PENGGALIAN DATA DAN ANALITIKA PROSES BISNIS ANALISIS TWITTER SENTIMEN PADA WILLOW PROJECT



Disusun Oleh:

Arya Bhima Andika	2109116075
Fadillah Jaga Pratama	2109116065
Hadi Prasetiyo	2109116070
Rezky Putra Kurniawan	2109116056

PROGRAM STUDI SISTEM INFORMASI FAKULTAS TEKNIK UNIVERSITAS MULAWARMAN

2023

KATA PENGANTAR

Dengan rahmat Allah SWT., kami mempersembahkan laporan ini sebagai bagian dari

Ujian Akhir Semester dalam mata kuliah Penggalian Data dan Analitika Proses Bisnis. Laporan

ini bertujuan untuk melakukan analisis sentimen pada proyek "Willow Project" dengan

menggunakan data teks yang dikumpulkan.

Proses bisnis saat ini semakin terpengaruh oleh penggunaan data dan analitika. Dalam

konteks ini, pemahaman terhadap sentimen dan opini pengguna atau pelanggan menjadi kunci

penting dalam pengambilan keputusan yang tepat dan merumuskan strategi bisnis yang efektif.

Oleh karena itu, kami melakukan penggalian data teks terkait proyek "Willow Project" dan

menerapkan teknik analisis sentimen untuk mendapatkan wawasan yang berharga.

Semoga laporan ini memberikan kontribusi yang berarti dalam pemahaman sentimen

dan pengambilan keputusan dalam konteks proyek "Willow". Kami sadar bahwa laporan ini

belum sempurna, dan kami sangat menghargai masukan dan saran untuk perbaikan di masa

depan.

Akhir kata, semoga laporan ini memberikan manfaat dan inspirasi bagi pembaca dalam

memahami pentingnya analisis sentimen dalam proses bisnis dan pemanfaatan data secara

efektif.

Penulis

Kelompok 8

2

DAFTAR ISI

KATA PENGANTAR	2
DAFTAR ISI	3
BAB I PENDAHULUAN	4
1.1 Latar Belakang	4
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Manfaat	5
BAB II SENTIMEN ANALISIS	6
2.1 Pengumpulan Data	6
2.2 Preprocessing Data	7
2.3 Analisis Sentimen	10
BAB III HASIL ANALISIS	12
3.1 Statistik Deskriptif	12
3.2 Visualisasi	13
BAB IV PENUTUP	16
4.1 Kesimpulan	16
4.2 Saran	16

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam era media sosial yang semakin berkembang, Twitter menjadi salah satu platform utama di mana orang berbagi pendapat, informasi, dan pengalaman mereka. Analisis sentimen pada Twitter telah menjadi metode yang populer untuk memahami bagaimana masyarakat merespons suatu topik atau proyek tertentu. Dalam konteks ini, laporan ini akan menganalisis sentimen terkait dengan Willow Project melalui penggunaan hashtag Twitter yang relevan.

1.2 Rumusan Masalah

Dari latar belakang di atas dapat diambil poin rumusan masalah dalam laporan ini yaitu :

- 1. Bagaimana persepsi dan tanggapan umum terhadap Willow Project di platform Twitter?
- 2. Apakah sentimen yang terkait dengan Willow Project cenderung positif, negatif, atau netral?
- 3. Apakah ada tren atau pola tertentu dalam sentimen yang muncul sehubungan dengan Willow Project di Twitter?

1.3 Tujuan

Adapun tujuan dari laporan ini antara lain:

- 1. Mengumpulkan data yang relevan dari Twitter dengan menggunakan hashtag terkait Willow Project.
- 2. Menganalisis sentimen yang muncul dalam tweet terkait Willow Project.
- 3. Membuat kesimpulan tentang persepsi umum dan sentimen terkait Willow Project di Twitter.
- 4. Memberikan pemahaman yang lebih baik tentang bagaimana proyek Willow Project diterima di kalangan pengguna Twitter.

1.4 Manfaat

Adapun beberapa manfaat dari laporan ini antara lain:

- 1. Memberikan wawasan tentang persepsi dan tanggapan publik terhadap Willow Project di platform Twitter.
- 2. Memberikan informasi berharga bagi tim pengelola Willow Project untuk mengevaluasi dan meningkatkan inisiatif mereka berdasarkan sentimen publik.
- 3. Menyoroti kekuatan analisis sentimen dalam mendapatkan wawasan dari media sosial dan pentingnya pemantauan respons publik.
- 4. Memberikan panduan dan contoh bagi peneliti atau praktisi lain yang tertarik dengan analisis sentimen dan respons publik dalam konteks proyek serupa.

