

Classification of Book Collections Based on DDC 23 Using Text Mining Algorithm at UNIDA Gontor Library

Klasifikasi Koleksi Buku Berbasis DDC 23 Menggunakan Algoritma Teks Mining di Perpustakaan UNIDA Gontor

Muhammad Alwi¹, Oddy Virgantara Putra², Dihin Muriyatmoko³
{ muhammadalwi@unida.gontor.ac.id ¹, oddy@unida.gontor.ac.id ², dihin@unida.gontor.ac.id ³}

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor^{1,2,3}

Abstract. *The collection of books in a library is a means of information that has become the main actor as a supporter of the existence of a library. UNIDA Gontor library uses the 23rd edition of the Dewey Decimal Classification (DDC 23) classification system, as a reference for the classification numbering system for each book collection. However, in the classification numbering there is no automatic system that helps librarians in assigning classification numbering to each collection. So it is necessary to select a suitable model system to be applied to the automatic classification system. The data used in this study is in the form of blurb data on each collection of Indonesian public books in the UNIDA Gontor Library. In this study, four methods of text mining algorithms were applied to be tested and compared. The algorithm used in testing this research is Multinomial Nb, Logistic Regression, Random Forest, and Support Vector Classifier. From the test results, the highest accuracy results are the Support Vector Classifier algorithm of 72%, while the Logistic Regression algorithm is 69%, Random Forest 69%, and Multinomial Nb 59%. Further research is recommended to apply the support vector classifier algorithm into the UNIDA Gontor library information system.*

Keywords – classification; library; dewey decimal classification; text mining

Abstrak. *Koleksi buku dalam sebuah perpustakaan merupakan sarana informasi yang telah menjadi aktor utama sebagai pendukung keberadaan sebuah perpustakaan. Perpustakaan UNIDA Gontor menggunakan sistem klasifikasi Dewey Decimal Classification edisi 23 (DDC 23), sebagai acuan sistem penomoran klasifikasi pada tiap koleksi buku. Namun, dalam penomoran klasifikasi tersebut belum ada sistem otomatis yang membantu para pustakawan dalam menetapkan penomoran klasifikasi pada tiap koleksi. Maka diperlukannya pemilihan sebuah sistem model yang cocok untuk diterapkan kedalam sistem klasifikasi otomatis tersebut. Data yang digunakan pada penelitian ini adalah berupa data blurb (konten promosi) pada tiap koleksi buku umum berbahasa Indonesia yang ada di Perpustakaan UNIDA Gontor. Dalam penelitian ini diterapkannya empat metode algoritma teks mining untuk diuji dan dibandingkan. Algoritma yang digunakan dalam pengujian penelitian ini adalah Multinomial Nb, Logistic Regression, Random Forest, dan Support Vector Classifier. Dari hasil pengujian mendapatkan hasil akurasi tertinggi yaitu pada algoritma Support Vector Classifier sebesar 72%, sedangkan pada algoritma Logistic Regression 69%, Random Forest 69%, dan Multinomial Nb 59%. Penelitian selanjutnya disarankan untuk menerapkan algoritma Support Vector Classifier kedalam sistem informasi perpustakaan UNIDA Gontor.*

Kata Kunci – klasifikasi; perpustakaan; dewey decimal classification; teks mining

I. PENDAHULUAN

Koleksi buku dalam sebuah perpustakaan merupakan sarana informasi yang menjadi aktor utama sebagai pendukung layanan perpustakaan [1]. Koleksi buku dalam perpustakaan adalah buku yang dikumpulkan, diproses, dan disimpan di perpustakaan sehingga pemustaka dapat menemukannya Kembali dan dibaca berulang-ulang, oleh karena itu perpustakaan menjadi pusat sumber informasi serta pengetahuan. Dengan perpustakaan, seseorang dapat mempelajari berbagai ide yang di simpan dari berbagai bentuk dan format. Jenis koleksi yang beragam dan dengan jumlah yang banyak dikelola berdasarkan bentuk klasifikasi. Klasifikasi dalam perpustakaan adalah kegiatan pengelompokan dari berbagai koleksi berdasarkan subjek tertentu dengan ciri-ciri yang menyerupai. Adanya klasifikasi ini dapat memudahkan penataan berbagai jenis koleksi buku yang ada di perpustakaan, sehingga tersusun rapi dan sesuai dengan jenis subjeknya. Pengelompokan koleksi tersebut menggunakan sistem yang terinci dan sistematis. Adapun sistem klasifikasi yang umumnya digunakan di perpustakaan adalah *Dewey Decimal Classification* (DDC), *Library Congress Classification* (LCC) dan *Universal Decimal Classification* (UDC) [2], dari sekian jenis sistem klasifikasi tersebut, DDC merupakan jenis sistem klasifikasi yang paling banyak digunakan di dunia dan diterjemahkan ke dalam 30 bahasa.

