



# Pengkategorian Komentar Instagram Terhadap Layanan Akademik dan Non-Akademik Universitas Terbuka

Rhini Fatmasari <sup>1</sup>, Alda Zevana Putri Widodo <sup>2\*</sup>, Valianda Farradillah Hakim <sup>3</sup>, Windu Gata <sup>4</sup>, Dedi Dwi Saputra <sup>5</sup>

<sup>1</sup> Fakultas Keguruan dan Ilmu Pendidikan, Universitas Terbuka.

<sup>2,3,4,5</sup> Fakultas Sistem Informasi, Universitas Nusa Mandiri.

## article info

### Article history:

Received 12 June 2022

Received in revised form

5 October 2022

Accepted 10 November 2022

Available online January 2023

### DOI:

<https://doi.org/10.35870/jtik.v7i1.669>

### Keywords:

Universitas Terbuka;

Instagram; Rapidminer.

### Kata Kunci:

Universitas Terbuka;

Instagram; Rapidminer.

## abstract

Instagram is one of the social media that has many users in Indonesia, where users are free to comment on whatever is going on, including being a form of online communication between campuses and their students. The number of topics and comments on an official Instagram account can be used as evaluation or learning material. The Open University is one of the campuses that has an official Instagram account with thousands of followers. In order to get an evaluation of academic and non-academic services, in this study a categorization analysis was carried out with 10,000 comment data taken from the official @univterbuka Instagram account. The data is categorized into 7 categories, then processed using 4 algorithms, namely SVM, Naïve Bayes, Random Forest and KNN. The highest accuracy in the category of teachers with the KNN method is 98.97% and the highest AUC is in the module category with the SVM method of 94.60%.

## abstract

Instagram merupakan salah satu media sosial yang memiliki banyak pengguna di Indonesia, dimana penggunaannya bebas berkomentar mengenai apapun yang sedang terjadi, termasuk menjadi salah satu bentuk komunikasi online antar kampus dengan mahasiswanya. Banyaknya topik dan komentar pada suatu akun official Instagram dapat dimanfaatkan sebagai bahan evaluasi atau pembelajaran. Universitas Terbuka adalah salah satu kampus yang memiliki akun official instagram dengan ribuan pengikut. Agar mendapatkan evaluasi terhadap layanan akademik dan non akademik, maka dalam penelitian ini dilakukan analisis pengkategorian dengan 10.000 data komentar yang diambil dari akun instagram official @univterbuka. Data tersebut dikategorikan menjadi 7 kategori, lalu diolah dengan menggunakan 4 algoritma yaitu SVM, Naïve Bayes, Random Forest dan KNN. Akurasi tertinggi pada kategori pengajar dengan metode KNN yaitu sebesar 98,97% serta AUC tertinggi pada kategori modul dengan metode SVM sebesar 94,60%.

\*Corresponding author. Email: [14210191@nusamandiri.ac.id](mailto:14210191@nusamandiri.ac.id) <sup>2</sup>.

## 1. Latar Belakang

Universitas Terbuka adalah universitas yang berbasis pada studi mandiri dan inisiatif. Pembelajaran dilakukan secara individu atau kelompok menggunakan perpustakaan, mengikuti bimbingan dari internet, radio, televisi, tutorial komputer, program audio atau video [1]. Dalam Webometrics.info, Universitas Terbuka memperoleh urutan di 71 untuk seluruh Universitas di Indonesia, dan dalam ranking dunia menempati urutan di 5041.

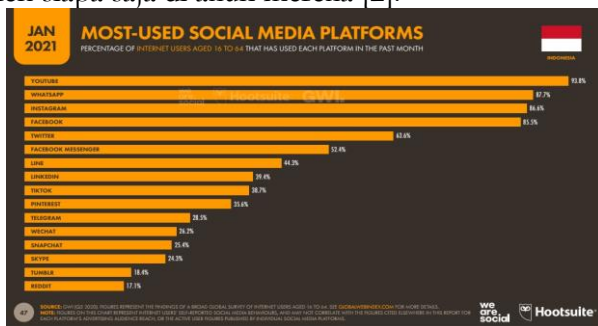
Ranking	World Ranking	University	Det.	Country	Impact	Openness	Excellence
71	5041	Universitas Terbuka			3106	2553	6775

Gambar 1. Ranking Universitas Terbuka di seluruh Universitas di Indonesia data Webometrics.info)

World Rank	University	Det.	Country	Impact	Openness	Excellence
5041	Universitas Terbuka			3106	2553	6775

