

# Analisis Sentimen Opini Publik pada Instagram mengenai *Covid-19* dengan SVM

Iin Kurniasari<sup>1</sup>, Kusrini<sup>2</sup>, Hanif Al Fatta<sup>3</sup>

<sup>1,2</sup>Magister Teknik Informatika, Universitas Amikom Yogyakarta

<sup>3</sup>Manajemen Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

E-mail: \*<sup>1</sup>[iinkurniasari@gmail.com](mailto:iinkurniasari@gmail.com), <sup>2</sup>[kusrini@amikom.ac.id](mailto:kusrini@amikom.ac.id), <sup>3</sup>[hanif.a@amikom.ac.id](mailto:hanif.a@amikom.ac.id)

**Abstrak** – Perkembangan teknologi dewasa ini mendorong masyarakat untuk selalu tanggap teknologi, terlebih di era pandemi covid-19 yang selalu mengedepankan *social distancing*. Media sosial digunakan sebagai suatu alat untuk menyampaikan opini masyarakat kepada khalayak. Dalam penelitian ini, penulis melakukan penelitian tentang opini masyarakat pada media sosial instagram dengan menggunakan *Support Vector Machine*. Setelah dilakukan uji akurasi dan presisi ternyata SVM belum sesuai digunakan sebagai algoritma yang dapat menangkap urutan karena susunan kata yang dibolak-balik meskipun maknanya berbeda tetap bermakna sama oleh mesin SVM, hal ini dibuktikan juga dengan jumlah akurasi yang kecil yaitu 59%. Sehingga diperlukan langkah untuk bisa diteliti dengan algoritma lain misalnya algoritma HRRN (*Highest Response Ratio Next*) atau LSTM (*Long Short-Term Memory*) yang memperhatikan urutan dan proses dengan rasio respon paling tinggi. Jika berdasarkan pendekatan ekstraksi fitur SVM dengan pendekatan *count vector*, *tf-idf word level*, *tf-idf ngram level* dan *tf-idf char level*. Dalam skenario ini nilai akurasi tertinggi terdapat pada perhitungan dengan menggunakan ekstraksi fitur *count vector* dan *tf-idf ngram level*.

**Kata Kunci** — analisis sentimen, instagram, Covid-19, SVM, media sosial

**Abstract** – Today's technological developments encourage people to always be responsive to technology, especially in the era of the Covid-19 pandemic which always prioritizes *social distancing*. Social media is used as a tool to convey public opinion to audiences. In this study, the authors conducted research on public opinion on social media Instagram using the *Support Vector Machine*. After testing the accuracy and precision, it turns out that SVM is not suitable for use as an algorithm that can capture sequences because the word arrangement is reversed even though the meaning is different, it still means the same by the SVM machine, this is also proven by a small amount of accuracy, namely 59%. So that steps are needed to be investigated with other algorithms such as the HRRN (*Highest Response Ratio Next*) algorithm or LSTM (*Long Short-Term Memory*) which takes into account the sequence and process with the highest response ratio. If it is based on the SVM feature extraction approach with the *count vector* approach, *tf-idf word level*, *tf-idf ngram level* and *tf-idf char level*. In this scenario the highest accuracy value is found in calculations using the *count vector* feature extraction and the *tf-idf ngram level*.

**Keywords** — sentiment analysis, instagram, covid-19, SVM, social media

## 1. PENDAHULUAN

Perkembangan teknologi digital yang sangat pesat saat ini mendorong masyarakat harus tanggap teknologi digital, tidak hanya masyarakat elit perkotaan atau pelaku usaha. Masyarakat pinggiran pun juga harus tanggap sehingga tidak tertindas oleh era digital itu sendiri. Di era digital komunikasi tidak harus dilakukan dengan tatap muka, terlebih pada masa Pandemi *Covid-19* yang mengedepankan *social distancing*. Hal ini dijadikan salah satu langkah pemerintah dalam menanggulangi penyebaran *pandemic Covid-19* dengan memanfaatkan teknologi digital berupa media sosial dalam upaya *social distancing*.

