

**ANALISIS SENTIMEN REVIEW APLIKASI BERITA *ONLINE*  
PADA *GOOGLE PLAY* MENGGUNAKAN METODE  
ALGORITMA *NAIVE BAYES CLASSIFIER*  
DAN *SUPPORT VECTOR MACHINES***

**THESIS**

**Oleh:  
ULFA KUSNIA  
NIM. 19841003**



**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK  
IBRAHIM  
MALANG  
2022**

***SENTIMENT ANALYSIS OF NEWS APP REVIEWS ONLINE ON GOOGLE PLAY  
USING NAIVE BAYES CLASSIFIER ALGORITHM  
AND SUPPORT VECTOR MACHINES METHODS***

***THESIS***

***By:***  
**ULFA KUSNIA**  
**NIM. 19841003**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2022**

**ANALISIS SENTIMEN REVIEW APLIKASI BERITA *ONLINE* PADA *GOOGLE PLAY*  
MENGUNAKAN METODE ALGORITMA *NAIVE BAYES CLASSIFIER* DAN *SUPPORT VECTOR*  
*MACHINES***

**THESIS**

**Diajukan Kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:  
ULFA KUSNIA  
NIM. 19841003**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2022**

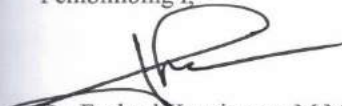
**ANALISIS SENTIMEN REVIEW APLIKASI BERITA *ONLINE* PADA  
GOOGLE PLAY MENGGUNAKAN METODE ALGORITMA NAIVE  
BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINES**

**THESIS**


Oleh :  
**ULFA KUSNIA**  
NIM. 19841003

Telah diperiksa dan disetujui untuk diuji:  
Tanggal: .....

Pembimbing I,

  
Dr. Fachrul Kurniawan, M.MT., IPM  
NIP. 19771020 200912 1 001

Pembimbing II,

  
Dr. M. Faisal M.T M.Si  
NIP. 19740510 200501 1 007

Mengetahui,  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Fawo Crysdiyan  
NIP. 19940424 200901 1 008

**ANALISIS SENTIMEN REVIEW APLIKASI BERITA ONLINE PADA  
GOOGLE PLAY MENGGUNAKAN METODE ALGORITMA NAIVE  
BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINES**

**THESIS**

Oleh :  
**ULFA KUSNIA**  
**NIM. 19841003**

Telah Dipertahankan di Depan Dewan Penguji Thesis  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Magister Komputer (M.Kom)  
Tanggal: .....

**Susunan Dewan Penguji**

Penguji Utama : Dr. Cahyo Crysdian  
NIP. 19740424 200901 1 008  
Ketua Penguji : Ririen Kusumawati, M.Kom  
NIP. 19720309 200501 2 002  
Sekertaris Penguji : Dr. Fachrul Kurniawan, M.MT  
NIP. 19771020 200912 1 001  
Anggota Penguji : Dr. M. Faisal M.T M.Si  
NIP. 19740510 200501 1 007

**Tanda Tangan**

()  
()  
()  
()

Mengetahui dan Mengesahkan  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Maulana Malik Ibrahim Malang



Crysdian  
NIP. 19740424 200901 1 008

## PERNYATAAN KEASLIAN TULISAN

### PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Ulfa Kusnia  
 NIM : 19841003  
 Program Studi : Magister Informatika  
 Fakultas : Sains dan Teknologi

arya saya  
 agai hasil  
 l.

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

bersedia

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 05 Agustus 2022  
 Yang membuat pernyataan,



Ulfa Kusnia  
 NIM. 19841003

## **HALAMAN PERSEMBAHAN**

Terimakasih kepada dosen pembimbing yang telah sabar mendampingi saya. Dosen Pembimbing yang telah mengarahkan saya dalam melakukan penulisan karya ilmiah ini.

Karya ini saya persembahkan untuk kedua orang tua saya. Orang Tua yang telah mendukung secara maksimal dalam penulisan karya ilmiah ini.

Terima Kasih pada keluarga saya yang telah mendukung saya untuk menyelesaikan karya tulis ilmiah ini.

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Syukur alhamdulillah penulis hanturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan jazakumullah ahsanal jaza' kepada semua pihak yang telah membantu terselesaikannya Thesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Fachrul Kurniawan, M.MT dan Bapak Dr. M. Faisal M.T M.Si selaku dosen pembimbing Thesis, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
2. Segenap sivitas akademika Program Studi Magister Informatika, terutama seluruh Bapak/ Ibu dosen, terima kasih atas segenap ilmu dan bimbingannya.
3. Ayahanda dan Ibunda tercinta yang senantiasa memberikan doa dan restunya kepada penulis dalam menuntut ilmu.
4. Semua pihak yang ikut membantu dalam menyelesaikan Thesis ini baik berupa materiil maupun moril.

Penulis menyadari bahwa dalam penyusunan Thesis ini masih terdapat kekurangan dan penulis berharap semoga Thesis ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Amin Ya Rabbal Alamin.*

*Wassalamu'alaikum Wr. Wb.*

Malang, 05 Agustus 2022  
Penulis



## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN .....</b>	<b>iv</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>v</b>
<b>KATA PENGANTAR .....</b>	<b>vi</b>
<b>DAFTAR ISI .....</b>	<b>vii</b>
<b>DAFTAR GAMBAR .....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xii</b>
<b>ABSTRAK .....</b>	<b>xiii</b>
<b>ABSTRACT .....</b>	<b>xiv</b>
<b>الملخص.....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah .....	3
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.5 Batasan Masalah .....	3
1.6 Sistematika Penulisan .....	4
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>6</b>
2.1 Analisis Sentimen .....	6
2.3 Kerangka Teori .....	9
<b>BAB III DESAIN PENELITIAN .....</b>	<b>13</b>
3.1 Prosedur Penelitian .....	13
3.2 Pengumpulan Data .....	13
3.3 <i>System Development</i> .....	14
3.5 Uji Coba .....	20
<b>BAB IV KLASIFIKASI MENGGUNAKAN <i>NAIVE BAYES</i> .....</b>	<b>23</b>
4.1 Desain Metode Algoritma <i>Naive Bayes</i> .....	23
4.2 Uji Coba Algoritma <i>Naive Bayes</i> .....	24
4.3 Kesimpulan Klasifikasi Metode Algoritma <i>Naive Bayes</i> .....	32

## **BAB V KLASIFIKASI MENGGUNAKAN *SUPPORT VECTOR MACHINE* (SVM) 33**

5.1 Desain Metode Algoritma *Support Vector Machine* ..... 33

5.2 Training *Support Vector Machine* ..... 34

5.3 Uji Coba Algoritma *Support Vector Machine* ..... 36

5.3 Kesimpulan Klasifikasi Metode *Support Vector Machine* ..... 41

## **BAB VI PEMBAHASAN..... 43**

6.1 Perbandingan Metode Algoritma *Naive Bayes* dan *Support Vector Machine* 43

6.2 Prepektif Al-Qur'an ..... 48

## **BAB VII KESIMPULAN ..... 48**

7.1 Kesimpulan ..... 48

7.2 Saran ..... 48

## DAFTAR GAMBAR

<b>Gambar 2.1 Kerangka Teori .....</b>	<b>9</b>
<b>Gambar 3.1 Diagram Alir Prosedur Penelitian.....</b>	<b>13</b>
<b>Gambar 3.2 Desain <i>System Development</i>.....</b>	<b>15</b>
<b>Gambar 4.1 Desain Alur Tahapan Klasifikasi <i>Naive Bayes</i> .....</b>	<b>23</b>
<b>Gambar 5.1 Desain Alur Tahapan Klasifikasi <i>Support Vector Machine</i> .....</b>	<b>33</b>
<b>Gambar 5.2 <i>Margin Hyperplane</i> .....</b>	<b>35</b>

