

KLASTERISASI SENTIMEN ULASAN PENGGUNA APLIKASI BCA MOBILE PADA PLATFORM GOOGLE PLAY STORE DENGAN ALGORITMA K-MEANS CLUSTERING

Gilbert Sanko Sunarko ¹⁾ Wasino ²⁾ Tri Sutrisno ³⁾

¹⁾²⁾³⁾ Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara

Jl. Letjen S Parman No. 1, Jakarta 11440 Indonesia

email : gilbert.825190084@stu.untar.ac.id ¹⁾, email : wasino@fti.untar.ac.id ²⁾, email : tris@fti.untar.ac.id ³⁾

ABSTRACT

Financial banking app is increasingly becomes part of daily lifestyle. From survey conducted by Top Brand Index showed one of the most popular banking app is BCA Mobile. This study was conducted to analyze BCA Mobile app user feedback from App reviews in google play store platform. The data is collected using scrapping method with google-play-scrapper library in python. Using K-Means Clustering algorithm to analyze 662 reviews in 26 December 2022 until 30 December 2022 time period. The clustering process is supported by Term Frequency-Inverse Document Frequency weighting method to help calculate the importance of a word in a set of document. This clustering process produces ten cluster, with silhouette score of 0.1027277 and average star rating of 2.65.

Key words

K-Means Clustering, BCA Mobile, Google Play Store, Reviews, Tf-Idf

1. Pendahuluan

Perkembangan teknologi kini semakin canggih seiring dengan perkembangan zaman. Teknologi mengubah segala aspek kehidupan manusia, salah satu aspek yang mengalami perubahan yaitu sistem transaksi pembayaran dari pembayaran tunai menjadi pembayaran elektronik yang menggunakan perkembangan teknologi finansial. Keseharian masyarakat yang selalu menggunakan smartphone dan internet dalam melakukan aktivitas, membuat pembayaran serta transaksi elektronik mulai diterima karena dapat dilakukan secara mudah dan cepat.

Selama pandemi Jumlah pengguna internet di Indonesia meningkat dari yang sebelumnya 175 juta hingga mencapai 210 juta pada tahun 2022 [1]. Kurang lebih 77% penduduk Indonesia sudah menggunakan internet, waktu yang dihabiskan orang Indonesia untuk

mengakses internet per hari rata-rata yaitu 8 jam 52 menit. Meningkatnya jumlah pengguna internet di Indonesia juga berbanding dengan peningkatan penggunaan aplikasi finansial atau keuangan. Seiring dengan transformasi digital industri keuangan, banyak bank yang saat ini telah mengembangkan layanan produk digital. Dengan adanya pandemi Covid-19 juga mendorong transformasi ini semakin masif. Sehingga terjadi peningkatan luar biasa terhadap jumlah pengguna *mobile* dan internet *banking*.

Berdasar riset yang dilakukan oleh *Populix* pada tahun 2022 berjudul “*Consumer Preference Towards Banking and e-Wallet Apps*” [2], Aplikasi finansial keuangan telah mejadi gaya hidup sehari – hari. Dari hasil riset *Populix* tersebut ditemukan bahwa lebih dari 80 persen responden nya menggunakan aplikasi perbankan digital setidaknya satu kali dalam seminggu. Dari survei yang dilakukan oleh *Top Brand Awards*, ditemukan bahwa aplikasi perbankan digital yang paling populer [3] berdasar kan *Top Brand Index* adalah m-BCA, disusul oleh BRI Mobile pada posisi kedua, m-Banking Mandiri pada posisi ketiga serta BNI Mobile pada posisi keempat.

