

ANALISIS SENTIMEN PADA POSTINGAN MEDIA SOSIAL TWITTER DAN INSTAGRAM TERKAIT MANGROVE SURABAYA

abstract

Social media is one of the platforms where people express their opinions and experiences more freely. This also applies to opinions and experiences regarding tourist attractions, including mangroves in Surabaya. Sentiment analysis is the process of defining perceptions towards a topic, and through this analysis, we can assess how the public perceives the mangrove tourist attraction in Surabaya. Topic modeling, on the other hand, is the process of extracting topics from a series of texts. This can be used to search for evaluation aspects of mangroves in Surabaya itself. The approach used in sentiment analysis is a list of sentiment words and shadow labeling. The results showed that the approach of sentiment word lists had a more even distribution of sentiments, but there were many misclassifications. This is due to the limited amount of data in manual labeling. The approach in topic modeling is LDA, where three topics can be identified: an attractive tourist area with food exploration, the presence of plastic waste and the discovery of a baby's corpse, and the presence of plastic waste and deforestation. The best-performing prediction model is LSTM with an accuracy of 0.78.

Keywords: Surabaya mangrove, social media sentiment analysis, topic extraction

Abstraksi

Sosial media merupakan salah satu tempat masyarakat dalam menyampaikan pendapat dan pengalamannya dengan lebih bebas. Hal ini tidak luput dari pendapat dan pengalaman mengenai suatu tempat wisata, termasuk mangrove di Surabaya. Analisis sentimen merupakan proses pendefinisian persepsi terhadap suatu topik, dengan analisis ini kita dapat menilai bagaimana persepsi masyarakat terhadap tempat wisata mangrove di Surabaya. Topic modelling sendiri merupakan proses ekstraksi topik dari rangkaian teks. Hal ini dapat dimanfaatkan sebagai pencarian aspek evaluasi dari mangrove di Surabaya sendiri. Pendekatan yang dilakukan dalam analisis sentimen adalah daftar sentimen kata-kata dan shadow labelling. Didapatkan hasil bahwa, pendekatan daftar sentimen kata-kata memiliki distribusi sentimen yang lebih merata, tetapi banyak terjadi misklasifikasi. Hal ini disebabkan oleh jumlah data pada manual labelling yang terlalu sedikit. Pendekatan dalam topic modelling adalah LDA, dimana bisa ditarik 3 topik, yakni kawasan wisata yang menarik dan mencoba makanan, adanya sampah plastik dan penemuan jenazah bayi, adanya sampah plastik dan penebangan pohon. Adapun model prediksi dengan nilai akurasi terbaik adalah LSTM dengan nilai akurasi 0.78.

Kata Kunci: *mangrove surabaya, analisis sentimen sosial media, ekstraksi topik*

PENDAHULUAN

Latar Belakang Penelitian

Berwisata atau Traveling merupakan salah satu kebutuhan manusia yang tidak dapat dipisahkan dalam kehidupan manusia. Berwisata memiliki manfaat yang besar bagi

keberlangsungan hidup manusia. Tujuan dari wisata bukan hanya untuk mendapatkan kesenangan atau menenangkan pikiran dari aktivitas yang membuat manusia lelah, namun berwisata memungkinkan untuk membantu dalam pekerjaan seperti pramuwisata, fotografer perjalanan, dan lainnya. Namun tidak semua tempat wisata memiliki kondisi yang baik. Ada juga tempat wisata yang tidak dikelola dengan baik yang menyebabkan para traveler tidak ingin berkunjung. Perbaikan sektor pariwisata memiliki keuntungan yang signifikan. Perbaikan sektor pariwisata dapat mendorong pelestarian alam, budaya dan warisan suatu daerah. Dengan adanya pengembangan pariwisata yang berkelanjutan, sumber daya alam dan kebudayaan dapat dijaga dan dilestarikan untuk generasi mendatang. Perbaikan pariwisata juga dapat meningkatkan kesadaran akan pentingnya konservasi lingkungan dan memberikan insentif bagi pengembangan ekowisata. Selain itu, perbaikan sektor pariwisata juga mendorong pertumbuhan ekonomi. Sektor pariwisata yang baik akan mendorong wisatawan atau traveller untuk menghabiskan uang mereka. Hal ini akan meningkatkan pertumbuhan ekonomi baik di wilayah itu dan juga negara. Dengan ada banyak wisatawan yang berkunjung, maka akan membuka lapangan pekerjaan langsung dan tidak langsung. Banyak orang yang bekerja di sektor pariwisata, seperti pramuwisata, pemandu wisata, petugas hotel, sopir taksi, dan lainnya. Selain itu, industri pariwisata juga memberikan kesempatan bagi pelaku usaha lokal untuk mengembangkan bisnis mereka, seperti pengelolaan akomodasi, restoran, toko souvenir, dan jasa transportasi.

