

**PERBANDINGAN ALGORITMA KNN, NAÏVE BAYES, DAN  
SUPPORT VECTOR MACHINE PADA ULASAN  
USER DI E-COMMERCE**



**UMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

**Gregorius Dinov Adityasmara**

**00000043810**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG**

**2024**

**PERBANDINGAN ALGORITMA KNN, NAÏVE BAYES, DAN  
SUPPORT VECTOR MACHINE PADA ULASAN  
USER DI E-COMMERCE**



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom)

**Gregorius Dinov Adityasmara**

**00000043810**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

### HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Gregorius Dinov Adityasmara

Nomor Induk Mahasiswa : 00000043810

Program studi : Sistem Informasi

Skripsi dengan judul:

PERBANDINGAN ALGORITMA KNN, NAÏVE BAYES, DAN SUPPORT VECTOR MACHINE PADA ULASAN USER DI E-COMMERCE merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 24 Mei 2024



Gregorius Dinov Adityasmara

## HALAMAN PENGESAHAN

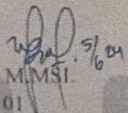

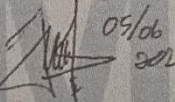
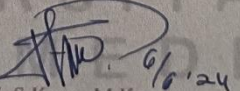
**HALAMAN PENGESAHAN**

Skripsi dengan judul  
"Perbandingan Algoritma KNN, Naive Bayes, Dan Support Vector Machine Pada Ulasan User Di E-Commerce"

Oleh

Nama : Gregorius Dinov Adityasmara  
NIM : 00000043810  
Program Studi : Sistem Informasi  
Fakultas : Teknik Dan Informatika

Telah diujikan pada hari Jumat, 24 Mei 2024  
Pukul 13.00 s.d 15.00 dan dinyatakan  
**LULUS**  
Dengan susunan penguji sebagai berikut.

Ketua Sidang	Penguji
 Wella, S.Kom., M.MSI. 0305119101	 Dr. Friska Wiratama, S.Kom., M.T. 0105128307
Pembimbing	
 Jansen Wiratama, S.Kom., M.Kom. 0409019301	
Ketua Program Studi Sistem Informasi	
 Ririn Ikana Desanti, S.Kom., M.Kom. 0313058001	

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Gregorius Dinov Adityasmara

NIM : 00000043810

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah :

**PERBANDINGAN ALGORITMA KNN, NAÏVE BAYES, DAN  
SUPPORT VECTOR MACHINE PADA ULASAN USER DI E-  
COMMERCE**

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial. Saya tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: Dalam proses pengajuan penerbitan ke dalam jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)\*.

Tangerang, 07 Juni 2024

Yang menyatakan,



Gregorius Dinov Adityasmara

## KATA PENGANTAR

Puji Syukur atas selesainya penulisan Skripsi ini dengan judul:” **PERBANDINGAN ALGORITMA KNN, NAÏVE BAYES, DAN SUPPORT VECTOR MACHINE PADA ULASAN USER DI E-COMMERCE**” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar (S.Kom) Jurusan Sistem Informasi Pada Teknik Dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, M.A. selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng Niki Prastomo, S.T.,M.Sc, selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom, selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Jansen Wiratama, S.Kom., M.Kom, sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman-teman seperjuangan saya dari awal sampai akhir yaitu Aloysius Michael, Alvin Lukita, Yeremia Steven, Marcellius Cahyadi, Rico Tanuwijaya Saputra, Muhamad Calvin Syah Putra, Celine Kurnia, Tralya Dharmada, Vincentius Juan

Semoga laporan ini dapat bermanfaat untuk dijadikan referensi atau informasi bagi pembaca.

