ANALISIS SENTIMEN MENGGUNAKAN METODE *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY* PADA ULASAN APLIKASI EDLINK

AMIRUL
NIM H1091211052

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

ANALISIS SENTIMEN MENGGUNAKAN METODE *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY* PADA ULASAN APLIKASI EDLINK

AMIRUL
NIM H1091211052

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Analisis Sentimen Menggunakan Metode *Term Frequency-Inverse Document Frequency* Pada Ulasan Aplikasi Edlink
Nama Mahasiswa	: Amirul
NIM	: H1091211052
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 30 Januari 2025
SK Pembimbing	: No. 2627/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	: No. 184/UN22.8/TD.06/2025/Tanggal 23 Januari 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Hendra Perdana, M.Sc.	Neva Satyahadewi, M.Sc., CRA., CRP., CRMP.
NIP. 198810102019031020	NIP. 198212042005012001

Dosen Penguji

Ketua Penguji	Anggota Penguji
Nurfitri Imro'ah, M.Si.	Shantika Martha, M.Si.
NIP. 198907182019032021	NIP. 198403082008122003

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Hendra Perdana, M.Sc.	Neva Satyahadewi, M.Sc., CRA., CRP., CRMP.
NIP. 198810102019031020	NIP. 198212042005012001

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP. 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Amirul

NIM : H1091211052

Program Studi/ Jurusan : Statistika/ Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyaatan ini untuk digunakan sebagaimana mestinya.

Pontianak, 30 Januari 2025

Amirul
NIM H1091211052

Analisis Sentimen Menggunakan Metode *Term Frequency-Inverse Document Frequency* Pada Ulasan Aplikasi Edlink

Abstrak

Aplikasi Edlink sebagai salah satu *platform* pembelajaran elektronik karya anak bangsa yang menyediakan layanan untuk menunjang kegiatan pembelajaran yang digunakan oleh pengajar dan pelajar. Dengan menganalisis ulasan aplikasi Edlink, pengembang aplikasi dapat meningkatkan kualitas layanan berdasarkan pengalaman pengguna. Analisis sentimen dilakukan pada ulasan aplikasi Edlink dari Google *Play Store* menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk menentukan kata-kata yang paling penting dalam ulasan dengan memberikan bobot pada setiap kata. Bobot tersebut dihitung dengan mengalikan nilai frekuensi kemunculan suatu kata dalam sebuah ulasan (TF) dan invers frekuensi ulasan yang mengandung kata tersebut (IDF), sehingga diketahui apakah kata tersebut penting pada setiap ulasan. Data yang digunakan sebanyak 1.800 ulasan pengguna aplikasi Edlink yang telah melalui proses *text preprocessing* yang terdiri dari *cleansing data*, *case folding*, *spelling normalization*, *stopword removal* dan *stemming*. Setelah melalui tahapan *text preprocessing*, ulasan diberi label menggunakan *library transformer* dan skor bintang. Pada proses pelabelan menghasilkan 646 ulasan berlabel positif dan 1.154 ulasan berlabel negatif, sedangkan pelabelan skor bintang menghasilkan 923 ulasan berlabel positif dan 877 ulasan berlabel negatif dengan selisih kesalahan sebesar 15,38%. Hasil pembobotan kata menggunakan metode TF-IDF pada ulasan positif menunjukkan bahwa kata “sangat”, dan “bantu” merupakan kata yang paling penting dalam menunjukkan tingkat kepuasan pengguna terhadap aplikasi, sedangkan pada ulasan negatif, kata “tidak” dan “aplikasi” merupakan kata yang paling penting dalam menunjukkan ketidakpuasan pengguna terhadap aplikasi.

Kata kunci: Kata penting, Frekuensi, Pelabelan, Pembobotan.

