

Klasifikasi Komentar Pengunjung Hotel Menggunakan Algoritma *Support Vector Machine* dan Seleksi Fitur *Information Gain*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Prayogi Notonegoro
NIM: 09021181722001

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN TUGAS AKHIR

KLASIFIKASI KOMENTAR PENGUNJUNG HOTEL
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*
DAN SELEKSI FITUR *INFORMATION GAIN*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika

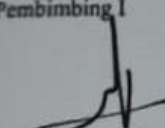
Oleh:

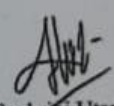
Prayogi Notonegoro (09021181722001)

Indralaya, 21 Juni 2021


Pembimbing I

Pembimbing II


Dr. Abdiansah, S.Kom., M.Cs
NIP. 198410012009121005


Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

Mengetahui,
Ketua Jurusan Teknik Informatika

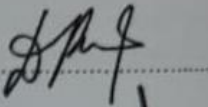
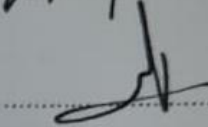
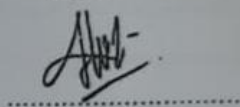
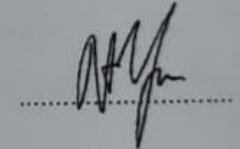
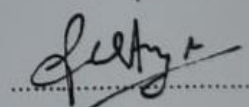

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



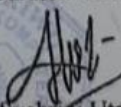
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis tanggal 10 Juni 2021 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Prayogi Notonegoro
NIM : 09021181722001
Judul : Klasifikasi Komentar Pengunjung Hotel menggunakan Algoritma *Support Vector Machine* dan Seleksi Fitur *Information Gain*.

1. Ketua Penguji
Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002 
2. Pembimbing 1
Dr. Abdiansah, S.Kom., M.Cs
NIP. 198410012009121005 
3. Pembimbing 2
Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003 
4. Penguji 1
Novi Yusliani, M.T.
NIP. 198211082012122001 
5. Penguji 2
Desty Rodiah, M.T.
NIP. 198912212020122011 

Mengetahui,
Ketua Jurusan Teknik Informatika


Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Prayogi Notonegoro
NIM : 09021181722001
Program Studi : Teknik Informatika
Judul Skripsi : Klasifikasi Komentar Pengunjung Hotel menggunakan
Algoritma *Support Vector Machine* dan Seleksi Fitur
Information Gain

Hasil Pengecekan Software *ithenticate/Turnitin* : 17 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Indralaya, 21 Juni 2020



Prayogi Notonegoro
NIM: 09021181722001

Motto dan Persembahan

“Tidak ada Kesuksesan yang diraih tanpa perjuangan, kerja keras, doa dan restu orang tua. Yakinkan diri bahwa semua impian yang dapat dicapai akan segera terwujud, *I Can & I Will Do It until I Win*”.

Kupersembahkan karya tulis ini kepada:

- Kedua Orang Tua
- Kakak kandungku
- Eyang dan Eyang Uti
- Keluarga Besar
- Teman-teman Seperjuangan
- Para guru dan dosen-dosen
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

HOTEL VISITOR COMMENT CLASSIFICATION USING SUPPORT VECTOR MACHINE ALGORITHM AND INFORMATION GAIN FEATURE SELECTION

By:

Prayogi Notonegoro (09021181722001)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email:yogiinformatics16@gmail.com

ABSTRACT

Hotel visitor comments is one of the basics of the hotel in understanding the needs of visitors to improve hotel services and facilities. This study discussed how to understand the needs of visitors in order to compete with other competitors. Sentiment analysis was used to solve this problem by classifying comments into positive and negative comments using a classification algorithm. Support Vector Machine is one of the classification algorithms that is commonly used, but it lacks the problem of feature selection. Feature selection is needed to solve this problem. Information Gain is one of the filtering feature selection methods in selecting relevant features. The results of the comments classification research using the Support Vector Machine get the best accuracy value in the linear kernel, parameter C: 1 with accuracy results: 0.98, and computation time 0.16 seconds. The addition of the Information Gain feature selection at the value of K: 0.2 was proven to be able to accelerate the computation time from 0.16 seconds to 0.13 seconds without reducing the accuracy value. The addition was succeed in reducing the number of features which can reduce the error of the algorithm model in classifying features. This was proven by the increased accuracy in the Kernel Rbf C: 0.1 K: 0.2 with an increase of 0.5, from 0.86 to 0.91.

Keyword: Sentiment analysis, Hotel, *Information Gain*, *Support Vector Machine*, Feature selection.