Untuk mengunduh sebuah aplikasi biasanya pengguna aplikasi mengunduh dari sebuah marketplace aplikasi. Ada beberapa marketplace aplikasi yang tersedia seperti *Apple App Store* untuk sistem operasi iOS dan *Google Play Store* untuk sistem operasi android. Di Indonesia sendiri, mayoritas pengguna internet menggunakan sistem operasi android. Dalam sistem operasi android untuk menggunakan sebuah aplikasi, pengguna perlu mengunduh aplikasi melalui toko aplikasi *Google Play Store*. Setiap aplikasi yang tersedia di *Google Play Store* memiliki kolom komentar agar pengguna dapat memberikan ulasan terhadap aplikasi tersebut. Setelah mengunduh aplikasi banyak pengguna aplikasi yang menuliskan opini atau *feedback* nya setelah menggunakan aplikasi. Opini tersebut dapat bernilai positif, negatif maupun netral.

Banyaknya *feedback* yang banyak tersebut tentunya sangat sulit untuk dapat dibaca satu persatu, sehingga dengan menggunakan *text mining* serta algoritma *K-Means Clustering* untuk melakukan analisis sentimen, data ulasan tersebut dapat di klusterisasi agar didapatkan gambaran besar. Hasil klusterisasi yang didapatkan dapat menjadi masukan dari pengguna tiap aplikasi perbankan.

Beberapa penelitian terdahulu sudah pernah dilakukan terkait analisis sentiment dengan data ulasan. Penelitian dilakukan [4] dengan menggunakan metode *K-Means Clustering* untuk melakukan klusterisasi sentiment pada ulasan dari pengguna maskapai penerbangan penerbangan Garuda Indonesia, Air Asia dan Lion Air pada website Skytrax dengan tujuan untuk mengelompokan opini positif dan opini negatif dari ulasan tersebut. Hasil penelitian tersebut didapat tingkat keberhasilan klusterisasi mencapai Lion Air 60,5% untuk Lion Air, 52,8% untuk Air Asia, dan 71,8% untuk Garuda Indonesia. Selanjutnya penelitian dilakukan [5] menggunakan algoritma K-Means Clustering untuk menganalisa data *tweet* konsumen pengguna *provider* Indosat. Dari penelitian tersebut didapatkan hasil tiga kluster yaitu kluster 1 mengenai jaringan Indosat yang parah, kluster 2 tentang permintaan perbaikan jaringan sinyal Indosat, dan pada kluster 3 tentang jaringan sinyal parah Indosat pada daerah Bogor.

2. Tinjauan Pustaka

2.1 Analisis Sentimen

Analisis sentimen adalah studi komputasional yang berkaitan dengan opini, sentimen, emosi, dan sikap seseorang [6]. Dalam beberapa tahun terakhir ini, analisis sentimen telah diterapkan dalam berbagai industri seperti produk – produk konsumen, jasa, layanan kesehatan, dan layanan keuangan.[7] Jutaan komentar yang dihasilkan oleh media sosial serta *website review* menjadi salah satu sumber data yang besar untuk analisis sentimen. Meskipun begitu, informasi yang terkandung dalam data tersebut sulit untuk di ekstrak dan digunakan secara langsung karena data tersebut berbentuk teks yang termasuk salah satu bentuk data tidak terstruktur.

2.2 Text Preprocessing

Dalam *text mining*, data yang didapatkan dalam bentuk teks merupakan data tidak terstruktur. Data tidak terstruktur dan kompleks yang berbentuk teks tersebut sulit untuk di analisis langsung oleh komputer. Untuk data teks tersebut dapat dipahami oleh komputer, diperlukan sebuah proses yaitu *Natural Language Processing*. *Text preprocessing* adalah salah satu

tahapan yang perlu dilakukan sebelum data teks dapat dimengerti oleh komputer.[8]

2.3 Tf-Idf

Setiap kata dalam sebuah dokumen memiliki tingkat kepentingan yang berbeda, maka untuk itu diperlukan sebuah indikator [9]. Salah satu indikator yang sering digunakan untuk menilai sebuah kata dalam dokumen adalah dengan metode *Term Frequency Inverse Document Frequency* (Tf-Idf). Tf-Idf merupakan metode statistik numerik yang dapat menunjukkan tingkat kepentingan sebuah kata dalam dokumen. Proses ini berguna untuk memproses data teks menjadi vektor dalam tahap yang dikenal sebagai *Feature Extraction*. Tf atau *Term Frequency* merupakan rasio jumlah sebuah huruf dibandingkan jumlah semua huruf dalam sebuah dokumen. Sedangkan *Inverse Document Frequency* merupakan rasio jumlah dokumen teks yang mengandung sebuah kata tersebut.

