

SKRIPSI



disusun oleh

BUDI OCTAVIANDY

16.11.0283

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOUR DAN
NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN TWITTER
TERHADAP PENGGUNAAN PROVIDER BY.U**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana pada
Program Studi Informatika



disusun oleh

Budi Octaviandy

16.11.0283

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOUR DAN NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN TWITTER TERHADAP PENGGUNAAN PROVIDER BY.U

yang dipersiapkan dan disusun oleh

Budi Octaviandy

16.11.0283

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 2020

Dosen Pembimbing,

Anggit Dwi Hartanto, M.Kom.

NIK. 190302163

PENGESAHAN

SKRIPSI

PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOUR DAN NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN TWITTER TERHADAP PENGGUNAAN PROVIDER BY.U

yang dipersiapkan dan disusun oleh

Budi Octaviandy

16.11.0283

telah dipertahankan di depan Dewan Penguji
pada tanggal

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto M.Kom

NIK. 190302163

Wiwi Widayani, M.Kom

NIK. 190302227

Lilis Dwi Farida, S.Kom, M.Eng

NIK. 190302288

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 September 2020
Dekan Fakultas Ilmu Komputer

Krisnawati, S.Si, M.T.

NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 16 September 2020



Budi Octaviandy
NIM. 16.11.0283

MOTTO

“When life gets you down, do you wanna know what you’ve gotta do? *Just keep swimming.*” – Finding Nemo

“Orang – orang sukses di dunia bukanlah orang-orang yang berbakat, melainkan orang-orang yang berjuang tanpa kenal menyerah” – Fiersa Besari

“Sesungguhnya Bersama kesulitan itu ada kemudahan.” – QS: Alam Nasyrh [94]



PERSEMBAHAN

Saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

1. Kepada Allah SWT yang telah memberikan rahmat dan membimbing serta memudahkan saya.
2. Orang tua saya yang selalu menjadi tempat berpulang saya, selalu menjadi penyemangat saya, serta memberika segalanya kepada saya.
3. Bapak Anggit Dwi Hartanto, M.Kom yang telah membimbing saya dari awal sampai akhir pembuatan skripsi saya.
4. Kepada dosen – dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
5. Kepada teman – teman kelas IF-05 2016 yang menemani saya selama menempuh Pendidikan disini.
6. Kepada teman – teman kepengurusan AEC 2017, maupun IEEE serta kepanitiaan yang sudah menjadi keluarga melepas lelah setelah melakukan kegiatan perkuliahan. Memberikan banyak pengalaman, canda tawa, kesedihan serta kegembiraan di waktu tersebut.
7. Kepada orang – orang yang rela mendengar keluh kesah saya selama perkuliahan maupun selama pengerjaan skripsi ini, terima kasih telah menyediakan telinga nya untuk mendengarkan.

Kata Pengantar

Puji dan syukur saya panjatkan kepada Allah SWT yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi saya yang berjudul Perbandingan Algoritma K-Nearest Neighbors dan Naïve Bayes Classifier Dalam Analisis Sentimen Twitter Terhadap Penggunaan Provider By.U. Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata satu dan untuk memperoleh gelas Sarjana Komputer. Dengan selesainya skripsi ini, maka pada kesempatan ini saya mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta
3. Bapak Anggit Dwi Hartanto, M.Kom selaku dosen pembimbing yang selalu bijaksana dalam memberikan bimbingan, nasehat serta waktunya selama pembuatan skripsi ini.
4. Kedua orang tua saya yang telah mendoakan, mendukung dan memberikan semangat kepada saya.
5. Manusia - manusia baik yang senantiasa mendoakan, memberi semangat dan bantuan kepada saya.
6. Teman – teman saya di organisasi maupun diluar kampus yang memberi semangat dan rela mendengar keluh kesah saya.
7. Teman – teman IF 05 2016 yang telah menemani selama proses perkuliahan.
8. Teman – teman kontrakan yang selalu memberikan semangat dan masukan dalam pembuatan skripsi.
9. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Allah SWT memberikan balasan yang lebih kepada semua yang telah ikut membantu saya hingga menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik membangun diterima dengan senang hati dan terima kasih. Semoga skripsi ini dapat bermanfaat bagi saya dan kita semua.

