

## ENSEMBLE STACKING DALAM ANALISA SENTIMEN REAKSI VETERAN MILITER AS TERHADAP PENGAMBILALIHAN AFGHANISTAN OLEH TALIBAN

Henny Leidiyana

Program Studi Sistem Informasi  
Universitas Bina Sarana Informatika  
Email: [henny.hnl@bsi.ac.id](mailto:henny.hnl@bsi.ac.id)



**Abstract**—Sentiment analysis can be used to glean information about user opinions and identify social or political trends. There have been many studies on sentiment analysis using machine learning or lexicon-based methods that have been quite impressive. However, machine learning models often have difficulty generalizing to new data due to various reasons, such as overfitting and limited training data. These models are also prone to bias and variance, which negatively affect the accuracy of their predictions. This study discusses the application of the ensemble stacking method in sentiment analysis with the topic of the takeover of Afghanistan by the Taliban. By monitoring social media, the author uses a dataset in the form of comments on YouTube news channels related to the topic raised. Several studies have shown how the ensemble stacking method predicts better than the single model. The research was carried out by creating a sentiment classification model with logistic regression machine learning algorithms, SVM, KNN, and CART then the ensemble stacking classifier formed by the base learner of the four algorithms. As a result, for a single classifier, the highest average accuracy is the logistic regression algorithm of 74.6 percent. The four algorithms are compiled and predicted by logistic regression, and the stacking ensemble classifier that is applied produces better accuracy than the stand-alone classifier, which is 75.3 percent.

**Keywords:** Ensemble, Stacking, base learner

**Abstrak**—Analisis sentimen dapat digunakan untuk memperoleh informasi tentang pendapat pengguna serta mengidentifikasi tren sosial atau pun politik. Terdapat banyak penelitian tentang analisis sentimen menggunakan metode pembelajaran mesin maupun berbasis leksikon yang berhasil dibuat dengan cukup mengesankan. Namun model pembelajaran mesin sering kesulitan dalam generalisasi ke data baru yang disebabkan oleh berbagai hal seperti *overfitting* dan data pelatihan yang terbatas. Model-model ini juga rentan terhadap bias dan varians, yang berdampak negatif terhadap akurasi prediksinya. Penelitian ini membahas tentang penerapan metode *ensemble stacking* dalam analisa sentimen dengan topik pengambilalihan Afghanistan oleh Taliban. Dengan melakukan pemantauan terhadap media sosial, penulis menggunakan *dataset* berupa komentar pada kanal berita *YouTube* yang terkait topik yang diangkat. Beberapa penelitian telah menunjukkan bagaimana metode *ensemble stacking* memprediksi lebih baik daripada model tunggal. Penelitian yang dilakukan yaitu membuat model klasifikasi sentimen dengan algoritma pembelajaran mesin regresi logistik, SVM, KNN, dan CART kemudian pengklasifikasi *ensemble stacking* yang dibentuk oleh *base learner* empat algoritma tersebut. Hasilnya, untuk pengklasifikasi tunggal, akurasi rata-rata paling tinggi yaitu algoritma regresi logistik sebesar 74,6 persen. Keempat algoritma tersebut disusun dan diprediksi dengan regresi logistik, dan pengklasifikasi *ensembl stacking* yang diterapkan menghasilkan akurasi lebih baik daripada pengklasifikasi yang berdiri sendiri yaitu sebesar 75,3 persen.

**Kata Kunci:** Ensemble, Stacking, base learner

### PENDAHULUAN

Setelah selama satu bulan Taliban terus menggempur tentara pemerintah dan merebut

kota-kota penting, akhirnya berhasil menguasai Kabul dan mengambil alih Istana Kepresidenan Afghanistan pada 15 Agustus 2021. Taliban menyatakan perang di Afghanistan sudah berakhir

setelah peperangan selama 20 tahun (*Taliban Duduki Istana Presiden: Perang Afghanistan Berakhir*, n.d.). Informasi ini tersebar melalui pemberitaan di berbagai media termasuk media sosial. Kejadian ini menimbulkan berbagai reaksi yang membawa perasaan rumit khususnya bagi veteran militer AS. Banyak veteran AS mengungkapkan frustrasi dan kekecewaan dengan seberapa cepat Taliban menguasai setelah hampir 20 tahun perang (Andone, 2021). Beberapa veteran siap untuk mengakhiri perang sementara yang lain ingin AS tetap tinggal. Media sosial menjadi tempat dimana mereka menyatakan opini dan mencurahkan perasaan.

