

**IMPLEMENTASI MODEL FLAN-T5 DALAM PENILAIAN
SKOR URAIAN LITERASI MEMBACA INDIVIDU**

SKRIPSI

Diajukan untuk Memenuhi sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer



oleh
Fajri Maulana Iskandar
2009576

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

2024

IMPLEMENTASI MODEL FLAN-T5 DALAM PENILAIAN SKOR URAIAN LITERASI MEMBACA INDIVIDU

Disusun Oleh:

Fajri Maulana Iskandar

NIM 2009576

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana
Komputer pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Fajri Maulana Iskandar 2024

Universitas Pendidikan Indonesia

Juni 2024

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difotokopi, atau cara lainnya tanpa izin dari penulis

FAJRI MAULANA ISKANDAR

2009576

IMPLEMENTASI MODEL FLAN-T5 DALAM PENILAIAN SKOR
URAIAN LITERASI MEMBACA INDIVIDU

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



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

NIP. 197809262008121001

Pembimbing II,



Rizky Rahman J., M.Kom.

NIP. 197711252006041002

Mengetahui,

Ketua Program Studi Ilmu Komputer



Dr. Muhammad Nursalman, M.T.

NIP. 197909292006041002

PERNYATAAN

Dengan ini penulis menyatakan bahwa skripsi dengan judul “Implementasi Model *FLAN-T5* dalam Penilaian Skor Uraian Literasi Membaca Individu” 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, Agustus 2024

Yang Membuat Pernyataan



Fajri Maulana Iskandar

KATA PENGANTAR

Bismillahirrahmanirrahim, Assalamualikum Wr. Wb.

Puji dan syukur penulis panjatkan kehadiran Allah swt. karena hanya dengan kehendak, berkat, serta karunia-Nya lah penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Model *FLAN-T5* dalam Penilaian Skor Uraian Literasi Membaca Individu” ini dapat terselesaikan.

Alhamdulillahirobbilalamin, puji dan syukur kehadiran Allah SWT Yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis diberikan kelancaran dalam menyelesaikan penulisan skripsi ini. Dalam proses menyelesaikan penelitian dan penyusunan skripsi ini, peneliti banyak mendapat bimbingan, dorongan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terimakasih serta penghargaan yang setinggi-tingginya, kepada:

1. Kedua orang tua yaitu Iskandar Zulkarnain dan Yuli Armiaty yang selalu memberi dukungan, motivasi, serta doa serta selalu menjadi penyemangat utama dalam menempuh pendidikan tinggi sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Prof. Dr. Lala Septem Riza, M.T. selaku pembimbing I atas segala waktu dan tenaga yang dicurahkan untuk membimbing penulis demi terselesaikannya skripsi ini.
3. Bapak Rizky Rahman J, M.Kom. selaku selaku pembimbing II atas saran dan masukan yang diberikan kepada penulis selama proses menyelesaikan penelitian dan penulisan skripsi.
4. Ibu Dr. Rani Megasari, M.T. selaku Mantan Ketua Program Studi Ilmu Komputer Universitas Pendidikan Indonesia Periode 2019 – 2023 yang telah memberikan dukungan topik skripsi ini dan memberikan banyak arahan saat pengujian seminar proposal di skripsi ini.
5. Bapak Dr. Muhammad Nursalman, M.T. selaku Ketua Program Studi Ilmu Komputer Universitas Pendidikan Indonesia yang telah memberikan dukungan atas penulisan skripsi ini.

6. Bapak Yaya Wihardi, S.Kom., M.Kom. selaku dosen pendamping akademik yang telah membimbing serta memberi motivasi penulis dari awal sampai dengan akhir perkuliahan.
7. Seluruh Dosen dan Staf Program Studi Ilmu Komputer dan Pendidikan Ilmu Komputer UPI, yang telah memberikan ilmu dan kesempatan bagi penulis untuk menyelesaikan pendidikan ini.
8. Bapak dan Ibu Dosen Prodi Pendidikan Ilmu Komputer dan Ilmu Komputer yang telah berbagi ilmu yang bermanfaat kepada penulis dan mahasiswa lainnya.
9. Seorang guru SMA Bani Saleh Bekasi Timur yang telah bersedia menjadi ahli untuk penilaian hasil tes literasi membaca yang ter validasi untuk penelitian ini.
10. Teman-teman dan rekan dari DPM & BEM KEMAKOM yang telah memberi pengalaman berharga.
11. Teman-teman seperjuangan Ilmu Komputer 2020 dan Pendidikan Ilmu Komputer 2020.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat kekurangan dan kelemahan. Oleh karena itu, penulis memohon maaf atas kesalahan dan kelemahan yang ada dalam penulisan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi semua pihak yang memerlukannya dan dapat menjadi referensi untuk pengembangan ilmu pengetahuan di bidang yang terkait.

