

PERBANDINGAN ALGORITMA *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN TERHADAP PELAYANAN GORIDE PADA *TWITTER*

SKRIPSI



Oleh
PUTI UTARI MAHARANI
NIM. 18337058

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2023**

PERBANDINGAN ALGORITMA *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN TERHADAP PELAYANAN GORIDE PADA *TWITTER*

SKRIPSI

Diajukan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana Statistika



Oleh
PUTI UTARI MAHARANI
NIM. 18337058

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2023**

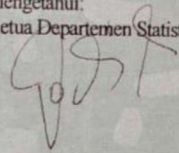
PERSETUJUAN SKRIPSI

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR
MACHINE DALAM ANALISIS SENTIMEN TERHADAP PELAYANAN
GORIDE PADA TWITTER**

Nama : Puti Utari Maharani
NIM : 18337058
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

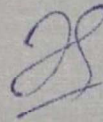
Padang, 8 Juni 2023

Mengetahui:
Ketua Departemen Statistika



Dodi Vionanda, M.Si., Ph.D
NIP. 197806112005011002

Disetujui Oleh:
Pembimbing



Dra. Nonong Amalita, M.Si
NIP. 196906151993032001

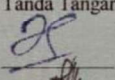

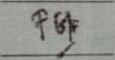
PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Puti Utari Maharani
NIM : 18337058
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN TERHADAP PELAYANAN GORIDE PADA TWITTER

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 8 Juni 2023

Tim Penguji	Nama	Tanda Tangan
Ketua	: Dra. Nonong Amalita, M.Si	
Anggota	: Drs. Atus Amadi Putra, M.Si	
Anggota	: Fadhilah Fitri, S.Si., M.Stat	

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

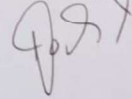
Nama : Puti Utari Maharani
NIM : 18337058
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Perbandingan Algoritma Naive Bayes dan Support Vector Machine dalam Analisis Sentimen terhadap Pelayanan Goride pada Twitter”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,
Ketua Departemen Statistika,



Dodi Vionanda, Ph.D
NIP. 197806112005011002

Saya yang menyatakan,



Puti Utari Maharani
NIM. 18337058

Perbandingan Algoritma *Naive Bayes* dan *Support Vector Machine* dalam Analisis Sentimen Terhadap Pelayanan *Goride* pada *Twitter*

Puti Utari Maharani

ABSTRAK

Goride merupakan inovasi dari teknologi transportasi berbasis aplikasi. Kehadiran *Goride* menimbulkan pro dan kontra di tengah masyarakat. *Goride* menawarkan kemudahan dalam aktivitas sehari-hari, harga yang terjangkau, dan menjamin keamanan bagi pengguna. Namun, *Goride* juga menimbulkan kemacetan di ibu kota karena memanfaatkan bahu jalan sebagai tempat istirahat serta menunggu pesanan dari pengguna aplikasi. Serta menimbulkan konflik horizontal di kalangan pengemudi transportasi karena kehadiran *Goride* mengurangi pendapatan transportasi konvensional. Masyarakat dapat menyampaikan opininya terhadap layanan *Goride* melalui *tweet* pada *Twitter*. *Twitter* sebagai media yang bebas untuk menyampaikan opini. Tujuan penelitian ini adalah ingin mengetahui sentimen masyarakat pada pelayanan *Goride* dan membandingkan tingkat keakuratan algoritma *Naive Bayes* dan SVM.

Algoritma yang digunakan untuk pengklasifikasian *tweet* adalah *Naive Bayes* dan *Support Vector Machine* (SVM). Data yang digunakan pada penelitian ini adalah sekumpulan *tweet* dengan kata kunci *Goride* yang diperoleh dari pengguna *Twitter* di Indonesia. Pengumpulan data dimulai pada tanggal 10 September sampai dengan 10 Oktober 2022.

Hasil dari penelitian ini adalah klasifikasi opini pengguna *Twitter* pada pelayanan *Goride* menjadi dua kelas sentimen yaitu sentimen positif sebanyak 725 *tweet* dan sentimen negatif sebanyak 490 *tweet*. Algoritma yang terbaik dalam mengklasifikasikan teks mengenai *Goride* adalah SVM. Tingkat keakuratan algoritma ini baik dengan nilai akurasi sebesar 81,48% dan nilai AUC 0.832.

