

**ANALISIS SENTIMEN WARGANET TWITTER TERHADAP PROVIDER  
BY.U MENGGUNAKAN METODE NAIVE BAYES CLASSIFICATION**

**SKRIPSI**

untuk memenuhi sebagian persyaratan mencapai derajat Sarjana S-1

Program Studi Teknik Informatika



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

Disusun oleh:

MANAARUL HIDAYAT

16650071

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA  
2020**

## HALAMAN PENGESAHAN



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

### PENGESAHAN TUGAS AKHIR

Nomor : B-321/Un.02/DST/PP.00.9/02/2021

Tugas Akhir dengan judul : ANALISIS SENTIMEN WARGANET TWITTER TERHADAP PROVIDER BY.U  
MENGUNAKAN METODE NAIVE BAYES CLASSIFICATION

yang dipersiapkan dan disusun oleh:

Nama : MANAARUL HIDAYAT  
Nomor Induk Mahasiswa : 16650071  
Telah diujikan pada : Rabu, 03 Februari 2021  
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

#### TIM UJIAN TUGAS AKHIR



Ketua Sidang

Rahmat Hidayat, S.Kom., M.Cs.  
SIGNED

Valid ID: 602a0ce55ba32



Penguji I

Muhammad Didik Rohmad Wahyudi, S.T.,  
MT.  
SIGNED

Valid ID: 6020f2d9cbfba



Penguji II

Dr. Agung Fatwanto, S.Si., M.Kom.  
SIGNED

Valid ID: 602a466c46126



Yogyakarta, 03 Februari 2021  
UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi  
Dr. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 602b9c2fceda1

## SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

### SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

*Assalamu 'alaikum wr. wb.*

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Manaarul Hidayat

NIM : 16650071

Judul Skripsi : Analisis Sentimen Warganet Twitter terhadap Provider By.U

Menggunakan Metode Naive Bayes Classification


sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

*Wassalamu 'alaikum wr. wb.*

Yogyakarta, 22 Januari 2021

Pembimbing

  
Rahmat Hidayat, S.Kom., M.Cs.  
NIP. 19850514 201503 1 002

## SURAT PERNYATAAN KEASLIAN SKRIPSI/TUGAS AKHIR

### PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini:

Nama : Manaarul Hidayat  
NIM : 16650071  
Program Studi : Teknik Informatika  
Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul **"Analisis Sentimen Warganet Twitter terhadap Provider By.U Menggunakan Metode Naive Bayes Classification"** merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 28 Januari 2021

STATE ISLAMIC UNIVERSITY  
SUNAN KALIDIGRA  
YOGYAKARTA



Manaarul Hidayat  
NIM. 16650071

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

*Alhamdulillahirabbil'alamin*, puji dan syukur atas kehadiran Allah SWT., yang senantiasa melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Analisis Sentimen Warganet Twitter terhadap *Provider* By.U menggunakan Metode *Naïve Bayes Classification*”. Sholawat serta salam semoga senantiasa tercurahkan kepada junjungan kita Nabi Muhammad SAW yang telah membawa kita dari zaman jahiliyah menuju zaman yang terang benderang. Semoga kita termasuk umatnya yang mendapatkan syafa'at pada Hari Kiamat nanti, aamiin.

Skripsi ini disusun dengan tujuan untuk memperoleh gelar sarjana Teknik Informatika di Universitas Islam Negeri Sunan Kalijaga. Penulis sangat menyadari bahwa skripsi ini tidak dapat terwujud dengan tenaga seorang diri. Oleh karena itu, penulis mengucapkan terimakasih kepada semua pihak yang ikut andil dalam proses penyusunan skripsi ini secara langsung maupun tidak langsung. Ucapan terimakasih penulis sampaikan kepada :

1. Bapak Prof. Dr. Phil. Al Makin S.Ag., M.A. selaku Rektor UIN Sunan Kalijaga
2. Ibu Dr. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi.
3. Ibu Maria Ulfah Siregar, S.Kom. MIT., Ph.D, selaku Ketua Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta
4. Bapak Rahmat Hidayat, S.Kom., M.Cs., selaku Dosen Pembimbing Skripsi.
5. Bapak Muhammad Didik Rohmad Wahyudi, S.T., MT, selaku Dosen Pembimbing Akademik
6. Seluruh Bapak/Ibu Dosen Teknik Informatika UIN Sunan Kalijaga Yogyakarta.
7. Kedua Orang tua penulis yang senantiasa memberikan motivasi, kepercayaan, serta dukungan dalam bentuk moril maupun materiil kepada penulis.
8. Seseorang yang selalu menemani, selalu mendengarkan keluh kesah dan hadir saat suka maupun duka penulis, Aprida Wati. Terima kasih atas dorongan semangatnya dan asupan-asupan yang bermanfaat bagi jiwa dan hati penulis.



9. Teman-teman seperjuangan Program Studi Teknik Informatika 2016 UIN Sunan Kalijaga.
10. Teman-teman seperjuangan dan keluarga Kos Pak Yuli.
11. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu dalam penyusunan skripsi ini.