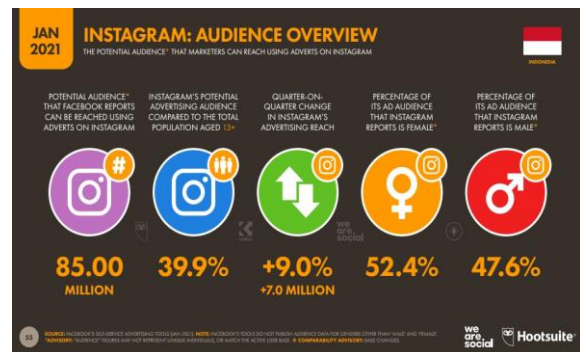
Gambar 2. Ranking Universitas Terbuka di seluruh dunia (data Webometrics.info)

Instagram adalah *platform* media sosial yang digunakan oleh orang Indonesia untuk berkomunikasi dengan orang-orang melalui internet, dengan banyak foto dan video yang dapat diposting oleh siapa saja di akun mereka [2].



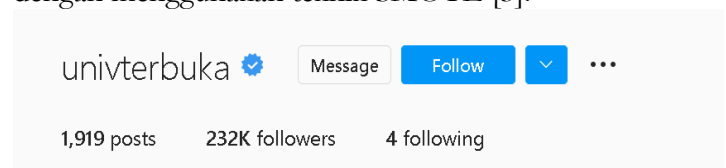
Gambar 3. Media sosial yang paling aktif 2021 (data survei We are Social)

Berdasarkan data dari wearesocial.com tahun 2021, pengguna aplikasi Instagram menempati urutan ketiga sebanyak 86,6% dari jumlah populasi. Pengguna Youtube di Indonesia masih menempati urutan pertama yaitu 93,8% dari jumlah populasi di Indonesia.



Gambar 4. Ringkasan pengguna Instagram tahun 2021 (data survei We are Social)

Ringkasan pengguna Instagram di Indonesia tahun 2021 dari data wearesocial.com, jumlah pengguna Instagram sebanyak 85 juta jiwa, dimana dengan jenis kelamin perempuan 52,4% dan laki-laki 47,6%. Penelitian dalam hal pengkategorian ini sudah dilakukan sebelumnya yaitu pada pengkategorian keluhan pengguna XL pada media sosial twitter dengan menggunakan teknik SMOTE [3].



Gambar 5. Instagram Universitas Terbuka

Universitas Terbuka memiliki akun resmi pada aplikasi Instagram dengan id @univterbuka dengan followers 232 ribu pengguna (per tanggal 1 Juni 2022). *Dataset* adalah sekumpulan data *record* atau data training dimana record dimiliki oleh atribut [4]. *Dataset* penelitian penulis didapat dari hasil scraping pada komentar-komentar beberapa postingan di akun instagram Universitas Terbuka. *Dataset* yang kami hasilkan memiliki 7 atribut diantaranya adalah Modul, Tutorial, Pengajar, Ujian, Aplikasi, Non Akademik, Other.

- 1) Modul : Jika ada kata dalam komentar yang terkait modul, bahan ajar, pdf, ppt, dan lain-lain.
- 2) Tutorial : Jika ada kata dalam komentar yang terkait tuton, tuweb, tutorial, online, e-learning, dan lain-lain.
- 3) Pengajar : Jika ada kata dalam komentar yang terkait dosen, tutor, pembimbing, penguji, dan lain-lain.
- 4) Ujian : Jika ada kata dalam komentar yang terkait UAS, MID, Take Home, ujian, dan lain-lain.
- 5) Aplikasi : Jika ada kata dalam komentar yang

terkait aplikasi, lelet, tidak bisa dibuka, dan lain-lain.

- 6) Non Akademik : Jika ada kata dalam komentar yang terkait layanan di UPBJJ, registrasi, pendaftaran, wisuda, ujian ulang, cetak kartu mahasiswa, lama pelayanan, keramahan petugas, informasi tentang kuliah di UT, contact centre, dan lain-lain.
- 7) Other : Jika ada kata dalam komentar yang terkait lainnya selain yang disebutkan dalam a sampai f.

Peningkatan kualitas pelayanan merupakan salah satu aspek penting untuk menentukan kepuasan mahasiswa pada Universitas Terbuka. untuk memperoleh hasil evaluasi yang baik diperlukan metode yang tepat dalam memproses data evaluasi sehingga hasilnya menggambarkan keadaan yang sebenarnya, dimana evaluasi ini menggunakan multi atribut dimana pemilihan atribut telah ditentukan agar berpengaruh pada evaluasi yang lebih baik lagi [5].

