

Implementasi Algoritma *BERT* Pada Komentar Layanan Akademik dan Non Akademik Universitas Terbuka di Media Sosial

Rhini Fatmasari^{a,1}, Riska Kurnia Septiani^{b,2}, Tuahta Hasiolan Pinem^{b,3}, Dedik Fabiyanto^{b,4}, Windu Gata^{b,5}

^a Fakultas Keguruan dan Ilmu Pendidikan, Universitas Terbuka

^b Fakultas Teknologi Informasi, Universitas Nusa Mandiri

¹ riens@ecampus.ut.ac.id; ² 14220021@nusamandiri.ac.id; ³ 14220020@nusamandiri.ac.id;

⁴ 14220030@nusamandiri.ac.id; ⁵ windu@nusamandiri.ac.id

INFORMASI ARTIKEL

Histori Artikel

Diterima : 03 Agustus 2023
Direvisi : 15 Agustus 2023
Diterbitkan : 30 Agustus 2023

Kata Kunci:

Universitas Terbuka
Twitter (X)
TikTok
Analisa
BERT

ABSTRAK

Media sosial TikTok dan Twitter (X) merupakan dua media sosial yang memiliki banyak pengguna di Indonesia. Berdasarkan lembaga survei *We Are Social* pengguna TikTok di Indonesia mencapai 109.9 juta sedangkan pengguna Twitter (X) di Indonesia mencapai 24 juta. Media sosial TikTok dan Twitter (X) seringkali menjadi tempat untuk menyampaikan pendapat atau komentar terhadap suatu hal. Universitas Terbuka merupakan suatu kampus yang memiliki media sosial TikTok dan Twitter (X) dengan ribuan pengikut. Penelitian ini dilakukan untuk mengetahui tingkat layanan bidang akademik dan non-akademik Universitas Terbuka. Data yang dianalisis sebanyak 685 data komentar pada media sosial TikTok dan Twitter (X) dengan kata kunci Universitas Terbuka. Metode yang digunakan adalah analisis menggunakan model *pre-trained BERT*. Pada model ini diperoleh nilai akurasi sebesar 90% dengan proporsi data latih dan data uji 80:20.

2023 SAKTI – Sains, Aplikasi, Komputasi dan Teknologi Informasi.

Hak Cipta.

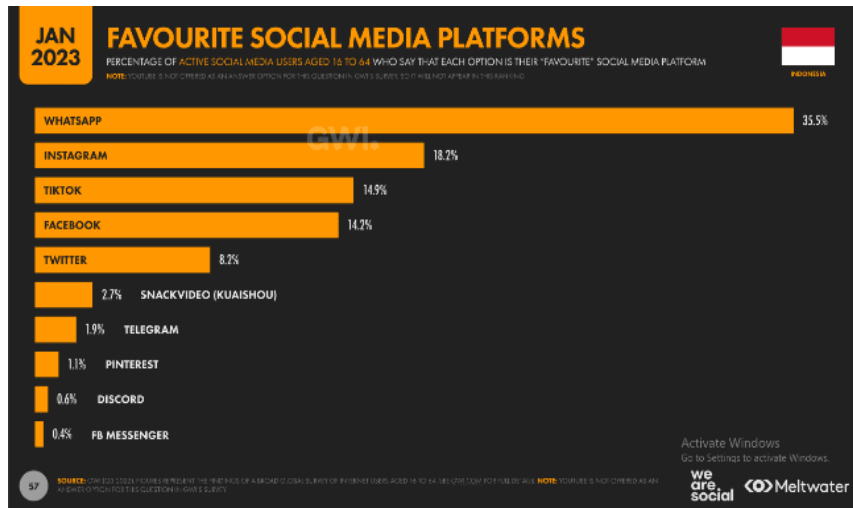
I. Pendahuluan

Universitas Terbuka merupakan salah satu perguruan tinggi yang melaksanakan proses belajar mengajar. Sebagai Perguruan Tinggi Negeri, Universitas Terbuka memiliki visi pendidikan tinggi terkemuka pada tahun 2021 yang memiliki kualitas daya saing bertaraf internasional. Untuk mewujudkan hal tersebut, Universitas Terbuka harus menjaga kualitas pembelajaran serta kualitas para mahasiswa agar bisa bersaing dengan lulusan perguruan tinggi lain[1]. Salah satu lembaga pemeringkatan universitas terkemuka yang digunakan dan menjadi acuan oleh 31.000 universitas dari lebih 200 negara di seluruh dunia adalah Webometrics. Universitas Terbuka berhasil meningkatkan peringkat Webometrics perguruan tinggi di Indonesia dan dunia. Pada rilis hasil pemeringkatan Webometrics bulan Januari 2023 Universitas Terbuka menempati peringkat 68 (tingkat Indonesia) dan 4.931 (tingkat dunia) sedangkan berdasarkan rilis hasil pemeringkatan Webometrics bulan Juli 2023 Universitas Terbuka menempati peringkat 60 (tingkat Indonesia) dan 4.259 (tingkat dunia).

