

Optimasi *Customer Experience* Menggunakan *Sentiment Analysis* dengan Algoritma Boyer-Moore

Aisyah Farras Aqila - 13519054

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jalan Ganesha 10 Bandung

E-mail (gmail): aisyahfarrasaqila@gmail.com · 13519054@std.stei.itb.ac.id

Abstrak—*Sentiment analysis* merupakan salah satu teknik yang dapat digunakan untuk menganalisis *customer experience* dari sebuah bisnis dengan melihat sentimen konsumen mengenai merk tertentu. Pendekatan yang digunakan adalah *sentiment analysis* berbasis kamus. Data sentimen merupakan sebuah kalimat dengan bahasa natural. Penggunaan *string matching* khususnya dengan algoritma Boyer-Moore digunakan untuk mendeteksi sentimen sesuai dengan kata yang ada pada kalimat dan kamus. Teknik ini sangat sederhana dan mudah untuk digunakan namun tidak memberikan akurasi yang baik untuk beberapa kasus.

Kata kunci: *customer experience*, *sentiment analysis*, *string matching*

I. PENDAHULUAN

Dalam dunia bisnis, *customer experience* menjadi suatu hal yang sangat signifikan untuk bersaing di pasar. Menurut survey yang dilakukan oleh Accenture, 64% *customer* yang beralih dari satu merk ke merk yang lainnya berpendapat bahwa alasan dari peralihan tersebut adalah produk, jasa, atau pengalaman yang lebih relevan. Untuk mendapatkan *customer experience* yang baik tentunya produk atau jasa yang diberikan harus sesuai dengan apa yang diharapkan *customer*. Accenture berpendapat bahwa untuk mencapai hal tersebut, dibutuhkan data yang sesuai. Dalam hal ini data yang dimaksud adalah data *feedback* dari *customer* yang akurat.

Untuk menghasilkan *insight* dari data yang disebutkan di atas, dapat digunakan suatu teknik yang disebut *sentiment analysis*. Dengan pendekatan ini, produsen dapat mengetahui seberapa positif atau negatifnya *feedback* dari *customer* atas produk mereka. Hal ini kemudian dapat dianalisis lebih lanjut untuk dapat mengembangkan produk agar lebih sesuai dengan harapan pasar.

Pada *sentiment analysis*, data yang digunakan adalah data dalam bentuk bahasa natural. Natural Language Processing (NLP) merupakan sebuah bidang yang mempelajari hubungan antara komputer dan bahasa sehari-hari yang digunakan manusia. Salah satu implementasinya adalah *string matching*. Pada penelitian ini digunakan algoritma *string matching* yaitu Boyer-Moore karena algoritma ini cukup baik untuk memproses bahasa sehari-hari.

II. DASAR TEORI

A. *Customer Experience*

1. Definisi *Customer Experience*

Customer experience adalah respons internal dan subjektif dari konsumen secara langsung atau tidak langsung terhadap sebuah perusahaan. Hubungan langsung biasanya terjadi pada saat pembelian, pemakaian, dan layanan yang diinisiasi oleh konsumen. Hubungan tidak langsung terjadi ketika terlibat sebuah hubungan yang tidak direncanakan yang merepresentasikan produk, jasa, atau merk. Biasanya dalam bentuk rekomendasi atau kritik dari mulut ke mulut, periklanan, berita, *review*, dan lain-lain.

2. Pentingnya *Customer Experience*

- Customer Experience* sangat dibutuhkan pada zaman sekarang. Konsumen yang sangat puas menghasilkan *revenue* 2,6 kali lebih banyak dibandingkan konsumen yang sekedar puas saja.
- Konsumen baru lebih mungkin untuk menjadi konsumen loyal.
- Sebuah komunitas yang mendukung/merekomendasikan produk merupakan alat yang sangat efektif. 84% konsumen sudah tidak percaya dengan iklan, mereka lebih memilih untuk melihat validasi pihak ketiga ketika ingin membeli suatu produk.
- Menjadi merk yang dominan bukan lagi sebuah masalah. Konsumen tidak lagi hanya membandingkan produk dan harga, namun juga layanan dan *user experience*.
- Konten yang personal merupakan hal yang sangat penting. Kuncinya adalah dalam membuat konsumen merasa spesial dan unik.