Kehadiran media sosial telah mengubah lanskap informasi dalam berbagai cara. Media sosial sekarang menjadi salah satu jalur paling umum di mana orang mendapatkan berita politik (Matalon et al., 2021). Data berbentuk teks saat ini membanjiri internet dalam bentuk situs, blog, media sosial, forum berisi berbagai informasi. Informasi yang tidak terstruktur dapat diubah menjadi data yang lebih terstruktur dengan analisa sentimen. Topik penelitian tentang analisa sentimen terhadap data dari media sosial saat ini banyak dilakukan di berbagai bidang bisnis (Sri, 2021), politik, seperti penelitian yang dilakukan Mohd Zeeshan Ansaria yang menggunakan data dari twitter untuk analisa sentimen politik nasional partai politik India sehubungan dengan pemilihan umum di India (Ansari et al., 2020).

Analisa sentimen menggunakan media sosial *YouTube* juga sudah banyak digunakan antara lain penelitian analisa sentimen pada komentar *YouTube* untuk melihat tren, musim, dan prakiraannya dapat memberikan gambaran transparan tentang pengaruh peristiwa dunia nyata terhadap sentimen public (Singh, 2021). Penelitian yang berkaitan dengan Taliban pun sudah pernah dilakukan, yaitu pada penelitian tentang analisa sentimen mengenai situasi Afghanistan saat ini di bawah kendali Taliban, menggunakan data dari twitter untuk pemodelan topik dengan metode LDA. (Lee et al., 2022).

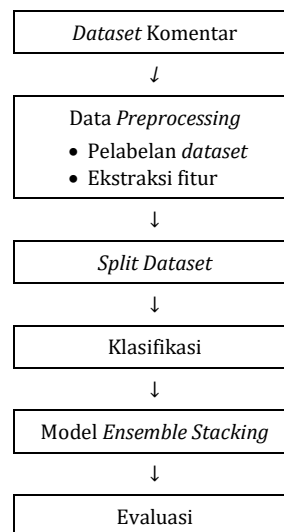
Banyak algoritma yang diusulkan untuk analisa sentimen termasuk metode *ensemble*, yang memiliki potensi untuk mengurangi tingkat kesalahan dari sebuah model klasifikasi (Görmez et al., 2020). Salah satu penelitian yang menggunakan metode ensemble dimana pengklasifikasi *ensemble* menggabungkan pengklasifikasi tunggal, dengan tujuan untuk meningkatkan kinerja dan akurasi teknik klasifikasi sentiment menghasilkan

pengklasifikasi *ensemble* yang diusulkan memiliki kinerja lebih baik daripada pengklasifikasi tunggal (Ankit & Saleena, 2018). Peningkatan kinerja model analisa sentimen juga dapat ditemukan pada penelitian (Muhammad Fikri et al., 2022)(BAŞARSLAN & KAYAALP, 2022)(Priyanka & Ashok Kumar, 2020)(Mahakul, 2020). Dari penelitian mereka diperoleh bahwa performansi metode *ensemble* terbukti dapat mengungguli metode pembelajaran mesin *Naive Bayes*, knn, pohon keputusan, *MultinomialNB*, *BernoulliNB*, SVM, dan regresi logistik.

Pada penelitian ini akan dilakukan analisa sentimen menggunakan data komentar *YouTube* kanal berita AS berbahasa Inggris terkait reaksi veteran militer AS yang pernah bertugas saat AS menduduki Afghanistan, menggunakan model pembelajaran mesin Regresi logistic, *Support Vector Machine* (SVM), K-Nearest Neighbor (kNN, dan pohon keputusan (*Decision Tree*). Selanjutnya penelitian ini juga menggunakan metode *ensemble* yaitu *stacking* dengan basis empat model di atas untuk membuktikan apakah penerapan *ensemble stacking* dapat meningkatkan kinerja pengklasifikasi sentiment seperti pada penelitian-penelitian terkait sebelumnya.

