**PROJECT UAS**

**INFORMATION RETRIEVAL**

**(Dr.Drs.ERI ZULIARSO,M.Kom)**

**Review Jurnal Clustering**



**PENYUSUN:**

**Dibyo Panggalih**

**(15.01.55.0022)**

**UNIVERSITAS STIKUBANK (UNISBANK)**

**SEMARANG 2018**

**LITERATURE RIVIEW**

|  |  |
| --- | --- |
| **Judul artikel** | SISTEM KLASIFIKASI DAN PENCARIAN JURNAL DENGAN MENGGUNAKAN METODE NAIVE BAYES |
| **Penulis** | Ainalia Indranandita , Budi Susanto , Antonius Rachmat C |
| **Judul Jurnal & Tahun Penerbitan** | JURNAL INFORMATIKA, NOVEMBER 2OO8 |
| **Masalah utama yang diangkat** | Permasalahan yang muncul adalah bagaimana sistem dapat  Melakukan pengelompokan dan pencarian jurnal yang relevan untuk memenuhi kebutuhan konsumen. |
| **Metodologi** | Metode Vector Space Model : Metode NaiVe Bayes atau NaiVe Bayes Classifier (NBC) adalah salah satu metode yang digunakan untuk klasifikasi teks. NBC menggunakan teori probabilitas sebagai dasar teori. |
| **Hasil Penelitian** | Terdapat 5 kategori, yaitu Health, Mltsic, Politics, Sport dan Technology. Jumlah data training adalah 250 jurnal Proquest yang terdiri dari 50 jurnal untuk setiap kategorinya. Sedangkan data fesfer yang telah dipersiapkan adalah sebanyak 50 jurnal Proquest, masingmasing 10 jurnal pada setiap kategori. Setiap jurnal training dan tester melewati proses text mining terlebih dahulu, yaitu fexf preprocessing, text transformation, feature selection dan pattern discovery. Setelah melewati proses text preprocessrng, diperoleh jumlah training = 249 jurnal, karena terdapat 1 jurnal berkategori Potitics yang sama dalam database (id jurnal harus unik, tidak boleh kembar). Kemudian setelah melewati proses text transformafion, terbentuk 3763 token yang unik. Rancangan pengujian sistem klasifikasi akan difokuskan terhadap beberapa pilihan feature selection yang diambil, kemudian dilihat berdasarkan nilai precision yang diperoleh. Asumsi yang dipakai adalah jumlah vektor ideal adalah kurang dari 40% dari total token dan feature selection dikatakan baik apabila dapat menghasilkan nilai preclslon lebih dari 60%.. |
| **Kesumpulan & Saran** | Berdasarkan hasil analisis dan implementasi sistem, maka dapat disimpulkan:   * Sistem klasifikasi dengan metode Nai've Bayes dengan FS-4 menghasilkan precision sebesar 64%. * Feature selection dengan bantuan pembobotan token menghasilkan precision kurang dari 60%, jadi proses pembobotan token tidak mempengaruhi dalam sistem klasifikasi.   Penggunaan metadata (hasil klasifikasi) dalam proses pencarian Oapit meningkatkan tingkat recall. Saran untuk pengembangan dan perbaikan sistem ini adalah: r' '   * Perlu adanya perbaikan struktur data untuk mempercepat proses text mining, karena semakin banyak jumlah data, semakin lama pula proses yang dibutuhkannya. * Pengembangan sistem, yaitu sistem dapat menambahkan kategori baru.   Pengembangan pencarian dengan kata kunci berupa frasa atau dengan penggunaan Boolean Operator (misal: OR, AND). |
| **Kekurangan** | Semakin besar tingkatthreshotd yang diambil dalam setiap feature selection (dilihat dari sumbu X), maka nilaiprecrbion akan semakin berkurang |

|  |  |
| --- | --- |
| **Judul artikel** | **PENERAPAN SISTEM KLASIFIKASI KEARSIPAN PADA KANTOR ARSIP DAERAH** |
| **Penulis** | H.M. Nawawi Dg. Sibali |
| **Judul Jurnal & Tahun Penerbitan** | JURNAL EKSIS , Agustus 2010 |
| **Masalah utama yang diangkat** | Permasalahan yang muncul adalah bagaimana sistem dapat  Melakukan pengelompokan dan pencarian jurnal yang relevan untuk Pengelolaan Arsip Kantor Arsip Daerah |
| **Metodologi** | Metode: **Filing**, Menurut Nomor Urut dan Sistem Wilayah (daerah/geografi). Sistem penyimpanan arsip yang digunakan harus disesuaikan dengan jumlah volume dan jenis arsip, artinya sistem penyimpanan tidak terpaku pada satu sistem penyimpanan saja, misalnya pada sistem klasifikasi berdasarkan masalah saja melainkan dapat diganti dengan sistem lainnya, misalnya sistem kronologis ataupun dapat dikombinasikan antara dua sistem, misalnya sistem wilayah digabungkan dengan sistem kronologis bahkan dapat pula dilakukan penggabungan tiga sistem sekaligus misalnya antara nomor urut, wilayah dan kronologis. |
| **Hasil Penelitian** | bahwa pemilik Kartu Keluarga berasal dari berbagai kecamatan, di mana memerlukan sistem kearsipan yang lebih optimal dalam penemuan kembali arsip, maka untuk menyimpan arsip Kartu Keluarga sebaiknya bagian Seksi Pengelolaan Arsip menggunakan sistem nomor urut yang digabungkan dengan sistem wilayah (daerah/geografi), tujuannya agar lebih mudah dalam hal penyimpanan dan penemuan kembali arsip, karena kode lacinya sudah disesuaikan dengan wilayah asal Kartu Keluarga tersebut, sehingga bila ada nama Kepala Keluarga yang sama, maka dapat dilihat dari kode arsip yang menunjukkan wilayah ataupun alamat kepala keluarga serta nomor urut file yang dicari, sehingga akan lebih mudah bila dilakukan proses pencarian dan penemuan kembali. |
| **Kesumpulan & Saran** | **Kesimpulan :**Berdasarkan pembahasan yang telah penulis uraikan, maka penulis dapat menarik suatu kesimpulan yang sesuai dengan permasalahan dan hipotesis yang telah penulis kemukakan:  1. Pekerjaan administrasi khususnya penanganan Kartu Keluarga pada Bagian Seksi Pengelolaan Arsip Kantor Arsip Daerah belumlah optimal karena penyimpanan arsipnya menggunakan Sistem Klasifikasi Masalah tanpa menyebutkan asal wilayah Kartu Keluarga sehingga sulit jika dilakukan pencarian kembali.  2. Peralatan yang digunakan untuk menyimpan Kartu Keluarga pada Bagian Seksi Pengelolaan Arsip Kantor Arsip Daerah hanya menggunakan kotak arsip, folder dan rak arsip.  3. Sistem penyimpanan arsip yang digunakan dalam proses penyimpanan tidak dikembangkan atau dikombinasikan serta tidak disesuaikan dengan jumlah volume dan jenis arsip yang akan disimpan.  **Saran :** Untuk masalah Kartu Keluarga sebaiknya sistem penyimpanan yang digunakan adalah gabungan antara Sistem nomor urut dan Sistem wilayah agar dapat mempermudah proses penyimpanan dan ketika dilakukan pencarian dengan sistem komputerisasi dapat dengan segera ditemukan. |
| **Kelebihan & Kekurangan** | **Kelebihan:** bisa lebih efesien dalam mencari data kartu keluarga di setiap daerah  **Kekurangan:** minim keter sediaan komputer / internet dalam wilayah tertentu |

|  |  |
| --- | --- |
| **Judul artikel** | SISTEM KLASIFIKASI DAN PENCARIAN JURNAL DENGAN MBNGGUNAKAN VECTOR SPACE MODEL |
| **Penulis** | Ainalia Indranandita , Budi Susanto , Antonius Rachmat C |
| **Judul Jurnal & Tahun Penerbitan** | JURNAL INFORMATIKA, NOVEMBER 2OO8 |
| **Masalah utama yang diangkat** | Permasalahan yang muncul adalah bagaimana sistem dapat  Melakukan pengelompokan dan pencarian jurnal yang relevan untuk memenuhi kebutuhan konsumen. |
| **Metodologi** | Metode Vector Space Model : Metode Vector Space Modelalau Term Vector Model adalah sebuah model aljabar untuk menggambarkan dokumen teks (beberapa objek) sebagai vektor dari identifier. Biasanya digunakan dalam penyaringan informasi (information filtering), penemuan informasi (information retrieval), indexing dan pemberian ranking yang saling relevan. |
| **Hasil Penelitian** | hasil pengujian dari sistem pencarian dengan VSM ditambah dengan suatu label kategori yang diasumsikan sebagai hasil sistem klasifikasi dengan NB. Kemudian label kategori akan dipergunakan dalam pencarian detilterhadap metadata = kategori. Analisis sistem klasifikasi dan pencarian akan difokuskan pada nilai recalldan precision yang diperoleh. |
| **Kesumpulan & Saran** | Berdasarkan hasil analisis dan implementasi sistem, maka dapat disimpulkan:   * Sistem pencarian dengan metode Vector Space Modetpendekatan Cosrne menghasilkan recal/ sebesar 54.8o/o dan preclsion sebesar 60.7%. * Feature Selection dari sumber data pelatihan dan fesfer yang berbeda menghasilkan precision kurang dari 60%, karena jangkauan topic   pembicaraan yang cukup berbeda.  Penggunaan metadata (hasil klasifikasi) dalam proses pencarian Oapit meningkatkan tingkat recall. Saran untuk pengembangan dan perbaikan sistem ini adalah: r' '   * Perlu adanya perbaikan struktur data untuk mempercepat proses text mining, karena semakin banyak jumlah data, semakin lama pula proses yang dibutuhkannya. * Pengembangan sistem, yaitu sistem dapat menambahkan kategori baru.   Pengembangan pencarian dengan kata kunci berupa frasa atau dengan penggunaan Boolean Operator (misal: OR, AND). |
| **Kelebihan** | tingkat ketepptan , kumpulan dokumen yang relevan,kumpulan dokumen yang ditemukan |
| **Judul artikel** | SISTEM KLASIFIKASI DAN PENCARIAN JURNAL DENGAN MBNGGUNAKAN VECTOR SPACE MODEL |
| **Penulis** | Ainalia Indranandita , Budi Susanto , Antonius Rachmat C |
| **Judul Jurnal & Tahun Penerbitan** | JURNAL INFORMATIKA, NOVEMBER 2OO8 |
| **Masalah utama yang diangkat** | Permasalahan yang muncul adalah bagaimana sistem dapat  Melakukan pengelompokan dan pencarian jurnal yang relevan untuk memenuhi kebutuhan konsumen. |
| **Metodologi** | Metode Vector Space Model : Metode Vector Space Modelalau Term Vector Model adalah sebuah model aljabar untuk menggambarkan dokumen teks (beberapa objek) sebagai vektor dari identifier. Biasanya digunakan dalam penyaringan informasi (information filtering), penemuan informasi (information retrieval), indexing dan pemberian ranking yang saling relevan. |
| **Hasil Penelitian** | hasil pengujian dari sistem pencarian dengan VSM ditambah dengan suatu label kategori yang diasumsikan sebagai hasil sistem klasifikasi dengan NB. Kemudian label kategori akan dipergunakan dalam pencarian detilterhadap metadata = kategori. Analisis sistem klasifikasi dan pencarian akan difokuskan pada nilai recalldan precision yang diperoleh. |
| **Kesumpulan & Saran** | Berdasarkan hasil analisis dan implementasi sistem, maka dapat disimpulkan:   * Sistem pencarian dengan metode Vector Space Modetpendekatan Cosrne menghasilkan recal/ sebesar 54.8o/o dan preclsion sebesar 60.7%. * Feature Selection dari sumber data pelatihan dan fesfer yang berbeda menghasilkan precision kurang dari 60%, karena jangkauan topic   pembicaraan yang cukup berbeda.  Penggunaan metadata (hasil klasifikasi) dalam proses pencarian Oapit meningkatkan tingkat recall. Saran untuk pengembangan dan perbaikan sistem ini adalah: r' '   * Perlu adanya perbaikan struktur data untuk mempercepat proses text mining, karena semakin banyak jumlah data, semakin lama pula proses yang dibutuhkannya. * Pengembangan sistem, yaitu sistem dapat menambahkan kategori baru.   Pengembangan pencarian dengan kata kunci berupa frasa atau dengan penggunaan Boolean Operator (misal: OR, AND). |
| **Kelebihan** | tingkat ketepptan , kumpulan dokumen yang relevan,kumpulan dokumen yang ditemukan |