**KLASIFIKASI KOMENTAR PENGUNJUNG HOTEL MENGGUNAKAN
ALGORITMA *SUPPORT VECTOR MACHINE* DAN SELEKSI FITUR
*INFORMATION GAIN***

Oleh:

Prayogi Notonegoro (09021181722001)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email:yogiinformatics16@gmail.com

ABSTRAK

Komentar pengunjung Hotel merupakan salah satu dasar pihak hotel dalam memahami kebutuhan pengunjung untuk melakukan peningkatan pelayanan dan fasilitas hotel. Penelitian ini membahas bagaimana melakukan pemahaman kebutuhan pengunjung guna bersaing dengan kompetitor lainnya. Analisa sentimen digunakan dalam memecahkan masalah tersebut dengan cara mengklasifikasi komentar menjadi komentar positif dan negatif menggunakan algoritma klasifikasi. *Support Vector Machine* merupakan salah satu algoritma klasifikasi yang umum digunakan, Namun memiliki kekurangan terhadap masalah pemilihan fitur. Seleksi fitur dibutuhkan dalam menyelesaikan masalah tersebut, *Information Gain* merupakan salah satu metode seleksi fitur *filtering* dalam memilih fitur yang relevan. Hasil penelitian klasifikasi komentar menggunakan *Support Vector Machine* mendapatkan nilai akurasi terbaik pada kernel linier, parameter C: 1 dengan hasil akurasi: 0.98, presisi: 0.96, recall: 0.98 dan waktu komputasi 0.16 detik. Penambahan Seleksi fitur *Information gain* pada nilai K: 0.2 terbukti mampu mempercepat waktu komputasi yakni awalnya 0.16 detik menjadi 0.13 detik tanpa mengurangi nilai akurasi. Serta dengan penambahan tersebut berhasil melakukan pengurangan jumlah fitur yang mampu mengurangi kesalahan model algoritma dalam mengklasifikasikan fitur. Hal ini dibuktikan dengan hasil akurasi yang meningkat pada Kernel Rbf C: 0.1 K: 0.2 dengan peningkatan sebesar 0.5 yakni dari 0.86 menjadi 0.91.

Kata Kunci: Analisa Sentimen, Hotel, *Information Gain*, *Support Vector Machine*, Seleksi fitur.

KATA PENGANTAR

Selama penelitian dan penyusunan laporan penelitian skripsi ini, penulis tidak luput dari kendala dan hambatan. Namun demikian kendala dan hambatan tersebut dapat penulis atasi berkat bantuan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini, penulis ingin menyampaikan rasa terimakasih sebesar-besarnya kepada:

1. Kedua Orang tua saya terutama Ibu saya Ns. Rosidah dwi putri Rivai, S, Kep dan Kakak saya Melati Ocatvia. S.Ikom serta keluarga besar yang telah memberikan dukungan moril maupun materil serta doa dan restu yang tiada henti-hentinya pada penulis.
2. Bapak Dr. Abdiansah, S. Kom., M.Cs selaku dosen pembimbing 1 dan Ibu Alvi Syahrini Utami, M.Kom selaku dosen pembimbing 2 dan selaku ketua jurusan Teknik Informatika yang telah membimbing, menyediakan fasilitas,memberikam motivasi terutama waktu berharganya kepada penulis dalam proses pengerjaan Tugas Akhir, hingga penulis dapat menyelesaikan Tugas Akhir dengan baik dan waktu yang cepat.
3. Ibu Yunita, S.Si, M.Cs Selaku dosen pembimbing akademik, yang telah selalu menyediakan waktu, memberikan motivasi, bimbingan dan mengarahkan penulis dalam proses pengerjaan Tugas Akhir dan Perkuliahan selama ini.
4. Ibu Dian Palupi Rini, M.Kom., Ph.D. Ibu Novi Yusliani M.T dan Ibu Desty rodiah M.T selaku ketua penguji, penguji 1 dan penguji 2 yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.

5. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Kak Cokro selaku kepala admin lab indralaya yang telah memberikan fasilitas tanpa pamrih kepada penulis dan Kak Ricy selaku admin Indralaya yang selalu direpotkan serta seluruh staf dan pegawai di Fakultas Ilmu Komputer yang telah membantu dalam kelancaran pengerjaan Tugas Akhir penulis dan pada masa perkuliahan.
7. Sahabatku Revan Muhammad dafa yang telah memberikan dukungan, motivasi dan kebersamaan selama proses pengerjaan Tugas Akhir serta teman-teman seperjuangan IF Reguler Indralaya maupun teman-teman kelas yang membantu pengerjaan proses Tugas Akhir ini.
8. Teman-teman SMA maupun SMP yakni Abiyyu, Riswan, Rika, Indri, Maudi, serta teman-teman lainnya yang telah membantu dan memberikan dukungan dalam proses pengerjaan Tugas Akhir ini.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, 21 Juni 2021

