(JUSS) Jurnal Sains dan Sistem Informasi ISSN 2614-8277
Vol.7 No.1, Juli 2024

Analisis Sentimen Tingkat Entitas di Twitter Menggunakan Metode TF-IDF

Cecilyani Saragih Sitio [1)], Anna Dina Kalifia, S.Kom., M.Cs.[2)]
[1]Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta
email: cecilyani.5221811033@student.uty.ac.id[1)], anna.dina.kalifia@staff.uty.ac.id[2)].

Abstract

This journal discusses entity sentiment level analysis on Twitter using the TF-IDF method. Sentiment analysis is a process for identifying and classifying opinions or sentiments contained in text on Twitter. Focusing on the entity level allows for a more specific understanding of the sentiment associated with a particular entity, such as a person, place, organization, product, or other topic.
The method used in this journal is TF-IDF (Term Frequency-Inverse Document Frequency). This method is used to calculate sentiment scores related to entities on Twitter. TF-IDF calculates the frequency of occurrence of a word in the document as well as the frequency of occurrence of that word in the entire corpus. Using this method, journal authors attempt to identify key words that contribute to positive or negative sentiment towards the entity being analyzed.
This research aims to provide insight into the use of the TF-IDF method in sentiment analysis at the entity level on Twitter. The results of sentiment analysis can provide a deeper understanding of Twitter users' opinions and sentiments towards the entities they discuss. It is hoped that this research can contribute to the further development of sentiment analysis, especially in the context of social media such as Twitter.

Abstrak

Jurnal ini membahas analisis tingkat sentimen entitas di Twitter dengan menggunakan metode TF-IDF. Analisis sentimen adalah proses untuk mengidentifikasi dan mengklasifikasikan opini atau sentimen yang terkandung dalam teks yang ada di Twitter. Fokus pada tingkat entitas memungkinkan pemahaman yang lebih spesifik terhadap sentimen yang terkait dengan entitas tertentu, seperti orang, tempat, organisasi, produk, atau topik lainnya.
Metode yang digunakan dalam jurnal ini adalah TF-IDF (Term Frekuensi-Invers Dokumen Frekuensi). Metode ini digunakan untuk menghitung skor sentimen yang berkaitan dengan entitas yang ada di Twitter. TF-IDF menghitung frekuensi kemunculan kata dalam dokumen tersebut serta frekuensi kemunculan kata tersebut di seluruh korpus. Dengan menggunakan metode ini, penulis jurnal berusaha untuk mengidentifikasi kata-kata kunci yang berkontribusi pada sentimen positif atau negatif terhadap entitas yang sedang dianalisis.
Penelitian ini bertujuan untuk memberikan wawasan tentang penggunaan metode TF-IDF dalam analisis sentimen pada tingkat entitas di Twitter. Hasil analisis sentimen dapat memberikan pemahaman yang lebih mendalam mengenai opini dan sentimen pengguna Twitter terhadap entitas-entitas yang mereka bahas. Diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan analisis sentimen lebih lanjut, terutama dalam konteks media sosial seperti Twitter.

1. PENDAHULUAN

Analisis sentimen telah menjadi fokus perhatian yang signifikan dalam beberapa tahun terakhir, terutama karena ketersediaan platform media sosial yang semakin meningkat. Twitter, sebagai situs mikroblog yang populer, menyediakan akses ke sejumlah besar data real-time yang

dapat digunakan untuk menganalisis opini publik tentang berbagai topik. Penelitian ini bertujuan untuk menyelidiki analisis sentimen tingkat entitas dalam data Twitter menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency).

Analisis sentimen tingkat entitas melampaui sekadar menganalisis sentimen setiap kata atau frasa untuk memahami sentimen yang terkait dengan entitas tertentu yang disebutkan dalam teks. Pendekatan ini menjadi penting dalam menganalisis topik yang kompleks, di mana terdapat banyak entitas dengan sentimen yang berbeda-beda. Metode TF-IDF adalah teknik yang umum digunakan dalam pengambilan informasi dan penambangan teks, yang mengukur pentingnya sebuah kata dalam sebuah dokumen relatif terhadap frekuensinya dalam sebuah korpus.

