

**ANALISIS OPINI MASYARAKAT MENGENAI UNGGAHAN
ENDORSMENT PRODUK TERHADAP INFLUENCER
SELEBRITI PADA MEDIA SOSIAL INSTAGRAM
MENGUNAKAN METODE KLASIFIKASI NAÏVE BAYES
CLASSIFIER**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat
Sarjana Program Studi Informatika



disusun oleh

ICHA NURA NUGRAHA

19.11.2646

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM

YOGYAKARTA YOGYAKARTA

2023

**ANALISIS OPINI MASYARAKAT MENGENAI UNGGAHAN
ENDORSMENT PRODUK TERHADAP INFLUENCER
SELEBRITI PADA MEDIA SOSIAL INSTAGRAM
MENGUNAKAN METODE KLASIFIKASI NAÏVE BAYES
CLASSIFIER**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ICHA NURA NUGRAHA

19.11.2646

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

PERSETUJUAN

SKRIPSI

**ANALISIS OPINI MASYARAKAT MENGENAI UNGGAHAN
ENDORSMENT PRODUK TERHADAP INFLUENCER
SELEBRITI PADA MEDIA SOSIAL INSTAGRAM
MENGUNAKAN METODE KLASIFIKASI NAÏVE BAYES
CLASSIFIER**

yang disusun dan diajukan oleh

ICHA NURA NUGRAHA

19.11.2646

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Desember 2022

Dosen Pembimbing

Prof. Dr. Kusrini, M.Kom

NIK. 190302106

PENGESAHAN

SKRIPSI

**ANALISIS OPINI MASYARAKAT MENGENAI UNGGAHAN
ENDORSMENT PRODUK TERHADAP INFLUENCER
SELEBRITI PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN
METODE KLASIFIKASI NAÏVE BAYES
CLASSIFIER**

yang disusun dan diajukan oleh

ICHA NURA NUGRAHA

19.11.2646

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Desember 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Prof. Dr. Kusriani, M.Kom

NIK. 190302106

Anggit Dwi Hartanto, M.Kom

NIK. 190302163

Norhikmah M.Kom

NIK. 190302245

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Desember 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **ICHA NURA NUGRAHA**
NIM : **19.11.2646**

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS OPINI MASYARAKAT MENGENAI UNGGAHAN
ENDORSMENT PRODUK TERHADAP INFLUENCER SELEBRITI PADA
MEDIA SOSIAL INSTAGRAM MENGGUNAKAN METODE KLASIFIKASI
NAÏVE BAYES CLASSIFIER**

Dosen Pembimbing : Prof. Dr. Kusriani, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Desember 2022

Yang Menyatakan,



ICHA NURA NUGRAHA

PERSEMBAHAN

Alhamdulillahirobbil'alamin segala puji syukur kehadiran Allah SWT yang telah memberi segala nikmat sempurna dan selawat kepada Nabi kita Muhammad SAW yang kita rindukan syafaatnya di yaumul akhir. Sehingga penulis dapat menyelesaikan tugas akhir skripsi dengan baik. Maka sesungguhnya beserta kesulitan ada kemudahan, sesungguhnya beserta kesulitan ada kemudahan Surat Al-Insyirah Ayat 5-6. Maka tugas akhir ini saya persembahkan untuk :

