

DISERTASI

**KLASIFIKASI PATEN DENGAN PEMBOBOTAN
RINGKASAN MENGGUNAKAN PENDEKATAN MACHINE
LEARNING**



Nama : Slamet Widodo
NIM : 03013622025005
BKU : Teknik Informatika
Promotor : Prof. Dr. Ermatita, M.Kom
Ko - Promotor : Prof. Ir. Deris Stiawan, M.T., Ph.D

**PROGRAM STUDI DOKTOR ILMU TEKNIK
FAKULTAS TEKNIK
UNIVERSITAS SRIWIJAYA
2025**

DISERTASI

**KLASIFIKASI PATEN DENGAN PEMBOBOTAN RINGKASAN
MENGUNAKAN PENDEKATAN *MACHINE LEARNING***

**Diajukan Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Doktor
Dalam Bidang Ilmu Teknik Informatika**



SLAMET WIDODO

03013622025005

PROGRAM STUDI ILMU TEKNIK PROGRAM DOKTOR

FAKULTAS TEKNIK

UNIVERSITAS SRIWIJAYA

2025

HALAMAN PENGESAHAN

DISERTASI

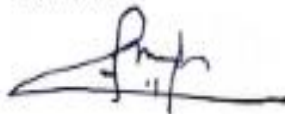
(TEK010317)

KLASIFIKASI PATEN DENGAN PEMBOBOTAN RINGKASAN MENGGUNAKAN PENDEKATAN MACHINE LEARNING

Oleh:
SLAMET WIDODO
03013622025005

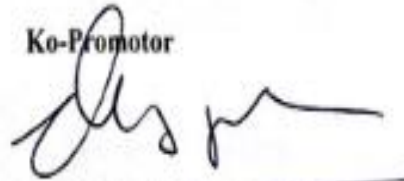
Telah disetujui
Pada Tanggal 15 Mei 2025

Promotor



Prof. Dr. Ermatita, M.Kom
NIP. 196709132006042001

Ko-Promotor



Prof. Ir. Deris Stiawan, M.T., Ph.D
NIP. 197806172006041002

Mengetahui,

Dekan Fakultas Teknik,



Dr. Ir. Bhakni Yudho Suprpto, S.T., M.T., IPM.
NIP. 197502112003121002

Koordinator Program Studi



Prof. Dr. Ir. Nukman, M.T.
NIP. 195903211987031001

HALAMAN PERSETUJUAN

Disertasi berjudul "Klasifikasi Paten Dengan Pembobotan Ringkasan Menggunakan Pendekatan Machine Learning " telah dipresentasikan dihadapan Tim Penguji Disertasi pada Program Studi Doktor Ilmu Teknik Fakultas Teknik Universitas Sriwijaya pada Kamis, 24 April, 2025.

Palembang, 15 Mei 2025

Ditandatangani oleh:

Ketua Tim Penguji:

Dr. Ir. Bhakti Yudho Suprpto, S.T., M.T., IPM.
NIP. 197502112003121002



Anggota Tim Penguji:

1. Dian Palupi Rini, S.Si., M.Kom., Ph.D
NIP.197806172006041002

2. Dr.Fathoni, S.T., MMSI
NIP. 197210182008121001

3. Dr.Rendra Gustriansyah,S.T.,M.Kom
NIDN. 0220087301



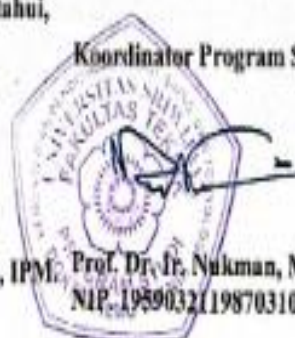
Mengetahui,

Dekan Fakultas Teknik,

Koordinator Program Studi

Dr. Ir. Bhakti Yudho Suprpto, S.T., M.T., IPM.
NIP. 197502112003121002

Prof. Dr. Ir. Nukman, M.T.
NIP. 195903211987031001



HALAMAN PERNYATAAN INTEGRITAS

Yang bertanda tangan di bawah ini:

Nama : Slamet Widodo
NIM : 03013622025005
Tempat,Tanggal Lahir : Sukoharjo,16 Mei 1973
Jurusan/Program Studi: S3 Ilmu Teknik – Bidang Teknik Informatika
Fakultas : Teknik Universitas Sriwijaya
Alamat Rumah : Komplek Multiwahana Jl. Bambu Runcing 1, Blok
N1 NO.06 RT.14 RW.04 Kel.Lebung Gajah,
Kecamatan Sematang Borang, Palembang 30168
Alamat E-mail : slametwidodo@polsri.ac.id

Dengan ini menyatakan dengan sesungguhnya bahwa disertasi yang berjudul
**“KLASIFIKASI PATEN DENGAN PEMBOBOTAN RINGKASAN
MENGUNAKAN PENDEKATAN MACHINE LEARNING”**

Bebas dari plagiarism dan bukan hasil karya orang lain

Apabila dikemudian hari ditemukan seluruh atau sebagian dari disertasi tersebut
terdapat indikasi plagirisme, saya bersedia menerima sanksi dengan peraturan
perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari
siapapun dan untuk dipergunakan sebagaimana mestinya.



Palembang, Mei 2025

Yang membuat pernyataan,



Slamet Widodo
NIM. 03013622025005

KATA PENGANTAR

Puji syukur dipanjatkan ke hadirat Allah SWT, atas limpahan rahmat dan hidayah-Nya dapat menyelesaikan penelitian Disertasi dengan judul” KLASIFIKASI PATEN DENGAN PEMBOBOTAN RINGKASAN MENGGUNAKAN PENDEKATAN MACHINE LEARNING”, Penyelesaian penyusunan Disertasi ini dapat diselesaikan dengan dukungan,bimbingan serta arahan dari berbagai pihak. Dalam kesempatan ini diucapkan terimakasih yang sedalam-dalamnya Kepada Yth:

1. Prof.Dr.Taufik Marwa,S.T.,M.T., selaku Rektor Universitas Sriwijaya.
2. Dr.Bhakti Yudho Suprpto,M.T.,IPM., selaku Dekan Fakultas Teknik Universitas Sriwijaya.
3. Prof.Dr.Nukman,M.T., selaku Koordinator Program Studi Ilmu Teknik Program Doktor, Fakultas Teknik Universitas Sriwijaya yang telah memberikan motivasi untuk menyelesaikan Disertasi.
4. Ibu Prof.Dr.Ermatita,M.Kom., sebagai promotor atas bimbingan,arahan dan serta motivasi yang diberika dalam penyusunan Disertasi.
5. Bapak Prof.Ir. Deris Stiawan.,M.T,Ph.D., sebagai ko-promotor atas bimbingan,arahan dan serta motivasi yang diberika dalam penyusunan Disertasi.
6. Dosen-dosen Program Studi Ilmu Teknik(S3) BKU Teknik Informatika beserta staf yang telah mengajarkan,berbagai ilmu, dan juga membantu penulis dalam pengurusan persyaratan administrasi.
7. Pimpinan, rekan kerja dan staf Politeknik Negeri Sriwijaya atas dukungannya selama menempuh pendidikan S3.
8. Rekan-rekan jurusan Teknik Komputer Politeknik Negeri Sriwijaya, terima kasih atas dukungannya.
9. Rekan-rekan di Program Studi Ilmu Teknik (S3) BKU Teknik Informatika khususnya angkatan 2019 Semester Genap yang telah bersama-sama berjuang menyelesaikan pendidikan S3.
10. Istri dan anak-anakku tercinta, keluargaku tercinta, yang telah memberikan dorongan,pengertian dan pengorbanan.

