

**IMPLEMENTASI *NATURAL LANGUAGE PROCESSING*
DAN ALGORITMA *ECLAT* UNTUK SISTEM REKOMENDASI PADA
SOAL *SENTENCE COMPLETION* DALAM *ENGLISH AS A SECOND
LANGUAGE***

SKRIPSI

Diajukan untuk memenuhi sebagian syarat
untuk memperoleh gelar Sarjana Komputer
Program Studi Ilmu Komputer



Oleh:

Muhammad Naufal Wicaksono
1904802

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
BANDUNG
2024**

**IMPLEMENTASI *NATURAL LANGUAGE PROCESSING*
DAN ALGORITMA *ECLAT* UNTUK SISTEM REKOMENDASI PADA
SOAL *SENTENCE COMPLETION* DALAM *ENGLISH AS A SECOND
LANGUAGE***

Oleh
Muhammad Naufal Wicaksono
NIM 1904802

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu
Pengetahuan Alam

© Muhammad Naufal Wicaksono 2024
Universitas Pendidikan Indonesia
Agustus 2024

Hak Cipta Dilindungi Undang-Undang
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difoto kopi atau cara lainnya tanpa izin dari penulis

MUHAMMAD NAUFAL WICAKSONO

1904802

**IMPLEMENTASI *NATURAL LANGUAGE PROCESSING*
DAN ALGORITMA *ECLAT* UNTUK SISTEM REKOMENDASI PADA
SOAL *SENTENCE COMPLETION* DALAM *ENGLISH AS A SECOND
LANGUAGE***

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



Prof. Dr. Lala Septem Riza, M.T.

NIP. 197809262008121001

Pembimbing II,



Yaya Wihardi, S.Kom., M.Kom.

NIP. 198903252015041001

Mengetahui,

Ketua Program Studi Ilmu Komputer



Dr. Muhammad Nursalman, M.T.

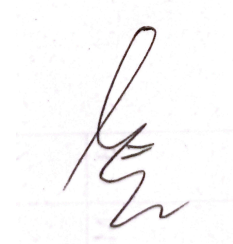
NIP. 197909292006041002

PERNYATAAN

Dengan ini penulis menyatakan bahwa skripsi dengan judul “Implementasi *Natural Language Processing* dan Algoritma *ECLAT* Untuk Sistem Rekomendasi Pada Soal *Sentence Completion* Dalam *English As a Second Language*” ini beserta seluruh isinya adalah benar-benar karya penulis sendiri. Penulis tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, penulis siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya penulis ini.

Bandung, Juli 2024

Yang Membuat Pernyataan,



Muhammad Naufal

Wicaksono

1904802

**IMPLEMENTASI *NATURAL LANGUAGE PROCESSING*
DAN *ECLAT* UNTUK SISTEM REKOMENDASI PADA SOAL *SENTENCE*
COMPLETION DALAM ESL**

Oleh

Muhammad Naufal Wicaksono — naufa58@upi.edu

1904802

ABSTRAK

Penelitian ini mengembangkan sistem rekomendasi yang mengintegrasikan model machine learning ECLAT dan teknik Natural Language Processing (NLP) untuk meningkatkan hasil tes Bahasa Inggris sebagai Bahasa Kedua (ESL). Sistem ini menganalisis jawaban peserta secara mendalam, menggunakan ECLAT untuk efisiensi analisis asosiasi dan NLP untuk mengidentifikasi pola kesalahan serta kelemahan dalam bahasa Inggris. Melalui pengolahan data tekstual, sistem secara otomatis mengklasifikasikan peserta berdasarkan kesamaan masalah bahasa dan menyediakan rekomendasi yang disesuaikan berupa materi pembelajaran, aktivitas, dan sumber daya. Pengujian sistem menggunakan dataset historis dan umpan balik peserta menunjukkan bahwa sistem ini efektif dalam meningkatkan hasil belajar. Rekomendasi yang dihasilkan sesuai dengan kompetensi yang tidak dikuasai siswa menurut para ahli. Rekomendasi diberikan dalam bentuk kalimat teks paragraf dengan urutan prioritas krusial, penting, dan disarankan, sehingga mudah dipahami dan diikuti siswa. Berdasarkan hasil penelitian, disimpulkan bahwa sistem rekomendasi ini mampu memberikan rekomendasi yang relevan dan bermanfaat. Saran untuk penelitian selanjutnya meliputi penambahan variasi data pada pelatihan dan mencoba metode lain untuk meningkatkan kinerja dan akurasi sistem.

