

**PERBANDINGAN KINERJA *MACHINE LEARNING*  
BERBASIS ALGORITMA *SUPPORT VECTOR MACHINE* DAN  
*NAIVE BAYES***

(Studi Kasus: Data Tanggapan Mengenai Traveloka Melalui Media Sosial  
*Twitter*)

**TUGAS AKHIR**



**Evi Fitria Umi Latifah**

**14 611 064**

**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

**PERBANDINGAN KINERJA *MACHINE LEARNING*  
BERBASIS ALGORITMA *SUPPORT VECTOR MACHINE* DAN  
*NAIVE BAYES***

(Studi Kasus: Data Tanggapan Mengenai Traveloka Melalui Media Sosial  
*Twitter*)

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Jurusan Statistika**



**Evi Fitria Umi Latifah**

**14 611 064**

**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA**

**2018**

## HALAMAN PERSETUJUAN PEMBIMBING

### TUGAS AKHIR

Judul : Perbandingan Kinerja *Machine Learning* Berbasis  
Algoritma *Support Vector Machine* dan *Naïve Bayes*  
(Studi Kasus: Data Tanggapan Mengenai Traveloka  
Melalui Media Sosial *Twitter*)

Nama Mahasiswa : Evi Fitria Umi Latifah

Nomor Mahasiswa : 14 611 064



TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK

DIUJIKAN

Yogyakarta, 29 Mei 2018

Pembimbing

A handwritten signature in black ink, appearing to read 'Kariyam', is written over a faint circular stamp. The signature is fluid and cursive.

(Kariyam, S.Si., M.Si.)

**HALAMAN PENGESAHAN  
TUGAS AKHIR**

**PERBANDINGAN KINERJA MACHINE LEARNING BERBASIS  
ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES**

(Studi Kasus: Data Tanggapan Mengenai Traveloka Melalui Media Sosial



**Nama Mahasiswa : Evi Fitria Umi Latifah**

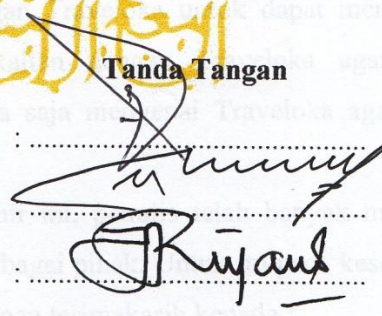
**Nomor Mahasiswa : 14 611 064**

**TUGAS AKHIR INI TELAH DIUJIKAN  
PADA TANGGAL 6 JUNI 2018**

**Nama Penguji**

1. Ir. Sukirman, M.M
2. Muhammad Muhajir, S.Si., M.Sc
3. Kariyam, S.Si., M.Si

**Tanda Tangan**



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Dr. Alhwar, M.Sc., Ph.D)

## KATA PENGANTAR



*Assalamu'alaikum Wr. Wb.*

Alhamdulillah rabbil'aalamiin, puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya kepada penulis sehingga penulis diberikan keimanan, kekuatan, kesehatan, kesabaran, kelancaran, serta keselamatan selama penyusunan Tugas Akhir ini hingga laporan Tugas Akhir ini dapat terselesaikan. Shalawat serta salam semoga selalu tercurah kepada junjungan kita Nabi Muhammad SAW beserta keluarga dan para pengikut-pengikutnya.

Tugas Akhir yang berjudul **“Perbandingan Kinerja *Machine Learning* Berbasis Algoritma *Support Vector Machine* dan *Naive Bayes*”** dengan studi kasus Data Tanggapan Mengenai Traveloka Melalui Media Sosial *Twitter* ini disusun sebagai salah satu persyaratan yang harus dipenuhi dalam menyelesaikan Program Strata Satu di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, serta memberikan pengetahuan khususnya pihak Traveloka dan Pelanggan Traveloka untuk dapat mengetahui tanggapan positif apa saja yang berkaitan dengan Traveloka agar dapat dipertahankan dan tanggapan negatif apa saja mengenai Traveloka agar dapat dilakukan perbaikan.

Dalam menyelesaikan Tugas Akhir ini, penulis telah banyak mendapat bimbingan, masukan dan bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penulis bermaksud menyampaikan ucapan terimakasih kepada :

1. Bapak Drs. Allwar., M.Sc., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
2. Bapak Dr. RB. Fajriya Hakim, S.Si., M.Si. selaku Ketua Jurusan Statistika beserta seluruh jajarannya.

3. Ibu Kariyam, S.Si., M.Si. selaku dosen pembimbing Tugas Akhir yang telah membimbing dan memberikan arahan kepada penulis selama penyusunan Tugas Akhir ini.
4. Dosen-dosen Statistika Universitas Islam Indonesia yang telah mendidik dan memberikan ilmu kepada penulis.
5. Bapak, Ibu, Simbah, Budhe dan Adik tercinta yang selalu mendoakan dan memberi dukungan kepada penulis baik moril maupun materiil. *I love you.*
6. Agus Mardika Najib, terimakasih atas do'a, semangat, dan dukungannya.
7. Sahabat-sahabat seperjuangan Nabilla, Ulfa, Avit, Indri, Ita, Emil, Kia, Yulina, Woro, Rabi, Wilda, Hafizan, Riza, dan Kiki yang selalu memberikan semangat, berbagi ilmu dan pengalaman.
8. Kakak tingkat Mbak Yulinda dan Mas Ulwan yang bersedia berbagi ilmu dan pengalaman.
9. Sahabat Statistika 2014 yang telah memberikan motivasi, dukungan, dan semangat kepada penulis.
10. Semua pihak yang turut membantu penulis dalam penyelesaian Tugas Akhir ini, penulis mengucapkan terima kasih semoga Allah SWT selalu memberikan rahmat dan anugerah-Nya kepada mereka semua tanpa henti. Aamiin ya robbal 'alamiin.

Penulis menyadari bahwa Tugas Akhir ini masih banyak kekurangan dan jauh dalam kesempurnaan karena keterbatasan pengetahuan yang dimiliki penulis. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi berbagai pihak. Akhir kata, penulis mohon maaf apabila dalam penyusunan Tugas Akhir ini terdapat banyak kesalahan.

***Wassalamu'alaikum Wr. Wb.***

Yogyakarta, 29 Mei 2018

Penulis

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN PEMBIMBING .....	i
HALAMAN PENGESAHAN.....	ii
KATA PENGANTAR .....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR LAMPIRAN.....	xiii
PERNYATAAN.....	xiv
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian .....	4
1.5. Manfaat Penelitian .....	5
BAB II TINJAUAN PUSTAKA.....	6
BAB III LANDASAN TEORI .....	11
3.1. Tanggapan.....	11
3.2. Traveloka.....	12
3.3. <i>Twitter</i> .....	12
3.4. Statistika Deskriptif.....	14
3.5. <i>Data Mining</i> .....	16
3.6. <i>Machine Learning</i> .....	18
3.7. <i>Text Mining</i> .....	19
3.7.1. <i>Case Folding</i> .....	20

3.7.2.	<i>Tokenizing dan Filtering</i> .....	20
3.7.3.	<i>Stemming</i> .....	21
3.8.	Pembobotan Kata ( <i>Term Weighting</i> ).....	22
3.8.1.	Metode TF-IDF ( <i>Term Frequency-Inverse Document Frequency</i> ) .....	22
3.8.2.	Simulasi Metode TF-IDF( <i>Term Frequency-Inverse Document Frequency</i> ) .....	23
3.9.	Klasifikasi .....	25
3.9.1.	Gambaran Fase <i>Learning</i> dan Fase <i>Testing</i> .....	25
3.9.2.	Data <i>Learning</i> .....	26
3.9.3.	Simulasi Perhitungan Data <i>Learning</i> dengan Rumus Slovin .....	26
3.9.4.	<i>Support Vector Machine</i> pada Data Terpisah Secara Linear .....	27
3.9.5.	Simulasi <i>Support Vector Machine</i> pada Data Terpisah Secara Linear .....	28
3.9.6.	Klasifikasi dengan <i>Naive Bayes</i> .....	30
3.9.7.	Simulasi Klasifikasi dengan <i>Naive Bayes</i> .....	32
3.10.	Evaluasi Performa .....	35
3.11.	<i>Word Cloud</i> .....	37
3.12.	Asosiasi ( <i>Association Rules</i> ) .....	37
BAB IV	METODOLOGI PENELITIAN .....	39
4.1.	Populasi dan Sampel Penelitian .....	39
4.2.	Tempat dan Waktu Penelitian .....	39
4.3.	Variabel Penelitian .....	40
4.4.	Metode Analisis .....	40
4.5.	Tahapan Penelitian .....	40
BAB V	HASIL DAN PEMBAHASAN .....	44
5.1.	Analisis Deskriptif .....	44
5.2.	Tahap <i>Preprocessing</i> .....	45
5.2.1	<i>Case Folding</i> .....	46



5.2.2	<i>Tokenizing dan Filtering</i> .....	47
5.2.3	<i>Stemming</i> .....	48
5.3.	Klasifikasi .....	49
5.3.1	Data <i>Training</i> dan Data <i>Testing</i> .....	49
5.3.2	Perbandingan Kinerja <i>Support Vector Machine</i> dan <i>Naive Bayes</i> .....	52
5.3.3	Perbandingan Kinerja Klasifikasi Tanggapan pada bulan Maret dan bulan April .....	68
5.4.	<i>Word Cloud</i> .....	69
5.5.	Asosiasi Kata.....	71
BAB VI KESIMPULAN DAN SARAN .....		75
6.1.	Kesimpulan .....	75
6.2.	Saran.....	75
DAFTAR PUSTAKA .....		77
LAMPIRAN		

## DAFTAR TABEL

Nomor	Judul	Halaman
1	Penelitian Terdahulu	8
2	Penelitian Terdahulu (Lanjutan 1)	9
3	Penelitian Terdahulu (Lanjutan 2)	10
4	Nilai UAS Mahasiswa	15
5	Nilai UTS dan UAS Mahasiswa	15
6	Nilai UTS dan UAS Mahasiswa	15
7	Hasil Perhitungan Bobot Masing-masing Dokumen	24
8	Data Simulasi SVM Linear	29
9	Frekuensi Kemunculan Kata	33
10	Probabilitas Kata Kelas Positif	34
11	Probabilitas Kata Kelas Negatif	34
12	Nilai Probabilitas Tanggapan Baru	35
13	<i>Confusion Matrix</i>	36
14	Interpretasi Nilai Koefisien Korelasi	38
15	Jumlah Sampel ( <i>Data Training</i> )	39
16	Variabel dan Definisi Operasional Variabel	40
17	<i>Tweet</i> Tanggapan Bulan Maret dan April 2018	46
18	<i>Tweet</i> Tanggapan Bulan Maret dan April 2018 Melalui Proses <i>Case Folding</i>	47
19	<i>Tweet</i> Tanggapan Bulan Maret dan April 2018 Melalui Proses <i>Case Folding</i> , <i>Tokenizing</i> , dan <i>Filtering</i>	48
20	<i>Tweet</i> Tanggapan Bulan Maret dan April 2018 Melalui Proses <i>Case Folding</i> , <i>Tokenizing</i> , <i>Filtering</i> , dan <i>Stemming</i>	48
21	<i>Tweet</i> Tanggapan Bulan Maret dan April 2018 Melalui Proses <i>Case Folding</i> , <i>Tokenizing</i> , <i>Filtering</i> , dan <i>Stemming</i> (Lanjutan)	49

22	<i>Tweet</i> Tanggapan Bulan Maret dan April 2018 Beserta Label	50
23	Pembagian Data <i>Training</i> dan Data <i>Testing Tweet</i> Bulan Maret-April 2018	50
24	Pembagian Data <i>Training</i> dan Data <i>Testing Tweet</i> Bulan Maret-April 2018 (Lanjutan)	51
25	Pembagian Data <i>Training</i> dan Data <i>Testing Tweet</i> Bulan Maret 2018	51
26	Pembagian Data <i>Training</i> dan Data <i>Testing Tweet</i> Bulan April 2018	51
27	<i>Confusion Matrix</i> Algoritma SVM 1	52
28	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 1	52
29	<i>Confusion Matrix</i> Algoritma SVM 2	54
30	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 2	54
31	<i>Confusion Matrix</i> Algoritma SVM 3	55
32	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 3	55
33	<i>Confusion Matrix</i> Algoritma SVM 4	56
34	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 4	56
35	<i>Confusion Matrix</i> Algoritma SVM 5	57
36	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 5	57
37	<i>Confusion Matrix</i> Algoritma SVM 6	58
38	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 6	58
39	<i>Confusion Matrix</i> Algoritma SVM 7	59
40	<i>Confusion Matrix</i> algoritma <i>Naive Bayes</i> 7	59
41	<i>Confusion Matrix</i> Algoritma SVM 8	60
42	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 8	60
43	<i>Confusion Matrix</i> Algoritma SVM 9	61
44	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 9	61
45	<i>Confusion Matrix</i> Algoritma SVM 10	62
46	<i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> 10	62

47	Perbandingan Tingkat Akurasi Algoritma SVM dan <i>Naive Bayes</i>	63
48	Pengacakan Kombinasi 1 sampai Kombinasi 4	64
49	Pengacakan Kombinasi 5 sampai Kombinasi 7	65
50	Pengacakan Kombinasi 8 sampai Kombinasi 10	66
51	Perbandingan Rata-Rata Akurasi SVM dan <i>Naive Bayes</i>	67
52.	Perbandingan Tingkat Akurasi Algoritma SVM dan <i>Naive Bayes</i> Data Bulan Maret dan April	68
53	Asosiasi Data Keseluruhan Bulan Maret-April	72
54.	Asosiasi Data Bulan Maret	73
55	Asosiasi Data Bulan April	74

## DAFTAR GAMBAR

Nomor	Judul	Halaman
1	Rata-Rata Waktu Mengakses Media Sosial	1
2	Skema Proses Tanggapan	11
3	Produk dan Layanan Traveloka	12
4	Tahapan dalam Proses <i>Knowledge Discovery in Database</i> (KDD)	17
5	Ilustrasi Cakupan <i>Data Mining</i>	18
6	Tahapan <i>Preprocessing</i> pada <i>Text Mining</i>	20
7	Proses <i>Case Folding</i>	20
8	Proses <i>Tokenizing</i> dan <i>Filtering</i>	21
9	Proses <i>Stemming</i>	21
10	Fase <i>Learning</i>	25
11	Fase <i>Testing</i>	26
12	Ilustrasi SVM Menemukan <i>Hyperplane</i> Terbaik	27
13	Grafik <i>Hyperplane</i>	28
14	<i>Word Cloud</i> Data Ahok 17 April 2017	37
15	Tahapan Penelitian	41
16	Tahapan Penelitian (Lanjutan)	42
17	Persentase dan Jumlah Tanggapan tentang Traveloka	44
18	Tanggapan tentang Traveloka pada Bulan Maret dan Bulan April	45
19	Grafik Perbandingan Akurasi SVM dan <i>Naive Bayes</i>	67
20	<i>Word Cloud</i> Data Keseluruhan	69
21	<i>Word Cloud</i> Data Bulan Maret	70
22	<i>Word Cloud</i> Data Bulan April	71

## DAFTAR LAMPIRAN

Lampiran 1	<i>Output</i> Klasifikasi SVM Pada Program R
Lampiran 2	<i>Output</i> Klasifikasi <i>Naive Bayes</i> Pada Program R
Lampiran 3	<i>Script R</i> <i>Web Scraping</i>
Lampiran 4	<i>Packages R</i> yang Digunakan
Lampiran 5	<i>Script R</i> <i>Preprocessing Data</i> dengan <i>Text Mining</i>
Lampiran 6	<i>Script R</i> Klasifikasi dengan SVM
Lampiran 7	<i>Script R</i> Klasifikasi dengan <i>Naive Bayes</i>
Lampiran 8	<i>Script R</i> <i>Word Cloud</i> dan Asosiasi Kata
Lampiran 9	Makalah

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu di dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 29 Mei 2018



Penulis

**PERBANDINGAN KINERJA *MACHINE LEARNING* BERBASIS  
ALGORITMA *SUPPORT VECTOR MACHINE* DAN *NAIVE BAYES***  
(Studi Kasus: Data Tanggapan Mengenai Traveloka Melalui Media Sosial  
*Twitter*)

Oleh: Evi Fitria Umi Latifah  
Program Studi Statistika Fakultas MIPA  
Universitas Islam Indonesia

**INTISARI**

*Seiring kemajuan teknologi, data yang dapat dikumpulkan bukan hanya data yang berbentuk numerik namun juga data yang berbentuk teks. Dalam melakukan analisis terhadap data teks digunakan text mining, oleh karena text mining lebih bersifat subjek umum maka diperlukan klasifikasi. Salah satu algoritma dalam machine learning yang dapat digunakan untuk melakukan klasifikasi teks adalah Support Vector Machine (SVM) dan Naive Bayes. SVM bekerja dengan cara mencari hyperplane terbaik yang memisahkan dua buah kelas, sedangkan Naive Bayes bekerja dengan cara mencari nilai probabilitas bersyarat terbesar. Berdasarkan perbedaan cara kerja kedua algoritma tersebut, peneliti ingin membandingkan kinerja kedua algoritma tersebut pada studi kasus baru yaitu data tanggapan mengenai Traveloka yang diperoleh melalui media sosial Twitter pada bulan Maret dan bulan April 2018. Berdasarkan tingkat akurasi terbesar dengan menggunakan 10 kombinasi dan pada masing-masing kombinasi dilakukan 10 kali pengacakan, diperoleh kesimpulan bahwa algoritma SVM memiliki tingkat akurasi lebih tinggi dibandingkan algoritma Naive Bayes dengan rata-rata besarnya tingkat akurasi sebesar 79,14% pada SVM dan 75,61% pada Naive Bayes.*

**Kata Kunci:** *Machine learning, Support Vector Machine, Naive Bayes, Twitter, Traveloka, Sentimen, Akurasi, Asosiasi, Word cloud*



**COMPARISON OF MACHINE LEARNING PERFORMANCE BASED ON  
SUPPORT VECTOR MACHINE AND NAIVE BAYES ALGORITHM**

*(Case Study: Response Data About Traveloka Through Social Media Twitter)*

By: Evi Fitria Umi Latifah

*Department of Statistics, Faculty of Mathematics and Science*

*Islamic University of Indonesia*

**ABSTRACT**

*As technology advances, data that can be collected is not just numerical data but also text-shaped data. The analysis of text data used text mining, because text mining is more general subject then required classification. One of the algorithms in machine learning that can be used to perform text classification is Support Vector Machine (SVM) and Naive Bayes. SVM works by finding the best hyperplane that separates two classes, while Naive Bayes works by finding the largest conditional probability value. Based on differences in the workings of the two algorithms, the researchers wanted to compare the performance of the two algorithms in a new case study of responses data on Traveloka obtained via Twitter social media in March and April 2018. Based on the greatest accuracy level using 10 combinations and at each combination was performed 10 times randomization, the conclusion is that the SVM algorithm has a higher accuracy than Naive Bayes algorithm with average accuracy rate of 79.14% in SVM and 75, 61% in Naive Bayes.*

**Keywords:** *Machine learning, Support Vector Machine, Naive Bayes, Twitter, Traveloka, Sentiment, Accuracy, Association, Word cloud*

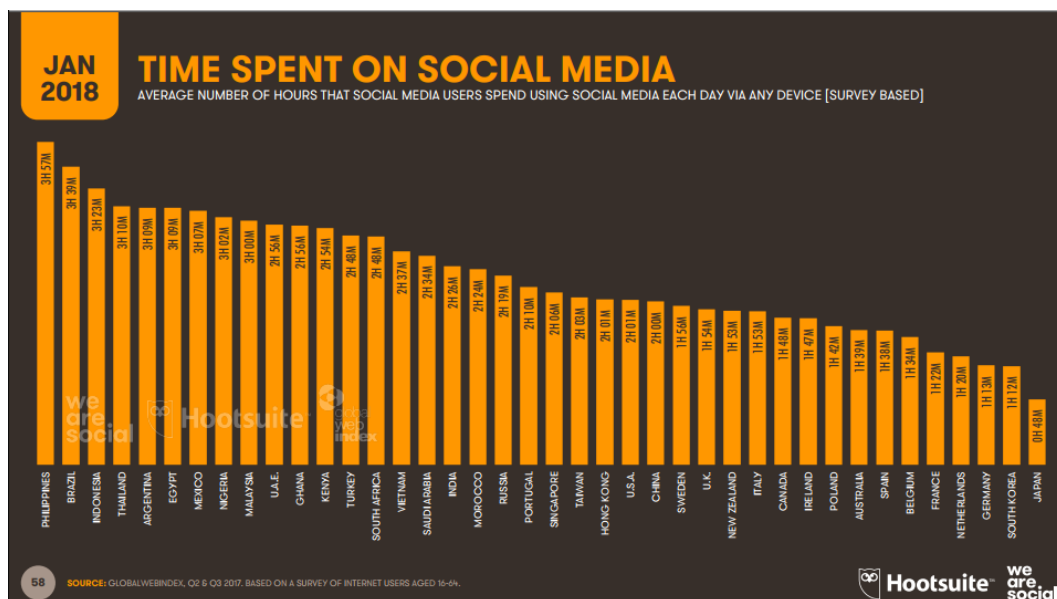
# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Kemajuan teknologi yang semakin pesat dalam dunia digital membuat jumlah informasi yang dihasilkan semakin besar. Informasi tersebut dapat berupa angka maupun teks, salah satu informasi yang berupa teks adalah tanggapan. Berdasarkan hasil sensus penduduk tahun 2010 diketahui jumlah penduduk Indonesia pada tahun 2010 sebanyak 237.641.326 jiwa (Badan Pusat Statistika, 2010). Bagi dunia digital, hal tersebut menjadikan Indonesia sebagai pasar yang sangat potensial, terutama dalam media sosial. Berdasarkan penelitian pada bulan Januari 2018 yang dilakukan *We Are Social*, sebuah perusahaan media asal Inggris yang bekerja sama dengan *Hootsuite* memperoleh hasil bahwa dalam hal lama waktu yang dihabiskan untuk mengakses media sosial, Indonesia menempati peringkat ketiga dengan durasi rata-rata yang dihabiskan 3 jam 23 menit sehari.

Berikut grafik yang menunjukkan rata-rata waktu yang dihabiskan untuk mengakses media sosial.



(Sumber: Pertiwi, 2018)

**Gambar 1.** Rata-Rata Waktu Mengakses Media Sosial

Selain itu, seperti yang tercantum dalam kompas.com (Pertiwi, 2018) berdasarkan laporan berjudul "*Essential Insights Into Internet, Social Media, Mobile, and E-Commerce Use Around The World*" yang diterbitkan tanggal 30 Januari 2018, pengguna aktif sosial media di Indonesia mencapai 130 juta dari total populasi Indonesia sebanyak 265,4 juta jiwa atau mencapai 49 persen.

Besarnya persentase pengguna media sosial dan lama waktu yang dihabiskan untuk mengakses sosial media tersebut tentu akan menghasilkan informasi yang tak terhitung jumlahnya yang dituangkan kedalam media sosial. Salah satu informasi yang dituangkan ke dalam media sosial adalah pendapat atau tanggapan seseorang terhadap suatu hal. Dalam menggali informasi yang terkandung dalam ribuan tanggapan tersebut maka perlu dilakukan sebuah analisis.

Tanggapan yang terdapat dalam sosial media merupakan data yang berbentuk teks, maka digunakan *text mining* dalam melakukan analisis. Menurut Ratna Maria (Damanik, 2014) *text mining* memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antara dokumen.

Dalam *text mining* suatu kelas lebih bersifat area subjek umum, maka diperlukan klasifikasi. Klasifikasi merupakan suatu pekerjaan menilai suatu objek data untuk memasukannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Klasifikasi disini akan mengelompokkan *tweet* atau tanggapan mana yang akan masuk ke dalam kelas tanggapan positif dan kelas tanggapan negatif, yang nantinya dapat dijadikan bahan pertimbangan untuk melakukan perbaikan.

Salah satu media sosial yang sering digunakan untuk menuangkan tanggapan atau keluhan adalah *Twitter*. Tanggapan yang terdapat di dalam *Twitter* merupakan tanggapan yang berisi kalimat atau kata yang tidak terstruktur dan mengandung *noise* yang besar, maka sebelum dilakukan klasifikasi diperlukan tahapan *preprocessing*. Menurut Raymond J. Mooney (Damanik, 2014), tahapan *preprocessing* meliputi *case folding*, *tokenizing*, *filtering*, dan *stemming*.

Terdapat beberapa algoritma dalam klasifikasi, antara lain *Support Vector Machine* dan *Naive Bayes*. Penelitian dengan algoritma *Support Vector Machine* pernah dilakukan (Naradhipa & Purwarianti, 2012) dalam klasifikasi pesan berbahasa Indonesia pada sosial media menghasilkan kesimpulan bahwa metode SVM memiliki tingkat akurasi paling tinggi dalam hal klasifikasi teks. Jurnal penelitian (Santoso, Ratnawati, & Indriati, 2014) dalam pengkategorian buku komputer berbahasa Indonesia berdasarkan judul dan sinopsis, menghasilkan kesimpulan bahwa metode *Naive Bayes* memiliki tingkat akurasi tertinggi dibandingkan metode KNN dan metode gabungan *k-means* dan LVQ dalam mengkategorikan buku komputer berbahasa Indonesia. Pada dasarnya SVM bekerja dengan cara mencari *hyperplane* terbaik yang memisahkan dua buah kelas, sedangkan *Naive Bayes* bekerja dengan cara mencari nilai probabilitas bersyarat terbesar dari masing-masing kelas.

Berdasarkan uraian tersebut, maka penulis ingin membandingkan kinerja dari kedua metode tersebut apabila diterapkan dalam suatu studi kasus yaitu tanggapan mengenai traveloka. Traveloka merupakan salah satu *online travel agent* terbaik di Indonesia seperti yang tercantum dalam [strategimanajemen.net](http://strategimanajemen.net) (Antariksa, 2017) “Traveloka yang didirikan pada tahun 2012 kini telah menjelma menjadi situs pemesanan hotel dan tiket pesawat terbang dan hotel no. 1 se-Indonesia”, namun terdapat berita dalam [tirto.id](http://tirto.id) (Zuhra, 2017) yang menyebutkan bahwa Tiket.com diakusisi Bliblicom, bersiap menyalip Traveloka. Oleh karena itu perlu dilakukan suatu analisis agar Traveloka tetap menjadi situs pemesanan hotel dan tiket pesawat terbang no.1 di Indonesia. Salah satu cara yang dilakukan adalah mengetahui tanggapan mengenai Traveloka.

Oleh karena itu, berdasarkan uraian di atas judul yang diambil adalah “Perbandingan Kinerja *Machine Learning* Berbasis Algoritma *Support Vector Machine* dan *Naive Bayes* (Studi kasus: Data Tanggapan Mengenai Traveloka Melalui Media Sosial *Twitter*)”.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, permasalahan yang akan dikaji dalam penelitian ini adalah sebagai berikut.

1. Bagaimana perbandingan kinerja *machine learning* dengan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi tanggapan mengenai Traveloka?
2. Apakah terdapat perbedaan kinerja algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi terhadap tanggapan mengenai Traveloka pada bulan Maret dan bulan April 2018?
3. Bagaimana asosiasi kata berdasarkan hasil tampilan *word cloud* data tanggapan mengenai Traveloka?

## 1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data tanggapan masyarakat yang digunakan dalam penelitian ini diambil dari postingan media sosial *Twitter* selama satu minggu pada masing-masing bulan Maret dan bulan April 2018 dengan menggunakan kata kunci “traveloka”.
2. Alat analisis yang digunakan adalah *software Microsoft PowerPoint 2007*, *Microsoft Excel 2007* dan *R 3.4.1*.

## 1.4. Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah sebagai berikut.

1. Mengetahui perbandingan kinerja *machine learning* dengan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi data tanggapan mengenai Traveloka.
2. Mengetahui perbedaan kinerja algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi terhadap tanggapan mengenai Traveloka pada bulan Maret dan bulan April 2018.
3. Mengetahui hasil asosiasi kata berdasarkan hasil tampilan *word cloud* data tanggapan mengenai Traveloka.

### 1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

1. Dengan diketahuinya perbandingan kinerja *machine learning* dengan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi, maka dapat diketahui algoritma yang sesuai untuk melakukan klasifikasi data tanggapan mengenai *online travel agent* Traveloka.
2. Dengan diketahuinya kinerja algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi terhadap tanggapan mengenai Traveloka pada bulan Maret dan bulan April 2018, maka dapat diketahui apakah terdapat perbedaan kinerja *machine learning* untuk melakukan klasifikasi tanggapan mengenai Traveloka pada data secara keseluruhan, data bulan Maret dan data bulan April 2018.
3. Dengan diketahuinya hasil asosiasi kata berdasarkan hasil tampilan *word cloud* maka:
  - a. Dapat mengetahui kata apa yang sering muncul dalam tanggapan positif maupun tanggapan negatif.
  - b. Dapat mengetahui kata apa yang menyertai kata dominan dalam tanggapan, baik tanggapan positif maupun tanggapan negatif.

## **BAB II**

### **TINJAUAN PUSTAKA**

Dalam bab ini memuat penelitian terdahulu yang berguna sebagai bahan kajian bagi penulis dalam penelitian yang dilakukan.

Beberapa penelitian yang pernah dilakukan terkait dengan studi kasus yang penulis gunakan dan metode klasifikasi yang penulis gunakan, antara lain penelitian (Naradhipa & Purwarianti, 2012) dengan judul “*Sentiment Classification for Indonesian Message in Social Media*” yang menghasilkan kesimpulan bahwa metode *Support Vector Machine* memiliki tingkat akurasi paling tinggi dalam hal klasifikasi teks dengan akurasi 86,66%.

Penelitian (Santoso, Ratnawati, & Indriati, 2014) dengan judul “Perbandingan Kinerja Metode *Naive Bayes*, *K-Nearest Neighbor*, dan Metode Gabungan *K-Means* dan LVQ dalam Pengkategorian Buku Komputer Berbahasa Indonesia Berdasarkan Judul dan Sinopsis” menghasilkan kesimpulan bahwa metode terbaik dalam mengkategorikan buku komputer berbahasa Indonesia adalah metode *Naive Bayes*.

Penelitian (Saleh, 2015) dengan judul “Implementasi Metode Klasifikasi *Naive Bayes* Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga” menghasilkan keakuratan sebesar 78.33%.

Penelitian (Ulwan, 2016) dengan judul “*Pattern Recognition* Pada *Unstructured* Data Teks Menggunakan *Support Vector Machine* dan *Association* (Studi Kasus: Portal Layanan Aspirasi dan Pengaduan *Online* Rakyat)” menghasilkan kesimpulan bahwa metode *Support Vector Machine* menghasilkan akurasi sebesar 96,7% dalam melakukan klasifikasi dan dilakukan visualisasi dan asosiasi untuk menemukan informasi yang dianggap penting.