BAB II

SENTIMEN ANALISIS

2.1 Pengumpulan Data

1. Sumber Data

Dalam laporan Anda, sumber data yang digunakan adalah Twitter dengan mengambil data yang terkait dengan hashtag Willow Project. Data ini terdiri dari tweet-tweet yang dikirim oleh pengguna Twitter yang menggunakan hashtag tersebut. Twitter menyediakan API (Application Programming Interface) yang memungkinkan pengambilan data dari platform mereka. Dalam link Colab kami, kami telah mengumpulkan tweet yang sesuai dengan hashtag Willow Project.

2. Proses Pengumpulan

Pada tahap pengumpulan data, kami telah melakukan permintaan pencarian berdasarkan hashtag Willow Project, menentukan jumlah tweet maksimal yang ingin diambil, dan menyimpan data tersebut dalam format yang sesuai, seperti file CSV atau DataFrame.

Bagian ini adalah langkah awal dalam proses pengumpulan data. Dalam kode di atas, kami menggunakan perpustakaan sntwitter untuk membuat objek TwitterSearchScraper. Objek ini digunakan untuk melakukan pencarian di Twitter dengan menggunakan hashtag "#willowproject". Dengan menggunakan TwitterSearchScraper, kami dapat mengumpulkan tweet-tweet yang terkait dengan hashtag tersebut.

Pada tahap ini, kode hanya menginisialisasi objek TwitterSearchScraper dan menentukan hashtag yang akan dicari. Namun, untuk benar-benar mengumpulkan data, kami perlu menjalankan operasi pencarian atau pengambilan data tambahan dengan menggunakan metode atau fungsi yang disediakan oleh perpustakaan tersebut.

2.2 Preprocessing Data

1. StopWords

Bagian ini berfungsi untuk mengumpulkan informasi penting dari setiap tweet yang ditemukan melalui pencarian hashtag "#willowproject". Berikut adalah penjelasan bagian tersebut:

Proses:

- A. Menginisialisasi sebuah variabel tweet_properties_list yang akan digunakan untuk menyimpan properti-properti dari setiap tweet yang relevan.
- B. Melakukan perulangan untuk setiap tweet yang ditemukan melalui hasilSearch.get_items().
- C. Membuat sebuah objek dictionary tweet_properties untuk menyimpan informasi tentang tweet tersebut.
- D. Menyimpan tanggal tweet, nama pengguna, dan isi tweet ke dalam tweet_properties.
- E. Membersihkan isi tweet dari karakter-karakter khusus, tautan, mention, dan simbol yang tidak diperlukan menggunakan ekspresi reguler.
- F. Menerjemahkan teks tweet dari bahasa Indonesia ke bahasa Inggris (namun kode tersebut dijelaskan dalam komentar).
- G. Menganalisis sentimen teks tweet menggunakan library TextBlob dan menyimpan hasil sentimen (positif atau negatif) ke dalam tweet_properties.
- H. Menyimpan tweet_properties ke dalam tweet_properties_list jika tweet tersebut bukan retweet, atau jika tweet tersebut merupakan retweet tetapi belum ada di dalam tweet_properties_list.
- I. Menghentikan proses pengambilan tweet setelah mencapai batas maksimal jumlah tweet (500 tweet dalam kode ini).

Bagian ini merupakan bagian penting dalam proses pengumpulan data dan analisis sentimen pada proyek Willow.

2. Pembersihan Data

Dalam langkah ini, karakter-karakter khusus, tanda baca, dan angka yang tidak relevan dalam teks dihapus. Hal ini dilakukan untuk memastikan bahwa teks yang digunakan dalam analisis sentimen hanya terdiri dari kata-kata yang relevan dan dapat memberikan informasi yang berguna dalam menentukan sentiment

a. Penghapusan StopWords

```
1 # remove stopwords
2 from nltk import word_tokenize, sent_tokenize
3
4 def remove_stopword(text, stop_words=stop_words):
5 word_tokens = word_tokenize(text)
6 filtered_sentence = [w for w in word_tokens if not w in stop_words]
7 return ' '.join(filtered_sentence)
```

b. Menggunakan pustaka sastrawi serta Stemming dan Lemmatization

```
Looking in indexes: <a href="https://pydi.org/simple">https://pydi.org/simple</a>, <a href="https://pydi.org/simple</a>, <a href="https://pydi.org/simple/">https://pydi.org/simple</a>, <a href="https://pydi.org/simple/">https://pydi.org/simple</a>,
```