Kata Kunci : *ESL, Bahasa Inggris, Sistem Rekomendasi, Natural Language Processing, Machine Learning, ECLAT Algorithm, Association Rule.*

**IMPLEMENTASI *NATURAL LANGUAGE PROCESSING*
DAN ALGORITMA *ECLAT* UNTUK SISTEM REKOMENDASI PADA
SOAL *SENTENCE COMPLETION* DALAM *ENGLISH AS A SECOND
LANGUAGE***

By

Muhammad Naufal Wicaksono — naufa58@upi.edu

1904802

ABSTRACT

This research develops a recommendation system that integrates the ECLAT machine learning model and Natural Language Processing (NLP) techniques to improve English as a Second Language (ESL) test results. The system analyzes participants' answers in depth, using ECLAT for association analysis efficiency and NLP to identify error patterns and weaknesses in English. Through textual data processing, the system automatically classifies participants based on similar language problems and provides customized recommendations on learning materials, activities and resources. Testing the system using historical datasets and participant feedback shows that the system is effective in improving learning outcomes. Recommendations are generated according to the competencies that students do not master according to experts. Recommendations are given in the form of paragraph text sentences with the priority order of crucial, important, and recommended, making it easy for students to understand and follow. Based on the research results, it is concluded that this recommendation system is able to provide relevant and useful recommendations. Suggestions for future research include adding data variations to training and trying other methods to improve system performance and accuracy.

Keywords: *ESL, English, Recommendation System, Natural Language Processing, Machine Learning, ECLAT Algorithm, Association Rule.*

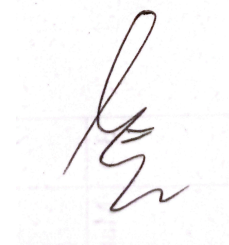
KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT., atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi dengan judul “Implementasi *Natural Language Processing* dan Algoritma *ECLAT* Untuk Sistem Rekomendasi Pada Soal *Sentence Completion* Dalam *English As a Second Language*”. Tidak lupa shalawat serta salam semoga tercurah kepada Nabi Muhammad SAW. beserta keluarganya, sahabatnya hingga akhir zaman.

Skripsi ini disusun untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer atas jenjang S1 pada Program Studi Ilmu Komputer, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia.

Penulis menyadari bahwa dalam penyusunan skripsi ini, tidak luput dalam kesalahan. Maka dari itu penulis terbuka untuk kritik dan saran yang membangun, penulis nantikan kehadirannya, semoga skripsi ini bermanfaat bagi kemajuan Ilmu Pengetahuan dan Teknologi, aamiin.

Bandung, Juli 2024
Penulis,



Muhammad Naufal

Wicaksono

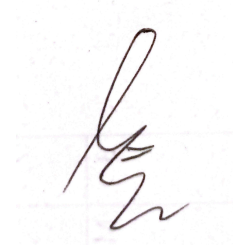
UCAPAN TERIMAKASIH

Segala puji bagi Allah SWT, yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis diberikan kelancaran dalam menyelesaikan penulisan skripsi ini. bagi Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari dukungan, bantuan dan do'a dari berbagai pihak yang terlibat secara langsung maupun tidak langsung selama penyusunan skripsi ini. Oleh karena itu, penulis ingin menyampaikan rasa terimakasih yang sebesar-besarnya kepada:

1. Kedua orang tua, adik dan keluarga penulis yang senantiasa mendoakan dan memberi dukungan baik moral maupun materi kepada penulis supaya diberi kelancaran selama kuliah dan dalam proses penyusunan skripsi ini.
2. Bapak Lala Septem Riza, M.T., Ph.D., sebagai pembimbing I, atas semua ilmu dan waktu yang telah dicurahkan kepada penulis selama penyusunan skripsi ini.
3. Bapak Yaya Wihardi, S.Kom., M.Kom., sebagai pembimbing II, atas ilmu, waktu dan motivasi yang diberikan kepada penulis selama perkuliahan dan penyusunan skripsi ini.
4. Bapak Dr. Muhammad Nursalman, M.T., selaku ketua Program Studi Ilmu Komputer.
5. Para dosen Program Ilmu Komputer atas ilmu yang sangat bermanfaat yang telah ditanamkan kepada penulis.
6. Teman teman dan tim penelitian yaitu Mlepin, Pak Aji, Aryoppa, Aththar, Joel the Joy, Zainul, Jenderal Yudi, Juliansyah, Ajay, Alip, Lazu dan Wafff yang senantiasa memberikan dukungan, semangat, dan kegiatan positif lainnya selama proses pengerjaan penelitian dan skripsi.
7. Para responden yang tidak bisa saya sebutkan satu satu atas waktunya dalam mengisi kuesioner.

Semoga jasa dan kebaikan kalian menjadi ladang pahala dan dibalas oleh Allah SWT., dan semoga dimudahkan dalam segala urusannya.

Bandung, Juli 2024
Penulis,

A handwritten signature in black ink, appearing to be 'Muhammad Naufal Wicaksono', written over a faint, light-colored grid background.

Muhammad Naufal

Wicaksono

DAFTAR ISI

ABSTRAK.....	1
ABSTRACT.....	2
KATA PENGANTAR.....	3
UCAPAN TERIMA KASIH.....	3
DAFTAR ISI.....	6
DAFTAR TABEL.....	9
DAFTAR GAMBAR.....	10
DAFTAR LAMPIRAN.....	11
BAB I	
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	7
1.5 Batasan Masalah.....	7
BAB II	
KAJIAN PUSTAKA.....	8
2.1. Evaluasi Pembelajaran.....	8
2.1.1 Deskripsi Evaluasi Pembelajaran.....	8
2.1.2 Tujuan dan Fungsi Evaluasi Pembelajaran.....	9
2.2. English As A Second Language.....	10
2.2.1. Konsep dan Jenis ESL.....	10
2.2.2. Testing Points.....	11
2.3. Machine Learning.....	13

2.3.1. Supervised Learning.....	16
2.3.2. Unsupervised Learning.....	18
2.3.3. Semi Supervised Learning.....	19
2.3.4. Reinforcement Learning.....	22
2.4. Natural Language Processing.....	24
2.4.1. Tokenization.....	26
2.4.2. Part-of-Speech Tagging (POS Tagging).....	28
2.4.3. Named Entity Recognition (NER).....	30
2.4.4. Dependency Parsing.....	32
2.4.5. Sentiment Analysis.....	34
2.4.6. Machine Translation.....	36
2.4.7. Text Summarization.....	38
2.5. Sistem Rekomendasi.....	40
2.6. Algoritma ECLAT.....	43
2.6.1. dECLAT (diffset ECLAT).....	51
2.6.2. ECLAT-BIT.....	52
2.6.3. VIPER (Vertical Itemset Partitioning for ECLAT Reduction)..	53
2.6.4. Parallel ECLAT.....	54
2.6.5. ECLAT with Dynamic Reordering.....	56
2.7. Penelitian Terkait.....	57
BAB III	
METODE PENELITIAN.....	59
3.1. Desain Penelitian.....	59
3.2. Alat dan Bahan Penelitian.....	65
Berikut alat dan bahan penelitian yang dibutuhkan:.....	65
BAB IV	
HASIL DAN PEMBAHASAN.....	66