## DAFTAR TABEL

Tabel 2.1 Daftar Jurnal .....	10
Tabel 3.1 Data Ulasan.....	14
Tabel 3.2 Pembersihan Dokumen .....	15
Tabel 3.3 <i>Tokenizing</i> .....	16
Tabel 3.4 <i>Stemming</i> .....	17
Tabel 3.5 <i>Stopword</i> .....	17
Tabel 3.6 Perhitungan TF-IDF.....	19
Tabel 3.7 Hasil Pelabelan Kelas Sentiment.....	20
Tabel 3.8 <i>Confusion Matrix</i> .....	20
Tabel 3.9 Split Data <i>Training</i> dan Data <i>Testing</i> .....	22
Tabel 4.1 <i>Split Data Training</i> dan Data <i>Testing Naive Bayes</i> .....	24
Tabel 4.2 Pembagian Jumlah Data <i>Training</i> 50% dan <i>Testing</i> 50% NB.....	25
Tabel 4.3 Hasil Perhitungan Data <i>Training</i> 50% dan <i>Testing</i> 50% NB .....	25
Tabel 4.4 Pembagian Jumlah Data <i>Training</i> 70% dan <i>Testing</i> 30% NB.....	25
Tabel 4.5 Hasil Perhitungan Data <i>Training</i> 70% dan <i>Testing</i> 30% NB .....	26
Tabel 4.6 Pembagian Jumlah Data <i>Training</i> 75% dan <i>Testing</i> 25% NB.....	26
.....Tabel 4.5 Hasil Perhitungan Data <i>Training</i> 75% dan <i>Testing</i> 25% NB	26
Tabel 4.8 Pembagian Jumlah Data <i>Training</i> 80% dan <i>Testing</i> 20% NB.....	27
Tabel 4.9 Hasil Perhitungan Data <i>Training</i> 80% dan <i>Testing</i> 20% NB .....	27
Tabel 4.10 Contoh Kasus Data <i>Training</i> .....	28
Tabel 4.11 Perhitungan Probabilitas Data Training Dokumen 1 .....	29
Tabel 4.12 Perhitungan Probabilitas Data Training Dokumen 2 .....	29
Tabel 4.13 Perhitungan Probabilitas Data Training Dokumen 3 .....	29
Tabel 4.14 Contoh Kasus Data <i>Testing</i> .....	30
Tabel 4.15 Nilai Probabilitas Data <i>Testing</i> .....	31
Tabel 4.16 Hasil Performa Algoritma <i>Naive Bayes</i> .....	32
Tabel 5.1 <i>Split Data Training</i> dan Data <i>Testing SVM</i> .....	37
Tabel 5.2 Pembagian Jumlah Data Training 50% dan Testing 50% SVM... 38	
Tabel 5.3 Hasil Perhitungan Data <i>Training</i> 50% dan <i>Testing</i> 50% SVM.....	38
Tabel 5.4 Pembagian Jumlah Data Training 70% dan Testing 30% SVM... 39	
Tabel 5.5 Hasil Perhitungan Data <i>Training</i> 70% dan <i>Testing</i> 30% SVM.....	39
Tabel 5.6 Pembagian Jumlah Data Training 75% dan Testing 25% SVM... 39	

<b>Tabel 5.7 Hasil Perhitungan Data <i>Training</i> 75% dan <i>Testing</i> 25% SVM.....</b>	<b>40</b>
<b>Tabel 5.8 Pembagian Jumlah Data <i>Training</i> 80% dan <i>Testing</i> 20% SVM...</b>	<b>40</b>
<b>Tabel 5.9 Hasil Perhitungan Data <i>Training</i> 80% dan <i>Testing</i> 20% SVM.....</b>	<b>40</b>
<b>Tabel 5.10 Hasil Hasil Performa Algoritma SVM .....</b>	<b>41</b>
<b>Tabel 6.1 Hasil Performa <i>Algoritma Naive Bayes</i> .....</b>	<b>43</b>
<b>Tabel 6.2 Hasil Performa <i>Algoritma Support Vector Machine</i>.....</b>	<b>44</b>
<b>Tabel 6.3 Perbandingan Metode <i>Algoritma Naive Bayes</i> dan <i>Algoritma Support Vector Machine</i> .....</b>	<b>45</b>

## DAFTAR LAMPIRAN

<b>Lampiran 1 Data Ulasan .....</b>	<b>52</b>
<b>Lampiran 2 Proses Preprocessing.....</b>	<b>52</b>
<b>Lampiran 3 Proses Pembobotan Kata .....</b>	<b>53</b>
<b>Lampiran 4 Sintaks <i>Machine Learning</i>.....</b>	<b>53</b>
<b>Lampiran 5 Hasil Klasifikasi Pada Iterasi Dataset .....</b>	<b>54</b>

## ABSTRAK

Kusnia, Ulfa. 2022. **Analisis Sentimen Review Aplikasi Berita Online Pada Google Play Menggunakan Metode Algoritma Naive Bayes Classifier Dan Support Vector Machines**, Program Magister Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Fachrul Kurniawan, M.MT. (II) Dr. M. Faisal M.T M.Si.

Kata kunci: Analisis Sentimen, *Support Vector Machine*, *Naive Bayes*

Media berita *online* sebagai media massa yang paling banyak dikonsumsi publik yang bisa mengungguli media era sebelumnya misalnya media elektronika dan media cetak. Keunggulan media *online* dibanding media cetak biasanya yaitu praktis, *real time* dan *up to date*. Penilaian umum atas layanan dan berita yang diberikan sangat penting untuk menjaga dan meningkatkan kinerja media berita *online*. Adapun evaluasi publik dapat dilihat melalui page *Google Play* dalam kolom *opini user*. Analisis sentimen bisa menganalisis opini tersebut, dengan proses menganalisa dan mengekstraksi data teks yang tidak terstruktur untuk menghasilkan informasi sentimen yang terdapat dalam kalimat *opini* pada aplikasi. Dalam penelitian analisis sentimen ini, mengimplementasikan algoritma *Naive Bayes* dan *SVM (Support Vector Machines)*. Hasil eksperimen menunjukkan bahwa keakuratan *SVM (Support Vector Machines)* 94.06 % lebih mengungguli dari pada *Naive Bayes* 91.58%.

## ABSTRACT

Kusnia, Ulfa. 2022. **SENTIMENT ANALYSIS OF NEWS APP REVIEW ONLINE ON GOOGLE PLAY USING NAIVE BAYES CLASSIFIER ALGORITHM AND SUPPORT VECTOR MACHINES METHODS**, Masters Program in Informatics, Maulana Ibrahim State Islamic University, Advisors: (1) Dr. Fachrul Kurniawan, M.MT. (2) Dr. M. Faisal M.T M.Si.

Online news media is the most widely consumed mass media by the public, which can outperform the media of the previous era, such as electronic media and print media. The advantage of online media compared to print media is that it is practical, real time and up to date. General assessment of services and news provided is very important to maintain and improve the performance of online news media. The public evaluation can be seen via the Google Play page in the user opinion column. Sentiment analysis can analyze these opinions, with the process of analyzing and extracting unstructured text data to produce sentiment information contained in opinion sentences in the application. In this sentiment analysis research, implementing the Naive Bayes algorithm and SVM (Support Vector Machines). The experimental results show that the accuracy of SVM (Support Vector Machines) 94.06% outperforms Naive Bayes 91.58%.

