

TINJAUAN PUSTAKA SISTEMATIS: SISTEM TEMU KEMBALI INFORMASI DENGAN METODE VECTOR SPACE MODEL

Ryas Widyana Kartika¹, Aang Alim Murtopo², Nurul Fadilah³

^{1,2,3}Program Studi Teknik Informatika, STMIK Tegal

E-mail : ryaskartika@gmail.com, aang.alim@gmail.com, nurulfadilah18@gmail.com

Jl. Pendidikan No.1, Pesurungan Lor, Kec. Margadana, Kota Tegal, Jawa Tengah 52122

Kata Kunci:

*Pandemi Covid-19,
Aplikasi E-learning,
Analisis Sentimen,
Support Vector
Machine, Quipper*

Abstrak

E-Learning atau Pembelajaran Online adalah pembelajaran formal dan informal yang dilakukan dengan menggunakan teknologi, sehingga siswa dan guru bisa melakukan proses belajar mengajar melalui media elektronik. Pembelajaran online memiliki kendala tersendiri dan berbeda-beda dari setiap pengguna. Ada yang merasa kesulitan untuk memahami materi atau penjelasan yang diberikan oleh dosen atau gurunya di saat pendidikan menjadi online, ada juga yang merasa mudah dalam memahami materi atau penjelasan saat melakukan pendidikan online. Pada penelitian ini metode SVM dipilih untuk mengklasifikasikan sentimen, karena metode SVM terbukti cukup efektif dalam melakukan klasifikasi sentimen. Alur awal yang dilakukan pada analisis sentimen di penelitian ini yaitu pengumpulan data ulasan dengan cara scrapping pada play store aplikasi Quipper. Selanjutnya adalah tahap preprocessing di mana proses ini memiliki lima tahapan yaitu case folding, text cleaning, tokenization, stemming, dan filtering (stopword removal). Setelah mendapatkan hasil preprocessing, akan dilakukan pelabelan dan transformation data. Hasil transformation data, yaitu data train akan digunakan dalam pembentukan model algoritma SVM. Selanjutnya data test dimasukkan ke dalam model algoritma SVM yang sudah dibuat. Proses terakhir adalah melakukan perhitungan akurasi metode SVM. Sentimen masyarakat atau pengguna terhadap aplikasi Quipper menghasilkan 410 ulasan negatif 590 dan ulasan positif. Hasil evaluasi kinerja algoritma Support Vector Machine kernel linier dalam klasifikasi atau analisis sentimen ulasan aplikasi e-learning Quipper menggunakan tiga skenario split data, diperoleh hasil terbaik di skenario ketiga dengan rasio perbandingan data training 60% dan data testing 40% dan menghasilkan accuracy sebesar 90,25%, precision 86,85%, recall 94,39%, dan f1-score 90,46%.

Abstract:

*Covid-19 Pandemic,
E-learning
Application,
Sentiment Analysis,
Support Vector
Machine, Quipper*

E-Learning or Online Learning is formal and informal learning carried out using technology, so that students and teachers can carry out the teaching and learning process through electronic media. Online learning has its own constraints and varies from user to user. There are those who find it difficult to understand the material or explanations given by the lecturer or teacher when education becomes online, there are also those who find it easy to understand the material or explanations when conducting online education. In this study, the SVM method was chosen to classify sentiment, because the SVM method proved to be quite effective in classifying sentiment. The initial flow of sentiment analysis in this study is collecting review data by scrapping on the Quipper application play store. Next is the preprocessing stage where this process has five stages, namely case folding, text cleaning, tokenization, stemming, and filtering (stopword removal). After getting the preprocessing

results, labeling and data transformation will be carried out. The results of data transformation, namely train data will be used in the formation of the SVM algorithm model. Furthermore, the test data is entered into the SVM algorithm model that has been created. The last process is to calculate the accuracy of the SVM method. The community or user sentiment towards the Quipper app resulted in 410 negative reviews, 590 positive reviews. The results of the evaluation of the performance of the Support Vector Machine linear kernel algorithm in the classification or sentiment analysis of the Quipper e-learning application reviews using three data split scenarios, the best results are obtained in the third scenario with a ratio of 60% training data and 40% testing data and produces an accuracy of 90, 25%, precision 86.85%, recall 94.39%, and f1-score 90.46%.

