



**ANALISIS SENTIMEN *BRAND AMBASSADOR* BTS TERHADAP  
TOKOPEDIA MENGGUNAKAN KLASIFIKASI *BAYESIAN NETWORK*  
DENGAN EKSTRAKSI FITUR TF-IDF**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh  
REGINA  
1711016120018**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**APRIL 2023**



**ANALISIS SENTIMEN *BRAND AMBASSADOR* BTS TERHADAP  
TOKOPEDIA MENGGUNAKAN KLASIFIKASI *BAYESIAN NETWORK*  
DENGAN EKSTRAKSI FITUR TF-IDF**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh  
REGINA  
1711016120018**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**APRIL 2023**

# SKRIPSI

## ANALISIS SENTIMEN *BRAND AMBASSADOR* BTS TERHADAP TOKOPEDIA MENGUNAKAN KLASIFIKASI *BAYESIAN NETWORK* DENGAN EKSTRAKSI FITUR TF-IDF

Oleh:

REGINA

NIM 1711016120018

Telah dipertahankan di depan Dosen Penguji pada Tanggal 10 April 2023

Susunan Dosen Penguji:

Pembimbing I



**Triando Hamonangan Saragih,**  
**S.Kom., M.Kom**  
NIP. 199308242019031012

Dosen Penguji I



**Radityo Adi Nugroho, S.T., M.Kom**  
NIP. 198212042008011006

Pembimbing II



**Dwi Kartini, S.Kom., M.Kom.**  
NIP. 198704212012122003

Dosen Penguji II



**Friska Abadi, S.Kom., M.Kom.**  
NIP.1 9880913201612110001



## PERNYATAAN

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

Banjarbaru, 10 April 2023

Yang Menyatakan



**Regina**

**NIM. 1711016120018**

## ABSTRAK

### **ANALISIS SENTIMEN *BRAND AMBASSADOR* BTS TERHADAP TOKOPEDIA MENGGUNAKAN KLASIFIKASI *BAYESIAN NETWORK* DENGAN EKSTRAKSI FITUR TF-IDF**

(Oleh: Regina; Pembimbing: Triando Hamonangan Saragih, S.Kom., M.Kom dan Dwi Kartini, S.Kom., M.Kom; 2023; 70 halaman)

BTS (*Bangtan Boys*) adalah salah satu *boy* grup asal korea selatan yang ditunjuk oleh Tokopedia untuk menjadi *brand ambassador* Tokopedia di Indonesia, BTS merupakan salah satu *boy* grup yang sangat mendunia dalam bidang musik dan Tokopedia merupakan salah satu *E-commerce* terkenal yang banyak digunakan oleh masyarakat indonesia untuk melakukan jual beli online. Kerjasama ini tentu saja memberikan pengaruh terhadap Tokopedia serta memperoleh banyak respon berupa opini masyarakat terutama pada media sosial *twitter*, karena hal tersebut maka dilakukan penelitian analisis sentiment. Data yang digunakan yaitu 900 data *tweet* dan terbagi menjadi 3 kelas yaitu positif, negatif, dan netral. Tahapan penelitian terdiri dari pengambilan dan pengumpulan data, *preprocessing* data, ekstraksi fitur dengan *Term Frequency - Inverse Document Frequency* (TF-IDF), klasifikasi dengan *Bayesian network*, evaluasi kinerja menggunakan *K-fold cross validation* (K-10) dan *confusion matrix*. Perbandingan terjadi pada tahap *preprocessing* data, yaitu saat menggunakan normalisasi data dan tidak menggunakan normalisasi data, dari hasil perbandingan tersebut diperoleh nilai akurasi jika tidak menggunakan normalisasi data sebesar 66,6667%, presisi sebesar 68,1%, dan recall sebesar 66,7%. Sedangkan hasil akurasi dengan menggunakan normalisasi data sebesar 76,5556%, *presisi* sebesar 77,4%, dan *recall* sebesar 76,6%. Selisih nilai akurasi dari kedua percobaan sebesar 9,8889 %, hal ini membuktikan bahwa menggunakan normalisasi data lebih baik.