Kata Kunci: *Goride*, *Naive Bayes*, SVM, *Twitter*

Comparison of Naive Bayes Algorithm and Support Vector Machine in Sentiment Analysis of *Goride* Services on Twitter

Puti Utari Maharani

ABSTRACT

Goride is an innovation of application-based transport technology. Goride's presence led to pros and cons in the community. Goride offers ease of day activities, affordable prices, and ensures safety for users. However, Goride also caused congestion in the capital because it used the shoulder of the road as a resting place and waited for orders from application users. As well as causing horizontal conflict among transportation drivers due to the presence of Gorides reducing conventional transportation revenue. The public can express their opposition to the Goride service by tweeting on Twitter. Twitter as a free medium to express opinions. The purpose of this study is to find out the public sentiment in Goride services and compare the accuracy of the Naive Bayes and SVM algorithms.

The algorithms used for the classification of tweets are Naive Bayes and Support Vector Machine (SVM). The data used in this study are a set of tweets with Goride keywords obtained from Twitter users in Indonesia. Data collection began from September 10 to October 10, 2022.

The result of this study is the classification of Twitter users' opinions on Goride services into two classes of sentiment: positive sentiment of 725 tweets and negative sentiment of 490 tweets. The best algorithm for classifying tweets about Goride is SVM. The accuracy of this algorithm is good with an accuracy of 81.48% and an AUC of 0.832.

Keywords: *Goride, Naive Bayes, SVM, Twitter*

KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadiran kepada Allah SWT atas berkat rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul “**Perbandingan Algoritma *Naive Bayes* dan *Support Vector Machine* dalam Analisis Sentimen Terhadap Pelayanan *Goride* Pada *Twitter*”**. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang. Dalam penyusunan dan penulisan ini, penulis mendapatkan banyak bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis ucapkan terima kasih kepada:

1. Ibu Dra. Nonong Amalita, M.Si., selaku pembimbing skripsi sekaligus Sekretaris Departemen Statistika FMIPA UNP.
2. Bapak Dodi Vionanda, M.Si., Ph.D., selaku Kepala Departemen Statistika sekaligus Koordinator Program Studi S1 Statistika.
3. Bapak Drs. Atus Amadi Putra, M.Si., selaku dosen penguji skripsi.
4. Ibu Fadhilah Fitri, S.Si., M.Stat selaku dosen penguji skripsi.
5. Bapak dan Ibu Dosen, Staf Pengajar dan Karyawan Departemen Statistika FMIPA UNP yang telah membantu penulis selama menimba ilmu di Program Studi Statistika.
6. Teristimewa kepada kedua orang tua, yang telah memberikan kasih sayang, semangat, nasehat, dukungan kepada penulis dalam menyelesaikan skripsi ini.

7. Kepada Sinta, Indah, Widya, Syifa dan Hana yang telah menemani hari-hari penulis. Terima kasih telah memberikan semangat dan menjadi pendengar yang baik untuk segala keluh kesah penulis.
8. Kepada semua orang tercinta yang selalu memberikan semangat dan mendoakan penulis dalam penyelesaian skripsi.

Semoga rahmat dan kasih sayang Allah SWT selalu tercurah pada kita semua serta usaha dan kerja keras kita bernilai ibadah di hadapan Allah SWT, Amin Ya Rabbal ‘Alamin. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan masukan dan saran dari pembaca agar skripsi ini bermanfaat dikemudian harinya.

Padang, Mei 2023

Penulis

Puti Utari Maharani

DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN.....	viii
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah	5
C. Rumusan Masalah.....	5
D. Tujuan Penelitian	6
E. Manfaat Penelitian	6
BAB II KERANGKA TEORITIS.....	7
A. Kajian Teori.....	7
B. Penelitian Relevan	22
BAB III METODOLOGI PENELITIAN	24
A. Jenis Penelitian.....	24
B. Jenis dan Sumber Data	24
C. Variabel dalam Penelitian	24
D. Teknik Analisis Data	25
BAB IV HASIL DAN PEMBAHASAN	28
A. Hasil.....	28
B. Pembahasan.....	50
BAB V PENUTUP.....	52
A. Kesimpulan.....	52
B. Saran	52
DAFTAR PUSTAKA	54
LAMPIRAN	57

