

**SENTIMENT ANALYSIS KOMENTAR YOUTUBE SAMSUNG S20  
MENGUNAKAN METODE MAJORITY VOTE  
SAMSUNG S20 YOUTUBE COMMENTS SENTIMENT ANALYSIS USING MAJORITY  
VOTE METHOD**

**Risang Putra Pradana<sup>1)</sup>, Deni Arifianto<sup>2)</sup>, Habibatul Azizah Al Faruq<sup>3)</sup>**

<sup>1</sup>Fakultas Teknik, Universitas Muhammadiyah Jember  
email: [risangard@gmail.com](mailto:risangard@gmail.com)

<sup>2</sup>Fakultas Teknik, Universitas Muhammadiyah Jember  
email: [deniarifianto@unmuhjember.ac.id](mailto:deniarifianto@unmuhjember.ac.id)

<sup>3</sup>Fakultas Teknik, Universitas Muhammadiyah Jember  
email: [habibatulazizah@unmuhjember.ac.id](mailto:habibatulazizah@unmuhjember.ac.id)

**Abstrak**

Samsung S20 merupakan *flagship* keluar Samsung, meskipun Samsung Galaxy S20 hadir dengan banyak fitur, namun informasi tentang tingkat penerimaan konsumen terhadap produk ini tetap dibutuhkan. Dalam penelitian ini, penulis akan melakukan analisis sentimen berdasarkan komentar yang terdapat dalam *YouTube review* Samsung S20, dengan menerapkan proses *text mining Majority vote* yang terdiri dari perbandingan metode *Support Vector Machine*, *K-Nearest Neighbors*, dan *Naïve Bayes*, untuk mengklasifikasikan apakah teks termasuk dalam sentimen positif, negatif dan netral. Data diklasifikasikan secara manual dengan mengelompokkan menjadi kelas sentimen positif, negatif dan netral kemudian secara otomatis data latih akan mengambil beberapa data untuk data uji dan menguji kemampuan sistem. Hasil yang didapatkan dari penelitian ini adalah dengan menggunakan Majority vote mendapatkan nilai akurasi 0.8 dengan menggunakan 100 data dan mendapatkan nilai akurasi 0.48 dengan menggunakan 50 data uji.

**Keywords:** *Samsung S20, YouTube, text mining, Majority Vote.*

**Abstract**

*Samsung S20 is the latest Samsung's flagship, even though Samsung Galaxy S20 comes with many features, information about the level of consumer acceptance of this product is necessary. In this study, the author will conduct a sentiment analysis based on the customer comments in the YouTube review of the Samsung S20, by applying the Majority vote text mining process which consists of a comparison of the Support Vector Machine, K-Nearest Neighbors, and Naïve Bayes methods, to classify whether the text is included in the positive, negative and neutral sentiment. Data is classified manually by grouping into positive, negative and neutral sentiment classes then automatically the accustomed data will take some data for test data and test the system's ability. The results obtained from this study are the data accuracy from 100 data is 0.8 and 0.48 from 50 data tests by using Majority Vote.*

**Keywords:** *Samsung S20, YouTube, text mining, Majority Vote.*

## 1. PENDAHULUAN

Samsung *Galaxy S20* adalah *smartphone flagship* keluaran terbaru Samsung pada tahun 2020. Memiliki ukuran yang hanya 6.2-inch membuatnya mudah digunakan dengan satu tangan, ringkas dan mudah dibawa. Tampilan dari Samsung *Galaxy S20* terkesan simple dan elegan karena bagian layar dan belakang pada Samsung *Galaxy S20* terbuat dari kaca. Menggunakan layar *Dynamic AMOLED* dengan fitur 120hz pastinya membuat mata nyaman saat memandang layar Samsung *Galaxy S20*. Samsung *Galaxy S20* memiliki *speaker stereo* yang lantang dan jelas, membuat pengguna akan nyaman ketika mendengarkan music. Dalam penggunaan sehari-hari dengan mengandalkan prosesor *Exynos 990*, RAM 8GB dan media penyimpanan internal sebesar 128 GB membuat Samsung *Galaxy S20* dapat digunakan dalam berbagai hal, seperti memainkan *game*, menjalankan berbagai jenis multimedia, *browsing*, atau membuka Instagram. Berkat baterai sebesar 4000 mAh, Samsung *Galaxy S20* dapat digunakan seharian penuh. Dan fitur andalan Samsung *Galaxy S20* adalah kamera. Samsung *Galaxy S20* memiliki tiga buah kamera yaitu kamera *Ultra-wide 12 MP*, kamera *Wide 12 MP* dan kamera *Telephoto 64 MP*, dengan tiga buah kamera yang ada, dipastikan kamera Samsung *Galaxy S20* dapat diandalkan dalam berbagai aplikasi *photography*, seperti untuk mengambil foto pemandangan, mengambil foto pada malam hari atau sekedar mengambil foto untuk di *upload* di Instagram.

