

**SISTEM ANALISIS UMPAN BALIK PENUMPANG  
SINGAPORE AIRLINES PADA LAMAN TRIPADVISOR  
MENGUNAKAN METODE RUMUS BAYES HIBRID DAN  
*SUPPORT VECTOR MACHINE***

**Tesis  
untuk memenuhi sebagian persyaratan  
mencapai derajat Sarjana S-2 Program Studi  
Magister Sistem Informasi**



**Aditya Tegar Satria**

**30000318410027**

**SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO**

**SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO  
SEMARANG**

**2020**

## HALAMAN PERSETUJUAN

Tesis dengan judul :

**SISTEM ANALISIS UMPAN BALIK PENUMPANG SINGAPORE  
AIRLINES PADA LAMAN TRIPADVISOR MENGGUNAKAN METODE  
RUMUS BAYES HIBRID DAN *SUPPORT VECTOR MACHINE***

Oleh:  
**Aditya Tegar Satria**  
30000318410027

Telah dilakukan pembimbingan tesis dan dinyatakan layak untuk mengikuti ujian tesis pada Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 2 Desember 2020  
Menyetujui,

Pembimbing I



Prof. Drs. Mustafid, M.Eng., Ph.D  
NIP. 195505281980031002

Pembimbing II



Dinar Mutiara K. N, S.T., M.Info., Ph.D  
NIP. 197601102009122002

**HALAMAN PENGESAHAN**

**TESIS**

**SISTEM ANALISIS UMPAN BALIK PENUMPANG SINGAPORE AIRLINES PADA  
LAMAM TRIPADVISOR MENGGUNAKAN METODE RUMUS BAYES HIBRID DAN  
SUPPORT VECTOR MACHINE**

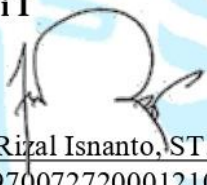
**Oleh:**

**Aditya Tegar Satria  
30000318410027**

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 3 Desember 2020 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 3 Desember 2020  
Mengetahui,

**Penguji I**



Dr. R. Rizal Isnanto, ST., MM., MT, IPM  
NIP. 197007272000121001

**Penguji II**



Dr. Catur Edi Widodo, MT.  
NIP. 196405181992031002

**Pembimbing I**



Prof. Drs. Mustafid, M.Eng., Ph.D  
NIP. 195505281980031002

**Pembimbing II**



Dinar Mutiara K. N., ST., M.InfoTech, Ph.D  
NIP. 197601102009122002

**Mengetahui :**

**Dekan Sekolah Pascasarjana  
Universitas Diponegoro**

Dr. R.B. Sularto, S.H., M.Hum  
NIP. 196701011991031005

**Ketua Program Studi  
Magister Sistem Informasi**



Dr. Budi Warsito, S.Si., M.Si  
NIP. 197508241999031003

## HALAMAN PERNYATAAN

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

Semarang, 2 Desember 2020



Aditya Tegar Satria

UNIVERSITAS DIPONEGORO  
SEMARANG  
SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO

**PERNYATAAN PERSETUJUAN  
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini :

Nama : Aditya Tegar Satria  
NIM : 30000318410027  
Program Studi : Magister Sistem Informasi  
Program : Sekolah Pascasarjana  
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul :

**SISTEM ANALISIS UMPAN BALIK PENUMPANG SINGAPORE  
AIRLINES PADA LAMAN TRIPADVISOR MENGGUNAKAN METODE  
RUMUS BAYES HIBRID DAN *SUPPORT VECTOR MACHINE***

beserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang  
Pada Tanggal : 2 Desember 2020  
Yang menyatakan



Aditya Tegar Satria  
NIM.30000318410027

## KATA PENGANTAR

Puji syukur ke hadirat Allah SWT, atas segala limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian tesis dengan judul “Sistem Analisis Umpan Balik Penumpang Singapore Airlines pada laman TripAdvisor dengan Metode Rumus Bayes hibrid dan *Support Vector Machine*”. Pada proses penyelesaian tesis ini, penulis mendapatkan banyak bantuan, dukungan dan bimbingan selama masa penelitian, pembuatan sistem informasi hingga penyusunan laporan. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Drs. Mustafid, M.Eng., Ph.D., selaku pembimbing pertama yang telah memberikan wawasan, petunjuk dan arahan mendalam dalam penyelesaian tesis ini.
2. Ibu Dinar Mutiara K. N, S.T., M.InfoTech.(Comp), Ph.D., selaku pembimbing kedua yang juga telah memberikan pendampingan, koreksi dan dukungannya dalam penyelesaian tesis ini.
3. Bapak Dr. R. B. Sularto, S.H., M.Hum., selaku Dekan Sekolah Pascasarjana Universitas Diponegoro Semarang.
4. Bapak Dr. Budi Warsito, S.Si., M.Si. selaku Ketua Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Akhir kata, suatu kehormatan dan kebanggaan bagi penulis telah menjadi bagian dalam keluarga besar sivitas akademika Universitas Diponegoro, dan telah berhasil menyelesaikan penelitian tesis ini. Penulis menyadari masih terdapat kekurangan dalam penelitian ini, namun penulis berharap dapat bermanfaat bagi pembaca. Kritik dan saran yang bersifat membangun sangat penulis harapkan bagi kemajuan dan pengembangan ilmu pengetahuan ke depan.