2.4 Algoritma K-Means Clustering

K-Means Clustering adalah salah satu algoritma *unsupervised learning* dan salah satu metode yang populer untuk klusterisasi data yang tidak memiliki label ke dalam *cluster*. *K-Means Clustering* adalah metode partisi yang diterapkan untuk menganalisa data dimana data yang diamati, dimasukkan dalam sebuah *cluster* berdasar jarak terdekat nya ke titik pusat sebuah *cluster* atau yang disebut dengan *centroid*. [10]. Langkah – langkah algoritma *K-Means Clustering* menurut Kanungo dan Mount (2002) adalah sebagai berikut:

1. Tentukan jumlah K cluster.
2. *Initialization* - Yaitu proses menentukan titik acak sebagai titik awal *centroid* untuk masing masing *cluster*.
3. *Classification* - Setiap titik data di tentukan masuk ke sebuah cluster berdasar jarak terdekat titik data tersebut ke *centroid cluster*.
4. *Centroid calculation* - Setelah setiap titik data dimasukkan ke dalam sebuah cluster, maka titik centroid dihitung ulang.
5. *Convergence criteria* – Langkah 3 dan 4 diulangi hingga tidak ada perubahan anggota cluster dan titik *centroid* tidak bergeser kembali.

2.5 Silhouette Score

Silhouette adalah alat yang digunakan untuk menilai validitas dari klusterisasi. Pengujian dilakukan untuk melihat bagaimana sample data dalam sebuah *cluster* memiliki kemiripan satu sama lain [11]. Skor *Silhouette* dihitung untuk setiap sample dari *cluster* yang berbeda beda. Skor bervariasi antara -1 hingga 1. Jika skor mendekati 1 maka *cluster* yang ada sudah padat dan

terpisah dengan baik dari *cluster* lain. Sedangkan nilai mendekati 0 berarti *cluster* saling timpang tindih dengan sample berada di ujung batas sebuah *cluster*. Skor *silhouette* yang mendekati -1 mengindikasikan bahwa sampel mungkin berada di *cluster* yang salah.

3. Metodologi

3.1 Data

Proses pengumpulan data dilakukan dengan teknik *scrapping*, dimana data yang di *scrapping* merupakan data ulasan pengguna aplikasi perbankan BCA Mobile pada situs Google Play Store dengan periode waktu 26 Desember 2022 hingga 30 Desember 2022. Proses *scrapping* dilakukan dengan menggunakan *library google-play-scraper* yang tersedia pada *python*. *Library* ini akan melakukan *scrapping* dengan mengambil data aplikasi yang sudah ditentukan dari situs *Google Play Store*.

```

{
  "appId": "com.bca",
  "at": "2022-08-31 17:08:53",
  "content": "Dari awal buka rekening sampai sekarang. Untuk saya sih semuanya oke saja. Tapi setelah ganti hape knn hilang. Semuanya jadi ribet knn tdk dpt verifikasi SMS. Pengaturan pesan, pengaturan aplikasi udah sesuai, sampe cari tau lewat yutub. Tapi msh blm bisa juga. Ada yg menyarankan pake sms premium. Apa hrs punya SMS premium yg bebayar biar dapatkan verifikasi SMSnya. Mungln yg punya saldo banyak itu bukan jd masalah, tp yang saldonya limit buat transaksi kecil2an. Itu jadi mikir2 lagi.....",
  "replyDate": "2022-08-31 17:19:53",
  "replyContent": "Mohon maaf atas kendala yang Bapak/Ibu alami, apabila kode OTP saat melakukan verifikasi n-BCA tidak diterima maka pastikan nomor HP yang terhubung dalam keadaan aktif dan tersedia pulsa. Tidak harus SMS premium ya.\n",
  "reviewCreatedVersion": "2.9.5",
  "reviewId": "e3cc0968-513a-468d-854d-3bf9afd5b4c5",
  "score": 1,
  "sortOrder": "most_relevant",
  "thumbsUpCount": 42,
  "userImage": "https://play-lh.googleusercontent.com/a-/ACNPEu_dFbo-ayKd88CQz7q491QxdVgVwamIo1XH79P",
  "userName": "Farid penelope"
}
    
```