*Covid-19* merupakan sebuah ujian, ujian penguasa semesta terhadap penduduk dunia. *Covid-19* bermula di Kota Wuhan China pada akhir 2019 dan kini telah menjangkiti lebih dari 100

negara. Lebih dari 100.000 orang di dunia dinyatakan positif *Covid-19*. Beberapa kasus *Covid-19* di China dilaporkan menurun, tetapi lonjakan di Korea Selatan, Italia, Iran, Indonesia dan beberapa Negara lain. [21]

Media sosial digunakan sebagai media dalam penyampaian informasi khususnya dalam masa pandemi *Covid-19* yang mengedepankan *social distancing*. Salah satu media di Kota Kediri yang aktif dalam menginformasikan *Covid-19* adalah AG243. AG243 merupakan akun resmi yang dimiliki oleh Radio Andika FM yang terletak di Jl. Semeru 243 Kediri. AG adalah symbol Plat Nomor Kendaraan Kediri sedangkan 243 adalah nomor alamat Radio Andika. Dalam kesehariannya Radio Andika FM yang terhubung dengan Radio Suara Surabaya (SS) selalu menginformasikan kondisi keadaan jalan raya di area Karisedanan Kediri (Kediri, Madiun, Tulungagung, Trenggalek, Blitar dan sekitarnya).

AG 243 memiliki pengikut *facebook* 497.554, dan 78.300 pengikut IG dengan 12.000 postingan. Hampir setiap jam update info selalu dilakukan. Sehingga tidak heran dengan masa sekian banyak, setiap kali update informasi selalu banyak yang berkomentar. Khususnya dalam pandemi *Covid-19* saat ini. Begitu banyak komentar baik positif, negatif ataupun netral yang diberikan khalayak terhadap update informasi yang disampaikan.

*Support Vector Machine* dalam melakukan proses klasifikasi dengan nilai akurasi yang sangat signifikan yaitu 90% berdasarkan penelitian Wanda [17]. Tetapi berdasarkan penelitian Krusnoslav, dkk [18] dalam *Comparison of Naïve Bayes and SVM Classifiers in Categorization of Concept Maps*, SVM hanya memiliki tingkat akurasi 63,74%. Sehingga berdasarkan fenomena di atas, penulis tertarik meneliti dan berinisiatif mengangkat judul “Analisis Sentimen Opini Publik Mengenai *Covid-19* Pada *Facebook* dan *Instagram* Menggunakan *Support Vector Machine*” dalam penelitiannya. Sehingga berdasarkan fenomena di atas, penulis tertarik meneliti dan berinisiatif mengangkat judul “Analisis Sentimen Opini Publik pada *Instagram* mengenai *Covid-19* dengan *Support Vector Machine*” dalam penelitiannya.

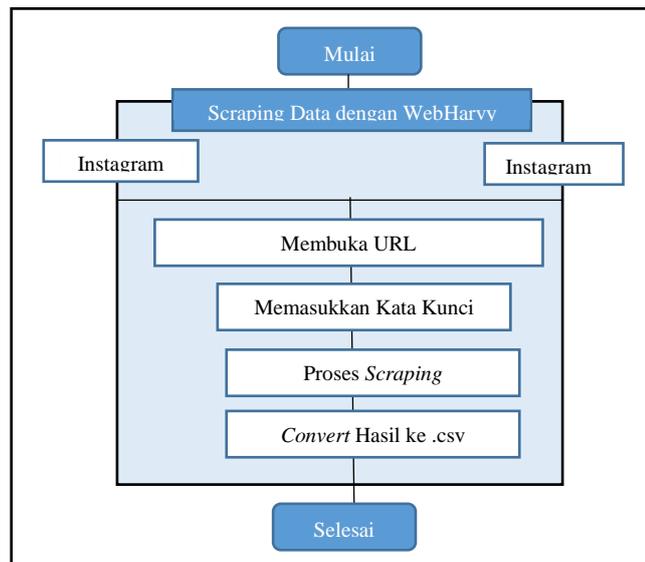
## 2. METODE PENELITIAN

### 2.1. Jenis, Sifat dan Pendekatan Penelitian

Jenis penelitian yang digunakan adalah penelitian eksperimen dan studi kasus. Penelitian eksperimen adalah penelitian yang dilakukan dengan menerapkan serangkaian tindakan untuk membuktikan suatu konsep [11]. Dalam hal ini penulis melakukan eksperimen melalui proses *scraping* data *instagram* kemudian mengklasifikasikannya menggunakan *Support Vector Machine*. Studi kasus dilaksanakan pada media sosial *instagram* akun AG243. Penelitian yang dilakukan penulis bersifat deskriptif, dimana data dijelaskan dalam deskriptif angka dan tabel atau diagram. Setelah mengolah data, penulis menganalisisnya dengan pendekatan kuantitatif yang dijelaskan dengan hasil penghitungan angka dan tabel atau diagram.