Indonesia merupakan wilayah yang kaya akan sumber daya alamnya. Negara ini kaya akan keindahan alam, budaya yang beragam, warisan sejarah, dan destinasi wisata yang menarik. Dari pantai-pantai eksotis di Bali, kekayaan kehidupan laut di Raja Ampat, keajaiban alam di Taman Nasional Komodo, hingga candi-candi kuno di Yogyakarta, Indonesia menawarkan pengalaman wisata yang memukau. Selain itu, budaya Indonesia yang kaya dan keramahan penduduknya menjadi daya tarik bagi wisatawan. Ragam adat, tarian, seni, dan festival yang unik merupakan daya tarik tersendiri bagi mereka yang ingin menjelajahi dan memahami keberagaman budaya.

Salah satu destinasi wisata adalah Mangrove Surabaya. Semakin bertambahnya jumlah penduduk, hutan mangrove di Kawasan pantai timur Surabaya terus mengalami penurunan baik dalam luas, panjang maupun jenis vegetasi mangrovenya. Hal ini membuat hasil tangkapan laut oleh nelayan ikut turun, saat ini nelayan hanya dapat menangkap 25 – 30 kg hasil laut per hari, jauh berbeda dengan perkiraan hasil tangkapan lima tahun sebelumnya, nelayan dapat menangkap 80 – 90 kg hasil tangkapan laut tiap harinya atau terjadi penurunan hasil tangkapan sebesar 66% - 69 %. Selain itu, tingkat kesadaran masyarakat akan menjaga kebersihan juga relatif rendah. Menurunnya luas hutan mangrove di Kawasan Pantai Timur Surabaya disebabkan oleh beberapa faktor yaitu Reklamasi lahan mangrove menjadi areal pertambakan dan pemukiman, mudahnya pengurusan izin penggunaan lahan, adanya timbunan sampah, serta penebangan liar oleh masyarakat.

Dengan adanya perkembangan teknologi informasi, Pengunjung wisata dapat memberikan komentar terhadap wisata yang dikunjungi. Salah satu aplikasi dan media sosial yang sering digunakan untuk melihat dan memberikan komentar wisata yaitu Twitter dan Instagram. Dalam aplikasi ini terdapat banyak informasi wisata dan komentar pengunjung yang sangat membantu untuk mengetahui kualitas dari wisata yang akan dikunjungi. Dengan demikian pengguna lain dapat mudah menemukan informasi apa yang sedang diinginkan.

Penelitian ini akan memanfaatkan analisis sentimental untuk mengengolah kata untuk melacak mood pengunjung wisata dari komentar pengunjung terhadap destinasi wisata Mangrove Surabaya. Metode yang digunakan dalam penelitian ini meliputi model individu dan LSTM.

Tujuan Penelitian

Adapun tujuan dalam melakukan penelitian ini adalah sebagai berikut :

- a. Mengetahui persepsi unggahan/cuitan terhadap wisata mangrove di Surabaya
- b. Mengetahui perbandingan metode analisis sentimen dengan metode dasar daftar kata negatif dan *shadow labelling*
- c. Mengetahui performa dari masing masing metode yang digunakan dalam analisis sentimen
- d. Mengetahui topik atau aspek negatif dalam wisata mangrove di Surabaya

Manfaat Penelitian

Adapun Manfaat dalam melakukan penelitian ini adalah sebagai berikut :

- a. bagi pemerintah: sebagai dasar pengambilan keputusan terkait kelanjutan wisata mangrove di Surabaya
- b. bagi pengelola tempat wisata: sebagai media untuk mengetahui aspek yang perlu diperbaiki dan evaluasi keberhasilan pengelolaan

Batasan Penelitian

Adapun batasan masalah dalam melakukan penelitian ini yaitu :

- a. Data yang digunakan adalah data unggahan/cuitan pada instagram dan twitter dengan kata kunci yang digunakan, sebagai berikut:
 - i. Instagram: mangrove gunung anyar, mangrove medokan ayu, mangrove wonorejo, mangrove surabaya, lokasi wisata mangrove gunung anyar
 - ii. twitter: mangrove surabaya, mangrove gunung anyar, mangrove medokan ayu, mangrove wonorejo
- b. Data pendukung yang digunakan, antara lain:
 - i. sentimen kata-kata dalam bahasa Indonesia:
https://www.researchgate.net/publication/321757985_InSet_Lexicon_Evaluati_on_of_a_Word_List_for_Indonesian_Sentiment_Analysis_in_Microblogs
 - ii. sentimen kata-kata dalam bahasa Inggris:
[sentiment-analysis/SemEval2015-English-Twitter-Lexicon.txt at master · alyaxey/sentiment-analysis · GitHub](https://github.com/alyaxey/sentiment-analysis/SemEval2015-English-Twitter-Lexicon.txt)
 - iii. daftar kata alay: <https://ieeexplore.ieee.org/abstract/document/8629151>
 - iv. daftar stopwords: nltk dan sastrawi

BAB II TINJAUAN PUSTAKA

Machine Learning

Machine learning merupakan salah satu ilmu yang dimana membuat sistem yang dapat belajar dengan sendiri. Machine Learning sendiri merupakan salah satu disiplin ilmu

dalam kecerdasan buatan atau yang sering biasa dikenal dengan Artificial Intelligent (AI). Machine Learning juga sering disebut dengan Artificial Intelligent (AI) konvensional karena merupakan kumpulan metode-metode yang digunakan dalam penerapannya. Sebelum sistem dapat Machine learning dapat bekerja, terlebih dahulu dibutuhkan data untuk melatih (training) yang kemudian hasil dari latihan (*training*) akan diuji atau di test (*testing*) dengan data yang ada.