Gambar 1 Json Hasil *Scrapping*

Data yang sudah di *scrapping* dapat disimpan dalam bentuk *file csv* ataupun dilanjutkan ke proses selanjutnya yaitu *Text Pre Processing.data*.

Table 1 Contoh Hasil Data *Scrapping*

No.	Ulasan
1	Sudah kirim sms untuk verifikasi, tidak direspon. Sedang minta otp dari aplikasi lain langsung dapat OTP. sayang sekali sudah puluhan kali alhasil cuma buang2 pulsa saja. Hub cs HALO BCA juga sama saja gak ada solusi!. TOLONG DIPERBAIKI!! Hanya karena ganti HP saja, RIBETnya bukan main!! Sudah isi pulsa 10rb, sudah pakai data pribadi (bukan wifi) dan sim memang ada di SIM 1 . Tp ttp aja. TOLONG PERBAIKANNYA TIDAK MEMPERSULIT.
2	Setelah d update . Aplikasi ini semakin jelek .

	Sering logout d tengah proses . Atau ga bisa masuk sama sekali . Beberapa kali sering kembali le menu utama . Apah perbaikkan malah menurunkan kualitas ?
3	interface jadul, buat pemula agak membingungkan, buat saldo mungkin lebih baik ditampilkan pada layar beranda, dan cara transfer mohon untuk dipermudah, soalnya agak membingungkan juga.
4	Mantap pelayanannya
5	Tambah fitur fingerprint dong, biar kalo login tinggal tap biar efisien juga.

Setelah melakukan proses Pra Pemrosesan Data Teks kemudian hasil data yang sudah di *scrapping* melalui *google play scraper* dan telah di lakukan *preprocessing* terhadap data ulasan akan disimpan dalam *database*. *Database* yang digunakan adalah *PostgreSQL*, dengan spesifikasi sebagai berikut.

Table 2 Struktur *Database PostgreSQL*

No	Field	Tipe	Keterangan
1	<i>Index</i>	Bigint	Primary Key
2	<i>appId</i>	Text	Id Aplikasi
3	<i>reviewId</i>	Text	Id Ulasan
4	<i>Content</i>	Text	Isi Ulasan
5	<i>score</i>	Bigint	<i>Rating</i>
6	<i>At</i>	Timestamp	Tanggal
7	<i>reviewCreatedversion</i>	Text	versi Ulasan
8	<i>preprocessing</i>	text	Ulasan yang sudah di preproce sing

3.2 Pra Pemrosesan Data Teks

Pada tahap ini data ulasan yang sudah didapatkan perlu diproses terlebih dahulu karena data teks termasuk bentuk data tidak terstruktur. Tahap ini mempersiapkan data agar mudah untuk diolah. Tahap *preprocessing* ini dapat dilakukan salah satunya dengan menggunakan *library* di *python* yang bernama *NLTK (Natural Language Toolkit)*. Berapa tahapan dalam *Text Pre Processing* adalah sebagai berikut:

$$w = tf_{t,d} \times idf = tf_{t,d} \times \log(N/df) \quad (1)$$

1. Cleansing

Pada tahap ini akan dilakukan pembersihan terhadap data dengan menghilangkan noise atau karakter yang tidak berarti. Ulasan: “Tambah fitur fingerprint dong, biar kalo login tinggal tap biar efisien juga.” Cleansing: “Tambah fitur fingerprint dong biar kalo login tinggal tap biar efisien juga”.

