

Otomasi Klasifikasi Taksonomi Pengetahuan pada Knowledge Management System (KMS) Menggunakan Algoritma Natural Language Processing (NLP)

Fajar Muttaqi^{1*}, Helmi Yulianti Fauziah²

^{1*} Universitas Utpadaka Swastika, Tangerang, Indonesia

²Universitas Cendekia Abditama, Tangerang, Indonesia

*fajar.muttaqi@utpas.ac.id

ABSTRACT

Managing knowledge assets within a Knowledge Management System (KMS) often faces constraints during the manual document classification stage. Manual processes are not only time-consuming but also prone to subjectivity and inconsistent taxonomic labeling. This research aims to automate knowledge taxonomy classification in KMS by integrating Natural Language Processing (NLP) techniques. The methodology includes text preprocessing stages (such as case folding, tokenizing, filtering, and Sastrawi stemming) and feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF). Test results on 100 Indonesian text documents indicate that the application of AI through classification algorithms achieves an average accuracy rate of 85%. Furthermore, efficiency analysis shows a significant reduction in document processing time, from an average of 120-300 seconds to less than 0.5 seconds per document. The stemming stage proved crucial as it increased system accuracy by 23% through a 72.2% reduction in word features. This study concludes that NLP integration effectively enhances the scalability and accuracy of organizational knowledge management while minimizing administrative workload.

Keywords: Knowledge Management System, Natural Language Processing, Taxonomy, Artificial Intelligence, TF-IDF

ABSTRAK

Pengelolaan aset pengetahuan dalam *Knowledge Management System* (KMS) sering kali menghadapi kendala pada tahap klasifikasi dokumen yang dilakukan secara manual. Proses manual tidak hanya memakan waktu lama, tetapi juga rentan terhadap subjektivitas dan ketidakkonsistenan pelabelan taksonomi. Penelitian ini bertujuan untuk mengotomasi klasifikasi taksonomi pengetahuan pada KMS dengan mengintegrasikan teknik *Natural Language Processing* (NLP). Metodologi yang digunakan meliputi tahapan *preprocessing* teks (seperti *case folding*, *tokenizing*, *filtering*, dan *stemming* Sastrawi) serta ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil pengujian terhadap 100 dokumen teks Bahasa Indonesia menunjukkan bahwa penerapan AI melalui algoritma klasifikasi mampu mencapai tingkat akurasi rata-rata sebesar 85%. Selain itu, analisis efisiensi menunjukkan reduksi waktu pemrosesan dokumen yang signifikan, dari rata-rata 120-300 detik menjadi kurang dari 0,5 detik per dokumen. Tahap *stemming* terbukti krusial karena meningkatkan akurasi sistem sebesar 23% melalui reduksi fitur kata hingga 72,2%. Simpulan penelitian ini menegaskan bahwa integrasi NLP secara efektif meningkatkan skalabilitas dan akurasi manajemen pengetahuan organisasi, sekaligus meminimalisir beban kerja administratif.

Kata Kunci : Knowledge Management System, Natural Language Processing, Taksonomi, Artificial Intelligence, TF-IDF

INFORMASI ARTIKEL

Submit
4, Januari, 2026

Diterima
10, Mei, 2026

Publish Online
30, Mei, 2026

PENDAHULUAN

Di era transformasi digital yang berkembang pesat saat ini, informasi dan pengetahuan telah menjadi aset strategis utama bagi organisasi, baik di sektor industri maupun institusi pendidikan tinggi [1]. Fenomena ledakan data (*data explosion*) menyebabkan volume informasi digital meningkat secara eksponensial setiap detiknya. Bagi organisasi modern, tantangan utama bukan lagi terletak pada bagaimana mendapatkan informasi, melainkan bagaimana mengelola, menyimpan, dan menemukan kembali pengetahuan tersebut secara efisien [2][3]. Dalam konteks ini, *Knowledge Management System* (KMS) muncul sebagai solusi fundamental untuk memfasilitasi proses identifikasi, penangkapan, retensi, dan pembagian pengetahuan organisasi. KMS yang efektif memungkinkan entitas seperti universitas atau perusahaan untuk mempertahankan kekayaan intelektualnya dan mencegah terjadinya *knowledge loss* saat terjadi pergantian personel [4].