3. Implementasi Preprocessing

```
[] 1 # implement preprocessing
2 preprocessed_data = df.copy()
3 preprocessed_data = df.copy()
3 preprocessed_data = df.copy()
3 preprocessed_data = df.copy()
4 y = preprocessed_data = limit =
```

4. Columning

```
1 tweet_data = []
2
3 for i, tweet in enumerate(hasilSearch.get_items()):
4    tweet_list=[tweet.date, tweet.user.username, tweet.content]
5    tweet_data.append(tweet_list)
6    if i == 500:
7     break
```

Bagian Columning mengacu pada proses mengorganisasi dan menyimpan data dari setiap tweet yang ditemukan dalam struktur kolom yang terpisah.

Berikut adalah penjelasan lebih lanjut mengenai bagian Columning dalam kode tersebut:

- A. Membuat sebuah list kosong yang disebut tweet_data yang akan digunakan untuk menyimpan data dari setiap tweet dalam bentuk kolom terpisah.
- B. Dalam perulangan for yang mengiterasi setiap tweet yang ditemukan melalui hasilSearch.get_items(), dilakukan langkah-langkah berikut:
 - a. Membuat sebuah list baru bernama tweet_list yang berisi tiga elemen: tanggal tweet (tweet.date), nama pengguna (tweet.user.username), dan isi tweet (tweet.content).
 - b. Menambahkan tweet_list ke dalam tweet_data. Dengan demikian, setiap tweet akan disimpan sebagai satu baris dalam tweet_data, dengan masing-masing elemen kolom terpisah.
 - c. Dalam kode yang Anda berikan, jika jumlah tweet yang telah diproses mencapai 500 (i == 500), proses perulangan dihentikan menggunakan pernyataan break.

Dengan cara ini, data dari setiap tweet yang ditemukan disimpan dalam list tweet_data, dengan setiap tweet mewakili satu baris dalam struktur data tersebut. Nantinya, kami dapat menggunakan tweet_data untuk melakukan analisis lanjutan atau menyimpan data tersebut dalam format yang sesuai, seperti file CSV.

```
1 tweet_df= pd.DataFrame(tweet_data, columns=['date', 'user', 'tweet'])
```

Lalu pada gambar di atas adalah kelanjutan dari proses Columning. Dalam kode tersebut, variabel tweet_data digunakan untuk membuat DataFrame menggunakan library pandas.

Berikut adalah penjelasan mengenai kode tersebut:

- 1. Import library pandas (biasanya diwakili dengan alias pd): import pandas as pd.
- 2. Membuat DataFrame baru dengan menggunakan pd.DataFrame(). DataFrame ini menggunakan tweet_data sebagai sumber datanya.
- 3. Menentukan kolom-kolom dalam DataFrame yang akan disesuaikan dengan struktur yang diinginkan. Dalam kode tersebut, kolom-kolom DataFrame ditentukan sebagai ['date', 'user', 'tweet']. Kolom-kolom ini akan menyimpan tanggal tweet, nama pengguna, dan isi tweet yang telah dikumpulkan sebelumnya.
- 4. Variabel tweet_df akan berisi DataFrame yang dibuat dengan kolom-kolom yang telah ditentukan. DataFrame ini dapat digunakan untuk melakukan analisis lanjutan, pemrosesan data lebih lanjut, atau menyimpan data dalam format yang sesuai, seperti file CSV.

Langkah selanjutnya adalah membuat DataFrame menggunakan library pandas dengan menggunakan variabel tweet_data sebagai sumber datanya. DataFrame ini memungkinkan pengolahan, eksplorasi, dan analisis data lebih lanjut.