2. Case Folding

Proses mengkonversi teks menjadi bentuk yang standar, misalnya menjadi huruf kecil. Dalam python untuk melakukan proses ini dapat menggunakan fungsi `lower()`.

3. Stemming

Proses stemming adalah proses menghilangkan imbuhan pada awal dan akhir kalimat serta merubah bentuk kata menjadi kata dasarnya. Ada Beberapa contoh algoritma stemming yang dapat digunakan untuk kata dengan Bahasa Indonesia, seperti Porter Stemmer (English & Indonesia), Sastrawi Nazief-Adriani (Indonesia), Stemming Arifin-Setiono (Indonesia). Proses stemming ini dapat menggunakan library Sastrawi yang berada pada python.

4. Tokenizing

Tokenizing adalah proses pemecahan kalimat menjadi bagian yang lebih kecil yaitu kata kata yang disebut token. 15 Ulasan: “tambah fitur fingerprint dong biar kalo login tinggal tap biar efisien juga” Tokenizing: [‘tambah’, ‘fitur’, ‘fingerprint’, ‘dong’, ‘biar’, ‘kalo’, ‘login’, ‘tinggal’, ‘tap’, ‘biar’, ‘efisien’, ‘juga’].

5. Stopwords Removal

Pada tahap ini, dilakukan penghilangan kata yang dianggap sering muncul dan tidak penting. Dalam proses ini diperlukan sebuah daftar kata yang perlu dihilangkan atau yang disebut dengan *Stoplist*. Dalam library NLTK sudah terdapat *stoplist* dalam Bahasa Indonesia yang dapat digunakan.

3.3 Feature Extraction

Feature Extraction merupakan proses dimana data teks diambil fiturnya agar dapat dianalisa lebih lanjut oleh algoritma *machine learning*. Karena bentuk data teks sulit untuk dipahami oleh komputer maka data kata perlu direpresentasikan dalam bentuk *vector*. Dalam proses ini biasa digunakan metode yang bernama Tf-Idf. *Term Frequency-Inverse Document Frequency* (Tf-Idf) merupakan metode pembobotan kata, dimana setiap kata diberi nilai untuk menilai pentingnya sebuah kata dalam dokumen [12]. Proses perhitungan Tf-Idf dilakukan dengan rumus:

Keterangan:

- w = Merupakan bobot kata t terhadap dokumen d
- $tf_{t,d}$ = Jumlah kemunculan kata dalam dokumen
- N = Jumlah dokumen
- df = Jumlah dokumen yang mengandung kata t

Dalam perancangan aplikasi ini, untuk menentukan bobot Tf-Idf digunakan *TfidfVectorizer* yang terdapat pada library sklearn untuk melakukan 16 perhitungan secara otomatis terhadap data ulasan yang sudah melalui tahap *text pre processing*.

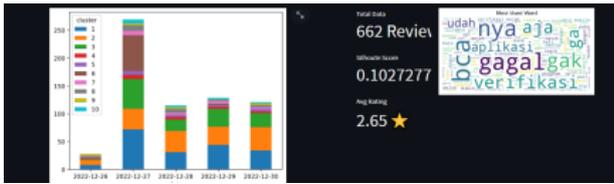
4. Hasil Percobaan

Pengujian terhadap algoritma *K-Means Clustering* untuk melakukan klasterisasi sentimen dilakukan dengan cara menguji secara langsung pada sistem. Pengujian dilakukan dengan menggunakan aplikasi m-BCA dengan rentang waktu 26 Desember 2022 hingga 30 Desember 2022. Untuk menguji *validitas* hasil klaster dilakukan pengujian dengan rentang klaster antara dua hingga 10 klaster kemudian jumlah klaster dengan *silhouette score* paling optimal akan digunakan untuk pengujian. Berikut Tabel hasil pengujian jumlah klaster dengan *silhouette score*:

Table 3 Hasil Pengujian Jumlah Klaster dengan *Silhouette Score*

Jumlah Cluster	Skor Silhouette
2	0.0285101
3	0.0340880
4	0.0611070
5	0.0712578
6	0.0835124
7	0.0800757
8	0.0878692
9	0.0862617
10	0.1027277

Dari Tabel diatas dapat disimpulkan bahwa jumlah *cluster* paling optimal untuk data pengujian adalah 10 *cluster*. Berikut pada gambar dibawah ini dapat dilihat hasil *clustering* pada data pengujian dari ulasan aplikasi m-Bca pada rentang waktu 26 Desember 2022 hingga 30 Desember 2022.