Preprocessing adalah teknik untuk mengolah data mentah menjadi data bersih untuk digunakan dalam proses analisis selanjutnya [6]. WebHarvy menggunakan bot otomatis untuk melakukan *scraping* web, memecah halaman target menjadi teks HTML, yang merupakan format paling dasar, mengekstraksi data informasi, memindainya, dan menyimpannya dalam bentuk XLSX, CSV, TSV, XML, JSON dan juga dalam bentuk *database*. Proses pengumpulan data sesuai dengan parameter yang telah ditentukan, yang menghasilkan format XLSX, CSV, TSV, XML dan JSON [7]. Rapidminer Studio adalah paket perangkat lunak penambangan data yang kuat yang menggunakan ratusan persiapan data dan algoritma pembelajaran mesin untuk membangun model prediktif dengan cepat dengan memasukkan data mentah seperti database dan teks yang dianalisis secara otomatis untuk mendukung proyek data mining [8]. KNN adalah model algoritma yang digunakan untuk mengklasifikasikan objek berdasarkan data dengan menganalisis jarak terpendek ke objek dengan tujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan data latih [9]. Naive Bayes adalah algoritma data mining untuk mengklasifikasikan data dengan kemungkinan probabilitas untuk keputusan berbasis formula [10].

$$p(X|H) = \frac{P(X|H)P(H)}{P(X)}$$

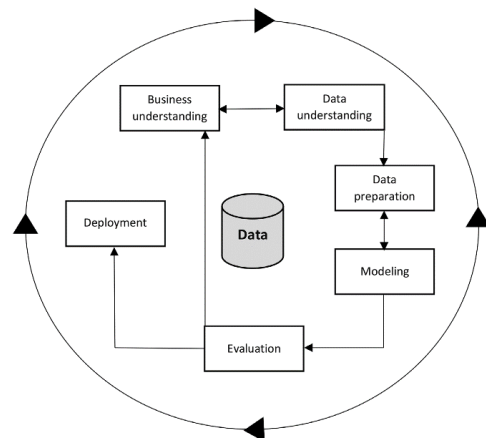
Keterangan:

- X : Data dengan class yang belum diketahui  
 H : Hipotesis data x merupakan suatu class spesifik  
 P(H|X) : Probabilitas hipotesis H berdasarkan kondisi x  
 P(H) : Probabilitas hipotesis H  
 P(X|H) : Probabilitas X berdasarkan kondisi tersebut  
 P(X) : Probabilitas dari X

SVM adalah teknik prediksi, klasifikasi, dan regresi yang memisahkan dua kelas data dengan mencari hyperplane terbaik [11]. Random Forest adalah kombinasi dari beberapa algoritma decision tree, dimana setiap tree bergantung pada vektor acak sampel dengan distribusi yang seragam dalam satu forest. Jika nilai F terlalu kecil, korelasi tree akan kecil [12].

## 2. Metode Penelitian

Dalam penelitian ini penulis menggunakan model yang bernama CRISP-DM (*Cross-Industry Standard Process for Data Mining*), CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah model proses independen industri untuk *data mining*, model ini terdiri dari enam fase berulang dari Business understanding sampai dengan Deployment [13], berikut adalah fase - fase dari model CRISP-DM pada gambar 6.



Gambar 6. Model CRISP-DM ([14])

### Business Understanding

Pada tahap awal ini adalah mengklarifikasi tujuan klien dan mendefinisikan kriteria prioritas dan keberhasilan [15].

### Data Understanding

Pada tahap ini adalah dimulai dari data awal fase pengumpulan. Kegiatannya meliputi: mendeskripsikan data, mengeksplorasi data, dan mengidentifikasi masalah yang terkait dengan kualitas data, dan mendeteksi subset data yang menarik sebagai hipotesis awal [16].

### Data Preparation

Pada tahap ini sebuah *dataset* disiapkan untuk pelatihan model. Fase ini termasuk pembersihan, imputasi data yang hilang, seleksi, penggabungan, dan/atau pemformatan data [17].

### Modeling

Pada tahap ini melibatkan pemilihan dan pengembangan analitik teknik dan model. Selain itu, bagian dari kumpulan data sering diatur selain untuk pelatihan dan validasi model [13].

### Evaluation

Pada tahap ini hasilnya diperiksa terhadap tujuan bisnis yang ditentukan. Oleh karena itu, hasilnya harus tindakan yang ditafsirkan dan lebih lanjut harus didefinisikan. Poin lainnya adalah, bahwa proses tersebut harus ditinjau secara umum [14].