Prayogi Notonegoro

DAFTAR ISI

Halaman

LEMBAR PENGESAHAN PROPOSAL SKRIPSI	i
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	ii
HALAMAN PERNYATAAN	iii
Motto dan Persembahan	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-4
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.1.1 Hotel	II-1
2.1.2 Klasifikasi Teks	II-3
2.1.3 Seleksi Fitur	II-6
2.1.4 <i>Support Vector Machine</i>	II-7
2.1.5 <i>Information Gain</i>	II-12
2.1.6 Pengukuran Hasil Klasifikasi	II-14
2.1.7 <i>Rational Unified Process</i>	II-15
2.3 Penelitian Lain yang Relevan	II-18

2.4	Kesimpulan.....	II-21
BAB III METODE PENELITIAN		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis dan Sumber Data	III-1
3.2.2	Metode Pengumpulan Data	III-3
3.3	Tahapan Penelitian	III-6
3.3.1	Kerangka Kerja Penelitian.....	III-8
3.3.2	Kriteria Pengujian.....	III-11
3.3.3	Format Data Pengujian	III-11
3.3.4	Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-12
3.3.5	Pengujian Penelitian	III-13
3.3.6	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-14
3.4	Metode Pengembangan Perangkat Lunak	III-14
3.4.1	Fase Insepsi	III-15
3.4.2	Fase Elaborasi.....	III-15
3.4.3	Fase Konstruksi	III-15
3.4.4	Fase Transisi.....	III-15
3.5	Kesimpulan.....	III-16
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem.....	IV-2
4.2.3	Analisis dan Perancangan.....	IV-4
4.2.3.1	Analisis Kebutuhan Perangkat Lunak	IV-5
4.2.3.2	Analisis Data	IV-6
4.2.3.3	Analisis Pra-Pengolahan	IV-6
4.2.3.4	Analisis Proses Klasifikasi.....	IV-15
4.2.3.5	Analisis Hasil Klasifikasi	IV-39
4.2.4	Implementasi	IV-39
4.3	Fase Elaborasi.....	IV-45
4.3.1	Pemodelan Bisnis	IV-46
4.3.1.1	Perancangan Data	IV-46
4.3.1.2	Perancangan Antarmuka	IV-46
4.3.2	Kebutuhan.....	IV-47

4.3.3	Analisis dan Perancangan.....	IV-47
4.3.3.1	Diagram Aktivitas	IV-47
4.3.3.2	Diagram Alur.....	IV-51
4.4	Fase Konstruksi	IV-56
4.4.1	Kebutuhan.....	IV-56
4.4.2	Implementasi	IV-57
4.4.2.1	Implementasi Kelas	IV-58
4.4.2.2	Implementasi Antarmuka	IV-59
4.5	Fase Transisi.....	IV-60
4.5.1	Pemodelan Bisnis	IV-60
4.5.2	Kebutuhan.....	IV-60
4.5.3	Analisis dan Perancangan.....	IV-60
4.5.3.1	Rencana Pengujian	IV-61
4.5.4	Implementasi	IV-63
4.6	Kesimpulan.....	IV-66
BAB V HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan	V-1
5.2	Hasil Program.....	V-1
5.3	Hasil Penelitian.....	V-4
5.4	Analisis Hasil Penelitian.....	V-47
5.4.1	Analisis Kernel, Nilai C dan Nilai K.....	V-47
5.4.1.1	Kernel Linier.....	V-47
5.4.1.2	Kernel Polinomial	V-49
5.4.1.3	Kernel Rbf.....	V-51
5.4.2	Analisis Jumlah Fitur.....	V-53
5.4.3	Analisis Hasil Prediksi Pengujian Klasifikasi	V-53
5.5	Kesimpulan.....	V-54
BAB VI KESIMPULAN DAN SARAN.....		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA		xviii
LAMPIRAN:		xxii