Ranking	World Ranking	University	Del.	Country	Impact	Openness	Excellence
60	4259	Universitas Terbuka	UN	Indonesia	1609	2345	6291

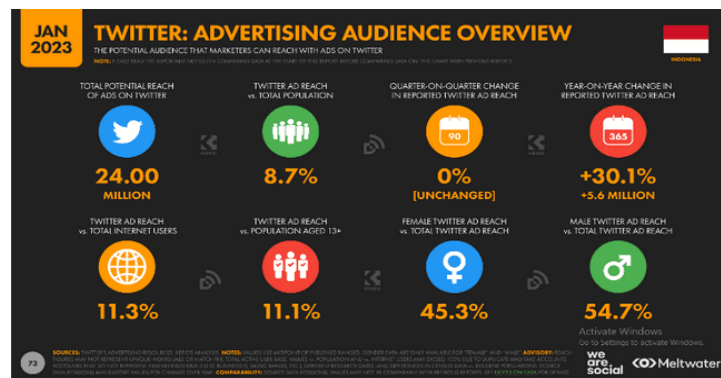
Gambar 1. Ranking Universitas Terbuka tingkat Indonesia

Menurut laporan “We Are Social”, jumlah pengguna aktif media sosial di Indonesia pada Januari 2023 mencapai 167 juta orang, setara 60,4% dari populasi di Indonesia. Jumlah pengguna media sosial Twitter (X) mencapai 24 juta yang menempati peringkat ke 5 dan pengguna media sosial TikTok mencapai 109,9 juta yang menempati peringkat ke 3.



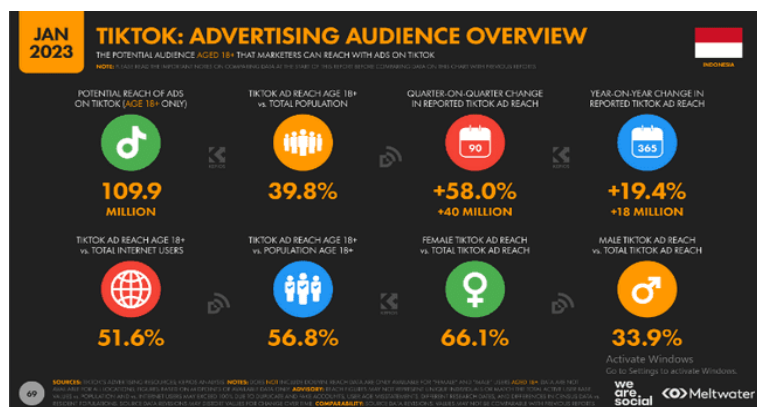
Gambar 2. Favourite Social Media Platforms di Indonesia (data We Are Social)

Twitter (X) merupakan sebuah layanan microblogging yang dalam bahasa Inggris artinya adalah kicauan. Twitter (X) membatasi para pengguna untuk menulis sebuah ide ataupun gagasan sebanyak 280 karakter. Twitter dapat memperlihatkan Trending Topic yang memudahkan penggunaannya untuk melihat suatu topik baik dalam negeri maupun luar negeri yang sedang ramai dibicarakan[2]. Berdasarkan We Are Social, pengguna Twitter di Indonesia mencapai 24 juta dimana dengan jenis kelamin perempuan sebanyak 45.3% dan laki-laki sebanyak 54.7%.



Gambar 3. Ringkasan Pengguna Twitter Tahun 2023 di Indonesia (data We Are Social)

TikTok merupakan salah satu media sosial yang sering digunakan oleh masyarakat Indonesia. TikTok merupakan aplikasi primadona, menarik, dan digandrungi oleh para milenial. TikTok diperkenalkan pada tahun 2016 tepatnya di Bulan September. TikTok dipublikasikan oleh ByteDance yaitu salah satu perusahaan asal Tiongkok. Pada mulanya aplikasi ini bernama Douyin[3]. Berdasarkan We Are Social, pengguna TikTok di Indonesia mencapai 109,9 juta dimana dengan jenis kelamin perempuan sebanyak 66,1% dan laki-laki sebanyak 33,9%.



Gambar 4. Ringkasan Pengguna TikTok Tahun 2023 di Indonesia (data We Are Social)



Gambar 5. Twitter (X) Universitas Terbuka (Data Twitter, November 2023)

Universitas Terbuka memiliki akun resmi pada media sosial Twitter dengan username @UnivTerbuka dengan followers 812,9 ribu pengguna. Universitas Terbuka juga memiliki beberapa akun resmi berdasarkan kota untuk media sosial TikTok pada penelitian ini menggunakan akun dengan id @utpalembang, @utsemarang, @utpurwokerto, @utpadang, @ut_kupang, @utpekanbaru.

Berdasarkan latar belakang di atas, penelitian ini fokus membahas mengenai apakah kualitas pelayanan akademik dan non akademik berpengaruh terhadap kepuasan masyarakat serta secara simultan berpengaruh terhadap kepuasan masyarakat? Untuk itu tujuan dari penelitian ini adalah untuk menguji dan menganalisis pengaruh kualitas pelayanan akademik dan non akademik terhadap kepuasan masyarakat serta menguji dan menganalisis secara simultan pengaruh kualitas pelayanan akademik dan non akademik terhadap kepuasan masyarakat.

Layanan akademik sendiri memiliki peranan yang besar dalam bidang pendidikan. Layanan akademik yang dimaksud yaitu meliputi peraturan akademik, perkuliahan, kurikulum, bimbingan akademik, praktikum, tugas akhir, evaluasi termasuk bantu perkuliahan seperti perpustakaan, Laboratorium dan lain-lain. Berdasarkan hal tersebut, layanan akademik merupakan layanan yang banyak berkaitan secara langsung dengan mahasiswa dan masyarakat[4].

Sedangkan layanan non akademik adalah layanan yang berorientasi kepada layanan yang mendukung terciptanya atmosfer akademik. Layanan non akademik yang dimaksud antara lain layanan di Unit Program Belajar Jarak Jauh (UPBJJ), registrasi, pendaftaran, wisuda, ujian ulang, cetak kartu mahasiswa, lama pelayanan, keramahan petugas, informasi tentang kuliah di UT, contact centre, dan lain-lain. Indikator tersampainya layanan akademik dan non akademik kepada masyarakat adalah melalui kepuasan masyarakat terhadap pelayanan tersebut yaitu dengan pelayanan prima.