Akhir kata, semoga Allah SWT senantiasa melimpahkan rahmat dan karunia-Nya kepada kita semua. Aamiin.

Bandung, Agustus 2024



Fajri Maulana Iskandar

Implementasi Model FLAN-T5 dalam Penilaian Skor Uraian Literasi Membaca Individu

oleh

Fajri Maulana Iskandar — fajrimaulana402@gmail.com

2009576

ABSTRAK

Penilaian literasi membaca saat ini menghadapi berbagai tantangan, termasuk keterbatasan dalam interpretasi skor dan kurangnya sumber daya untuk melibatkan profesional berpengalaman dalam menilai uraian literasi membaca individu. Penelitian ini bertujuan untuk mengatasi tantangan tersebut dengan mengimplementasikan model *FLAN-T5* yang telah dilatih sebelumnya (*pretrained*) menggunakan teknik *machine learning* dalam penilaian uraian literasi membaca untuk menghasilkan skornya. Model ini dipilih karena kemampuannya dalam mengolah tugas-tugas pemrosesan bahasa alami dengan memahami hubungan antarfrasa. Pendekatan penelitian ini melibatkan beberapa tahapan, mulai dari pengumpulan data melalui pembuatan soal tes literasi membaca, pembersihan data, transformasi data menjadi *dictionary*, tokenisasi, pembangunan model *FLAN-T5*, *fine-tuning* model, hingga pengujian model yang telah di-tuning. Metode ini dirancang untuk meningkatkan efektivitas dalam pelatihan dan evaluasi model, serta menghasilkan prediksi yang lebih akurat dalam tugas-tugas pemrosesan bahasa alami. Data yang digunakan meliputi teks dan pertanyaan, dengan jawaban sebagai komponen utama yang diproses, sementara skor bertindak sebagai label target untuk mendukung pengembangan model. Hasil penelitian menunjukkan bahwa model ini efektif dalam mengidentifikasi jawaban dengan tingkat akurasi 74% yang konsisten tinggi pada setiap kategori label target. Meskipun terdapat beberapa kesalahan penilaian, terutama pada jawaban dengan skor yang sangat tinggi atau rendah, model ini tetap mampu memberikan gambaran yang cukup akurat mengenai kemampuan literasi membaca individu.

Kata Kunci: Literasi Membaca, *FLAN-T5*, Penilaian, *Machine Learning*, *Natural Language Processing*.

Implementation of the FLAN-T5 Model in the Assessment of Individual Reading Literacy Description Scores

Arranged by

Fajri Maulana Iskandar— fajrimaulana402@gmail.com

2009576

ABSTRACT

The current assessment of reading literacy faces various challenges, including limitations in interpreting scores and a lack of resources to engage experienced professionals in evaluating individual reading literacy descriptions. This research aims to address these challenges by implementing the pre-trained FLAN-T5 model using machine learning techniques in the assessment of reading literacy descriptions to generate scores. This model was chosen for its ability to handle natural language processing tasks by understanding the relationships between phrases. This research approach involves several stages, starting from data collection through the creation of reading literacy test questions, data cleaning, transforming data into a dictionary, tokenization, building the FLAN-T5 model, fine-tuning the model, and finally testing the tuned model. This method is designed to enhance the effectiveness of training and evaluating models, as well as to produce more accurate predictions in natural language processing tasks. The data used includes text and questions, with answers as the main component being processed, while scores serve as target labels to support model development. The research results indicate that this model is effective in identifying answers with an accuracy rate of 74%, consistently high across each category of target labels. Although there are some assessment errors, especially in responses with very high or very low scores, this model is still able to provide a fairly accurate picture of an individual's reading literacy skills.

Keywords: Reading Literacy, FLAN-T5, Assessment, Machine Learning, Natural Language Processing.