Meskipun Samsung *Galaxy S20* hadir dengan banyak fitur, namun informasi tentang tingkat penerimaan konsumen terhadap produk ini tetap dibutuhkan. Informasi itu disebut *Consumer Preference*. *Consumer Preference* adalah penilaian subjektif konsumen individu diukur dari kepuasan mereka dalam menggunakan produk yang telah mereka beli. *Consumer Preference* penting diketahui oleh produsen dan konsumen potensial. Konsumen potensial adalah konsumen yang berminat

melakukan pembelian terhadap produk yang ditawarkan. Berdasarkan penelitian terdahulu (Næs et al., 2010; Rintyarna, 2016), *Consumer preference* umumnya di ekstrak dengan metode *Conjoint Analysis*. *Conjoint Analysis* adalah metode riset pemasaran yang paling banyak digunakan untuk menganalisis preferensi konsumen. Data yang digunakan pada teknik ini adalah data yang dikumpulkan dengan survei.

Teknik survei dianggap *time-consuming* dan *costly* (perlu waktu lama dan biaya yang besar). Penelitian Tugas Akhir ini mengusulkan sebuah pendekatan berbasis Teknik *Sentiment analysis* untuk mengetahui tingkat *Consumer Preference* tanpa melakukan survei. Analisis sentimen adalah bidang studi yang menganalisis pendapat seseorang, opini, evaluasi, penilaian, sikap dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik dan atribut (Balya, 2019). Teknik *Sentiment analysis* akan diaplikasikan pada data komentar *YouTube* yang berkaitan dengan produk Samsung *Galaxy S20*.

Banyak macam metode untuk menghitung *Sentiment analysis* seperti SVM, KNN dan *Naïve Bayes*. Metode *Majority vote* yang digunakan oleh (Zamahsyari & Nurwidyanoro, 2017) dalam menghitung sentimen komentar dapat menghasilkan nilai presisi dan akurasi yang lebih baik yaitu sebesar 75% dan 72%, sedangkan metode KNN yang digunakan oleh (Rosdiansyah, 2014) menghasilkan nilai akurasi 70%, dan *Naïve Bayes* berdasarkan penelitian yang telah dilakukan oleh (Sipayung et al., 2016) menunjukkan *Naïve Bayes* dengan membagi perhitungan dengan dua bagian, mendapatkan hasil akurasi sistem memiliki nilai *accuracy* kategori 77.14% dan untuk *precision sentiment* 99.12%, *recall sentiment* 72.9%, dan *accuracy sentiment* 75.42%.

Untuk menyeimbangkan kelemahan dari masing – masing metode terdapat metode yang dapat membandingkan ketiga metode tersebut yaitu metode *Sentiment analysis* yang akan digunakan berbasis teknik *Majority Vote*. *Majority vote* adalah teknik membandingkan hasil klasifikasi dari sejumlah algoritma

*Machine Learning*. Menurut Zamahsyari & Nurwidyantoro (2017) berpendapat teknik *Majority vote* mampu meningkatkan kinerja algoritma *Machine Learning*.

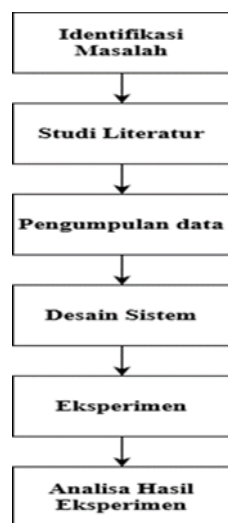
Penelitian dengan topik *Sentiment analysis* pada data komentar *YouTube* pernah dilakukan oleh (Balya, 2019) menggunakan *Gaussian Naïve Bayes* dan *Multinomial Gaussian Naïve Bayes*. Penelitian ini salah satunya dimaksudkan untuk mengetahui tingkat akurasi dari teknik *Majority Vote*. Algoritma yang akan digunakan dalam perbandingan dalam *Majority vote* adalah SVM, KNN dan *Naïve Bayes*.

Data yang akan digunakan dalam penelitian ini berasal dari *channel YouTube "GSMarena Official"*. Channel tersebut dibuat pada tanggal 6 Februari 2007, di *channel* itu berisi tentang video *hands-on, unboxing* dan berbagai video yang terkait dengan dunia teknologi *smartphone*.