Yogyakarta, 16 September 2020

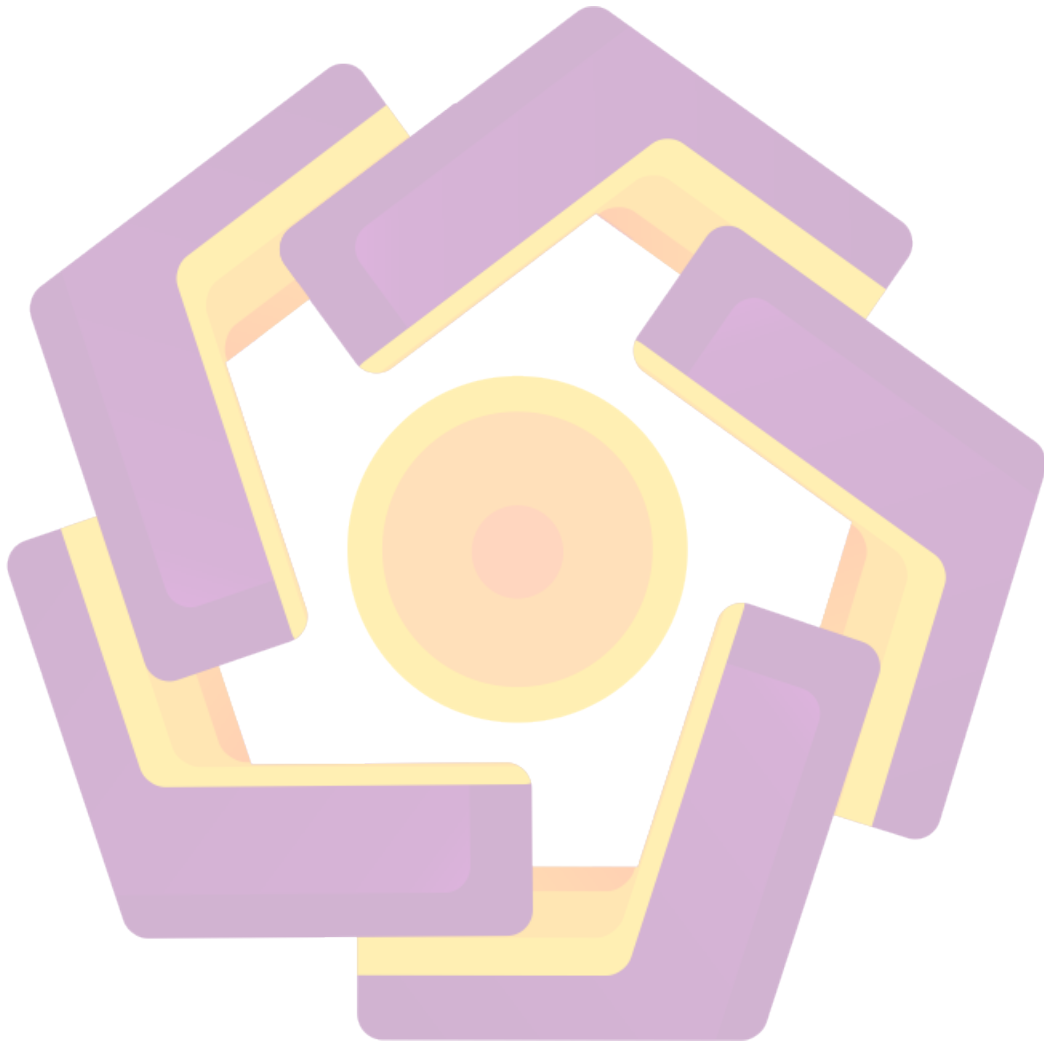
Budi Octaviandy

DAFTAR ISI

JUDUL.....	I
PERSETUJUAN.....	II
PENGESAHAN.....	III
PERNYATAAN.....	IV
MOTTO.....	VI
KATA PENGANTAR.....	VIII
DAFTAR ISI.....	IX
INTISARI.....	XIV
ABSTRACT.....	XV
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH.....	4
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	4
1.5 MANFAAT PENELITIAN.....	2
1.6 METODE PENELITIAN.....	2
1.6.1 Metode Pengumpulan Data.....	2
1.6.2 Metode Analisis.....	3
1.7 SISTEMATIKA PENULISAN.....	4
BAB II LANDASAN TEORI.....	6
2.1 KAJIAN PUSTAKA.....	6
2.2 DATA MINING.....	9
2.3 SENTIMEN ANALYSIS.....	10
2.4 TEXT MINING.....	12
2.5 PRE-PROCESSING.....	12
2.6 TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY (TF-IDF).....	13
2.7 K-NEAREST NEIGHBOR.....	15
2.8 NAÏVE BAYES CLASSIFIER.....	16
2.9 CONFUSION MATRIX.....	18