Semoga Allah S.W.T. membalas segala kebaikan semua pihak yang ikut andil dalam penyusunan skripsi ini. Penulis menyadari dalam penyusunan skripsi ini, penulis melakukan banyak kesalahan, dan jauh dari kata sempurna. Oleh karena itu penulis dengan senang hati menerima saran dan kritik agar menjadi lebih baik. Semoga penelitian ini dapat bermanfaat bagi pembaca. Akhir kata penulis ucapkan terimakasih,

*Wassalamu'alaikum Wr. Wb.*

Yogyakarta, 31 Januari 2021



Manaarul Hidayat



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan kebahagiaan, skripsi ini saya persembahkan untuk:

1. Orang tua yang memberikan saya “kehidupan”, Bapak Mad Sofan dan Ibu Mudrikah. Terimakasih telah menjadi orang tua yang selalu mendukung, selalu mendo’akan, selalu memberikan kebahagiaan, selalu menjadi alasan dalam setiap pencapaian yang penulis peroleh, dan selalu menjadi *support system* dalam setiap kondisi yang dihadapi penulis tanpa pamrih dan mengeluh sedikitpun. Tentu semua pemberian tersebut tidak dapat saya balas dengan apapun.
2. Segenap keluarga besar Bani Munadi, Terimakasih telah menjadi panutan, yang selalu memberi dukungan serta motivasi yang tiada terhenti untuk penulis.
3. Bapak Rahmat Hidayat, S.Kom., M.Cs., yang telah menyediakan waktunya untuk membimbing penulis dalam penelitian ini dengan sepenuh hati, sabar dan penuh totalitas.
4. Teruntuk dambaan hati, Aprida Wati yang selalu ada dan menjadi tempat keluh kesah serta *support system* penulis, baik diwaktu senang maupun susah, terimakasih yang selalu meluangkan waktu untuk penulis.
5. Sahabat-sahabat masa kecil penulis, Abdullah, dan Huda Nur Rafik. Terimakasih sudah memberikan warna selama masa kecil penulis.
6. Hernanto Feriansyah dan Muhammad Rizaldi sebagai salah satu sahabat pertama penulis di masa-masa awal perantauan yang senantiasa memberi bantuan. Penulis tidak akan pernah lupa atas semua yang sudah diberikan.
7. Rio Walua dan Adien Tsaqif sebagai sahabat penulis yang senantiasa memberikan dukungan baik dalam bentuk bantuan serta moril di fase-fase sulit penulis.

8. Teman-teman seperjuangan dan keluarga Kos Pak Yuli yang sudah memberikan banyak pelajaran hidup selama di perantauan.
9. Teman seperjuangan di Teknik Informatika 2016. Terimakasih sudah menjadi bagian dari pembelajaran selama ini, terimakasih supportnya, terimakasih semua sarannya, semoga yang terbaik kembali kepada kalian semua.
10. Teman-teman KKN Kelompok 243 Angkatan 99, Mas Jack, Zul, Rifda, Tari, May, Retno, Devi, terimakasih telah memberikan pengalaman dan pembelajaran hidup yang sangat berharga.

Kepada seluruh rekan yang tidak sempat saya tuliskan, terimakasih atas segala pengalaman berharganya. Semoga Allah SWT senantiasa memberikan keberkahan dan kebahagiaan. Terimakasih.



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA



## Halaman Motto

*“3 Simple Rules of Life: DO, ADAPT or DIE “*

*“3 Aturan Sederhana dalam Hidup: LAKUKAN, BERADAPTASI atau MATI”*



## DAFTAR ISI

<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR .....</b>	<b>iii</b>
<b>SURAT PERNYATAAN KEASLIAN SKRIPSI/TUGAS AKHIR.....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vii</b>
<b>Halaman Motto .....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>DAFTAR GAMBAR.....</b>	<b>xvi</b>
<b>DAFTAR RUMUS .....</b>	<b>xviii</b>
<b>INTISARI .....</b>	<b>xix</b>
<b>ABSTRACT.....</b>	<b>xx</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
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	Kebaruan Penelitian .....	6
1.7	Sistematika Penulisan.....	6
<b>BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....</b>		<b>8</b>
2.1	Tinjauan Pustaka .....	8
2.2	Landasan Teori.....	18
2.2.1	Internet Service Provider (ISP).....	18
2.2.2	By.U .....	18
2.2.3	Twitter.....	19
2.2.4	Python .....	19
2.2.5	<i>Text Mining</i> .....	20
2.2.6	Analisis Sentimen .....	21
2.2.7	TF-IDF .....	21
2.2.8	N-Gram .....	22
2.2.9	<i>Naïve Bayes Classification</i> .....	23
2.2.10	<i>Split Validation</i> .....	25
2.2.11	<i>Confusion Matrix</i> .....	26
<b>BAB III METODE PENELITIAN.....</b>		<b>28</b>
3.1	Metode Penelitian.....	28
3.2	Alur Penelitian.....	28
3.2.1	Studi Pustaka.....	28

3.2.2	Pengumpulan Data .....	29
3.2.3	Seleksi Data .....	29
3.2.4	<i>Preprocessing</i> .....	30
3.2.5	Analisa .....	31
3.2.6	Pembuatan Laporan .....	32
3.3	Perangkat Penelitian .....	32
3.3.1	Perangkat Keras : .....	32
3.3.2	Perangkat Lunak .....	32
3.4	<i>Classification Method</i> .....	33
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>34</b>
4.1	Pengumpulan Data .....	35
4.2	Pelabelan dan Seleksi Data .....	37
4.3	<i>Preprocessing</i> Data .....	39
4.3.1	<i>Cleansing</i> .....	39
4.3.2	<i>Case Folding</i> .....	41
4.3.3	<i>Slangword Removal</i> .....	42
4.3.4	<i>Stopword Removal</i> .....	43
4.3.5	<i>Stemming</i> .....	45
4.4	Pembobotan Kata ( <i>Term Weighting</i> ) .....	46
4.4.1	<i>Term Frequency – Invers Document Frequency (TF-IDF)</i> .....	47

