Klasifikasi Genre Buku Berbasis Judul dan Sinopsis Menggunakan Metode Support Vector Machine

Ni Wayan Emmy Rosiana Dewi [1], I Made Suwija Putra[2], Dwi Putra Githa[3]
[1,2]Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana
Email: [1] emmyrosiana@unud.ac.id, [2] putrasuwija@unud.ac.id, [3]dwiputragitha@unud.ac.id

Abstract. *A major challenge in traditional library management is creating methods to categorize books according to their genres. This study suggests applying the Support Vector Machine (SVM) technique to classify books based on their titles and synopses. This research's dataset was taken from the CMU Book Summary Dataset., which includes various genres of books. The research results show that the SVM method with a linear kernel performs the task of classifying book genres better than the K-Nearest Neighbors (K-NN) method, as well as text preprocessing and feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF). SVM achieves the highest accuracy of 47.7% when 90 percent of the data is training and 10 percent is testing, while K-NN only achieves 31.53% when 80 percent of the data is training and 20 percent is testing.. The results show that SVM is better than K-NN in handling text classification based on titles and synopses. However, for further validation, additional research is needed with a larger dataset and in various languages.*
Keywords: *classification, book, title, synopsis, Support Vector Machine*

Abstrak. *Tantangan besar dalam manajemen perpustakaan tradisional adalah menciptakan metode untuk mengkategorikan buku sesuai dengan genrenya. Penelitian ini mengusulkan penggunaan metode Support Vector Machine (SVM) untuk mengklasifikasikan buku berdasarkan judul dan sinopsisnya. Dataset penelitian ini diambil dari CMU Book Summary Dataset, yang mencakup berbagai genre buku. Hasil penelitian menunjukkan bahwa metode SVM dengan kernel linier melakukan tugas klasifikasi genre buku dengan lebih baik dibandingkan metode K-Nearest Neighbors (K-NN), serta preprocessing teks dan ekstraksi fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Akurasi tertinggi SVM adalah 47,7% dalam skenario 90 persen data pelatihan dan 10 persen data pengujian, sedangkan K-NN hanya memiliki 31,53% dalam skenario 80 persen dan 20 persen data pelatihan dan pengujian. Hasil menunjukkan bahwa SVM lebih baik daripada K-NN dalam menangani klasifikasi teks berdasarkan judul dan sinopsis. Namun, untuk validasi lebih lanjut, penelitian lebih lanjut diperlukan dengan dataset yang lebih besar dan dalam berbagai bahasa.*
Kata Kunci: *klasifikasi, buku, judul, sinopsis, Support Vector Machine*

1. Pendahuluan

Teknologi informasi telah mengubah banyak hal, salah satunya dalam pengelolaan perpustakaan. Perpustakaan menyediakan berbagai koleksi buku untuk memenuhi kebutuhan informasi pengguna (Amalia & Yustanti, 2021). Proses klasifikasi buku berdasarkan genre adalah salah satu tantangan yang dihadapi oleh pustakawan. Klasifikasi ini penting untuk membantu pembaca menemukan buku yang sesuai dengan minat dan kebutuhan mereka. Namun, metode manual yang sering digunakan masih memiliki beberapa kelemahan, seperti subjektivitas pustakawan dalam menentukan genre, ketidakkonsistenan dalam penggolongan, dan waktu yang cukup lama untuk mengklasifikasikan koleksi buku yang terus bertambah (Winiarti et al., 2022).

Jumlah buku yang tersedia di perpustakaan semakin banyak dan beragam sementara sumber daya pustakawan terbatas. Proses klasifikasi yang bergantung pada penilaian manual menjadi tidak efisien, terutama ketika pustakawan harus membaca dan memahami judul dan sinopsis setiap buku sebelum menentukan genre yang tepat, sehingga diperlukan alat otomatis yang dapat mempercepat proses klasifikasi. Support Vector Machines (SVM) adalah salah satu metode klasifikasi yang paling umum dan dikenal memiliki tingkat akurasi yang tinggi dalam memproses data teks (Winiarti et al., 2022).

