

**KLASIFIKASI DOKUMEN PADA *DATABASE*
PERPAJAKAN: PENERAPAN *ARTIFICIAL*
INTELLIGENCE DAN *NATURAL LANGUAGE*
*PROCESSING***

**Dewi Putrie Lestari
Nurma Nugraha
Sayidati Karima
Rifiana Arief
Beny Susanti
Feni Andriani
Nola Marina**

2026



PENERBIT GUNADARMA

**KLASIFIKASI DOKUMEN PADA *DATABASE PERPAJAKAN:*
PENERAPAN ARTIFICIAL INTELLIGENCE DAN NATURAL
*LANGUAGE PROCESSING***

PENULIS

Dewi Putrie Lestari, Nurma Nugraha, Sayidati Karima, Rifiana Arief, Beny Susanti, Feni Andriani, Nola Marina

(ix + 184)

ISBN

CETAKAN PERTAMA

Januari 2026

PENYUNTING

Dewi Putrie Lestari

DESAIN DAN TATA LETAK

Sayidati Karima

PENERBIT

Penerbit Gunadarma

Jl. Margonda Raya No. 100, Pondok Cina, Depok 16424

Telp. +62-21-78881112, 7863819

Faks. +62-21-7872829

Email: penerbit@gunadarma.ac.id

Hak cipta dilindungi undang-undang.

Dilarang mengutip atau memperbanyak dalam bentuk apa pun, sebagian atau seluruh isi buku, tanpa ijin tertulis dari Penerbit.

KATA PENGANTAR

Bismillahirrahmanirrahiim

Assalammu'alaykum warrahmatullahi wabarakatuuh

Alhamdulillahirabbil'aalamiin

Puji dan syukur kami panjatkan ke hadirat Allah Yang Maha Esa, karena atas limpahan rahmat dan karunia-Nya, buku referensi yang berjudul “Klasifikasi Dokumen pada *Database* Perpajakan: Penerapan *Artificial Intelligence* dan *Natural Language Processing*” ini dapat diselesaikan dengan baik. Buku ini disusun untuk memberikan wawasan serta panduan bagi akademisi, peneliti, dan praktisi yang ingin memahami serta menerapkan teknologi kecerdasan buatan (*artificial intelligence*) dan pemrosesan bahasa alami (*natural language processing*) dalam sistem klasifikasi dokumen perpajakan untuk meningkatkan efisiensi dan akurasi pengelolaan *database* perpajakan.

Dalam era digital, pengelolaan dokumen perpajakan dalam basis data (*database*) menjadi tantangan besar terutama dalam hal klasifikasi dokumen. Metode klasifikasi konvensional yang selama ini digunakan tidak lagi efisien dalam menangani volume data yang terus meningkat. Volume dokumen perpajakan yang semakin besar, seperti laporan pajak, surat ketetapan pajak, dan dokumen administratif lainnya, menuntut sistem yang mampu mengklasifikasikan dokumen secara otomatis dan akurat.

Penerapan *Artificial Intelligence* (AI) dan *Natural Language Processing* (NLP) menjadi solusi yang efektif dalam mengatasi tantangan ini. Penggunaan AI dan NLP memungkinkan klasifikasi dokumen secara otomatis, pencarian informasi yang lebih relevan, serta ekstraksi data yang lebih cepat dan akurat. Kemampuan tersebut sangat penting dalam menangani volume dokumen perpajakan yang besar dan beragam, sehingga dapat meningkatkan efisiensi pengelolaan dokumen pada *database* perpajakan.

Buku ini membahas berbagai metode dan teknik dalam penerapan AI dan NLP untuk klasifikasi dokumen perpajakan. Beberapa pendekatan yang umum digunakan mencakup metode

berbasis pembelajaran mesin (*machine learning*) seperti *Support Vector Machine* (SVM), serta pendekatan berbasis pembelajaran mendalam (*deep learning*) seperti *Recurrent Neural Network* (RNN) dengan arsitektur *Long Short-Term Memory* (LSTM) yang mampu menangkap hubungan kontekstual dalam teks secara lebih efektif. Model berbasis *embedding* kata seperti *Global Vectors* (GloVe) juga telah menunjukkan keunggulan dalam representasi teks yang lebih bermakna untuk keperluan klasifikasi dokumen. Metode-metode ini mampu menangani teks berukuran besar dan memahami pola linguistik dalam dokumen pajak secara lebih efektif, sehingga implementasi sistem ini dapat mengotomatisasi proses klasifikasi dokumen, mempercepat pencarian informasi, serta mengurangi kesalahan dalam pengelolaan *database* perpajakan.