Penelitian (Farid, 2017) dengan judul “Pengaruh Kualitas Pelayanan dan Persepsi Harga Terhadap Kepuasan Pelanggan dalam Meningkatkan Loyalitas Pelanggan Pengguna Aplikasi Traveloka Berbasis Internet (Studi Pada Pengguna Aplikasi *Online* Traveloka Universitas Lampung)” menghasilkan kesimpulan

bahwa variabel kualitas pelayanan dan variabel persepsi harga berpengaruh signifikan terhadap kepuasan pelanggan pengguna aplikasi Traveloka, loyalitas pelanggan pengguna aplikasi Traveloka, serta loyalitas pelanggan dalam meningkatkan loyalitas pelanggan pengguna aplikasi Traveloka.

Penelitian (Nuraini & Mudiantono, 2017) dengan judul “Faktor-Faktor yang Mempengaruhi Kepuasan Pelanggan dan Dampaknya Terhadap Minat Beli Ulang Pada Situs Traveloka (Studi Pada Konsumen Traveloka di Kota Semarang)”. Penelitian tersebut menghasilkan kesimpulan bahwa faktor yang mempengaruhi kepuasan pelanggan adalah kualitas pelayanan dengan baik, sedangkan proses yang paling berpengaruh terhadap minat beli ulang adalah dengan kualitas pelayanan yang mempengaruhi citra merek sebagai salah satu penentu keberhasilan untuk dapat meningkatkan minat beli ulang.

Penelitian (Risyanidi & Zuliestiana, 2017) dengan judul “Pengaruh Kualitas *Website* Traveloka Terhadap Kepuasan Pengguna” memberikan kesimpulan bahwa secara parsial maupun simultan variabel yang berpengaruh signifikan terhadap kualitas website traveloka atau kepuasan pengguna adalah variabel *Webqual* dan sub variabel *Usability*, *Information Quality*, *Service Interaction Quality*.

Penelitian (Syafitri, 2017) dengan judul “Penerapan Teknik *Web Scraping* dan Klasifikasi Sentimen Menggunakan Metode *Naive Bayes Classifier* dan Asosiasi (Studi Kasus: Data Ulasan *The House Of Raminten* Pada Situs TripAdvisor)”. Penelitian tersebut menghasilkan kesimpulan bahwa metode *Naive Bayes Classifier* menghasilkan tingkat akurasi 84% pada ulasan bahasa Indonesia dan pada ulasan bahasa Inggris sebesar 77%.

Selanjutnya, penelitian (Ghifary, 2018) dengan judul “Analisis Kualitas Layanan pada Perusahaan *E-Commerce* Traveloka” menghasilkan kesimpulan bahwa *ESQual* berpengaruh signifikan terhadap *Perceived Value*, *ESQual* berpengaruh signifikan terhadap *Loyalty Intentions*, sedangkan *ESQual* melalui *Perceived Value* berpengaruh signifikan terhadap *Loyalty Intentions*.



Rangkuman penelitian terdahulu tersebut dapat dilihat pada **Tabel 1**, **Tabel 2**, dan **Tabel 3**.

**Tabel 1.** Penelitian Terdahulu

No.	Nama/Tahun	Metode	Judul Penelitian	Hasil Penelitian
1.	(Naradhipa & Purwarianti, 2012)	<i>Support Vector Machine</i> dan <i>Maximum Entropy</i>	<i>Sentiment Classification for Indonesian Message in Social Media</i>	Metode <i>Support Vector Machine</i> memiliki tingkat akurasi paling tinggi dalam hal klasifikasi teks dengan akurasi 86,66%
2.	(Santoso, Ratnawati, & Indriati, 2014)	<i>Naive Bayes</i> , <i>K-Nearest Neighbor</i> , dan Metode Gabungan <i>K-Means</i> dan LVQ	Perbandingan Kinerja Metode <i>Naive Bayes</i> , <i>K-Nearest Neighbor</i> , dan Metode Gabungan <i>K-Means</i> dan LVQ dalam Pengkategorian Buku Komputer Berbahasa Indonesia Berdasarkan Judul dan Sinopsis	Metode terbaik dalam mengkategorikan buku komputer berbahasa Indonesia adalah metode <i>Naive Bayes</i>
3.	(Saleh, 2015)	<i>Naive Bayes</i>	Implementasi Metode Klasifikasi <i>Naive Bayes</i> Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga	Metode <i>Naive Bayes</i> menghasilkan keakuratan sebesar 78.33% dalam memprediksi besarnya pengguna listrik rumah tangga
4.	(Ulwan, 2016)	<i>Support Vector Machine</i> dan <i>Association</i>	<i>Pattern Recognition</i> Pada <i>Unstructured Data</i> Teks Menggunakan <i>Support Vector Machine</i> dan <i>Association</i> (Studi Kasus: Portal Layanan Aspirasi dan Pengaduan <i>Online</i> Rakyat)	Metode <i>Support Vector Machine</i> menghasilkan akurasi sebesar 96,7% dalam melakukan klasifikasi dan dilakukan visualisasi dan asosiasi untuk menemukan informasi yang dianggap penting.

Tabel 2. Penelitian Terdahulu (Lanjutan 1)

No.	Nama/Tahun	Metode	Judul Penelitian	Hasil Penelitian
5.	(Farid, 2017)	<i>Partial Least Square</i> (PLS)	Pengaruh Kualitas Pelayanan dan Persepsi Harga Terhadap Kepuasan Pelanggan dalam Meningkatkan Loyalitas Pelanggan Pengguna Aplikasi Traveloka Berbasis Internet (Studi Pada Pengguna Aplikasi Online Traveloka Universitas Lampung)	Variabel kualitas pelayanan dan variabel persepsi harga berpengaruh signifikan terhadap kepuasan pengguna aplikasi Traveloka, loyalitas pelanggan pengguna aplikasi Traveloka, serta loyalitas pelanggan dalam meningkatkan loyalitas pelanggan pengguna aplikasi Traveloka.
6.	(Nuraini & Mudiantono, 2017)	<i>Partial Least Square</i> (PLS)	Analisis Faktor-Faktor yang Mempengaruhi Kepuasan Pelanggan dan Dampaknya Terhadap Minat Beli Ulang pada Situs Traveloka (Studi Pada Konsumen Traveloka di Kota Semarang)	Faktor yang mempengaruhi kepuasan pelanggan adalah kualitas pelayanan dengan baik, sedangkan proses yang paling berpengaruh terhadap minat beli ulang adalah dengan kualitas pelayanan yang mempengaruhi citra merek sebagai salah satu penentu keberhasilan untuk dapat meningkatkan minat beli ulang
7.	(Risyandi & Zuliestiana, 2017)	<i>Webqual</i>	Pengaruh Kualitas Website Traveloka Terhadap Kepuasan Pengguna	Secara parsial maupun simultan variabel yang berpengaruh signifikan terhadap kualitas website traveloka atau kepuasan pengguna adalah variabel <i>Webqual</i> dan sub variabel <i>Ussability</i> , <i>Information Quality</i> , <i>Service Interaction Quality</i>

**Tabel 3.** Penelitian Terdahulu (Lanjutan 2)

<b>No.</b>	<b>Nama/Tahun</b>	<b>Metode</b>	<b>Judul Penelitian</b>	<b>Hasil Penelitian</b>
8.	(Syafitri, 2017)	<i>Naive Bayes</i>	Penerapan Teknik <i>Web Scraping</i> dan Klasifikasi Sentimen Menggunakan Metode <i>Naive Bayes Classifier</i> dan Asosiasi (Studi Kasus: Data Ulasan <i>The House Of Raminten</i> Pada Situs <i>TripAdvisor</i> )	Metode <i>Naive Bayes Classifier</i> menghasilkan tingkat akurasi 84% pada ulasan bahasa Indonesia dan pada ulasan bahasa Inggris sebesar 77%.
9.	(Ghifary, 2018)	Regresi Linier Berganda	Analisis Kualitas Layanan pada Perusahaan <i>E-Commerce</i> Traveloka	<i>ESQual</i> berpengaruh signifikan terhadap <i>Perceived Value</i> , <i>ESQual</i> berpengaruh signifikan terhadap <i>Loyalty Intentions</i> , sedangkan <i>ESQual</i> melalui <i>Perceived Value</i> berpengaruh signifikan terhadap <i>Loyalty Intentions</i>

## BAB III

### LANDASAN TEORI

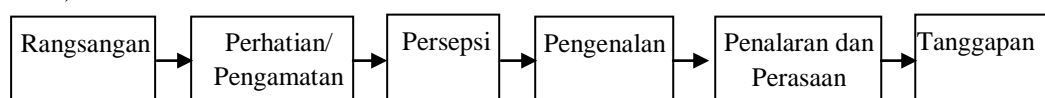
#### 3.1. Tanggapan

Dalam sebuah komunikasi, tanggapan merupakan suatu hasil yang ingin dicapai baik komunikasi primer maupun komunikasi sekunder. Komunikasi primer adalah proses penyampaian perasaan dan atau pikiran seseorang kepada orang lain menggunakan lambang (bahasa, kial, isyarat, gambar, warna, dan sebagainya) yang dapat menerjemahkan secara langsung pikiran atau perasaan komunikator, sedangkan komunikasi sekunder adalah proses penyampaian pesan seseorang kepada orang lain melalui alat atau sarana (surat, telepon, radio, televisi, dan sebagainya) sebagai media kedua setelah memakai lambang sebagai media pertama (Effendy, 2001).

Beberapa pengertian tentang tanggapan antara lain, menurut Syafrin (Hartanti, 2013), “Tanggapan adalah perilaku yang muncul dikarenakan adanya rangsangan dari lingkungan. Jika rangsangan dan respon dipasangkan atau dikondisikan maka akan membentuk tingkah laku baru terhadap rangsangan yang dikondisikan”.

Onong Uchjana Effendy (2001) mengemukakan bahwa tanggapan adalah seperangkat reaksi pada komunikan setelah diterima pesan. Tanggapan seorang komunikan yang telah disampaikan kepada komunikator disebut umpan balik. Umpan balik dapat bersifat positif maupun negatif. Umpan balik positif adalah tanggapan atau reaksi komunikan yang menyenangkan komunikator, sedangkan umpan balik negatif adalah tanggapan komunikan yang tidak menyenangkan komunikatornya.

Proses terjadinya tanggapan ditunjukkan oleh gambar berikut (Erwin, 2014).

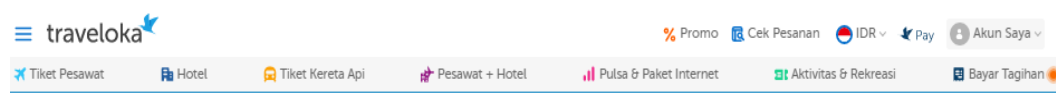


**Gambar 2.** Skema Proses Tanggapan

Pada **gambar 2** terlihat bahwa sebelum tercapai sebuah tanggapan harus terdapat rangsangan terlebih dahulu, setelah terdapat rangsangan muncul sebuah perhatian/pengamatan yang kemudian menimbulkan persepsi yang menyebabkan pengenalan rangsangan. Pengenalan sendiri berarti cara seseorang memberikan arti terhadap lingkungan, selanjutnya muncul penalaran dan perasaan yang kemudian memunculkan sebuah tanggapan.

### 3.2. Traveloka

Traveloka merupakan salah satu *online travel agent* di Indonesia yang menyediakan layanan pemesanan hotel, tiket pesawat, tiket kereta, pembelian pulsa dan paket internet, pembayaran tagihan listrik PLN, BPJS kesehatan, Telkom, dan juga PDAM.



(Sumber: Traveloka, 2018)

**Gambar 3.** Produk dan Layanan Traveloka

Traveloka didirikan oleh Ferry Unardi pada bulan Maret 2012, namun *launching* secara resmi pada bulan Oktober 2012. Sebelum mendirikan traveloka, Ferry sempat bekerja sebagai *software engineer* di kantor cabang perusahaan *Microsoft* yang berada di kota Seattle (Khalidi, 2014).

Pada awal konsepnya traveloka berfungsi sebagai mesin pencari yang digunakan untuk membandingkan harga tiket pesawat dari berbagai situs yang ada. Pada pertengahan tahun 2013, traveloka berubah menjadi situs reservasi tiket pesawat yang menyediakan pemesanan melalui situs resminya [www.traveloka.com](http://www.traveloka.com) dan pada bulan Maret 2014, Ferry Unardi menyatakan bahwa traveloka akan segera masuk ke bisnis reservasi kamar hotel. Pada bulan Juli 2014, jasa pemesanan hotel telah tersedia di situs Traveloka (Wikiwand, 2017).

### 3.3. Twitter

*Twitter* merupakan salah satu media jejaring sosial yang dapat digunakan dengan mendaftarkan diri sebagai pengguna *twitter* melalui *website* [www.twitter.com](http://www.twitter.com) dan memasukkan email untuk melakukan registrasi. Awalnya

*twitter* membatasi penggunaannya untuk mengirim dan membaca teks hanya 140 karakter, namun pada akhir bulan September 2017 *twitter* memperbanyak jumlah karakternya menjadi 280 karakter (Bohang, 2017). *Twitter* didirikan oleh Jack Dorsey pada tanggal 21 Maret 2006 di San Francisco, California, Amerika dan mulai diluncurkan situs jejaring sosialnya pada tanggal 15 Juli 2006.

*Twitter* menyediakan wadah bagi penggunaannya untuk dapat membuat *tweet* pada akun profil pengguna dan *tweets* tersebut dapat ditampilkan pada beranda teman yang mengikutinya (*follower*). Selain membuat *tweets*, pengguna *twitter* juga dapat melihat *tweets* dari akun teman yang diikutinya (*following*). Pengguna *twitter* juga dapat mencari *tweets* mengenai topik tertentu dengan cara menuliskan *keyword*, misalnya “Traveloka”, maka akan muncul beberapa postingan pengguna-pengguna *twitter* yang memuat kata traveloka (Wikipedia, 2017).

Beberapa istilah atau simbol yang digunakan dalam *Twitter* antara lain (Pakar Online Indonesia, 2012):

1. *Tweet*, yaitu artikel atau tulisan atau postingan di dalam *Twitter*.
2. *DM/Direct Message*, yaitu pesan pribadi antara kedua pengguna/pemilik akun *Twitter*.
3. *Follow*, yaitu mengikuti pengguna *Twitter* lainnya. Pengguna akun *Twitter* yang mengikuti akan melihat semua *tweet* terbaru dari pengguna akun *Twitter* yang diikutinya.
4. *Following*, yaitu memilih untuk melihat *tweet* orang lain pada *timeline*.
5. *Handle*, yaitu akun *Twitter* disebut dalam pegangan dan biasanya diidentifikasi dengan simbol @.
6. *Lists*, yaitu cara untuk menggabungkan beberapa *tweet* dari orang-orang atau perusahaan yang diikuti di satu tempat. *List* bisa berisikan *tweet* dari teman dekat, kompetitor, orang-orang dari komunitas tertentu ataupun yang lainnya. *List* adalah cara yang bagus untuk fokus ke berita atau *tweet* tertentu.
7. *RT/Retweet*, yaitu melanjutkan (*forward*) *tweet* yang diterima di *timeline Twitter*.

8. *Search*, yaitu kolom untuk mencari sesuatu yang dibutuhkan di dunia *Twitter*, karena banyak *tweet* bersifat umum, maka *search* dapat digunakan untuk mencari apapun.
9. *Timeline*, yaitu daftar kronologi dari *tweet*, orang yang diikuti atau *list*.
10. *Trending Topics*, yaitu *tweet* yang sedang *tren* yang ada di *Twitter* berbasis web.
11. *Tweetup*, yaitu bahasa yang digunakan oleh pengguna *Twitter* jika mereka ingin bertemu di dunia nyata, yang biasanya pemimpin *Tweetup* membuat *hashtag* tertentu.
12. *Unfollow*, yaitu melakukan *unfollow* sehingga tidak akan menerima *tweet* terbaru lagi di timeline dari pengguna *Twitter* yang sebelumnya diikuti.
13. *Hash Tag*, disimbolkan dengan # yang mana digunakan untuk menandai kata kunci atau topik dalam sebuah *Tweet* dan diciptakan secara organis oleh pengguna *Twitter* sebagai cara untuk mengkategorikan pesan *tweet*.

### 3.4. Statistika Deskriptif

Statistika deskriptif disebut juga statistika deduktif yaitu bagian dari statistik yang mempelajari cara pengumpulan dan penyajian data sehingga mudah dipahami (Hasan, 1999).

Menurut (Rasyad, 2003) jenis penyajian data secara umum dapat dibagi ke dalam tiga bentuk yaitu :

#### 1. Narasi

Narasi atau penyajian secara teks adalah cara menyajikan data hasil penelitian ke dalam suatu bentuk kalimat yang memuat gambaran umum tentang kesimpulan hasil pengamatan. Penyajian dalam bentuk teks banyak digunakan dalam bidang ekonomi, sosial, dan psikologi atau berperan sebagai laporan hasil penelitian kualitatif.

#### 2. Tabel/Daftar

Tabel/daftar merupakan kumpulan angka yang disusun menurut suatu kategori. Tabel mempunyai beberapa arah yaitu:

- Satu arah (*one way table*) yaitu tabel yang memuat keterangan tentang satu hal atau satu karakteristik saja. Misalnya nilai UAS (Ujian Akhir Semester) 5 orang mahasiswa.

**Tabel 4.** Nilai UAS Mahasiswa

Nama	Nilai UAS
A	78
B	87
C	77
D	90
E	83

- Dua arah (*two way table*) yaitu tabel yang menunjukkan dua hal atau dua karakteristik. Misalnya nilai UAS (Ujian Akhir Semester) 5 orang mahasiswa dipengaruhi nilai UTS (Ujian Tengah Semester).

**Tabel 5.** Nilai UTS dan UAS Mahasiswa

Nama	Nilai UTS	Nilai UAS
A	80	78
B	76	87
C	78	77
D	82	90
E	85	83

- Tiga arah (*three way table*) yaitu tabel yang menunjukkan tiga hal atau tiga karakteristik. Misalnya nilai UAS (Ujian Akhir Semester) 5 orang mahasiswa berdasarkan asal daerah yang dipengaruhi nilai UTS (Ujian Tengah Semester).

**Tabel 6.** Nilai UTS dan UAS Mahasiswa

Nama	Asal Daerah	Nilai UTS	Nilai UAS
A	Jawa tengah	80	78
B	Yogyakarta	76	87
C	Kalimantan Selatan	78	77
D	Jawa Tengah	82	90
E	Jambi	85	83

### 3. Diagram/Grafik

Penyajian data dengan grafik dianggap lebih komunikatif karena dapat diketahui karakteristik dari data yang disajikan dalam waktu singkat. Terdapat beberapa jenis grafik yaitu:



- Grafik garis (*line chart*) dipakai untuk menggambarkan data berkala. Grafik garis dibagi menjadi grafik garis tunggal, grafik garis berganda, garis batangan komponen berganda, grafik garis presentase komponen berganda dan grafik garis berimbang neto.
- Grafik batang (*bar chart*) yaitu grafik yang memiliki kesamaan fungsi dengan grafik garis namun garis pada grafik baris diganti dengan balok/batangan. Grafik batang dibagi menjadi grafik batangan tunggal, grafik batangan berganda, grafik batangan komponen berganda, grafik batangan presentase komponen berganda dan grafik batangan berimbang neto.
- Grafik lingkaran (*pie chart*) yaitu grafik yang cocok digunakan untuk menyajikan data *cross section*, yang nantinya data tersebut disajikan dalam bentuk persentase.
- Grafik gambar (*pictogram*) yaitu grafik yang berupa gambar atau lambang yang menunjukkan jumlah benda yang dilambangkan.
- Grafik peta (*cartogram*) yaitu grafik yang banyak digunakan oleh BMG untuk menunjukkan cuaca di beberapa daerah.

### 3.5. *Data Mining*

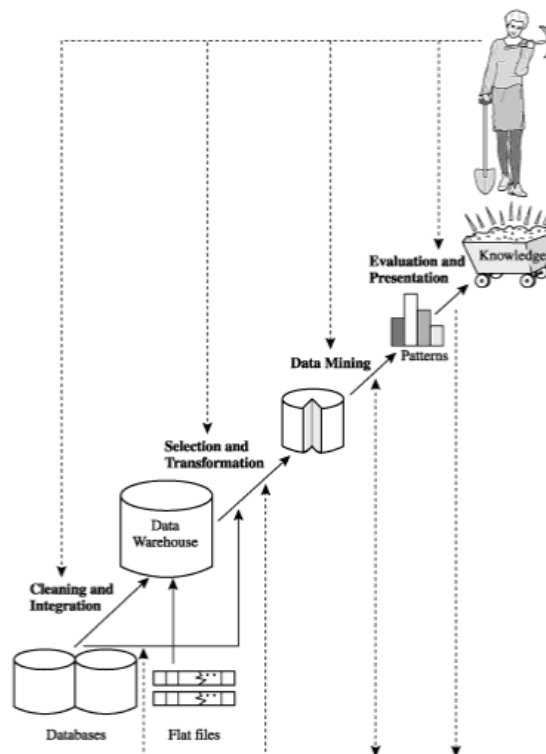
*Data Mining* atau sering disebut *knowledge discovery in database* (KDD) merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam suatu set data yang berukuran besar. Keluaran atau *output* dari data mining dapat digunakan sebagai perbaikan dalam pengambilan keputusan di masa depan (Santosa, 2007).

Tahapan dalam proses *knowledge discovery* terdiri dari langkah-langkah sebagai berikut (Han & Kamber, 2006):

1. *Data cleaning* (pembersihan data) adalah proses menghapus data yang tidak konsisten serta menghilangkan noise.
2. *Data integration* (integrasi data) adalah proses menggabungkan data dari berbagai sumber data dalam sistem data mining.

3. *Data selection* (seleksi data) adalah proses pengambilan data yang relevan yang berkaitan dengan tugas analisis yang akan digunakan dalam proses data mining.
4. *Data transformation* (transformasi data) adalah proses dimana data ditransformasikan ke dalam bentuk yang sesuai untuk *mining* (penambangan).
5. *Data mining* (penambangan data) adalah inti pada proses KDD yang melibatkan teknik tertentu untuk menghasilkan suatu pola dari data yang terpilih.
6. *Pattern evaluation* (evaluasi pola) adalah proses untuk menguji kebenaran dari pola data yang mewakili *knowledge* pada data.
7. *Knowledge representation* (representasi pengetahuan) adalah proses representasi secara visual kepada pengguna agar membantu mempermudah pemahaman pengguna terhadap hasil *data mining*.

Tahapan tersebut ditunjukkan oleh gambar berikut.



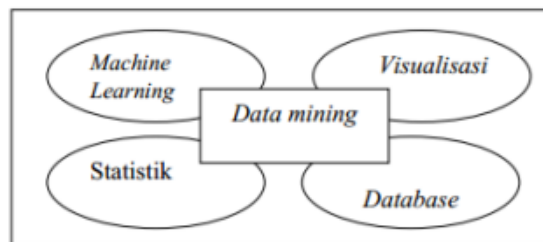
(Sumber: Jiawei Han dan Micheline Kamber, 2006)

**Gambar 4.** Tahapan dalam Proses *Knowledge Discovery in Database* (KDD)

### 3.6. *Machine Learning*

*Machine Learning* adalah suatu area dalam kecerdasan buatan (*artificial intelligence*) yang berhubungan dengan teknik-teknik yang dapat dikembangkan melalui pemrograman dan proses pembelajaran menggunakan data masa lalu (Santosa, 2007).

Kecerdasan buatan pada *machine learning* dibuat oleh *user* terhadap komputer dan algoritma dalam *machine learning* tersebut akan menghasilkan sebuah model atau pola dari masukan (*input*). Pengenalan pola, *machine learning* dan data mining sering disebutkan dalam artian yang sama, bidang ini bersinggungan dengan ilmu probabilitas dan statistik terkadang juga optimasi. *Machine learning* merupakan sebuah alat analisis dalam data mining. Hubungan antara bidang-bidang tersebut dapat dilihat pada **gambar 5** (Santosa, 2007).



(Sumber: Budi Santoso, 2007)

**Gambar 5.** Ilustrasi Cakupan *Data Mining*

Perbedaan antara *machine learning*, statistik, dan data mining adalah statistik lebih berdasarkan teori dan lebih fokus pada pengujian hipotesis, *machine learning* lebih bersifat heuristik dan fokus pada perbaikan performansi dari suatu teknik *learning*, selain itu juga meliputi *real-time learning* dan *robotic area* yang tidak termasuk dalam data mining, sedangkan *data mining* merupakan gabungan antara teori dan heuristik dan fokus pada seluruh proses penemuan pola yang termasuk data *cleaning*, *learning*, dan visualisasi hasil (Santosa, 2007).

Tiga teknik didalam data mining yang populer (Pramudiono, 2003) adalah:

#### 1. *Association Rule Mining*

*Association Rule Mining* merupakan suatu teknik yang digunakan untuk menemukan aturan assosiasi antara suatu kombinasi item. Menurut Ather dan Lobo (Ulwan, 2016) terdapat dua langkah terpisah dalam aturan asosiasi: Pertama, mencari minimum support yang diterapkan untuk menemukan

semua pengulangan itemset dalam basis data. Kedua, melakukan pengulangan itemset dan menentukan batasan minimum *confidence* yang digunakan untuk membentuk aturannya.

## 2. Klasifikasi (*Classification*)

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.

## 3. *Clustering*

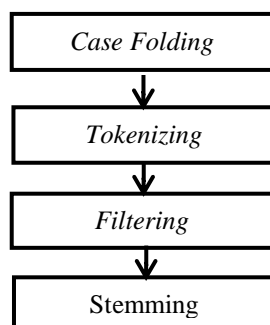
*Clustering* merupakan pengelompokan data tanpa berdasarkan kelas data tertentu, berbeda dengan *association rule mining* dan *classification*. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota dalam satu kelas dan meminimumkan kesamaan antar kelas.

### 3.7. *Text Mining*

Menurut (Han & Kamber, 2006), *text mining* adalah satu langkah dari analisis teks yang secara otomatis dilakukan oleh komputer dengan tujuan menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen. Ratna Maria (Damanik, 2014) mengemukakan bahwa *text mining* adalah proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan memiliki tujuan mencari kata-kata yang dapat mewakili isi dari dokumen tersebut, sehingga dapat dilakukan analisa keterhubungan antara dokumen.

Perbedaan antara *data mining* dan *text mining* adalah pada data yang digunakan, *data mining* menggunakan data yang terstruktur (*structured data*) sedangkan *text mining* menggunakan data yang umumnya adalah *unstructured data* atau *semistructured*. Pada *unstructured data* atau *semistructured data* tentu akan sulit menggali suatu informasi, selain itu permasalahan yang terdapat dalam *text mining* adalah jumlah data yang besar, dimensi yang tinggi, serta adanya data *noise* pada teks. Oleh karena itu, langkah yang dapat dilakukan agar teks dapat dianalisis adalah dengan melakukan tahap *preprocessing*.

Tahap *preprocessing* dilakukan agar data *input* mentah dapat diubah menjadi format yang sesuai untuk analisis selanjutnya. Menurut Raymond J. Mooney (Damanik, 2014) tahap *preprocessing* yang dilakukan dalam *text mining* pada dokumen secara umum adalah *case folding*, *tokenizing*, *filtering*, dan *stemming*. Diagram tahapan tersebut ditunjukkan oleh **gambar 6**.

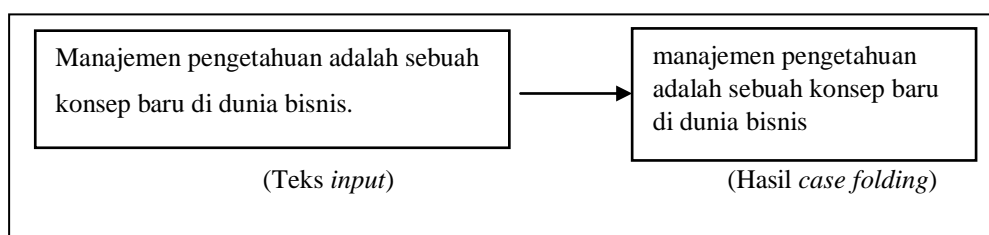


(Sumber: Ruth Mega Ulina Damanik, 2013)

**Gambar 6.** Tahapan *Preprocessing* pada *Text Mining*

### 3.7.1. *Case Folding*

*Case folding* adalah tahap mengubah semua huruf kapital dalam dokumen menjadi huruf kecil meliputi huruf ‘a’ sampai dengan ‘z’ yang diterima, karakter selain huruf akan dihilangkan karena dianggap sebagai delimiter. Proses *case folding* ditunjukkan oleh **gambar 7**.



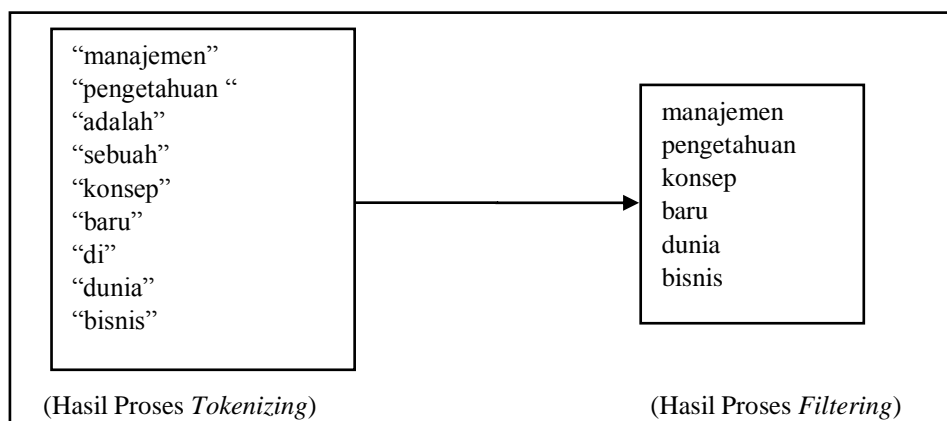
(Sumber: Ruth Mega Ulina Damanik, 2014)

**Gambar 7.** Proses *Case Folding*

### 3.7.2. *Tokenizing dan Filtering*

*Tokenizing* adalah tahap pemotongan *string input* dalam suatu kalimat berdasarkan kata-kata yang menyusunnya, sedangkan *filtering* adalah tahap pengambilan kata-kata penting yang dibutuhkan untuk dianalisis. Hasil dari *tokenizing* adalah kata-kata yang sudah berdiri sendiri, sedangkan pada tahap *filtering* kata-kata tersebut akan disaring dan kata-kata yang tidak deskriptif (kata

penghubung dan kata-kata yang tidak diinginkan penulis) akan dihilangkan. Proses filtering bisa dilakukan dengan menggunakan algoritma *stoplist/stopword* (membuang kata-kata yang tidak deskriptif) atau *wordlist* (menyimpan kata-kata yang penting). Proses *tokenizing* dan *filtering* ditunjukkan oleh **gambar 8**.

