

**ANALISIS SENTIMEN PADA LAYANAN *STREAMING*
FILM DI TWITTER MENGGUNAKAN METODE
LATENT DIRICHLET ALLOCATION DAN SUPPORT
VECTOR MACHINE**

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



Noorhanida Royani

30000420026

**SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG**

2023

HALAMAN PENGESAHAN

TESIS

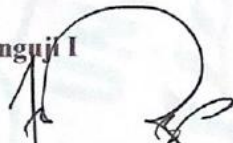
**ANALISIS SENTIMEN PADA LAYANAN *STREAMING* FILM DI TWITTER
MENGUNAKAN METODE LDA DAN SVM**

Oleh:
Noorhanida Royani
30000321420026

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 29 Desember 2023 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

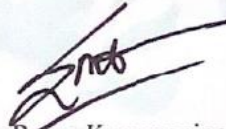
Semarang, 27 Desember 2023
Mengetahui,

Penguji I



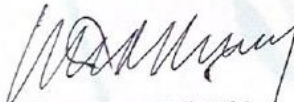
Prof. Dr. Ir. R. Riza Isnanto, S.T.,
M.M., M.T., IPU, ASEAN Eng.
NIP 197007272000121001

Penguji II



Dr. Retno Kusumaningrum, S.Si., M.Kom.
NIP 198104202005012001

Pembimbing I



Dr. Drs. Catur Edi Widodo, M.T.
NIP 196405181992031002

Pembimbing II



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

Mengetahui :

**Dekan Sekolah Pascasarjana
Universitas Diponegoro**



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

**Ketua Program Studi
Magister Sistem Informasi**



Dr. Bayu Surarso, M.Sc., Ph.D.
NIP 196311051988031001

**PERNYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

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

Nama : Noorhanida Royani
NIM : 30000321420026
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 :

**ANALISIS SENTIMEN PADA LAYANAN STREAMING FILM DI
TWITTER MENGGUNAKAN METODE LATENT DIRICHLET
ALLOCATION 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 : 07 Desember 2023
Yang menyatakan



Noorhanida Royani
NIM 30000321420026

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, 07 Desember 2023



Noorhanida Royani

SEKOLAH PASCASARJANA

KATA PENGANTAR

Alhamdulillah *Rabbil' Alamin* saya ucapkan sebagai tanda syukur yang dalam kepada Allah SWT atas segala karunia dan rahmat yang diberikan-Nya, sehingga penulis dapat menyelesaikan penulisan tesis ini dengan judul “Analisis Sentimen Pada Layanan Streaming Film di Twitter Menggunakan Metode Latent Dirichlet Allocation dan Support Vector Machine”.

Tesis ini guna melengkapi salah satu persyaratan untuk memperoleh Magister Komputer (M.Kom) pada Program Studi Magister Sistem Informasi Universitas Diponegoro Semarang. Banyak pihak yang telah membantu penulis dalam menyusun penelitian tesis ini, baik berupa materi maupun moril atau motivasi. Untuk itu pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. R.B. Sularto, S.H., M.Hum selaku Dekan Sekolah Pascasarjana Universitas Diponegoro.
2. Bapak Drs.Bayu Surarso, M.Sc., Ph.D selaku Ketua Program Studi Sekolah Pascasarjana Magister Sistem Informasi Universitas Diponegoro Semarang.
3. Bapak Dr. Catur Edi Widodo, MT sebagai Pembimbing I, terima kasih atas waktu, ilmu, saran dan semangat yang Bapak berikan selama bimbingan.
4. Bapak Dr. Budi Warsito, S.Si., M.Si selaku Pembimbing II yang penuh kesabaran memberikan pengarahan dan motivasi, terima kasih atas waktu, ilmu, saran dan semangat yang selalu Bapak berikan selama bimbingan.
5. Semua pihak yang telah membantu yang tidak dapat disebutkan satu persatu.

Penulis menyadari sepenuhnya bahwa dalam penyusunan tesis ini masih terdapat banyak kekurangan, untuk itu penulis mengharap kritik dan saran dari berbagai pihak untuk sempurnanya sebuah karya tulis.

