**RANCANG BANGUN WEB *AUTOMATED SHORT ANSWER SCORING* DENGAN METODE TF-IDF DAN *COSINE SIMILARITY* GUNA MENILAI KEMAMPUAN *GRAMMAR* SISWA DENGAN *SDLC* MODEL PXP**

**TUGAS AKHIR**

Diajukan sebagai syarat menyelesaikan jenjang strata Satu (S-1) di Program Studi Teknik Informatika, Jurusan Teknologi, Produksi dan Industri, Institut Teknologi Sumatera

**Oleh:**

**Markus Togi Fedrian Rivaldi Sinaga**

**118140037**

****

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNOLOGI, PRODUKSI, DAN INDUSTRI**

**INSTITUT TEKNOLOGI SUMATERA**

**LAMPUNG SELATAN**

**2022**

# BAB I PENDAHULUAN

## Latar Belakang Masalah

Bahasa Inggris merupakan bahasa yang paling banyak digunakan di dunia. Per 2022, jumlah penutur bahasa Inggris adalah sebanyak kurang lebih 1,5 miliar orang yang tersebar di 118 negara, bahkan terdapat lebih dari 60 negara yang menjadikan bahasa Inggris sebagai bahasa resmi nasional [1][2].Sementara bila dilihat dari keseluruhan pengguna internet, per akhir Maret 2020, bahasa Inggris juga menjadi bahasa yang paling banyak digunakan dengan nilai kontribusi sebesar 25%, yaitu sebanyak 1,18 miliar pengguna dari total 4,58 miliar jumlah pengguna internet di seluruh dunia [3]**.** Di Indonesia sendiri banyak lembaga dan instansi yang menjadikan kemampuan berbahasa Inggris sebagai parameter pengukuran untuk berbagai tujuan tertentu. Saya ambil contoh, hampir seluruh perguruan tinggi di Indonesia menjadikan kemampuan berbahasa Inggris sebagai parameter pengujian kelayakan seorang calon mahasiswa untuk diterima di perguruan tinggi tersebut. Hal ini lumrah bila mengingat bahwa bahasa Inggris juga layak dikatakan sebagai bahasa ilmu pengetahuan, yang diperkuat oleh fakta bahwa sebanyak 1,26 juta judul buku yang berbeda di seluruh dunia, tidak termasuk edisi-edisi lanjutannya, ditulis dalam bahasa Inggris ketika dicetak perdana, juga lebih dari 28 ribu jurnal ilmiah ditulis dalam bahasa Inggris, terpaut jauh dari peringkat kedua yaitu bahasa Jerman, yang hanya mencapai kurang dari 6800 jurnal ilmiah [4], terlebih lagi, saat ini nilai tes kemampuan bahasa Inggris TOEFL (*Test of English as a Foreign Language*), maupun IELTS (*International English Language Testing System*), dijadikan salah satu prasyarat kelulusan di berbagai perguruan tinggi di Indonesia.Di samping itu semua, dalam persaingan dunia kerja pun, kemampuan berbahasa Inggris sangat dihargai para perekrut di berbagai perusahaan di Indonesia.

Mengetahui berbagai fakta tersebut, rasanya realistis bila dikatakan bahwa kemampuan bahasa Inggris merupakan hal yang wajib dimiliki setiap orang yang ingin bersaing dalam perkembangan zaman ini. Akibatnya perlu adanya dorongan yang kuat, khususnya dari penyelenggara pendidikan di Indonesia untuk memajukan indeks kecakapan berbahasa Inggris masyarakat Indonesia. Dalam upaya untuk mewujudkan hal tersebut, dibutuhkan peningkatan baik secara kuantitas yang meliputi jumlah tenaga pengajar, bobot mata pelajaran bahasa Inggris di sekolah dan perguruan tinggi, dan dukungan bagi sebanyak mungkin penyelenggara pelatihan bahasa Inggris; maupun secara kualitas yang meliputi peningkatan materi dalam mata pelajaran bahasa Inggris di sekolah maupun di perguruan tinggi, kemudahan untuk menjangkau penyedia jasa pelatihan dan pengujian kemampuan bahasa Inggris, serta peningkatan efisiensi baik dari segi harga maupun waktu dalam berbagai hal yang telah disebutkan.