4.4.2	TF-IDF dengan Bi-Grams .....	49
4.4.3	TF-IDF dengan Bi-Grams tanpa <i>Stopword Removal</i> .....	51
4.5	<i>Naïve Bayes Classifier</i> .....	54
4.6	Evaluasi Model Klasifikasi .....	58
4.6.1	Evaluasi TF-IDF dengan Bi-Grams .....	64
4.6.2	TF-IDF dengan Bi-Grams tanpa <i>Stopword Removal</i> .....	70
4.7	Implementasi .....	77
<b>BAB V</b>	<b>PENUTUP</b> .....	<b>89</b>
5.1	Kesimpulan .....	89
5.2	Saran .....	93
<b>DAFTAR PUSTAKA</b> .....		<b>94</b>
<b>LAMPIRAN</b> .....		<b>97</b>
<b>CURRICULUM VITAE</b> .....		<b>117</b>

STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
 YOGYAKARTA

## DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka.....	15
Tabel 2. 2 <i>Confusion Matrix</i> (Han & Kamber, 2011).....	26
Tabel 4. 1 Contoh Data Hasil <i>Crawling</i> .....	36
Tabel 4. 2 Contoh Data dengan Label .....	38
Tabel 4. 3 Contoh Data Sebelum Tahap <i>Cleansing</i> .....	40
Tabel 4. 4 Contoh Data Setelah Tahap <i>Cleansing</i> .....	41
Tabel 4. 5 Contoh Penerapan <i>Case Folding</i> .....	42
Tabel 4. 6 Contoh Penerapan <i>Slangword Removal</i> .....	43
Tabel 4. 7 Contoh Penerapan <i>Stopword Removal</i> .....	45
Tabel 4. 8 Contoh Penerapan <i>Stemming</i> .....	46
Tabel 4. 9 Contoh Data untuk Pembobotan Kata .....	47
Tabel 4. 10 Contoh Perhitungan IDF.....	48
Tabel 4. 11 Contoh Perhitungan TF-IDF.....	49
Tabel 4. 12 Contoh Perhitungan IDF dengan Bi-Grams.....	50
Tabel 4. 13 Contoh Perhitungan TF-IDF dengan Bi-Grams.....	51
Tabel 4. 14 Contoh Perhitungan IDF dengan Bi-Grams tanpa <i>Stopword Removal</i> .....	52
Tabel 4. 15 Contoh Perhitungan TF-IDF dengan Bi-Grams tanpa <i>Stopword Removal</i> .....	53
Tabel 4. 16 Contoh Perhitungan <i>Laplace Smoothing</i> .....	55



Tabel 4. 17 Contoh Perhitungan Kelas Negatif .....	57
Tabel 4. 18 Contoh Perhitungan Kelas Positif.....	57
Tabel 4. 19 Contoh Skenario <i>Split Validation</i> .....	59
Tabel 4. 20 Hasil Pengujian <i>Split Validation</i> .....	63
Tabel 4. 21 Perbandingan .....	69
Tabel 4. 22 Perbandingan Split Validation dengan dan tanpa.....	75



## DAFTAR GAMBAR

Gambar 4. 1 <i>Flowchart</i> Proses Bisnis.....	34
Gambar 4. 2 Grafik <i>Confusion Matrix</i> Skenario Pertama.....	59
Gambar 4. 3 Grafik <i>Confusion Matrix</i> Skenario Kedua .....	61
Gambar 4. 4 Grafik <i>Confusion Matrix</i> Skenario Ketiga .....	62
Gambar 4. 5 Grafik <i>Confusion Matrix</i> Skenario Keempat .....	62
Gambar 4. 6 Grafik <i>Confusion Matrix</i> Skenario Kelima.....	63
Gambar 4. 7 Grafik <i>Confusion Matrix</i> Bi-Grams Skenario Pertama.....	64
Gambar 4. 8 Grafik <i>Confusion Matrix</i> Bi-Grams Skenario Kedua .....	66
Gambar 4. 9 Grafik <i>Confusion Matrix</i> Bi-Grams Skenario Ketiga .....	66
Gambar 4. 10 Grafik <i>Confusion Matrix</i> Bi-Grams Skenario Keempat .....	67
Gambar 4. 11 Grafik <i>Confusion Matrix</i> Bi-Grams Skenario Kelima .....	68
Gambar 4. 12 Grafik <i>Confusion Matrix</i> tanpa <i>Stopword</i> Skenario Pertama.....	70
Gambar 4. 13 Grafik <i>Confusion Matrix</i> tanpa <i>Stopword</i> Skenario Kedua .....	72
Gambar 4. 14 Grafik <i>Confusion Matrix</i> tanpa <i>Stopword</i> Skenario Ketiga .....	72
Gambar 4. 15 Grafik <i>Confusion Matrix</i> tanpa <i>Stopword</i> Skenario Keempat .....	73
Gambar 4. 16 Grafik <i>Confusion Matrix</i> tanpa <i>Stopword</i> Skenario Kelima.....	74