Kami mengucapkan terima kasih yang sebesar-besarnya kepada Direktorat Jenderal Riset dan Pengembangan (Ditjen Risbang), Kementerian Pendidikan Tinggi, Sains, dan Teknologi (Kemdiktisaintek) atas dukungan pendanaan penelitian ini. Dukungan tersebut memberikan kesempatan kepada kami untuk mengembangkan solusi inovatif di bidang teknologi informasi, khususnya dalam penerapan AI dan NLP untuk pengelolaan *database* dokumen perpajakan yang semakin kompleks di era digital.

Ucapan terima kasih juga kami sampaikan kepada Universitas Gunadarma, Lembaga Penelitian Universitas Gunadarma, dan Tax Center Universitas Gunadarma yang telah memberikan dukungan penuh selama proses penelitian dan penyusunan buku ini. Dukungan tersebut tidak hanya berupa fasilitas penelitian, tetapi juga dorongan akademis dan teknis yang sangat membantu kami dalam mencapai hasil yang optimal. Selain itu, kami ingin mengucapkan terima kasih kepada para narasumber yang berasal dari Direktorat Jenderal Pajak (DJP), Kantor Wilayah DJP Jawa Barat III, KPP Pratama Depok Cimanggis, KPP Pratama Depok Sawangan, semua anggota tim, serta pihak-pihak yang berperan dalam membantu penyelesaian buku ini.

Kami menyadari bahwa buku ini masih memiliki kekurangan, baik dari segi isi maupun penyajiannya. Oleh karena itu, kami sangat terbuka untuk menerima masukan, kritik, dan saran yang membangun untuk penyempurnaan buku ini di masa mendatang. Semoga buku ini

dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, serta menjadi referensi bagi akademisi, peneliti, mahasiswa, serta praktisi yang tertarik dalam pengembangan sistem klasifikasi dokumen dan pengelolaan *database* perpajakan berbasis kecerdasan buatan. Dengan menggabungkan pendekatan teoritis dan aplikasi praktis, buku ini diharapkan mampu menjembatani kebutuhan akademis dan kebutuhan dunia industri dalam menghadapi tantangan era digitalisasi yang semakin berkembang pesat.

Akhir kata, kami mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penyusunan buku ini. Semoga buku ini dapat memberikan manfaat dan wawasan yang luas bagi pembaca.

Wassalammu'alaykum warrahmatullahi wabarakatuuh

Depok, Januari 2026
Penulis

Penerbit Gunadarma

DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI.....	vi
SINOPSIS	ix

BAB 1 PENDAHULUAN	1
1.1 Permasalahan Pengelolaan <i>Database</i> Perpajakan	3
1.2 Pencarian Dokumen dalam <i>Database</i> Perpajakan.....	5
1.2.1 Metode Pencarian Dokumen Perpajakan.....	5
1.2.2 Kendala Pencarian Dokumen Perpajakan	7
1.2.3 Teknologi Pencarian Dokumen Perpajakan	8
1.3 Klasifikasi pada Pencarian Dokumen Perpajakan	10
1.4 Penerapan AI dan NLP Klasifikasi Dokumen Perpajakan	11

BAB 2 DATABASE PERPAJAKAN	14
2.1 Pengertian dan Peran <i>Database</i> Perpajakan.....	14
2.2 Tujuan Penggunaan <i>Database</i> Perpajakan	16
2.3 Struktur dan Komponen <i>Database</i> Perpajakan	18
2.4 Fungsi dan Manfaat Database Perpajakan.....	19
2.5 Tantangan dalam Pengelolaan <i>Database</i> Perpajakan	20
2.6 Solusi Teknologi untuk Pengelolaan Database Perpajakan.....	21
2.7 Optimalisasi Pengelolaan Database Perpajakan	22
2.7.1 Pentingnya Peningkatan Sistem Pencarian	22
2.7.2 Penggunaan Teknologi AI dan NLP	23
2.7.3 Rekomendasi untuk Pengoptimalan Database	25

BAB 3 KLASIFIKASI DOKUMEN PERPAJAKAN	27
3.1 Konsep Dasar Klasifikasi Dokumen	27
3.1.1 Definisi Klasifikasi Dokumen	28
3.1.2 Tujuan Klasifikasi Dokumen	32
3.1.3 Penerapan Klasifikasi Dokumen dalam Kehidupan Nyata	34
3.2 Jenis Klasifikasi Dokumen.....	40

3.2.1 Klasifikasi Dokumen Berbasis Konten	40
3.2.2 Klasifikasi Dokumen Berbasis Konteks.....	44
3.2.3 Klasifikasi Dokumen Berbasis Metadata	46
3.3 Metode Klasifikasi Dokumen	50
3.3.1 Metode <i>Machine Learning</i>	50
3.3.1.1 Metode <i>Supervised Learning</i>	51
3.3.1.1 Metode <i>Unsupervised Learning</i>	53
3.3.1.2 Metode <i>Semi-Supervised Learning</i>	54
3.3.2 Metode <i>Deep Learning</i>	57
3.4 Evaluasi Model pada Klasifikasi Dokumen.....	62
3.4 Penerapan Klasifikasi Dokumen pada Database Perpajakan.....	64