Dalam penelitian ini, kami akan menerapkan metode ini untuk mengekstrak fitur yang merepresentasikan sentimen entitas yang disebutkan dalam data Twitter. Metode yang diusulkan melibatkan pra-pemrosesan data teks untuk menghilangkan kata-kata berhenti, tanda baca, dan melakukan stemming. Selanjutnya, kami akan memisahkan token teks menjadi kata-kata individual dan menghitung skor TF-IDF untuk setiap entitas yang disebutkan dalam tweet. Entitas akan diidentifikasi menggunakan teknik Named Entity Recognition (NER). Terakhir, kami akan melatih model pembelajaran mesin menggunakan fitur ini untuk memprediksi polaritas sentimen setiap entitas yang disebutkan dalam tweet.

Penelitian ini bertujuan untuk berkontribusi pada bidang analisis media sosial dengan memberikan pendekatan baru untuk analisis sentimen tingkat entitas menggunakan metode TF-IDF. Metode yang diusulkan ini dapat memberikan wawasan tentang sentimen entitas yang disebutkan dalam data Twitter, yang dapat berguna untuk berbagai aplikasi seperti manajemen reputasi merek, analisis umpan balik pelanggan, dan pengumpulan opini politik.

2. METODE PENELITIAN

Metode penelitian ini meliputi pengumpulan data dari sumber primer yaitu aplikasi twitter. Data yang saya ambil ini adalah Kumpulan data Analisis Sentiment Tingkat entitas di Twitter. Tugasnya adalah menilai sentiment pesan tentang entitas tersebut. Ada tiga kelas dalam Kumpulan data ini: Positif, Negative, dan Netral. Hasil dan pembahasan yang saya gunaka menggunakan RapidMiner,dan analisis sentiment menggunakan algoritma TF-IDF. Metode penelitian yang dilakukan adalah sebagai berikut :

A. Pengumpulan Data
   - Menggunakan API Twitter untuk mengumpulkan data tweet terkait topik tertentu.
   - Menentukan entitas yang menjadi focus studi, misalnya merek, produk, tokoh public, atau topik spesifik lainnya.

B. Pra- Pemrosesan Data:
   - Menghilangkan karakter khusus, tanda baca, dan kata-kata penghubung dari teks tweet.

C. Identifikasi Entitas
   - Melakukan klasifikasi entitas ke dalam kategori tertentu, seperti merek, produk, atau orang.

D. Pembobotan dengan TF-IDF
   - Menghitung skor Term Frequency-Inverse Document Frequency (TF-IDF) untuk setiap kata yang muncul dalam tweet yang berkaitan dengan entitas yang diteliti.
   - Melatih model klasifikasi mesin pembelajaran mengunakan skor TF-IDF dan label sentiment yang di tugaskan pada set data pelabelan

E. Pengklasifikasi Sentimen
   - Menggunakan metode klasifikasi memprediksi sentiment (positif, negative, netral) untuk setiap entitas.

F. Evaluasi dan Analisis

 ISSN 2614-8277

- Menganalisis hasil analisis sentiment untuk entitas yang di teliti, termasuk perbandingan sentiment antara entitas, tren sentiment seiring waktu, dan pola sentiment yang signifikan.

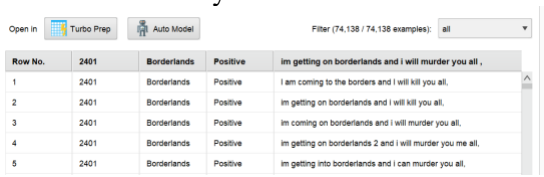
Metode penelitian ini menggabungkan pendekatan NLP (Natural Language Processing) dengan menggunakan TF-IDF untuk menganalisis sentiment Tingkat entitas di Twitter.