Gambar 4. 17 Grafik Proporsi Sentimen.....	77
Gambar 4. 18 <i>Wordcloud</i> sentimen positif .....	78
Gambar 4. 19 <i>Wordcloud</i> sentimen negatif .....	79
Gambar 4. 20 Kata Paling Sering Muncul Sentimen Positif .....	80
Gambar 4. 21 Kata Paling Sering Muncul Sentimen Positif .....	81
Gambar 4. 22 Topik pada Sentimen Positif.....	83
Gambar 4. 23 Topik pada Sentimen Negatif .....	84
Gambar 4. 24 Sumber <i>Tweet</i> Pengguna by.U .....	85
Gambar 4. 25 <i>Wordcloud</i> Kelas Positif dan Negatif Pengguna Android .....	86
Gambar 4. 26 <i>Wordcloud</i> Kelas Positif dan Negatif Pengguna Iphone.....	86
Gambar 4. 27 <i>Wordcloud</i> Kelas Positif dan Negatif Pengguna Web .....	87
Gambar 4. 28 <i>Wordcloud</i> Kelas Positif dan Negatif Pengguna Ipad.....	87
Gambar 4. 29 <i>Wordcloud</i> kelas Positif dan Negatif Pengguna Mac.....	88

## DAFTAR RUMUS

TF-IDF ( 1 ) .....	22
N-Grams ( 2 ) .....	23
Probabilitas ( 3 ) .....	23
<i>Prior Probability</i> ( 4 ) .....	24
<i>Conditional Probability</i> ( 5 ) .....	24
<i>Conditional Probability dengan Laplace Smoothing</i> ( 6 ) .....	25
<i>Accuracy</i> ( 7 ) .....	26
<i>Precision</i> ( 8 ) .....	27
<i>Recall</i> ( 9 ) .....	27
<i>F1-Score</i> ( 10 ) .....	27

# ANALISIS SENTIMEN WARGANET TWITTER TERHADAP PROVIDER BY.U MENGUNAKAN METODE NAIVE BAYES CLASSIFICATION

Manaarul Hidayat

16650071

## INTISARI

Dengan pertumbuhan jumlah pemuda yang pesat, beberapa *provider* meluncurkan inovasi layanan *provider* digital. By.U sebagai pelopornya sudah melayani segmen ini selama setahun belakangan dan kehadirannya mendapatkan berbagai respon dari warganet. Respon tersebut jika diteliti dapat meningkatkan layanan dan memunculkan inovasi lainnya. Akan tetapi belum ada penelitian yang membahas tentang hal tersebut. Maka dari itu, dilakukan penulisan tugas akhir mengenai Analisis Sentimen terhadap *provider* by.U.

Dalam tugas akhir ini dibuat model klasifikasi dengan 3804 data menggunakan *Naive Bayes Classifier* dan *TF-IDF* dengan *Bi-Grams*. Dilakukan perbandingan dengan menghilangkan salah satu tahapan *preprocessing: stopwords removal*. Dari hasil perbandingan ditemukan *TF-IDF* dengan *Bi-Grams* tanpa penerapan *stopword removal* memiliki nilai performa terbaik. Nilai performa terbaik didapatkan pada skenario *split validation* 90:10 dengan *accuracy* 86,88%, *Precision* 88,24%, *recall* 83,33% dan *f1-score* 85,71%.

Model tersebut digunakan untuk klasifikasi data baru sejumlah 1075 data yang diambil pada 20-27 Desember 2020 dengan hasil prediksi sentimen positif sebesar 54,9% dan sentimen negatif sebesar 45,1%. Kata yang sering muncul pada sentimen positif meliputi “25gb”, “paket”, “promo”, “beli”, “kuota”, “murah”. Kata yang sering muncul pada sentimen negatif meliputi “sinyal”, “lelet”, “jelek”, “admin”, “jaring”. Ditemukan 842 *tweet* yang berasal dari Android, 157 *tweet* berasal dari Iphone.

**Kata Kunci:** Analisis Sentimen, By.U, Klasifikasi, *Naive Bayes*, *TF-IDF*, *Bi-Grams*

# **ANALISIS SENTIMEN WARGANET TWITTER TERHADAP PROVIDER BY.U MENGUNAKAN METODE NAIVE BAYES CLASSIFICATION**

**Manaarul Hidayat**

**16650071**

## **ABSTRACT**

With the rapid growth of the number of youth, several providers have launched digital service provider innovations. By.U as a pioneer has been serving this segment for the past year and its presence has received various responses from netizens. This response, if researched, can improve services and lead to other innovations. However, there is no research that addresses this. Therefore, a final project was written regarding Sentiment Analysis on provider by.U.

In this final project, a classification model was made with 3804 data using Naive Bayes Classifier and TF-IDF with Bi-Grams. Comparison was made by eliminating one of the preprocessing steps: stopword removal. From the comparison, it was found that TF-IDF and Bi-Grams without the application of stopword removal had the best performance values. The best performance value is obtained in the split validation scenario 90:10 with accuracy 86.88%, Precision 88.24%, recall 83.33% and f1-score 85.71%.

The model is used to classify 1075 new data from 20-27 December 2020 with a positive sentiment prediction of 54.9% and a negative sentiment of 45.1%. Words that often appear on positive sentiment include “25gb”, “package”, “promo”, “buy”, “quota”, “cheap”. Words that often appear in negative sentiment include “signal”, “slow”, “bad”, “admin”, “net”. There were 842 tweets from Android, 157 of them came from Iphone.