2.3 Analisis Sentimen

1. Pembagian Label

```
[] 1 tweet_positif = [t for t in tweet_properties_list if t['sentiment'] == 'positive']
2 tweet_negatif = [t for t in tweet_properties_list if t['sentiment'] == 'negative']

[] 1 print(f*""Hasil Sentimen
2 Positif : [len(tweet_positif))
3 Negatif : [len(tweet_negatif))
4 """]

Hasil Sentimen
Positif : 148
Negatif : 352

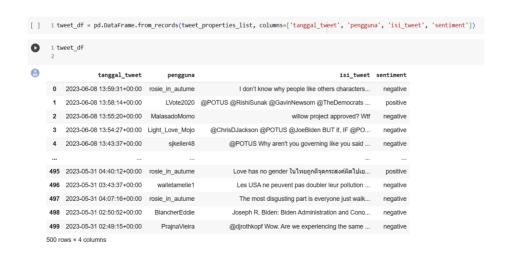
[] 1 print(tweet_properties_list)
[{'tanggal_tweet': datetime.datetime(2023, 6, 8, 13, 59, 31, tzinfo=datetime.timezone.utc), 'pengguna': 'rosie_in_autume', 'isi_tweet': 'I don't know why people like others characters
```

tweet_positif = [t for t in tweet_properties_list if t['sentiment'] == 'positive'] Pada baris ini, digunakan list comprehension untuk membuat list tweet_positif yang berisi tweet-tweet yang memiliki sentimen positif. Setiap tweet (t) dalam tweet_properties_list diperiksa, dan hanya tweet yang memiliki sentimen 'positive' yang akan dimasukkan ke dalam list tweet_positif.

tweet_negatif = [t for t in tweet_properties_list if t['sentiment'] == 'negative']

Pada baris ini, juga digunakan list comprehension untuk membuat list tweet_negatif yang berisi tweet-tweet yang memiliki sentimen negatif. Mirip dengan langkah sebelumnya, tweet-tweet yang memiliki sentimen 'negative' akan dimasukkan ke dalam list tweet_negatif.

Dengan menggunakan pemfilteran ini, dapat memisahkan tweet-tweet berdasarkan sentimen, sehingga memungkinkan untuk melakukan analisis yang lebih spesifik atau pengambilan keputusan berdasarkan sentimen yang terkait dengan setiap tweet.



BAB III

HASIL ANALISIS

3.1 Statistik Deskriptif

Pada tahap ini, dilakukan analisis statistik deskriptif untuk menggambarkan karakteristik data yang ada. Hal ini meliputi perhitungan statistik seperti rata-rata, median, modus, dan deviasi standar dari variabel yang relevan. Tujuannya adalah untuk memberikan pemahaman tentang sebaran data dan menganalisis tren atau pola yang muncul.

les.	tanggal_tweet	pengguna	isi_tweet	sentiment
count	500	500	500	500
unique	500	340	478	2
top	2023-06-09 12:53:53+00:00	rosie_in_autume	Joseph R. Biden: Biden Administration and Cono	negative
freq	1	37	7	351
first	2023-06-02 11:09:36+00:00	NaN	NaN	NaN
last	2023-06-09 12:53:53+00:00	NaN	NaN	NaN

Berikut adalah penjelasan yang lebih mudah dan sederhana untuk setiap baris dalam tabel:

- 1. tanggal_tweet : Ini menunjukkan tanggal dan waktu saat tweet diposting, dengan rentang dari 2 Juni 2023 hingga 9 Juni 2023.
- 2. pengguna: Ini mencantumkan nama pengguna Twitter yang membuat tweet. Terdapat 342 pengguna yang berbeda yang terlibat dalam analisis, artinya ada banyak orang yang berpartisipasi dalam percakapan ini.
- 3. isi_tweet : Ini adalah teks dari setiap tweet yang dianalisis. Ada 475 teks tweet yang berbeda dalam data yang dikumpulkan.
- 4. Sentiment: Ini mencerminkan sentimen yang terkait dengan setiap tweet. Terdapat dua kategori sentimen yang digunakan: "negative" (negatif) dan mungkin ada juga "positive" (positif) atau "neutral" (netral) yang tidak ditampilkan dalam tabel.

Informasi tambahan tentang setiap kolom:

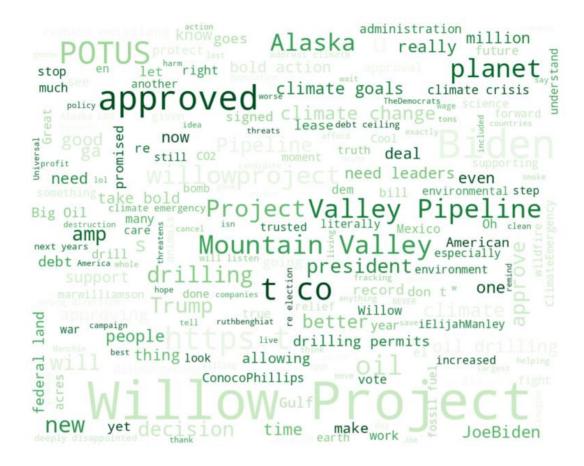
1. Count : Jumlah total entri dalam setiap kolom. Pada kasus ini, ada 500 entri untuk setiap kolom, artinya data tweet sejumlah 500.