Perpustakaan perguruan tinggi merupakan unit penunjang untuk melakukan Pendidikan, penelitian dan pengabdian masyarakat berdasarkan PP No.30 tahun 1990 pasal 34 PPT (Perpustakaan Perguruan Tinggi) [3]. Salah satu fungsi perpustakaan pada perguruan tinggi adalah sebagai sumber informasi pada sebuah perguruan tinggi, maka dari itu

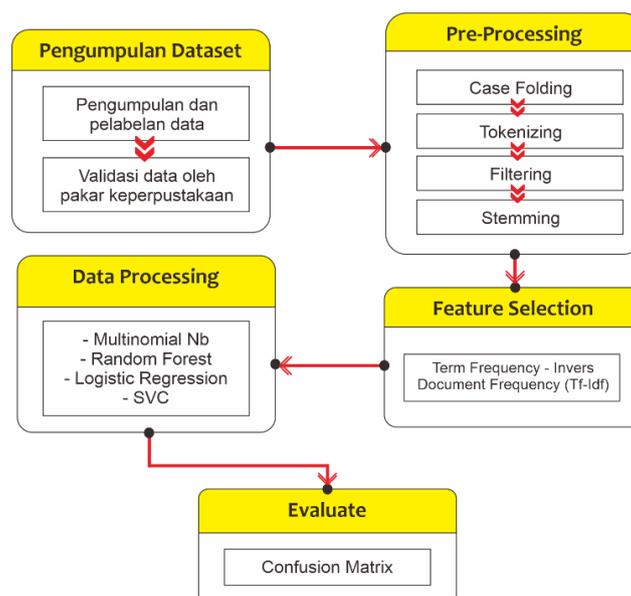
perpustakaan perguruan tinggi sering disebut dengan jantungnya perguruan tinggi, karena proses Pendidikan pada perguruan tinggi tidak terlepas dari penelitian dan pengembangan. Dengan adanya pernyataan diatas dapat disimpulkan bahwa keberadaan perpustakaan pada sebuah perguruan tinggi sangat penting adanya karena bukan hanya sebagai pusat sumber informasi pada perguruan tinggi, tapi juga sebagai pendukung kinerja pada sebuah perguruan tinggi.

Perkembangan perpustakaan dari tahun ketahun mengalami banyak peningkatan khususnya dibidang sistem teknologi yang dapat membantu penyusunan koleksi buku-buku di perpustakaan. Sistem klasifikasi koleksi buku otomatis dapat diterapkan dengan model algoritma teks mining telah banyak mengundang para peneliti untuk ikut andil dalam perkembangan teknologi perpustakaan, penelitian tentang penerapan model algoritma text mining pada sistem klasifikasi otomatis mulai banyak dilakukan oleh para peneliti sebelumnya, diantaranya : penelitian yang membahas tentang klasifikasi judul buku perpustakaan menggunakan metode K-NN oleh Muhammad Kadafi (2020), dalam penelitian tersebut melakukan pengujian penerapan algoritma K-NN untuk menentukan jenis buku dengan memanfaatkan data peminjaman buku dan disusun berdasarkan nomor klasifikasi, adapun penelitian tersebut mendapatkan akurasi 72.50% [4]. Lalu pada tahun setelahnya Esti Mulyani dkk (2021) melakukan penelitian mengenai pengaruh N-Gram terhadap klasifikasi buku dengan menggunakan ekstraksi dan seleksi fitur, penelitian tersebut melakukan pengujian metode seleksi fitur N-Gram (*Unigram, Bigram, Trigram*) dan Algoritma Multinomial naïve bayes untuk sistem klasifikasi otomatis dengan menggunakan 514 data judul buku berbahasa inggris yang disusun berdasarkan subjek pada DDC dan menghasilkan akurasi tertinggi pada metode *Unigram* sebesar 74,4% [5].