Pelayanan prima mengandung tiga hal pokok, yaitu: adanya pendekatan sikap yang berkaitan dengan kepedulian dalam melayani dengan tindakan yang baik adanya tujuan untuk memuaskan pelanggan dengan berorientasi pada standar layanan. Konsep dari pelayanan prima adalah melakukan pelayanan sebaik mungkin kepada setiap pelanggan yang membutuhkan, sehingga dapat menimbulkan kepuasan, melebihi apa yang diharapkan dalam diri[5]. Menurut Barata (2003) pelayanan prima adalah pelayan terbaik yang diberikan kepada konsumen yang berorientasi kepada pelanggan sehingga menimbulkan kepuasan.

A. Algoritma SVM

Support vector machines (SVM) di perkenalkan pertama kali oleh Vapnik pada tahun 1992 sebagai konsep unggulan dalam bidang pattern recognition, algoritma ini dapat memilih model otomatis dan tidak memiliki masalah overfitting.

B. Algoritma Naïve Bayes

Naïve Bayes adalah sebuah teknik klasifikasi menggunakan probabilitas yang simple berakar pada Teorema Bayes serta mempunyai anggapan ketidak ketergantungan (independent) yang tinggi dari masing - masing kondisi atau kejadian[6].

C. Algoritma KNN

Algoritma KNN merupakan algoritma berbasis memori yang menggunakan iterasi pada data sampai atribut atau parameter data yang terdekat ditemukan. Jarak minimum data yang diproses pada data testing akan dibandingkan dengan data training dengan jarak yang terdekat.

D. Algoritma Decision Tree

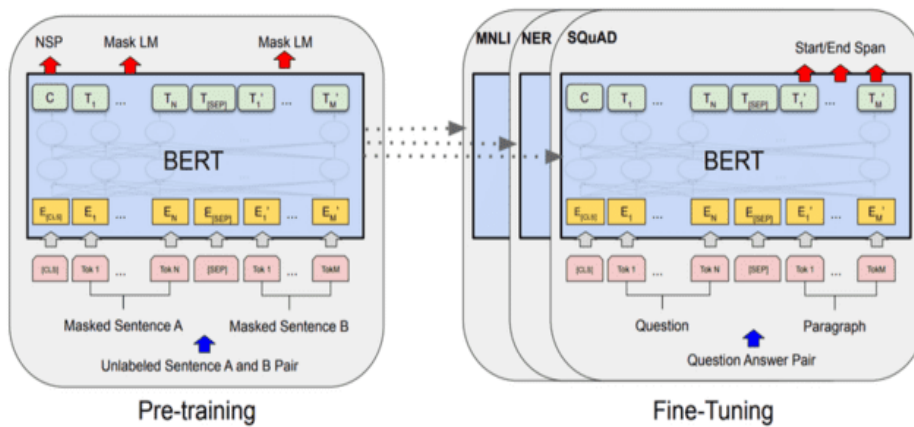
Pohon (tree) adalah sebuah struktur data yang terdiri dari simpul (node) dan rusuk (edge). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (root/node), simpul percabangan/internal (branch/internal node) dan simpul daun (leaf node).

E. Algoritma Logistic Regretion

Regresi logistik (Logistic Regression) adalah bagian dari analisis regresi yang digunakan ketika variabel dependen (respon) merupakan variabel dikotomi. Variabel dikotomi biasanya hanya terdiri atas dua nilai yang mewakili kemunculan atau tidak adanya suatu kejadian yang biasanya diberi angka 0 atau 1.

F. BERT Pre-Trained Model

Menurut Devlin, BERT merupakan singkatan dari *Bidirectional Encoder Representation from Transformer* yang didesain untuk melakukan pre-training dengan representasi *deep bidirectional* dari teks tidak berlabel[7]. Encoder pada BERT membaca semua urutan dari teks yang masuk secara menyeluruh. Model ini dibuat untuk dapat mengerti makna yang terkandung sebelum dan sesudah. BERT dapat meningkatkan performa pada berbagai kegiatan pada bidang NLP seperti *Natural Language Inference* (NLI) dan *Question Answering* (QA) karena kemampuan BERT yang dapat memahami makna[8]. Secara lengkap BERT Pre-Trained Model dijelaskan pada Gambar 6.



Gambar 6. Ilustrasi pre-training dan fine-tuning BERT

Metrik performa yang digunakan adalah *F1 score*, *Precision*, *Recall*, dan *Accuracy*. *F1-Score* secara definisi adalah *harmonic mean* dari *precision* dan *recall*, berikut metrik yang menggabungkan presisi dan perolehan.

$$F1\ Score = 2 \times \frac{recall \times precision}{recall + precision} \tag{1}$$

Precision adalah rasio observasi positif yang diprediksi dengan benar terhadap total observasi positif yang diprediksi. Berikut rumus untuk *precision* :

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$Precision = \frac{True\ Positive}{Total\ Predicted\ Positive} \tag{3}$$

Recall diartikan sebagai rasio hasil yang diprediksi dengan benar terhadap semua prediksi. Ini juga dikenal sebagai sensitivitas atau spesifisitas. Perumusan *Recall* dituliskan pada persamaan 4 dan 5.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{4}$$

$$Recall = \frac{True\ Positive}{Total\ Actual\ Positive} \tag{5}$$

Selanjutnya untuk melihat hasil yang baik dilihat dari tingkat keakuratannya. Akurasi adalah perbandingan prediksi yang benar dari seluruh prediksi yang dibuat oleh suatu algoritma. Hal ini dapat dihitung dengan membagi presisi dengan perolehan atau dengan 1 dikurangi tingkat negatif palsu (FNR) dibagi dengan tingkat positif palsu (FPR). Untuk menghitung akurasi, dapat dilihat pada persamaan 6.

$$Accuracy = \frac{(TP+TN)}{(TP+FN+TN+FP)} \quad (6)$$

G. Confusion Matrix

Variabel TN, TP, FN, dan FP pada Confusion Matrix merupakan pengukuran performa untuk klasifikasi machine learning. Hasilnya berupa dua kelas atau lebih. Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual seperti pada tabel 1 berikut.