1. Allah SWT yang selalu memberi harapan dan memberi rahmat hidayah segala sesuatu terhadap hambanya agar selalu senantiasa bersabar, gigih dan pantang menyerah
2. Nabi Muhammad SAW yang selalu mengajarkan contoh keteladanan untuk selalu berperilaku jujur, amanat, kesabaran, rendah hati, ketegaran, kedermawanan dan selalu memaafkan.
3. Alm Ibu yang telah membuat penulis selalu tabah sabar dan ikhlas dalam menjalani kehidupan untuk menjadi selalu kebanggaan keluarga besar.
4. Keluarga besar yang telah membuat penulis tangguh dalam menjalani hidup dan tidak terlena dengan kesenangan waktu muda.
5. Ibu. Prof. Dr. Kusriani, M.Kom selaku dosen Pembimbing yang selalu membuat penulis termotivasi dengan segala pencapaian beliau, besar hormat penulis dapat di bimbing oleh Prof.Dr Kusriani, M.Kom.
6. Bpk. Muh Mudzakir, selaku Konsultan Bisnis dan Wirausaha, Pusat Pengembangan Bisnis UMKM dan Koperasi. Joglo UMKM Business Center, Klaten.
7. Bpk. Eko Nugroho Cahyono S.H, Ibu. Wiwin Nidhianingrum Amd.Keb , Ibu.Sumini S.pd, Bpk. Ponidi, Bpk. Heri Nidhianto yang sudah mendanai.
8. Keluarga besar Universitas Amikom Yogyakarta.
9. Seluruh mantan kekasih yang selalu memotivasi “ Mendang Mending Aku”
10. Group Takeshi dan Informatika 1 yang selalu memberi hiburan canda tawa berbagi cerita selama perkuliahan.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT dan Junjungan Nabi Muhammad SAW atas segala nikmat berkat dan rahmat yang diberikan sehingga penulis dapat menyelesaikan penelitian yang diambil dengan judul “Analisis Sentimen Opini Masyarakat Mengenai Endorsment Produk Influencer Selebriti Pada Media Sosial Instagram Menggunakan Metode Klasifikasi Naïve Bayes Classifier “.

Penelitian yang dilakukan oleh penulis dan dituangkan ke dalam naskah sebagai skripsi digunakan untuk memenuhi salah satu syarat kelulusan dan gelar jenjang sarjana Program Sastra 1 Universitas Amikom Yogyakarta. Terimakasih atas semua pihak yang telah membantu, membimbing, memberi fasilitas penulis khususnya kepada :

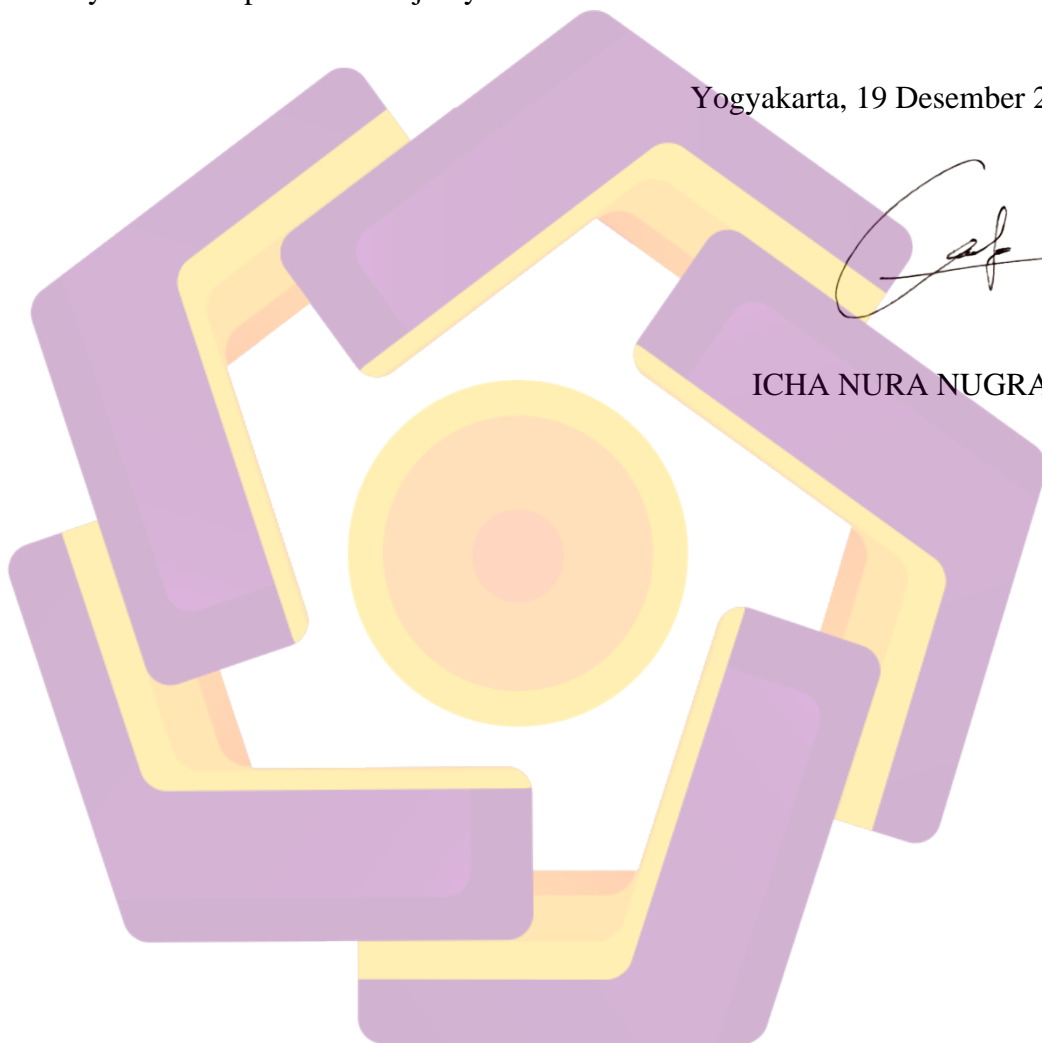
1. Bapak Prof. Dr. M. Suyanto , MM. selaku Rektor Universitas Amikom Yogyakarta
2. Ibu Prof. Dr. Kusrini , M.Kom. selaku dosen pembimbing yang telah membimbing, mengarahkan, memberi saran dengan sabar, perhatian, semangat, dan motivasi kepada penulis.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom. Selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak. Muh Mudzakir, Selaku Konsultan Bisnis dan Wirausaha, Pusat Pengembangan Bisnis UMKM dan Koperasi. Joglo UMKM Business Center, Klaten.
5. Alm. Ibu Ratna Nidhi Astuti S.Ak Selaku Ibu yang selalu memotivasi agar menjadi seorang yang Tangguh.
6. Bapak. Eko Nugroho Cahyono S.H, Ibu. Wiwin Nidhianingrum Amd.Keb , Ibu.Sumini S.pd, Bapak. Ponidi, Bapak. Heri Nidhianto yang sudah mendanai.
7. Keluarga Besar Universitas Amikom Yogyakarta.
8. Group Takeshi dan Informatika 1 yang selalu memberi hiburan.