DAFTAR TABEL

Tabel	Halaman
1. <i>Confusion Matrix</i>	17
2. Kriteria Keakuratan Klasifikasi.....	18
3. Variabel Penelitian	24
4. Hasil <i>Crawling Data</i>	31
5. Hasil Proses <i>Case Folding</i>	33
6. Hasil Proses <i>Cleaning</i>	34
7. Hasil Proses <i>Tokenizing</i>	35
8. Hasil Proses <i>Filtering</i>	36
9. Kamus Data	36
10. Hasil Proses Normalisasi Data	37
11. Hasil Proses <i>Stemming</i>	38
12. Hasil Pelabelan Data <i>Tweet</i>	39
13 Proses Perhitungan Manual TF-IDF	40
14. Hasil Pembagian Data <i>Training</i> dan Data <i>Testing</i>	42
15. Kemunculan Frekuensi Kata pada Data <i>Training</i>	42
16. Hasil Perhitungan Peluang Kemunculan Kata (Contoh)	43
17. Matrik $x_i x_j^T$	46
18. Hasil <i>Confusion Matrix</i> pada Algoritma <i>Naive Bayes</i>	47
19. Hasil <i>Confusion Matrix</i> pada Algoritma SVM	48
20. Perbandingan Evaluasi Kinerja Algoritma <i>Naive Bayes</i> dan SVM	49

DAFTAR GAMBAR

Gambar	Halaman
1. Proses SVM dalam Menemukan <i>Hyperplane</i>	14
2. <i>Word Cloud</i>	19
3. Diagram Alir Teknik Analisis Data	27
4. Hasil Persentase Kategori <i>Tweet</i>	28
5. Halaman Twitter Developer	29
6. Token Akses <i>Twitter</i> API	29
7. Hasil proses TF- IDF	40
8. Hasil <i>Word Cloud</i>	50

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Hasil Proses <i>Text Preprocessing</i> dan Pelabelan Data	57
2. Kode Program untuk Pelabelan.....	87
3. Kode Program Menghitung <i>Term</i> Positif dan Negatif.....	87
4. Kode Program untuk Proses <i>Naive Bayes</i> dan SVM	88
5. Kode Program Untuk <i>Confusion Matrix</i>	89

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Perkembangan teknologi semakin berkembang pesat dan semakin canggih seiring dengan perkembangan zaman. Kemajuan teknologi adalah sesuatu yang tidak bisa dihindari dalam kehidupan ini, karena kemajuan teknologi memberikan dampak dan pengaruh di segala aspek kehidupan manusia (Ngafifi, 2014). Perkembangan di bidang teknologi turut berperan serta dalam perkembangan transportasi. Teknologi transportasi merupakan sebuah teknologi yang memfasilitasi pergerakan manusia atau barang dari suatu tempat ke tempat lainnya, dengan menggunakan kendaraan yang digerakkan oleh manusia atau mesin. Dalam era globalisasi, kemajuan teknologi transportasi memberikan manfaat yang signifikan bagi kehidupan masyarakat. Hal ini tercermin dari beragamnya jenis angkutan umum yang tersedia serta ketersediaan angkutan yang diperlukan. (Watung, 2020).

Sarana transportasi yang menjadi pilihan masyarakat salah satunya adalah ojek. Kemampuan yang dimiliki ojek untuk menjangkau tempat- tempat yang sulit diakses oleh mobil atau transportasi umum lainnya, memberikan solusi transportasi alternatif dengan cepat dan mudah. Hal ini penggunaan ojek semakin diminati dikalangan masyarakat. Keberadaan ojek konvensional (ojek pangkalan) terbatas karena hanya bisa ditemukan di satu tempat saja. Di samping itu, tarif ojek yang tidak memiliki standar tertentu dan keamanan yang kurang menjanjikan. Hal

tersebut mengurangi ketertarikan masyarakat untuk menggunakannya (Prasetya dan Legowo, 2016).