Text Mining

Salah satu ilmu yang memanfaatkan Machine learning adalah text mining . Text mining merupakan proses menambang data berupa informasi dan pengetahuan yang berguna dimana sumber data didapatkan dari dokumen atau teks, seperti dokumen Word, PDF, kutipan teks, atau sebagainya. Text mining memiliki tujuan untuk mencari kata-kata dan mendapatkan informasi yang berguna dimana informasi tersebut dapat mewakili isi dari dokumen yang berkaitan sehingga dapat dilakukan analisa yang berhubungan antar dokumen. Text mining merupakan penerapan konsep dan teknik data mining untuk mencari pola dalam teks, yaitu proses penganalisisan teks guna mendapatkan informasi yang bermanfaat untuk tujuan tertentu. Ada beberapa teknik dalam text mining diantaranya adalah :

a. Information Extraction (IE)

Langkah awal untuk mengambil informasi yang ada yaitu dengan Identifikasi terhadap hubungan dan frase-frase kunci dalam teks dengan mencari urutan yang sudah ditetapkan dalam text menggunakan pencocokan pola.

b. Topic Tracking (Pelacakan Topik)

Berdasarkan pada profil user dan berbagai dokumen yang dilihat user, text mining bisa memprediksi dokumen-dokumen lain yang menjadi perhatian/minat user tersebut.

c. Summarization (Peringkasan)

Meringkas suatu dokumen untuk menghemat waktu dari sisi pembaca.

d. Clustering (Penggugusan)

Mengelompokkan dokumen-dokumen yang mirip tanpa memiliki kategori yang sudah ditetapkan sebelumnya.

e. Concept Linking (Penautan Konsep)

Menghubungkan berbagai dokumen terkait dengan mengidentifikasi konsep yang digunakan bersama dan dengan demikian membantu para user untuk menemukan informasi yang barangkali mereka tidak akan temukan dengan menggunakan metode-metode pencarian tradisional.

f. Question Answering (Penjawab Otomatis)

Menemukan jawaban terbaik pada pertanyaan yang diberikan melalui pencocokan pola berbasis knowledge

Sentimen Analisis

Sentimen analisis adalah proses menganalisis teks digital untuk mendapatkan informasi sentimen yang terkandung dalam kalimat atau teks. Tujuan dari sentimen analisis adalah untuk menentukan apakah kalimat atau teks tersebut cenderung berisikan pandangan negatif ,netral, atau positif. Sentiment Analysis dapat dibedakan berdasarkan sumber datanya, beberapa level yang sering digunakan dalam penelitian analisis sentimen adalah analisis

sentimen pada level dokumen dan Sentiment Analysis pada level kalimat (Clayton, 2011). Berdasarkan level sumber datanya, analisis sentimen terbagi menjadi 2 kelompok besar yaitu:

1. Coarse-grained Sentiment Analysis

Pada Sentiment Analysis jenis ini, Sentiment Analysis yang dilakukan adalah pada level dokumen. Secara garis besar fokus utama dari Sentiment Analysis jenis ini adalah menganggap seluruh isi dokumen sebagai sebuah sentimen positif atau sentimen negatif (Clayton, 2011).

2. Fined-grained Sentiment Analysis

Fined-grained Sentiment Analysis adalah Sentiment Analysis pada level kalimat. Fokus utama fined-grained Sentiment Analysis adalah menentukan sentimen pada setiap kalimat pada suatu dokumen, dimana kemungkinan yang terjadi adalah terdapat sentiment pada level kalimat yang berbeda pada suatu dokumen (Clayton, 2011).

Penelitian Terdahulu

Hans Juwiantho (2020) dalam jurnalnya melakukan analisis sentimen pada 999 data media sosial twitter. Metode yang digunakan adalah deep-CNN. Akurasi terbaik yang didapatkan adalah 76,40%. Adapun metode penelitian yang dilakukan, sebagai berikut:

1. Preprocessing: dilakukan dengan manual labelling, tokenisasi, dan *stopword removal*
2. Penerapan *Word2Vec*
3. Pembentukan model deep-CNN dengan optimasi adam

Didapatkan hasil berupa distribusi sentimen adalah 419 positif, 453 negatif, dan 127 netral.