4.1. Pengumpulan Data.....	66
4.2. Perancangan Model Komputasi.....	73
4.2.1. Data Preprocessing.....	75
4.2.2. Modelling.....	88
4.2.3. Testing.....	96
4.3. Eksperimen.....	108
4.3.1. Skenario Eksperimen.....	108
4.3.2. Hasil Eksperimen.....	109
4.4. Analisis Hasil Eksperimen.....	115
BAB V	
KESIMPULAN DAN SARAN.....	121
5.1. Kesimpulan.....	121
5.2. Saran.....	121
DAFTAR PUSTAKA.....	122
LAMPIRAN.....	130

DAFTAR PUSTAKA

- Alderson, J. C. (2000). *"Assessing Reading."* Cambridge: Cambridge University Press.
- Agrawal, R., & Srikant, R.. *Fast algorithms for mining association rules*. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc..
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
- Black, P., & Wiliam, D.. Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *"Taxonomy of Educational Objectives: The Classification of Educational Goals."* New York: David McKay Company.
- Blum, A., & Mitchell, T. (1998). *Combining labeled and unlabeled data with co-training*. In Proceedings of the eleventh annual conference on Computational learning theory (pp. 92-100).
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *"How People Learn: Brain, Mind, Experience, and School."* Washington, D.C.: National Academy Press.

- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Longman.
- Buck, G. (2001). *"Assessing Listening."* Cambridge: Cambridge University Press.
- Chapelle, O., Scholkopf, B., & Zien, A. (2006). *Semi-Supervised Learning*. MIT Press.
- Chen, Y., Huang, J., Xia, L., Xu, Y., Dai, P., Chen, Y., Bo, L., Zhao, J., & Huang, J. X. (2021). Graph-enhanced multi-task learning of multi-level transition dynamics for session-based recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 4123-4130.
- Cummins, J. (2000). *"Language, Power, and Pedagogy: Bilingual Children in the Crossfire."* Clevedon: Multilingual Matters.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186).
- Duong, Q. H., & Nguyen, P. T. (2021). Enhancing ESL learner's literacy by peer-assisted learning strategy of online English news. *Frontiers in Psychology*, 14, 1172099.
- Duong, Q. H., & Nguyen, P. T. (2021). Impact of immersive technologies on language learning. *Journal of Computer Assisted Learning*.

- Ezugwu, A. E., Ikotun, A. M., Oyelade, O. N., Abualigah, L. M., Agushaka, J. O., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110, 104743.
- Fi-Eclat: An Enhancement of Incremental Eclat Algorithm. (2023). *AIP Conference Proceedings*.
- Fournier-Viger, P. (2023). The Eclat Algorithm. *Data Mining Blog*.
- Galpin, L. (2023). English language teaching trends to watch in 2023. *Ellii Blog*.
- Gardner, J., O’Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: ‘Breakthrough? Or buncombe and ballyhoo?’. *Journal of Computer Assisted Learning*, 37(5), 1207–1216.
- Gass, S. M., & Selinker, L. (2008). *"Second Language Acquisition: An Introductory Course."* New York: Routledge.
- Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*. Synthesis Lectures on Human Language Technologies, 10(1), 1-309.
- Gong, W. (2023). A systematic review of AI-driven educational assessment in STEM education. *Journal for STEM Education Research*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