Sentiment Analysis Using Term Frequency-Inverse Document Frequency Method on Edlink Application Reviews

Abstract

Edlink application as one of the nation's electronic learning platforms that provides services to support learning activities used by teachers and students. By analyzing Edlink app reviews, app developers can improve service quality based on user experience. Sentiment analysis is conducted on Edlink app reviews from Google Play Store using Term Frequency-Inverse Document Frequency (TF-IDF) method to determine the most important words in the reviews by giving weight to each word. The weight is calculated by multiplying the value of the frequency of occurrence of a word in a review (TF) and the inverse of the frequency of reviews containing that word (IDF), so that it is known whether the word is important in each review. The data used is 1,800 reviews of Edlink application users who have gone through a text preprocessing process consisting of data cleansing, case folding, spelling normalization, stopword removal and stemming. After going through the text preprocessing stage, the reviews are labeled using library transformers and star scores. The labeling process resulted in 646 positively labeled reviews and 1,154 negatively labeled reviews, while the star score labeling resulted in 923 positively labeled reviews and 877 negatively labeled reviews with an error difference of 15.38%. The results of word weighting using the TF-IDF method on positive reviews show that the words “very”, and “help” are the most important words in indicating the level of user satisfaction with the application, while on negative reviews, the words “no” and “application” are the most important words in indicating user dissatisfaction with the application.

Keywords: *Frequency, Important word, Labeling, Weighting*

PRAKATA

Puji dan syukur penulis panjatkan kehadirat Tuhan yang Maha Esa karena atas segala berkat, rahmat, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi. Skripsi yang berjudul **"Analisis Sentimen Menggunakan Metode *Term Frequency-Inverse Document Frequency* Pada Ulasan Aplikasi Edlink"** merupakan salah satu syarat untuk memperoleh gelar Sarjana Statistika pada Program Studi Statistika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.

Penulisan skripsi ini tidak terlepas dari dukungan dan bimbingan dari beberapa pihak, oleh karena itu penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada

1. Kedua orang tua saya, saudara laki-laki saya, saudara perempuan saya, adik laki-laki saya dan keluarga saya yang saya cintai yang telah memberikan dorongan, nasehat, serta dukungan moril dan materil.
2. Ibu Wirda Andani, S.Si., M.Si., selaku Dosen Pembimbing Akademik yang telah memberikan bimbingan, arahan, dukungan, serta motivasi untuk penilis selama perkuliahan.
3. Bapak Hendra Perdana, M.Sc., selaku Dosen Pembimbing I yang telah memberikan arahan dan motivasi untuk menyelesaikan penulisan skripsi ini.
4. Ibu Neva Satyahadewi, M.Sc., CRA., CRP., CRMP., selaku Dosen Pembimbing II yang juga telah memberikan arahan serta motivasi untuk menyelesaikan penulisan skripsi ini.
5. Ibu Nurfitri Imro'ah, M.Si selaku Dosen Penguji I yang telah memberikan saran dan masukan untuk perbaikan skripsi ini sehingga menjadi lebih baik.
6. Ibu Shantika Martha, M.Si selaku Dosen Penguji II yang telah memberikan saran dan masukan untuk perbaikan skripsi ini sehingga menjadi lebih baik.
7. Teman-teman Statistika Angkatan 2021 yang telah membersamai ketika perkuliahan serta memberikan semangat dalam penulisan skripsi ini.
8. Semua orang yang saya kenal yang memberikan dukungan, nasehat, motivasi, dan doa yang tidak dapat saya sebutkan satu-satu.

Penulis berharap dapat memperoleh saran dan masukan untuk kesempurnaan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkannya. Akhir kata penulis mengucapkan terima kasih kepada semua pihak yang telah berperan dari awal hingga akhir dalam penulisan skripsi ini.

Pontianak, 30 Januari 2025

Amirul

DAFTAR ISI

LEMBAR PENGESAHAN ii
PERNYATAAN INTEGRITAS AKADEMIK ii
Abstrak iv
Abstract v
PRAKATA vi
DAFTAR ISI viii
DAFTAR TABEL x
DAFTAR SIMBOL xi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 2
1.3 Tujuan Penelitian 2
1.4 Batasan Masalah 3
1.5 Tinjauan Pustaka 3
1.6 Metodologi Penelitian 4
BAB II LANDASAN TEORI 6
2.1 Ulasan Aplikasi 6
2.2 *Data Mining* 6
2.3 *Text Mining* 7
2.4 Analisis Sentimen 8
2.5 *Machine Learning* 9
2.6 Metode TF-IDF 10
BAB III *WEB SCRAPING* DAN ANALISIS SENTIMEN MENGGUNAKAN METODE *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY* ... 11
3.1 *Web Scraping* 11
3.2 *Text Preprocessing* 11
3.3 Pembobotan TF-IDF 12
BAB IV HASIL DAN PEMBAHASAN 15