11. Semua pihak yang telah memberikan bantuan sehingga selesainya disertasi ini.

Semoga Allah SWT, memberikan balasan atas semua amal kebaikan bapak/ibu dan mejadikan ladang amal yang baik.

Salam Hormat,

Slamet Widodo

NIM. 03013622025005

**SURAT KETERANGAN PENGECEKAN
SIMILARITY**

Saya yang bertanda tangan dibawah ini:

Nama : Slamet Widodo
NIM : 03013622025008
Prodi : Doktoral Ilmu Teknik

Menyatakan bahwa benar hasil pengecekan similarity Disertasi yang berjudul “ Klasifikasi Paten Dengan Pembobotan Ringkasan Menggunakan Pendekatan Machine Learning” adalah dibawah 20%

Dicek Oleh* : 1. Dosen pembimbing
2. Dichek oleh operator UPT Perpustakaan Universitas Sriwijaya melalui laman website:<https://digilib.unsri.ac.id/>.

Demikian surat keterangan ini saya buat dengan sebenarnya dan dapat dipertanggung jawabkan

Menyetujui,
Dosen Pembimbing


Prof. Dr. Ermatita, M. Kom
NIP. 196709132006042001

Palembang, Mei 2025
Yang Menyatakan


Slamet Widodo
NIM. 03013622025005

***Lingkari salah satu jawaban, tempat anda melakukan pengecekan similarity**

RINGKASAN

Klasifikasi Paten Dengan Pembobotan Ringkasan Menggunakan Pendekatan Machine Learning

Paten merupakan hak eksklusif inventor atas invensi di bidang teknologi pada saat melakukan sendiri atau memberikan persetujuan kepada pihak lain untuk melaksanakan invensinya. Invensi merupakan ide inventor yang dijabarkan dalam kegiatan untuk memecahkan permasalahan khusus pada bidang teknologi, bisa dalam bentuk produk maupun cara melakukan prosesing atau penyempurnaan untuk pengembangannya produk. Permasalahan rumitnya usulan invensi paten bagi inventor menjadi pembahasan yang sangat penting melalui tahapan-tahapan proses menentukan bidang teknik invensi, latar belakang invensi, ringkasan invensi, uraian lengkap invensi, klaim, pencarian referensi, dan abstrak. Masalah redundansi ringkasan paten adalah untuk menghindari duplikasi atau berulang kali menyimpan informasi yang sama pada label atau kelas paten *IPC (International Patent Classification)* di beberapa dokumen paten melalui media internet, sehingga menyimpan informasi yang sama di lebih dari satu kelas atau label. Tujuan dari penelitian ini adalah untuk membuat dataset paten Indonesia sebagai model pembelajaran mesin untuk memprediksi kelas atau label dari data masukan berdasarkan informasi yang diberikan oleh data paten *IPC (International Patent Classification)* pelatihan sebelumnya. Tujuan utamanya untuk mengidentifikasi pola atau hubungan dalam data yang dapat digunakan untuk mengklasifikasikan data baru usulan paten ke dalam salah satu dari banyak kategori yang telah ditentukan data paten *IPC* yang relevan. Hasil prediksi model bisa satu, dua atau tiga, tergantung kategori mana yang paling mirip dengan model yang dibuat. Kontribusinya adalah membuat model prediksi dokumen paten baru untuk mendapatkan kutipan paten yang signifikan berdasarkan kategori dokumen paten., klasifikasi dapat membantu pengambilan keputusan dengan memberikan prediksi tentang kelas atau kategori paten yang paling relevan berdasarkan fitur data paten. Kontribusi penelitian ini adalah diperolehnya bahan yang dapat dijadikan acuan untuk mengukur dan mengevaluasi model klasifikasi yang diterapkan dalam dokumen paten Indonesia; dapat menghemat waktu dan tenaga manusia dalam tugas-tugas seperti menyortir data kemiripan paten, memeriksa dokumen paten.

Kata Kunci : Paten, Klasifikasi Paten, *IPC (International Patent Classification)*

Palembang, Mei 2025,

Promotor

Mahasiswa



Prof. Dr. Ernaita, M. Kom
NIP. 196709132006042001



Slamet Widodo



SUMMARY

Classification of Patents By Weighting Summary of Use Machine Learning Approach

Patents are the exclusive rights of inventors to inventions in the field of technology when they create their own inventions or authorize others to implement their inventions. An invention is an inventor's idea that leads to activities aimed at solving specific problems in the field of technology, either in the form of a product or how to treat or refine it for product development. The issue of the complexity of patent proposals for inventors becomes a very important discussion through the process steps that define the patent engineering domain, the foundation of the invention. invention, patent abstract, complete description of the invention, claim, reference search, and abstract. The problem of redundancy of patent abstracts is to avoid duplication or repeated storage of the same information on the IPC (International Patent Classification) label or patent group in several patent documents. over the Internet, thus storing the same information in multiple layers or tags. The aim of this study is to generate a dataset of Indonesian patents for a machine learning model to predict classes or labels from the input data based on the information provided by the patent data. IPC (International Patent Classification) of the previous training course. The primary objective is to identify patterns or relationships in the data that can be used to classify proposed new patent data into one of several predefined IPC-related patent data types. . The model's prediction results can be one, two, or three, depending on which category is most similar to the generated model. Its contribution is to create a predictive model for new patent documents to obtain meaningful patent citations based on patent document types. Classification can help decision-making by providing predictions of the most likely type or class of inventions based on new input data documents. the study is to obtain documents that can be used as a reference to measure and evaluate the taxonomy pattern applied in Indonesian patent documents; can save time and effort in tasks such as sorting patent similarity data, checking patent documents .

