

IMPLEMENTASI ALGORITMA *K – MEANS* UNTUK *CLUSTERING* SENTIMEN PADA OPINI KUALITAS PELAYANAN JASA PENERBANGAN

Syarifah Iin Safitri¹, Cucu Suhery², Syamsul Bahri³

^{1,2,3}Jurusan Rekayasa Sistem Komputer, Fakultas MIPA Universitas Tanjungpura

Jalan Prof. Dr. H. Hadari Nawawi Pontianak

Telp./Fax : (0561) 577963

e-mail: ¹syarifahiinsafitri@student.untan.ac.id.com, ²csuhery@siskom.untan.ac.id,

³syamsul.bahri@siskom.untan.ac.id.

ABSTRAK

Skytrax adalah perusahaan yang melakukan riset mengenai peringkat dan kualitas maskapai penerbangan. Skytrax memiliki kolom komentar tempat para pengguna maskapai penerbangan memberikan *review* mengenai kualitas layanan maskapai penerbangan yang ada di seluruh dunia. *Review* dari pengguna maskapai tersebut dapat membantu calon penumpang maskapai untuk mengetahui kualitas dari layanan maskapai. Membaca *review* tersebut secara keseluruhan tentu memerlukan waktu yang lama. Hal yang dapat dilakukan adalah mengumpulkan semua *review* tersebut dan mengolahnya menjadi sesuatu yang dapat menjawab keingintahuan para pengguna maskapai penerbangan mengenai kualitas pelayanan jasa maskapai penerbangan yang akan dipilih. Penelitian ini bertujuan untuk membangun sebuah sistem yang digunakan untuk mengelompokkan opini positif dan opini negatif yang terdapat pada *website* Skytrax menggunakan algoritma *K-Means*. Data yang digunakan merupakan data opini maskapai penerbangan Garuda Indonesia, Air Asia dan Lion Air yang terdapat pada *website* Skytrax dengan jumlah keseluruhan 1060 data opini dari tahun 2015 sampai dengan 2019. Keluaran yang dihasilkan berupa data opini yang telah dikelompokkan menjadi kelompok negatif dan kelompok positif. Perhitungan akurasi didapat dengan membandingkan nilai rating opini pada *website* Skytrax dengan hasil *clustering* sentimen pada algoritma *K-Means*. Persentase keberhasilan sistem *clustering* sentimen maskapai Lion Air 60,5%, Air Asia 52,8%, dan Garuda Indonesia 71,8%.

Kata Kunci: Skytrax, *Review*, *K-Means*, *Clustering*

1. PENDAHULUAN

Pesawat terbang merupakan salah satu sarana transportasi yang banyak digunakan oleh masyarakat Indonesia. Tingginya minat masyarakat dalam menggunakan pesawat terbang dapat dilihat pada data statistik jumlah penumpang pesawat terbang di Indonesia yang terus meningkat [1].

Penyedia jasa penerbangan mempunyai 2 kategori layanan yaitu *full service* dan *low cost carrier* (LCC). *Full service* merupakan penerbangan yang mengutamakan pelayanan penuh kepada penumpang baik dari segi kenyamanan keamanan,

pelayanan konsumsi, kelebihan bagasi, dan pelayanan lainnya yang tidak terdapat pada maskapai penerbangan *low cost carrier* (LCC). *Full service* memiliki banyak *additional service* yang menjadi nilai tambah dari *main service* yang ditawarkan. LCC adalah konsep penerbangan dengan biaya murah dan pelayanan yang terbatas.

Kualitas layanan maskapai merupakan suatu kemampuan untuk memenuhi kebutuhan internal dan eksternal pengguna maskapai secara konsisten sesuai prosedur. Dalam hal ini penyedia jasa penerbangan dituntut untuk berusaha mengerti apa yang diinginkan pelanggan, sehingga pelanggan mendapatkan kepu-

asan. Kepuasan pelanggan merupakan perasaan senang atau kecewa seseorang setelah membandingkan antara kinerja produk atau realitas yang dirasakan dengan yang diharapkan [2].

Untuk mengetahui kepuasan pelanggan terhadap kualitas layanan jasa penerbangan dari segi keselamatan, keamanan, maupun kenyamanan dapat diketahui dari opini atau *review* penumpang yang telah menggunakan jasa penerbangan. Saat ini pengguna jasa penerbangan yang menulis opini dan pengalaman secara *online* terus meningkat, opini tersebut banyak dituangkan pada forum-forum yang terdapat di internet, salah satu diantaranya adalah sebuah situs *website* yang bernama Skytrax.

Skytrax adalah perusahaan yang melakukan riset mengenai peringkat dan kualitas maskapai penerbangan. Skytrax memiliki kolom komentar tempat dimana para pengguna maskapai penerbangan bisa memberikan ulasan atau *review* mengenai kualitas layanan dari maskapai penerbangan yang ada di seluruh dunia. Membaca *review* tersebut secara keseluruhan tentu memerlukan waktu yang lama, namun jika hanya sedikit *review* yang dibaca tentu evaluasi yang dihasilkan tidak maksimal. Hal yang dapat dilakukan adalah usaha untuk mengumpulkan semua *review* tersebut dan mengolahnya menjadi sesuatu yang dapat menjawab keingintahuan para pengguna maskapai penerbangan mengenai kualitas maskapai penerbangan yang akan dipilih.