Segala bentuk yang ada pada naskah skripsi masih banyak kekurangan yang perlu untuk dibenahi agar menjadi lebih baik harapanya. Dilain kesempatan semoga penulis dapat memberi penelitian lebih baik untuk selanjutnya. Akhir kata semoga naskah skripsi yang telah ditulis dapat bermanfaat bagi pembaca khususnya masyarakat dan peneliti selanjutnya.

Yogyakarta, 19 Desember 2022



ICHA NURA NUGRAHA



DAFTAR ISI

SKRIPSI.....	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI.....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	vi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xv
INTISARI.....	xvi
ABSTRAK.....	xvii
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan masalah.....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan penelitian	5
1.5. Manfaat Penelitian.....	5
1.6. Metode Penelitian.....	5
1.6.1. Studi Literatur.....	6
1.6.2. Pengumpulan Data.....	6
1.6.3. Data-Preprocessing.....	6
1.6.4. Metode Klasifikasi.....	6
1.6.5. Testing hasil dan pembahasan	6
BAB II.....	8
TINJAUAN PUSTAKA	8

2.1 Tinjauan Pustaka.....	8
2.2 Landasan Teori	15
2.2.1. Kecerdasan Buatan	15
2.2.1.1. Natural Language Processing	15
2.2.1.2. Document Classification.....	16
2.2.2. Analisis Sentimen.....	16
2.2.3. Media Sosial	16
2.2.3.1. Instagram	17
2.2.3.2. Digital Marketing	17
2.2.3.3. Influencer Marketing	18
2.2.3.4. Scraping	18
2.2.3.4.1. Selenium	19
2.2.3.4.2. Webdriver	19
2.2.4. Pre-processing	19
2.2.4.1. Unicode CLDR	19
2.2.4.2. Conversion	20
2.2.4.3. Case Folding.....	20
2.2.4.4. Data Cleaning	20
2.2.4.5. Tokenizing	20
2.2.4.6. Normalization	20
2.2.4.7. Stopword Removal	21
2.2.4.8. Stemming.....	21
2.2.5. Ekstrasi Fitur dan Pembobotan Kata	21