4.1 Pengambilan Data Ulasan Menggunakan Teknik *Web Scraping*.....15
4.2 *Text Preprocessing*.....15
4.2.1 *Cleansing*.....16
4.2.2 *Case Folding*.....16
4.2.3 *Tokenizing*.....17
4.2.4 *Spelling Normalization*.....17
4.2.5 *Stopword Removal*.....18
4.2.6 *Stemming*.....19
4.3 Pelabelan Data.....19
4.4 Pembobotan Kata TF-IDF.....21
4.4.1 *Term Frequency* (TF).....21
4.4.2 *Inverse Document frequency* (IDF).....22
4.4.3 *Term Frequency – Inverse Document Frequency* (TF-IDF).....22
BAB V KESIMPULAN DAN SARAN..... 24
5.1 Kesimpulan.....24
5.2 Saran.....24
DAFTAR PUSTAKA..... 25

DAFTAR TABEL

Tabel 4. 1 Proses *Cleansing* 16
Tabel 4. 2 Proses *Case Folding* 16
Tabel 4. 3 Proses *Tokenizing* 17
Tabel 4. 4 Proses *Spelling Normalization* 18
Tabel 4. 5 Proses *Stopword Removal* 18
Tabel 4. 6 Proses *Stemming* 19
Tabel 4. 7 Pelabelan *Library Transformer* 20
Tabel 4. 8 Perbandingan Pelabelan 20
Tabel 4. 9 *Term Frequency* 21
Tabel 4. 10 *Inverse Document frequency* 22
Tabel 4. 11 *Term Frequency – Inverse Document Frequency* 22
Tabel 4. 12 Kata-Kata Paling Penting Pada Ulasan Positif 23
Tabel 4. 13 Kata-Kata Paling Penting Pada Ulasan Negatif 23

DAFTAR SIMBOL

$W_{tf_{(p,n)}}$	: Bobot frekuensi kata ke-*p* pada ulasan ke-*n*
$n_{p,n}$	: Banyak kata ke-*p* pada ulasan ke-*n*
n_n	: Banyak kata pada ulasan ke-*n*
idf_p	: Invers bobot frekuensi ulasan pada kata ke-*p*
N	: Banyak ulasan
df_p	: Banyak ulasan yang mengandung kata ke-*p*
$W_{p,n}$	: Bobot kata ke-*p* pada ulasan ke-*n*
$N_{positif}$	: Banyak ulasan aplikasi Edlink label positif
$N_{negatif}$	: Banyak ulasan aplikasi Edlink label negatif

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi khususnya internet saat ini telah mengalami pertumbuhan yang pesat. Perkembangan teknologi informasi dan komunikasi yang begitu pesat memiliki dampak yang signifikan terhadap proses pembelajaran di berbagai lembaga pendidikan (Rusydi, 2019). Suatu bentuk pemanfaatan teknologi informasi yang telah diterapkan dalam sektor pendidikan adalah *e-learning*.

E-Learning merupakan sebuah pembelajaran berbasis media elektronik, *e-learning* memungkinkan proses belajar dan mengajar dikelas virtual dengan memanfaatkan teknologi informasi seperti komputer, ataupun handphone dan internet sebagai medianya (Astuti & Febrian, 2019). Beberapa aplikasi yang digunakan pembelajaran elektronik seperti *Moodle, Schoolgy,* EdLink, Google *Classroom*, dan aplikasi pembelajaran berbasis media elektronik lainnya.

Edlink merupakan aplikasi yang dikembangkan oleh Perseroan Terbatas (PT) Sentra Vidya Utama (Sevima). Berbeda dengan aplikasi pembelajaran elektronik lainnya yang kebanyakan berasal dari luar negeri, aplikasi Edlink merupakan karya anak bangsa yang dirancang dengan menyesuaikan kultur pembelajaran yang ada di Indonesia. Aplikasi Edlink dalam pembelajaran berbasis elektronik memungkinkan dosen bertindak sebagai pengelola materi perkuliahan, tugas, dan nilai sedangkan mahasiswa dapat melihat materi perkuliahan, mengerjakan tugas, serta dapat melihat nilai yang diperoleh (Darwanto & Khasanah, 2021).