K-Means merupakan salah satu algoritma pengelompokan yang melakukan partisi set data kedalam sejumlah *cluster* yang sudah ditetapkan di awal dengan mengambil nilai minimal dari jarak titik data terhadap *cluster*. Proses perhitungan jarak dihitung menggunakan rumus *euclidian distance* [3].

Dalam penelitian yang berjudul “*Text Mining* untuk Analisis Sentimen *Review* Film Menggunakan Algoritma *K-Means*” di dapat hasil penelitian dengan dataset berupa 300 sentimen positif dan

300 sentimen negatif mempunyai akurasi 57,83%, 700 sentimen negatif dan 700 sentimen positif mendapatkan akurasi sebesar 56,71% dan 1000 sentimen positif serta 1000 sentimen negatif memiliki akurasi sebesar 50,40% data diambil dari cornell.edu [4].

Dalam penelitian yang berjudul “*Analisis Sentimen pada Opini Pengguna Maskapai Penerbangan Menggunakan Hybrid Cuckoo Search*” pada penelitian ini metode *Hybrid Cuckoo Search* digunakan untuk melakukan analisis sentimen negatif yang didapat oleh maskapai yang mengindikasikan buruknya kualitas layanan maskapai penerbangan. Dari hasil uji coba diperoleh nilai rata-rata akurasi, *precision*, dan *recall* dari 7 maskapai dengan 1.000 iterasi masing-masing sebesar 69,24%; 70,88%; dan 77,57% [3].

Penelitian yang berjudul “*Text Mining* dengan *K-Means Clustering* pada Tema LGBT dalam Arsip *Tweet* Masyarakat Kota Bandung” penelitian ini menggunakan 5 *Cluster* untuk mengelompokkan komentar-komentar di *twitter* yang berhubungan dengan LGBT di kota Bandung. Dari lima *cluster* yang dibentuk pada proses *K-Means* diperoleh bahwa kecenderungan cuitan pengguna *twitter* kota Bandung terkait LGBT secara umum masih berhubungan dengan *perspektif* religi yang ditandai dengan kata agama yang sering muncul [5].

Berdasarkan penelitian-penelitian tersebut, maka dilakukan penelitian tentang “*Implementasi Algoritma K-Means Untuk Clustering Sentimen pada Opini Kualitas Pelayanan Jasa Penerbangan Berbasis Web*”. Hasil penelitian ini dapat digunakan untuk memudahkan pengguna jasa maskapai penerbangan dalam memilih maskapai yang memiliki kualitas pelayanan terbaik yang mampu memberikan kenyamanan dari tempat keberangkatan sampai tiba ke tempat tujuan.

2. LANDASAN TEORI

2.1 Text Mining

Text mining adalah proses menggali informasi dimana *user* berinteraksi dengan sekumpulan dokumen dengan menggunakan *tools* analisis. *Text mining* berusaha untuk mendapatkan informasi yang berguna dari sumber data berupa dokumen atau teks yang biasanya menggunakan *unstructured data*, atau minimal *semi-structured data* [6].

2.2 Analisis Sentimen

Analisis sentimen merupakan bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang-orang terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa dan topik tertentu [7].

Analisis sentimen dapat dibedakan berdasarkan levelnya, beberapa level yang sering digunakan dalam penelitian analisis sentimen adalah analisis sentimen pada level dokumen dan analisis sentimen pada level kalimat.

2.2.1 Analisis Sentimen Level Dokumen

Analisis sentimen level dokumen menganggap seluruh isi dokumen sebagai sebuah sentimen positif atau sentimen negatif. Tingkat analisis ini mengasumsikan bahwa setiap dokumen mengungkapkan pendapat pada satu entitas dan tidak berlaku untuk dokumen yang mengevaluasi atau membandingkan beberapa entitas.

2.2.2 Analisis Sentimen Level Kalimat

Analisis sentimen pada level kalimat memiliki fokus utama untuk menentukan sentimen pada setiap kalimat, apakah pada setiap kalimat tersebut mengekspresikan opini positif atau negatif, atau netral.

2.3 Clustering

Clustering adalah proses pengelompokan satu set objek data menjadi beberapa kelompok atau *cluster* sehingga objek dalam suatu *cluster* memiliki kesamaan yang tinggi, namun sangat berbeda dengan objek

dalam *cluster* lain. Kesamaan dan ketidaksamaan dinilai berdasarkan nilai atribut yang menggambarkan objek dan sering melibatkan pengukuran jarak atau *measures distance*. Pembelajaran ini termasuk dalam *unsupervised learning* yang tidak memiliki data latih, sehingga data yang ada di kelompokan menjadi 2 bagian atau tiga bagian dan seterusnya [3].