Tangerang, 15 Mei 2024



Gregorius Dinov Adityasmara



# PERBANDINGAN ALGORITMA KNN, NAÏVE BAYES, DAN SUPPORT VECTOR MACHINE PADA ULASAN USER DI E-COMMERCE

Gregorius Dinov Adityasmara

## ABSTRAK

Dalam era digital saat ini, *e-commerce* telah menjadi pilar utama dalam bisnis global. Teknologi *e-commerce* memungkinkan proses jual beli barang dan jasa melalui internet, mendorong pelaku usaha untuk mengadopsi sistem penjualan elektronik guna menjaga daya saing. Namun, perkembangan *e-commerce* di Indonesia juga menimbulkan masalah, seperti munculnya toko online palsu yang merugikan konsumen. Meskipun ada masalah seperti tersebut, dengan memanfaatkan analisis sentimen dapat dilihat toko yang memiliki ulasan yang bagus dan toko yang memiliki ulasan buruk. Maka dari itu diperlukan penelitian analisis sentimen ulasan user di-commerce pada platform google play. Dalam penelitian ini metode analisis data yang digunakan yaitu metode CRISP-DM. Sentimen analisis yang digunakan Dalam perbandingan ini yaitu membandingkan tiga algoritma yang ada dari tiap e-commerce. Algoritma yang dilakukan perbandingan yaitu Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor. Dalam melakukan perbandingan algoritma dan pengumpulan data tools yang digunakan yaitu google collabs. Teknik pengumpulan data yang dilakukan yaitu dengan teknik scrapping data ulasan google play store pada aplikasi shopee, bukalapak, lazada, dan Tokopedia. Berdasarkan hasil penelitian menghasilkan perbandingan algoritma antara *KNN*, *Naïve Bayes*, dan *SVM* dari ketiga perbandingan yang sudah dilakukan terhadap empat *e-commerce* di Indonesia didapatkan hasil algoritma yang paling tertinggi dan memiliki performa yang baik dengan hasil yang selalu tinggi yaitu algoritma naïve bayes dengan hasil 93% dan 91%. Shopee memiliki jumlah ulasan positif yang signifikan, menunjukkan bahwa pelanggan cenderung puas dengan produk, layanan, atau merek. Ini bisa menunjukkan bahwa Shopee berhasil memenuhi atau bahkan melebihi harapan pelanggan mereka dalam beberapa hal.

**Kata kunci:** Analisis Sentimen, Big Data Analisis, e-commerce

## **COMPARISON OF KNN, NAIVE BAYES, AND SUPPORT VECTOR MACHINE ALGORITHM IN REVIEW USERS IN E-COMMERCE**

Gregorius Dinov Adityasmara

### **ABSTRACT (English)**

*In the digital era, e-commerce has become a cornerstone of global business, enabling the buying and selling of goods and services via the Internet. This technological advancement encourages businesses to adopt electronic sales systems to stay competitive. However, the rise of e-commerce in Indonesia has also led to issues like fake online stores, which harm consumers. Despite these problems, sentiment analysis can identify which stores have good or bad reviews. Therefore, research on sentiment analysis of user reviews on the Google Play platform is essential. This study employs the CRISP-DM method and compares three algorithms Naïve Bayes, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) for e-commerce sentiment analysis. Tools like Google Collab facilitate algorithm comparison and data collection, utilizing the Google Play Store review data scraping technique for applications such as Shopee, Bukalapak, Lazada, and Tokopedia. The findings reveal that among KNN, Naïve Bayes, and SVM, the Naïve Bayes algorithm consistently performs the best, with accuracy rates of 93% and 91%. Shopee stands out with a significant number of positive reviews, suggesting high customer satisfaction with its products, services, or brand. This indicates that Shopee successfully meets or exceeds customer expectations in several areas. The comparative analysis highlights the importance of using robust algorithms in sentiment analysis to ensure accurate insights. The results can help e-commerce platforms improve their services by understanding customer feedback better, ultimately enhancing user experience and trust in online shopping.*