DAFTAR TABEL

Tabel II-1.	<i>Confusion Matrix</i>	II-14
Tabel III-1.	Contoh komentar yang dikumpulkan.....	III-2
Tabel III-2.	Daftar Hotel Tempat Pengambilan Komentar di Tripadvisor	III-3
Tabel III-3.	Rancangan Tabel <i>Confusion Matrix</i> Hasil Klasifikasi	III-12
Tabel III-4.	Rancangan Tabel Hasil Pengujian	III-12
Tabel III-5.	Tabel Hasil Analisis Klasifikasi	III-14
Tabel IV-1.	Kebutuhan Fungsional Perangkat Lunak.....	IV-4
Tabel IV-2.	Kebutuhan Non Fungsional Perangkat Lunak.....	IV-4
Tabel IV-3.	Data Komentar.....	IV-6
Tabel IV-4.	Hasil Pra-Pengolahan <i>Cleaning</i>	IV-8
Tabel IV-5.	Hasil Pra-Pengolahan <i>Case Folding</i>	IV-9
Tabel IV-6.	Hasil Pra-Pengolahan <i>Tokenizing</i>	IV-10
Tabel IV-7.	Hasil Pra-Pengolahan <i>Formalization</i>	IV-12
Tabel IV-8.	Hasil Pra-Pengolahan <i>Stopword Removal</i>	IV-13
Tabel IV-9.	Hasil Pra-Pengolahan <i>Stemming</i>	IV-14
Tabel IV-10.	Perhitungan pembobotan TF-IDF	IV-16
Tabel IV-11.	Perhitungan Nilai Information Gain dan Pengurutan Fitur	IV-30
Tabel IV-12.	Hasil Seleksi Fitur Kata.....	IV-34
Tabel IV-13.	Pembobotan TF-IDF Hasil Seleksi Fitur	IV-35
Tabel IV-14.	Tabel Definisi Aktor.....	IV-40
Tabel IV-15.	Definisi Use Case	IV-41
Tabel IV-16.	Skenario Use Case Memproses Data.....	IV-42
Tabel IV-17.	Skenario Use Case Melakukan Proses Klasifikasi	IV-43
Tabel IV-18.	Skenario Use Case Melakukan Proses Klasifikasi	IV-44
Tabel IV-19.	Skenario Use Case Melakukan Proses Pengujian Klasifikasi	IV-45
Tabel IV-20.	Keterangan Implementasi Kelas.....	IV-58
Tabel IV-21.	Rencana Use Case Memproses Data	IV-61
Tabel IV-22.	Rencana Use Case Proses Klasifikasi SVM.....	IV-61
Tabel IV-23.	Rencana Use Case Proses Klasifikasi SVM+IG	IV-61
Tabel IV-24.	Rencana Use Case Proses Pengujian Klasifikasi.....	IV-62
Tabel IV-25.	Pengujian Use Case Memproses Data.....	IV-63
Tabel IV-26.	Pengujian Use Case Proses Klasifikasi SVM.....	IV-64
Tabel IV-27.	Pengujian Use Case Proses Klasifikasi SVM+IG	IV-65
Tabel IV-28.	Pengujian Proses Pengujian Klasifikasi	IV-66
Tabel V-1.	Hasil <i>Confusion Matrix</i> Klasifikasi SVM Kernel Linier dan C: 0.1	V-6
Tabel V-2.	Hasil Evaluasi Klasifikasi SVM Kernel Linier dan C: 0.1	V-6
Tabel V-3.	Hasil Confusion Matrix Klasifikasi SVM Kernel Linier dan C: 1	V-7
Tabel V-4.	Hasil Evaluasi Klasifikasi SVM Kernel Linier dan C: 1	V-8
Tabel V-5.	Hasil Confusion Matrix Klasifikasi SVM Kernel Polinomial dan C: 0.1.....	V-9
Tabel V-6.	Hasil Evaluasi Klasifikasi SVM Kernel Polinomial dan C: 0.1	V-9
Tabel V-7.	Hasil Confusion Matrix Klasifikasi SVM Kernel Polinomial dan C:1	V-10

Tabel V-8. Hasil Evaluasi Klasifikasi SVM Kernel Polinomial dan C: 1	V-11
Tabel V-9. Hasil Confusion Matrix Klasifikasi SVM Kernel Rbf dan C: 0.1	V-12
Tabel V-10. Hasil Evaluasi Klasifikasi SVM Kernel Rbf dan C: 0.1	V-12
Tabel V-11. Hasil Confusion Matrix Klasifikasi SVM Kernel Rbf dan C: 1 ...	V-13
Tabel V-12. Hasil Evaluasi Klasifikasi SVM Kernel Rbf dan C: 1	V-14
Tabel V-13. Hasil <i>Confusion Matrix</i> Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.2	V-15
Tabel V-14. Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.2	V-15
Tabel V-15. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.2	V-16
Tabel V-16. Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.2	V-17
Tabel V-17. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.6	V-18
Tabel V-18. Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.6	V-18
Tabel V-19. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.6	V-19
Tabel V- 20. Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.6	V-20
Tabel V-21. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.8	V-21
Tabel V-22. Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.8	V-21
Tabel V-23. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.8	V-23
Tabel V-24. Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.8	V-23
Tabel V-25. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Polinomial C: 0.1 dan K: 0.2	V-24
Tabel V-26. Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 0.1 dan K: 0.2	V-25
Tabel V-27. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Polinomial C: 1 dan K: 0.2	V-26
Tabel V-28. Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 1 dan K: 0.2	V-26
Tabel V-29. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Polinomial C: 0.1 dan K: 0.6	V-27
Tabel V-30. Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 0.1 dan K: 0.6	V-28
Tabel V-31. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel Polinomial C: 1 dan K: 0.6	V-29
Tabel V-32. Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 1 dan K: 0.6	V-29
Tabel V-33. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	

