

Naskah Publikasi

PROYEK TUGAS AKHIR

**ANALISIS SENTIMEN KOMENTAR WARGANET
TERHADAP POSTINGAN INSTAGRAM MENGGUNAKAN
METODE *NAÏVE BAYES CLASSIFIER* DAN TF-IDF
(Studi Kasus: Instagram Gubernur Jawa Barat Ridwan Kamil)**



**MEGA KURNIA MAULIDINA
5150411382**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN ELEKTRO
UNIVERSITAS TEKNOLOGI YOGYAKARTA
YOGYAKARTA
2020**

Naskah Publikasi

**ANALISIS SENTIMEN KOMENTAR WARGANET
TERHADAP POSTINGAN INSTAGRAM MENGGUNAKAN
METODE *NAÏVE BAYES CLASSIFIER* DAN TF-IDF
(Studi Kasus: Instagram Gubernur Jawa Barat Ridwan Kamil)**

Disusun Oleh

Mega Kurnia Maulidina

5150411382

Telah disetujui oleh pembimbing



Dr. Enny Itje Sela, S.Si., M.Kom.

Tanggal 30 September 2020

ANALISIS SENTIMEN KOMENTAR WARGANET TERHADAP POSTINGAN INSTAGRAM MENGGUNAKAN METODE *NAÏVE* *BAYES CLASSIFIER* DAN TF-IDF (Studi Kasus: Instagram Gubernur Jawa Barat Ridwan Kamil)

Mega Kurnia Maulidina, Dr. Enny Itje Sela, S.Si., M.Kom.

Program Studi Informatika, Fakultas Teknologi Informasi dan Elektro

Universitas Teknologi Yogyakarta

Jl. Ringroad Utara Jombor, Sleman, Yogyakarta

E-Mail: megakurniamaulidina@gmail.com, ennysela@uty.ac.id

ABSTRAK

*Instagram merupakan media sosial yang cukup populer saat ini. Pengguna yang dimulai dari anak-anak, remaja hingga orang dewasa turut mendongkrak popularitas instagram. Dalam suatu postingan instagram, semua orang dapat dengan bebas menulis komentar. Tidak jarang publik berkomentar dengan kata-kata kasar bahkan tidak segan mengeluarkan ujaran kebencian. Begitu pula pada akun instagram Gubernur Jawa Barat Ridwan Kamil yaitu @ridwankamil biasanya kritikan, pujian, hinaan warganet tertuang dalam kolom komentar yang ada pada setiap postingannya. Untuk dapat menggali informasi dan melakukan klasifikasi sebuah teks diperlukan analisis sentiment. Analisis sentimen merupakan suatu cabang ilmu dari text mining yang digunakan untuk mengekstrak, memahami, dan mengolah data teks. Langkah pertama dengan melakukan export komentar dalam suatu postingan instagram dan kemudian untuk dilakukan preprocessing pada data yang meliputi, case folding, stemming, stopwords removing, URL and uniq character removing dan tokenizing. Untuk kemudian dilakukan pembobotan kata dengan term frequency–invers document frequency (TF-IDF). Dalam penelitian ini analisis sentimen berupa proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen negatif dan positif. Untuk mengetahui klasifikasi setiap sentimen pada komentar digunakan metode *Naïve Bayes Classifier*.*

Kata Kunci: *Analisis Sentimen, Naïve Bayes Classifier, TF-IDF, Instagram.*

1. PENDAHULUAN

Di era digital ini, media sosial merupakan salah satu tempat atau media untuk berekspresi dan berpendapat tentang berbagai macam topik. Sosial media adalah satu set baru komunikasi dan alat kolaborasi yang memungkinkan banyak jenis interaksi yang sebelumnya tidak tersedia. Instagram merupakan media sosial yang pada dasarnya berfungsi untuk berbagi foto maupun video pada sesama pengguna. Sejak September 2017 pengguna aktif instagram telah tercatat kurang lebih 800 juta orang. Dalam suatu postingan instagram, semua orang dapat dengan bebas menulis komentar guna mengekspresikan pendapat terhadap postingan tersebut. Tidak jarang publik berkomentar dengan kata-kata kasar bahkan tidak segan mengeluarkan ujaran kebencian. Begitu pula pada akun instagram Gubernur Jawa Barat Ridwan Kamil yaitu @ridwankamil yang postingannya tidak pernah luput dari komentar warganet baik itu sebuah kritikan yang membangun atau menjatuhkan. Dengan banyaknya pengguna instagram yang menyampaikan opini-opininya melalui komentar dapat dimanfaatkan untuk mencari sebuah informasi. Namun dalam pemanfaatannya

membutuhkan analisis yang tepat sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan. Analisis sentimen sangat diperlukan dalam menyaring komentar-komentar di media sosial. Analisis sentimen pada komentar dilakukan untuk mengetahui komentar yang bersifat negatif dan komentar yang bersifat positif. Salah satu metode yang dapat digunakan untuk menganalisis opini – opini komentar pada sebuah postingan instagram adalah analisa sentimen.

Analisa sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Dalam penelitian ini analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek mengandung sentimen negatif atau positif menggunakan algoritma *Naïve Bayes Classifier* (NBC).

Naïve Bayes Classifier (NBC) merupakan teknik pembelajaran mesin yang berbasis probabilistik. *Naïve Bayes Classifier* adalah metode sederhana

tetapi memiliki akurasi dan performansi yang tinggi dalam pengklasifikasian teks. Analisis sentimen yang dibangun menggunakan *Naïve Bayes Classifier* memperoleh akurasi sebesar 83% dan rata-rata harmonik sebesar 90,713% [1]. Perbandingan beberapa metode seperti metode *Naïve Bayes Classifier*, *K-nearest Neighbor*, dan gabungan *K-means* dan LVQ dalam mengklasifikasikan kategori buku berbahasa Indonesia dengan data yang digunakan berjumlah 200 buku, 150 buku digunakan sebagai data latih, sedangkan 50 buku digunakan sebagai data uji. Dari hasil penelitian yang dilakukan, metode KNN memperoleh akurasi sebesar 96%, kemudian *Naïve Bayes Classifier* sebesar 98%, lalu kombinasi *K-Means* dan LVQ menghasilkan akurasi sebesar 92,2%. Metode *Naïve Bayes Classifier* mendapatkan hasil akurasi yang tertinggi. Tingkat keakurasian pengklasifikasian sangat dipengaruhi oleh proses training. Kompleksitas data juga dapat mempengaruhi hasil akurasi.

