



PEMODELAN *LOGISTIC REGRESION* DAN *RANDOM FOREST* DENGAN METODE CRISP-DM PADA ANALISIS SENTIMEN ULASAN HOTEL DAFAM SEMARANG DI SITUS TRAVELOKA

Tugas Akhir

diajukan sebagai salah satu syarat untuk memperoleh gelar
Ahli Madya Program Studi Statistika Terapan dan Komputasi

oleh

Gabriella Novena Kinthananda Samoling

4112320043

**PROGRAM STUDI STATISTIKA TERAPAN DAN KOMPUTASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG
SEMARANG, 2023**



PEMODELAN *LOGISTIC REGRESION* DAN *RANDOM FOREST* DENGAN METODE CRISP-DM PADA ANALISIS SENTIMEN ULASAN HOTEL DAFAM SEMARANG DI SITUS TRAVELOKA

Tugas Akhir

diajukan sebagai salah satu syarat untuk memperoleh gelar
Ahli Madya Program Studi Statistika Terapan dan Komputasi

oleh

Gabriella Novena Kinthananda Samoling

4112320043

**PROGRAM STUDI STATISTIKA TERAPAN DAN KOMPUTASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG
SEMARANG, 2023**

PERSETUJUAN PEMBIMBING

Tugas akhir yang berjudul “Pemodelan *Logistic Regresion* dan *Random Forest* dengan Metode CRISP-DM pada Analisis Sentimen Ulasan Hotel Dafam Semarang di Situs Traveloka” yang disusun oleh

Nama : Gabriella Novena Kinthananda Samoling

NIM : 4112320043

Program Studi : Statistika Terapan dan Komputasi

Telah disetujui untuk diajukan ke sidang ujian tugas akhir.

Semarang, 6 September 2023

Pembimbing



Prof. Dr. Wardono, M.Si.

NIP: 196202071986011001

PENGESAHAN PENGUJI

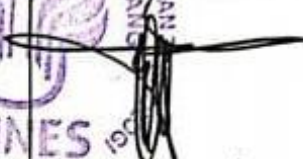



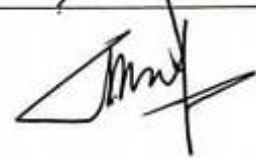
Tugas akhir yang berjudul “Pemodelan *Logistic Regresion* dan *Random Forest* dengan Metode CRISP-DM pada Analisis Sentimen Ulasan Hotel Dafam Semarang di Situs Traveloka” yang disusun oleh

Nama : Gabriella Novena Kinthananda Samoling

NIM : 4112320043

Program Studi : Statistika Terapan dan Komputasi

Telah dipertahankan dalam ujian tugas akhir pada hari Senin, 11 September 2023

Tim Penguji	
Ketua Panitia Prof. Dr. Edy Cahyono, M.Si. NIP 196412051990021001	
Sekretaris Panitia Dr. Putriaji Hendikawati, S.Si., M.Pd., M.Sc. NIP 198208182006042001	
Penguji 1 Drs. Mashuri, M.Si. NIP 196708101992031003	
Penguji 2 Dr. Scolastika Mariani, M.Si. NIP 196502101991022001	
Penguji 3/Pembimbing Prof. Dr. Wardono, M.Si. NIP: 196202071986011001	

PERNYATAAN

Tugas Akhir yang berjudul “Pemodelan *Logistic Regresion* dan *Random Forest* dengan Metode CRISP-DM pada Analisis Sentimen Ulasan Hotel Dafam Semarang di Situs Traveloka” merupakan karya ilmiah asli dan bukan hasil plagiasi dari karya ilmiah orang lain. Pendapat atau temuan orang lain yang dikutip di dalam Tugas Akhir ini telah ditulis berdasarkan kode etik ilmiah.