SVM menawarkan kinerja yang kuat dalam berbagai aplikasi klasifikasi multi-kelas, seperti pengkategorian teks dan pengenalan karakter tulisan tangan (Deng et al., 2012). Salah satu keunggulan SVM adalah kemampuannya untuk menghasilkan model klasifikasi yang baik meskipun dilatih dengan

jumlah data yang relatif sedikit. Hal ini sangat berguna dalam konteks perpustakaan, di mana tidak semua koleksi buku memiliki data pelatihan yang melimpah (Leli et al., 2023). Penerapan SVM tidak hanya meningkatkan pengalaman pengguna tetapi juga mendukung pengelolaan informasi di era digital saat ini (Monika & Furqon, 2018).

Penelitian sebelumnya mengenai teknik klasifikasi teks sudah banyak dilakukan oleh peneliti. Studi tentang penggunaan pembelajaran mesin dalam klasifikasi buku perpustakaan menemukan bahwa Support Vector Machine (SVM) berfungsi sebagai algoritma yang memahami pola dan karakteristik judul buku dan kemudian mengkategorikannya ke dalam Dewey Decimal Classification (DDC). Hasil penelitian menunjukkan bahwa kombinasi TF-IDF dan Word2Vec dapat digunakan untuk metode klasifikasi buku dengan SVM dengan akurasi tertinggi sebesar 73% (Cahyani & Saraswati, 2023).

Penelian terkait mengusulkan solusi optimasi representasi pengetahuan serta pemilihan fitur untuk meningkatkan klasifikasi buku dan menunjukkan algoritma klasifikasi yang sesuai. Beberapa percobaan telah dilakukan dan ditemukan bahwa Naïve Bayes dapat memberikan hasil prediksi terbaik. Akurasi dan performa Naïve Bayes dapat ditingkatkan dan mengungguli algoritme klasifikasi lainnya dengan menerapkan strategi yang sesuai dari pemilihan fitur, pemilihan tipe data, serta transformasi data (Nguyen & Do, 2020).

Penelitian ini mencoba mengatasi permasalahan mengenai kesulitan seorang pustakawan dalam menggolongkan buku yang ada di perpustakaan menggunakan metode SVM dan mencoba menerapkan ekstraksi fitur kata dengan pembobotan kata yang muncul pada judul dan sinopsis buku terserbut dimana data penelitian ini diperoleh langsung dari buku yang ada diperpustakaan, sehingga bisa memberikan manfaat bagi manajemen perpustakaan.

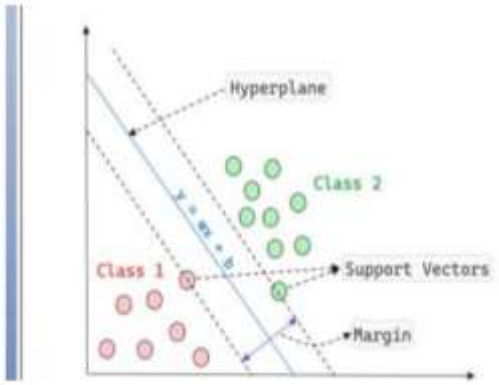
2. Tinjauan Pustaka

2.1 Machine Learning

Prinsip utama pembelajaran otomatis dan pengenalan pola terdiri dari pembelajaran mesin atau dikenal dengan machine learning, sebuah metode analisis data yang memungkinkan komputer membuat prediksi dan belajar dari data (Bernardes, 2024). Machine learning berfokus pada pengembangan program komputer yang dapat mengakses data dan menggunakannya untuk belajar sendiri. Proses pembelajaran dimulai dengan pengamatan atau data, seperti instruksi atau pengalaman langsung, untuk menemukan pola dalam data dan membuat keputusan yang lebih baik di masa depan berdasarkan contoh sebelumnya. Pembelajaran mesin dapat diklasifikasikan sebagai pembelajaran terbimbing (supervised learning) atau pembelajaran tidak terbimbing. Pelatihan model machine learning ini memungkinkan untuk menganalisis informasi dalam jumlah besar. Meskipun itu dapat bermanfaat dalam hal memberikan hasil lebih cepat dan lebih presisi dalam menjaga dan menciptakan peluang, serta dalam mengidentifikasi ancaman, ada kalanya itu juga akan memakan biaya dan dikendalikan dengan sumber daya yang tepat. Pembelajaran mesin (ML) menyediakan mekanisme bagi manusia untuk memproses data dalam jumlah besar, memperoleh wawasan tentang perilaku data, dan membuat keputusan yang lebih tepat di berbagai bidang (Injadat et al., 2021).