3. Cara Mendapatkan Data *Customer Experience*

Data mengenai *customer experience* didapat dari interaksi konsumen dengan produk/perusahaan. Di

antaranya adalah data mengenai transaksi, profil, sikap, preferensi produk, aktivitas sosial media, sentimen, dan lokasi. Selain itu juga bisa digunakan metode survey kepada konsumen untuk mendapatkan *feedback* secara langsung.

B. Sentiment Analysis

1. Definisi Sentiment Analysis

Sentiment analysis adalah *contextual mining* dari teks yang mengidentifikasi dan mengekstrak informasi subjektif dari sumber dengan guna membantu bisnis mengetahui sentimen sosial dari merknya. Percakapan konsumen dapat diklasifikasi berdasarkan aspek utama dari sebuah produk yang konsumen pedulikan dan niat serta reaksi konsumen terhadap aspek-aspek tersebut.

2. Langkah-Langkah Sentiment Analysis

- a. Koleksi data: tahap pertama dalam melakukan *sentiment analysis* adalah mengumpulkan data dari konsumen yang terdapat pada *blog*, forum, dan sosial media. Karena data tersebut tidak terorganisir dan menggunakan konteks penulisan yang berbeda-beda, dibutuhkan alat untuk melakukan analisis teks dan *natural language processing* untuk mengekstraknya.
- b. Persiapan teks: terdiri dari pembersihan data yang sudah diekstrak sebelum dianalisis. Konten yang tidak relevan akan dihilangkan pada tahap ini.
- c. Deteksi sentimen: tahap ini dilakukan untuk menyaring kalimat-kalimat faktual sehingga yang lanjut ke tahap selanjutnya hanya kalimat-kalimat subjektif saja.
- d. Klasifikasi sentimen: kalimat-kalimat subjektif akan diklasifikasi menjadi positif, negatif, atau netral.
- e. Presentasi hasil: objektif utama dari *sentiment analysis* adalah mengubah teks tidak berstruktur menjadi informasi yang bermanfaat. Ketika analisis sudah dilaksanakan, hasilnya akan divisualisasikan dengan *pie chart*, *bar chart*, ataupun *line graph*.

3. Pendekatan Sentiment Analysis

Tabel berikut menunjukkan beberapa pendekatan terhadap *sentiment analysis*.

Tabel 2.1 Pendekatan *Sentiment Analysis*

SENTIMENT CLASSIFICATION APPROACHES		FEATURES/TECNQUES	ADVANTAGES AND LIMITATIONS
Machine learning	Bayesian Networks Naive Bayes Classification Maximum Entropy Neural	Term presence and frequency Part of speech information	ADVANTAGES the ability to adapt and create trained models for specific purposes

	Networks Support Vector Machine	Negations Opinion words and phrases	and contexts LIMITATIONS the low applicability to new data because it is necessary the availability of labeled data that could be costly or even prohibitive
Lexicon based	Dictionary based approach Novel Machine Learning Approach Corpus based approach Ensemble Approaches	Manual construction, Corpus-based Dictionary-based	ADVANTAGES wider term coverage LIMITATIONS finite number of words in the lexicons and the assignation of a fixed sentiment orientation and score to words
Hybrid	Machine learning Lexicon based	Sentiment lexicon constructed using public resources for initial sentiment detection Sentiment words as features in machine learning method	ADVANTAGES lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain LIMITATIONS noisy reviews

Sumber: D’Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications* (0975 – 8887), 125(13), 29-31.

C. String Matching/Pattern Matching

1. Definisi String Matching/Pattern Matching

Jika T (teks) adalah *string* dengan panjang n karakter dan P (*pattern*) adalah *string* dengan panjang m karakter (asumsi $m \lll n$), maka *string matching* adalah proses mencari lokasi pertama pada teks yang bersesuaian dengan *pattern*.

Terdapat dua konsep dari *string*, yaitu *prefix* dan *suffix*. *Prefix* adalah *substring* $S[0..k]$ dengan k adalah indeks manapun di antara 0 sampai m-1. Pada *string* “andrew” semua kemungkinan *prefix* adalah: “a”, “an”, “and”, “andr”, “andre”, dan “andrew”. Sedangkan *suffix* adalah *substring* $S[k..m-1]$. Semua

kemungkinan *suffix* dari *string* tersebut adalah: “w”, “ew”, “rew”, “drew”, “ndrew”, dan “andrew”.