## BAHAN DAN METODE

Tahapan penelitian dapat dilihat pada Gambar 1. Penelitian diawali dengan Pengumpulan *dataset* komentar. *Dataset* yang digunakan untuk percobaan yaitu komentar video-video dari kanal-kanal *YouTube* berita televisi tentang reaksi veteran militer AS yang bertugas di Afghanistan Terhadap Pengambilalihan Afghanistan oleh Taliban.



Gambar 1. Tahapan Penelitian

Tahap pertama yaitu persiapan data sebanyak 2803 komentar yang belum memiliki label. Dataset tidak berlabel karena menggunakan API sehingga hanya dapat mengekstrak komentar. Tahap kedua yaitu *preprocessing* yang diawali dengan pelabelan data komentar dengan teknik VADER (*valence aware dictionary and sentiment reasoned*). Setelah pelabelan maka dilakukan ekstraksi fitur yaitu vektorisasi dataset dengan Count Vectorizer. Tahap ketiga yaitu data yang telah divektorisasi digunakan untuk pembuatan model klasifikasi menggunakan Regresi logistik, Support Vector Machine (SVC), K-Nearest Neighbor, dan Decision Tree yang dilanjutkan dengan evaluasi. Berikutnya penerapan metode *ensemble Stacking* dan dilakukan evaluasi kemudian penulis melakukan analisa hasil setelah penerapan *Stacking*. Evaluasi menggunakan metode *K-Fold Cross Validation*.

*Stacking* adalah algoritma pembelajaran mesin ensemble yang menggabungkan prediksi dari beberapa model pembelajaran mesin. Arsitektur *model Stacking* menggunakan dua atau lebih base model yang sering disebut dengan model level-0, dan meta model yang menggabungkan prediksi base model yang disebut dengan model level-1. (Görmez et al., 2020) Percobaan dalam penelitian menggunakan meta-model Regresi logistik dengan lima base-model, yaitu Regresi logistik, *Support Vector Machine* (SVM), K-Nearest Neighbor, Decision Tree.

Tabel 1. *Sample Dataset* Komentar

Name	Komentar
PT08	<i>US Spy in foreign countries</i>
Bashir Sardar	<i>Come on now help Ukraine</i>
motherhoods beauty	<i>It's USA fault for sticking their nose in other people business</i>
Hamid Ashrafi	<i>Coward evil terroriste nation</i>
Elpapiiriin tuyo	<i>Has anyone ever thought that the USA government deliverate</i>
M Ansari	<i>I fully agree and understand Joe's frustration anger</i>
Stuff With Cleatus Automotive	<i>It's a bad situation all together been past year</i>

Sumber: (BBC, 2021)

**HASIL DAN PEMBAHASAN**

Data untuk pembuatan model harus berkualitas. Data komentar merupakan data teks yang sering kali memiliki informasi yang tidak perlu dan berlebihan sehingga tidak ada kontribusinya terhadap prediksi kelas target. Oleh karena itu perlu langkah *preprocessing* untuk

membersihkan data. Langkah-langkah berikut diikuti dalam *preprocessing*:

1. Hastag dan penghapusan tautan: Komentar berisi tag dan tautan yang tidak diperlukan untuk melatih model sehingga dihapus dari kumpulan data untuk mengurangi kerumitan. Tag dan tautan dihapus menggunakan ekspresi reguler.
2. Penghapusan tanda baca dan nomor: Tanda baca adalah bagian penting dari kalimat agar lebih mudah dipahami dan bermakna bagi pembaca manusia. Namun untuk model pembelajaran mesin, tanda baca menciptakan komplikasi ekstra dan meningkatkan ukuran dari fitur vektor. Tanda baca dihapus dari komentar karena tidak berkontribusi pada proses pelatihan.
3. Konversi ke huruf kecil: Model pembelajaran mesin peka terhadap huruf besar/kecil sehingga perbedaan huruf besar/kecil menyebabkan kerumitan dalam pelatihan model. Fungsi bawaan Python digunakan untuk konversi ke huruf kecil.
4. *Stemming* dan lemmatisasi: *Stemming* mengubah kata ke bentuk dasarnya (Bonta et al., 2019). Lemmatisasi mirip dengan *stemming*. *Stemming* menghapus beberapa karakter yang dapat menyebabkan istilah yang salah. Sebaliknya, lemmatisasi mempertimbangkan konteks kalimat dan mengubah kata menjadi bentuk dasar yang tepat. Penelitian ini melakukan *stemming* dan lemmatisasi.
5. Penghapusan *Stopwords*: *Stopwords* adalah bagian penting dari kalimat untuk meningkatkan keterbacaan dan kebermaknaan. Namun untuk mengurangi kompleksitas ruang fitur *stopword* perlu dihapus

