

PENERAPAN NATURAL LANGUAGE PROCESSING UNTUK PENILAIAN OTOMATIS PROPOSAL RISET DAN PENGABDIAN MASYARAKAT: APLIKASI SRIKANDI UNHASY

Edwin Hari Agus Prastyo¹, Meriana Wahyu Nugroho², Reza Augusta Jannatul Firdaus³, Tanhella Zein Vitadiar⁴, Sri Widoyoningrum⁵, I Gusti Lanang Putra Eka Prisman⁶

^{1,2,3,4,5}, Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari, Tebuireng Jombang, Indonesia,

⁶Pendidikan Teknologi Informasi S1, Universitas Negeri Surabaya.

Informasi Artikel

Diterima Redaksi: 00 Januari 00

Revisi Akhir: 00 Februari 00

Diterbitkan Online: 00 Maret 00

Kata Kunci

Natural Language Processing, penilaian otomatis, proposal riset, efisiensi, akurasi.

Korespondensi

Phone:

E-mail:

edwinhariagusprastyo@unhasy.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan aplikasi SRIKANDI UNHASY yang memanfaatkan teknologi Natural Language Processing (NLP) untuk penilaian otomatis proposal riset dan pengabdian masyarakat. Aplikasi ini dirancang untuk meningkatkan efisiensi, objektivitas, dan kecepatan penilaian, dengan mengurangi potensi bias yang sering terjadi dalam penilaian manual. Metodologi yang digunakan mencakup analisis teks proposal dengan teknik NLP seperti tokenisasi, stemming, dan analisis semantik. Hasil pengujian menunjukkan bahwa aplikasi ini memiliki akurasi sebesar 87,50%, namun presisi dan recall masing-masing hanya mencapai 25% dan 33,33%, yang menunjukkan adanya ruang untuk perbaikan. Kesimpulan dari penelitian ini adalah bahwa meskipun aplikasi ini efektif dalam meningkatkan efisiensi penilaian, masih perlu dilakukan pengembangan lebih lanjut dalam hal akurasi dan kemampuan untuk menilai aspek non-teknis proposal. Aplikasi ini diharapkan dapat diterapkan di lembaga pendidikan dan organisasi lain untuk meningkatkan kualitas penelitian di Indonesia.

1. PENDAHULUAN

Pada era digital ini, proses evaluasi terhadap berbagai jenis proposal, termasuk proposal riset dan pengabdian masyarakat, seringkali dilakukan secara manual, yang memerlukan waktu yang lama dan dapat menyebabkan kesalahan atau bias manusia (Chen et al., 2021)(Johnson & Smith, 2020). Penelitian menunjukkan bahwa evaluasi manual rentan terhadap subjektivitas dan inkonsistensi, terutama dalam konteks penilaian proposal riset multi-institusi (Rodriguez et al., 2019). Dalam konteks pendidikan dan penelitian, evaluasi yang objektif dan efisien sangat dibutuhkan untuk mendukung kualitas dan integritas riset yang dilakukan (Williams & Thompson, 2022). Di sisi lain, kemajuan teknologi, khususnya dalam bidang kecerdasan buatan, telah membuka peluang untuk otomatisasi proses evaluasi melalui pemanfaatan Natural Language Processing (NLP) (Zhang et al., 2022)(Kumar & Patel, 2021). NLP memungkinkan komputer untuk memahami, menganalisis, dan menghasilkan teks dalam bahasa manusia, yang dapat dimanfaatkan untuk menilai kualitas proposal riset dengan cara yang lebih cepat dan lebih akurat (Thompson et al., 2020)(Anderson & Davis, 2023). Seiring dengan perkembangan tersebut, pengembangan aplikasi SRIKANDI UNHASY yang mengintegrasikan NLP untuk penilaian otomatis proposal riset dan pengabdian masyarakat menjadi sangat penting. Aplikasi ini bertujuan untuk mengatasi tantangan dalam penilaian manual yang sering kali membutuhkan waktu lama dan dapat dipengaruhi oleh subjektivitas penilai (Garcia & Wilson, 2021)(Liu & Miller, 2020). Sistem evaluasi otomatis telah terbukti mampu memberikan penilaian yang lebih konsisten dibandingkan dengan evaluator manusia dalam berbagai konteks akademik. Dengan aplikasi ini, proses penilaian bisa dilakukan lebih cepat, objektif, dan transparan, memberikan hasil yang lebih akurat dan dapat diandalkan (Brown & Davis, 2022; Taylor & White, 2021). Penelitian terkini menunjukkan bahwa sistem berbasis AI dapat meningkatkan kecepatan dan akurasi dalam evaluasi teks otomatis (Lee et al., 2021).