2. Jenis-Jenis Algoritma String Matching

a. Brute Force Algorithm

Algoritma *brute force* pada *string matching* dilakukan dengan membandingkan setiap posisi pada teks T untuk melihat apakah *pattern* P mulai pada posisi tersebut.

Teks: NOBODY NOTICED HIM

Pattern: NOT

NOBODY **NOT**ICED HIM

1 NOT

2 NOT

3 NOT

4 NOT

5 NOT

6 NOT

7 NOT

8 **NOT**

Gambar 2.1 Brute Force Algorithm

Sumber:

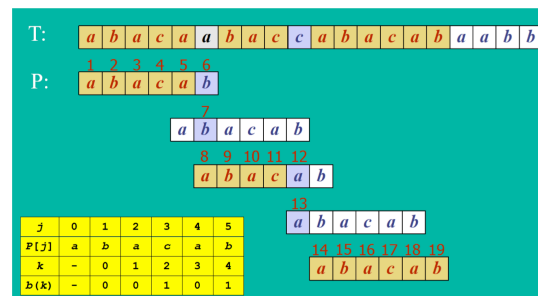
<https://informatika.stei.itb.ac.id/~rinaldi.munir/Smik/2020-2021/Pencocokan-string-2021.pdf>

Pada kasus terburuk, algoritma ini memiliki kompleksitas waktu $O(mn)$. Pada kasus terbaik kompleksitas waktu adalah $O(n)$. Sedangkan pada kasus rata-rata kompleksitas waktu adalah $O(m+n)$.

b. Knutt-Morris-Pratt Algorithm

Algoritma Knuth Morris-Pratt mencari *pattern* pada teks dari kiri ke kanan seperti *brute force* namun dengan pergerakan *pattern* yang lebih pintar. Jika *mismatch* terjadi pada *pattern* $P[j]$ dengan $T[i] \neq P[j]$, pergeseran paling banyak pada *pattern* adalah sebesar *prefix* terbesar $P[0..j-1]$ yang juga merupakan *suffix* $P[1..j-1]$.

Algoritma ini membutuhkan *border function* sebelum pencarian *pattern* dilaksanakan. Fungsi ini berguna untuk menemukan ukuran *prefix* $P[0..k]$ terbesar yang juga merupakan *suffix* $P[1..k]$ dengan k adalah $j-1$ dan j adalah indeks terjadinya *mismatch* pada *pattern*.



Gambar 2.2 Knutt-Morris-Pratt Algorithm

Sumber:

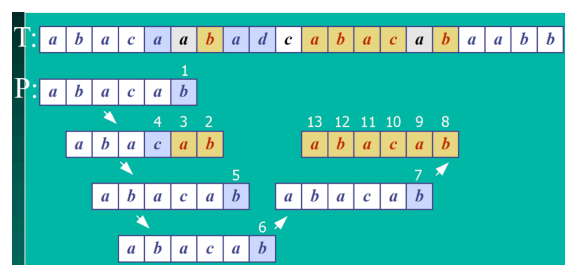
<https://informatika.stei.itb.ac.id/~rinaldi.munir/Smik/2020-2021/Pencocokan-string-2021.pdf>

Algoritma ini baik untuk memproses *file* yang sangat besar namun kurang baik ketika ukuran *alphabet* besar karena kemungkinan *mismatch* lebih besar. Sehingga algoritma ini kurang efektif untuk memproses bahasa natural.

c. Boyer-Moore Algorithm

Algoritma Boyer-Moore diimplementasikan berdasarkan dua teknik. Yang pertama adalah teknik *looking-glass*, yaitu mencari *pattern* P di teks T dengan berjalan mundur pada P, mulai dari indeks terakhir. Yang kedua adalah teknik *character-jump*, yaitu jika *mismatch* terjadi pada $T[i] = x$ dan karakter pada *pattern* $P[j]$ tidak sama dengan $T[i]$, akan ada tiga skenario. Yang pertama jika x terakhir pada *pattern* berada di kiri indeks *pattern*, maka P digerakkan ke kanan sehingga x terakhir pada P sesuai dengan $T[i]$. Skenario kedua yaitu ketika x terakhir berada di kanan indeks *pattern*, maka P digerakkan ke kanan sebanyak 1 karakter ke $T[i+1]$. Skenario ketiga adalah jika x tidak ada pada *pattern*, maka P digeser ke kanan sehingga $P[0]$ bersesuaian dengan $T[i+1]$.

