

DAFTAR PUSTAKA

- Adinugroho, R. (2022). Perbandingan Rasio *Split Data Training* dan *Data Testing* Menggunakan Metode LSTM dalam Memprediksi Harga Indeks Saham Asia. *Skripsi*. Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta.
- Aditya, B. R., No, J. T., Buah, T., & Bandung, B. (2015). Penggunaan Web Crawler Untuk Menghimpun *Tweets* dengan Metode Pre-Processing Text Mining. *Jurnal Infotel*, 7(2), 93-100.
- Ambarwari, A., Adrian, Q. J., & Herdiyeni, Y. (2017). Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 1(3), 117–122.
- Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(2), 183-190.
- Baita, A., Pristyanto, Y., & Cahyono, N. (2021). Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). *Information System Journal*, 4(2), 42-46.
- Barro, R. A., Sulvianti, I. D., & Afendi, F. M. (2013). Penerapan Synthetic Minority Oversampling Technique (Smote) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu. *Xplore: Journal of Statistics*, 1(1), 1-6.
- Briliani, A., Irawan, B., & Setianingsih, C. (2019). Hate Speech Detection in Indonesian Language on Instagram Comment Section using K-Nearest Neighbor Classification Method. *IEEE international conference on internet of things and intelligence system (IoTais)*. 98-104.
- Buda, M., Maki, A., & Mazurowski, M. A. (2018). A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks. *Elsevier*, 106, 249–259.
- Chawla, N. v, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

- Choirunnisa, S. (2019). Metode Hibrida Oversampling dan Undersampling Untuk Menangani Ketidakseimbangan Data Kegagalan Akademik Universitas XYZ. *Tesis*. Fakultas Teknologi Informasi dan Komunikasi Institut Teknologi Sepuluh Nopember, Surabaya.
- Damarta, R., Hidayat, A., & Abdullah, A. S. (2021). The Application of K-Nearest Neighbors Classifier for Sentiment Analysis of PT PLN (Persero) Twitter Account Service Quality. *Journal of Physics: Conference Series*, 1722(1), 1-7.
- Dang, N. C., Moreno-García, M. N., & de la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics (Switzerland)*, 9(3), 1-29.
- Dang, Y., Jiang, N., Hu, H., Ji, Z., & Zhang, W. (2018). Image classification based on quantum K-Nearest-Neighbor algorithm. *Quantum Information Processing*, 17(9), 1-18.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131-145.
- DataIndonesia.id. (2023, Januari 1). Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022. <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>
- Dikiyanti, T. D., Rukmi, A. M., & Irawan, M. I. (2021). Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm. *Journal of Physics: Conference Series*, 1821(1), 1-11.
- Douzas, G., Bacao, F., & Last, F. (2018). Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Information Sciences*, 465, 1–20.
- Fajri Nur Ika, & Astuti Dwi Femi. (2022). Pengaruh SMOTE DAN Forward Selection Dalam Menangani Ketidakseimbangan Kelas pada Algoritma Klasifikasi. *Jurnal Informasi Interaktif*, 7(1), 45–49.
- Fairuz, A.L., Ramadhani, R.D., & Tanjung, N.A. (2021). Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter. *Journal of*

- Dinda : Data Science, Information Technology, and Data Analytics*, 1(1), 1-10.
- Habibi, A. M., & Santika, R. R. (2020). Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Jurusan Menggunakan Metode Euclidean Distance Berbasis Web Pada SMP Setia Gama. *SKANIKA*, 3(4), 7–14.
- Hairani, H., Anggrawan, A., & Priyanto, D. (2023). Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link. *JOIV: International Journal on Informatics Visualization*, 7(1), 258-264.
- Ihsan, Arinal. M. (2018). Reduksi Atribut pada Algoritma K-Nearest Neighbor (KNN) dengan Menggunakan Algoritma Genetika. *Thesis*. Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara Medan.
- Indrawati, A. (2021). Penerapan Teknik Kombinasi Oversampling Dan Undersampling Untuk Mengatasi Permasalahan Imbalanced Dataset. *JIKO (Jurnal Informatika dan Komputer)*, 4(1), 38-43.
- Indrayanti, Sugianti, D., & Karomi, A. A. M. (2017). Optimasi Parameter K Pada Algoritma K-Nearest Neighbour untuk Klasifikasi Penyakit Diabetes Mellitus. *Prosiding SNATIF*, 823-829.
- Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(2), 121-130.
- Jananto, A., Sulastri, S., Wahyudi, E. N., & Sunardi, S. (2021). Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 10(1), 71-78.
- Jaya, U. B., Nasri, E., & Aw, A. S. (2020). Aplikasi Seleksi Penentuan Nasabah untuk Penjualan Barang Secara Kredit dengan Algoritma K-Nearest Neighbor. *Jurnal Sains dan Teknologi*. 4(1), 1-11.
- Joang, I., Kusriani, & Emha, T. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, 6(1), 28–36.

- Juba, B., & Le, H. S. (2019). Precision-Recall versus Accuracy and the Role of Large Data Sets. *Proceedings of the AAAI Conference on Artificial Intelligence*, 4039–4048.
- Kementerian Kesehatan Direktorat Promosi Kesehatan dan Pemberdayaan Masyarakat. (2023, Januari 1). Sebelum Melepas Masker di Keramaian, Kamu Wajib Ketahui 3 Hal Penting Ini! <https://ayosehat.kemkes.go.id/sebelum-melepas-masker-di-keramaian-kamu-wajib-ketahui-3-hal-penting-ini>
- Khaulasari, H. (2016). Combine Sampling-Least Square Support Vector Machine untuk Klasifikasi Multi Class Imbalanced Data. *Thesis*. Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.
- Krisdiyanto, T., Maricha, E., & Nurharyanto, O. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT*, 7(1), 27-32.
- Kurniawan, T. (2017). Implementasi Text Mining pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier dan Support Vector Machine. *Disertasi*. Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.
- Liu, C., Wu, J., Mirador, L., Song, Y., & Hou, W. (2018). Classifying DNA methylation imbalance data in cancer risk prediction using SMOTE and Tomek link methods. *International Conference of Pioneering Computer Scientists, Engineers and Educators*, 902, 1–9.
- Mulyani, S. (2022). Perbandingan Kinerja Metode Mesin Vektor Pendukung Tanpa dan Dengan Fitur Berbasis Leksikon Pada Analisis Sentimen Vaksinasi Universitas Hasanuddin 19 Covid-19 di Indonesia. *Skripsi*. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar.
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *J. Inform*, 6(2), 226-235.
- Nikmatul Kasanah, A., Muladi, & Pujiyanto, U. (2017). Penerapan Teknik SMOTE

- untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI*, 1(3), 196–201.
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi data mining untuk klasifikasi masa studi mahasiswa menggunakan algoritma K-Nearest Neighbor. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 10(2), 421-432.
- Nugraha, W., Risdiansyah, D., Purwaningtiyas, D., Hidayatulloh, T., & Suhada, S. (2022). Kombinasi Tomek-Link Dan Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Credit Card Fraud. *Jurnal Larik: Ladang Artikel Ilmu Komputer*, 2(2), 32-40.
- Nugroho A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram. *Jurnal Sains Komputer & Informatika*, 2(2), 201-209.
- Nurjanah, W. E., Setya Perdana, R., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah *Retweet*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1750-1757.
- Permana, A. P., Ainiyah, K., & Holle, K. F. H. (2021). Analisis perbandingan algoritma decision tree, kNN, dan Naive Bayes untuk prediksi kesuksesan start-up. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(3), 178-188.
- Prasetio, R. T. (2020). Seleksi Fitur dan Optimasi Parameter k-NN Berbasis Algoritma Genetika pada Dataset Medis. *Jurnal Responsif*, 2(2), 213–221.
- Putra, A. T., Kardinata, E., Junaedi, H., Chandra, F., & Santoso, J. (2021). Ekstraksi Relasi Antar Entitas di Bahasa Indonesia Menggunakan Neural Network. *Journal of Information System, Graphics, Hospitality and Technology*, 3(02), 49-54.
- Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor dalam Pengklasifikasian Follower Twitter yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), 31-37.
- Rofiqoh, U., Perdana, R. S., & 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. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1725-1732.
- Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., & Mononen, J. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behavioural Processes*, 148, 56–62.
- Sabilla, I. W., & Vista, C. B. (2021). Implementasi SMOTE dan Under Sampling pada Imbalanced Dataset untuk Prediksi Kebangkrutan Perusahaan. *Jurnal Komputer Terapan*, 7(2), 329–339.
- Saifudin, A., & Wahono, R. S. (2015). Pendekatan Level Data untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *IlmuKomputer.com Journal of Software Engineering*, 1(2), 76-85.
- Sain, H., & Purnami, S. W. (2015). Combine Sampling Support Vector Machine for Imbalanced Data Classification. *Procedia Computer Science*, 72, 59-66.
- Satgas Covid-19. (2023, Januari 1). Pemerintah Longgarkan Kebijakan Pemakaian Masker. <https://covid19.go.id/artikel/2022/05/17/pemerintah-longgarkan-kebijakan-pemakaian-masker>
- Shintia, R. (2018). Penerapan Combine Undersampling pada Klasifikasi Data Imbalanced Biner (STUDI KASUS : Desa Tertinggal di Jawa Timur Tahun 2014). *Skripsi*. Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.
- Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor. *Jurnal ISD*. 3(1), 44-49.
- Spelmen Vimalraj S, & Porkodi R. (2018). A Review on Handling Imbalanced Data. *International Conference on Current Trends toward Converging Technologies*, 1–11.
- Swana, E. F., Doorsamy, W., & Bokoro, P. (2022). Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors*, 22(9), 1-21.
- Tresnawati, Y. (2017). Analisis Sentimen pada Twitter Menggunakan Pendekatan Agglomerative Hierarchical Clustering. *Skripsi*. Fakultas Sains dan Teknologi

Universitas Sanata Dharma Yogyakarta.

- Utami, H. (2022). Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network. *Indonesian Journal of Applied Statistics*, 5(1), 31-38.
- Widhiyasana, Y., Semiawan, T., Mudzakir, I. G. A., & Noor, M. R. (2021). Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi/ Vol, 10(4)*, 354-361.
- Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems*. 9(2), 1-29.

LAMPIRAN

Lampiran 1 Struktur data sebelum dan setelah praproses data teks

No	Teks Sebelum Praproses Data Teks	Teks Setelah Praproses Data Teks
1	@txtdrjkt udah saatnya kita lepas masker	['udah', 'saat', 'kita', 'lepas', 'masker']
2	@Eka4Liong Sampai skrg belum berani ke mall disaat weekend . Belum berani lepas masker. Belum berani ke tempat keramaian .	['sampai', 'skrg', 'belum', 'berani', 'ke', 'mall', 'saat', 'weekend', 'belum', 'berani', 'lepas', 'masker', 'belum', 'berani', 'ke', 'tempat', 'ramai']
3	Dimanapun, kapanpun kamu berada, selalu gunakan masker! Jangan lepas maskermu. Dan selalu patuhi #ProkesPutusRantaiKovid di setiap aktifitasmu https://t.co/UrZicjLqLw	['mana', 'kapan', 'kamu', 'ada', 'selalu', 'guna', 'masker', 'jangan', 'lepas', 'masker', 'dan', 'selalu', 'patuh', 'di', 'tiap', 'aktifitasmu']
4	@elleanor12 @DokterTifa Saya beberapa kali ke kantor client, swasta, BUMN, kementerian, semua pada santai, gak taat prokes di kantor, lepas masker, ngopi sambil ngobrol, ya gimana pada gak sharing virus.	['saya', 'beberapa', 'ke', 'kantor', 'client', 'swasta', 'bumn', 'tri', 'semua', 'pada', 'santai', 'gak', 'taat', 'prokes', 'di', 'kantor', 'lepas', 'masker', 'ngopi', 'sambil', 'ngobrol', 'ya', 'gimana', 'pada', 'gak', 'sharing', 'virus']
5	@Tristan_ST18 @helloworldail @Pecel_lover @syaif38 @priaseratuskg @raydanadya @whoa_luthfi @Askrlfess Saya sarankan anda lepas masker ajaðŸ~ðŸ~-	['saya', 'saran', 'anda', 'lepas', 'masker', 'aja']
:	:	:
1265	Wah aktivitas luar ruangan sekarang boleh lepas masker nih.	['wah', 'aktivitas', 'luar', 'ruang', 'sekarang', 'boleh', 'lepas', 'masker', 'nih']
1266	belum siap sm kebijakan lepas masker, skrg kalo lepas masker kayak aib bgt dah ðŸ~” kalo pake masker w bisa ngatain orang yg gk w suka tanpa keliatan bibirnya gt wakakak, trs jg kalo lg males make up jg ga ketahuan sih soalnya ketutup masker ðŸ~”	['bom', 'siap', 'sm', 'bijak', 'lepas', 'masker', 'skrg', 'lepas', 'masker', 'kayak', 'aib', 'dah', 'pake', 'masker', 'bisa', 'ngatain', 'orang', 'yg', 'gk', 'suka', 'tanpa', 'liat', 'bibir', 'gt', 'wakakak', 'jg', 'lg', 'males', 'make', 'up', 'jg', 'ga', 'tahu', 'sih', 'soal', 'tutup', 'masker']
1267	@pandji yuk sdh saatnya lepas masker, muka ganteng gue ketutup mulu... bodo amat apa yg diomongin mereka berdua	['yuk', 'sdh', 'saat', 'lepas', 'masker', 'muka', 'ganteng', 'gue', 'tutup', 'mulu', 'bodo', 'amat', 'apa', 'yg', 'diomongin', 'mereka', 'dua']

Lampiran 2 *Count vectorizer* dataset

Dokumen	Abis	Aib	Aja	Udah	Aktivitas	Akut	Alas	Alhamdulillah	Alias	Allah	...	Yuk
1	0	0	0	1	0	0	0	0	0	0	...	0
2	0	0	0	0	0	0	0	0	0	0	...	0
3	0	0	0	0	0	0	0	0	0	0	...	0
4	0	0	0	0	0	0	0	0	0	0	...	0
5	0	0	1	0	0	0	0	0	0	0	...	0
6	0	0	0	2	0	0	0	0	0	0	...	0
7	0	0	0	0	0	0	0	0	0	0	...	0
8	0	0	2	0	0	0	0	0	0	0	...	0
9	0	0	0	1	0	0	0	0	0	0	...	0
10	0	0	0	0	0	0	0	0	0	0	...	0
⋮
1263	0	0	0	0	0	0	0	0	0	0	...	0
1264	0	0	0	0	0	0	0	0	0	0	...	0
1265	0	0	0	0	1	0	0	0	0	0	...	0
1266	0	1	0	0	0	0	0	0	0	0	...	0
1267	0	0	0	0	0	0	0	0	0	0	...	1

Lampiran 3 Pembobotan dengan TF-IDF

Dokumen	Abis	Aib	Aja	Udah	Aktivitas	Akut	Alas	Alhamdulillah	Alias	Allah	...	Yuk
1	0	0	0	2,44	0	0	0	0	0	0	...	0
2	0	0	0	0	0	0	0	0	0	0	...	0
3	0	0	0	0	0	0	0	0	0	0	...	0
4	0	0	3,14	0	0	0	0	0	0	0	...	0
5	0	1,06	0	0	0	0	0	0	0	0	...	0
6	0	0	0	4,87	0	0	0	0	0	0	...	0
7	0	0	0	0	0	0	0	0	0	0	...	0
8	0	0	6,27	0	0	0	0	0	0	0	...	0
9	0	0	0	2,43	0	0	0	0	0	0	...	0
10	0	0	0	0	0	0	0	0	0	0	...	0
⋮
1263	0	0	0	0	0	0	0	0	0	0	...	0
1264	0	0	0	0	0	0	0	0	0	0	...	0
1265	0	0	0	0	6,07	0	0	0	0	0	...	0
1266	0	5,25	0	0	0	0	0	0	0	0	...	0
1267	0	0	0	0	0	0	0	0	0	0	...	5,95

Lampiran 4 Data latih setelah *combine sampling*

No	Abis	Aib	Aja	Udah	Aktivitas	Akut	Alas	Alhamdulillah	Alias	Allah	...	Yuk	Sentimen
1	0	0	0	0	0	0	0	0	0	0	...	0	Positif
2	0	0	0	0	0	0	0	0	0	0	...	0	Negatif
3	0	0	0	0	0	0	0	0	0	0	...	0	Negatif
4	0	0	3,13	0	0	5,44	0	0	0	0	...	0	Negatif
5	0	0	0	0	0	0	0	0	0	0	...	0	Negatif
6	0	0	0	0	0	0	0	0	0	0	...	0	Negatif
7	0	0	3,13	2.43	0	0	0	0	0	0	...	0	Negatif
8	0	0	0	0	0	0	0	0	0	0	...	0	Negatif
9	0	0	0	0	0	0	0	0	0	0	...	0	Positif
10	5,95	0	0	0	0	0	5,44	0	0	0	...	0	Positif
⋮
1222	0	0	0	0	0	0	0	0	0	0	...	0	Positif
1223	0	0	0	0	0	0	0	0	0	0	...	0	Positif
1224	0	0	0	0	0	0	0	0	0	0	...	0	Positif
1225	0	0	0	1.92	0	0	0	0	0	0	...	0	Positif
1226	0	0	0	0.39	0	0	0	0	0	0	...	0	Positif

Lampiran 5 Nilai akurasi untuk setiap parameter k

Nilai k	Akurasi
2	0,60
3	0,50
4	0,53
5	0,48
6	0,53
7	0,49
8	0,52
9	0,49
10	0,53