2. LANDASAN TEORI

2.1. Analisis Sentimen

Analisis sentimen juga disebut penambangan opini, adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap dan emosi orang-orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik dan atributnya [1]. Terdapat banyak nama dan tugas yang sedikit berbeda, misalnya analisis sentimen, penambangan opini, ekstrasi pendapat, penambangan sentimen, analisis subjektivitas, analisis pengaruh, analisis emosi, penambangan peninjauan, dan lain-lain namun sekarang semuanya berada dibawah payung analisis sentimen atau penambangan opini.

2.2. PreProcessing

Preprocessing merupakan salah satu tahapan menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses pengolahan data. Dalam kasus klasifikasi dokumen yang menggunakan data bertipe teks, terdapat beberapa macam proses yang dilakukan umumnya diantaranya *case folding*, *stemming*, *stopword removal*, *Uniq character removing* dan *tokenizing*. *Preprocessing* dilakukan pada komentar terlebih dahulu, karena tidak semua atribut yang terdapat pada kolom komentar dapat digunakan untuk menganalisis masalah [2].

Preprocessing dilakukan pada komentar terlebih dahulu, karena tidak semua atribut yang terdapat pada kolom komentar dapat digunakan untuk menganalisis masalah [2]. Didalam *preprocessing* terdapat beberapa tahap, yaitu :

a. Case Folding

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

b. Stemming

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

c. Stopword Removing

Pada tahap ini akan dilakukan pengecekan disetiap kata pada komentar, selanjutnya proses menghilangkan kata-kata yang dianggap tidak penting, semisal terdapat kata sambung, kata depan, kata ganti maka akan dihapus.

d. Uniq Character Removing

Pada tahap ini akan menghapus *uniq character* biasanya berupa emoji atau serangkaian karakter yang mendefinisikan sebuah pola.

e. Tokenizing

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

2.3. TF-IDF

TF-IDF adalah suatu metode yang bisa digunakan untuk pembobotan kata. *Term Weighting* atau pembobotan kata bertujuan untuk memberikan bobot nilai pada setiap kata. Perhitungan bobot ini memerlukan dua hal yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) [3]. *Term Frequency* merupakan banyaknya jumlah kata atau term tertentu yang ada dalam suatu dokumen. Sementara *Inverse Document Frequency* adalah frekuensi kemunculan kata atau term pada seluruh dokumen. Sebelum menghitung bobot TF-IDF terlebih dahulu menghitung bobot IDF. Rumus perhitungan *Inverse Document Frequency* (IDF) dapat dilihat pada persamaan berikut :

$$IDF_t = \log\left(\frac{d}{df_t}\right) \quad (2.1)$$

Keterangan :

IDF_t = Bobot IDF ke t dokumen d.

df_t = Jumlah dokumen yang mengandung *term* t.

d = Jumlah dokumen keseluruhan.

Setelah itu barulah menghitung bobot *Term Frequency* dan *Inverse Document Frequency* (TF-IDF) dengan rumus seperti pada persamaan berikut :

$$W_{dt} = tf_t \times IDF_t \quad (2.2)$$

Keterangan:

W_{dt} = Bobot *term* ke t dokumen d.

tf_t = Jumlah kemunculan *term* t pada dokumen t.

IDF_t = Bobot IDF yang telah dihitung sebelumnya.

2.4. Naïve Bayes Classifier

Menurut Rossi, A. dkk. [4] *Naïve Bayes Classifier* adalah konsep probabilitas penentuan kelompok kelas dokumen. Algoritma klasifikasi ini dapat mengolah data dalam jumlah besar dengan hasil akurasi yang tinggi. Sedangkan menurut Munitasri, I. dkk. [5] *Naïve Bayes Classifier* adalah metode klasifikasi berdasarkan probabilitas dan teorema bayesian dengan asumsi bahwa setiap variable X bersifat bebas (*independence*), dengan mengasumsikan bahwa keberadaan sebuah atribut (*variable*) tidak ada kaitannya dengan atribut (*variable*) yang lain. Performa *Naïve Bayes* yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi ke-indepenan atribut (tidak ada kaitan antar atribut).

Naïve Bayes Classifier dilakukan dengan melakukan perhitungan *prior probability*, perhitungan *conditional probability*, dan perhitungan *posterior probability*. Rumus perhitungan *prior probability* menurut Rahman [3] dapat dilihat pada persamaan berikut :

$$P(c) = \left(\frac{d}{d_c}\right) \quad (2.3)$$

Keterangan :

C = Hipotesis data merupakan suatu *class* Spesifik.
 $P(c)$ = *Prior Probability* dari *class* c .
 d = Jumlah keseluruhan dokumen.
 d_c = Jumlah keseluruhan dokumen dari *class* c

Perhitungan *conditional probability* dilakukan dengan model multinomial. Multinomial *Naïve Bayes* digunakan untuk menghitung *conditional probability* setiap kata yang ada. Menurut Rahman [3], model multinomial memperhitungkan frekuensi setiap kata yang muncul pada dokumen. Rumus perhitungan model multinomial dapat dilihat pada persamaan berikut :

$$P(\text{Term } d_n | c) = \frac{\text{count}(d_n | c) + 1}{\text{count}(c) + |V|} \quad (2.4)$$

Keterangan :

W = *Weight* atau bobot.
 $P(\text{Term } d_n | c)$ = *Conditional Probability* kata ke- n dengan diketahui kelas c .
 W_{ct} = Nilai bobot TF-IDF atau W dari *term* d dikategori c .
 $\sum W' \in VW'_{ct}$ = Jumlah total bobot IDF kelas c atau W dari keseluruhan *term* yang berada dikategori c .
 B' = Jumlah total bobot (nilai IDF yang tidak dikali dengan TF) atau W kata unik pada seluruh dokumen.

Sementara menurut Rahman [3], perhitungan *posterior probability* dilakukan dengan rumus yang dapat dilihat pada persamaan berikut :

$$P(\text{Term } d | c) = P(c) \times P(\text{Term } d_1 | c) \times P(\text{Term } d_2 | c) \times \dots \times P(\text{Term } d_n | c) \quad (2.5)$$

Keterangan:

$P(\text{Term } d | c)$ = *Posterior Probability* suatu dokumen termasuk kelas c .
 $P(c)$ = *Prior Probability* dari kelas c .
 $P(\text{Term } d_n | c)$ = *Conditional Probability* kata ke- n dengan diketahui kelas c .