Algoritma ini membutuhkan *last occurrence function* sebelum pencarian *pattern* dilaksanakan. Fungsi ini berguna untuk mencari indeks terbesar pada *pattern* dari masing-masing karakter yang ada pada teks.



Gambar 2.3 Boyer-Moore Algorithm

Sumber:

Algoritma ini baik untuk memproses teks dengan *alphabet* besar sehingga cocok untuk digunakan dalam pemrosesan bahasa natural.

III. APLIKASI ALGORITMA BOYER-MOORE PADA SENTIMENT ANALYSIS

Sentiment Analysis dengan *string matching* akan menggunakan algoritma Boyer-Moore karena data teks berada dalam bahasa natural dan algoritma tersebut efektif untuk memproses teks dengan bahasa natural.

Pendekatan yang dilakukan pada proses ini adalah *sentiment analysis* berbasis lexicon, lebih tepatnya berbasis kamus atau *dictionary-based*. Hal ini dilakukan karena untuk membandingkan teks dengan suatu *pattern*, kita sudah harus memiliki data mengenai *pattern*-nya terlebih dahulu. Sehingga kamus berisi seluruh *pattern* yang ingin dibandingkan dengan teks *input*.

A. Tahap Preprocessing

Pada tahap *preprocessing*, akan disimpan semua *pattern* yaitu kata-kata yang menggambarkan suatu sentimen. Masing-masing kata tersebut akan diklasifikasikan sebagai kata positif atau negatif. Tabel di bawah merupakan contoh dari data kamus *pattern*.

Tabel 3.1 Kamus *Pattern*

Kata	Tipe
Suka	Positif
Bagus	Positif
Menarik	Positif
Jelek	Negatif
Membosankan	Negatif
Cepat rusak	Negatif

B. Tahap Persiapan Teks

Pada tahap ini, teks akan diekstrak dari sumber sesuai dengan dengan langkah-langkah yang sudah dijelaskan pada bagian II. Berikut merupakan contoh kalimat-kalimat dari sumber untuk melakukan *sentiment analysis* terhadap merk A.

1. Produk buatan merk A sangat menarik.
2. Merk A berdiri pada tahun 1967.
3. Keluaran baru merk A cepat rusak. Lebih mending beli yang lama karena lebih bagus.
4. Tidak usah nonton film B, sangat membosankan.
5. Saya sangat suka inovasi yang dilakukan merk A.
6. Aku tidak merekomendasikan produk dari merk A karena tidak bagus.
7. Kualitas merk A sangat buruk.

Kemudian setelah dilakukan pembersihan data dan deteksi sentimen, yaitu hanya memproses kalimat-kalimat yang subjektif saja, maka data kalimat akan menjadi seperti di bawah ini.

1. Produk buatan merk A sangat menarik.
2. Keluaran baru merk A cepat rusak. Lebih mending beli yang lama karena lebih bagus.
3. Saya sangat suka inovasi yang dilakukan merk A.
4. Aku tidak merekomendasikan produk dari merk A karena tidak bagus.
5. Kualitas merk A sangat buruk.

C. Tahap Klasifikasi Sentimen dengan String Matching

String matching digunakan pada tahap deteksi sentimen. Setiap kalimat akan diperiksa dengan setiap *pattern* dan akan dihitung jumlah kata positif dan negatifnya. Kalimat yang memiliki kata positif lebih banyak akan diklasifikasikan sebagai kalimat bersentimen positif sedangkan kalimat dengan kata negatif lebih banyak akan diklasifikasikan sebagai kalimat bersentimen negatif. Kalimat yang memiliki jumlah kata positif dan negatif sama akan diklasifikasikan sebagai kalimat bersentimen netral.

Tabel di bawah menunjukkan klasifikasi sentimen untuk setiap kalimat.