**Keywords:** Sentiment Analysis, By.U, Classification, Naive Bayes, TF-IDF, Bi-Grams



# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Memasuki era *Industri 4.0* ini internet menjadi bagian penting dalam kehidupan dan seakan menjadi gaya hidup baru. Bahkan internet sudah menjadi kebutuhan primer bagi sebagian orang. Hampir seluruh lapisan masyarakat secara langsung maupun tidak langsung sudah memanfaatkan teknologi ini mulai dari penduduk usia lanjut, orang dewasa, remaja hingga anak-anak. Kegunaan internet sangat beragam, mulai dari yang paling dasar sebagai media bertukar informasi, komunikasi, sampai jual beli, berbisnis dan sebagai media hiburan. Kegunaan yang beragam ini juga yang menjadikan internet sebagai ujung tombak perkembangan zaman sejak pertama kali diperkenalkan pada tahun 1980-an sebagai tulang punggung untuk interkoneksi jaringan akademik dan militer regional Amerika Serikat.

Menurut data terbaru dari survey yang diselenggarakan Asosiasi Penyelenggara Jasa Internet Indonesia (2020) pengguna internet di Indonesia terus meningkat dari tahun ke tahun. Pada tahun 2019, sebanyak 73,7% penduduk Indonesia sudah memanfaatkan internet dalam kesehariannya. Pertumbuhan pengguna internet di Indonesia mengalami pertumbuhan sebesar 8,9% dari periode sebelumnya menjadi berjumlah 196,71 juta jiwa dari total 266,91 juta penduduk Indonesia. Dari jumlah tersebut ada sekitar 42,3% pengguna internet Indonesia

berasal dari kelompok usia 10-29 tahun. Sejalan dengan survey APJII, menurut survey Badan Pusat Statistik (2019) ada sekitar 81,22% penduduk Indonesia dari kalangan pemuda sudah tersentuh layanan internet. Karakteristik pemuda sebagai kelompok masyarakat yang haus akan informasi dan sosialisasi dengan sekitar ini mengakibatkan begitu dekatnya internet dengan kehidupan mereka.

Berdasarkan data tersebut, berbagai *provider* internet di Indonesia merespon positif dengan pertumbuhan pengguna internet di kalangan pemuda. Dalam setahun terakhir, berbagai *provider* internet di Indonesia sudah meluncurkan layanan *provider digital*. Sebut saja telkomsel sebagai pionir dalam bidang ini yang meluncurkan by.U pada bulan oktober 2019 sebagai *provider digital* pertama di Indonesia. Kemudian disusul Smartfren dan XL Axiata yang masing-masing meluncurkan Switch Mobile dan Live.ON. Dan yang terbaru, Indosat Ooredoo juga tak mau kalah dengan masuk ke industri *provider digital* dan meluncurkan MPWR. *Provider-provider digital* ini diluncurkan untuk memenuhi kebutuhan internet kalangan pemuda sehingga layanan dan nilai yang ditawarkan sudah disesuaikan dengan karakteristik pemuda.

Layanan *provider* berbasis digital ini tentu memiliki beberapa perbedaan jika dibandingkan dengan layanan *provider konvensional*. Berbeda dengan kartu konvensional yang bisa dengan mudah ditemukan di outlet terdekat, untuk bisa menggunakan layanan *provider digital* seperti by.U, calon pengguna harus membeli dan mendaftarkan diri secara digital melalui aplikasi yang disediakan di Google Play Store. Selain itu *provider* internet by.U menawarkan kemudahan

dalam menentukan paket internet berdasarkan berbagai parameter: waktu (harian, mingguan, bulanan), kecepatan bandwidth (1 mbps, 2 mbps, dll) dan berdasarkan aplikasi yang sering diakses (Youtube, TikTok, Instagram, dll). Pengguna juga dimudahkan dalam memilih nomor. Pengguna bisa memasukkan kombinasi nomor sesuai yang diinginkan. Dengan berbagai kemudahan yang ditawarkan, selama setahun terakhir *provider* by.U telah menarik banyak pengguna baru yang kebanyakan berasal dari generasi Z.

Dengan semakin maraknya pengguna internet yang beralih ke layanan digital by.U banyak diantaranya yang berbagi kesan positif maupun negatif melalui media sosial. Twitter sebagai salah satu sosial media dengan pengguna harian terbanyak di Indonesia menjadi salah satu opsi utama pengguna by.U untuk berbagi cerita, kesan hingga keluhan mengenai layanan yang ditawarkan. Beberapa pengguna menceritakan kesan positif mengenai *provider digital* baru ini seperti ketika mendapatkan promo. Pengguna yang lain menyampaikan keluhan seperti mengenai jaringan yang terganggu. Sebagai *brand* baru di ranah *provider internet digital*, tentu suara-suara pelanggan mengenai citra dan layanan menjadi sangat penting untuk diperhatikan. Dengan mendengarkan suara penggunanya, by.U bisa lebih mudah dalam mengambil keputusan bisnis seperti hal-hal apa saja yang seharusnya ditingkatkan dan hal-hal apa saja yang wajib diperbaiki. Dengan begitu, nantinya strategi dan kebijakan yang dibuat akan lebih tepat berdasarkan kesan, saran, dan keluhan dari pengguna dan calon pengguna.

