

# Analisis Sentimen Siswa Terhadap Pelajaran Informatika di SMPK St. Yohanes Kalembo Lona Dengan Metode Naive Bayes Classifier

Sopiana Lende<sup>1\*</sup>, Yulius Nahak Tetik<sup>2</sup>, Mitra Permata Ayu<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, STIMIKO Stella Maris Sumba, Tambolaka, Indonesia

Email: <sup>1</sup> [sopianalende@gmail.com](mailto:sopianalende@gmail.com), <sup>2</sup> [yuliussteti@gmail.com](mailto:yuliussteti@gmail.com), <sup>3</sup> [mitrapermataayu@gmail.com](mailto:mitrapermataayu@gmail.com)

Email Penulis Korespondensi: <sup>1</sup> [wsurasih@email.com](mailto:wsurasih@email.com) @email.com

**Abstrak**—Semakin pesatnya perkembangan ilmu pengetahuan dan teknologi saat ini, proses pembelajaran yang berkaitan dengan teknologi informasi dan komunikasi tidak hanya dipelajari di Perguruan Tinggi saja tetapi sudah di ajarkan kepada peserta didik di Sekolah Lanjutan Tingkat Pertama (SLTP) maupun Lanjutan Tingkat Atas (SLTA), dengan suatu harapan bahwa pada mata pelajaran Teknologi Informasi dan Komunikasi (TIK) bisa memberikan kontribusi untuk terwujudnya daya nalar yang kritis dan kreatif. Sejak diterapkannya pada Kurikulum Tingkat Satuan Pendidikan (KTSP), mata pelajaran TIK menjadi salah satu pelajaran yang nilainya turut serta dijadikan sebagai kriteria kenaikan kelas siswa. Dalam penelitian tentang sentimen siswa terhadap pelajaran informatika ini, akan dilakukan pengelompokan pendapat-pendapat positif, negatif dan netral menggunakan metode Naive Bayes Classifier. proses penelitian yang dilakukan oleh penulis pada penelitian ini dilakukan dengan pengumpulan data, pelabelan pada data set, pre-processing dataset yang terdiri atas proses cleaning, case folding, tokenizing, stopword removal, stemming kemudian dilanjutkan dengan proses perhitungan TF-IDF. Tujuan diterapkannya Naive Bayes Classifier pada penelitian ini agar dapat dikelompokkan sentimen yang kemudian dilanjutkan dengan validasi dan evaluasi sistem dengan metode K-Fold Cross Validation serta Confusion Matrix. data set yang digunakan merupakan opini dari siswa di SMPK St. Yohanes Kalembo Lona yang dikumpulkan menggunakan kuisioner yang disebar secara acak kepada siswa kelas VII, VIII dan Kelas IX sebanyak 5 pertanyaan terkait pelajaran informatika kepada masing-masing responden sebanyak 40 siswa/i yang dipilih secara acak. Tingkat akurasi yang dihasilkan dengan metode Naive Bayes Classifier berdasarkan hasil pengujian menggunakan teknik splitting dataset diadaptasi hasil dengan persentasi yang ditetapkan yaitu sebesar 80% berbanding 20% diperoleh hasil sebesar 85% sedangkan pengujian dengan menggunakan sub dataset diperoleh hasil masing-masing 84%, 83%, 80%, 82% dan 84 % untuk ke 5 sub dataset sedangkan untuk tingkat akurasi dengan teknik K-fold Cross Validation diperoleh hasil untuk masing-masing dataset yaitu 84%, 83%, 82%, 78%, dan 80%

**Kata Kunci:** Classifier, Naive Bayes, informatika, dataset, sentimen

**Abstract**— With the increasingly rapid development of science and technology today, the learning process related to information and communication technology is not only studied in universities but has been taught to students in junior secondary schools (SLTP) and senior secondary schools (SLTA), with There is a hope that Information and Communication Technology (ICT) subjects can contribute to the realization of critical and creative reasoning power. Since its implementation in the Education Unit Level Curriculum (KTSP), ICT subjects have become one of the subjects whose grades are also used as criteria for students' grade promotion. In this research on student sentiment toward informatics lessons, positive, negative, and neutral opinions will be grouped using the Naive Bayes Classifier method. The research process carried out by the author in this study was carried out by collecting data, labeling the data set, and pre-processing the dataset which consisted of cleaning, case folding, tokenizing, stopword removal, stemming, and then continued with the TF-IDF calculation process. The aim of applying the Naive Bayes Classifier in this research is to group sentiments which are then followed by system validation and evaluation using the K-Fold Cross Validation and Confusion Matrix methods. The data set used is the opinion of students at SMPK St. Yohanes Kalembo Lona collected using a questionnaire that was distributed randomly to students in Class VII, VIII, and Class IX with 5 questions related to informatics lessons to each of the 40 students selected at random. The level of accuracy produced by the Naive Bayes Classifier method is based on test results using the splitting dataset technique, the results are adapted to the specified percentage, namely 80% compared to 20%, the results obtained are 85%, while testing using sub-datasets produces results of 84%, 83% respectively. , 80%, 82%, and 84% for the 5 sub-datasets while for the level of accuracy using the K-fold Cross Validation technique the results obtained for each dataset are 84%, 83%, 82%, 78%, and 80%.