2.2.5.1. Term Frequency (TF)	22
2.2.5.2. Normalization Term Frequency.....	22
2.2.5.3. Inverse Document Frequency (IDF).....	22
2.2.5.4. Term Frequency – Inverse Document Frequency (TF-IDF)	23
2.2.6. Naïve Bayes Classifier.....	23
2.2.6.1. Multinomial Naïve Bayes Classifier.....	24
2.2.7. K-Fold Cross Validation.....	25
2.2.8. Evaluasi dan Visualisasi	26
2.2.8.1. WordCloud	26
2.2.8.2. Confusion Matrix.....	26
BAB III	29
METODE PENELITIAN	29
3.1. Object Penelitian.....	29
3.2. Alur Metode Penelitian.....	29
3.3. Analisis Kebutuhan Perangkat.....	31
3.3.1. Hardware dan Software	31
3.4. Uraian Metodologi.....	32
3.4.1. Menentukan Target Influencer	32
3.4.2. Unicode CLDR.....	34
3.4.3. Preprocessing.....	36
3.4.3.1. Conversion ASCII.....	36
3.4.3.2. Case Folding	37
3.4.3.3. Cleaning.....	38
3.4.3.4. Tokenizing.....	39

3.4.3.5. Normalization Words	40
3.4.3.6. Stopwords Removal.....	41
3.4.3.7. Stemming.....	42
3.4.3.8. Labeling.....	43
3.4.4. Feature Ekstrasi dan Pembobotan (Term Weight)	43
3.4.4.1. Term Frequency.....	44
3.4.4.2. Term Normalization	44
3.4.4.3. Invers Document Frequency.....	45
3.4.4.4. TF-IDF.....	46
3.4.5. Classification Naïve Bayes Multinomial.....	46
3.4.5.1. Perancangan System Klasifikasi Naïve Bayes Multinomial	46
3.4.6. K-Fold Cross Validation.....	50
3.4.7. Evaluasi	50
3.4.7.1. WordCloud Visualisasi.....	50
3.4.7.2. Confusion Matrix Visual	51
3.4.7.2.1. Accuracy.....	52
3.4.7.2.2. Precision	52
3.4.7.2.3. Recall.....	53
3.4.7.2.4. F1-Score.....	53
BAB IV	54
HASIL DAN PEMBAHASAN.....	54
4.1.Scraping.....	54
4.2.Preprocessing.....	59
4.2.1. Kode Prose Cleaning	59

4.2.2. Kode Proses Unicode Emoji to CLDR dan Terjemah	65
4.2.3. Kode Proses Case Folding	68
4.2.4. Kode Proses Tokenization	68
4.2.5. Kode Proses Normalization Word Dictionary	68
4.2.6. Kode Program Proses Stopwords Removal	70
4.2.7. Kode Program Proses Stemming	71
4.3. Kode Proses Pembobotan Kata (Term Weighting)	72
4.4. Implementasi Classification Naïve Bayes Multinomial	73
4.5. Kode Implementasi K-Fold Cross Validation	73
4.6. Evaluasi Performa Model	74
4.6.1. Data Balance	74
4.6.2. Evaluasi Model Train test Split	75
4.6.3. Evaluasi WordCloud	76
4.6.4. Performance Model Confusion Matrix	77
BAB V	82
PENUTUP	82
5.1. KESIMPULAN	82
5.2. SARAN	83
DAFTAR PUSTAKA	84