Sehingga berdasarkan latar belakang tersebut penelitian ini dibuat untuk menganalisis sentimen pengguna dan calon pengguna *provider digital* by.U di media sosial twitter. Penelitian ini mengambil data *tweet* dari pengguna maupun calon pengguna by.U sejumlah 5596 *tweet* dengan kata kunci byu, by.u, dan hup beta. Analisis sentimen dalam penelitian ini menggunakan pemodelan pembelajaran mesin dengan metode *Naïve Bayes Classifier* yang diharapkan bisa mengklasifikasikan opini pengguna maupun calon pengguna by.U. Dengan adanya penelitian ini diharapkan dapat memberikan wawasan baru berupa hasil penelitian dan selanjutnya bisa digunakan pihak-pihak terkait yang membutuhkan informasi mengenai opini pengguna dan calon pengguna by.U dalam media sosial twitter baik digunakan untuk membuat kebijakan maupun penelitian lanjutan.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, masalah penelitian ini adalah bagaimana penerapan metode *Naive Bayes Classifier* dalam analisis sentimen pengguna dan calon pengguna *provider* by.U di media sosial twitter.

## 1.3 Batasan Masalah

Agar penelitian ini lebih fokus, maka perlu adanya batasan masalah penelitian. Adapun Batasan masalah penelitian ini adalah:

1. Data yang digunakan sejumlah 3804 *tweet*
2. Kata Kunci yang digunakan: byu, by.u, hup beta

3. Penelitian ini hanya membahas klasifikasi sentimen pengguna dan calon pengguna terhadap *provider digital by.U*
4. Klasifikasi *tweet*, mencakup sentimen positif dan negatif
5. Klasifikasi dibuat dengan metode pemodelan *Naïve Bayes Classifier*
6. Proses klasifikasi menggunakan bahasa pemrograman Python

#### 1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Mengetahui sentimen pengguna dan calon pengguna *provider digital by.U* di media sosial twitter
2. Menerapkan metode *Naive Bayes Classifier* dalam menganalisis sentimen pengguna dan calon pengguna *provider digital by.U* di media sosial twitter dan melihat sejauh mana tingkat akurasi pemodelan menggunakan *Confusion Matrix*.
3. Memberikan bahan masukan kepada pihak terkait berdasarkan hasil analisis sentimen pengguna dan calon pengguna *provider digital by.U* di media sosial twitter

#### 1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan memiliki manfaat penelitian sebagai berikut:

1. Memberikan wawasan berupa hasil penelitian yang bisa digunakan sebagai sumber pembelajaran atau sebagai acuan untuk mengambil kebijakan.

2. Menambah wawasan dalam implementasi algoritma *Naïve Bayes* dalam pengklasifikasian sentimen warganet twitter terhadap *provider by.U*.

### 1.6 Kebaruan Penelitian

Penelitian mengenai analisis sentimen menggunakan metode *Naïve Bayes Classifier* sampai saat ini sudah dilakukan oleh beberapa peneliti sebelumnya. Namun berdasarkan referensi dan tinjauan pustaka, penelitian Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta mengenai analisis sentimen menggunakan media sosial twitter menggunakan metode *Naive Bayes Classifier* mengenai opini pengguna dan calon pengguna terhadap *provider by.U* belum pernah ditemukan oleh peneliti.

### 1.7 Sistematika Penulisan

Sebagai gambaran dan kerangka yang jelas mengenai pokok bahasan setiap bab dalam penelitian ini, maka diperlukan sistematika penulisan. Berikut adalah penjelasan pada tiap-tiap bab dalam laporan penelitian ini:

#### BAB I PENDAHULUAN

Pada bab ini dijelaskan latar belakang, rumusan masalah, batasan, tujuan, manfaat, dan sistematika penulisan.



## BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Pada bab ini dijelaskan teori-teori dan penelitian-penelitian terdahulu yang digunakan sebagai acuan atau referensi dalam melakukan penelitian

## BAB III METODE PENELITIAN

Pada bab ini dijelaskan tentang metode ataupun algoritma yang digunakan serta tahapan-tahapan yang akan dilakukan dalam penelitian ini.

## BAB IV HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan hasil dan pembahasan dari implementasi sistem dan juga hasil pengujian sistem

## BAB V PENUTUP

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan data tweet yang diambil dengan menggunakan kata kunci “byu”, “by.u” dan “hup beta” dengan *filter-out tweet* dari akun @BYU dan @BYUfootball dan filter Bahasa Indonesia yang kemudian dilakukan penyeleksian data dengan membuang data yang sumbernya tidak jelas dan mengambil hanya yang bersumber dari perangkat android, iphone, web, ipad, dan mac kemudian diberi label secara manual oleh 2 orang, membuang data yang terduplikasi dan membuang data yang tidak relevan menghasilkan data yang siap diolah berjumlah 3804 data diantara data tersebut sejumlah 1807 data berlabel positif dan sejumlah 2167 data berlabel negatif. Data tersebut diolah menggunakan metode *Naïve Bayes Classifier* dengan ekstraksi fitur TF-IDF dan dievaluasi menggunakan *split validation* kemudian dihitung akurasi dengan *confusion matrix* untuk mendapatkan model terbaik yang selanjutnya diimplementasikan ke dalam data baru untuk mengklasifikasikan *tweet* secara otomatis ke dalam label positif dan negatif.