**Keywords:** Classifier, Naive Bayes, informatika, dataset, sentimen

## 1. PENDAHULUAN

Seiring dengan semakin pesatnya perkembangan ilmu pengetahuan dan teknologi saat ini, proses pembelajaran yang berkaitan dengan teknologi informasi dan komunikasi tidak hanya dipelajari di Perguruan Tinggi saja tetapi sudah di ajarkan kepada peserta didik di Sekolah Lanjutan Tingkat Pertama (SLTP) maupun Lanjutan Tingkat Atas (SLTA), dengan suatu harapan bahwa pada mata pelajaran Teknologi Informasi dan Komunikasi (TIK) bisa memberikan kontribusi untuk terwujudnya daya nalar yang kritis dan kreatif. Sejak diterapkannya pada Kurikulum Tingkat Satuan

Pendidikan (KTSP), mata pelajaran TIK menjadi salah satu pelajaran yang nilainya turut serta dijadikan sebagai kriteria kenaikan kelas siswa.

Mata pelajaran Teknologi Informasi dan Komunikasi (TIK) sejak diberlakukannya kurikulum 2013 tidak lagi dimasukkan sebagai mata pelajaran yang diajarkan di kelas secara tatap muka berdasarkan Permendikbud nomor 45 pada tahun 2015 tentang perubahan atas peraturan menteri pendidikan dan kebudayaan Republik Indonesia tahun 2014 nomor 68 tentang Peran Guru Teknologi Informasi dan Komunikasi dan Guru Keterampilan Komputer dan Pengelolaan Informasi dalam Implementasi Kurikulum 2013 [1]. Pada tahun 2018 melalui Permendikbud Republik Indonesia Nomor 35 Tahun 2018 tentang Perubahan Atas Peraturan Menteri Pendidikan dan Kebudayaan Nomor 58 Tahun 2014 tentang Kurikulum 2013 Sekolah Menengah Pertama/Madrasah Tsanawiyah kembali memberlakukan mata pelajaran TIK namun berubah nama menjadi Informatika dan Pelaksanaan pembelajaran Informatika sebagai mata pelajaran pilihan dilaksanakan mulai tahun ajaran 2019/2020 sesuai dengan kesiapan sekolah [2]. Menurut Wisnubhadra, dkk (2021) dalam Buku Panduan Guru Informatika untuk SMP kelas VII berpendapat bahwa, untuk mengikuti perkembangan teknologi, sistem pendidikan Indonesia perlu memasukkan informatika sebagai mata pelajaran dasar yang mengajarkan tentang pengetahuan dan kompetensi guna membentuk manusia yang cerdas dan mampu bersaing di kawasan regional maupun global [3]. Pengintegrasian TIK pada proses pembelajaran sangat dibutuhkan untuk meningkatkan kemampuan berpikir siswa, mengembangkan keterampilan dalam bidang teknologi informasi dan komunikasi [4]. Menurut Pratasik dan Ahyar pada penelitiannya tentang Pengembangan Media Pembelajaran Pada Mata Pelajaran Informatika MTs pada tahun 2022 mendeskripsikan bahwa untuk membantu keberhasilan pendidikan pada mata pelajaran informatika tidak terlepas dari penggunaan media pembelajaran yang sangat berkaitan erat dengan proses komunikasi. Oleh karena itu, pada penelitiannya, menghasilkan sebuah aplikasi media pembelajaran untuk mata pelajaran informatika [5]. Diberlakukannya kembali mata pelajaran TIK walaupun berubah nama menjadi Informatika tentunya akan memunculkan berbagai tanggapan atau pendapat dari peserta didik terhadap materi pelajaran yang di pelajari di kelas, hal ini karena tidak semua sekolah di tingkat Sekolah Menengah Pertama terutama di daerah yang tergolong daerah afirmasi tidak memiliki fasilitas laboratorium komputer yang lengkap. SMPK St. Yohanes Kalembu Lona merupakan lembaga pendidikan menengah pertama swasta yang didirikan pada Tahun 2022 di Desa Reda Pada, Kecamatan Wewewa Barat, Kabupaten Sumba Barat Daya. Pada pelaksanaan pembelajarannya, SMPK St. Yohanes Kalembu Lona menerapkan kurikulum 2013, dengan demikian mata pelajaran Informatika juga di pelajari di sekolah ini.