- Guimarães, M., Palumbo, G., Carneiro, D., Novais, P., & Alves, V. (2022). Real-Time Algorithm Recommendation Using Meta-Learning. *International Symposium on Ambient Intelligence*, 249-258.
- Han, J., Pei, J., & Yin, Y. (2000). *Mining frequent patterns without candidate generation*. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data.
- Hariyanto, S. (2016). *Belajar dan Pembelajaran Teori dan Konsep Dasar*. PT Remaja Rosdakarya.
- Hastie, T., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Hattie, J., & Timperley, H.. The power of feedback. *Review of educational research*, 77(1), 81-112.
- He, X., & Fu, S. (2022). Data analysis and processing application of deep learning in engineering cost teaching evaluation. *Journal of Mathematics*, 2022(2), 1-12.
- Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 328-339).
- Hutchinson, T., & Waters, A. (1987). *"English for Specific Purposes: A Learning-Centred Approach."* Cambridge: Cambridge University Press.
- Jain, A. K. (2010). *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters, 31(8), 651-666.

- Jones, H. R., Smith, J. T., & Taylor, E. F. (2023). Impact of formative assessments on student engagement and learning outcomes. *Educational Research Review*.
- Jurafsky, D., & Martin, J. H. (2018). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed.). Pearson.
- Lightbown, P. M., & Spada, N. (2013). *"How Languages are Learned."* Oxford: Oxford University Press.
- Luoma, S. (2004). *"Assessing Speaking."* Cambridge: Cambridge University Press.
- Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). *Human-level control through deep reinforcement learning*. *Nature*, 518(7540), 529-533.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Necula, S.-C., & Păvăloaia, V.-D. (2023). AI-Driven Recommendations: A Systematic Review of the State of the Art in E-Commerce. *Applied Sciences*, 13(9), 5531.
- Nunan, D. (2015). *Teaching English to speakers of other languages: An introduction*. Routledge.

- Ogihara, M., & Zaki, M. J. (2023). A Survey on the Use of Association Rules Mining Techniques in Textual Social Media. *Artificial Intelligence Review*.
- Priya, V. (2021). ECLAT. Assistant Professor, Nehru Memorial College, Puthanampatti.
- Richards, J. C. (2001). *"Curriculum Development in Language Teaching."* Cambridge: Cambridge University Press.
- Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- Simpson, E. J. (1972). *"The Classification of Educational Objectives in the Psychomotor Domain."* Washington, D.C.: Gryphon House.
- Smith, J. L., Brown, C. A., & Johnson, M. W. (2023). The integration of technology in classroom assessment. *Journal of Educational Measurement and Technology*.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems* (pp. 3104-3112).
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.
- Tan, Y., et al. (2023). Heterogeneous Federated Learning: State-of-the-art and Research Challenges. *arXiv*.
- Umar, H. (2002). *Evaluasi kinerja perusahaan / Husein Umar*. Jakarta : Gramedia Pustaka Utama,.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. In *Advances in Neural Information Processing Systems* (pp. 5998-6008).
- Watkins, C. J. C. H., & Dayan, P. (1992). *Q-learning*. Machine Learning, 8(3-4), 279-292.
- Weigle, S. C. (2002). *"Assessing Writing."* Cambridge: Cambridge University Press.
- Wen, G. (2023). Enhancing ESL learner's literacy by peer-assisted learning strategy of online English news. *Frontiers in Psychology*. Retrieved from
- Williams, P. R., & Davis, L. E. (2023). Effective practices in educational evaluation. *Journal of Learning Analytics and Assessment*.
- Wu, X., et al. (2020). Federated Learning for Personalization and User-Driven Privacy: Perspectives, Research, and Applications. *IEEE Internet of Things Journal*.
- Zaki, M. J. (1997). *Fast algorithms for mining association rules*. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB).
- Zaki, M. J. (2014). Efficiently mining frequent trees in a forest: Algorithms and applications. *IEEE Transactions on Knowledge and Data Engineering*, 17(8), 1021-1035.
- Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3).

Zaki, M. J., & Goethals, B. (2003). *Advances in frequent itemset mining implementations: Report on FIMI'03* (6th ed., Vol. 1). SIGKDD Explorations Newsletter.

Zhu, X. (2005). *Semi-supervised learning literature survey*.