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa metode *Naïve Bayes Classifier* dan ekstraksi fitur TF-IDF dapat digunakan dengan cukup baik dalam klasifikasi sentimen pengguna *provider by.U*. Dari hasil evaluasi diketahui terdapat perbedaan nilai performa antara TF-IDF tanpa bigrams dan TF-

IDF dengan bigrams. TF-IDF dengan penerapan bigrams mengalami peningkatan nilai performa pada beberapa scenario pembagian data akan tetapi peningkatannya tidak terlalu konsisten. Kemudian dilakukan perbandingan antara TF-IDF tanpa bigrams dan TF-IDF dengan bigrams tanpa penerapan *stopword removal*. Hasilnya TF-IDF dengan bigrams tanpa penerapan *stopword* mengalami peningkatan nilai performa secara konsisten di hampir semua skenario pembagian data. Rasio *split validation* yang digunakan pada masing masing model menunjukkan performa yang berbeda dengan menggunakan evaluasi *confusion matrix*. Pada ekstraksi fitur dengan bigrams tanpa penerapan *stopword removal*, model dengan rasio 90:10 menghasilkan nilai akurasi sebesar 86,88%, *precision* sebesar 88,24%, *recall* sebesar 83,33% dan *f1-score* sebesar 85,71%. Rasio 85:15 menghasilkan nilai akurasi sebesar 84,41%, *precision* sebesar 86,51%, *recall* sebesar 79,85% dan *f1-score* sebesar 83,05%. Rasio 80:20 menghasilkan nilai akurasi sebesar 84,49%, *precision* sebesar 85,96%, *recall* sebesar 80,77% dan *f1-score* sebesar 83,29%. Rasio 75:25 menghasilkan nilai akurasi sebesar 85,17%, *precision* sebesar 85,61%, *recall* sebesar 81,51% dan *f1-score* sebesar 83,51%. Rasio 70:30 menghasilkan nilai akurasi sebesar 84,41%, *precision* sebesar 84,94%, *recall* sebesar 80,42% dan *f1-score* sebesar 82,62%. Dapat disimpulkan model dengan rasio *split* 90:10 mempunyai performa paling tinggi. Selanjutnya model dengan nilai performa tertinggi diimplementasikan terhadap data baru untuk melakukan klasifikasi secara otomatis dengan jumlah data 1075 data *tweet* yang diambil dalam rentang tanggal 20 Desember 2020 - 27 Desember 2020.

Model tersebut menghasilkan prediksi *tweet* yang mengandung sentimen positif sebesar 54,88% dan prediksi *tweet* yang mengandung sentimen negatif sebesar 45,12%. Selanjutnya dilakukan visualisasi dari masing-masing kelas dalam bentuk *wordcloud*, visualisasi kata-kata yang paling sering muncul dari masing-masing kelas, dan visualisasi jumlah sentimen dari masing-masing kelas berdasarkan perangkat yang digunakan.

Dari grafik visualisasi didapat beberapa kata yang paling sering muncul dan cenderung memiliki sentimen positif diantaranya “25gb”, “paket”, “promo”, “beli”, “kuota”, “murah”. Hal ini bisa diasumsikan bahwa pengguna by.U yang memiliki sentimen positif merasa senang dengan salah satu paket yang ditawarkan yaitu paket kuota 25gb. Kemudian bisa diasumsikan juga bahwa pengguna by.U merasa senang dengan promo yang diadakan oleh provider by.U. Pengguna by.U merasa bahwa paket kuota yang ditawarkan cukup murah.

Dan beberapa kata yang paling sering muncul dan cenderung memiliki sentimen negatif diantaranya “sinyal”, “lelet”, “jelek”, “admin”, “jaring”. Dari hal tersebut bisa diasumsikan bahwa pengguna by.U merasa tidak puas atau tidak senang dengan masalah sinyal dan jaringan yang lambat dan buruk. Kemudian bisa diasumsikan juga bahwa pengguna by.U merasa kurang puas dengan layanan pelanggan. Hal tersebut bisa dilihat dari kata “admin” menjadi kata yang paling sering muncul dalam kelas sentimen negatif.

Dari visualisasi dapat dilihat beberapa topik di kelas sentimen positif. Topik tersebut meliputi *personality test* yang diadakan by.U di *platform* twitter, layanan unggulan by.U yang bisa mengatur sendiri jumlah dan jangka waktu kuota yang akan dibeli yang diberi nama “Yang Semuanya Semaunya”, kebijakan kuota *anniversary* yang diberikan by.U kepada penggunanya secara gratis ketika pemakaiannya sudah mencapai jangka waktu tertentu seperti dalam tiga bulan, enam bulan, satu tahun, dan lain-lain, dan promo kuota 25 GB dengan harga Rp 50.000 yang diadakan oleh by.U.

Dari visualisasi dapat dilihat beberapa topik di kelas sentimen negatif. Topik tersebut meliputi jeleknya sinyal by.U ketika digunakan untuk bermain *game*, ketidaksenangan pengguna by.U yang sudah terlanjur membeli kuota ketika by.U mengadakan promo kuota murah, dan masalah yang muncul ketika pembelian pulsa dan atau kuota.

Kemudian dari grafik visualisasi jumlah sentimen dari masing-masing kelas berdasarkan perangkat yang digunakan didapati bahwa sebagian besar pengguna by.U menggunakan perangkat genggam dengan mobilitas tinggi, dan sebagian besarnya memakai perangkat android. Dengan data ini bisa diasumsikan bahwa peningkatan layanan untuk ponsel android adalah hal yang patut dipertimbangkan mengingat proporsi pengguna android yang memakai by.U cenderung lebih banyak.

## 5.2 Saran

Pada penelitian ini, peneliti tentu tidak jauh dari kekurangan. Oleh karena itu, terdapat beberapa saran untuk dijadikan perbaikan untuk penelitian selanjutnya.