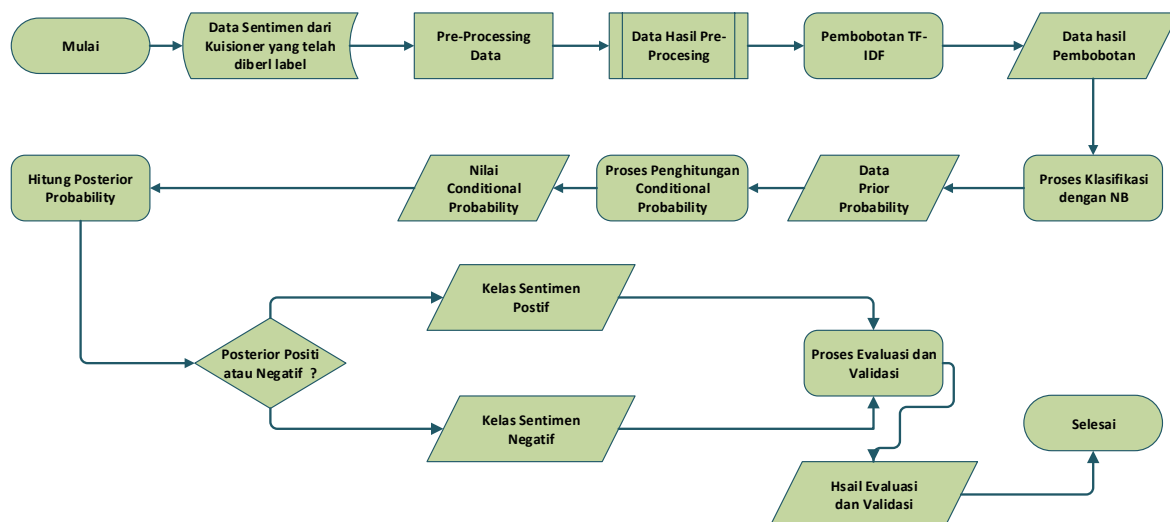
Pada penelitian yang dilakukan oleh Santoso, dkk (2022) tentang analisis sentimen mahasiswa terkait pembelajaran tatap muka, dimana pada penelitian tersebut, proses pengklasifikasian terhadap opini mahasiswa menggunakan metode *Naive Bayes Classifier* dan untuk proses pengujiannya peneliti menerapkan *splitting dataset*. Berdasarkan penerapan metode dan pengujiannya akurasi yang dihasilkan dalam melakukan kalasifikasi opini (sentimen) sangat baik yaitu 84% dengan tingkat akurasi pada setiap sub *dataset* sebesar 83%, 88%, 87%, 89% dan 79% [6]. Penelitian oleh Zusrotun, dkk (2022) yang berjudul sentimen analisis belajar online di twitter menggunakan metode . Pada peneltian yang dilakukan, pengambilan data atau proses scraping dari media sosial twitter kemudian dimasukkan kedalam bentuk file excel untuk proses *scraping* meggunakan alat bantu analisis yaitu *Naive Bayes Classifier* di mana pada proses pengujian menggunakan algoritma *Naive Bayes Classifier* menghasilkan nilai akurasi sebesar 74,08% dan untuk memastikan hasil yang didapatkan peneliti melakukan lagi pengujian dengan *K-Fold Cross Validation* dengan K sebesar 15 dengan nilai akurasinya sebesar 76,39 % [7]. Penelitian yang dilakukan oleh Rahmatullah pada tahun 2021 yang berjudul Analisis sentimen mahasiswa terhadap perkuliahan dalam jaringan menggunakan metode *Naive Bayes Classifier*. Pada penelitian tersebut pendekatan yang digunakan ialah deskriptif kuantitatif dimana pada teknik pengumpulan data disebarkan kepada responden dengan memanfaatkan *Google Form* dengan menrapkan variabel bebas yaitu berupa pendapat atau opini mahaisswa tentang proses pembalejaran secara daring dan variabel terikat berupa label positif dan negatif. Dari hasil pengelompokkan yang dilakukan menggunakan metode *Naive Bayes Classifier* diperoleh hasil dengan akurasi sebanyak 75% dan untuk memastikan bahwa hasil dari penelitiannya peneliti melakukan pengujian dengan *k-fold cross validation* dengan k yaitu 10 dan mendapatkan hasil dengan nilai akurasi sebesar 61% [8].