2.5. Sastrawi

Sastrawi adalah *library* yang dikembangkan untuk melakukan proses pencarian kata dasar. Menurut Rachmat C, A. dan Lukito, Y [6], kemampuan Sastrawi *stemming* tergolong baik dalam mengambil kata dasar dari setiap kata-kata yang menjadi data masukan. Instalasi Sastrawi menggunakan PHP *Composer* sehingga *library* Sastrawi dapat terunduh dan digunakan. Sastrawi juga dapat digunakan untuk proses *stopword removal*, yaitu menghilangkan kata *stopword*. Sastrawi sudah dilengkapi dengan daftar kata dalam bahasa Indonesia dan daftar kata *stopword* dalam bahasa Indonesia.

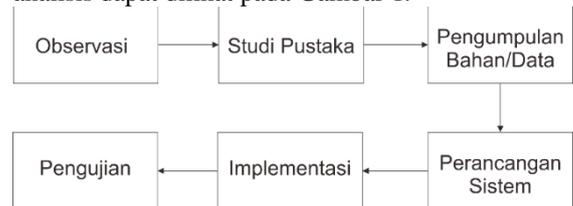
3. METODOLOGI PENELITIAN

3.1. Objek Penelitian

Objek yang digunakan dalam penelitian ini adalah berupa data komentar. Aspek penelitian yang diteliti adalah mengklasifikasikan komentar bersentimen positif dan negative.

3.2. Tahapan Proses Penelitian

Metode penelitian adalah langkah-langkah yang dilakukan oleh peneliti dalam rangka untuk mengumpulkan informasi atau data serta melakukan investigasi pada data yang telah didapatkan tersebut. Metode penelitian yang dilakukan penulis dalam penelitian yang dilakukan untuk membangun sistem analisis dapat dilihat pada Gambar 1.



Gambar 1 Alur Metode Penelitian

- a. Observasi
 Observasi merupakan proses yang dilakukan untuk mengamati data yang akan digunakan. Observasi dilakukan dengan melakukan pengamatan langsung dengan mengakses media sosial instagram untuk mengetahui komentar-komentar yang dibuat oleh pengguna instagram. Observasi dilakukan untuk memberikan

gambaran mengenai data yang akan digunakan dalam penelitian ini.

b. Studi Pustaka

Studi pustaka adalah proses pengumpulan informasi yang dibutuhkan untuk melakukan penelitian. Studi pustaka dilakukan dengan melakukan pencarian bahan-bahan dan pengambilan informasi yang relevan. Sumber studi pustaka mengambil informasi dari jurnal, skripsi dan sumber lain dengan kasus mengenai analisis sentimen.

c. Pengumpulan Bahan/Data

Proses pengumpulan data adalah proses yang dilakukan untuk mengumpulkan data yang akan digunakan sebagai data latih dan data uji. Hal ini dilakukan dengan mengambil data komentar dari instagram terutama pada postingan yang akan dianalisis. Pengambilan data dilakukan meng-*export* komentar yang ada pada postingan tersebut melalui situs <https://exportcomments.com/>.

d. Pengujian

Sistem yang telah selesai dibangun kemudian dilakukan pengujian untuk mengetahui apakah sistem telah berjalan seperti yang diharapkan dan untuk mengetahui kekurangan sistem. Pengujian juga dilakukan untuk mengetahui akurasi yang dihasilkan oleh sistem. Pengujian dilakukan dengan mengolah data yang telah dikumpulkan menggunakan sistem yang telah dibuat. Data terdiri dari 313 data latih dan 100 data uji.

e. Implementasi

Implementasi dilakukan untuk mengimplementasikan hasil perancangan ke dalam kode atau bahasa yang dimengerti oleh mesin komputer dengan Bahasa pemrograman tertentu. Dalam sistem ujian yang akan dibangun, sistem dibangun menggunakan menggunakan bahasa pemrograman PHP. Selain menggunakan bahasa pemrograman PHP, penyimpanan data untuk sistem ujian disimpan dalam basis data MySQL dan dimunculkan dalam *Website*.

f. Perancangan Sistem

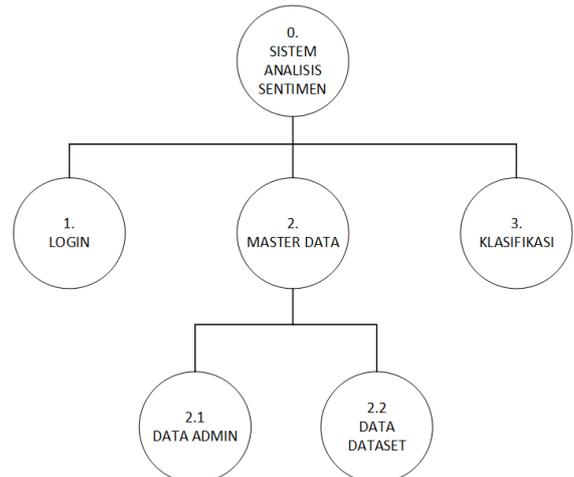
Perancangan dilakukan untuk mengetahui gambaran bagaimana sistem yang akan dibangun. Pada perancangan beberapa perancangan utama dibuat, seperti rancangan alur sistem, rancangan basis data, maupun rancangan antarmuka pengguna.

3.3. Perancangan Sistem

Perancangan sistem meliputi Diagram Jenjang, *Data Flow Diagram* (DFD), dan *Entitas Relationship Diagram* (ERD).

a. Diagram Jenjang

Diagram ini menggambarkan hubungan dan seluruh dari fungsi-fungsi didalam sistem secara berjenjang. Rancangan diagram jenjang untuk sistem pada sistem analisis sentimen ini dapat dilihat pada Gambar 2.