### 2.2 Metode Pengumpulan Data

---



Gambar 1. Metode Pengumpulan Data

Dalam pengumpulan data atau *scraping* menggunakan *Webharvy* dengan langkah membuka URL instagram kemudian memasukkan kata kunci. Setelah kata kunci dimasukkan maka akan melakukan proses *scraping*. Setelah proses selesai maka akan otomatis bisa ter-convert ke .csv.

### 2.3 Analisis Data

Evaluasi dilakukan dengan uji coba fungsional terhadap layanan yang disediakan oleh sistem satu per satu dan uji coba terhadap tingkat akurasi klasifikasi yang dihasilkan oleh sistem. Uji coba fungsional terhadap layanan yang disediakan oleh sistem yang memiliki tujuan untuk mengetahui hasil dari setiap proses dalam klasifikasi telah sesuai dengan analisis dan perancangan yang dilakukan. Uji coba terhadap tingkat akurasi klasifikasi yang dihasilkan bertujuan untuk mengetahui nilai *accuracy* dan *precision* klasifikasi menggunakan algoritma SVM berdasarkan kesesuaian hasil klasifikasi dan label. Pengujian akurasi sistem Pengujian akurasi sistem dilakukan dengan cara menghitung nilai dari *accuracy* dan *precision*. Persaman 1 digunakan untuk menghitung nilai *accuracy* sebagai berikut.

$$Acc = \frac{\sum v}{n} \times 100 \dots\dots\dots (1)$$

Keterangan:

- $v$  : Jumlah data benar
- $n$  : Jumlah dokumen

Sedangkan untuk menghitung *precision* menggunakan rumus sebagai berikut:

$$Prec = \frac{\sum vP}{nP} \times 100 \dots\dots\dots (2)$$

Keterangan:

- $vP$  : Jumlah data positif, negatif atau netral benar
- $nP$  : Jumlah dokumen positif, negatif atau netral

Pengujian dilakukan dengan jumlah data training yang berbeda. Pada setiap data training akan dilakukan 5 kali pengujian, sehingga tingkat akurasi sistem akan dihitung berdasarkan rata-rata dari setiap jumlah data training.

Data latih atau data training digunakan algoritma *Support Vector Machine* untuk mempelajari pola data berdasarkan ciri-ciri pada masing-masing kelas. Hasil pembelajaran tersebut kemudian diuji menggunakan data uji untuk mengetahui tingkat akurasi dalam memprediksi kelas pada data baru, proses ini disebut sebagai *machine learning*. Disini penulis menggunakan 1027 *datasheet* komentar positif, negatif dan netral dari facebook dan 1000 *datasheet* komentar positif,

negatif dan netral pada instagram yang dibagi menjadi 5 (lima) kali pengujian dengan hasil akurasi klasifikasi algoritma *Support Vector Machine* menggunakan 60%, 70%, 80% dan 90% data training.

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini menerapkan *text pre processing* dalam melakukan pengolahan data yang akan digunakan dalam proses analisis *sentiment*. Data yang diperoleh dari opini pengguna Instagram memiliki bentuk yang tidak terstruktur karena masih banyak mengandung kata-kata yang tidak ada maknanya sehingga sulit untuk dilakukan analisis secara langsung. Sehingga hal ini perlu dilakukan seleksi dalam mempermudah analisis data dan data yang digunakan lebih terstruktur serta memiliki keseragaman untuk proses *mining* lebih lanjut. Penulis menggunakan script *Python* untuk mempermudah dalam melakukan proses preprocessing. Pada tahap *text preprocessing*, metode *text mining* akan dilakukan untuk pembersihan data diantaranya adalah *case folding*, *tokenizing*, *filtering* atau *stopword* removal dan *stemming*.