Dilakukan juga beberapa skenario pengujian, antara lain:

1. dimensi pada *Word2Vec* dengan ukuran filter 100: didapatkan hasil bahwa dimensi 100 memiliki akurasi terbaik
2. dimensi pada *Word2Vec* dengan ukuran filter 50: didapatkan hasil bahwa dimensi 100 memiliki akurasi terbaik
3. model NB, SVM, dan DCNN: didapatkan hasil DCNN memiliki nilai akurasi terbaik

Keterbaharuan

Keterbaharuan yang dilakukan dalam penelitian ini, sebagai berikut:

1. Data yang lebih variatif, yakni dari Instagram dan Twitter. Selain itu, jumlah data yang diproses adalah 2853
2. Preprocessing pada data: dilakukan beberapa preprocessing tambahan, yakni konversi emoji, alay-removal (merubah kata-kata alay dengan kata formal), penghapusan atau perubahan kata-kata *slang*
3. Labelling dilakukan dengan dua cara:
 - a. shadow labelling: dilakukan labelling secara manual terhadap 200 data, berikutnya data akan digunakan sebagai training model untuk menentukan sentimen pada data lain
 - b. basis daftar kata positif dan negatif: menggunakan daftar kata bersentimen positif dan negatif, baik dalam bahasa Indonesia, maupun bahasa Inggris
4. Pengambilan topik dengan metode lda
5. Model training yang digunakan adalah LSTM, B-LSTM, model individu (SVM, GNB, LR, LDA, dan DTREE), ensemble learning (max voting, averaging, dan weighted averaging)

BAB III METODE PENELITIAN

Dataset

Dataset didapatkan dengan melakukan *scraping* instagram dan twitter dengan keyword yang digunakan, antara lain:

1. Instagram: mangrove gunung anyar, mangrove medokan ayu, mangrove wonorejo, mangrove surabaya, lokasi wisata mangrove gunung anyar
2. twitter: mangrove surabaya, mangrove gunung anyar, mangrove medokan ayu, mangrove wonorejo

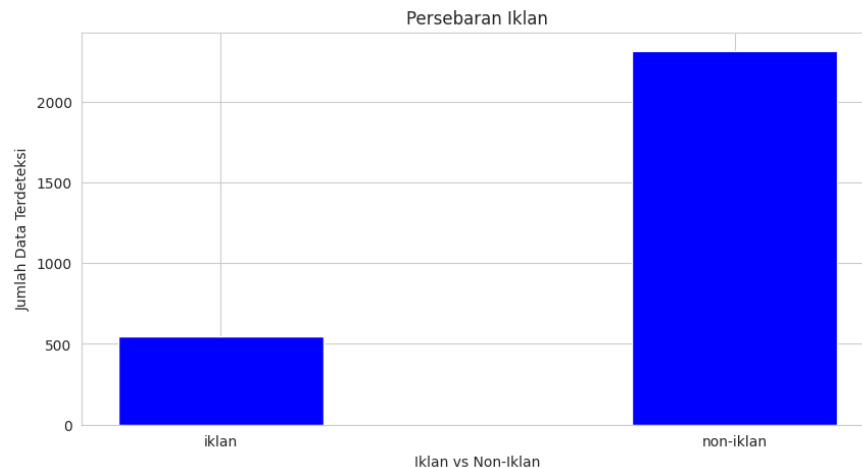
Dari scraping tersebut didapatkan sebanyak 2186 data instagram dan 723 data twitter. Berikut merupakan salah satu contoh data hasil *scraping*

Kolom	Pengertian	Contoh	Makna
link	url untuk scraping komentar dari twitter dan instagram	https://www.instagram.com/p/CNjbER3guqZ/;	Link untuk scraping komentar dari platform instagram
hasil	Hasil scraping komentar yang belum di cleaning	Awalnya nyaman, Ujung2 nya paylater 😂 #😂\n\n#wisatagratis #wisatasurabaya #wisatamangrove #mangrovegununganyar #ekowisatamangrove #mangrovesurabaya #bosemmewah #wisatadekatrumah #reelinstagram #video #photo\n\n#	Scraping komentar