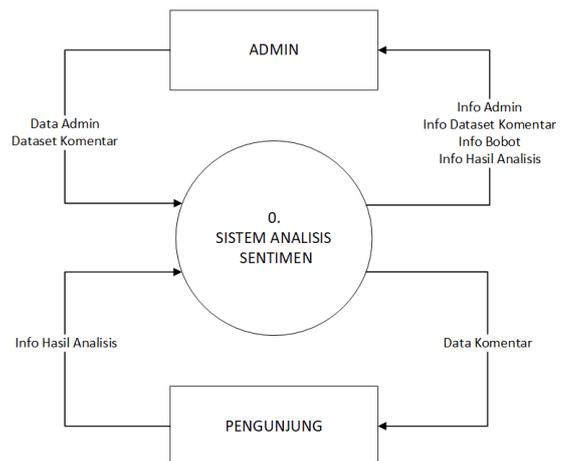


Gambar 2 Diagram Jenjang

b. *Data Flow Diagram*

1. Diagram Konteks

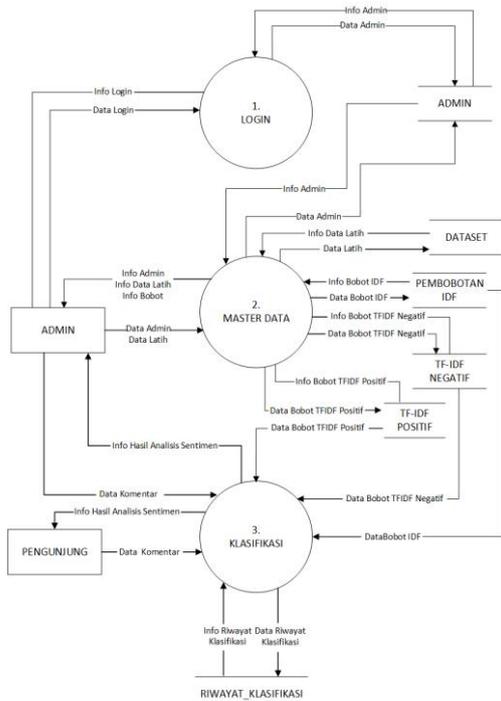
Diagram konteks adalah gambaran sistem secara garis besar. Dalam Diagram konteks digambarkan entitas apa saja yang berinteraksi dengan sistem. Entitas yang terlibat dalam sistem yang akan dibangun ada dua buah entitas, yaitu admin dan pengunjung. Rancangan diagram konteks dapat dilihat pada Gambar 3.



Gambar 3 Diagram Konteks

2. *Data Flow Diagram* Level 1

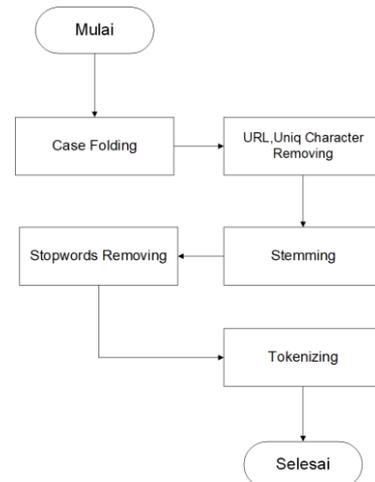
DAD Level 1 merupakan suatu gambaran yang dibuat untuk menggambarkan asal dan tujuan data yang masuk dan keluar dari sistem, serta proses yang terjadi di dalam sistem. Pada DAD Level 1 ini akan dijelaskan mengenai hubungan entitas dengan login, pengelolaan master data dan klasifikasi. Rancangan DAD Level 1 sistem analisis sentimen dapat dilihat pada



Gambar 4 DAD Level 1

Entitas terdiri dari admin dan pegunjung. Admin bisa melakukan *login*, pengelolaan data dan juga klasifikasi sedangkan pegunjung hanya bisa melakukan klasifikasi. Login dilakukan untuk masuk ke dalam sistem pengelolaan master data untuk melakukan pengelolaan terhadap master data dan klasifikasi dilakukan untuk menentukan sentimen terhadap data komentar.

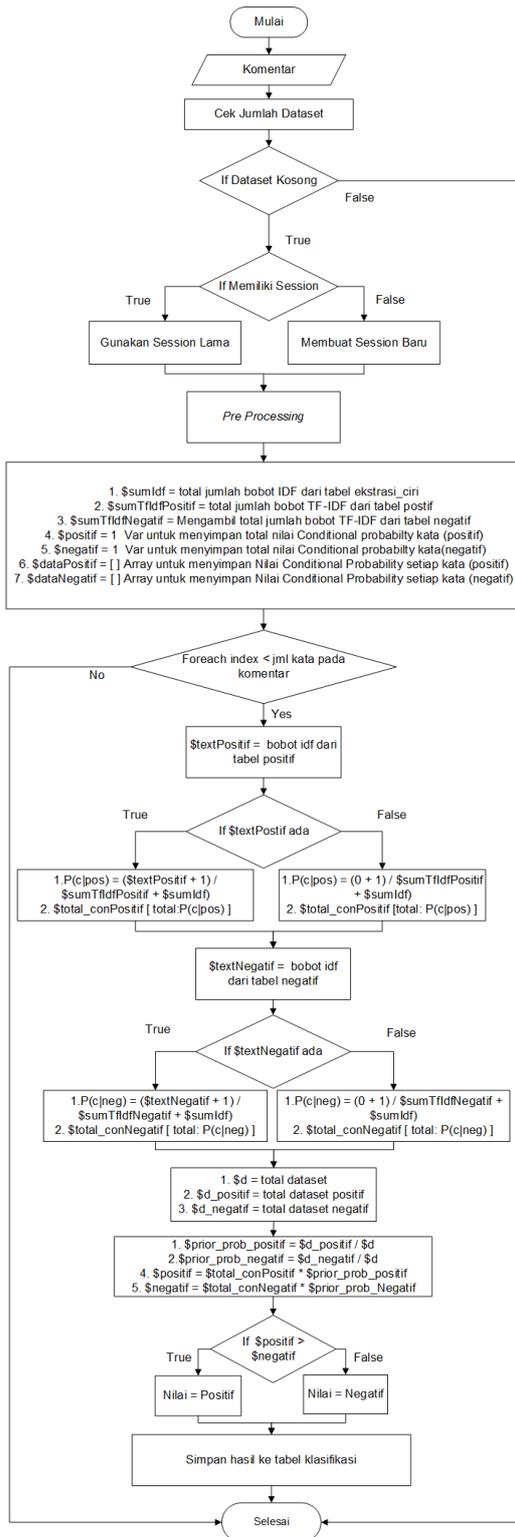
Sebelum proses klasifikasi dilakukan proses *preprocessing* yaitu pertama mengubah semua karakter menjadi huruf kecil melalui proses *casefolding*. Setelah *casefolding*, dilakukan proses menghilangkan url dan *uniq character*. Proses *stemming* dilakukan setelah kedua proses sebelumnya, proses *stemming* dilakukan untuk mendapatkan kata dasar dari suatu kata. Setelah *stemming*, dilakukan proses *stopword removal* untuk menghilangkan beberapa kata keterangan dan *tokenizing* untuk memecah kalimat berdasarkan spasi agar menjadi kata per kata. Proses *stemming* dan *stopword removal* menggunakan *library* sastra yang juga telah memiliki korpus kata dalam kamus dan juga korpus kata *stopword*. Flowchart proses *pre-processing* dapat dilihat pada