Key word: Patent, Patent Classification, IPC (International Patent Classification)

DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN INTEGRITAS	iv
KATA PENGANTAR	v
RINGKASAN	vii
SUMMARY	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	4
1.4. Batasan Masalah	5
1.5. Kontribusi Penelitian.....	5
1.6. Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	7
2.1. Teks Mining	7
2.2. International Patent Classification (IPC)	8
2.3. Text Summarization	13
2.4. Pembelajaran Mesin (<i>Machine Learning</i>).....	17
2.5. Pengambilan Kembali (<i>Information Retrieval</i>).....	22
2.6. Literatur Review	23
2.7. Penelitian Terkait	34
2.8. Evaluasi Kinerja	42
2.9. Hipotesis.....	44
BAB III METODOLOGI PENELITIAN	45
3.1. Kerangka Kerja Penelitian	45
3.2. Langkah Penelitian.....	47
3.3. Rancangan Model	49
BAB IV HASIL DAN PEMBAHASAN	56
4.1. Dataset.....	56
4.2. Kerangka Kerja	58
4.3. Tahapan Preprocessing	59
4.4. Tahapan Klasifikasi.....	60
4.5. Analisis Hasil	77
4.6. Modul Pengumpulan dan Pemrosesan Data.....	77
4.7. Hasil Multi Label Klasifikasi Menggunakan Supervised Learning.....	78
4.8. Pengujian Algoritma Machine Learning.....	82
4.9. Pengujian Model	84

4.10. Evaluasi Model	96
4.11. Analisis Kinerja Model Berdasarkan F1-score	108
4.12. One-vs-Rest (OvR) Classification Model	109
BAB V KESIMPULAN DAN SARAN	122
5.1. Kesimpulan	122
5.2. Saran.....	123
DAFTAR PUSTAKA.....	124

DAFTAR TABEL

Tabel 2.1 Perbandingan metode algoritma klasifikasi dan pengelompokan	28
Tabel 2.2 Penelitian Klasifikasi dan ringkasan teks	34
Tabel 4.1 Dataset paten	56
Tabel 4.2 Sample Clean Dataset Paten Indonesia	56
Tabel 4.3 Sample Paten Indonesia.....	57
Tabel 4.4 Hasil analisis tag (kategori) data paten.....	69
Tabel 4.5 Hasil penilaian akurasi klasifikasi berdasarkan label klasifikasi IPC	70
Tabel 4.6 Hasil klasifikasi kelas paten IPC	72
Tabel 4.7 Menjelaskan 10 (sepuluh) label IPC teratas :.....	78
Tabel 4.8 Menjelaskan Label IPC MAX :.....	78
Tabel 4.9 Menjelaskan Label IPC MIN :	79
Tabel 4.10 Menghapus Tabel Label Kecuali 10 Top Label Utama :	79
Tabel 4.11 Tabel Jumlah Dataset Label	80
Tabel 4.12 Pengujian Menghitung Akurasi 10 (sepuluh) label teratas	83
Tabel 4. 13 Pengujian Menghitung Akurasi AVG (Rata-rata).....	83
Tabel 4.14 Pengujian dengan beberapa model Algoritma Machine Learning	84
Tabel 4.15 Dokumen Paten	86
Tabel 4.16 Fitur Dokumen Paten.....	89
Tabel 4.17 Kelas Logit Prediksi	93
Tabel 4.18 Hasil Paten Klasifikasi IPC	94
Tabel 4.19 Dataset Quantity	97
Tabel 4.20 Jumlah Quantity Label	99
Tabel 4.21 Perhitungan Jumlah Quantity label	100
Tabel 4.22 Tiga kelas (A, B, C) dan hasil prediksi model	102
Tabel 4.23 Dokumen Paten	114
Tabel 4. 24 Dokumen bigram Pasangan.....	114
Tabel 4.25 Dokumen Bigram Term Frekuensi.....	114

DAFTAR GAMBAR

Gambar 2.1 Contoh cara kerja peringkasan berbasis abstraksi [48].	14
Gambar 2.2 BERT-Rewriter.....	16
Gambar 2.3 BART-JointSR.....	16
Gambar 2.4 Contoh artikel dan ringkasan dari Liputan6.	18
Gambar 2.5 Metode Taxonomi Paten Retrieval	23
Gambar 2.6 Diagram Tabel Review Penelitian Sebelumnya	32
Gambar 3.1 Kerangka kerja penelitian.....	47
Gambar 3.2 Flowchart Sistem Kerja Klasifikasi Dokumen Paten	48
Gambar 3.3 Model Klasifikasi Paten.....	51
Gambar 4.1 Kerangka Kerja Model	58
Gambar 4.2 Preprocessing Steps	59
Gambar 4.3 Classification Model.....	60
Gambar 4.4 Hasil Klasifikasi Label Paten IPC	69
Gambar 4.5 Grafik Hasil Perhitungan Nilai Akurasi Klasifikasi Paten Setiap Label Klas.....	71
Gambar 4.6 Threshold F1 Score.....	74
Gambar 4.7 F1 Score IPC Class	75
Gambar 4.8 Distribusi IPC Paten Mutibahasa.....	76
Gambar 4.9 Visualisasi Kelompok Dataset Paten All Text	80
Gambar 4.10 Visualisasi Stopword Kelompok Data Label IPC	81
Gambar 4.11 Grafik hasil Kemunculan kata dalam dokumen	81
Gambar 4.12 Hasil Prediksi kelas F1-Score.....	103
Gambar 4.13 Grafik Hasil Macro F1 Score.....	107
Gambar 4.14 Grafik Hasil F1 Score Klas IPC	108

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam Peraturan Menteri Hukum dan Hak Asasi Manusia Republik Indonesia Nomor 38 Tahun 2018 menyatakan Paten adalah hak eksklusif yang diberikan oleh negara kepada inventor atas hasil invensinya di bidang teknologi untuk jangka waktu tertentu melaksanakan sendiri invensi tersebut atau memberikan persetujuan kepada pihak lain untuk melaksanakannya [1]. Selanjutnya menurut [2], paten adalah jenis kekayaan intelektual dengan kepemilikan dan hak monopoli yang dapat diakses publik. Dokumen paten didaftarkan dan diterbitkan oleh pemerintah atau organisasi internasional . Dalam pengajuan paten dituangkan dalam bentuk invensi dari ide calon inventor. Inventor adalah seorang atau beberapa orang yang secara bersama- sama melaksanakan ide yang dituangkan ke dalam kegiatan yang menghasilkan Invensi. Invensi merupakan ide-ide baru yang berada di pusat transformasi masyarakat adalah penemuan secara historis dilindungi oleh sistem hukum kekayaan intelektual di mana paten berada [3]. Oleh karena itu diperkuat [4], tertolaknya pemohon paten dalam menyusun abstrak sering sekali tidak mencamtumkan kedalaman isi deskripsi dan klaim. Dokumen paten memiliki format yang terdefinisi dengan baik termasuk halaman depan, deskripsi, klaim, dan gambar.