K-Means merupakan salah satu algoritma pengelompokan yang melakukan partisi set data ke dalam sejumlah *k cluster* yang sudah ditetapkan di awal dengan mengambil nilai minimal dari jarak titik data terhadap *k cluster* [3]. Langkah-langkah untuk melakukan pengelompokan menggunakan algoritma *K-Means* adalah sebagai berikut:

1. Menentukan *k* sebagai jumlah *cluster* yang ingin dibentuk.
2. Membangkitkan nilai acak untuk pusat *cluster* awal (*centroid*) sebanyak *k*.
3. Menghitung jarak setiap data terhadap masing-masing *centroid* menggunakan rumus jarak *euclidian distance* hingga ditemukan jarak yang terdekat dari setiap data dengan *centroid*.

Berikut rumus *euclidian distance* :

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Keterangan:

x_i = Data kriteria ke-*i*, $i = 1, 2, \dots, n$

μ_j = Data kriteria *centroid* pada *cluster* ke-*j*, $j = 1, 2, 3, \dots, m$

d = Jarak terdekat antara data kriteria

4. Kelompokan data berdasarkan kedekatan dengan *centroid*. Data yang mempunyai nilai terdekat dengan *centroid* akan menjadi *cluster* yang diikuti.
5. Memperbaharui nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* dengan menggunakan rumus :

$$c_j = \frac{1}{N_k} \sum_{j=1}^{N_k} x_j \quad (2)$$

Keterangan:

c_j = Centroid Baru

N_k = Jumlah keseluruhan data yang tergabung dalam *cluster*

j = Data kriteria *centroid* pada *cluster* ke- j

6. Melakukan perulangan dari langkah 3 hingga langkah 5 hingga *cluster* yang diikuti dan *centroid* tidak ada yang berubah, antara *centroid* terakhir dan *centroid* sebelum terakhir. Iterasi terakhir akan digunakan sebagai parameter untuk menentukan *cluster* data.

2.4 Kepuasan Pelanggan

Kepuasan pelanggan adalah tingkat perasaan seseorang setelah membandingkan kinerja yang dirasakan dengan harapan. Apabila persepsi terhadap kinerja tidak dapat memenuhi harapan pelanggan maka akan terjadi ketidakpuasan demikian juga sebaliknya [2]. Indikator kepuasan pelanggan ada-lah sebagai berikut:

1. Rasa Senang
2. Kepuasan terhadap pelayanan: menunjukkan sejauh mana konsumen jasa merasa puas dengan cara dan sikap penyedia jasa selama masa pelayanan.
3. Kepuasan terhadap sistem: Kepuasan terhadap sistem menunjukkan bahwa seberapa besar konsumen merasa puas dengan kecepatan dan ketepatan dalam pelayanan yang dijalankan.

2.5 Preprocessing

Preprocessing adalah tahap awal dalam *text mining*, tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi *knowledge discovery* atau penggalian informasi pada *text mining* [3]. *Preprocessing* yang dilakukan pada penelitian ini adalah *case folding*, *data cleaning* dan *stopword removal*.

1. *Case Folding* adalah proses mengubah semua kata yang ditulis dengan huruf kapital menjadi huruf kecil.

2. *Stopword Removal* adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Misalnya "as", "a", "is", "the", dan lain sebagainya. Sebelum proses *stopword removal* dilakukan, harus dibuat daftar *stoplist*, dalam penelitian ini daftar *stoplist* diambil dari *stopword dictionary* [9].
3. *Data Cleaning* adalah fase dimana akan menghilangkan elemen *noise* yang tidak diinginkan dari data yang akan di proses, diantaranya sebagai berikut:

- a. Menghapus semua simbol seperti at (@), *hashtag*, tanda kurung (), garis miring maju (/), garis miring terbalik (\), tanda hubung (-) dan lain sebagainya [3].
- b. Hilangkan semua URL melalui pen-cocokan ekspresi reguler. Ekspresi reguler adalah pola teks yang mendefinisikan pola pencarian untuk teks.
- c. Menghapus *double space* dan di ganti dengan *single space*.
- d. Menghapus huruf berulang. Ganti urutan huruf berulang (tiga atau lebih) dalam satu kata.
- e. Menghapus kata yang tidak dimulai dengan alfabet.
- f. Ganti semua bentuk kata pendek atau singkatan ke dalam bentuk kata penuh.

2.6 Ekstraksi Fitur

Ekstraksi Fitur adalah salah satu teknik pengurangan dimensionalitas. Ekstraksi Fitur mengekstrak fitur asli menjadi fitur baru yang ditetapkan melalui beberapa pemetaan fungsional yang menyimpan sebanyak mungkin informasi dalam data. Menurut Agarwal [9] dalam bukunya yang berjudul *Prominent Feature Extraction for Sentiment Analysis*, menjelaskan bahwa ada empat jenis fitur dasar, diantaranya *Unigrams*, *Bigrams*, *Bi-Tagged* dan

Dependency Features. Pada penelitian ini fitur dasar yang digunakan adalah *Bi-Tagged*.