Semarang, Desember 2020  
Penulis

SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO

## DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN.....	iv
PERNYATAAN PERSETUJUAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	x
ABSTRAK.....	xii
ABSTRACT.....	xiii
BAB I.....	<b>Error! Bookmark not defined.</b>
1.1 Latar Belakang.....	<b>Error! Bookmark not defined.</b>
1.2 Tujuan Penelitian.....	<b>Error! Bookmark not defined.</b>
1.3 Manfaat Penelitian.....	<b>Error! Bookmark not defined.</b>
BAB II.....	<b>Error! Bookmark not defined.</b>
2.1 Tinjauan Pustaka.....	<b>Error! Bookmark not defined.</b>
2.2 Dasar Teori.....	<b>Error! Bookmark not defined.</b>
2.2.1 <i>Data Mining</i> dan <i>Text Mining</i> .....	<b>Error! Bookmark not defined.</b>
2.2.2 Prapengolahan Teks dan <i>Bag-of-Words model</i> .....	<b>Error! Bookmark not defined.</b>
2.2.3 Rumus Bayes untuk Klasifikasi Kategori.....	<b>Error! Bookmark not defined.</b>
2.2.4 <i>Support Vector Machine</i> untuk Analisis Sentimen.....	<b>Error! Bookmark not defined.</b>
2.2.5 Pengukuran Kinerja Klasifikasi.....	<b>Error! Bookmark not defined.</b>
BAB III.....	<b>Error! Bookmark not defined.</b>

3.1	Bahan dan Alat Penelitian .....	<b>Error! Bookmark not defined.</b>
3.2	Prosedur Penelitian.....	<b>Error! Bookmark not defined.</b>
3.3	Perancangan Sistem.....	<b>Error! Bookmark not defined.</b>
3.3.1	Kerangka Perancangan Sistem.....	<b>Error! Bookmark not defined.</b>
3.3.2	Diagram Alir Sistem .....	<b>Error! Bookmark not defined.</b>
3.4	Desain Sistem .....	<b>Error! Bookmark not defined.</b>
3.4.1	Perancangan Diagram Aliran Data...	<b>Error! Bookmark not defined.</b>
3.4.2	Pemodelan Sistem .....	<b>Error! Bookmark not defined.</b>
3.4.3	Rancangan Basis Data.....	<b>Error! Bookmark not defined.</b>
3.4.4	Desain Antarmuka Sistem.....	<b>Error! Bookmark not defined.</b>
BAB IV	.....	<b>Error! Bookmark not defined.</b>
4.1	Hasil Penelitian.....	<b>Error! Bookmark not defined.</b>
4.1.1	Pemilihan Data .....	<b>Error! Bookmark not defined.</b>
4.1.2	Prapengolahan Teks .....	<b>Error! Bookmark not defined.</b>
4.1.3	Pelabelan Kelas Kategori dan Sentimen .....	<b>Error! Bookmark not defined.</b>
4.1.4	Implementasi Algoritma Bayes untuk Klasifikasi Kategori ....	<b>Error! Bookmark not defined.</b>
4.1.5	Implementasi Algoritma SVM untuk Analisis Sentimen .....	<b>Error! Bookmark not defined.</b>
4.1.6	Implementasi Komputasi Algoritma.....	<b>Error! Bookmark not defined.</b>
4.1.7	Implementasi Sistem .....	<b>Error! Bookmark not defined.</b>
4.2	Evaluasi Sistem .....	<b>Error! Bookmark not defined.</b>
4.2.1	Pengujian Kotak Hitam .....	<b>Error! Bookmark not defined.</b>
4.2.2	Pengujian Validasi dan Akurasi .....	<b>Error! Bookmark not defined.</b>
4.3	Pembahasan .....	<b>Error! Bookmark not defined.</b>
BAB V	.....	<b>Error! Bookmark not defined.</b>
5.1	Kesimpulan.....	<b>Error! Bookmark not defined.</b>
5.2	Saran.....	<b>Error! Bookmark not defined.</b>
DAFTAR PUSTAKA	.....	<b>Error! Bookmark not defined.</b>