Bukan hal mudah bagi pemohon invensi paten untuk menemukan informasi yang relevan untuk mengatasi masalah penulisan deskripsi paten. Salah satu cara dengan peringkasan dokumen otomatis. Representasi dokumen yang gagal dikarenakan tumpang tindih kata. Dengan metode penyisipan kata pembobotan untuk menghitung nilai matriks. Skema pembobotan baru dievaluasi pada tugas peringkasan dokumen, untuk mengekstrak informasi penting dari dokumen sangat bergantung pada fitur rekayasa manusia [5]. Selanjutnya [6], dengan pesatnya pertumbuhan aplikasi paten telah menjadi masalah akurasi rendah dan memiliki kemampuan generalisasi yang buruk pada rekayasa fitur dengan metode pembelajaran mesin SVM klasifikasi paten otomatis.

Disisi lain seleksi fitur bertujuan untuk memilih subset fitur diskriminan merupakan strategi yang efektif untuk menangani masalah HC (*Hierarchy Classification*) skala besar. Ini mempercepat proses pelatihan, mengurangi waktu prediksi dan meminimalkan persyaratan memori dengan mengompresi ukuran total vektor bobot model yang dipelajari [7]. Perlunya mengeksplorasi fitur tekstual dan implikasinya terhadap perilaku organisasi [8]. Menurut [9], salah satu fitur dokumen teks bahasa Indonesia akan diekstraksi menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) lalu diklasifikasikan menggunakan algoritma K-Means.

Sesuai penjelasan [10], bahasa Indonesia memiliki populasi peringkat terbesar keempat didunia dengan jumlah penduduk Indonesia lebih dari 250 juta penduduk saat ini. Walaupun mempunyai tingkat populasi bahasa yang sangat besar tetapi bahasa Indonesia belum terwakilkan dalam NLP (*Natural Language Processing*) sehingga menyebabkan minimnya dataset besar pada tugas parsing penjumlahan serta klasifikasi teks. Sistem klasifikasi paten banyak digunakan dalam analisis inovasi. Sistem klasifikasi paten saat ini umumnya bergantung pada teknologi atau berbasis TRIZ (*Teoriya Resheniya Izobretatelskikh Zadatch*) [11]. Sistem klasifikasi yang bergantung pada teknologi seperti IPC dan UPC biasanya dikembangkan oleh kantor paten di berbagai negara terutama untuk tujuan penelitian paten dan aplikasi paten, sedangkan yang terakhir ditujukan untuk pengguna TRIZ dan analisis hingga 40 kelas. Klasifikasi statis ini terlalu rumit dan terlalu umum untuk memenuhi kebutuhan klasifikasi paten yang komprehensif di area atau organisasi teknologi tertentu. Distribusi IPC juga membantu mengelompokkan paten berdasarkan kesamaan subjek, yang dapat membantu dalam analisis pemetaan teknologi atau penelitian literature[12].

Dalam melakukan drafting paten masalah “Sulitnya proses pencarian, mengakses database yang sesuai, sulitnya pencarian property dan pembobotan, sulitnya mengintegrasikan pencarian kembali sehingga bermakna dan sulitnya struktur taksonomi IPC yang rumit didasarkan per sub group”. IPC adalah standar **klasifikasi paten yang digunakan di seluruh dunia di lebih dari 100 negara**[13]. Paten (IPC) menggunakan pengklasifikasi yang dimodifikasi model untuk mengoptimalkan kelas IPC dan kode grup serta hasil klasifikasi paten

dengan literatur akademik yang relevan dan melakukan analisis pemilihan fitur untuk klasifikasi paten otomatis[14].

Klasifikasi multilabel adalah masalah klasifikasi di mana tidak hanya satu, tetapi beberapa kelas dapat ditugaskan ke sebuah *instance*. Ada dua jenis masalah yang sering dibahas dalam klasifikasi multilabel, yaitu datar dan hierarkis. Pada klasifikasi multilabel yang tanpa struktur tingkatan antar label (*Flat Multilabel Classification*), setiap sampel data dapat memiliki lebih dari satu kemungkinan label, sedangkan dalam klasifikasi multilabel hierarkis (*Hierarchical Multilabel Classification*), hierarki label digunakan untuk mengatur hubungan antara label yang ada [15].

Dalam konteks klasifikasi paten, beberapa bagian dari dokumen paten dapat memuat informasi yang tidak relevan atau informasi yang diulang (*redundan*) mengulangi pernyataan yang sama dalam dokumen lain atau mengulang kata-kata yang mirip di bagian dokumen yang berbeda [6]. Penelitian Zhang, mengembangkan pengklasifikasi biner dengan memanfaatkan deep learning dan teknik NLP (*Natural Language Processing*) untuk secara otomatis mengidentifikasi paten untuk mengubah data tekstual paten menjadi vektor numerik. Kemudian dikembangkan model *supervised deep learning* untuk mempelajari hubungan antara vektor masukan dan keluaran [16].

Algoritma klasifikasi teks diperlukan untuk model prediksi. Tujuan klasifikasi dimana **data yang diolah belum diketahui kategorinya** [17] Algoritma klasifikasi yang berbeda seperti Naive Bayes, SVM, C4.5 dan kNN diterapkan pada korpus Reuters-21578 dan Ohsumed[18]. Untuk mengetahui kesamaan dokumen paten menggunakan algoritma Cosine Similarity[17].

Algoritma TF-IDF (*term frequency-inverse document frequency*)[19],[20],[21]. Dari beberapa penelitian diatas penggunaan algoritma teknik klasifikasi dengan **permasalahan klasifikasi teks set data yang tidak seimbang** [18]. Untuk mencari Akurasi dan nilai rata-rata F1[11] Pemilihan metode dan algoritma akan sangat berpengaruh pada hasil yang diharapkan. Metode *Support Vektor Machine* (SVM)[22], [23],[24],[25],[26] sangat sesuai dengan beberapa kategori atau klas yang tumpang tindih satu dengan menggunakan *machine learning* untuk mengekstrak fitur yang berguna dari dokumen. Penelitian[27] menggunakan model

stacking untuk klasifikasi paten dengan pengukuran *Micro averaging* untuk masalah klasifikasi multi-kelas atau multi-label berguna menyelesaikan masalah **ketidak seimbangan kategori atau kelas** yang signifikan dalam kumpulan data paten dengan bahasa Rusia dan Inggris.