Semarang, 8 September 2023

SEKOLAH PASCASARJANA

Noorhanida Royani

DAFTAR ISI

	Halaman
HALAMAN PENGESAHAN.....	i
PERNYATAAN PERSETUJUAN	ii
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS	ii
PERNYATAAN.....	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
DAFTAR ARTI LAMBANG DAN SINGKATAN	ix
ABSTRAK.....	x
ABSTRACT.....	xi
BAB I PENDAHULUAN.....	12
1.1 Latar Belakang	12
1.2 Tujuan Penelitian	14
1.3 Manfaat Penelitian	14
BAB II LANDASAN TEORI.....	15
2.1 Tinjauan Pustaka	15
2.2 Layanan <i>Streaming</i>	16
2.3 Analisis Sentimen	17
2.4 Text Preprocessing	17
2.5 <i>Topic Modelling</i>	18
2.6 Coherence Score.....	20
2.7 Latent Dirichlet Allocation	20
2.8 <i>Support Vector Machine</i>	22
2.9 Evaluasi Performa	24
BAB III METODE PENELITIAN.....	26
3.1 Bahan dan Alat Penelitian.....	26
3.2 Prosedur Penelitian.....	26
3.3 Kerangka Sistem Informasi.....	29
3.3.1 Preprocessing Text.....	30
3.3.2 Count Vectorize	34
3.3.3 Pemodelan Topik menggunakan LDA.....	34
3.3.4 Pembagian Data	38
3.3.5 Klasifikasi Support Vector Machine.....	38
3.4 Perancangan Sistem	39

BAB IV PEMBAHASAN.....	40
4.1 Pengembangan Hasil Website.....	40
4.2 Pembahasan.....	44
4.2.1 Pengumpulan Data	44
4.2.2 <i>Preprocessing Text</i>	45
4.2.3 Count Vectorizer	46
4.2.4 Topic Modelling dengan LDA	47
4.2.5 Klasifikasi Support Vector Machine (SVM).....	50
4.2.6 Optimalisasi Model	51
4.2.7 Hasil Penelitian	53
BAB V KESIMPULAN DAN SARAN.....	55
5.1 Kesimpulan	55
5.2 Saran.....	55
DAFTAR PUSTAKA	56



SEKOLAH PASCASARJANA

DAFTAR GAMBAR

Gambar 1. <i>Text Preprocessing</i>	18
Gambar 2. <i>Topic Modelling</i>	19
Gambar 3. Ilustrasi Metode LDA.....	20
Gambar 4. Garis pemisah antar kelas atau hyperplane di metode SVM.....	23
Gambar 5. Diagram alur prosedur penelitian	28
Gambar 6. Kerangka Sistem Informasi	29
Gambar 7. <i>Punctuatuon Removal</i>	30
Gambar 8. <i>Lowercasing</i>	31
Gambar 9. <i>Tokenizing</i>	32
Gambar 10. <i>Stemming</i>	33
Gambar 11. <i>Stopword removal</i>	34
Gambar 12. Langkah-langkah LDA.....	37
Gambar 13. Alur perancangan	39
Gambar 14. Tampilan Halaman Beranda.....	40
Gambar 15. Tampilan Halaman Data.....	41
Gambar 16. Tampilan Hasil dari Preprocessing text	41
Gambar 17. Tampilan Hasil dari <i>Count Vectorize</i>	42
Gambar 18. Tampilan Halaman <i>Coherence</i>	42
Gambar 19. Tampilan Halaman <i>Topic Modelling</i>	43
Gambar 20. Tampilan Halaman Klasifikasi SVM	44
Gambar 21. Grafik Nilai Coherence	47

SEMARANG
SEKOLAH PASCASARJANA

DAFTAR TABEL

Tabel 1. <i>Confusion matrix</i>	24
Tabel 2. Contoh Daftar Kata Positif dan Negatif.....	27
Tabel 3. Contoh Data Sebelum dan Sesudah <i>Punctuation Removal</i>	30
Tabel 4. Contoh Data Sebelum dan Sesudah Lowercasing.....	31
Tabel 5. Contoh Data sebelum dan sesudah tokenizing.....	32
Tabel 6. Contoh Data sebelum dan sesudah stemming.....	33
Tabel 7. Contoh Data sebelum dan sesudah <i>stopword removal</i>	34
Tabel 8. Persiapan Data.....	35
Tabel 9. Inisialisasi topik secara acak.....	35
Tabel 10. Inisialisasi topik baru.....	35
Tabel 11. Topik terhadap dokumen.....	37
Tabel 12. Kata terhadap topik.....	37
Tabel 13. <i>Preprocessing text</i>	45
Tabel 14. Hasil dari <i>Count Vectorize</i>	47
Tabel 15. Identifikasi 2 Topik.....	48
Tabel 16. Hasil Topik Dokumen.....	49
Tabel 17. Hasil Gridsearchcv.....	50
Tabel 18. Hasil Prediksi Klasifikasi.....	51
Tabel 19. Detail Data.....	53
Tabel 20. <i>Confusion matrix</i>	53
Tabel 21. Lampiran inisialisasi topik secara acak.....	59
Tabel 22. Lampiran inisialisasi topik baru.....	67
Tabel 23. Lampiran topik terhadap dokumen.....	71
Tabel 24. Lampiran kata terhadap topik.....	72

