dengan lawan kata setelah kata negasi tersebut.

2.3.7 Filtering

Proses *filtering* dilakukan untuk menghilangkan setiap *stopword* yang terkandung dalam kalimat yang ada pada dokumen. Kalimat yang akan diolah dalam sistem temu kembali informasi adalah kalimat yang tidak lagi mengandung *stopword*. *Stopword* adalah kata-kata yang sangat umum yang tidak terlalu membantu memilih dokumen yang sesuai dengan kebutuhan pengguna.

2.4 Feature Selection

Feature selection merupakan tahapan untuk mengurangi dimensi dari sebuah data tekstual sehingga hasil dari proses *Text Mining* memiliki kualitas yang lebih baik. (M.Thampi et al., 2018) Pemilihan fitur memainkan peran utama dalam penambahan atau pengambilan data dan tugas pembelajaran mesin. (M.Thampi et al., 2018) Telah menggunakan ambang varian dan teknik pemilihan fitur skor *laplacian* sebagai faktor evaluasi lebih lanjut. (M.Thampi et al., 2018) Memilih fitur berdasarkan ambang menggunakan kedua metode. *Threshold* (Ambang) yang dipilih adalah nilai rata-rata dalam kedua kasus. Kedua metode pemilihan fitur telah meningkatkan kualitas dari hasil dengan data *cancer_tweets_2017*. Tetapi metode pemilihan fitur ini menjadi tidak konsisten dengan pertumbuhan data, dan karenanya, (M.Thampi et al., 2018) menjelajahi kelayakannya kombinasi fitur yang dipilih menggunakan kedua metode pemilihan fitur. kombinasi fitur dieksplorasi dalam tiga cara:

1. *Union: set* (fitur yang dipilih menggunakan *variance threshold*) *union set* (fitur yang dipilih menggunakan skor *laplacian*).
2. *Persimpangan: set* (fitur yang dipilih menggunakan ambang batas varians) *set persimpangan* (fitur yang dipilih menggunakan skor *laplacian*).
3. *Kombinasi proporsional: persimpangan fitur bersama dengan persentase fitur yang berbeda dalam penyatuan fitur. Bersama dengan semua fitur persimpangan, menggabungkan 10,20,30,40 dan 50% dari fitur yang ada di serikat pekerja.*

Pada penelitian analisis sentimen pada maskapai Garuda Indonesia kali ini menggunakan fitur *threshold* dalam menyeleksi data analisis.

2.5 Pembobotan Kata

Dalam pembobotan kata terdapat beberapa metode yang dapat digunakan diantaranya adalah *document frequency* (DF), *term frequency* (TF), dan *inverse document frequency* (IDF). Pada penelitian ini *feature selection* yang digunakan adalah gabungan dari *term frequency* dan *inverse document frequency*.

1. *Document Frequency*

Dokumen frekuensi (DF) adalah jumlah dokumen yang mengandung suatu *term* tertentu. Dokumen frekuensi merupakan metode *feature selection* yang paling sederhana dengan waktu komputasi yang rendah (Herdi, 2018).

2. *Term Frequency*

Term frequency (Tf) merupakan salah satu metode untuk menghitung bobot tiap *term* dalam teks. Dalam metode ini tiap *term* diasumsikan memiliki nilai kepentingan yang sebanding dengan jumlah kemunculan *term* tersebut pada teks.

3. *Inverse Document Frequency* (IDF)

Inverse Document Frequency (IDF) merupakan metode untuk menghitung kemunculan *term* dalam keseluruhan koleksi teks. Dalam hal ini, *term* yang jarang muncul pada koleksi keseluruhan *term* dinilai lebih berharga. Nilai kepentingan tiap *term* diasumsikan berbanding terbalik dengan jumlah teks yang mengandung *term* tersebut.

4. *Term Frequency Inverse Document Frequency* (TF-IDF)

Pembobotan TF-IDF merupakan pembobotan pengukuran statistik yang digunakan untuk mengukur seberapa penting satu kata dalam kumpulan dokumen. Semakin sering suatu kata muncul dalam suatu dokumen maka kata tersebut dinilai penting dalam sebuah dokumen seiring dengan frekuensi kemunculan kata tersebut dalam kumpulan dokumen”(Wisnu & Hetami, 2015). Algoritma TF-IDF dirumuskan sebagai berikut:

$$TF - IDF(t_k, d_j) = TF(t_k, d_j) \times IDF(t_k) \quad (2.1)$$

Keterangan:

t_k = *term* ke-k

d_j = dokumen ke-j

Untuk mendapatkan nilai TF-IDF maka harus didapatkan nilai TF dan IDF terlebih dahulu. TF adalah *term frequency* yang berarti frekuensi kemunculan *term* dalam setiap dokumen, TF dirumuskan sebagai berikut:

$$TF(t_k, d_j) = f(t_k, d_j) \quad (2.2)$$

Keterangan:

TF = jumlah frekuensi *term*

t_k = *term* ke-k

d_j = dokumen ke-j

f = jumlah frekuensi kemunculan

Setelah didapatkan nilai TF maka harus didapatkan nilai IDF yang merupakan nilai bobot suatu *term* yang didapatkan dengan menghitung seringnya suatu term muncul di beberapa dokumen. IDF dihitung dengan menggunakan rumus berikut:

$$IDF(t_k) = \frac{1}{df(t)} \quad (2.3)$$

atau

$$IDF(t_k) = \log \frac{N}{df(t)} \quad (2.4)$$

Keterangan:

IDF = bobot *term*

N = jumlah semua dokumen

df = jumlah kemunculan dokumen

t_k = *term* ke-k

d_j = dokumen ke-j

Menurut penelitian yang dilakukan oleh (Salton & Buckley, 1988) kombinasi antara TF dan IDF untuk menghitung bobot *term* menunjukkan bahwa gabungan keduanya menghasilkan performansi yang lebih baik. Faktor TF dan IDF dapat berkontribusi untuk memperbaiki nilai *recall* dan *precision*.

2.6 Klasifikasi

Menurut (E. Prasetyo, 2012) klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi terdapat dua proses yang dilakukan yaitu dengan membangun model untuk disimpan sebagai memori dan menggunakan model tersebut

untuk melakukan pengenalan atau klasifikasi atau prediksi pada suatu data lain supaya diketahui di kelas mana objek data tersebut dimasukkan berdasarkan model yang telah disimpan dalam memori.

Beberapa metode klasifikasi yang telah digunakan oleh peneliti diantaranya adalah *Support Vector Machine* (Buntoro, 2016), *Naïve Bayes* (Darujati & Gumelar, 2012), dan *K-Nearest Neighbor* (Claudy, Perdana, & Fauzi, 2018). Dalam penelitian ini peneliti menggunakan metode *K-Nearest Neighbor* (KNN) untuk melakukan tahap klasifikasi.

2.6.1 Algoritma K-Nearest Neighbor

K-Nearest Neighbor adalah algoritma untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples* (data latih). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru (Krisandi, 2013). Data latih akan dibangun dengan memperhatikan keseimbangan dokumen satu sama lain.

Algoritma metode *K-Nearest Neighbor* (KNN) bekerja berdasarkan jarak terpendek dari query instance ke *training sample* untuk menentukan KNN-nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai kelas jika kelas merupakan klasifikasi yang paling banyak ditemui pada KNN dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*.

Jarak *Euclidean* paling sering digunakan menghitung jarak. Jarak *Euclidean* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek. Yang direpresentasikan sebagai berikut:

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2} \quad \dots \dots \dots (2.5)$$

Dimana matriks $D(a,b)$ adalah jarak scalar dari kedua vector a dan b dari matriks dengan ukuran dua dimensi. Semakin besar nilai D akan semakin jauh tingkat keserupaan antara kedua individu dan sebaliknya jika nilai D semakin kecil maka Akan semakin dekat tingkat keserupaan antar individu tersebut.

Langkah-langkah untuk menghitung metode *K-Nearest Neighbor* adalah sebagai berikut:

1. Menentukan parameter K (Jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclidean (query instance)* masing-masing obyek terhadap data sample yang diberikan.
3. Mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak *Euclid* terkecil.
4. Mengumpulkan kateogri Y (klasifikasi *nearest neighbor*).
5. Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksikan nilai *query instance* yang telah dihitung.

2.6.2 Confusion Matrix

Confusion matrix merupakan metode untuk menghitung tingkat akurasi, dengan menghitung jumlah prediksi benar dan salah dari sebuah metode klasifikasi berbanding dengan data sesungguhnya atau prediksi target.

Berikut ini contoh perhitungan akurasi menggunakan dua kelas yaitu positif dan negatif.