Analisis Sentimen terhadap pembelajaran daring dengan algoritma *Naive Bayes Classifier* juga dilakukan oleh Sidik, dkk (2022) dimana pada proses Ekstraksi *Dataset* dilakukan setelah tahap *word frequency* untuk pencarian kata, selanjutnya dilakukakan penambahan bobot kata dengan teknik *Term Frequency-Inverse Document Frequency*. Setelah dilakukan tahap prmbobotan, tahap selanjutnya ialah tahapan implementasi algoritma dimana data terlebih dahulu dibagi 80:20 dengan 80% data *training* dan 20% data *validation* kemudian *fitting* dengan algoritma *Naive Bayes Classifier*. Berdasarkan penerapan metode yang digunakan diperoleh hasil yang cukup baik karena sebanyak 62,5% memberikan jawaban yang dapat divisualisasikan sebagai jawaban positif dan 37,5% jawaban negatif dari 40 data kuisisioner dari total 40 data kuisisioner jawaban responden yang dievaluasi didapatkan hasil dengan nilai akurasi sebesar

65% [9]. Penelitian oleh Lopez, dkk (2023) tentang sistem pendidikan atau pembelajaran dimana pada penelitian yang dilakukannya, peneliti mengumpulkan data dari twiter. Berdasarkan data yang berhasil dikumpulkan sebanyak 85% siswa pada negara-negara yang tergabung dalam organisasi ekonomi eropa (OECD) merasa senang dengan sistem pendidikan yang ada saat ini dan berdasarkan hasil survei sebanyak 6% menyatakan bahwa tidak senang hal ini dikarenakan mereka lebih senang menghaiskan waktu diluar sekolah dengan bermain internet [10]. Berdasarkan penjelasan yang telah di jabarkan diatas maka pada penelitian ini penulis akan melakukan analisis sentimen terhadap pelajaran informatika guna mengetahui seberapa besar minat belajar siswa/i terhadap mata pelajaran informatika dengan dengan metode *Naïve Bayes Classiver*.

## 2. METODOLOGI PENELITIAN

Dalam penelitian tentang sentimen siswa terhadap pelajaran informatika ini, akan dilakukan pengelompokan pendapat-pendapat positif, negatif dan netral menggunakan metode *Naïve Bayes Classiver*. proses penelitian yang dilakukan oleh penulis pada penelitian ini dilakukan dengan pengumpulan data, pelabelan pada data set, pre-processing dataset yang terdiri atas proses cleaning, case folding, tokenizing, stopwords removal, stemming kemudian dilanjutkan denagn proses perhitungan TF-IDF. Tujuan diterapkannya *Naïve Bayes Classifier* pada penelitian ini agar dapat dikelompokkan sentimen yang kemudian dilanjutkan dengan validasi dan evaluasi sistem dengan metode *K-Fold Cross Validation* serta *Confusion Matrix*. Tahapan-tahapan yang digunakan oleh penulis pada penelitian ini disajikan pada gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1 Pengumpulan data aset

Pada penelitian tentang analisis sentimen ini, data set yang digunakan merupakan opini dari siswa di SMPK St. Yohanes Kalembu Lona yang dikumpulkan menggunakan kuisioner yang disebar secara acak kepada siswa kelas VII, VIII dan Kelas IX sebanyak 5 pertanyaan terkait pelajaran informatika kepada masing-masing responden sebanyak 40 siswa/i yang dipilih secara acak dan data set yang berhasil dikumpulkan yaitu sebanyak 200 opini yang nantinya akan digunakan untuk proses analisis dan untuk pengklasifikasiannya, data set yang dikumpulkan di pisah sesuai dengan jawaban setiap pertanyaan yang terdapat pada kuisioner.

Kuisiomer Pelajaran Informatika SMPK St. Yohanes Kalembo Lona					
Nama : .....					
Kelas : .....					
NO	Pertanyaan/Pernyataan	Ya	Tidak	Ragu-Ragu	Alasan
1	Apakah anda sangat suka pelajaran Informatika	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ya saya suka karena bisa mengetik di word.
2	Saya tidak mengerti setiap materi pada mata pelajaran Informatika	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	karena ada yang saya mengerti dan tidak dengan perintah bahasa inggris di komputer
3	Saya sering belajar Informatika di luar pelajaran sekolah	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ya saya sering belajar karena kaka saya ada laptop di rumah
4	Apakah pelajaran Informatika adalah mata pelajaran yang membosankan	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Tidak membosankan karena saya suka belajar komputer dan bermain game di laptop
5	Saya ingin meningkatkan pengetahuan saya dalam mengoperasikan komputer	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ya karena saya mau lebih lancar ketik di komputer

Gambar 2. Data Aset Opini siswa pada kuisiomer

## 2.2 Pelabelan

Merupakan teknik yang digunakan untuk menentukan teks sentimen(opini) bermakna atau bersifat positif, negatif maupun netral yang dapat dilakukan secara manual agar dapat menghasilkan nilai presisi yang cukup baik. Suatu opini dapat dilabeli positif apabila sentimen(opini) mengandung tujuan yang sifatnya membangun sedangkan suatu opini dilabeli negatif apabila sentimen(opini) bermakna hinaan atau mencela.