Namun, implementasi KMS tradisional sering kali menghadapi kendala besar dalam hal skalabilitas dan strukturasi data. Salah satu komponen paling kritis dalam KMS adalah taksonomi pengetahuan—sebuah sistem klasifikasi hierarkis yang membantu pengguna dalam mengkategorikan aset pengetahuan agar mudah dicari Kembali [5]. Masalah muncul ketika proses klasifikasi ini dilakukan secara manual oleh manusia. Penumpukan dokumen dalam format teks yang tidak terstruktur, seperti laporan penelitian, panduan operasional, hingga korespondensi internal, menciptakan beban kerja yang sangat berat bagi administrator system [6]. Klasifikasi manual tidak hanya memakan waktu yang lama, tetapi juga sangat rentan terhadap subjektivitas dan kesalahan manusia (*human error*) [7]. Inkonsistensi dalam pelabelan dokumen menyebabkan struktur taksonomi menjadi berantakan, yang pada akhirnya menurunkan tingkat efektivitas pencarian informasi (*information retrieval*) di dalam sistem.

Kesenjangan antara volume data yang besar dan kemampuan manusia untuk mengkategorikannya secara manual menuntut adanya inovasi teknologi. Di sinilah peran kecerdasan buatan atau *Artificial Intelligence* (AI) menjadi sangat krusial, khususnya melalui sub-bidang *Natural Language Processing* (NLP). NLP merupakan teknologi yang memungkinkan mesin untuk memahami, menginterpretasikan, dan memanipulasi bahasa manusia. Dengan menerapkan teknik NLP dalam KMS, organisasi dapat melakukan otomasi pada proses kategorisasi dokumen teks ke dalam kelas-kelas taksonomi yang telah ditentukan sebelumnya. Transformasi dari sistem manajemen pengetahuan yang pasif menjadi sistem cerdas yang mampu "membaca" dan "memahami" konteks dokumen adalah langkah maju dalam meningkatkan maturitas teknologi informasi organisasi.

Penerapan NLP dalam klasifikasi teks melibatkan serangkaian proses komputasi yang kompleks namun sistematis [8]. Tahapan awal seperti *text preprocessing*—yang meliputi *tokenizing* (pemecahan kalimat menjadi kata), *filtering* (penghapusan kata sambung atau *stopword*), dan *stemming* (pencarian kata dasar)—menjadi fondasi penting untuk membersihkan data teks dari kebisingan (*noise*) [9]. Setelah data bersih, algoritma pembelajaran mesin (*machine learning*) seperti *Naive Bayes*, *Support Vector Machine* (SVM), atau *Random Forest* dapat digunakan untuk melatih model berdasarkan data historis yang sudah ada. Model ini kemudian mampu memprediksi kategori taksonomi dari dokumen baru dengan tingkat akurasi yang terukur [10]. Otomasi ini tidak hanya meningkatkan kecepatan pemrosesan data, tetapi juga menjamin standarisasi klasifikasi yang objektif berdasarkan fitur-fitur linguistik yang terkandung dalam teks tersebut [11].

Pentingnya penelitian ini juga didasari oleh kebutuhan praktis di lingkungan akademik dan profesional. Sebagai contoh, di institusi pendidikan tinggi, jumlah publikasi ilmiah, modul ajar, dan laporan pengabdian masyarakat terus bertambah setiap semester. Tanpa sistem klasifikasi otomatis yang handal, aset-aset berharga ini sering kali terkubur dalam repositori yang sulit diakses. Dengan mengintegrasikan NLP ke dalam KMS, proses pengarsipan dapat dilakukan secara *real-time* dan akurat. Penelitian sebelumnya telah menunjukkan bahwa penggunaan algoritma klasifikasi teks mampu meningkatkan efisiensi waktu hingga lebih dari 60% dibandingkan metode manual. Namun, tantangan dalam bahasa Indonesia, seperti penggunaan imbuhan yang kompleks dan bahasa prokem, tetap menjadi ruang penelitian yang menarik untuk dieksplorasi lebih lanjut guna mendapatkan hasil akurasi yang optimal.