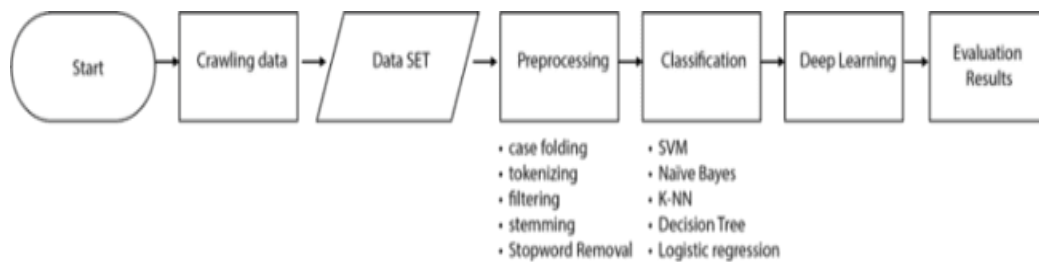
Tabel 1. Tabel Confusion Matrix

	Prediction Yes	Prediction No
True Yes	TP	FN
True No	FP	TN

Informasi yang dikumpulkan diisolasi menjadi informasi penyiapan dan pengujian informasi. Berbagi informasi dilakukan dengan menggunakan teknik persetujuan silang N-lipatan untuk menghilangkan kecenderungan kata. N-peningkatan persetujuan silang mempartisi laporan menjadi n bagian[9].

II. Metode Penelitian

Metode penelitian yang dalam penelitian ini adalah model yang bernama BERT yang dikembangkan oleh Google AI merupakan salah satu model yang memanfaatkan arsitektur Transformer.



Gambar 7. Tahapan Analisis Sentimen

Gambar 7 merupakan tahapan penelitian untuk melakukan Analisis Sentimen Komentar Layanan Akademik dan Non Akademik Universitas Terbuka. Penjelasan dari Gambar 7 adalah sebagai berikut:

A. Crawling Data

Crawling data di Twitter adalah suatu proses untuk mengambil atau mengunduh data dari server Twitter[10]. Proses crawling data dengan bantuan Application Programming Interface (API) Twitter baik berupa data user maupun data tweet menggunakan program python dan API Twitter(X) untuk melakukan scraping komentar, mengekstrak data informasi, memindainya, dan menyimpan dalam bentuk XLSX, CSV dan menggunakan software web base scraping yaitu apify. Web scraping adalah teknik untuk mendapatkan informasi dari website secara otomatis tanpa harus melakukan penyalinan secara manual. Tujuan dari web scraper adalah untuk mencari informasi tertentu dan kemudian mengumpulkannya dalam web yang baru. Web scraping berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi[11].

B. Analisa atau Labeling

Analisa atau Labeling adalah proses untuk menentukan kelas pada sebuah himpunan data atau dataset. Himpunan data atau dataset akan dianotasi menjadi kelas sentimen positif dan sentimen negatif melalui tiga cara, yaitu metode VADER, Text Blob, dan manual.

C. Preprocessing

Data preprocessing adalah teknik yang digunakan untuk mengubah data yang melibatkan validasi dan imputasi data, untuk menilai tingkat kelengkapan dan akurasi data untuk digunakan dalam proses analisis selanjutnya. Tahapan preprocessing yang dilakukan meliputi *cleansing*, *case folding*, *tokenizing*, *filtering*, *stemming* dan *Stopword Removal*. Berikut untuk penjelasan dari tahapan preprocessing:

- *Cleansing*

Proses membersihkan data dari kata – kata yang tidak menjelaskan apapun. Pada tahapan cleansing meliputi, menghapus emot, menghapus url/link, menghapus mentions, menghapus hashtags, dan menghapus spasi berlebih.

- *Case Folding*

Mengubah semua kata dalam teks isi pantun menjadi berhuruf kecil dan karakter lain selain huruf dihilangkan.

- *Tokenizing*

Memecah kalimat menjadi token-token atau kata-kata tunggal. Pemecahan ini dilakukan berdasarkan tanda spasi.

- *Stemming*

Mengembalikan kata ke dalam kata dasarnya atau kata baku (Irmanda & Ria Astriratma, 2020).

- *Stopword Removal*

Menghilangkan kata-kata yang dianggap tidak memiliki makna.

D. Klasifikasi

Tahapan selanjutnya adalah analisis sentimen. Metode yang dapat digunakan untuk menganalisis sentimen dari media social Twitter(X) dan TikTok yang berupa data komentar dengan membandingkan metode klasifikasi statistik berupa SVM, Naïve Bayes, K-NN, Decision Tree, dan Logistic Regression[12].

E. Model Deep Learning

Deep learning adalah metode dalam kecerdasan buatan (AI) mengajarkan komputer untuk memproses data dengan cara yang terinspirasi otak manusia. Algoritma yang memiliki fitur yang memungkinkan untuk secara otomatis menangkap fitur yang relevan. Secara teoritis, model digunakan dalam deep learning memiliki fungsi matematika dengan $f: X \rightarrow Y$. *Deep learning* merupakan pengembangan mendalam *Artificial Neural Network* (ANN) yang menggunakan lebih dari satu lapisan tersembunyi dalam memodelkan dataset yang diberikan .