PENDAHULUAN

Perubahan jaman hingga saat ini sangat berpengaruh besar terhadap kehidupan sekarang. Salah satunya adalah cara menggunakan data sebagai informasi pada era big data sekarang ini. Dengan semakin banyaknya jumlah dokumen yang beredar saat ini memicu sebuah persoalan untuk melakukan pencarian yang diinginkan dengan cepat dan akurat baik pada media online melalui internet ataupun offline. Meskipun saat ini beberapa e-library sudah menerapkan algoritma untuk pencariannya seperti algoritma boolean search namun hal itu belum cukup akurat untuk proses pencarian karena tidak dapat mengidentifikasi dokumen yang paling relevan atau cukup relevan[1].

Proses pencarian informasi sekarang ini dengan menggunakan mesin pencarian Sistem Temu Kembali Informasi (STKI)[2]. Sistem Temu Kembali Informasi (STKI) atau yang dikenal sebagai Information Retrieval System (IRS) adalah sebuah sistem yang berfungsi untuk dapat menemukan (Retrieve) informasi yang dibutuhkan oleh User secara otomatis. Prinsip kerja sistem temu kembali informasi adalah dengan mencari sebuah dokumen yang relevan dari sekumpulan dokumen yang ada pada sistem dan membuang dokumen yang tidak relevan [3].

Dalam metode Vector Space Model, dokumen dipresentasikan untuk vektor yang mempunyai jarak (magnitude) dan arah (direction). Dalam metode Vector Space Model dapat membentuk kesamaan secara spesifik antara kata, frase dan dokumen. Teknik model ruang vektor memiliki keunggulan dalam penyesuaian query karena hanya dapat disesuaikan sebagian query dengan dokumen yang sudah ada. selain itu, pendekatan model ruang vektor dapat dengan mudah disesuaikan melalui penyetelan parameter, termasuk parameter pembobotan[4].

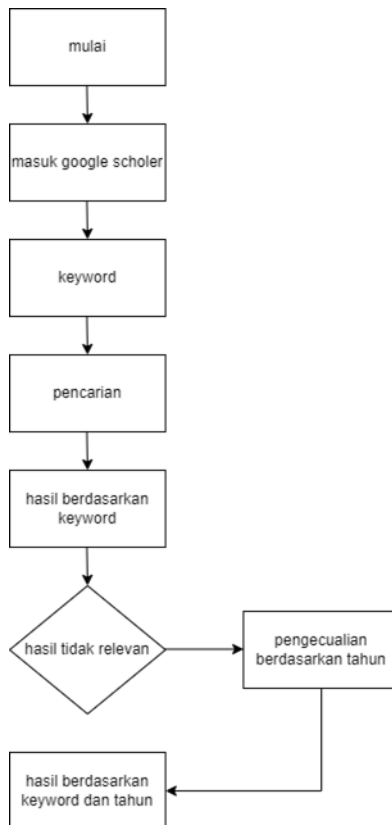
LANDASAN TEORI

Studi ini digunakan dalam pencarian literatur atau studi literatur yang sistematis. Metode yang digunakan untuk metode pengumpulan data. Yaitu, melakukan identifikasi, evaluasi, interpretasi, literatur, dan laporan yang tersedia. Peneliti membuat review, mengidentifikasi jurnal, dan secara sistematis mengikuti langkah-langkah dari setiap proses yang ditunjukkan oleh prosedur[5].

Ketika survei ini selesai peneliti terkait. Kata kunci dari penelitian ini adalah hasil belajar, keterampilan komunikasi, metode membaca dan pendekatan terbuka. Artikel majalah, 14 artikel jurnal dikumpulkan diterbitkan antara 2018-2022. Peneliti dari berbagai artikel jurnal kumpulkan artikel majalah kata kunci untuk penelitian ini. Dengan kata kunci “SISTEM TEMU KEMBALI INFORMASI” dan “VECTOR SPACE MODEL”.

Alur penggunaan system temu kembali, Pertama kita masuk ke google scholar, kemudian masukan keyword/kata kunci dokumen yang ingin dicari, kemudian lakukan pencarian berdasarkan keyword. Setelah hasil dari pencarian keyword muncul dan jika hasil ingin berdasarkan tahun maka lakukan pencarian berdasarkan tahun.

Metode pengumpulan data merupakan langkah awal bagi peneliti yang perlu memperoleh data terlebih dahulu, dimana data tersebut dijadikan sebagai acuan saat melakukan penelitian. Jenis kegiatan dalam proses pendataan[6]. Studi literature Membantu mendukung penelitian yang sedang dilakukan. Teori yang digunakan berasal dari buku, jurnal, dan penelitian sejenis yang dapat membantu memecahkan masalah dalam penelitian yang dilakukan [7]. Dokumentasi adalah metode pengumpulan data dengan membaca, mencatat, mengutip, dan mengumpulkan data teoritis dari internet sebagai dasar untuk menyusun hasil penelitian [8]. Peneliti mencari data di Internet juga dilakukan sehubungan dengan laporan ini, yang teorinya didasarkan pada teori [9].