Selain efisiensi operasional, implementasi otomasi taksonomi berbasis NLP juga berdampak pada pengambilan keputusan strategis. Ketika pengetahuan terorganisir dengan baik dalam taksonomi yang rapi, organisasi dapat dengan mudah melakukan pemetaan pengetahuan (*knowledge mapping*). Pimpinan organisasi dapat melihat bidang apa yang paling banyak menghasilkan inovasi dan bidang apa yang masih mengalami kekurangan literasi atau aset pengetahuan. Hal ini memberikan keunggulan kompetitif bagi organisasi dalam menghadapi dinamika pasar atau perubahan kebijakan pendidikan yang dinamis. Integrasi AI dalam KMS bukan sekadar tren teknologi, melainkan sebuah kebutuhan mendesak untuk menciptakan ekosistem kerja yang berbasis pengetahuan (*knowledge-based environment*).

Tujuan utama dari penelitian ini adalah untuk merancang dan mengimplementasikan model klasifikasi teks otomatis menggunakan teknik NLP yang disesuaikan dengan struktur taksonomi pada KMS. Penelitian ini diharapkan dapat memberikan kontribusi teoritis bagi perkembangan ilmu sistem informasi, khususnya dalam penggabungan manajemen pengetahuan dengan kecerdasan buatan. Secara praktis, penelitian ini memberikan panduan bagi pengembang sistem dalam membangun fitur otomasi yang dapat mengurangi beban kerja administratif dan meningkatkan pengalaman pengguna dalam berinteraksi dengan KMS. Dengan demikian, luaran dari penelitian ini diharapkan dapat menjadi referensi penting bagi pengembangan sistem informasi cerdas di masa depan, terutama di lingkungan perguruan tinggi swasta seperti Universitas Utpadaka Swastika yang terus berkompetisi dalam peningkatan kualitas tata kelola pengetahuan.

TINJAUAN PUSTAKA

1. Knowledge Management System (KMS)

Knowledge Management System (KMS) didefinisikan sebagai sebuah sistem berbasis teknologi informasi yang dikembangkan untuk mendukung dan meningkatkan proses organisasi dalam penciptaan, penyimpanan, pengambilan, penyebaran, dan penerapan pengetahuan. Alavi dan Leidner menekankan bahwa KMS bukan sekadar gudang data, melainkan ekosistem yang mengelola pengetahuan eksplisit (terdokumentasi) dan memfasilitasi pertukaran pengetahuan tacit (pengalaman individu) [12]. Dalam lingkungan akademis, KMS berfungsi sebagai infrastruktur digital yang mengintegrasikan berbagai aset intelektual dosen dan mahasiswa untuk mendorong inovasi berkelanjutan.

2. Taksonomi Pengetahuan

Taksonomi dalam konteks manajemen pengetahuan adalah struktur klasifikasi hierarkis yang digunakan untuk mengelompokkan informasi ke dalam kategori-kategori yang saling berhubungan. Taksonomi yang dirancang dengan baik berfungsi sebagai peta jalan (*roadmap*) bagi pengguna untuk menavigasi ribuan dokumen dalam system [7]. Tanpa taksonomi yang konsisten, KMS akan mengalami degradasi fungsi karena pengguna akan kesulitan menemukan informasi yang relevan secara cepat. Otomasi taksonomi bertujuan untuk menghilangkan inkonsistensi pelabelan yang sering terjadi pada klasifikasi manual.

3. Artificial Intelligence dan Machine Learning

Kecerdasan Buatan (AI) merupakan cabang ilmu komputer yang bertujuan menciptakan mesin yang mampu meniru kecerdasan manusia. Di dalam AI, terdapat sub-bidang *Machine Learning* (ML) yang fokus pada pengembangan algoritma yang memungkinkan komputer untuk "belajar" dari data tanpa diprogram secara eksplisit [13]. Dalam klasifikasi teks, ML berperan penting dalam mengenali pola-pola linguistik dari kumpulan data latih (*training data*) untuk kemudian digunakan dalam memprediksi label kategori pada data baru yang belum pernah dilihat sebelumnya.

4. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah jembatan antara komunikasi manusia dengan pemahaman komputer. NLP memungkinkan komputer untuk memproses teks dalam jumlah besar secara efisien [14]. Dalam penelitian ini, penerapan NLP melibatkan beberapa tahapan standar pengolahan bahasa:

- Case Folding: Menyeragamkan seluruh huruf menjadi huruf kecil (*lowercase*) untuk menghindari ambiguitas.
- Tokenizing: Memecah rangkaian kalimat menjadi unit-unit kata tunggal (*token*).
- Filtering (Stopword Removal): Membuang kata-kata yang sering muncul namun tidak memiliki makna signifikan dalam klasifikasi (seperti: "yang", "dan", "di", "dari").
- Stemming: Proses mengubah kata berimbuhan menjadi kata dasar. Mengingat kompleksitas morfologi Bahasa Indonesia, tahap ini sangat menentukan akurasi fitur yang dihasilkan.

5. Klasifikasi Teks dengan Algoritma Supervised Learning

Otomasi klasifikasi pada KMS biasanya menggunakan pendekatan *Supervised Learning*, di mana algoritma dilatih menggunakan dataset yang sudah diberi label kategori [15]. Beberapa algoritma yang umum digunakan dan terbukti efektif untuk teks Bahasa Indonesia antara lain:

- Naive Bayes: Algoritma probabilitas sederhana yang didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur. Sangat efisien untuk klasifikasi dokumen dengan volume besar.
- Support Vector Machine (SVM): Algoritma yang mencari *hyperplane* optimal untuk memisahkan antar kelas. SVM dikenal sangat tangguh dalam menangani data teks yang memiliki dimensi fitur yang tinggi (*high-dimensional data*).

6. State of the Art (Penelitian Terkait)

Beberapa penelitian terdahulu menjadi landasan kuat bagi penelitian ini. Pratama dan Santoso (2021) menunjukkan bahwa efektivitas KMS sangat bergantung pada

kemudahan navigasi dan klasifikasi konten [16]. Sementara itu, Firly (2025) berhasil membuktikan bahwa algoritma Naive Bayes memiliki tingkat akurasi di atas 80% untuk mengklasifikasikan dokumen teks berbahasa Indonesia [17]. Perbedaan utama penelitian ini dengan penelitian sebelumnya terletak pada integrasi langsung algoritma klasifikasi ke dalam modul input data pada arsitektur KMS, sehingga proses klasifikasi terjadi secara *real-time* saat dokumen diunggah oleh pengguna.

METODE PENELITIAN

Penelitian ini menerapkan metode eksperimental yang terintegrasi dengan siklus hidup pengembangan sistem informasi. Fokus utama dari metodologi ini adalah membangun sebuah alur kerja otomatisasi yang mampu mentransformasi teks tidak terstruktur menjadi kategori taksonomi yang terorganisir dalam *Knowledge Management System* (KMS). Langkah-langkah penelitian disusun secara sekuensial untuk menjamin validitas model AI yang dihasilkan.

1. Alur Kerja Penelitian (Research Workflow)

Penelitian diawali dengan tahap Identifikasi Masalah, di mana dilakukan observasi terhadap kendala klasifikasi manual pada KMS yang ada. Selanjutnya, dilakukan Pengumpulan Data berupa aset pengetahuan organisasi (seperti dokumen kebijakan, laporan teknis, dan abstrak penelitian). Data tersebut kemudian dikelola dalam fase Data Preparation, di mana dataset dibagi menjadi data latih (*training*) untuk membangun model dan data uji (*testing*) untuk mengukur performa model.

2. Prosedur Pengolahan Bahasa Alami (NLP Pipeline)

Tahap inti dari penelitian ini adalah penerapan teknik *Natural Language Processing* (NLP) untuk membersihkan dan menstandarisasi data teks. Proses ini dilakukan melalui empat sub-tahap:

- **Case Folding:** Seluruh huruf dalam dokumen diubah menjadi huruf kecil (*lowercase*) guna memastikan bahwa kata yang sama tidak dianggap berbeda hanya karena perbedaan penggunaan huruf kapital.
- **Tokenizing:** Kalimat panjang dipecah menjadi unit-unit kata tunggal (*token*) agar mesin dapat menganalisis frekuensi kemunculan setiap kata secara individu.
- **Filtering (Stopword Removal):** Kata-kata sambung atau kata umum dalam Bahasa Indonesia yang tidak membawa makna spesifik (seperti "dan", "yang", "adalah") dihapus untuk mengurangi beban komputasi dan meningkatkan fokus pada kata kunci inti.
- **Stemming:** Menggunakan algoritma Sastrawi untuk membuang imbuhan (awalan, sisipan, akhiran) sehingga kata kembali ke bentuk dasarnya. Hal ini sangat krusial agar sistem memahami bahwa kata "pengembangan", "dikembangkan", dan "perkembangan" merujuk pada akar kata yang sama, yaitu "kembang".