**Kata Kunci:** Analisis sentiment, Data mining, Brand ambassador BTS, Tokopedia, TF-IDF, *Bayesian Network*

## ABSTRACT

### ***SENTIMENT ANALYSIS OF BTS AS A BRAND AMBASSADORS OF TOKOPEDIA USING BAYESIAN NETWORK CLASSIFICATION WITH TF-IDF FEATURE EXTRACTION***

(Oleh: Regina; Pembimbing: Triando Hamonangan Saragih, S.Kom., M.Kom dan Dwi Kartini, S.Kom., M.Kom; 2023; 70 halaman)

*BTS (Bangtan Boys) is one of the boy groups from South Korea appointed by Tokopedia to become Tokopedia's brand ambassador in Indonesia, BTS is one of the boy groups that is very global in the field of music and Tokopedia is one of the famous E-commerce which is widely used by Indonesian people to buy and sell online. This collaboration certainly has an influence on Tokopedia and has received many responses in the form of public opinion, especially on social media twitter, because of this, sentiment analysis research was conducted. The data used is 900 tweet data and is divided into 3 classes, namely positive, negative, and neutral. The research stages consist of data collection and retrieval, data preprocessing, feature extraction with Term Frequency - Inverse Document Frequency (TF-IDF), classification with Bayesian network, performance evaluation using K-fold cross validation (K-10) and confusion matrix. Comparison occurs at the data preprocessing stage, namely when using data normalization and not using data normalization, from the results of this comparison, the accuracy value obtained if not using data normalization is 66.6667%, precision is 68.1%, and recall is 66.7%. While the accuracy result using data normalization is 76.5556%, precision is 77.4%, and recall is 76.6%. The difference in the accuracy value of the two experiments is 9.8889%, this proves that using data normalization is better.*

**Kata Kunci:** *Sentiment analysis, Data Mining, Brand ambassador BTS, Tokopedia, TF-IDF, Bayesian Network*

## PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh. Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen *Brand Ambassador* BTS Terhadap Tokopedia Menggunakan Klasifikasi *Bayesian Network* Dengan Ekstraksi Fitur TF-IDF” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiyamah*.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga terutama kepada orang tua dan adik perempuan saya juwita maharani yang selalu memberikan bantuan berupa semangat, doa, serta dukungan berupa moril dan materil dalam proses penyelesaian skripsi ini.
2. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
3. Ibu Dwi Kartini, S.Kom., M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom selaku dosen pembimbing akademik serta Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA UNLAM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman keluarga Ilmu Komputer angkatan 2017, terutama Noor Hidayah, Nurhuda, Sartika Dewi, Sulastri Nor Indah Sari, dan Mia Lidiyani, Maria Ulfah, dan Muhammad Rofiq, terima kasih untuk canda, tawa, perjuangan yang sudah dilewati bersama, untuk semua kenangan manis yang telah terukir selama ini. Senang bisa menjadi salah satu bagian dari kehidupan kalian.

7. Esti Ayu Mahrani, Puji Rahayu, Vera Yuniar, Yuyun Suyaningsih, Alimatus Sya' Diah, Linpirlin, Rohima Rani, Celine Xenia Prajnaparamita, Neelriver Kirkley Hursepuny, dan Shella Tyara atas bantuan dan dukungannya selama ini.
8. Kepada semua pihak yang tidak dapat disebutkan satu persatu, yang telah turut membantu dalam penyelesaian skripsi ini. Saya juga ingin mengucapkan terimakasih yang sebanyak-banyaknya.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 10 April 2023



Regina



## DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN .....	ii
PERNYATAAN.....	iii
ABSTRAK .....	iv
ABSTRACT .....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xii
DAFTAR LAMPIRAN.....	xiii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan masalah .....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>5</b>
2.1 Kajian Terdahulu .....	5
2.2 Keaslian Penelitian .....	6
2.3 Twitter .....	8
2.4 <i>E-commerce</i> (Tokopedia) .....	8
2.5 Brand Ambassador (BTS) .....	9
2.6 Data Mining.....	10
2.7 Analisis Sentimen .....	10
2.8 <i>Preprocessing</i> .....	11
2.9 Ekstraksi fitur .....	12
2.10 TF-IDF.....	12
2.11 Klasifikasi.....	13
2.12 <i>Cross Validation</i> .....	14
2.13 <i>Bayesian Network</i> .....	15
2.14 Confution Matriks .....	17

<b>BAB III METODE PENELITIAN .....</b>	<b>19</b>
3.1 Alat Penelitian .....	19
3.2 Bahan Penelitian.....	19
3.3 Variabel Penelitian .....	19
3.4 Prosedur Penelitian.....	20
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>26</b>
4.1 Hasil.....	26
4.1.1 Pengumpulan Data .....	26
4.1.2 Preprocessing Data.....	29
4.1.3 Ekstraksi Fitur .....	37
4.1.4 Klasifikasi .....	41
4.1.5 Evaluasi Kinerja.....	49
4.2 Pembahasan.....	59
<b>BAB V PENUTUP.....</b>	<b>70</b>
5.1 Kesimpulan.....	70
5.2 Saran.....	70
DAFTAR PUSTAKA	
LAMPIRAN	