Tabel 2. 1 Confusion Matrix

<i>Confusion Matrix</i>		<i>Target</i>			
		<i>Positif</i>	<i>Negatif</i>		
<i>Model</i>	<i>Positif</i>	A	B	<i>Positif predictive value</i>	$a/(a+b)$
	<i>Negatif</i>	C	D	<i>Negatif predictive value</i>	$d/(c+d)$
		<i>Sensitivity</i> $a/(a+c)$	<i>Specificity</i> $d/(b+d)$	<i>Accuracy</i> = $(a+d)/(a+b+c+d)$	

Keterangan:

Accuracy : Jumlah prediksi benar (a+d)/ jumlah seluruh data (a+b+c+d)

Positive predictive Value/precision :jumlah prediksi benar positive (a)/ jumlah target data positive (a+b).

Negative predictive Value :jumlah prediksi benar positive (d)/ jumlah target data negative (b+d).

Sensitivity/ Recall :jumlah prediksi benar positif (a)/ jumlah seluruh prediksi benar (a+c).

Specificity :Jumlah prediksi salah negative (d) / jumlah seluruh prediksi salah (c+d).

2.7 Penelitian Terkait

Berikut penelitian terkait mengenai analisis sentimen yang akan digunakan oleh penulis dalam penelitian ini.

Tabel 2. 2 Penelitian Terkait

No	Penulis	Judul	Hasil
1.	(Kharisman, 2017)	Analisis Sentimen Pada <i>Review</i> Konsumen Makapai Penerbangan Menggunakan Kombinasi Lexicon Berbasis Sentiwordnet Dan Supervised Model	Hasil pengujian menunjukkan bahwa kata adjektif pada dokumen pelatihan sebagai data pelatihan memberikan persentase yang lebih besar dibandingkan menggunakan keseluruhan kata dalam dokumen pelatihan sebagai data pelatihan pada metode NBC dan SVM.
2	(Windasari, Uzzi, & Satoto, 2018)	Sentiment Analysis On Twitter Post: An Analysis Of Positive Or Negative Opinion On Gojek	Dari hasil pengujian didapatkan akurasi 86%, tingkat kesalahan prediksi 14%, tingkat prediksi yang benar untuk 100% positif sentimen, dan tingkat prediksi yang benar untuk sentimen negatif 67,44%.
3	(Nurjanah et al., 2017)	Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Pada Media Sosial Twitter Menggunakan Metode K-Nearest Neighborn Dan Pembobotan Jumlah Retweet	Hasil pengujian akurasi menggunakan pembobotan tekstual diperoleh 82,50%, menggunakan pembobotan non-tekstual 60%, dan menggunakan penggabungan keduanya 83,33% dengan nilai $k=3$ dan konstanta perkalian yang tepat $\alpha=0,8$ dan $\beta=0,2$.

No	Penulis	Judul	Hasil
4	(Mentari et al., 2018)	Analisis sentimen Kurikulum 2013 pada sosial media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking	Hasil penelitian ini memiliki 4 proses utama yang dilakukan sistem yaitu, <i>text pre-processing</i> , kemudian melakukan perhitungan <i>term weighting</i> (TF-IDF) pada seluruh fitur yang ada pada data uji dan latih, melakukan <i>feature selection</i> dengan menggunakan metode <i>Query Expansion Ranking</i> , dan klasifikasi dengan <i>K-Nearest Neighbor</i> pada setiap dokumen uji apakah termasuk kelas “opini positif” atau “opini negatif”.
5	(Wati, 2016)	Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen <i>Review</i> Jasa Maskapai Penerbangan Menggunakan Naive Bayes	Penelitian ini menghasilkan akurasi 60.00% dan AUC sebesar 0.512 sedangkan akurasi setelah menggunakan pemilihan fitur Algoritma Genetika sebesar 89.50 dan AUC sebesar 0.919 termasuk kedalam <i>excellent classification</i> terjadi peningkatan akurasi sebesar 29.5% dan AUC sebesar 0.407.
6	(Indriati & Ridok, 2016)	Sentiment Analysis For <i>Review</i> Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwkn)	Berdasarkan hasil pengujian pengaruh data latih tidak seimbang terhadap akurasi metode NWKNN dengan membandingkan metode NWKNN dan KNN, didapatkan bahwa metode NWKNN lebih baik 0.27 dibandingkan metode KNN dalam melakukan klasifikasi terhadap dokumen <i>review</i> aplikasi mobile dikarenakan adanya penambahan bobot pada klas minoritas sehingga kesalahan klasifikasi data pada klas minoritas dapat dikurangi.
7	(Putri et al., 2014)	Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012	Status pekerjaan dari metode pemrosesan data Naïve Bayes dengan akurasi yang diperoleh adalah sebesar 94,09% dan metode K-Nearest Neighbor yang diperoleh adalah dengan akurasi 96,06%. Untuk mengevaluasi hasil klasifikasi digunakan perhitungan Tekan Q dan APER. Berdasarkan analisis, nilai Q Press yang diperoleh menunjukkan bahwa kedua metode sudah baik dalam klasifikasi data status pekerjaan di Demak. Berdasarkan perhitungan APER, klasifikasi data dalam status pekerjaan Demak menggunakan metode K-Nearest Neighbor memiliki tingkat kesalahan lebih kecil dari metode Naïve Bayes. Dari analisis ini dapat disimpulkan bahwa metode K-Nearest Neighbor bekerja lebih baik dibandingkan dengan Naïve Bayes untuk data status pekerjaan dalam kasus Demak untuk 2012.

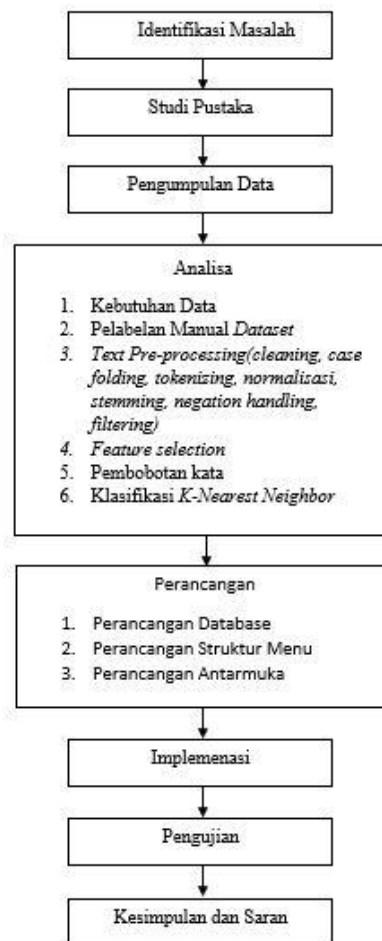
No	Penulis	Judul	Hasil
8	(Utami, 2017)	Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma <i>Support Vector Machine</i> Dan <i>K-Nearest Neighbor</i> Berbasis <i>Particle Swarm Optimization</i>	Dalam penelitian ini, hasil perhitungan metode SVM memiliki <i>Accuracy</i> sebesar 80.83% dan AUC sebesar 0.947 sedangkan Metode SVM+PSO menghasilkan <i>Accuracy</i> sebesar 86.11% dan AUC sebesar 0.922. Pengujian juga telah dibandingkan dengan metode k-NN. Hasil perhitungan yang diperoleh dari pengujian data dengan metode k-NN yaitu <i>Accuracy</i> sebesar 85.00% dan AUC sebesar 0.918, kemudian dibandingkan dengan k-NN+PSO menghasilkan nilai <i>Accuracy</i> sebesar 73.06% dan AUC sebesar 0.500.
9	(Lidya, Sitompul, & Efendi, 2015)	<i>Sentiment Analysis</i> Pada Teks Bahasa Indonesia Menggunakan <i>Support Vector Machine (Svm)</i> Dan <i>K-Nearest Neighbor (K-NN)</i>	Berdasarkan seluruh hasil pengujian, bahwa pengaruh nilai k pada k-fold cross validation yang terlalu kecil menghasilkan akurasi yang rendah, sedangkan nilai k yang terlalu besar menghasilkan nilai akurasi yang besar, kemudian Pengaruh nilai K pada KNN terhadap akurasi, jika n memiliki akurasi rendah pada saat nilai k kecil. Hal ini dikarenakan, data yang masuk pada k tetangga terdekat terlalu sedikit dan tidak dapat merepresentasikan kelas pada data uji.
10	(Rosdiansyah, 2014)	Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor Dan Pendekatan Lexicon	Sistem telah dapat menentukan orientasi sentimen pada media Twitter dengan tingkat akurasi k-NN adalah 70 %, Lexicon 77 % , Kombinasi A 78 % dan Kombinasi B 82 %
11	(Chandani & Wahono, 2015)	Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen <i>Review Film</i>	Hasil dari komparasi <i>feature selection</i> , information gain mendapatkan hasil yang paling baik dengan <i>average accuracy</i> 84.57% dan <i>average AUC</i> 0.899. Hasil integrasi algoritma klasifikasi terbaik dan algoritma <i>feature selection</i> terbaik menghasilkan <i>accuracy</i> 81.50% dan AUC 0.929.