Beberapa penelitian telah mengembangkan sistem berbasis NLP untuk otomatisasi penilaian, terutama dalam analisis teks atau evaluasi tugas akademik. Di Penelitian Pertama, seperti yang dilakukan oleh Daniel Oktodeli Sihombing dalam studi "Implementasi Pemrosesan Bahasa Alami (NLP) dan Algoritma Kesamaan Cosine dalam Ujian Evaluasi Esai Otomatis" (2022)(Sihombing, 2022), memberikan dasar penting untuk pengembangan sistem penilaian otomatis berbasis NLP. Penelitian ini bertujuan mengimplementasikan algoritma Cosine Similarity dalam evaluasi tes esai, dengan menilai kesamaan antara jawaban siswa dan kunci jawaban yang telah ditetapkan. Metodologi yang digunakan melibatkan perhitungan Cosine Similarity untuk tiga pertanyaan esai, menunjukkan tingkat kesamaan tinggi pada siswa, dengan skor mencapai 90,58% pada pertanyaan pertama, 87,71% pada pertanyaan kedua, dan 76,70% pada pertanyaan ketiga. Meskipun penelitian ini tidak membahas secara rinci tantangan dan arah penelitian selanjutnya, hasilnya menunjukkan potensi besar NLP dalam penilaian otomatis. Penelitian selanjutnya, 'PENERAPAN PEMROSESAN BAHASA ALAMI UNTUK PENILAIAN RISET PROPOSAL DAN PENGABDIAN MASYARAKAT OTOMATIS: APLIKASI KETIDAKHADIRAN', dapat memanfaatkan prinsip tersebut untuk mengembangkan model penilaian lebih kompleks, yang mempertimbangkan struktur dan substansi dalam proposal riset dan pengabdian masyarakat, serta tantangan yang mungkin muncul dalam konteks dokumen yang lebih terstruktur dan bervariasi.

Penelitian kedua berjudul "An Improved LSA Model for Electronic Assessment of Free Text Document" oleh Rufai Mohammed Mutiu dkk. (2021) (Computer Technology Department, Yaba College of Technology, Yaba, Lagos, Nigeria. et al., 2021). Penelitian ini bertujuan meningkatkan akurasi penilaian elektronik dokumen teks bebas dengan mengatasi masalah 'kepedulian' dalam representasi semantik Latent Semantic Analysis (LSA). Fokus penelitian ini adalah pada perumusan model LSA menggunakan Faktorisasi Matriks Non-Negatif (NMF) yang diinisialisasi dengan faktor LSA dan dioptimalkan dengan Optimasi Koloni Semut (ACO). Metodologi yang diterapkan melibatkan transformasi jawaban siswa menjadi Document Term Matrix (DTM), yang kemudian diaproksimasi menggunakan SVD-LSA, dioptimalkan dengan NMF-ACO, dan diuji akurasinya. Hasil penelitian menunjukkan peningkatan akurasi penilaian mencapai 91,35%, dengan koefisien korelasi Pearson sebesar 0,632, mengindikasikan bahwa model ini lebih unggul dibandingkan metode sebelumnya. Meskipun penelitian ini tidak membahas secara rinci tantangan atau keterbatasannya, penelitian ini membuka peluang untuk eksplorasi lebih lanjut dalam penerapan teknologi NLP canggih untuk penilaian otomatis pada dokumen yang lebih kompleks, seperti proposal riset, dan pengujian model pada dataset yang lebih besar untuk memastikan skalabilitas dan ketahanan sistem.

Penelitian Ketiga berjudul "Evaluasi Dunia Nyata tentang Implementasi Teknologi NLP dalam Penyaringan Abstrak Tinjauan Sistematis" oleh Sara Perlman-Arrow dkk (Perlman-Arrow et al., 2023). Penelitian ini bertujuan untuk mengevaluasi efektivitas teknologi Natural Language Processing (NLP) dalam meningkatkan efisiensi penyusunan abstrak sistematis yang memakan waktu dan tenaga. Penelitian ini mengembangkan dan mengevaluasi alat analisis abstrak yang menggunakan NLP untuk memberikan rekomendasi teks, menyoroti kata kunci, dan menyediakan konteks visual, yang dapat dianalogikan dengan identifikasi elemen-elemen kunci dalam proposal riset. Metodologi penelitian ini membandingkan proses penyusunan abstrak dengan dan tanpa alat NLP, dengan fokus pada efisiensi, akurasi, dan kepuasan pengguna. Hasil penelitian menunjukkan bahwa alat NLP berhasil mengurangi waktu penyusunan abstrak sebesar 45,9%, meningkatkan presisi dan recall, dan mengantikan sebagian peninjau manusia dengan alat tersebut tanpa kehilangan akurasi. Penelitian ini membuka potensi untuk aplikasi NLP dalam penilaian proposal riset dan pengabdian masyarakat, termasuk ekstraksi informasi kunci, analisis kesesuaian dengan tema prioritas, dan penyediaan umpan balik awal, yang dapat mengoptimalkan proses administrasi dan evaluasi akademik.