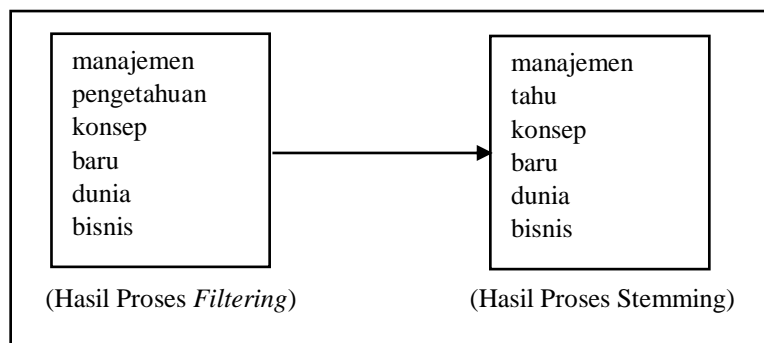


(Sumber: Ruth Mega Ulina Damanik, 2014)

**Gambar 8.** Proses *Tokenizing* dan *Filtering*

### 3.7.3. Stemming

*Stemming* adalah tahap mencari *root* kata (kata dasar) dari setiap kata hasil proses *filtering* yang mengandung imbuhan, baik imbuhan deduktif maupun imbuhan induktif. Pada tahap ini dilakukan penghapusan imbuhan agar kata yang terbentuk memiliki representasi yang sama dengan kata yang tidak mengandung imbuhan. *Stemming* kebanyakan dipakai pada teks berbahasa Inggris dan lebih sulit diterapkan pada teks berbahasa Indonesia, dikarenakan pada teks berbahasa Indonesia tidak memiliki bentuk baku yang permanen. Proses *stemming* ditunjukkan oleh **gambar 9**.



(Sumber: Ruth Mega Ulina Damanik, 2014)

**Gambar 9.** Proses *Stemming*

### 3.8. Pembobotan Kata (*Term Weighting*)

*Term weighting* merupakan hal yang penting dalam text mining. *Term weighting* diperlukan karena masing-masing kata yang berbeda di dalam suatu dokumen memiliki tingkat kepentingan yang berbeda, karena itu *term weighting* diperlukan untuk mencari informasi dari koleksi dokumen yang bersifat heterogen (beragam). Menurut Mandala (Pratiwi & Widodo, 2017) dijelaskan bahwa *term weighting* dipengaruhi oleh hal-hal berikut:

1. *Term Frequency (TF) factor*, yaitu faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*Term Frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu *term* (TF tinggi) dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.
2. *Inverse Document Frequency (IDF) factor*, yaitu pengurangan dominansi *term* yang sering muncul di berbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul di berbagai dokumen, dapat dianggap sebagai *term* umum (*common term*) sehingga tidak penting nilainya. Sebaliknya faktor kata (*term scarcity*) yang jarang muncul dalam koleksi dokumen harus diperhatikan dalam pemberian bobot. Menurut Mandala, kata yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting (*uncommon terms*) daripada kata yang muncul pada banyak dokumen. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*Inverse Document Frequency*). Hal ini merupakan usulan dari George Zipf. Zipf mengamati bahwa frekuensi dari sesuatu cenderung kebalikan secara proporsional dengan urutannya.

#### 3.8.1. Metode TF-IDF (*Term Frequency-Inverse Document Frequency*)

Metode pembobotan yang paling banyak digunakan dalam pembobotan adalah metode penggabungan antara *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. Metode tersebut disebut metode TF-IDF (*Term Frequency-Inverse Document Frequency*). TF menunjukkan seberapa sering sebuah kata

muncul dalam suatu dokumen tertentu, sedangkan IDF menjelaskan seberapa penting arti sebuah kata dalam sebuah dokumen. IDF dapat dihitung dengan menggunakan rumus berikut:

$$IDF(A) = \log \frac{D}{df_A} \quad (1)$$

dimana,

$A$  : suatu contoh kata

$IDF A$  : pentingnya kata  $A$  dalam dokumen (*termfrequency*)

$D$  : jumlah seluruh dokumen

$df_A$  : jumlah dokumen yang memuat kata  $A$  di dalamnya

Apabila nilai  $D$  sama dengan  $df_j$  maka akan didapatkan hasil nol untuk perhitungan IDF sehingga perlu ditambahkan nilai 1 untuk rumus IDF menjadi

$$IDF(A) = \log \frac{D}{df_A} + 1 \quad (2)$$

$$W A = TF A \times IDF(A) \quad (3)$$

dimana,

$W A$  : bobot *term* dari kata  $A$

$TF A$  : jumlah kata  $A$  dalam dokumen (*termfrequency*)

### 3.8.2. Simulasi Metode TF-IDF (*Term Frequency-Inverse Document Frequency*)

Misalkan terdapat empat buah dokumen yang diambil dari kelas positif sebagai berikut.

Dokumen 1: Traveloka Tambahkan Layanan Pemesanan Tiket Bus

Dokumen 2: traveloka sangat disarankan buat pesan tiket

Dokumen 3: MEMBELI TIKET PESAWAT MURAH DI  
SITUS TRAVELOKA

Dokumen 4: untuk periode promo tersebut dapat digunakan 17 - 21 Maret  
2018

Keempat dokumen tersebut dilakukan perhitungan pembobotan kata/*query* menggunakan metode TF-IDF. Misal kata/*query* yang digunakan adalah “tiket”, “bus”, “pesawat”, dan “promo”. Keempat dokumen tersebut dilakukan proses *preprocessing* maka akan mengalami perubahan kata seperti berikut.



Dokumen 1 : layan pesan **tiket bus**

Dokumen 2 : saran pesan **tiket**

Dokumen 3 : beli **tiket pesawat** murah situs

Dokumen 4 : periode **promo** maret

Berdasarkan keempat dokumen tersebut diperoleh beberapa *document term* sebagai berikut.

-layan	-bus	-pesawat	-periode
-pesan	-saran	-murah	-promo
-tiket	-beli	-situs	-maret

Nilai bobot *term* kata “tiket” atau  $W(\text{tiket})$  dalam dokumen 1 dapat dihitung dengan mengetahui:

1. Jumlah kata tiket dalam dokumen 1 yaitu 1, maka  $TF(\text{tiket})=1$ .
2. Jumlah seluruh dokumen yaitu 4, maka  $D=4$ .
3. Jumlah dokumen yang memuat kata tiket yaitu 3 dokumen, maka  $df_{(\text{tiket})}=3$ .

Oleh karena itu, dengan menggunakan rumus pada Persamaan 2 dan Persamaan 3 diperoleh nilai bobot *term* untuk kata “tiket” pada dokumen 1 sebagai berikut.

$$W \text{ tiket} = 1 \times \left( \log \frac{4}{3} + 1 \right)$$

$$W \text{ tiket} = 1,12$$

Bobot *term* kata dalam masing-masing dokumen ditunjukkan oleh tabel berikut.

**Tabel 7.** Hasil Perhitungan Bobot Masing-masing Dokumen

Query	Dokumen				D	df(A)	W			
	1	2	3	4			1	2	3	4
tiket	1	1	1	0	4	3	1.12	1.12	1.12	0
bus	1	0	0	0		1	1.6	0	0	0
pesawat	0	0	1	0		1	0	0	1.6	0
promo	0	0	0	1		1	0	0	0	1.6
Nilai bobot setiap dokumen ( $\sum W_{(di)}$ )							2.73	1.12	2.73	1.60

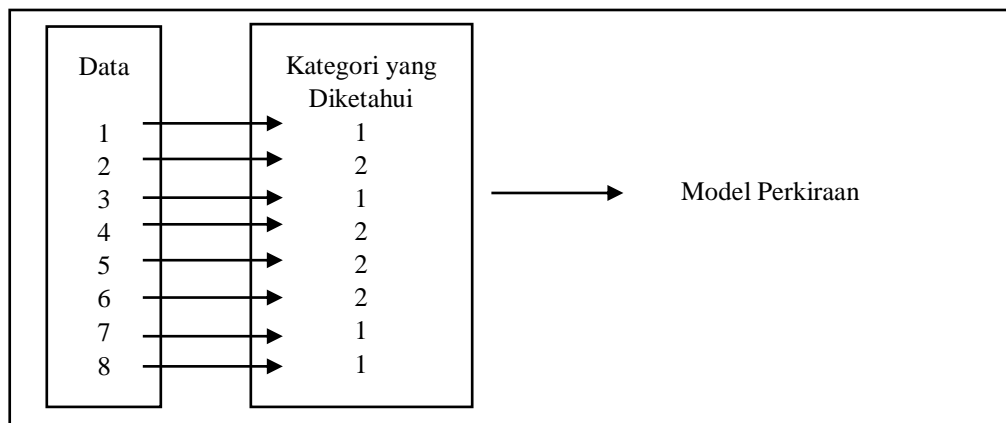
Nilai bobot pada dokumen menunjukkan tinggi rendahnya kesesuaian antara dokumen dengan *query*. Berdasarkan **tabel 7** diketahui bahwa dokumen yang memiliki tingkat similiaritas paling tinggi terhadap *query* “tiket”, “bus”, “pesawat”, dan “promo” adalah dokumen 1 dan dokumen 3, setelah itu dokumen 4, dan dokumen yang memiliki similiaritas terendah terhadap keempat *query* adalah dokumen 2.

### 3.9. Klasifikasi

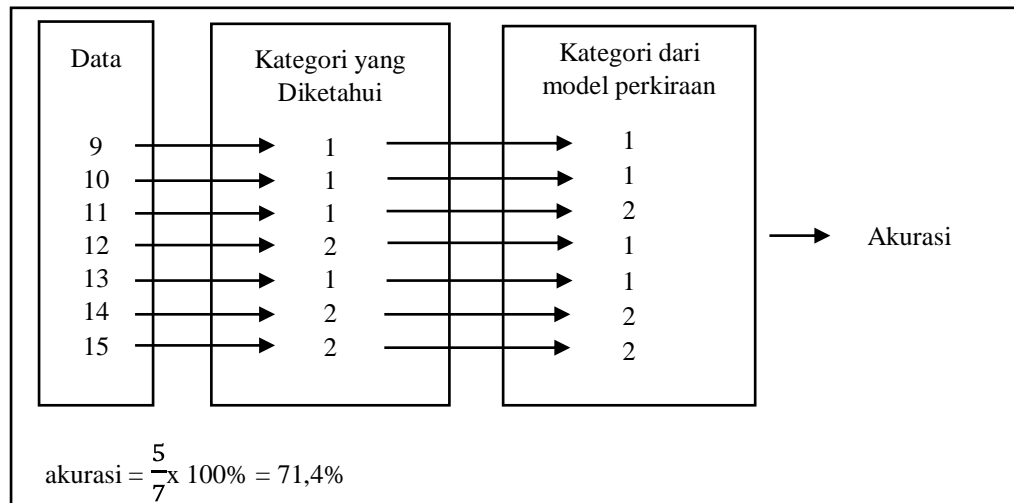
Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Pramudiono, 2003). Proses klasifikasi dibagi kedalam dua fase yaitu fase *learning* dan fase *testing*. Fase *learning* dilakukan pada sebagian data yang telah diketahui kelas datanya kemudian diumpungkan untuk membentuk model perkiraan, sedangkan fase *testing* model yang sudah terbentuk diuji dengan data lainnya untuk mengetahui akurasi dari model tersebut.

#### 3.9.1. Gambaran Fase *Learning* dan Fase *Testing*

Terdapat 15 data tanggapan masyarakat mengenai “Traveloka” yang terbagi kedalam dua kelas kategori yaitu 1 untuk kelas positif dan 2 untuk kelas negatif. Peneliti ingin mengetahui metode klasifikasi yang paling tepat untuk mengklasifikasikan tanggapan tersebut, maka dilakukan pembagian 15 data yang diperoleh ke dalam dua fase sehingga diperoleh tingkat akurasi. Fase *learning* ditunjukkan oleh **gambar 10** dan fase *testing* ditunjukkan oleh **gambar 11**.



**Gambar 10.** Fase *Learning*



**Gambar 11.** Fase *Testing*

### 3.9.2. Data *Learning*

Banyaknya data yang digunakan pada fase *learning* dihitung dengan menggunakan rumus slovin, sedangkan sisanya digunakan pada data *testing*. Menurut Sevilla et. al (Pratiwi & Widodo, 2017), rumus slovin digunakan untuk menentukan jumlah sampel dalam suatu penelitian. Berikut rumus slovin yang digunakan dalam melakukan perhitungan jumlah sampel.

$$n = \frac{N}{1+Ne^2} \quad (4)$$

dimana,

$N$  : ukuran populasi (dalam kasus ini populasi adalah total data yang diambil)

$n$  : ukuran sampel (dalam kasus ini sampel adalah jumlah *data training*)

$e$  : batas toleransi kesalahan

### 3.9.3. Simulasi Perhitungan Data *Learning* dengan Rumus Slovin

Pada suatu penelitian untuk mengetahui tanggapan masyarakat mengenai situs jejaring sosial Traveloka, dilakukan pengumpulan data-data yang mengandung kata Traveloka yang ditulis dalam jejaring sosial *Twitter* dan diperoleh total tanggapan sebanyak 500 data. Dalam kasus ini, peneliti ingin melakukan klasifikasi terhadap 500 tanggapan tersebut, maka diperlukan sampel data yang digunakan pada fase pembelajaran (*learning*). Peneliti menggunakan

rumus Slovin untuk menentukan jumlah data yang digunakan pada fase *learning* dengan batas toleransi kesalahan sebesar 5%.

$$n = \frac{N}{1 + Ne^2}$$

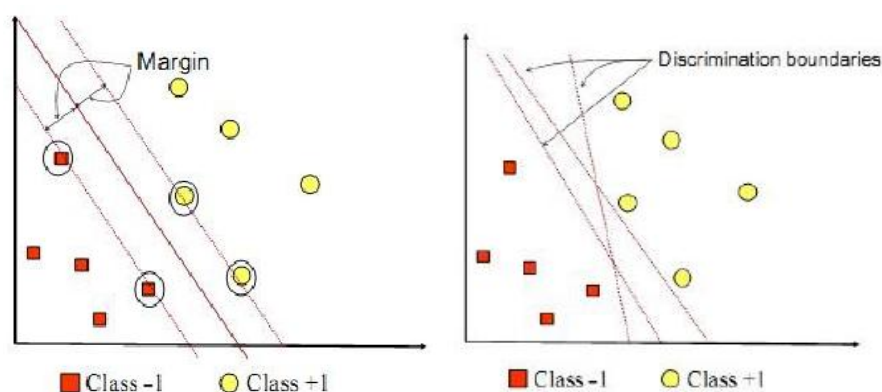
$$n = \frac{500}{1 + 500 \times (0,05)^2}$$

$$n = 222,22 \approx 222$$

Berdasarkan rumus *Slovin*, maka diperoleh jumlah data yang digunakan pada fase *learning* sebanyak 222 data.

#### 3.9.4. *Support Vector Machine* pada Data Terpisah Secara Linear

*Support Vector Machine* (SVM) adalah metode yang dikembangkan oleh Boser, Guyon, Vapnik, dan dipresentasikan pertama kali pada acara *Annual Workshop on Computational Learning Theory* tahun 1992. *Support Vector Machine* (SVM) bekerja menggunakan prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik untuk memisahkan dua kelas data.



(Sumber: Susilowati, Sabariah, & Gozali, 2015)

**Gambar 12.** Ilustrasi SVM Menemukan *Hyperplane* Terbaik

Pada **gambar 12** menunjukkan beberapa *pattern* dari dua kelas yaitu negatif (-1) dan positif (+1). *Hyperplane* pemisah kedua kelas terbaik dapat ditemukan dengan memaksimalkan *margin*. *Margin* merupakan jarak antara *hyperplane* dengan *pattern* terdekat dari masing-masing kelas. *Pattern* terdekat tersebut disebut *support vector*, pada **gambar 12** *support vector* ditunjukkan oleh *pattern* merah dan kuning yang berada dalam lingkaran hitam.

Fungsi pemisah linear didefinisikan sebagai (Santosa, 2007):

$$g(x) := \text{sgn}(f(x)) \quad (5)$$

dengan  $f(x) = wx + b$ , dimana  $x, w \in \mathbb{R}^n$  dan  $b \in \mathbb{R}$ .

*Hyperplane* pendukung dari kelas +1 (positif):

$$w x_i + b = +1 \quad (6)$$

Sedangkan *hyperplane* pendukung kelas -1 (negatif):

$$w x_i + b = -1 \quad (7)$$

Secara matematika, formulasi problem SVM untuk kasus klasifikasi linear di dalam *primal space* adalah

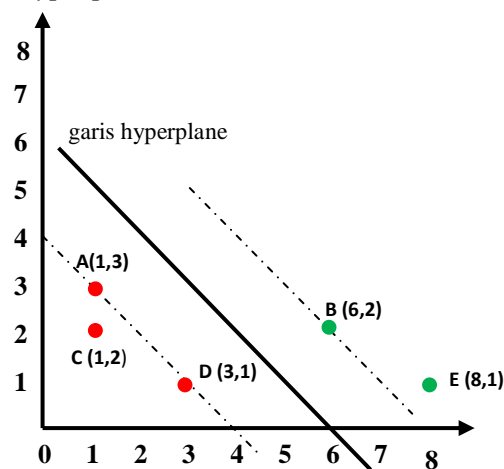
$$\min_w (w) = \frac{1}{2} \|w\|^2 \quad (8)$$

$$y_i (w \cdot x_i + b) \geq 1 \quad (9)$$

dimana,  $x_i$  adalah data input;  $y_i$  adalah keluaran dari data  $x_i$ ;  $w$  dan  $b$  adalah parameter yang dicari nilainya. Pada formula Persamaan 8 dan Persamaan 9 ingin meminimalkan *obyektif function* (fungsi tujuan)  $\frac{1}{2} \|w\|^2$  dengan memperhatikan pembatas pada Persamaan 9.

### 3.9.5. Simulasi *Support Vector Machine* pada Data Terpisah Secara Linear

Simulasi metode *SVM* pada kasus linier ini, menggunakan lima buah titik yang akan diklasifikasikan dengan metode *SVM*. Prinsip kerja *SVM* adalah mencoba menemukan *hyperplane* terbaik.



Gambar 13. Grafik *Hyperplane*

Berdasarkan **gambar 13** dapat disajikan dalam **tabel 8** berikut.

**Tabel 8.** Data Simulasi SVM Linear

Titik	$x_1$	$x_2$	$y$
A	1	3	-1
B	6	2	1
C	1	2	-1
D	3	1	-1
E	8	1	1

Formulasi titik minimal fungsi tujuan  $\frac{1}{2}\|w\|^2$  dengan memperhatikan pembatas, berdasarkan **tabel 8** diperoleh persamaan *constraint* (pembatas) sebagai berikut.

$$-w_1 - 3w_2 - b \geq 1 \quad (10)$$

$$6w_1 + 2w_2 + b \geq 1 \quad (11)$$

$$-w_1 - 2w_2 - b \geq 1 \quad (12)$$

$$-3w_1 - w_2 - b \geq 1 \quad (13)$$

$$8w_1 + w_2 + b \geq 1 \quad (14)$$

Eliminasi Persamaan 10 dan Persamaan 11 diperoleh.

$$-w_1 - 3w_2 - b = 1 \quad (10)$$

$$6w_1 + 2w_2 + b = 1 \quad (11)$$

$$\frac{5w_1 - w_2}{5w_1 - w_2} = 1 \quad +$$

$$w_1 = 0,4 + 0,2w_2 \quad (15)$$

Eliminasi Persamaan 12 dan Persamaan 13, serta substitusi Persamaan 15 diperoleh.

$$-w_1 - 2w_2 - b = 1 \quad (12)$$

$$-3w_1 - w_2 - b = 1 \quad (13)$$

$$\frac{2w_1 - w_2}{2w_1 - w_2} = 0 \quad -$$

$$2*(0,4 + 0,2w_2) - w_2 = 0$$

$$0,8 + 0,8w_2 - w_2 = 0$$

$$w_2 = 4$$

$w_2=4$ , maka  $w_1 = 0,4+0,2w_2 = 1,2$

Substitusi nilai  $w_1$  dan  $w_2$  kedalam Persamaan 14 diperoleh.

$$\begin{aligned} 8w_1+w_2+b &= 1 & (14) \\ 8*1,2+4+b &= 1 \\ 9,6+4+b &= 1 \\ b &= -12,6 \end{aligned}$$

Berdasarkan nilai  $w_1$ ,  $w_2$ , dan  $b$  yang diperoleh, maka persamaan fungsi pemisahannya yaitu:

$$f x = 1,2x_1 + 4x_2 - 12,6$$

Klasifikasi untuk data selanjutnya dapat dilakukan dengan menggunakan fungsi pada Persamaan 5, dengan fungsi *sgn* tersebut semua nilai  $f(x)<0$  masuk ke kelas -1 (negatif) dan nilai  $f(x)>0$  masuk ke dalam kelas +1 (positif).

### 3.9.6. Klasifikasi dengan *Naive Bayes*

*Naive Bayes* merupakan metode yang dikemukakan oleh Thomas Bayes seorang ilmuwan inggris, *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik (Saleh, 2015). *Naive Bayes* merupakan metode yang didasarkan pada sebuah asumsi penyederhanaan dimana nilai atribut secara kondisional saling bebas jika diberikan nilai *output* (Santosa, 2007).

Secara umum, teorema Bayes ditunjukkan oleh Persamaan 16 berikut (Ariadi & Fithriasari, 2015).

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (16)$$

dimana,

$P(A|B)$  : Peluang suatu kejadian A apabila B terjadi

$P(B|A)$  : Peluang suatu kejadian B apabila A terjadi

$P(A)$  : Peluang kejadian A

$P(B)$  : Peluang kejadian B

Menurut Manning, dkk (Syafitri, 2017), “setiap ulasan pada klasifikasi *Naive Bayes* dipresentasikan kedalam pasangan atribut dimana  $a_1$  merupakan kata pertama,  $a_2$  merupakan kata kedua, dan seterusnya, sedangkan himpunan kelas disimbolkan dengan  $V$ ”. Metode *Naive Bayes* akan menghasilkan kelas atau kategori tertinggi probabilitasnya ( $V_{MAP}$ ) dengan memasukkan atribut ( $a_1, a_2, a_3, \dots, a_n$ ), rumus  $V_{MAP}$  ditunjukkan oleh Persamaan 17.

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2, a_3, \dots, a_n) \quad (17)$$

dimana,

$V_{MAP}$  : Output nilai hasil klasifikasi *Naive Bayes*

$\operatorname{argmax}$  : titik maksimum probabilitas kelas

$P(v_j | a_1, a_2, a_3, \dots, a_n)$  : probabilitas bukti  $v_j$  terjadi apabila atribut  $a_1, a_2, a_3, \dots, a_n$ .

Berdasarkan teorema Bayes pada Persamaan 16, maka Persamaan 17 dapat ditulis sebagai berikut.

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2, a_3, \dots, a_n | v_j) \times P(v_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (18)$$

$P(a_1, a_2, a_3, \dots, a_n)$  bernilai konstan untuk semua  $v_i$  sehingga Persamaan 18 menjadi Persamaan 19.

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2, a_3, \dots, a_n | v_j) \times P(v_j) \quad (19)$$

Menghitung  $P(a_1, a_2, a_3, \dots, a_n | v_j)$  akan memiliki tingkat kesulitan yang tinggi karena jumlah *term*  $P(a_1, a_2, a_3, \dots, a_n | v_j)$  tergantung jumlah kombinasi posisi kata dikali jumlah kelas. Menurut Tan, dkk (Syafitri, 2017) Klasifikasi dengan metode *Naive Bayes* menyederhanakan hal tersebut dengan mengasumsikan bahwa dalam setiap kategori, setiap atribut bebas bersyarat satu sama lain, dengan kata lain,

$$P(a_1, a_2, a_3, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \quad (20)$$



Persamaan 20 tersebut disubstitusikan ke dalam Persamaan 19, sehingga diperoleh Persamaan 21.

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \times \prod_i P(a_i | v_j) \quad (21)$$

Nilai  $P(v_j)$  dan  $P(a_i | v_j)$  dapat dihitung dengan menggunakan Persamaan 22 dan Persamaan 23 (Ariadi & Fithriasari, 2015).

$$P(v_j) = \frac{doc_j}{training} \quad (22)$$

$$P(a_i | v_j) = \frac{n_i + 1}{n + kosakata} \quad (23)$$

dimana,

$doc_j$  : jumlah dokumen memiliki kategori  $j$  pada fase *training/learning*

$training$  : jumlah dokumen pada fase *training*

$n_i$  : jumlah kemunculan kata  $a_i$  dalam dokumen berkategori  $v_j$

$n$  : jumlah seluruh kata pada dokumen dengan kategori  $v_j$

$kosakata$  : jumlah kata dalam fase *training*

### 3.9.7. Simulasi Klasifikasi dengan *Naive Bayes*

Misalkan terdapat empat buah dokumen yang telah melalui tahapan *preprocessing*, dua dokumen diambil dari kelas positif dan dua dokumen diambil dari kelas negatif. Dokumen tersebut adalah sebagai berikut.

Dokumen 1 : beli tiket pesawat murah situs

Dokumen 2 : beli tiket periode promo maret

Dokumen 3 : server tiket habis

Dokumen 4 : habis beli tiket buka twitter

Selanjutnya akan dihitung frekuensi kemunculan kata pada setiap dokumen seperti pada **tabel 9**.

**Tabel 9.** Frekuensi Kemunculan Kata

<i>Query</i>	<b>Dokumen</b>			
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
beli	1	1	0	1
tiket	1	1	1	1
pesawat	1	0	0	0
murah	1	0	0	0
situs	1	0	0	0
periode	0	1	0	0
promo	0	1	0	0
maret	0	1	0	0
server	0	0	1	0
habis	0	0	1	1
buka	0	0	0	1
twitter	0	0	0	1
<b>Kelas</b>	<b>Positif</b>	<b>Positif</b>	<b>Negatif</b>	<b>Negatif</b>

Berdasarkan **tabel 9** diketahui bahwa kelas positif terdiri dari 2 dokumen dengan jumlah kata sebanyak 10 kata dari 12 kosakata yang ada, sedangkan 2 dokumen kelas negatif terdiri dari 8 kata dari 12 kosakata yang ada. Berdasarkan jumlah kata tersebut, dapat dihitung nilai probabilitas untuk setiap kelasnya dengan menggunakan Persamaan 23.

1. Probabilitas kata kelas positif

Contoh perhitungan probabilitas untuk kata “beli” yang terdapat dalam kelas positif.

$$P_{a_i v_j} = \frac{n_i + 1}{n + \text{kosakata}}$$

$$P_{\text{beli positif}} = \frac{2+1}{10+12} = 0,136$$

dimana,

$n_i$  : jumlah kata “beli” dalam dokumen kelas positif

$n$  : jumlah seluruh kata pada dokumen kelas positif

*kosakata* : jumlah kata dalam fase *training*

Nilai probabilitas untuk kata yang lain, dengan menggunakan cara perhitungan yang sama ditunjukkan oleh **tabel 10**.

**Tabel 10.** Probabilitas Kata Kelas Positif

Query	Probabilitas	Query	Probabilitas
beli	0.136	promo	0.091
tiket	0.136	maret	0.091
pesawat	0.091	server	0.045
murah	0.091	habis	0.045
situs	0.091	buka	0.045
periode	0.091	twitter	0.045

## 2. Probabilitas kata kelas negatif

Contoh perhitungan kata “beli” pada kelas negatif.

$$P_{a_i v_j} = \frac{n_i+1}{n+kosakata}$$

$$P_{beli\ positif} = \frac{1+1}{8+12} = 0,1$$

dimana,

$n_i$  : jumlah kata “beli” dalam dokumen kelas negatif

$n$  : jumlah seluruh kata pada dokumen kelas negatif

$kosakata$  : jumlah kata dalam fase *training*

Nilai probabilitas untuk kata yang lain dalam kelas negatif ditunjukkan oleh **tabel 11**.

**Tabel 11.** Probabilitas Kata Kelas Negatif

Query	Probabilitas	Query	Probabilitas
beli	0.10	promo	0.05
tiket	0.15	maret	0.05
pesawat	0.05	server	0.10
murah	0.05	habis	0.15
situs	0.05	buka	0.10
periode	0.05	twitter	0.10

Nilai probabilitas kata pada masing-masing kelas tersebut, kemudian disimpan pada *database* yang nantinya akan digunakan untuk menguji data baru. Misal ingin diketahui kelas data dari tanggapan baru “tiket pesawat promo murah”. Langkah pertama yang dilakukan untuk melakukan klasifikasi adalah memecah kalimat dalam tanggapan tersebut menjadi kata per kata kemudian menghitung nilai probabilitas dari kata pada masing-masing kelas dengan

menggunakan tabel probabilitas kata yang telah diperoleh sebelumnya, sedangkan probabilitas masing-masing kelas ditentukan dengan menggunakan Persamaan 22.

$$P v_j = \frac{doc_j}{training}$$

$$P \text{ positif} = \frac{2}{4} = 0,5$$

$$P \text{ negatif} = \frac{2}{4} = 0,5$$

Probabilitas untuk tanggapan baru yang ingin diklasifikasikan ditentukan dengan menggunakan Persamaan 21. Nilai hasil perhitungan ditunjukkan oleh **tabel 12**.

**Tabel 12.** Nilai Probabilitas Tanggapan Baru

Kelas	tiket	pesawat	promo	murah	Nilai Probabilitas
Positif (P=0,5)	0.136	0.091	0.091	0.091	0.000051
Negatif (P=0,5)	0.15	0.05	0.05	0.05	0.000009

Hasil klasifikasi dari kelas tanggapan baru tersebut adalah kelas atau kategori yang memiliki probabilitas tertinggi. Berdasarkan **tabel 12** diketahui bahwa nilai probabilitas tertinggi adalah probabilitas kelas positif, maka tanggapan baru “tiket pesawat promo murah” masuk ke dalam kelas positif.

### 3.10. Evaluasi Performa

*Confusion matrix* merupakan metode yang digunakan sebagai evaluasi. Menurut Manning, dkk (Syafitri, 2017), “*Confusion matrix* merupakan salah satu *tools* penting dalam metode visualisasi yang digunakan pada mesin pembelajaran yang biasanya memuat dua kategori atau lebih”. Terdapat empat kemungkinan yang terjadi dalam evaluasi klasifikasi, yaitu *true positive* (TP) untuk data positif yang berhasil diprediksi positif, *false negative* (FN) untuk data positif yang diprediksi negatif. *True negative* (TN) untuk data negatif yang diprediksi negatif, dan *false positive* (FP) untuk data negatif yang diprediksi positif.

Hasil klasifikasi biner suatu dataset tersebut dipresentasikan oleh matriks berukuran 2x2 dan disebut *confusion matrix*, matriks tersebut ditunjukkan oleh **tabel 13**.

**Tabel 13. Confusion Matrix**

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>Positive</i>	<i>TP</i>	<i>FN</i>
	<i>Negative</i>	<i>FP</i>	<i>TN</i>

Terdapat beberapa nilai yang dapat digunakan untuk mengukur performa dalam klasifikasi. Menurut Rianto (Syafitri, 2017), hasil dari nilai akurasi, presisi, dan recall biasa ditampilkan dalam persentase.