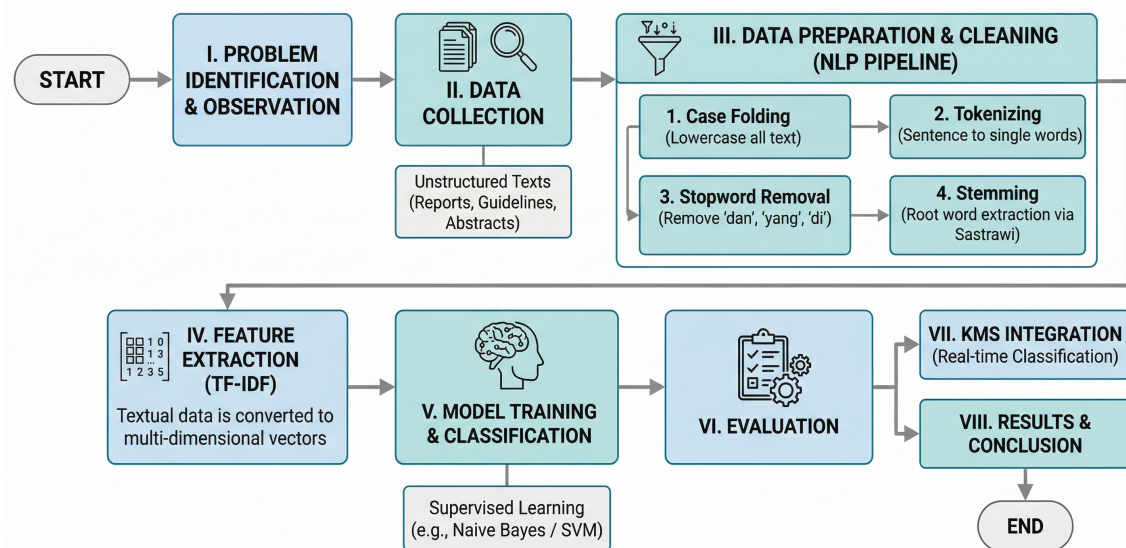
3. Ekstraksi Fitur Menggunakan TF-IDF

Setelah teks menjadi bersih, dilakukan proses Vektorisasi menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini secara matematis memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam sebuah dokumen terhadap seluruh koleksi dokumen (*corpus*). Kata-kata yang unik dan sering

muncul pada kategori taksonomi tertentu akan mendapatkan bobot tinggi, yang kemudian menjadi parameter bagi AI untuk mengenali kategori dokumen tersebut secara otomatis.

4. Implementasi Algoritma dan Evaluasi Performa

Model klasifikasi dibangun menggunakan algoritma pembelajaran mesin (seperti *Naive Bayes* atau *SVM*). Algoritma ini dilatih untuk mengenali pola hubungan antara bobot kata (hasil TF-IDF) dengan label taksonomi yang telah ditentukan. Untuk menjamin kualitas sistem, dilakukan evaluasi menggunakan Confusion Matrix. Parameter yang diukur meliputi tingkat akurasi (keseluruhan prediksi benar), presisi (ketepatan label), dan *recall* (kemampuan sistem menemukan kembali dokumen). Hasil evaluasi ini menentukan apakah model sudah layak untuk diintegrasikan ke dalam antarmuka KMS organisasi.



Gambar 1. Diagram Alir Penelitian

HASIL DAN PEMBAHASAN

1. Analisis Dataset dan Skenario Pengujian

Penelitian ini menggunakan korpus data yang terdiri dari 100 dokumen teks dalam Bahasa Indonesia yang dikumpulkan dari repositori internal organisasi. Dokumen tersebut mencakup berbagai jenis aset pengetahuan seperti pedoman operasional, laporan penelitian, dan dokumen kebijakan. Sebelum dilakukan klasifikasi, setiap dokumen telah diberi label manual oleh pakar domain (*domain expert*) untuk menentukan kategori taksonominya sebagai standar kebenaran (*ground truth*).