## 2.3 Pre-Processing

Merupakan tahap yang dilakukan untuk mengolah dataset sebelum diproses menggunakan metode yang ditentukan. Data set akan diubah kedalam bentuk yang terstruktur sehingga dapat diproses dengan lebih mudah [11]. Adapun tahapan Proses *Pre-Processing* meliputi:

### 1. Cleaning

Merupakan proses untuk menghilangkan atau menghapus data yang memiliki *noise*, *hashtag*, dan *url*.

### 2. Case folding

Merupakan proses yang dilakukan untuk merubah seluruh bentuk huruf yang terdapat pada *dataset* menjadi huruf kecil

### 3. Tokenizing

Merupakan proses yang dilakukan untuk memotong kata atau kalimat yang bertipe *string* pada input setiap kata.

### 4. Filtering

Merupakan proses yang dilakukan untuk menghilangkan kata yang tidak diperlukan dari hasil *tokenizing*.

### 5. Normalization

Merupakan proses untuk melengkapi kata yang tidak lengkap

### 6. Stemming

Merupakan proses untuk menghapus kata yang memiliki imbuhan sehingga kata tersebut menjadi kata dasar.

## 2.4 Pembobotan (TF-IDF)

Merupakan proses pemberian nilai pada setiap kata (*term*) yang berhasil di ekstrak dan akan dijadikan sebagai inputan ketika proses klasifikasi data dengan *Naïve Bayes Classifier* [6] [12]. Adapun rumus yang akan digunakan pada proses perhitungan ini adalah sebagai berikut:

$$TF = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \dots\dots\dots(1)$$

Keterangan rumus:

TF : *Term Frequency*

d : Dokumen

t : token (*term*)

$f_{t,d}$  : jumlah term (kata) di setiap dokumen

$\sum_{t' \in d} f_{t',d}$  : jumlah dokumen yang memuat t

Selanjutnya rumas atau persamaan matematis untuk menghitung *IDF* ( $idf_t$ ) adalah sebagai berikut:

$$idf_t = \log \frac{N}{df_t} \dots\dots\dots(2)$$

Keterangan rumus :

---

$Idf_t$	: nilai $idf$ dari term ( $t$ )
$N$	: jumlah dokumen yang tersedia
$Dft$	: volume kemunculan kata yang terdapat pada dokumen

---

Dan rumus atau persamaan matematis untuk menghitung TF-IDF sebagai berikut:

$$tf - idf_{t,d} = tf_{td} \times idf_t \dots\dots\dots(3)$$

Dengan keterangan rumus:

$idf_{t,d}$	: bobot TF – IDF yang berasal dari kata $t$ dan terdapat pada dokumen $d$
$tf_{td}$	: frekuensi intensitas kemunculan kata $t$ pada dokumen $d$
$idf_t$	: nilai dari IDF ( $idf$ )

## 2.5 Naïve Bayes Classifier

*Naïve Bayes Classifier* merupakan metode pengelompokan yang sederhana dimana pada proses pengelompokan dilakukan berdasarkan perhitungan dari peluang distribusi kata yang terdapat pada dokumen [13]. Adapun rumus atau persamaan matematis dari *Naïve Bayes Classifier* sebagai berikut :

$$P(P|X) = \frac{P(P|X)P(H)}{P(X)} \dots\dots\dots(4)$$

Keterangan rumus :

$X$  = Sampel data yang memiliki kelas yang belum diketahui

$H$  : Hipotesa bahwa  $x$  merupakan data kelas

$P(H)$  : Peluang hipotesa  $H$  (*prior probability*)

$P(X)$  : Peluang data yang diamati (*conditional probability*)

$P(X|H)$  : Peluang data sampel  $X$  jika dianggap hipotesa tersebut benar

$P(H|X)$  : Peluang hipotesis  $H$  berdasarkan kondisi  $X$  (*posteriori probability*)

Dengan demikian dapat dijabarkan persamaan matematis dari *Naïve Bayes Classifier*, sebagai berikut :

$$1. \text{ Prior probability : } P(X) = \frac{d_x}{d} \dots\dots\dots(5)$$

dengan keterangan rumus sebagai berikut :

$X$  : Hipotesa suatu kategori atau kelas

$P(X)$  : *Prior probability*

$D$  : Jumlah seluruh dokumen

$d_x$  : Jumlah seluruh data dengan kategori atau kelas  $X$

$$2. \text{ Conditional probability : } P(\text{Term } H_n|X) = \frac{(H_n|X)+1}{(c)+|V|} \dots\dots\dots(6)$$

Keterangan rumus :

$P(\text{Term } H_n|X)$  : *Conditional probability* kata ke- $n$  dengan kelas  $c$

$(H_n|X)$  : Frekuensi kemunculan kata yang terdapat pada kelas  $c$

$(c)$  : Bobot (TF-IDF) dari kata yang terdapat pada kelas  $c$

$|V|$  : Nilai TF dari seluruh kata yang terdapat pada dokumen

Pada rumus diatas, terdapat angka 1 yang merupakan angka pembilang yang dapat digunakan untuk mengantisipasi jika tidak ditemukan token atau term pada data training sehingga tidak bernilai 0 pada pembilang.