**Keywords:** Big Data Analytics, e-commerce, Sentiment Analysis



## DAFTAR ISI

<b>HALAMAN PERNYATAAN TIDAK PLAGIAT</b> .....	ii
<b>HALAMAN PENGESAHAN</b> .....	iii
<b>HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS</b> .....	iv
<b>KATA PENGANTAR</b> .....	v
<b>ABSTRAK</b> .....	vi
<b>ABSTRACT (English)</b> .....	vii
<b>DAFTAR ISI</b> .....	viii
<b>DAFTAR TABEL</b> .....	x
<b>DAFTAR GAMBAR</b> .....	xi
<b>DAFTAR LAMPIRAN</b> .....	xiv
<b>BAB I PENDAHULUAN</b> .....	1
<b>1.1 Latar Belakang</b> .....	1
<b>1.1.1 Rumusan Masalah</b> .....	4
<b>1.2 Batasan Masalah</b> .....	4
<b>1.3 Tujuan dan Manfaat Penelitian</b> .....	5
<b>1.3.1 Tujuan Penelitian</b> .....	5
<b>1.3.2 Manfaat Penelitian</b> .....	5
<b>1.4 Sistematika Penulisan</b> .....	5
<b>BAB II LANDASAN TEORI</b> .....	7
<b>2.1 Penelitian terdahulu</b> .....	7
<b>2.2 Sentimen Analisis</b> .....	13
<b>2.3 Teori Algoritma yang digunakan</b> .....	14
<b>2.3.1 Support Vector Machine</b> .....	14
<b>2.3.2 Naïve Bayes</b> .....	15
<b>2.3.3 K-Nearest Neighbor (K-NN)</b> .....	15
<b>2.4 Google Collabs dan Web Scrapping</b> .....	16
<b>BAB III METODOLOGI PENELITIAN</b> .....	17
<b>3.1 Gambaran Umum Objek Penelitian</b> .....	17
<b>3.2 Metode Penelitian</b> .....	18

3.2.1	Alur Penelitian.....	18
3.3	Teknik Pengumpulan Data .....	19
3.3.1	Scraping data.....	20
3.4	Variabel Penelitian.....	20
3.5	Teknik Analisis Data .....	20
3.5.1	Bussiness Understanding.....	21
3.5.2	Data Understanding.....	21
3.5.3	Data preparation .....	21
3.5.4	Modeling .....	22
3.5.5	Evaluation.....	22
3.5.6	Deployment.....	22
3.6	Algoritma yang digunakan untuk pengujian .....	22
3.6.1	Support Vector Machine .....	23
3.6.2	Naïve Bayes.....	23
3.6.3	K-Nearest Neighbor (K-NN) .....	23
<b>BAB IV ANALISIS DAN HASIL PENELITIAN.....</b>		<b>24</b>
4.1	Analisa Masalah dan Kebutuhan Penelitian .....	24
4.1.1.	Bussiness Understanding .....	24
4.1.3	Data Preparation .....	28
4.2	Kebutuhan Sistem.....	28
4.3	Hasil Analisis Data .....	79
4.3.1	Hasil observasi .....	79
<b>BAB V SIMPULAN DAN SARAN.....</b>		<b>81</b>
5.1	Kesimpulan.....	81
5.2	Saran.....	81
<b>DAFTAR PUSTAKA.....</b>		<b>82</b>
<b>LAMPIRAN.....</b>		<b>87</b>

## DAFTAR TABEL

<b>Tabel 1. 1</b> Dampak positif dan negatif e-commerce [13].....	3
<b>Tabel 2. 1</b> Penelitian Terdahulu.....	7
<b>Tabel 3. 1</b> E-commerce terbaik di indonesia[25].....	17
<b>Tabel 3. 2</b> Perbandingan metode data mining untuk analisis data[26].....	20

UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR GAMBAR

<b>Gambar 1. 1</b> List Pengguna E-Commerce yang sering digunakan [4] .....	1
<b>Gambar 2. 1</b> Hyperplane memisahkan dua kelas .....	14
<b>Gambar 2. 2</b> Rumus Naive Bayes .....	15
<b>Gambar 2. 3</b> Rumus KNN .....	16
<b>Gambar 3.1</b> Alur penelitian .....	19
<b>Gambar 4. 1</b> Proses input library dan pensortiran data berdasarkan Most Relevant .....	25
<b>Gambar 4. 2</b> Hasil scrape ulasan Shopee.....	25
<b>Gambar 4. 3</b> Hasil scrape ulasan Bukalapak.....	26
<b>Gambar 4. 4</b> Hasil scrape ulasan Lazada .....	27
<b>Gambar 4. 5</b> Hasil scrape ulasan Tokopedia.....	28
<b>Gambar 4. 6</b> Proses input library dan input data shopee sebelum preprocessing	29
<b>Gambar 4. 7</b> Proses pengecekan data duplicated dan kolom kosong sebelum cleaning data.....	29
<b>Gambar 4. 8</b> Proses cleaning data scrapping shopee.....	30
<b>Gambar 4. 9</b> Proses Normalisasi data shopee yang sudah dilakukan cleaning....	31
<b>Gambar 4. 10</b> Install Sastrawi untuk persiapan tahap StopWord .....	31
<b>Gambar 4. 11</b> Proses StopWord data shopee yang sudah di normalisasi .....	32
<b>Gambar 4. 12</b> Proses Tokenize data shopee .....	32
<b>Gambar 4. 13</b> Proses Stemming data Shopee sebagai tahap preprocessing terakhir .....	33
<b>Gambar 4. 14</b> Hasil preprocessing data shopee .....	34
<b>Gambar 4. 15</b> Proses input library dan input data Bukalapak sebelum preprocessing.....	34
<b>Gambar 4. 16</b> Proses pengecekan data duplicated dan kolom kosong sebelum cleaning data.....	35
<b>Gambar 4. 17</b> Proses cleaning data scrapping Bukalapak .....	36
<b>Gambar 4. 18</b> Proses Normalisasi data Bukalapak yang sudah dilakukan cleaning .....	37
<b>Gambar 4. 19</b> Install Sastrawi untuk persiapan tahap StopWord .....	37
<b>Gambar 4. 20</b> Proses StopWord data Bukalapak yang sudah di normalisasi .....	38

<b>Gambar 4. 21</b> Proses Tokenize data Bukalapak.....	38
<b>Gambar 4. 22</b> Proses Stemming data Bukalapak sebagai tahap preprocessing terakhir .....	39
<b>Gambar 4. 23</b> Hasil preprocessing data Bukalapak.....	40
<b>Gambar 4. 24</b> Proses input library dan input data Lazada sebelum preprocessing .....	40
<b>Gambar 4. 25</b> Proses pengecekan data duplicated dan kolom kosong sebelum cleaning data.....	41
<b>Gambar 4. 26</b> Proses cleaning data scrapping Lazada .....	41
<b>Gambar 4. 27</b> Proses Normalisasi data Lazada yang sudah dilakukan cleaning .	42
<b>Gambar 4. 28</b> Install Sastrawi untuk persiapan tahap StopWord .....	42
<b>Gambar 4. 29</b> Proses StopWord data Lazada yang sudah di normalisasi.....	43
<b>Gambar 4. 30</b> Proses Tokenize data Lazada .....	43
<b>Gambar 4. 31</b> Proses Stemming data Lazada sebagai tahap preprocessing terakhir .....	44
<b>Gambar 4. 32</b> Hasil preprocessing data Lazada .....	44
<b>Gambar 4. 33</b> Proses input library dan input data Tokopedia sebelum preprocessing .....	45
<b>Gambar 4. 34</b> Proses pengecekan data duplicated dan kolom kosong sebelum cleaning data.....	45
<b>Gambar 4. 35</b> Proses cleaning data scrapping Tokopedia .....	46
<b>Gambar 4. 36</b> Proses Normalisasi data Tokopedia yang sudah dilakukan cleaning .....	47
<b>Gambar 4. 37</b> Install Sastrawi untuk persiapan tahap StopWord .....	47
<b>Gambar 4. 38</b> Proses StopWord data Tokopedia yang sudah di normalisasi .....	48
<b>Gambar 4. 39</b> Proses Tokenize data Tokopedia.....	48
<b>Gambar 4. 40</b> Proses Stemming data Tokopedia sebagai tahap preprocessing terakhir .....	49
<b>Gambar 4. 41</b> Hasil preprocessing data Tokopedia.....	49
<b>Gambar 4. 42</b> Proses input library dan input data shopee sebelum Labelling .....	50
<b>Gambar 4. 43</b> Proses input library untuk analisis data .....	51
<b>Gambar 4. 44</b> Proses analisis ulasan dikelompokkan menjadi Positif, Negatif... ..	52
<b>Gambar 4. 45</b> Proses install library matplotlib .....	52
<b>Gambar 4. 46</b> Proses WordCloud untuk melihat kata kata terbanyak pada ulasan .....	53
<b>Gambar 4. 47</b> Hasil WordCloud Shopee .....	53
<b>Gambar 4. 48</b> Proses input library dan input data Bukalapak sebelum Labelling.....	54
<b>Gambar 4. 49</b> Proses input library untuk analisis data .....	55
<b>Gambar 4. 50</b> Proses analisis ulasan dikelompokkan menjadi Positif, Negatif... ..	55
<b>Gambar 4. 51</b> Proses WordCloud untuk melihat kata kata terbanyak pada ulasan .....	56
<b>Gambar 4. 52</b> Hasil WordCloud Bukalapak .....	57