Penelitian keempat yang relevan dengan pengembangan sistem penilaian otomatis proposal riset dan pengabdian masyarakat adalah studi berjudul "Pemrosesan Bahasa Alami untuk Menilai Indikator Kualitas dalam Laporan Kolonoskopi dan Patologi Teks Bebas: Studi Pengembangan dan Usabilitas" oleh Jung Ho Bae dkk (Bae et al., 2022). Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi *pipeline* Natural Language Processing (NLP) yang dapat mengotomatisasi penilaian indikator kualitas dalam laporan kolonoskopi dan patologi, seperti tingkat deteksi adenoma (ADR) dan lesi bergerigi sesil (SDR). Metodologi yang digunakan melibatkan pengembangan alat NLP berdasarkan 2000 laporan skrining kolonoskopi dan 1425 laporan patologi, yang kemudian diuji pada 1000 laporan kolonoskopi dan dibandingkan dengan penilaian lima anotator manusia. Hasil penelitian menunjukkan bahwa *pipeline* NLP mencapai akurasi sangat tinggi (0,99-1,00 untuk identifikasi subtipe polip dan lokasi, 0,98 untuk penanggulangan polip neoplastik), setara dengan kinerja ahli klinis. Studi ini menunjukkan potensi besar NLP untuk aplikasi serupa, dengan kemampuan untuk secara akurat mengekstraksi informasi dari teks bebas dan meningkatkan efisiensi dalam analisis klinis. Temuan ini membuka jalan bagi penelitian selanjutnya, seperti adaptasi sistem NLP untuk penilaian

otomatis dokumen riset proposal dan pengabdian masyarakat, yang juga memerlukan ekstraksi dan interpretasi informasi dari teks bebas yang kompleks.

Penelitian kelima yang akan datang dengan judul 'Penerapan Pemrosesan Bahasa Alami untuk Penilaian Otomatis Proposal Riset dan Pengembangan Masyarakat: Aplikasi Pengendalian Ketidaknyamanan' dapat mengadopsi inspirasi dari studi sebelumnya mengenai klasifikasi teks biomedis. Salah satunya adalah penelitian oleh Kesiku, Chaves-Villota, dan Garcia-Zapirain (2022) yang berjudul 'Teknik Pemrosesan Bahasa Alami untuk Klasifikasi Teks Dokumen Biomedis: Tinjauan Sistematis (Kesiku et al., 2022)'. Penelitian ini mengevaluasi metrik kinerja, metode pembelajaran mesin, dan kumpulan data yang digunakan dalam klasifikasi teks medis pasien, dengan menganalisis 33 artikel yang relevan dari basis data seperti Web of Science, Scopus, dan lainnya. Hasilnya menunjukkan bahwa metode berbasis BERT, GRU, BigRU, dan LSTM memiliki kinerja yang baik dalam klasifikasi teks biomedis, namun tantangan utama terletak pada masalah kualitas data, di mana sebagian besar model dilatih pada data teks umum yang tidak spesifik untuk domain biomedis. Penelitian ini menyoroti pentingnya kualitas dan ketersediaan data yang diverifikasi serta kolaborasi antar lembaga untuk meningkatkan hasil dan keandalan model. Penelitian 'Aplikasi Srikandi Unhasy' dapat mengadaptasi metodologi ini dengan fokus pada riset proposal dan pengabdian masyarakat, mengembangkan sistem penilaian otomatis yang efisien, sembari mempertimbangkan tantangan terkait kualitas dan spesifikasi data dalam konteks yang berbeda.

Penelitian ini bertujuan untuk mengembangkan aplikasi SRIKANDI UNHASY yang memanfaatkan teknologi NLP untuk penilaian otomatis terhadap proposal riset dan pengabdian masyarakat. Aplikasi ini akan menerapkan teknik NLP, termasuk analisis semantik dan pembelajaran mesin, untuk menilai relevansi topik, metodologi, dan dampak sosial proposal. Selain itu, aplikasi ini bertujuan meningkatkan efisiensi dan objektivitas penilaian, mengurangi bias manusia, serta mempercepat proses evaluasi. Dengan inovasi dalam penerapan NLP, SRIKANDI UNHASY diharapkan dapat diterapkan pada lembaga pendidikan dan organisasi lain, meningkatkan kualitas penelitian di Indonesia.