#### 3.1 Case Folding

Tahap *preprocessing* yang pertama adalah *case folding*. *Case folding* adalah proses mengubah semua karakter dalam dokumen menjadi tema yang sama. Contohnya, dalam dokumen teks tidak semua penggunaan huruf kapital dilakukan dengan konsisten. Disini penulis menggunakan script *python* di bawah ini seperti pada gambar 2

```
train_X = train_csv['0'].str.lower()
train_y = train_csv['1'].str.lower()
test_X = test_csv['0'].str.lower()
test_y = test_csv['1'].str.lower()
```

Gambar 2. *case folding*

Peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil). Dalam *case folding*, hanya karakter huruf a sampai z yang diterima, sedangkan karakter lain tidak diterima dan dianggap pembatas.

#### 3.2 Tokenizing

Tahap selanjutnya yaitu *tokenizing* atau bisa juga disebut *parsing*. *Tokenizing* yang script-nya pada gambar 3 adalah proses pemotongan dokumen menjadi bagian-bagian kata yang disebut token. Spasi digunakan untuk memisahkan antar kata tersebut. Sedangkan kata-kata yang tidak dibutuhkan akan dihilangkan melalui proses *filtering* dari hasil *tokenizing*.

```
for k in range(len(train_X)):
    train_X[k] = nltk.tokenize.word_tokenize(train_X[k])

for k in range(len(test_X)):
    test_X[k] = nltk.tokenize.word_tokenize(test_X[k])
```

Gambar 3. *tokenizing*

#### 3.3 Stopword Removal atau Filtering

Pada proses ini menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Kata-kata yang dihilangkan yaitu: - penghubung antar kata, seperti: dan, atau, serta - preposisi, seperti: di, ke, pada - kata-kata yang tidak diinginkan. Dalam hal ini script *Python* yang digunakan adalah seperti pada gambar 4

```
factoryStopRemove = StopWordRemoverFactory()
stopword = factoryStopRemove.create_stop_word_remover()
for z in range(len(train_X)):
    train_X[z] = stopword.remove(train_X[z])

for z in range(len(test_X)):
    test_X[z] = stopword.remove(test_X[z])
```

Gambar 4. Proses *Filtering* atau *Stopword Removal*

3.4 *Stemming*

Setelah proses *tokenizing* dan *filtering* selanjutnya adalah proses *stemming* yang digunakan untuk mencari *root* dari kata hasil proses *filtering*. Contoh proses *stemming* dapat dilihat pada gambar 5 di bawah ini. Dalam proses ini ada pengembalian kata berimbuhan ke kata dasarnya (*root*).

```
factoryStemm = StemmerFactory()
stemmer = factoryStemm.create_stemmer()
for y in range(len(train_X)):
    train_X[y] = stemmer.stem(train_X[y])
for y in range(len(test_X)):
    test_X[y] = stemmer.stem(test_X[y])
```