- 2. Unique : Jumlah nilai unik dalam setiap kolom. Contohnya, ada 498 tanggal unik, 342 pengguna unik, 475 teks tweet unik, dan 2 kategori sentimen unik dalam data yang dianalisis.
- 3. Top: Nilai yang paling sering muncul dalam setiap kolom. Sebagai contoh, pengguna dengan nama "rosie_in_autume" muncul paling sering dalam data, dan sentimen negatif adalah yang paling umum dalam sampel ini.
- 4. Freq: Frekuensi kemunculan nilai yang paling sering dalam setiap kolom. Dalam contoh ini, tweet dengan sentimen negatif muncul sebanyak 352 kali dalam sampel.
- 5. First dan last: Menunjukkan tanggal dan waktu tweet pertama dan terakhir dalam data yang dianalisis. Namun, tidak ada informasi khusus yang diberikan untuk kolom "pengguna" dan "isi_tweet" pada baris ini.

3.2 Visualisasi

Visualisasi data adalah representasi grafis dari data atau informasi yang bertujuan untuk memperjelas, memahami, atau menyajikan pola, hubungan, atau karakteristik yang terkandung dalam data.

Word Cloud adalah representasi visual dari teks yang menampilkan kata-kata yang paling sering muncul dalam teks tersebut. Dalam word cloud, ukuran kata yang ditampilkan diatur berdasarkan frekuensi kemunculan kata tersebut dalam teks. Kata-kata yang lebih sering muncul akan ditampilkan dengan ukuran yang lebih besar, sementara kata-kata yang jarang muncul akan ditampilkan dengan ukuran yang lebih kecil atau bahkan tidak ditampilkan sama sekali. Word Cloud dapat memberikan gambaran visual yang intuitif tentang kata-kata kunci yang paling penting atau sering muncul dalam teks tersebut. Word cloud sentiment positif adalah sebagai berikut:



Word cloud sentiment negatif adalah sebagai berikut:

BAB IV

PENUTUP

4.1 Kesimpulan

Laporan ini menganalisis sentimen terkait proyek Willow melalui data Twitter dengan hashtag yang relevan. Penelitian mencakup persepsi dan tanggapan umum, sentimen, dan tren dalam sentimen di Twitter terkait Willow Project. Analisis data Twitter memberikan gambaran tentang penerimaan dan tanggapan publik terhadap Willow Project. Hasil analisis sentimen akan menunjukkan apakah sentimen cenderung positif, negatif, atau netral. Manfaatnya adalah memberikan wawasan kepada tim pengelola Willow Project untuk evaluasi dan peningkatan berdasarkan sentimen publik. Laporan ini juga menyoroti pentingnya analisis sentimen dan pemantauan respons publik dalam proyek serupa. Visualisasi data seperti word cloud digunakan untuk memperlihatkan kata-kata kunci yang sering muncul dalam konteks proyek tersebut. Word cloud memberikan pemahaman intuitif tentang kata-kata penting dalam proyek tersebut.

4.2 Saran

Dalam pembuatan laporan ini, Penulis menghargai konstruktifitas saran dan kritik dari para pembaca terhadap laporan ini. Permintaan penulis kepada pembaca yang telah membaca hasil analisis ini adalah untuk memberikan masukan yang berharga guna meningkatkan kualitas laporan ini baik saat ini maupun di masa mendatang. Penulis sangat mengharapkan kontribusi pembaca dalam memperbaiki penulisan, penyajian, dan penggunaan kosa kata yang lebih proporsional. Masukan yang diberikan akan sangat berarti dalam upaya peningkatan kualitas karya ini. Dengan demikian, laporan ini diharapkan dapat memberikan manfaat yang lebih baik bagi pembaca yang telah meluangkan waktu untuk membacanya. Terima kasih atas perhatian dan kerjasama pembaca dalam meningkatkan kualitas laporan ini.