**Keywords:** Sentiment Analysis, Support VectorMachine, Naive Bayes



### الملخص

آلات المتجهات ، برنامج الماجستير في جامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف: (١) د. فخر الكورنياون. (٢) د. م. فيصل،

الكلمات المفتاحية: تحليل المشاعر ، آلة المتجهات الداعمة ، بايز السذاجة وسائل الإعلام الإخبارية على

الإنترنت هي أكثر وسائل الإعلام استخدامًا من قبل الجمهور ، والتي يمكن أن تتفوق على وسائل الإعلام في الحقبة السابقة ، على سبيل المثال ، الوسائط الإلكترونية والوسائط المطبوعة. تتمثل ميزة الوسائط عبر الإنترنت مقارنة بالوسائط المطبوعة في أنها عملية وفي الوقت الفعلي ومحدثة. التقييم العام للخدمات والأخبار المقدمة مهم للغاية للحفاظ على أداء وسائل الإعلام الإخبارية على الإنترنت وتحسينه. يمكن مشاهدة التقييم العام عبر صفحة في عمود رأي المستخدم. يمكن لتحليل المشاعر تحليل هذه الآراء ، من خلال عملية تحليل واستخراج البيانات النصية غير المهيكلة لإنتاج معلومات المشاعر الواردة في جمل الرأي في التطبيق. في بحث تحليل المشاعر هذا ، تم تنفيذ خوارزمية و (آلات المتجهات الداعمة). أظهرت النتائج التجريبية أن دقة (آلات المتجهات الداعمة) تبلغ ٦٩٤٪. تفوقت على ٥٨,٩١٪.



“Sebenarnya orang yang membawa berita palsu juga dari grup Anda. Jangan menganggap berita palsu itu buruk bagi Anda, itu sebenarnya baik untuk Anda. Masing-masing dari mereka akan menerima upah atas dosa mereka. Dan siapa yang mengambil bagian terbesar dari penyebaran berita bohong akan mendapatkan hukuman yang besar”.

(QS. An-Nur: 11)

Aplikasi media berita online yang terdapat disitus Google Play masih tercatat 50 aplikasi berita. Google Play yaitu platfon digital Google yang mencakup e-commerce produk buku – buku, aplikasi, game & Musik. (Mustopa et al., 2020). Dalam fitur Google Play yaitu adanya kolom umpan balik yang memungkinkan pengguna untuk memberikan peringkat aplikasi dalam bentuk peringkat dan ulasan. Ini digunakan untuk menampilkan peringkat publik berdasarkan peringkat pengguna untuk aplikasi media berita online. Review dari pengguna aplikasi sering opini positif, seperti suka dan rekomendasi, atau opini negatif, seperti kritikan. Ini mungkin kurang lebih merupakan cerminan bagaimana media online bekerja di mata pengguna. Jika Anda memiliki jumlah review media online yang sangat banyak, mengumpulkan dan mengkategorikan review tersebut tentu bukan tugas yang mudah. Oleh karena itu, pengumpulan data informasi dalam jumlah besar memerlukan metode atau teknik khusus, dan teknik untuk mengumpulkan data informasi yaitu menggunakan teknik scraping. Teknik Scraping menangkap dan menganalisis dokumen semi-terstruktur dan mengambil data spesifik dari web page di Internet. Penelitian ini scraper menggunakan scraper setara Google Chrome ntuk mengumpulkan data tentang rating dan opini yang diberikan oleh pengguna aplikasi media berita online. Teknik crawling dan scraping digunakan untuk mendapatkan informasi dari play store. Melalui menambang pada konten untuk mendapatkan informasi. Jadi, konten dari Google Play Store dan metadata akan diukur dan dianalisis (Latif et al., 2019).

Dalam analisis opini merupakan studi komputer tentang opini, tindakan, dan perasaan orang terhadap suatu entitas. Entitas ini dapat menjabarkan individu, peristiwa, atau subjek. Sentiment analysis yaitu pengumpulan data serupa cara pandang terhadap suatu subjek.. Dapat juga diartikan untuk menghasilkan informasi sentimental yang terkandung dalam teks dengan cara memproses data tekstual secara otomatis. Sentimen Analisis

membawa hasil yang luar biasa pengaruh dan manfaat, menyebabkan penelitian dan aplikasi berdasarkan analisis sentimen tumbuh pesat (Mustopa et al., 2020). Metode yang digunakan untuk analisis sentimen yaitu metode klasifikasi machine learning. Klasifikasi adalah proses mendapatkan pola dalam kumpulan data yang membantu memprediksi data yang kategori data tertentu belum tersedia. Sebagai algoritma klasifikasi yang mencakup metode statistik, pembelajaran mesin, dan manajemen basis data, ini juga diklasifikasikan sebagai bagian penting dari interpretasi dan visualisasi data. (Berry, 2010). Dalam riset ini, kami menggunakan Naive Bayesian Algoritma (NB) dan Support Vector Machine Algoritma (SVM) untuk melakukan proses klasifikasi machine learning. Pengklasifikasi Naive Bayesian Algoritma (NB) adalah metode klasifikasi statistik untuk memprediksi probabilitas suatu kelas (Mustopa et al., 2020) dan Metode Support Vector Machine Algoritma (SVM) adalah Sebagai pengklasifikasi diskriminan yang lebih akurat dari pada kebanyakan model klasifikasi lainnya. Generalisasi karakteristik SVM adalah karena penerapan Prinsip Minimalisasi Risiko Struktural (SRM), yang mencakup menemukan hyper-plane pemisah yang optimal (Rizaldy & Santoso, 2017).

Setelah klasifikasi selesai, dilakukan proses penggalian dan pencarian informasi sebanyak-banyaknya dari klasifikasi sentiment kelas positif dan negatif. Tahapan analisis diproses dengan cara statistik deskriptif dan kata-kata untuk menemukan wawasan tentang aplikasi media berita online dan informasi yang terkait dengan topik atau topik yang umum digunakan, penilaian pengguna, dan topik atau diskusi tersebut. Analisis membantu mengidentifikasi kekuatan dan kelemahan aplikasi Anda. Hasil analisis penelitian ini perlu mendukung pengolahan data uji semaksimal mungkin agar dapat memberikan informasi yang selengkap mungkin, dan bisa bermanfaat bagi untuk pemangku kepentingan yang memerlukan.

## **1.1 Pernyataan Masalah**

1. Bagaimana cara mengklasifikasikan review aplikasi media berita online menggunakan Naive Bayesian Algorithmi (NB) dan Support Vector Machiines Algorithmi (SVM)
2. Bagaimana ketepatan performa dari metode Naive Bayesian Algorithmi (NB) dan Support Vector Machiines Algorithmi (SVM) dalam pengklasifikasian teks ?

## 1.2 Tujuan Penelitian

1. Bagi Peneliti
  - a. Mengetahui pengklasifikasian ulasan aplikasi media berita online menggunakan metode *Naive Bayesian Algorithmi* (NB) dan *Support Vector Machiines Algorithmi* (SVM)?
  - b. Mengetahui Akurasi, Precision, Recall, machine learning dari metode *Naive Bayesian Algoriithm* (NB) dan *Support Vector Machiines Algoriithm* (SVM) dalam pengklasifikasian teks?
  - c. Pengetahuan tentang kinerja aplikasi media berita online berdasarkan informasi dari ulasan pengguna aplikasi?
2. Bagi Ilmu Pengetahuan
 

Sebagai referensi untuk perbandingan dan pertanyaan sejenis untuk meningkatkan kualitas penelitian selanjutnya.