Metode Penelitian

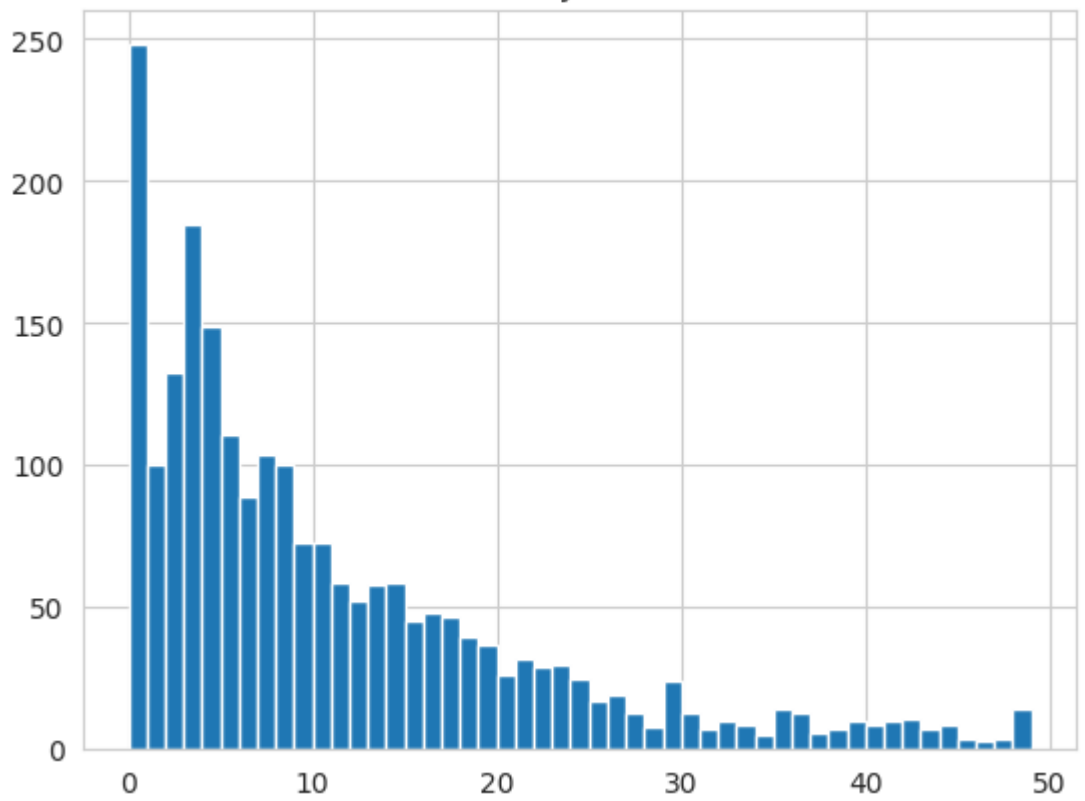
1. Preprocessing data
 - a. konversi emoji: digunakan library emoji dimana akan melakukan konversi emoji menjadi teks. Adapun, hasil yang didapatkan menggunakan bahasa Inggris. Sehingga, dalam preproses berikutnya dilakukan penyesuaian, yakni menyertakan penghapusan stopwords bahasa Inggris dan daftar sentimen bahasa Inggris
 - b. penghapusan beberapa karakter: sesuai pada tabel di atas, terlihat hasil scraping yang masih kotor, dimana terdapat beberapa karakter yang tidak diperlukan, sebagai contoh \n.

- c. konversi kata alay: dilakukan konversi kata alay menjadi formal dengan menggunakan data kamus alay. Hal ini diharapkan pemodelan dapat menangkap semua kata dalam data.
- d. penghapusan stopwords: stopwords merupakan kata-kata yang dirasa tidak memiliki makna yang kuat, kebanyakan kata yang termasuk stopwords adalah kata konjungsi atau kata depan. Contoh stopwords adalah “dari” dan “yang”
- e. dilakukan analisis terhadap kata-kata yang sering muncul. Dari analisis ini, dilakukan penghapusan kata-kata yang tidak berkaitan dengan analisis, sebagai contoh nama tempat mangrove, “mangrove”, “surabaya”, “gunung”, dan “wonorejo”
- f. penghapusan data tidak terkait: dalam *scraping* didapatkan data-data yang tidak berkaitan, utamanya terkait online shop. Sebagai contoh tautan berikut <https://www.instagram.com/p/CtLOTj3PxNm/> . Sehingga, perlu dilakukan penghapusan. Adapun, pendekatan yang digunakan dalam penelitian ini adalah deteksi keberadaan nomor telepon dan beberapa kata yang merujuk pada promosi. Ditemukan sebanyak 544 data terdeteksi promosi.