Pada penelitian ini sendiri, yang menjadi fokus utama adalah bagaimana ilmu yang diajarkan pada program studi Teknik Informatika dapat dimanfaatkan untuk meningkatkan kualitas dan efisiensi dalam pengujian kemampuan bahasa Inggris seseorang, khususnya siswa, dengan melihat dari kemampuan *grammar* (tata bahasa) melalui penilaian jawaban teks singkat secara otomatis, dalam rangka memberikan evaluasi, sehingga dapat digunakan untuk tujuan lebih lanjut, yaitu meningkatkan kemampuan berbahasa Inggris siswa tersebut. *Grammar* sendiri, merupakan komponen terpenting dalam kemampuan berbahasa Inggris baku/standar, yang mana kemampuan ini dapat digunakan sekaligus dalam kemampuan memahami dari membaca (*reading*), dan mendengar (*listening*), serta kemampuan menyampaikan makna melalui berbicara (*speaking*), dan menulis (*writing*), sehingga dapat dikatakan, kemampuan *grammar* yang baik adalah kombinasi dari seluruh keempat kemampuan berbahasa Inggris tersebut [5]. Untuk menjawab hal yang menjadi fokus utama penelitian ini, dibutuhkan adanya suatu sistem yang dapat melakukan penilaian otomatis tersebut. Sehingga diharapkan, penelitian ini dapat memberi kontribusi dalam peningkatan mutu dan memberikan kemudahan dalam pelaksanaan tes kemampuan bahasa Inggris, namun perlu diperhatikan, bahwa penelitian ini hanya berfokus pada proses perancangan dan pembangunan *Automated Short Answer Scoring System* berbasis web untuk mengevaluasi kemampuan *grammar* siswa melalui penilaian teks jawaban singkat. Sistem yang dikembangkan dan diteliti berupa aplikasi berbasis web, dengan fitur ujian. Berbeda dengan ujian bertipe objektif seperti pilihan berganda yang dapat ditentukan kunci jawabannya terlebih dahulu, dan kemungkinan jawaban peserta juga terbatas, serta saat ini sudah banyak sistem yang dapat melakukan penilaiannya secara otomatis, ujian pada sistem yang dikembangkan pada penelitian ini berfokus pada ujian bertipe subjektif seperti esai atau isian singkat, dimana jawaban yang diberikan peserta dapat sangat bervariasi berdasarkan tingkat pemahaman mereka. Hal ini juga yang menjadi faktor penting yang mendorong perlu dilakukannya penelitian ini, karena saat ini, masih sangat sedikit sistem yang menawarkan kemampuan penilaian jawaban berupa teks dari ujian bertipe esai atau isian singkat. Untuk tipe soal seperti ini, biasanya penilaiannya dilakukan secara manual oleh guru atau tenaga pengajar yang memberikan ujian, dimana hal ini sangat berpotensi mengakibatkan inkonsistensi dalam penilaian, karena dalam waktu singkat guru diberi kewajiban untuk menilai banyak jawaban yang beragam.

Agar proses perancangan dan pembangunan *Automated Short Answer Scoring System* ini dapat berhasil, tentunya dibutuhkan pemahaman yang baik terkait terminologi, definisi, dan konsep dari *Automated Scoring*. *Automated Scoring* dapat dimaknai sebagai proses konversi dari performa dalam penyelesaian tugas – umumnya di bidang pendidikan dan penelitian – menjadi berbagai level atau karakteristik kualitas kemampuan, yang mana proses ini umumnya dilakukan secara otomatis menggunakan bantuan komputer [6]. Kemudian, karena yang menjadi fokus dari penelitian ini adalah penilaian teks jawaban singkat otomatis, maka bidang terfokus yang paling sesuai adalah *Automated Essay Scoring*, yang secara sederhana dapat dimaknai sebagai pemanfaatan kemampuan komputer untuk melakukan evaluasi dan penilaian pada kalimat yang ditulis atau diketik secara otomatis [7]. *Automated Essay Scoring* sendiri merupakan aplikasi dari Pengolahan Bahasa Alami (*Natural Language Processing*) yang merupakan salah satu fokus bidang dari Kecerdasan Buatan (*Artificial Intelligence*) [7]. Dalam praktiknya, proses *Automated Essay Scoring* ini menggunakan teknik pembobotan kata (*Term Weighting*) yang merupakan teknik yang digunakan untuk mengukur dan memberi peringkat kecocokan suatu dokumen terhadap suatu kueri, kemudian dilanjutkan dengan menerapkan teknik pengukuran kemiripan (*Similarity Measurement*) [8].