SEMARANG
SEKOLAH PASCASARJANA

DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
p	Probabilitas
M	Banyaknya dokumen
n	Banyaknya kata dalam suatu dokumen
α	Nilai distribusi topik pada setiap dokumen
β	Nilai distribusi kata pada setiap topik
θ_d	Probabilitas distribusi topik dalam suatu dokumen
p	Nilai probabilitas
M	Banyaknya dokumen

DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
LDA	<i>Latent Dirichlet Allocation</i>
SVM	<i>Support Vector Machine</i>
NLP	<i>Natural Language Processing</i>
SVM	<i>Support Vectore Machine</i>
CSV	<i>Comma-Separated Values</i>

SEMARANG
SEKOLAH PASCASARJANA

ABSTRAK

Seiring dengan berkembangnya teknologi, memunculkan banyak *platform online* untuk *streaming* film. *Streaming platform* banyak digunakan masyarakat seperti netflix, disney+, hbo go, we tv, vidio. Selama ini permasalahan *streaming* banyak menimbulkan opini di masyarakat. Banyaknya perbandingan antar *streaming platform* menjadi perbincangan di media sosial yaitu twitter. Opini yang disampaikan pengguna tentang *streaming platform* berisi komentar positif dan komentar negatif yang mempengaruhi pengguna lainnya yang ingin menonton film. Oleh karena itu, opini tersebut dapat dijadikan analisis sentimen untuk mengungkapkan dan memberi informasi dari tanggapan masyarakat terhadap *platform streaming*. LDA dapat digunakan sebagai *topic modelling* dan SVM untuk klasifikasi. Pada tahapan pengambilan data dengan menggunakan *tools framework scrapy* dengan *python*, data diambil sebanyak 5.000 dilakukan *preprocessing text* dan *count vectorize*. LDA dapat mempresentasikan topik dan dokumen serta klasifikasi menggunakan SVM mendapatkan hasil bahwa komentar positif lebih banyak dari pada komentar negatif. Hasil evaluasi performa didapatkan nilai akurasi 91%, *precision* 90%, *recall* 91% dan *f1 score* 90%. Hasil penelitian ini mendapatkan bahwa jumlah komentar positif yang lebih dominan dari pada komentar negatif. Oleh karena itu, perlunya data yang seimbang dalam melakukan klasifikasi komentar *streaming platform* agar pada saat klasifikasi dapat memberikan hasil prediksi klasifikasi yang baik.

Kata Kunci: analisis sentimen, *platform streaming*, LDA, SVM

SEMARANG
SEKOLAH PASCASARJANA

ABSTRACT

Along with the advancement of technology, many online platforms for streaming movies have emerged. Streaming platforms are widely used by the public, such as Netflix, Disney+, HBO Go, WeTV, and Vidio. Streaming issues have been a topic of discussion in society. The comparison between streaming platforms has become a conversation on social media, particularly on Twitter. User opinions on streaming platforms consist of both positive and negative comments, influencing other users who want to watch movies. Therefore, these opinions can be used for sentiment analysis to reveal and provide information about the public's response to streaming platforms. LDA can be used for topic modeling, and SVM for classification. In the data collection stage using the Scrapy framework with Python, 5,000 data points were gathered text preprocessing and count vectorize was performed. LDA can represent topics and documents, and classification using SVM yielded results that positive comments were more abundant than negative comments. The performance evaluation results showed an accuracy of 91%, precision of 90%, recall of 91%, and an F1 score of 90%. This research found that the number of positive comments is more dominant than negative comments. Therefore, there is a need for balanced data in classifying streaming platform comments so that the classification can provide good predictive results.

Keyword: sentiment analysis, platform streaming, LDA, SVM

SEMARANG
SEKOLAH PASCASARJANA