Pada penggunaannya, aplikasi Edlink tentu memiliki kekurangan dan kelebihan. Pengalaman pengguna aplikasi Edlink sebagai media pembelajaran dapat dilihat pada kolom ulasan di situs Google *Play Store*, ulasan berupa kritik maupun kepuasan karena setiap pengguna memiliki pemikiran dan pendapat yang berbeda terhadap aplikasi Edlink. Untuk menghadapi permasalahan opini atau sentimen pengguna terhadap layanan yang disediakan oleh aplikasi Edlink

diperlukan pendekatan yaitu analisis sentimen, yang mana analisis ini dapat memberikan informasi tentang pengalaman pengguna dan membantu perusahaan untuk meningkatkan kualitas produk atau layanan.

Analisis sentimen bertujuan untuk menentukan pandangan dari masyarakat mengenai suatu topik (Siregar, Ladayya, Albaqi, & Wardana, 2023). Analisis sentimen dilakukan dengan mengklasifikasikan pandangan masyarakat misalnya ulasan aplikasi Edlink positif dan negatif. Hasil analisis sentimen dapat digunakan pemilik aplikasi untuk melakukan evaluasi atau strategi yang lebih baik kedepannya. Salah satu metode yang digunakan untuk analisis sentimen ialah teknik pembobotan kata menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF).

Metode TF-IDF adalah metode statistik yang digunakan untuk mengukur pentingnya sebuah kata dalam suatu ulasan (Annisa, Kalifia, Bisnis, Humaniora, & Yogyakarta, 2024). Metode TF-IDF menghasilkan bobot atau skor pada setiap kata, kata dengan skor TF-IDF yang tinggi menunjukkan bahwa kata tersebut penting dalam memberikan informasi yang lebih relevan pada semua ulasan.

Berdasarkan latar belakang tersebut, pada penelitian ini akan dilakukan analisis sentimen terhadap ulasan aplikasi Edlink. Analisis sentimen ini menggunakan metode TF-IDF untuk mengidentifikasi kata-kata yang paling penting pada setiap ulasan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan, rumusan masalah dalam penelitian ini, yaitu bagaimana analisis sentimen dalam menentukan kata-kata penting dengan metode TF-IDF pada ulasan aplikasi Edlink?.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan dalam penelitian ini, yaitu menganalisis sentimen untuk mencari kata-kata penting dengan metode TF-IDF pada ulasan aplikasi Edlink.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan merupakan data ulasan pengguna aplikasi Edlink di Google *Play Store* pada tanggal 23 Februari 2017 sampai 24 September 2024.
2. Pengambilan data ulasan pada Google *Play Store* dilakukan dengan menggunakan *softwere python* pada tanggal 29 September 2024.
3. Ulasan yang diambil merupakan ulasan aplikasi Edlink yang menggunakan Bahasa Indonesia.
4. Pelabelan data ulasan berdasarkan dua kategori yaitu positif dan negatif menggunakan *Library Transformer* dan skor bintang.

1.5 Tinjauan Pustaka

Rofiqi, Fauzan, Agustin, dan Saputra (2019) melakukan penelitian menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mencari relevansi dokumen berdasarkan *Query*. Penelitian ini melibatkan tokenisasi, penghapusan kata-kata umum, stemming, dan pembobotan berdasarkan frekuensi kemunculan kata dalam dokumen. Dengan mengalikan nilai TF dan IDF, menghasilkan peringkat skor untuk setiap dokumen dengan skor tertinggi dianggap paling relevan. Hasil penelitian menujukan bahwa dokumen pertama memiliki skor tertinggi sebesar 3,90847, sehingga dokumen pertama paling relevan dibandingkan dokumen lainnya.

Kosasih dan Alberto (2021) melakukan penelitian menggunakan metode *Naïve Bayes* dan TF-IDF. Penelitian ini menunjukkan efektivitas kombinasi metode *Naïve Bayes* dan TF-IDF dalam mengklasifikasi sentimen ulasan produk game di Shopee. Jumlah data yang digunakan sebanyak 1000 data ulasan. Berdasarkan hasil penelitian diperoleh tingkat akurasi sebesar 80,2223% dan skor f1 sebesar 0,691372.