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	x
BAB I PENDAHULUAN	11
1.1 Latar Belakang	11
1.2 Rumusan Masalah	14
1.3 Tujuan Penelitian	14
1.4 Manfaat Penelitian	14
1.5 Batasan Masalah	14
BAB II KAJIAN PUSTAKA	16
2.1 Literasi	16
2.1.1 Konsep Dasar Literasi	16
2.1.2 Dimensi Literasi	19
2.1.3 Relevansi Literasi Dalam Masyarakat	27
2.1.4 Metodologi Penilaian Literasi Membaca	34
2.1.5 Pemahaman Teks	43
2.1.6 Literasi Tradisional dan Digital	46
2.2 Machine learning	47
2.3 Model FLAN-T5	52
2.4 Natural Language Processing	62
2.5 Penelitian Terkait	67

BAB III METODOLOGI PENELITIAN	74
3.1 Desain Penelitian	74
3.2 Metode Penelitian	76
3.2.1 Metode Pengumpulan Data.....	76
3.2.2 Metode Pengembangan Perangkat Lunak	77
3.3 Alat dan Bahan Penelitian	78
BAB IV HASIL DAN PEMBAHASAN	80
4.1 Pengumpulan Data	80
4.1.1 Penyusunan Soal	80
4.1.2 Prosedur Pengumpulan Data	81
4.2 Perancangan Model Komputasi dengan menggunakan <i>FLAN-T5</i>..	85
4.2.1 Data Collection.....	86
4.2.2 Cleansing Data dan Train/Test Split	86
4.2.3 Transformasi Dataset Menjadi <i>Dictionary</i>.....	90
4.2.4 Tokenisasi	91
4.2.5 Model <i>FLAN-T5</i> Constructions	93
4.2.6 Fine Tuning Model <i>FLAN-T5</i>	94
4.2.7 Testing Fine Tuned Model.....	95
4.3 Pengembangan dan Implementasi Perangkat Lunak	96
4.3.1 Analisis	96
4.3.2 Desain	98
4.3.3 Implementasi	101
4.3.4 Pengujian	115
4.4 Skenario Eksperimen	116
4.5 Hasil	119
4.6 Pembahasan	121

BAB V KESIMPULAN DAN SARAN	131
5.1 Kesimpulan	131
5.2 Saran.....	131
DAFTAR PUSTAKA	133
LAMPIRAN.....	144
Lampiran 1 : Rincian Hasil Skenario Eksperimen <i>Cross Validation fold 10</i>	144

DAFTAR GAMBAR

Gambar 2.1 Pilar Literasi	18
Gambar 2.2 Kerangka heuristik untuk memahami literasi.....	19
Gambar 2.3 Dimensi Literasi.....	20
Gambar 2.4 Diagram Konseptual: Dimensi Kognitif.....	21
Gambar 2.5 Peran Literasi Dalam Mengembangkan Berbagai Aspek Kehidupan Individu.....	33
Gambar 2.6 Contoh Teks pertanyaan.....	40
Gambar 2.7 Diagram Hubungan antara Pemahaman Teks dan Keterampilan Literasi Keseluruhan	45
Gambar 2.8 <i>Supervised Learning</i> Process.....	50
Gambar 2.9 Penyetelan <i>Prompt Tuning FLAN-T5</i>	52
Gambar 2.10 Diagram Alir <i>Fine tuning FLAN-T5</i>	56
Gambar 2.11 Contoh Penggunaan <i>FLAN-T5</i> (HuggingFace, 2022).....	62
Gambar 2.12 Contoh Alur Penggunaan <i>Natural Processing Language</i>	63
Gambar 2.13 Proses <i>Tokenization</i>	66
Gambar 3.1 Desain Penelitian.....	74
Gambar 3.2 Model <i>Waterfall</i> dalam Pengembangan Perangkat Lunak.....	77
Gambar 4.1 Pengumpulan Informasi Dasar Responden	81
Gambar 4.2 Form Pengumpulan Jawaban Responden.....	82
Gambar 4.3 Model Komputasi	85
Gambar 4.4 Sebelum dan Sesudah Menghilangkan Karakter Spesial	87
Gambar 4.5 Menghapus Spasi Kosong	88
Gambar 4.6 Mengubah Ke Huruf Kecil	88
Gambar 4.7 Konversi Kolom Label menjadi <i>String</i>	89
Gambar 4.8 <i>Pseudocode Cleansing Data</i>	89
Gambar 4.9 <i>Pseudocode Split data</i>	89
Gambar 4.10 <i>Pseudocode</i> Transformasi Dataset Menjadi <i>Dictionary</i>	90
Gambar 4.11 <i>Input dan Output</i> Konversi ke <i>Dictionary</i>	91
Gambar 4.12 <i>Input dan Output Combine Column</i> Instruksi dan Jawaban sebagai <i>Prompt Input</i>	91