Terdapat beberapa jenis metode/skema dalam teknik pembobotan kata, seperti TF (*Term Frequency*), IDF (*Inverse Document Frequency*), TF-IDF (*Term Frequency-Inverse Document Frequency*), RF (*Relevance Frequency*), TF-RF (*Term Frequency-Relevance Frequency*), dsb. TF dan IDF merupakan skema pembobotan bertipe tanpa pengawasan (*unsupervised*), dimana TF digunakan untuk menghitung bobot suatu kata/istilah dalam suatu dokumen dengan melihat frekuensi kemunculannya, sementara IDF digunakan untuk mereduksi kekuatan diskriminasi suatu kata/istilah dengan memperhatikan jumlah dokumen yang tersedia [8][9][10]. Di lain sisi, RF merupakan skema pembobotan bertipe dengan pengawasan (*supervised*), dimana RF ini digunakan untuk meningkatkan kekuatan diskriminasi suatu kata/istilah dengan memperhatikan perbandingan antara jumlah dokumen yang dikategorikan sebagai positif dengan yang dikategorikan sebagai negatif (pengkategorian ini berkaitan dengan pelabelan pada teknik pembelajaran mesin dengan pengawasan) [9][10]. Dalam penelitian ini, skema dari teknik pembobotan kata yang dipilih adalah TF-IDF dengan pertimbangan bahwa TF dan IDF, keduanya merupakan teknik pembobotan kata bertipe tanpa pengawasan, yang mana hal ini sesuai dengan tujuan *Automated* *Essay Scoring* itu sendiri. Sementara untuk teknik pengukuran kemiripan (*Similarity Measurement*) sendiri, dipilih salah satu skema yang sudah sangat banyak digunakan, yaitu skema *Cosine Similarity* karena tidak seperti salah satu skema alternatif lainnya yaitu *Euclidean* Similarity, pada skema ini, ukuran dokumen (jumlah kata/istilah) yang dibandingkan tidak begitu mempengaruhi nilai kemiripan yang didapatkan, dengan demikian, akurasi yang dapat dilakukan menjadi lebih tinggi [11], [12].

Untuk pengembangan aplikasi berbasis web-nya sendiri, digunakan SDLC (*System Development Life Cycle*, Siklus Hidup Pengembangan Sistem) model *Personal eXtreme Programming* (PXP), yang merupakan pengembangan dari model terdahulunya yaitu *eXtreme Programming* (XP). XP berfokus pada empat hal, yaitu : keterlibatan klien, pengujian berkelanjutan, pemrograman dilakukan berpasangan, serta siklus iterasi yang singkat . Sementara untuk model PXP, dapat dikatakan sebagai variasi atau pengembangan dari model XP itu sendiri. Meski tidak jauh berbeda dengan XP, PXP lebih menekankan pada keotonomian [13]. Untuk pengembang yang bekerja sendirian PXP lebih menguntungkan karena pengembang dapat bekerja dengan menggunakan caranya sendiri, dan pengembang menjadi lebih mudah dalam melacak serta memprediksi perubahan yang akan terjadi [14]. Selain itu kelebihan dari PXP adalah, pada tahap inisialisasi proyek pengembang tidak diharuskan untuk menyusun perencanaan yang begitu dalam sekaligus begitu menyeluruh, hal ini dikarenakan sifat fleksibel yang juga menjadi karakteristik dari metode SDLC model ini.

Sementara untuk berbagai teknologi yang akan digunakan dalam proses pengembangannya, antara lain bahasa pemrograman Python dengan beberapa *library* tambahan yang dibutuhkan untuk integrasi antara aplikasi dengan model *Automated Scoring*, serta HTML, CSS, dan Javascript, dengan bantuan *framework* Vue.js dan beberapa *package* tambahan sesuai kebutuhan untuk sisi *front-end* sistem, juga bahasa pemrograman PHP, dengan bantuan *framework* Laravel untuk sisi *back-end*. Untuk kebutuhan basis data, peneliti menggunakan teknologi basis data MySQL dengan bantuan XAMPP. Sementara untuk mendukung model SDLC yang dipilih, untuk metode pengujian sistem (*system testing*), akan digunakan metode *Unit Testing* dan *Black Box Testing*.