Setelah data dibersihkan selanjutnya dilakukan pelabelan berbasis leksikon VADER yang merupakan paket yang disediakan oleh python. Teknik pelabelan ini dipilih karena memiliki kinerja sangat baik di domain media sosial (Preeti et al., 2020). VADER adalah aturan dan alat analisis sentimen berbasis leksikon, yang secara khusus diselaraskan dengan sentimen yang diungkapkan di media sosial, dikembangkan oleh Hutto dan Gilbert. VADER mengambil kalimat sebagai input dan memberikan nilai persen untuk tiga kategori positif, netral dan negatif. (Gata & Bayhaqy, 2020). VADER membuat skor dan label lalu menambahkan kolom ke dataset untuk menyimpan skor sentimen. Jika Nilai pada kolom skor sentiment lebih dari 0 artinya positif, nilai

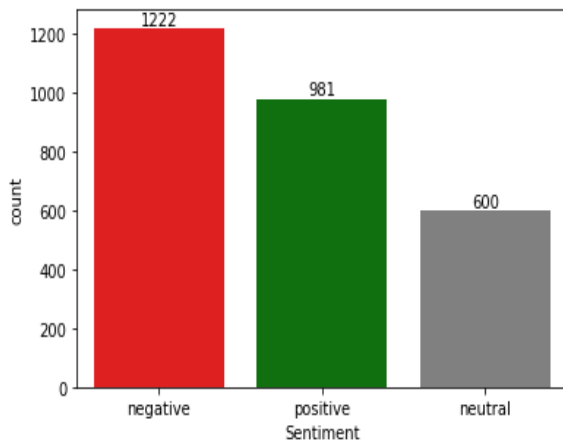
kurang dari nol artinya negatif, dan nilai nol berarti netral, hasilnya terlihat pada Tabel 2. Perbandingan jumlah komentar positif (1), negatif (-1), dan netral (0) ditunjukkan pada Tabel 2.

Tabel 2. Sampel Dataset Setelah Pelabelan dengan VADER

Komentar	Skor Sentimen	Sentimen
<i>foreign countries</i>	0	0
<i>come help Ukraine</i>	0.3612	1
<i>fault sticking their nose in other people business</i>	-0.9652	-1
<i>coward evil terroriste nation</i>	-0.8126	-1
<i>anyone ever thought that government deliverate</i>	-0.34	-1
<i>I fully agree and understand frustration anger</i>	-0.8211	-1
<i>situation all together</i>	0	0

Sumber: (Leidiyana, 2023)

Pelabelan dengan VADER membagi 2803 data komentar ke dalam label positif sebanyak 981 komentar, label negatif 1222 komentar, dan label netral 600 komentar. Data berlabel inilah yang nanti akan digunakan dalam pembuatan model klasifikasi.



Sumber: (Leidiyana, 2023)

Gambar 2. Grafik Jumlah Sentimen Positif, Negatif, dan Netral

Berikutnya adalah mengubah penyajian data supaya bisa komputasikan dan dipelajari oleh komputer. Tugas klasifikasi menggunakan data dalam bentuk vektor yang setiap elemennya menyajikan fitur dari sebuah data komentar. Dalam penelitian ini menggunakan python yang mana memiliki library bernama Scikit Learn yang dapat digunakan dalam Machine Learning. Pada

library ini terdapat algoritma CountVectorizer yang bekerja dengan cara mengubah teks yang diberikan menjadi vektor berdasarkan frekuensi (jumlah) dari setiap kata yang muncul di seluruh teks.