Gambar 4.13 <i>Input dan Output Tokenisasi Input dan Label</i>	92
Gambar 4.14 <i>Pseudocode Tokenisasi</i>	93
Gambar 4.15 <i>Pseudocode Model FLAN-T5 Constructions</i>	94
Gambar 4.16 <i>Pseudocode Fine tuning Model FLAN-T5</i>	95
Gambar 4.17 <i>Pseudocode Testing Model</i>	96
Gambar 4.18 <i>Diagram Overview Sistem Perangkat Lunak</i>	98
Gambar 4.19 <i>Rancangan Antarmuka Login</i>	99
Gambar 4.20 <i>Rancangan Antarmuka Dashboard Admin</i>	100
Gambar 4.21 <i>Rancangan Antarmuka Jawaban Peserta</i>	100
Gambar 4.22 <i>Rancangan Antarmuka Penilaian Automatis</i>	101
Gambar 4.23 <i>Confusion Matrix Antara Skor Aktual dan Skor Model</i>	121
Gambar 4.24 <i>Grafik Rata-Rata Peningkatan Akurasi Setiap Epoch</i>	125
Gambar 4.25 <i>Diagram Rata - rata Peningkatan Akurasi Setiap Epoch per Parameter (Learning Rate, Weight Decay, dan Batch Size)</i>	126
Gambar 4.26 <i>Diagram Rata - rata Peningkatan Akurasi Setiap Fold per Parameter (Learning Rate, Weight Decay, dan Batch Size)</i>	127
Gambar 4.27 <i>Rata - rata peningkatan Error rate dan Waktu Komputasi Setiap Epoch</i>	128
Gambar 4.28 <i>Perbandingan Error Rate dan Train Time per Epoch pada Setiap Fold</i>	129

DAFTAR TABEL

Tabel 2.1 Poin Rangkuman Dimensi Literasi.....	26
Tabel 2.2 Aspek Penilaian Kognitif	35
Tabel 2.3 Dimensi dan Jenis Task	36
Tabel 2.4 Tingkat Kompetensi Literasi Membaca.....	42
Tabel 2.5 Perbedaan Literasi Tradisional dan Digital	47
Tabel 2.6 Penelitian Terkait	68
Tabel 4.1 Dataset	84
Tabel 4.2 Pengujian Metode Black Box	115
Tabel 4.3 Skenario Pengujian Model Komputasi <i>FLAN-T5</i>	119
Tabel 4.4 Hasil Simulasi Model Sebagai <i>Output</i> Program	120
Tabel 4.5 Hasil <i>Cross Validation</i> fold 10.....	124

DAFTAR PUSTAKA

- Ahmed, M. (2011). Defining and measuring literacy: Facing the reality. *Int Rev Educ*, 179–195.
- Ai, L., Hui, Z., Liu, Z., & Hirschberg, J. (2024). Enhancing Pre-Trained Generative Language Models with Question Attended Span Extraction on Machine Reading Comprehension. *arXiv preprint arXiv*.
- Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, 235–251.
- Alsariera, Y., Baashar, Y., Alkawsy, G., Mustafa, A., Alkahtani, A., & Ali, N. (2022). Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. *Computational Intelligence and Neuroscience*.
- Altin, S., Finke, I., Kautz-Freimuth, S., & Stock, S. (2014). The evolution of health literacy assessment tools: a systematic review. *BMC Public Health*, 1207–.
- Badcock, C. (1988). Culture and the Evolutionary Process. *Man (New Series)*, , 204–205.
- Barton, D. (2001). *Literacy and Motivation*. Routledge.
- Beatie, B. A. (1986). Literacy, Language, and Learning. *The Modern Language Journal*, 174–175. .
- Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low Health Literacy and Health Outcomes: An Updated Systematic Review. *Annals of Internal Medicine*,.
- Borgonovi, F., Choi, A., & Paccagnella, M. (2021). The evolution of gender gaps in numeracy and literacy between childhood and young adulthood. *Economics of Education Review*.
- Bosma, M., & Jason Wei. (2021, Oktober 6). *Introducing FLAN: More generalizable Language Models with Instruction Fine-Tuning*. Retrieved from Google Research: <https://research.google/blog/introducing-flan-more-generalizable-language-models-with-instruction-fine-tuning/>
- Bruthiaux, P. (2002). Hold Your Courses: Language Education, Language Choice, and Economic Development. *TESOL Quarterly*, 275–296.

- Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2018). Malleability, plasticity, and individuality: How children learn and develop in context. *Applied Developmental Science*, 1–31.
- Carpenter, D., Min, W., Lee, S., Ozogul, G., Z. X., & Lester, J. (2024). Assessing Student Explanations with Large Language Models Using Fine-Tuning and Few-Shot Learning. *In Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, 403-413.
- Clark, J., Garrette, D., Turc, I., & Wieting, J. (2021). Canine: Pre-training an Efficient Tokenization-Free Encoder for Language Representation. *Transactions of the Association for Computational Linguistics*, 73-91.
- Contemporary Educational Psychology Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 293–316.
- Cunningham, P., Cord, M., & Delany, S. (2008). Supervised Learning. *Springer*, 21-49.
- Datacamp. (2023, November). *FLAN-T5 Tutorial: Guide and Fine-Tuning*. Retrieved from datacamp: <https://www.datacamp.com/tutorial/flan-t5-tutorial>
- David, D., & Tolchinsky, I. (2002). *Developing linguistic literacy: a comprehensive model*. Cambridge: Cambridge University.
- Dietz, W. H. (1994). Critical periods in childhood for the development of obesity. *The American Journal of Clinical Nutrition*, 955–959.
- Edwards, A., & Camacho-Collados, J. (2024). Language Models for Text Classification: Is In-Context Learning Enough? *arXiv*.
- Edwards, L., Bryant, A., Keegan, R., Morgan, K., Cooper, S., & Jones, A. (2017). ‘Measuring’ Physical Literacy and Related Constructs: A Systematic Review of Empirical Findings. *Sports Medicine (Auckland, N.z.)*, 659 - 682.
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The Impact of Vocabulary Instruction on Passage-Level Comprehension of School-Age Children: A Meta-Analysis. *Journal of Research on Educational Effectiveness*, 1–44.

- Erkaya, E., & Güngör, T. (2023). Analysis of Subword Tokenization Approaches for Turkish Language. *2023 31st Signal Processing and Communications Applications Conference (SIU)*, 1-4.
- Fahd, K., Venkatraman, S., & Miah, S. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Educ Inf Technol*, 3743–3775.
- Farrar, M. J., & Maag, L. (2002). Early language development and the emergence of a theory of mind. *First Language*, 197-213.
- Ferstl, E., Neumann, J., & Bogler, C. a. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Hum. Brain Mapp*, 581-593.
- Frankel, K. K., Becker, B. L., Rowe, M. W., & Pearson, P. D. (2016). From “what is reading?” to what is literacy? *Journal of education*, 7-17.
- Fransman, J. (2005). *Understanding literacy: a concept paper*. UNESCO.
- Friedman, R. (2023). Tokenization in the Theory of Knowledge. *Encyclopedia*, 380-386.
- Funtowicz, M. (2020). *Hugging Face Tutorials - Training Tokenizer*. Retrieved from kaggle: <https://www.kaggle.com/code/funtowiczmo/hugging-face-tutorials-training-tokenizer>.
- Gazette, G. (2013). *four-dimensions-of-literacy*. Retrieved from gutierre gazette: <https://gutierrezgazette.weebly.com/four-dimensions-of-literacy.html>
- Gee, J. P., & Hayes, E. R. (2011). *Language and Learning in the Digital Age*. Canada: Routledge.
- Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol: O'Reilly Media.
- Ghosal, D., Majumder, N., Mehrish, A., & Poria, S. (2023). Text-to-Audio Generation using Instruction-Tuned LLM and Latent Diffusion Model. *ArXiv*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Optimization for Training Deep Models. In I. Goodfellow, Y. Bengio, & A. Courville, *Deep Learning* (p. 276). Massachusetts: MIT Press.