Gambar 5 Flowchart *Pre-Processing*

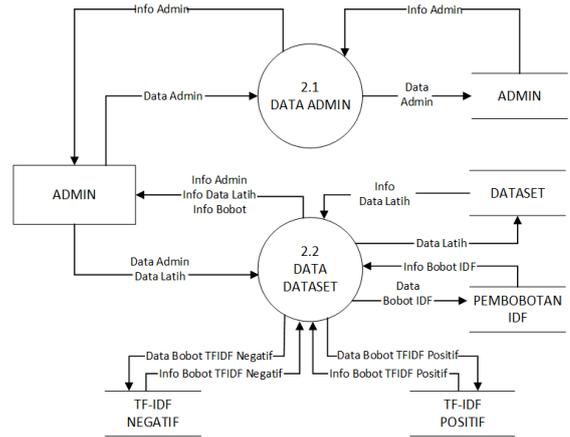
Proses klasifikasi pertama dilakukan dengan menghitung bobot setiap kata. Bobot kata dihitung dengan metode TF-IDF. Perhitungan bobot pertama dilakukan dengan mengambil semua kata dan bobot idf. Kata kemudian dihitung total kemunculannya (TF) kemudian kemuculan kata dikalikan bobot kata (IDF). Bobot TF-IDF telah didapat, langkah selanjutnya adalah menghitung nilai *prior probability* tiap kelas. *Prior Probability* didapat dengan membagi banyaknya data pada suatu kelas dengan total data pada data latih. *Prior Probability* akan digunakan kembali untuk menghitung nilai *posterior probability*.

Proses klasifikasi bisa dilakukan oleh admin dan pegunjung. Proses klasifikasi digunakan untuk menentukan kelas sentimen suatu komentar apakah positif atau negatif. Proses klasifikasi dilakukan dengan metode *Naive Bayes Classifier*



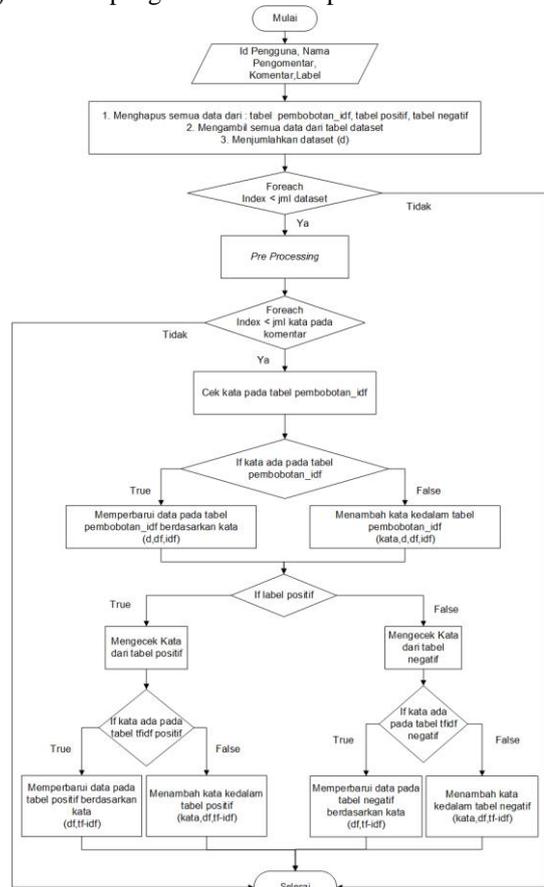
Gambar 6 Flowchart Klasifikasi

3. Data Flow Diagram Level 2 Proses 2
 DAD Level 2 Proses 2 adalah proses pengelolaan master data. Pengelolaan master data terdiri dari pengelolaan dataset dan pengelolaan data admin. Rancangan diagram aliran data level 2 proses 2 dapat dilihat pada Gambar 7.



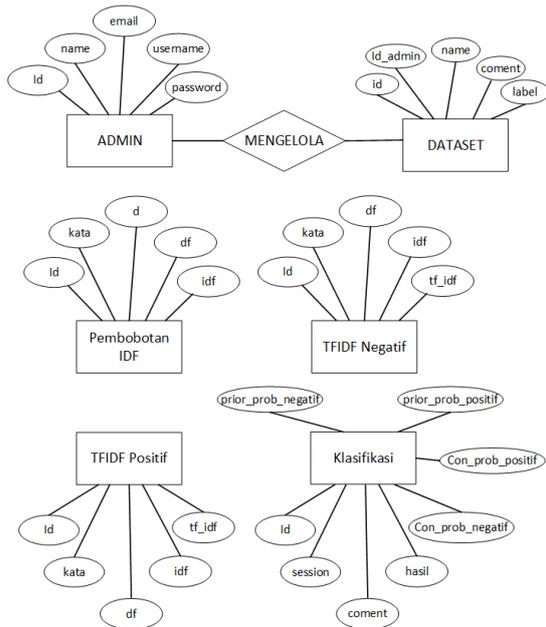
Gambar 7 DAD Level 2 Proses 2

Pengelolaan data admin dilakukan untuk mengelola data mengenai admin. Pengelolaan dataset dilakukan untuk mengelola data mengenai data latih. Proses pengelolaan dataset lebih jelasnya dapat dilihat pada flowchart pengelolaan dataset pada Gambar 8.



Gambar 8 Flowchart Pengelolaan Dataset.

c. Entitas Relationship Diagram
 Entitas Relationship Diagram (ERD) menunjukkan hubungan antar entitas. Perancangan ERD dapat dilihat pada Gambar 9.



Gambar 9 Entity Relationship Diagram.

4. HASIL DAN PEMBAHASAN

Hasil sistem yang telah dibuat diuji dengan skenario untuk pembuktian kebenaran perhitungan yang dilakukan oleh sistem. Pembuktian dilakukan dengan membandingkan hasil perhitungan manual dengan hasil yang dihasilkan oleh perhitungan sistem. Pembuktian dilakukan dengan mengerjakan perhitungan dengan data latih yang bisa dilihat pada Tabel 1 Tabel Data Latih Tabel 1 Tabel 2

Tabel 1 Tabel Data Latih

No	Komentar	Sentimen
1	Terharu ..., semoga berkah ya pak ...	Positif
2	Bandung Perlu 1000 orang Kaya kang @ridwankamil	Positif
3	heran deh kok suka tolol >:0	Negatif
4	gak tau emang suka tolol	Negatif

Tabel 2 Tabel Data Uji

No	Komentar	Sentimen
1	bikin terharu ya kang emil :(Positif

1. Perhitungan Manual

Perhitungan manual dilakukan dari proses *preprocessing* hingga menghasilkan hasil analisis sentimen. Terdapat beberapa langkah yang harus dilakukan antara lain :

- a. Melakukan *preprocessing* terhadap data latih. Proses *preprocessing* dilakukan untuk mengolah data komentar sebelum dilakukan perhitungan. Hasil dari *preprocessing* data latih dapat dilihat pada Tabel 3.