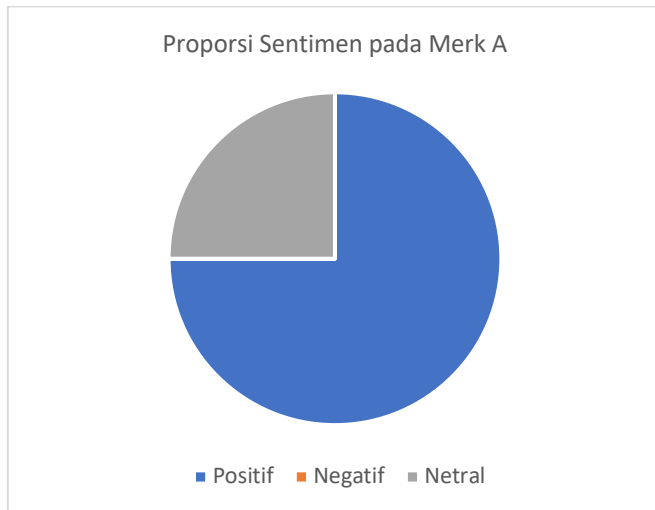
Tabel 3.2 Klasifikasi Sentimen

Nomor	Positif	Negatif	Sentimen
1	1	0	Positif
2	1	1	Netral
3	1	0	Positif
4	1	0	Positif
5	0	0	Netral

D. Visualisasi Data Kesimpulan

Setelah klasifikasi sentimen sudah dilakukan, maka dapat dilihat kesimpulan dari sentimen konsumen terhadap merk A dengan *pie chart* di bawah.

Grafik 3.1 Proporsi Sentimen pada Merk A



Dari visualisasi tersebut dapat terlihat bahwa 60% sentimen konsumen adalah sentimen positif dan 40% adalah sentimen netral. Merk A dapat menyimpulkan bahwa kondisi produknya cukup baik di pasar sehingga tidak perlu melakukan evaluasi atau perubahan yang signifikan.

IV. ANALISIS PENGGUNAAN ALGORITMA BOYER-MOORE PADA SENTIMENT ANALYSIS

Penggunaan *string matching* khususnya dengan algoritma Boyer-Moore untuk melakukan *sentiment analysis* sangat simpel dan mudah untuk dilakukan. Namun terdapat beberapa kekurangan yang membuat analisis kurang akurat sehingga *insight* yang diberikan dapat melenceng atau kurang lengkap.

Pada kalimat nomor 1 dan 3 analisis dengan teknik ini memberikan hasil yang akurat. Dapat kita lihat bahwa kedua kalimat tersebut menggambarkan kepuasan konsumen akan merk A dan kedua kalimat tersebut terdeteksi sebagai kalimat bersentimen positif.

Kalimat nomor 2 menunjukkan sentimen netral karena terdapat jumlah opini positif dan negatif yang sama. Hal ini tidak masalah jika analisis hanya dilakukan untuk melihat sentimen umum dari konsumen. Namun kinerja dapat ditingkatkan jika analisis dapat melihat bahwa kalimat mengandung sentimen positif/negatif terhadap suatu aspek tertentu. Penggunaan *string matching* sendiri tidak dapat melakukan hal tersebut, dibutuhkan algoritma tambahan untuk menyimpan data-data aspek dari suatu produk/merk dan kaitannya dengan opini pada suatu kalimat.

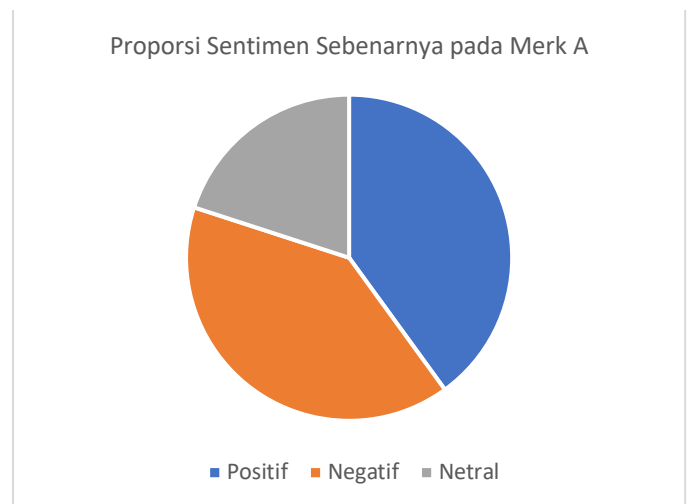
Pada kalimat nomor 4 terdapat satu kata positif sehingga kalimat ini diklasifikasikan sebagai kalimat bersentimen positif. Padahal jika kita lihat, terdapat kata negasi sebelum kata "bagus" sehingga sebenarnya kalimat ini merupakan kalimat bersentimen negatif. Hal ini terjadi karena dengan pendekatan *string matching* sederhana, tidak dapat dideteksi kata-kata negasi yang berkorelasi dengan kata yang berada pada kamus. Kasus ini merupakan contoh dari analisis yang