3. HASIL DAN PEMBAHASAN

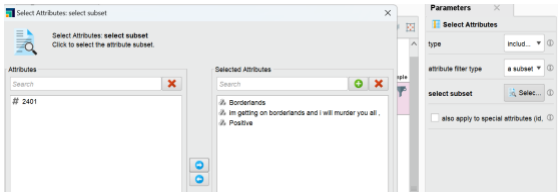
A. Pengumpulan data

Pada tahap ini saya mengambil data dari Kaggle. Data yang saya ambil ini adalah Kumpulan data analis sentiment Tingkat entitas di twitter. Mengingat pesan dan entitas, tugasnya adalah menilai sentiment pesan tentang entitas tersebut. Ada tig akelas dalam Kumpulan data ini: positif, negative, Netral.

Yang menjadi fokus studi untuk menganalisis TF-IDF di data ini adalah atribut: Borderlands, Positive, Im getting borderlands and I will murder you all



B. Pra-Pemrosesan Data



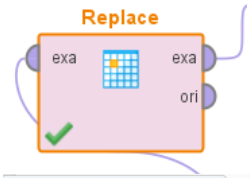
D. Pembobotan dengan TF-IDF

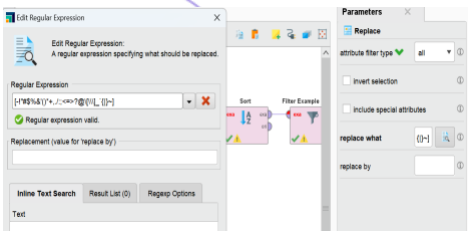
- Menghitung skor Term Frequency-Inverse Document Frequency (TF-IDF) untuk setiap kata yang muncul dalam tweet yang berkaitan dengan entitas yang diteliti.

Tahap ini merupakan tahap yang peling penting dalam penelitian ini. Pada tahap ini saya menggunakan operator procces document from data

- Menghilangkan karakter khusus, tanda baca, dan kata-kata penghubung dari teks tweet.

Pada proses ini saya menggunakan operator replace pada rapid minner. Operator ini adalah salah satu operator yang di gunakan untuk mengganti nilai atau memodifikasi teks datam suatu dataset. Berikut akan di tampilkan gambar operatornya serta parameternya :

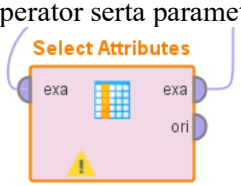




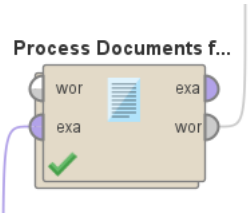
C. Identifikasi Entitas

- Melakukan klasifikasi entitas ke dalam kategori tertentu, seperti merek, produk, atau orang.

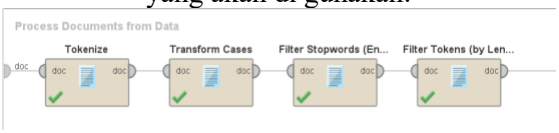
Pada tahap ini saya menggunakan Select attributes. Select attributes ini adalah salah satu operator yang di gunakan untuk memilih atribut atau mengklasifikasikan atribut atau kolom tersentu dari data set yang akan di gunakan dalam analisis atau pemodelan data. Berikut akan di tampilkan operator serta parameternya:



 ISSN 2614-8277



Setelah ini di tampilkan dan jika dii klik 2 kali maka akan muncul proses selanjutnya menuju TF-IDF, berikut operator operator yang akan di gunakan:



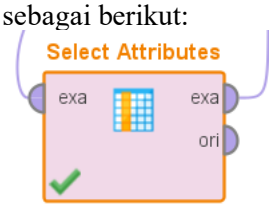
- Operator Tokenize: digunakan untuk memecah teks menjadi unit-unit yang lebih kecil yang di sebut sebagai "Token". Tokenisasi sering di gunakan dalam pemrosesan bahasa alami untuk memisahkan kata kata atau bagian bagian dari teks yang lebih besar. Operator ini membantu dalam analisis teks dengan memungkinkan perhitungan statistik kata, pemodelan, dan pemrosesan teks lainya.
- Operatoor Transform Cases: di gunakan untuk mengubah kasus teks menjadi kasus tertentu, seperti huruf kecil, huruf besar, atau gata kapitalisasi yang lain. Ini membantu dalam memastikan keseragaman dalam teks dan memfasilitasi pemrosesan dan analisis teks yang lebih menarik.
- Operator Stopwords: digunakan untuk menghapus kata kata umum yang tidak memberikan banyak informasi dalam analisis teks. Kata-kata seperti "a", "an", "the", "is", "are", dan lain sebagainya sering di anggap sebagai stop words. Menghapus stopwords membantu dalam fokus pada kata kata kunci yang lebih bermakna dalam analisis teks, seperti kata benda, kata kerja, atau kata-kata kunci lainnya.
- Operator Filter Token: digunakan untuk menyaring atau memfilter token-token dari teks berdasarkan kriteria tetentu.operator ini juga berguna untuk menghapus token berdasarkan panjangnya, frekuensi kemunculanya, jenis kkarakter yang di perbolehkan, atau menggunakan ekspresi reguler untuk memfilter token dengan pola tertentu. Operator ini membatu dalam membersihkan dan mempersiapkan teks untuk analisis lebih lanjut dengan menyingkirkan token yang tidak relevan atau tidak di inginkan.

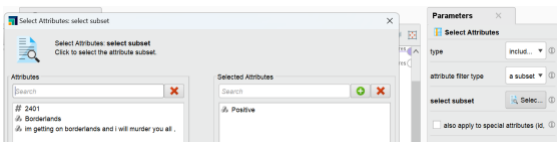
Dengan mengunakan kombinasi dari operator-tokenisasi transformasi kasus, pembuangan stopword, dan penyaringan token, dapat mempersiapkan teks secara efektif untuk analisis teks lebih lanjut seperti klasifikasi, analisis sentimen, atau pengelompokan.

E. Pengklasifikasi Sentimen

- Menggunakan metode klasifikasi memprediksi sentiment (positif, negative, netral) untuk setiap entitas.

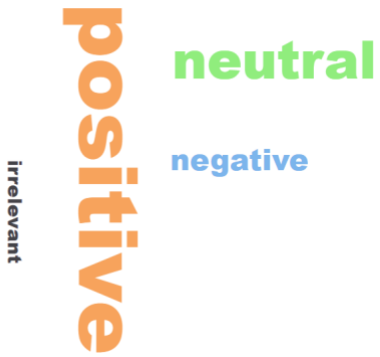
  Pada tahap ini klasifikasi sentimen adalah proses mengidentifikasi dan mengkategorikan sentimen atau opini dari teks atau data yang di berikan. Pada proses ini saya menggunakan operator Select Attributes, dengan parameternya sebagai berikut:





Dan dengan hasil:

 ISSN 2614-8277



F. Evaluasi dan Analisis

➢ Menganalisis hasil analisis sentiment untuk entitas yang di teliti, termasuk perbandingan sentiment antara entitas, tren sentiment seiring waktu, dan pola sentiment yang signifikan. Proses ini adalah proses yang menentukann hasis alkir dari data yang di analisis. Pada proses ini saya menggunakan plot type word cloud. Untuk melihat kata-kata yang sering muncul atau kata-kata yang sering di pakai dalam data analisis twitter tersebut.

4. KESIMPULAN

Analisis Sentimen Tingkat Entitas di Twitter Menggunakan mmetode TF-IDF ini membahas tentang penggunakaan metode TF-IDF dalam melakukan analisis sentiment terhadap entitas di Twitter.