1. *Accuracy*

Akurasi merupakan nilai proporsi prediksi yang benar. Menurut Lim, dkk (Syafitri, 2017), formula perhitungan nilai akurasi ditunjukkan oleh Persamaan 24.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (24)$$

2. *Precision*

*Precision* merupakan nilai proporsi jumlah dokumen yang relevan terkenal di antara semua dokumen teks yang terpilih oleh sistem. Menurut Lim, dkk (Syafitri, 2017), formula untuk menghitung nilai *precision* ditunjukkan oleh Persamaan 25.

$$Precision = \frac{TP}{TP+FP} \quad (25)$$

3. *Recall*

*Recall* merupakan nilai proporsi jumlah dokumen teks yang relevan terkenal di antara semua dokumen teks relevan yang ada pada koleksi. Menurut Lim, dkk (Syafitri, 2017), formula untuk menghitung nilai *recall* ditunjukkan oleh Persamaan 26.

$$Recall = \frac{TP}{TP+FN} \quad (26)$$



**Tabel 14.** Interpretasi Nilai Koefisien Korelasi

<b>Interval Koefisien</b>	<b>Tingkat Korelasi</b>
0-0,199	Sangat rendah
0,2-0,299	Rendah
0,3-0,599	Cukup
0,6-0,799	Kuat
0,8-1	Sangat Kuat

## BAB IV

### METODOLOGI PENELITIAN

#### 4.1. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah tanggapan mengenai “Traveloka” pada media sosial *twitter* selama satu minggu pada masing-masing bulan Maret dan April tahun 2018. Data tanggapan yang diperoleh sebanyak 1489 yang terdiri dari 732 tanggapan pada bulan Maret dan 757 tanggapan pada bulan April 2018.

Sampel dalam penelitian ini merupakan data yang digunakan pada fase *learning* (data *training*). Penentuan jumlah data *training* menggunakan rumus slovin, pada penelitian ini untuk seluruh data (data bulan Maret-April) dibagi ke dalam 10 kombinasi sehingga dengan menggunakan rumus slovin Persamaan 4 pada tingkat signifikansi 5% diperoleh jumlah data sampel (data *training*) seperti pada **tabel 15**.

**Tabel 15.** Jumlah Sampel (Data *Training*)

Bulan	N	Data Latih
Maret-April	150	109
	300	171
	450	212
	600	240
	750	261
	900	277
	1050	290
	1200	300
	1350	309
	1400	311
Maret	732	259
April	757	262

#### 4.2. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di perpustakaan pusat dan di Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia dengan memanfaatkan



buku referensi dan fasilitas internet yang tersedia. Waktu yang digunakan untuk penelitian ini mulai dari bulan Desember 2017 sampai bulan April 2018.

#### 4.3. Variabel Penelitian

Variabel dan definisi operasional variabel ditunjukkan pada **tabel 16**.

**Tabel 16.** Variabel dan Definisi Operasional Variabel

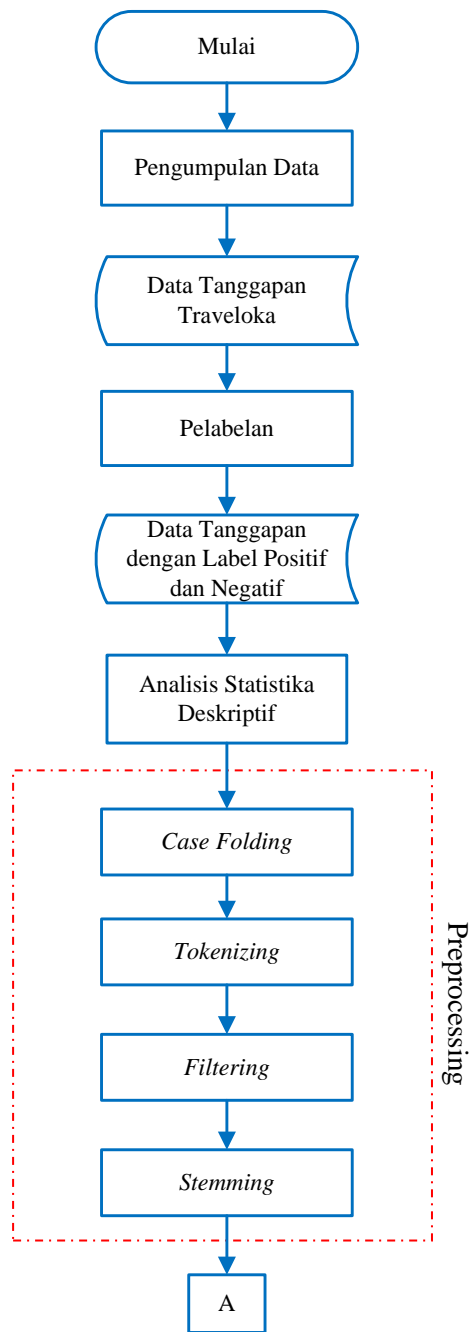
Variabel	Definisi Operasional Variabel
Teks	Teks yang dimaksud merupakan tanggapan mengenai Traveloka yang terdapat pada media sosial <i>Twitter</i> .
Label	Label yang dimaksud merupakan kategori dari tanggapan yang telah diperoleh. Peneliti mengkategorikan tanggapan ke dalam dua kelas, yaitu tanggapan positif dan tanggapan negatif.

#### 4.4. Metode Analisis

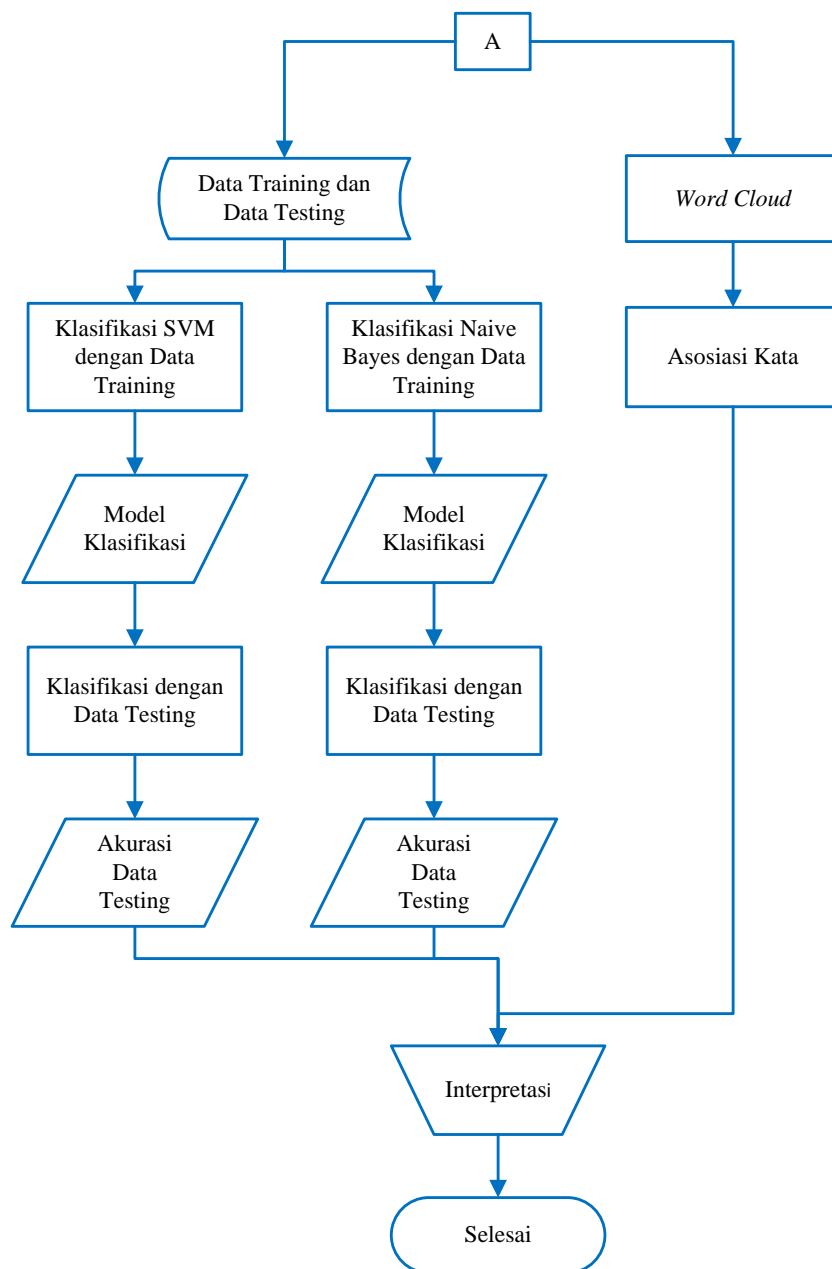
Penelitian ini menggunakan metode analisis Statistika Deskriptif untuk mendeskripsikan data tanggapan mengenai Traveloka pada media sosial *Twitter*. Klasifikasi dengan algoritma *Support Vector Machine* dan *Naive Bayes* untuk mengetahui metode klasifikasi yang memiliki akurasi tertinggi dalam melakukan klasifikasi tanggapan mengenai Traveloka, serta *Word Cloud* dan Asosiasi Kata untuk menampilkan kata yang sering muncul atau sering dibahas tentang Traveloka, sedangkan asosiasi kata digunakan untuk mengetahui kata yang mengikuti kata yang sering muncul tersebut. Alat analisis yang digunakan adalah *software Microsoft PowerPoint 2007, Microsoft Excel 2007 dan R 3.4.1*

#### 4.5. Tahapan Penelitian

*Flowchart* yang menunjukkan tahapan penelitian dari mulai pengumpulan data hingga diperoleh kesimpulan ditunjukkan oleh **gambar 15** dan **gambar 16**.



**Gambar 15.** Tahapan Penelitian



**Gambar 16.** Tahapan Penelitian (Lanjutan)

Tahapan dimulai dengan mengumpulkan data, data dikumpulkan melalui media sosial *Twitter* dengan menggunakan *keyword* “traveloka” kemudian data tersebut disimpan dan dilakukan pelabelan data secara manual oleh peneliti. Pelabelan data dibagi ke dalam dua kelas kategori yaitu tanggapan positif dan tanggapan negatif. Data hasil pelabelan tersebut disimpan dan dilakukan analisis statistika deskriptif untuk mengetahui persentase tanggapan positif dan tanggapan negatif terhadap Traveloka dari seluruh data yang diperoleh, selain itu juga

dilakukan analisis deskriptif pada tanggapan masing-masing bulan untuk mengetahui perbedaan tanggapan pada bulan Maret dan pada bulan April.

Setelah itu, data yang telah melalui proses pelabelan akan dilakukan *preprocessing* dimana tahapan *preprocessing* tersebut meliputi *case folding*, *tokenizing*, *filtering*, dan *stemming*. Data yang telah melalui tahapan *preprocessing* akan dibagi ke dalam data *training* dan data *testing*, data *training* digunakan untuk melatih *machine learning* agar dapat bekerja dengan baik dan *data testing* digunakan untuk menguji kinerja *machine learning*. Selain itu, data yang telah melalui tahapan *preprocessing* juga divisualisasikan ke dalam sebuah *word cloud*.

Data yang telah dibagi menjadi dua bagian yaitu data *training* dan data *testing* dilakukan klasifikasi menggunakan algoritma *Support Vector Machine* dan *Naive Bayes*. Proses klasifikasi dimulai dengan menggunakan data *training* sehingga diperoleh model klasifikasi, selanjutnya model klasifikasi yang telah diperoleh digunakan untuk melakukan klasifikasi pada data *testing* sehingga diperoleh nilai akurasi dari hasil klasifikasi.

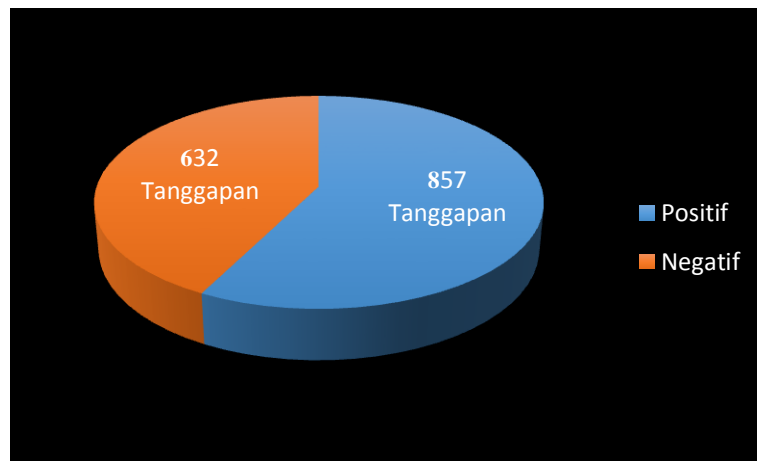
Berdasarkan data yang telah divisualisasikan ke dalam *word cloud* akan diketahui kata dominan yang sering dikaitkan dengan Traveloka dan dari beberapa kata dominan tersebut akan dilakukan asosiasi untuk mengetahui kata apa yang sering diperbincangkan bersamaan dengan kata dominan yang telah diketahui. Setelah semua tahapan telah terlewati, maka dilakukan interpretasi terhadap hasil yang diperoleh.

## BAB V

### HASIL DAN PEMBAHASAN

#### 5.1. Analisis Deskriptif

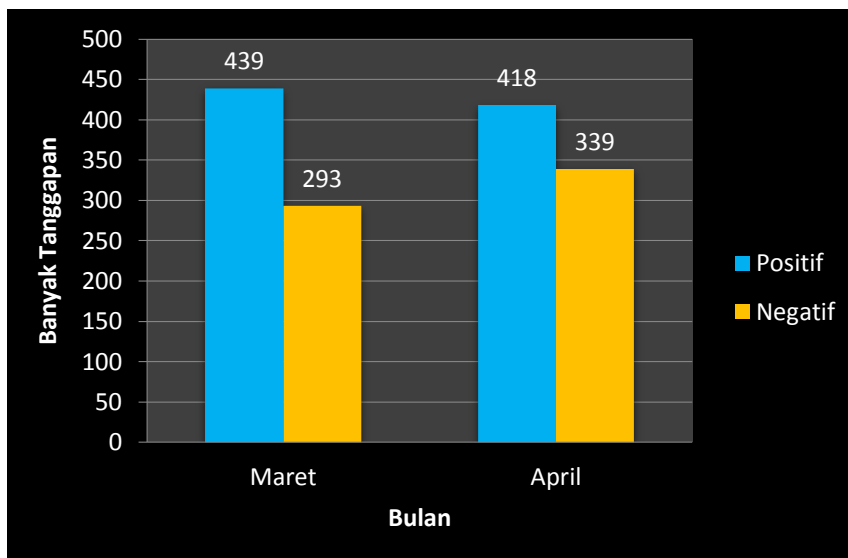
Analisis deskriptif digunakan untuk mengetahui gambaran umum mengenai data tanggapan tentang Traveloka yang telah diperoleh. Data tanggapan terdiri dari data tanggapan pada bulan Maret dan bulan April yang dikategorikan kedalam dua kelas kategori, yaitu tanggapan positif dan tanggapan negatif. Secara keseluruhan, gambaran mengenai tanggapan yang diperoleh berdasarkan masing-masing kategori ditunjukkan oleh **gambar 17**.



**Gambar 17.** Jumlah Tanggapan tentang Traveloka

Berdasarkan **gambar 17** diketahui bahwa dari total tanggapan yang diperoleh yaitu 1489 tanggapan, sebanyak 857 tanggapan merupakan tanggapan positif yaitu tanggapan yang mendukung atau menyukai Traveloka sedangkan 632 tanggapan merupakan tanggapan negatif yaitu tanggapan yang kurang menyukai suatu hal yang berhubungan dengan Traveloka.

Perbandingan tanggapan yang diperoleh pada **gambar 17** merupakan perbandingan dari seluruh data tanggapan yang diperoleh yaitu data tanggapan pada bulan Maret dan bulan April. Oleh karena itu, untuk melihat perbedaan perbandingan tanggapan pada masing-masing bulan, peneliti menampilkan perbandingan pada masing-masing bulan ke dalam **gambar 18**.



**Gambar 18.** Tanggapan tentang Traveloka pada Bulan Maret dan Bulan April

Berdasarkan **gambar 18** diketahui bahwa pada masing-masing bulan, jumlah tanggapan positif selalu lebih besar dibandingkan tanggapan negatif, hal tersebut mengindikasikan bahwa sebagian besar masyarakat pengguna *twitter* menyukai Traveloka. Pada bulan Maret dari 732 data tanggapan, sebanyak 439 tanggapan merupakan tanggapan positif dan sebanyak 293 tanggapan merupakan tanggapan negatif. Pada bulan April dari 757 tanggapan, sebanyak 418 tanggapan merupakan tanggapan positif dan sisanya sebanyak 339 tanggapan merupakan tanggapan negatif.

Secara umum jumlah tanggapan positif pada bulan Maret lebih besar dibandingkan tanggapan positif pada bulan April namun jumlah tanggapan yang diperoleh pada bulan April lebih besar daripada tanggapan yang diperoleh pada bulan Maret.

## 5.2. Tahap *Preprocessing*

Data dalam penelitian ini merupakan data *tweet* tanggapan yang diambil dari media sosial *Twitter* yang memiliki berbagai macam gaya penulisan sehingga data yang diperoleh merupakan data yang tidak terstruktur, oleh karena itu sebelum dilakukan klasifikasi, data perlu diubah menjadi data yang lebih terstruktur. Tahapan untuk mengubah data yang tidak terstruktur menjadi data yang lebih terstruktur disebut tahap *preprocessing*.

**Tabel 17.** *Tweet* Tanggapan Bulan Maret dan April 2018

No	Bulan	<i>Tweet</i> Tanggapan
1	Maret	Fitur Uangku traveloka Pay ini ribet banget. Katanya lagi promo, aku top up, udah tau terbatas waktu bayarnya, tinggal klik Pay pake verifikasi, lama pula.
2	Maret	Halo traveloka sedang promo
3	Maret	Lagi-lagi udah booking tiket traveloka bulak-balik direfund
4	Maret	Mau buka twitter kebukanya traveloka
5	Maret	Traveloka Tambahkan Layanan Pemesanan Tiket Bus, Traveling Semakin Mudah! <a href="https://t.co/oB477pgr1I">https://t.co/oB477pgr1I</a> via @TribunTravel
6	April	PrambananJazz tiket partner penjualan tiket nya ada traveloka kok belum ada ya di cek di app nya?
7	April	Ferry Unardi, sosok tangguh dibalik situs web Traveloka. Ferry cukup hebat memulai di industri e-ticketing pada usia 23 tahun.
8	April	Buka traveloka ngecek tiket balik minggu depan. Kog mahal ya
9	April	@ANGELVIP___ Lain kali pake traveloka aja.. tergiur dgn harga murah,
10	April	Langsung buka traveloka, cari promo tiket ke medan.

Pada **tabel 17** menunjukkan sebagian *tweet* tanggapan mengenai Traveloka yang diperoleh pada bulan Maret dan bulan April. *Tweet* tersebut diambil secara acak pada masing-masing bulan. Selanjutnya data *tweet* yang diperoleh melalui tahap *preprocessing* yang meliputi *case folding*, *tokenizing* dan *filtering*, serta *stemming*.

### 5.2.1 Case Folding

Pada proses *case folding*, semua huruf dirubah menjadi huruf kecil yang ditunjukkan dengan karakter yang diberi warna biru dan dilakukan penghapusan nomor (*remove number*) ditunjukkan oleh karakter yang diberi warna kuning, penghapusan tanda baca (*remove punctuation*) ditunjukkan oleh oleh yang diberi warna hijau, penghapusan *retweet/RT* (*remove retweet*) ditunjukkan oleh karakter berwarna coklat, penghapusan URL (*remove URL*) ditunjukkan oleh karakter berwarna abu-abu, dan penghapusan *username* (*remove username*) ditunjukkan oleh karakter yang siberi warna merah. Hasil *tweet* setelah melalui proses *case folding* ditunjukkan oleh **tabel 18**.

**Tabel 18.** *Tweet* Tanggapan Bulan Maret dan April 2018 Melalui Proses *Case Folding*

No	<i>Tweet</i> Tanggapan	<i>Tweet</i> Setelah Melalui Proses <i>Case Folding</i>
1	Fitur Uangku traveloka Pay ini ribet banget. Katanya lagi promo, aku top up, udah tau terbatas waktu bayarnya, tinggal klik Pay pake verifikasi, lama pula.	fitur uangku traveloka pay ini ribet banget katanya lagi promo aku top up udah tau terbatas waktu bayarnya tinggal klik pay pake verifikasi lama pula
2	Halo traveloka sedang promo	halo traveloka sedang promo
3	Lagi-lagi udah booking tiket traveloka bulak-balik direfund	lagilagi udah booking tiket traveloka bulakbalik direfund
4	Mau buka twitter kebukanya traveloka	mau buka twitter kebukanya traveloka
5	Traveloka Tambahkan Layanan Pemesanan Tiket Bus, Traveling Semakin Mudah! <a href="https://t.co/oB477pgr1I">https://t.co/oB477pgr1I</a> via @TribunTravel	traveloka tambahkan layanan pemesanan tiket bus traveling semakin mudah via
6	Prambananjazz tiket partner penjualan tiket nya ada traveloka kok belum ada ya di cek di app nya?	prambananjazz tiket partner penjualan tiket nya ada traveloka kok belum ada ya di cek di app nya
7	Ferry Unardi, sosok tangguh dibalik situs web Traveloka, Ferry cukup hebat memulai di industri e-ticketing pada usia 23 tahun.	ferry unardi sosok tangguh dibalik situs web traveloka ferry cukup hebat memulai di industri eticketing pada usia
8	Buka traveloka ngecek tiket balik minggu depan, Kog mahal ya <ed>U+00A0<U+00BD><ed>U+00B8<U+00B0>	buka traveloka ngecek tiket balik minggu depan kog mahal ya eduaubdedubub
9	RT @ANGELVIP Lain kali pake traveloka aja, tergiur dgn harga murah, ed>U+00A0<U+00BD><ed>U+00B8<U+0081>	lain kali pake traveloka aja tergiur dgn harga murah eduaubdedubu
10	Langsung buka traveloka, cari promo tiket ke medan.	langsung buka traveloka cari promo tiket ke medan

### 5.2.2 *Tokenizing dan Filtering*

Pada proses *tokenizing* kalimat akan dipecah berdasarkan kata-kata yang menyusunnya dan kata-kata tersebut akan berdiri sendiri, selanjutnya pada proses *filtering* kata-kata tersebut akan disaring dan dilakukan penghapusan kata yang tidak dibutuhkan menggunakan *stopwords (remove stopwords)* serta kata tambahan yang tidak dibutuhkan dan tidak terdapat di dalam *stopwords* yaitu kata “traveloka”, “kok”, dll.



**Tabel 19.** *Tweet* Tanggapan Bulan Maret dan April 2018 Melalui Proses *Case Folding, Tokenizing, dan Filtering*

No.	<i>Tweet</i> Setelah Melalui Proses <i>Case Folding</i>	<i>Tweet</i> Setelah Melalui Proses <i>Case Folding, Tokenizing, dan Filtering</i>
1	fitur uangku traveloka pay ini ribet banget katanya lagi promo aku top up udah tau terbatas waktu bayarnya tinggal klik pay pake verifikasi lama pula	fitur uangku pay ribet promo top up terbatas bayarnya tinggal klik pay verifikasi
2	halo traveloka sedang promo	promo
3	lagilagi udah booking tiket traveloka bulakbalik direfund	booking tiket bulakbalik direfund
4	mau buka twitter kebukanya traveloka	buka twitter kebukanya
5	traveloka tambahkan layanan pemesanan tiket bus traveling semakin mudah via	layanan pemesanan tiket bus traveling mudah
6	prambananjazz tiket partner penjualan nya ada traveloka kok belum ada ya di cek di app nya	prambananjazz tiket partner penjualan tiket cek app
7	ferry unardi sosok tangguh dibalik situs web traveloka ferry cukup hebat memulai di industri eticketing pada usia	ferry unardi sosok tangguh situs web ferry hebat industri eticketing
8	buka traveloka ngecek tiket balik minggu depan kog mahal ya eduaubdedubub	buka ngecek tiket minggu mahal
9	lain kali pake traveloka aja tergiur dgn harga murah eduaubdedubu	tergiur harga murah
10	langsung buka traveloka cari promo tiket ke medan	buka cari promo tiket medan

### 5.2.3 *Stemming*

Pada proses *stemming* dilakukan penghapusan imbuhan agar kata yang diperoleh merupakan kata dasar dan dapat direpresentasi sama dengan kata yang tidak mengandung imbuhan.

**Tabel 20.** *Tweet* Tanggapan Bulan Maret dan April 2018 Melalui Proses *Case Folding, Tokenizing, Filtering, dan Stemming*

No.	<i>Tweet</i> Setelah Melalui Proses <i>Case Folding, Tokenizing, dan Filtering</i>	<i>Tweet</i> Setelah Melalui Proses <i>Case Folding, Tokenizing, Filtering, dan Stemming</i>
1	fitur uangku pay ribet promo top up terbatas bayarnya tinggal klik pay verifikasi	fitur uang pay ribet promo top up batas bayar tinggal klik pay verifikasi
2	promo	promo
3	booking tiket bulakbalik direfund	booking tiket bulakbalik refund
4	buka twitter kebukanya	buka twitter buka
5	layanan pemesanan tiket bus traveling mudah	layan pesan tiket bus traveling mudah
6	prambananjazz tiket partner penjualan tiket cek app	prambananjazz tiket partner jual tiket cek app

**Tabel 21.** *Tweet* Tanggapan Bulan Maret dan April 2018 Melalui Proses *Case Folding, Tokenizing, Filtering, dan Stemming* (Lanjutan)

No.	<i>Tweet</i> Setelah Melalui Proses <i>Case Folding, Tokenizing, dan Filtering</i>	<i>Tweet</i> Setelah Melalui Proses <i>Case Folding, Tokenizing, Filtering, dan Stemming</i>
7	ferry unardi sosok tangguh situs web ferry hebat industri eticketing	ferry unardi sosok tangguh situs web ferry hebat industri eticketing
8	buka ngecek tiket minggu mahal	buka cek tiket minggu mahal
9	tergiur harga murah	giur harga murah
10	buka cari promo tiket medan	buka cari promo tiket medan

Hasil dari proses *preprocessing* yang terdapat pada kolom *tweet* setelah melalui proses *case folding, tokenizing, filtering, dan stemming* **tabel 20** dan **tabel 21** tersebut telah memiliki struktur yang lebih baik, sehingga akan mempermudah pemahaman *machine learning* pada proses klasifikasi.

### 5.3. Klasifikasi

Dalam penelitian ini klasifikasi dilakukan dengan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes*. Klasifikasi dilakukan untuk mengetahui algoritma apa yang paling baik digunakan untuk melakukan klasifikasi terhadap data tanggapan mengenai Traveloka. Total data tanggapan yang telah diperoleh dibagi ke dalam 10 kombinasi, untuk selanjutnya dilakukan klasifikasi dengan kedua algoritma yang telah ditentukan. Hasil klasifikasi dari tersebut disajikan dalam bentuk tabel *confusion matrix*

#### 5.3.1 Data Training dan Data Testing

Data *training* merupakan data yang digunakan pada fase *training*, sedangkan data uji merupakan data yang digunakan pada fase *testing*. Data *training* maupun data *testing* terbagi ke dalam dua kategori yaitu kategori tanggapan positif dan tanggapan negatif. Pelabelan tanggapan positif dan tanggapan negatif dilakukan secara manual oleh peneliti. Contoh pelabelan data ditunjukkan oleh **tabel 22**.

**Tabel 22.** *Tweet* Tanggapan Bulan Maret dan April 2018 Beserta Label

No.	Tweet Tanggapan	Label
1	Halo traveloka sedang promo	Positif
2	Traveloka Tambahkan Layanan Pemesanan Tiket Bus, Traveling Semakin Mudah! <a href="https://t.co/oB477pgr1I">https://t.co/oB477pgr1I</a> via @TribunTravel	
3	RT @ANGELVIP Lain kali pake traveloka aja.. tergiur dgn harga murah, ed><U+00A0><U+00BD><ed><U+00B8><U+0081>	
4	Buka traveloka ngecek tiket balik minggu depan. Kog mahal ya <ed><U+00A0><U+00BD><ed><U+00B8><U+00B0>	Negatif
5	Fitur Uangku @traveloka Pay ini ribet banget. Katanya lagi promo, udah tau terbatas waktu bayarnya, tinggal klik Pay pake verifikasi, lama pula.	
6	Mau buka twitter kebukanya traveloka	

Selanjutnya data yang telah diberikan label tersebut dibagi ke dalam data *training* dan data *testing*. Penentuan banyaknya data latih ditentukan dengan rumus Slovin pada Persamaan 4 menggunakan tingkat signifikansi 5%.

**Tabel 23.** Pembagian Data *Training* dan Data *Testing* *Tweet* Bulan Maret-April 2018

Bulan	N	Klasifikasi	Data Training	Data Testing	Jumlah
Maret-April	150	Positif	63	28	91
		Negatif	46	13	59
	<b>Total</b>		<b>109</b>	<b>41</b>	<b>150</b>
	300	Positif	105	73	178
		Negatif	66	56	122
	<b>Total</b>		<b>171</b>	<b>129</b>	<b>300</b>
	450	Positif	125	144	269
		Negatif	87	94	181
	<b>Total</b>		<b>212</b>	<b>238</b>	<b>450</b>
	600	Positif	139	205	344
		Negatif	101	155	256
	<b>Total</b>		<b>240</b>	<b>360</b>	<b>600</b>
	750	Positif	152	285	437
		Negatif	109	204	313
	<b>Total</b>		<b>261</b>	<b>489</b>	<b>750</b>
	900	Positif	162	362	524
Negatif		115	261	376	
<b>Total</b>		<b>277</b>	<b>623</b>	<b>900</b>	
1050	Positif	160	449	609	
	Negatif	130	311	441	
<b>Total</b>		<b>290</b>	<b>760</b>	<b>1050</b>	

**Tabel 24.** Pembagian Data *Training* dan Data *Testing Tweet* Bulan Maret-April 2018 (Lanjutan)

Bulan	N	Klasifikasi	Data Training	Data Testing	Jumlah
Maret-April	1200	Positif	176	519	695
		Negatif	124	381	505
	<b>Total</b>		<b>300</b>	<b>900</b>	<b>1200</b>
	1350	Positif	164	612	776
		Negatif	145	429	574
	<b>Total</b>		<b>309</b>	<b>1041</b>	<b>1350</b>
	1400	Positif	180	465	857
		Negatif	131	624	632
	<b>Total</b>		<b>311</b>	<b>1089</b>	<b>1400</b>

**Tabel 23** dan **Tabel 24** menunjukkan pembagian data *training* dan data *testing* pada total tanggapan (tanggapan bulan Maret-April), sedangkan pembagian data *training* dan data *testing* tanggapan bulan Maret ditunjukkan oleh **tabel 25** dan pembagian data *training* dan data *testing* tanggapan bulan April ditunjukkan oleh **tabel 26**.