Gambar 2 Hasil Clustering Pada Data Pengujian Rentang Waktu 26 Desember

Hasil *clustering* yang dilakukan terhadap 662 ulasan dengan jumlah *cluster* sepuluh menghasilkan skor silhouette sebesar 0.1027277. Dengan skor yang berada pada nilai positif menjadikan hasil kluster cukup valid. Skor yang bernilai positif menandakan cluster tidak saling tumpang tindih. Hanya saja dengan skor yang mendekati 0 juga menandakan bahwa cluster saling berdempetan sehingga ada kemungkinan bahwa ada beberapa data yang masuk ke cluster yang salah.



Gambar 3 Cluster 1

Pada Gambar cluster 1 dapat dilihat terdiri dari 189 ulasan dari 662 ulasan secara keseluruhan. Pada cluster ini mayoritas ulasan berbicara mengenai ‘update’ aplikasi m-Bca yang menyebabkan pengguna tidak bisa masuk ke aplikasi. Klaster ini memiliki rata rata *rating* ulasan 3.06.



Gambar 4 Cluster 2

Pada gambar cluster 2 menunjukkan banyaknya ulasan yang berbicara mengenai gagal verifikasi pada aplikasi m-Bca. Banyaknya keluhan ini juga tercermin dari rata rata *rating* yang terdapat pada klaster ini yang bernilai 1.29.



Gambar 5 Cluster 3

Pada kluster 3 yang berisi 122 reviews dengan rata rata *rating* 2.08, banyak yang memberikan ulasan susahny login aplikasi m-Bca.



Gambar 6 Cluster 4 dan Cluster 5

Pada Cluster 4 dan 5 dapat disimpulkan berisi sentiment yang positif dengan rata rata *rating* yang tinggi dan kata kata yang bernilai positif sebagai cerminan klasternya.



Gambar 7 Cluster 6

Pada kluster 6 mayoritas ulasan berbicara mengenai indikator lampu merah pada aplikasi m-Bca yang menyala terus sehingga pengguna tidak dapat menggunakan aplikasi tersebut. Hal ini menyebabkan rata rata *rating* pada klaster ini cukup rendah pada angka 2.33



Gambar 8 Cluster 7

Pada kluster 7 ini, banyak ulasan yang bernilai positif namun jika dilihat satu persatu ada beberapa ulasan yang berbicara mengenai susahny verifikasi pada aplikasi m-Bca. Ulasan tersebut seharusnya berada pada klaster 2 yang juga berbicara mengenai verifikasi. Hal ini kemungkinan terjadi dikarenakan batas klaster yang saling berdekatan sehingga menyebabkan ulasan masuk ke klaster yang salah. Hal yang sama juga dapat dilihat pada klaster 8.



Gambar 9 Cluster 8

Pada kluster 9 dan 10 memiliki ulasan dengan sentimen yang positif hal ini dapat dilihat dari wordcloud tiap kluster yang hanya berisi kata kata yang positif serta rata rata *rating* tiap kluster berada pada nilai tertinggi yaitu 5.