2.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. SVM adalah metode pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi dengan memanfaatkan prinsip minimisasi risiko struktural dan transformasi ruang fitur menggunakan kernel untuk memisahkan data dalam dimensi yang lebih tinggi (Cervantes et al., 2020). SVM dijabarkan sebagai algoritma pencarian hyperplane terbaik yang memisahkan dua kelas dari input ruang. Pada dasarnya, data dari dua kelas dapat dipisahkan oleh hyperplane ini. SVM menggunakan vektor pendukung (support vectors) dan margin untuk menentukan posisi optimal dari hyperplane tersebut (Sulistilawati et al., 2024).



Gambar 1. Hyperplane SVM

SVM berupaya untuk menemukan hyperplane yang paling optimal yang memisahkan kelas-kelas data dengan margin terbesar. Ada dua jenis SVM yang dapat digunakan, yaitu:

a. SVM Linier: Diterapkan ketika data dapat dipisahkan menggunakan hyperplane linier.
b. SVM Nonlinier: Diterapkan ketika data tidak dapat dipisahkan menggunakanhyperplane linier.

Dalam situasi ini, SVM memanfaatkan trik kernel untuk mentransformasikan data ke dalam ruang dimensi yang lebih tinggi di mana data dapat dipisahkan dengan cara linier.

3. Metode Penelitian

Dataset yang dipakai dalam penelitian ini merupakan dataset yang diperoleh dari CMU Book Summary Dataset yang disusun oleh David Bamman and Noah Smith dengan jumlah data sebanyak 11.599 data yang memiliki informasi lengkap berupa judul, sinopsis, dan genre yang valid (Bamman & Smith, 2013). Detail jumlah dataset dapat dilihat pada Table 1. Untuk mengevaluasi sensitivitas model terhadap variasi jumlah data latih, dilakukan lima skenario pembagian data training dan testing: 50:50, 60:40, 70:30, 80:20, dan 90:10. Setiap skenario, data dibagi agar distribusi label genre tetap seimbang. Model yang sama digunakan di tiap skenario untuk menjaga konsistensi hasil. Hasil eksperimen ini akan menunjukkan seberapa besar ketergantungan model terhadap ukuran data latih dalam klasifikasi genre berbasis judul dan sinopsis.

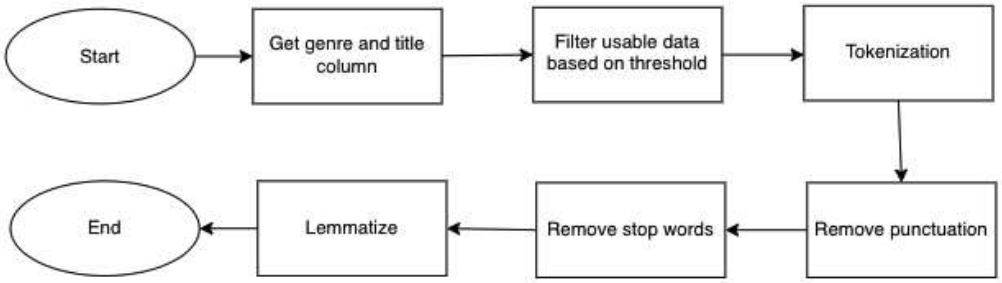
Tabel 1. Dataset Size

Genre	***Total Size***
Science Fiction	2551
Speculative fiction	1438
Children's literature	1152
Fiction	943
Novel	952
Mystery	734
Crime Fiction	627
Fantasy	624
Thriller	568
Young adult literature	321
Historical fiction	258
Historical novel	207
Alternate history	198
Non-fiction	166
Spy fiction	112
Autobiography	94
Horror	88
Gothic fiction	86
Autobiographical novel	76
Comic novel	73