DAFTAR GAMBAR

Gambar 3. 1 Alur Metode Penelitian	29
Gambar 3. 2 Detail Alur Metode Penelitian.....	31
Gambar 3. 3 Halaman Instagram Login	32
Gambar 3. 4 Halaman Profile Influencer	32
Gambar 3. 5 Halaman unggahan influencer.....	33
Gambar 3. 6 Inspect Element Unggahan Influencer Selebriti.....	33
Gambar 3. 7 Hasil Scraping	34
Gambar 3. 8 Ilustrasi K-Fold Cross Validation	50
Gambar 3. 9 Visual WordCloud	51
Gambar 4. 1 Proses Scraping Import Multi File Selenium	54
Gambar 4. 2 Proses Scraping Otomatisasi Element Halaman Login.....	55
Gambar 4. 3 Proses Scraping Otomatisasi Load More Comment	56
Gambar 4. 4 Proses Scraping Otomatisasi Load Replies View	57
Gambar 4. 5 Proses Scraping Blok Kode Username dan Comment.....	57
Gambar 4. 6 Proses Scraping Menyimpan Hasil	58
Gambar 4. 7 Kode Proses Cleaning Tag User Replies	59
Gambar 4. 8 Kode Proses Cleaning duplikat emoji	60
Gambar 4. 9 Kode Proses Cleaning add space from emoji.....	61
Gambar 4. 10 Kode Proses Cleaning replace karakter khusus.....	62
Gambar 4. 11 Kode Proses Cleaning Punctuation	62
Gambar 4. 12 Kode Proses Normalization Unicodedata	63
Gambar 4. 13 Kode Proses Replace Single String dan Double Whitespace.....	64
Gambar 4. 14 Kode Proses Insert Data NaN pada Data Only Whitespace.....	64
Gambar 4. 15 Kode Proses Cleaning Data NaN dan Reset Index	65
Gambar 4. 16 Kode Proses Unicode Emoji to CLDR.....	66
Gambar 4. 17 Kode Proses Terjemah Bahasa Inggris ke Indonesia	66
Gambar 4. 18 Kode Proses Case Folding	68

Gambar 4. 19 Kode Proses Tokenization.....	68
Gambar 4. 20 Kode Proses Normalization Word Dictionary	69
Gambar 4. 21 Kode Proses Removing Bracket Token	69
Gambar 4. 22 Kode Proses Stopwords Removal	70
Gambar 4. 23 Kode Program Proses Stemming.....	71
Gambar 4. 24 Kode Program Proses Term Weighting	72
Gambar 4. 25 Kode Implementasi Classification Multinomial	73
Gambar 4. 26 Implementasi K-Fold Cross Validation.....	74
Gambar 4. 27 Nested Scoring K-Fold Cross Validation.....	74
Gambar 4. 28 Data Balance CLDR dan Non-CLDR	75
Gambar 4. 29 Evaluasi Model Train/Test Split	75
Gambar 4. 30 Kode Proses WordCloud Sentimen Comment	76
Gambar 4. 31 Hasil Evaluasi WordCloud CLDR	77
Gambar 4. 32 Hasil Evaluasi WordCloud Non-CLDR.....	77
Gambar 4. 33 Score Model Classification CLDR No-Balance.....	78
Gambar 4. 34 Rata-rata K-fold CLDR No Balance	78
Gambar 4. 35 Score Model Classification CLDR Balance.....	79
Gambar 4. 36 Rata-rata K-fold CLDR Balance	79
Gambar 4. 37 Score Model Classification No-CLDR No-Balance	80
Gambar 4. 38 Rata-rata K-fold No-CLDR No-Balance.....	80
Gambar 4. 39 Score Model Classification No-CLDR Balance.....	81
Gambar 4. 40 Rata-Rata K-fold No-CLDR Balance	81

DAFTAR TABEL

Tabel 2.1. 1 Tinjauan Pustaka.....	12
Tabel 2.2.8.2. 1 Confusion Matrix	26
Tabel 3.4.2. 1 Cleaning Double Emoji.....	34
Tabel 3.4.2. 2 Emoji to CLDR	35
Tabel 3.4.2. 3 Preprocessing Clean Punctuation.....	35
Tabel 3.4.3.1. 1 Preprocessing Normalization ASCII.....	36
Tabel 3.4.3.2. 1 Preprocessing Case Folding.....	37
Tabel 3.4.3.3. 1 Preprocessing Cleaning	38
Tabel 3.4.3.4. 1 Preprocessing Tokenizing	39
Tabel 3.4.3.5. 1 Preprocessing Normalization Words	40
Tabel 3.4.3.6. 1 Preprocessing Stopwords Removal.....	41
Tabel 3.4.3.7. 1 Preprocessing Stemming.....	42
Tabel 3.4.3.8 1 Preprocessing Labeling	43
Tabel 3.4.4.1. 1 Term Frequency	44
Tabel 3.4.4.2. 1 Term Normalization Frequency	45
Tabel 3.4.4.2. 2 Total dokumen mengandung kata i.....	45
Tabel 3.4.4.3. 1 Invers Dokument Frequency (IDF).....	45
Tabel 3.4.4.4. 1 TF-IDF Term Weighting.....	46
Tabel 3.4.5.1. 1 Comment Data Train.....	47
Tabel 3.4.5.1. 2 Prior probabilitas.....	47
Tabel 3.4.5.1. 3 Hitungan Klasifikasi	48
Tabel 3.4.5.1. 4 Data Test	49
Tabel 3.4.7.2. 1 Visual Confusion Matrix.....	51