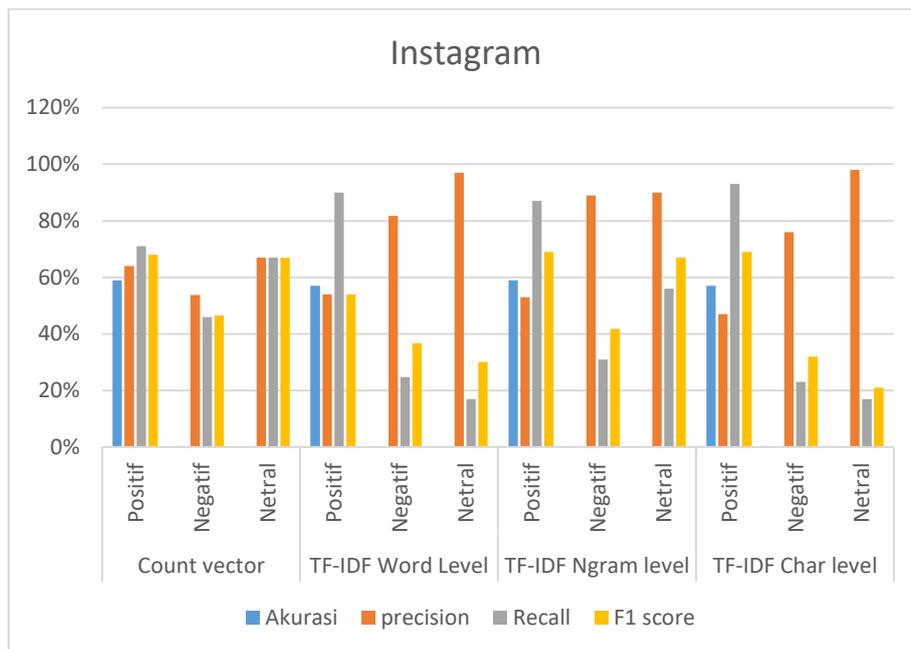
Gambar 5. *Stemming*

3.5 Uji Coba dan Evaluasi

Tabel 1. Pengujian *Count Vector* dan TF IDF pada Instagram

Nama	Komentar	Akurasi	precision	Recall	F1 score
<i>Count vector</i>	Positif	59%	64%	71%	68%
	Negatif		54%	46%	47%
	Netral		67%	67%	67%
<i>TF-IDF Word Level</i>	Positif	57%	54%	90%	54%
	Negatif		82%	25%	37%
	Netral		97%	17%	30%
<i>TF-IDF Ngram level</i>	Positif	59%	53%	87%	69%
	Negatif		89%	31%	42%
	Netral		90%	56%	67%
<i>TF-IDF Char level</i>	Positif	57%	47%	93%	69%
	Negatif		76%	23%	32%
	Netral		98%	17%	21%

Pada tabel 1 dapat diketahui bahwa penghitungan pada media sosial instagram dengan ekstraksi *Count Vector* terdapat nilai akurasi sebesar 59% dengan tingkat presisi komentar positif, negatif dan netral adalah 64%, 54% dan 67%. Nilai *recall* 71%, 46% dan 67% serta *F1 score* 68%, 47% dan 67%. Sedangkan dengan *TF IDF Word Level* nilai akurasi sebesar 57% dengan tingkat presisi komentar positif, negatif dan netral adalah 54%, 82% dan 97%. Nilai *recall* 90%, 25% dan 17% serta *F1 score* 54%, 37% dan 30%. Dengan *TF-IDF N-Gram Level* memiliki nilai akurasi sebesar 59% dengan tingkat presisi komentar positif, negatif dan netral adalah 53%, 89% dan 90%. Nilai *recall* 87%, 31% dan 56% serta *F1 score* 69%, 42% dan 67%. Serta nilai akurasi pada *TF-IDF Char Level* sebesar 57% dengan tingkat presisi komentar positif, negatif dan netral adalah 47%, 76% dan 98%. Nilai *recall* 93%, 23% dan 17% serta *F1 score* 69%, 32% dan 21%.



Gambar 6. Diagram Perbandingan Nilai Akurasi dengan Pendekatan Fitur *TF-IDF* dan *Count Vector*

Pada gambar 6 pada media instagram dapat dilihat bahwa nilai akurasi tertinggi terdapat pada *Count Vector* sedangkan nilai presisi tertinggi terdapat pada *TF-IDF* baik *word level*, *count level* dan *char level*. Pada pengujian dengan klasifikasi *Support Vector Machine* dengan pendekatan *Count Vector* dan *TF-IDF* dapat diketahui bahwa nilai akurasi pada instagram adalah 59%.

#### 4. KESIMPULAN

Kesimpulan yang dihasilkan pada uji coba dan evaluasi penelitian ini diantaranya adalah:

- Jika berdasarkan pendekatan ekstraksi fitur SVM dengan pendekatan *count vector*, *TF-IDF word level*, *TF-IDF ngram level* dan *TF-IDF char level*. Dalam skenario ini nilai akurasi tertinggi terdapat pada perhitungan dengan menggunakan ekstraksi fitur *count vector* dan *TF-IDF ngram level* dengan nilai akurasi 59% pada instagram.
- Support vector machine* belum bisa digunakan sebagai algoritma yang menangkap urutan karena kata yang dibolak-balik meskipun maknanya berbeda tetap diakui sama oleh mesin dengan klasifikasi *SVM*, hal ini juga terbukti dengan jumlah akurasi yang sangat kecil.

#### 5. SARAN

Diperlukan algoritma klasifikasi yang bisa menangkap urutan kata dengan makna yang sesuai misalnya dengan menggunakan algoritma HRRN (*Highest Response Ratio Next*) yang memperhatikan proses dengan rasio respon paling tinggi atau LSTM (*Long Short-Term Memory*).