## 1.3 Manfaat Penelitian

1. Meningkatkan pengetahuan tentang klasifikasi machine learning, terutama untuk metode Naive Bayesian Algorithm (NB) dan Support Vector Machiines Algorithm (SVM).
2. Memberikan informasi tentang gambaran pola ulasan media berita online.
3. Media berita online memudahkan untuk menemukan umpan balik publik tentang kekuatan dan kelemahan aplikasi.

4. Memberikan berbagai kemungkinan solusi untuk meningkatkan kualitas dan reputasi layanan media berita online di masa depan.

#### **1.4 Batasan Masalah**

1. Data yang dianalisis berasal dari ulasan pada Google Play untuk aplikasi media berita online.
2. Media berita online yang digunakan situs Kompas.com

#### **1.5 Sistematik Penulisan**

Dalam thesis ini, dibagi dalam beberapa bab :

1. BAB - I PENDAHULUAN

Beriisi latar belakang, pernyataan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

2. BAB- II LANDASAN TEORI / DAFTAR PUSTAKA

Membahas penelitian sebelumnya ke dalam masalah yang diteliti dan berfungsi sebagai referensi bagi penulis.

3. BAB III METODOLOGI

Menguraikan analisis data, design system sampai tahapan penelitian.

4. BAB IV HASIL DAN PEMBAHASAN

Menjelaskan hasil analisis penelitian dan pembahasan

5. BAB V KESIMPULAN DAN SARAN

Berisikan kesimpulan dari penelitian serta saran yang bisa menjadi masukan untuk digunakan di masa mendatang.

## BAB II

### TINJAUAN PUSTAKA

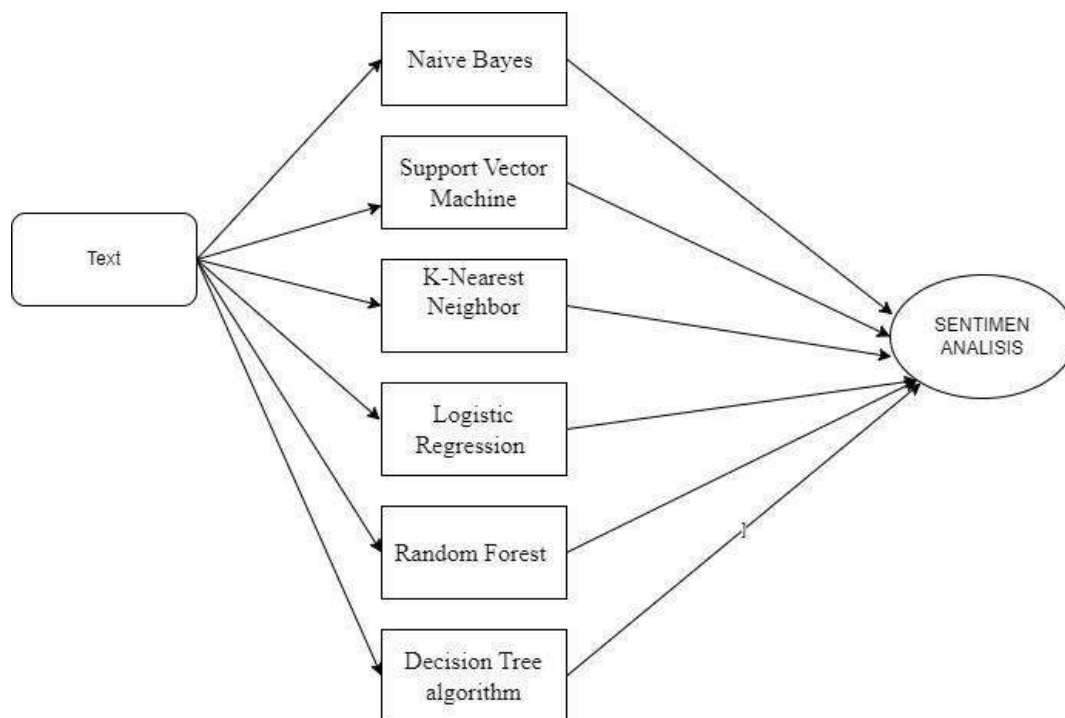
#### 1.1 Sentiment Analysis

*Soumik et al* (2019) algoritma Naive Bayes Classifier and Support Vector Machine. Hasil eksperimen menunjukkan bahwa algoritma garis dasar, setelah pra-pemrosesan yang tepat, dapat menunjukkan hasil yang menjanjikan pada dataset Bangla.

Rizaldy & Santoso (2017) melakukan klasifikasi berita Indonesia penggunaan perolehan informasi sebagai pemilihan fitur meningkatkan akurasi dari pada tanpa fitur apa pun pilihan. Model ini memberikan hasil yang memuaskan dengan akurasi 98,057% dari klasifikasi berita Indonesia. Peningkatan 2,9 poin dari 95,11% dengan teknik SVM tanpa seleksi fitur.

#### 1.2 Kerangka Teori

Penelitian ini memiliki kerangka teori yang mengacu pada jurnal – jurnal dan tergambar jelas pada Gambar 2.1 Kerangka Teorii berikut ini :



Gambar 2.1 Kerangka Teorii



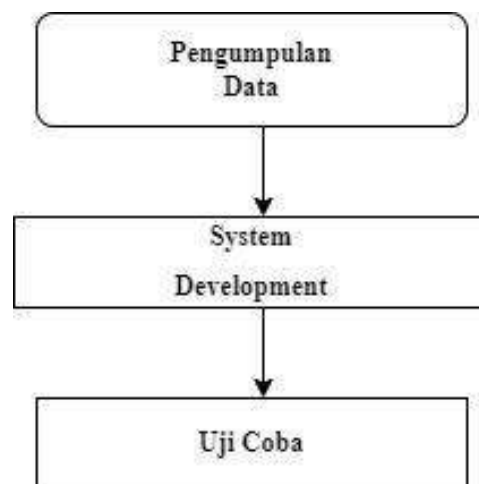
Pada Gambar 2.1 Kerangka Teori dijelaskan bahwa untuk analisis sentimen dapat menggunakan metode algoritma Naiive Bayes, Support Vector Machiine, KNN, Random Forest dan Deciision Tree Algoritma.