- Gotch, C., & French, B. (2014). A Systematic Review of Assessment Literacy Measures. *Educational Measurement: Issues and Practice*, 14-18.
- Guo, S., Armstrong, R., Waters, E., Sathish, T., Alif, S., Browne, G., & Yu, X. (2018). Quality of health literacy instruments used in children and adolescents: a systematic review. *BMJ Open*, 8.
- Haberlandt, K. (1982). Reader Expectations in Text Comprehension. *Language And Comprehension*, 239–249.
- Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of Child Poverty, Brain Development, and Academic Achievement. *JAMA Pediatrics*.
- Hashimi, M. B., & Momand, R. (2010). A hybrid approach to identify and eliminate stop words from the Pashto text.
- Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenom Cogn Sci*, 577-598.
- Hensch, T. K. (2004). Critical Period Regulation. *Annual Review of Neuroscience*, 549–579.
- Hicken, A., & Simmons, J. W. (2008). The Personal Vote and the Efficacy of Education Spending. *American Journal of Political Science*, 109–124.
- Hiraoka, T., Takase, S., Uchiumi, K., Keyaki, A., & Okazaki, N. (2021). Joint Optimization of Tokenization and Downstream Model. *ArXiv*.
- Hodge, M. A., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., . . . Mathieson, S. (2018). Literacy Assessment Via Telepractice Is Comparable to Face-to-Face Assessment in Children with Reading Difficulties Living in Rural Australia. *Telemedicine and e-Health*.
- Horning, A. S. (2007). Defining Literacy And Illiteracy. *The Reading Matrix*.
- Hsieh, W.-Y., Hemmeter, M. L., McCollum, J. A., & Ostrosky, M. M. (2009). Using coaching to increase preschool teachers' use of emergent literacy teaching strategies. *Early Childhood Research Quarterly*, 0–247.
- Hu, J., Peng, Y., & Chen, X. (2023). Decoding Contextual Factors Differentiating Adolescents' High, Average, and Low Digital Reading Performance Through Machine-Learning Methods. *IEEE Transactions on Learning Technologies*, 516-527.

- HuggingFace. (2022). *huggingface*. Retrieved from T5: https://huggingface.co/docs/transformers/main/en/model_doc/t5#transformers.T5ForConditionalGeneration.
- Hughes, G., & Dobbins, C. (2015). The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs). *Research and Practice in Technology Enhanced Learning*, 10-10.
- Hull, G. A., & Moje, E. B. (2012). What is the Development of Literacy the Development Of? In S. University, *Commissioned Papers on Language and Literacy Issues in the Common Core State Standards and Next Generation Science Standards* (p. 64). California: Understanding Language.
- HyungWonChung, LeHou, ShayneLongpre, BarretZoph, YiTay, Fedus, W., . . . Pellat, M. (2022). Scaling Instruction-Finetuned Language Models. *Google*.
- Issah, I., Appiah, O., Appiahene, P., & Fuseini Inusah. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, 2772-6622.
- Jacobi, M., & Moran, M. G. (1990). *Research in Basic Writing: A Bibliographic Sourcebook*. Westport: Library of Congress.
- Jaiswal, A., Babu, A., Zadeh, M., Banerjee, D., & Makedon, F. (2020). A Survey on Contrastive Self-supervised Learning. *ArXiv*.
- Jobard, G., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *Elsevier*, 784–800.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 14-26.
- Jolly, S., & Agrawal, R. (2019). Anatomizing Lexicon With Natural Language Tokenizer Toolkit 3. *Advances in Data Mining and Database Management*.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 255-260.
- Kadhim, A. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, 273-292.

- Kavlakoglu, E. (2020, November 12). *NLP vs. NLU vs. NLG: the differences between three natural language processing concepts*. Retrieved from IBM: <https://www.ibm.com/blog/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts>.
- Korkmaz, C., & Correia, A.-P. (2019). A review of research on machine learning in educational technology. *Educational Media International*, 1469-5790.
- Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica (Slovenia)*, 249-268.
- Krogh, A., & Hertz, J. (1991). A simple weight decay can improve generalization. *Advances in neural information processing systems*, 4.
- Kucer, S., & Silva, C. (2012). *Teaching the Dimensions of Literacy*. New York: Routledge.
- Kuleto, V., Ilić, M., M. D., Ranković, M., Martins, O. M., Păun, D., & Mihoreanu, d. (2021). Exploring Opportunities and Challenges of Artificial Intelligence and Machine Learning in Higher Education Institutions. *Sustainability*, 13-18.
- L. Taylor, H. A. (2000). Connecting Schools, Families, and Communities. *Professional school counseling*, 298-307.
- Lee, J., & Shute, V. J. (2010). Personal and Social-Contextual Factors in K–12 Academic Performance: An Integrative Perspective on Student Learning. *Educational Psychologist*, 185-2-2.
- Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic Status and Child Development: A Meta-Analysis. . *Journal of Emotional and Behavioral Disorders* , 211–224.
- Lin, L.-M., & Zabrocky, K. M. (1998). Calibration of Comprehension: Research and Implications for Education and Instruction. *Contemporary Educational Psychology*, 345–391.
- Longpre, S., Hou, L., Vu, T., Webson, A., Chung, H. W., Tay, Y., . . . Roberts, A. (2023). The Flan Collection: Designing Data and Methods for Effective Instruction Tuning. *Google Research*.

- Mackey, L. M., Doody, C., Werner, E. L., & Fullen, B. (2016). Self-Management Skills in Chronic Disease Management: What Role Does Health Literacy Have?. . *Medical Decision Making*.
- Maulud, D. H., Ameen, S. Y., Omar, N., Kak S. F., R. Z., Yasin, H. M., & Ahmed, D. M. (2021). Review on natural language processing based on different techniques. *Asian Journal of Research in Computer Science*, 1-17.
- Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*.
- Merchant, G. (2007). Writing the future in the digital age. *UKLA*, 118–128.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Moran, A. (2014). Sub-Saharan Africa. *Global Heart*, 23–28.
- Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 544-551.
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*, 51-62.
- Nicula, B., Dascalu, M., Arner, T., Balyan, R., & McNamara, D. (2023). Automated Assessment of Comprehension Strategies from Self-Explanations Using LLMs. *Information* , 10-567.
- Nutbeam, D. (2000). Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. . *Health Promotion International*, 259–267.
- OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. Paris: PISA, OECD.
- Ofori, F., Maina, E., & Gitonga, R. (2020). Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review. *Journal of Information and Technology*, 33-55.
- Okan, O., Lopes, E., Bollweg, T., Bröder, J., Messer, M., Bruland, D., . . . Pinheiro, P. (2018). Generic health literacy measurement instruments for children and adolescents: a systematic review of the literature. *BMC Public Health*, 18.

- Ortaokulu, K. S., Bağlar, Diyarbakır, & Turkey. (2022). The Importance of Literacy. *Scholars Journal of Arts, Humanities and Social Sciences*, 2347-5374.
- Osisanwo, F. Y., Akinsola, J. E., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 128-13.
- Oza, J., & Yadav, H. (2023). Enhancing Question Prediction with Flan T5- A Context-Aware Language Model Approach. *TechRxiv*.
- Panel, N. R. (2000). *Teaching Children To Read : An Evidence-Based Assessment of The Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Chicago: National Reading Panel .
- Pangakis, N., & Wolken, S. (2024). Knowledge Distillation in Automated Annotation: Supervised Text Classification with LLM-Generated Training Labels. *arXiv preprint*.
- Parker, A. J., & Newsome, W. T. (1998). Sense And The Single Neuron: Probing the Physiology of Perception . *Annual Review of Neuroscience*, 227–277.
- Parker, R. M., Ratzan, S. C., & Lurie, N. (2003). Health Literacy: A Policy Challenge For Advancing High-Quality Health Care. *Health Affairs*, 147–153.
- Perfetti, C., Yang, C.-L., & Schmalhofer, F. (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology*, 303–318.
- Perry, K. H. (2012). What is Literacy? – A Critical Overview of Sociocultural. *Journal of Language & Literacy Education*, 50-71.
- Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicide note classification using natural language processing: A content analysis. *Biomedical informatics insights*.
- Piper, B., Simmons Zuilkowski, S., Dubeck, M., Jepkemei, E., & King, S. J. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development*, 324-336.