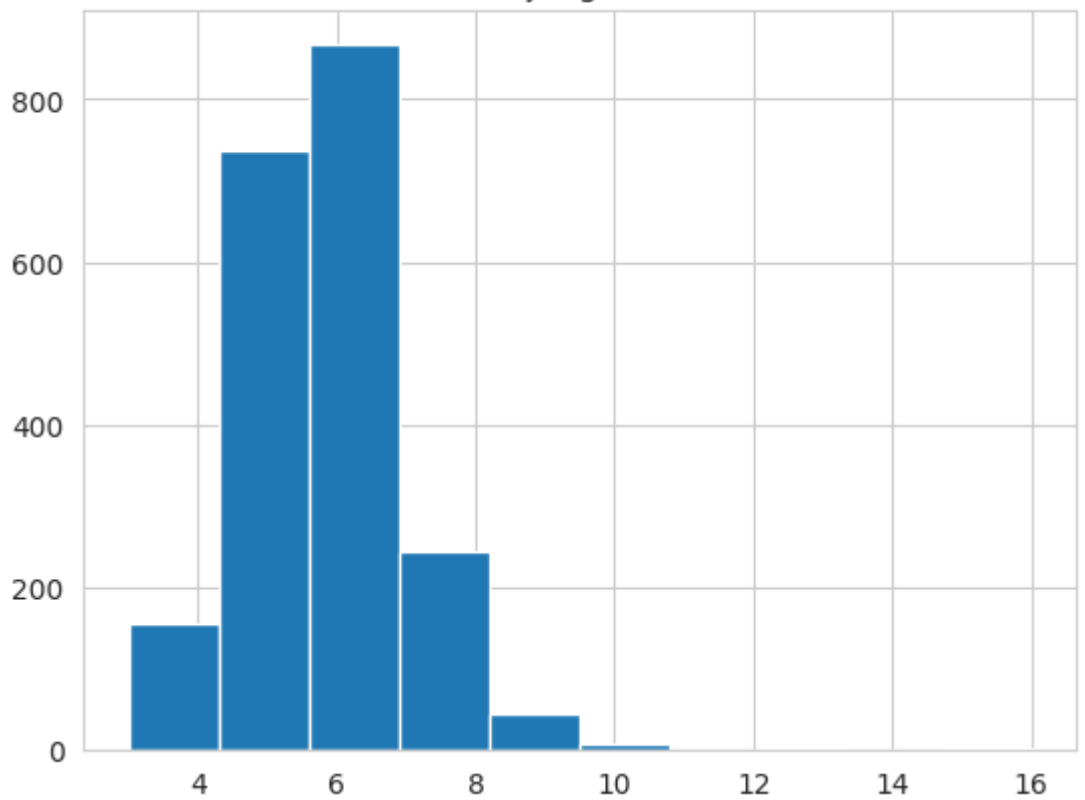


- g. dilakukan analisis kuantitatif terhadap data, terdiri dari persebaran jumlah huruf, jumlah kata, dan panjang kata rata-rata. Distribusi dapat dilihat pada grafik berikut:

Distribusi Jumlah Kata



Distribusi Panjang Kata Rata-Rata



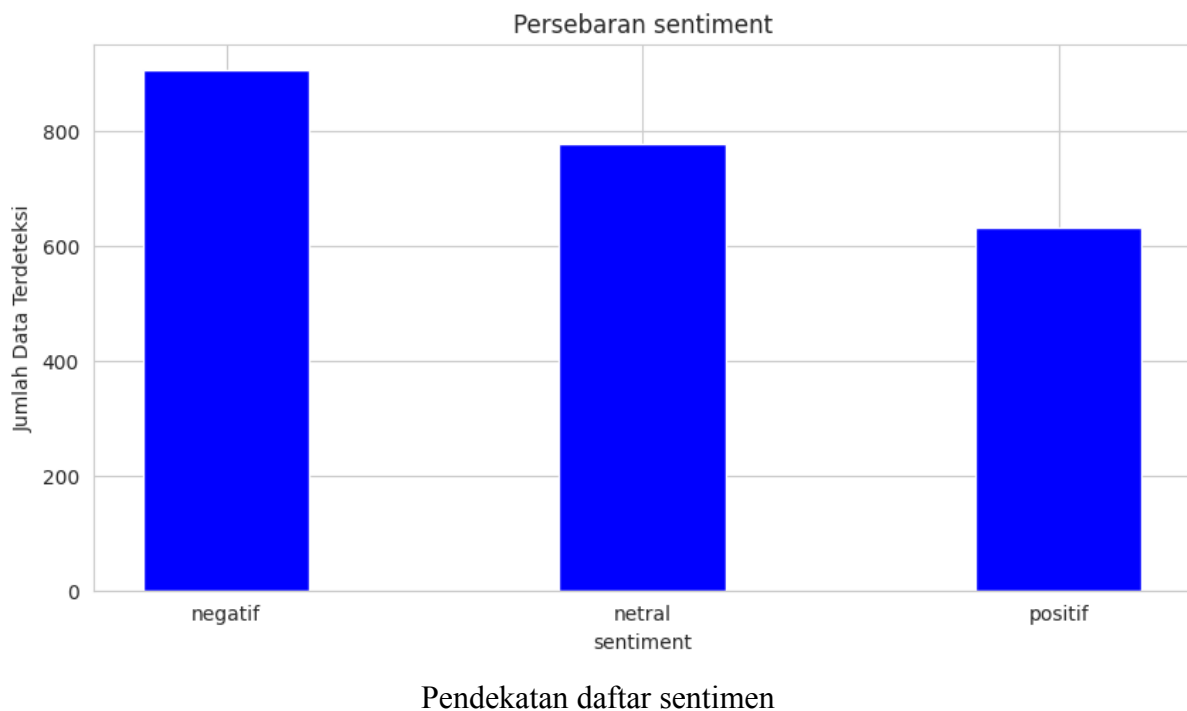
- b. Pendekatan shadow labelling
Dilakukan labeling manual terhadap 200 data. Dikembangkan model BLSTM dengan 200 data tadi, yang akan digunakan sebagai model prediksi untuk data lainnya.
3. Topik modelling
Digunakan metode lda (latent dirichlet allocation) merupakan metode yang populer dan sering digunakan dalam topic modelling. Dilakukan hyperparameter tuning terhadap parameter jumlah topik, alpha, beta, validation sets, dan corpus title. Proses ini diharapkan bisa mendapatkan topik dengan nilai koherensi terbaik.
4. Pembuatan model prediksi
Model yang dibuat adalah LSTM, BLSTM, SVM, DTREE, GNB, LDA, LR, max voting, averaging, dan weighted averaging. Nilai evaluasi yang digunakan adalah akurasi.