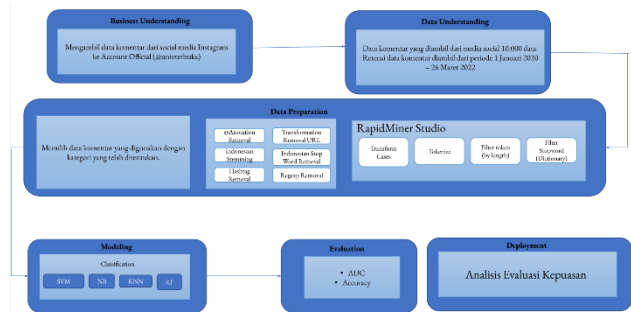
### Deployment

Pada tahap ini akan dijelaskan secara umum dalam panduan pengguna. Ini bisa berupa laporan akhir atau komponen perangkat lunak. Pengguna panduan menjelaskan bahwa fase penyebaran terdiri dari perencanaan penyebaran, pemantauan, dan pemeliharaan [14].

## 3. Hasil dan Pembahasan

### Model CRISP-DM

Model yang digunakan dalam pengkategorian ini menggunakan model CRISP-DM, berikut adalah implementasi dari model tersebut pada gambar 7.



Gambar 7. Model CRISP~DM (dokumen pribadi)

### Business Understanding

Data komentar ini diambil dari instagram resmi Universitas Terbuka (@univterbuka), data ini diambil dengan menggunakan *software* WebHarvy versi 6.30.

### Data Understanding

Data komentar ini diambil dengan rentang waktu 1 Januari 2020 sampai dengan 26 Maret 2022 dengan total data yang diambil sebanyak 10.000 data.

### Data Preparation

Data komentar ini akan ditentukan berdasarkan 7 kategori masing - masing yaitu Modul, Tutorial, Pengajar, Ujian, Aplikasi, Non Akademik dan Other berdasarkan kata kunci tertentu pada setiap komentarnya, kemudian dari data komentar tersebut akan dilakukannya *pre-processing* pertama di *software* RapidMiner Studio dengan melalui 4 tahap (Transform Case, Tokenize, Filter token (by length) dan Filter Stopword (Dictionary)), kemudian setelah dilakukan *pre-processing* pertama, maka akan ada 6 tahap untuk menghilangkan beberapa kata (Transformation Remove Url, @Annotation Removal, Hashtag Removal, Regexp Removal, Indonesian Stemming, Indonesian Stopword Removal).

### Modeling

Data komentar ini akan diproses dengan menggunakan 4 algoritma *machine learning* yaitu Naive Bayes, SVM, KNN (K = 5) dan Random Forest

### Evaluation

Data komentar yang sudah diproses dengan algoritma *machine learning* akan ditentukan nilai akurasi dan AUC dari setiap algoritma *machine learning*.

### Deployment

Data komentar akan dilakukan analisis untuk komentar terbanyak dalam kategori tertentu yang sering ditanyakan di media sosial Instagram.

### Data Komentar

Data komentar ini diambil dari instagram resmi Universitas Terbuka (@univterbuka), data komentar ini diambil menggunakan *software* WebHarvy versi 6.30 dengan rentang waktu 1 Januari 2020 sampai dengan 26 Maret 2022 dengan total data sebanyak 10.000 data, kemudian data komentar ini akan dilakukan pengkategorian dengan 7 kategori yaitu Modul, Tutorial, Pengajar, Ujian, Aplikasi dan Non Akademik, data komentar tersebut dicocokkan sesuai dengan kriteria tertentu dari masing - masing algoritma, seperti contohnya jika data komentar tersebut membahas mengenai ujian maka data tersebut berada di kategori ujian maka data tersebut di kategori komentar diketik 1, untuk kategori lainnya diketik 0 karena data tersebut tidak membahas kata - kata yang berhubungan dengan kategori lainnya, namun dalam kasus ini, terdapat beberapa data komentar yang berada dalam 2 kategori yang berbeda, dalam melakukan pengkategorian membutuhkan 2 user atau disebut juga manual.

Berikut adalah jumlah komentar dari masing - masing kategori yang ditanyakan di instagram resmi Universitas terbuka, yaitu :

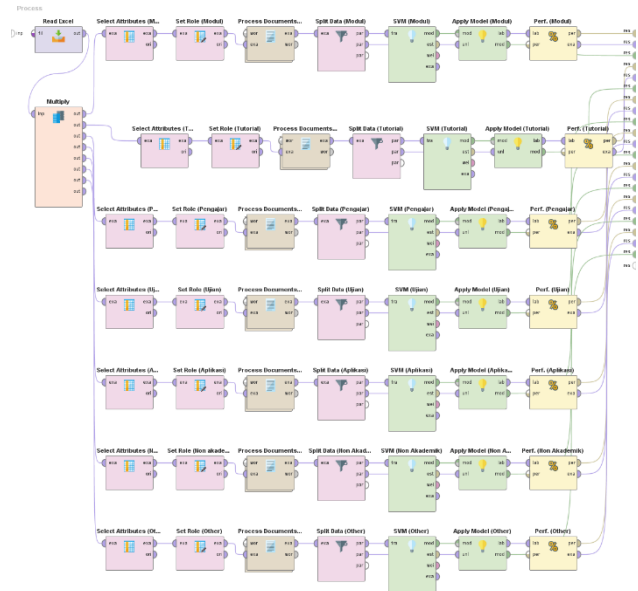
- 1) Modul : 187 data
- 2) Tutorial : 381 data
- 3) Pengajar : 181 data
- 4) Ujian : 781 data
- 5) Aplikasi : 1.054 data
- 6) Non Akademik : 5.490 data
- 7) Other : 2.490 data