Berdasarkan hasil studi literatur, metode yang digunakan untuk **mengatasi redudansi dan duplikasi dokumen paten yang sering muncul berulang** kali kategori atau kelas yang sama pada paten *IPC (International Patent Classification)*. Penelitian ini mengusulkan solusi klasifikasi dokumen paten DJKI ke dalam kategori atau kelas yang berbeda menggunakan algoritma klasifikasi dengan pembobotan ringkasan yang benar. Kontribusi penelitian ini adalah membantu dalam pengambilan keputusan dengan memberikan prediksi tentang kelas atau kategori paten yang paling relevan berdasarkan dokumen data baru masukan, dapat dijadikan acuan untuk mengukur dan mengevaluasi model klasifikasi yang diterapkan dalam dokumen paten Indonesia.

1.2. Rumusan Masalah

Pertanyaan penelitian yang dirumuskan dari latar belakang permasalahan adalah sebagai berikut:

1. **Bagaimana mengukur kinerja model dalam menangani dataset bilingual pada usulan paten baru untuk melakukan prediksi kode IPC?**
2. Bagaimana membangun model klasifikasi dokumen paten untuk memprediksi usulan paten yang baru masuk dikelompok atau kategori dokumen paten yang mana ?
3. Bagaimana membangun framework pembobotan text summarization pada fitur abstraksi paten yang teruji nilai validasi untuk mengatasi redudansi dan meningkatkan akurasi model pada submit usulan paten ?

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu:

1. Melakukan pengembangan model klasifikasi teks mining dokumen paten Indonesia yang menguji bobot dan nilai indeks dari dokumen paten.

2. Menggunakan dan menguji model yang dibuat untuk mengklasifikasikan data baru usulan paten ke dalam banyak kategori atau kelas paten IPC yang relevan.
3. Meningkatkan kinerja evaluasi klasifikasi paten multilabel setiap klas IPC dapat di terapkan pada submit usulan paten.

1.4. Batasan Masalah

Dengan peningkatan kompleksitas dan volume pengajuan paten maka pada penelitian dibatasi pada :

1. Basis data paten yang digunakan dalam penelitian ini adalah **Google Patents** (GP) dan Patents Indonesia (pdki).
2. Dokumen data paten yang digunakan fitur ID_paten, Judul_Paten, Abstrak dan IPC.
3. Penelitian hanya membahas klasifikasi model untuk memberikan keseimbangan antara kinerja model dan efisiensi komputasi.

1.5. Kontribusi Penelitian

Penelitian ini memberikan kontribusi dalam pengembangan sistem klasifikasi paten otomatis yang berpotensi meningkatkan efisiensi dalam proses klasifikasi paten. Dengan penyempurnaan lebih lanjut, metodologi yang dikembangkan dalam penelitian ini dapat menjadi alat pendukung yang berharga dalam proses klasifikasi paten, memfasilitasi manajemen kekayaan intelektual yang lebih efektif. Penelitian ini membuka jalan untuk pengembangan lebih lanjut dalam bidang klasifikasi paten otomatis, dengan tujuan meningkatkan efisiensi dan akurasi dalam sistem paten internasional.

Aspek keterbaruan (*novelty*) penelitian ini adalah **penanganan dataset bilingual**, dokumen paten dari sumber Indonesia dan Inggris. Kemampuan model untuk berkinerja cukup baik pada dataset gabungan ini menunjukkan **potensinya dalam menangani tugas klasifikasi paten multibahasa, yang sangat relevan dalam konteks sistem paten internasional.**

1.6. Sistematika Penulisan

Sistematika penulisan disertasi terdiri dari empat bab sebagai berikut: Bab 1 merupakan bagian pendahuluan yang menguraikan latar belakang, perumusan masalah, ,

tujuan, batasan masalah, kontribusi penelitian, dan sistematika penulis. Bab 2 Menjelaskan tinjauan pustaka yang menjadi dasar penelitian, tingkat kesiapan teknologi dan roadmap penelitian dan evaluasi kinerja. Bab 3 menguraikan metodologi penelitian. Bab 4 yang menyajikan hasil dan pembahasan yang diharapkan pada penelitian ini, dan Bab 5 membahas kesimpulan dan saran.

DAFTAR PUSTAKA

- [1] Indonesian Government, “Act Number 13 / 2016 about Patent,” no. 13, 2016, [Online]. Available: http://www.dgip.go.id/images/ki-images/pdf-files/uu_pp1/UU-nomor-13-tahun-2016-tentang-paten.pdf
- [2] A. J. C. Trappey, C. V. Trappey, J. L. Wu, and J. W. C. Wang, “Intelligent compilation of patent summaries using machine learning and natural language processing techniques,” *Adv. Eng. Informatics*, vol. 43, no. November 2018, p. 101027, 2020, doi: 10.1016/j.aei.2019.101027.
- [3] G. Chimuka, “Impact of artificial intelligence on patent law. Towards a new analytical framework – [the Multi-Level Model],” *World Pat. Inf.*, vol. 59, no. June 2018, p. 101926, 2019, doi: 10.1016/j.wpi.2019.101926.
- [4] A. Abbas, L. Zhang, and S. U. Khan, “A literature review on the state-of-the-art in patent analysis,” *World Pat. Inf.*, 2014, doi: 10.1016/j.wpi.2013.12.006.
- [5] C. S. Yadav, “Automatic Text Document Summarization using Semantic-based Analysis,” 2018, [Online]. Available: <http://arxiv.org/abs/1811.06567>
- [6] H. Zhu, C. He, Y. Fang, B. Ge, M. Xing, and W. Xiao, “Patent automatic classification based on symmetric hierarchical convolution neural network,” *Symmetry (Basel)*, vol. 12, no. 2, pp. 1–12, 2020, doi: 10.3390/sym12020186.
- [7] A. Naik and H. Rangwala, “Embedding feature selection for large-scale hierarchical classification,” *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016*, pp. 1212–1221, 2016, doi: 10.1109/BigData.2016.7840725.
- [8] L. Hickman, S. Thapa, and L. Tay, “Text Preprocessing for Text Mining in Organizational Research : Review and Recommendations,” pp. 1–33, 2020, doi: 10.1177/1094428120971683.
- [9] A. A. P. Ratna, N. A. Wulandari, A. Kaltsum, I. Ibrahim, and P. D. Purnamasari, “Answer categorization method using K-means for Indonesian language automatic short answer grading system based on latent semantic analysis,” *2019 16th Int. Conf. Qual. Res. QIR 2019 - Int. Symp. Electr.*