Tabel 3 Hasil *PreProcessing* Data Latih

No	Komentar	Hasil <i>PreProcessing</i>
1	Terharu ..., semoga berkah ya pak ...	haru moga berkah pak
2	Bandung Perlu 1000 orang Kaya kang @ridwankamil	bandung perlu orang kaya kang ridwankamil
3	heran deh kok suka tolol >:0	heran deh kok suka tolol
4	gak tau emang suka tolol	gak tau emang suka tolol

- b. Proses setelah *preprocessing* adalah ekstraksi ciri yaitu untuk memperoleh daftar kata unik dan nilai idf. Kata unik secara keseluruhan dan nilai idf dapat dilihat pada Tabel 4 dan Tabel 5. Kata unik kelas positif dan bobot TF-IDF nya dapat dilihat pada Tabel 6 dan Tabel 7. sedangkan kata unik kelas positif dan bobot TF-IDF nya dapat dilihat pada Tabel 8 dan Tabel 9.

Tabel 4 Kata Unik dan Nilai IDF

No	Kata	d	df	idf
1	bandung	4	1	$\text{Log}(4/1)$ = 0.60205999132796
2	berkah	4	1	$\text{Log}(4/1)$ = 0.60205999132796
3	deh	4	1	$\text{Log}(4/1)$ = 0.60205999132796
4	emang	4	1	$\text{Log}(4/1)$ = 0.60205999132796
5	gak	4	1	$\text{Log}(4/1)$ = 0.60205999132796
6	haru	4	1	$\text{Log}(4/1)$ = 0.60205999132796
7	heran	4	1	$\text{Log}(4/1)$ = 0.60205999132796
8	kang	4	1	$\text{Log}(4/1)$ = 0.60205999132796
9	kaya	4	1	$\text{Log}(4/1)$ = 0.60205999132796
10	kok	4	1	$\text{Log}(4/1)$ = 0.60205999132796

Tabel 5 Lanjutan

No	Kata	d	df	idf
11	moga	4	1	Log(4/1) = 0.60205999132 796
12	orang	4	1	Log(4/1) = 0.60205999132 796
13	pak	4	1	Log(4/1) = 0.60205999132 796
14	perlu	4	1	Log(4/1) = 0.60205999132 796
15	ridwan kamil	4	1	Log(4/1) = 0.60205999132 796
16	suka	4	2	Log(4/2) = 0.30102999566 398
17	tau	4	1	Log(4/1) = 0.60205999132 796
18	tolol	4	2	Log(4/2) = 0.30102999566 398
Total				10.2350198525 75322

Tabel 6 Kata Unik Kelas Positif dan Bobotnya

No	Kata	df	idf	tf-idf
1	bandung	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
2	berkah	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
3	haru	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796

Tabel 7 Lanjutan

No	Kata	df	idf	tf-idf
4	kang	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
5	kaya	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
6	moga	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
7	orang	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
8	pak	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
9	perlu	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
10	ridwank amil	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796
Total				6.0205999 13279599

Tabel 8 Kata Unik Kelas Negatif dan Bobotnya

No	Kata	df	idf	tf-idf
1	deh	1	0.60205 9991327 96	= 1 * 0.6020599 9132796 = 0.6020599 9132796

Tabel 9 Lanjutan

No	Kata	df	idf	tf-idf
2	emang	1	$\frac{0.60205}{999132796}$	$= 1 * \frac{0.6020599}{9132796}$
3	gak	1	$\frac{0.60205}{999132796}$	$= 1 * \frac{0.6020599}{9132796}$
4	heran	1	$\frac{0.60205}{999132796}$	$= 1 * \frac{0.6020599}{9132796}$
5	kok	1	$\frac{0.60205}{999132796}$	$= 1 * \frac{0.6020599}{9132796}$
6	suka	2	$\frac{0.30102}{999566398}$	$= 2 * \frac{0.3010299}{9566398}$
7	tau	1	$\frac{0.60205}{999132796}$	$= 1 * \frac{0.6020599}{9132796}$
8	tolol	2	$\frac{0.30102}{999566398}$	$= 2 * \frac{0.3010299}{9566398}$
Total				4.8164799 3062368

c. Setelah didapatkan daftar kata dan bobot TF-IDFnya kemudian dilakukan perhitungan naive bayes pada data uji. Sebelum diuji data uji juga harus di *preprocessing* terlebih dahulu. Hasil dari *preprocessing* data uji dapat dilihat pada

No	Komentar	Hasil PreProcessing
1	bikin terharu ya kang emil :(bikin haru kang emil

- d. Setelah *dipreprocessing* data uji kemudian dihitung probabilitasnya terhadap kelas positif dan negatif dengan rumus naive bayes yang telah dijelaskan pada landasan teori.
- e. Menghitung probabilitas komentar terhadap kelas positif dimulai dengan menghitung prior probability dengan cara sebagai berikut:

$$P(\text{Positif}) = \left(\frac{2}{4}\right) = 0,5$$

Setelah menghitung *prior probability* kemudian menghitung masing-masing kata unik yaitu bikin, haru, kang, emil dengan perhitungan *conditional probability* seperti pada Tabel 10 berikut:

Tabel 10 Tabel Perhitungan *Conditional Probability Positif*

$P(\text{bikin} \text{positif})$ $= \left(\frac{0 + 1}{6.020599913279599 + 10.235019852575322}\right)$ $= 0.061517186942358804$
$P(\text{haru} \text{positif})$ $= \left(\frac{0.60205999132796 + 1}{6.020599913279599 + 10.235019852575322}\right)$ $= 0.09855422397939584$
$P(\text{kang} \text{positif})$ $= \left(\frac{0.60205999132796 + 1}{6.020599913279599 + 10.235019852575322}\right)$ $= 0.09855422397939584$
$P(\text{emil} \text{positif})$ $= \left(\frac{0 + 1}{6.020599913279599 + 10.235019852575322}\right)$ $= 0.061517186942358804$

Langkah selanjutnya adalah menghitung *posterior probability* dengan perhitungan seperti pada Tabel 11 berikut:

Tabel 11 Tabel Perhitungan *Posterior Probability Positif*

$P(\text{Positif} \text{bikin haru kang emil})$
$= (0,5 \times 0,061517186942358804 \times 0,09855422397939584 \times 0,09855422397939584 \times 0,061517186942358804)$ $= 0,000036757284601186965$