## Rumusan Masalah

Berdasarkan latar belakang permasalah yang telah diidentifikasi, berikut adalah rumusan masalah yang dapat saya susun :

1. Bagaimana proses pengembangan model *Automated Scoring* menggunakan metode TF-IDF dan *Cosine Similarity* untuk melakukan penilaian jawaban teks singkat secara otomatis, untuk mengevaluasi kemampuan *grammar* seorang siswa?
2. Bagaimana aplikasi berbasis web yang dibangun dapat mengoptimalkan kemampuan model *Automatic Scoring* yang telah dikembangkan?
3. Bagaimana proses integrasi aplikasi berbasis web yang dibangun dengan model kecerdasan buatan yang ada?
4. Bagaimana proses metode SDLC model Personal Extreme Programming digunakan dalam perancangan dan pembangunan sistem?
5. Bagaimana metode *Unit Testing* dan *Black Box Testing* digunakan dalam pengujian untuk menunjang penggunaan metode SDLC model Personal Extreme Programming?
6. Bagaimana kesesuaian sistem yang dibuat dengan tujuan meningkatkan kemudahan dan efisiensi dalam pengujian kemampuan bahasa Inggris seorang siswa?

## Tujuan Penelitian

Berikut adalah tujuan dari penelitian ini berdasarkan masalah-masalah yang telah dirumuskan :

* 1. Mengembangkan model *Automated Scoring* menggunakan metode TF-IDF dan *Cosine Similarity* untuk melakukan penilaian jawaban teks singkat secara otomatis, untuk mengevaluasi kemampuan *grammar* seorang siswa.
  2. Merancang dan membangun sistem sebagai wadah penggunaan model kecerdasan buatan yang ada.
  3. Membangun aplikasi berbasis web yang dapat terintegrasi dengan baik dengan model kecerdasan buatan yang ada.
  4. Menganalisis penggunaan metode SDLC model *Personal eXtreme Programming* dalam perancangan dan pembangunan sistem.
  5. Menganalisis penggunaan metode *Unit Testing* dan *Black Box Testing* dalam pengujian sistem sebagai upaya untuk menunjang penggunaan metode SDLC model *Personal eXtreme Programming*.
  6. Menganalisis tingkat kemudahan dan efisiensi yang ditawarkan sistem dalam pengujian kemampuan bahasa Inggris siswa.

## Batasan Masalah

Batasan masalah dalam penelitian ini perlu ditetapkan untuk kespesifikan tujuan dari penelitian, dan tidak membebani berbagai pihak yang terkait dengan penelitian ini, adapun rumusan masalahnya sebagai berikut :

1. Penelitian hanya bertujuan untuk membuat model *Automated Scoring* dan aplikasi berbasis web untuk menjadi wadah implementasinya.
2. Penelitian hanya berfokus dalam melakukan analisis, untuk perbandingan maupun peningkatan pengembangan lebih lanjut, disesuaikan dengan kemampuan peneliti dan waktu yang tersedia.
3. Penelitian hanya terbatas pada analisis kemampuan *grammar* siswa berdasarkan jawaban teks singkat.
4. Sistem yang dirancang dan dibangun hanya akan dapat digunakan oleh admin, tim manajemen pengembangan, serta pengguna yang telah mendaftar ke dalam sistem.
5. Aplikasi berbasis web hanya menyesuaikan dengan karakteristik masukan dan luaran yang dibutuhkan oleh model kecerdasan buatan yang dikembangkan.

## Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut :

1. Meningkatkan kemudahan dan efisiensi pengujian kemampuan tata bahasa Inggris siswa.
2. Membantu melakukan evaluasi terhadap kualitas kemampuan dan tingkat pemahaman tata bahasa Inggris siswa di sekolah tertentu.
3. Memperkenalkan secara singkat metode-metode yang digunakan dalam penelitian ini kepada pembaca.
4. Membantu memudahkan pengembangan dan penelitian lebih lanjut di kemudian hari.
5. Bagi peneliti, sebagai sarana untuk memperdalam pengetahuan dan kemampuan di bidang teknologi informasi, khususnya pengolahan bahasa alami, dan pengembangan web.
6. Memenuhi tanggung jawab menyelesaikan tugas akhir sebagai prasyarat kelulusan.