## 2. METODE PENELITIAN

### 2.1. Alur Penelitian

Metodologi penelitian merupakan tahapan-tahapan yang dilakukan selama penelitian berlangsung agar pelaksanaan dapat tersusun dengan baik dan sistematis agar dapat mencapai tujuan seperti yang diharapkan. Pada penelitian ini diperlukan langkah-langkah kegiatan penelitian untuk mendapatkan hasil yang maksimal seperti yang ditunjukkan pada Gambar 1.



Gambar 1 Alur penelitian

### 2.2. Identifikasi Masalah

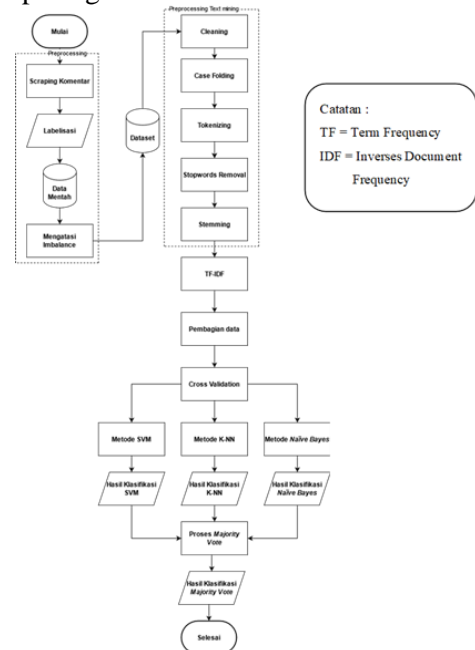
Tahapan awal setelah menentukan topik penelitian yang akan diambil adalah mengidentifikasi permasalahan yang akan dipelajari. Tahapan Identifikasi Masalah dimaksud sebagai penegasan batas-batas permasalahan, sehingga cakupan penelitian tidak keluar dari tujuan. Dalam penelitian ini, proses identifikasi masalah tentang komentar penonton video *review Samsung Galaxy S20 di YouTube*, yang berupa singkatan, Bahasa asing, dan berbagai masalah dalam melakukan *sentiment analysis* pada *YouTube*. Untuk itu diperlukan Algoritma yang mampu menemukan pesan yang terkandung dalam komentar penonton di *YouTube* dan sekaligus dapat menyeleksi kata yang tidak diperlukan dan mengubahnya menjadi data yang dapat dijadikan fitur.

### 2.3. Studi Literatur

Pada tahap ini akan dilakukan dengan cara mengumpulkan bahan referensi yaitu berupa buku, artikel, paper, jurnal, dan makalah yang berkaitan dengan metode TF-IDF, Sentiment Analysis, Algoritma Support Vector Machine, Algoritma K-Nearest Neighbors, Algoritma Naïve Bayes dan Metode Majority Vote.

### 2.4. Desain Sistem

Proses perancangan desain sistem dapat dilihat pada gambar 3.2 berikut.



Gambar 2 Alur sistem

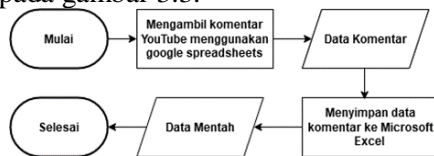
a. Preprocessing

Preprocessing dilakukan untuk mendapatkan data yang akan digunakan dalam penelitian ini dan mengolah data menjadi dataset.

Dalam preprocessing ada beberapa langkah-langkah yang harus dilakukan yaitu:

a. Scraping Komentar

Komentar yang akan di scraping untuk penelitian ini diambil dari video komentar YouTube review Samsung s20. Proses scraping komentar pada YouTube dapat dilihat pada gambar 3.3.



Gambar 3 Scraping komentar

Sebelum melakukan scraping, penulis mengunjungi video YouTube review Samsung s20 dan kemudian data komentar di scraping menggunakan google spreadsheets dengan menyalin video id pada link video. Proses scraping menggunakan google spreadsheets. Proses scraping menggunakan google spreadsheets dapat dilihat pada gambar 4.