Fenomena munculnya transportasi *online* di tengah masyarakat telah menarik perhatian, karena memberikan kemudahan dalam menggunakan layanan transportasi untuk kegiatan sehari - hari. Secara umum transportasi *online* sama halnya dengan kendaraan pribadi baik roda dua maupun roda empat. Perbedaan antara transportasi konvensional dan transportasi *online* adalah pemesanan, karena hanya dilakukan melalui aplikasi *online* dengan sistem Android maupun iOS. Transportasi *online* yang memanfaatkan teknologi aplikasi adalah ojek *online* (Djamhari, 2017 :24).

Jasa transportasi *online* memiliki keunggulan dibandingkan transportasi konvensional yaitu pengguna bisa mengetahui identitas *driver* yang akan mengantarkan pengguna ke tempat tujuan, harga yang relatif murah dan menawarkan fitur diskon juga menarik perhatian masyarakat untuk menggunakan ojek *online* (Anwar, 2017). Kondisi ini memicu ketertarikan masyarakat pada kualitas layanan yang akan diberikan. Kepuasan pelanggan terhadap layanan mempengaruhi kepada perusahaan yang menyediakan jasa transportasi *online*. Saat ini terdapat beberapa aplikasi layanan transportasi *online* yang ada di Indonesia saat ini seperti Gojek, Grab, dan Uber (Djamhari, 2017: 27). Gojek merupakan ojek *online* yang berasal dari Indonesia yang berdiri pada tahun 2010. Gojek memiliki daya tarik tersendiri bagi penggunanya karena merupakan *start up* lokal pertama di Indonesia. Alasan utama masyarakat menggunakan layanan Gojek karena mudah

diakses melalui aplikasi, memiliki harga yang terjangkau, dan menjamin keamanan bagi pengguna (Djamhari, 2017: 32).

Dibalik kehadiran Gojek sebagai *start up* memunculkan permasalahan baru melalui salah satu fitur ojek *online*, yaitu *Goride*. Meningkatnya kemacetan di ibu kota yang disebabkan oleh *driver Goride* karena memanfaatkan bahu jalan sebagai tempat istirahat serta menunggu pesanan dari pengguna aplikasi. Selanjutnya menimbulkan konflik horizontal di kalangan pengemudi transportasi karena kehadiran *Goride* menurunkan pendapatan transportasi konvensional (Djamhari, 2017: 29). Berbagai spekulasi di tengah masyarakat yang muncul terhadap *Goride*. Mendorong masyarakat untuk mengekspresikan pendapat mereka dan memberi penilaian terbuka tentang objek melalui media sosial, salah satunya *Twitter*. Mengutip dari Databoks, pengguna *Twitter* di Indonesia pada bulan Januari 2022 mencapai 18,45 juta. *Twitter* merupakan sebuah situs web yang dioperasikan oleh *Twitter Inc.* yang memungkinkan penggunanya untuk mengirim dan membaca pesan yang disebut *tweet* (Emeraldien, 2019).

Twitter juga memberikan ruang kepada pengguna untuk berkeluh kesah, beropini, mengungkapkan kepuasan atau kekecewaan terhadap sesuatu hal (Rezeki, 2020). Efektivitas dari kehadiran *Goride* tidak terlepas dari permasalahan dan perdebatan. Pro dan kontra terus disampaikan lewat media sosial (Gusnita, 2019). Diperlukan suatu metode untuk mendapatkan informasi dari sekumpulan teks secara efektif dan efisien, salah satunya dengan *text mining*. Salah satu teknik analisis *text mining* yaitu analisis sentimen.

Analisis sentimen digunakan untuk mendeteksi opini berupa pandangan, penilaian, evaluasi, sikap, dan emosi seseorang terhadap suatu topik seperti layanan, produk, atau kegiatan lainnya (Liu, 2012:16). Analisis sentimen bagi pihak perusahaan dapat membantu mengetahui informasi mengenai tanggapan dan sikap dari suatu kelompok atau individu terhadap suatu topik bahasan kontekstual keseluruhan dokumen. Informasi yang didapatkan dari proses tersebut dapat dijadikan sebagai bahan evaluasi untuk melakukan inovasi maupun perbaikan. Algoritma klasifikasi digunakan untuk mengelompokkan data opini ke dalam dua kategori yaitu positif dan negatif. Di antaranya adalah algoritma yang bisa digunakan *Naive Bayes*, *Support Vector Machine*, *K-Nearest-Neighbor* dan *Decision Tree* (Han dan Kamber, 2006:24).