Perpustakaan UNIDA Gontor menggunakan *Senayan Library Management Sytem (SLiMS)* sebagai sistem automasi yang digunakan untuk mengelola koleksi buku umum maupun buku arab yang menjadi kebutuhan para mahasiswa. Perpustakaan UNIDA Gontor memiliki 8 cabang yang masing-masing terletak di tiap kampus cabang UNIDA Gontor. Saat ini perpustakaan kampus pusat maupun cabang menggunakan aplikasi Electronic – Dewey Decimal Classification (e-DDC) sebagai panduan dalam melakukan pengklasifikasian dan penomoran koleksi pada buku yang berbentuk aplikasi dan diimplementasikan kedalam sistem SLiMS, dimana para pustakawan mencari kelas atau subjek secara manual dengan membuka satu persatu pada tiap subjek untuk mengetahui nomor klasifikasi tersebut. Dengan melihat situasi dan kondisi yang terjadi pada pepustakaan, diperlukannya sistem klasifikasi otomatis untuk membantu para pustakawan dalam melakukan penomoran pada tiap koleksi buku tersebut. Namun, dalam membuat sistem klasifikasi otomatis tersebut, memerlukan model yang cocok untuk diterapkan kedalam sistem yang akan dibuat. Adapun metode yang digunakan dalam klasifikasi ini adalah algoritma teks mining yaitu *multinomial Nb, random forest, logistic regression, dan Support Vector Classifier*. Penelitian ini bertujuan untuk mencari model yang cocok dari pengujian empat algoritma teks mining tersebut.

II. METODE

Dalam melakukan pencarian model yang cocok pada penelitian ini, untuk diterapkan kedalam sistem klasifikasi koleksi buku umum otomatis memiliki beberapa tahapan yang dilakukan sebelum mendapatkan model yang cocok.



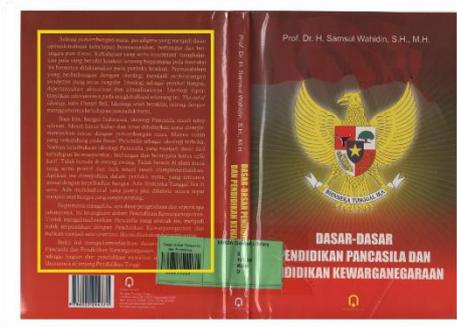
Gambar 1 Alur Penelitian

Alur penelitian yang tercantum pada **Gambar 1**, merupakan langkah-langkah yang dilakukan dalam melakukan pencarian model yang cocok untuk diterapkan kedalam sistem klasifikasi koleksi buku umum otomatis. Dimana dalam alur penelitian tersebut terdapat 5 langkah yakni, Pengumpulan dan validasi dataset, *Pre-Processing*, *Feature Selection*, *Data Processing*, dan *Evaluate/evaluasi*.

A. Pengumpulan Dataset

Pada tahap pengumpulan dataset terdapat 2 langkah yang dilakukan, yakni tahap pengumpulan dan pelabelan data, dan validasi data oleh pakar bidang keperpustakaan UNIDA Gontor.

- **Pengumpulan dan Pelabelan Data**



Gambar 2 Contoh Data Blurb

Data yang diolah dan diproses dalam penelitian ini adalah berupa kumpulan data blurb (konten promosi) pada tiap koleksi buku umum berbahasa Indonesia yang dikumpulkan satu persatu dan disusun berdasarkan subjek utama pada sistem klasifikasi DDC untuk diproses. Blurb merupakan konten promosi yang menceritakan sekilas tentang isi dalam buku tanpa menceritakan akhir cerita dari buku tersebut, dimana konten ini dibuat dengan semenarik mungkin, agar dapat menarik minat para pembaca. Beberapa koleksi buku meletakkan blurb disampul belakang buku atau cover belakang, namun sebagian buku juga tidak mempunyai blurb dan menggantinya dengan *sinopsis*, *review*, dan *resensi* bahkan sebagian koleksi buku tidak ada yang dicantumkan sama sekali pada sampul belakang buku. Adapun contoh data blurb terdapat pada **Gambar 2**. Pada proses pengambilan data blurb, peneliti menggunakan bantuan aplikasi *Google Lens*, *Google Chrome*, dan *Google Spreadsheet* untuk memudahkan dan mempercepat pengumpulan dataset.