Gambar 1. Pencarian dan Seleksi Studi.

Proses pengumpulan berbagai jurnal akan saya bahas dari satu acuan yakni Google Scholar. Langkah pertama adalah menggabungkan jurnal yang ada kaitannya untuk penelitian di jurnal ini. Pencarian pustaka yang berhubungan dengan penelitian jurnal. Pengambilan data sistem temu kembali informasi dengan metode vektor space model. Pengambilan data sistem temu kembali informasi dalam mencari sebuah dokumen.

Langkah awal sebelum melakukan review terlebih dahulu membuat pertanyaan penelitiannya (Research Question) untuk direview yaitu:

RQ1 : apa kegunaan dari system temu kembali informasi?

RQ2 : apa metode yang digunakan pada system temu kembali informasi?

METODE PENELITIAN

Systematic Literatur Review dikenal sebagai cara untuk menggabungkan, mengkategorikan, menyusun, dan menerapkan banyak penelitian unik berdasarkan kriteria, memperdalam revisi, dan menggabungkan kreativitas ke dalam data penelitian[10]. Metode ini mencoba untuk mengevaluasi dan mengintegrasikan semua bukti empiris yang relevan untuk memberikan penjelasan lengkap dari temuan. Tinjauan sistematis juga dapat membantu mengidentifikasi kesenjangan studi area[11].

HASIL DAN PEMBAHASAN

Pada penelitian ini merupakan hasil penerapan algoritma *Vektor Space Model* pada sistem temu kembali informasi perpustakaan. *Systematic literature rievew* didefinisikan sebagai proses mengidentifikasi, mengevaluasi, dan menafsirkan semua bukti penelitian yang tersedia untuk tujuan memberikan jawaban atas pertanyaan penelitian tertentu.

Tabel 1. Perbandingan refensi jurnal

Penelitian & Tahun	Judul Penelitian	Hasil Penelitian
--------------------	------------------	------------------

Suprianto, Sunardi, Abdul Fadlil,2018	Aplikasi sistem temu kembali angket mahasiswa menggunakan metode <i>generalized vector space model</i>	Dari hasil penelitian Pencarian dokumen diurutkan berdasarkan nilai similaritas hasil perhitungan dengan nilai > 0 dan diurutkan dari nilai terbesar ke nilai tertinggi. Pencarian akan dilakukan sampai tahap akhir yaitu mencari nilai similaritas dokumen, jika terdapat kata pada kata kunci didalam dokumen. Sebaliknya jika tidak terdapat kata tersebut maka, proses pencarian terhenti sampai proses preprocessing saja. Berdasarkan uji coba yang dilakukan terhadap 5 kata kunci dengan 50 data set, diperoleh rata-rata nilai precision sebesar 72% dan recall sebesar 100%.
Bayu Sugara, Dody, Donny,2018	Sistem Temu Kembali Informasi Pada Gejala Autisme Dengan Metode <i>Vector Space Model</i>	<p>Pada jurnal ini dihasilkan sebuah kesimpulan, metode <i>Vector Space Model</i> (VSM) dapat dijadikan sebagai mesin pencari dalam mencari data gejala - gejala gangguan pada autisme.</p> <p>Pada hasil pengujian dari jurnal ini menggunakan <i>Cosine Similarity</i> dengan 10 data gejala pada gangguan autisme mengungkapkan bahwa GJ04 adalah gejala seseorang yang sering mengucapkan bahasa/kalimat yang aneh secara terus-menerus atau berulang-ulang dan mempunyai tingkat kemiripan bahasa/kalimat yang tinggi dengan <i>query</i> sebesar 0.37647.</p>
Muhammad Syafiq H, Sisilia Thya Safitri, Rifki Adhitama,2020	Sistem Temu Kembali Informasi Dinas Pemuda Olahraga dan Pariwisata Kabupaten Purbalingga dengan Metode <i>Vector Space Model</i>	Hasil pengujian sistem menggunakan <i>blackbox testing</i> menunjukkan bahwa seluruh parameter pengujian yaitu sesuai dengan harapan dan tujuan, dari Hasil pada penelitian ini pencarian yang ditunjukkan mempunyai nilai yang relevan dengan <i>query</i> yang diinputkan oleh <i>user</i> . Dalam proses tampilan pada dokumen dari <i>query</i> yang diinputkan <i>user</i> memperoleh kecepatan sebesar 1,45 detik. Tingkat pda relevansi dari suatu sistem temu kembali informasi yang akan di hitung berdasarkan nilai <i>precision</i> dan <i>recall</i> menampilkan yaitu 65% dan 100%. Di simpulkan bahwa sistem dianggap relevan dengan nilai diatas 50%
Erik Ratu Prawira Negara, Fatkhul Amin,2021	System temu kembali cerita rayat nusantara menggunakan metode vector space model	Hasil uji pencarian pada sistem temu kembali informasi dengan satu istilah memiliki bobot tertinggi sebesar 0.0629182. Pengujian dengan dua istilah memiliki bobot tertinggi sebesar 0.704718. Pengujian dengan tiga term memiliki bobot tertinggi 0.233981. Bobot hasil penelusuran untuk tiga istilah atau lebih paling tinggi di 0,771766.
Amalia Beladonna Arifa, Gita Fadila Fitriana, Ananda Rifkiy Hasan,2021	Temu Kembali Informasi pada Soal Ujian dengan Rencana Pembelajaran	Hasil dari pada penelitian ini menunjukkan penggunaan metode <i>Vector Space Model</i> bisa diterapkan untuk pencarian sistem dalam menemukan kemiripan antara dokumen dengan soal ujian perencanaan dalam pembelajaran. dalam pencarian diawali berdasarkan nilai kesamaan hasil