Penelitian ini menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM), dan kemudian membandingkan akurasi kedua algoritma tersebut. Algoritma *Naive Bayes* dan SVM merupakan algoritma yang sama-sama memiliki hasil tingkat akurasi tinggi saat diterapkan dalam suatu kasus (Han dan Kamber, 2006). Namun *Naive Bayes* juga memiliki kekurangan yaitu sensitif terhadap fitur yang terlalu banyak sehingga membuat akurasi menjadi rendah (Chen, 2009). Sedangkan, SVM memiliki kekurangan yaitu salah pemilihan parameter dapat mempengaruhi hasil tingkat akurasi (Nugroho, 2003).

Analisis sentimen dengan metode *Naive Bayes* dan SVM telah banyak digunakan dalam berbagai penelitian sebelumnya, seperti penelitian yang dilakukan Alizah, *et al* (2020) mengenai klasifikasi komentar pada *Twitter* tentang isu “*lockdown*”. Hasil penelitian tersebut diperoleh metode SVM memiliki nilai akurasi

lebih efektif yaitu sebesar 86,54% dibandingkan dengan metode *Naive Bayes* sebesar 84,69%. Selanjutnya penelitian Dwianto dan Sadikin, (2021) tentang perbandingan metode *Naive Bayes* dan SVM mengenai penggunaan dua transportasi *online* yaitu GrabId dan Gojek. Hasil penelitian tersebut diperoleh metode SVM memiliki akurasi terbaik dengan nilai 84,08% untuk GrabId dan 69,50% untuk Gojek dibandingkan dengan metode *Naive Bayes*.

Berdasarkan paparan di atas maka akan dilakukan penelitian analisis sentimen yang berjudul **“Perbandingan Algoritma *Naive Bayes* dan *Support Vector Machine* dalam Analisis Sentimen terhadap Pelayanan *Goride* pada *Twitter*”**

B. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan yaitu data *tweet* berbahasa Indonesia dari pengguna *Twitter* terkait *hashtag* “*Goride*”.
2. Data *tweet* diambil dari tanggal 10 September sampai 10 Oktober 2022.

C. Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka didapatkan rumusan masalah yaitu sebagai berikut.

1. Bagaimana klasifikasi opini pengguna *Twitter* terhadap pelayanan *Goride* menggunakan analisis sentimen?
2. Bagaimana hasil akurasi algoritma *Naive Bayes* dalam pengklasifikasian sentimen pengguna *Twitter* terhadap pelayanan *Goride*?
3. Bagaimana hasil akurasi algoritma *Support Vector Machine* dalam pengklasifikasian sentimen pengguna *Twitter* terhadap pelayanan *Goride*?

4. Algoritma manakah yang terbaik diantara *Naive Bayes* dan *Support Vector Machine* dalam pengklasifikasian sentimen terhadap pelayanan *Goride* di media sosial *Twitter* ?

D. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut.

1. Mengetahui klasifikasi opini pengguna *Twitter* terhadap pelayanan *Goride* menggunakan analisis sentimen.
2. Mengetahui hasil akurasi algoritma *Naive Bayes* dalam pengklasifikasian sentimen pengguna *Twitter* terhadap pelayanan *Goride*.
3. Mengetahui hasil akurasi algoritma *Support Vector Machine* dalam pengklasifikasian sentimen pengguna *Twitter* terhadap pelayanan *Goride*.
4. Mengetahui algoritma terbaik di antara *Naive Bayes* dan *Support Vector Machine* dalam pengklasifikasian sentimen terhadap pelayanan *Goride* di media sosial *Twitter*.

E. Manfaat Penelitian

1. Bagi peneliti, dapat menambah pengetahuan peneliti mengenai analisis sentimen menggunakan algoritma *Naive Bayes* dan *Support Vector Machine*.
2. Bagi mahasiswa, diharapkan dapat menambah pengetahuan mengenai analisis sentimen dan dapat dijadikan referensi untuk penelitian selanjutnya
3. Bagi perusahaan, hasil sentimen dapat dijadikan sebagai bahan evaluasi guna untuk pengembangan layanan *Goride*.