Tabel 1 Kelas subjek dan Jumlah Data yang di proses

NO	Subjek	Jumlah Data
1	Karya Umum	130 Data
2	Filsafat dan Psikologi	130 Data
3	Agama	130 Data
4	Ilmu-ilmu Sosial	130 Data
5	Bahasa	105 Data
6	Ilmu-Ilmu Alam dan Matematika	130 Data
7	Teknologi dan Ilmu-Ilmu Terapan	130 Data
8	Geografi dan Sejarah	130 Data
TOTAL		1.015 Data

Adapun dari data yang terdapat pada **Tabel 1**, adalah data-data yang diolah dan diproses pada penelitian ini. Subjek adalah kelas yang menjadi target dan jumlah data adalah kumpulan data blurb yang disusun rapi dengan masing-masing subjek tertentu. 10 kelas DDC tidak digunakan sepenuhnya, karena terdapat beberapa subjek yang jumlah koleksinya yang tidak mencapai minimum target, diantaranya adalah subjek Kesenian, hiburan, dan olahraga dan subjek Kesastraan. Setiap kelas diisi dengan 130 data blurb, namun pada kelas Bahasa terdapat 105 data, karena jumlah koleksi yang terbatas akan tetapi data tersebut mencapai minimum target. Maka data yang diproses pada penelitian ini berjumlah **1.015** data blurb pada tiap koleksi buku umum berbahasa Indonesia.

- **Validasi Oleh Pakar Pustakawan**

Setelah terkumpulnya data beserta label, data divalidasi oleh pakar bidang keperpustakaan yang saat ini menjadi pustakawan Perpustakaan UNIDA Gontor. Adapun penyeleksian ini berguna untuk memvalidasi data yang telah disusun berdasarkan subjek-subjeknya dan menghindari terjadinya kesalahan dalam penyusunan blurb dimasing-masing subjeknya.

B. Pre-Processing

Merupakan tahap terpenting dalam teknik pembelajaran mesin [6], dimana pada tahap ini melakukan pembersihan data-data mentah yang diolah dan diproses, gunanya agar data yang telah dikumpulkan menjadi data yang lebih terstruktur. Pada penelitian ini, tahap *pre-processing* terbagi menjadi empat tahap, yakni tahap *Case Folding*, *Tokenizing*, *Stopword*, dan *Stemming*.

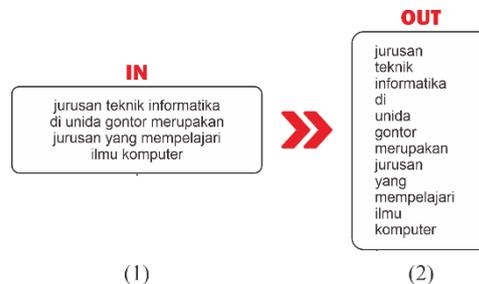
- **Case Folding**



Gambar 3 Proses Case Folding

Adalah proses yang salah satunya bertujuan untuk mengkonversi atau mengubah seluruh kalimat *Uppercase* (huruf kapital) dan *Propercuse* (huruf kapital diawal) menjadi *lowercase* (huruf kecil), proses *case folding* juga digunakan untuk menghapus angka, tanda baca, dan *whitespace*. Ilustrasi proses case folding dapat dilihat **Gambar 3**, dimana pada **Gambar 3 (1)** adalah ilustrasi kalimat sebelum di proses yang memiliki huruf *Uppercase* maupun *Propercuse* dan pada **Gambar 3 (2)** adalah ilustrasi kalimat yang sudah melalui proses *case folding*, yang semua hurufnya diubah menjadi *lowercase*.