Semarang, 6 September 2023

Yang menyatakan



Gabriella Novena K. S

4112320043

MOTTO DAN PERSEMBAHAN

MOTTO

"I can do all things through Christ who strengthens me"

- Phillipians 4:13

PERSEMBAHAN

Saya sampaikan terima kasih sebesar-besarnya untuk seluruh pihak yang telah membantu saya dari awal hingga akhir penyusunan tugas akhir ini. Tugas akhir ini saya persembahkan untuk Orang tua, kakak, adik, serta teman-teman. Terima kasih atas doa, semangat, motivasi, pengorbanan, nasihat serta kasih sayang yang selalu diberikan dari proses penyusunan tugas akhir ini. Semoga tugas akhir ini dapat bermanfaat bagi semua pihak yang membacanya.

ABSTRAK

Novena, Gabriella. 2023. analisis sentimen Data Ulasan Hotel Dafam Semarang Pada Situs Traveloka Dengan Model CRISP-DM. Tugas Akhir, Statistika Terapan dan Komputasi, Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang.

Kata kunci: Hotel Dafam, Traveloka, Analisis Sentimen, *Logistic Regression*, *Random Forest*

Traveloka merupakan salah satu situs perjalanan terbesar di dunia. Hotel dafam membutuhkan hasil data ulasan pada situs traveloka untuk melakukan analisis sentimen mengenai kepuasan pelanggan. Untuk mengukur kinerja pelayanan hotel, pihak Hotel Dafam meminta ulasan para pengunjung hotel untuk menanggapi pelayanan yang telah diberikan dengan cara pengunjung dapat memberikan ulasan melalui *website* Traveloka. Sering bertambahnya waktu, membaca seluruh ulasan hotel tentu akan memakan waktu yang lama. Diperlukan usaha untuk mengumpulkan ulasan tersebut dan mengolahnya menjadi informasi yang berguna bagi berbagai pihak, yaitu menggunakan analisis sentimen. Penelitian ini bertujuan untuk mengetahui kecenderungan sentimen dari ulasan Hotel Dafam Semarang yang diambil dari situs Traveloka. Data ulasan akan diklasifikasikan ke dalam sentimen positif, netral, dan negatif dengan pelabelan lexicon.

Tahap selanjutnya dilakukan prediksi sentimen menggunakan algoritma Logistic Regression dan Random Forest. Algoritma tersebut kemudian dievaluasi menggunakan confusion matrix, yang dibandingkan untuk menentukan nilai akurasi terbaik. Hasil penelitian ini menunjukkan bahwa algoritma Random Forest menghasilkan nilai akurasi tertinggi, yaitu sebesar 100%. Berdasarkan hasil dari visualisasi dengan *word cloud* dapat disimpulkan bahwa hasil ulasan bersentimen positif, dimana banyak pengguna yang merasa nyaman dan puas dengan pelayanan yang diberikan oleh pihak hotel. Hasil analisis ini bagi pihak hotel kedepannya untuk mempertahankan konsistensi dari Hotel Dafam dan pertimbangan untuk meningkatkan kualitas.

PRAKATA

Puji dan syukur kepada Allah Bapa di surga yang memberikan kemudahan dan kekuatan bagi penulis, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Pemodelan *Logistic Regresion* dan *Random Forest* dengan Metode CRISP-DM pada Analisis Sentimen Ulasan Hotel Dafam Semarang Situs Traveloka”. Adapun maksud dari penyusunan tugas akhir ini adalah untuk gelar Ahli Madya Program Studi Statistika Terapan dan Komputasi Fakultas Matematika dan Ilmu Pengetahuan Alam. Penulis menyadari bahwa dalam penulisan tugas akhir ini penulis mendapat bantuan dari berbagai pihak, maka dalam kesempatan ini penulis ingin menyampaikan ucapan terimakasih atas segala bantuan, bimbingan dan dukungan yang telah diberikan sehingga tugas akhir ini dapat terselesaikan kepada :