No	Penulis	Judul	Hasil
12	(Claudy et al., 2018)	Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritme K-Nearest Neighbor (KNN)	Data yang di dapatkan dari perusahaan berupa tweet dari calon karyawan ini agar menghasilkan hasil pengujian yang baik, maka dibagi menjadi dua jenis dengan rasio 50% untuk data latih dan 50% untuk data uji. Dengan memasukkan Nilai K yaitu 4 sebagai nilai yang diuji. maka di dapatkan Hasil akurasi sistem yang diperoleh dari klasifikasi karakter calon karyawan berdasarkan tweet adalah 66%. Hasil ini adalah hasil dimana ada 53 hasil data uji yang benar dan 27 hasil data uji yang salah dalam proses pengujian.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian adalah prosedur digunakan untuk mencapai tujuan dari suatu penelitian. Berikut tahapan-tahapan penelitian yang akan dilakukan agar tujuan penelitian dapat tercapai. Pada penelitian ini, terdapat beberapa tahapan penelitian yang dijabarkan pada gambar 3.1 berikut ini.



Gambar 3. 1 Tahapan Metodologi Penelitian

3.1 Identifikasi Masalah

Pada tahapan ini berisikan latar belakang, rumusan masalah, dan batasan masalah. Latar belakang merupakan landasan dari penelitian ini dibuat. Kemudian rumusan masalah merupakan permasalahan yang dibahas pada latar belakang. Kemudian batasan masalah merupakan batasan yang buat untuk membatasi penelitian ini.

3.2 Studi Pustaka

Studi Pustaka ialah tahapan setelah dari identifikasi masalah. Pada tahapan ini menjelaskan bagaimana mengumpulkan data dan informasi dari banyak sumber. Tahapan ini dilaksanakan dengan maksud mencari referensi yang sesuai dengan permasalahan yang diteliti. Referensi yang ditemukan dapat berupa *paper*, jurnal, buku dan beberapa situs *website* dengan topik terkait.

3.3 Pengumpulan Data

Pengumpulan data merupakan tahap yang bertujuan untuk memperoleh data-data dan informasi yang berhubungan dengan penelitian yang akan dilakukan. Data yang diambil dari komentar pada media sosial Twitter sebanyak 1000 komentar yang akan dimasukkan ke dalam *Database*.

3.4 Analisa

Setelah melakukan langkah-langkah di atas maka langkah selanjutnya dilakukan tahapan analisa sehingga akan dapat diketahui tentang gambaran yang jelas mengenai penelitian yang dilakukan. Adapun rincian analisisnya sebagai berikut:

3.4.1 Kebutuhan Data

Analisa kebutuhan data merupakan suatu proses menganalisa data yang telah dikumpulkan. Jumlah dataset yang akan digunakan yaitu 1000 komentar dengan pembagian 70% data latih 30% data uji, 80% data uji 20% data latih, dan 90% data

latih 10% data uji. Tujuan dari analisa kebutuhan data yaitu akan memudahkan dalam proses pengolahan data untuk diolah pada tahap selanjutnya.

3.4.2 Pelabelan Manual Dataset

Melakukan pelabelan manual yang dilakukan oleh peneliti untuk menyelesaikan permasalahan klasifikasi pada komentar agar data dapat masuk ke dalam kelompok yang benar berdasarkan informasi yang dikandungnya.

3.4.3 Text Pre-processing

Pada tahap ini merupakan langkah awal setelah melakukan proses analisa data. Pada *text pre-processing* ini akan dijelaskan proses dari pengolahan data untuk diolah pada proses selanjutnya. Berikut merupakan langkah-langkah pada proses *text pre-processing*:

1. Cleaning

Pada tahap cleaning dilakukan pembersihan atribut-atribut yang tidak berhubungan dengan informasi yang ada pada data seperti *hashtag*, *URL*, *mention* dan *emoticon*.

2. Case Folding

Pada tahap ini semua *tweet* akan dirubah menjadi huruf kecil. Karakter yang diproses pada tahap *case folding* yaitu "a" sampai "z".

3. Tokenizing

Tokenizing merupakan proses pemisahan antar kata pada tiap kata yang menyusunnya. Setiap huruf pada kata akan diubah menjadi huruf kecil dan selain huruf alfabet akan dihapus seperti tanda baca.

4. Normalisasi

Normalisasi merupakan proses mengubah kata yang salah eja ataupun kata-kata tidak baku kedalam bahasa baku dalam kamus bahasa Indonesia seperti kata “ngeliat” menjadi “melihat”.

5. *Stemming*

Stemming dilakukan untuk mengubah semua kata menjadi kata dasar.

Adapun contohnya yaitu "memakan" menjadi "makan".

6. *Negation Handling*

Proses *negation handling* ini yang dilakukan adalah mencari kata yang bersifat negasi, jika ditemukan kata negasi maka kata negasi tersebut digabungkan dengan kata selanjutnya dan selanjutnya digantikan dengan lawan kata setelah kata negasi tersebut. Adapun contohnya yaitu “tidak suka”, menjadi “benci”.

7. *Filtering*

Merupakan proses penghapusan kata yang termasuk ke dalam *stopword*, seperti “dan”, “di”, “dari”, “dengan” dan lain-lain.

3.4.4 **Feature Selection**

Feature selection digunakan untuk meningkatkan akurasi pada penelitian ini dengan menyeleksi hasil bobot kata yang telah di dapatkan, *feature selection* yang digunakan pada penelitian ini menggunakan *threshold* atau ambang batas.

3.4.5 **Pembobotan Kata**

Pada proses pembobotan kata, akan mencari dan menghitung tiap kata dalam komentar yang kemudian di cari jumlah bobot kata tersebut dengan menggunakan TF-IDF pada rumus (2.1)

3.4.6 **Klasifikasi**

Tahapan pada klasifikasi K-Nearest Neighbor adalah sebagai berikut:

1. Menentukan parameter K (Jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclidean* (query instance) masing-masing obyek terhadap data sample yang diberikan.
3. Mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak Euclid terkecil.

4. Mengumpulkan kateogri Y (klasifikasi *nearest neighbor*).
5. Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksikan nilai *query instance* yang telah dihitung.

3.5 Perancangan

Dalam tahapan perancangan terdapat tiga tahapan yaitu:

1. Perancangan struktur menu yang akan dibuat dan digunakan.
2. Tahapan rancangan *database*, yakni tahapan merancang tabel dan atribut yang dibutuhkan.
3. Tahapan merancang antarmuka atau *user interface* yang merupakan suatu bentuk perancangan menu dan tampilan yang akan digunakan.

3.6 Implementasi

Tahapan implementasi adalah tahapan dimana sistem akan dijalankan dan digunakan. Berikut penjelasan mengenai perangkat keras dan perangkat lunak yang digunakan:

1. Perangkat Keras (*Hardware*)

Processor : Intel(R) Core (TM) i3-5005U CPU @ 2.00 GHz 2.00 GHz
System Type : 64-bit *Operating System*, x64-based *processor*
RAM : 4,00 GB
Harddisk : 500 GB

2. Perangkat Lunak (*Software*)

Sistem Operasi : *Microsoft Windows 10 Enterprise*
Bahasa Pemrograman : PHP
Text Editor : *Sublime Text 3*
Browser : *Google Chrome*

3.7 Pengujian

Untuk melakukan pengujian terhadap sistem, dilakukan evaluasi akurasi sistem dalam mengklasifikasikan sentimen pada dataset dengan menggunakan *white box* dan *confusion matrix* pada tabel 2.1.

3.8 Kesimpulan Dan Saran

Tahapan ini berisi tentang kesimpulan yang dapat diambil dari penelitian yang telah dilakukan. Pada bagian saran berisi saran-saran yang penulis berikan untuk mengembangkan aplikasi agar kedepannya menjadi lebih baik.

BAB IV

ANALISA DAN PERANCANGAN

Tahap analisa dan perancangan akan menentukan hasil dari penelitian ini. Analisa berisi tentang tahapan-tahapan yang akan dilakukan pada aplikasi yang bertujuan untuk melakukan proses analisis sentimen menggunakan metode *K-Nearest Neighbor*. Selanjutnya, proses perancangan dilakukan untuk merancang aplikasi yang akan dibuat berdasarkan analisa yang telah dilakukan.

4.1 Analisa

Pada proses analisa ini dilakukan proses-proses analisis sentimen meliputi analisa kebutuhan data, pelabelan data, *text pre-processing*, *feature selection*, pembobotan kata, dan klasifikasi *K-Nearest Neighbor*.

4.1.1 Analisa Kebutuhan Data

Penelitian ini akan mengklasifikasikan sentimen-sentimen pada maskapai Garuda Indonesia, dimana sentimen tersebut merupakan sentimen negatif atau positif dengan menggunakan metode KNN dimana data sentimen yang digunakan dari media sosial Twitter. Data yang di butuhkan pada penelitian ini adalah sentimen dalam bahasa Indonesia dan tentang maskapai Garuda Indonesia, data yang ambil sejumlah 1000 data dengan pembagian 500 sentimen positif dan 500 sentimen negatif. Proses pengambilan data menggunakan Twitter API yang kemudian dimasukkan kedalam basis data.