- Comput. Eng.*, no. 1, pp. 1–5, 2019, doi: 10.1109/QIR.2019.8897845.
- [10] F. Koto, J. H. Lau, and T. Baldwin, “Liputan6: A Large-scale Indonesian Dataset for Text Summarization,” no. 1, 2020, [Online]. Available: <http://arxiv.org/abs/2011.00679>
 - [11] F. Zhu, X. Wang, D. Zhu, and Y. Liu, “A Supervised Requirement-oriented Patent Classification Scheme Based on the Combination of Metadata and Citation Information,” *Int. J. Comput. Intell. Syst.*, vol. 8, no. 3, pp. 502–516, 2015, doi: 10.1080/18756891.2015.1023588.
 - [12] J. X. Zhu, M. Sun, S. X. Wei, and F. Y. Ye, “Characterizing patent big data upon IPC: a survey of triadic patent families and PCT applications,” *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00778-5.
 - [13] S. Li, J. Hu, Y. Cui, and J. Hu, “DeepPatent: patent classification with convolutional neural networks and word embedding,” *Scientometrics*, vol. 117, no. 2, pp. 721–744, 2018, doi: 10.1007/s11192-018-2905-5.
 - [14] C. Cassidy, “Parameter tuning Naïve Bayes for automatic patent classification,” *World Pat. Inf.*, vol. 61, no. May, p. 101968, 2020, doi: 10.1016/j.wpi.2020.101968.
 - [15] A. Melo and H. Paulheim, “Local and global feature selection for multilabel classification with binary relevance: An empirical comparison on flat and hierarchical problems,” *Artif. Intell. Rev.*, vol. 51, no. 1, pp. 33–60, 2019, doi: 10.1007/s10462-017-9556-4.
 - [16] B. Zhang, H. Li, Clyde ZheWu, G. Shen, X. Lin, and M. Li, “Screening patents of ICT in construction using deep learning and NLP techniques,” *Eng. Constr. Archit. Manag.*, vol. 27, no. 8, pp. 1891–1912, 2020, doi: 10.1108/ECAM-09-2019-0480.
 - [17] R. T. Wahyuni, D. Prastiyanto, and E. Suprptono, “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi,” vol. 9, no. 1, 2017.
 - [18] C. Anne, A. Mishra, M. T. Hoque, and S. Tu, “Multiclass patent document classification,” *Artif. Intell. Res.*, vol. 7, no. 1, p. 1, 2017, doi: 10.5430/air.v7n1p1.
 - [19] S. Qaiser and R. Ali, “Text Mining: Use of TF-IDF to Examine the

- Relevance of Words to Documents Text Mining : Use of TF-IDF to Examine the Relevance of Words to Documents,” no. July, 2018, doi: 10.5120/ijca2018917395.
- [20] M. Krallinger, O. Rabal, A. Lourenço, J. Oyarzabal, and A. Valencia, “Information retrieval and text mining technologies for chemistry,” *Chem. Rev.*, vol. 117, no. 12, pp. 7673–7761, 2017, doi: 10.1021/acs.chemrev.6b00851.
- [21] F. A. Nugroho, F. Septian, D. A. Pungkastyo, and J. Riyanto, “Penerapan Algoritma Cosine Similarity untuk Deteksi Kesamaan Konten pada Sistem Informasi Penelitian dan Pengabdian Kepada Masyarakat,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 529, 2021, doi: 10.32493/informatika.v5i4.7126.
- [22] B. Nugroho and M. Aritsugi, “Application of k-step random walk paths to graph kernel for automatic patent classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10647 LNCS, pp. 14–29, 2017, doi: 10.1007/978-3-319-70232-2_2.
- [23] J. Yun and Y. Geum, “Automated classification of patents: A topic modeling approach,” *Comput. Ind. Eng.*, vol. 147, p. 106636, 2020, doi: 10.1016/j.cie.2020.106636.
- [24] J. Hu, S. Li, Y. Yao, L. Yu, G. Yang, and J. Hu, “Patent keyword extraction algorithm based on distributed representation for patent classification,” *Entropy*, vol. 20, no. 2, 2018, doi: 10.3390/e20020104.
- [25] N. Dogru and A. Subasi, “Traffic accident detection using random forest classifier,” *2018 15th Learn. Technol. Conf. L T 2018*, pp. 40–45, 2018, doi: 10.1109/LT.2018.8368509.
- [26] M. Pal, “Random forest classifier for remote sensing classification,” *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005, doi: 10.1080/01431160412331269698.
- [27] V. V Yadrntsev and I. V Sochenkov, “The Hybrid Method for Accurate Patent Classification,” *Lobachevskii J. Math.*, 2019, [Online]. Available: <https://link.springer.com/article/10.1134/S1995080219110325>
- [28] C. C. Aggarwal and C. X. Zhai, *Mining text data*, vol. 9781461432. 2013. doi: 10.1007/978-1-4614-3223-4.

- [29] Rahul, S. Adhikar, and Monika, “NLP based Machine Learning Approaches for Text Summarization,” *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020*, no. Iccmc, pp. 535–538, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-00099.
- [30] A. Trappey, C. V. Trappey, and A. Hsieh, “An intelligent patent recommender adopting machine learning approach for natural language processing: A case study for smart machinery technology mining,” *Technol. Forecast. Soc. Change*, vol. 164, no. February 2020, p. 120511, 2021, doi: 10.1016/j.techfore.2020.120511.
- [31] P. Ampornphan and S. Tongngam, “Patent knowledge discovery using data analytics,” *ACM Int. Conf. Proceeding Ser.*, pp. 42–46, 2017, doi: 10.1145/3176653.3176721.
- [32] L. Xiao, G. Wang, and Y. Zuo, “Research on Patent Text Classification Based on Word2Vec and LSTM,” *Proc. - 2018 11th Int. Symp. Comput. Intell. Des. Isc. 2018*, vol. 1, pp. 71–74, 2018, doi: 10.1109/ISCID.2018.00023.
- [33] A. Kravets, N. Shumeiko, B. Lempert, N. Salnikova, and N. Shcherbakova, “‘Smart Queue’ Approach for new technical solutions discovery in patent applications,” *Commun. Comput. Inf. Sci.*, vol. 754, pp. 37–47, 2017, doi: 10.1007/978-3-319-65551-2_3.
- [34] J. Shin, S. Lee, and T. Wang, “Semantic Patent Analysis System Based on Big Data,” *Proc. - IEEE 11th Int. Conf. Semant. Comput. ICSC 2017*, pp. 284–285, 2017, doi: 10.1109/ICSC.2017.20.
- [35] M. Okamoto, Z. Shan, and R. Orihara, “Applying information extraction for patent structure analysis,” *SIGIR 2017 - Proc. 40th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 989–992, 2017, doi: 10.1145/3077136.3080698.
- [36] D. Seneviratne, S. Geva, G. Zuccon, G. Ferraro, and M. Meireles, “Linking patents to knowledge sources: A context matching technique using automatic patent classification,” *ACM Int. Conf. Proceeding Ser.*, pp. 10–13, 2018, doi: 10.1145/3291992.3291994.
- [37] Y. Kim, J. Lee, J. Kang, S. Park, S. Jun, and D. Jang, “Study on the prediction of patent hiding company using patent information analysis,”