1. Prof. Dr. S Martono, M.Si., selaku Rektor Universitas Negeri Semarang;
2. Prof. Dr. Edy Cahyono, M.Si., selaku Dekan FMIPA UNNES;
3. Dr. Putriaji Hendikawati, M.Pd., M.Sc., selaku Ketua Program Studi Statistika Terapan dan Komputasi FMIPA UNNES;
4. Prof. Dr. Wardono, M.Si., selaku dosen pembimbing tugas akhir yang telah meluangkan banyak waktu untuk membimbing penulis dalam pembuatan tugas akhir ini;
5. Bapak dan Ibu Dosen serta Tenaga Pendidik Jurusan Matematika yang telah memberikan bekal ilmu kepada penulis dalam penyusunan tugas akhir ini;
6. Orang tua, kakak, adik serta teman teman yang telah memberikan dukungan, doa dan semangat dari awal proses penyusunan tugas akhir ini;
7. Nanno, anjing saya tercinta yang selalu menemani saya dari awal proses penyusunan tugas akhir ini, dan selalu menjadi pendengar yang baik untuk mendengarkan keluh kesah saya;
8. Gabriella Novena Kinthananda Samoling, *last but no least*, ya! Diri saya sendiri. Apresiasi sebesar-besarnya karena telah bertanggung jawab untuk menyelesaikan apa yang telah dimulai. Terima kasih karena terus berusaha dan tidak menyerah, serta senantiasa menikmati setiap prosesnya yang bisa dibbilang tidak mudah. Terima kasih sudah bertahan.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna, karena dengan segala keterbatasan pengetahuan dan pengalaman yang masih harus penulis tingkatkan lagi agar bisa lebih baik ke depannya. Untuk itu, penulis sangat menerima kritikan dan saran yang membangun dari pihak mana pun. Semoga tugas akhir ini bermanfaat bagi siapapun yang membacanya.

Semarang, 6 September 2023

DAFTAR ISI

HALAMAN JUDUL	i
PERSETUJUAN PEMBIMBING	ii
PENGESAHAN PENGUJI	Error! Bookmark not defined.
PERNYATAAN	Error! Bookmark not defined.
MOTTO DAN PERSEMBAHAN.....	v
ABSTRAK	vi
PRAKATA	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian	4
1.6 Keaslian Penelitian	4
BAB II KAJIAN PUSTAKA	6
2.1 Tinjauan Pustaka	6
2.2 Landasan Teori	7
2.2.1 <i>Machine Learning</i>	7
2.2.2 <i>Data Mining</i>	7
2.2.3 Analisis Sentimen	8
2.2.4 CRISP-DM.....	9
2.2.5 Hotel Dafam Kota Semarang	9
2.2.6 <i>Online Travel Agents</i>	10
2.2.7 Traveloka.....	11
2.2.8 Pengumpulan Data dengan Web Scraping	12
2.2.9 Praproses Data Teks.....	13
2.2.10 <i>Data Labelling</i>	15

2.2.11 Pembobotan TF-IDF	16
2.2.12 Metode Klasifikasi	16
2.2.13 <i>Logistic Regression</i>	17
2.2.14 <i>Random Forest</i>	17
2.2.15 <i>Confusion Matrix</i>	18
2.2.16 <i>Word Cloud</i>	19
2.3 Kerangka Berpikir	19
BAB III METODE PENELITIAN	22
3.1 Metode Pengumpulan Data	22
3.2 Populasi dan Sampel Data.....	22
3.3 Variabel Penelitian.....	23
3.4 Analisis Data	23
3.5 Alur Penelitian.....	23
BAB IV HASIL DAN PEMBAHASAN	25
4.1 <i>Business Understanding</i>	25
4.2 <i>Data Understanding</i>	25
4.3 <i>Data Preparation</i>	26
4.3.1 Praproses data.....	27
4.3.2 <i>Data Labelling</i>	31
4.4 <i>Modelling</i>	33
4.4.1 Pembagian data.....	33
4.4.2 Klasifikasi.....	34
4.5 <i>Evaluation</i>	34
4.5.1 Visualisasi	37
4.6 <i>Deployment</i>	39
BAB V PENUTUP	41
5.1 Kesimpulan	41
5.2 Saran	42
DAFTAR PUSTAKA	43
LAMPIRAN	50