Polinomial C: 0.1 dan K: 0.8.....	V-30
Tabel V-34. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Polinomial C: 0.1 dan K: 0.8.....	V-31
Tabel V-35. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	
Polinomial C: 1 dan K: 0.8.....	V-32
Tabel V-36. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Polinomial C: 1 dan K: 0.8.....	V-32
Tabel V-37. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	
Rbf C: 0.1 dan K: 0.2	V-33
Tabel V-38. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Rbf C: 0.1 dan K: 0.2	V-34
Tabel V-39. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	
Rbf C: 1 dan K: 0.2	V-35
Tabel V-40. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Rbf C: 1 dan K: 0.2	V-35
Tabel V-41. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	
Rbf C: 0.1 dan K: 0.6	V-37
Tabel V-42. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Rbf C: 0.1 dan K: 0.6	V-37
Tabel V-43. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	
Rbf C: 1 dan K: 0.6	V-38
Tabel V-44. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Rbf C: 1 dan K: 0.6	V-39
Tabel V-45. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	
Rbf C: 0.1 dan K: 0.8	V-40
Tabel V-46. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Rbf C: 0.1 dan K: 0.8	V-40
Tabel V-47. Hasil Confusion Matrix Klasifikasi SVM+IG Kernel	
Rbf C: 1 dan K: 0.8	V-41
Tabel V-48. Hasil Evaluasi Klasifikasi SVM+IG Kernel	
Rbf C: 1 dan K: 0.8	V-42
Tabel V-49. Perbandingan Hasil Evaluasi Klasifikasi SVM Linier	V-43
Tabel V-50. Perbandingan Hasil Evaluasi Klasifikasi SVM Kenel	
Polinomial	V-44
Tabel V-51. Perbandingan Hasil Evaluasi Klasifikasi SVM Kenel Rbf.....	V-44
Tabel V-52. Perbandingan Pengujian Prediksi Klasifikasi Linier	V-45
Tabel V-53. Perbandingan Pengujian Prediksi Klasifikasi Polinomial	V-46
Tabel V-54. Perbandingan Pengujian Prediksi Klasifikasi Rbf.....	V-46

DAFTAR GAMBAR

Gambar II-1. Gambar pola dan garis <i>hyperplane</i>	II-8
Gambar II-2. Arsitektur <i>Rational Unified Process</i>	II-17
Gambar III-1. Proses mengunjungi akun hotel	III-4
Gambar III-2. Komentar pada kolom Review.....	III-4
Gambar III-3. Contoh gambar komentar terlabelisasi.....	III-5
Gambar III-4. Diagram Tahapan Penelitian	III-6
Gambar III-5. Diagram Alur Proses Umum Perangkat Lunak	III-8
Gambar IV-1. Use Case Perangkat Lunak	IV-40
Gambar IV-2. Rancangan Antarmuka Perangkat Lunak	IV-46
Gambar IV-3. Diagram Aktivitas Memproses Data	IV-48
Gambar IV-4. Diagram Aktivitas Melakukan Proses Klasifikasi SVM	IV-49
Gambar IV-5. Diagram Aktivitas Melakukan Proses Klasifikasi SVM+IG....	IV-50
Gambar IV-6. Diagram Aktivitas Melakukan Proses Pengujian Klasifikasi...	IV-51
Gambar IV-7. Diagram Alur Memproses data.....	IV-52
Gambar IV-8. Diagram Alur Melakukan Proses Klasifikasi SVM.....	IV-53
Gambar IV-9. Diagram Alur Melakukan Proses Klasifikasi SVM+IG.....	IV-54
Gambar IV-10. Diagram Alur Melakukan Proses Pengujian Klasifikasi	IV-55
Gambar IV-11. Diagram Kelas Perangkat Lunak	IV-57
Gambar IV-12. Implementasi Antarmuka Perangkat Lunak	IV-59
Gambar V-1. Hasil Tampilan Awal Perangkat Lunak.....	V-2
Gambar V-2. Tampilan Fitur Proses Data Perangkat Lunak	V-3
Gambar V-3. Tampilan Fitur Klasifikasi Perangkat Lunak	V-3
Gambar V-4. Tampilan Fitur Prediksi klasifikasi Pengujian Komentar Pengguna.	V-4
Gambar V-5. Grafik Hasil Evaluasi Klasifikasi SVM Kernel Linier dan C: 0.1	V-7
Gambar V-6. Grafik Hasil Evaluasi Klasifikasi SVM Kernel Linier dan C: 1	V-8
Gambar V-7. Grafik Hasil Evaluasi Klasifikasi SVM Kernel Polinomial dan C: 0.1	V-10
Gambar V-8. Grafik Hasil Evaluasi Klasifikasi SVM Kernel Polinomial dan C: 1.....	V-11
Gambar V-9. Grafik Hasil Evaluasi Klasifikasi SVM Kernel Rbf dan C: 0.1	V-13
Gambar V-10. Grafik Hasil Evaluasi Klasifikasi SVM Kernel Rbf dan C: 1	V-14
Gambar V-11. Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.2.....	V-16
Gambar V-12. Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.2.....	V-17
Gambar V-13. Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 0.1 dan K: 0.6.....	V-19
Gambar V-14. Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.6.....	V-20
Gambar V-15. Grafik Hasil Evaluasi Klasifikasi SVM+IG	