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Satire	68
Romance novel	68
History	64
Adventure novel	60
War novel	54
Detective fiction	52

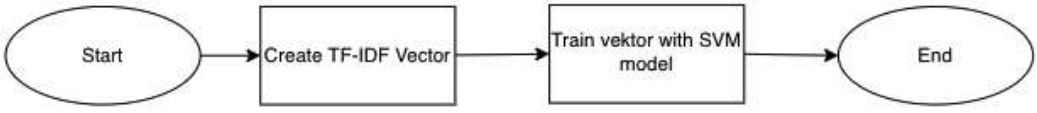
Penelitian ini dibagi menjadi dua tahapan. Tahapan pertama yaitu preprocessing berguna dalam mengekstraksi fitur yang terdapat pada sinopsis dan judul buku. Tahapan kedua adalah tahapan training yang berguna dalam melatih model SVM dalam mengklasifikasi genre dari sebuah buku berdasarkan data latih yang dihasilkan dari proses preprocessing. Detail tahapan preprocessing dapat dilihat pada Gambar 2.



Gambar 2. Alur Tahap Preprocessing

Alur tahap preprocessing dalam penelitian ini dimulai dengan tahapan pengambilan kolom yang dibutuhkan dengan memilih kolom yang akan digunakan untuk membuat fitur latihan, seperti genre dan judul. Selanjutnya filtering berdasarkan threshold dimana set data difilter dengan memperhatikan jumlah sampel dalam setiap genre. Hanya genre yang jumlah sampelnya lebih besar dari ambang yang dipertahankan yang dapat difilter. Threshold yang digunakan dalam penelitian ini sebanyak lima puluh dataset. Tokenisasi dilakukan pada kolom synopsis setelah proses filtering selesai. Tujuan tokenisasi adalah untuk memecah teks menjadi bagian kecil seperti token, contohnya kata atau frasa sehingga lebih mudah diproses di tingkat berikutnya. (Daniel & Martin, 2023).

Tanda baca dihapus setelah tokenisasi selesai. Tahapan ini bertujuan untuk menghilangkan karakter yang tidak diperlukan sehingga data menjadi lebih bersih dan siap digunakan untuk pembuatan fitur latih dalam klasifikasi genre buku. Selanjutnya dilakukan lemmatisasi yang merupakan proses mengubah kata menjadi bentuk dasar atau lema. Langkah ini berfungsi untuk meningkatkan akurasi model dan mengurangi ukuran kamus data dataset (Sanderson, 2010). Setelah preprocessing selesai, fitur data yang telah diproses digunakan untuk melakukan training model SVM. SVM adalah algoritma klasifikasi yang baik untuk menangani data teks karena dapat menangani data besar (Wang, 2022). Gambar 3 menunjukkan detail proses pelatihan model SVM.



Gambar 3. Alur Proses Training

Tahap training dimulai dengan membangun vector data latih menggunakan metode TF-IDF. Metode TF-IDF digunakan untuk menggambarkan teks dalam bentuk numerik dengan

mempertimbangkan frekuensi kata dalam dokumen dan relevansinya untuk korpus secara keseluruhan (Jalilifard et al., 2021). Selanjutnya, fitur latih TF-IDF digunakan dalam model SVM. Eksperimen dilakukan dengan berbagai skenario perubahan proporsi data pelatihan dan pengujian untuk mengevaluasi kinerja model. Tujuan skenario ini adalah untuk mengetahui bagaimana akurasi model dipengaruhi oleh perubahan ukuran dataset. Tabel 2 menunjukkan detail skenario yang digunakan.