#### DAFTAR PUSTAKA

- [1] Abram, C. 2016. *Facebook For Dummies: Edition 6 (Vol. 6)*. John Wiley & Sons.
- [2] Faisal, M. R. 2017. *Seri Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R. Indonesia*. Net Developer Community.

- [3] Feldman, R., & Sanger, J. 2007. *The Text Mining Handbook Advanced Approaches In Analyzing Unstructured Data*. New York : Cambridge University Press. (Jamil, 2017).
- [4] Kusrini. 2006. *Sistem Pakar Teori dan Aplikasi*. Andi Offset
- [5] ----- . 2008. *Aplikasi sistem Pakar Lanjut*. Andi Offset
- [6] Laksana, E. A., & Sulianta, F. 2017. *Analisis Dan Studi Komparatif Algoritma Klasifikasi Genre Musik*. 67-72
- [7] Lee, L., & Pang, B. 2008. *Opinion Mining and Sentiment Analysis. Foundation and Trends in Information Retrieval*, 2(1-2): 1-135.
- [8] Liu, B. 2012. *Sentiment Analysis and Subjectivity. Synthesis Lectures on Human Language Technologies*. USA: Morgan & Claypool Publishers.
- [9] Mohri, M., Rostamizadeh, A., & Talwalkar, A. 2012. *Foundations of Machine Learning*. MIT Press
- [10] Prasetyo, Eko. 2014. *Data Mining Mengolah Data Menjadi Informasi menggunakan Matlab*. Yogyakarta: ANDI
- [11] Sugiyono. 2015. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. 22 ed. Bandung: Alfabeta.
- [12] Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Bandung: Informatika Bandung.
- [13] Tan, P., Steinbach, M., & Karpatne, A. 2006. *Introduction To Data Mining*. USA: Addison-Wesley.
- [14] Turban, E., Aronson, J., & Liang, T. 2005. *Decision Support System And Intelligent System*. Upper Saddle River, New Jersey USA: Prentice Hall.
- [15] ----- 2005. *Service Quality Satisfaction*. Yogyakarta: Andi.
- [16] Alhajji, Moehammad. 2020. *Sentiment Analysis of Tweets in Saudi Arabia Regarding Governmental Preventive Measures to Contain COVID-19*. preprints
- [17] Athira, Wanda. 2018. *Analisis Sentimen Cyberbullying Pada Komentar Instagram Dengan Metode Klasifikasi Support Vector Machine*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer
- [18] Budi, Eko. 2019. *Analisis Sentimen Calon Presiden Indonesia 2019. Berdasarkan Komentar Publik Di Facebook*. Jurnal Eksplora Informatika
- [19] Burhanudin, Ikhsan. 2020. *Ancaman Krisis Ekonomi Global Dari Dampak Penyebaran Virus Corona (Covid 19)*. Akmen
- [20] Fatmawati. 2017. *Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus: Akun Fb Group Iraise Helpdesk)*. Jurnal CoreIT
- [21] Fauzi, Akhmad. 2019. *Sentimen Analisis Berinternet Pada Media Sosial Dengan Menggunakan Algoritma Bayes*. Jurnal Informatika
- [22] Janoušek, V., Moyen, J.-F., Martin, H., Erban, V., & Farrow, C. 2015. *Geochemical Modelling of Igneous Processes—Principles And Recipes in R Language: Bringing the Power of R to a Geochemical Community*. Springer Geochemistry
- [23] Lappeman, James. 2020. *Studying Social Media Sentiment Using Human Validated*. Elsevier

- [24] Maulana, Akbar. 2018. *Analisis Sentimen Media Sosial Univ.Amikom Yogyakarta Sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi SVM*. Seminar Nasional Teknologi Informasi dan Multimedia
  - [25] Murty, M., & Raghava, R. 2016. *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks*.Springer.
  - [26] Puspita. 2014. *Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi SD Di Kab. Magelang*. Jurnal Gaussian
  - [27] Putri, D.U.K. 2016. *Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten*. Program Studi Ilmu Komputer FMIPA UGM Yogyakarta.
  - [28] Salam, Abu. 2018. *Analisis Sentimen Data Komentar Sosial Media Fb Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)*. Prosiding Sintak
  - [29] Saraswati, N.S. 2011. *Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis*. Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.
  - [30] Krunoslav, dkk. 2018. *Comparison of Naïve Bayes and SVM Classifiers in Categorization of Concept Map*
-