	Kernel Linier C: 0.1 dan K: 0.6.....	V-22
Gambar V-16.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Linier C: 1 dan K: 0.8.....	V-24
Gambar V-17.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 0.1 dan K: 0.2.....	V-25
Gambar V-18.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 1 dan K: 0.2.....	V-27
Gambar V-19.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 0.1 dan K: 0.6.....	V-28
Gambar V-20.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 1 dan K: 0.6.....	V-30
Gambar V-21.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 0.1 dan K: 0.8.....	V-31
Gambar V-22.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Polinomial C: 1 dan K: 0.8.....	V-33
Gambar V-23.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Rbf C: 0.1 dan K: 0.2	V-34
Gambar V-24.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Rbf C: 1 dan K: 0.2	V-36
Gambar V-25.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Rbf C: 0.1 dan K: 0.6	V-38
Gambar V-26.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Rbf C: 1 dan K: 0.6	V-39
Gambar V-27.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Rbf C: 0.1 dan K: 0.8	V-41
Gambar V-28.	Grafik Hasil Evaluasi Klasifikasi SVM+IG Kernel Rbf C: 1 dan K: 0.8	V-42
Gambar V-29.	Grafik Perbandingan Evaluasi hasil data klasifikasi Kernel Linier dan C: 0.1	V-47
Gambar V-30.	Grafik Perbandingan Evaluasi hasil data klasifikasi Kernel Linier dan C: 0.1	V-48
Gambar V-31.	Grafik Perbandingan Evaluasi hasil data klasifikasi Kernel Polinomial dan C: 0.1	V-49
Gambar V-32.	Grafik Perbandingan Evaluasi hasil data klasifikasi Kernel Polinomial dan C: 1	V-49
Gambar V-33.	Grafik Perbandingan Evaluasi hasil data klasifikasi Kernel Rbf dan C: 0.1	V-51
Gambar V-34.	Grafik Perbandingan Evaluasi hasil data klasifikasi Kernel Rbf dan C: 1	V-51
Gambar V-35.	Grafik Perbandingan Jumlah Fitur Pada Algoritma Klasifikasi	V-53

Daftar Lampiran

1. Algoritma Pemrograman

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian serta batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian.

1.2 Latar Belakang

Industri Hotel merupakan usaha yang menyediakan tempat penginapan dan pelayanan kepada konsumen. (Widyawati, 2018). Seiring meningkatnya pelaku usaha perhotelan, maka persaingan tidak bisa dihindari, setiap hotel berusaha melakukan peningkatan kualitas yang dimiliki dengan tujuan meningkatkan keuntungan (Cahyani dan Indriati, 2019). Pelaku bisnis perhotelan memanfaatkan komentar sebagai pemahaman kebutuhan pengunjung hotel dalam mengetahui peningkatan layanan yang sesuai. (Phillips, et al., 2017). Namun berbagai komentar yang diberikan membuat pihak hotel mengalami kesulitan dalam memahami ulasan dan membutuhkan waktu yang cukup lama (Saputra dan Mubarak, 2017). Pemahaman perlu dilakukan dengan cara mengklasifikasikan jenis ulasan menjadi sebuah ulasan positif atau negatif, proses ini biasa disebut Sentimen Analisis (Cahyani, dan Indriati, 2019).

Beberapa penelitian umumnya menggunakan metode *K-Nearest Neighbor* (K-NN), *Support Vector Machine* (SVM) dan *Naive Bayes* (NB) dalam melakukan klasifikasi sentimen.. Setiowati dan Helen (2018) melakukan penelitian mengenai klasifikasi sentimen komentar pengunjung hotel di Yogyakarta pada website

*traveloka*¹. Dari penelitian tersebut dihasilkan algoritma *Support Vector Machine* mendapatkan nilai *recall*, *precision*, dan *f-measure* yang paling baik diikuti dengan metode *Naive Bayes* dan *K-Nearest Neighbor*.

Menurut Moraes et al (2013) banyak penelitian yang membahas bahwa penggunaan metode *Support Vector Machines* merupakan metode yang paling akurat pada klasifikasi teks. Namun, *Support Vector Machine* memiliki kekurangan terhadap masalah pemilihan fitur yang sesuai (Basari et al., 2013). Metode seleksi fitur bertujuan dalam mengurangi fitur dan berdampak pada peningkatan proses serta waktu komputasi. Seleksi fitur merupakan solusi terhadap masalah tersebut dan berfungsi sebagai pengoptimal kinerja dari *classifier* (Chandani, Wahono, dan Purwanto, 2015). Seleksi fitur dapat digunakan untuk mengeliminasi atribut yang kurang relevan.