**Tabel 25.** Pembagian Data *Training* dan Data *Testing Tweet* Bulan Maret 2018

Bulan	Label	Data Training	Data Testing	Jumlah
Maret	Positif	159	280	439
	Negatif	100	193	293
<b>Total</b>		<b>259</b>	<b>473</b>	<b>732</b>

Berdasarkan **tabel 25** diketahui bahwa untuk data tanggapan mengenai Traveloka pada bulan Maret yaitu sebanyak 732 tanggapan, data *training* yang digunakan sebanyak 259 yang terdiri dari 159 data tanggapan positif dan 100 data tanggapan negatif sedangkan pada fase *testing* data yang digunakan sebanyak 473 yang terdiri dari 280 data tanggapan positif dan 193 data tanggapan negatif.

**Tabel 26.** Pembagian Data *Training* dan Data *Testing Tweet* Bulan April 2018

Bulan	Label	Data Training	Data Testing	Jumlah
April	Positif	142	276	418
	Negatif	120	219	339
<b>Total</b>		<b>262</b>	<b>495</b>	<b>757</b>

Berdasarkan **tabel 26** diketahui data tanggapan mengenai Traveloka yang diperoleh pada bulan April sebanyak 757 data yang terdiri dari 418 data tanggapan positif dan 339 data tanggapan negatif. Data *training* yang digunakan sebanyak 262 yang terdiri dari data tanggapan positif sebanyak 142 dan data tanggapan negatif sebanyak 120. Total data *testing* sebanyak 495 terdiri dari 276 data tanggapan positif dan 219 data tanggapan negatif.

### 5.3.2 Perbandingan Kinerja *Support Vector Machine* dan *Naive Bayes*

Berikut hasil klasifikasi dengan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* pada 10 kombinasi yang telah ditentukan.

**Tabel 27.** *Confusion Matrix* Algoritma SVM 1

Predicted	Actual		Precision
	negatif	positif	
negatif	9	2	0,8182
positif	4	26	0,8667
<b>Recall</b>	0,6923	0,9286	
<b>Accuracy</b>	0,8537		

**Tabel 28.** *Confusion Matrix* Algoritma *Naive Bayes* 1

Predicted	Actual		Precision
	negatif	positif	
negatif	5	3	0,6250
positif	8	25	0,7576
<b>Recall</b>	0,3846	0,8929	
<b>Accuracy</b>	0,7317		

Berdasarkan **tabel 27** dan **tabel 28** diketahui bahwa jumlah data tanggapan positif mengenai Traveloka sebanyak 28 tanggapan, dengan menggunakan algoritma *Support Vector Machine* terdapat 26 tanggapan yang diklasifikasikan secara benar ke dalam kategori tanggapan positif dan 2 tanggapan diklasifikasikan ke dalam kategori tanggapan negatif sedangkan dengan algoritma klasifikasi *Naive Bayes* data tanggapan positif yang berhasil diklasifikasikan ke dalam kategori tanggapan positif sebanyak 25 data tanggapan dan 3 data yang lain diklasifikasikan ke dalam kategori tanggapan negatif.

Data tanggapan negatif mengenai Traveloka sebanyak 13 tanggapan, dengan algoritma *Support Vector Machine* terdapat 9 tanggapan yang berhasil

diklasifikasikan secara benar ke dalam kategori tanggapan negatif dan sisanya sebanyak 4 data diklasifikasikan ke dalam kelas tanggapan positif, sedangkan dengan algoritma *Naive Bayes* banyaknya data tanggapan negatif yang berhasil diklasifikasikan ke dalam kelas tanggapan negatif sebanyak 5 data tanggapan dan sisanya salah diklasifikasikan ke dalam kelas tanggapan positif. Nilai akurasi untuk algoritma *Support Vector Machine* sebesar 85,37% dan nilai akurasi algoritma *Naive Bayes* sebesar 73,17%.

Nilai *recall* yang diperoleh dengan menggunakan algoritma *Support Vector Machine* untuk data tanggapan positif sebesar 0,9286 dan untuk data tanggapan negatif sebesar 0,6923 sedangkan nilai *precision* yang diperoleh untuk data tanggapan positif sebesar 0,8667 dan untuk data tanggapan negatif sebesar 0,8182. Algoritma *Naive Bayes* menghasilkan nilai *recall* sebesar 0,8929 untuk data tanggapan positif dan 0,3846 untuk data tanggapan negatif, sedangkan nilai *precision* yang diperoleh sebesar 0,7576 untuk data tanggapan positif dan 0,6250 untuk data tanggapan negatif. Berikut perhitungan manual nilai akurasi, *recall*, dan *precision* yang telah diperoleh.

$$Akurasi_{SVM} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{26+9}{26+4+9+2} = 0,8537$$

$$Akurasi_{NBC} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{25+5}{25+8+5+3} = 0,7317$$

### **Kelas Positif**

$$Recall_{SVM} = \frac{TP}{TP+FN} = \frac{26}{26+2} = 0,9286$$

$$Recall_{NBC} = \frac{TP}{TP+FN} = \frac{25}{25+3} = 0,8929$$

$$Precision_{SVM} = \frac{TP}{TP+FP} = \frac{26}{26+4} = 0,8667$$

$$Precision_{NBC} = \frac{TP}{TP+FP} = \frac{25}{25+8} = 0,7576$$

### **Kelas Negatif**

$$Precision_{SVM} = \frac{TP}{TP+FP} = \frac{9}{9+2} = 0,8182$$

$$Precision_{NBC} = \frac{TP}{TP+FP} = \frac{5}{5+3} = 0,6250$$

$$Recall_{SVM} = \frac{TP}{TP+FN} = \frac{9}{9+4} = 0,6923$$

$$Recall_{NBC} = \frac{TP}{TP+FN} = \frac{5}{5+8} = 0,3846$$

**Tabel 29.** *Confusion Matrix* Algoritma SVM 2

Predicted	Actual		Precision
	negatif	positif	
negatif	32	5	0,8649
positif	24	68	0,7391
<b>Recall</b>	0,5714	0,9315	
<b>Accuracy</b>	0,7752		

**Tabel 30.** *Confusion Matrix* Algoritma Naive Bayes 2

Predicted	Actual		Precision
	negatif	positif	
negatif	28	8	0,7778
positif	28	65	0,6989
<b>Recall</b>	0,5000	0,8904	
<b>Accuracy</b>	0,7209		

Berdasarkan tabel *confusion matrix* **tabel 29** dan **tabel 30** diketahui bahwa dari 73 data tanggapan positif mengenai traveloka, menggunakan algoritma *Support Vector Machine* sebanyak 68 data tanggapan berhasil diklasifikasikan secara benar sebagai tanggapan positif dan 5 data tanggapan diklasifikasikan ke dalam tanggapan negatif sedangkan dengan algoritma *Naive Bayes* data tanggapan positif yang berhasil diklasifikasikan ke dalam kelas kategori positif sebanyak 65 data dan sisanya sebanyak 8 data tanggapan positif diklasifikasikan ke dalam kelas tanggapan negatif.

Pada data tanggapan negatif yaitu sebanyak 56 tanggapan, menggunakan algoritma *Support Vector Machine* sebanyak 32 data tanggapan berhasil diklasifikasikan secara benar ke dalam kelas tanggapan negatif dan 24 tanggapan diklasifikasikan ke dalam kelas tanggapan positif, sedangkan dengan menggunakan algoritma *Naive Bayes* terdapat 28 data tanggapan negatif yang berhasil diklasifikasikan dengan tepat ke dalam kelas tanggapan negatif dan sisanya sebanyak 28 data tanggapan negatif diklasifikasikan ke dalam kelas tanggapan positif.

Nilai akurasi yang diperoleh dengan menggunakan algoritma *Support Vector Machine* sebesar 77,52% dengan nilai *recall* untuk data tanggapan positif sebesar 0,9315 dan untuk data tanggapan negatif sebesar 0,5714 dan nilai *precision* sebesar 0,7391 untuk data tanggapan positif dan 0,8649 untuk data tanggapan negatif.

Algoritma *Naive Bayes* menghasilkan nilai akurasi sebesar 72,09% dan nilai *recall* yang diperoleh untuk data tanggapan positif sebesar 0,8904 dengan nilai *precision* yang dimiliki sebesar 0,6989, sedangkan untuk data tanggapan negatif memiliki nilai *recall* sebesar 0,5000 dan nilai *precision* sebesar 0,8904.

**Tabel 31.** *Confusion Matrix* Algoritma SVM 3

Predicted	Actual		Precision
	negatif	positif	
negatif	68	12	0,8500
positif	26	132	0,8354
<b>Recall</b>	0,7234	0,9167	
<b>Accuracy</b>	0,8403		

**Tabel 32.** *Confusion Matrix* Algoritma *Naive Bayes* 3

Predicted	Actual		Precision
	negatif	positif	
negatif	59	15	0,7973
positif	35	129	0,7866
<b>Recall</b>	0,6277	0,8958	
<b>Accuracy</b>	0,7899		

Berdasarkan **tabel 31** dan **tabel 32** diketahui bahwa nilai akurasi yang diperoleh dengan menggunakan algoritma *Support Vector Machine* sebesar 84,03% dan dengan algoritma *Naive Bayes* sebesar 78,99%. Jumlah data tanggapan positif mengenai Traveloka yang berhasil diklasifikasikan secara benar menggunakan algoritma *Support Vector Machine* sebanyak 132 tanggapan dari 144 tanggapan positif dan sisainya yaitu 12 tanggapan diklasifikasikan ke dalam kelas tanggapan negatif dengan nilai *recall* sebesar 0,9167 dan nilai *precision* sebesar 0,8354. Algoritma *Naive Bayes* berhasil mengklasifikasikan data tanggapan positif secara benar ke dalam kelas data tanggapan positif sebanyak 129 data dan 15 data tanggapan positif yang lain diklasifikasikan ke dalam kelas



tanggapan negatif sehingga nilai *recall* yang diperoleh sebesar 0,8958 dan nilai *precision* sebesar 0,7866.

Total data tanggapan negatif sebanyak 94 data tanggapan, menggunakan algoritma *Support Vector Machine* sebanyak 68 data tanggapan berhasil diklasifikasikan secara benar ke dalam kelas tanggapan negatif dan sisanya yaitu sebanyak 26 data tanggapan diklasifikasikan ke dalam kelas positif dengan nilai *recall* sebesar 0,7234 dan nilai *precision* sebesar 0,8500 sedangkan dengan algoritma *Naive Bayes* data tanggapan negatif yang berhasil diklasifikasikan secara benar ke dalam kelas tanggapan negatif sebanyak 59 dan sisanya sebanyak 35 data tanggapan negatif diklasifikasikan ke dalam kelas tanggapan positif dan menghasilkan nilai *recall* sebesar 0,6277 dan nilai *precision* sebesar 0,7973.

**Tabel 33.** *Confusion Matrix* Algoritma SVM 4

Predicted	Actual		Precision
	negatif	positif	
negatif	110	24	0,8209
positif	45	181	0,8009
<b>Recall</b>	0,7097	0,8829	
<b>Accuracy</b>	0,8083		

**Tabel 34.** *Confusion Matrix* Algoritma Naive Bayes 4

Predicted	Actual		Precision
	negatif	positif	
negatif	92	24	0,7931
positif	63	181	0,7418
<b>Recall</b>	0,5935	0,8829	
<b>Accuracy</b>	0,7583		

Berdasarkan tabel *confusion matrix* yang diperoleh diketahui bahwa dari jumlah data *testing* yang digunakan yaitu sebanyak 360 data, 205 data merupakan data tanggapan positif dan 155 data merupakan data tanggapan negatif. Data tanggapan positif sebanyak 205 tersebut dengan menggunakan algoritma *Support Vector Machine* sebanyak 181 data berhasil diklasifikasikan secara benar ke dalam kelas data tanggapan positif dan sisanya sebanyak 24 data tanggapan diklasifikasikan ke dalam tanggapan negatif sehingga diperoleh nilai *recall* sebesar 0,8829 dan nilai *precision* sebesar 0,8009. Algoritma klasifikasi *Naive*

*Bayes* berhasil mengklasifikasikan secara benar 181 data tanggapan positif kedalam kelas tanggapan positif dan sisanya sebanyak 24 data tanggapan positif diklasifikasikan kedalam kelas tanggapan negatif dengan nilai *recall* sebesar 0,8829 dan nilai *precision* sebesar 0,7418.

Data tanggapan negatif sebanyak 155, menggunakan algoritma *Support Vector Machine* berhasil diklasifikasikan secara benar ke dalam kelas tanggapan negatif sebanyak 110 data tanggapan dan sisanya sebanyak 45 data tanggapan diklasifikasikan ke dalam kelas tanggapan positif sedangkan dengan algoritma klasifikasi *Naive bayes* data tanggapan negatif yang berhasil diklasifikasikan ke dalam kelas tanggapan negatif sebanyak 92 data dan sisanya sebanyak 63 data tanggapan negatif diklasifikasikan ke dalam kelas tanggapan positif. Nilai *recall* tanggapan negatif dengan menggunakan algoritma *Support Vector Machine* sebesar 0,7097 dan nilai *precision* sebesar 0,8209 sedangkan dengan algoritma *Naive Bayes* pada data tanggapan negatif diperoleh nilai *recall* sebesar 0,5935 dan nilai *precision* sebesar 0,7931. Nilai akurasi yang diperoleh dengan algoritma *Support Vector Machine* sebesar 80,83% dan dengan algoritma *Naive Bayes* sebesar 75,83%.

**Tabel 35. Confusion Matrix Algoritma SVM 5**

Predicted	Actual		Precision
	negatif	positif	
negatif	120	43	0,7362
positif	84	242	0,7423
<b>Recall</b>	0,5882	0,8491	
<b>Accuracy</b>	0,7403		

**Tabel 36. Confusion Matrix Algoritma Naive Bayes 5**

Predicted	Actual		Precision
	negatif	positif	
negatif	123	24	0,8367
positif	81	261	0,7632
<b>Recall</b>	0,6029	0,9158	
<b>Accuracy</b>	0,7853		

Berdasarkan **tabel 35** dan **tabel 36** dengan menggunakan kombinasi kelima, yaitu jumlah data *training* sebanyak 261 dan jumlah data *testing* 489.

Hasil pengujian data *testing* dengan jumlah data tanggapan positif mengenai Traveloka sebanyak 285, dengan algoritma SVM diketahui bahwa data yang berhasil diklasifikasikan secara tepat kedalam kelas positif sebanyak 242 dan sisanya sebanyak 43 data diklasifikasikan ke dalam kelas negatif sehingga diperoleh nilai *recall* sebesar 0,8491 dan nilai *precision* sebesar 0,7423. Algoritma *Naive Bayes* berhasil mengklasifikasikan 261 data tanggapan positif secara benar dan sisanya sebanyak 24 salah diklasifikasikan dengan nilai *recall* sebesar 0,9158 dan nilai *precision* sebesar 0,7632.

Data tanggapan negatif sebanyak 204 dengan algoritma SVM sebanyak 120 data tanggapan berhasil diklasifikasikan secara benar dan sisanya diklasifikasikan ke dalam kelas positif. Data tanggapan negatif memiliki nilai *recall* pada algoritma SVM sebesar 0,5882 dan pada algoritma *Naive Bayes* sebesar 0,6029, sedangkan nilai *precision* pada algoritma SVM sebesar 0,7362 dan 0,8367 pada algoritma *Naive Bayes*. Nilai akurasi dengan menggunakan algoritma SVM sebesar 74,03 dan pada algoritma *Naive Bayes* sebesar 78,53.

**Tabel 37.** *Confusion Matrix* Algoritma SVM 6

Predicted	Actual		Precision
	negatif	positif	
negatif	191	46	0,8059
positif	70	316	0,8187
<b>Recall</b>	0,7318	0,8729	
<b>Accuracy</b>	0,8138		

**Tabel 38.** *Confusion Matrix* Algoritma *Naive Bayes* 6

Predicted	Actual		Precision
	negatif	positif	
negatif	150	39	0,7937
positif	111	323	0,7442
<b>Recall</b>	0,5747	0,8923	
<b>Accuracy</b>	0,7592		

Berdasarkan **tabel 37** dan **tabel 38** *confusion matrix* diketahui bahwa total tanggapan positif mengenai Traveloka sebanyak 362 tanggapan, dengan algoritma SVM terdapat 316 tanggapan yang berhasil diklasifikasikan secara benar ke dalam kelas tanggapan positif dan dengan algoritma *Naive Bayes* sebanyak 323 data

yang berhasil diklasifikasikan secara benar. Data tanggapan negatif mengenai Traveloka sebanyak 261, hanya 191 data yang berhasil diklasifikasikan secara benar ke dalam kelas tanggapan negatif dengan menggunakan algoritma SVM dan 150 data dengan algoritma *Naive Bayes* sedangkan sisanya diklasifikasikan ke dalam kelas positif.

Nilai akurasi untuk algoritma SVM dan algoritma *Naive Bayes* berturut-turut adalah 81,38% dan 75,92%. Pada algoritma SVM nilai *recall* untuk kelas data positif sebesar 0,8729 dan nilai *precision* sebesar 0,8187 sedangkan nilai *recall* pada kelas data negatif sebesar 0,7318 dan nilai *precision* sebesar 0,8059. Nilai *recall* algoritma *Naive Bayes* untuk data tanggapan positif sebesar 0,8923 dan nilai *precision* sebesar 0,7442 sedangkan untuk kelas data negatif nilai *recall* sebesar 0,5747 dan nilai *precision* yang dimiliki sebesar 0,7937.

**Tabel 39.** *Confusion Matrix* Algoritma SVM 7

Predicted	Actual		Precision
	negatif	positif	
negatif	219	52	0,8081
positif	92	397	0,8119
<b>Recall</b>	0,7042	0,8842	
<b>Accuracy</b>	0,8105		

**Tabel 40.** *Confusion Matrix* algoritma *Naive Bayes* 7

Predicted	Actual		Precision
	negatif	positif	
negatif	214	71	0,7509
positif	97	378	0,7958
<b>Recall</b>	0,6881	0,8419	
<b>Accuracy</b>	0,7789		

Berdasarkan **tabel 39** dan **tabel 40** diketahui jumlah data tanggapan positif mengenai Traveloka yang berhasil diklasifikasikan secara benar ke dalam kelas tanggapan positif dengan algoritma SVM sebanyak 397 data tanggapan dan dengan algoritma *Naive Bayes* sebanyak 378 data tanggapan sedangkan sisanya diklasifikasikan ke dalam data tanggapan negatif. Nilai *recall* dan *precision* algoritma SVM pada kelas data positif berturut-turut sebesar 0,8842 dan 0,8119 sedangkan pada algoritma *Naive Bayes* sebesar 0,8419 dan 0,7958.

Data tanggapan negatif yang berhasil diklasifikasikan ke dalam kelas data tanggapan negatif menggunakan algoritma SVM sebanyak 219 dan dengan algoritma *Naive Bayes* sebanyak 214 data, dengan nilai *recall* dan *precision* pada kelas data negatif berturut-turut pada algoritma SVM sebesar 0,7042 dan 0,8081 sedangkan pada algoritma *Naive Bayes* sebesar 0,6881 dan 0,7509. Nilai akurasi algoritma SVM lebih besar dibandingkan *Naive Bayes* yaitu 81,05% dan 77,89%.

**Tabel 41.** *Confusion Matrix* Algoritma SVM 8

Predicted	Actual		Precision
	negatif	positif	
negatif	233	48	0,8292
positif	148	471	0,7609
<b>Recall</b>	0,6115	0,9075	
<b>Accuracy</b>	0,7822		

**Tabel 42.** *Confusion Matrix* Algoritma *Naive Bayes* 8

Predicted	Actual		Precision
	negatif	positif	
negatif	204	46	0,8160
positif	177	473	0,7277
<b>Recall</b>	0,5354	0,9114	
<b>Accuracy</b>	0,7522		

Berdasarkan **tabel 41** dan **tabel 42** diketahui jumlah data tanggapan positif pada kombinasi delapan dengan menggunakan jumlah data *training* sebanyak 300 dan data *testing* sebanyak 900 diperoleh jumlah data positif yang berhasil diklasifikasikan ke dalam kelas tanggapan positif sebanyak 471 dengan menggunakan algoritma SVM dan sebanyak 473 dengan menggunakan algoritma *Naive Bayes* dari 519 data tanggapan positif. Nilai *recall* dan *precision* yang dihasilkan berturut-turut sebesar 0,9075 dan 0,7609 pada algoritma SVM dan 0,9114 dan 0,7277 pada algoritma *Naive Bayes*.

Data tanggapan negatif mengenai Traveloka sebanyak 381, sebanyak 233 data tanggapan yang berhasil diklasifikasikan secara benar ke dalam kelas data tanggapan negatif dengan menggunakan algoritma SVM dan 204 data menggunakan algoritma *Naive Bayes* sedangkan sisanya diklasifikasikan ke dalam kelas data tanggapan positif sehingga nilai *recall* dan *precision* yang diperoleh

secara berturut-turut sebesar 0,6115 dan 0,8292 dengan algoritma SVM dan dengan algoritma *Naive Bayes* sebesar 0,5354 dan 0,8160. Tingkat akurasi menggunakan algoritma SVM sebesar 78,22% dan dengan algoritma *Naive Bayes* sebesar 75,22%.

**Tabel 43.** *Confusion Matrix* Algoritma SVM 9

Predicted	Actual		Precision
	negatif	positif	
negatif	317	96	0,7676
positif	112	516	0,8217
<b>Recall</b>	0,7389	0,8431	
<b>Accuracy</b>	0,8002		

**Tabel 44.** *Confusion Matrix* Algoritma *Naive Bayes* 9

Predicted	Actual		Precision
	negatif	positif	
negatif	303	104	0,7445
positif	126	508	0,8013
<b>Recall</b>	0,7063	0,8301	
<b>Accuracy</b>	0,7791		

Berdasarkan tabel *confusion matrix* menggunakan data *training* sebesar 309 dan 1041 data *testing* diketahui data tanggapan positif yang digunakan pada fase *testing* tersebut sebanyak 612 data tanggapan positif dan 429 data tanggapan negatif. Data tanggapan positif tentang Traveloka yang berhasil diklasifikasikan secara benar ke dalam kelas tanggapan positif sebanyak 516 data dengan menggunakan algoritma SVM dan 508 data dengan menggunakan algoritma *Naive Bayes* sedangkan sisanya diklasifikasikan ke dalam kelas data negatif. Nilai *recall* dan *precision* yang diperoleh secara berturut-turut sebesar 0,8431 dan 0,8217 dengan algoritma SVM dan dengan algoritma *Naive Bayes* sebesar 0,8301 dan 0,8013.

Data tanggapan negatif yang berhasil diklasifikasikan ke dalam kelas tanggapan negatif sebanyak 317 data tanggapan dengan menggunakan algoritma SVM dan sebanyak 303 dengan menggunakan algoritma *Naive Bayes* sehingga diperoleh nilai *recall* untuk algoritma SVM dan *Naive Bayes* berturut-turut 0,7389 dan 0,7063 dan nilai *precision* sebesar 0,7676 dan 0,7445. Tingkat akurasi dengan

algoritma SVM sebesar 80,02% dan dengan algoritma *Naive Bayes* sebesar 77,91%.

**Tabel 45.** *Confusion Matrix* Algoritma SVM 10

Predicted	Actual		Precision
	negatif	positif	
negatif	331	72	0.8213
positif	134	552	0.8047
<b>Recall</b>	0.7118	0.8846	
<b>Accuracy</b>	0.8108		

**Tabel 46.** *Confusion Matrix* Algoritma *Naive Bayes* 10

Predicted	Actual		Precision
	negatif	positif	
negatif	274	52	0.8405
positif	191	572	0.7497
<b>Recall</b>	0.5892	0.9167	
<b>Accuracy</b>	0.7769		

Berdasarkan **tabel 45** dan **tabel 46** diketahui jumlah data tanggapan positif yang berhasil diklasifikasikan ke dalam kelas data tanggapan positif sebanyak 552 menggunakan algoritma SVM dan sebanyak 572 menggunakan algoritma *Naive Bayes*. Data tanggapan negatif yang berhasil diklasifikasikan ke dalam kelas data tanggapan negatif sebanyak 331 data tanggapan menggunakan algoritma SVM dan 274 tanggapan menggunakan algoritma *Naive Bayes*. Nilai *recall* dan *precision* yang diperoleh dengan menggunakan algoritma SVM adalah 0,8846 dan 0,8047 untuk data tanggapan positif dan 0,7118 dan 0,8213 untuk data tanggapan negatif, sedangkan nilai *recall* dan *precision* yang diperoleh dengan menggunakan algoritma *Naive Bayes* adalah 0,9167 dan 0,7497 untuk data tanggapan positif dan 0,5892 dan 0,8405 untuk data tanggapan negatif. Tingkat akurasi dengan menggunakan algoritma SVM sebesar 81,08% dan dengan algoritma *Naive Bayes* sebesar 77,69%.

Selanjutnya akan dilakukan perbandingan tingkat akurasi yang telah diperoleh dari tabel *confusion matrix*. Perbandingan nilai akurasi tersebut ditunjukkan oleh **tabel 47**.

**Tabel 47.** Perbandingan Tingkat Akurasi Algoritma SVM dan *Naive Bayes*

Kombinasi	N	Akurasi	
		SVM	Naive Bayes
1	150	85,37%	73,17%
2	300	77,52%	72,09%
3	450	84,03%	78,99%
4	600	80,83%	75,83%
5	750	74,03%	78,53%
6	900	81,38%	75,92%
7	1050	81,05%	77,89%
8	1200	78,22%	75,22%
9	1350	80,02%	77,91%
10	1400	81,08%	77,69%
<b>Rata-Rata</b>		<b>80,35%</b>	<b>76,32%</b>

Berdasarkan **tabel 47** diketahui bahwa dengan menggunakan 10 kombinasi jumlah data *training* dan data *testing* yang berbeda, algoritma *Support Vector Machine* memiliki tingkat akurasi yang lebih tinggi pada 9 kombinasi yang diberikan dibandingkan algoritma *Naive Bayes* dalam melakukan klasifikasi data tanggapan mengenai Traveloka, namun pada kombinasi kelima yaitu dengan jumlah data *training* sebanyak 261 dan data *testing* sebanyak 489 algoritma *Naive Bayes* memiliki tingkat akurasi lebih tinggi dibandingkan algoritma *Support Vector Machine*. Secara keseluruhan, rata-rata tingkat akurasi yang dihasilkan algoritma *Support Vector Machine* sebesar 80,35% dan *Naive Bayes* sebesar 76,32% maka dapat dikatakan bahwa dalam melakukan pengklasifikasian data tanggapan mengenai Traveloka yang diambil melalui media sosial *Twitter*, algoritma *Support Vector Machine* lebih baik dibandingkan *Naive Bayes*.

Berdasarkan hasil klasifikasi dengan menggunakan 10 kombinasi jumlah data *training* dan data *testing* yang berbeda **tabel 47** telah diketahui bahwa algoritma *Support Vector Machine* memiliki kinerja yang lebih baik dibandingkan algoritma *Naive Bayes*, namun agar hasil yang diberikan lebih terpercaya peneliti menambah 9 pengacakan data pada masing-masing kombinasi sehingga akan dihasilkan 10 nilai akurasi pada masing-masing kombinasi yang nantinya akan dihitung nilai rata-rata dari 10 pengacakan pada masing-masing kombinasi tersebut dan hasil dari nilai rata-rata tersebut yang akan dibandingkan untuk



menentukan kinerja terbaik antara algoritma *Support Vector Machine* dan *Naive Bayes*. Nilai akurasi dari 10 pengacakan data tanggapan untuk 10 kombinasi adalah sebagai berikut.

**Tabel 48.** Pengacakan Kombinasi 1 sampai Kombinasi 4

Pengacakan	Banyak Data							
	150		300		450		600	
	SVM	NB	SVM	NB	SVM	NB	SVM	NB
1	85,37%	73,17%	77,52%	72,09%	84,03%	78,99%	80,83%	75,83%
2	87,8%	75,61%	80,62%	72,87%	81,09%	74,79%	80,56%	75,56%
3	80,49%	73,17%	79,84%	75,19%	77,73%	73,53%	79,44%	75,56%
4	80,49%	78,05%	82,95%	75,19%	80,67%	75,21%	80,83%	76,67%
5	78,05%	78,05%	82,95%	77,52%	77,73%	73,11%	84,17%	73,61%
6	78,05%	73,17%	82,17%	73,64%	80,25%	76,89%	76,67%	75,83%
7	90,24%	78,05%	77,52%	72,87%	76,89%	74,37%	80,00%	75,83%
8	73,17%	73,17%	72,87%	64,34%	76,05%	75,21%	77,50%	77,50%
9	75,61%	63,41%	77,52%	65,89%	79,83%	74,37%	77,78%	73,89%
10	65,85%	46,34%	74,42%	67,44%	68,91%	73,53%	69,44%	77,78%
<b>Rata-Rata</b>	<b>79,51%</b>	<b>71,22%</b>	<b>78,84%</b>	<b>71,70%</b>	<b>78,32%</b>	<b>75,00%</b>	<b>78,72%</b>	<b>75,81%</b>

Berdasarkan **tabel 48** pada kombinasi pertama yaitu dengan jumlah data 150 diketahui bahwa algoritma SVM memiliki tingkat akurasi lebih tinggi dibandingkan algoritma *Naive Bayes* pada 8 pengacakan, sedangkan pada pengacakan kelima dan kedelapan tingkat akurasi yang dimiliki algoritma SVM dan *Naive Bayes* bernilai sama. Pada kombinasi kedua atau dengan jumlah data sebanyak 300 diketahui tingkat akurasi yang dihasilkan algoritma SVM pada setiap pengacakan selalu lebih besar dibandingkan algoritma *Naive Bayes*. Pada kombinasi ketiga dengan menggunakan jumlah data *training* dan data *testing* berturut-turut sebanyak 212 dan 238, diperoleh rata-rata tingkat akurasi algoritma SVM sebesar 78,32% dan tingkat akurasi algoritma *Naive Bayes* sebesar 75,00% sedangkan pada kombinasi keempat diketahui bahwa tingkat akurasi algoritma SVM lebih besar daripada algoritma *Naive Bayes* dengan rata-rata tingkat akurasi dari 10 pengacakan data yang berbeda sebesar 78,72%, sedangkan tingkat akurasi algoritma *Naive Bayes* sebesar 75,81%.