Tabel 2. Skenario untuk Training dan Testing

Scenario	Genre Size	Training %	Test %	Training Size	Test Size
1	27	90%	10%	1,215	135
2	27	80%	20%	1,080	270
3	27	70%	30%	945	405
4	27	60%	40%	810	540
5	27	50%	50%	675	675

4. Hasil dan Pembahasan

4.1 Pengujian Sistem dengan Metode SVM

Metode SVM utama akan diterapkan untuk mengevaluasi kemampuannya dalam tugas mengklasifikasikan buku berdasarkan karakteristik judul dan sinopsis. Kernel linear dipilih karena fitur judul dan sinopsis akan dikonversi menjadi nilai TF-IDF, sehingga bentuk penyelesaiannya harus linear. Berdasarkan pengujian yang telah dilakukan, didapatkan hasil pengukuran kinerja seperti terlihat pada Tabel 3.

Tabel 3. Hasil Training dan Testing dengan Metode SVM

Training Dataset	Testing Dataset	Training Accuracy	Test Acuraccy
90%	10%	100%	47.7%
80%	20%	100%	45%
70%	30%	100%	36.6%
60%	40%	100%	42.7%
50%	50%	100%	36.15%

Ketika dataset terdiri dari 90% data pelatihan dan 10% data pengujian, model mencapai akurasi 47,7%. Namun, ketika dataset terdiri dari 50% data pelatihan dan 50% data pengujian, akurasi menurun menjadi 36,15%. Perbedaan ini menunjukkan bahwa lebih banyak data pelatihan cenderung membuat model lebih akurat. Ini disebabkan oleh fakta bahwa model memiliki lebih banyak data untuk dipelajari sebelum diuji, yang memungkinkannya untuk menggeneralisasi pola dalam data dengan lebih baik.

4.2 Pengujian Sistem dengan Metode KNN

Metode K-NN diuji sebagai metode pembanding untuk menentukan apakah metode SVM memiliki kualitas yang lebih baik dalam mengklasifikasikan buku dengan fitur judul dan sinopsis dibandingkan dengan metode lain seperti K-NN. Metode K-NN digunakan karena merupakan metode klasik yang memiliki kinerja yang stabil dalam tugas klasifikasi, sehingga dapat digunakan sebagai metode pembanding SVM (Boateng et al., 2020). Proses pengujian metode K-NN menggunakan nilai k yang sama dengan jumlah genre hasil threshold normalisasi, yaitu 20. Hasil pengujian terhadap kinerja metode K-NN dapat dilihat pada Tabel 4.

Tabel 4. Hasil Training dan Testing dengan Metode KNN

Training Dataset	Testing Dataset	Training Accuracy	Test Acuraccy
90%	10%	100%	30 %
80%	20%	100%	31.53%
70%	30%	100%	29.48%
60%	40%	100%	30.19%
50%	50%	100%	28%

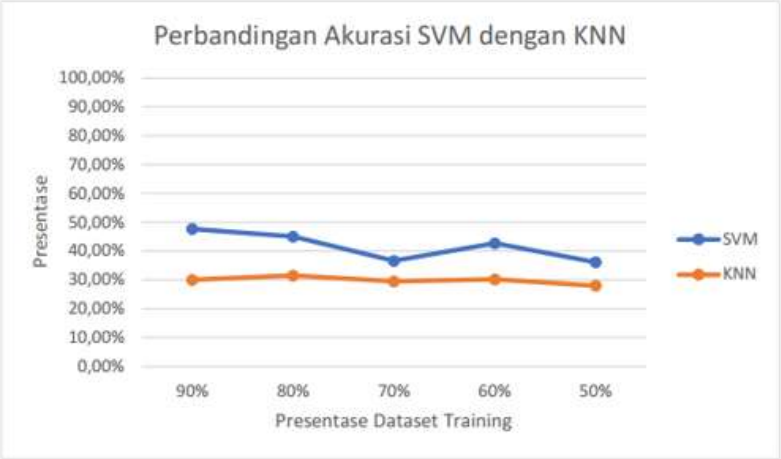
Berdasarkan Tabel 4, nilai akurasi yang paling tertinggi adalah 31.53% yang dimana ini dihasilkan dari komposisi dataset, yaitu 80% data training dan 20% sebagai data testing. Nilai akurasi