### Model RapidMiner Studio

Pemodelan yang dilakukan dalam analisis ini menggunakan *software* RapidMiner Studio versi 9.10.

### Model Pengkategorian Komentar

Berikut ini adalah model dari pengkategorian komentar dalam gambar 8 yaitu:



Gambar 8. Model pengkategorian komentar

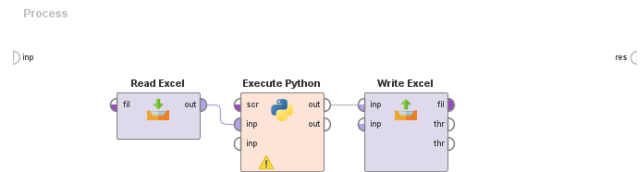
Model pengkategorian komentar yang terdapat dalam *software* RapidMiner Studio ini berisi banyak proses didalamnya dengan menggunakan 7 kategori (Modul, Tutorial, Pengajar, Ujian, Aplikasi, Non Akademik, dan Other) dalam bentuk tabel 1, yaitu :

Tabel 1. Proses model pengkategorian komentar

Nama Process	Keterangan
Read Excel	Membaca file excel dalam bentuk .xls / .xlsx
Multiply	Membuat salinan dari objek
Select Attribute	Memilih attribute yang akan diproses, contohnya kategori modul dengan komentar
Set Role	Memilih role yang akan diproses contohnya kategori modul
Process Document	Melakukan proses pada data
Split Data	Melakukan pembagian data menjadi data train dan testing (70:30)
SVM	Nama algoritma yang digunakan

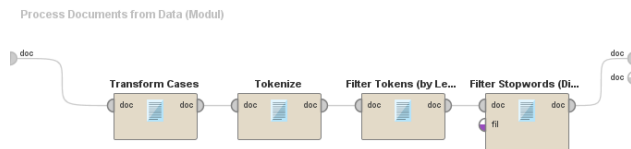


Apply Model	Mengaplikasikan model dari data komentar
Perf (Performance)	Menganalisis data dari komentar, seperti akurasi, AUC, dan lain sebagainya



Gambar 10. Model Processing Data

Khusus untuk proses Process Document terdapat 4 tahap yang dijalankan dalam bentuk gambar 9, yaitu:



Gambar 9. Model Process Document  
Penjelasan masing - masing proses dijelaskan dalam bentuk tabel 2, yaitu :

Nama Proses	Keterangan
Transform Cases	Mengubah struktur data komentar, dalam kasus ini menjadi <i>lowercase</i> (huruf kecil semua)
Tokenize	Memisahkan dari kalimat di data menjadi sekuen dari kata
Filter Tokens (by Length)	Melakukan filter data komentar, dalam kasus ini data komentar yang akan diproses untuk minimal kata sebanyak 4 kata dan maksimal 25 kata
Filter Stopword (Dictionary)	Melakukan filter stopwords data komentar, dalam kasus ini stopwords yang diambil dari kata kata yang tidak penting dalam data

#### Model Processing Data

Berikut ini adalah model dari pengkategorian komentar dalam gambar 10 yaitu:

Model *processing* data ini diproses pada situs web [gataframework.com](http://gataframework.com) dengan menggunakan bahasa pemrograman Python, model ini berisi tiga proses dalam bentuk tabel 3, yaitu :

Nama Proses	Keterangan
Read Excel	Membaca file excel dalam bentuk .xls / .xlsx
Execute Python	Melakukan proses koding menggunakan bahasa pemrograman Python
Write Excel	Menulis file excel dalam bentuk .xls / .xlsx

*Pre-processing* data komentar ini melalui 6 tahapan yang dilakukan, yaitu :

#### Transformation Remove URL

Tahapan awal ini setiap data komentar yang memiliki link url, maka link url tersebut akan dihilangkan. berikut adalah contoh salah satu datanya dalam tabel 4, yaitu :