BAB IV

HASIL DAN PEMBAHASAN

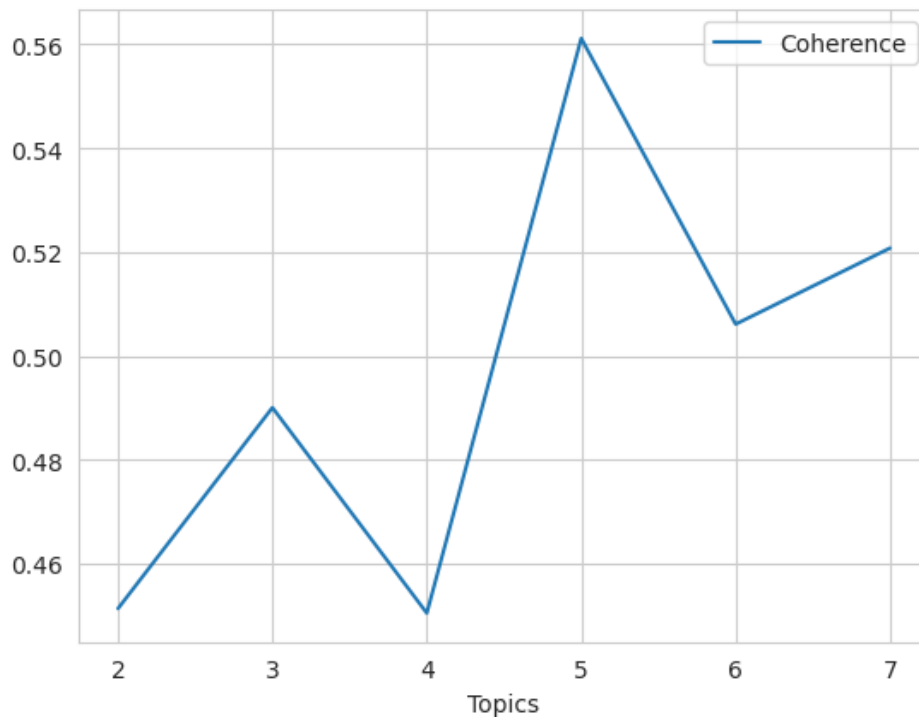
Distribusi Sentimen

Sebagaimana dijelaskan pada bab sebelumnya, dilakukan dua pendekatan dalam menganalisis sentimen, yakni dengan daftar sentimen dan shadow labelling. Distribusi sentimen dari masing-masing pendekatan sebagai berikut:



Ekstraksi Topik

Dari proses hyper parameter tuning yang dilakukan didapatkan grafik koherensi sebagai berikut:



Dari grafik tersebut jumlah topik 5 memiliki koherensi paling tinggi, sehingga akan dilakukan ekstraksi topik sebanyak 5 topik. Didapatkan topik sebagai berikut:

```
[ (0,
  '0.022*"wisata" + 0.014*"face" + 0.013*"ayo" + 0.011*"kawasan" + '
  '0.010*"coba" + 0.010*"video" + 0.010*"menarik" + 0.010*"mencoba" + '
  '0.010*"cek" + 0.009*"makan"'),
(1,
  '0.028*"minus" + 0.023*"preservation" + 0.012*"sih" + 0.012*"looking" + '
  '0.012*"support" + 0.012*"denmark" + 0.012*"carbon" + 0.012*"climate" + '
  '0.012*"works" + 0.012*"change"'),
(2,
  '0.017*"sampah" + 0.016*"plastik" + 0.014*"ditemukan" + 0.013*"bayi" + '
  '0.013*"jenazah" + 0.007*"heart" + 0.006*"fire" + 0.006*"face" + '
  '0.004*"bahan" + 0.004*"sambal"'),
(3,
  '0.014*"foto" + 0.014*"take" + 0.014*"sampah" + 0.013*"plastik" + '
  '0.013*"penebangan" + 0.012*"terancam" + 0.012*"nasib" + 0.008*"face" + '
  '0.007*"pohon" + 0.007*"smiling"'),
(4,
  '0.009*"wisata" + 0.008*"gym" + 0.006*"time" + 0.005*"arah" + 0.005*"pemkot" '
  '+ 0.004*"nby" + 0.004*"kawasan" + 0.004*"nsurabaya" + 0.004*"kota" + '
  '0.004*"xx"') ]
```