INTISARI

Perkembangan teknologi tidak dapat dicegah dan dihindari setiap manusia. Revolusi Industri 4.0 dimanfaatkan pelaku bisnis dan wirausaha dalam mendukung sektor bisnis dalam migrasi tradisional ke teknologi digital untuk menjangkau pasar yang luas. Teknologi informasi berkembang pesat salah satunya media sosial Instagram mudahnya penggunaan dalam bertukar informasi dan saling mengikuti dapat menyebabkan setiap pengguna dapat menjadi idola maupun selebriti bagi pengguna lain. Popularitas memberi peluang dan berkolaborasi dengan suatu produk sebagai strategi marketing.

Peneliti melakukan analisis sentimen terhadap influencer selebriti dan produk merek ternama dengan metode Naïve Bayes Classifier. Pengukuran model yang dibantu metode Naïve Bayes mengambil dari data komentar sejumlah unggahan produk endorse yang serupa dari influencer sebagai dataset. Metode Naïve Bayes Classifier menghasilkan akurasi 77% presisi 78% recall 77% f1-score 77% dengan rata-rata cross validation akurasi 73% presisi 71% recall 71% f1-score 71% k-5 iterasi.

Peneliti melakukan uji pada data apabila terdapat seimbang data menghasilkan akurasi 86% presisi 86% recall 86% f1-score 86% dengan rata-rata cross validation akurasi 87% presisi 88% recall 87% f1-score 87% k-5 iterasi. Sedangkan pada data yang telah ditambah Unicode CLDR menghasilkan akurasi 74% presisi 77% recall 73% f1 score 73% dengan rata rata cross validation k-5 iterasi akurasi 73% presisi 75% recal 72% f1 score 71% apabila data dibagi seimbang 1000 data accuracy 81% precision 83% recall 81% f1 score 80 dengan K-fold 5 Iterasi rata-rata accuracy 77% precision 79% recall 77% f1-score 76% . Data 1500 akurasi 82% presisi 84% recall 82% f1 score 82% dengan rata-rata cross validation k-5 akurasi 79% presisi 81% recall 79% f1 score 79%.

Kata Kunci : Analisis Sentimen, Endorsment Produk, Strategi Marketing Selebriti Influencer, Naïve Bayes Classifier, Instagram.

ABSTRAK

Technological developments cannot be prevented and avoided by every human being. The Industrial Revolution 4.0 is used by business people and entrepreneurs to support the business sector in migrating from traditional to digital technology to reach a broad market. Information technology is developing rapidly, one of which is Instagram social media, the ease of use in exchanging information and following each other can cause each user to become an idol or celebrity for other users. Popularity provides opportunities and collaborates with a product as a marketing strategy.

Researchers conducted a sentiment analysis of celebrity influencers and well-known brand products using the Naïve Bayes Classifier method. The measurement model assisted by the Naïve Bayes method takes from the comment data a number of uploads of similar endorsed products from influencers as a dataset. The Naïve Bayes Classifier method produces an accuracy of 77% precision 78% recall 77% f1-score 77% with an average cross validation accuracy of 73% precision 71% recall 71% f1-score 71% k-5 iterations,

The researcher tests the data when there is balanced data yielding accuracy 86% precision 86% recall 86% f1-score 86% with an average cross validation accuracy 87% precision 88% recall 87% f1-score 87% k-5 iterations. Whereas the data that has been added to Unicode CLDR produces an accuracy of 74% precision 77% recall 73% f1 score 73% with an average cross validation k-5 iterations accuracy 73% precision 75% recall 72% f1 score 71% if the data is divided equally 1000 data accuracy 81% precision 83% recall 81% f1 score 80 with K-fold 5 Iterations average accuracy 77% precision 79% recall 77% f1-score 76%. Data 1500 accuracy 82% precision 84% recall 82% f1 score 82% with average cross validation k-5 accuracy 79% precision 81% recall 79% f1 score 79%.

Keywords : Sentiment Analysis, Product Endorsement, Celebrity Influencer Marketing Strategy, Naïve Bayes Classifier, Instagram.