Information Gain adalah salah satu metode seleksi fitur yang menghitung seberapa pentingnya fitur tersebut untuk proses klasifikasi. Chandani, Wahono dan Purwanto (2015) melakukan komparasi algoritma klasifikasi *machine learning* dan seleksi fitur pada analisis sentimen review film hasil komparasi algoritma klasifikasi, SVM mendapatkan hasil terbaik dengan akurasi 81.10% dan AUC 0.904 sedangkan hasil dari komparasi seleksi fitur, *information gain* mendapatkan hasil yang paling baik dengan akurasi rata-rata 84.57%. Selain itu Sahuri (2018) meneliti dengan melakukan komparasi antara metode seleksi fitur dengan metode klasifikasi. Hasil menunjukkan bahwa algoritma *Support Vector Machine*

¹ <https://www.traveloka.com/id-id/>

menghasilkan akurasi terbaik dengan penambahan metode seleksi fitur *Information gain*.

Berdasarkan uraian dan referensi penelitian sebelumnya, algoritma *Information Gain* dan *Support Vector Machine* digunakan pada penelitian klasifikasi komentar pengunjung hotel. Hasil penelitian ini diharapkan dapat menjadi rujukan dan memberikan solusi terhadap pihak hotel dalam mengembangkan sistem klasifikasi komentar pengunjung.

1.3 Rumusan Masalah

Berdasarkan permasalahan pada latar belakang yang telah diuraikan maka rumusan masalah dari penelitian ini adalah

1. Bagaimana melakukan klasifikasi komentar pengunjung hotel menggunakan metode *Support Vector Machine*?
2. Bagaimana pengaruh seleksi fitur *Information Gain* terhadap hasil klasifikasi komentar pengunjung hotel?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun perangkat lunak klasifikasi komentar pengunjung hotel menggunakan metode *Support Vector Machine*.
2. Mengetahui pengaruh penggunaan seleksi fitur *Information Gain* pada klasifikasi menggunakan metode *Support Vector Machine*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Mendapatkan informasi akurasi dan hasil klasifikasi komentar pengunjung hotel menggunakan seleksi fitur *Information Gain* dan *Support Vector Machine* sebagai metode klasifikasi.
2. Menjadi rujukan pihak hotel dalam mengembangkan perangkat lunak.
3. Hasil penelitian dapat menjadi rujukan penelitian yang relevan.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan merupakan komentar pengunjung hotel pada situs website *tripadvisor*².
2. Data komentar dikumpulkan berjumlah 1.000 komentar berbahasa Indonesia terbagi menjadi 500 komentar positif dan 500 negatif.
3. Klasifikasi terdiri dari 2 kelas, yaitu positif (1) dan negative (-1).
4. Penggunaan fungsi kernel SVM terdiri dari kernel Linear, Polinomial dan RBF.
5. Penggunaan parameter C pada setiap kernel SVM adalah 0,1 dan 1.
6. Treshold Nilai K adalah 0.2, 0.6 dan 0.8.

² <https://www.tripadvisor.co.id/>

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini akan membahas landasan dari penelitian, seperti latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas literatur pada penelitian, seperti pengertian Hotel, Klasifikasi Teks, Seleksi Fitur, Seleksi fitur *Information gain* dan algoritma *Support Vector Machine*, dan penelitian yang relevan.

BAB III. METODOLOGI PENELITIAN

Pada Bab ini menjelaskan pelaksanaan alur penelitian. yakni pengumpulan data dan perancangan pembangunan perangkat lunak. Serta tahapan dijelaskan secara detail berdasarkan kerangka yang dibuat.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada Bab ini membahas analisa dan rancangan pengembangan perangkat lunak. Dimulai dengan kebutuhan analisis, perancangan dan konstruksi, serta diakhiri dengan melakukan pengujian yang bertujuan untuk memastikan pengembangan sistem sesuai rancangan dan kebutuhan yang telah dibuat.

BAB V. HASIL DAN ANALISA PENELITIAN

Pada bab ini menjelaskan hasil pengujian sesuai perancangan. Tabel hasil evaluasi pengujian dan analisis serta grafik akan menjadi patokan dari kesimpulan yang akan diambil dalam penelitian.

BAB VI. KESIMPULAN DAN SARAN

Pada Bab ini membahas mengenai kesimpulan dari semua uraian pada bab sebelumnya serta saran yang diuraikan dari hasil penelitian.