- **Tokenizing**



Gambar 4 Proses Tokenizing

Pada proses ini, seluruh kalimat yang ada didalam dataset dipecah menjadi bagian-bagian yang disebut sebagai “token” seperti yang terjadi pada **Gambar 4**, dimana proses ini memecah sekumpulan kalimat kedalam satuan kata, yang memiliki tanda pemisah seperti space, enter, tanda titik (.), semicolon (;), titik dua (:), dan lain sebagainya. Ilustrasi pada **Gambar 4 (1)** merupakan kalimat sebelum diproses (hasil dari *case folding*) dan **Gambar 4 (2)** adalah kalimat yang sudah melalui proses tokenizing, dimana kalimat semua kalimat dipecah menjadi token-token.

- **Stopword**

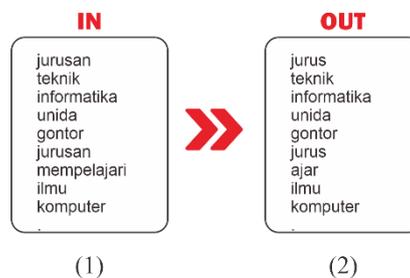
Proses ini merupakan proses lanjutan dari *tokenizing*, dimana proses ini mengambil kata-kata penting dari hasil token yang didapatkan dari proses *tokenizing*, proses *stopword* juga didefinisikan sebagai kata yang tidak mempunyai makna dan tidak terlalu berpengaruh terhadap sebuah kalimat [7]. Proses yang terjadi pada **Gambar 5** merupakan ilustrasi penerapan *stopword*, yakni menghilangkan kata seperti “di”, “yang”, “dari” dan lain sebagainya.



Gambar 5 Proses Stopword

Kata-kata yang termasuk dalam proses stopwords merupakan kata-kata yang sering muncul dalam setiap dokumen, sehingga tidak dapat digunakan sebagai pencari pada suatu dokumen. Gambar 5 (1) adalah token yang dihasilkan pada proses *tokenizing* sebelum diproses dan pada Gambar 5 (2) merupakan hasil dari proses *stopword*, dimana kata seperti “di”, atau “yang” dihilangkan.

- **Stemming**



Gambar 6 Proses Stemming

Proses ini adalah rangkaian terakhir dari tahap *pre-processing*, ilustrasi pada Gambar 6 menggambarkan transformasi kata menjadi kata dasarnya (*root word*) dengan menyesuaikan Bahasa yang digunakan [8], yakni dengan menghilangkan semua imbuhan kata dari hasil stopwords. Gambar 6 (1) merupakan bentuk keluaran yang dihasilkan pada proses *stopword* sebelum melalui proses *stemming* dan pada Gambar 6 (2) adalah hasil dari proses *stemming*, yang mengubah semua bentuk kata menjadi bentuk kata dasarnya (*root word*).

C. Feature Selection

Tahap *Feature Selection* merupakan tahap memilih atau menyeleksi fitur yang sangat dibutuhkan dari inti teks setelah melalui tahap *pre-processing*. Pada tahap ini digunakannya metode Tf-Idf, yakni metode yang meniadakan efek dari kata yang memiliki frekuensi tinggi dalam menentukan tingkat kepentingan pada dokumen [9]. Tahap *Feature Selection* mengubah kalimat menjadi vector numerik, metode *Term Frequency – Invers Document Frequency* (Tf-Idf). Metode ini merupakan metode yang sederhana dalam melakukan pembobotan kata pada sebuah kalimat, metode tf-idf berasal dari bidang *Information Retrieval* [10]. Tf-Idf terbagi menjadi 2, yakni *Term frequency* (Tf) dan *Invers Document Frequency* (IDF). Tf adalah hasil frekuensi dokumen berdasarkan kemunculan sebuah term pada dokumen yang di proses. Semakin sering kemunculan sebuah kata, maka semakin pula tinggi bobot pada dokumen tersebut. Lalu Idf adalah metode pembobotan kata yang dipadukan dengan Tf yang menghitung banyaknya istilah pada dokumen tersebut, Idf menunjukkan hubungan ketersediaan sebuah term pada seluruh dokumen [11], semakin sedikit jumlah dokumen yang mengandung *term*, maka nilai Idf semakin tinggi.