BAB III METODOLOGI PENELITIAN	20
3.1 ALAT DAN BAHAN PENELITIAN.....	20
3.1.1 <i>Alat Penelitian.....</i>	20
3.1.2 <i>Bahan Penelitian.....</i>	21
3.2 ALUR PENELITIAN.....	21
3.2.1 <i>Pengumpulan Data.....</i>	21
3.2.2 <i>Pre-Processing Data.....</i>	22
3.2.3 <i>TF-IDF.....</i>	30
3.3 K-NEAREST NEIGHBOUR (KNN)	38
3.4 NAÏVE BAYES CLASSIFIER	47
3.5 EVALUASI.....	59
BAB IV HASIL DAN PEMBAHASAN.....	60
4.1 DESKRIPSI IMPLEMENTASI	60
4.2 IMPLEMENTASI PENGUMPULAN DATA TRAINING.....	60
4.3 IMPLEMENTASI PELABELAN DATA TRAINING	61
4.4 IMPLEMENTASI PERANCANGAN FUNGSI DAN SISTEM.....	62
4.4.1 <i>Mengakses Dataset Training.....</i>	62
4.4.2 <i>Implementasi Fungsi Preprocessing dan Stemming Data Training.....</i>	62
4.4.3 <i>Implementasi TF-IDF.....</i>	63
4.4.4 <i>Implementasi Fungsi Klasifikasi K-Nearest Neighbors.....</i>	63
4.4.5 <i>Implementasi Fungsi Klasifikasi Naive Bayes Classifier</i>	64
4.4.6 <i>Mengakses Data Testing</i>	64
4.4.7 <i>Implementasi fungsi Cleaning data testing</i>	65
4.4.8 <i>Implementasi Fungsi Tokenize</i>	66
4.4.9 <i>Implementasi fungsi Stemming data testing</i>	66
4.4.10 <i>Implementasi fungsi prediksi sentiment.....</i>	66
4.5 PEMBUATAN ANTARMUKA.....	67
4.5.1 <i>Halaman Utama</i>	67
4.5.2 <i>Halaman Testing Algoritma KNN</i>	68
4.5.3 <i>Halaman Testing Algoritma NBC</i>	68
4.5.4 <i>Halaman Hasil Klasifikasi KNN</i>	69
4.5.5 <i>Halaman Hasil Klasifikasi NBC.....</i>	70
4.6 EVALUASI PROGRAM.....	70
BAB V PENUTUP	74

5.1	KESIMPULAN	74
5.2	SARAN	74
	DAFTAR PUSTAKA	76



DAFTAR TABEL

TABEL 2.1 PERBANDINGAN PENELITIAN	7
TABEL 2.2 CONFUSION MATRIX	18
TABEL 3.1 DATA TWEET DARI HASIL CRAWLING	22
TABEL 3.2 DATA TWEET PEMBERIAN LABEL	23
TABEL 3.3 DATA TWEET SEBELUM DILAKUKAN CLEANING	24
TABEL 3.4 DATA SETELAH DILAKUKAN CLEANING	25
TABEL 3.5 CASE FOLDING	26
TABEL 3.6 TOKENIZING	27
TABEL 3.7 STEMMING	28
TABEL 3.8 FEATURE LIST	29
TABEL 3.9 DATA TWEET	31
TABEL 3.10 TERM KATA	32
TABEL 3.11 PERHITUNGAN IDF	34
TABEL 3.12 PERHITUNGAN TF-IDF	36
TABEL 3.13 DATA TWEET	39
TABEL 3.14 DATA SETELAH PREPROCESSING	40
TABEL 3.15 TF-IDF	41
TABEL 3.16 PANJANG VECTOR	45
TABEL 3.17 KEMIRIPAN HASIL	46
TABEL 3.18 HASIL KEMIRIPAN TERDEKAT	46
TABEL 3.19 DATA TWEET	48
TABEL 3.20 DATA TWEET SETELAH PREPROCESS	49
TABEL 3.21 SELURUH TERM DATA	50
TABEL 3.22 TERM FREQUENCY POSITIF	52
TABEL 3.23 TERM FREQUENCY NEGATIF	53
TABEL 3.24 VMAP DATA UJI	58
TABEL 4.1 CONFUSION MATRIX	70