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Rancangan Penelitian.....	7
Tabel 3. Confusion Matrix Multiclass .....	17
Tabel 4. Kelas Positif.....	21
Tabel 5. Kelas Negatif.....	22
Tabel 6. Kelas Netral.....	22
Tabel 7. Contoh Data Tweet .....	27
Tabel 8. Jumlah Data.....	28
Tabel 9. Penerapan Cleaning Data .....	29
Tabel 10. Penerapan Case Folding Data .....	30
Tabel 11. Penerapan Normalisasi Data .....	31
Tabel 12. Kamus Stopword (Tala & Wibisono, 2016) .....	33
Tabel 13. Tambahan Kamus Stopword.....	33
Tabel 14. Penerapan Stopword Removal Data (Tanpa Normalisasi Data).....	34
Tabel 15. Penerapan Stopword Removal Data (Menggunakan Normalisasi Data) ...	34
Tabel 16. Penerapan Stemming Data (Tanpa Normalisasi Data) .....	35
Tabel 17. Penerapan Stemming Data (Menggunakan Normalisasi Data).....	36
Tabel 18. Contoh Hasil Dari Tokenizing .....	37
Tabel 19. Rumus TF-IDF.....	38
Tabel 20. Hasil Pembobotan TF-IDF.....	39
Tabel 21. Perhitungan TF-IDF.....	40
Tabel 22. Penerapan Pembagian Data Testing Dan Data Training.....	42
Tabel 23. Pembagian Data Training Dan Data Testing Setiap Kelas (Class).....	42
Tabel 24. Perhitungan Prior Probability .....	44
Tabel 25. Frekuensi Kemunculan Term.....	45
Tabel 26. Perhitungan Conditional Probability.....	45
Tabel 27. Hasil Conditional Probability Setiap Term.....	46
Tabel 28. Perhitungan Joint Probability.....	47
Tabel 29. Hasil Joint Probability Setiap Term .....	47

Tabel 30. Perhitungan Posterior Probability .....	48
Tabel 31. Hasil Posterior Probability Setiap Term .....	48
Tabel 32. Hasil Confution Matrix Fold 1-10 .....	50
Tabel 33. Hasil Akurasi Fold 1-10 (Tanpa Normalisasi).....	51
Tabel 34. Hasil Perhitungan Akurasi .....	52
Tabel 35. Hasil Confution Matrix Fold 1-10 .....	52
Tabel 36. Hasil Akurasi Fold 1-10 (Menggunakan Normalisasi) .....	53
Tabel 37. Hasil Perhitungan Akurasi .....	54
Tabel 38. Hasil Perhitungan Presisi .....	55
Tabel 39. Hasil Perhitungan Presisi .....	56
Tabel 40. Hasil Perhitungan Recall.....	56
Tabel 41. Hasil Perhitungan Recall.....	57
Tabel 42. Hasil Kinerja Banyesian Network Dengan TF-IDF.....	58
Tabel 43. Hasil Kinerja Banyesian Network Dengan TF-IDF.....	59
Tabel 44. Confution Matrix 3x3.....	68

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 1. Struktur Bayesian Network .....	16
Gambar 2. Alur Penelitian.....	20
Gambar 3. Contoh Diagram Acyclic Graph (DAG) .....	43
Gambar 4. Diagram Acyclic Graph (DAG) .....	44
Gambar 5. Hasil kinerja aplikasi WEKA (tanpa normalisasi) .....	51
Gambar 6. Hasil kinerja aplikasi WEKA (menggunakan normalisasi) .....	54
Gambar 7. Grafik kinerja Bayesian network dengan pembobotan TF-IDF.....	58
Gambar 8. Grafik kinerja Bayesian network dengan pembobotan TF-IDF.....	58
Gambar 9. Pembagian dan penyimpanan data prediksi .....	66
Gambar 10. Grafik perbandingan kinerja.....	69

## DAFTAR LAMPIRAN

### Lampiran

Lampiran 1. Kamus Slang.....	74
Lampiran 2. Kamus Tala dan Wibisono .....	90
Lampiran 3. Kamus Tambahan Stopword .....	97
Lampiran 4. Sorce code Cleaning .....	100
Lampiran 5. Sorce code Case Folding .....	101
Lampiran 6. Sorce code Normalisasi .....	102
Lampiran 7. Sorce code Stopword Removal.....	103
Lampiran 8. Sorce code Stemming .....	104
Lampiran 9. Sorce code Tokenizing & TF-IDF .....	104