Ada beberapa point yang dapat di ambil dalam jurnal ini yaitu :

1. Analisis sentiment: jurnal ini focus pada analisis sentiment, yaitu proses identifikasi dan klasifikasi opini atau sentiment yang terkandung dalam teks yang ada pada Twitter. Analisis sentiment bertujuan untuk memahami perasaan atau pendapat pengguna Twitter terhadap suatu entitas atau topik tertentu.
2. Tingkat Entitas: jurnal ini membahas tentang analisis sentiment pada Tingkat entitas. Entitas dapat merajuk pada orang, tempat, organisasi, produk, atau topik lain yang diidentifikasi dalam teks di Twitter. Focus dalam Tingkat entitas dapat memberikan wawasan yang lebih spesifik tentang sentiment yang terkait dengan entitas tersebut.
3. Metode TF-IDF: jurnal ini menggunakan metode TF-IDF dalam analisis sentiment. TF-IDF adalah metode umum yang di gunakan dalam pengolahan Bahasa alami untuk mengevaluasi pentingnya sebuah kata dalam suatu document atau korpus berdasarkan frekuensi kemunculan kata tersebut dalam seluruh korpus.

Dengan menggunakan metode TF-IDF, penulis jurnal berusaha untuk menghitung skor sentiment yang berkaiatn dengan entitas ynag ada di Twitter. Metode ini dapat membantu dalam mengidentifikasi kata-kata kunci yang berkontribusi pada sentiment positif atau negative terhadap entitas tersebut.

Kesimpulan utama pada jurnal ini adalah bahwa penulis menggunakan metode TF-IDf untuk melakukan analisis sentiment Tingkat Entitas di Twitter. Dengan demikian, jurnal ini dapat memberikan wawasan tentang penggunaan metode ini dalam memahami sentiment yang terkait dengan entitas di platform media social tersebut.

 ISSN 2614-8277

5.REFERENSI

1. *ANALISIS SENTIMEN KOMENTAR NETIZEN TERHADAP CALON PRESIDEN INDONESIA 2019 DARI*. (n.d.).

2. Atika, D., Ari Aldino, A., Informasi, S., Pagar Alam No, J., Ratu, L., & Kedaton, K. (2022). TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP TEKANAN MENTAL PADA MEDIA SOSIAL TWITTER. In *Jurnal Teknologi dan Sistem Informasi (JTSI)* (Vol. 3, Issue 4). http://jim.teknokrat.ac.id/index.php/JTSI

3. Darwis, D., Shintya Pratiwi, E., Ferico, A., & Pasaribu, O. (2020). PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA. In *Jurnal Ilmiah Edutic* (Vol. 7, Issue 1).

4. Deviyanto, A., Didik Wahyudi, M. R., & Informatika UIN Sunan Kalijaga Yogyakarta Jl Marsda Adi Sucipto No, T. (2018). PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR. *Jurnal Informatika Sunan Kalijaga)*, *3*(1), 1–13. https://twitter.com/search?l=id&q=AHY%20since%3A2017-01-01%20until%3A2017-01-

5. *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION 43*. (n.d.). https://t.co/9WloaWpfD5

6. Kurnia Sutrisno, F., & Muslim Lhaksmana, K. (2021). Analisis Sentimen Destinasi Wisata Kuliner di Twitter Menggunakan TF-IDF dan Complement Naïve Bayes pada Dataset Tidak Seimbang. *Agustus*, *8*(4), 4115.

7. Rizkia, S., Budi, E., Si, S. S., Puspandari, D., & Pd, M. (n.d.). *Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF*.

8. Rizkina, N. Q., & Hasan, F. N. (2023). Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes. *Journal of Information System Research (JOSH)*, *4*(4), 1136–1144. https://doi.org/10.47065/josh.v4i4.3803

9. Rofiqoh, U., Setya Perdana, R., & Fauzi, M. A. (2017). *Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features* (Vol. 1, Issue 12). http://j-ptiik.ub.ac.id

10. Sari, F. V., & Wibowo, A. (2019). ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI. *Jurnal SIMETRIS*, *10*(2).