**Tabel 49.** Pengacakan Kombinasi 5 sampai Kombinasi 7

Pengacakan	Banyak Data					
	750		900		1050	
	SVM	NB	SVM	NB	SVM	NB
<b>1</b>	74,03%	78,53%	81,38%	75,92%	81,05%	77,89%
<b>2</b>	80,78%	75,46%	81,06%	76,24%	80,53%	77,63%
<b>3</b>	80,98%	77,71%	80,58%	76,73%	80,00%	76,32%
<b>4</b>	79,75%	76,48%	79,78%	76,57%	81,05%	76,32%
<b>5</b>	79,55%	74,23%	78,81%	75,12%	80,66%	76,32%
<b>6</b>	77,71%	77,51%	77,85%	76,4%	78,68%	78,68%
<b>7</b>	81,6%	77,3%	80,26%	77,53%	80,39%	76,97%
<b>8</b>	79,55%	79,14%	79,61%	80,58%	79,47%	79,21%
<b>9</b>	79,35%	73,82%	80,58%	75,6%	79,74%	74,74%
<b>10</b>	73,42%	79,14%	75,44%	78,97%	75,26%	79,87%
<b>Rata-Rata</b>	<b>78,67%</b>	<b>76,93%</b>	<b>79,54%</b>	<b>76,97%</b>	<b>79,68%</b>	<b>77,40%</b>

Berdasarkan **tabel 49** diketahui bahwa dengan jumlah data sebanyak 750 terdapat 2 tingkat akurasi SVM yang nilainya lebih kecil dibandingkan algoritma *Naive Bayes*, namun secara keseluruhan rata-rata tingkat akurasi algoritma SVM lebih besar dibandingkan algoritma *Naive Bayes*. Pada kombinasi keenam atau dengan jumlah data sebanyak 900 menunjukkan rata-rata tingkat akurasi dari kombinasi dengan jumlah data training sebanyak 277 data tanggapan dan data testing sebanyak 623 data tanggapan. Rata-rata tingkat akurasi dengan menggunakan algoritma SVM sebesar 79,54% dan dengan menggunakan algoritma *Naive Bayes* sebesar 76,97% sedangkan tingkat akurasi algoritma SVM pada kombinasi 7 dengan jumlah data sebanyak 1050 data tanggapan pada delapan pengacakan yang berbeda nilainya lebih besar dibandingkan algoritma *Naive Bayes*, sedangkan pada pengacakan keenam bernilai sama dan pada pengacakan kesepuluh tingkat akurasi algoritma SVM lebih rendah dibandingkan algoritma *Naive Bayes* dengan rata-rata tingkat akurasi algoritma SVM dan *Naive Bayes* berturut-turut sebesar 79,68% dan 77,40%.

**Tabel 50.** Pengacakan Kombinasi 8 sampai Kombinasi 10

Pengacakan	Banyak Data					
	1200		1350		1400	
	SVM	NB	SVM	NB	SVM	NB
<b>1</b>	78,22%	75,22%	80,02%	77,91%	81,08%	77,69%
<b>2</b>	80,11%	77,56%	81,08%	78,00%	79,8%	76,4%
<b>3</b>	79,44%	76,00%	79,63%	76,08%	80,17%	75,85%
<b>4</b>	80,56%	76,00%	79,83%	76,08%	78,88%	77,13%
<b>5</b>	74,11%	77,78%	78,96%	77,33%	78,7%	77,13%
<b>6</b>	78,67%	77,89%	78,19%	79,83%	79,71%	77,69%
<b>7</b>	81,22%	77,67%	79,15%	75,6%	78,97%	76,12%
<b>8</b>	79,67%	78,67%	79,63%	78,77%	81,36%	77,23%
<b>9</b>	79,78%	76,44%	80,5%	75,98%	75,11%	76,22%
<b>10</b>	78,22%	76,00%	79,15%	77,33%	81,18%	77,32%
<b>Rata-Rata</b>	<b>79,00%</b>	<b>76,92%</b>	<b>79,61%</b>	<b>77,29%</b>	<b>79,50%</b>	<b>76,88%</b>

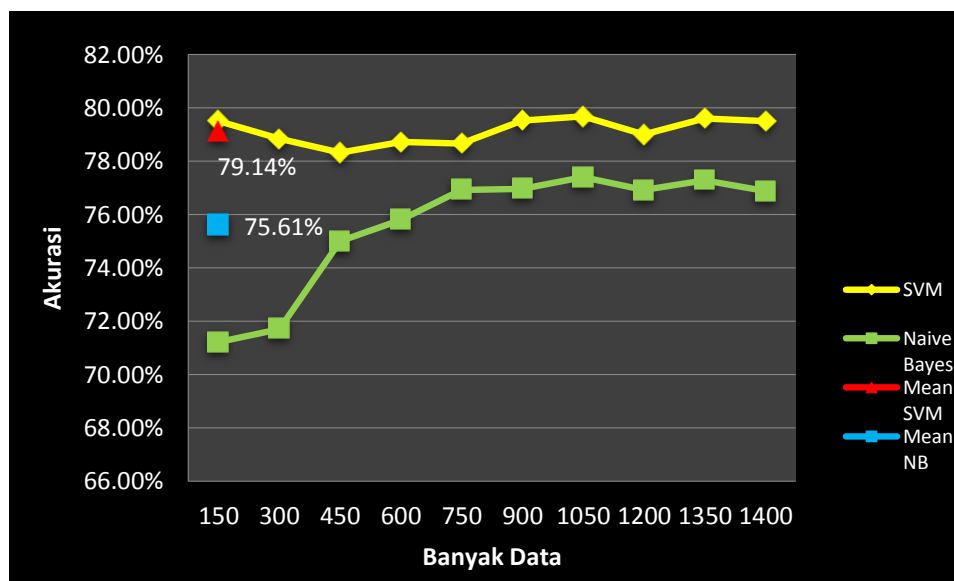
Berdasarkan **tabel 50** pada pengacakan dengan jumlah data yang digunakan sebanyak 1200 diketahui bahwa rata-rata tingkat akurasi algoritma SVM lebih besar dibandingkan algoritma *Naive Bayes* dengan rata-rata tingkat akurasi algoritma SVM sebesar 79,00% dan pada algoritma *Naive Bayes* sebesar 76,92%. Rata-rata tingkat akurasi algoritma SVM pada 10 pengacakan data yang berbeda dengan kombinasi jumlah data *training* dan data *testing* kesembilan atau jumlah data yang digunakan sebanyak 1350 sebesar 79,61% dan pada algoritma *Naive Bayes* sebesar 77,29%. Berdasarkan **tabel 50** juga diketahui bahwa dengan menggunakan 10 data acak yang berbeda, tingkat akurasi yang dimiliki algoritma SVM pada kombinasi kesepuluh bernilai lebih besar dibandingkan algoritma *Naive Bayes* kecuali pada pengacakan kesembilan, namun nilai rata-rata akurasi algoritma SVM lebih tinggi dibandingkan algoritma *Naive Bayes* dengan nilai sebesar 79,50% dan 76,88%.

Berdasarkan nilai rata-rata hasil pengacakan pada 10 kombinasi jumlah data *training* dan data *testing* tersebut akan dilakukan perbandingan tingkat akurasi algoritma SVM dan algoritma *Naive Bayes*. Perbandingan tingkat akurasi tersebut ditunjukkan oleh **tabel 51**.

**Tabel 51.** Perbandingan Rata-Rata Akurasi SVM dan *Naive Bayes*

Banyak Data	SVM	<i>Naive Bayes</i>
150	79.51%	71.22%
300	78.84%	71.70%
450	78.32%	75.00%
600	78.72%	75.81%
750	78.67%	76.93%
900	79.54%	76.97%
1050	79.68%	77.40%
1200	79.00%	76.92%
1350	79.61%	77.29%
1400	79.50%	76.88%
<b>Rata-Rata</b>	<b>79.14%</b>	<b>75.61%</b>

Berdasarkan **tabel 51** tersebut dapat disajikan dalam bentuk grafik **gambar 19**.

**Gambar 19.** Grafik Perbandingan Akurasi SVM dan *Naive Bayes*

Berdasarkan **gambar 19** diketahui bahwa tingkat akurasi algoritma SVM lebih tinggi dibandingkan algoritma *Naive Bayes*. Tingkat akurasi algoritma SVM berkisar antara 78%-80% atau rata-rata tingkat akurasi sebesar 79,14% sedangkan tingkat akurasi algoritma *Naive Bayes* hanya berkisar diantara 70%-78% atau rata-rata sebesar 75,61%. Algoritma *Naive Bayes* cenderung mengalami peningkatan nilai akurasi ketika jumlah data yang digunakan semakin banyak meskipun tingkat

akurasinya tidak lebih besar dari algoritma SVM. Oleh karena itu, dapat disimpulkan bahwa algoritma *Support Vector Machine* memiliki kinerja lebih baik dalam melakukan klasifikasi data tanggapan mengenai Traveloka melalui media sosial *Twitter* dibandingkan algoritma *Naive Bayes*.

### 5.3.3 Perbandingan Kinerja Klasifikasi Tanggapan pada bulan Maret dan bulan April

Perbandingan kinerja akan difokuskan pada tingkat akurasi dan besarnya nilai *error*. Dalam menentukan algoritma yang memiliki kinerja terbaik didasarkan pada tingkat akurasi yang tinggi dan didukung dengan nilai *error* yang rendah. Perbandingan kinerja algoritma *Support Vector Machine* dan *Naive Bayes* ditunjukkan oleh **tabel 52**.

**Tabel 52.** Perbandingan Tingkat Akurasi Algoritma SVM dan *Naive Bayes* Data Bulan Maret dan April

Algoritma	Accuracy		Error		Correctly Classified		Incorrectly Classified	
	Maret	April	Maret	April	Maret	April	Maret	April
<i>Support Vector Machine</i>	83,51%	79,60%	16,49%	20,40%	395	394	78	101
<i>Naive Bayes</i>	79,28%	76,57%	20,72%	23,43%	375	379	98	116

Berdasarkan **tabel 52** diketahui bahwa algoritma *Support Vector Machine* pada pengklasifikasian data tanggapan mengenai Traveloka pada bulan Maret memiliki tingkat akurasi sebesar 83,51% atau data yang dapat diklasifikasikan secara benar sebanyak 395 dari 473 data dengan nilai *error* sebesar 16,49% atau data yang tidak dapat diklasifikasikan secara benar sebanyak 78 data. Pada pengklasifikasian data tanggapan bulan April tingkat akurasi algoritma *Support Vector Machine* sebesar 79,60% atau sebanyak 394 data berhasil diklasifikasikan dengan benar dan nilai *error* sebesar 20,40% atau sebanyak 101 data tidak dapat diklasifikasikan secara benar dari 495 data.

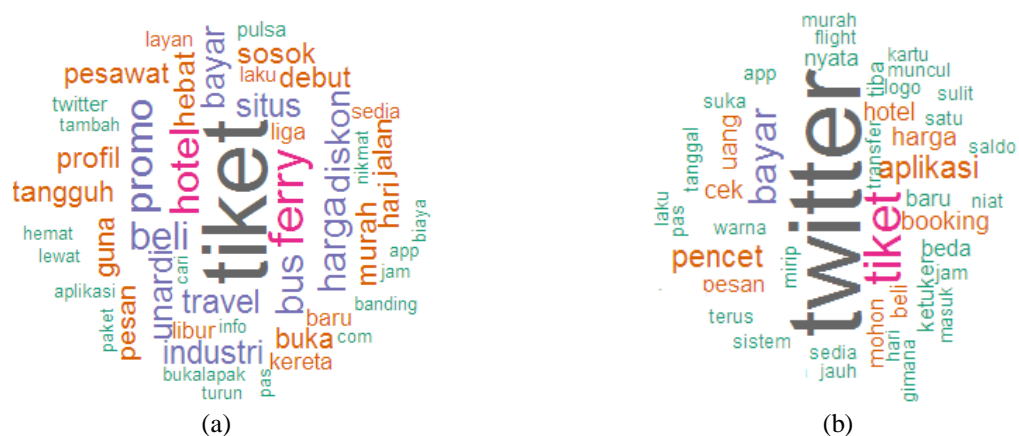
Algoritma *Naive Bayes* berhasil melakukan klasifikasi data tanggapan mengenai Traveloka pada bulan Maret dengan tingkat akurasi sebesar 79,28% dan nilai *error* sebesar 20,72% atau data yang berhasil diklasifikasikan secara benar sebanyak 375 dan data yang tidak dapat diklasifikasikan secara benar sebanyak 98

data. Pada pengklasifikasian data tanggapan bulan April, algoritma *Naive Bayes* memiliki tingkat akurasi sebesar 76,57% dan nilai *error* sebesar 23,43% atau sebanyak 379 data dapat diklasifikasikan secara benar dan 116 data tidak dapat diklasifikasikan secara benar.

Secara garis besar dapat dikatakan bahwa baik pada pengklasifikasian data tanggapan mengenai Traveloka pada bulan Maret maupun data tanggapan Traveloka pada bulan April, metode *Support Vector Machine* memiliki tingkat akurasi yang lebih tinggi dibandingkan metode *Naive Bayes* dengan kata lain tidak terdapat perbedaan kinerja *machine learning* dalam melakukan klasifikasi tanggapan pada bulan Maret dan bulan April 2018.

#### 5.4. *Word Cloud*

*Word cloud* merupakan salah satu bentuk penyajian data yang berupa teks, melalui *word cloud* akan diketahui kata dominan yang sering diperbincangkan mengenai Traveloka. Pada penelitian ini akan ditampilkan *word cloud* untuk data tanggapan mengenai Traveloka menurut kelas sentimennya yaitu positif dan negatif secara keseluruhan (data bulan Maret-April) dan pada masing-masing bulan yaitu bulan Maret dan bulan April.



**Gambar 20.** *Word Cloud* Data Keseluruhan

Pada **gambar 20 (a)** merupakan *word cloud* dari keseluruhan tanggapan mengenai Traveloka yang diperoleh pada bulan Maret dan April 2018 melalui media sosial Twitter serta masuk kedalam kategori kelas positif, sedangkan **(b)** merupakan *word cloud* untuk tanggapan yang masuk kedalam kategori tanggapan







**Tabel 53.** Asosiasi Data Keseluruhan Bulan Maret-April

Tiket (Positif)		Beli (Positif)		Twitter (Negatif)		Tiket (Negatif)		
pesawat	0,40	airasia	0,32	ketuker	0,29	pesawat	0,22	
pesan	0,26	paket	0,24	mirip	0,20	beda	0,17	
kereta	0,25	token	0,20	logo	0,19	kereta	0,15	
layan	0,24	kereta	0,19	mencet	0,19	mahal	0,15	
Ferry(Positif)		Bus (Positif)		Aplikasi (Negatif)		selisih		0,15
debut	1,00	layan	0,49	twitter	0,33	prambananjazz	0,15	
hebat	1,00	antarkota	0,44	ketuker	0,24	cetak	0,15	
profil	1,00	pesan	0,41	loading	0,15	Bayar (Negatif)		
sosok	1,00	tambah	0,35	dua	0,15	metode	0,32	
tangguh	1,00	sedia	0,34	Harga (Negatif)		jumlah	0,27	
unardi	0,98	Harga (Positif)		website	0,46	tetap	0,24	
Promo (Positif)		turun	0,52	beda	0,43	kendala	0,23	
maret	0,39	banding	0,41	maskapai	0,37	bpjs	0,21	
periode	0,39	website	0,30	janggal	0,66	transaksi	0,18	
menarik	0,33	good	0,21	selisih	0,66	pulsa	0,17	
link	0,25	nice	0,21					
kode	0,24	potong	0,17					
kunjung	0,22	Hotel (Positif)						
informasi	0,21	rekomendasi	0,27					
lokal	0,19	review	0,21					
		jangkau	0,18					
		recomended	0,15					

Berdasarkan **tabel 53** diketahui kata yang sering diperbincangkan bersamaan dengan kata “tiket”, “beli”, “ferry”, “bus”, “harga”, “promo”, dan “hotel” pada kategori data tanggapan positif mengenai Traveloka. Berdasarkan nilai asosiasi tersebut informasi yang dapat diperoleh untuk tanggapan positif antara lain pengguna *twitter* merasa senang atau setuju dengan adanya layanan pemesanan tiket pesawat dan kereta, pengguna *twitter* juga setuju dengan kehebatan sosok Ferry Unardi yang merupakan pendiri Traveloka. Promo yang sering pengguna *twitter* perbincangkan di media sosial adalah promo pada bulan maret yang menarik yang disertai dengan *link* dan informasi promo lainnya. Pengguna *twitter* juga senang dengan layanan penambahan bus antarkota, selain itu ada juga pengguna *twitter* yang membandingkan harga pada aplikasi Traveloka

dengan aplikasi lainnya, serta untuk hotel banyak diperbincangkan mengenai keterjangkauan dan rekomendasi.

Pada tanggapan negatif banyak diperbincangkan hal yang berkaitan dengan *twitter* dimana pelanggan Traveloka sering mengalami kesalahan pilih ketika ingin membuka Traveloka justru yang terbuka adalah *twitter* atau sebaliknya dikarenakan terdapat kemiripan logo aplikasi *twitter* dan Traveloka, selain itu aplikasi Traveloka sering dikeluhkan dalam hal *loading*. Tanggapan negatif lainnya muncul pada kata harga yang sering dikaitkan dengan perbedaan antara *website* Traveloka dengan maskapai, hal ini bisa dikarenakan harga tiket yang dapat berubah sewaktu-waktu. Selain itu, tiket prambananjazz dan metode pembayaran. Pembayaran sering mengalami keluhan atau kendala pada pembayaran bpjs dan pulsa.

**Tabel 54.** Asosiasi Data Bulan Maret

Tiket (Positif)		Promo (Positif)		Tiket (Negatif)		Beli (Negatif)	
layan	0,28	periode	0,55	cari	0,29	paket	0,43
tambah	0,27	tarik	0,43	pesawat	0,24	internet	0,31
sedia	0,26	kode	0,42	bus	0,19	masuk	0,28
kereta	0,26	maret	0,38	ganggu	0,19	konfirmasi	0,24
cetak	0,26	lebaran	0,38	iklan	0,19	respon	0,22
lebaran	0,24	<b>Diskon (Positif)</b>		jual	0,19	<b>Aplikasi (Negatif)</b>	
antarkota	0,20	istimewa	0,28	kapal	0,19	access	0,41
bis	0,17	untung	0,24	laut	0,19	loading	0,34
mudik	0,17	murah	0,20	<b>Bayar (Negatif)</b>		browser	0,24
<b>Bus (Positif)</b>		hemat	0,19	versi	0,30	sabar	0,24
layan	0,47	<b>Beli (Positif)</b>		hangus	0,27	ribet	0,16
tambah	0,45	paket	0,28	transaksi	0,27	uangkuid	0,16
sedia	0,45	desak	0,25	internet	0,27		
antarkota	0,33	token	0,25	pemberitahuan	0,19		
<b>Hotel (Positif)</b>		listrik	0,20	handphone	0,19		
rekomendasi	0,24	api	0,20				
murah	0,23	kereta	0,19				
untung	0,20						

Berdasarkan **tabel 54** diketahui bahwa tanggapan mengenai Traveloka pada bulan Maret 2018, hal positif banyak memperbincangkan layanan penyediaan tambahan layanan tiket, tiket kereta dan bis, serta tiket untuk mudik

lebaran. Promo dan diskon juga sering disebutkan dalam tanggapan pada bulan Maret 2018, diskon pada bulan maret dianggap istimewa, sedangkan tanggapan negatif sering disebutkan dengan kata tiket yang memuat iklan dan keluhan karena tidak terdapat tiket kapal laut. Tanggapan negatif lainnya mengenai jangka waktu pembayaran yang terlalu pendek sehingga pemesanan hangus, karena memang untuk situs Traveloka pembayaran hanya diberikan batas 24 jam sejak pemesanan dilakukan. Aplikasi uangkuid yang merupakan aplikasi yang bekerjasama dengan Traveloka untuk melakukan pembayaran juga sering dikeluhkan.

**Tabel 55.** Asosiasi Data Bulan April

Ferry (Positif)		Twitter (Negatif)		Bayar (Negatif)	
tanggguh	1	ketuker	0,32	kendala	0,38
unardi	1	suka	0,27	metode	0,34
Tiket (Positif)		pencet	0,25	bpjs	0,24
pesawat	0,36	niat	0,23	kartu	0,24
pesan	0,36	logo	0,19	kredit	0,24
bagus	0,35	mirip	0,19	uangkuid	0,24
turun	0,28	jauh	0,18	sistem	0,12
antarkota	0,25	layar	0,18	durasi	0,11
beli	0,24	posisi	0,18	Aplikasi (Negatif)	
kereta	0,23	nyasar	0,18	ketuker	0,39
bus	0,23	lambang	0,16	twitter	0,13
Harga (Positif)		biru	0,25	koneksi	0,12
turun	0,55	keliru	0,15	sinyal	0,12
banding	0,39	Tiket (Negatif)		install	0,06
website	0,37	beda	0,28	mirip	0,01
rekomendasi	0,20	kereta	0,22		
		pesawat	0,21		
		mahal	0,20		
		prambananjazz	0,18		

Tanggapan positif mengenai Traveloka pada bulan April sering dikaitkan dengan ketangguhan Ferry Unardi yang merupakan pendiri Traveloka, pemesanan tiket, serta rekomendasi perbandingan harga. Tanggapan negatif sering dikaitkan dengan *twitter* yang memiliki kesamaan logo sehingga pelanggan Traveloka sering salah pilih atau salah pencet antara Traveloka dan *twitter*, perbedaan tiket juga muncul sebagai tanggapan negatif, selain itu kendala dalam metode pembayaran juga muncul sebagai tanggapan negatif.

## BAB VI

### KESIMPULAN DAN SARAN

#### 6.1. Kesimpulan

Berdasarkan hasil analisis dan rumusan masalah yang telah disebutkan, diperoleh kesimpulan untuk menjawab rumusan masalah tersebut, yaitu:

1. Berdasarkan tingkat akurasi yang dihasilkan, *machine learning* dengan menggunakan algoritma *Support Vector Machine* memiliki kinerja yang lebih baik dibanding algoritma *Naive Bayes* dalam melakukan klasifikasi tanggapan mengenai Traveloka yang diperoleh melalui media sosial *Twitter*.
2. Tidak terdapat perbedaan kinerja *machine learning*, baik ketika diaplikasikan pada data keseluruhan (bulan Maret-April), data bulan Maret, maupun data bulan April 2018 kinerja *machine learning* dengan menggunakan algoritma *Support Vector Machine* memiliki tingkat akurasi lebih tinggi dibandingkan algoritma *Naive Bayes*.
- 3.a. Tanggapan positif sering dikaitkan dengan Ferry Unardi, penambahan layanan tiket pesawat, kereta dan bis, layanan penambahan bus antarkota, serta promo dan diskon istimewa.
  - b. Tanggapan negatif sering dikaitkan dengan *twitter* yang memiliki kesamaan logo, perbedaan harga antara *website* Traveloka dengan maskapai, iklan yang cukup mengganggu, serta kendala jangka waktu pembayaran yang terlalu pendek sehingga pemesanan hangus.

#### 6.2. Saran

1. Bagi Traveloka, berdasarkan hasil analisis yang dilakukan diharapkan dapat dilakukan perbaikan-perbaikan terutama hal yang paling sering dikeluhkan, seperti logo yang mirip dengan aplikasi *twitter*, perbedaan harga antara *website* Traveloka dengan maskapai, iklan yang cukup mengganggu, serta kendala jangka waktu pembayaran yang terlalu pendek sehingga pemesanan hangus.

2. Bagi peneliti selanjutnya, sebaiknya digunakan algoritma klasifikasi lain agar dapat diketahui perbandingan kinerja algoritma tersebut dibandingkan kedua algoritma yang telah digunakan, jumlah kombinasi dan pengacakan yang lebih banyak, serta dapat memperbanyak periode waktu pengumpulan data agar informasi yang diperoleh lebih mendalam.

## DAFTAR PUSTAKA

- Antariksa, Y. (2017). *Kenapa Layanan Online Trael Traveloka Makin Melesat?* Strategimanajemen.net.
- Ariadi, D., & Fithriasari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *Jurnal Sains dan Seni ITS* , 2337-3520.
- Badan Pusat Statistika. (2010). Dipetik 2018, dari <http://sp2010.bps.go.id/>
- Bohang, F. K. (2017). *Twitter 280 Karakter Resmi di Seluruh Dunia*. Kompas.com.
- Damanik, R. M. (2014). Pembangunan Aplikasi Pencarian Dokumen Menggunakan Text Mining Berbasis Web. *Jurnal Universitas Atma Jaya Yogyakarta* .
- Effendy, O. U. (2001). *Ilmu Komunikasi Teori dan Praktek*. Bandung: PT Remaja Rosdakarya.
- Erwin. (2014). *Tanggapan Mahasiswa Ilmu Komunikasi Universitas Hasanuddin Terhadap Program Hitam Putih di Trans 7*. Skripsi Program Studi Komunikasi, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Hasanuddin.
- Farid, M. H. (2017). *Pengaruh Kualitas Pelayanan dan Perpsepsi Harga Terhadap Kepuasan Pelanggan dalam Meningkatkan Loyalitas Pelanggan Pengguna Aplikasi Traveloka Berbasis Internet (Studi Pada Pengguna Aplikasi Online Traveloka Universitas Lampung)*. Skripsi Program Studi Ilmu Administrasi Bisnis, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Lampung.
- Ghifary, R. A. (2018). *Analisis Kualitas Layanan pada Perusahaan E-Commerce Traveloka*. Skripsi Program Studi Manajemen, Fakultas Ekonomi, Universitas Islam Indonesia.
- Hadi, A. F., Bagus C. W., D., & Hasan, M. (2017). Text Mining pada Media Sosial Twitter Studi Kasus: Masa Tenang Pilkada DKI 2017 Putaran 2. *Seminar Nasional Matematika dan Aplikasinya* (hal. 324-331). Surabaya: math.fst.unair.ac.id.
- Han, J., & Kamber, M. (2006). *Data Mining : Concepts and Techniques Second Edition*. USA: Elsevier.
- Hartanti, D. C. (2013). *Tanggapan Masyarakat Pengguna Terminal Terhadap Relokasi Terminal Dhaksinarga Baru di Desa Selang Kecamatan*

Wonosari Kabupaten Gunungkidul. Skripsi Program Studi Pendidikan Geografi, Fakultas Ilmu Sosial, Universitas Negeri Yogyakarta.

- Hasan, I. (1999). *Pokok-Pokok Materi Statistik 1 [Statistik Deskriptif]*. Jakarta: Bumi Aksara.
- Khalidi, F. (2014). *Ferry Unardi, Mengibarkan Traveloka dari Titik Nol*. swa.co.id.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Masithoh, N. (2016). *Analisis Klasifikasi Topik Menggunakan Metode Naive Bayes Classifier, Naive Bayes Multinomial Classifier, dan Maximum entropy pada Artikel Berita*. Skripsi Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada.
- Naradhipa, A. R., & Purwarianti, A. (2012). Sentiment Classification for Indonesian Message in Social Media. *Cloud Computing and Social Networking (ICCCSN), 2012 International Conference* (hal. 1-5). IEEE.
- Nuraini, I. S., & Mudiantono. (2017). Faktor-Faktor yang Mempengaruhi Kepuasan Pelanggan dan Dampaknya Terhadap Minat Beli Ulang Pada Situs Traveloka (Studi Pada Konsumen Traveloka di Kota Semarang). *Diponegoro Journal Of Management* , 1-11.
- Pakar Online Indonesia. (2012,12 30). *Mengenal Twitter dan Daftar Istilah yang Digunakan*. Diambil kembali dari PakarOnline.com: <https://www.pakaronline.com/twitter/mengenal-twitter-dan-daftar-istilah-yang-digunakan/>
- Pertiwi, W. K. (2018). *Riset Ungkap Pola Pemakaian Medsos Orang Indonesia*. Kompas.com.
- Pramudiono, I. (2003). *Pengantar Data Mining: Menambang Permata Pengetahuandi Gunung Data. Materi Kuliah Umum IlmuKomputer.com*. www.academia.edu.
- Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: ANDI.
- Pratiwi, Y. R., & Widodo, E. (2017). Comparison Of Maximum Entropy and Support Vector Machine Methods For Sentiment Analysis Of Peralite Product Through Twitter Social Network. *International Journal of Advances in Electronics and Computer Science* , 10-14.
- Rasyad, R. (2003). *Metode Statistik Deskriptif untuk Umum*. Jakarta: Grasindo.
- Risyandi, O., & Zuliestiana, D. A. (2017). Pengaruh Kualitas Website Traveloka Terhadap Kepuasan Pengguna. *e-Proceeding of Management*, (hal. 2244).