F. Modelling (Pemodelan)

Tahap ketiga pemodelan (Modelling) yaitu data yang telah diperiapkan, dibersihkan dan diberi label akan dimasukkan ke dalam model. Pada penelitian ini model yang digunakan adalah *Pre-Trained BERT Classificatoin*[13]. Tahap pemodelan ini adalah sebagai berikut:

- *Set-up BERT Pre-Trained Model*

Pada tahap ini dilakukan instalasi BERT pre- trained model sebelum dapat digunakan untuk keperluan sentimen analisis.

- *Input Data*

Setelah model di input, proses selanjutnya yaitu memasukkan data ke dalam model yang sudah di encode menjadi token pada proses sebelumnya.

- *Data Loaders*

Di dalam proses sentimen analisis yang termasuk kategori deep learning, data loaders menjadi salah satu komponen penting faktor kesuksesan. Data loaders membuat proses pemuatan data lebih efisien untuk skala data ribuan hingga jutaan dikarenakan algoritma yang dimiliki dapat memuat data secara terstruktur dan sistematis dan meminimalisir terjadinya data loss.

- *Set-up Optimiser and Scheduler*

Dalam melakukan komputasi data, pada umumnya menggunakan gradient descent untuk melakukan proses penjumlahan dan turunan dengan tujuan menemukan titik yang meminimumkan suatu fungsi yang dapat diturunkan.

- *Mendefinisikan Performa Metrik*

Pada penelitian ini menggunakan Scikit-Learn sebagai *software non-berbayar machine learning library* untuk pemrograman python. Pada tahap ini mendefinisikan metrik performa yang menjadi tolok ukur hasil perhitungan komputasi data sentimen analisis. Hal ini meliputi *F1 score, Precision, Recall, Accuracy*.

III. Hasil dan Pembahasan

Perancangan sistem analisis sentimen akan dibangun sesuai dengan hasil penelitian analisis.

A. Crawling Data

Data komentar ini diambil dari media sosial Twitter(X) dan TikTok. Pada media sosial Twitter(X) data diambil dengan menggunakan python pada google collab crawling data Twitter(X) berdasarkan kata kunci "Universitas Terbuka" menggunakan tweet-harvest. Selanjutnya, pada media sosial TikTok data komentar diambil dari TikTok resmi Universitas terbuka (@utpalembang, @utsemarang, @utpurwokerto, @utpadang, @ut_kupang, @utpekanbaru), data ini diambil dengan menggunakan software WebHarvy. Data proses tersebut diperoleh data komentar sebanyak 685 data.

Tabel 2. Hasil Crawling Data Komentar

No	Username	Komentar
1	Nanaalza	Open joki tugas buat mahasiswa/i Universitas Terbuka. diskusi : 30k-50k (sesuai tingkat kesulitannya) tugas matkul : 65-85k (sesuai tingkat kesulitannya) TMK/matkul : 35k pilgan : 15k order? wa in bio #universitasterbuka #UT #jokitugas #zonauang #zonajajan #jokimurah
2	UnivTerbuka	Sahabat UT, yuk ikuti Klinik Literasi Perpustakaan UT LIB-TALK" bersama Irmayati, S.IP. pada hari Rabu, 15 November 2023 pukul 13.30 WIB via Zoom meeting pada link berikut: https://t.co/L0esT3UNMx #Literasi #Perpustakaan #UniversitasTerbuka https://t.co/QG51zOqgcO "
3	utpalembang	Direktur dan Seluruh Sivitas Akademika Universitas Terbuka Palembang mengucapkan Selamat Memperingati Hari Pahlawan Nasional - 10 November 2023 - Semangat Pahlawan untuk Masa Depan Bangsa dalam Memerangi Kemiskinan dan Kebodohan." #universitasterbuka #utpalembang #infout https://t.co/17kWF59b9B "
...
684	jokitugas_bysya	Eka desy indah fitriani nomor induk mahasiswa 043839784 nomor hp 081217530649 Universitas Terbuka prodi manajemen penipu joki tugas
685	ziggyyyz	@thisisgreeny tauukk iih tp kaa, rerata yg bilang alih kredit tu lama dee prosesnya.. ada yg bilang sampe 3 bulanan ada grupnya di tele lebih aptudet ketik aja universitas terbuka.. aku di kick adminnya gegara bikin promo joki tugas disitu wkwkwkwkw

B. Analisa atau Labeling

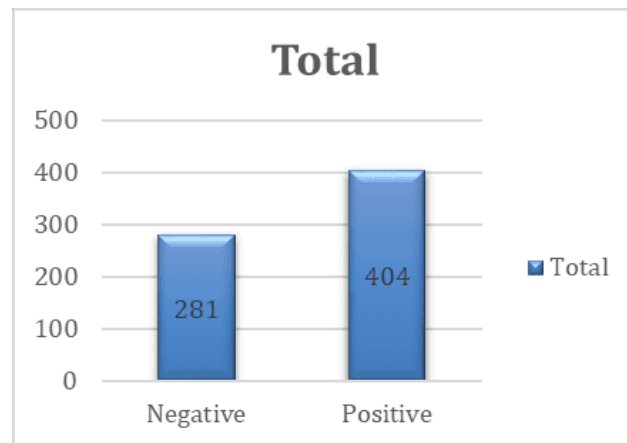
Data harus dilabelkan dan dibagi menjadi 2 kelas sentiment yaitu kelas sentiment positif dan kelas sentiment negatif berdasarkan kata – kata yang terdapat pada data komentar. melalui tiga cara, yaitu metode VADER, Text Blob, dan Manual. Contoh data yang sudah dilabelkan dapat dilihat pada tabel 3.