## BAB III

### DESAIN PENELITIAN

#### 3.1 Prosedur Penelitian

Prosedur dalam penelitian ini t dipaparkan melalui gambar diagram alir prosedur penelitian sebagai berikut :



Gambar 3.1 Diagram Alir Prosedur Penelitian

#### 3.2 Pengumpulan data

Data ulasan pengguna Aplikasi Media Berita Onliine pada kolom opini di Google Play, pada halaman situs Kompas.com yang telah terdaftar pada Dewan Pers Indosenia, dimana. Komponen yang diambil nama user, rating, dan ulasan dengan proses scraping dengan bantuan tools scrape similar pada Google Crome sebanyak 5894. Berikut sebagian hasil scraping yang diperoleh sebagai berikut :

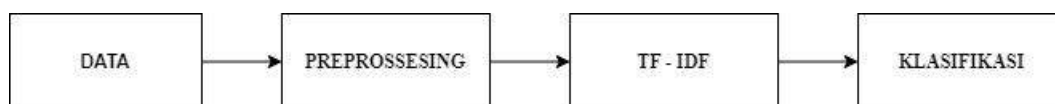
Tabel 3.1 Data Ulasan

No.	Peringkat	User	Ulasan	Tanggal	Bulan	Tahun	Like
1	1	Bassuni Buay Pemuka	jelek	6	3	2021	0
2	2	putra raja	Beritanya yang itu itu aja. ..	20	3	2020	4
3	3	herpian leo	Yang penting pada sehat	21	2	2021	3
4	4	Trisuntara Tara	Kompas media aktual terpercaya.	2	3	2021	0
5	5	indri dwiyatno	bagus banget, bravo kompas	12	6	2021	0
6	4	Budi Suhartono	Ok berita	22	7	2020	4
7	5	Angga Tri S.	Top 3 news	9	3	2021	0
8	5	Nyoman Budiwartawan	Semoga terus sukses dengan berita n info terbaik	6	1	2021	25
9	5	krishna dwipangga	Luar biasa berguna bagi masyarakat yg membutuhkan berita akurat dan terpercaya	3	3	2021	11
10	5	Ramadhan Hafid	Akurat dan independen	13	3	2020	2
11	5	Hasyim Ado	Akurat aktual dan terverifikasi...	4	3	2020	0
12	5	Pengguna Google	akurat , netral , terpercaya	11	3	2019	0

Bedasarkan tabel 3.1 pengguna yaitu pengguna aplikasi media berita online'. opini yaitu asumsi pengguna terhadap aplikasi. Pada tahap ini, data masih real atau belum dibersihkan, sehingga data teks yang diproses tetap bercampur dengan karakter lain yang masih melekat pada data tersebut. Proses selanjutnya adalah proses preprocessing.

### 3.1 System Development

Sistem yang dibangun untuk menganalisis sentimen pada ulasan opini aplikasi berita online sebagai berikut :



Gambar 3.2 Desain System Development

#### 1. Preprocessing Data

Tahapan pengelola data real atau utuh untuk menghilangkan noise, memperjelas fitur, mengkonversi data asli agar sesuai kebutuhan serta memperbesar atau memperkecil data. Berikut beberapa tahapan dalam preprocessing.

a. Proses Data (Cleansing)

Tabel 3.2 Dokumen setelah dibersihkan

No	Ulasan
1	jelek
2	bagus sih berita lengkap cuma buka baca tiba aplikasi keluar sendiri ulang jadi makin kesel paksa uninstal aplikasi baca berita kompas
3	dah lah berita sifat jatuh padahal dulu netral terus
4	sy blm pernah intal kompas com
5	mai yg lincah
6	muka akurat
7	baik
8	gagus akutrak
9	lumayan bagus
10	berita gak imbang bukan sumber asli muat berita hoax
11	bagus sitpi ga pilih buat ngilangin berita hype karna ga suka berita tema artis
12	copot buang ribet
13	aplikasi bagus buat cari tahu berita hari hari
14	baru coba kalo media netral kasih bintang
15	ada kompas sangat bantu saya ada informasi berita kini yang tambah wawasan tahu ilmu yg bermanfaat
16	beritanya akurat lugas moga dpt d percaya publik tentunyn sangat butuh
17	saji berita nya bagus

Pada Tabel 3.2 dijelaskan dokumen diperoleh dibersihkan dari karakter seperti hastag, usernam(@username), tanda baca (.,!?,()/%::[]), angka(1234567) serta karakter – karakter lain selain alphabet. Cleansing bertujuan mengurangi noise.

### b. Proses Tokenizing

Tabel 3.3 Tokenizing

D1	berita gak imbang bukan sumber asli muat berita hoax	berita gak imbang	bukan sumber asli	muat berita hoax
D2	aplikasi bagus buat cari tahu berita hari hari	aplikasi bagus	buat cari tahu	berita hari hari
D3	beritany akurat lugas moga dpt d percaya publik tentuny sangat butuh	beritany akurat lugas	moga dpt d percaya	publik tentuny sangat butuh

Pada Tabel 3.3 Tokenizing bertujuan untuk mempermudah perhitungan frekuensi kemunculan kata dalam dokumen.

### c. Proses Stemming

Tabel 3.4 Stemming

D1	aplikasi	bagus	buat	cari	tahu	berita	hari	hari
D2	saji	berita	bagus					
D3	baru	coba	kalo	media	netral	kasih	bintang	

Pada Tabel 3.4 dijelaskan untuk menghilangkan imbuhan menjadi kata dasar.

### d. Proses Stopword

D1 : berita gak imbang bukan sumber asli muat berita hoax

D2 : beritany akurat lugas moga dpt d percaya publik tentuny sangat butuh

D3 : bagus sitpi ga pilih buat ngilangin berita hype karna ga suka berita tema artis

Tabel 3.5 *Stopword*

D1	berita		imbang	bukan	sumber	asli	muat	berita	hoax	
D2	beritanya	akurat	lugas	moga		percaya	publik	tentunya	sangat	butuh
D3	bagus			pilih	buat		berita	hype	karna	
	suka	berita	tema	artis						

Pada tabel 3.5 dijelaskan tahapan menghapus kata – kata yang tidak penting berdasarkan kamus stopwords seperti kata penghubung (yang, di, ke, ya) atau kata-kata yang tidak mempunyai makna.

## 2. Proses Pembobotan Kata (TF-IDF)

Tahapan proses untuk menghitung dan informasi TF (Term Frequency), DF (Document Frequency) dan IDF (Inverse Document Frequency) menghitung dokumen atau term ini berdasarkan frekuensi kemunculan term atau dokumen tersebut. Kata/term dihitung probabilitas kemunculan dalam satu dokumen (D1 sampai D3). Untuk mendapatkan IDF digunakan persamaan dengan menggunakan persamaan. Setelah dilakukan proses preprocessing, Nilai IDF dapat dirumuskan dalam persamaan sebagai berikut :

$$idf = \log\left(\frac{n}{df}\right)$$

N = jumlah data ulasan

df = keseluruhan ulasan suatu kata (term) yang muncul

Terdapat tiga ulasan (sudah melewati preprocessing) seperti berikut:

Doc.1 : Akurat dan independen

Doc.2 : Akurat Aktual terverifikasi

Doc.3 : Akurat , netral , terpercaya

Tabel 3.6 Hasil Perhitungan TF-IDF

No .	Teks	Dok. 1	Dok. 2	Dok. 3	D F	IDF	(TF.IDF)
------	------	--------	--------	--------	-----	-----	----------

							Dok.1	Dok.2	Dok.3
1.	Akurat	1	1	1	3	$\log(3/3)=0$	0	0	0
2.	Independen	1	0	0	1	$\log(3/1)=0.477$ 1	0.477 1	0	0
3.	Aktual	0	1	0	1	$\log(3/1)=0.477$ 1	0	0.477 1	0
4.	Terverifikasi	0	1	0	1	$\log(3/1)=0.477$ 1	0	0.477 1	0
5.	Netral	0	0	1	1	$\log(3/1)=0.477$ 1	0	0	0.477 1
6.	Terpercaya	0	0	1	1	$\log(3/1)=0.477$ 1	0	0	0.477 1