Fitur *Bi-Tagged* secara selektif diekstraksi menggunakan pola *Part Of Speech* (POS). *Bi-Tagged* banyak mengandung kata sifat dan kata keterangan karena kata sifat dan kata keterangan dianggap lebih mengekspresikan suatu sentimen. Informasi berbasis POS digunakan untuk mengekstrak fitur yang kaya sentimen, seperti yang sudah dijelaskan dalam buku yang berjudul *Prominent Feature Extraction for Sentiment Analysis*, bahwa kata sifat dan kata keterangan bersifat subjektif [3]. Dalam penelitian ini fitur *Bi-Tagged* yang diekstrak adalah sebagai berikut

1. Total Kata: Untuk menghitung jumlah total kata yang terdapat pada setiap kalimat yang akan diproses.
2. Seruan Positif, Dihitung dengan membandingkan setiap kata dengan Positive Exclamation Dictionary.
3. Seruan negatif, Dihitung dengan membandingkan setiap kata pada opini dengan Negative Exclamation Dictionary [10].
4. Penyangkalan, Umumnya digunakan untuk ekspresi penolakan. Dihitung dengan membandingkan setiap kata pada opini dengan kata penyangkalan pada kamus [11].
5. Kata Positif, Dihitung dengan membandingkan setiap kata pada opini dengan kata pada kamus Positive Words [12].
6. Kata Negatif, Dihitung dengan membandingkan setiap kata pada opini dengan kata pada kamus Negatif Words [13].
7. Kata Intens, Dihitung dengan membandingkan setiap kata opini dengan kata pada kamus Intense Word Dictionary [11].

2.7 Skytrax

Skytrax adalah perusahaan konsultan Britania Raya yang melakukan riset mengenai maskapai penerbangan. Per-

sahaan ini melakukan survei untuk menentukan maskapai, bandar udara, hiburan dalam pesawat, staf, dan elemen perjalanan udara terbaik lainnya. Selain survei, Skytrax juga memiliki forum maskapai penerbangan yang merupakan sarana bagi penumpang pesawat untuk memberikan ulasan yang dapat dilihat oleh calon penumpang lain, Skytrax juga dikenal sebagai perusahaan yang memberikan penghargaan kepada maskapai penerbangan dunia dan bandara dunia setiap tahun.

2.8 Aplikasi Berbasis Web

Aplikasi web adalah suatu jenis aplikasi komputer yang dapat diakses dengan web browser, selama pemakai dapat mengakses web server. Web server adalah server yang melayani permintaan aplikasi web. Aplikasi yang dibuat dalam penelitian ini merupakan aplikasi berbasis website [15]. Dalam penelitian ini, untuk mendukung dalam pembuatan aplikasi maka menggunakan beberapa perangkat lunak pendukung terutama dalam hal pembuatan database dan program. Perangkat lunak yang digunakan antara lain:

2.8.1 Hypertext Preprocessor

PHP (*Hypertext Preprocessor*) adalah bahasa *server-side scripting* yang menyatu dengan HTML untuk membuat halaman website. PHP mampu berjalan di Windows dan beberapa versi linux. Hampir seluruh aplikasi berbasis website dapat dibuat dengan PHP. Kelebihan-kelebihan dari PHP diantaranya adalah PHP secara mendasar dapat mengerjakan beberapa tugas, seperti mendapatkan data dari form, menghasilkan isi halaman website yang dinamik dan menerima cookies.

2.8.2 Hypertext Markup Language

HTML (*Hypertext Markup Language*) merupakan salah satu format yang digunakan dalam pembuatan dokumen dan aplikasi yang berjalan di halaman website. HTML merupakan protokol yang digunakan untuk mentransfer data antara web server ke

web browser. Dikatakan *markup language* karena HTML berfungsi untuk memformat *file* dokumen sehingga bisa ditampilkan pada web browser dengan bantuan kode-kode yang sudah ditentukan dengan menambahkan elemen atau disebut *tag* [14].

2.9 Confusion Matrix

Confusion Matrix dapat digunakan untuk menganalisis hasil dari proses *clustering*. Matriks yang ditunjukkan pada Tabel 1 berisi tentang nilai aktual dan nilai prediksi hasil analisis sentimen yang didapat dari proses *Clustering*.

Tabel 1 *Cofusion Matrix*

Nilai Aktual	Nilai Prediksi	
	Negatif	Positif
Negatif	TN	FP
Positif	FN	TP

Dimana :

1. *True Positive* (TP) adalah jumlah prediksi yang tepat bersifat positif.
2. *True Negative* (TN) adalah jumlah prediksi yang tepat bersifat negatif.
3. *False Positive* (FP) adalah jumlah prediksi yang salah bersifat positif.
4. *False Negative* (FN) adalah jumlah prediksi yang salah bersifat negatif.