Tabel 3. Labeling Data Komentar

Komentar	VADER	TextBlob	Manual
Open joki tugas buat mahasiswa/i Universitas Terbuka. diskusi : 30k-50k (sesuai tingkat kesulitannya) tugas matkul : 65-85k (sesuai tingkat kesulitannya) TMK/matkul : 35k pilgan : 15k order? wa in bio #universitasterbuka #UT #jokitugas #zonauang #zonajajan #jokimurah	Neutral	Neutral	Negative
Sahabat UT, yuk ikuti Klinik Literasi Perpustakaan UT LIB-TALK" bersama Irmayati, S.IP. pada hari Rabu, 15 November 2023 pukul 13.30 WIB via Zoom meeting pada link berikut: https://t.co/L0esT3UNMx #Literasi #Perpustakaan #UniversitasTerbuka https://t.co/QG51zOqgcO "	Neutral	Positive	Positive

Direktur dan Seluruh Sivitas Akademika Universitas Terbuka Palembang mengucapkan Selamat Memperingati Hari Pahlawan Nasional - 10 November 2023 - Semangat Pahlawan untuk Masa Depan Bangsa dalam Memerangi Kemiskinan dan Kebodohan." #universitasterbuka #utpalembang #infout https://t.co/17kWF59b9B"	Positive	Positive	Positive
...
Eka dessy indah fitriani nomor induk mahasiswa 043839784 nomor hp 081217530649 Universitas Terbuka prodi manajemen penipu joki tugas	Neutral	Negative	Negative
@thisisgreeny tauukk iiii tp kaa, rerata yg bilang alih kredit tu lama dee prosesnya.. ada yg bilang sampe 3 bulanan ada grupnya di tele lebih aptudet ketik aja universitas terbuka.. aku di kick adminnya negara bikin promo joki tugas disitu wkwkwkkkw	Positive	Positive	Negative

Dari 685 data komentar yang diperoleh melalui media social TikTok dan X(Twitter) melalui proses labeling secara manual diperoleh sentimen negatif sebanyak 281 data komentar dan sentiment positif sebanyak 404 data komentar.



Gambar 8. Grafik Hasil Proses Labeling

C. Pre-processing

Pre-processing adalah proses mengubah data mentah menjadi berguna dan efisien. Pada tahapan ini, data akan melewati beberapa proses, yaitu proses *cleansing*, *case folding*, *tokenizing*, *steming*, *stopword removal*.

- *Cleansing*

Pada tahapan *cleansing* meliputi, menghapus emot, menghapus *url/link*, menghapus mentions, menghapus hashtags, dan menghapus spasi berlebih. Berikut perbandingan data komentar sebelum dan sesudah proses *cleansing*.

Tabel 4. Data Komentar Setelah Proses *Cleansing*

Sebelum	Sesudah
Pendidikan Masyarakat : Pondasi Pengetahuan MasaPendidikan Masyarakat Pondasi Pengetahuan Depan https://t.co/q7aovz8THi lewat @ut_hkg	Masa Depan lewat uthkg
Open joki tugas buat mahasiswa/i UniversitasOpen joki tugas buat mahasiswai Universitas Terbuka. diskusi : 30k-50k (sesuai tingkatTerbuka diskusi kk sesuai tingkat kesulitannya kesulitannya) tugas matkul : 65-85k (sesuai tingkat tugas matkul k sesuai tingkat kesulitannya kesulitannya) TMK/matkul : 35k pilgan : 15k order?TMKmatkul k pilgan k order wa in bio wa in bio #universitasterbuka #UT #jokitugasuniversitasterbuka UT jokitugas zonauang #zonauang #zonajajan #jokimurah	zonajajan jokimurah

- *Case Folding*

Mengubah semua kata dalam teks isi pantun menjadi berhuruf kecil. Berikut perbandingan data komentar sebelum dan sesudah proses *cleansing*.

Tabel 5. Data Komentar Setelah Proses Case Folding

Sebelum	Sesudah
Pendidikan Masyarakat Pondasi Pengetahuan Masa Depan lewat uthkg	pendidikan masyarakat pondasi pengetahuan masa depan lewat uthkg
Open joki tugas buat mahasiswai Universitas Terbuka diskusi kk sesuai tingkat kesulitannya tugas matkul k sesuai tingkat kesulitannya TMKmatkul k pilgan k order wa in biotmkmatkul k pilgan k order wa in bio universitasterbuka UT jokitugas zonauang zonajajan jokimurah	open joki tugas buat mahasiswai universitas terbuka diskusi kk sesuai tingkat kesulitannya tugas matkul k sesuai tingkat kesulitannya tmkmatkul k pilgan k order wa in biotmkmatkul k pilgan k order wa in bio universitasterbuka ut jokitugas zonauang zonajajan jokimurah

- *Tokenizing*

Memecah kalimat menjadi token-token atau kata-kata tunggal. Berikut perbandingan data komentar sebelum dan sesudah proses *tokenizing*.

Tabel 6. Data Komentar Setelah Proses Tokenizing

Sebelum	Sesudah
pendidikan masyarakat pondasi pengetahuan masa depan lewat uthkg	['pendidikan', 'masyarakat', 'pondasi', 'pengetahuan', 'masa', 'depan', 'lewat', 'uthkg']
open joki tugas buat mahasiswai universitas terbuka diskusi kk sesuai tingkat kesulitannya tugas matkul k sesuai tingkat kesulitannya tmkmatkul k pilgan k order wa in bio universitasterbuka ut jokitugas zonauang zonajajan jokimurah	['open', 'joki', 'tugas', 'buat', 'mahasiswai', 'universitas', 'terbuka', 'diskusi', 'kk', 'sesuai', 'tingkat', 'kesulitannya', 'tugas', 'matkul', 'k', 'tmkmatkul', 'k', 'pilgan', 'k', 'order', 'wa', 'in', 'bio', 'universitasterbuka', 'ut', 'jokitugas', 'zonauang', 'zonajajan', 'jokimurah']

- *Stemming*

Mengembalikan kata ke dalam kata dasarnya atau kata baku. Berikut perbandingan data komentar sebelum dan sesudah proses *steming*.