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

paling rendah adalah 28% dengan dataset, yaitu 50% data training dan 50% sebagai data testing. Lebih banyak data untuk dipelajari oleh model, yang memungkinkan pembelajaran pola yang lebih baik dalam dataset, proporsi data pembelajaran yang lebih besar (80%) menghasilkan akurasi yang lebih tinggi dibandingkan dengan proporsi yang lebih seimbang (50% training, 50% testing). Kualitas data juga penting dalam menentukan akurasi, jika data pelatihan mengandung suara atau tidak representatif, maka akurasi dapat rendah meskipun proporsi data pelatihan tinggi.

4.3 Perbandingan Hasil Pengujian Sistem Metode SVM dengan KNN

Berdasarkan pengujian terhadap dua metode, yaitu SVM dan K-NN didapatkan hasil semua skenario pembagian data training dan testing yang jika digambarkan dengan bentuk grafik garis maka akan terlihat seperti pada Gambar 4.



Gambar 4. Grafik perbandingan hasil kinerja dari metode SVM dan K-NN

Nilai akurasi metode SVM Linear Kernel berada di atas nilai akurasi metode K-NN secara keseluruhan dengan nilai k=20. Hasilnya menunjukkan bahwa metode SVM bergantung pada jumlah data pelatihan yang tersedia. Semakin banyak data instruksi yang dialokasikan, semakin baik kinerja klasifikasi SVM. Sebaliknya, hal ini akan terjadi sebaliknya. Ini dibuktikan dengan fakta bahwa nilai akurasi tertinggi dari percobaan metode SVM diperoleh pada komposisi sembilan puluh persen data pelatihan dan sepuluh persen data pengujian. Hal ini disebabkan oleh fakta bahwa SVM menggunakan pendekatan vektor terhadap garis hyperplane selama proses mesin pembelajarannya. Dengan demikian, semakin banyak data yang dijadikan proses pelatihan, semakin banyak garis hyperplane yang menginduksi data terdekatnya. Akibatnya, pengenalan setiap garis hyperplane menjadi lebih jelas dan berdampak pada peningkatan akurasi selama proses pengujian.

Sementara nilai akurasi tertinggi, 31,53% yang berasal dari komposisi dataset, 80% data pelatihan, dan 20% data pengujian, menunjukkan bahwa metode K-NN dari hasil pegujian tidak terpengaruh oleh jumlah data pelatihan yang ada saat ini. Namun, pelatihan tidak akan meningkatkan akurasi bahkan mungkin kembali menurun. Hal ini disebabkan oleh fakta bahwa metode K-NN sangat stabil dalam dataset yang terbatas. Namun, metode ini sangat bergantung pada nilai k yang diinisialisasi sebelumnya. Nilai k akan kurang akurat jika kurang dari idealnya dan lebih besar dari idealnya. Dalam kasus ini, ketika suatu model sangat dekat dengan set instruksi datanya dan digunakan pada data lain, akan menyebabkan akurasi yang lebih rendah. Ini karena K-NN secara sederhana mengelompokkan data baru berdasarkan jarak data dengan tetangganya sebanyak nilai k. Artinya, jika nilai k sudah ideal untuk klasifikasi, penambahan data baru tidak akan terlalu berpengaruh.

5. Kesimpulan

Penelitian ini membahas penerapan metode SVM untuk mengklasifikasikan buku menjadi genre berdasarkan judul dan sinopsisnya. Hasil pengujian menunjukkan bahwa SVM lebih baik daripada K-Nearest Neighbors (K-NN). SVM mencapai akurasi tertinggi 47,7% pada skenario 90 persen dan 10 persen data pelatihan, sedangkan K-NN hanya mencapai akurasi 31,53% pada skenario 80 persen dan 20 persen data pelatihan. SVM memiliki keunggulan dalam menangani jumlah data yang relatif kecil, tetapi tetap menghasilkan model dengan kinerja yang lebih baik dibandingkan K-NN dalam klasifikasi teks. Ini terbukti dengan fakta bahwa semakin banyak data pelatihan yang digunakan, semakin akurat model SVM. Terlepas dari fakta bahwa temuan penelitian ini menunjukkan SVM efektif dalam klasifikasi genre buku, penelitian lebih lanjut diperlukan dengan dataset yang lebih besar, termasuk buku dalam bahasa Indonesia. Penelitian masa depan mungkin melihat pendekatan tambahan atau kombinasi teknik pemrosesan bahasa alami untuk meningkatkan akurasi model.