D. Data Processing

Pada tahap ini dilakukan proses *training* data untuk mengenali pola-pola dari data teks yang sudah melalui tahap *pre-processing* dan *feature selection*. Dalam tahap ini mulai diterapkannya empat algoritma yang diuji coba lalu dibandingkan berdasarkan tingkat akurasi, untuk mendapatkan model yang cocok yang akan diterapkan pada sistem klasifikasi koleksi buku umum otomatis. Adapun empat algoritma yang diuji coba adalah Algoritma *Multinomial NB*, *Random Forest*, *Logistic Regression*, dan *Support Vector Classifier*.

E. Evaluasi

Pada tahap ini menggunakan metode *Confusion Matrix* yang menampilkan hasil evaluasi model dengan menggunakan table matrix, yakni tabel yang menyatakan klasifikasi jumlah data uji yang benar dan salah [12], Evaluasi menggunakan metode *confusion matrix* menghasilkan nilai Akurasi, *Precision*, *Recall*, dan *F1 Score*. Nilai akurasi pada klasifikasi merupakan presentasi ketepatan *record* data yang diklasifikasikan secara benar setelah melakukan pengujian pada hasil klasifikasi. Nilai *Precision* adalah proposikasi yang diprediksi positif dan benar pada kumpulan data sebenarnya. Nilai *Recall* adalah proporsi positif yang diprediksi positif dan benar [13]. *Confusion matrix* menghasilkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan nilai *False Negative* (FN) dari model yang diuji. TP adalah jumlah *record* positif yang diprediksi positif, begitu sebaliknya jika jumlah *record* positif diprediksi negatif maka disebut dengan TN. Berbeda dengan FP yang memiliki jumlah *record false* namun diprediksi positif, begitu sebaliknya dengan nilai FN yang memiliki jumlah *record false* diprediksi juga dengan *false*.

III. HASIL DAN PEMBAHASAN

Data yang diproses berjumlah **1.015** data, lalu data tersebut dibagi menjadi 2 data yakni data *Train* dan data *Test*, dimana dibagi dengan skala **80 x 20** secara random, sehingga data train berjumlah **812 data** dan data test berjumlah **203 data**. Adapun pengujian ini menggunakan laptop *Toshiba Satellite AMD A4-5000 APU with Radeon(TM) HD Graphics 1.50 GHz, RAM 6 GB, 64-bit operating system*. Berikut adalah hasil pengujian yang dilakukan dengan menerapkan algoritma *Multinomial Nb*, *Random Forest*, *Logistic Regression*, dan *Support Vector Classifier*.

A. Hasil Pengujian Model Dengan Algoritma Multinomial Nb

Tabel 2 Hasil Pengujian Algoritma Multinomial Nb

Kelas	Precision	Recall	F1 Score
Karya Umum	74%	64%	68%
Filsafat dan Psikologi	100%	11%	20%
Agama	54%	29%	38%
Ilmu Sosial	82%	75%	78%
Bahasa	59%	80%	68%
Ilmu Alam dan Matematika	27%	91%	42%
Teknologi Ilmu Terapan	95%	60%	74%
Geografi dan Sejarah	83%	69%	75%
AVG	72%	60%	58%
Accuracy	59%		

Hasil pengujian dengan menggunakan model *Multinomial Nb* yang dipresentasikan pada **Tabel 2**, dimana hasil pengujian tersebut mendapatkan nilai akurasi sebesar 59%, *Precision* 72%, *Recall* 60%, dan *F1-Score* 58%. Adapun pada **Tabel 2** merupakan bentuk hasil dari perhitungan *confusion matrix*, dimana pada tabel tersebut memberikan informasi bahwa terdapat *False Positif* (FP) yang dominan tinggi pada subjek filsafat dan psikologi dan subjek agama yang diprediksi sebagai subjek ilmu alam dan matematika.

B. Hasil Pengujian Model Dengan Algoritma Random Forest

Tabel 3 Hasil Pengujian Algoritma Random Forest

Kelas	Precision	Recall	F1 Score
Karya Umum	60%	68%	64%
Filsafat dan Psikologi	80%	89%	84%
Agama	60%	25%	35%
Ilmu Sosial	59%	71%	64%
Bahasa	48%	75%	59%
Ilmu Alam dan Matematika	64%	82%	72%
Teknologi Ilmu Terapan	96%	74%	84%
Geografi dan Sejarah	83%	66%	73%
AVG	69%	69%	67%
Accuracy	69%		

Pada hasil pengujian dengan Algoritma *Random Forest* telah dijabarkan pada **Tabel 3**, dimana hasil pengujian tersebut mendapatkan hasil nilai akurasi sebesar 69%, *Precision* 69%, *Recall* 69%, dan *F1-Score* 67%. Adapun hasil akurasi yang didapatkan pada algoritma *random forest* lebih besar 10% dari tingkat akurasi yang didapatkan pada algoritma *Multinomial Nb*. **Tabel 3** mempresentasikan bentuk perhitungan dari hasil *confusion matrix* dan menghasilkan informasi bahwa hasil prediksi hampir tidak memiliki nilai kesalahan prediksi yang tinggi.

C. Hasil Pengujian Model Dengan Algoritma Logistic Regression

Tabel 4 Hasil Pengujian Algoritma Logistic Regression

Kelas	Precision	Recall	F1 Score
Karya Umum	58%	64%	61%
Filsafat dan Psikologi	80%	74%	77%
Agama	60%	25%	35%
Ilmu Sosial	63%	79%	70%
Bahasa	59%	85%	69%
Ilmu Alam dan Matematika	53%	82%	64%
Teknologi Ilmu Terapan	96%	69%	80%
Geografi dan Sejarah	85%	76%	80%
AVG	69%	69%	67%
Accuracy	69%		

Dengan melakukan pengujian terhadap Algoritma *Logistic Regression* yang dipresentasikan pada **Tabel 4**, yakni menghasilkan tingkat akurasi sebesar 69%, *Precision* 69%, *Recall* 69%, dan nilai *F1-Score* 67%. Adapun hasil dari nilai akurasi, *precision*, *recall*, dan nilai *f1-score* sama persis dengan algoritma *random forest*. Namun, yang menjadi perbedaan telah disajikan pada perhitungan hasil *confusion matrix* dalam **Tabel 4**, dimana terdapat nilai kesalahan prediksi lebih tinggi pada subjek agama yang diprediksi sebagai subjek ilmu alam dan matematika.

D. Hasil Pengujian Model Dengan Algoritma Support Vector Classifier

Tabel 5 Hasil Pengujian Algoritma Support Vector Classifier

Kelas	Precision	Recall	F1 Score
Karya Umum	60%	55%	57%
Filsafat dan Psikologi	80%	89%	84%
Agama	69%	38%	49%
Ilmu Sosial	63%	79%	70%
Bahasa	62%	75%	68%
Ilmu Alam dan Matematika	59%	77%	67%
Teknologi Ilmu Terapan	93%	71%	81%
Geografi dan Sejarah	83%	86%	85%
AVG	71%	71%	70%
Accuracy	72%		

Pada pengujian terakhir dengan menggunakan Algoritma *Support Vector Classifier* yang dipresentasikan pada **Tabel 5**, mendapatkan hasil nilai akurasi sebesar 72%, *Precision* 71%, *Recall* 71%, dan nilai *F1-Score* 70%. Adapun nilai akurasi yang didapatkan pada pengujian dengan algoritma *support vector classifier* lebih unggul dibandingkan dengan nilai pengujian dari 3 algoritma sebelumnya. Namun pada hasil perhitungan yang terdapat pada **Tabel 5**, terdapat sedikit kesalahan prediksi pada subjek agama yang diprediksi sebagai subjek ilmu alam dan matematika.

E. Hasil Perbandingan Akurasi

Table 6 Data hasil akurasi pengujian algoritma

Multinomial Nb	Random Forest	Logistic Regression	Support Vector Classifier
59%	69%	69%	72%

Berdasarkan dari hasil pengujian yang telah dilakukan pada penelitian ini, dapat disimpulkan bahwa pengujian dengan empat algoritma untuk mencari model yang cocok dalam penerapan model pada sistem klasifikasi koleksi buku umum otomatis, menghasilkan tingkat akurasi pada masing-masing algoritma yang terdapat pada **Tabel 6**.