Pembuatan Model Prediksi

Dilakukan pembuatan model dan didapatkan hasil akurasi sebagai berikut:

Model	Akurasi	Model	Akurasi
-------	---------	-------	---------

LSTM	0.779697597026825	LR	0.723404255319149
BLSTM	0.758099377155304	GNB	0.624113475177305
SVM	0.7186761229314421	Max voting	0.712431931531292
DTREE	0.7163120567375887	Averaging	0.7446808510638298
LDA	0.6643026004728132	Weighted Averaging	0.7423167848699763

Dari hasil tersebut didapatkan 3 model dengan akurasi terbaik adalah LSTM, BLSTM, dan averaging. Meski memiliki akurasi terbaik, model LSTM dan BLSTM dapat dikatakan sebagai overfitting. Hal ini dapat dilihat dengan kondisi nilai loss yang justru semakin menguat. Sehingga dikhawatirkan tidak dapat menangani analisis sentimen pada kasus ril.

BAB V KESIMPULAN

Distribusi persepsi pengguna media sosial, yakni 905 negatif, 631 netral, dan 776 positif menunjukkan bahwa terdapat persepsi negatif yang sangat besar terhadap wisata mangrove di Surabaya. Namun, dibalik itu pendekatan daftar sentimen kata-kata bahasa Indonesia dan bahasa Inggris perlu dipertimbangkan. Pada pendekatan ini seringkali terjadi misklasifikasi terutama terhadap data netral dan positif yang terklasifikasikan menjadi negatif. Adapun, pendekatan shadow labelling dirasa masih kurang dikarenakan jumlah labelling manual yang terlalu sedikit, sehingga terdapat ketimpangan yang sangat kuat. Oleh sebab itu, dalam penelitian berikutnya dapat dilakukan shadow labelling dengan data label manual yang lebih banyak.

Dari topic modelling didapatkan 5 topik dengan koherensi terbaik, yakni:

```
[0,
  '0.022*"wisata" + 0.014*"face" + 0.013*"ayo" + 0.011*"kawasan" + '
  '0.010*"coba" + 0.010*"video" + 0.010*"menarik" + 0.010*"mencoba" + '
  '0.010*"cek" + 0.009*"makan"'),
(1,
  '0.028*"minus" + 0.023*"preservation" + 0.012*"sih" + 0.012*"looking" + '
  '0.012*"support" + 0.012*"denmark" + 0.012*"carbon" + 0.012*"climate" + '
  '0.012*"works" + 0.012*"change"'),
(2,
  '0.017*"sampah" + 0.016*"plastik" + 0.014*"ditemukan" + 0.013*"bayi" + '
  '0.013*"jenazah" + 0.007*"heart" + 0.006*"fire" + 0.006*"face" + '
  '0.004*"bahan" + 0.004*"sambal"'),
(3,
  '0.014*"foto" + 0.014*"take" + 0.014*"sampah" + 0.013*"plastik" + '
  '0.013*"penebangan" + 0.012*"terancam" + 0.012*"nasib" + 0.008*"face" + '
  '0.007*"pohon" + 0.007*"smiling"'),
(4,
  '0.009*"wisata" + 0.008*"gym" + 0.006*"time" + 0.005*"arah" + 0.005*"pemkot" '
  '+ 0.004*"nby" + 0.004*"kawasan" + 0.004*"nsurabaya" + 0.004*"kota" + '
  '0.004*"xx"')]
```

Sehingga, topik yang dapat ditarik adalah:

1. kawasan wisata yang menarik dan mencoba makanan
2. *tidak dapat ditarik topik*
3. adanya sampah plastik dan penemuan jenazah bayi
4. adanya sampah plastik dan penebangan pohon
5. *tidak dapat ditarik topik*

Dari hasil ini, dapat ditarik kesimpulan permasalahan utama yang terjadi adalah sampah plastik dan adanya berita penemuan jenazah bayi. Kedepannya, hal ini bisa digunakan sebagai evaluasi dari pihak pengurus wisata.

Model LSTM mendapatkan hasil yang terbaik, yakni dengan nilai akurasi 0.779697597026825 . Meski begitu, perlu digaris bawahi dalam model ini terjadi overfitting. Sehingga penggunaan model averaging lebih dianjurkan dengan harapan mampu menangani masalah ril.

REFERENSI

- https://www.researchgate.net/publication/349628900_Sentiment_Analysis_using_Neural_Network_and_LSTM
- <https://jurnal.ugm.ac.id/ijccs/article/view/41236/25073>
- <https://www.mdpi.com/2076-3417/12/5/2694>
- https://www.academia.edu/43056511/IJERT_Text_based_Sentiment_Analysis_using_LSTM
- <https://jtiik.ub.ac.id/index.php/jtiik/article/view/1758/pdf>
- <https://www.sleepinggiantmedia.co.uk/why-do-people-complain-on-social-media/>
- http://eprints.upnjatim.ac.id/1261/2/2._Jurnal_Hendra.pdf