Sebelum	Sesudah
Guys, sorry for asking, website elearning.ut.ac.id lagi error? terus, Ruang Baca Virtual UT juga sama, lagi error kah? mau login baca ebook yang kemarin malam username dan password nya masih bisa login, malam ini ada tulisan "username dan password tidak ditemukan." jadi heran... ada yang alami juga? ini link nya <a href="https://www.pustaka.ut.ac.id/lib/rb">https://www.pustaka.ut.ac.id/lib/rb</a>	guys, sorry for asking, website elearning.ut.ac.id lagi error? terus, ruang baca virtual ut juga sama, lagi error kah?

v/fhisip/sastra-inggris-s1-  
penerjemahan/

mau login  
baca ebook  
yang  
kemarin  
malam  
username  
dan  
password  
nya masih  
bisa login,  
malam ini  
ada tulisan  
"username  
dan  
password  
tidak  
ditemukan.  
" jadi  
heran... ada  
yang alami  
juga? ini  
link  
nya

adalah salah satu contoh datanya dalam bentuk tabel 6, yaitu:

Tabel 6. Preprocessing Hashtag Removal

Sebelum	Sesudah
Sedih iyaah kecewa juga iyaah.. Saat orang lain berusaha berikhtiar mendaftar cpns kita yg masih menggantung makin digantung sampai entah kapan katanya.. Maaf bapak ibu UT yg terhormat.. Kami para mahasiswa memahami keadaan negeri ini bagaimna.. Namun ada selalu solusi disetiap masalah dan alternatif demi keadilan kami yg juga ingin berusaha seperti mereka.. Jaman sekarang masih ada kurir yg ingin mengantarkan barang2 jikalau tak bisa pun masih ada yg namanya teknologi dmna mengirim dokumen hanya menunggu menit bahkan detik langsung terkirim.. Mohon pertimbangannya para petinggi @univterbuka Kami hargai usaha kerja kalian.. Kalian juga hargai kepercayaan kami.. #ijazahUTbagikan #wisudaonline2021 @univterbuka @ut_bogor	sedih iyaah kecewa juga iyaah.. saat orang lain berusaha berikhtiar mendaftar cpns kita yg masih menggantung makin digantung sampai entah kapan katanya.. maaf bapak ibu ut yg terhormat.. kami para mahasiswa memahami keadaan negeri ini bagaimna.. namun ada selalu solusi disetiap masalah dan alternatif demi keadilan kami yg juga ingin berusaha seperti mereka.. jaman sekarang masih ada kurir yg ingin mengantarkan barang2 jikalau tak bisa pun masih ada yg namanya teknologi dmna mengirim dokumen hanya menunggu menit bahkan detik langsung terkirim.. mohon pertimbangannya para petinggi @univterbuka kami hargai usaha kerja kalian.. kalian juga hargai kepercayaan kami.. @univterbuka @ut_bogor

#### @Annotation Removal

Tahapan ini setiap data komentar yang memiliki simbol @ (add) maka akan dihilangkan, berikut adalah contoh salah satu datanya dalam tabel 5, yaitu :

Tabel 5. Preprocessing @Annotation Removal

Sebelum	Sesudah
@univterbuka Kak mau tanya, saya lulusan November th 2020 sudah mengumpulkan foto dll tp sampe sekarang belum ada info lagi. Kapan ya saya bisa menerima ijazahnya? 🙏 terimakasih	kak mau tanya, saya lulusan november th 2020 sudah mengumpulkan foto dll tp sampe sekarang belum ada info lagi. kapan ya saya bisa menerima ijazahnya? 🙏 terimakasih

#### Hashtag Removal

Tahapan ini setiap data komentar yang memiliki simbol # (Hashtag) maka akan dihilangkan, berikut

*Regex Removal*

Tahapan ini setiap data komentar yang memiliki angka di data, maka akan dihilangkan, berikut adalah salah satu contoh data dalam bentuk tabel 7, yaitu:

Tabel 7. Preprocessing Remove Regexp

Sebelum	Sesudah
Lulusan sma tahun 2019 bisa daftar gk min?	lulusan sma tahun bisa daftar gk min

*Indonesian Stemming*

Tahapan ini setiap data komentar yang memiliki kata imbuhan (me-, di-, dan imbuhan lainnya) maka akan dihilangkan, berikut salah satu contohnya dalam bentuk tabel 8, yaitu:

Tabel 8. Preprocessing Indonesian Stemming

Sebelum	Sesudah
Saya Sudah Mendaftar online Tapi gagal terus... 😞	Saya sudah daftar online tapi gagal terus

*Stopword*

Tahapan akhir ini setiap data komentar yang ada kata umum yang muncul dengan jumlah besar dan tidak memiliki makna maka akan dihilangkan, berikut adalah salah satu contohnya dalam bentuk tabel 9, yaitu :