	Name	Comment	Time	Likes	Reply Count
1	GSMArena Offi	Our detailed Gal	2020-03-20T22:5	17	3
2	bob no name	Watching from m	2020-12-25T16:2	0	0
3	Alen Dechev	<a href="http://b	2020-12-23T14:3	0	0
4	Lachlan Mckenzi	<a href="https://	2020-12-19T09:2	1	0
5	you U	Nice voice.	2020-12-13T16:2	1	0
6	Sid Ga	Bring back curv	2020-12-10T22:1	1	0
7	Bimobile Cast	What shooting g	2020-12-04T01:2	1	0
8	Jesse James Pa	Can you please t	2020-11-27T11:4	0	0
9	Night Hood	ve seen and tou	2020-11-24T17:1	0	0
10	Imran Momin	Exactly...battery	2020-11-22T06:5	0	0

Gambar 4 Scraping dengan google spreadsheets

b. Labelisasi

Labelisasi dilakukan untuk memberikan nilai pada setiap teks komentar, dimana terdapat tiga kategori sentimen yaitu sentiment positif, negatif dan netral.

c. Mengatasi imbalance

Teknik yang digunakan untuk mengatasi imbalance adalah menggunakan Teknik random undersampling.

d. Preprocessing Text mining

Dalam preprocessing text mining ada beberapa langkah-langkah yang harus

dilakukan yaitu: *Cleaning, Case Folding, Tokenizing, Stopword Removal, Stemming.*

e. TF-IDF

Setelah melakukan preprocessing text mining akan dilakukan proses untuk mengekstraksi fitur variabel pada term yang ada pada komentar. Proses ini menggunakan metode TF-IDF.

f. Pembagian Data

Setelah mendapatkan hasil dari TF-IDF untuk mengetahui seberapa penting kata tersebut sering digunakan pada suatu teks. Langkah selanjutnya adalah membagi data menjadi 2 bagian, 100 digunakan dalam proses cross validation untuk mencari Langkah uji terbaik dan 50 data untuk menguji Langkah uji terbaik yang telah didapatkan melalui proses cross validation.

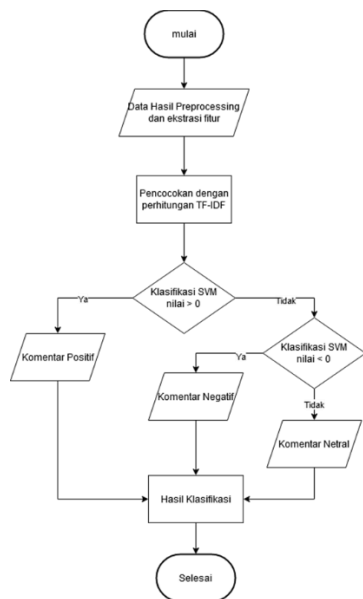
g. Cross Validation

Setelah melakukan pembagian data selanjutnya adalah Cross validation dan data yang digunakan sebanyak 100 data

h. Algoritma Support Vector Machine (SVM)

Setelah mendapatkan hasil dari TF-IDF untuk mengetahui seberapa penting kata tersebut sering digunakan pada suatu teks. Langkah selanjutnya adalah mengubah teks menjadi vector. Kemudian SVM akan menemukan garis untuk membagi 2 kelas dan mengklasifikasi teks testing berdasarkan di sisi mana garis yang ditemukan.

Flowchart proses klasifikasi menggunakan Support Vector Machine dapat dilihat pada gambar 5.



Gambar 5 Alur SVM

Berdasarkan teks diatas, diambil salah satu document (D3) untuk dilakukan contoh perhitungan yaitu sebagai berikut:

D7: [samsung, best, world]

Kata	TF	DF/N	IDF
samsung	2	7	0,544
best	2	7	0,544
world	1	7	0,845

Catatan:

- TF : Term frequency
- DF : Document frequency
- N : Banyaknya kata
- IDF : Invers document frequency

TF pada kata “iphone” bernilai 1 karena dari 6 teks yang ada kata “iphone” hanya ada 1 begitu seterusnya pada kata yang lain. DF adalah 7 karena jumlah teks dibagi banyak data. Nilai DF/N untuk kata “best” = 7/2, begitu seterusnya untuk kata yang lain. Nilai IDF diperoleh dari Log (DF/N). Kemudian nilai IDF diurutkan berdasarkan nilai terbesar sampai terkecil,

Kemudian dilakukan perhitungan untuk mencari nilai bobot(w) digunakan rumus pada persamaan (3.1) sebagai berikut:

$$T1=T2=2/0,544=3,676$$

$$T3=1/0,845=1,183$$

Setelah Nilai Bobot (w) sudah didapatkan, dihitung nilai bias (b) dengan menggunakan rumus persamaan (1) yaitu sebagai berikut:

$$b = \frac{-1}{2} \begin{pmatrix} 3,676 \\ 3,676 \\ 1,183 \end{pmatrix} \begin{bmatrix} -0,544 \\ -0,544 \\ -0,845 \end{bmatrix} + \begin{pmatrix} 3,676 \\ 3,676 \\ 1,183 \end{pmatrix} \begin{bmatrix} -0,544 \\ -0,544 \\ -0,845 \end{bmatrix} \quad (1)$$

$$b = \frac{-1}{2} ([-15,985] + [-15,985]) = 15,985$$

Dari hasil perhitungan diatas, dapat mencari nilai prediksi SVM berdasarkan persamaan (2) yaitu sebagai berikut:

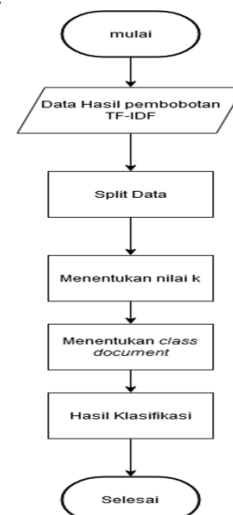
$$= \text{sign} \left( ((-19,4663) + (-0,544) + (-0,544) + (-0,845) + 15,985) \right) \quad (2)$$

$$= \text{sign} (-5,414)$$

$$= -1$$

Hasil prediksi SVM pada teks 7 (D7) adalah -1 yaitu “Negatif”.

i. Algoritma K-Nearest Neighbors (K-NN) proses pengklasifikasian. Flowchart untuk algoritma K-Nearest Neighbors dapat dilihat pada gambar 6.



Gambar 6 Alur K-NN

teks akan dibagi menjadi 3 kelas yaitu kelas positif(K1), kelas netral (K2), dan kelas negatif(K3). Dalam kasus ini D1-D6 akan digunakan sebagai training dan D7 digunakan



sebagai testing, metode K-NN yang digunakan hanya untuk mencari nilai dari data testing. Masing-masing teks akan memiliki 1 kelas, Hasilnya adalah sebagai berikut:

Tabel 1 KNN

	D1	D2	D3	D4	D5	D6	D7
K Positif	✓	✓					?
K2 Negatif			✓	✓			?
K Netral					✓	✓	?

Proses perhitungan dimulai dengan memilih salah satu teks untuk dilakukan pengujian. Data D7 akan dijadikan data uji, Kemudian data yang sudah di TF-IDF dihitung kesamaan kata antar teks menggunakan Cosine Similarity. Rumus dari cosine similarity adalah sebagai berikut:

$$WD7 \times WD1 = 0.544 \times 0.544 = 0.29601 \quad (3)$$

Hitung kesamaan kata antara teks D7 dengan teks D1, D2, D3, D4, D5, dan D6 Dengan menghitung perkalian skalar dengan rumus  $WD7 * Wdi$  dan panjang setiap teks dengan mengkuadratkan tiap term teks TF-IDF.

Contoh perhitungan  $WD7 * WD1$  term "best".

j. Algoritma Naïve Bayes

Setelah mendapatkan hasil dari TF-IDF, langkah selanjutnya adalah data akan melalui tahap pengklasifikasian menggunakan algoritma Naïve Bayes.

Berikut merupakan contoh kalimat yang akan diklasifikasikan menggunakan algoritma Naïve Bayes yang dapat dilihat pada Tabel 3.11.

Tabel 2 Algoritma Naive Bayes

No	Komentar	Label
1	best ever phone	Positif
2	Nice design	Positif
3	hate curve screens not buy Samsung again	Negatif
4	3 5mm jack	Negatif
5	Nice shots Sofia Bulgaria	Netral

Dapat dilihat pada tabel 3.11 contoh kalimat 1 sampai dengan 6 sudah diberi label (sentimen). Sedangkan pada contoh kalimat 7 belum diberi label (sentimen). Maka dari itu untuk memberi label (sentimen) pada contoh kalimat 7 akan digunakan metode klasifikasi algoritma Naïve Bayes.

Dapat dilihat pada tabel 3.11 contoh kalimat 1 sampai dengan 6 sudah diberi label (sentimen). Sedangkan pada contoh kalimat 7 belum diberi label (sentimen). Maka dari itu untuk memberi label (sentimen) pada contoh kalimat 7 akan digunakan metode klasifikasi algoritma Naïve Bayes. Adapun tahapan-tahapan klasifikasi dalam algoritma Naïve Bayes adalah sebagai berikut.

$$y(\alpha|\beta) * (y(\beta | \alpha)) / y(\beta)$$

Menghitung jumlah kelas/label:

$y(\text{Positif}) = 2/6$  (jumlah kalimat dengan kategori "positif")

$y(\text{Negatif}) = 2/6$  (jumlah kalimat dengan kategori "negatif")

$y(\text{Netral}) = 2/6$  (jumlah kalimat dengan kategori "netral")