- Saleh, A. (2015). Implementasi Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal* , 207-217.
- Samsubar, S. (1990). *Statistik Deskriptif*. Yogyakarta: AMP-YKPN.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Santoso, D., Ratnawati, D. E., & Indriati. (2014). Perbandingan Kinerja Metode Naive Bayes, K-Nearest Neighbor, dan Metode Gabungan K-Means dan LVQ dalam Pengkategorian Buku Komputer Berbahasa Indonesia. *Mahasiswa PTIIK UB* .
- Susilowati, E., Sabariah, M. K., & Gozali, A. A. (2015). Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas pada Twitter. *e-Proceeding of Engineering*, (hal. 1478).
- Syafitri, F. (2017). *Penerapan Teknik Web Scraping dan Klasifikasi Sentimen Menggunakan Metode Naive Bayes Classifier dan Asosiasi (Studi Kasus: Data Ulasan The House Of Raminten Pada Situs TripAdvisor)*. Skripsi Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- Ulwan, M. N. (2016). *Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association (Studi Kasus: Portal Layanan Aspirasi dan Pengaduan Online Rakyat)*. Skripsi Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- Wikipedia. (2017). *Twitter*. Dipetik April 09, 2018, dari <https://id.wikipedia.org/wiki/Twitter>
- Wikiwand. (2017). *Traveloka*. Dipetik April 12, 2018, dari <http://www.wikiwand.com/id/Traveloka>
- Zuhra, W. U. (2017). *Tiket.com Diakuisi Bliblicom, Bersiap Menyalip Traveloka*. Jakarta: <https://tirto.id>



# LAMPIRAN

## Lampiran 1. Output Klasifikasi SVM Pada Program R

### Kombinasi 1

```
> conf.matSVM_1
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      9      2
positif      4     26

      Accuracy : 0.8537
      95% CI : (0.7083, 0.9443)

> conf.matSVM_150_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      9      1
positif      4     27

      Accuracy : 0.878
      95% CI : (0.738, 0.9592)

> conf.matSVM_150_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     10      1
positif      7     23

      Accuracy : 0.8049
      95% CI : (0.6513, 0.9118)

> conf.matSVM_150_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     10      0
positif      8     23

      Accuracy : 0.8049
      95% CI : (0.6513, 0.9118)

> conf.matSVM_150_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     14      3
positif      6     18

      Accuracy : 0.7805
      95% CI : (0.6239, 0.8944)

> conf.matSVM_150_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      8      3
positif      6     24

      Accuracy : 0.7805
      95% CI : (0.6239, 0.8944)

> conf.matSVM_150_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     16      3
positif      1     21

      Accuracy : 0.9024
      95% CI : (0.7687, 0.9728)

> conf.matSVM_150_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     11      6
positif      5     19

      Accuracy : 0.7317
      95% CI : (0.5706, 0.8578)

> conf.matSVM_150_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     14      3
positif      7     17

      Accuracy : 0.7561
      95% CI : (0.597, 0.8764)

> conf.matSVM_150_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     10      4
positif     10     17

      Accuracy : 0.6585
      95% CI : (0.4941, 0.7992)
```

## Kombinasi 2

```
> conf.matSVM_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  32     5
positif  24    68

      Accuracy : 0.7752
      95% CI : (0.6934, 0.844)

> conf.matSVM300_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  36     7
positif  18    68

      Accuracy : 0.8062
      95% CI : (0.7274, 0.8705)

> conf.matSVM300_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  34     7
positif  19    69

      Accuracy : 0.7984
      95% CI : (0.7188, 0.8639)

> conf.matSVM300_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  36     9
positif  13    71

      Accuracy : 0.8295
      95% CI : (0.7533, 0.8899)

> conf.matSVM300_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  41     6
positif  16    66

      Accuracy : 0.8295
      95% CI : (0.7533, 0.8899)

> conf.matSVM300_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  44    14
positif   9    62

      Accuracy : 0.8217
      95% CI : (0.7446, 0.8835)

> conf.matSVM300_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  36    12
positif  17    64

      Accuracy : 0.7752
      95% CI : (0.6934, 0.844)

> conf.matSVM300_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  43    13
positif  22    51

      Accuracy : 0.7287
      95% CI : (0.6434, 0.8032)

> conf.matSVM300_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  32     4
positif  25    68

      Accuracy : 0.7752
      95% CI : (0.6934, 0.844)

> conf.matSVM300_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  40    20
positif  13    56

      Accuracy : 0.7442
      95% CI : (0.6599, 0.8169)
```

## Kombinasi 3

```
> conf.matSVM_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  68    12
positif  26   132

      Accuracy : 0.8403
      95% CI : (0.7875, 0.88)

> conf.matSVM450_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  79    19
positif  26   114

      Accuracy : 0.8109
      95% CI : (0.7553, 0.8586)
```

```

> conf.matSVM450_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  63    18
positif  35   122

      Accuracy : 0.7773
      95% CI : (0.7191, 0.8285)

> conf.matSVM450_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  63    16
positif  30   129

      Accuracy : 0.8067
      95% CI : (0.7507, 0.8549)

> conf.matSVM450_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  56    13
positif  40   129

      Accuracy : 0.7773
      95% CI : (0.7191, 0.8285)

> conf.matSVM450_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  82    25
positif  22   109

      Accuracy : 0.8025
      95% CI : (0.7462, 0.8511)

> conf.matSVM450_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  64    22
positif  33   119

      Accuracy : 0.7689
      95% CI : (0.7101, 0.8209)

> conf.matSVM450_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  72    19
positif  38   109

      Accuracy : 0.7605
      95% CI : (0.7011, 0.8133)

> conf.matSVM450_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  62     9
positif  39   128

      Accuracy : 0.7983
      95% CI : (0.7417, 0.8474)

> conf.matSVM450_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  63    37
positif  37   101

      Accuracy : 0.6891
      95% CI : (0.6261, 0.7473)

```

## Kombinasi 4

```

> conf.matSVM_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  110   24
positif   45  181

      Accuracy : 0.8083
      95% CI : (0.7638, 0.8477)

> conf.matSVM600_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  110   25
positif   45  180

      Accuracy : 0.8056
      95% CI : (0.7608, 0.8452)

> conf.matSVM600_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  103   24
positif   50  183

      Accuracy : 0.7944
      95% CI : (0.749, 0.835)

> conf.matSVM600_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  112   27
positif   42  179

      Accuracy : 0.8083
      95% CI : (0.7638, 0.8477)

```

```
> conf.matSVM600_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	116	30
positif	27	187

Accuracy : 0.8417  
95% CI : (0.7998, 0.8778)

```
> conf.matSVM600_7
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	100	25
positif	47	188

Accuracy : 0.8  
95% CI : (0.7549, 0.8401)

```
> conf.matSVM600_9
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	98	25
positif	55	182

Accuracy : 0.7778  
95% CI : (0.7312, 0.8197)

```
> conf.matSVM600_6
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	120	47
positif	37	156

Accuracy : 0.7667  
95% CI : (0.7195, 0.8094)

```
> conf.matSVM600_8
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	110	31
positif	50	169

Accuracy : 0.775  
95% CI : (0.7283, 0.8171)

```
> conf.matSVM600_10
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	88	48
positif	62	162

Accuracy : 0.6944  
95% CI : (0.644, 0.7416)

## Kombinasi 5

```
> conf.matSVM_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	120	43
positif	84	242

Accuracy : 0.7403  
95% CI : (0.699, 0.7786)

```
> conf.matSVM750_3
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	142	33
positif	60	254

Accuracy : 0.8098  
95% CI : (0.7722, 0.8437)

```
> conf.matSVM750_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	129	36
positif	64	260

Accuracy : 0.7955  
95% CI : (0.757, 0.8304)

```
> conf.matSVM750_2
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	155	40
positif	54	240

Accuracy : 0.8078  
95% CI : (0.77, 0.8418)

```
> conf.matSVM750_4
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	162	50
positif	49	228

Accuracy : 0.7975  
95% CI : (0.7592, 0.8323)

```
> conf.matSVM750_6
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	160	52
positif	57	220

Accuracy : 0.7771  
95% CI : (0.7376, 0.8132)

```

> conf.matSVM750_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  138    27
positif   63   261

      Accuracy : 0.816
      95% CI : (0.7787, 0.8493)

> conf.matSVM750_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  156    45
positif   55   233

      Accuracy : 0.7955
      95% CI : (0.757, 0.8304)

> conf.matSVM750_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  135    28
positif   73   253

      Accuracy : 0.7935
      95% CI : (0.7548, 0.8285)

> conf.matSVM750_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  120    50
positif   80   239

      Accuracy : 0.7342
      95% CI : (0.6926, 0.7728)

```

## Kombinasi 6

```

> conf.matSVM_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  191    46
positif   70   316

      Accuracy : 0.8138
      95% CI : (0.781, 0.8436)

> conf.matSVM900_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  189    46
positif   72   316

      Accuracy : 0.8106
      95% CI : (0.7776, 0.8406)

> conf.matSVM900_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  184    40
positif   81   318

      Accuracy : 0.8058
      95% CI : (0.7725, 0.8361)

> conf.matSVM900_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  202    58
positif   68   295

      Accuracy : 0.7978
      95% CI : (0.764, 0.8286)

> conf.matSVM900_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  167    53
positif   79   324

      Accuracy : 0.7881
      95% CI : (0.7539, 0.8196)

> conf.matSVM900_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  208    74
positif   64   277

      Accuracy : 0.7785
      95% CI : (0.7438, 0.8105)

> conf.matSVM900_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  170    42
positif   81   330

      Accuracy : 0.8026
      95% CI : (0.7691, 0.8331)

> conf.matSVM900_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  192    60
positif   67   304

      Accuracy : 0.7961
      95% CI : (0.7623, 0.8271)

```

```
> conf.matSVM900_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  178   35
positif   86  324

Accuracy : 0.8058
95% CI : (0.7725, 0.8361)
```

```
> conf.matSVM900_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  150   47
positif  106  320

Accuracy : 0.7544
95% CI : (0.7187, 0.7877)
```

## Kombinasi 7

```
> conf.matSVM_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  219   52
positif   92  397

Accuracy : 0.8105
95% CI : (0.7808, 0.8378)
```

```
> conf.matSVM1050_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  221   53
positif   95  391

Accuracy : 0.8053
95% CI : (0.7753, 0.8328)
```

```
> conf.matSVM1050_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  223   52
positif  100  385

Accuracy : 0.8
95% CI : (0.7698, 0.8279)
```

```
> conf.matSVM1050_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  235   62
positif   82  381

Accuracy : 0.8105
95% CI : (0.7808, 0.8378)
```

```
> conf.matSVM1050_5
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif  212   48
positif   99  401

Accuracy : 0.8066
95% CI : (0.7767, 0.8341)
```

```
> conf.matSVM1050_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  237   80
positif   82  361

Accuracy : 0.7868
95% CI : (0.756, 0.8155)
```

```
> conf.matSVM1050_7
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif  216   46
positif  103  395

Accuracy : 0.8039
95% CI : (0.7739, 0.8316)
```

```
> conf.matSVM1050_8
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif  233   75
positif   81  371

Accuracy : 0.7947
95% CI : (0.7642, 0.8229)
```

```
> conf.matSVM1050_9
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif  214   40
positif  114  392

Accuracy : 0.7974
95% CI : (0.767, 0.8254)
```

```
> conf.matSVM1050_10
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif  185   58
positif  130  387

Accuracy : 0.7526
95% CI : (0.7204, 0.7829)
```

## Kombinasi 8

```
> conf.matSVM_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  233    48
positif  148   471

      Accuracy : 0.7822
      95% CI : (0.7538, 0.8088)

> conf.matSVM1200_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  262    58
positif  121   459

      Accuracy : 0.8011
      95% CI : (0.7735, 0.8267)

> conf.matSVM1200_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  260    60
positif  125   455

      Accuracy : 0.7944
      95% CI : (0.7665, 0.8204)

> conf.matSVM1200_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  270    61
positif  114   455

      Accuracy : 0.8056
      95% CI : (0.7782, 0.8309)

> conf.matSVM1200_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  231    86
positif  147   436

      Accuracy : 0.7411
      95% CI : (0.7112, 0.7695)

> conf.matSVM1200_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  268    86
positif  106   440

      Accuracy : 0.7867
      95% CI : (0.7584, 0.813)

> conf.matSVM1200_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  257    54
positif  115   474

      Accuracy : 0.8122
      95% CI : (0.7851, 0.8372)

> conf.matSVM1200_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  274    91
positif   92   443

      Accuracy : 0.7967
      95% CI : (0.7689, 0.8225)

> conf.matSVM1200_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  253    57
positif  125   465

      Accuracy : 0.7978
      95% CI : (0.77, 0.8236)

> conf.matSVM1200_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  233    48
positif  148   471

      Accuracy : 0.7822
      95% CI : (0.7538, 0.8088)
```

## Kombinasi 9

```
> conf.matSVM_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  317    96
positif  112   516

      Accuracy : 0.8002
      95% CI : (0.7746, 0.8241)

> conf.matSVM1350_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  314    69
positif  128   530

      Accuracy : 0.8108
      95% CI : (0.7856, 0.8341)
```

```

> conf.matSVM1350_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  303    70
positif  142   526

      Accuracy : 0.7963
      95% CI : (0.7706, 0.8204)

> conf.matSVM1350_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  324    81
positif  129   507

      Accuracy : 0.7983
      95% CI : (0.7726, 0.8223)

> conf.matSVM1350_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  301    78
positif  141   521

      Accuracy : 0.7896
      95% CI : (0.7636, 0.814)

> conf.matSVM1350_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  321   103
positif  108   509

      Accuracy : 0.7973
      95% CI : (0.7716, 0.8213)

> conf.matSVM1350_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  286    59
positif  158   538

      Accuracy : 0.7915
      95% CI : (0.7656, 0.8159)

> conf.matSVM1350_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  331   107
positif  105   498

      Accuracy : 0.7963
      95% CI : (0.7706, 0.8204)

> conf.matSVM1350_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  304    60
positif  143   534

      Accuracy : 0.805
      95% CI : (0.7796, 0.8287)

> conf.matSVM1350_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  300    75
positif  142   524

      Accuracy : 0.7915
      95% CI : (0.7656, 0.8159)

```

## Kombinasi 10

```

> conf.matSVM1400_1
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  331    72
positif  134   552

      Accuracy : 0.8108
      95% CI : (0.7863, 0.8337)

> conf.matSVM1400_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  317    73
positif  147   552

      Accuracy : 0.798
      95% CI : (0.7729, 0.8215)

> conf.matSVM1400_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  343    82
positif  134   530

      Accuracy : 0.8017
      95% CI : (0.7767, 0.825)

> conf.matSVM1400_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  315    82
positif  148   544

      Accuracy : 0.7888
      95% CI : (0.7633, 0.8127)

```



```
> conf.matSVM1400_5
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif    324     93
positif    139    533
```

```
Accuracy : 0.787
95% CI : (0.7614, 0.8109)
```

```
> conf.matSVM1400_7
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif    296     62
positif    167    564
```

```
Accuracy : 0.7897
95% CI : (0.7643, 0.8136)
```

```
> conf.matSVM1400_9
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif    277    105
positif    166    541
```

```
Accuracy : 0.7511
95% CI : (0.7244, 0.7766)
```

```
> conf.matSVM1400_6
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif    336    107
positif    114    532
```

```
Accuracy : 0.7971
95% CI : (0.7719, 0.8206)
```

```
> conf.matSVM1400_8
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif    347     91
positif    112    539
```

```
Accuracy : 0.8136
95% CI : (0.7892, 0.8363)
```

```
> conf.matSVM1400_10
Confusion Matrix and Statistics
```

```
      Reference
Prediction negatif positif
negatif    315     65
positif    140    569
```

```
Accuracy : 0.8118
95% CI : (0.7872, 0.8346)
```

## Lampiran 2. Output Klasifikasi Naive Bayes Pada Program R

### Kombinasi 1

```
> conf.matNB_1
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      5      3
positif      8     25

      Accuracy : 0.7317
      95% CI : (0.5706, 0.8578)

> conf.matNB_150_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      6      3
positif      7     25

      Accuracy : 0.7561
      95% CI : (0.597, 0.8764)

> conf.matNB_150_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      7      1
positif     10     23

      Accuracy : 0.7317
      95% CI : (0.5706, 0.8578)

> conf.matNB_150_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     10      1
positif      8     22

      Accuracy : 0.7805
      95% CI : (0.6239, 0.8944)

> conf.matNB_150_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     12      1
positif      8     20

      Accuracy : 0.7805
      95% CI : (0.6239, 0.8944)

> conf.matNB_150_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      7      4
positif      7     23

      Accuracy : 0.7317
      95% CI : (0.5706, 0.8578)

> conf.matNB_150_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     12      4
positif      5     20

      Accuracy : 0.7805
      95% CI : (0.6239, 0.8944)

> conf.matNB_150_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      9      4
positif      7     21

      Accuracy : 0.7317
      95% CI : (0.5706, 0.8578)

> conf.matNB_150_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     10      4
positif     11     16

      Accuracy : 0.6341
      95% CI : (0.4694, 0.7788)

> conf.matNB_150_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif      6      8
positif     14     13

      Accuracy : 0.4634
      95% CI : (0.3066, 0.6258)
```

### Kombinasi 2

```
> conf.matNB_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     28      8
positif     28     65

      Accuracy : 0.7209
      95% CI : (0.6352, 0.7963)

> conf.matNB300_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif     25      6
positif     29     69

      Accuracy : 0.7287
      95% CI : (0.6434, 0.8032)
```

```

> conf.matNB300_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  25   4
positif  28  72

      Accuracy : 0.7519
      95% CI : (0.6682, 0.8237)
> conf.matNB300_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  36   8
positif  21  64

      Accuracy : 0.7752
      95% CI : (0.6934, 0.844)
> conf.matNB300_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  28  10
positif  25  66

      Accuracy : 0.7287
      95% CI : (0.6434, 0.8032)
> conf.matNB300_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  19   6
positif  38  66

      Accuracy : 0.6589
      95% CI : (0.5703, 0.7401)
> conf.matNB_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  59  15
positif  35 129

      Accuracy : 0.7899
      95% CI : (0.7326, 0.8399)
> conf.matNB300_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  37  18
positif  16  58

      Accuracy : 0.7364
      95% CI : (0.6516, 0.8101)
> conf.matNB300_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  24   5
positif  41  59

      Accuracy : 0.6434
      95% CI : (0.5543, 0.7258)
> conf.matNB300_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  25  14
positif  28  62

      Accuracy : 0.6744
      95% CI : (0.5864, 0.7543)

```

### Kombinasi 3

```

> conf.matNB_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  59  15
positif  35 129

      Accuracy : 0.7899
      95% CI : (0.7326, 0.8399)
> conf.matNB450_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  58  13
positif  47 120

      Accuracy : 0.7479
      95% CI : (0.6877, 0.8018)
> conf.matNB450_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  51  16
positif  47 124

      Accuracy : 0.7353
      95% CI : (0.6744, 0.7902)
> conf.matNB450_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  43   9
positif  50 136

      Accuracy : 0.7521
      95% CI : (0.6922, 0.8056)

```

```
> conf.matNB450_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	51	19
positif	45	123

Accuracy : 0.7311  
95% CI : (0.67, 0.7863)

```
> conf.matNB450_7
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	52	16
positif	45	125

Accuracy : 0.7437  
95% CI : (0.6833, 0.7979)

```
> conf.matNB450_9
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	54	14
positif	47	123

Accuracy : 0.7437  
95% CI : (0.6833, 0.7979)

```
> conf.matNB450_6
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	69	20
positif	35	114

Accuracy : 0.7689  
95% CI : (0.7101, 0.8209)

```
> conf.matNB450_8
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	61	10
positif	49	118

Accuracy : 0.7521  
95% CI : (0.6922, 0.8056)

```
> conf.matNB450_10
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	53	16
positif	47	122

Accuracy : 0.7353  
95% CI : (0.6744, 0.7902)

## Kombinasi 4

```
> conf.matNB_4
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	92	24
positif	63	181

Accuracy : 0.7583  
95% CI : (0.7107, 0.8017)

```
> conf.matNB600_3
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	82	17
positif	71	190

Accuracy : 0.7556  
95% CI : (0.7078, 0.7991)

```
> conf.matNB600_2
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	86	19
positif	69	186

Accuracy : 0.7556  
95% CI : (0.7078, 0.7991)

```
> conf.matNB600_4
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	80	10
positif	74	196

Accuracy : 0.7667  
95% CI : (0.7195, 0.8094)

```
> conf.matNB600_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	73	25
positif	70	192

Accuracy : 0.7361  
95% CI : (0.6874, 0.7809)

```
> conf.matNB600_6
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	106	36
positif	51	167

Accuracy : 0.7583  
95% CI : (0.7107, 0.8017)

```
> conf.matNB600_7
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	83	23
positif	64	190

Accuracy : 0.7583  
95% CI : (0.7107, 0.8017)

```
> conf.matNB600_9
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	81	22
positif	72	185

Accuracy : 0.7389  
95% CI : (0.6903, 0.7835)

```
> conf.matNB600_8
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	96	17
positif	64	183

Accuracy : 0.775  
95% CI : (0.7283, 0.8171)

```
> conf.matNB600_10
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	90	20
positif	60	190

Accuracy : 0.7778  
95% CI : (0.7312, 0.8197)

## Kombinasi 5

```
> conf.matNB_5
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	123	24
positif	81	261

Accuracy : 0.7853  
95% CI : (0.7462, 0.8209)

```
> conf.matNB750_3
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	121	28
positif	81	259

Accuracy : 0.7771  
95% CI : (0.7376, 0.8132)

```
> conf.matNB750_5
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	116	49
positif	77	247

Accuracy : 0.7423  
95% CI : (0.7012, 0.7806)

```
> conf.matNB750_7
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	115	25
positif	86	263

Accuracy : 0.773  
95% CI : (0.7333, 0.8094)

```
> conf.matNB750_2
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	119	30
positif	90	250

Accuracy : 0.7546  
95% CI : (0.714, 0.7921)

```
> conf.matNB750_4
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	118	22
positif	93	256

Accuracy : 0.7648  
95% CI : (0.7247, 0.8017)

```
> conf.matNB750_6
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	149	42
positif	68	230

Accuracy : 0.7751  
95% CI : (0.7354, 0.8113)

```
> conf.matNB750_8
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	142	33
positif	69	245

Accuracy : 0.7914  
95% CI : (0.7527, 0.8266)

```
> conf.matNB750_9
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	109	29
positif	99	252

Accuracy : 0.7382  
95% CI : (0.6969, 0.7767)

```
> conf.matNB750_10
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	128	30
positif	72	259

Accuracy : 0.7914  
95% CI : (0.7527, 0.8266)

## Kombinasi 6

```
> conf.matNB_6
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	150	39
positif	111	323

Accuracy : 0.7592  
95% CI : (0.7237, 0.7923)

```
> conf.matNB900_2
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	146	33
positif	115	329

Accuracy : 0.7624  
95% CI : (0.727, 0.7954)

```
> conf.matNB900_3
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	159	39
positif	106	319

Accuracy : 0.7673  
95% CI : (0.732, 0.7999)

```
> conf.matNB900_4
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	153	29
positif	117	324

Accuracy : 0.7657  
95% CI : (0.7304, 0.7984)

```
> conf.matNB900_5
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	154	63
positif	92	314

Accuracy : 0.7512  
95% CI : (0.7153, 0.7847)

```
> conf.matNB900_6
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	178	53
positif	94	298

Accuracy : 0.764  
95% CI : (0.7287, 0.7969)

```
> conf.matNB900_7
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	152	41
positif	99	331

Accuracy : 0.7753  
95% CI : (0.7404, 0.8075)

```
> conf.matNB900_8
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	183	45
positif	76	319

Accuracy : 0.8058  
95% CI : (0.7725, 0.8361)

```
> conf.matNB900_9
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	143	31
positif	121	328

Accuracy : 0.756  
95% CI : (0.7203, 0.7893)

```
> conf.matNB900_10
Confusion Matrix and Statistics
```

	Reference	
Prediction negatif positif		
negatif	160	35
positif	96	332

Accuracy : 0.7897  
95% CI : (0.7556, 0.8211)

## Kombinasi 7

```
> conf.matNB_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  214    71
positif   97   378

      Accuracy : 0.7789
      95% CI : (0.7477, 0.808)

> conf.matNB1050_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  177    31
positif  139   413

      Accuracy : 0.7763
      95% CI : (0.745, 0.8055)

> conf.matNB1050_3
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  189    46
positif  134   391

      Accuracy : 0.7632
      95% CI : (0.7313, 0.793)

> conf.matNB1050_4
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  169    32
positif  148   411

      Accuracy : 0.7632
      95% CI : (0.7313, 0.793)

> conf.matNB1050_5
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  198    67
positif  113   382

      Accuracy : 0.7632
      95% CI : (0.7313, 0.793)

> conf.matNB1050_6
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  215    58
positif  104   383

      Accuracy : 0.7868
      95% CI : (0.756, 0.8155)

> conf.matNB1050_7
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  195    51
positif  124   390

      Accuracy : 0.7697
      95% CI : (0.7381, 0.7992)

> conf.matNB1050_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  223    67
positif   91   379

      Accuracy : 0.7921
      95% CI : (0.7615, 0.8204)

> conf.matNB1050_9
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  176    40
positif  152   392

      Accuracy : 0.7474
      95% CI : (0.7149, 0.7779)

> conf.matNB1050_10
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  197    35
positif  118   410

      Accuracy : 0.7987
      95% CI : (0.7684, 0.8266)
```

## Kombinasi 8

```
> conf.matNB_8
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  204    46
positif  177   473

      Accuracy : 0.7522
      95% CI : (0.7227, 0.7801)

> conf.matNB1200_2
Confusion Matrix and Statistics

      Reference
Prediction negatif positif
negatif  221    40
positif  162   477

      Accuracy : 0.7756
      95% CI : (0.7469, 0.8024)
```

```
> conf.matNB1200_3
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	225	56
positif	160	459

Accuracy : 0.76  
95% CI : (0.7307, 0.7876)

```
> conf.matNB1200_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	237	59
positif	141	463

Accuracy : 0.7778  
95% CI : (0.7492, 0.8045)

```
> conf.matNB1200_7
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	217	46
positif	155	482

Accuracy : 0.7767  
95% CI : (0.748, 0.8035)

```
> conf.matNB1200_9
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	219	53
positif	159	469

Accuracy : 0.7644  
95% CI : (0.7353, 0.7918)

```
> conf.matNB1200_4
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	207	39
positif	177	477

Accuracy : 0.76  
95% CI : (0.7307, 0.7876)

```
> conf.matNB1200_6
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	250	75
positif	124	451

Accuracy : 0.7789  
95% CI : (0.7503, 0.8056)

```
> conf.matNB1200_8
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	256	82
positif	110	452

Accuracy : 0.7867  
95% CI : (0.7584, 0.813)

```
> conf.matNB1200_10
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	210	45
positif	171	474

Accuracy : 0.76  
95% CI : (0.7307, 0.7876)

## Kombinasi 9

```
> conf.matNB_9
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	303	104
positif	126	508

Accuracy : 0.7791  
95% CI : (0.7526, 0.8039)

```
> conf.matNB1350_3
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	262	66
positif	183	530

Accuracy : 0.7608  
95% CI : (0.7337, 0.7864)

```
> conf.matNB1350_2
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	261	48
positif	181	551

Accuracy : 0.78  
95% CI : (0.7536, 0.8048)

```
> conf.matNB1350_4
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	250	46
positif	203	542

Accuracy : 0.7608  
95% CI : (0.7337, 0.7864)



```
> conf.matNB1350_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	287	81
positif	155	518

Accuracy : 0.7733  
95% CI : (0.7466, 0.7984)

```
> conf.matNB1350_7
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	253	63
positif	191	534

Accuracy : 0.756  
95% CI : (0.7287, 0.7818)

```
> conf.matNB1350_9
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	261	64
positif	186	530

Accuracy : 0.7598  
95% CI : (0.7327, 0.7855)

```
> conf.matNB1350_6
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	290	88
positif	139	524

Accuracy : 0.7819  
95% CI : (0.7556, 0.8067)

```
> conf.matNB1350_8
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	300	85
positif	136	520

Accuracy : 0.7877  
95% CI : (0.7616, 0.8122)

```
> conf.matNB1350_10
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	287	81
positif	155	518

Accuracy : 0.7733  
95% CI : (0.7466, 0.7984)

## Kombinasi 10

```
> conf.matNB_10
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	274	52
positif	191	572

Accuracy : 0.7769  
95% CI : (0.7509, 0.8013)

```
> conf.matNB1400_2
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	275	68
positif	189	557

Accuracy : 0.764  
95% CI : (0.7376, 0.7889)

```
> conf.matNB1400_3
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	263	49
positif	214	563

Accuracy : 0.7585  
95% CI : (0.7319, 0.7836)

```
> conf.matNB1400_4
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	302	88
positif	161	538

Accuracy : 0.7713  
95% CI : (0.7452, 0.796)

```
> conf.matNB1400_5
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	302	88
positif	161	538

Accuracy : 0.7713  
95% CI : (0.7452, 0.796)

```
> conf.matNB1400_6
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	314	107
positif	136	532

Accuracy : 0.7769  
95% CI : (0.7509, 0.8013)

```
> conf.matNB1400_7
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	259	56
positif	204	570

Accuracy : 0.7612  
95% CI : (0.7348, 0.7863)

```
> conf.matNB1400_9
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	310	126
positif	133	520

Accuracy : 0.7622  
95% CI : (0.7357, 0.7872)

```
> conf.matNB1400_8
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	313	102
positif	146	528

Accuracy : 0.7723  
95% CI : (0.7462, 0.7969)

```
> conf.matNB1400_10
Confusion Matrix and Statistics
```

	Reference	
Prediction	negatif	positif
negatif	269	61
positif	186	573

Accuracy : 0.7732  
95% CI : (0.7471, 0.7978)

### **Lampiran 3. Packages R yang Digunakan**

```
library(devtools)
library(twitterR)
library(plyr)
library(tm)
library(wordcloud)
library(RColorBrewer)
library(stringr)
library(ggplot2)
library(katadasaR)
require(twitterR)
library(e1071)
library(RTextTools)
library(maxent)
require(parallel) # parallel processing
require(tau)      # tokenise function
require(katadasaR)
library(caret)
library(dplyr)
```

#### **Lampiran 4. Script R Web Scraping**

```
api_key="NSr*****"  
api_secret="UBqT*****"  
access_token="94962*****"  
access_token_secret="l4ikIN*****"  
  
#LOGIN TWITTER#  
setup_twitter_oauth(api_key, api_secret, access_token,  
access_token_secret)  
tweets <- searchTwitter("traveloka", lang="id", n = 100000, since  
= "2018-01-01", resultType="recent")  
write.csv(twListToDF(tweets), "d:/travelokamaret.csv")
```

## Lampiran 5. Script R Preprocessing Data dengan Text Mining

```
#Input Data
cf<-read.csv("D://REKAP JADI/Lengkap/Traveloka_Maret_April.csv" ,
stringsAsFactors = TRUE)
glimpse(cf)

#Randomize Dataset
set.seed(1)
cf <- cf[sample(nrow(cf)),]
glimpse(cf)

#Preparing corpus
corpus <- Corpus(VectorSource(cf$text))
corpus
inspect(corpus[1:5])

#PREPROCESSING

#CASE FOLDING
#Remove punctuation, to lower
corpus.clean <- tm_map(corpus, removePunctuation)
corpus.clean <- tm_map(corpus.clean, tolower)
corpus.clean
inspect(corpus.clean[1:5])

#Menghapus URL, rt, username
removeURL <- function(x) gsub("http[^[:space:]]*", "", x)
corpus.clean <- tm_map(corpus.clean, removeURL)
removeRT <- function(y) gsub("RT ", "", y)
corpus.clean <- tm_map(corpus.clean, removeRT)
removeUN <- function(z) gsub("@\\w+", "", z)
corpus.clean<- tm_map(corpus.clean, removeUN)
inspect(corpus.clean[1:5])

#Menghapus nomor, strip white space
corpus.clean<- tm_map(corpus.clean, removeNumbers)
corpus.clean<- tm_map(corpus.clean, stripWhitespace)
```

```

corpus.clean<- tm_map(corpus.clean, PlainTextDocument)
inspect(corpus.clean[1:5])

#TOKENIZING#
tokenizing <- function(x) strsplit(as.character(x), ";")
corpus.clean<- tm_map(corpus.clean, tokenizing)
corpus.clean[[1]]
inspect(corpus.clean[1:5])

#FILTERING
#Replace words
corpus.clean<-      tm_map(corpus.clean,      gsub,      pattern="ctk",
replacement="cetak")
corpus.clean<-      tm_map(corpus.clean,      gsub,      pattern="dlm",
replacement="dalam")
corpus.clean<-      tm_map(corpus.clean,      gsub,      pattern="lbh",
replacement="lebih")
corpus.clean<-      tm_map(corpus.clean,      gsub,      pattern="klo",
replacement="kalau")
corpus.clean<-      tm_map(corpus.clean,      gsub,      pattern="ksh",
replacement="kasih")

#Menghapus stopwords
myStopwords = readLines("stopwordID.csv")
corpus.clean <- tm_map(corpus.clean, removeWords, myStopwords)

#Remove words
corpus.clean<-      tm_map(corpus.clean,      removeWords,
c("traveloka","halo","lagilagi","fthmzhrxx","yugyeom","eduaubdedub
u","kog","kali","eduaubdedubub","dgn","lnnuxhc"))
inspect(corpus.clean[1:5])

#STEMMING
stem_text <- function(text,mc.cores=1) {
  #stem each word in a block of text
  stem_string <- function(str)
  {
    str <- tokenize(x=str)