Saran-saran tersebut meliputi :

1. Pertimbangan untuk mengikutsertakan pakar terkait seperti pakar dalam bidang bahasa atau psikologi dalam proses pelabelan data sehingga resiko bias dapat ditekan dengan lebih maksimal.
2. Penerapan metode dan ekstraksi fitur lain yang lebih dieksplorasi sehingga dapat membandingkan performa antara satu metode dengan metode yang lain dan menemukan metode mana yang memiliki performa yang paling baik.
2. Penggunaan kamus *slangword* yang lebih lengkap sehingga pengolahan data menjadi lebih baik dan lebih terhindar dari *noise*.
3. Pengaplikasian model ke dalam sistem analisis sentimen sehingga dapat digunakan secara praktis.
4. Visualisasi data yang lebih dieksplorasi sehingga diharapkan dapat ditemukan *insight* menarik lebih banyak lagi.



## DAFTAR PUSTAKA

- Ali, W. P., & Sibaroni, Y. (2019). Analisis Sentimen Masyarakat Terhadap Kinerja Presiden Indonesia Dalam Aspek Ekonomi, Kesehatan, Pembangunan Berdasarkan Opini dari Twitter. *e-Proceeding of Engineering: Vol.6, No.2*, 8637-8649.
- Apriandi dkk. (2016). Analisis Sentimen Pelanggan Wifi.id pada Twitter dengan Support Vector Machine. *Jurnal Sistem Informasi Dan Bisnis Cerdas (SIBC) Vol. 9, No. 1*, 41-49.
- Asosiasi Penyelenggara Jasa Internet Indonesia. (2020, Juni). Laporan Survei Internet APJII 2019 - 2020 (Q2). Jakarta Selatan, DKI Jakarta, Indonesia.
- Badan Pusat Statistik. (2019). Statistik Pemuda Indonesia 2019. Jakarta, Indonesia.
- Bonzanini, M. (2016). *Mastering Social Media Mining with Python*. Birmingham: Packt Publishing Ltd.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal, Vol 2, No 1*, 32-41.
- Chatterjee, S., & Krystyanczuk, M. (2017). *Python Social Media Analytics*. Birmingham: Packt Publishing Ltd.
- Fitriyyah dkk. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi dan Penelitian Informatika), Vol. 5, No. 3*, 279-285.
- Hamidah, N. N. (2019). Analisis Sentimen Masyarakat terhadap Fintech Menggunakan Metode Naive Bayes Classifier.

- Han, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. San Fransisco: Morgan Kaufman Publisher.
- Hidayatullah, A. F. (2016). Pengaruh Stopword terhadap Performa Klasifikasi Tweet Berbahasa Indonesia. *JISKa, Vol. 1, No. 1*, 1-4.
- Hormansyah, D. S., & Aulia, I. (2018). Penerapan Metode TF-IDF dan N-Gram pada Pengembangan Aplikasi Chatbot Berbasis Line untuk Layanan Publik Kesehatan di Kota Malang. *Jurnal Informatika Polinema*, 7-11.
- Luqyana dkk. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 2, No. 11*, 4704-4713.
- Maning dkk. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge Univeristy Press.
- Nooraeni dkk. (2019). Analisis Sentimen Publik Terhadap Sistem Zonasi Sekolah Menggunakan Data Twitter dengan Metode Naive Bayes Classification. *Factor Exacta*, 315-322.
- Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram. *Jurnal Sains Komputer & Informatika (J-SAKTI) Volume (2) No.2*, 200-209.
- Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 200-209.

- Pudjajana, A. M., & Manongga, D. (2018). Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia di Twitter dengan Naive Bayes. *Jurnal SIMETRIS*, Vol. 9 No. 1, 313-318.
- RapidMiner GmbH. (2020, November 20). *Split Validation - RapidMiner Documentation*. Diambil kembali dari RapidMiner Documentation: [https://docs.rapidminer.com/latest/studio/operators/validation/split\\_validation.html](https://docs.rapidminer.com/latest/studio/operators/validation/split_validation.html)
- Robbani, H. A. (2020, November 20). *Sastrawi . PyPi*. Diambil kembali dari PyPi . The Python Package Index: <https://pypi.org/project/Sastrawi/>
- Wikipedia. (2020, November 20). *by.U - Wikipedia bahasa Indonesia, ensiklopedia bebas*. Diambil kembali dari Wikipedia bahasa Indonesia, ensiklopedia bebas: <https://id.wikipedia.org/wiki/By.U>
- Wikipedia. (2020, November 20). *Penyedia jasa Internet - Wikipedia bahasa Indonesia, ensiklopedia bebas*. Diambil kembali dari Wikipedia bahasa Indonesia, ensiklopedia bebas: [https://id.wikipedia.org/wiki/Penyedia\\_jasa\\_Internet](https://id.wikipedia.org/wiki/Penyedia_jasa_Internet)
- Wikipedia. (2020, November 20). *Twitter - Wikipedia bahasa Indonesia, ensiklopedia bebas*. Diambil kembali dari Wikipedia bahasa Indonesia, ensiklopedia bebas: <https://id.wikipedia.org/wiki/Twitter>
- Zuhri, F. N., & Alamsyah, A. (2017). Analisis Sentimen Masyarakat terhadap Brand Smartfren Menggunakan Naive Bayes Classifier di Forum Kaskus. *e-Proceeding of Management: Vol.4, No.1*, 242-251.