LAMPIRAN.....**Error! Bookmark not defined.**



**SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO**

## DAFTAR GAMBAR

Gambar 1.1 Tangkapan layar ulasan penumpang pada laman TripAdvisor .....	5
Gambar 2.1 Tahapan proses di dalam <i>text mining</i> .....	10
Gambar 2.2 <i>Optimal separating hyperplane</i> .....	19
Gambar 2.3 Persamaan <i>hyperplane</i> pada SVM .....	20
Gambar 2.4 <i>Hyperplane</i> klasifikasi linear vektor $x_1$ dan $x_2$ .....	22
Gambar 3.1 SDLC dengan metode <i>waterfall</i> .....	28
Gambar 3.2 Prosedur penelitian.....	30
Gambar 3.3 Kerangka perancangan sistem.....	32
Gambar 3.4 Diagram alir sistem informasi .....	35
Gambar 3.5 DFD level 0 sistem analisis umpan balik penumpang <i>airlines</i> .....	36
Gambar 3.6 DFD level 1 sistem analisis umpan balik penumpang <i>airlines</i> .....	38
Gambar 3.7 <i>Use case diagram</i> sistem informasi analisis yang dirancang.....	40
Gambar 3.8 Diagram sekuen sistem informasi analisis yang dirancang.....	40
Gambar 3.9 Diagram aktivitas pada sistem informasi analisis yang dirancang...	41
Gambar 3.10 Rancangan desain <i>mockup</i> sistem informasi.....	44
Gambar 3.11 Tampilan halaman <i>dashboard</i> sistem informasi .....	45
Gambar 3.12 Tampilan halaman hasil sistem informasi .....	46
Gambar 4.1 <i>Pie chart</i> hasil perhitungan algoritma Bayes .....	58
Gambar 4.2 Grafik pencapaian nilai <i>cost function</i> terendah iterasi ke-50 .....	62
Gambar 4.3 Tampilan tahap masukan sistem dengan <i>training data</i> .....	67
Gambar 4.4 Tampilan tahap masukan sistem dengan <i>real-time</i> data.....	68
Gambar 4.5 Tampilan <i>preview</i> dan statistik data masukan.....	69
Gambar 4.6 Tampilan tombol proses sistem untuk masukan <i>training data</i> .....	69
Gambar 4.7 Tampilan modul keluaran sistem dengan masukan <i>training data</i> ....	70
Gambar 4.8 Tampilan tabel evaluasi matriks konfusi.....	71
Gambar 4.9 Tampilan modul keluaran sistem dengan masukan <i>real-time</i> data ..	72
Gambar 4.10 Tampilan prediksi kategori dan sentimen .....	73

## DAFTAR TABEL

Tabel 2.1 Penerapan model <i>bag-of-words</i> dalam dokumen $D_1$ dan $D_2$ .....	11
Tabel 2.2 Contoh penerapan tokenisasi .....	12
Tabel 2.3 Contoh penerapan <i>stop words removal</i> .....	13
Tabel 2.4 Contoh proses <i>Porter stemming</i> .....	13
Tabel 2.5 Contoh proses <i>lemmatization</i> .....	14
Tabel 2.6 Penerapan <i>N-grams</i> model .....	14
Tabel 2.7 Daftar kemunculan kata dan peluang untuk perhitungan Bayes.....	17
Tabel 2.8 Contoh perhitungan SVM untuk klasifikasi linear dua vektor .....	21
Tabel 2.9 Perhitungan SVM vektor $x_1$ dan $x_2$ .....	22
Tabel 2.10 Perhitungan <i>hinge loss</i> $D_1$ hingga $D_{10}$ .....	23
Tabel 2.11 Transformasi data baru <i>delta w</i> dan $b$ .....	24
Tabel 2.12 Contoh matriks konfusi 2 kelas.....	25
Tabel 3.1 Tabel data masukan.....	42
Tabel 3.2 Tabel hasil algoritma.....	42
Tabel 4.1 Sampel data ulasan.....	48
Tabel 4.2 Tabel hasil prapengolahan .....	49
Tabel 4.3 Hasil pelabelan manual kelas label kategori dan sentimen.....	51
Tabel 4.4 <i>Vocabulary</i> data latih dokumen $D_1$ dan $D_2$ .....	53
Tabel 4.5 Parameter label kategori ( $C_i$ ).....	54
Tabel 4.6 Hasil perhitungan $Pr(W_j/C_1, C_2 \dots C_5)$ .....	54
Tabel 4.7 Hasil perhitungan peluang $Pr(C_i)$ .....	55
Tabel 4.8 Hasil perhitungan peluang $Pr(W_j)$ .....	56
Tabel 4.9 Tabel Hasil perhitungan peluang Rumus Bayes .....	57
Tabel 4.10 Data masukan untuk pemrosesan algoritma SVM.....	59
Tabel 4.11 Hasil prapengolahan teks dan pembagian tipe data latih atau uji .....	59
Tabel 4.12 Hasil vektorisasi data latih dan data uji pada algoritma SVM.....	60
Tabel 4.13 Nilai $w$ dan $b$ pada data latih.....	61
Tabel 4.14 Nilai $t$ , $y'$ dan <i>average loss</i> pada data latih iterasi pertama.....	61
Tabel 4.15 Nilai <i>delta w</i> pada data latih iterasi pertama .....	61
Tabel 4.16 Nilai <i>weight</i> baru ( $w'$ ) pada iterasi pertama.....	62
Tabel 4.17 Nilai <i>weight</i> baru ( $w'$ ) pada iterasi ke-50.....	62
Tabel 4.18 Nilai $t$ , $y'$ dan <i>average loss</i> pada iterasi ke-50.....	63
Tabel 4.19 Nilai atribut data latih $D_1$ .....	63
Tabel 4.20 Hasil perhitungan $f(D_i)$ pada data latih .....	64
Tabel 4.21 Label hasil perhitungan $f(D_i)$ pada data latih .....	65
Tabel 4.22 Nilai <i>weight</i> baru ( $w'$ ) pada data uji.....	65