Tabel 7. Data Komentar Setelah Proses Stemming

Sebelum	Sesudah
['pendidikan', 'masyarakat', 'pondasi', 'pengetahuan', 'masa', 'depan', 'lewat', 'uthkg']	['didik', 'masyarakat', 'pondasi', 'tahu', 'masa', 'depan', 'lewat', 'uthkg']
['open', 'joki', 'tugas', 'buat', 'mahasiswai', 'universitas', 'terbuka', 'diskusi', 'kk', 'sesuai', 'tingkat', 'kesulitannya', 'tugas', 'matkul', 'k', 'sesuai', 'tingkat', 'kesulitannya', 'tmkmatkul', 'k', 'pilgan', 'k', 'order', 'wa', 'in', 'bio', 'universitasterbuka', 'ut', 'jokitugas', 'zonauang', 'zonajajan', 'jokimurah']	['open', 'joki', 'tugas', 'buat', 'mahasiswa', 'universitas', 'buka', 'diskusi', 'kk', 'sesuai', 'tingkat', 'sulit', 'tugas', 'matkul', 'k', 'sesuai', 'tingkat', 'sulit', 'tmkmatkul', 'k', 'pilgan', 'k', 'order', 'wa', 'in', 'bio', 'universitasterbuka', 'ut', 'jokitugas', 'zonauang', 'zonajajan', 'jokimurah']

- *Stopword Removal*

Menghilangkan kata-kata yang dianggap tidak memiliki makna. Berikut perbandingan data komentar sebelum dan sesudah proses *stopword removal*.

Tabel 8. Data Komentar Setelah Proses Stopword Removal

Sebelum	Sesudah
[didik', 'masyarakat', 'pondasi', 'tahu', 'masa',didik masyarakat pondasi uthkg 'depan', 'lewat', 'uthkg']	
[open', 'joki', 'tugas', 'buat', 'mahasiswa',open joki tugas mahasiswa universitas buka 'universitas', 'buka', 'diskusi', 'kk', 'sesuai', 'tingkat',diskusi kk sesuai tingkat sulit tugas matkul k 'sulit', 'tugas', 'matkul', 'k', 'sesuai', 'tingkat', 'sulit',sesuai tingkat sulit tmkmatkul k pilgan k order 'tmkmatkul', 'k', 'pilgan', 'k', 'order', 'wa', 'in', 'bio',wa in bio universitasterbuka ut jokitugas 'universitasterbuka', 'ut', 'jokitugas', 'zonauang',zonauang zonajajan jokimurah 'zonajajan', 'jokimurah']	

D. TF-IDF

Algoritma TF-IDF atau Term-Frequency-Inverse Document merupakan salah satu algoritma yang berfungsi untuk menghitung bobot setiap kata. Proses TF-IDF dapat dilihat pada Gambar 8.

```
#Training Testing - TF-IDF (menampilkan kata yg paling banyak muncul)
vectorizer = TfidfVectorizer(min_df = 5,
                             max_df = 0.8,
                             sublinear_tf = True,
                             use_idf = True)
vectorizer.fit_transform(df['komentar'])
Train_X_Tfidf = vectorizer.transform(Train_X)
Test_X_Tfidf = vectorizer.transform(Test_X)
print(Train_X_Tfidf)
```

Gambar 9. Proses TF-IDF

E. Accuracy

Berdasarkan tahapan validasi yang dilakukan, diperoleh nilai akurasi sesuai dengan tabel dibawah ini.

Tabel 9. Nilai Akurasi

Algoritma	Akurasi
Super Vector Machine (SVM)	88.34%
Naïve Bayes – MultinomialNB	87.37%
Naïve Bayes – Gausiyannb	87.37%
K-NN	85.43%
Decision Tree	83.00%
Logistic Regresion	83.00%

F. Pre-Trained BERT Model

Pada tahap ini data telah melalui proses encoding dan tokenisasi. Selanjutnya melalui proses pembuatan model pre-trained BERT yang bertujuan memasukkan dataset ke dalam model untuk diuji dan dilatih.

Uji coba dataset ditentukan 3 epoch karena untuk menguji dan memastikan tidak terjadi underfitting (suatu kondisi ketika model tidak mampu untuk mempelajari beberapa data latih yang telah di masukkan) dan akan berakhir dengan performa yang buruk. Terdapat beberapa parameter untuk menentukan apakah model underfitting atau overfitting (suatu kondisi dimana model melewati hasil yang diharapkan pada beberapa iterasi

pertama yang dinilai terlalu optimis). Penentuan semua parameter diatas pada dasarnya berdasarkan “Hit-and-trial error” hingga mendapatkan performa yang diharapkan.