4.1.2 Analisa pelabelan manual data

Pada proses pelabeian manual ,data sentimen akan dilabelkan untuk kebutuhan pelatihan data pada penelitian ini, dimana data tersebut sebagai acuan atau pedoman dalam menentukan proses klasifikasi data yang akan di uji. Berikut contoh data dan pelabelan data pada media sosial Twitter tentang maskapai Garuda Indonesia.

Tabel 4. 1 Kebutuhan Data Dan Pelabelan

No	Dataset	Pelabelan data
1	Garuda Indonesia selalu menjadi pilihan utama penerbangan keluarga saya 😎	Positif
2	Payah @IndonesiaGaruda sdh tdk puja urat malu itu pesawat punya negara bukan punya kokowi knp dicorat coret pakai nama seseorang.	Negatif
3	Garuda Indonesia keren 👍	Positif
4	Trimakasi @IndonesiaGaruda captionnya menyejukkan hati sekali pas...tiada yg membahagiakan selain bisa berkumpul bersama keluarga dan saling memaafkan	Negatif
5	Selalu memilih mudik dgn @IndonesiaGaruda tp tahun ini campaign #PulangNyaman dgn penambahan jumlah armada & jadwal GA hanya sebatas campaign. Flight kita yg tadinya lebih awal diresechedule seenaknya ke last flight. Kita milih jadwal jg udah ada pertimbangannya Bambang, Maimunah	Negatif
6	Alhamdulillah,,, mantap,,, maju dan jaya terus,,, Hutang" juga segera lunas, agar makin sehat BUMN ini	Positif

4.1.3 Analisa Text Pre-processing

Pada analisa text pre-processing data yang sudah dikumpulkan akan diolah agar memudahkan untuk proses selanjutnya. Berikut tahapan yang dilakukan pada Text pre-processing.

1. Cleaning

Pada proses cleaning akan menghilangkan atribut pada data, seperti karakter atau simbol (!@#\$%^&*():{ }.,?~/[]), angka, *hashtag*, *URL*, *mention*, *emoticon*. Tabel (.....) merupakan contoh hasil dari tahapan *cleaning*.

Tabel 4. 2 Cleaning

D	Data	Cleaning
D1	Garuda Indonesia selalu menjadi pilihan utama penerbangan keluarga saya 😎	Garuda Indonesia selalu menjadi pilihan utama penerbangan keluarga saya

D	Data	Cleaning
D2	Payah @IndonesiaGaruda sdh tdk puja urat malu itu pesawat punya negara bukan punya kokowi knp dicorat coret pakai nama seseorang.	Payah sdh tdk puja urat malu itu pesawat punya negara bukan punya kokowi knp dicorat coret pakai nama seseorang
D3	Garuda Indonesia keren 👍	Garuda Indonesia keren
D4	Trimakasi @IndonesiaGaruda captionnya menyejukkan hati sekali pas...tiada yg membahagiakan selain bisa berkumpul bersama keluarga dan saling memaafkan	Trimakasi captionnya menyejukkan hati sekali pas tiada yg membahagiakan selain bisa berkumpul bersama keluarga dan saling memaafkan
D5	Selalu memilih mudik dgn @IndonesiaGaruda tp tahun ini campaign #PulangNyaman dgn penambahan jumlah armada & jadwal GA hanya sebatas campaign. Flight kita yg tadinya lebih awal diresechedule seenaknya ke last flight. Kita milih jadwal jg udah ada pertimbangannya Bambang, Maimunah	Selalu memilih mudik dgn tp tahun ini campaign dgn penambahan jumlah armada jadwal GA hanya sebatas campaign Flight kita yg tadinya lebih awal diresechedule seenaknya ke last flight. Kita milih jadwal jg udah ada pertimbangannya Bambang Maimunah
D6	Alhamdulillah,,, mantap,,, maju dan jaya terus,,, Hutang" juga segera lunas, agar makin sehat BUMN ini	Alhamdulillah mantap maju dan jaya terus Hutang juga segera lunas agar makin sehat BUMN ini

2. Case Folding

Case folding yang merupakan proses mengubah semua huruf yang ada dalam dokumen menjadi huruf kecil. Karakter yang diproses pada tahap *case folding* yaitu "a" sampai "z". Berikut merupakan hasil dari case folding.

Tabel 4. 3 Case Folding

D	Cleaning	Case folding
D1	Garuda Indonesia selalu menjadi pilihan utama penerbangan keluarga saya	garuda indonesia selalu menjadi pilihan utama penerbangan keluarga saya
D2	Payah sdh tdk puja urat malu itu pesawat punya negara bukan punya kokowi knp dicorat coret pakai nama sesorang	payah sdh tdk puja urat malu itu pesawat punya negara bukan punya kokowi knp dicorat coret pakai nama sesorang
D3	Garuda Indonesia keren	garuda indonesia keren
D4	Trimakasi captionnya menyejukkan hati sekali pas tiada yg membahagiakan selain bisa berkumpul bersama keluarga dan saling memaafkan	trimakasi captionnya menyejukkan hati sekali pas tiada yg membahagiakan selain bisa berkumpul bersama keluarga dan saling memaafkan
D5	Selalu memilih mudik dgn tp tahun ini campaign dgn penambahan jumlah armada jadwal GA hanya sebatas campaign Flight kita yg tadinya lebih awal diresechedule seenaknya ke last flight Kita milih jadwal jg udah ada pertimbangannya Bambang Maimunah	selalu memilih mudik dgn tp tahun ini campaign dgn penambahan jumlah armada jadwal ga hanya sebatas campaign flight kita yg tadinya lebih awal diresechedule seenaknya ke last flight kita milih jadwal jg udah ada pertimbangannya bambang maimunah
D6	Alhamdulillah mantap maju dan jaya terus Hutang juga segera lunas agar makin sehat BUMN ini	alhamdulillah mantap maju dan jaya terus hutang juga segera lunas agar makin sehat bumn ini

3. Tokenizing

Setelah melakukan *case folding* dan *cleaning* pemrosesan teks dilakukan dengan *tokenizing* yang berarti sebuah proses pemotongan kalimat menjadi perkata yang menyusun kalimat tersebut. Berikut merupakan conoth hasil dari tokenizing.

Tabel 4. 4 Tokenizing

D	Case Folding	Tokenizing
D1	garuda indonesia selalu menjadi pilihan utama penerbangan keluarga saya	garuda
		indonesia
		selalu
		menjadi
		pilihan
		utama
		penerbangan
		keluarga
		saya
D2	payah sdh tdk pujya urat malu itu pesawat punya negara bukan punya kokowi knp dicorat coret pakai nama sesorang	payah
		sdh
		tdk
		pujya
		urat
		malu
		itu
		pesawat
		punya
		negara
		bukan
		punya
		kokowi
		knp
		dicoret
		coret
		pakai
		nama
		sesorang
D3	garuda indonesia keren	garuda
		indonesia
		keren
D4	trimakasi captionya menyejukkan hati sekali pas tiada yg membahagiakan selain bisa berkumpul bersama keluarga dan saling memaafkan	trimakasi
		captionya
		menyejukkan
		hati
		sekali
		pas
		tiada
		yg
		membahagiakan

D	Case Folding	Tokenizing
D4		selain
		bisa
		berkumpul
		bersama
		keluarga
		dan
		saling
		memaafkan
D5	selalu memilih mudik dgn tp tahun ini campaign dgn penambahan jumlah armada jadwal ga hanya sebatas campaign flight kita yg tadinya lebih awal diresechedule seenaknya ke last flight kita milih jadwal jg udah ada pertimbangannya bambang maimunah	selalu
		memilih
		mudik
		dgn
		tp
		tahun
		ini
		campaign
		dgn
		penambahan
		jumlah
		armada
		jadwal
		ga
		hanya
		sebatas
		campign
		flight
		kita
		yg
		tadinya
		lebih
		awal
		diresechedule
		seenaknya
		ke
		last
		fight
		kita
		milih
		jadwal
		jg
		udah

D	Case Folding	Tokenizing
D5		ada
		pertimbangannya
		bambang
		maimunah
D6	alhamdulillah mantap maju dan jaya terus hutang juga segera lunas agar makin sehat bumh ini	alhamdulillah
		mantap
		maju
		dan
		jaya
		terus
		hutang
		juga
		segera
		lunas
		agar
		makin
		sehat
		bumh
		ini

4. Normalisasi

Normalisasi merupakan proses mengubah kata yang salah eja ataupun kata-kata tidak baku kedalam bahasa baku menggunakan kamus normalisasi dari penelitian sebelumnya. Berikut contoh normalisasi.