	Menggunakan Vector Space Model	perhitungannya, dinyatakan sebagai persentase dari nilai maksimum ke nilai minimum. Nilai yang dihasilkan oleh sistem memiliki arti bahwa query butir soal ujian tersebut telah mempunyai persamaan dengan suatu materi pokok pada dokumen rencana pembelajaran.
Astrid Noviriandini, Diah Ayu Ambarsari, Fahmi Aprian, 2019	Information Retrieval System Pada Pencarian File Dokumen Berbasis Teks Dengan Metode Vector Space Model	dari hasil penelitian yang telah dilakukan adalah STKI yang dibuat dapat mencari informasi dari isi file dokumen yang tersimpan dalam sistem. Proses pengindeksan dokumen pada aplikasi STKI yang dikembangkan melalui beberapa tahapan pengolahan teks seperti tokenisasi, filtering, dan stemming. Sementara itu proses pencarian juga melalui beberapa proses yaitu penghitungan bobot dengan tf-idf, menghitung jarak tiap dokumen dan query, menghitung dot product, menghitung similaritas dan perangkungan.

RQ1: Sistem temu kembali informasi yang dirancang untuk mendapatkan informasi atau dokumen yang relevan dan memenuhi kebutuhan informasi pengguna dari sekumpulan besar dokumen.[12] Sistem temu kembali informasi dapat juga diukur dengan seberapa baik sistem mendapati sebuah koleksi yang relevan dan menyimpan sebuah koleksi yang tidak relevan. Relevansi ini merupakan tolak ukur keberhasilan pada sistem yang mempertemukan pencari dan koleksi yang dicari. Ukuran relevansi dapat digunakan sebagai cara untuk menentukan penilai sistem temu kembali informasi. *Recall* dan *precision* adalah pendekatan yang digunakan untuk menentukan penilaian efektivitas sistem temu kembali informasi. Penelitian sebelumnya telah menunjukkan bahwa sistem temu kembali informasi efektif ketika persentase *precision* nya tinggi sekalipun *recall* nya rendah.[13]

RQ2: pada *system* temu kembali informasi ini menggunakan metode *vector space model*. *Vector space model* adalah representasi aljabar dari dokumen teks sebagai pengenalan vektor, seperti indeks. Ini digunakan dalam pengindeksan informasi, peringkat relevansi, dan penyortiran informasi. Untuk membuat dokumen lebih sederhana dan lebih mudah ditangani. Dokumen yang diambil diurutkan dalam kesamaan, model vektor mempertimbangkan dokumen yang terkait dengan permintaan pengguna [14]. *Vector Space Model* (VSM) adalah cara dalam melihat suatu kesamaan (*smilarity*) atau kedekatan konsep dengan bobotnya. Sebuah dokumen dilihat sebagai vektor dengan besar jarak (*magnitude*) dan arah (*direction*). Dalam *Vector Space Model*, istilah ini digambarkan dengan dimensi ruang vektor. Relevansi *query* dokumen didasarkan pada kesamaan atau similaritas vektor *query* dan dokumen. [15].

