

BIBLIOGRAPHY

- Alasadi S. & Bhaya W. (2017) Review of Data Preprocessing Techniques in Data Mining: *Journal of Engineering and Applied Sciences* 12 (16): 4102-4107, 2017 ISSN: 1816-949X
- Asosiasi Penyelenggara Jasa Internet Indonesia. (2023). Hasil Survei Internet Indonesia: *Press Conference Survei Internet Indonesia 2023 Tahap 1*. <https://survei.apjii.or.id/survei/> Accessed 2nd of November 2023
- Astuti L., Sari Y., and Suprpto. (2023) Code-Mixed Sentiment Analysis using Transformer for Twitter Social Media Data, *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(10), 2023. <http://dx.doi.org/10.14569/IJACSA.2023.0141053>
- Barik A., Mahendra R., & Adriani M. (2019) Normalization of Indonesian-English Code-Mixed Twitter Data: *Proceedings of the 2019 EMNLP Workshop W-NUT: The 5th Workshop on Noisy User-generated Text*, pages 417–424 Hong Kong, Nov 4, 2019, Association for Computational Linguistics
- Bird, S., Klein, E., and Loper, E., (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. 1st ed. Sebastopol, CA: O'Reilly Media. ISBN: 978-0-596-51649-9.
- Brennan S., Glenberg A., Wider K. (2018). The Role of Language in Intelligence: Laying the Groundwork for the Language Construct: *A Journal of Psychological Science*
- Carley K., Malik M., Pfeffer M., Landwehr P. (2015), Twitter Usage in Indonesia: *CMU-ISR-15-109 Center for the Computational Analysis of Social and Organizational Systems CASOS technical report*
- Dhar M., Kumar V., & Shriyastaya M. (2018) Enabling Code-Mixed Translation Parallel Corpus Creation and MT Augmentation Approach: *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing*, pages 131–140 Santa Fe, New Mexico, USA, August 20, 2018.
- Gaye, B., Zhang, D., & Wulamu, A. (2021). Improvement of Support Vector Machine Algorithm in Big Data Background. *Hindawi*. Received 21 Feb 2021; Revised 30 Apr 2021; Accepted 03 Jun 2021; Published 17 Jun 2021.
- Harris, C.R., Millman, K.J., van der Walt, S.J. et al. (2020). Array programming with

- NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.
- Isik F., Ozden G., & Kuntalp M. (2012), Importance of data preprocessing for neural networks modeling: The case of estimating the compaction parameters of soils: *Energy Education Science and Technology Part A: Energy Science and Research 2012 Volume (issues) 29(2): 871-882*
- Javdan, S., ataei, T. S., & Minaei-Bidgoli, B. (2020). IUST at SemEval-2020 Task 9: *Sentiment Analysis for Code-Mixed Social Media Text using Deep Neural Networks and Linear Baselines. arXiv preprint. doi:https://doi.org/10.48550/arXiv.2007.12733*
- Mahata S., Mandal S., Das D., & Bandyopadhyay S. (2019) Code-Mixed to Monolingual Translation Framework Association for Computing Machinery. *ACM ISBN 978-1-4503-7750-8/19/12...\$15.00* <https://doi.org/10.1145/3368567.3368579>
- Mishra, P., Danda, P., & Dhakras, P. (2018). Code-Mixed Sentiment Analysis Using Machine Learning and Neural Network Approaches. *Language Technologies Research Center, IIIT Hyderabad, Kohli Center On Intelligent Systems*.
- Mejova Y. (2009), Sentiment Analysis an Overview: *Comprehensive Exam Paper at Computer Science Department at University of Iowa*
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825-2830.
- Pratapa A., Choudhury M., & Sitaram S. (2018), Word Embeddings for Code-Mixed Language Processing: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3067–3072 Brussels, Belgium, October 31 - November 4, 2018, Association for Computational Linguistics*
- Rizal A. & Stymne S. (2020), Evaluating Word Embeddings for Indonesian-English Code-Mixed Text Based on Synthetic Data: *Proceedings of the LREC 2020 – 4th Workshop on Computational Approaches to Code Switching, pages 26–35 Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020 European Language Resources Association (ELRA), licensed under CC-BY-NC*

- Sadia A., Khan F., & Bashir F. (2018) An Overview of Lexicon-Based Approach For Sentiment Analysis https://ieec.neduet.edu.pk/2018/Papers_2018/15.pdf
- Salsabila, N. A., Winatmoko, Y. A., Septiandri, A. A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. *International Conference on Asian Language Processing (IALP)*, 236-239.
- Shuyo, N., (2014). langdetect. Available at: <https://pypi.org/project/langdetect/#description> [Last Accessed: 3rd of April, 2024]
- Shokirjon X. & Robiya S. (2023) Globalization and Spread of English: *Выпуск журнала № – 22 Часть–6 Июнь –2023*.
- Singh, G. (2021). Sentiment Analysis of Code-Mixed Social Media Text (Hinglish). *School of Computing, University of Leeds*.
- Symeonidis S., Effrosynidis D., & Arampatzis A. (2018), A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis: *S. Symeonidis et al. / Expert Systems With Applications 110 (2018) 298–310*
- Tho C., Heryadi Y., Lukas L., & Wibowo A. (2021), Code-Mixed Sentiment Analysis of Indonesian Language and Javanese Language using Lexicon Based Approach, *J. Physics Conference Series 1869012084*
- Varma K. & Mamidi P. (2021) Sentiment Analysis in Code-Mixed Telugu-English Text with Unsupervised Data Normalization. *Proceedings of Recent Advances in Natural Language Processing, pages 753–760, Sep 1–3, 2021*.
https://doi.org/10.26615/978-954-452-072-4_086
- Wijaya, A. D., & Bram, B. (2021). A Sociolinguistic Analysis of Indoglish Phenomenon in South Jakarta. *Professional Journal of English Education*, 672-684.
- Yuang C., Bachs R., & Siong C., (2012), An Empirical Evaluation of Stop Word Removal, *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*
- Yulianti E., Kurnia A., Adriani M., & Duto Y. (2021), Normalisation of Indonesian-English Code-Mixed Text and its Effect on Emotion Classification. *(IJACSA) International Journal of Advance Computer Science and Applications, Vol. 12, No. 11, 2021*

ATTACHMENTS

Attachment 1: Keywords for Obtaining Tweets (Astuti *et al*, 2023):

Keywords	Example of Tweet
Kepo	A: hey kamu udah hamil? " Kok follow akun parenting? B : lah emang ga boleh follow akun kek gitu? A : oh, kirain. Aku kan kepo B : doain yah. Aamiin in aja 🙏
Santuy	Abis baca berita soal pasien covid dikucilkan dan dibully alhamdulillah orang terdekatku baik hati. Kemaren rumah camer dtg mantan pasien covid dan uda Negatif malah ditanya camer "samean wes waras tenan a pak? Ojo kesel2" santuy aja gitu camerku ❤️
Jujurly	Dear mas ² abang ² bpak ² pengendara motor, tlgah klo lgi nyetir motor jgn sambil ngerokok. Kasihanilah kami yg ada diblknng klean yg lebih beresiko kena bara dri rokok klean. Klean mah enak ngebul santuy lha kita celaka klo lgi apes itu bara kena mata 🙄🙄
Normally	Oh god it hurts so bad I can't breathe normally I can't even breathe it's hard God help me to find anything like air oil or can relieve the feeling My pain is God help me I need help I'm hard to breathe for too long it's hard for me God Daddy's home I need your help for 🏠👨👧👦🌱
Flexing	Seru banget dengerin orang lagi flexing kekayaan wkwkw
Self proclaimed	2 minggu guling2 stress sendiri. ternyata obatnya cuma geser meja kerja di kamar sm belanja bulanan di mall. (terakhir -/+6bln lalu, biasanya go shop dsb) langsung mood membaik drastis. hhh pandemi pandemi. cepet kelar dong. gw self proclaimed introvert aja udh berasa mo gila ni.
Which is	29 Juli 2020 - which is 1th lalu. gembar gembor Bioskop akan buka serentak.in fact.... zonk29 Juli 2021 - which is sekarang. semua bioskop kembali tutup lagi, 🎬
Literally	Akhirnya punya brimo, literally pak bos forced me 🍌...

Actually	<p>baru sempet baca hari ini</p> <p>this chapter is actually really good, we now have Osaragi's background story and all the things leading up to her actions.</p> <p>and of course, her somewhat twisted idea of love. i believe "twisted" is a good word to use to describe it. https://t.co/GMMjc1qKDd</p>
Prefer	<p>daripada nonton start up yang begitu boomingnya kalo aku sih prefer nonton search yg tayangnya brengan start up lebih menarik</p>
Probably	<p>Subuh tahun lalu. Mengejar epic sunrise di Danau Kelimutu, Kabupaten Ende, Nusa Tenggara Timur (NTT). Cuaca cerah di Tanah Flores, probably salah satu momen sunrise ter- epic yg pernah saya saksikan. https://t.co/XdIp96RzEd</p>
Around	<p>Suni Lee yang gantiin Simone Biles di nomer beregu 2 hari lalu sekarang lagi mimpin di nomer individu all around. https://t.co/0NAa8rNLLV</p>
Honestly	<p>udah hampir seharian Miku gamau makan sampe dibeliin royelkenin gamau juga 🍷 taunnya ? di pancing wf whiskas baru makannya kesetanan..</p> <p>akhirnya tu dicampur jg dfnya. yaAllah mbabu udah pusing setegah mampus:)</p> <p>honestly,Miku blm blh mam wf dr dokternya krn tkt mencrit lg:”) https://t.co/uCnXN3AMfz</p>
Whatever	<p>Whatever 🍌🍌🍌 !! Biar tua2 gini sumbangsih utk negara gede..pajaknya mahal 🍌🍌🍌🍌🍌🍌🍌</p> <p>https://t.co/iwoezLWMtT https://t.co/B5Krh1ZISH</p>
Clingy	<p>Ya Allah, aku masih nunggu kiriman black/tuxedo kitten yg sehat lincah aktif bisa makan minum pup sendiri dan clingy ya.</p> <p>(Doa spesifik) 🙏</p>

Confuse	"Ijazah tanda orang pernah sekolah, bukan berpikir"-Cak Nun-"I hate when ppl confuse education with intelligent, you can have bachelor's degree and still be an idiot"-Elon Musk-
Shoutout	[A THREAD] Favourite Foods In BatamHalooo, cuma mau share dikiiit. Inspirasi ide ini came up while I'm having some chit-chats a.k.a ghibah with my roommate, shoutout to @tamarineka. These foods are my fav selama beratus-ratus hari aku tinggal di Batam 🍲👀There you go.
While	31 Des 2020 lepas behel (just for a while)
The point	AKU. duku gasuka kl punya pacar "berseragam" krn mostly sosoan. tapi kemakan omongan sendiri, pacaran sama IPDN sampe nikah punya anak 🍌*the point is jgn benci, nnti ditakdirin kebalikannya https://t.co/wNLaUBeoEU
Surely	He's surely soo excited! Siapa sih yg nggak? 🍌🍌 Baru pulang kerumah trus dsambut senyum manis ayank beb & kado indah segede gaban 🍌 ✂ BrightWin bener2 y, set the standard soo damn high in loving each other 🍌🍌🍌 God, please always give them happiness 🍌🍌🍌 #ใบพรวิน https://t.co/iDeIzDhKQj
Staycation	Jadi pengen staycation sendirian. Literally sendirian, pengen bisa self healing. Capek gaksi lo?
Healing	jam segini baru selesai deeptalk ama mamskuy, healing bgt ga sih bonus juga buat hari ini
Spill	Kalau di Twitter pada baik semua takut kena spill soalnya
Deep talk	Malem - malem gini udah jam 23:36 hujan pula enak nya deep talk nih hmm 🍌
Bestie	Malam thn baru pada kemana bestie?

Red flag	Oh iya ya, waktu itu dgn pedenya gue blg ke Pak Wayan (pemilik hostel yg gue tempatin di Sanur) kalo gue bakal balik lg di bulan November. Gue juga ngejanjiin temen gue yg baru bgt buka cafe bakal ngunjungin cafe dia di bulan November. Red flag bgt gue, tukang php 🍷
Salty	Pada salty bio di clubhouse kayak linkedin. Abisnya di platform itu gimana bisa ngasih atensi ke orang at the 1st time kalo nggatau backgroundnya apa?
Manifesting	🟩 Garasi udah ada 🟩 Jalanannya ada 🟩 Nyetir udah bisa 🟩 SIM adaPengen upgrade mobil ke versi yg lebih baru biar dapet fitur safety yg lebih mumpuni. Prefer beli cash biar ga bnyk utang dan krn mobil nilai jualnya turun. #ManifestingIn2022 https://t.co/a9aEjj9L0n
Relatable	Ketawa-Nangis-Ketawa-Nangis-Ketawa- Nangis-Nangis-Nangis-Nangis Maraton How I Met Your Mother udah kaya olahraga, saking banyaknya tenaga yang keluar buat... nangis. Too many relatable stories.
Fomo	Jalan ninja ku sebagai wanita beranjak tua yg mulai ga tau perkembangan jaman. -keep in touch sama anak2 muda, anak magang, dedek2 gemes yg sekarang fomo abezz biar kalo apa2 tinggal, " Dek, ini apaan sikk"-
Overthinking	Kebangun gara2 tiba2 overthinking nanti lebaran mau makan opor sm gule dimana. Ada yg open house g y
Thrifting	Masih belum nyerah di thrifting.. Doakan guys.. Semoga kali ini, lancar Jaya.. 🙏🙏
Trust issues	susah bgt percaya lg sama org yg udh bikin kecewa + trust issues gue gede bgt yaallah suudzon mulu sm org 🤔
Anxiety	Tadi sesengkukan trs telpon si Dinda. Covid + anxiety bikin perasaan hari ini down. Terus dikirimin buku sama Alquran 📖 https://t.co/Ui0ixuvX1j
Overwhelmed	Bisa nggak, sih, merasa 'overwhelmed' dari paparan krisis di sekitar?

Mood swing	Diri ini mudah mood swing skali apa aja yg sedih bisa teikut nangis, hbs itu ketawa
Meanwhile	Kok bisa bisa nya ya manager gue sendiri ngomongnya "kerjain aja besok" ????? Meanwhile gue yg overthinking karna kerjaan gue yg harusnya kelar hari ini malah disuruh kerjain besok??????
Sexist	sexist joke is not a joke. sungguh perilaku memalukan dan menggambarkan betapa dangkalnya isi otak.
Feminist	We all a feminist, disadari ataupun tidak.
Support system	31 Desember 2020, waw tidak menyangka bisa survive di 2020. Terima kasih kepada semua pihak yang udah jadi support system saat aku up and down. Semoga 2021 lebih baik
Gaslighting	Yang banyak dibilang orang "bagus dari segi manapun..." ya gatau kan kalo dibalik itu ternyata suka gaslighting?
Open minded	Yang kek gini Tuh yang harus Dijadiin Gerakan Para Open Minded bukan Konten konten Gak jelas sok sok EDGY https://t.co/rlnYqjHYU8
Narsis	Cowo yg teges, berani ngomong depan public siapapun itu, manajemen waktu nya bagus, ontime, ga narsis, wangii, rapi, i have crush on you
Crush	crush aku dingin banget aku yg selalu nyari topik sekarang lagi abiss voba dong kaish topik anti mainstream. org nya sat set gitu dia juga suka ngegame dia org nya juga singkat coba bantuin aku supaya di ngebales cukup panjang.btw orgnya slowrepond.Terimakasih
Even	Dari akun ini Not even to ugliest woman in USA you Will touch for real Weirdos like you all Will just underground forever with fake touch screen Marriage And an old grandmother for real Marriage together with some men in cult https://t.co/8rzvp1sgRa



UNIVERSITAS
GADJAH MADA

**COMPARATIVE STUDY OF DATA PREPROCESSING METHODS FOR CODE-MIXED SENTIMENT
ANALYSIS IN BAHASA
INDONESIA**

GHIFARI NUGRAHA PRADANA, Yunita Sari, S.Kom., M.Sc., Ph.D.

Universitas Gadjah Mada, 2024 | Diunduh dari <http://etd.repository.ugm.ac.id/>

Burnout	Keknya emang negara ini belum bisa bersaing karena kebanyakan libur hari raya. Dan kurang libur yg diambil untuk diri sendiri, jadinya orang-orang yg ada di kota industri terutama gampang bgt burnout lalu semua capaian hanya terpatok pada apa yg bisa ia raih ketika hari raya.
Denial	Masih denial besok dah 2022 berarti nambah umur tapi bersyukur juga masih dikasih umur sama allah :)
Sugar coating	Dih gw sih ogah. No sugar coating lah,yg bikin gede ga enakan tuh krn sugar coating. Klo ga enak ya blg aja ga enak kenapa mesti maksa di makan. Aneh
Somehow	Fungsi membeli dagangan tetangga adalah selain membantu beliau2, juga bisa "nonggo" in a good ways. Ukhuwah islaminya tetep terjaga, bisa ngobrol ngalor ngidul, dan somehow mreka duluan yg dtg klo ada kesusahan 🤔