Sehingga Akurasi, *Specifity* dan *Precision* dapat dicari dengan menggunakan rumus sebagai berikut:

- a. Akurasi (AC) adalah proporsi jumlah prediksi dataset yang benar.

$$AC = \frac{TP + TN}{TP + TN + FN + FP} \times 100\% \quad (3)$$

- b. *Specifity* (S) adalah Porposi prediksi jumlah opini negatif yang benar.

$$S = \frac{TN}{FN + TN} \times 100\% \quad (4)$$

- c. *Precision* (P) adalah proporsi prediksi jumlah opini positif yang benar.

dihitung dengan menggunakan persamaan:

$$P = \frac{TP}{FP + TP} \times 100\% \quad (5)$$

3. METODE PENELITIAN

3.1 Studi Literatur dan Pengumpulan Data

Studi literatur yang digunakan pada penelitian ini berupa jurnal ilmiah, penelitian sebelumnya, buku-buku, artikel, dan data-data yang dapat digunakan untuk mendukung penelitian tercapainya tujuan penelitian.

Data yang digunakan merupakan data yang berupa opini pengguna maskapai penerbangan yang diambil dari sebuah situs *website* yang bernama Skytrax. Data yang dikumpulkan berjumlah 1.060 data opini dari pengguna maskapai penerbangan dari tahun 2015 sampai dengan 2019.

3.2 Analisis Kebutuhan

Analisa kebutuhan dalam pengerjaan ini meliputi kebutuhan perangkat keras dan perangkat lunak.

3.2.1 Kebutuhan Perangkat Keras

Kebutuhan perangkat keras yang dibutuhkan dalam mengerjakan penelitian ini meliputi:

1. Prosesor: Intel(R) Core i3
2. RAM: 4 GB
3. Hardisk: 500 GB
4. Monitor: LCD Monitor

3.2.2 Kebutuhan Perangkat Lunak

Kebutuhan perangkat lunak yang dibutuhkan dalam mengerjakan penelitian ini meliputi:

1. Sublime Text Editor 3 digunakan sebagai tempat untuk pengetikan kode program.
2. Web Sever (XAMPP) versi 3.2.2 digunakan sebagai server yang berdiri sendiri (*localhost*) untuk membantu design *website* secara lokal yang

terdiri dari beberapa aplikasi pendukung.

3. MySql sebagai *database server* untuk penyimpanan data.
4. Google Chrome versi 65.0.3325.181 digunakan dalam melakukan proses design *website*.
5. Balsamiq Mock-Up digunakan untuk merancang tampilan sistem.

3.3 Perancangan Sistem, Implementasi dan pengujian

Perancangan sistem pada penelitian ini terdiri dari desain *flowchart* keseluruhan sistem, desain *flowchart* proses *data cleaning*, desain *flowchart* ekstraksi fitur, desain *flowchart* algoritma *k-means*, desain model sistem, rancangan basis data, desain antarmuka aplikasi dan perancangan pengujian.

Sistem yang telah dibuat pada proses implementasi diuji menggunakan pengujian *Black Box* untuk melihat sejauh mana sistem berfungsi sesuai dengan apa yang diharapkan. Sedangkan akurasi pada sistem dilakukan dengan cara membandingkan hasil dari pengelompokan menggunakan algoritma *K-Means* dengan rating pada tiap opini di *website* Skytrax.

4. IMPLEMENTASI, PENGUJIAN DAN PEMBAHASAN

4.1 Implementasi

4.1.1 Persiapan Data

Dalam penelitian ini proses pengelompokan data menggunakan metode *K-Means* memerlukan data berupa angka, jadi perlu ada proses perubahan data yang semula berbentuk kata ke dalam bentuk angka agar dapat diproses. Data yang digunakan berjumlah 1.060 data yang berisikan opini dari pengguna maskapai penerbangan dari tahun 2015 sampai dengan 2019.

4.1.2 Perhitungan Algoritma *K-Means*

Pada perhitungan ini dilakukan pengelompokan 5 data, nilai dari data tersebut dapat dilihat pada Tabel 2. Langkah-

langkah yang digunakan adalah sebagai berikut:

1. Tentukan jumlah *cluster*. Pada penelitian ini digunakan 2 *cluster*, yaitu *cluster* negatif dan *cluster* positif. Data *centroid* awal terdapat pada Tabel 3.
2. Inisialisasi. Dilakukan pemilihan data sebagai *centroid* awal secara acak. Dipilih data ke 1 dan data ke 2.
3. Hitung jarak setiap data ke *centroid*. Berikut ini contoh perhitungan jarak data numerik hasil proses ekstraksi fitur:
 - a. Jarak antara data kriteria pada opini ke 1 dengan *centroid* Negatif menggunakan Persamaan 1.

$$d(x_1, \mu_1) = \sqrt{(21 - 21)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 0$$

- b. Jarak data kriteria pada opini ke 1 dengan *centroid* Positif menggunakan Persamaan 1.

$$d(x_1, \mu_2) = \sqrt{(21 - 34)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 1)^2} = 13,03840$$

4. Kelompokkan setiap data yang diperoleh dari perhitungan menggunakan Persamaan 1 diatas berdasarkan kedekatan dengan *centroid*. *Centroid* terdekat menjadi *cluster* yang diikuti oleh data tersebut. Anggota *cluster* negatif dan positif berdasarkan jarak data dapat dilihat pada Tabel 4.
5. Memperbarui nilai *centroid*. *Centroid* baru digunakan untuk perhitungan pada iterasi berikutnya, yang diperoleh dari nilai rata-rata data kriteria yang telah melewati proses perhitungan jarak menggunakan Persamaan 2.1. Perhitungan rata-rata data kriteria didapat dengan menggunakan Persamaan 2.2. Hasil dari perhitungan terdapat pada Tabel 5.