- ACM Int. Conf. Proceeding Ser.*, pp. 123–126, 2019, doi: 10.1145/3357419.3357447.
- [38] V. Raghupathi, Y. Zhou, and W. Raghupathi, “Legal Decision Support: Exploring Big Data Analytics Approach to Modeling Pharma Patent Validity Cases,” *IEEE Access*, vol. 6, pp. 41518–41528, 2018, doi: 10.1109/ACCESS.2018.2859052.
- [39] F. Chiarello, A. Cimino, G. Fantoni, and F. Dell’Orletta, “Automatic users extraction from patents,” *World Pat. Inf.*, vol. 54, no. July, pp. 28–38, 2018, doi: 10.1016/j.wpi.2018.07.006.
- [40] E. Marttin and A. C. Derrien, “How to apply examiner search strategies in Espacenet. A case study,” *World Pat. Inf.*, vol. 54, pp. S33–S43, 2018, doi: 10.1016/j.wpi.2017.06.001.
- [41] K. Loveniers, “How to interpret EPO search reports,” *World Pat. Inf.*, vol. 54, pp. S23–S28, 2018, doi: 10.1016/j.wpi.2017.03.008.
- [42] C. S. Hoo, “Impacts of patent information on clustering in Derwent Innovation’s ThemeScape map,” *World Pat. Inf.*, vol. 63, no. February, p. 102001, 2020, doi: 10.1016/j.wpi.2020.102001.
- [43] M. Samandar Ali Eshtehardi and M. Goodarzi, “Improving the Iranian Industrial Property System (IPS): A comparative case study of Iran and South Korea,” *World Pat. Inf.*, vol. 63, no. April 2019, p. 101991, 2020, doi: 10.1016/j.wpi.2020.101991.
- [44] R. Jain, M. Tripathi, V. Agarwal, and J. Murthy, “Patent data analytics for technology benchmarking: R-based implementation,” *World Pat. Inf.*, vol. 60, no. April 2019, p. 101952, 2020, doi: 10.1016/j.wpi.2020.101952.
- [45] M. E. Leusin, J. Günther, B. Jindra, and M. G. Moehrle, “Patenting patterns in Artificial Intelligence: Identifying national and international breeding grounds,” *World Pat. Inf.*, vol. 62, no. August, 2020, doi: 10.1016/j.wpi.2020.101988.
- [46] N. Deng, S. Lin, C. Xiong, and D. Li, “A Clustering Algorithm of Four Character Medicine Effect Phrases in TCM Patents,” *Proc. 2018 IEEE 8th Int. Conf. Electron. Inf. Emerg. Commun. ICEIEC 2018*, pp. 198–201, 2018, doi: 10.1109/ICEIEC.2018.8473529.

- [47] M. Lupu, “Information retrieval, machine learning, and Natural Language Processing for intellectual property information,” *World Pat. Inf.*, vol. 49, pp. A1–A3, 2017, doi: 10.1016/j.wpi.2017.06.002.
- [48] E. Gustafsson and I. Technology, “Automatic Text Summarization of Patent Documents Automatic Text Summarization of Patent Documents,” 2020.
- [49] J. L. Neto, A. A. Freitas, and C. A. A. Kaestner, “Automatic text summarization using a machine learning approach,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2507, pp. 205–215, 2002, doi: 10.1007/3-540-36127-8_20.
- [50] A. Khurana and V. Bhatnagar, “Investigating Entropy for Extractive Document Summarization,” *Expert Syst. Appl.*, vol. 187, 2022, doi: 10.1016/j.eswa.2021.115820.
- [51] L. Abualigah, M. Q. Bashabsheh, H. Alabool, and M. Shehab, “Text Summarization: A Brief Review,” *Stud. Comput. Intell.*, vol. 874, no. January, pp. 1–15, 2020, doi: 10.1007/978-3-030-34614-0_1.
- [52] G. Bao and Y. Zhang, “A General Contextualized Rewriting Framework for Text Summarization,” vol. XX, no. X, pp. 1–12, 2022.
- [53] K. Munir and M. Sheraz Anjum, “The use of ontologies for effective knowledge modelling and information retrieval,” *Appl. Comput. Informatics*, vol. 14, no. 2, pp. 116–126, 2018, doi: 10.1016/j.aci.2017.07.003.
- [54] W. Shalaby and W. Zadrozny, “Patent retrieval: a literature review,” *Knowl. Inf. Syst.*, vol. 61, no. 2, pp. 631–660, 2019, doi: 10.1007/s10115-018-1322-7.
- [55] D. H. Lin, “A Comparison of Statistical Models for Information Retrieval in Patent Translation : A Probability Distribution Approach,” vol. 28, no. 1, 2016.
- [56] A. Abood and D. Feltenberger, “Automated patent landscaping,” *Artif. Intell. Law*, vol. 26, no. 2, pp. 103–125, 2018, doi: 10.1007/s10506-018-9222-4.
- [57] S. Q. Wu, C. C. Tsao, P. C. Chang, C. Y. Fan, M. H. Chen, and X. Zhang, “A study of patent analysis for stock price prediction,” *Proc. - 2017 4th Int. Conf. Inf. Sci. Control Eng. ICISCE 2017*, pp. 115–119, 2017, doi: 10.1109/ICISCE.2017.34.

- [58] D. Yin, K. Motohashi, and J. Dang, “Large-scale name disambiguation of Chinese patent inventors (1985–2016),” *Scientometrics*, vol. 122, no. 2, pp. 765–790, 2020, doi: 10.1007/s11192-019-03310-w.
- [59] D. T. Jo, “Optimizing index of news articles by table based version of k nearest neighbors,” *Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018*, pp. 1238–1241, 2018, doi: 10.1109/CSCI46756.2018.00239.
- [60] J. Maillo, J. Luengo, S. García, F. Herrera, and I. Triguero, “A preliminary study on Hybrid Spill-Tree Fuzzy k-Nearest Neighbors for big data classification,” *IEEE Int. Conf. Fuzzy Syst.*, vol. 2018-July, 2018, doi: 10.1109/FUZZ-IEEE.2018.8491595.
- [61] J. Maillo, S. Garcia, J. Luengo, F. Herrera, and I. Triguero, “Fast and Scalable Approaches to Accelerate the Fuzzy k-Nearest Neighbors Classifier for Big Data,” *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 874–886, 2020, doi: 10.1109/TFUZZ.2019.2936356.
- [62] C. Lee and G. Lee, “Technology opportunity analysis based on recombinant search: patent landscape analysis for idea generation,” *Scientometrics*, vol. 121, no. 2, pp. 603–632, 2019, doi: 10.1007/s11192-019-03224-7.
- [63] Y. Zhang, P. Yuan, and B. Yu, “Evaluating cities’ independent innovation capabilities based on patent using data analysis methods,” *Commun. Comput. Inf. Sci.*, vol. 727, pp. 696–707, 2017, doi: 10.1007/978-981-10-6385-5_58.
- [64] M. Metode *et al.*, “Pendeteksian Jenis Dan Kelas Aroma Dengan,” vol. 7, no. 3, pp. 15–25, 2003.
- [65] I. I. Leong, R. Intan, and L. W. Santoso, “Pembuatan Aplikasi Complaint Management System pada Universitas Kristen Petra dengan menggunakan Metode Support Vector Machine Multiclass One vs Rest,” *J. Infra*, 2020, [Online]. Available: <http://publication.petra.ac.id/index.php/teknik-informatika/article/view/10500>
- [66] A. I. Chandra, Y. Yulia, and R. Adipranata, “Aplikasi Penentu Subyek Skripsi Menggunakan Metode Support Vector Machine,” *J. Infra*, 2020, [Online]. Available: <http://publication.petra.ac.id/index.php/teknik-informatika/article/view/10518>
- [67] A. B. Raharjo and M. Quafafou, “Penggabungan Keputusan Pada Klasifikasi