Tabel 9. Indonesian Stopword

Sebelum	Sesudah
Untuk pendaftaran mahasiswa pada fakultas pendidikan, apakah harus mempunyai surat	Pendaftaran mahasiswa fakultas pendidikan, surat mengajar min? mohon informasinya,

mengajar min? Mohon  
informasinya,  
terimakasih 🙏



Berikut adalah contoh dalam melakukan *pre-processing* data di gataframework.com dalam gambar 11, yaitu :

No	textbefore	preprocessingtechniques	textafter	Entrytime	EntryIP
1	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
2	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
3	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
4	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
5	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
6	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
7	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
8	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
9	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215
10	lulusan sma tahun 2019 bisa daftar gk min?	7 - Indonesian Stemming	lulusan sma tahun bisa daftar gk min	2022-09-10 12:11:20	192.204.130.215

Gambar 11. Proses data di gataframework (dokumen pribadi &amp; gataframework)

Saat melakukan *pre-processing* data komentar di gataframework.com terdapat tabel “No” untuk melihat urutan data komentar yang diproses, “textbefore” untuk data komentar yang asli atau belum dilakukan tahap preprocessing, “preprocessingtechniques” untuk nama tahap yang dilakukan dalam melakukan preprocessing data komentar, “textafter” untuk data komentar yang sudah di preprocessing, “Entrytime” yaitu tanggal dan jam berapakah data komentar tersebut diproses, dan yang terakhir “entryIP” untuk melihat IP Address yang melakukan *preprocessing* data komentar.

*Akurasi dan AUC algoritma*

Akurasi dan AUC dari masing - masing kategori dengan empat algoritma yang digunakan yaitu SVM, Naive Bayes, KNN (K=1), dan Random Forest, berikut adalah akurasi dan AUC dari masing - masing kategori dan algoritma dari data komentar yang belum dilakukan pre-processing dalam bentuk tabel 10, yaitu :



Tabel 10. Akurasi &amp; AUC data komentar

Kategori	Nama Algoritma	Akurasi	AUC
Modul	SVM	98.43%	92.6%
Tutorial		96.47%	88.2%
Pengajar		98.77%	91.3%
Ujian		92.73%	83.3%
Aplikasi		91.17%	83.6%
Non akademik		71.77%	82%
Other		84.83%	88.1%
Modul	Naive Bayes	83.43%	50%
Tutorial		73.83%	50%
Pengajar		85.77%	50%
Ujian		64.97%	51.3%
Aplikasi		56.90%	45.3%
Non akademik		66.70%	51.2%
Other		70.53%	60.7%
Modul	KNN	98.40%	50%
Tutorial		43.03%	50%
Pengajar		98.97%	50%
Ujian		93.30%	50%
Aplikasi		89.90%	50%
Non akademik		57.20%	50%
Other		78.63%	50%
Modul	Random Forest	98.13%	82.8%
Tutorial		96.20%	77.1%
Pengajar		98.83%	85.2%
Ujian		92.87%	65.6%
Aplikasi		89.46%	89.4%
Non akademik		54.93%	75.9%
Other		75.33%	71.4%

Kemudian terdapat juga Akurasi dan AUC dari data (Remove URL - Stopword), berikut adalah hasilnya komentar yang sudah dilakukan *pre-processing* dalam bentuk tabel 11, yaitu :

Tabel 11. Akurasi dan AUC hasil pre-processing data komentar

Kategori	Nama Algoritma	Akurasi	AUC
Modul	SVM	98.63%	94.60%
Tutorial		96.27%	88.90%
Pengajar		98.73%	81%
Ujian		92.90%	80.30%
Aplikasi		90%	85%
Non akademik		73.10%	82%
Other		85.77%	87%
Modul	Naive Bayes	83.43%	50%
Tutorial		73.83%	50%
Pengajar		85.77%	50%
Ujian		64.97%	51.3%
Aplikasi		56.90%	45.3%
Non akademik		66.70%	51.2%
Other		70.53%	60.7%
Modul	KNN	98.40%	50%
Tutorial		58.40%	50%
Pengajar		98.87%	50%
Ujian		92.33%	50%
Aplikasi		89.67%	50%
Non akademik		62.10%	50%
Other		76.63%	50%
Modul	Random Forest	98.13%	71.3%
Tutorial		96.20%	83%
Pengajar		98.83%	90%
Ujian		92.87%	72%
Aplikasi		89.46%	88.8%