Kemudian mengkalkulasi nilai maksimum likelihood dengan menggunakan rumus:

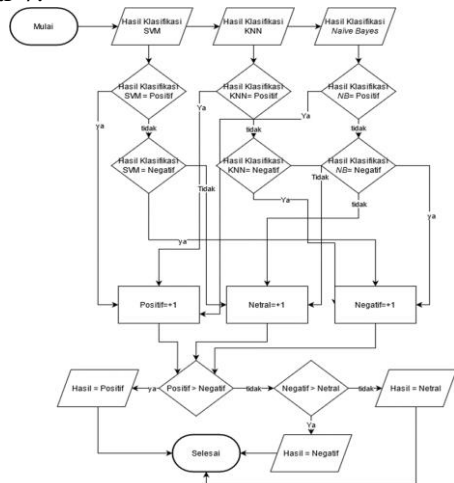
$$\begin{aligned} y(\text{samsung} | \text{pos}) &= (0+1)/(5+20) = 1/25 \\ y(\text{best} | \text{pos}) &= (1+1)/(5+20) = 2/25 \\ y(\text{world} | \text{pos}) &= (0+1)/(5+20) = 1/25 \\ y(\text{samsung} | \text{net}) &= (0+1)/(6+20) = 1/26 \\ y(\text{best} | \text{net}) &= (0+1)/(6+20) = 1/26 \\ y(\text{world} | \text{net}) &= (0+1)/(6+20) = 1/26 \\ y(\text{samsung} | \text{neg}) &= (1+1)/(10+20) = 2/30 \\ y(\text{best} | \text{neg}) &= (0+1)/(10+20) = 1/30 \\ y(\text{world} | \text{neg}) &= (0+1)/(10+20) = 1/30 \\ y(\text{pos} | d7) &= 2/6 * 1/25 * 2/25 * 1/25 = 0,04106 \\ y(\text{net} | d7) &= 2/6 * 1/27 * 1/26 * 1/26 = 1,89653 \times 10^{-10} \\ y(\text{neg} | d7) &= 2/6 * 2/30 * 1/30 * 1/30 = 2,46914 \times 10^{-5} \\ y(\text{pos} | d7) &> y(\text{neg} | d7) \text{ dan } y(\text{net} | d7) \end{aligned}$$

Dari hal di atas, nilai probabilitas tertinggi ada pada variabel "Positif", sehingga dapat

disimpulkan bahwa kalimat tersebut termasuk ke dalam kategori komentar netral. Maka pada contoh kalimat 7 diberi label (sentimen) “Positif”.

k. Algoritma Majority vote

Setelah mendapatkan hasil dari ketiga algoritma langkah selanjutnya adalah memasukkan hasil dari Ketiga algoritma tersebut ke dalam Majority Vote. Flowchart untuk algoritma Majority vote dapat dilihat pada gambar 7.



Gambar 7 Majority Vote

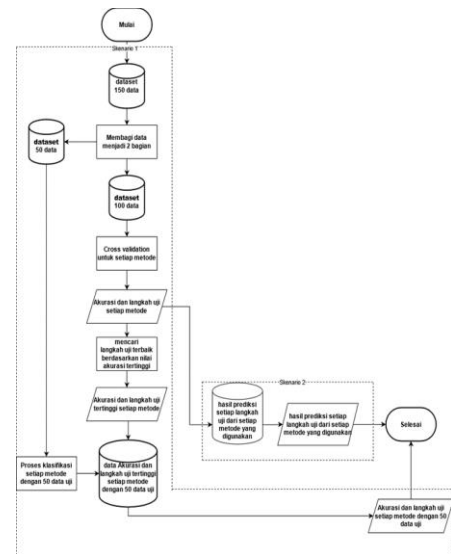
Pada proses Majority vote hasil dari klasifikasi SVM, K-NN dan Naïve Bayes dikumpulkan dan diambil suara yang mayoritas paling banyak.

Tabel 3 Penggunaan Majority Vote

SVM	K-NN	Naïve Bayes	Majority Vote
Negatif	Positif	Positif	Positif

2.5. Skenario pengujian

Skenario pengujian adalah skenario yang akan dilakukan untuk menjawab rumusan masalah pada bab 1 dengan menggunakan 150 data. Pada penelitian ini terdapat 2 skenario yaitu:



Gambar 8 Skenario pengujian

Skenario perbandingan akurasi yang dihasilkan Majority vote dibandingkan metode SVM, KNN dan Naïve Bayes.