Noer (2023) dalam penelitiannya menggunakan metode *Naïve Bayes* dan TF-IDF untuk analisis sentimen terhadap data ulasan Aplikasi Shopee di Google *Play Store.* Dari hasil analisis, akurasi yang didapatkan adalah 90.76%, dimana akurasi ini sudah sangat baik untuk suatu model *machine learning.* Dari hasil analisis dapat

disimpulkan bahwa Algoritma *Naïve Bayes* dan TF-IDF menghasilkan akurasi yang sangat baik.

Penelitian lainnya dilakukan oleh Hidayah dan Dodiman (2024) menggunakan algoritma *Naïve Bayes*, TF-IDF dan *Confusion Matrix* untuk pengklasifikasian saran monitoring dan evaluasi mahasiswa terhadap dosen. Dari hasil analisis nilai *precision* yang didapatkan adalah 89%, *recall* adalah 85% dan akurasinya adalah 85%. Dari hasil analisis, menunjukkan bahwa algoritma memberikan hasil klasifikasi yang sangat baik.

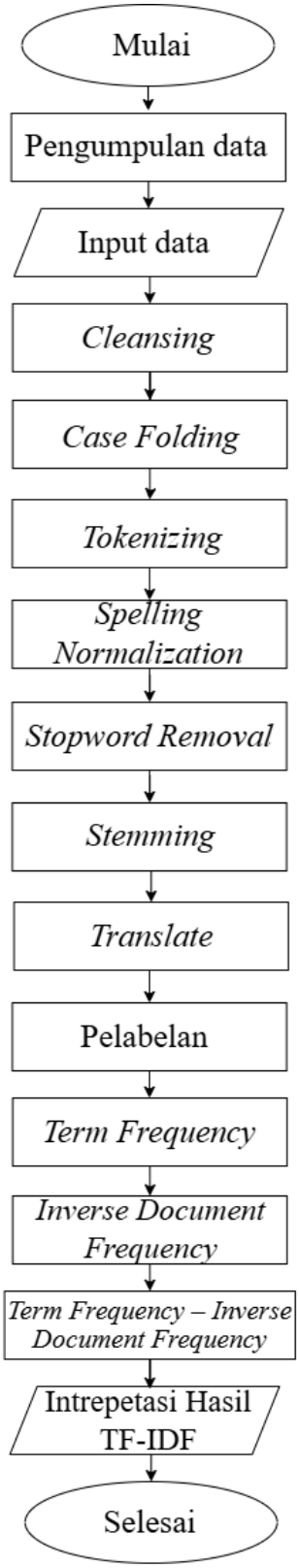
1.6 Metodologi Penelitian

Penelitian ini diawali dengan studi literatur untuk mencari referensi yang relevan dengan permasalahan yang akan diteliti. Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Google *Play Store*. Data yang digunakan merupakan data ulasan aplikasi Edlink. Data ulasan diambil menggunakan *web scrapping* pada tanggal 25 September 2024 berjumlah 2.200 data ulasan aplikasi Edlink dari Google *Play Store*.

Data ulasan yang sudah diambil kemudian masuk tahap preprocessing. Tahapan *preprocesing* yang dilakukan pada proses ini yaitu *cleansing data*, *case folding*, *tokenizing*, *spelling normalization*, *stopword removal*, *stemming*.

Setelah data ulasan melalui tahap preprocessing, data ulasan akan masuk pada tahap pelabelan. Data ulasan akan diberi positif dan negatif. Pelabelan dilakukan secara otomatis menggunakan *library python* yang disebut dengan *transformer*. Pelabelan juga dilakukan dengan dengan skor bintang untuk melihat perbandingannya.

Proses selanjutnya ialah pembobotan kata menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). Metode TF-IDF digunakan untuk menghitung bobot setiap kata pada masing-masing ulasan, sehingga dapat menunjukkan seberapa penting kata tersebut didalam ulasan. Adapun alur dalam penelitian dapat dilihat pada *flowchart* berikut.



Gambar 1. 1 *Flowchart* Penelitian