tidak sesuai sehingga memberikan *insight* yang tidak sesuai pula.

Kalimat 5 terklasifikasi sebagai kalimat netral karena tidak ada kata pada kalimat tersebut yang *match* dengan kata yang ada pada kamus. Padahal kita mengetahui bahwa kalimat tersebut seharusnya bersentimen negatif. Kejadian ini membuktikan kekurangan dari *sentiment analysis* berbasis kamus yaitu saat kata bersentimen pada suatu kalimat tidak ada pada kamus *pattern*. Pendekatan ini mengharuskan analisis untuk memiliki data kamus yang sangat besar untuk dapat bekerja efektif.

Dari penjabaran di atas, manusia dapat mendeteksi bahwa jumlah sebenarnya untuk sentimen positif adalah 2 atau 40%, jumlah sentimen negatif adalah 2 atau 40%, dan jumlah sentimen netral adalah 1 atau 20%.

Grafik 4.1 Proporsi Sentimen Sebenarnya pada Merk A



Jika analisis akurat, merk A akan mendapat *insight* yang berbeda dan akan melakukan tindakan yang berbeda juga karena dapat disimpulkan bahwa hanya 50% pasar yang memiliki sentimen positif sehingga diperlukan evaluasi terhadap produknya agar dapat lebih bersaing.

V. KESIMPULAN

Algoritma *string matching* Boyer-Moore sangat sederhana dan mudah digunakan untuk melakukan *sentiment analysis*. Namun akurasi yang diberikan oleh teknik ini kurang baik karena terdapat banyak *corner case* yang tidak dapat ditangani.

Ketidakakurasi dari hasil analisis menggambarkan *insight* yang tidak sesuai sehingga tidak menggambarkan sentimen konsumen yang sebenarnya. Hal ini kemudian berdampak pada keputusan yang diambil oleh produsen yang tidak sesuai pula sehingga optimasi *customer experience* memiliki kemungkinan untuk dilakukan secara tidak efektif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih dan syukur kepada Tuhan Yang Maha Esa atas segala nikmat yang telah

diberikan-Nya, yang dengan itu penulis dapat menulis makalah ini. Selain itu penulis juga mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Rinaldi Munir, M.T. atas bimbingan dalam perkuliahan Strategi Algoritma.
2. Bapak Dr. Ir. Rila Mandala, M.Eng. atas pengajaran dalam perkuliahan Strategi Algoritma.
3. Teman-teman IF'19 yang senantiasa membantu penulis dalam menjalani perkuliahan.

REFERENSI

- [1] <https://www.accenture.com/acnmedia/PDF-109/Accenture-Effortless-Experiences.pdf> diakses pada 8 Mei 2021 pukul 20.20.
- [2] <https://hbr.org/2007/02/understanding-customer-experience> diakses pada 10 Mei 2021 pukul 23.27.
- [3] <https://business.trustpilot.com/reviews/learn-from-customers/5-reasons-why-customer-experience-is-the-pulse-of-every-business> diakses pada 11 Mei 2021 pukul 00.00.
- [4] <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/prediction-the-future-of-cx> diakses pada 11 Mei 2021 pukul 00.21.
- [5] <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17> diakses pada 11 Mei 2021 pukul 00.29.

- [6] D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications (0975 – 8887)*, 125(13), 29-31.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Jakarta, 11 Mei 2021



Aisyah Farras Aqila - 13519054