3. *Posterior Probability* :  $P(\text{Term } H|X) = P(X) \times P(\text{Term } H_1|X) \times P(\text{Term } H_2|X) \times \dots \times P(\text{Term } H_n|X)$

Keterangan :

$P(\text{Term } H|X)$  : *Posterior probability*

$P(X)$  : *Prior probability*

$P(\text{Term } H_n|X)$  : *Conditional probability ke-n dengan kategori atau kelas X*

## 2.6 K-fold Cros Validation

*K-fold Cross validation* adalah suatu metode validasi yang pada prinsipnya bahwa proses validasi dilakukan akan dilakukan secara berulang sebanyak nilai K berdasarkan catatan data yang dimasukkan. Dari data yang diinput tersebut akan diacak sehingga diperoleh data *testing* dan data *training* dimana pada kedua data tersebut memiliki jumlah data yang sama. Berikut ini ilustrasi *K-fold Cross validation* yang tunjukkan pada gambar dibawah ini.

Pengujian Ke	Data Set			
1				
2				
3				
4				
...				
Keterangan Warna :				

Gambar 3. Ilustrasi Metode K-Fold Cross Validation

## 2.7 Confusion Matrix

*Confusion Matrix* adalah sebuah metode yang dapat digunakan untuk mengevaluasi kinerja dan menentukan seberapa mungkin tingkat akurasi dari metode klasifikasi yang dipakai dengan hasil prediksinya berupa *true positive*, *false negative* dan *true negative* dengan rumus:

1. *Precession*

$$\text{Precession} = \frac{TP}{TP+FP} \dots\dots\dots(7)$$

2. *Recall*

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots(8)$$

3. *F-Measure*

$$\text{F-Measure} = \frac{2(\text{precession}+\text{recall})}{TP+FP} \dots\dots\dots(9)$$

4. *Accuracy*

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(10)$$

# 3. HASIL DAN PEMBAHASAN

## 3.1 Pelabelan

Proses pelabelan pada penelitian ini dilakukan secara manual dan diperoleh jumlah data set dengan data keseluruhan postif sebanyak 168 dan dataset negatif sebanyak 32 dengan jumlah sentimen seperti yang ditunjukkan pada tabel berikut ini.

Tabel 1. Hasil pemberian label

Sentimen	Positif	Negatif
----------	---------	---------



Sub Data set		
1	160	40
2	175	25
3	169	31
4	180	20
5	165	35

### 3.2 Pre-Processing

Pada tahapan ini meliputi cleanig, case folding, tokenizing, stopwords removal dan stemming hingga evaluasi terhadap sistem dengan alat bantu *python* dan *text editor* visual studio code. Adapun hasil dari proses pre-processing sentimen atau opini terhadap teks alis “*Tidak Membosankan, karena saya suka belajar komputer dan bermain game di laptop*” dapat dilihat pada tabel berikut:

Tabel 2. Tahapan Hasil *Pre-Processing*

Tahap Pre-Processing	Hasil Pre-Processing
<i>Cleaning</i>	Tidak Membosankan karena saya suka belajar komputer dan bermain game di laptop
<i>Case Folding</i>	tidak membosankan karena saya suka belajar komputer dan bermain game di laptop
<i>Tokenizing</i>	‘tidak’, ‘membosankan’, ‘karena’, ‘saya’, ‘suka’, ‘belajar’, ‘komputer’, ‘dan’, ‘bermain’, ‘game’, ‘di’, ‘laptop’
<i>Stopword Removal</i>	‘saya’, ‘suka’, ‘belajar’, ‘komputer’, ‘dan’, ‘bermain’, ‘game’, ‘di’, ‘laptop’
<i>Stemming</i>	‘saya’, ‘suka’, ‘belajar’, ‘komputer’, ‘dan’, ‘main’, ‘game’, ‘di’, ‘laptop’

### 3.3 Pembobotan TF-IDF

Pada tahapan ini, digunakan untuk mengetahui bobot dari kata yang terdapat pada suatu sentimen menggunakan rumus yang sudah ditentukan yaitu rumus atau persamaan matematis pada no (1) untuk menghitung nilai TF, (2) untuk menghitung nilai IDF dan persamaan matematis no (3) untuk menghitung nilai bobot *term* (kata). ‘saya’, ‘suka’, ‘belajar’, ‘komputer’, ‘dan’, ‘main’, ‘game’, ‘di’, ‘laptop’