BAB 4 PEMROSESAN BAHASA ALAMI UNTUK KLASIFIKASI DOKUMEN PERPAJAKAN

4.1 Teknik Pra-pemrosesan Data	70
4.1.1 Tokenisasi	72
4.1.2 <i>Stopword Removal</i>	73
4.1.3 <i>Stemming</i> dan <i>Lemmatisasi</i>	73
4.2 Representasi Teks dalam Klasifikasi Dokumen	74
4.2.1 <i>Bag-of-Words</i>	75
4.2.2 <i>Term Frequency-Inverse Document Frequency</i>	77
4.2.3 <i>Word Embeddings</i>	79
4.3 Ekstraksi Fitur untuk Dokumen Perpajakan	81
4.3.1 Pemilihan Fitur yang Relevan	82
4.3.2 Reduksi Dimensi dan Pengaruhnya pada Klasifikasi.....	84

BAB 5 METODE KLASIFIKASI BERBASIS KECERDASAN BUATAN.....

5.1 Metode Klasifikasi Berbasis Pembelajaran Mesin.....	86
5.1.1 Metode <i>Naïve Bayes</i>	87
5.1.2 Metode <i>Support Vector Machine</i>	101
5.1.3 Metode <i>K-Nearest Neighbors</i>	111
5.1.4 Metode <i>Decision Tree</i>	121

5.2 Metode Klasifikasi Berbasis Pembelajaran Mendalam	131
5.2.1 Metode <i>Convolutional Neural Network</i>	132
5.2.2 Metode <i>Long Short-Term Memory</i>	140
5.2.3 Metode <i>Recurrent Neural Network</i>	147
 BAB 6 IMPLEMENTASI DAN EVALUASI MODEL KLASIFIKASI PADA DOKUMEN PAJAK	154
6.1 Pengumpulan dan Persiapan Data.....	155
6.1.1 Sumber Data Perpajakan	155
6.1.2 Pra-Pemrosesan Data dan Pengelolaan Data	157
6.2 Implementasi Model Klasifikasi Dokumen.....	158
6.2.1 Model Klasifikasi SVM dengan TF-IDF	159
6.2.2 Model Klasifikasi CNN dengan GloVe	160
6.2.3 Model Klasifikasi RNN dengan LSTM dan	161
6.3 Pengujian Model dan Analisis Hasil.....	162
6.3.1 Evaluasi Kinerja Model Klasifikasi	163
6.3.2 Perbandingan Kinerja Model Klasifikasi	166
6.4 Hasil Klasifikasi Dokumen Perpajakan	167
 DAFTAR PUSTAKA	170

SINOPSIS

"Klasifikasi Dokumen pada *Database Perpajakan: Penerapan Artificial Intelligence dan Natural Language Processing*" menghadirkan solusi inovatif untuk otomatisasi klasifikasi dokumen di era digital. Buku ini mengintegrasikan teknologi *Artificial Intelligence (AI)* dan *Natural Language Processing (NLP)* guna meningkatkan akurasi serta efisiensi dalam pengelolaan dokumen perpajakan yang semakin kompleks seiring dengan pertumbuhan data dan kebutuhan digitalisasi. Dengan pendekatan yang menggabungkan teori dan praktik, buku ini membahas berbagai teknik modern untuk mengoptimalkan proses klasifikasi dokumen.

Buku ini terdiri atas enam bab yang disusun secara sistematis untuk memudahkan pembaca dalam memahami setiap pembahasan. **Bab 1** menguraikan latar belakang, urgensi, dan manfaat otomatisasi pengelolaan *database* perpajakan. **Bab 2** membahas konsep dasar *database* perpajakan. **Bab 3** menjelaskan prinsip klasifikasi dokumen, termasuk teknik dan model untuk meningkatkan akurasi dan efisiensi. **Bab 4** membahas teknik NLP, seperti tokenisasi, *stopword removal*, *stemming*, lemmatisasi, serta metode ekstraksi fitur seperti **TF-IDF** dan **GloVe**. **Bab 5** membahas berbagai pendekatan AI, seperti **SVM**, **CNN**, dan **RNN**. **Bab 6** menyajikan implementasi model klasifikasi dokumen perpajakan serta evaluasi kinerjanya menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*.

Dengan pendekatan ilmiah yang tetap mudah dipahami, buku ini menjadi referensi penting bagi akademisi, mahasiswa, peneliti, dan praktisi di bidang **teknologi informasi, perpajakan**, serta **manajemen data** yang mencari solusi inovatif dalam menghadapi tantangan digitalisasi. Selain itu, melalui buku ini, pembaca akan memperoleh wawasan mendalam mengenai pemanfaatan AI dan NLP dalam pengelolaan dokumen perpajakan secara lebih efektif dan efisien.