1.8 Kesimpulan

Pada Bab ini telah menjelaskan dasar dan patokan pada penelitian , seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

DAFTAR PUSTAKA

- Akbari, M. I., Novianty, A., & Setianingsih, C. (2017). Analisis Sentimen Menggunakan Metode Learning Vector Quantization. *eProceedings of Engineering*, 4(2).
- B.Azhagusundari, & Thanamani, A. S. (2013). Feature selection based on information gain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2(2), 18-21.
- Basari, A. S., Hussin, B., Ananta, I. G., & Zeniarja, J. (2013). Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Engineering*, 53(7), 453-462.
- Bramer, M. (2007). *Principles of data mining* (Vol. 180). Springer.
- Cahyani, R., Indriati, & Adikara, P. P. (2019). Analisis Sentimen terhadap Ulasan Hotel menggunakan Boosting Weighted Extreme Learning Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X*.
- Chandani, V., Wahono, R. S., & Purwanto. (2015). Komparasi algoritma klasifikasi Machine Learning dan feature selection pada analisis sentimen review film. *Journal of Intelligent Systems*, 1(1), 56-60.
- Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, 36(3), 5432–5435.
- Darujati, C., & Gumelar, A. B. (2012). Pemanfaatan teknik supervised untuk klasifikasi teks bahasa indonesia. *Jurnal Bandung Text Mining*, 16(1), 5-1.
- Endah, S. N., & KN, D. M. (2012). Klasifikasi Ucapan Kata dengan Support Vector Machine. *Jurnal Masyarakat Informatika*, 3(6), 7-14.

- Gallager, R. G. (2001). Claude E. Shannon: A retrospective on his life, work, and impact. *IEEE Transactions on Information Theory*, 47(7), 2681-2695.
- Goronescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. New. Elsevier.
- Krutchén, P. (2000). *The Rational Unified Process An Introduction Second Edition*. USA: Addison Wesley.
- Lawson, F. (1976). *Hotels, motels and condominiums: design, planning and maintenance*. London: Architectural Press.
- Maulida, I., Suyatno, A., & Hatta, H. R. (2016). Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. *Jurnal SIFO Mikroskil*, 17(2), 249-258.
- Moraes, R., Valiati, J. F., & Neto, W. P. (2013). Document-level sentiment classification: An empirical comparison between. *Expert Systems with Applications*, 40(2), 621-633.
- Mubaroq, I. M., & Setiawan, E. B. (2020, september). The Effect of Information Gain Feature Selection For Hoax Identification in Twitter Using Classification Method Support Vector Machine. *Ind. Journal on Computing*, 5(2), 107-118. doi:10.21108/indojc.2020.5.2.499
- Negara, A. B., Muhandi, H., & Putri, I. M. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 7(3).

- Phillips, P., Barnes, S., Zigan, K., & Schegg, R. (2017). Understanding the impact of online reviews on hotel performance: an empirical analysis. *Journal of Travel Research*, 56(2), 235--249.
- Prasanti, A. A., Fauzi, M. A., & Furqon, M. T. (2018). Klasifikasi teks pengaduan pada sambat online menggunakan metode n-gram dan neighbor weighted k-nearest neighbor (NW-KNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 946X.
- Ridok, A., & Latifah, R. (2015). Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN. *Proceedings Konferensi Nasional Sistem dan Informatika (KNS&I)*.
- Sahuri, G. (2018). Studi Perbandingan Penggabungan Metode Pemilihan Fitur dengan Metode Klasifikasi dalam Klasifikasi Teks. *Information System Application*, 1(2).
- Saputra, A., Adiwijaya, & Mubarak, M. (2017). Klasifikasi Sentimen Pada Level Aspek Terhadap Ulasan Produk Berbahasa Inggris Menggunakan Bayesian Network (case Study: Data Ulasan Produk Amazon). *eProceedings of Engineering*, 4(3), 4882-4891.
- Setiowati, Y., & Helen, A. (2018). Klasifikasi Analisis Sentimen Mengenai Hotel di Yogyakarta. *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, 13(1), 1-10.
- Suprawoto, T. (2016). Klasifikasi data mahasiswa menggunakan metode k-means untuk menunjang pemilihan strategi pemasaran. *JIKO (Jurnal Informatika dan Komputer)*, 1(1).
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data classification: Algorithms and applications*, 37.
- Tarmoezi, T. (2000). *Professional Hotel Front Liner (Hotel Front Office)*. (H. Sudrajat, Ed.) Jakarta: Kesaint Blanc.

Uysal, A. K., & Gunal, S. (2012). A novel probabilistic feature selection method for text classification. *Knowledge-Based Systems*, 36, 226-235.

Widyawati, N. (2018). Pengaruh Kepercayaan dan Komitmen serta bauran Pemasaran Jasa Terhadap Loyalitas Konsumen Di Hotel Zakiah Medan. *EKUITAS (Jurnal Ekonomi dan Keuangan)*, 12(1), 74-96.