Tabel 4.23 Hasil perhitungan $f(D_i)$ pada data uji .....	65
Tabel 4.24 Label hasil perhitungan $f(x_i)$ pada data uji .....	66
Tabel 4.25 Label hasil perhitungan $f(D_i)$ pada data uji .....	66
Tabel 4.26 Hasil pengujian kotak hitam sistem informasi .....	74
Tabel 4.27 Tabel matriks konfusi 2 kelas untuk analisis sentimen.....	76
Tabel 4.28 Kompilasi evaluasi matriks konfusi 2 kelas.....	77
Tabel 4.29 Tabel hasil matriks konfusi 5 kelas untuk klasifikasi kategori .....	78
Tabel 4.30 Kompilasi evaluasi matriks konfusi 5 kelas.....	80



## SEKOLAH PASCASARJANA UNIVERSITAS DIPONEGORO

## ABSTRAK

Algoritma Bayes dan *Support Vector Machine* merupakan metode yang umum digunakan untuk menyelesaikan permasalahan klasifikasi linier termasuk dalam penambangan data teks, namun pada umumnya kedua metode algoritma tersebut digunakan secara terpisah. Pada penelitian ini kedua metode algoritma tersebut digabungkan dalam satu sistem informasi untuk menganalisis kumpulan data teks dari ulasan penumpang pesawat Singapore Airlines pada laman TripAdvisor. Sistem yang dibangun ini bertujuan untuk klasifikasi kategori dan menganalisis aspek sentimen yang terkandung pada data ulasan penumpang dengan mengintegrasikan algoritma Bayes dan SVM pada modul inti pemrosesan dengan modul sistem lainnya seperti modul data masukan dan modul prapengolahan teks, serta mampu melakukan evaluasi sistem dengan data yang diujikan. Dari hasil proses terhadap 1000 dokumen yang diujikan, sistem menunjukkan tingkat akurasi algoritma Bayes mencapai 88,79% dan model algoritma SVM mencapai 84,76%. Penelitian ini telah mengimplementasikan metode baru dimana dua metode algoritma diterapkan secara hibrid yakni bekerja bersamaan pada satu sistem informasi dengan tetap mengoptimalkan kinerja masing-masing sehingga dapat menyajikan hasil yang lebih efektif, efisien, serta menunjukkan tingkat akurasi dan performa yang baik.

Kata kunci: penambangan teks, sistem hibrid, rumus Bayes, *support vector machine*



SEMARANG

SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO

## ABSTRACT

Bayes dan Support Vector Machine algorithms was commonly used to solve some linear classification problems such as text mining, but usually each of this method was particularly used as a single and independent method. On this research, both algorithms will be gathered and combine in a single information system to analyse large text dataset of Singapore Airline passengers review documents on TripAdvisor website. The main purpose of this information system is to perform a classification of document's categories and analysing sentiment aspect that contain on it by integrating Bayes and SVM algorithm method as main core module with other modules such as data input and text pre-processing, and also capable to perform system's evaluation. By processing 1000 amount of testing documents, system was able to show evaluation results of Bayes algorithm as 88,79% and SVM algorithm as 84,76%. This research has implemented a new method where two different algorithm methods has been combined and working simultaneously as a hybrid method on one single information system while optimizing each other's capabilities and performances so that it able to manage results that increase effectiveness, efficiency and show a good level of accuracy and performance.

Keywords: text mining, hybrid system, Bayes formula, support vector machine



SEMARANG

SEKOLAH PASCASARJANA  
UNIVERSITAS DIPONEGORO

