

ANALISA SENTIMEN JELANG PILPRES 2024 MENGUNAKAN NAÏVE BAYES CLASSIFIER BERDASARKAN OPINI PUBLIK DI TWITTER

Oleh:

Heri Prasetyo

Arif Senja Fitriani

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Agustus, 2023

Pendahuluan

Ramainya antusiasme masyarakat terhadap isu pilpres yang semakin berkejang, hingga media sosial (twitter) pun jadi media yang banyak digemari oleh masyarakat untuk bertukar argumen mendebatkan kandidat calon presiden yang layak.

Hal ini yang membuat peneliti tertarik untuk menganalisa sentimen yang timbul didalamnya.

Pertanyaan Penelitian (Rumusan Masalah)

1. Bagaimana mengelola data publik berupa text ?
2. Bagaimana performa naïve bayes classifier terhadap analisa sentiment ?

Metode

Metode Penelitian

Algoritma Naive Bayes Classifier adalah salah satu metode klasifikasi yang populer dalam pembelajaran mesin. Ini didasarkan pada Teorema Bayes dengan asumsi "naif" bahwa semua fitur dalam dataset adalah independen secara kondisional.

Hasil

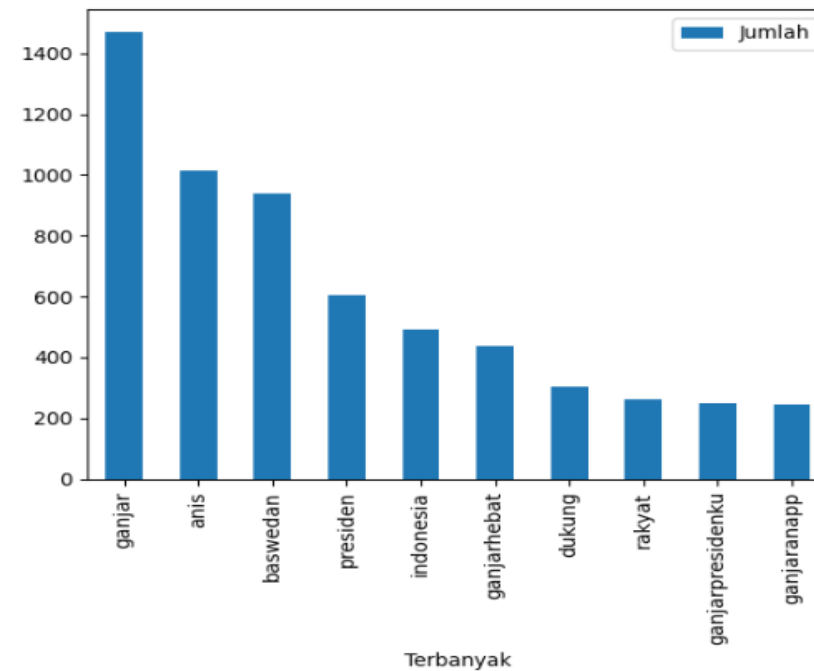
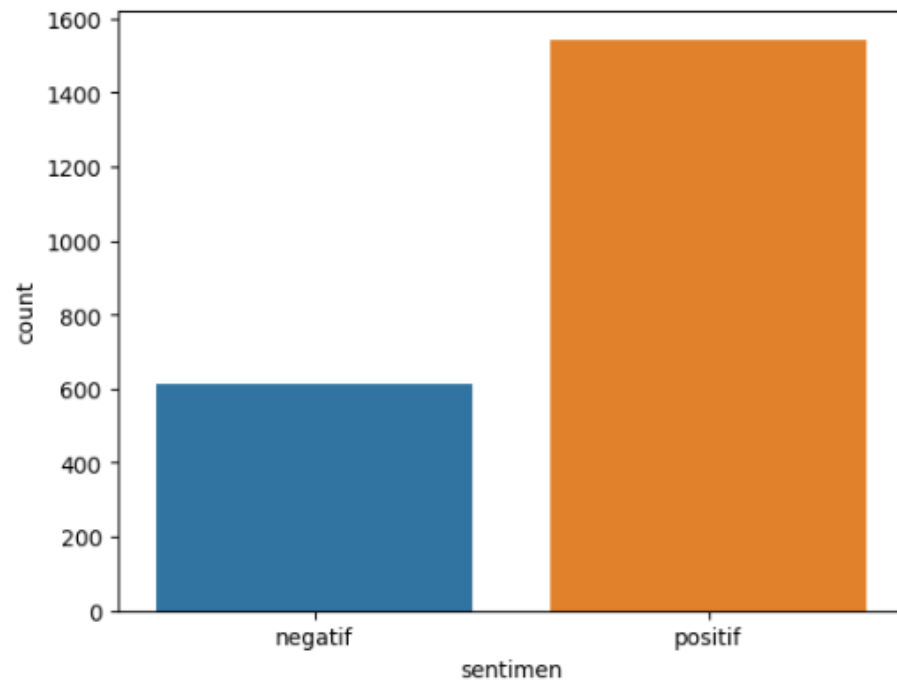


Berdasarkan proses tahapan pengolahan data yang dilakukan pada penelitian ini, proses penambangan data yang berhasil dilakukan berjumlah 3021 dengan parameter *query* “pilpres 2024”, “anis”, “ganjar” untuk mengetahui secara umum sentiment opini public apa saja yang sedang dibicarakan di ruang publik, “anis” seorang politikus dan intelektual Indonesia yang saat ini menjabat sebagai Gubernur DKI Jakarta sejak 2017, “ganjar” yang merupakan seorang politikus dan gubernur Jawa Tengah yang saat ini menjabat untuk periode kedua sejak 2018, masing – masing dari kata kunci tersebut adalah kandidat calon presiden pada pemilihan presiden 2024 yang sedang ramai dibicarakan di ruang public, media social dan media cetak, data di ekstrak menjadi format .csv yang kemudian dilakukan proses selanjutnya.

Pembahasan

Labeling

Hasil proses di dapatkan sentiment yang bersifat positif (1543) 71.5% dan negatif (614) 28.5%.



Pembahasan

Splitting Data

Data dibagi menjadi data uji dan data latih dengan 3 kali skema/pengujian untuk didapatkan hasil terbaik.

- Pada percobaan pertama menggunakan komposisi pembagian data : 90% data latih dan 10% data uji, menghasilkan 71% keakurasian menggunakan metode NBC.
- Pada percobaan selanjutnya komposisi data yang digunakan adalah 80% data latih dan 20% data uji. Keakurasian pada percobaan ini didapatkan nilai 68%.
- Pada percobaan terakhir menggunakan pembagian data 70% data latih dan 30% data uji sehingga mendapatkan 69% akurasi data.

Temuan Penting Penelitian

precision recall f1-score support

negatif 0.45 0.85 0.59 53
positif 0.93 0.66 0.77 163

Percobaan ke-1

accuracy 0.71 216
macro avg 0.69 0.76 0.68 216
weighted avg 0.81 0.71 0.73 216

precision recall f1-score support

negatif 0.46 0.82 0.59 179
positif 0.90 0.63 0.74 469

Percobaan ke-2

accuracy 0.68 648
macro avg 0.68 0.72 0.67 648
weighted avg 0.78 0.68 0.70 648

negatif 0.48 0.82 0.60 246
positif 0.90 0.64 0.75 617

Percobaan ke-3

accuracy 0.69 863
macro avg 0.69 0.73 0.68 863
weighted avg 0.78 0.69 0.71 863

Hasil Penguujian

Temuan Penting Penelitian

RINGKASAN HASIL PERCOBAAN

Percobaan	Kelas	Akurasi	Precision	Recall	f-measure
Ke – 1	Negatif	0.71	0.45	0.85	0.59
	Positif		0.93	0.66	0.77
Ke – 2	Negatif	0.68	0.46	0.82	0.59
	Positif		0.90	0.63	0.74
Ke – 3	Negatif	0.69	0.48	0.82	0.60
	Positif		0.90	0.64	0.75

Manfaat Penelitian

Berdasarkan percobaan yang telah dilakukan sehingga mendapatkan hasil yang ingin dicari, hasil percobaan pengujian metode Naïve Bayes Classifier terhadap data hasil *scrapping* opini publik pada sosial media twitter dengan topik pilpres 2024 dan parameter pencarian kata “pilpres 2024”, “anis”, “ganjar” dan kata pendekatan atau yang berhubungan lainnya. Data hasil scrapping 3021 dilakukan *pre-processing* sehingga menjadi 2157 kata, kemudian diberikan label sentimen yang bermuatan positif berjumlah 1543 (71,5%) dan negatif berjumlah 614 (28,5%). Percobaan yang telah dilakukan berturut – turut sebanyak 3 kali percobaan dengan hasil terbaik diperoleh pada percobaan ke – 1 sehingga di dapatkan *accuracy scored* 71% dengan nilai *precission* 93% class positif 45% negatif, *recall* 66% class negatif 85% positif dan *f - measure scored* 59% pada class negative 77% pada class positif. Sehingga dapat disimpulkan bahwa percobaan terbaik yang dilakukan adalah pada percobaan ke – 1 dengan pembagian 10% data uji dan 90% data latih. Pada penelitian selanjutnya diharapkan menggunakan algoritma lain dengan data yang lebih banyak sehingga dapat mendapatkan hasil yang berbeda.

Referensi

- [1] M. A. Firmansyah, D. Mulyana, S. Karlinah, and S. Sumartias, "Kontestasi Pesan Politik dalam Kampanye Pilpres 2014 di Twitter: Dari Kultwit Hingga Twitwar," *JIK*, vol. 16, no. 1, p. 42, Jan. 2018, doi: 10.31315/jik.v16i1.2681.
- [2] A. Septiana, "Analisis Fungsi Partai Politik Pada Pilkada Musi Banyuasin 2017 (Studi Terhadap Partai Politik Pengusung Pasangan Dodi Reza Dan Beni Hernedi)," *jssp*, vol. 3, no. 1, pp. 28–41, Jun. 2019, doi: 10.19109/jssp.v3i1.4066.
- [3] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, "Who says what to whom on twitter," in *Proceedings of the 20th international conference on World wide web*, Hyderabad India: ACM, Mar. 2011, pp. 705–714. doi: 10.1145/1963405.1963504.
- [4] P. Patmawati and M. Yusuf, "Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara," *bits*, vol. 3, no. 3, pp. 122–129, Dec. 2021, doi: 10.47065/bits.v3i3.1012.
- [5] I. W. D. Gafatia and N. Hadinata, "Analisis Pro Kontra Vaksin Covid 19 Menggunakan Sentiment Analysis Sumber Media Sosial Twitter," *JPSII*, vol. 2, no. 1, pp. 34–42, Nov. 2021, doi: 10.47747/jpsii.v2i1.544.
- [6] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," *Procedia Computer Science*, vol. 87, pp. 44–49, 2016, doi: 10.1016/j.procs.2016.05.124.
- [7] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Jurnal Eksplora Informatika*, vol. 9, no. 1, pp. 1–10, Sep. 2019, doi: 10.30864/eksplora.v9i1.237.
- [8] E. I. Saptanti, "Analisis Manajemen Impresi Ma'ruf Amin dalam Debat Pilpres 2019," *ULTIMA Comm*, vol. 12, no. 2, pp. 262–284, Dec. 2020, doi: 10.31937/ultimacomm.v12i2.1573.
- [9] Y. Sahria, "Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia," 2020.
- [10] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *RESTI*, vol. 5, no. 1, pp. 123–131, Feb. 2021, doi: 10.29207/resti.v5i1.2587.
- [11] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, Aug. 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [12] D. T. Anggraeni, "FORECASTING HARGA SAHAM MENGGUNAKAN METODE SIMPLE MOVING AVERAGE DAN WEB SCRAPPING," *jurnalmatrik*, vol. 21, no. 3, pp. 234–241, Dec. 2019, doi: 10.33557/jurnalmatrik.v21i3.726.
- [13] D. Gunawan, "Metode Klasifikasi pada Data Preprocessing Data," no. 1, 2016.
- [14] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *EduTic*, vol. 7, no. 1, Nov. 2020, doi: 10.21107/edutic.v7i1.8779.
- [15] S. Raschka, "Naive Bayes and Text Classification I - Introduction and Theory," *arXiv preprint arXiv:1701.01923*, 2017. Accessed May 25, 2023. [Online]. Available: <https://arxiv.org/abs/1701.01923>