Tabel 2 Contoh Data Opini Maskapai Penerbangan yang Telah Diubah dalam Bentuk Numerik

Opini ke-	Total Kata	Seruan Positif	Seruan Negatif	Penyangkalan	Kata Positif	Kata Negatif	Kata Intens
1	21	0	0	0	3	2	0
2	34	0	0	0	3	2	1
3	19	0	0	0	1	2	0
4	44	0	0	0	2	2	1
5	33	0	0	0	3	2	2

Tabel 3 Data *Centroid* Awal

Opini ke-	<i>Centroid</i>	Total Kata	Seruan Positif	Seruan Negatif	Penyangkalan	Kata Positif	Kata Negatif	Kata Intens
1	Negatif	21	0	0	0	3	2	0
2	Positif	34	0	0	0	3	2	1

Tabel 4 Hasil Perhitungan Jarak Antara Data ke *Centroid*

Opini ke-	Jarak <i>Centroid</i>		Terdekat	Cluster yang Diikuti
	Negatif	Positif		
1	0	13,03840	0	Negatif
2	13,03840	0	0	Positif
3	2,82843	15,16575	2,82843	Negatif
4	23,04344	10,04988	10,04988	Positif
5	12,16553	1,41421	1,41421	Positif

Tabel 5 *Centroid* Baru pada Iterasi Ke 2

<i>Centroid</i>	Total Kata	Seruan Positif	Seruan Negatif	Penyangkalan	Kata Positif	Kata Negatif	Kata Intens
Negatif	20	0	0	0	2	2	0
Positif	37	0	0	0	2,66667	2	1,33333

Contoh perhitungan rata-rata data kriteria pada iterasi ke-2 sebagai berikut:

- a. Nilai rata-rata data kriteria ke 1 pada *centroid* Negatif menggunakan Persamaan 2.

$$c_{11} = \frac{1}{2} \times 40 \rightarrow c_{11} = 20$$

- b. Nilai rata-rata data kriteria ke 2 pada *centroid* Negatif menggunakan Persamaan 2.

$$c_{12} = \frac{1}{2} \times 0 \rightarrow c_{12} = 0$$

- c. Nilai rata-rata data kriteria ke 1 pada *centroid* Positif menggunakan Persamaan 2.

$$c_{21} = \frac{1}{3} \times 111 \rightarrow c_{12} = 37$$

- d. Nilai rata-rata data kriteria ke 2 pada *centroid* Positif menggunakan Persamaan 2.

$$c_{22} = \frac{1}{3} \times 0 \rightarrow c_{22} = 0$$

Centroid baru ini digunakan untuk perhitungan jarak pada iterasi berikutnya. Setelah dilakukan perhitungan jarak kembali

dari data kriteria pada Tabel 2 dengan *centroid* baru pada Tabel 5 menggunakan Persamaan 2.1 diperoleh anggota *cluster* baru pada iterasi ke 2 yang terdapat pada Tabel 6.

6. Melakukan perulangan dari langkah 3 sampai dengan langkah 5 sehingga *centroid* dan anggota *cluster* yang diikuti tidak ada yang berubah.

Iterasi ketiga pada per-hitungan ini menghasilkan *centroid* dan *cluster* yang diikuti sama dengan iterasi ke dua. Hal ini menunjukan bahwa perhitungan algoritma *K-Means* yang digunakan untuk mengelompokan data opini pada Tabel 2 berhenti pada iterasi ke tiga. *Centroid* baru pada iterasi ketiga dapat dilihat pada Tabel 7 dan *cluster* yang diikuti dapat dilihat pada Tabel 8. Tabel 7 dan 8 merupakan hasil akhir perhitungan algoritma *K-Means*, sehingga dapat disimpulkan bahwa data opini maskapai Air Asia pada Tabel 2 memperoleh hasil *clustering* dengan data opini ke 1 dan ke 3 mengikuti *cluster* Negatif dan data opini ke 2, 4 dan 5 mengikuti *cluster* Positif.

4.2 Pengujian

Pengujian dilakukan dengan membandingkan hasil dari perhitungan yang dilakukan secara manual dan dengan menggunakan sistem. Pada tahap pengujian ini didapatkan hasil yang sama antara perhitungan secara manual dan perhitungan menggunakan sistem. Tabel *centroid* awal, *centroid* akhir, anggota *cluster* hasil perhitungan secara manual dan sistem dan Tabel Akurasi, *Specifity* serta *Precision* terdapat pada Tabel 9, Tabel 10, Tabel 11 dan Tabel 12.