Referensi

Amalia, D. H., & Yustanti, W. (2021). Klasifikasi Buku Menggunakan Metode Support Vector Machine pada Digital Library. *Journal of Informatics and Computer Science (JINACS)*, *3*(01). https://doi.org/10.26740/jinacs.v3n01.p55-61

Bamman, D., & Smith, N. A. (2013). *New Alignment Methods for Discriminative Book Summarization*. http://arxiv.org/abs/1305.1319

Bernardes, R. (2024). Machine learning - Basic principles. *Acta Ophthalmologica*, *102*(S279). https://doi.org/10.1111/aos.16281

Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *Journal of Data Analysis and Information Processing*, *08*(04). https://doi.org/10.4236/jdaip.2020.84020

Cahyani, S. N., & Saraswati, G. W. (2023). Implementation of Support Vector Machine Method in Classifying School Library Books with Combination of TF-IDF and Word2VEC. *Jurnal Teknik Informatika (Jutif)*, *4*(6). https://doi.org/10.52436/1.jutif.2023.4.6.1536

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Daniel, J., & Martin, J. H. (2023). Chapter 5 - Speech and Language Processing. *Speech and Language Processing*.

Deng, N., Tian, Y., & Zhang, C. (2012). Support vector machines: Optimization based theory, algorithms, and extensions. In *Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions*. https://doi.org/10.1201/b14297

Injadat, M. N., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Machine learning towards intelligent systems: applications, challenges, and opportunities. *Artificial Intelligence Review*, *54*(5). https://doi.org/10.1007/s10462-020-09948-w

Jalilifard, A., Caridá, V. F., Mansano, A. F., Cristo, R. S., & da Fonseca, F. P. C. (2021). Semantic Sensitive TF-IDF to Determine Word Relevance in Documents. *Lecture Notes in Electrical Engineering*, *736 LNEE*. https://doi.org/10.1007/978-981-33-6987-0_27

Leli, N., Rakhmawati, F., & Widyasari, R. (2023). Penerapan Metode Support Vector Machine (SVM) untuk Klasifikasi Uang Kuliah Tunggal di Universitas Islam Negeri Sumatera. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *4*(2). https://doi.org/10.46306/lb.v4i2.354

Monika, I. P., & Furqon, M. T. (2018). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(10).

Nguyen, T. T. S., & Do, P. M. T. (2020). Classification optimization for training a large dataset with Naïve Bayes. *Journal of Combinatorial Optimization*, *40*(1). https://doi.org/10.1007/s10878-020-00578-0

Sanderson, M. (2010). Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. ISBN-13 978-0-521-86571-5, xxi + 482 pages. *Natural Language Engineering*, *16*(1). https://doi.org/10.1017/s1351324909005129

Sulistilawati, I., Musyafa, A., Zain, R. M., Informatika, T., Pamulang, U., & Selatan, T. (2024). *Penerapan Data Mining Dalam Menentukan Pelajaran yang Diminati Dengan Metode Support Vector Mechine (SVM)* (Vol. 2, Issue 1).

Wang, Q. (2022). Support Vector Machine Algorithm in Machine Learning. *2022 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2022*. https://doi.org/10.1109/ICAICA54878.2022.9844516

Winiarti, S., Widayanti, D., Ahdiani, U., & Ismail, T. (2022). Klasifikasi Jenis Buku Berdasarkan Cover dan Judul Buku Menggunakan Metode Support Vector Machine dan Cosine Similarity. *Sainteks*, *19*(1). https://doi.org/10.30595/sainteks.v19i1.13423