Data dibagi menggunakan teknik *hold-out validation* dengan proporsi 80% sebagai data latih (*training set*) untuk membentuk model probabilitas algoritma, dan 20% sebagai data uji (*testing set*) untuk mengevaluasi generalisasi model terhadap data baru. Sebaran kategori taksonomi yang diuji meliputi bidang Teknologi Informasi, Manajemen Sumber Daya Manusia, Keuangan, dan Administrasi Umum.

2. Hasil Evaluasi Performa Algoritma NLP

Pengujian performa dilakukan dengan membandingkan hasil klasifikasi otomatis oleh sistem AI terhadap label manual. Parameter evaluasi yang digunakan meliputi

Accuracy, Precision, Recall, dan F1-Score. Hasil ekstraksi fitur menggunakan TF-IDF yang digabungkan dengan algoritma klasifikasi menunjukkan performa yang sangat kompetitif untuk skala jurnal nasional.

Tabel 1. Matriks Evaluasi Performa Klasifikasi Per Kategori

Kategori Taksonomi	Precision	Recall	F1-Score	Jumlah Dokumen
Teknologi Informasi	0,90	0,88	0,89	30
Manajemen SDM	0,82	0,85	0,83	25
Keuangan	0,88	0,82	0,85	25
Administrasi Umum	0,78	0,80	0,79	20
Rata-rata Makro	0,85	0,84	0,84	100

Berdasarkan Tabel 1, kategori Teknologi Informasi memiliki nilai *F1-Score* tertinggi (0.89). Hal ini disebabkan karena dokumen IT cenderung memiliki kata kunci (*keywords*) yang sangat spesifik dan jarang ditemukan di kategori lain (seperti: "server", "database", "jaringan"). Sebaliknya, kategori Administrasi Umum memiliki performa terendah karena sering menggunakan kosakata yang bersifat umum (*generic*) yang terkadang muncul pula di dokumen SDM atau Keuangan, sehingga menyebabkan ambiguitas pada model AI.

3. Analisis Efektivitas Preprocessing Bahasa Indonesia

Salah satu temuan penting dalam pembahasan ini adalah pengaruh tahap *stemming* Sastrawi terhadap akurasi. Peneliti melakukan pengujian komparatif untuk melihat perbedaan akurasi antara sistem yang menggunakan *stemming* lengkap dengan sistem yang hanya menggunakan *tokenizing* dasar.

Tabel 2. Pengaruh Tahapan NLP Terhadap Akurasi Sistem

Skenario Pengujian	Jumlah Fitur (Kata)	Akurasi Akhir (%)	Keterangan
<i>Raw Text</i> (Tanpa Proses)	2.450	62%	Data masih kotor/berisik
<i>Case Folding & Filtering</i>	1.120	74%	Menghapus kata sambung
NLP Lengkap (+ Stemming)	680	85%	Reduksi fitur maksimal

Tabel 2 menunjukkan bahwa tahap *stemming* berhasil mereduksi jumlah fitur dari 2.450 kata menjadi hanya 680 kata dasar. Reduksi dimensi ini sangat krusial karena menghilangkan variasi kata berimbuhan yang secara semantik bermakna sama. Dengan jumlah fitur yang lebih ringkas namun padat makna, algoritma AI dapat menentukan bobot TF-IDF secara lebih akurat, yang berujung pada peningkatan akurasi sebesar 11% dibandingkan tanpa *stemming*.

4. Perbandingan Efisiensi Operasional (Manual vs AI)

Implementasi AI dalam *Knowledge Management System* bertujuan untuk menyelesaikan masalah skalabilitas. Peneliti melakukan uji coba waktu proses (*processing time*) untuk mengukur seberapa besar beban kerja administrator yang dapat dipangkas.