Tabel 4. 5 Normalisasi

D	Tokenizing	Normalisasi
D1	garuda	garuda
	indonesia	indonesia
	selalu	selalu
	menjadi	menjadi
	pilihan	pilihan
	utama	utama
	penerbangan	penerbangan
	keluarga	keluarga
	saya	saya

D	Tokenizing	Normalisasi
D2	payah	payah
	sdh	sudah
	tdk	tidak
	pujya	punya
	urat	urat
	malu	malu
	itu	itu
	pesawat	pesawat
	punya	punya
	negara	negara
	bukan	bukan
	punya	punya
	kokowi	jokowi
	knp	kenapa
	dicoret	dicoret
	coret	coret
	pakai	pakai
	nama	nama
	sesorang	sesesorang
D3	garuda	garuda
	indonesia	indonesia
	keren	keren
D4	trimakasi	terima kasih
	captionya	captionnya
	menyejukkan	menyejukkan
	hati	hati
	sekali	sekali
	pas	pas
	tiada	tiada
	yg	yang
	membahagiakan	membahagiakan
	selain	selain
	bisa	bisa
	berkumpul	berkumpul
	bersama	bersama
	keluarga	keluarga
	dan	dan
	saling	saling
	memaafkan	memaafkan
D5	selalu	selalu
	memilih	memilih
	mudik	mudik

D	Tokenizing	Normalisasi
D5	dgn	dengan
	tp	tapi
	tahun	tahun
	ini	ini
	campaign	campaign
	dgn	dengan
	penambahan	penambahan
	jumlah	jumlah
	armada	armada
	jadwal	jadwal
	ga	tidak
	hanya	hanya
	sebatas	sebatas
	campign	campign
	flight	flight
	kita	kita
	yg	yang
	tadinya	tadinya
	lebih	lebih
	awal	awal
	diresechedule	diresechedule
	seenaknya	seenaknya
	ke	ke
	last	last
	fight	fight
	kita	kita
	milih	milih
	jadwal	jadwal
	jg	juga
	udah	udah
	ada	ada
	pertimbangannya	pertimbangannya
	bambang	bambang
	maimunah	maimunah
D6	alhamdulillah	alhamdulillah
	mantap	mantap
	maju	maju
	dan	dan
	jaya	jaya
	terus	terus
	hutang	hutang
	juga	juga

D	Tokenizing	Normalisasi
D6	segera	segera
	lunas	lunas
	agar	agar
	makin	makin
	sehat	sehat
	bumn	badan
	ini	usaha
		milik
		negara
		ini

5. Stemming

Stemming merupakan proses yang digunakan untuk menemukan kata dasar dari sebuah kata dengan menghilangkan semua imbuhan (*affixes*) baik yang merupakan awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*), dan kombinasi dari awalan serta akhiran (*cofixes*) pada kata turunan. Berikut contoh hasil dari stemming menggunakan ECS.

Tabel 4. 6 Steaming ECS

D	Normalisasi	Stemming
D1	garuda	garuda
	indonesia	indonesia
	selalu	selalu
	menjadi	jadi
	pilihan	pilih
	utama	utama
	penerbangan	terbang
	keluarga	keluarga
	saya	saya
D2	payah	payah
	sudah	sudah
	tidak	tidak
	punya	punya
	urat	urat
	malu	malu
	itu	itu
	pesawat	pesawat

D	Normalisasi	Stemming
	punya	punya
	negara	negara
	bukan	bukan
	punya	punya
	jokowi	jokowi
	kenapa	kenapa
	dicoret	coret
	coret	coret
	pakai	pakai
	nama	nama
	sesesorang	orang
D3	garuda indonesia	garuda
	Indonesia	indonesia
	keren	keren
D4	terima kasih	terima kasih
	captionnya	caption
	menyejukkan	sejuk
	hati	hati
	sekali	sekali
	pas	pas
	tiada	tiada
	yang	yang
	membahagiakan	bahagia
	selain	selain
	bisa	bisa
	berkumpul	kumpul
	bersama	sama
	keluarga	keluarga
	dan	dan
	saling	saling
	memaafkan	maaf
D5	selalu	selalu
	memilih	pilih
	mudik	mudik
	dengan	dengan
	tapi	tapi
	tahun	tahun
	ini	ini
	campaign	campaign
	dengan	dengan
	penambahan	tambah

D	Normalisasi	Stemming
D5	jumlah	jumlah
	armada	armada
	jadwal	jadwal
	tidak	tidak
	hanya	hanya
	sebatas	batas
	campign	campign
	flight	flight
	kita	kita
	yang	yang
	tadinya	tadi
	lebih	lebih
	awal	awal
	diresechedule	resechedule
	seenaknya	enak
	ke	ke
	last	last
	fight	flight
	kita	kita
	milih	pilih
	jadwal	jadwal
	juga	juga
	udah	sudah
	ada	ada
	pertimbangannya	timbang
	bambang	bambang
	maimunah	maimunah
D6	alhamdulillah	alhamdulillah
	mantap	mantap
	maju	maju
	dan	dan
	jaya	jaya
	terus	terus
	hutang	hutang
	juga	juga
	segera	segera
	lunas	lunas
	agar	agar
	makin	makin
	sehat	sehat
	badan	badan

D	Normalisasi	Stemming
D6	usaha	usaha
	milik	milik
	negara	negara
	ini	ini

6. Negation Handling

Proses *negation handling* ini yang dilakukan adalah mencari kata yang bersifat negasi, jika ditemukan kata negasi maka kata negasi tersebut digabungkan dengan kata selanjutnya dan selanjutnya digantikan dengan lawan kata setelah kata negasi tersebut. Berikut merupakan contoh hasil dari *negation handling*.

Tabel 4. 7 Negation Handling

D	Stemming	Negation handling
D1	garuda	garuda
	indonesia	indonesia
	selalu	selalu
	jadi	jadi
	pilih	pilih
	utama	utama
	terbang	terbang
	keluarga	keluarga
	saya	saya
D2	payah	payah
	sudah	sudah
	tidak	tiada
	punya	
	urat	urat
	malu	malu
	itu	itu
	pesawat	pesawat
	punya	punya
	negara	negara
	bukan	tiada
	punya	
	jokowi	jokowi
	kenapa	kenapa

D	Stemming	Negation handling
D2	coret	coret
	coret	coret
	pakai	pakai
	nama	nama
	orang	orang
D3	garuda	garuda
	indonesia	indonesia
	keren	keren
D4	terima kasih	terima kasih
	caption	caption
	sejuk	sejuk
	hati	hati
	sekali	sekali
	pas	pas
	tiada	tiada
	yang	yang
	bahagia	bahagia
	selain	selain
	bisa	bisa
	kumpul	kumpul
	sama	sama
	keluarga	keluarga
	dan	dan
	saling	saling
	maaf	maaf
D5	selalu	selalu
	pilih	pilih
	mudik	mudik
	dengan	dengan
	tapi	tapi
	tahun	tahun
	ini	ini
	campaign	campaign
	dengan	dengan
	tambah	tambah
	jumlah	jumlah
	armada	armada
	jadwal	jadwal

D	Stemming	Negation handling
D5	tidak	semua
	hanya	
	batas	batas
	campaign	campaign
	flight	flight
	kita	kita
	yang	yang
	tadi	tadi
	lebih	lebih
	awal	awal
	resechedule	resechedule
	enak	enak
	ke	ke
	last	last
	flight	flight
	kita	kita
	pilih	pilih
	jadwal	jadwal
	juga	juga
	sudah	sudah
	ada	ada
	timbang	timbang
	bambang	bambang
	maimunah	maimunah
D6	alhamdulillah	alhamdulillah
	mantap	mantap
	maju	maju
	dan	dan
	jaya	jaya
	terus	terus
	hutang	hutang
	juga	juga
	segera	segera
	lunas	lunas
	agar	agar
	makin	makin
	sehat	sehat
	badan	badan
	usaha	usaha
	milik	milik
	negara	negara

7. Filtering

Proses *filtering* dilakukan untuk menghilangkan setiap *stopword* yang terkandung dalam kalimat yang ada pada dokumen. *Stopword* adalah kata-kata yang sangat umum yang tidak membantu memilih dokumen yang sesuai dengan kebutuhan pengguna. Berikut hasil dari proses filtering.