IV. KESIMPULAN

Berdasarkan nilai akurasi yang didapatkan dari pengujian algoritma, disimpulkan bahwa algoritma *support vector classifier* mendapatkan nilai akurasi tertinggi dari pada algoritma *Multinomial Nb*, *Random Forest* dan *Logistic Regression*. Sehingga model algoritma *support vector classifier* cocok untuk diterapkan pada sistem klasifikasi koleksi buku umum otomatis. Adapun pada penelitian selanjutnya disarankan untuk membuat sistem klasifikasi koleksi buku otomatis dengan Algoritma *support vector classifier* berbasis sistem klasifikasi DDC sebagai panduannya dan mengimplementasikannya kedalam sistem automasi informasi Perpustakaan UNIDA Gontor.

UCAPAN TERIMA KASIH

Penelitian ini didanai oleh Program Studi Teknik Informatika Universitas Darussalam Gontor, Ponorogo Indonesia. Penelitian ini didukung oleh pihak UPT Perpustakaan Universitas Darussalam Gontor Ponorogo Indonesia.

REFERENSI

- [1] T. Yuliani, "Analisis kebutuhan pemustaka pada kegiatan layanan pengembangan koleksi buku Perpustakaan IAIN Batusangkar," *Al-Kuttab J. Kaji. Perpustakaan, Inf. dan kearsipan*, vol. 2, no. 1, pp. 41–52, 2020, doi: 10.24952/ktb.v2i1.2328.
- [2] B. L. H. Harahap and J. Husna, "Penerapan Sistem Klasifikasi Mandala Di Perpustakaan Gelaran Indonesia Buku Yogyakarta," *J. Ilmu Perpust.*, vol. 7, no. 1, pp. 181–190, 2018, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/jip/article/view/22829>.
- [3] R. Pajaransyah, T. Darmanto, and D. E. Saputra, "Penambahan Fitur Buku Tamu pada Aplikasi Open Source Slims di Perpustakaan STMIK 'AMIK BANDUNG,'" *MIND J.*, vol. 3, no. 2, pp. 1–15, 2019, doi: 10.26760/mindjournal.v3i2.1-15.
- [4] M. Kadafi, "Klasifikasi Text Judul Buku Perpustakaan untuk Menentukan Kategori Buku Menggunakan K-Nearest Neighbor," *Syst. Inf. Syst. Informatics J.*, vol. 6, no. 2, pp. 47–53, 2020.
- [5] E. Mulyani, F. Pralienka, B. Muhamad, and K. A. Cahyanto, "Pengaruh N-Gram terhadap Klasifikasi Buku menggunakan Ekstraksi dan Seleksi Fitur pada Multinomial Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 264–272, 2021, doi: 10.30865/mib.v5i1.2672.
- [6] N. M. Samsudin, C. F. B. Mohd Foozy, N. Alias, P. Shamala, N. F. Othman, and W. I. S. Wan Din, "Youtube spam detection framework using naïve bayes and logistic regression," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 3, pp. 1508–1517, 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.
- [7] O. V. Putra, F. M. Wasmanson, T. Harmini, and S. N. Utama, "Sundanese Twitter Dataset for Emotion Classification," *CENIM 2020 - Proceeding Int. Conf. Comput. Eng. Network, Intell. Multimed. 2020*, no. Cenim 2020, pp. 391–395, 2020, doi: 10.1109/CENIM51130.2020.9297929.
- [8] J. Ipmawati, Kusriani, and E. Taufiq Luthfi, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," *Indones. J. Netw. Secur.*, vol. 6, no. 1, pp. 28–36, 2017.
- [9] A. Tanari, A. Handojo, and J. Andjarwirawan, "Aplikasi Pencarian Jurnal Ilmiah dengan Term Frequency-Inverse Document Frequency," *J. Infra*, vol. 7, no. 1, 2019.
- [10] S. W. U. Vitandy, A. A. Supianto, and F. A. Bachtiar, "Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency-Inverse Document Frequency dan Naïve Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 6082–6083, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5645>.
- [11] M. Sulehu, Juhar, W. Rimalia, and A. Iskandar, "Implementasi Metode Term Frequency-Inverse Document Frequency- Class Frequency untuk Peringkasan Berita Online," *Celeb. Eng. J.*, vol. 1, no. April, pp. 54–61, 2019, [Online]. Available: <https://journal.ildikti9.id/CEJ/article/view/294>.
- [12] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [13] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," *SENARIS 2020*, vol. 2, pp. 84–92, 2020.