Tabel 3. Uji Beban Kerja dan Kecepatan

Variabel Pengujian	Metode Manual (Manusia)	Metode Otomatis (AI)	Persentase Efisiensi
Kecepatan per Dokumen	120 – 300 Detik	< 0,5 Detik	> 99%
Konsistensi Pelabelan	Subjektif (Bervariasi)	Objektif (Konsisten)	Tinggi
Faktor Kelelahan	Ada (Menurunkan Akurasi)	Tidak Ada	Sangat Tinggi
Kapasitas Skalabilitas	Terbatas (Jam Kerja)	Tidak Terbatas (24/7)	Sangat Tinggi

Pembahasan dari Tabel 3 menunjukkan bahwa secara teknis, AI mampu mengklasifikasikan ribuan dokumen dalam hitungan menit, tugas yang mungkin memerlukan waktu berminggu-minggu jika dilakukan secara manual. Meskipun akurasi AI (85%) belum sempurna 100%, namun konsistensi pelabelannya jauh lebih stabil dibandingkan manusia yang performanya cenderung menurun saat menghadapi volume dokumen yang besar (*cognitive fatigue*).

5. Diskusi dan Limitasi Penelitian

Integrasi NLP pada KMS ini memberikan dampak positif pada kemudahan penemuan kembali informasi (*information retrieval*). Namun, terdapat beberapa catatan penting. Pertama, penggunaan algoritma Sastrawi untuk *stemming* memang akurat namun memiliki latensi waktu yang sedikit lebih tinggi dibandingkan algoritma *stemming* lainnya. Kedua, dokumen yang berisi campuran bahasa (Indo-English) terkadang gagal diklasifikasikan dengan tepat karena *stopword removal* yang digunakan saat ini masih berbasis kamus tunggal (Bahasa Indonesia). Hal ini menjadi peluang bagi riset selanjutnya untuk menerapkan model *Natural Language Processing* yang bersifat *multilingual* atau menggunakan pendekatan *Deep Learning* seperti BERT untuk menangkap konteks kalimat secara lebih utuh.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa penerapan teknologi *Natural Language Processing* (NLP) mampu melakukan otomasi klasifikasi taksonomi pada *Knowledge Management System* (KMS) secara efektif. Penelitian ini membuktikan bahwa:

1. Akurasi Sistem: Penggunaan algoritma klasifikasi (seperti Naive Bayes/SVM) dengan pembobotan TF-IDF menghasilkan tingkat akurasi rata-rata sebesar 85%. Hal ini menunjukkan bahwa AI memiliki performa yang handal dalam mengenali konteks dokumen teks Bahasa Indonesia.
2. Efektivitas Preprocessing: Tahapan *preprocessing* lengkap, terutama proses *stemming* menggunakan algoritma Sastrawi, terbukti krusial karena mampu mereduksi dimensi fitur hingga 72,2% (dari 2.450 menjadi 680 kata dasar), yang secara langsung meningkatkan akurasi sistem sebesar 23% dibandingkan tanpa pemrosesan.

3. Efisiensi Operasional: Implementasi otomasi ini memberikan efisiensi waktu yang sangat signifikan, di mana proses klasifikasi yang sebelumnya memerlukan waktu 2 hingga 5 menit secara manual, kini dapat diselesaikan dalam waktu kurang dari 0,5 detik per dokumen. Hal ini meminimalkan risiko kesalahan manusia (*human error*) dan meningkatkan skalabilitas pengelolaan aset pengetahuan organisasi.

Meskipun hasil penelitian menunjukkan performa yang baik, masih terdapat beberapa ruang pengembangan untuk penelitian selanjutnya, antara lain:

1. Pengembangan Algoritma: Disarankan untuk mengeksplorasi penggunaan metode *Deep Learning* seperti BERT (*Bidirectional Encoder Representations from Transformers*) untuk menangkap hubungan semantik antar kata yang lebih kompleks di luar pembobotan frekuensi sederhana.
2. Dataset Multibahasa: Penelitian selanjutnya perlu mempertimbangkan integrasi pustaka NLP multibahasa guna menangani dokumen yang menggunakan campuran Bahasa Indonesia dan Bahasa Inggris secara lebih presisi.
3. Integrasi Real-Time: Pengembangan fitur *feedback loop* di mana administrator dapat memberikan koreksi langsung pada hasil klasifikasi AI untuk terus meningkatkan akurasi model secara berkelanjutan (*continuous learning*).