Tabel 3. TF-IDF Sentimen

Term	TF			DF	IDF	TF-IDF		
	D1	D2	D3			D1	D2	D3
saya	1	0	0	1	0,453	0,453	0	0
suka	1	0	0	1	0,453	0,453	0	0
belajar	1	0	0	1	0,453	0,453	0	0
komputer	1	0	0	1	0,453	0,453	0	0
dan	1	0	0	1	0,453	0,453	0	0
main	1	0	0	1	0,453	0,453	0	0
game	1	0	0	1	0,453	0,453	0	0
di	1	0	0	1	0,453	0,453	0	0
laptop	1	0	0	1	0,453	0,453	0	0

### 3.4 Pengklasifikasian menggunakan *Naïve Bayes Classifier*

Pada penelitian ini, proses klsifikasi terhadap opini atau sentimen menggunakan *Naïve Bayes Classifier* Dan proses pengujiannya dilakukan dengan cara splitting data set sebesar 80% sebagai data training dan 20% sebgai data testing berdasarkan jumlah dari masing-masing sub dataset yang digunakan dan berisikan 40 sentimen (opini) yang didapatkan melalui kuisioner yang disebarakan secara acak kepada siswa di SMPK St. Yohanes Kalembo Lona Kelas VII, VIII,

dan IX. Setelah proses pengujian, tahap selanjutnya ialah validasi menggunakan K-fold Cross Validation dengan nilai K=10. Adapun hasil pengujian dan validasi yang dilakukan dapat dilihat pada tabel dibawah ini.

**Tabel 4.** Hasil pengujian dan validasi

Dataset	Sub Dataset	Prediksi	Split		K-Fold Cross Validation	
			Positif	Negatif	Positif	Negatif
Opini Keseluruhan (200)		Positif	25	78	3	33
		Negatif	6	112	34	73
	1	Positif	26	16	20	14
		Negatif	4	117	1	64
	2	Positif	1	145	1	10
		Negatif	0	95	0	74
	3	Positif	4	20	1	9
		Negatif	2	170	1	85
	4	Positif	1	30	2	13
		Negatif	1	165	0	87
	5	Positif	129	6	70	1
		Negatif	14	15	10	4

### 3.5 Evaluasi Sistem

Evaluasi terhadap sistem dilakukan guna mengetahui tingkat keakuratan dan kemampuan pada sistem yang dikembangkan dalam melakukan pengelompokan atau klasifikasi. Pada proses evaluasi ini dilakukan menggunakan confusion matrix agar lebih mudah untuk mengetahui nilai dari accuracy, precision, recall dan f-measure yang didapatkan dari pengujian dan validasi yang telah dilakukan. Berikut ini, tabel nilai yang diperoleh pada proses evaluasi sistem seperti yang ditunjukkan pada tabel 5.

**Tabel 5.** Hasil Evaluasi dengan *confusion matrix*

Dataset Split	Dataset	Accuracy	Precision	Recall	f-Measure
Split 80% -20%	Total sentimen (200)	85%	82%	53%	53%
	Sub Data set 1 (40)	84%	86%	70%	79%
	Sub Data set 2 (40)	83%	92%	45%	50%
	Sub Data set 3 (40)	80%	78%	55%	59%
	Sub Data set 4 (40)	82%	83%	65%	65%
	Sub Data set 5 (40)	84%	83%	70%	69%
K-Fold Cross Validation	Total sentimen (200)	84%	60%	45%	78%
	Sub Data set 1 (40)	84%	55%	90%	57%
	Sub Data set 2 (40)	83%	45%	99%	45%
	Sub Data set 3 (40)	82%	25%	45%	56%
	Sub Data set 4 (40)	78%	20%	98%	59%
	Sub Data set 5 (40)	80%	95%	78%	64%

## 4. KESIMPULAN

Adapaun kesimpulan yang dapat dijabarkan dari hasil penelitian yang diperoleh sebagai berikut:

1. Dataset yang terdapat pada penelitian ini sebanyak 200 sentimen yang dibagi menjadi 5 sub dataset dari jawaban responden pada kuisioner yang dibagikan secara acak kepada 40 orang siswa dikelas VII, VIII, dan kelas IX.
2. Pada proses pemberian label yang sudah dilakukan, terdapat 168 opini yang dinyatakan positif dan 32 opini yang dinyatakan negatif dari 200 sehingga dapat disimpulkan bahwa siswa/i di SMPK St. Yohanes Kalembo Lona senang dengan mata pelajaran Informatika dikelas.
3. Dengan menerapkan metode Naïve Bayes Classifier dalam penelitian ini, didapatkan hasil yang sesuai dengan yang diharapkan
4. Tingkat akurasi yang dihasilkan dengan metode Naïve Bayes Classifier berdasarkan hasil pengujian menggunakan teknik splitting dataset diadaptkan hasil dengan persentasi yang ditetapkan yaitu sebesar 80% berbanding 20% diperoleh hasil sebesar 85% sedangkan pengujian dengan menggunakan sub dataset diperoleh hasil masing-masing