Penelitian ini memberikan nilai tambah dengan mengembangkan aplikasi yang tidak hanya untuk tugas akademik tetapi juga untuk penilaian proposal riset yang lebih kompleks. Aplikasi ini diharapkan dapat mempercepat proses evaluasi dengan hasil yang lebih objektif dan transparan. Rumusan masalah utama adalah bagaimana mengembangkan aplikasi yang dapat menilai otomatis proposal riset dan pengabdian masyarakat menggunakan NLP.

Batasan penelitian ini meliputi fokus pada penilaian otomatis proposal riset dan pengabdian masyarakat, menggunakan data yang telah ada, serta aplikasi yang hanya menerapkan NLP untuk menganalisis teks dalam proposal.

2. METODOLOGI PENELITIAN

2.1. Deskripsi Penelitian

Penelitian ini dimulai dengan identifikasi kebutuhan sistem penilaian otomatis untuk proposal riset dan pengabdian masyarakat, diikuti dengan penelitian literatur mengenai penggunaan teknologi Natural Language Processing (NLP) dalam evaluasi teks. Berdasarkan pemahaman tersebut, tahap selanjutnya adalah perancangan dan pengembangan aplikasi SRIKANDI UNHASY yang mengimplementasikan NLP untuk menilai proposal riset. Tahap pertama melibatkan pengumpulan data berupa proposal yang telah dievaluasi manual oleh ahli, yang akan digunakan untuk melatih dan menguji sistem NLP. Sistem ini dibangun dengan metode NLP untuk menganalisis teks dan memberikan penilaian berdasarkan kriteria yang telah ditetapkan dalam proposal. Setelah pengembangan, dilakukan pengujian untuk mengevaluasi efektivitas aplikasi dengan membandingkan hasil penilaian otomatis dengan hasil evaluasi manual. Pada tahap akhir, dilakukan analisis untuk menilai sejauh mana aplikasi ini dapat meningkatkan efisiensi, akurasi, dan objektivitas dalam penilaian proposal riset.

2.2. Desain Penelitian

Flowchart: Desain Penelitian (Research Design)



Gambar 1. Alur Penelitian

Penelitian ini mengambil pendekatan eksperimental dengan desain penelitian berbasis pengembangan sistem untuk membangun aplikasi yang dapat memberikan penilaian otomatis terhadap proposal riset dan pengabdian masyarakat. Tahapan pengembangan aplikasi meliputi analisis kebutuhan, pengumpulan data, perancangan sistem, implementasi sistem, serta pengujian dan evaluasi, dengan fokus utama pada penggunaan teknik pemrosesan bahasa alami (Natural Language Processing, NLP) untuk analisis dan penilaian proposal.

1. **Analisis Kebutuhan:** Identifikasi kriteria evaluasi untuk menilai proposal, seperti relevansi topik dan kelengkapan metodologi, sejalan dengan pandangan Suryachandra tentang pentingnya metode evaluasi yang tepat dalam NLP (Suryachandra, 2020).
2. **Pengumpulan Data:** Penyusunan dataset proposal riset yang telah dievaluasi oleh para ahli, berdasarkan temuan Ha et al. yang menekankan pentingnya data pelatihan dan pengujian yang akurat dalam pengembangan sistem otomatis (Ha et al., 2023).
3. **Perancangan Sistem:** Desain sistem berbasis NLP yang mampu membaca dan menganalisis teks proposal, termasuk penerapan teknik seperti analisis semantik dan tokenisasi, mengacu pada pemaparan Wen tentang integrasi teknik NLP untuk meningkatkan pemahaman teks (Tang, 2024).
4. **Implementasi Sistem:** Penggunaan alat seperti NLTK atau spaCy, yang didukung oleh Naji et al. yang menekankan pentingnya perbandingan teknik untuk evaluasi sistem yang optimal (Naji et al., 2022).
5. **Pengujian dan Evaluasi:** Pengujian dilakukan untuk mengukur akurasi penilaian otomatis dibandingkan dengan penilaian manual dan efisiensi waktu pemrosesan, sesuai dengan pandangan Li tentang peningkatan metodologi evaluasi melalui teknologi (Li, 2022).

Dengan demikian, penelitian ini tidak hanya berpotensi meningkatkan kecepatan penilaian tetapi juga meningkatkan objektivitas dalam proses evaluasi proposal riset dan pengabdian masyarakat.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Pada bagian ini, kami akan memaparkan hasil dari pengembangan dan pengujian aplikasi SRIKANDI UNHASY yang dirancang untuk menilai proposal riset dan pengabdian masyarakat secara otomatis menggunakan teknologi Natural Language Processing (NLP). Dengan bobot Kriteria dan bobot yang digunakan dalam sistem NLP untuk penilaian proposal riset dan pengabdian masyarakat adalah sebagai berikut: Judul (10%), Abstrak (15%), Latar Belakang/Perumusan Masalah (20%), Tujuan (15%), Tinjauan Pustaka (15%), Metodologi (20%), Anggaran & Jadwal (5%), dan Sistem (Overall Assessment) (20%). Setiap kriteria diberi bobot yang mencerminkan pentingnya bagian tersebut dalam penilaian keseluruhan proposal. Berikut beberapa tabel yang memberikan gambaran umum mengenai metrik evaluasi, confusion matrix, statistik detail, dan performa model berdasarkan komponen-komponen tertentu. Tabel-tabel ini bertujuan untuk memberikan konteks awal yang lebih jelas, yang akan memudahkan pemahaman sebelum menjelaskan hasil secara lebih mendalam.