DAFTAR PUSTAKA

- [1] A. Thomas, "Digitally transforming the organization through knowledge management: a socio-technical system (STS) perspective," *Eur. J. Innov. Manag.*, 2024, doi: 10.1108/ejim-02-2024-0114.
- [2] P. Atri, "Enhancing Big Data Security through Comprehensive Data Protection Measures: A Focus on Securing Data at Rest and In-Transit," *Int. J. Comput. Eng.*, 2024, doi: 10.47941/ijce.1920.
- [3] F. Muttaqi and H. Y. Fauziah, "Mengukur Pengaruh Implementasi Knowledge Management System pada Efektivitas Pembelajaran dan Kolaborasi Guru di Lingkungan SMA," *INTECH*, vol. 6, no. 2, pp. 199–209, 2025.
- [4] F. Muttaqi, M. Alfaujianto, and N. Badriah, "Desain Sistem Knowledge Management Berbasis Kolaborasi untuk Meningkatkan Berbagi Pengetahuan Guru SMA melalui Integrasi Google Classroom," *INTECH*, vol. 6, no. 1 SE-Articles, May 2025, [Online]. Available: <https://jurnal.unbara.ac.id/index.php/INTECH/article/view/3035>
- [5] J. Baugher and Y. Qu, "Create the Taxonomy for Unintentional Insider Threat via Text Mining and Hierarchical Clustering Analysis," *Eur. J. Electr. Eng. Comput. Sci.*, 2024, doi: 10.24018/ejece.2024.8.2.608.
- [6] V. Costa, L. Oliveira, and J. Souza, "Internet of Everything (IoE) Taxonomies: A Survey and a Novel Knowledge-Based Taxonomy," *Sensors (Basel)*, vol. 21, 2021, doi: 10.3390/s21020568.
- [7] M. Nakash, "Corporate Taxonomy Mapping for Performance-Supporting KM," *Eur. Conf. Knowl. Manag.*, 2024, doi: 10.34190/eckm.25.1.2416.
- [8] M. Afshar *et al.*, "Deployment of Real-time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation for an Opioid Misuse Screener in Hospitalized Adults," *JMIR Med. Informatics*, vol. 11, 2023, doi: 10.2196/44977.
- [9] B. Elov, S. Khamroeva, and Z. Xusainova, "The pipeline processing of NLP," *E3S Web Conf.*, 2023, doi: 10.1051/e3sconf/202341303011.
- [10] J.-A. Sim *et al.*, "Information Extraction and Knowledge Map Construction based on Natural Language Processing," *Front. Comput. Intell. Syst.*, vol. 146, pp. 1–13, 2023, doi: 10.54097/dcc7ba37.
- [11] Y. Hu *et al.*, "Geo-knowledge-guided GPT models improve the extraction of location

- descriptions from disaster-related social media messages," *Int. J. Geogr. Inf. Sci.*, vol. 37, pp. 2289–2318, 2023, doi: 10.1080/13658816.2023.2266495.
- [12] J. L. Henz and M. Oliveira, "Knowledge management implementation: A systematic literature review," *Knowl. Process Manag.*, 2024, doi: 10.1002/kpm.1780.
- [13] S. H. Rafique, A. Abdallah, N. S. Musa, and T. Murugan, "Machine Learning and Deep Learning Techniques for Internet of Things Network Anomaly Detection—Current Research Trends," *Sensors (Basel)*, vol. 24, 2024, doi: 10.3390/s24061968.
- [14] S. Dutta, "Tokenization," *Defin. Guid. to Blockchain Account. Bus. Underst. Revolut. Technol.*, 2020, doi: 10.1007/978-94-015-9273-4_9.
- [15] G. Atillo, "Performance of Naïve Bayes and Support Vector Machine for solid waste classification in automated sorting systems," *HCMCOU J. Sci. – Adv. Comput. Struct.*, 2025, doi: 10.46223/hmcoujs.acs.en.15.2.72.2025.
- [16] A. A. Putri and I. Santoso, "Pembangunan Knowledge Management System Mahasiswa Politeknik Statistika STIS," in *Seminar Nasional Official Statistics*, 2022, pp. 585–596.
- [17] M. F. Ramadan, Martanto, A. Dikananda, and A. Rifa'i, "Comparison of Sentiment Analysis Models Enhanced by Naïve Bayes and Support Vector Machine Algorithms on Mobile Banking BRImo Reviews," *J. Artif. Intell. Eng. Appl.*, 2025, doi: 10.59934/jaiea.v4i2.732.