Gambar 10 Cluster 9 dan Cluster 10

5. Kesimpulan

Dari hasil percobaan diatas dapat disimpulkan beberapa poin sebagai berikut :

1. Hasil percobaan analisis sentimen ulasan pengguna aplikasi *BCA Mobile* berhasil dilakukan.
2. Percobaan klusterisasi dengan menggunakan data ulasan dari rentang waktu 26 desember 2022 hingga 30 desember 2022 menghasilkan sepuluh kluster dengan skor *silhouette* sebesar 0.1027277 dan rata rata *rating* sebesar 2.65.
3. Percobaan mendapatkan *silhouette score* dengan nilai positif yang berarti kluster yang dihasilkan cukup baik. Namun dari hasil klusterisasi yang didapatkan terdapat beberapa data ulasan yang masuk ke kluster yang salah.
4. Implementasi algoritma *K-Means Clustering* untuk menganalisa ulasan pengguna aplikasi perbankan sudah berhasil dibuat.

6. Saran

Berikut merupakan beberapa saran yang mungkin dapat membantu pengembangan aplikasi berikutnya:

1. Menambahkan informasi yang ditampilkan ke pengguna.
2. Menambahkan *filter* yang dapat digunakan oleh pengguna sehingga lebih interaktif.
3. Menggabungkan dengan metode klasifikasi terlebih dahulu sehingga didapatkan kesimpulan sentimen yang positif dan yang negatif
4. Menghilangkan data *outlier* dari kluster sehingga hasil klusterisasi lebih *valid*.

REFERENSI

- [1] Dewi, I. R. (2022, June 9). Data Terbaru! Berapa Pengguna internet Indonesia 2022?. CNBC Indonesia. Diakses 1 September 2022, dari <https://www.cnbcindonesia.com/tech/20220609153306-37-345740/data-terbaru-berapa-pengguna-internet-indonesia-2022>
- [2] Consumer preference towards banking and E-Wallet Apps. Populix. (2022, July 28). Diakses 28 August 2022, dari <https://info.populix.co/report/digital-banking-survey/>

- [3] Annur, C. M. (2022, June 22). Aplikasi Mobile banking terpopuler di Indonesia, Siapa Juaranya?. Databoks. Diakses 28 September 2022, dari <https://databoks.katadata.co.id/datapublish/2022/06/22/aplikasi-mobile-banking-terpopuler-di-indonesia-siapa-juaranya>
- [4] Liu B. (2015). Sentiment Analysis : Mining Opinions, Sentiments, and Emotions. New York, NY : Cambridge University Press.
- [5] Safitri, S. I., Suhery, C., & Bahri, S. IMPLEMENTASI ALGORITMA K-MEANS UNTUK CLUSTERING SENTIMEN PADA OPINI KUALITAS PELAYANAN JASA PENERBANGAN. Coding Jurnal Komputer dan Aplikasi, 9(02).
- [6] Saputra, Try & Arianty, Rini. (2019). IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA ANALISIS SENTIMEN KELUHAN PENGGUNA INDOSAT. Jurnal Ilmiah Informatika Komputer. 24. 191-198. 10.35760/ik.2019.v24i3.2361.
- [7] Liu, B. (2012). Sentiment analysis and opinion mining. San Rafael: Morgan & Claypool Publishers
- [8] Feldman, R., & Sanger, J. (2007). The text mining handbook: Advanced approaches in analyzing unstructured data. Cambridge: Cambridge University Press.
- [9] Zafikri, A. 2008. Implementasi Metode Term Frequency Inverse Document Frequency (Tf-Idf) pada Sistem Temu Kembali Informasi. Universitas Sumatera Utara, (Online), (<http://repository.usu.ac.id/handle/123456789/16465>), diakses 21 Mei 2017.
- [10] Ghosh, S., & Dubey, S.K. (2013). Comparative Analysis of K-Means and Fuzzy C- Means Algorithms. International Journal of Advanced Computer Science and Applications, 4.
- [11] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. Vol.20, pp.53-65.
- [12] Abtohi, S. 2017. Implementasi Teknik Web Scraping dan Klasifikasi Sentimen Menggunakan Metode Support Vector Machine dan Asosiasi. Skripsi. Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam UII Yogyakarta.

Gilbert Sanko Sunarko, saat ini sebagai Mahasiswa program studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara