

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dalam penelitian ini dapat ditarik kesimpulan, yaitu pembuatan korpus dan pengembangan model NER dari komentar *TikTok* tentang *skincare* berhasil dilakukan. Tahapan *pre-processing* yang melibatkan proses *screening* manual dengan pembuatan panduan *keyword* berdasarkan delapan data set berhasil membantu menyaring komentar yang relevan serta membantu mempercepat proses verifikasi dengan *expert*. Selain itu proses pembersihan data dengan NLP dan *multilabel classification* turut mempercepat identifikasi komentar yang layak untuk dianotasi. Proses anotasi dilakukan dengan teratur walaupun menggunakan NER manual *Prodigy* dengan memanfaatkan panduan *keyword* yang telah diverifikasi oleh *expert* dari PT L'Essential. Sehingga entitas yang digunakan sudah sesuai dengan istilah dan batasan dalam industri kecantikan. Verifikasi ini juga berfungsi sebagai bentuk validasi untuk menjaga agar data yang digunakan tidak menyesatkan pengguna dan tetap aman secara informasi.

Pengembangan model NER juga berhasil dilakukan dengan menunjukkan performa yang sangat baik dengan *precision* sebesar 94.44%, *recall* sebesar 92.27%, dan *F1-score* sebesar 93.34% pada data *testing* yang tidak digunakan dalam proses *training*. Angka ini menunjukkan bahwa model mampu mengenali entitas delapan data set secara akurat dan stabil. Kinerja model tertinggi ditemukan pada entitas yang sering disebut secara eksplisit oleh pengguna, seperti KANDUNGAN_SKINCARE, JENIS_PRODUK, dan WAKTU PERAWATAN. Sedangkan entitas yang bersifat kontekstual seperti TIPS dan IMPACT sedikit lebih susah dikenali. Walaupun demikian, keseluruhan proses membuktikan bahwa pendekatan ini mampu menghasilkan model yang andal dan dapat digunakan untuk mengolah komentar media sosial sebagai sumber informasi *skincare* yang valid.

6.2 Saran

Dalam pembangunan model NER untuk ekstraksi informasi dari komentar *TikTok* tentang *skincare* ini, peneliti menyadari bahwa hasil yang dicapai masih memiliki potensi untuk dikembangkan lebih lanjut, baik dari sisi kualitas data maupun pemanfaatannya dalam bentuk aplikasi nyata. Oleh karena itu, beberapa saran yang dapat dilakukan untuk pengembangan penelitian selanjutnya agar hasilnya lebih optimal dan aplikatif adalah:

1. Menambah jumlah data pelatihan dari berbagai jenis konten aplikasi *TikTok* maupun media sosial lainnya agar model mampu menangani variasi konteks dan gaya bahasa yang lebih beragam.
2. Mengembangkan model menjadi asisten rekomendasi *skincare* otomatis yang dapat menyarankan produk berdasarkan jenis kulit, permasalahan, dan kandungan yang sesuai.
3. Membuat *dashboard* analisis tren komentar *TikTok* yang menampilkan data visual tentang topik yang sering dibahas pengguna seperti kandungan populer atau masalah kulit terbanyak.
4. Mengembangkan panduan interaktif pemilihan *skincare* berbasis entitas yang dapat menjawab pertanyaan pengguna secara otomatis dan memberikan edukasi yang aman dan terpercaya

DAFTAR PUSTAKA

- Alexander, D., & de Vries, A. P. (2021). *"This research is funded by...": Named Entity Recognition of financial information in research papers.*
- Anetta, K., & Arslan, M. (2021). *Transferability of General Polish NER to Electronic Health Records.*
- Azzouzi, M. El, Coatrieux, G., Bellaqira, R., Delamarre, D., Riou, C., Oubenali, N., Cabon, S., Cuggia, M., & Bouzillé, G. (2024). Automatic de-identification of French electronic health records: a cost-effective approach exploiting distant supervision and deep learning models. *BMC Medical Informatics and Decision Making*, 24(1). <https://doi.org/10.1186/s12911-024-02422-5>
- Bird, S. (2006). *NLTK: The Natural Language Toolkit.*
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python.*
- Buck, J. W., Perugini, S., & Nguyen, T. V. (2018, January 5). Natural language, mixed-initiative personal assistant agents. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3164541.3164609>
- Challita, S., Zalila, F., Gourdin, C., & Merle, P. (2018). *A Precise Model for Google Cloud Platform.* <https://cloud.google.com>
- Chowdhury, G. G. (2003). Natural language processing. In *Annual Review of Information Science and Technology* (Vol. 37). <http://eprints.cdlr.strath.ac.uk/2611/>
- Dash, G., Sharma, C., & Sharma, S. (2023). Sustainable Marketing and the Role of Social Media: An Experimental Study Using Natural Language Processing (NLP). In *Sustainability (Switzerland)* (Vol. 15, Issue 6). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/su15065443>
- Franchino, F., Migliorati, M., Pagano, G., & Vignoli, V. (2024). Concepts and measures of bureaucratic constraints in European Union laws from hand-coding to machine-learning. *Regulation and Governance*, 18(3), 688–703. <https://doi.org/10.1111/regu.12543>
- Guo, Q., Wang, S., & Wan, F. (2020). Research on named entity recognition for information extraction. *Proceedings - 2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture, AIAM 2020*, 121–124. <https://doi.org/10.1109/AIAM50918.2020.00030>
- Hilmi, I. L., Rianoor, N. P., & Gatera, V. A. (2022). Hubungan Pengetahuan dan Sikap terhadap Perilaku Pemilihan Skincare Wajah melalui Media Sosial pada Salah Satu Universitas di Karawang Jawa Barat. In *Jurnal Farmasi Indonesia* (Vol. 19, Issue 2). <http://journals.ums.ac.id/index.php/pharmacon>

- Hodorog, A., Petri, I., & Rezgui, Y. (2022). Machine learning and Natural Language Processing of social media data for event detection in smart cities. *Sustainable Cities and Society*, 85. <https://doi.org/10.1016/j.scs.2022.104026>
- Honnibal, M., & Johnson, M. (2015). *An Improved Non-monotonic Transition System for Dependency Parsing*. Association for Computational Linguistics.
- Jiang, H., Hua, Y., Beeferman, D., & Roy, D. (2022). *Annotating the Tweebank Corpus on Named Entity Recognition and Building NLP Models for Social Media Analysis*. <http://arxiv.org/abs/2201.07281>
- Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing*.
- Kocaman, V., & Talby, D. (2022). Accurate Clinical and Biomedical Named Entity Recognition at Scale[Formula presented]. *Software Impacts*, 13. <https://doi.org/10.1016/j.simpa.2022.100373>
- Ladas, N., Borchert, F., Franz, S., Rehberg, A., Strauch, N., Sommer, K. K., Marschollek, M., & Gietzelt, M. (2023). Programming techniques for improving rule readability for rule-based information extraction natural language processing pipelines of unstructured and semi-structured medical texts. *Health Informatics Journal*, 29(1). <https://doi.org/10.1177/14604582231164696>
- Lally, A., & Fodor, P. (2011). *Natural Language Processing With Prolog in the IBM Watson System*.
- Le Guillarme, N., & Thuiller, W. (2022). TaxoNERD: Deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature. *Methods in Ecology and Evolution*, 13(3), 625–641. <https://doi.org/10.1111/2041-210X.13778>
- Manning Hinrich, C. D. (1999). *Foundations of Statistical Natural Language Processing*.
- Mikheev, A., Moens, M., Grover, C., Moens, M., & ac Uk, ed. (1999). *Named Entity Recognition without Gazetteers*.
- Mirza, A. F. S. (2024). *ANALISIS PENGARUH SIKAP BEAUTY INFLUENCER TASYA FARASYA TERHADAP MINAT BELI KONSUMEN MELALUI SIKAP KONSUMEN*.
- Nadeau, D., & Sekine, S. (2007). *A survey of named entity recognition and classification*. <http://projects.ldc.upenn.edu/gale/>
- Narayanan, A., Rao, A., Prasad, A., & Das, B. (2020, February 1). Character Level Neural Architectures for Boosting Named Entity Recognition in Code Mixed Tweets. *International Conference on Emerging Trends in Information*

- Technology and Engineering, Ic-ETITE 2020*. <https://doi.org/10.1109/ic-ETITE47903.2020.089>
- Patton, D. U., Frey, W. R., McGregor, K. A., Lee, F. T., McKeown, K., & Moss, E. (2020). Contextual analysis of social media: The promise and challenge of eliciting context in social media posts with natural language processing. *AIES 2020 - Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 337–342. <https://doi.org/10.1145/3375627.3375841>
- Pitropakis, N., Kokot, K., Gkatzia, D., Ludwiniak, R., Mylonas, A., & Kandias, M. (2020). Monitoring Users' Behavior: Anti-Immigration Speech Detection on Twitter. *Machine Learning and Knowledge Extraction*, 2(3). <https://doi.org/10.3390/make2030011>
- Pratiwi, N., Asrina, A., & Hasan, C. (2023). HUBUNGAN PENGETAHUAN DENGAN PEMILIHAN SKINCARE PADA REMAJA PUTRI DI SMPN 1 AWANGPONE. *Window of Public Health Journal*, 4.
- Purnomo, Y. S., Kumar, Y. J., Zulkarnain, N. Z., & Raza, B. (2024). *Extraction and attribution of public figures statements for journalism in Indonesia using deep learning*.
- PUTRI, C. A. H. (2020). *INFLUENCER PRODUK KECANTIKAN DALAM MEMPROMOSIKAN PRODUK DARI PERSPEKTIF ETIKA BISNIS ISLAM*.
- Rothman, Denis. (2021). *Transformers for Natural Language Processing* Rothman, Denis. Packt Publishing.
- Satheesh, K., Jahnavi, A., Iswarya, L., Ayesha, K., Bhanusekhar, G., & Hanisha, K. (2020). Resume Ranking based on Job Description using SpaCy NER model. *International Research Journal of Engineering and Technology*. www.irjet.net
- Ścieszko, E., Budny, E., Rotsztejn, H., & Erkiert-Polguj, A. (2021). How has the pandemic lockdown changed our daily facial skincare habits? *Journal of Cosmetic Dermatology*, 20(12), 3722–3726. <https://doi.org/10.1111/jocd.14059>
- Shelar, H., Kaur, G., Heda, N., & Agrawal, P. (2020). Named Entity Recognition Approaches and Their Comparison for Custom NER Model. *Science and Technology Libraries*, 39(3), 324–337. <https://doi.org/10.1080/0194262X.2020.1759479>
- Son, J., & Kim, B. (2023). Translation Performance from the User's Perspective of Large Language Models and Neural Machine Translation Systems. *Information (Switzerland)*, 14(10). <https://doi.org/10.3390/info14100574>

- Srinivasa-Desikan, Bhargav. (2018). *Natural language processing and computational linguistics : a practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing.
- Tabassum, A., & Patil, R. R. (2020). A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing. *International Research Journal of Engineering and Technology*. www.irjet.net
- Tohma, K., & Kutlu, Y. (2020). Challenges Encountered in Turkish Natural Language Processing Studies. In *Natural and Engineering Sciences* (Vol. 5, Issue 3, pp. 204–211). Natural and Engineering Science. <https://doi.org/10.28978/nesciences.833188>
- Wang, B., Lu, W., Wang, Y., & Jin, H. (2018). *A Neural Transition-based Model for Nested Mention Recognition*. Association for Computational Linguistics.
- Yilmaz, S. F., Balaban, I., Tekin, S. F., & Kozat, S. S. (2020). *Hybrid Framework for Named Entity Recognition in Turkish Social Media*. IEEE.
- Zitouni, I. (2014). *Natural Language Processing of Semitic Languages*. <http://www.springer.com/librarians/e-content/mycopy>

Lampiran Bukti Verifikasi Data