KESIMPULAN DAN SARAN

Sistem Temu Kembali Informasi (STKI) dengan studi kasus menggunakan 6 dokumen jurnal skripsi mahasiswa. menghasilkan;

STKI dapat melakukan pemisahan term (tokenisasi), penyaringan term (pengurangan angka, tanda baca dan perhentian kata) dan turunan term (pembuatan kata dasar) dalam jangka waktu komputasi 4 menit 41 detik. STKI dapat melakukan pencarian teks dokumen bahasa Indonesia dan menampilkan hasil dari pencarian teks dokumen disertai pembobotan tiap-tiap dokumen beserta letak dokumen dilengkapi dengan bobot dan letak dokumen pada database.

REFERENSI

- [1] F. Sanjaya, "Pemanfaatan Sistem Temu Kembali Informasi dalam Pencarian Dokumen Menggunakan Metode Vector Space Model," J-INTECH (Journal Inf. Technol., vol. 05, no. 02, pp. 1689–1699, 2017.
- [2] A. Yudha, "Sistem Temu Kembali Index Berita," Sist. Temu Kembali Index Ber., pp. 44–52, 2020.

- [3] E. Ratu, P. Negara, and F. Amin, "Sistem Temu Kembali Cerita Rakyat Nusantara," *J. Din. Inform.*, vol. 13, no. 1, pp. 44–54, 2021.
- [4] A. B. Arifa, Gita Fadila Fitriana, and Ananda Rifkiy Hasan, "Temu Kembali Informasi pada Soal Ujian dengan Rencana Pembelajaran Menggunakan Vector Space Model," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 63–68, 2021, doi: 10.29207/resti.v5i1.2739.
- [5] E. Triandini, S. Jayanatha, A. Indrawan, G. W. Putra, and B. Iswara, "Metode systematic literature review untuk identifikasi platform dan metode pengembangan sistem informasi di Indonesia," *Indones. J. Inf. Syst.*, vol. 1, no. 2, pp. 63–77, 2019.
- [6] Y. Rahmanto, "Rancang Bangun Sistem Informasi Manajemen Koperasi Menggunakan Metode Web Engineering (Studi Kasus: Primkop Kartika Gatam)," *J. Data Min. Dan Sist. Inf.*, vol. 2, no. 1, pp. 24–30, 2021.
- [7] R. R. Suryono, I. Budi, and B. Purwandari, "Challenges and trends of financial technology (Fintech): a systematic literature review," *Information*, vol. 11, no. 12, p. 590, 2020.
- [8] D. Pasha, "Sistem Pengolahan Data Penilaian Berbasis Web Menggunakan Metode PIECIES," *J. Teknol. Dan Sist. Inf.*, vol. 1, no. 1, pp. 97–104, 2020.
- [9] S. Setiawansyah, Q. J. Adrian, and R. N. Devija, "Penerapan Sistem Informasi Administrasi Perpustakaan Menggunakan Model Desain User Experience," *J. Manaj. Inform.*, vol. 11, no. 1, pp. 24–36, 2021.
- [10] I. K. Crombie and H. T. Davies, "What is meta-analysis," *What is*, vol. 1, no. 8, 2009.
- [11] A. H. Eagly and W. Wood, "Using research syntheses to plan future research.," 1994.
- [12] S. Bahri, "Aplikasi Pencarian Bahan Pustaka Di Perpustakaan Menggunakan Metode Vector Space Model," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 5, no. 2, 2021, doi: 10.37438/jimp.v5i2.265.
- [13] T. T. Prabowo, "Efektivitas Sistem Temu Kembali Informasi Perpustakaan Digital Institut Seni Indonesia (ISI) Yogyakarta dalam Tinjauan Recall dan Precision," *Media Pustak.*, vol. 28, no. 1, pp. 37–48, 2021, doi: 10.37014/medpus.v28i1.1087.
- [14] F. Febriyanto, "Sistem Penilaian Otomatis Jawaban Esai Dengan Menggunakan Metode Vector Space Model Pada Beberapa Perkuliahan Di Stmik Indonesia Banjarmasin," *Respati*, vol. 14, no. 1, pp. 53–68, 2019, doi: 10.35842/jtir.v14i1.272.
- [15] R. Ridwan and T. A. Hermawan, "Penerapan mesin pencari informasi dengan menggunakan metode Vector Space Model," *JUTEKIN (Jurnal Tek. ...)*, vol. 7, no. 2, 2019, [Online]. Available: <http://jurnal.stmik-dci.ac.id/index.php/jutekin/article/view/372>