- f. Menghitung *probabilitas* komentar terhadap kelas negatif dimulai dengan menghitung *prior probability* dengan cara sebagai berikut:

$$P(\text{Negatif}) = \left(\frac{2}{4}\right) = 0,5$$

Setelah menghitung *prior probability* kemudian menghitung masing-masing kata unik yaitu bikin, haru, kang, emil dengan perhitungan *conditional probability* seperti pada Tabel 12 berikut:

Tabel 12 Tabel Perhitungan *Conditional Probability Negatif*

$P(\text{bikin} \text{negatif})$ $= \left(\frac{0 + 1}{4.81647993062368 + 10.235019852575322}\right)$ $= 0.0664385618977475$
$P(\text{haru} \text{negatif})$ $= \left(\frac{0 + 1}{4.81647993062368 + 10.235019852575322}\right)$ $= 0.0664385618977475$
$P(\text{kang} \text{negatif})$ $= \left(\frac{0 + 1}{4.81647993062368 + 10.235019852575322}\right)$ $= 0.0664385618977475$
$P(\text{emil} \text{negatif})$ $= \left(\frac{0 + 1}{4.81647993062368 + 10.235019852575322}\right)$ $= 0.0664385618977475$

Langkah selanjutnya adalah menghitung *posterior probability* dengan perhitungan seperti pada Tabel 13 berikut :

Tabel 13 Tabel Perhitungan *Posterior Probability Negatif*

$P(\text{Negatif} \text{kbikin haru kang emil})$ $= (0,5 \times 0,0664385618977475 \times 0,0664385618977475 \times 0,0664385618977475 \times 0,0664385618977475)$ $= 0,000019484124378963816$
--

- g. Probabilitas data uji terhadap masing-masing sentimen sudah diketahui, selanjutnya adalah mencari nilai tertinggi. Karena nilai *posterior probability* kelas positif

$$(0,000036757284601186965)$$

lebih besar daripada *posterior probability* kelas negatif

$$(0,000019484124378963816)$$

maka data uji termasuk dalam kelas **positif**.

2. Perhitungan Oleh Sistem

Perhitungan oleh sistem dilakukan dengan mengunggah data latih dan data uji yang sama dengan data latih dan data uji dalam perhitungan manual. Data latih dan data uji yang diunggah adalah berupa *file* yang berekstensi .xls. Langkah yang dilakukan untuk melakukan perhitungan dengan sistem adalah sebagai berikut

- a. Unggah data latih. Langkah pertama adalah mengunggah data latih, data latih yang diunggah dapat dilihat pada Gambar 10.

#	Name	Username	Label	Aktif
1	ibrahim	gpn-10000000000000000000	neg	aktif
2	ibrahim	gpn-10000000000000000000	neg	aktif
3	ibrahim	gpn-10000000000000000000	neg	aktif
4	ibrahim	gpn-10000000000000000000	neg	aktif

Gambar 10 Data Latih yang Diunggah

- b. Data latih yang diunggah juga telah dilakukan proses *preprocessing* dan perhitungan bobot IDF dan TF-IDF. Hasil dari *preprocessing* yang menghasilkan kata dan bobot IDF dapat dilihat pada Gambar 11 dan Gambar 12 sedangkan TF-IDF dapat dilihat pada Gambar 13 dan Gambar 14.

#	Kata	id	df	bobot idf	Perhitungan
1	bandung	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
2	banjar	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
3	gpn	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
4	emang	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
5	gpn	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
6	haru	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
7	haru	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
8	kang	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
9	kang	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$
10	kang	4	1	0.60205999132796	$IDF = \log\left(\frac{N}{df}\right) = \log\left(\frac{10}{1}\right) = 0.60205999132796$

Gambar 11 Hasil *Preprocessing* dan Perhitungan Nilai IDF

#	Kata	TF	IDF	Perhitungan
11	moga	4	1	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$
12	orang	4	1	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$
13	pak	4	1	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$
14	satu	4	1	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$
15	hidupkan	4	1	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$
16	satu	4	2	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$
17	satu	4	1	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$
18	satu	4	2	$IDF = \log(10)$ $= \log(10)$ $= 0.30103000000000004$

Gambar 12 Lanjutan

#	Kata	TF	IDF	Perhitungan
1	tanding	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
2	berkah	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
3	baru	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
4	kang	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
5	kang	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
6	moga	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
7	moga	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
8	perlu	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
9	hidupkan	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$

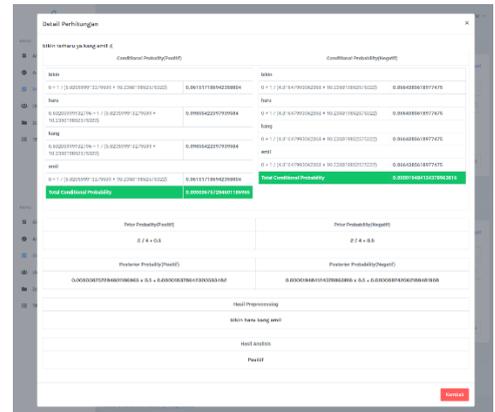
Gambar 13 Kata dan Bobot TF-IDF pada Kelas Positif

#	Kata	TF	IDF	Perhitungan
1	diti	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
2	manga	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
3	gan	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
4	heran	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
5	bah	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
6	satu	2	0.30103000000000004	$TF \cdot IDF = 2 \cdot 0.30103000000000004$ $= 0.60205999132796$
7	satu	1	0.60205999132796	$TF \cdot IDF = 1 \cdot 0.60205999132796$ $= 0.60205999132796$
8	satu	2	0.30103000000000004	$TF \cdot IDF = 2 \cdot 0.30103000000000004$ $= 0.60205999132796$

Gambar 14 Kata dan Bobot TF-IDF pada Kelas Negatif

- c. Unggah data uji. Setelah data latih sukses diunggah maka langkah selanjutnya adalah mengunggah data uji dan memilih data latih

yang digunakan. Setelah data uji diproses maka akan muncul hasil perhitungan. Hasil perhitungan oleh sistem dapat dilihat pada Gambar 15.



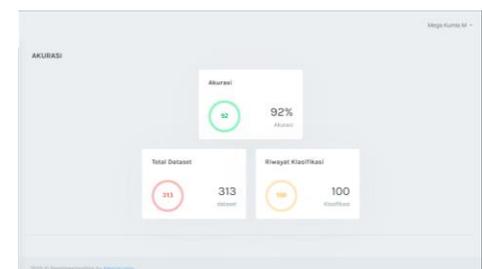
Gambar 15 Detail Hasil Perhitungan

Sistem yang telah melakukan perhitungan akan menampilkan detail perhitungan dan hasilnya. Perhitungan oleh sistem menghasilkan bahwa data uji bersentimen **positif**.