### 3. Klasifikasi

Secara keseluruhan, berdasarkan kinerja permoforma Naive Bayesian Algorithm dan Support Vector Machine Algorithm, metode terbaik yang dapat memberikan akurasi yang lebih baik untuk analisis sentimen ulasan aplikasi berita online pada google play dapat dibuktikan pada algoritma Support Vector Machine Algorithm dalam berbagai ukuran ulasan training. Ketika ukuran training maksimum, Support Vector Machine Algorithm mencapai puncak akurasi 94.09 persen. Berdasarkan hasil kinerja performa akurasi Naive Bayes dan Support Vector Machine menjadi lebih tinggi ketika ukuran training ditingkatkan. Pengklasifikasi Naive Bayes hanya menganggap setiap kata yang dimasukkan dalam ulasan sebagai kata independen tanpa mempertimbangkan posisi atau urutan kata dalam klasifikasi. Berbeda dengan Support Vector Machine dapat memprediksi kelas dengan mencari hyperplane optimal di antara kemungkinan hyperplane yang memisahkan dua kelas yang berbeda. Hasil klasifikasi sentiment sebagai berikut :

Tabel 3.7 Hasil klasifikasi sentiment

KELAS	JUMLAH
POSITIF	4655 ulasan
NEGATIF	421 ulasan

Setelah diperoleh hasil model dari pembelajaran mesin klasifikasi *Naive Bayes* (NB) dan SVM dilakukan skenario pengujian data uji yang telah displit sebanyak 4 kali. Hasil dari pada model pembelajaran tersebut dilakukan pengujian dengan menggunakan data baru yang belum dilakukan pembelajaran sebelumnya.

Tabel 3.8 *Tabel Kontigensi*

Aktual Class	Prediksi Class	
	+	-
+	Nilai Positif	False Negative
-	False Positive	Nilai Negatif

Banyaknya data pengamatan dalam kategori positif yang dapat diprediksi secara positif (prediksi benar) oleh pembelajaran mesin disebut True Positive (TP). True Negatif (TN) keseluruhan data observasi dalam kelas negatif yang dapat diprediksi secara negatif (prediksi secara benar) oleh algoritma machine learning. False Positive (FP) keseluruhan observasi yang tergolong positif tetapi memiliki kesalahan prediksi disebut. Negatif Palsu (FN) banyaknya pengamatan yang tergolong negatif tetapi memiliki kesalahan prediksi.

1. Akurasi merupakan persentase prediksi yang benar. Akurasi digunakan sebagai akurasi antara nilai aktual dan prediksi.

$$\text{Akurasi} = \frac{\text{True Positive} + \text{False Negative}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

2. *Precision* mengevaluasi kemampuan sistem temu kembali informasi untuk menemukan kembali data top-ranked yang paling relevan, dan didefinisikan



sebagai presentasi data yang dikembalikan yang benar – benar relevan terhadap queri pengguna.

$$Precision = \frac{\text{Relevant Document} \cap \text{Retrieved Document}}{\text{Retrieved Document}}$$

3. *Recall* untuk menemukan semua elemen yang relevan dalam pengumpulan data dan didefinisikan sebagai presentasi data yang diterima sehubungan dengan permintaan pengguna.

Split Data	
<i>Data Training</i>	<i>Data Testing</i>
80%	20%
75%	25%
70%	30%
50%	50%

$$Recall = \frac{\text{Relevant Document} \cap \text{Retrieved Document}}{\text{Relevant Document}}$$

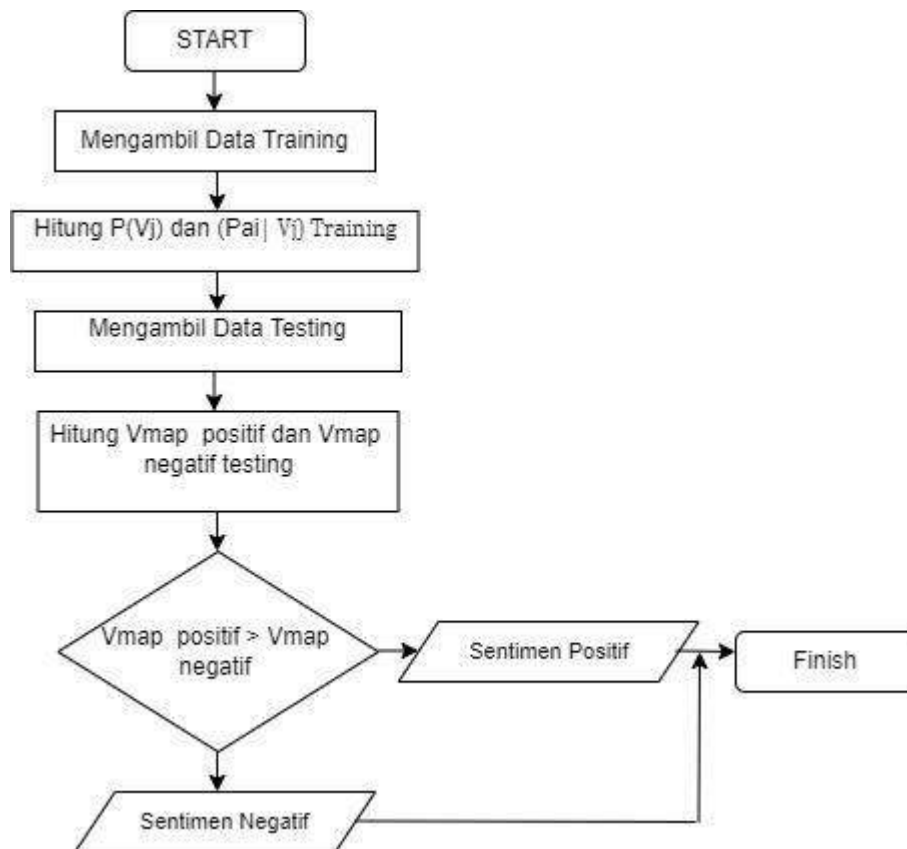
Tabel 3.9 skenario data training dan testing

## **BAB IV**

### **KLASIFIKASI MENGGUNAKAN METODE *NAIVE BAYES ALGORITHM* (NB)**

#### **4.1 Desain Metode *Naive Bayes Algorithm* (NB)**

Tahap klasifikasi *Naive Bayesian Algorithm* melalui proses pelatihan dan proses pengujian. Pada fase ini proses pelatihan terlebih dahulu, baru kemudian dilakukan proses pengujian dengan melihat pada probabilitas dari dataset pelatihan. Diagram alir tahap klasifikasi naive Bayes ditunjukkan pada Gambar 4.1 di bawah ini.



Gambar 4.1 Diagram Metode *Naive Bayes*

Pada pembahasan di bawah ini, penulis memberikan contoh perhitungan naive bayesian manual dengan menggunakan sampel 6 review data latih dan 3 review data uji.

#### 4.2 Uji Coba Metode *Naive Bayes Algorithm* (NB)

Model classifer dibentuk untuk algoritma klasifikasi pada data latih. Sebagai representasi pengetahuann Ini digunakan untuk memprediksi data untuk kelas baru yang tidak ada, dalam pembagian jumlah data latih dan uji dapat dilihat sebagai berikut.

Tabel 4.1 Split jumlah data latih dan uji

Kelas	Banyaknya	Training Data	Testing Data
		(80%)	(20%)

+	4655	3724	931
-	421	336.8≈336	84.2≈84
Total		4060	1015

Berdasarkan Tabel 3.11 diketahui perbandingan jumlah data training dan testing. Dengan proporsi perbandingan sebesar 80% : 20% dari keseluruhan 5.076 ulasan melakukan input data set dengan rasio perbandingan traning 80 % : testing 20%, traning 70% : testing 30%, traning 50 % : testing 50% dan traning 75% : testing 25%. Dari hasil pengklasifikasian data ulasan kelas sentimen positif 4655 dan negatif 421.