---

84%, 83%, 80%, 82% dan 84 % untuk ke 5 sub dataset sedangkan untuk tingkat akurasi dengan teknik K-fold Cross Validation diperoleh hasil untuk masing-masing data set yaitu 84%, 83%, 82%, 78%, dan 80%

---

## UCAPAN TERIMA KASIH

Dengan selesainya penulisan artikel ini, saya ucapkan limpah terimakasih kepada Bapak. Yulius Nahak Tetik, M.Kom selaku dosen pembimbing 1 dan Ibu Mitra Permata Ayu, M.Pd selaku dosen pembimbing 2 yang telah dengan sabar membimbing saya dalam menyelesaikan penulisan karya ilmiah ini, serta tak lupa juga saya mengucapkan limpah terimakasih kepada kedua orang tua, kaaka, dan adik yang telah mendukung saya dalam menyelesaikan penulisan karya ilmiah ini.

## REFERENCES

- [1] B. P. d. P. P. K. d. P. Kemenntrian Pendidikan dan Kebudayaan, "<https://www.kemdikbud.go.id/>," 2019.[Online].Available:[https://repositori.kemdikbud.go.id/18254/1/20190618\\_PedomanImplementasiIF\\_v22.pdf](https://repositori.kemdikbud.go.id/18254/1/20190618_PedomanImplementasiIF_v22.pdf). [Accessed 6 Oktober 2023].
- [2] K.R.Indonesia,"<https://jdih.kemdikbud.go.id/>,"2019.[Online].Available:<https://jdih.kemdikbud.go.id/sjdih/siperpu/dokumen/salinan/Permendikbud%20Nomor%2036%20Tahun%202018.pdf>. [Accessed 6 Oktober 2023].
- [3] I. Wisnubhadra, C. M. Wijanto, V. Natali, W. Wahyono, S. Mulyati, A. Wardhani, S. Sutardi, H. Pratiwi, B. Saputra, K. Astiani and S. Sumiati, Buku Panduan Guru Informatika untuk SMP Kelas VII, Jakarta Pusat: Pusat Kurikulum dan Perbukuan Badan Penelitian dan Pengembangan dan Perbukuan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2021.
- [4] H.Nissa,"<https://pusdatin.kemdikbud.go.id/>,"23juni2021.[Online].Available:<https://pusdatin.kemdikbud.go.id/pembelajaran-berbasis-tik-pembatik-dalam-meningkatkan-level-kompetensi-tik-guru-di-indonesia/>. [Accessed 6 Oktober 2023].
- [5] S. Pratasik and B. M. Ahyar, "Pengembangan Media Pembelajaran pada Mata Pelajaran Informatika MTs," *Jurnal Edutik*, vol. 2, no. 3, pp. 359-373, 2022.
- [6] H. Santoso, A. Armansyah and D. Desliani, "Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naï ve Bayes Classifier," *Techno.COM*, vol. 21, no. 3, pp. 644-654, 2022.
- [7] O. P. Zusrotun, A. Catur Murti and R. Fiati, "Sentimem Analsis Belajar online di twitter menggunakan Naive Bayes," *JANAPATI*, vol. 11, no. 3, pp. 310-320, 2022.
- [8] B. A. Rahmatullah, "Analisis Sentimen Mahasiswa Terhadap Perkuliahan Dalam Jaringan Menggunakan Metode Naï ve Bayes Classifier," Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, 2021.
- [9] F. Sidik, I. Suhada, A. H. Anwar and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring dengan Algoritma Naï ve Bayes Classifier," *Jurnal Linguistik Komputasional*, vol. 5, no. 1, pp. 34-43, 2022.
- [10] M. L. M. Lopez, J. S. Ceres and A. M. Columbrans, "Analysing the sentiments about the education system trough Twitter," *Educational and Information Technologies*, p. 10965–10994, 2023.
- [11] M. K. A. Muslim and Y. Yamasari, "Klasifikasi Opini Pengguna Twitter Terhadap Sekolah Daring dengan Metode Naï ve Bayes dan Support Vector Machine," *Journal of Informatics and Computer Science*, vol. 3, no. 2, pp. 171-179, 2021.
- [12] M. A. Khadafi, . P. K. Kartika and F. Febrinita, "Penerapan Metode Naï ve Bayes Classifier dan Lexicon Based untuk Analisis Sentimen Cyberbullying pada BPJS," *JATI*, vol. 6, no. 2, pp. 725-733, 2022.
- [13] E. Indrayuni, "Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film," *Jurnal PILAR Nusa Mandiri*, vol. 14, no. 2, pp. 175-180, 2028.