Tabel 4. 8 Filtering

D	Negation Handling	filtering
D1	garuda	garuda
	indonesia	indonesia
	selalu	
	jadi	
	pilih	pilih
	utama	
	terbang	terbang
	keluarga	keluarga
	saya	
D2	payah	payah
	sudah	
	tiada	tiada
	urat	urat
	malu	malu
	itu	
	pesawat	pesawat
	punya	
	negara	negara
	tiada	tiada
	jokowi	jokowi
	kenapa	
	coret	coret
	coret	coret
	pakai	pakai
	nama	
	orang	
D3	garuda	garuda
	indonesia	indonesia
	keren	keren

D	Negation Handling	filtering
D4	terima kasih	terima kasih
	caption	
	sejuk	sejuk
	hati	hati
	sekali	
	pas	pas
	tiada	tiada
	yang	
	bahagia	bahagia
	selain	
	bisa	
	kumpul	kumpul
	sama	
	keluarga	keluarga
	dan	
	saling	
	maaf	maaf
D5	selalu	
	pilih	pilih
	mudik	mudik
	dengan	
	tapi	
	tahun	
	ini	
	campaign	
	dengan	
	tambah	
	jumlah	
	armada	armada
	jadwal	jadwal
	semua	
	batas	batas
	campign	
	flight	
	kita	
	yang	
	tadi	

D	Negation Handling	filtering
D5	lebih	
	awal	
	resechedule	
	enak	enak
	ke	
	last	
	flight	
	kita	
	pilih	pilih
	jadwal	jadwal
	juga	
	sudah	
	ada	
	timbang	timbang
	bambang	bambang
	maimunah	
D6	alhamdulillah	alhamdulillah
	mantap	mantap
	maju	maju
	dan	
	jaya	jaya
	terus	
	hutang	hutang
	juga	
	segera	
	lunas	lunas
	agar	
	makin	
	sehat	sehat
	badan	badan
	usaha	usaha
	milik	milik
	negara	negara
	ini	

4.1.4 Analisa Feature Selection

Pada penelitian analisis sentimen pada maskapai Garuda Indonesia kali ini menggunakan fitur *threshold* dalam menyeleksi data analisis. Threshold yang digunakan untuk memberi ambang batas pada bobot data yang akan digunakan pada analisis pembobotan yaitu >0 dan >-1 . Tujuan dari feature selection untuk meningkatkan akurasi dengan menyeleksi bobot yang dianggap menyimpang.

4.1.5 Analisa Pembobotan Kata

Pada proses analisa pembobotan kata, setiap kata yang dihasil dari proses sebelumnya akan di berikan bobot. Pada penelitian ini pembobotan kata menggunakan tf-idf. Kata yang digunakan dari kumpulan data yang telah melewati proses *cleaning*, *case folding*, *tokenizing*, normalisasi, *stemming*, *negation handling*, dan *filtering*. Berikut merupakan hasil dari pembobotan kata.

Tabel 4. 9 Pembobotan Kata

NO	Kata	tf						DF	idf	Tf-idf					
		D1	D2	D3	D4	D5	D6			D1	D2	D3	D4	D5	D6
1	garuda	1		1				2	0.477121	0.477121	0	0.477121	0	0	0
2	indonesia	1		1				2	0.477121	0.477121	0	0.477121	0	0	0
3	pilih	1				2		2	0.477121	0.477121	0	0	0	0.954243	0
4	terbang	1						1	0.778151	0.778151	0	0	0	0	0
5	keluarga	1			1			2	0.477121	0.477121	0	0	0.477121	0	0
6	payah		1					1	0.778151	0	0.778151	0	0	0	0
7	tiada		2		1			2	0.477121	0	0.954243	0	0.477121	0	0
8	urat		1					1	0.778151	0	0.778151	0	0	0	0
9	malu		1					1	0.778151	0	0.778151	0	0	0	0

NO	Kata	Tf						DF	Idf	Tf-Idf					
		D1	D2	D3	D4	D5	D6			D1	D2	D3	D4	D5	D6
10	pesawat		1					1	0.778151	0	0.778151	0	0	0	0
11	negara		1				1	2	0.477121	0	0.477121	0	0	0	0.477121
12	jokowi		1					1	0.778151	0	0.778151	0	0	0	0
13	coret		2					1	0.778151	0	1.556303	0	0	0	0
14	pakai		1					1	0.778151	0	0.778151	0	0	0	0
15	keren			1				1	0.778151	0	0	0.778151	0	0	0
16	terima kasih				1			1	0.778151	0	0	0	0.778151	0	0
17	sejuk				1			1	0.778151	0	0	0	0.778151	0	0
18	hati				1			1	0.778151	0	0	0	0.778151	0	0
19	pas				1			1	0.778151	0	0	0	0.778151	0	0
20	bahagia				1			1	0.778151	0	0	0	0.778151	0	0
21	kumpul				1			1	0.778151	0	0	0	0.778151	0	0
22	maaf				1			1	0.778151	0	0	0	0.778151	0	0
23	mudik					1		1	0.778151	0	0	0	0	0.778151	0
24	armada					1		1	0.778151	0	0	0	0	0.778151	0
25	jadwal					2		1	0.778151	0	0	0	0	1.556303	0
26	batas					1		1	0.778151	0	0	0	0	0.778151	0
27	enak					1		1	0.778151	0	0	0	0	0.778151	0
28	timbang					1		1	0.778151	0	0	0	0	0.778151	0
29	bambang					1		1	0.778151	0	0	0	0	0.778151	0
30	alhamdulillah						1	1	0.778151	0	0	0	0	0	0.778151
31	mantap						1	1	0.778151	0	0	0	0	0	0.778151
32	maju						1	1	0.778151	0	0	0	0	0	0.778151
33	jaya						1	1	0.778151	0	0	0	0	0	0.778151
34	hutang						1	1	0.778151	0	0	0	0	0	0.778151

35	lunas						1	1	0.778151	0	0	0	0	0	0.778151
36	sehat						1	1	0.778151	0	0	0	0	0	0.778151
37	badan						1	1	0.778151	0	0	0	0	0	0.778151
38	usaha						1	1	0.778151	0	0	0	0	0	0.778151
39	milik						1	1	0.778151	0	0	0	0	0	0.778151

Berikut penjelasan dari proses pembobotan kata.

- a. Tf_{Di} = trem frekuensi (jumlah kata yang muncul pada data)
- b. Df = Document Frekuensi (jumlah data yang mengandung kata tersebut)
- c. N/df = jumlah data / df
- d. Idf = $\log(N/df)$
- e. $Tf-idf_i$ = $tf_i * idf_i$
- f. Bobot akhir = jumlah $tf-idf_{Di}$

4.1.6 Analisa Klasifikasi K-Nearest Neighbor

Setelah proses pembobotan kata sudah selesai, hasil data pembobotan kata akan di klasifikasikan menggunakan metode K-Nearest Neighbor. Pada proses klasifikasi ini membutuhkan data latih yang telah diketahui kelas nya sebagai panduan untuk menguji data baru. Tahapan dalam proses klasifikasi KNN sebagai berikut.

1. Menentukan nilai k

Nilai K pada proses Klasifikasi ini yaitu k=3.

2. Menghitung kuadrat jarak *euclidean* (query instance) masing-masing obyek terhadap data pada tabel pembobotan kata dengan menggunakan rumus (2.5).

- a. Berikut data uji yang akan di klasifikasikan dengan metode KNN, data berikut sudah melewati proses text pre-processing, feature selection, dan pembobotan kata. Untuk pembobotan kata data uji berpedoman pada dokumen frekuensi (df) data latih.

Tabel 4.10 Bobot Data Uji

No	Kata	Tf	Df	idf	Tf-Idf
1	garuda	1	2	-0.301029996	-0.301029996
2	indonesia	1	2	-0.301029996	-0.301029996
3	pilih	1	2	-0.301029996	-0.301029996
4	terbang	1	1	0	0
5	keluarga	1	2	-0.301029996	-0.301029996

b. Menghitung jarak *Euclidean*

Setelah mendapatkan bobot data latih dan data uji, selanjutnya ke tahapan algortima KNN dengan rumus sebagai berikut:

• **Data uji (garuda)**

$$\begin{aligned} D1 &= \sqrt{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0 - (-0.301029996))^2} \\ &= \sqrt{0.605519368 + 0.605519368 + 0.605519368 + 1.164632162 + 0.605519368 + 3.081047982} \\ D1 &= 2.462161296 \end{aligned}$$

$$\begin{aligned}
D2 &= \sqrt{\begin{aligned} &(0.77815125 - (-0.301029996))^2 + (0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \\ &+ (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 \\ &+ (0.77815125 - (-0.301029996))^2 + (1.556302501 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \\ &+ (0 - (-0.301029996))^2 \times 30 \end{aligned}} \\
&= \sqrt{1.164632162 + 1.575709062 + 1.164632162 + 1.164632162 + 1.164632162 + 0.605519368 + 1.164632162 \\
&\quad + 3.449684002 + 1.164632162 + 2.718571749} \\
&= 3.916283589
\end{aligned}$$

$$\begin{aligned}
D3 &= \sqrt{\begin{aligned} &(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \\ &+ (0 - (-0.301029996))^2 \times 36 \end{aligned}} \\
&= \sqrt{0.605519368 + 0.605519368 + 1.164632162 + 3.262286098} \\
&= 2.374438249
\end{aligned}$$

$$\begin{aligned}
D4 &= \sqrt{\begin{aligned} &(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 7 \\ &+ (0 - (-0.301029996))^2 \times 30 \end{aligned}} \\
&= \sqrt{0.605519368 + 0.605519368 + 8.152425133 + 2.718571749} \\
&= 3.475922269
\end{aligned}$$

$$\begin{aligned}
D5 &= \sqrt{\begin{aligned} &(0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 6 + (1.556302501 - (-0.301029996))^2 \\ &+ (0 - (-0.301029996))^2 \times 31 \end{aligned}} \\
&= \sqrt{1.575709062 + 6.987792971 + 3.449684002 + 2.809190807} \\
&= 3.849984005
\end{aligned}$$

$$\begin{aligned}
D6 &= \sqrt{(0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 10 + (0 - (-0.301029996))^2 \times 28} \\
&= \sqrt{0.605519368 + 11.64632162 + 2.537333632} \\
&= 3.845669593
\end{aligned}$$