DAFTAR GAMBAR

GAMBAR 3.1 FLOWCHART PENELITIAN	21
GAMBAR 3.2 PREPROCESSING	23
GAMBAR 3.3 FLOWCHART K-NN	39
GAMBAR 3.4 NAÏVE BAYES FLOWCHART	47
GAMBAR 4.1 SCRIPT SCRAPING TWITTERSCRAPE	60
GAMBAR 4.2 FILE DATA TRAINING	62
GAMBAR 4.3 SCRIPT AKSES DATA TRAINING	62
GAMBAR 4.4 SCRIPT PREPROCESSING DATA TES	63
GAMBAR 4.5 SCRIPT PREPROCESSING DATA TES	63
GAMBAR 4.6 SCRIPT IMPLEMENTASI TF-IDF	63
GAMBAR 4.7 SCRIPT KLASIFIKASI KNN	64
GAMBAR 4.8 SCRIPT KLASIFIKASI NBC	64
GAMBAR 4.9 SCRIPT CRAWLING DATA TESTING	65
GAMBAR 4.10 SCRIPT PREPROCESSING DATA TESTING	66
GAMBAR 4.11 SCRIPT TOKENIZING	66
GAMBAR 4.12 SCRIPT STEMMING	66
GAMBAR 4.13 SCRIPT PREDIKSI	67
GAMBAR 4.14 TAMPILAN HALAMAN UTAMA	68
GAMBAR 4.15 TAMPILAN HALAMAN TESTING KNN	68
GAMBAR 4.16 TAMPILAN HALAMAN TESTING NBC	69
GAMBAR 4.17 TAMPILAN HASIL KNN	69
GAMBAR 4.18 TAMPILAN HASIL NBC	70
GAMBAR 4.18 TAMPILAN HASIL KLASIFIKASI KNN	71
Gambar 4.18 Tampilan Hasil Klasifikasi NBC.....	70

INTISARI

Perkembangan internet di Indonesia cukup pesat, hal ini dibuktikan dengan semakin banyaknya provider yang memberikan beragam paket internet kepada pengguna. Hal ini dimanfaatkan oleh provider Telkomsel dimana provider tersebut belakangan mengeluarkan layanan seluler prabayar digital pertama mereka yang bernama By.U. Dengan membawa nama Telkomsel yang besar, provider By.U yang baru saja diluncurkan ini tentu saja akan dibandingkan dengan provider yang telah dikeluarkan sebelumnya oleh Telkomsel.

Banyaknya opini terhadap provider By.U ini tentu saja dapat menjadi evaluasi terhadap pihak By.U. Serta untuk calon konsumen juga dapat digunakan sebagai pertimbangan apakah provider ini cocok untuk digunakan atau tidak. Oleh karena itu, penelitian ini mencoba melakukan analisis sentiment mengenai penggunaan provider By.U. Banyaknya opini yang dituliskan di twitter akan dijadikan sebagai dataset yang berjumlah 1.000 data. Dataset tersebut akan diklasifikasikan apakah dataset tersebut cenderung beropini positif atau negative. Algoritma yang digunakan pada penelitian ini adalah Algoritma K-Nearest Neighbour dan Naïve Bayes Classifier.

Berdasarkan hasil eksperimen menggunakan Algoritma K-Nearest Neighbor tingkat akurasi yang dihasilkan mencapai 75% dan untuk algoritma Naïve Bayes Classifier menghasilkan tingkat akurasi 79%. Dari hasil penelitian tersebut dengan menggunakan objek by.u dan dataset training berjumlah 1.000 memperoleh hasil bahwa algoritma Naïve Bayes Classifier lebih tinggi tingkat akurasi nya dibandingkan dengan K-Nearest Neighbor.

Kata Kunci: KNN, NBC, Analisis Sentimen , Twitter

ABSTRACT

The development of the internet in Indonesia is quite fast, this is evidenced by the increasing number of providers who provide various internet packages to users. This is utilized by the Telkomsel provider where the provider recently issued their first digital prepaid cellular service called By.U. By carrying the big name Telkomsel, the provider By.U which has just been launched will of course be compared to the providers that have been issued previously by Telkomsel.

The number of opinions on By.U provider can of course be an evaluation of By.U. As well as for potential customers, it can also be used as a consideration whether this provider is suitable for use or not. Therefore, this study tries to analyze sentiment regarding the use of provider By.U. The number of opinions written on twitter will be used as a dataset of 1,000 data. The dataset will be classified whether the dataset tends to be positive or negative. The algorithm used in this study is the K-Nearest Neighbor Algorithm and the Naïve Bayes Classifier.

Based on the experimental results using the K-Nearest Neighbor algorithm, the resulting accuracy reaches 75% and the Naïve Bayes Classifier algorithm produces an accuracy rate of 79%. From the results of this study using the object by.u and a training dataset of 1,000, the results show that the Naïve Bayes Classifier algorithm has a higher level of accuracy than the K-Nearest Neighbor.

Keywords: *KNN, NBC, Sentiment Analysis, Twitter*