<b>Gambar 4. 53</b>	Proses input library dan input data Lazada sebelum Labelling ....	58
<b>Gambar 4. 54</b>	Proses input library untuk analisis data .....	59
<b>Gambar 4. 55</b>	Proses analisis ulasan dikelompokkan menjadi Positif, Negatif...	59
<b>Gambar 4. 56</b>	Proses WordCloud untuk melihat kata kata terbanyak pada ulasan .....	60
<b>Gambar 4. 57</b>	Hasil WordCloud Lazada .....	60
<b>Gambar 4. 58</b>	Proses input library dan input data Tokopedia sebelum Labelling	61
<b>Gambar 4. 59</b>	Proses input library untuk analisis data .....	61
<b>Gambar 4. 60</b>	Proses analisis ulasan dikelompokkan menjadi Positif, Negatif...	62
<b>Gambar 4. 61</b>	Proses WordCloud untuk melihat kata kata terbanyak pada ulasan .....	62
<b>Gambar 4. 62</b>	Hasil WordCloud Tokopedia .....	63
<b>Gambar 4. 63</b>	Training dan testing data shopee.....	64
<b>Gambar 4. 64</b>	Evaluasi model naive bayes .....	65
<b>Gambar 4. 65</b>	Evaluasi model SVM.....	65
<b>Gambar 4. 66</b>	Evaluasi model KNN.....	66
<b>Gambar 4. 67</b>	Training dan testing data Bukalapak .....	66
<b>Gambar 4. 68</b>	Evaluasi model naive bayes .....	67
<b>Gambar 4. 69</b>	Evaluasi model SVM.....	67
<b>Gambar 4. 70</b>	Evaluasi model KNN.....	68
<b>Gambar 4. 71</b>	Training dan testing data Lazada .....	68
<b>Gambar 4. 72</b>	Evaluasi model naive bayes .....	69
<b>Gambar 4. 73</b>	Evaluasi model SVM.....	69
<b>Gambar 4. 74</b>	Evaluasi model KNN.....	70
<b>Gambar 4. 75</b>	Training dan testing data Tokopedia .....	70
<b>Gambar 4. 76</b>	Evaluasi model naive bayes .....	71
<b>Gambar 4. 77</b>	Evaluasi model SVM.....	71
<b>Gambar 4. 78</b>	evaluasi model KNN .....	72
<b>Gambar 4. 79</b>	Visualisasi berdasarkan pengelompokan ulasan shopee .....	72
<b>Gambar 4. 80</b>	Visualisasi berdasarkan pengelompokan ulasan bukalapak .....	73
<b>Gambar 4. 81</b>	Visualisasi berdasarkan pengelompokan ulasan Lazada .....	74
<b>Gambar 4. 82</b>	Visualisasi berdasarkan pengelompokan ulasan Tokopedia.....	75
<b>Gambar 4. 83</b>	Confusion Matrix berdasarkan data shopee.....	76
<b>Gambar 4. 84</b>	Confusion Matrix berdasarkan data bukalapak.....	77
<b>Gambar 4. 85</b>	Confusion Matrix berdasarkan data Lazada .....	78
<b>Gambar 4. 86</b>	Confusion Matrix berdasarkan data Tokopedia.....	79





## DAFTAR LAMPIRAN

<b>Lampiran 1. Scrapping Shopee</b> .....	87
<b>Lampiran 2. Scrapping Bukalapak</b> .....	88
<b>Lampiran 3. Scrapping Lazada</b> .....	89
<b>Lampiran 4. Scrapping Tokopedia</b> .....	91
<b>Lampiran 5. Hasil Turnitin</b> .....	92
<b>Lampiran 6 Form Bimbingan</b> .....	97

UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA