|  |  |
| --- | --- |
| Jurnal 1 | |
| Judul | **Vector Space Model-based Information Retrieval Systems at South Sumatera Regional Libraries A R T I C L E I N F O** |
| Penulis | M. Akbar As Shiddiqi, A Sanmarino |
| Tahun | (2023) |
| Hasil dan Pembahasan | * Penelitian ini menyajikan gambaran penelitian yang bertujuan untuk mengoptimalkan pencarian informasi perpustakaan melalui pemanfaatan metode Vector Space Model (VSM) dalam konteks ilmu computer. * penelitian ini menggunakan teknik pengambilan informasi, khususnya metode VSM, yang menilai kesamaan istilah dengan memberikan bobot pada istilah, memungkinkan representasi dokumen dan kueri sebagai vektor. * Dengan menggunakan model Air Terjun untuk pengembangan sistem, penelitian ini menguraikan tahapan seperti analisis, desain, pengkodean, pengujian, dan implementasi. * Representasi numerik dokumen teks dalam metode VSM memfasilitasi perhitungan kesamaan yang tepat, didukung oleh nilai TF-IDF yang menunjukkan pentingnya istilah dalam dokumen relatif terhadap korpus. |
| Kesimpulan Penelitian | Secara keseluruhan, penelitian ini menyajikan pendekatan sistematis untuk meningkatkan pengambilan informasi di perpustakaan, menekankan peran penting VSM dalam mengoptimalkan pencarian dokumen dalam koleksi yang luas. |

|  |  |
| --- | --- |
| Jurnal 2 | |
| Judul | **Topic Modelling Using VSM-LDA For Document Summarization** |
| Penulis | Atikah, Luthfi Hasanah, Novrindah Alvi Arifin, Agus Zainal |
| Tahun | (2022) |
| Hasil dan Pembahasan | makalah ini mengusulkan metode pemodelan topik menggunakan kombinasi LDA dan VSM (Vector Space Model) untuk peringkasan otomatis.  Metode yang diusulkan dapat mengatasi order effect dan mengidentifikasi topik dokumen yang dihitung berdasarkan bobot TF-IDF pada VSM yang dihasilkan oleh LDA |
| Kesimpulan Penelitian | Hasil usulan metode pemodelan topik pada 1300 data Twitter menghasilkan nilai koherensi tertinggi mencapai 0,72. Hasil rangkuman diperoleh Rouge 1 sebesar 0,78, Rouge 2 sebesar 0,67 dan Rouge L sebesar 0,80. |

|  |  |
| --- | --- |
| Jurnal 3 | |
| Judul | **Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter** |
| Penulis | Darwis, Muhammad Tri Pranoto, Gatot Eka Wicaksana, Yusuf |
| Tahun | (2020) |
| Hasil dan Pembahasan | * Dalam tulisan ini, penulis mengelompokkan data tweet dengan algoritma TF-IDF dan Metode K-Mean menggunakan bahasa pemrograman python * Hasil clustering data tweet menunjukkan prediksi atau mungkin topik pembicaraan yang sedang banyak dibicarakan oleh netizen |
| Kesimpulan Penelitian | Hasil clustering data tweet menunjukkan prediksiatau mungkin topik pembicaraan yang sedang banyak dibicarakan oleh netizen. Akhirnya data tersebut dapat digunakan untuk membuat keputusan yang memanfaatkan sentimen masyarakat terhadap suatu peristiwa melalui media sosial seperti Twitter. |

|  |  |
| --- | --- |
| Jurnal 4 | |
| Judul | **Literatify : Trends in Library Developments A Literature Review: the Importance of Term Normalization in Vector Space Model** |
| Penulis | Hasyim, Fayyaz Mubarak  Fahmi, Faisal |
| Tahun | (2024) |
| Hasil dan Pembahasan | * Artikel ini menggarisbawahi peran penting VSM dalam mengelola data dalam jumlah besar dan meningkatkan akurasi pengambilan dengan memberi peringkat dokumen berdasarkan kesamaan kueri. Normalisasi istilah, bagian dari pengembangan VSM, menstandarkan kata untuk pengindeksan, meningkatkan akurasi dengan mengatasi variasi kata. * Metodologi penelitian ini melibatkan tinjauan literatur sistematis, pengumpulan data melalui database elektronik, dan analisis tematik. * Temuan penelitian menyoroti aspek-aspek penting: dasar-dasar sistem pencarian informasi, prinsip kerja VSM dalam penyortiran dokumen, dan proses normalisasi istilah. Berbagai metode dalam normalisasi term, seperti tokenisasi, pemfilteran, stemming, dan pembobotan term (misalnya, TF, IDF, Cosine Kemiripan), dijelaskan untuk menyempurnakan relevansi dokumen. * Diskusi menggarisbawahi dampak normalisasi istilah pada pengambilan informasi, menekankan peningkatan akurasi, efisiensi, dan pengurangan tingkat kesalahan. Dalam makalah penelitian, lima studi yang menunjukkan keberhasilan penerapan VSM di berbagai domain direferensikan. Domain tersebut meliputi penelusuran lagu karaoke, seleksi penguji skripsi, identifikasi hama pada tanaman padi, tafsir hadis, dan penelusuran bahan pustaka. Setiap studi menunjukkan efektivitas dan fleksibilitas VSM dalam memecahkan berbagai masalah di berbagai bidang. |
| Kesimpulan Penelitian | Kesimpulannya, VSM muncul sebagai alat yang ampuh dalam mengelola kelebihan informasi, terutama bila digabungkan dengan teknik normalisasi. Studi yang ditinjau menggambarkan kemanjuran VSM dalam memberikan hasil yang tepat, menegaskan statusnya sebagai metode pilihan dalam sistem pencarian informasi karena keakuratan dan efektivitasnya. |

|  |  |
| --- | --- |
| Jurnal 5 | |
| Judul | **Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF** |
| Penulis | Nuzul Hikmah  Dyah Ariyanti  Ferry Agus Pratama |
| Tahun | (2021) |
| Hasil dan Pembahasan | * Penelitian ini mengimplementasikan chatbot sebagai asisten virtual yang dapat digunakan sebagai layanan informasi akademik bagi masyarakat umum maupun civitas akademika kampus Universitas Panca Marga Probolinggo. * Tahapan pengembangan chatbot ini dengan *waterfall method* meliputi analisis, desain, kode, pengujian dan pemeliharaan. * Metode yang digunakan untuk pembelajaran chatbots menggunakan Tf-Idf dan VSM untuk pembobotan kata pada dokumen dan query serta Cosine kesamaan untuk menghitung kemiripan (similarity) antara dokumen dan query. * Hasil akhir dari penelitian ini adalah sebuah aplikasi chatbot yang dapat digunakan sebagai asisten virtual sebagai customer service dalam melayani dan memberikan informasi seputar sivitas akademika Universitas Panca Marga Probolinggo. |
| Kesimpulan Penelitian | Berdasarkan hasil pengujian akurasi dan UAT, tingkat akurasi yang diperoleh chatbot mencapai 85,7% dan pengujian UAT pada pengujian pertama mencapai 84,1% dengan jumlah responden 30 orang, pada pengujian kedua mencapai 82,1% dengan total dari 92 responden. |

|  |  |
| --- | --- |
| Jurnal 6 | |
| Judul | **Application of an Improved TF-IDF Method in Literary Text Classification** |
| Penulis | Xiang, Lin |
| Tahun | (2022) |
| Hasil dan Pembahasan | * tujuan mengklasifikasikan teks sastra dalam penelitian ini, karya ini mengusulkan metode IDF yang lebih baik untuk masalah kata fitur yang muncul beberapa kali dan memiliki makna beragam di berbagai bidang. * Arti kata fitur dalam domain berbeda dipisahkan untuk meningkatkan kepercayaan pada keluaran algoritma TF-IDF. Dengan menggunakan metode TF-IDF yang ditingkatkan yang disarankan dalam penelitian ini dengan pengklasifikasi hutan acak (RF) |
| Kesimpulan Penelitian | hasil eksperimen menunjukkan bahwa pengklasifikasi memiliki dampak klasifikasi yang baik, yang dapat memenuhi kebutuhan kerja sebenarnya, berdasarkan eksperimen komparatif pada pemilihan dimensi fitur, algoritma pemilihan fitur, algoritma bobot fitur, dan pengklasifikasi. Ini memiliki signifikansi sejarah yang cukup besar |

|  |  |
| --- | --- |
| Jurnal 7 | |
| Judul | **TF-IDF Method and Vector Space Model Regarding the Covid-19 Vaccine on Online News** |
| Penulis | Zen, Bita Parga Susanto, Irwan Finaliamartha, Dian |
| Tahun | (2021) |
| Hasil dan Pembahasan | * Pada jurnal ini digunakan Metode yang biasa untuk sistem pencarian \yang diharapkan dapat membantu dalam menemukan informasi yang diinginkan atau relevan dengan input query, dalam hal ini adalah TF-IDF dan VSM (Vector Space Model) yang digunakan dalam pembobotan untuk mengukur statistik dari kumpulan dokumen pencarian beberapa informasi tentang vaksin Covid 19 pada berita kompas.com kemudian melakukan tokenisasi. untuk memisahkan teks, stopword removal atau filtering untuk menghilangkan kata-kata yang tidak diperlukan yang biasanya mengandung kata sambung dan lain-lain. * Langkah selanjutnya adalah pembentukan kalimat yang bertujuan untuk menghilangkan infleksi kata pada bentuk dasarnya. Kemudian dilakukan perhitungan TF-IDF dan VSM |
| Kesimpulan Penelitian | hasil akhir berupa dokumen berita 3 (DOC 3) dengan bobot 5.914226424; dokumen berita 2 (DOC 2) dengan bobot 1.767692186; dokumen berita 5 (DOC 5) dengan bobot 1.550165096; dokumen berita 4 (DOC 4) dengan bobot 1.17141223;, dan terakhir dokumen berita 1 (DOC 1) dengan bobot 0.5244103739. |

|  |  |
| --- | --- |
| Jurnal 8 | |
| Judul | **Research of Text Classification Based on TF-IDF and CNN-LSTM** |
| Penulis | Hai Zhou 2022 J. Phys.: Conf. Ser. **2171** 012021 |
| Tahun | (2022) |
| Hasil dan Pembahasan | * Makalah ini menggunakan Term Frequency-inverse Document Frequency (TF-IDF) untuk menghapus fitur dengan bobot lebih rendah, mengekstrak fitur utama dalam teks, mengekstrak vektor kata yang sesuai melalui model Word2Vec, dan kemudian memasukkannya ke dalam model CNN-LSTM * Peneliti membandingkan model dengan metode perhatian CNN, LSTM, dan LSTM dan menemukan bahwa model dapat secara signifikan mengurangi parameter model dan waktu pelatihan dalam kumpulan data teks pendek dan panjang |
| Kesimpulan Penelitian | Makalah ini juga mengusulkan penggabungan fitur teks asli untuk menebus hilangnya akurasi yang disebabkan oleh metode ekstraksi fitur TF-IDF. |

|  |  |
| --- | --- |
| Jurnal 9 | |
| Judul | **BERTopic: Neural topic modeling with a class-based TF-IDF procedure** |
| Penulis | Maarten Grootendorst |
| Tahun | (2022) |
| Hasil dan Pembahasan | * Peneliti menyajikan BERTopic, model topik yang memperluas proses ini dengan mengekstraksi representasi topik yang koheren melalui pengembangan variasi TF-IDF berbasis kelasPeneliti membandingkan model dengan metode perhatian CNN, LSTM, dan LSTM dan menemukan bahwa model dapat secara signifikan mengurangi parameter model dan waktu pelatihan dalam kumpulan data teks pendek dan Panjang. * BERTopic menghasilkan penyematan dokumen dengan model bahasa berbasis transformator yang telah dilatih sebelumnya, mengelompokkan penyematan ini, dan akhirnya, menghasilkan representasi topik dengan prosedur TF-IDF berbasis kelas. |
| Kesimpulan Penelitian | BERTopic menghasilkan topik yang koheren dan tetap kompetitif di berbagai tolok ukur yang melibatkan model klasik dan yang mengikuti pendekatan pengelompokan pemodelan topik yang lebih baru. |

|  |  |
| --- | --- |
| Jurnal 10 | |
| Judul | **Identifying Emerging Trends in Scientific Texts Using TF-IDF Algorithm: A Case Study of Medical Librarianship and Information Articles** |
| Penulis | Meisam Dastani 1, Afshin Mousavi Chelak 2, Soraya Ziaei 3, Faeze Delghandi4 |
| Tahun | (2020) |
| Hasil dan Pembahasan | * Hasil yang diperoleh dari algoritma TF-IDF menunjukkan bahwa kata “Library”, “Patient”, dan “Inform” dengan bobot dari 95.087, 65.796, dan 63.386, masing-masing, merupakan kata kunci terpenting dalam artikel yang diterbitkan tentang perpustakaan medis dan informasi. * Selain itu, kata “Katalog”, “Buku”, dan “Jurnal” merupakan kata kunci terpenting yang digunakan dalam artikel yang diterbitkan antara tahun 1960 dan 1970, dan kata “Pasien”, “Toko Buku”, dan “Intervensi” adalah kata kunci terpenting yang digunakan dalam artikel tentang kedokteran kepustakawanan dan informasi yang diterbitkan dari tahun 2015 hingga 2020. Kata-kata “Blockchain”, “Telerehabilit”, “Instagram”, “WeChat”, dan “Komik” adalah kata kunci baru yang diamati dalam artikel tentang perpustakaan medis dan informasi antara tahun 2015 dan 2020. |
| Kesimpulan Penelitian | Hasil penelitian ini mengungkapkan bahwa kata kunci yang digunakan dalam artikel tentang kepustakawanan medis dan informasi adalah tidak konsisten dari waktu ke waktu dan mengalami perubahan pada periode yang berbeda-beda sehingga saat ini bidang ilmu tersebut juga mengalami perubahan kebutuhan masyarakat dengan munculnya dan pertumbuhan teknologi informasi. |