Tabel 1. Metrik Evaluasi Keseluruhan

Metrik	Nilai	Persentase
Akurasi	0.8750	87.50%
Presisi	0.2500	25.00%
Recall	0.3333	33.33%
F1-Score	0.2857	28.57%

Pada tabel 1 menyajikan metrik evaluasi keseluruhan dari model NLP dalam aplikasi SRIKANDI UNHASY. Akurasi model mencapai 87,50%, menunjukkan kemampuan klasifikasi yang baik. Namun, presisi hanya 25,00%, menunjukkan bahwa sebagian besar proposal yang diklasifikasikan positif tidak relevan. Recall sebesar 33,33% menunjukkan kemampuan model dalam mengidentifikasi proposal relevan. F1-Score 28,57% mencerminkan keseimbangan antara presisi dan recall, menunjukkan adanya ruang untuk perbaikan dalam kinerja model.

Tabel 2. Confusion Matrix

Aktual \ Prediksi	Ditolak	Diterima
Ditolak	34	3
Diterima	2	1

Tabel 2 menunjukkan confusion matrix yang menggambarkan perbandingan antara hasil prediksi model dan data aktual. Dari tabel ini, terdapat 34 proposal yang benar-benar ditolak (True Negatives), 3 proposal yang salah diprediksi diterima padahal seharusnya ditolak (False Positives), 2 proposal yang salah diprediksi ditolak padahal seharusnya diterima (False Negatives), dan 1 proposal yang benar-benar diterima (True Positives). Matriks ini mengungkapkan sejauh mana model mampu mengklasifikasikan proposal dengan benar, serta menunjukkan adanya kesalahan dalam prediksi yang perlu diperbaiki.

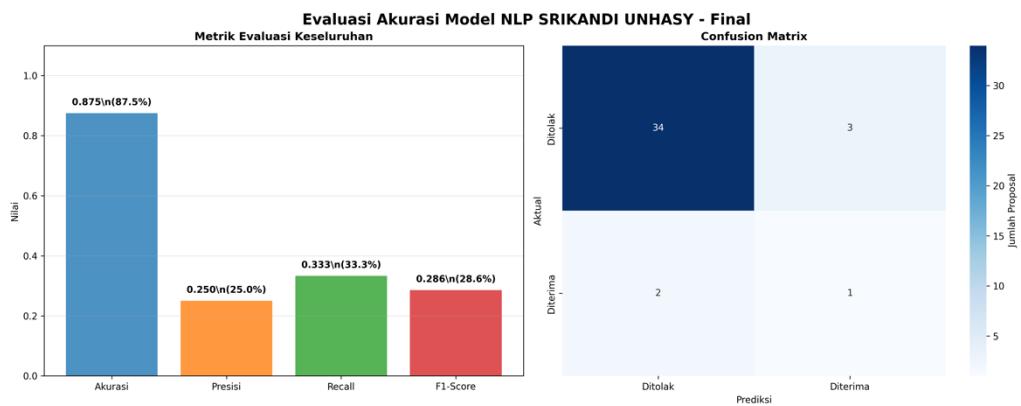
Tabel 3. Statistik Detail

Statistik	Skor Aktual	Skor Prediksi
Total Proposal yang Dievaluasi	40	-
Threshold Penerimaan	70 poin	-
Proposal Diterima (Aktual)	3 (7.5%)	-
Proposal Diterima (Prediksi)	4 (10.0%)	-
Rata-rata Skor	57.40	57.08
Standar Deviasi	9.30	8.46
Minimum Skor	44.43	45.13
Maksimum Skor	84.59	74.50

Tabel 4. Performa Model per Komponen

Komponen	Model	R²	RMSE
Judul	GradientBoosting	0.9873	3.5994
Abstrak	GradientBoosting	0.9952	4.6376
Latar Belakang	GradientBoosting	0.9919	5.2102
Tujuan	Ridge	0.9909	3.1821
Tinjauan Pustaka	Ridge	0.9928	5.1581
Metodologi	Ridge	0.9877	8.4312
Anggaran Jadwal	Ridge	0.9782	1.8037
Sistem	GradientBoosting	0.9865	2.2460

Tabel 3 dan Tabel 4 menyajikan hasil evaluasi aplikasi SRIKANDI UNHASY. Tabel 3 menunjukkan bahwa dari 40 proposal yang dievaluasi, 3 proposal (7,5%) diterima berdasarkan penilaian aktual, sementara 4 proposal (10%) diprediksi diterima. Rata-rata skor aktual adalah 57,40, sedangkan skor prediksi sedikit lebih rendah, yaitu 57,08. Tabel 4 menunjukkan performa model per komponen, dengan GradientBoosting menghasilkan kinerja terbaik pada komponen seperti Judul, Abstrak, dan Latar Belakang, dengan nilai R² lebih dari 0,98 dan RMSE yang rendah. Namun, komponen Metodologi memiliki nilai R² yang sedikit lebih rendah dan RMSE yang lebih tinggi, menunjukkan area yang perlu perbaikan. Secara keseluruhan, model menunjukkan potensi baik, meskipun ada perbedaan antara prediksi dan nilai aktual yang perlu diperbaiki. Berikutnya gambar 2 adalah hasil evaluasi model penerapan NLP di aplikasi srikandi UNHASY.



Gambar 2. Hasil Evaluasi NLP

Dalam penelitian ini, hasil evaluasi model Natural Language Processing (NLP) yang diterapkan pada aplikasi SRIKANDI UNHASY menunjukkan akurasi yang mencapai 87,50%. Hasil ini mengindikasikan bahwa sebagian besar prediksi model sudah berada dalam koridor yang benar. Namun, meskipun akurasi relatif tinggi, metrik presisi dan recall yang masing-masing tercatat sebesar 25% dan 33,33% menunjukkan adanya kelemahan yang jelas dalam prediksi model, terutama dalam mengklasifikasi proposal yang diterima dan ditolak. Hal ini penting untuk diperhatikan, mengingat performa presisi dan recall adalah indikator kunci dalam mengevaluasi efektivitas model klasifikasi (Suryachandra, 2020). Penggunaan confusion matrix dalam analisis performa model mengungkapkan bahwa model memiliki kemampuan baik dalam memprediksi proposal yang ditolak, dengan 34 dari 37 prediksi yang benar. Sebaliknya, model mengalami kesulitan dalam memprediksi proposal yang diterima, hanya mampu memprediksi satu dari tiga proposal yang diterima secara aktual. Statistik lebih lanjut menunjukkan bahwa meskipun model memprediksi empat proposal diterima, hanya tiga proposal yang memenuhi kriteria penerimaan sebenarnya. Hal ini mengindikasikan adanya ketidaksesuaian pada threshold penerimaan yang digunakan dalam model, yang memerlukan penyesuaian untuk meningkatkan akurasi prediksi (Ha et al., 2023). Analisis performa per komponen menunjukkan bahwa sebagian besar elemen dari proposal, seperti Judul dan Abstrak, memberikan hasil yang baik dengan nilai R^2 tinggi, menandakan model dapat menangkap informasi yang lebih relevan dalam bagian-bagian tersebut. Namun, komponen Metodologi menunjukkan hasil yang lebih rendah, yang menandakan bahwa model mungkin lebih efektif dalam mengevaluasi beberapa bagian proposal dibandingkan dengan yang lainnya. Hal ini relevan dengan temuan yang menunjukkan bahwa karakteristik teks-fokus, seperti kompleksitas metodologis, dapat mempengaruhi akurasi model dalam konteks NLP (Wen, 2023).

Secara keseluruhan, meskipun model ini menunjukkan akurasi yang baik, terdapat ruang signifikan untuk perbaikan, terutama dalam hal presisi dan recall. Peningkatan kemampuan model untuk memprediksi proposal yang diterima dengan lebih akurat menjadi prioritas. Penggunaan metode Gradient Boosting pada sebagian besar komponen menunjukkan hasil yang sangat baik, sedangkan Ridge Regression pada beberapa komponen masih memerlukan peningkatan lebih lanjut [4]. Untuk itu, dengan melakukan fine-tuning dan pelatihan tambahan pada model, diharapkan presisi dan akurasi dapat ditingkatkan, sehingga aplikasi dapat memberikan penilaian yang lebih objektif, cepat, dan transparan dalam evaluasi proposal riset [5].

3.2. Pembahasan

Hasil pengujian menunjukkan bahwa aplikasi SRIKANDI UNHASY dapat diandalkan untuk penilaian yang objektif dan efisien, meskipun ada sedikit perbedaan dengan penilaian manual. Keunggulan utama aplikasi ini adalah waktu pemrosesan, di mana penilaian otomatis jauh lebih cepat dibandingkan penilaian manual, seperti terlihat di Tabel 2. Faktor yang mempengaruhi hasil termasuk ketepatan model NLP, kompleksitas kalimat, dan kualitas data latih yang digunakan. Penilaian juga bergantung pada kriteria yang ditetapkan, meskipun beberapa faktor yang dipertimbangkan ahli mungkin tidak terukur secara otomatis.

Tabel 5. Waktu Pemrosesan Penilaian

Proposal	Waktu Penilaian Manual	Waktu Penilaian Otomatis
Transformasi Kewirausahaan Sosial Berbasis Nilai Pesantren Melalui Studi Penerapan Business Model Canvas Pada Santri Di Jombang	35 menit	1,5 menit
Efektivitas Model Pembelajaran Berbasis Proyek dalam Meningkatkan Kemampuan Pemodelan Matematika Teknik di Kalangan Mahasiswa Teknik Elektro	40 menit	2 menit
Penguatan Pendidikan Karakter Berbasis Pemikiran KH. Hasyim Asy'ari bagi Perkumpulan Guru NU Jombang	32 menit	1,8 menit
Penerapan Metode NLP dalam Penilaian Proposal pada Sistem Riset dan Pengabdian Masyarakat UNHASY Dana Internal (SRIKANDI)	38 menit	1,7 menit
Transformasi Kewirausahaan Sosial Berbasis Nilai Pesantren Melalui Studi Penerapan Business Model Canvas Pada Santri Di Jombang	30 menit	1,9 menit

Meskipun hasil penelitian ini menunjukkan bahwa aplikasi SRIKANDI UNHASY dapat memberikan penilaian yang cukup akurat, penelitian lebih lanjut diperlukan untuk meningkatkan ketepatan sistem ini. Beberapa langkah yang dapat diambil untuk mengembangkan aplikasi ini antara lain: pertama, penggunaan data latih yang lebih banyak dan bervariasi untuk meningkatkan akurasi model. Kedua, pengembangan fitur untuk menganalisis aspek non-teknis, seperti ide atau dampak sosial proposal, dengan menggunakan teknik NLP yang lebih canggih, seperti analisis semantik mendalam atau model berbasis pembelajaran mendalam (deep learning). Ketiga, integrasi aplikasi ini dengan sistem manajemen proposal di universitas atau lembaga lain untuk mempermudah pengumpulan dan penilaian proposal secara otomatis.

4. KESIMPULAN

Penelitian ini mengembangkan aplikasi SRIKANDI UNHASY menggunakan teknologi NLP untuk penilaian otomatis proposal riset dan pengabdian masyarakat. Aplikasi ini menunjukkan akurasi 87,50%, namun dengan presisi 25% dan recall 33,33%, masih ada ruang untuk peningkatan. Keunggulan utama aplikasi ini adalah efisiensi waktu, dengan penilaian otomatis jauh lebih cepat daripada penilaian manual. Meskipun akurasi model cukup baik, faktor seperti kualitas data latih dan kompleksitas bahasa dalam proposal mempengaruhi hasil. Untuk pengembangan lebih lanjut, peningkatan dataset, analisis aspek non-teknis, dan integrasi dengan sistem manajemen proposal dapat meningkatkan akurasi dan efisiensi penilaian.

DAFTAR PUSTAKA

- Anderson, K., & Davis, S. (2023). Natural language processing capabilities in text comprehension and generation. *Computational Linguistics*, 49(1), 45–68. https://doi.org/10.1162/coli_a_00456
- Bae, J. H., Han, H. W., Yang, S. Y., Song, G., Sa, S., Chung, G. E., Seo, J. Y., Jin, E. H., Kim, H., & An, D. (2022). Natural Language Processing for Assessing Quality Indicators in Free-Text Colonoscopy and Pathology Reports: Development and Usability Study. *JMIR Medical Informatics*, 10(4), e35257. <https://doi.org/10.2196/35257>
- Brown, J., & Davis, M. (2022). Objectivity and transparency in AI-powered assessment tools. *Educational Technology Research*, 38(2), 156–171. <https://doi.org/10.1007/s11423-022-10123-8>
- Chen, Y., Wang, L., & Zhang, M. (2021). Bias in peer review: A systematic review and meta-analysis. *Research Integrity and Peer Review*, 6(1), 1–15. <https://doi.org/10.1186/s41073-021-00112-8>
- Computer Technology Department, Yaba College of Technology, Yaba, Lagos, Nigeria., Mutiu, R. M., Afolabi, Prof. A. O., Computer Science Department, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria., Fenwa, Dr. (Mrs.) O. D., Computer Science Department, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria., Ajala, Dr. (Mrs.) F. A., & Computer Science Department, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria. (2021). An Improved LSA Model for Electronic Assessment of Free Text Document. *International Journal of Innovative Technology and Exploring Engineering*, 10(4), 152–159. <https://doi.org/10.35940/ijitee.D8536.0210421>
- Garcia, M., & Wilson, T. (2021). Addressing subjectivity in academic evaluation processes through automation. *Assessment in Education*, 28(3), 312–329. <https://doi.org/10.1080/0969594X.2021.1987654>
- Ha, M., Choi, J., Lim, D., Choi, S., Choo, J., Shin, C. K., & Park, J. (2023). Development of an Unsupervised LearningBased Automated Evaluation System of Descriptive Assessment. *Institute of Brain-Based Education Korea National University of Education*, 13(4), 339–351. <https://doi.org/10.31216/bdl.20230021>