- Progress, N. A. (2013, n.d. n.d.). *National Report Card*. Retrieved from National Report Card Question Tool: <https://www.nationsreportcard.gov/nqt/searchquestions>.
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained Models for Natural Language Processing: A Survey. *arXiv*.
- Ramdoss, S., Mulloy, A., Lang, R., O'Reilly, M., Sigafos, J., Lancioni, G., . . . Zein, F. (2011). Use of Computer-Based Interventions to Improve Literacy Skills in Students with Autism Spectrum Disorders: A Systematic Review. *Research in Autism Spectrum Disorders*, 1306-1318.
- Razaulla, S., Pasha, M., & Farooq, M. (2022). *Integration of Machine Learning in Education: Challenges, Issues and Trends*. Singapore: Springer.
- Resnick, D. P., & Ed. (1988). *Literacy of Congress Persepective*. Washington DC: Library of Congress.
- Rivlin, L. G., & Weinstein, C. S. (1984). Educational issues, school settings, and environmental psychology. . *Journal of Environmental Psychology*, 347–364.
- Roberts, B. W., Wood, D., & Smith, J. L. (2005). Evaluating Five Factor Theory and social investment perspectives on personality trait development. . *Research in Personality*, 166–184.
- Roth, W.-M. (2014). Socio-Cultural Perspectives on Learning Science. . *Encyclopedia of Science Education*, 1–12.
- Santafé, G., Inza, I., & Lozano, J. (2015). Dealing with the evaluation of supervised classification algorithms. *rtificial Intelligence Review*, 467-508.
- Sénéchal, M., & LeFevre, J.-A. (2002). Parental Involvement in the Development of Children's Reading Skill: A Five-Year Longitudinal Study. *Child Development*, 445–460.
- Shaip. (2023, April 11). *Demystifying NLU: A Guide to Understanding Natural Language Processing*. Retrieved from Shaip: <https://www.shaip.com/blog/demystifying-nlu-a-guide-to-understanding-natural-language-processing>.

- Sheikhaei, M. S., Tian, Y., Wang, S., & Xu, B. (2024). An Empirical Study on the Effectiveness of Large Language Models for SATD Identification and Classification. *arXiv*.
- Shetty, S., Shetty, S., Singh, C., & Rao, A. (2022). Supervised Machine Learning: Algorithms and Applications. In *Fundamentals and Methods of Machine and Deep Learning*.
- Sinclair, J., Jang, E., & Rudzicz, F. (2021). Using machine learning to predict children's reading comprehension from linguistic features extracted from speech and writing. *Journal of Educational Psychology*, 1088-1106.
- Sivakumaran, L. (2020, Februari 18). *Waterfall Model vs Agile Model*. Retrieved from Medium: <https://medium.com/@luxana2304/waterfall-model-vs-agile-model-a48bf2e85dc4>.
- Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *The future of children*, 57-74.
- Stöter, F.-R., Chakrabarty, S., Edler, B., & Habets, E. A. (2018). Classification vs. Regression in Supervised Learning for Single Channel Speaker Count Estimation. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 436-440.
- Stromswold, K. (2001). The Heritability of Language: A Review and Metaanalysis of Twin, Adoption, and Linkage Studies. *Project Muse*, 647-723.
- Sun, J., Shaib, C., & Wallace, B. C. (2023). Evaluating the zero-shot robustness of instruction-tuned language models. *arXiv preprint arXiv*.
- Tashakkori, A., & Teddlie, C. (2010). *Sage handbook of mixed methods in social & behavioral research*. California: sage.
- Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*.
- Thornton, A., & Fricke, T. E. (1987). Social change and the family: Comparative perspectives from the west, China, and South Asia. *University Of Michigan*, 746-779.
- Tsinarakli, C., & Kostic, U. (2020). *Historical Evolution of Artificial Intelligence*. Luxembourg: Publications Office of the European Union.

- UNESCO. (2013). *What you need to know about literacy*. Retrieved from UNESCO: <https://www.unesco.org/en/literacy/need-know>.
- Urraco, D., & Manuel, J. (2023). Efficient summarization of video messages for work. *E.T.S. de Ingenieros Informáticos (UPM)*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Wagner, D. (2003). Smaller, Quicker, Cheaper: Alternative Strategies for Literacy Assessment in the UN Literacy Decade. *International Journal of Educational Research*, 293-309.
- Zhang, D., & Tsai, J. (2003). Machine Learning and Software Engineering. *Software Quality Journal*, 87–119.
- Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discovery Today*, 1680-1685.
- Zhang, Y. J., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International journal of machine learning and cybernetics*, 43-52.
- Zhou, M., Duan, N., Liu, S., & Shum, H. (2020). Progress in Neural NLP: Modeling, Learning, and Reasoning. *Engineering*, 275-290.
- Zhou, Z. (2018). A brief introduction to weakly supervised learning. *National Science Review*, 44-53.