4.3 Pembahasan

Opini yang digunakan merupakan opini yang diambil dari *website* Skytrax dengan nama maskapai Garuda Indonesia, Air Asia dan Lion Air. Opini tersebut merupakan opini dari tahun 2015 sampai dengan 2019 dengan jumlah keseluruhan opini sebanyak 1060 opini.

Sebelum melewati proses *clustering* sentimen, opini terlebih dahulu melalui proses *preprocessing*. Hasil dari proses *preprocessing* adalah data bersih yang tidak mengandung *noise* dan merupakan isi dari opini. Kemudian data yang telah melewati proses *preprocessing* diubah terlebih dahulu dalam bentuk numerik dengan proses ekstraksi fitur agar dapat dikelompokan menggunakan algoritma *K-Means*.

Pada proses Ekstraksi Fitur, data opini maskapai penerbangan yang berupa teks diubah kedalam bentuk numerik. Proses ekstraksi fitur ini dilakukan dengan cara membandingkan antara data opini dengan data dari kamus yang terdapat pada tabel *dictionary*. Seluruh kata yang telah diubah dalam bentuk numerik kemudian disimpan pada tabel ekstraksi fitur.

Pengujian yang telah dilakukan pada 6 proses algoritma *K-Means*, berdasarkan pengujian tersebut diperoleh *centroid* awal secara acak terdapat pada Tabel 9, *centroid* akhir yang merupakan *centroid* dengan nilai dan anggota *cluster* yang tidak berubah terdapat pada Tabel 10 dan data opini maskapai penerbangan yang telah terkelompokan berdasarkan jarak terdekat dengan *centroid* akhir terdapat pada Tabel 11 serta akurasi, *Specifity* dan *Precision* terdapat pada Tabel 12

Tabel 6 Anggota *Cluster* Baru Hasil Perhitungan Jarak Antara Data Kriteria ke *Centroid* Untuk Iterasi ke 2

Data ke-	Jarak Ke <i>Centroid</i>		Terdekat	<i>Cluster</i> Yang Diikuti
	Negatif	Positif		
1	2	257,88889	2	Negatif
2	198	9,22222	9,22222	Positif

Tabel 6 (Lanjutan)

Data ke-	Jarak Ke Centroid		Terdekat	Cluster Yang Diikuti
	Negatif	Positif		
3	2	328,55556	2	Negatif
4	577	49,55556	49,55556	Positif
5	174	16,55556	16,55556	Positif

Tabel 7 Tabel Centroid Baru Pada Iterasi ke 3

Centroid	Total Kata	Seruan Positif	Seruan Negatif	Penyangkalan	Kata Positif	Kata Negatif	Kata Intens
Negatif	20	0	0	0	2	2	0
Positif	37	0	0	0	2,66667	2	1,33333

Tabel 8 Hasil Perhitungan Jarak Antara Data ke Centroid Untuk Iterasi ke 3

Data ke-	Jarak Ke Centroid		Terdekat	Cluster Yang Diikuti
	Negatif	Positif		
1	2	257,88889	2	Negatif
2	198	9,22222	9,22222	Positif
3	2	328,55556	2	Negatif
4	577	49,55556	49,55556	Positif
5	174	16,55556	16,55556	Positif

Tabel 9 Data Centroid Awal

Nama Makapai	Total Kata	Seruan Positif	Seruan Negatif	Penyangkalan	Kata Positif	Kata Negatif	Kata Intens	Centroid
Air Asia	70	0	0	0	3	3	1	Negatif
	35	0	0	0	3	4	0	Positif
Garuda Indonesia	130	0	0	0	9	12	2	Negatif
	45	0	0	0	7	1	2	Positif
Lion Air	44	0	0	0	1	7	0	Negatif
	115	0	0	0	8	9	3	Positif

Tabel 10 Centroid Akhir Hasil Perhitungan Algoritma K-Means

Nama Maskapai	Total Kata	Seruan Positif	Seruan Negatif	Penyangkalan	Kata Positif	Kata Negatif	Intens	Centroid
Air Asia	103,0807	0,01613	0	0	5,59677	6,06452	1,27419	Negatif
	34,86744	0	0	0,00288	2,70317	2,29107	0,63401	Negatif
Garuda Indonesia	98,37931	0	0,00862	0,00862	10,31034	3,55172	2,12931	Negatif
	39,77404	0	0	0,0024	5,1875	1,36538	1,10096	Negatif
Lion Air	37,07843	0,0098	0	0,02941	2,19608	3,44118	0,71569	Negatif
	107,6471	0	0	0,05882	4,41176	8,94118	1,58824	Negatif

Tabel 11 Perbandingan Anggota *Cluster* Perhitungan Manual dan Sistem

No	Nama Maskapai	Manual		Sistem	
		Opini Positif	Opini Negatif	Opini Positif	Opini Negatif
1	Garuda Indonesia	116	416	116	416
2	Lion Air	17	102	17	102
3	Air Asia	62	347	62	347