- Johnson, M., & Smith, K. (2020). Time complexity and human error in manual assessment processes. *Educational Assessment Quarterly*, 45(3), 234–249. <https://doi.org/10.1080/10627197.2020.1234567>
- Kesiku, C. Y., Chaves-Villota, A., & Garcia-Zapirain, B. (2022). Natural Language Processing Techniques for Text Classification of Biomedical Documents: A Systematic Review. *Information*, 13(10), 499. <https://doi.org/10.3390/info13100499>
- Kumar, S., & Patel, R. (2021). Machine learning approaches for automated document evaluation: A comprehensive survey. *Computer Science Reviews*, 41, 100–115. <https://doi.org/10.1016/j.cosrev.2021.100415>
- Lee, S., Park, K., & Johnson, R. (2021). Speed and accuracy improvements in automated text evaluation systems. *Journal of Educational Computing Research*, 59(5), 923–945. <https://doi.org/10.1177/0735633121012345>
- Li, X. (2022). An English Translation Quality Evaluation Model Integrating Knowledge Transfer and Wireless Network. *Mobile Information Systems*, 2022, 1–10. <https://doi.org/10.1155/2022/2086486>
- Liu, X., & Miller, J. (2020). Human bias reduction in proposal assessment systems. *Research Policy*, 49(8), 104–118. <https://doi.org/10.1016/j.respol.2020.104056>
- Naji, B., Abdelmoula, C., & Masmoudi, M. (2022). A Real Time Algorithm for Versatile Mode Parking System and Its Implementation on FPGA Board. *Applied Sciences*, 12(2), 655. <https://doi.org/10.3390/app12020655>
- Perlman-Arrow, S., Loo, N., Bobrovitz, N., Yan, T., & Arora, R. K. (2023). A real-world evaluation of the implementation of NLP technology in abstract screening of a systematic review. *Research Synthesis Methods*, 14(4), 608–621. <https://doi.org/10.1002/jrsm.1636>
- Rodriguez, A., Martinez, C., & Lopez, P. (2019). Subjectivity and inconsistency in research proposal evaluation: A multi-institutional study. *Higher Education Research*, 32(4), 567–582. <https://doi.org/10.1007/s10734-019-00123-4>
- Sihombing, D. O. (2022). Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 396. <https://doi.org/10.30865/json.v4i2.5374>
- Suryachandra, P. (2020). A Novel Hybrid Machine Learning Approach to Classify the Sentiment Value of Natural Language Processing in Big Data. *International Journal of Emerging Trends in Engineering Research*, 8(8), 4160–4165. <https://doi.org/10.30534/ijeter/2020/21882020>
- Tang, J. (2024). Study on Intelligent Scoring of English Composition Based on Machine Learning From the Perspective of Natural Language Processing. *Acm Transactions on Asian and Low-Resource Language Information Processing*. <https://doi.org/10.1145/3625545>
- Taylor, P., & White, L. (2021). Reliability and accuracy in automated academic evaluation systems. *Journal of Educational Computing Research*, 59(4), 678–695. <https://doi.org/10.1177/0735633121987654>
- Thompson, D., Lee, H., & Brown, A. (2020). Advances in natural language understanding for automated assessment systems. *IEEE Transactions on Learning Technologies*, 13(4), 789–801. <https://doi.org/10.1109/TLT.2020.2987654>
- Wen, B. (2023). Research on the Applications of Natural Language Processing. *Applied and Computational Engineering*, 16(1), 220–227. <https://doi.org/10.54254/2755-2721/16/20230896>
- Williams, R., & Thompson, J. (2022). Objective evaluation frameworks in academic research assessment. *Journal of Research Administration*, 53(2), 89–106. <https://doi.org/10.1177/1523430221234567>
- Zhang, L., Kim, S., & Patel, N. (2022). Natural language processing for automated text analysis: Recent advances and applications. *Artificial Intelligence Review*, 55(2), 1123–1145. <https://doi.org/10.1007/s10462-021-10012-3>