Non akademik		54.93%	75.9%
Other		75.33%	69.9%

#### 4. Kesimpulan

Berdasarkan dari pengkategorian dalam komentar Instagram Universitas Terbuka bahwa nilai akurasi tertinggi dari pengkategorian data komentar yang masih belum dilakukan *pre-processing* adalah kategori Pengajar pada algoritma KNN sebesar 98,97%, untuk nilai AUC tertinggi dari pengkategorian komentar adalah kategori modul pada algoritma Support Vector Machine (SVM) sebesar 94,60%, kemudian pada pengkategorian data komentar yang sudah dilakukan *pre-processing* untuk nilai akurasi tertinggi terdapat di kategori Pengajar sebesar 98.87% pada algoritma KNN dan nilai AUC tertinggi terdapat di kategori Modul sebesar 94.60% untuk kategori yang ditanyakan terbanyak adalah kategori Non akademik dengan jumlah 5490 komentar dari 10.000 komentar, untuk penelitian selanjutnya diharapkan dapat memperbanyak data komentar supaya nilai akurasi dan AUC lebih tinggi juga menambahkan beberapa algoritma lain seperti Neural Network, J48, dan algoritma lainnya.

#### 5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Lembaga kita.org dalam pengelolaan jurnal ilmiah serta Pak Windu Gata, Bu Rhini Fatmasari dan Pak Oky Kurniawan atas bantuannya dalam pembuatan jurnal ini sampai selesai, tak lupa teman-teman dari Magister Ilmu Komputer Universitas Nusa Mandiri.

#### 6. Daftar Pustaka

- [1] Dewi, S.S., 2018. *Hafal Mahir Materi Sejarah SMA/MA KELAS 11, 12, 13*. Gramedia Widiasarana Indonesia.
- [2] Helianthusonfri, J., 2020. *Instagram Marketing Untuk Pemula*. Elex Media Komputindo.
- [3] Saputra, D. 2020. Otomasi Pengkategorian Terhadap Keluhan Pelanggan Pada Twitter Operator Telekomunikasi Xl Axiata Di Indonesia, p. 120.
- [4] Arhami, and Nasir, M, 2020. *Data Mining- Algoritma dan Implementasi*. Penerbit Andi.
- [5] Ginting, R.U. and Wardoyo, R., 2019. *Sistem Evaluasi Kepuasan Mahasiswa Menggunakan Metode Fuzzy*. Zifatama Jawaara.
- [6] Putri, S.A., 2021. *Talend Data Preparation dan Talend Open Studio for Data Intregator untuk Prapengolahan Data*. Indego publisher.
- [7] Wahid, S.H., 2021. *A Tool Kit for Penulisan Karya Ilmiah (Skripsi dan Jurnal) Plus Analisis Data*. Q Media.
- [8] Jollyta, D., Siddik, M., Mawengkang, H. and Efendi, S., 2021. *Teknik Evaluasi Cluster Solusi Menggunakan Python Dan Rapidminer*. Deepublish.
- [9] M. Sadli, W. Fuadi, and Fajriana., 2019. *Aplikasi Penerapan Auto Matic Clustering dan Fuzzy Time Series*. Yayasan Kita Menulis.
- [10] Rahayu, W.I. and Saputra, M.H.K., 2020. *Penerapan Metode Naive Bayes dan Skala Likert Pada Aplikasi Prediksi Kelulusan Mahasiswa*. Bandung: Kreatif Industri Nusantara.
- [11] Werdiningsih, I., Kom, M., Nuqoba, B., Kom, M. and Muhammadun, S.S., 2020. *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press.
- [12] Muflikhah, L. and Mahmudy, W.F., 2021. *Machine Learning dalam Bioinformatika*. Universitas Brawijaya Press.

- [13] Jaggia, S., Kelly, A., Lertwachara, K. and Chen, L., 2020. Applying the CRISP-DM Framework for Teaching Business Analytics. *Decision Sciences Journal of Innovative Education*, 18(4), pp.612-634. DOI: 10.1111/dsji.12222.
- [14] Schröer, C., Kruse, F. and Gómez, J.M., 2021. A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, pp.526-534. DOI: 10.1016/j.procs.2021.01.199.
- [15] Nagashima, H. and Kato, Y., 2019, March. APREP-DM: a Framework for Automating the Pre-Processing of a Sensor Data Analysis based on CRISP-DM. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (pp. 555-560). IEEE. DOI: 10.1109/PERCOMW.2019.8730785.
- [16] Khumaidi, A., 2020. Data mining for predicting the amount of coffee production using CRISP-DM method. *Techno Nusa Mandiri: Journal of Computing and Information Technology*, 17(1), pp.1-8. DOI: 10.33480/techno.v17i1.1240.
- [17] Singh, V., Singh, A. and Joshi, K., 2022, January. Fair CRISP-DM: Embedding Fairness in Machine Learning (ML) Development Life Cycle. In *HICSS* (pp. 1-10). DOI: 10.24251/hicss.2022.190.