3. Pembahasan

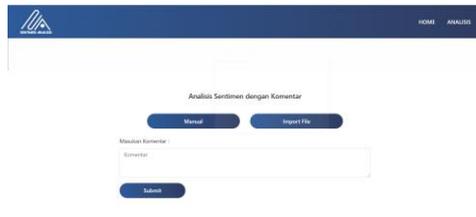
- a. Pengujian dengan dataset uji

Sistem analisis sentimen menggunakan naive bayes ini diuji dengan menggunakan 313 data latih dan 100 data uji. Kata unik yang dihasilkan dari 110 komentar adalah 1453 kata unik, sedangkan komentar bersentimen positif memiliki 766 kata unik dan komentar bersentimen negatif memiliki 953 kata unik. Akurasi yang dihasilkan menggunakan 313 data latih dan 100 data uji adalah 92%. Artinya sistem analisis sentimen mampu menganalisis 92 komentar dengan benar dari 100 komentar yang digunakan. Hasil pengujian dengan dataset uji dapat dilihat pada Gambar 16.



Gambar 16 Hasil pengujian dengan data uji.

- b. Pengujian dengan komentar oleh pengguna
 Pengujian ini dilakukan dengan meminta beberapa pengguna menilai postingan Instagram untuk kemudian menuliskan komentarnya. Setelah hasil keluar, pengguna diminta pendapatnya apakah hasil analisis sentimen sesuai dengan dengan pendapatnya. Contoh pengujian oleh pengguna bisa dilihat pada Gambar 17.



Gambar 17 Input komentar oleh pengguna

Pengujian ini mendapatkan beberapa komentar yang diuji oleh pengguna. Terdapat 10 komentar yang diuji. Hasil dari pengujian beragam, terdapat beberapa yang sesuai dan beberapa yang tidak sesuai. Komentar yang diuji dan hasilnya dapat dilihat pada Tabel 14 dan Gambar 18.

Tabel 14 Komentar yang Diuji oleh Pengguna

No	Komentar	Hasil	Kesesuaian
1	Bismillah smoga berguna. Sehat selalu Kang Emil.	Positif	Sesuai
2	kang emil coba kalo kerja tuh yang bener kerja kok males	Positif	Tidak Sesuai
3	duh gubernur kok kerjanya nya foto	Negatif	Sesuai
4	Bandung indah gara gara kang emil, terimakasih kang	Positif	Sesuai
5	Semoga sehat terus ya kang, kerjanya bagus	Positif	Sesuai
6	wih jembatan nya bagus kang	Negatif	Tidak Sesuai
7	kang emil mah pamer mulu ah sebel	Negatif	Sesuai
8	jembatan gitu doang mah sepele	Negatif	Sesuai
9	fotonya jelek kang	Negatif	Sesuai
10	tidak indah begitu kang emil coba dibuat lebih resik lagi	Positif	Tidak Sesuai

Gambar 18 Hasil Pengujian oleh pengguna

Hasil yang didapatkan dari pengujian terhadap 10 komentar oleh pengguna adalah 7 hasil analisis sesuai dengan pendapat pengguna sedangkan 3 hasil analisis tidak sesuai dengan pendapat pengguna. Pada hasil yang tidak sesuai terdapat kata beberapa yang tidak terdapat pada fitur data latih. Pada komentar 'kang emil gak jelek rajin juga' terdapat kata negasi 'tidak' yang diikuti oleh kata 'bagus' sehingga mengubah arti kalimat menjadi negatif, namun hasil dari sistem adalah positif. Kedua kesalahan yang terjadi karena tidak ditemukannya kata dalam fitur yang digunakan untuk perhitungan dan juga kata negasi yang mampu merubah arti ini kemungkinan bisa lebih diminimalisir lagi dengan cara menambahkan data latih yang lebih banyak dan bervariasi dan juga menambahkan algoritma deteksi kata negasi.

5. PENUTUP

5.1. Kesimpulan

Berdasarkan keseluruhan proses analisis, perancangan, dan implementasi, pada pembuatan sistem analisis sentimen menggunakan metode *naive bayes classifier* dan TF-IDF, diperoleh kesimpulan sebagai berikut :

- Sistem analisis sentimen komentar warganet menggunakan metode *naive bayes classifier* dan TF-IDF ini mampu melakukan analisis sentimen terhadap data yang berupa komentar tekstual.
- Tingkat keakuratan sistem analisis sentimen komentar warganet menggunakan metode *naive bayes classifier* dan TF-IDF yang diuji menggunakan data uji mendapatkan nilai 92%.
- Masih terdapat beberapa kesalahan ketika diuji oleh beberapa pengguna karena kalimat yang lebih bervariasi.

5.2. Saran

Berdasarkan kesimpulan yang telah diperoleh, maka terdapat beberapa saran agar kedepannya dapat diperoleh hasil yang lebih baik. Saran untuk penelitian selanjutnya adalah sebagai berikut.

- Pengembangan sistem selanjutnya, dapat diterapkan algoritma yang mampu menangani kesalahan penulisan atau algoritma perbaikan kata tidak baku.
- Disarankan untuk menambahkan data latih yang lebih banyak lagi dan bervariasi sehingga fitur kata yang dimiliki lebih banyak.

- c. Peneliti selanjutnya disarankan menambahkan fitur deteksi negasi (*convert negation*).

6. DAFTAR PUSTAKA

- [1] Liu, B. (2012), *Sentiment Analysis And Opinion Mining* no. May .
- [2] Salam, A., Zeniarja, J. and Khasanah, R. S. U. (2018), *Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)* Prosiding SINTAK pp. 480–486.
- [3] Rahman, A. (2017), *Online News Classification Using Multinomial Naive Bayes* vol. 6, no. 1 .
- [4] Rossi, A., Lestari, T., Setya Perdana, R. and Fauzi, M. A. (2017), *Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes Dan Pembobotan Emoji* Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 1, no. 12 pp. 1718–1724.
- [5] Munitasri, I., Santosa, S. and Supriyanto, C. (2018), *ALGORITMA NAIVE BAYES DENGAN SELEKSI* Jurnal Teknologi Informasi, vol. 14 .
- [6] Rachmat C, A. and Lukito, Y. (2016), *Klasifikasi Sentimen Komentar Politik Dari Facebook Page Menggunakan Naive Bayes* Jurnal Informatika dan Sistem Informasi Universitas Ciputra, vol. 02, no. 02 pp. 26–34.