- **Data uji (indonesia)**

$$\begin{aligned}
 D1 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2}{34}} \\
 &= \sqrt{0.605519368 + 0.605519368 + 0.605519368 + 1.164632162 + 0.605519368 + 3.081047982} \\
 &= 2.462161296
 \end{aligned}$$

$$\begin{aligned}
 D2 &= \sqrt{\frac{(0.77815125 - (-0.301029996))^2 + (0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (1.556302501 - (-0.301029996))^2 + (0 - (-0.301029996))^2}{30}} \\
 &= \sqrt{1.164632162 + 1.575709062 + 1.164632162 + 1.164632162 + 1.164632162 + 0.605519368 + 1.164632162 + 3.449684002 + 1.164632162 + 2.718571749} \\
 &= 3.916283589
 \end{aligned}$$

$$\begin{aligned}
 D3 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2}{36}} \\
 &= \sqrt{0.605519368 + 0.605519368 + 1.164632162 + 3.262286098} \\
 &= 2.374438249
 \end{aligned}$$

$$\begin{aligned}
 D4 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2}{30}} \\
 &= \sqrt{0.605519368 + 0.605519368 + 8.152425133 + 2.718571749} \\
 &= 3.475922269
 \end{aligned}$$

$$\begin{aligned}
 D5 &= \sqrt{\frac{(0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (1.556302501 - (-0.301029996))^2 + (0 - (-0.301029996))^2}{31}} \\
 &= \sqrt{1.575709062 + 6.987792971 + 3.449684002 + 2.809190807} \\
 &= 3.849984005
 \end{aligned}$$

$$\begin{aligned}
D6 &= \sqrt{(0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 10 + (0 - (-0.301029996))^2 \times 28} \\
&= \sqrt{0.605519368 + 11.64632162 + 2.537333632} \\
&= 3.845669593
\end{aligned}$$

• **Data uji (pilih)**

$$\begin{aligned}
D1 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2 \times 34}{3}} \\
&= \sqrt{0.605519368 + 0.605519368 + 0.605519368 + 1.164632162 + 0.605519368 + 3.081047982} \\
&= 2.462161296
\end{aligned}$$

$$\begin{aligned}
D2 &= \sqrt{\frac{(0.77815125 - (-0.301029996))^2 + (0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (1.556302501 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2 \times 30}{10}} \\
&= \sqrt{\frac{1.164632162 + 1.575709062 + 1.164632162 + 1.164632162 + 1.164632162 + 0.605519368 + 1.164632162 + 3.449684002 + 1.164632162 + 2.718571749}{10}} \\
&= 3.916283589
\end{aligned}$$

$$\begin{aligned}
D3 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2 \times 36}{4}} \\
&= \sqrt{0.605519368 + 0.605519368 + 1.164632162 + 3.262286098} \\
&= 2.374438249
\end{aligned}$$

$$\begin{aligned}
D4 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2 \times 30}{4}} \\
&= \sqrt{0.605519368 + 0.605519368 + 8.152425133 + 2.718571749} \\
&= 3.475922269
\end{aligned}$$

$$\begin{aligned}
D5 &= \sqrt{\frac{(0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 6 + (1.556302501 - (-0.301029996))^2}{+ (0 - (-0.301029996))^2 \times 31}} \\
&= \sqrt{1.575709062 + 6.987792971 + 3.449684002 + 2.809190807} \\
&= 3.849984005
\end{aligned}$$

$$\begin{aligned}
D6 &= \sqrt{(0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 10 + (0 - (-0.301029996))^2 \times 28} \\
&= \sqrt{0.605519368 + 11.64632162 + 2.537333632} \\
&= 3.845669593
\end{aligned}$$

- **Data uji (terbang)**

$$\begin{aligned}
D1 &= \sqrt{\frac{(0.477121255 - (0))^2 + (0.477121255 - (0))^2 + (0.477121255 - (-0))^2}{(0.77815125 - (0))^2 + (0.477121255 - (-0))^2}} \\
&= \sqrt{0.227644692 + 0.227644692 + 0.227644692 + 0.605519368 + 0.227644692} \\
&= 1.231299369
\end{aligned}$$

$$\begin{aligned}
D2 &= \sqrt{\frac{(0.77815125 - (0))^2 + (0.954242509 - (0))^2 + (0.77815125 - (0))^2}{+ (0.77815125 - (0))^2 + (0.77815125 - (0))^2 + (0.477121255 - (0))^2}{+ (0.77815125 - (0))^2 + (1.556302501 - (0))^2 + (0.77815125 - (0))^2}} \\
&= \sqrt{\frac{0.605519368 + 0.910578767 + 0.605519368 + 0.605519368 + 0.605519368}{0.227644692 + 0.605519368 + 2.422077474 + 0.605519368}} \\
&= 2.68205465
\end{aligned}$$

$$\begin{aligned}
D3 &= \sqrt{(0.477121255 - (0))^2 + (0.477121255 - (0))^2 + (0.77815125 - (0))^2} \\
&= \sqrt{0.227644692 + 0.227644692 + 0.605519368} \\
&= 1.029955704
\end{aligned}$$

$$\begin{aligned}
D4 &= \sqrt{(0.477121255 - (0))^2 + (0.477121255 - (0))^2 + (0.77815125 - (0))^2 \times 7} \\
&= \sqrt{0.227644692 + 0.227644692 + 4.238635579} \\
&= 2.166546783
\end{aligned}$$

$$\begin{aligned}
D5 &= \sqrt{(0.954242509 - (0))^2 + (0.77815125 - (0))^2 \times 6 + (1.556302501 - (0))^2} \\
&= \sqrt{0.910578767 + 3.633116211 + 2.422077474} \\
&= 2.639274986
\end{aligned}$$

$$\begin{aligned}
D6 &= \sqrt{(0.477121255 - (0))^2 + (0.77815125 - (0))^2 \times 10} \\
&= \sqrt{0.227644692 + 6.055193685} \\
&= 2.506559071
\end{aligned}$$

- **Data Uji (keluarga)**

$$\begin{aligned}
D1 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0 - (-0.301029996))^2 \times 34}{6.05519368 + 0.605519368 + 0.605519368 + 1.164632162 + 0.605519368 + 3.081047982}} \\
&= 2.462161296
\end{aligned}$$

$$\begin{aligned}
D2 &= \sqrt{\frac{(0.77815125 - (-0.301029996))^2 + (0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (1.556302501 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2 \times 30}{1.164632162 + 1.575709062 + 1.164632162 + 1.164632162 + 1.164632162 + 0.605519368 + 1.164632162 + 3.449684002 + 1.164632162 + 2.718571749}} \\
&= 3.916283589
\end{aligned}$$

$$\begin{aligned}
D3 &= \sqrt{\frac{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 + (0 - (-0.301029996))^2 \times 36}{0.605519368 + 0.605519368 + 1.164632162 + 3.262286098}} \\
&= 2.374438249
\end{aligned}$$

$$\begin{aligned}
D4 &= \sqrt{(0.477121255 - (-0.301029996))^2 + (0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 7} \\
&\quad + (0 - (-0.301029996))^2 \times 30 \\
&= \sqrt{0.605519368 + 0.605519368 + 8.152425133 + 2.718571749} \\
&= 3.475922269
\end{aligned}$$

$$\begin{aligned}
D5 &= \sqrt{(0.954242509 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 6 + (1.556302501 - (-0.301029996))^2} \\
&\quad + (0 - (-0.301029996))^2 \times 31 \\
&= \sqrt{1.575709062 + 6.987792971 + 3.449684002 + 2.809190807} \\
&= 3.849984005
\end{aligned}$$

$$\begin{aligned}
D6 &= \sqrt{(0.477121255 - (-0.301029996))^2 + (0.77815125 - (-0.301029996))^2 \times 10 + (0 - (-0.301029996))^2 \times 28} \\
&= \sqrt{0.605519368 + 11.64632162 + 2.537333632} \\
&= 3.845669593
\end{aligned}$$

c. Menjumlahkan nilai euclidean

Setelah mendapatkan nilai *euclidean* dari masing-masing kelas, maka langkah selanjutnya adalah menjumlahkan nilai euclidean dari masing-masing kelas.