```

```

    str <- sapply(str, katadasaR)
    str <- paste(str, collapse="")
    return(str)
  }
  #stem each text block in turn
  x <- mclapply(X=text, FUN=stem_string, mc.cores=mc.cores)
  #return stemmed text blocks
  return(unlist(x))
}
corpus.clean<- tm_map(corpus.clean, stem_text)
inspect(corpus.clean[1:10])

#Matrix representation
dtm <- DocumentTermMatrix(corpus.clean,
control=list(weighing=weightTfIdf))
inspect(dtm)

df<-data.frame(cf$label,text=unlist(sapply(corpus.clean, `[\`]),
stringsAsFactors=F)
write.csv(df,"D://REKAP JADI/SCORING/Traveloka_Maret_April.csv")

```

## Lampiran 6. Script R Klasifikasi dengan SVM

```
setwd("D://REKAP JADI/NEW ACAK")
positif = readLines("Ptrain150_10.csv")
negatif = readLines("Ntrain150_10.csv")
positiftes = readLines("Ptest150_10.csv")
negatiftes = readLines("Ntest150_10.csv")

tweet = c(positif, negatif)
tweet_test= c(positiftes, negatiftes)
tweet_all = c(tweet, tweet_test)
sentiment = c(rep("positif", length(positif) ),
              rep("negatif", length(negatif)))
sentiment_test = c(rep("positif", length(positiftes) ),
                  rep("negatif", length(negatiftes)))
sentiment_all = as.factor(c(sentiment, sentiment_test))

mat = create_matrix(tweet_all, language = "id", removeStopwords =
FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf)

mat = as.matrix(mat)

container      <-      create_container(mat,      sentiment_all,
trainSize=1:109,testSize=110:150, virgin=FALSE)
model <- train_model(container, 'SVM',kernel='linear')
results <- classify_model(container, model)

#Tabel Prediksi vs Asli
SVM_table_1=table("Prediction"
=as.character(results[, "SVM_LABEL"]), "Actual"=
as.character(sentiment_all[110:150]))
SVM_table_1

#confusion matriks
conf.matSVM_150_10<-
confusionMatrix(results[, "SVM_LABEL"],sentiment_all[110:150])
conf.matSVM_150_10
```



## Lampiran 7. Script R Klasifikasi dengan *Naive Bayes*

```
#Partitioning
df.train <- df[1:109,]
df.test <- df[110:150,]
write.csv(df.train,"D://REKAP JADI/Output/df_Train150_10.csv")
write.csv(df.test,"D://REKAP JADI/Output/df_Test150_10.csv")

dtm.train <- dtm[1:109,]
dtm.train
dtm.test <- dtm[110:150,]
dtm.test

corpus.clean.train <- corpus.clean[1:109]
corpus.clean.test <- corpus.clean[110:150]

#Featured Selection
fivefreq <- findFreqTerms(dtm.train,1)
length((fivefreq))
dtm.train.nb_1 <- DocumentTermMatrix(corpus.clean.train,
control=list(dictionary = fivefreq))
dim(dtm.train.nb_1)
dtm.test.nb_1<- DocumentTermMatrix(corpus.clean.test,
control=list(dictionary = fivefreq))
dim(dtm.test.nb_1)

#Boolean Naive Bayes
convert_countNB <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}

#Naive Bayes Model
trainNB <- apply(dtm.train.nb_1, 2, convert_countNB)
testNB <- apply(dtm.test.nb_1, 2, convert_countNB)

#Training
classifier <- naiveBayes(trainNB, df.train$cf.label, laplace = 1)
```

```
#Use the NB classifier we built to make predictions on the test
set
pred <- predict(classifier, testNB)
#Create a truth table by tabulating the predicted class labels
with the actual predicted class labels with the actual class
labels
NB_table_150_5=table("Prediction"=      pred,      "Actual"      =
df.test$cf.label)
NB_table_150_5

#confusion matriks
conf.matNB_150_10 <- confusionMatrix(pred, df.test$cf.label)
conf.matNB_150_10
```

## Lampiran 8. *Script R Word Cloud* dan Asosiasi Kata

```
#Build a term-document matrix
dtm <- TermDocumentMatrix(corpus.clean)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 350)

#Generate the Word cloud
set.seed(561)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=50, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

#asosiasi kata
v<-as.list(findAssocs(dtm, terms
=c("twitter","tiket","bayar","aplikasi","harga","booking","cek","h
otel","pesan","beda"),
corlimit = c(0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15)))
v
```

# COMPARISON OF MACHINE LEARNING PERFORMANCE BASED ON SUPPORT VECTOR MACHINE AND NAIVE BAYES ALGORITHM

(Case Study: Response Data About Traveloka Through Social Media Twitter)

Evi Fitria Umi Latifah<sup>1</sup>, Kariyam<sup>2</sup>

<sup>1</sup>Mahasiswa Program Studi Statistika, FMIPA, Universitas Islam Indonesia

<sup>2</sup>Dosen Program Studi Statistika, FMIPA, Universitas Islam Indonesia

E-mail: <sup>1</sup>[14611064@students.uui.ac.id](mailto:14611064@students.uui.ac.id), <sup>2</sup>[kariyam@uui.ac.id](mailto:kariyam@uui.ac.id)

## ABSTRACT

As technology advances, data that can be collected is not just numerical data but also text-shaped data. The analysis of text data used text mining, because text mining is more general subject than required classification. One of the algorithms in machine learning that can be used to perform text classification is Support Vector Machine (SVM) and Naive Bayes. SVM works by finding the best hyperplane that separates two classes, while Naive Bayes works by finding the largest conditional probability value. Based on differences in the workings of the two algorithms, the researchers wanted to compare the performance of the two algorithms in a new case study of responses data on Traveloka obtained via Twitter social media in March and April 2018. Based on the greatest accuracy level using 10 combinations and at each combination was performed 10 times randomization, the conclusion is that the SVM algorithm has a higher accuracy than Naive Bayes algorithm with average accuracy rate of 79.14% in SVM and 75, 61% in Naive Bayes.

**Keywords:** Machine learning; Support Vector Machine; Naive Bayes; Twitter; Traveloka; Sentiment; Accuracy; Association; Word cloud.

## ABSTRAK

Seiring kemajuan teknologi, data yang dapat dikumpulkan bukan hanya data yang berbentuk numerik namun juga data yang berbentuk teks. Dalam melakukan analisis terhadap data teks digunakan text mining, oleh karena text mining lebih bersifat subjek umum maka diperlukan klasifikasi. Salah satu algoritma dalam machine learning yang dapat digunakan untuk melakukan klasifikasi teks adalah Support Vector Machine (SVM) dan Naive Bayes. SVM bekerja dengan cara mencari hyperplane terbaik yang memisahkan dua buah kelas, sedangkan Naive Bayes bekerja dengan cara mencari nilai probabilitas bersyarat terbesar. Berdasarkan perbedaan cara kerja kedua algoritma tersebut, peneliti ingin membandingkan kinerja kedua algoritma tersebut pada studi kasus baru yaitu data tanggapan mengenai Traveloka yang diperoleh melalui media sosial Twitter pada bulan Maret dan bulan April 2018. Berdasarkan tingkat akurasi terbesar dengan menggunakan 10 kombinasi dan pada masing-masing kombinasi dilakukan 10 kali pengacakan, diperoleh kesimpulan bahwa algoritma SVM memiliki tingkat akurasi lebih tinggi dibandingkan algoritma Naive Bayes dengan rata-rata besarnya tingkat akurasi sebesar 79,14% pada SVM dan 75,61% pada Naive Bayes.

**Kata Kunci :** Machine learning, Support Vector Machine, Naive Bayes, Twitter, Traveloka, Sentimen, Akurasi, Asosiasi, Word cloud.

## **Pendahuluan**

Berdasarkan penelitian pada bulan Januari 2018 yang dilakukan *We Are Social*, sebuah perusahaan media asal Inggris yang bekerja sama dengan *Hootsuite* memperoleh hasil bahwa dalam hal lama waktu yang dihabiskan untuk mengakses media sosial, Indonesia menempati peringkat ketiga dengan durasi rata-rata yang dihabiskan 3 jam 23 menit sehari.

Selain itu, seperti yang tercantum dalam kompas.com (Pertiwi, 2018) berdasarkan laporan berjudul "*Essential Insights Into Internet, Social Media, Mobile, and E-Commerce Use Around The World*" yang diterbitkan tanggal 30 Januari 2018, pengguna aktif sosial media di Indonesia mencapai 130 juta dari total populasi Indonesia sebanyak 265,4 juta jiwa atau mencapai 49 persen.

Besarnya persentase pengguna media sosial dan lama waktu yang dihabiskan untuk mengakses sosial media tersebut tentu akan menghasilkan informasi yang tak terhitung jumlahnya yang dituangkan kedalam media sosial. Salah satu

informasi yang dituangkan ke dalam media sosial adalah pendapat atau tanggapan seseorang terhadap suatu hal. Dalam menggali informasi yang terkandung dalam ribuan tanggapan tersebut maka perlu dilakukan sebuah analisis.

Tanggapan yang terdapat dalam sosial media merupakan data yang berbentuk teks, maka digunakan *text mining* dalam melakukan analisis. Dalam *text mining* suatu kelas lebih bersifat area subjek umum, maka diperlukan klasifikasi. Klasifikasi merupakan suatu pekerjaan menilai suatu objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia.

Terdapat beberapa algoritma dalam klasifikasi, antara lain *Support Vector Machine* dan *Naive Bayes*. Penelitian dengan algoritma *Support Vector Machine* pernah dilakukan (Naradhipa & Purwarianti, 2012) dalam klasifikasi pesan berbahasa Indonesia pada sosial media menghasilkan kesimpulan bahwa metode SVM memiliki tingkat akurasi paling tinggi dalam hal klasifikasi teks. Jurnal penelitian

(Santoso, Ratnawati, & Indriati, 2014) dalam pengkategorian buku komputer berbahasa Indonesia berdasarkan judul dan sinopsis, menghasilkan kesimpulan bahwa metode *Naive Bayes* memiliki tingkat akurasi tertinggi dibandingkan metode KNN dan metode gabungan *k-means* dan LVQ dalam mengkategorikan buku komputer berbahasa Indonesia.

Berdasarkan uraian tersebut, maka penulis ingin membandingkan kinerja dari kedua metode tersebut apabila diterapkan dalam suatu studi kasus yaitu tanggapan mengenai Traveloka.

### **Tujuan Penelitian**

Tujuan dari penelitian yang dilakukan adalah sebagai berikut.

1. Mengetahui perbandingan kinerja *machine learning* dengan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi data tanggapan mengenai Traveloka.
2. Mengetahui perbedaan kinerja algoritma *Support Vector Machine* dan *Naive Bayes* dalam melakukan klasifikasi terhadap tanggapan mengenai Traveloka

pada bulan Maret dan bulan April 2018.

3. Mengetahui hasil asosiasi kata berdasarkan hasil tampilan *word cloud* data tanggapan mengenai Traveloka.

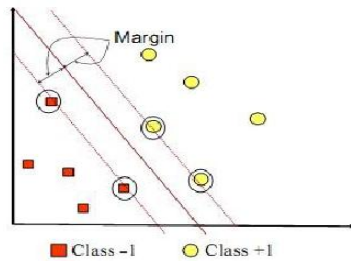
### **Landasan Teori**

#### ***Machine Learning***

*Machine Learning* adalah suatu area dalam kecerdasan buatan (*artificial intelligence*) yang berhubungan dengan teknik-teknik yang dapat dikembangkan melalui pemrograman dan proses pembelajaran menggunakan data masa lalu (Santosa, 2007).

#### ***Support Vector Machine (SVM)***

*Support Vector Machine (SVM)* adalah metode yang dikembangkan oleh Boser, Guyon, Vapnik, dan dipresentasikan pertama kali pada acara *Annual Workshop on Computational Learning Theory* tahun 1992. *Support Vector Machine (SVM)* bekerja menggunakan prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik untuk memisahkan dua kelas data.



(Sumber: Susilowati, Sabariah, & Gozali, 2015)

**Gambar 1.** Ilustrasi SVM Menemukan *Hyperplane* Terbaik

Fungsi pemisah linear didefinisikan sebagai (Santosa, 2007):

$$g(x) := \text{sgn}(f(x))$$

dengan  $f(x) = w^T x + b$ , dimana  $x, w \in \mathbb{R}^n$  dan  $b \in \mathbb{R}$ .

*Hyperplane* pendukung dari kelas +1 (positif):

$$w \cdot x_i + b = +1$$

Sedangkan *hyperplane* pendukung kelas -1 (negatif):

$$w \cdot x_i + b = -1$$

Secara matematika, formulasi problem SVM untuk kasus klasifikasi linear di dalam *primal space* adalah

$$\min_w (w) = \frac{1}{2} \|w\|^2$$

$$y_i (w \cdot x_i + b) \geq 1$$

dimana,  $x_i$  adalah data input;  $y_i$  adalah keluaran dari data  $x_i$ ;  $w$  dan  $b$  adalah parameter yang dicari nilainya.

## Naive Bayes

Secara umum, teorema Bayes ditunjukkan oleh Persamaan berikut (Ariadi & Fithriasari, 2015).

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

dimana,

$P(A|B)$  : Peluang suatu kejadian A apabila B terjadi

$P(B|A)$  : Peluang suatu kejadian B apabila A terjadi (5)

$P(A)$  : Peluang kejadian A

$P(B)$  : Peluang kejadian B

Menurut Manning, dkk (Syafitri, 2017), “setiap ulasan pada klasifikasi *Naive Bayes* dipresentasikan kedalam pasangan atribut dimana  $a_1$  merupakan kata pertama,  $a_2$  merupakan kata kedua, dan seterusnya, sedangkan himpunan kelas disimbolkan dengan  $V$ ”. Metode *Naive Bayes* akan menghasilkan kelas atau kategori tertinggi probabilitasnya ( $V_{MAP}$ ) dengan memasukkan atribut ( $a_1, a_2, a_3, \dots, a_n$ ), rumus  $V_{MAP}$  ditunjukkan oleh Persamaan berikut.

$$V_{MAP} = \underset{v_j \in V}{\text{argmax}} P(v_j | a_1, a_2, a_3, \dots, a_n)$$

dimana,

$V_{MAP}$  : Output nilai hasil klasifikasi Naive Bayes

$argmax$ : titik maksimum probabilitas kelas

$P(v_j | a_1, a_2, a_3, \dots, a_n)$  : probabilitas bukti  $v_j$  terjadi apabila atribut  $a_1, a_2, a_3, \dots, a_n$ .

Berdasarkan teorema Bayes, maka Persamaan diatas dapat ditulis sebagai berikut.

$$V_{MAP} = argmax_{v_j \in V} \frac{P(a_1, a_2, a_3, \dots, a_n | v_j) \times P(v_j)}{P(a_1, a_2, a_3, \dots, a_n)}$$

$P(a_1, a_2, a_3, \dots, a_n)$  bernilai konstan untuk semua  $v_i$  sehingga Persamaan sebelumnya menjadi Persamaan berikut.

$$V_{MAP} = argmax_{v_j \in V} P(a_1, a_2, a_3, \dots, a_n | v_j) \times P(v_j)$$

#### Menghitung

$P(a_1, a_2, a_3, \dots, a_n | v_j)$  akan memiliki tingkat kesulitan yang tinggi karena jumlah term  $P(a_1, a_2, a_3, \dots, a_n | v_j)$  tergantung jumlah kombinasi posisi kata dikali jumlah kelas. Menurut Tan, dkk (Syafitri, 2017) Klasifikasi dengan metode Naive Bayes menyederhanakan hal tersebut dengan mengasumsikan bahwa dalam setiap kategori, setiap atribut

bebas bersyarat satu sama lain, dengan kata lain,

$$P(a_1, a_2, a_3, \dots, a_n | v_j) = \prod_i P(a_i | v_j)$$

Persamaan tersebut akan disubstitusikan ke dalam Persamaan sebelumnya, sehingga diperoleh Persamaan berikut.

$$V_{MAP} = argmax_{v_j \in V} P(v_j) \times \prod_i P(a_i | v_j)$$

Nilai  $P(v_j)$  dan  $\prod_i P(a_i | v_j)$  dapat dihitung dengan menggunakan dua Persamaan berikut (Ariadi & Fithriasari, 2015).

$$P(v_j) = \frac{doc_j}{training}$$

$$P(a_i | v_j) = \frac{n_i + 1}{n + kosakata}$$

dimana,

$doc_j$  : jumlah dokumen memiliki kategori  $j$  pada fase *training/learning*

$training$  : jumlah dokumen pada fase *training*

$n_i$  : jumlah kemunculan kata  $a_i$  dalam dokumen berkategori  $v_j$

$n$  : jumlah seluruh kata pada dokumen dengan kategori  $v_j$

$kosakata$  : jumlah kata dalam fase *training*



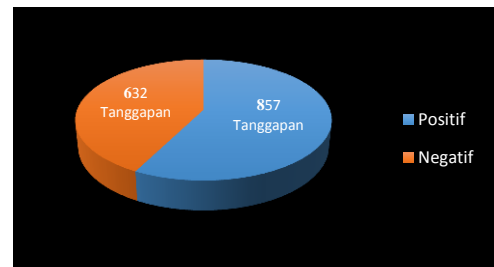
## Metodologi Penelitian

Data dalam penelitian ini merupakan data tanggapan mengenai “Traveloka” pada media sosial *twitter* selama satu minggu pada masing-masing bulan Maret dan April tahun 2018. Data tanggapan yang diperoleh sebanyak 1489 yang terdiri dari 732 tanggapan pada bulan Maret dan 757 tanggapan pada bulan April 2018.

Penelitian ini menggunakan metode analisis Statistika Deskriptif, klasifikasi dengan algoritma *Support Vector Machine* dan *Naive Bayes*, serta *Word Cloud* dan Asosiasi Kata. Alat analisis yang digunakan adalah *software Microsoft PowerPoint 2007*, *Microsoft Excel 2007* dan *R 3.4.1*.

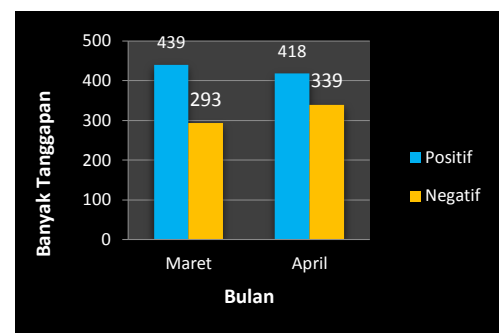
## Pembahasan

Data tanggapan terdiri dari data tanggapan pada bulan Maret dan bulan April yang dikategorikan kedalam dua kelas kategori, yaitu tanggapan positif dan tanggapan negatif. Secara keseluruhan, gambaran mengenai tanggapan yang diperoleh berdasarkan masing-masing kategori ditunjukkan oleh **gambar 2**.



**Gambar 2.** Jumlah Tanggapan tentang Traveloka

Perbandingan tanggapan yang diperoleh pada **gambar 2** merupakan perbandingan dari seluruh data tanggapan yang diperoleh yaitu data tanggapan pada bulan Maret dan bulan April. Oleh karena itu, untuk melihat perbedaan perbandingan tanggapan pada masing-masing bulan, peneliti menampilkan perbandingan pada masing-masing bulan ke dalam **gambar 3**.



**Gambar 3.** Tanggapan tentang Traveloka pada Bulan Maret dan Bulan April

## Perbandingan Kinerja *Support Vector Machine* dan *Naive Bayes*

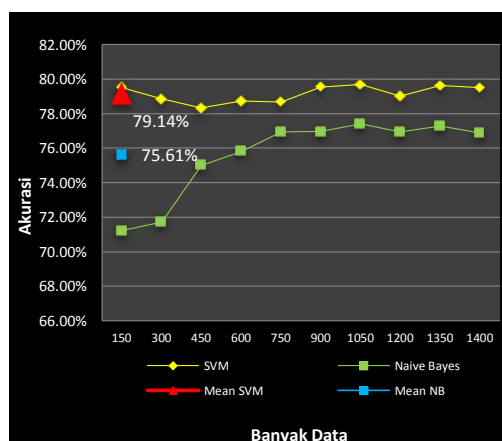
Berdasarkan nilai rata-rata hasil pengacakan pada 10 kombinasi jumlah data *training* dan data *testing*

tersebut akan dilakukan perbandingan tingkat akurasi algoritma SVM dan algoritma *Naive Bayes*. Perbandingan tingkat akurasi tersebut ditunjukkan oleh **tabel 1**.

**Tabel 1.** Perbandingan Rata-Rata Akurasi SVM dan *Naive Bayes*

Banyak Data	SVM	<i>Naive Bayes</i>
150	79.51%	71.22%
300	78.84%	71.70%
450	78.32%	75.00%
600	78.72%	75.81%
750	78.67%	76.93%
900	79.54%	76.97%
1050	79.68%	77.40%
1200	79.00%	76.92%
1350	79.61%	77.29%
1400	79.50%	76.88%
<b>Rata-Rata</b>	<b>79.14%</b>	<b>75.61%</b>

Berdasarkan **tabel 1** tersebut dapat disajikan dalam bentuk grafik **gambar 4**.



**Gambar 4.** Grafik Perbandingan Akurasi SVM dan *Naive Bayes*

Berdasarkan **gambar 4** diketahui bahwa tingkat akurasi algoritma SVM lebih tinggi dibandingkan algoritma *Naive Bayes*. Tingkat akurasi algoritma SVM berkisar antara 78%-80% atau rata-rata tingkat akurasi sebesar 79,14% sedangkan tingkat akurasi algoritma *Naive Bayes* hanya berkisar diantara 70%-78% atau rata-rata sebesar 75,61%. Oleh karena itu, dapat disimpulkan bahwa algoritma *Support Vector Machine* memiliki kinerja lebih baik dalam melakukan klasifikasi data tanggapan mengenai Traveloka melalui media sosial *Twitter* dibandingkan algoritma *Naive Bayes*.

## Perbandingan Kinerja Klasifikasi Tanggapan pada bulan Maret dan bulan April

**Tabel 2.** Perbandingan Tingkat Akurasi Algoritma SVM dan *Naive Bayes* Data Bulan Maret dan April

Algoritma	Accuracy		Error		Correctly Classified		Incorrectly Classified	
	Maret	April	Maret	April	Maret	April	Maret	April
<i>Support Vector Machine</i>	83,51%	79,60%	16,49%	20,40%	395	394	78	101
<i>Naive Bayes</i>	79,28%	76,57%	20,72%	23,43%	375	379	98	116

Berdasarkan tingkat akurasi pada **tabel 2**, secara garis besar dapat dikatakan bahwa baik pada pengklasifikasian data tanggapan mengenai Traveloka pada bulan Maret maupun data tanggapan Traveloka pada bulan April, metode *Support Vector Machine* memiliki tingkat akurasi yang lebih tinggi dibandingkan metode *Naive Bayes* dengan kata lain tidak terdapat perbedaan kinerja *machine learning* dalam melakukan klasifikasi tanggapan pada bulan Maret dan bulan April 2018.

### Word Cloud

*Word cloud* merupakan salah satu bentuk penyajian data yang berupa teks, melalui *word cloud* akan diketahui kata dominan yang sering

diperbincangkan mengenai Traveloka.



(a)



(b)

**Gambar 5.** *Word Cloud* Data Bulan Maret

Pada **gambar 5 (a)** diketahui bahwa perbincangan positif mengenai Traveloka pada bulan



**Tabel 3.** Asosiasi Data Bulan Maret

Tiket (Positif)		Promo (Positif)		Tiket (Negatif)		Beli (Negatif)	
layan	0,28	periode	0,55	cari	0,29	paket	0,43
tambah	0,27	tarik	0,43	pesawat	0,24	internet	0,31
sedia	0,26	kode	0,42	bus	0,19	masuk	0,28
kereta	0,26	maret	0,38	ganggu	0,19	konfirmasi	0,24
cetak	0,26	lebaran	0,38	iklan	0,19	respon	0,22
lebaran	0,24	<b>Diskon (Positif)</b>		jual	0,19	<b>Aplikasi (Negatif)</b>	
antarkota	0,20	istimewa	0,28	kapal	0,19	access	0,41
bis	0,17	untung	0,24	laut	0,19	loading	0,34
mudik	0,17	murah	0,20	<b>Bayar (Negatif)</b>		browser	0,24
<b>Bus (Positif)</b>		hemat	0,19	versi	0,30	sabar	0,24
layan	0,47	<b>Beli (Positif)</b>		hangus	0,27	ribet	0,16
tambah	0,45	paket	0,28	transaksi	0,27	uangkuid	0,16
sedia	0,45	desak	0,25	internet	0,27		
antarkota	0,33	token	0,25	pemberitahuan	0,19		
<b>Hotel (Positif)</b>		listrik	0,20	handphone	0,19		
rekomendasi	0,24	api	0,20				
murah	0,23	kereta	0,19				
untung	0,20						

Berdasarkan **tabel 3** diketahui bahwa tanggapan mengenai Traveloka pada bulan Maret 2018, hal positif banyak memperbincangkan layanan penyediaan tambahan layanan tiket, tiket kereta dan bis, serta tiket untuk mudik lebaran. Promo dan diskon juga sering disebutkan dalam tanggapan pada bulan Maret 2018, diskon pada bulan maret dianggap istimewa, sedangkan tanggapan negatif sering disebutkan dengan kata tiket yang memuat iklan dan keluhan karena tidak terdapat tiket

kapal laut. Tanggapan negatif lainnya mengenai jangka waktu pembayaran yang terlalu pendek sehingga pemesanan hangus. Aplikasi uangkuid yang merupakan aplikasi yang bekerjasama dengan Traveloka untuk melakukan pembayaran juga sering dikeluhkan.

**Tabel 4.** Asosiasi Data Bulan April

Ferry (Positif)		Twitter (Negatif)		Bayar (Negatif)	
tangguh	1	ketuker	0,32	kendala	0,38
unardi	1	suka	0,27	metode	0,34
Tiket (Positif)		pencet	0,25	bpjs	0,24
pesawat	0,36	niat	0,23	kartu	0,24
pesan	0,36	logo	0,19	kredit	0,24
bagus	0,35	mirip	0,19	uangkuid	0,24
turun	0,28	jauh	0,18	sistem	0,12
antarkota	0,25	layar	0,18	durasi	0,11
beli	0,24	posisi	0,18	Aplikasi (Negatif)	
kereta	0,23	nyasar	0,18	ketuker	0,39
bus	0,23	lambang	0,16	twitter	0,13
Harga (Positif)		biru	0,25	koneksi	0,12
turun	0,55	keliru	0,15	sinyal	0,12
banding	0,39	Tiket (Negatif)		install	0,06
website	0,37	beda	0,28	mirip	0,01
rekomendasi	0,20	kereta	0,22		
		pesawat	0,21		
		mahal	0,20		
		prambananjazz	0,18		

Tanggapan positif mengenai Traveloka pada bulan April sering dikaitkan dengan ketangguhan Ferry Unardi yang merupakan pendiri Traveloka, pemesanan tiket, serta rekomendasi perbandingan harga. Tanggapan negatif sering dikaitkan dengan *twitter* yang memiliki kesamaan logo sehingga pelanggan Traveloka sering salah pilih atau salah pencet antara Traveloka dan *twitter*, perbedaan tiket juga muncul sebagai tanggapan negatif, selain itu kendala dalam metode pembayaran

juga muncul sebagai tanggapan negatif.

## Kesimpulan dan Saran

### Kesimpulan

Berdasarkan hasil analisis dan rumusan masalah yang telah disebutkan, diperoleh kesimpulan untuk menjawab rumusan masalah tersebut, yaitu:

1. Berdasarkan tingkat akurasi yang dihasilkan, *machine learning* dengan menggunakan algoritma *Support Vector Machine* memiliki kinerja yang lebih baik dibanding algoritma

*Naive Bayes* dalam melakukan klasifikasi tanggapan mengenai Traveloka yang diperoleh melalui media sosial *Twitter*.

2. Tidak terdapat perbedaan kinerja *machine learning*, baik ketika diaplikasikan pada data keseluruhan (bulan Maret-April), data bulan Maret, maupun data bulan April 2018 kinerja *machine learning* dengan menggunakan algoritma *Support Vector Machine* memiliki tingkat akurasi lebih tinggi dibandingkan algoritma *Naive Bayes*.
- 3.b. Tanggapan positif sering dikaitkan dengan Ferry Unardi, penambahan layanan tiket pesawat, kereta dan bis, layanan penambahan bus antarkota, serta promo dan diskon istimewa.
- b. Tanggapan negatif sering dikaitkan dengan *twitter* yang memiliki kesamaan logo, perbedaan harga antara *website* Traveloka dengan maskapai, iklan yang cukup mengganggu, serta kendala jangka waktu

pembayaran yang terlalu pendek sehingga pemesanan hangus.

### **Saran**

1. Bagi Traveloka, berdasarkan hasil analisis yang dilakukan diharapkan dapat dilakukan perbaikan-perbaikan terutama hal yang paling sering dikeluhkan, seperti logo yang mirip dengan aplikasi *twitter*, perbedaan harga antara *website* Traveloka dengan maskapai, iklan yang cukup mengganggu, serta kendala jangka waktu pembayaran yang terlalu pendek sehingga pemesanan hangus.
2. Bagi peneliti selanjutnya, sebaiknya digunakan algoritma klasifikasi lain agar dapat diketahui perbandingan kinerja algoritma tersebut dibandingkan kedua algoritma yang telah digunakan, jumlah kombinasi dan pengacakan yang lebih banyak, serta dapat memperbanyak periode waktu pengumpulan data agar informasi yang diperoleh lebih mendalam.

## Pustaka

- Ariadi, D., & Fithriasari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *Jurnal Sains dan Seni ITS* , 2337-3520.
- Naradhipa, A. R., & Purwarianti, A. (2012). Sentiment Classification for Indonesian Message in Social Media. *Cloud Computing and Social Networking (ICCCSN), 2012 International Conference* (hal. 1-5). IEEE.
- Pertiwi, W. K. (2018). *Riset Ungkap Pola Pemakaian Medsos Orang Indonesia*. Kompas.com.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Santoso, D., Ratnawati, D. E., & Indriati. (2014). Perbandingan Kinerja Metode Naive Bayes, K-Nearest Neighbor, dan Metode Gabungan K-Means dan LVQ dalam Pengkategorian Buku Komputer Berbahasa Indonesia. *Mahasiswa PTIIK UB* .
- Susilowati, E., Sabariah, M. K., & Gozali, A. A. (2015). Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas pada Twitter. *e-Proceeding of Engineering*, (hal. 1478).
- Syafitri, F. (2017). *Penerapan Teknik Web Scraping dan Klasifikasi Sentimen Menggunakan Metode Naive Bayes Classifier dan Asosiasi (Studi Kasus: Data Ulasan The House Of Raminten Pada Situs TripAdvisor)*. Skripsi Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.