G. Hasil Klasifikasi Metode BERT

Epoch 1/3

Average Training Loss : 0.6099557945693749
 Average Validation Loss : 0.4045141165455182
 Validation Accuracy : 0.8686131386861314

Tabel 10. Hasil Evaluasi Pada Pre-Trained BERT Model Pada Epoch 1/3

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>Support</i>
0	0.91	0.75	0.82	56
1	0.85	0.95	0.90	81
Accuracy			0.87	137
Macro avg	0.88	0.85	0.86	137
Weighted avg	0.87	0.87	0.87	137

Epoch 2/3

Average Training Loss : 0.3851983586083288
 Average Validation Loss : 0.28419020399451256
 Validation Accuracy : 0.8978102189781022

Tabel 11. Hasil Evaluasi Pada Pre-Trained BERT Model Pada Epoch 2/3

	<i>Precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
0	0.89	0.86	0.87	56
1	0.90	0.93	0.91	81
Accuracy			0.90	137
Macro avg	0.90	0.89	0.89	137
Weighted avg	0.90	0.90	0.90	137

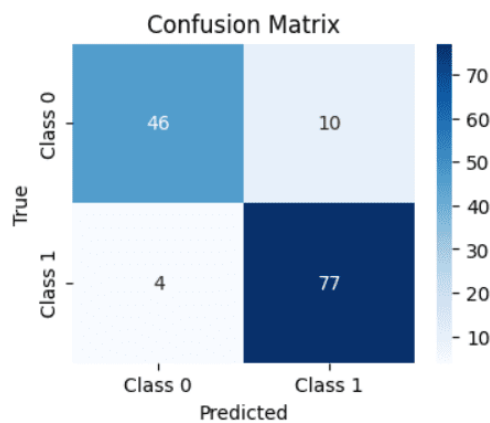
Epoch 3/3

Average Training Loss : 0.2838914982326653
 Average Validation Loss : 0.23185356044107014
 Validation Accuracy : 0.8978102189781022

Tabel 12. Hasil Evaluasi Pada Pre-Trained BERT Model Pada Epoch 3/3

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
0	0.92	0.82	0.87	56
1	0.89	0.95	0.92	81
Accuracy			0.90	137
Macro avg	0.90	0.89	0.89	137
Weighted avg	0.90	0.90	0.90	137

Parameter kinerja dianalisis untuk menilai kinerja sistem dengan matriks konfusi. Gambar 9 adalah hasilnya matriks konfusi berdasarkan akurasi kinerja tertinggi dalam pengujian kami.



Gambar 10. Hasil Confusion Matrix

IV. Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilaksanakan, dapat disimpulkan beberapa hal yaitu :

1. Dengan adanya analisis sentimen yang ditulis oleh penulis maka diketahui analisis sentiment terhadap layanan akademik dan non akademik Universitas Terbuka pada media sosial Twitter(X) dan TikTok dari 685 data diperoleh sentimen positif sebesar 281 data dan sentimen negatif 404 sebesar data.
2. Nilai akurasi tertinggi diperoleh dengan menggunakan algoritma SVM sebesar 88%.
3. Nilai akurasi yang diperoleh dari hasil *deep learning* BERT yaitu 90%.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Lembagakita.org dalam pengelolaan jurnal ilmiah serta teman – teman kelas 14.2A.10 program Magister Ilmu Komputer Universitas Nusa Mandiri Margonda

Daftar Pustaka

- [1] B. Basmi, Burhan, A. Sigit, and S. Musa, “Pengaruh Self Regulation Learning dan Efikasi Diri terhadap Hasil Belajar Melalui Fasilitas Belajar di UPBJJ Univesitas Terbuka Makassar Pokjar Kabupaten Wajo,” *Cokroaminoto J. Prim. Educ.*, vol. 5, no. 1, pp. 90–103, 2022, doi: 10.30605/cjpe.512022.1587.
- [2] F. A. Girnanfa and A. Susilo, “Studi Dramaturgi Pengelolaan Kesan Melalui Twitter Sebagai Sarana Eksistensi Diri Mahasiswa di Jakarta,” *J. New Media Commun.*, vol. 1, no. 1, pp. 58–73, 2022, doi: 10.55985/jnmc.v1i1.2.
- [3] A. D. Kusumandaru and F. P. Rahmawati, “Implementasi Media Sosial Aplikasi Tik Tok sebagai Media Menguatkan Literasi Sastra dalam Pembelajaran Tematik di Sekolah Dasar,” *J. Basicedu*, vol. 6, no. 3, pp. 4876–4886, 2022, doi: 10.31004/basicedu.v6i3.2972.
- [4] F. Hadi and D. Guswandi, “Penentuan Penerimaan Mahasiswa Baru Pascasarjana Menggunakan Simple Additive Weighting (SAW),” *Indones. J. Comput. Sci.*, vol. 8, no. 2, pp. 121–129, 2019, doi: 10.33022/ijcs.v8i2.175.
- [5] H. N. Irmanda and Ria Astriratma, “Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM),” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020, doi: 10.29207/resti.v4i5.2313.
- [6] J. Teknika, R. K. Septiani, S. Anggraeni, and S. D. Saraswati, “Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes,” *Teknika*, vol. 16, no. 2, pp. 245–254, 2022, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/teknika/article/view/4875>
- [7] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
- [8] A. N. Azhar and M. L. Khodra, “2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA) : September 8-9, 2020, online conference.,” 2020.

- [9] D. Ayu Muthia, "Komparasi Algoritma Klasifikasi Text Mining Untuk Analisis Sentimen Pada Review Restoran," *Maret*, vol. 14, no. 1, p. 69, 2018, [Online]. Available: <http://www.bsi.ac.id>
- [10] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," no. October 2018, pp. 11–16, 2016, doi: 10.21108/indosc.2016.111.
- [11] D. D. A. Yani, H. S. Pratiwi, and H. Muhandi, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, p. 257, 2019, doi: 10.26418/justin.v7i4.30930.
- [12] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [13] R. Mas, R. W. Panca, K. Atmaja¹, and W. Yustanti², "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *Jeisbi*, vol. 02, no. 3, pp. 55–62, 2021.