Berikut percobaan dengan data latih 50% dan uji 50%.

Tabel 4.2 Hasil Percobaan Data latih 50% dan uji 50%

Data Aktual	Data Prediksi		Precision	Recall
	Positif	Negatif		
+	2320 (TP)	0 (FN)	91.48%	100%
-	216 (FP)	2 (TN)	100%	0,91%
Akurasi	91.48%			

Dengan perhitungan persamaan sebagai berikut

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} = \frac{2320 + 0}{2320 + 216 + 2 + 0} = 91.48\%$$

$$\text{Precision}_{\text{Positif}} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{2320}{2320 + 216} = 91.48\%$$

$$\text{Precision}_{\text{Negatif}} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{216}{216 + 0} = 100\%$$

$$\text{Recall}_{\text{Positif}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{2320}{2320 + 0} = 100\%$$

$$Recall_{Negatif} = \frac{TN}{TN + FP} = \frac{2}{216 + 216} = 0,91\%$$

### 4.3 Kesimpulan

Kesimpulan mengenai hasil performa algoritma *Naive Bayes*. Akurasi prediksi dihitung dengan menggunakan matriks konfusi sebagai salah satu alat perhitungannya. Metode ini melakukan perhitungan pada tiga output: Akurasi, Precision dan Recall.

Tabel 4.3 Hasil Performa Algoritma *Naive Bayes*

Training Size	Testing Size	Akurasi	Precision	Recall
80%	20%	92.02 %	95.98%	54.49%
75%	25%	91.88%	87.64%	54.41%
70%	30%	91.00%	95.48%	52.43%
50%	50%	91.48%	95.74%	50.22%
Rata - rata		91.58	93.70%	50.45%

Pada Tabel 4.3 tentang perbandingan hasil pengujian menyatakan bahwa dari hasil pengamatan uji coba yang telah dilakukan 4 kali percobaan yaitu dengan merubah prosentase data training mendapatkan hasil yang menyatakan bahwa prosentase split data training dan data testing sangat mempengaruhi tingkat akurasi, presisi, dan recall dalam proses pengujian. Pada pengujian yang dilakukan menunjukkan bahwa nilai akurasi tertinggi didapat pada pengujian ke 1 dengan

prosentase 80% data training yaitu sebanyak 4060 data, sedangkan 20% yaitu sebanyak 1015 data testing dengan menghasilkan akurasi 92.02 %, presisi 95.98%, Recall 54.49%, serta rata-rata akurasi 91.58%.

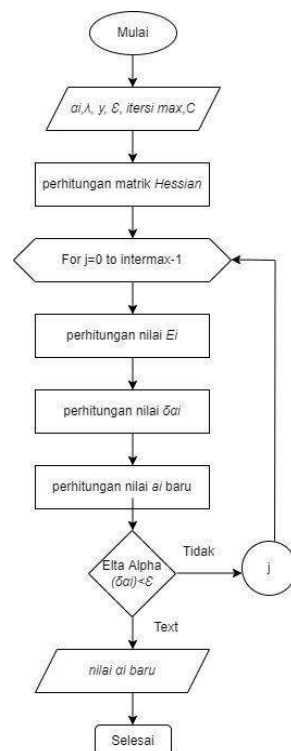
## BAB V

### KLASIFIKASI MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

#### ALGORITHM

##### 5.1 Desain Metode Support Vector Machine Algorithm

Pada desain metode Support Vector Machine terdapat uraian tentang pola dan rancangan klasifikasi sebagai berikut:



Gambar 5.1 Diagram Alir Metode Support Vector Machine

Bedasarkan Gambar 5.1 Untuk memperoleh hyperplane yang maksimal digunakan proses sequential training SVM. Inisialisasi nilai parameter SVM yaitu  $\alpha_i$ ,  $\lambda$ ,  $y$ ,  $\epsilon$ , itersi, fungsi kernel dan  $C$ . Fungsi kernel yang dipakai pada penelitian yaitu fungsi kernel Linier. Menghitung nilai matrik Hessian. Menghitung nilai  $E_i$ . Selanjutnya mencari nilai  $\delta \alpha_i$ . Setelah nilai  $\delta \alpha_i$  diperoleh, kemudian melakukan iterasi dengan jumlah iterasimax sesuai panjang data latih. Dilanjutkan dengan perhitungan nilai  $\alpha_i$ . Iterasi berjalan apabila  $(\delta \alpha_i) < \epsilon$  atau sudah memenuhi dari seluruh data latih yang ditentukan. Proses diatas dilakukan sampai

diperoleh nilai ai baru. Hasil dari nilai ai terakhir yang sudah di-update disebut pembatas yang optimal untuk kemudian digunakan pada proses learning dan testing SVM.

## 5.2 Uji Coba Metode Algoritma Support Vector Machine

Uji coba dilakukan dengan melakukan input data set dengan rasio perbandingan training 80 % : testing 20%, training 70% : testing 30%, training 50 % : testing 50% dan training 75% : testing 25%. Dari hasil pengklasifikasian data ulasan kelas sentimen positif 4655 dan negatif 421. Berikut percobaan dengan data latih 50% dan uji 50%.

Tabel 5.1 Hasil Percobaan Data latih 50% dan uji 50%

Data Aktual	Data Prediksi		Precision	Recall
	Positif	Negatif		
Positif	2298 (TP)	22 (FN)	94.52%	99.05%
Negatif	133 (FP)	85 (TN)	79.43%	38.99%
Akurasi	93.89%			

Dengan perhitungan persamaan sebagai berikut

$$\text{Akurasi} = \frac{TP + FN}{TP + FP + TN + FN} = \frac{2298 + 22}{2298 + 133 + 85 + 22} = 93.89\%$$

$$\text{Precision}_{\text{Positif}} = \frac{TP}{TP + FP} = \frac{2298}{2298 + 133} = 94.52\%$$

$$\text{Precision}_{\text{Negatif}} = \frac{TN}{TN + FN} = \frac{85}{85 + 22} = 79.43\%$$

$$\text{Recall}_{\text{Positif}} = \frac{TP}{TP + FN} = \frac{2298}{2298 + 22} = 99.05\%$$

$$\text{Recall}_{\text{Negatif}} = \frac{TN}{TN + FP} = \frac{85}{85 + 133} = 38.99\%$$



### 5.3 Kesimpulan

Kesimpulan mengenai hasil performa Support Vector Machine algoritma. Perhitungan ketepatan prediksi menggunakan Confusion Matriks sebagai alat perhitungannya. Metode ini menghitung dengan 3 keluaran yaitu Akurasi, presisi dan recall.

Tabel 5.2 Hasil Performa Algoritma *Support Vector Machine*

Training Size	Testing Size	Akurasi	Precision	Recall
80%	20%	94.09%	89.91%	69.33%
75%	25%	94.08%	87.87%	70.69%
70%	30%	93.76%	87.73%	71.67%
50%	50%	93.89 %	90.13 %	70.08 %
Rata - rata		94.06%	88.91%	70.44%

Pada Tabel 5.2 tentang perbandingan hasil pengujian menyatakan bahwa dari hasil pengamatan uji coba yang telah dilakukan 4 kali percobaan yaitu dengan merubah prosentase data training mendapatkan hasil yang menyatakan bahwa prosentase split data training dan data testing sangat mempengaruhi tingkat akurasi, presisi, dan recall dalam proses pengujian. Pada hasil pengujian yang dilakukan menunjukkan bahwa nilai akurasi tertinggi didapat pada pengujian ke 4 dengan prosentase 50% data training yaitu sebanyak 2538 data, sedangkan 50% yaitu sebanyak 2538 data testing dengan menghasilkan akurasi 94.06%, presisi 88.91%, Recall 70.44%, serta rata-rata akurasi 94.06% .