Pada skenario ini akurasi Majority vote, SVM, KNN dan Naïve Bayes yang dibandingkan berasal dari akurasi dari langkah uji terbaik masing - masing metode yang didapatkan melalui metode cross validation, sebelum menggunakan metode cross validation untuk mencari Langkah uji terbaik dan 50 data digunakan sebagai data uji, tujuan memisahkan data adalah untuk mengetahui apakah langkah uji terbaik yang telah didapatkan melalui proses cross validation tetap mendapatkan akurasi yang tetap mendapatkan baik atau tidak.

Skenario mencari metode yang terpilih dalam metode Majority vote.

Pada skenario ini pencarian metode didapatkan dari menghitung metode yang terpilih oleh Majority vote dari seluruh pengujian skenario 1.

3. HASIL DAN PEMBAHASAN

Data yang didapatkan sebanyak 240 yang terdiri dari komentar 50 komentar positif, 68 komentar negatif, 122 komentar netral, yang dikembangkan sehingga didapatkan data sebanyak 150 data yang terdiri dari komentar 50 komentar positif, 50 komentar negatif, 50

komentar netral. 150 data tersebut dilakukan berbagai proses seperti *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*. Selanjutnya proses ekstraksi fitur menggunakan TFIDF agar dapat dihitung menggunakan metode, setelah proses tersebut dibagi menjadi 2 bagian 1. Bagian berjumlah 100 data, bagian 2 berjumlah 50 data, 100 akan dilakukan pembagian data atau partisi data dengan menggunakan K Fold Cross Validation untuk mendapatkan Langkah uji terbaik. Setelah dilakukan pembagian data maka masuk kedalam tahap klasifikasi masing-masing metode ditampilkan Gambar 4

Langkah uji	Support Vector Machine	K-Nearest Neighbors	Naive Bayes	Majority Vote
2-fold langkah uji 1	0.24	0.54	0.44	0.42
2-fold langkah uji 2	0.36	0.32	0.4	0.34
4-fold langkah uji 1	0.36	0.56	0.48	0.52
4-fold langkah uji 2	0.16	0.36	0.28	0.24
4-fold langkah uji 3	0.4	0.28	0.36	0.36
4-fold langkah uji 4	0.28	0.28	0.38	0.28
5-fold langkah uji 1	0.5	0.5	0.45	0.45
5-fold langkah uji 2	0.45	0.6	0.6	0.55
5-fold langkah uji 3	0.4	0.6	0.55	0.65
5-fold langkah uji 4	0.45	0.35	0.3	0.3
5-fold langkah uji 5	0.3	0.2	0.4	0.3
10-fold langkah uji 1	0.5	0.3	0.5	0.5
10-fold langkah uji 2	0.3	0.3	0.5	0.4
10-fold langkah uji 3	0.5	0.7	0.6	0.6
10-fold langkah uji 4	0.5	0.8	0.5	0.8
10-fold langkah uji 5	0.3	0.2	0.4	0.2
10-fold langkah uji 6	0.5	0.7	0.4	0.5
10-fold langkah uji 7	0.6	0.6	0.4	0.6
10-fold langkah uji 8	0.1	0	0.3	0

Gambar 9 akurasi seluruh langkah uji

50 data digunakan untuk menguji Langkah uji terbaik yang didapatkan masing-masing metode. Setelah dilakukan perhitungan klasifikasi masing-masing metode berdasarkan Langkah uji terbaik yang diuji dengan 50 data uji ditampilkan Gambar 5

Metode	Akurasi
Support Vector Machine 10 fold langkah uji 7	0.36
K-Nearest Neighbors 10 fold langkah uji 4	0.48
Naive Bayes 5 fold langkah uji 2	0.52
Majority vote 10 fold langkah uji 4	0.48

Gambar 10 akurasi dari Langkah 50 data uji

#### 4. KESIMPULAN

Kesimpulan yang dapat penulis berikan selama melakukan penelitian dan pembuatan sistem mengenai *sentiment analysis* komentar YouTube samsung s20 menggunakan metode Majority vote yaitu:

a) Tingkat akurasi Majority vote berdasarkan langkah uji 10-fold langkah uji 4 menggunakan 100 data mendapatkan nilai akurasi 0.8 dan berdasarkan langkah uji 10-fold langkah uji 4 dengan 50 data uji mendapatkan nilai akurasi 0.48.

b) Tingkat akurasi Majority vote lebih baik dari Support Vector Machine, tetapi tidak lebih baik dari Naive Bayes dan memiliki tingkat akurasi yang sama dengan K-Nearest Neighbors.

c) Metode yang paling sering terpilih dalam metode Majority vote pada semua

langkah uji yang telah dilakukan adalah Metode K-Nearest Neighbors yaitu terpilih sebanyak 319 kali.