|  |  |
| --- | --- |
| **Judul artikel** | Klasifikasi Dokumen Karya Akhir Mahasiswa Menggunakan Naïve Bayes Classifier (NBC) Berdasarkan Abstrak Karya Akhir Di Jurusan Teknik Elektro Universitas Negeri Jakarta |
| **Penulis** | Nur Indah Pratiwi, Widodo |
| **Judul Jurnal & Tahun Penerbitan** | JURNAL PINTER ,1 JUNI 2017 |
| **Masalah utama yang diangkat** | 1. Jumlah dokumen digital semakin bertambah dari segi kuantitas. 2. Isi dokumen digital yang semakin banyak, sehingga diperlukan metode paling efektif untuk mengatur dan me-retriev kembali data yang telah disimpan. 3. Kesalahan dalam pencarian karena penggunaan metode yang digunakan tidak sesuai. |
| **Metodologi** | Eksperimen dan menggunakan 100 dokumen Abstrak |
| **Hasil Penelitian** | Pengujian sistem menggunakan metode K-Fold Cross Validation dengan membagi data menjadi 10 bagian (10 Fold) dengan mempertahankan perbandingan pembagian dokumen data train dan data test sebesar 9:1. |
| **Kesimpulan & Saran** | Kesimpulan: Berdasarkan hasil penelitian, dihasilkan sebuah software yang dapat mengklasifikasikan dokumen berbentuk isi abstrak skripsi untuk mengetahui termasuk ke dalam kategori apa isi abstrak tersebut.  Saran : Untuk penelitian selanjutnya diharapkan dapat mencoba algoritma yang lain, yang kemungkinan mempunyai tingkat akurasi yang lebih tinggi. Sehingga mempunya hasil prediksi yang jauh lebih baik. Serta penambahan jumlah data train akan mempengaruhi tingkat akurasi dari sistem. Maka, diharapkan dalam penelitian selanjutnya target dokumen, baik dokumen set untuk data training dan data test semakin banyak |