Tabel 12 Hasil Akurasi, *Specifity* dan *Precision*

Nama Maskapai	Akurasi	<i>Specifity</i>	<i>Precision</i>
Lion Air	62,1%	69,6%	17,6%
Air Asia	52,8%	67,7%	50,1%
Garuda Indonesia	71,8%	17,2%	87,02%

Pada penelitian ini dilakukan 3 kali *clustering* dengan menggunakan dataset yang sama. Pada 3 kali *clustering*, didapatkan *centroid* awal dengan nilai yang berbeda pada masing-masing *clustering*. Perbedaan ini dikarenakan setiap dilakukan *clustering* data, diperoleh *centroid* awal secara acak, sehingga hasil *centroid* akhir, anggota *cluster* dan akurasi sistem mengikuti letak *centroid* awal tersebut.

5. KESIMPULAN

Berdasarkan pengujian yang telah dilakukan pada sistem, maka dapat diambil kesimpulan sebagai berikut:

1. Presentase keberhasilan Algoritma *K-Means* pada maskapai Garuda Indonesia sebesar 71,8%, Air Asia sebesar 52,8% dan Lion Air sebesar 62,1%.
2. Faktor yang dapat mempengaruhi akurasi pada algoritma *K-Means* yaitu penentuan pusat *cluster* atau inisialisasi *centroid* awal yang dilakukan secara acak.

6. SARAN

Terdapat beberapa saran pada penelitian ini:

1. Menambahkan metode untuk mengoptimalkan dalam pemilihan *centroid* awal yang baik berdasarkan kedekatan dengan anggota *cluster*.
2. Data yang digunakan merupakan data yang didapat secara *realtime* dari *website* Skytrax agar hasil dari *clus-*

tering merupakan data terbaru dari *situs* tersebut.

DAFTAR PUSTAKA

- [1] Direktorat Jenderal Perhubungan Udara, "Tiga Tahun Tren Positif Pencapaian Penerbangan Nasional," Desember 2019. [Online]. Available: <http://hubud.dephub.go.id/?id/news/detail/3283>.
- [2] K. d. Kotler, Marketing management, Upper Saddle River: Pearson Education, Inc, 2006.
- [3] N. F. d. Rozi, "Analisis Sentimen Pada Opini Pengguna Maskapai Penerbangan Menggunakan Hybrid Cuckoo Search," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, pp. 321-326, 2018.
- [4] S. Budi, "Text Mining Untuk Analisis Sentimen Review Film," *Techno.Com*, pp. 1-8, 2017.
- [5] E. Yulian, "Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung," *JURNAL MATEMATIKA "MANTIK"*, p. Vol. 04 No. 01, 2018.

- [6] K. & P. Han, Data Mining: Concept and Techniques, Third Edition, Waltham: Morgan Kaufmann, 2012.
- [7] S. & Feldman, Text Mining Handbook: Advance in Analyzing Unstructured Data, New York: Cambrige University Press, 2007.
- [8] R. NL, "Stopword Dictionary," 05 Maret 2014. [Online]. Available: <http://www.ranks.nl/stopwords>.
- [9] B. Agarwal, Prominet Feature Extraction for Sentiment Analysis, Second Edition, Switzerland: Springer International, 2016.
- [10] Vidarholen, "Exclamation Word Dictionary," 21 September 2013. [Online]. Available: <http://www.vidarholen.net/contents/interjections/>.
- [11] Psychpage, "Psychological Feelings," 07 November 2015. [Online]. Available: <http://www.psychpage.com/learning/library/assess/feelings.html>.
- [12] Jeffreybreen, 29 Juli 2015. [Online]. Available: <https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107/blob/master/data/opinion-lexicon-English/positive-words.txt>.
- [13] Jeffreybreen, "Negative Word Dictionary," 29 Juli 2015. [Online]. Available: <https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107/blob/master/data/opinion-lexicon-English/negative-words.txt>.
- [14] A. Kadir, Membuat Aplikasi Web dengan PHP + Database MySQL, Yogyakarta: Andi, 2009.
- [15] Sunarfrihantono, PHP dan MySQL untuk Web, Yogyakarta: Andi, 2002.
- [16] A. d. Martin, Quality Service : What every hospitality manager need to know, USA: Prentice Hall, 2001.
- [18] B. Liu, Sentiment Analysis and Opinion Mining, California, USA: Morgan & Claypool Publisher, 2012.
- [19] A. d. Bastian, "Penerapan Algoritma K-Means Clustering Analysis pada Penyakit Menular Manusia," *Jurnal Sistem Informasi*, 2018.
- [20] F. & Sanger, Text Mining Handbook: Advances in Analyzing Unstructured Data, New York: Cambridge University Press, 2007.
- [21] NetLingo, "Emoticon Dictionary," 12 Desember 2015. [Online]. Available: <http://www.netlingo.com/smileys.php>.
- [22] NetLingo, "Acronym Dictionary," 12 Desember 2015. [Online]. Available: www.netlingo.com/acronyms.php.