Komentar yang diekstrak mengandung beberapa kata berbeda yang umum untuk sebagian besar komentar, misalnya, Taliban, Afghanistan, Biden. Gambar 5 *wordcloud* untuk seluruh dataset menunjukkan istilah yang paling umum ditemukan digunakan dalam komentar yang diekstraksi terkait dengan reaksi veteran militer AS yang bertugas di Afghanistan.

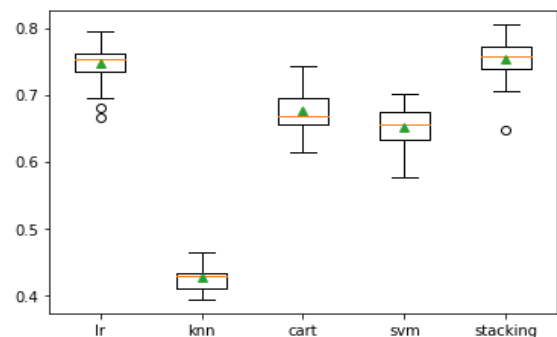


Sumber: (Leidiyana, 2023)

Gambar 3. Wordcloud

Dataset untuk meta-model disiapkan menggunakan validasi silang. Setiap model akan dievaluasi menggunakan validasi silang *k-fold* berulang. Contoh (*instance*) dari model diambil dan mengembalikan daftar skor dari tiga pengulangan validasi silang 10 kali lipat bertingkat (*stratified 10-fold cross-validation*). Gambar 4 menunjukkan kinerja rata-rata dari setiap algoritma klasifikasi dan *boxplot* untuk membandingkan distribusi skor akurasi untuk setiap algoritma.

```
>lr 0.746 (0.029)
>knn 0.426 (0.017)
>cart 0.676 (0.030)
>svm 0.652 (0.029)
>stacking 0.753 (0.033)
```



Sumber: (Leidiyana, 2023)

Gambar 4. Boxplot Akurasi Pengklasifikasi

Terlihat bahwa dalam kasus ini, untuk model tunggal, Regresi logistik menghasilkan akurasi rata-rata paling tinggi yaitu 74,6 persen. Untuk model tertinggi kedua yaitu CART dengan akurasi 67,6 persen, SVM 65,2 persen, dan terendah yaitu KNN 42,6 persen. Selanjutnya adalah percobaan menggabungkan kelima model tersebut menjadi satu model *ensemble* dengan menggunakan *stacking*. Model regresi logistik digunakan untuk menggabungkan prediksi dari masing-masing dari empat model yang terpisah lalu menggunakan validasi silang 5 kali lipat. Seperti yang terlihat pada Gambar 4 bahwa penerapan *stacking* menghasilkan akurasi sebesar 75,3 persen. Ini berarti akurasi di atas model tunggal dengan akurasi tertinggi yaitu regresi logistik.

### KESIMPULAN

Dalam penelitian ini pengklasifikasi *ensemble* dibentuk oleh *base learner* yang berbeda yaitu algoritma regresi logistik, SVM, KNN, dan CART. Untuk mengklasifikasi tunggal, mulai dari yang terendah yaitu algoritma KNN dengan akurasi rata-rata 42,6 persen, disusul CART dengan akurasi rata-rata 67,6 persen, SVM dengan akurasi rata-rata 65,2 persen, dan akurasi rata-rata paling tinggi yaitu algoritma regresi logistik sebesar 74,6 persen. Keempat algoritma tersebut disusun dan diprediksi dengan regresi logistik, dan hasilnya menunjukkan bahwa pengklasifikasi *ensembl stacking* yang diterapkan menghasilkan akurasi lebih baik daripada pengklasifikasi yang berdiri sendiri yaitu sebesar 75,3 persen. Dengan demikian terbukti bahwa teknik klasifikasi *ensemble* banyak digunakan di banyak bidang untuk memecahkan masalah klasifikasi karena dapat meningkatkan kinerja model klasifikasi. Untuk penelitian lebih lanjut dapat menambahkan algoritma klasifikasi lainnya. Jumlah dataset yang digunakan dalam penelitian ini juga tidak terlalu banyak. Menambah dataset untuk pembuatan model klasifikasi bisa jadi mempengaruhi akurasi model yang dihasilkan.

### REFERENSI

- Andone, D. (2021). *US veterans are disappointed with how the war in Afghanistan is ending and fearful for their Afghan allies*.
- Ankit, & Saleena, N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. *Procedia Computer Science*, 132(Iccids), 937-946.
- Ansari, M. Z., Aziz, M. B., Siddiqui, M. O., Mehra, H., & Singh, K. P. (2020). Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science*, 167, 1821-1828. <https://doi.org/10.1016/j.procs.2020.03.201>
- BAŞARSLAN, M. S., & KAYAALP, F. (2022). Sentiment Analysis with Ensemble and Machine Learning Methods in Multi-domain and Dataset. *Turkish Journal of Engineering*, 7(2), 141-148. <https://doi.org/10.31127/tuje.1079698>
- BBC. (2021). *Afghanistan Veteran MP Say Taliban Take Over Has Caused Anger, Grief, And Rage*. [https://youtu.be/chhy1Tdne\\_Q](https://youtu.be/chhy1Tdne_Q)
- Bonta, V., Kumares, N., & Janardhan, N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology*, 8(S2), 1-6. <https://doi.org/10.51983/ajcst-2019.8.s2.2037>
- Gata, W., & Bayhaqy, A. (2020). Analysis sentiment about islamophobia when Christchurch attack on social media. *Telkomnika (Telecommunication Computing Electronics and Control)*, 18(4), 1819-1827. <https://doi.org/10.12928/TELKOMNIKA.V18I4.14179>
- Görmez, Y., Işık, Y. E., Temiz, M., & Aydın, Z. (2020). FBSEM: A Novel Feature-Based Stacked Ensemble Method for Sentiment Analysis. *International Journal of Information Technology and Computer Science*, 12(6), 11-22. <https://doi.org/10.5815/ijitcs.2020.06.02>
- Lee, E., Rustam, F., Ashraf, I., Washington, P. B., Narra, M., & Shafique, R. (2022). Inquest of Current Situation in Afghanistan Under Taliban Rule Using Sentiment Analysis and Volume Analysis. *IEEE Access*, 10, 10333-10348. <https://doi.org/10.1109/ACCESS.2022.3144659>
- Mahakul, A. J. M. J. D. J. R. P. P. (2020). Global Perception of the Belt and Road Initiative: A Natural Language Processing Approach. *TrT Forum*, June, 400. <https://doi.org/10.54116/jbdtv.v1i1.18>

- Matalon, Y., Magdaci, O., Almozlino, A., & Yamin, D. (2021). Using sentiment analysis to predict opinion inversion in Tweets of political communication. *Scientific Reports*, *11*(1), 1–9. <https://doi.org/10.1038/s41598-021-86510-w>
- Muhammad Fikri, A. R., Jondri, J., & Astuti, W. (2022). Sentiment Analysis Against IndiHome and First Media Internet Providers Using Ensemble Stacking Method. *Building of Informatics, Technology and Science (BITS)*, *4*(2), 924–931. <https://doi.org/10.47065/bits.v4i2.1969>
- Preeti, Bhardwaj, P., & Kaur, R. (2020). Sentiment analysis using different techniques. *International Journal of Advanced Science and Technology*, *29*(10 Special Issue), 2439–2443.
- Priyanka, H. S., & Ashok Kumar, R. (2020). Sentiment Analysis using Machine Learning Based Ensemble Model for Food Reviews. *International Journal of Innovative Research in Applied Sciences and Engineering*, *4*(3), 690–694. <https://doi.org/10.29027/ijirase.v4.i3.2020.690-694>
- Singh, R. (2021). *Youtube comments sentiment analysis*. May, 0–11.
- Sri, M. (2021). Practical Natural Language Processing with Python. In *Practical Natural Language Processing with Python*. <https://doi.org/10.1007/978-1-4842-6246-7>
- Taliban Duduki Istana Presiden: Perang Afghanistan Berakhir.* (n.d.). <https://www.cnnindonesia.com/internasional/20210816123846-113-680967/taliban-duduki-istana-presiden-perang-afghanistan-berakhir>