## Sistematika Penulisan

Sistematika penulisan dokumen penelitian ini terdiri dari lima bab utama, yaitu sebagai berikut :

### Bab I Pendahuluan

Berisi gambaran umum terkait isi dari dokumen penelitian ini, antara lain, latar belakang, rumusan masalah, tujuan penelitian, batasan masalah penelitian, manfaat penelitian, dan sistematika penulisan dan penyusunan dokumen penelitian.

### Bab II Tinjauan Pustaka

Berisi informasi singkat terkait berbagai karya ilmiah yang dijadikan sumber acuan dalam penyusunan dokumen penelitian ini.

### Bab III Metode Penelitian

Berisi deskripsi rinci terkait berbagai metode yang digunakan dalam melakukan penelitian ini.

### Bab IV Hasil Penelitian dan Pembahasan

Berisi deskripsi rinci serta pembahasan menyeluruh terkait hasil dari penelitian yang telah dilakukan.

### Bab V Kesimpulan dan Saran

Berisi kesimpulan yang merangkum hasil analisis dari pembahasan hasil penelitian yang telah dilakukan

# DAFTAR PUSTAKA

[1] “The most spoken languages worldwide in 2022,” *Statista Research Department*, 2022. https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide (accessed Aug. 04, 2022).

[2] “Ethnologue 21st Edition,” *Ethnologue*, 2018. https://www.ethnologue.com/ethnoblog/gary-simons/welcome-21st-edition (accessed Aug. 04, 2022).

[3] “Internet World Users by Language,” *Internet World Stats*, 2020. https://www.internetworldstats.com/stats7.htm (accessed Aug. 04, 2022).

[4] S. Lobachev, “Top Languages in Global Information Production,” *Partnership: The Canadian Journal of Library and Information Practice and Research*, vol. 3, no. 2, Dec. 2008, doi: 10.21083/partnership.v3i2.826.

[5] R. Murphy, *English Grammar in Use*, 5th ed. Cambridge University Press, 2019.

[6] D. Yan, A. A. Rupp, and P. W. Foltz, *Handbook of Automated Scoring Theory into Practice*. Boca Raton: Taylor & Francis Group, LLC, 2020.

[7] M. D. Shermis and J. C. Burstein, *Automated Essay Scoring: A Cross-Disciplinary Perspective*. New Jersey: Lawrence Erlbaum Associates, Inc., 2003.

[8] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*. New York: Cambridge University Press, 2008.

[9] A. Onan, “Ensemble of Classifiers and Term Weighting Schemes for Sentiment Analysis in Turkish,” *Scientific Research Communications*, vol. 1, no. 1, pp. 1–12, Jul. 2021, doi: 10.52460/src.2021.004.

[10] M. Lan, C.-L. Tan, H.-B. Low, and S.-Y. Sung, “A Comprehensive Comparative Study on Term Weighting Schemes for Text Categorization with Support Vector Machines,” in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, pp. 546–551. doi: 10.1109/IJCNN.2005.1555890.

[11] S. Gupta, “Top 5 Distance Similarity Measures Implementation in Machine Learning,” Sep. 30, 2019. https://medium.com/@gshriya195/top-5-distance-similarity-measures-implementation-in-machine-learning-1f68b9ecb0a3 (accessed Oct. 18, 2022).

[12] M. Harmouch, “17 Types of Similarity and Dissimilarity Measures Used in Data Science,” *Towards Data Science*, Mar. 14, 2021. https://towardsdatascience.com/17-types-of-similarity-and-dissimilarity-measures-used-in-data-science-3eb914d2681 (accessed Oct. 18, 2022).

[13] J. F. Dooley, *Software Development, Design and Coding*. Berkeley, CA: Apress, 2017. doi: 10.1007/978-1-4842-3153-1.

[14] M. van Deurzen, “The anatomy of the modern window manager,” Bachelor thesis, Radboud University, Nijmegen, 2019.