- Multi-Label,” *JUTI J. Ilm. Teknol. Inf.*, vol. 13, no. 1, p. 12, 2015, doi: 10.12962/j24068535.v13i1.a384.
- [68] A. Hanafi, A. Adiwijaya, and W. Astuti, “Klasifikasi Multi Label pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan Mutual Information dan k-Nearest Neighbor,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 357–364, 2020, doi: 10.32736/sisfokom.v9i3.980.
- [69] J. M. Moyano, E. L. Gibaja, S. Ventura, and A. Cano, “Speeding up classifier chains in multi-label classification,” *IoT BDS 2019 - Proc. 4th Int. Conf. Internet Things, Big Data Secur.*, no. IoT BDS, pp. 29–37, 2019, doi: 10.5220/0007614200290037.
- [70] R. Senge, J. J. Del Coz, and E. Hüllermeier, “Rectifying classifier chains for multi-label classification,” *LWA 2013 - Lernen, Wissen Adapt. Work. Proc.*, pp. 151–158, 2013.
- [71] M. Pushpa and S. Karpagavalli, “Multi-label Classification: Problem Transformation methods in Tamil Phoneme classification,” *Procedia Comput. Sci.*, vol. 115, pp. 572–579, 2017, doi: 10.1016/j.procs.2017.09.116.
- [72] W. Shalaby and W. Zadrozny, “Toward an interactive patent retrieval framework based on distributed representations,” *41st Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR 2018*, vol. 10, pp. 957–960, 2018, doi: 10.1145/3209978.3210106.
- [73] Y. Vernanda, M. B. Kristanda, and S. Hansun, “Indonesian language email spam detection using n-gram and naïve bayes algorithm,” *Bull. Electr. Eng. Informatics*, vol. 9, no. 5, pp. 2012–2019, 2020, doi: 10.11591/eei.v9i5.2444.
- [74] M. H. Ramadhan, V. I. Malik, and T. Sjafrizal, “Artificial neural network approach for technology life cycle construction on patent data,” *2018 5th Int. Conf. Ind. Eng. Appl. ICIEA 2018*, pp. 499–503, 2018, doi: 10.1109/IEA.2018.8387152.
- [75] J. Liu *et al.*, “Data Mining and Information Retrieval in the 21st century: A bibliographic review,” *Comput. Sci. Rev.*, vol. 34, 2019, doi: 10.1016/j.cosrev.2019.100193.
- [76] J. Ramírez *et al.*, “Ensemble of random forests One vs. Rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature

- selection and partial least squares,” *J. Neurosci. Methods*, vol. 302, pp. 47–57, 2018, doi: 10.1016/j.jneumeth.2017.12.005.
- [77] X. Che, D. Chen, and J. Mi, “Feature distribution-based label correlation in multi-label classification,” *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 6, pp. 1705–1719, 2021, doi: 10.1007/s13042-020-01268-3.
- [78] R. Wang, S. Kwong, X. Wang, and Y. Jia, “Active k-labelsets ensemble for multi-label classification,” *Pattern Recognit.*, vol. 109, 2021, doi: 10.1016/j.patcog.2020.107583.
- [79] E. Montañes, R. Senge, J. Barranquero, J. Ramón Quevedo, J. José Del Coz, and E. Hüllermeier, “Dependent binary relevance models for multi-label classification,” *Pattern Recognit.*, vol. 47, no. 3, pp. 1494–1508, 2014, doi: 10.1016/j.patcog.2013.09.029.
- [80] T. Tran and R. Kavuluru, “Supervised approaches to assign cooperative patent classification (CPC) codes to patents,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10682 LNAI, no. 9, pp. 22–34, 2017, doi: 10.1007/978-3-319-71928-3_3.
- [81] J. C. Gomez, “Analysis of the effect of data properties in automated patent classification,” *Scientometrics*, vol. 121, no. 3, pp. 1239–1268, 2019, doi: 10.1007/s11192-019-03246-1.
- [82] H. Sasaki and I. Sakata, “Identifying potential technological spin-offs using hierarchical information in international patent classification,” *Technovation*, vol. 100, no. September 2019, p. 102192, 2021, doi: 10.1016/j.technovation.2020.102192.
- [83] D. Kim and S. Kim, “Role and challenge of technology toward a smart sustainable city: topic modeling, classification, and time series analysis using information and communication technology patent data,” *Sustain. Cities Soc.*, vol. 82, no. April, p. 103888, 2022, doi: 10.1016/j.scs.2022.103888.
- [84] S. S. Tandel, “A Survey on Text Mining Techniques,” *2019 5th Int. Conf. Adv. Comput. Commun. Syst.*, no. Icaccs, pp. 1022–1026, 2019.
- [85] L. Zhang, L. Li, and T. Li, “Patent Mining,” *ACM SIGKDD Explor. Newsl.*, vol. 16, no. 2, pp. 1–19, 2015, doi: 10.1145/2783702.2783704.
- [86] K. Kurniawan and S. Louvan, “IndoSum: A New Benchmark Dataset for

- Indonesian Text Summarization,” *Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018*, pp. 215–220, 2019, doi: 10.1109/IALP.2018.8629109.
- [87] D. Tomar and S. Agarwal, “A multilabel approach using binary relevance and one-versus-rest least squares twin support vector machine for scene classification,” *Proc. - 2016 2nd Int. Conf. Comput. Intell. Commun. Technol. CICT 2016*, vol. 37, no. c, pp. 37–42, 2016, doi: 10.1109/CICT.2016.17.