#### 5. REFERENSI

Alif, F. Z. (2020). Ekstraksi Fitur untuk Pemilihan Topik Spesifik Review Film dalam Menghasilkan Aspect-Based Sentiment Analysis.pdf. Institusi USU, Universitas Sumatera Utara.

Balya. (2019). Analisis Sentimen Pengguna Youtube di Indonesia pada Review Smartphone Menggunakan Naive Bayes.

Dedi Ary Prasetya, I. N. (2012). Deteksi wajah metode viola jones pada opencv menggunakan pemrograman python. Simposium Nasional RAPI XI FT UMS, 18–23.



- Diaz, M., & Rangkuti, E. (2020). Analisis topik komentar video beberapa akun youtube e-commerce indonesia menggunakan metode latent dirichlet allocation.
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, VI(1), 64–69. <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626>.
- Harismawan, A. F., Kharisma, A. P., & Afirianto, T. (2018). Analisis Perbandingan Performa Web Service Menggunakan Bahasa Pemrograman Python , PHP , dan Perl pada Client Berbasis Android. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 2(January), 237–245. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/781>
- Hustinawaty, Dwiputra, R. A. A., & Rumambi, T. (2019). Public Sentiment Analysis of Pasar Lama Tangerang Using K-Nearest Neighbor Method and Programming Language R. *Jurnal Ilmiah Informatika Komputer*, 24(2), 129–133. <https://doi.org/10.35760/ik.2019.v24i2.2367>
- Julia, W. (2018). Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode Random Forest Dengan Teknik Sampling Kelas Imbalanced. *Высшей Нервной Деятельности*, 2, 227–249.
- KBBI. (2016). Kamus Besar Bahasa Indonesia (KBBI ). In Kementerian Pendidikan dan Budaya.
- Kurniawan, I., & Susanto, A. (2019). Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019. *Eksplora Informatika*, 9(1), 1–10. <https://doi.org/10.30864/eksplora.v9i1.237>
- Loper, E., & Bird, S. (2002). NLTK: The Natural Language Toolkit. July, 69–72. <https://doi.org/10.3115/1225403.1225421>
- Maisarah, M. A. (2020). Sistem Analisis Sentimen pada Fanpage Facebook Kandidat Presiden 2019-2024.
- Næs, T., Lengard, V., Bølling Johansen, S., & Hersleth, M. (2010). Alternative methods for combining design variables and consumer preference with information about attitudes and demographics in conjoint analysis. *Food Quality and Preference*, 21(4), 368–378. <https://doi.org/10.1016/j.foodqual.2009.09.004>
- 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 (J-PTIHK) Universitas Brawijaya*, 1(12), 1750–1757. <https://doi.org/10.1074/jbc.M209498200>
- Onantya, I. D., Indriati, & Adikara, P. P. (2019). Analisis Sentimen Pada Ulasan Aplikasi BCA Mobile Menggunakan BM25 Dan Improved K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2575–2580.
- Rintyarna, B. S. (2016). Sentiment Analysis pada Data Twitter dengan Pendekatan Naïve Bayes Multinomial. 1–6.
- Rosdiansyah, D. (2014). Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor dan Pendekatan Lexicone. *Tugas Akhir Jurusan Teknik Informatika*, 1–15.
- Shi, H. X., & Li, X. J. (2011). A sentiment analysis model for hotel reviews based on supervised learning. *Proceedings - International Conference on Machine Learning and Cybernetics*, 3, 950–954. <https://doi.org/10.1109/ICMLC.2011.6016866>
- Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi*, 8(1), 958–965.

Tanesab, F. I., Sembiring, I., & Purnomo, H. D. (2017). Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine. *International Journal of Computer Science and Software Engineering (IJCSSE)*, 6(8), 180–185.  
<http://ijcsse.org/published/volume6/issue8/p2-V6I8.pdf>

Wibowo, A. P., & Jumiati, E. (2018). Sentiment Analysis Masyarakat Pekalongan Terhadap Pembangunan Jalan Tol Pemalang-Batang Di Media Sosial. *IC-Tech*, XIII(0285), 42–48.

Zamahsyari, & Nurwidyantoro, A. (2017). Sentiment analysis of economic news in Bahasa Indonesia using majority vote classifier. *Proceedings of 2016 International Conference on Data and Software Engineering, ICoDSE 2016*.  
<https://doi.org/10.1109/ICODSE.2016.7936123>

Zulfikar, W. B., & Lukman, N. (2016). Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata. *Jurnal Online Informatika*, 1(2), 82–86.  
<https://doi.org/10.15575/join.v1i2.33>