$$\begin{aligned}
D1 &= 2.462161296 + 2.462161296 + 2.462161296 + 1.231299369 + 2.462161296 \\
&= 11.079944553
\end{aligned}$$

$$\begin{aligned}
D2 &= 3.916283589 + 3.916283589 + 3.916283589 + 2.68205465 + 3.916283589 \\
&= 18.347189006
\end{aligned}$$

$$\begin{aligned}
D3 &= 2.374438249 + 2.374438249 + 2.374438249 + 1.029955704 + 2.374438249 \\
&= 10.5277087
\end{aligned}$$

$$\begin{aligned}
D4 &= 3.475922269 + 3.475922269 + 3.475922269 + 2.166546783 + 3.475922269 \\
&= 16.070235859
\end{aligned}$$

$$\begin{aligned}
D5 &= 3.849984005 + 3.849984005 + 3.849984005 + 2.639274986 + 3.849984005 \\
&= 18.039211006
\end{aligned}$$

$$\begin{aligned}
D6 &= 3.845669593 + 3.845669593 + 3.845669593 + 2.506559071 + 3.845669593 \\
&= 17.88923744
\end{aligned}$$

d. Mengurutkan Berdasarkan Nilai Euclidean

Kelas	<i>Term</i>					<i>Euclidean</i>	Urutkan Jarak	Apakah termasuk K=3
	garuda	indonesia	pilih	terbang	keluarga			
D1 (P)	2.462161296	2.462161296	2.462161296	1.231299369	2.462161296	11.079944553	2	YA K < 3
D2 (N)	3.916283589	3.916283589	3.916283589	2.68205465	3.916283589	18.347189006	6	TIDAK K > 3
D3 (P)	2.374438249	2.374438249	2.374438249	1.029955704	2.374438249	10.5277087	1	YA K < 3
D4 (N)	3.475922269	3.475922269	3.475922269	2.166546783	3.475922269	16.070235859	3	YA K = 3
D5 (N)	3.849984005	3.849984005	3.849984005	2.639274986	3.849984005	18.039211006	5	TIDAK K > 3
D6 (P)	3.845669593	3.845669593	3.845669593	2.506559071	3.845669593	17.88923744	4	TIDAK K > 3

*P = positif, N = negatif

e. Tentukan Kategori Tetangga Terdekat

Kelas	<i>Term</i>					<i>Euclidean</i>	Urutkan Jarak	Apakah termasuk K=3	Kategori YA untuk KNN
	garuda	indonesia	pilih	terbang	keluarga				
D1 (P)	2.462161296	2.462161296	2.462161296	1.231299369	2.462161296	11.079944553	2	YA	P
D2 (N)	3.916283589	3.916283589	3.916283589	2.68205465	3.916283589	18.347189006	6	TIDAK	
D3 (P)	2.374438249	2.374438249	2.374438249	1.029955704	2.374438249	10.5277087	1	YA	P
D4 (N)	3.475922269	3.475922269	3.475922269	2.166546783	3.475922269	16.070235859	3	YA	N
D5 (N)	3.849984005	3.849984005	3.849984005	2.639274986	3.849984005	18.039211006	5	TIDAK	
D6 (P)	3.845669593	3.845669593	3.845669593	2.506559071	3.845669593	17.88923744	4	TIDAK	

Hasil pengujian diatas termasuk kedalam kategori **sentimen Positif** dengan menggunakan K = 3.

Daftar Pustaka

- Aggrawal, N. (2018). Detection of Offensive Tweets: A Comparative Study, *I*(1), 75–89. Retrieved from <https://scholar.google.co.in/citations?user=tuoeHAAAAAJ&hl=en>
- Arsyam, I. (2018). Hasil pencarian untuk harga tiket lion air - Tribun Timur. Retrieved January 23, 2019, from <http://makassar.tribunnews.com/search?q=harga+tiket+lion+air&cx=partner-pub-7486139053367666%3A4965051114&cof=FORID%3A10&ie=UTF-8&siteurl=www.tribunnews.com>
- Blachly, L. (2011). Boeing, Lion Air pioneer precision satellite navigation technology | Aircraft & Engines content from ATWOnline. Retrieved January 23, 2019, from <http://atwonline.com/aircraft-amp-engines/boeing-lion-air-pioneer-precision-satellite-navigation-technology>
- Buntoro, G. A. (2016). Analisis Sentimen Hatespeech pada Twitter dengan Metode Naive Bayes Classifier dan Support Vector Machine. *Jurnal Dinamika Informatika*, 5(2).
- Chandani, V., & Wahono, R. S. (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. *Journal of Intelligent Systems*, 1(1), 56–60.
- Claudy, Y. I., Perdana, R. S., & Fauzi, M. A. (2018). Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritme K-Nearest Neighbor (KNN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*, 2548, 964X.
- Darujati, C., & Gumelar, A. B. (2012). Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia. *Jurnal Bandung Text Mining*, 16(1), 1–5.
- Erwanti, M. O. (2018). Penumpang JT 556 Turun karena AC-Lampu Mati, Ini Penjelasan Lion Air. Retrieved January 23, 2019, from https://news.detik.com/berita/d-4304182/penumpang-jt-556-turun-karena-ac-lampu-mati-ini-penjelasan-lion-air?_ga=2.164831935.1843896049.1548251126-218164827.1546276137
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
- Herdi, annisa ulkhairah. (2018). *Klasifikasi Tweet Penghinaan Agama Islam Pada*

Media Sosial Twitter Menggunakan Metode Support Vector Machine. Universitas islam negeri sultan syarif kasim riau.

- Indriati, I., & Ridok, A. (2016). Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwkn). *Journal of Environmental Engineering and Sustainable Technology*, 3(1), 23–32.
- Kementrian Komunikasi dan Informatika Republik Indonesia. (2013). Pengguna Internet di Indonesia 63 Juta Orang.
- Kharisman, O. (2017). Analisis Sentimen Pada Review Konsumen Maskapai Penerbangan Menggunakan Kombinasi Lexicon Berbasis Sentiwordnet Dan Supervised Model.
- Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) Dan K-Nearest Neighbor (K-NN). In *Seminar Nasional Teknologi Informasi dan Komunikasi*.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.
- M.Thampi, S., Trajkovic, L., Mitra, S., P.Nagabhushan, Mukhopadhyay, J., M.Corchado, J., ... Mishra, D. (2018). *Intelligent Systems, Technologies and Applications: Proceedings of ISTA 2018*. singapore. Retrieved from <https://link.springer.com/book/10.1007%2F978-981-13-6095-4>
- Mahendra, K., & Kerta, I. P. (2008). Penggunaan Algoritma Semut Dan Confix Stripping Stemmer Untuk Klasifikasi Dokumen Berita Berbahasa Indonesia. *Institut Teknologi Sepuluh Nopember*.
- Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*, 2548, 964X.
- Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 194–201). Springer.
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*, 2548,

964X.

- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135.
- Prasetyo, B., & Susana. (2012). Pengembangan Aplikasi Cerdas Berbasis Intelegensia Buatan Untuk Perbandingan Karya Ilmiah Hasil Penelitian Dalam Upaya Mencegah Plagiasi, 157–172.
- Prasetyo, E. (2012). Data Mining Konsep dan aplikasi menggunakan matlab. *Yogyakarta: Andi*.
- Putri, R. E., Suparti, & Rahmawati, R. (2014). Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012. *Jurnal Gaussian*, 3, 831–838. <https://doi.org/10.1016/j.jlumin.2012.12.015>
- Rosdiansyah, D. (2014). Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor dan Pendekatan Lexicon. Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523.
- Shende, S. B., & Deshpande, L. (2017). A computational framework for detecting offensive language with support vector machine in social communities. In *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1–4). <https://doi.org/10.1109/ICCCNT.2017.8204020>
- Simanjuntak, R. A. (2018). Analisis Sentimen pada Layanan Gojek Indonesia Menggunakan Multinomial Naive Bayes.
- Tahitoe, A. D., & Purwitasari, D. (2010). Implementasi Modifikasi Enhanced Confix Stripping Stemmer Untuk Bahasa Indonesia dengan Metode Corpus Based Stemming. *Institut Teknologi Sepuluh November (ITS)*.
- Utami, L. A. (2017). Analisis Sentimen Opini Publik Berita Kebakaran Hutan melalui Komparasi Algoritma Support Vector Machine dan K-nearest Neighbor Berbasis Particle Swarm Optimization. *Jurnal Pilar Nusa Mandiri*, 13(1), 103–112.
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes. *EVOLUSI-Jurnal Sains Dan Manajemen AMIK BSI Purwokerto*, 4(1).

- Windasari, I. P., Uzzi, F. N., & Satoto, K. I. (2018). Sentiment analysis on Twitter posts: An analysis of positive or negative opinion on GoJek. *Proceedings - 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2017*, 2018-Janua, 266–269. <https://doi.org/10.1109/ICITACEE.2017.8257715>
- Wisnu, D., & Hetami, A. (2015). Perancangan Information Retrieval (IR) Untuk Pencarian Ide Pokok Teks Artikel Berbahasa Inggris Dengan Pembobotan Vector Space Model. *Jurnal Ilmiah Teknologi Dan Informasi ASIA*, 9(1), 53–59.