## BAB VI

### PEMBAHASAN

#### 6.1 Perbandingan Metode *Naive Bayesian Algorithm* dan *Algoritma Support Vector Machine*

Tabel 6.1 Hasil Perbandingan Pengujian Metode *Naive Bayesian Algorithm* dan *Algoritma Support Vector Machine*

Trainin g	Testing	<i>Naive Bayesian Algorithm</i>			<i>Support Vector Machine</i>		
		Accurac y	Presisi	Recall	Accuracy	Presisi	Recall
80%	20%	92.02 %	95.98%	54.49%	94.09%	89.91%	69.33%
75%	25%	91.88%	87.64%	54.41%	94.08%	87.87%	70.69%
70%	30%	91.00%	95.48%	52.43%	93.76%	87.73%	71.67%
50%	50%	91.48%	95.74%	50.22%	93.89 %	90.13 %	70.08 %
Rata - rata		91.58%	93.70%	50.45%	94.06%	88.91%	70.44%

#### 6.2 Prepektif Al-Qur'an

Dalam perspektif kitab suci al-quran yang memaparkan tentang berita, dalam surah al-qur'ani an-nur ayat 11 menjelaskan tentang cara menerima suatu berita.

وَمِنْ أَشْرَارٍ أُولَئِكَ  
 الَّذِينَ يَنْفَرُونَ مِنْ أَهْلِ  
 الْقَرْيَةِ وَهُمْ لَا يُخْبِرُونَ  
 أَهْلَ الْقَرْيَةِ بِمَا  
 كَانُوا يَفْعَلُونَ

“Sebenarnya orang yang membawa berita palsu juga dari grup Anda. Jangan menganggap berita palsu itu buruk bagi Anda, itu sebenarnya baik untuk Anda. Masing-masing dari mereka akan menerima upah atas dosa mereka. Dan siapa yang mengambil

bagian terbesar dari penyebaran berita bohong akan mendapatkan hukuman yang besar”.

(QS. An-Nur: 11)

## **BAB VII**

### **KESIMPULAN**

#### **7.1 Kesimpulan**

Hasil Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play Menggunakan Support Vector Machine Algorithm (SVM) dan Naive Bayesian Algorithm. Hasil pengklasifikasi data ulasan kelas sentimen positif menjadi 4655 dan negatif menjadi 421. Hasil performa menghasilkan nilai akurasi dari Support Vector Machine Algorithm (SVM) 94.06% dan Naive Bayes Algorithm 91.58%.

Dari dua pengujian dalam empat skenario pengujian, kami menemukan bahwa akurasi dipengaruhi oleh perbandingan jumlah data training dan data testing. Jika jumlah data training yang dilakukan lebih besar dari jumlah data testing, akurasi hasil akan melebihi 90%, seperti yang terlihat dari hasil uji pada Skenario 1 sampai . Namun pada penelitian ini, skenario pertama, Support Vector Machine Algorithm (SVM) 94.09 % Naive Bayesian Algorithm 92.02%, memberikan akurasi terbaik ketika perbandingan angka 80:20.

#### **7.1 Saran**

Sebagai pembanding performa Support Vector Machine Algorithm dan algoritma Naive Bayesian Algorithm untuk mengklasifikasikan opini aplikasi media berita online ataupun pada situs lainya selain Google Play, untuk penulis selanjutnya bisa menggunakan algoritma machine learning lainnya.

Untuk peneliti berikutnya, sebagai bahan perbaikan perlu dikaji untuk perbandingan hasil analisis sentiment dengan menerapkan negasi dalam Bahasa Indonesia, misal dengan Algoritma First Sentiment Word (First Sentiment Word).

## DAFTAR PUSTAKA

- Adiwijaya. (2006). Text Mining dan Knowledge Discovery.
- Awad, M., & Khanna, R. (2015). Support vector machines for classification. In *Efficient Learning Machines* (pp. 39–66). Springer.
- Bayquni, B. (2018). Partisipasi khalayak media online terhadap liptan6. Com dalam memenangkan persaingan industri media massa di Indonesia. *Jurnal Pustaka Komunikasi*, 1(2), 228–237.
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189–215.
- Darujati, C., & Gumelar, A. B. (2012). Pemanfaatan teknik supervised untuk klasifikasi teks bahasa indonesia. *Jurnal Bandung Text Mining*, 16(1), 5–1.
- Day, M.-Y., & Lin, Y.-D. (2017). Deep learning for sentiment analysis on google play consumer review. *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, 382–388.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115–123.
- Handayani, F., & Pribadi, F. S. (2015). Implementasi algoritma naive bayes classifier dalam pengklasifikasian teks otomatis pengaduan dan pelaporan masyarakat melalui layanan call center 110. *Jurnal Teknik Elektro*, 7(1), 19–24.

Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. *IEEE Access*, 8, 90847–90861.

Husnun N., D. (2006). *Panduan Menulis Berita*. Umm.  
<https://www.google.com/search?client=ms-google&q=Berita+adalah+sebuah+laporan+atau+pemberitahuan+mengenai+terjadinya+sebuah+peristiwa+atau+>

Karima, R. C., Silfiani, S., Puspasari, R. A., Junaeni, R., & Fatoni, U. (2018). Reportase Wartawan dalam Berita Hukum. *Prosiding Konferensi Nasional Komunikasi*, 2(01), 721–736.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4), 150.

Latif, R. M. A., Abdullah, M. T., Shah, S. U. A., Farhan, M., Ijaz, F., & Karim, A. (2019). Data scraping from google play store and visualization of its content for analytics. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET)*, 1–8.

Ma, T. M., Yamamori, K., & Thida, A. (2020). A comparative approach to Naïve Bayes classifier and support vector machine for email spam classification. *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 324–326.

Muin, A. A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 2(1), 22–26.

Mustopa, A., Pratama, E. B., Hendini, A., & Risdiansyah, D. (2020). Analysis of user reviews for the pedulilindungi application on google play using the support vector

machine and naive bayes algorithm based on particle swarm optimization. 2020 Fifth International Conference on Informatics and Computing (ICIC), 1–7.

Okezone. (2017, February 19). Dewan Pers: Ada 43.000 Media Online yang Terdata, 500 Terdaftar, 7 Terverifikasi : Okezone Nasional. <https://Nasional.Okezone.Com/>.  
<https://nasional.okezone.com/read/2017/02/19/337/1622468/dewan-pers-ada-43-000-media-online-yang-terdata-500-terdaftar-7-terverifikasi>

Rizaldy, A., & Santoso, H. A. (2017). Performance improvement of Support Vector Machine (SVM) With information gain on categorization of Indonesian news documents. 2017 International Seminar on Application for Technology of Information and Communication (ISemantic), 227–232.

Romli, A. S. M. (2018). Jurnalistik online: Panduan mengelola media online. Nuansa Cendekia.

Santoso, S. (2021). Analisis Resepsi Audiens Terhadap Berita Kasus Meiliana di Media Online. Komuniti: Jurnal Komunikasi Dan Teknologi Informasi, 12(2), 140–154.

Wongkar, M., & Angdresey, A. (2019). Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter. 2019 Fourth International Conference on Informatics and Computing (ICIC), 1–5.

Yunus, S. (2010). Jurnalistik terapan. Bogor: Ghalia Indonesia.