|  |  |
| --- | --- |
| Judul | KLASIFIKASI DAN PENCARIAN BUKU REFERENSI AKADEMIK MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC) (STUDI KASUS: PERPUSTAKAAN DAERAH PROVINSI KALIMANTAN TIMUR) |
| Tahun | 1 Februari 2015 |
| Penulis | Agus Setiawan , Indah Fitri Astuti , Awang Harsa Kridalaksana |
| Masalah utama yang diangkat | * Data buku yang diuji berasal dari Badan Perpustakaan Daerah Provinsi Kalimantan Timur. * Pengklasifiasian buku berdasarkan 5 kategori, yaitu Pemrograman Komputer, Jaringan Komputer, Database, Multimedia, dan Sistem Operasi. * Algoritma yang akan diimplementasikan pada bahasan mengenai proses pengklasifikasian buku adalah Naïve Bayes Classifier (NBC). |
| Metode Penelitian | Naïve Bayes Classiffier |
| Objek Penelitian | Pencarian Buku di Perpustakaan |
| Hasil Penelitian | Pada sistem ini terdapat dua level hak pengguna yaitu admin dan pengunjung sebagai user. Level admin berfungsi memanajemen buku-buku pada aplikasi ini, seperti menambah dan menghapus buku, melakukan klasifikasi buku, serta menambah dan menghapus admin. Disamping itu admin juga dapat melihat statistik web dan daftar buku hasil pelatihan. Level user digunakan oleh pengunjung perpustakaan untuk melihat daftar buku terbaru, melakukan pencarian buku, dan melihat daftar seluruh buku hasil dari klasifikasi. |
| Kesimpulan | yang dapat diambil berdasarkan penelitian mengenai Klasifikasi dan Pencarian Buku Referensi Akademik Menggunakan Metode Naïve Bayes Classifier antara lain:  1. Metode Naïve Bayes Classifier dapat digunakan untuk melakukan klasifikasi buku berbahasa Indonesia yang berada di Perpustakaan Daerah Provinsi Kalimantan Timur.  2. Jumlah data latih sangat berpengaruh dalam proses klasifikasi, karena semakin banyak data latih maka tingkat keakuratan dalam penentuan kategori akan semakin tinggi begitu juga sebaliknya |
| Saran | Berdasarkan hasil penelitian dari Implementasi sistem Klasifikasi dan Pencarian Buku Referensi Akademik Menggunakan Metode Naïve Bayes Classifier, penulis memiliki beberapa saran untuk pengembangan selanjutnya, antara lain:  1. Perlu adanya penelitian yang lebih mendalam untuk mengetahui bagaimana pengaruh kinerja metode Naïve Bayes Classifier ini terhadap kategori lain.  2. Memperbanyak jumlah data latih sehingga tingkat akurasi pada proses klasifikasi semakin tinggi serta menghindari kemungkinan hasil klasifikasi yang tidak masuk dalam kategori manapun. |

|  |  |
| --- | --- |
| **Judul artikel** | SISTEM KLASIFIKASI DAN PENCARIAN JURNAL DENGAN MBNGGUNAKAN VECTOR SPACE MODEL |
| **Penulis** | Ainalia Indranandita , Budi Susanto , Antonius Rachmat C |
| **Judul Jurnal & Tahun Penerbitan** | JURNAL INFORMATIKA, NOVEMBER 2OO8 |
| **Masalah utama yang diangkat** | Permasalahan yang muncul adalah bagaimana sistem dapat  Melakukan pengelompokan dan pencarian jurnal yang relevan untuk memenuhi kebutuhan konsumen. |
| **Metodologi** | Metode Vector Space Model : Metode Vector Space Modelalau Term Vector Model adalah sebuah model aljabar untuk menggambarkan dokumen teks (beberapa objek) sebagai vektor dari identifier. Biasanya digunakan dalam penyaringan informasi (information filtering), penemuan informasi (information retrieval), indexing dan pemberian ranking yang saling relevan. |
| **Hasil Penelitian** | hasil pengujian dari sistem pencarian dengan VSM ditambah dengan suatu label kategori yang diasumsikan sebagai hasil sistem klasifikasi dengan NB. Kemudian label kategori akan dipergunakan dalam pencarian detilterhadap metadata = kategori. Analisis sistem klasifikasi dan pencarian akan difokuskan pada nilai recalldan precision yang diperoleh. |
| **Kesumpulan & Saran** | Berdasarkan hasil analisis dan implementasi sistem, maka dapat disimpulkan:   * Sistem pencarian dengan metode Vector Space Modetpendekatan Cosrne menghasilkan recal/ sebesar 54.8o/o dan preclsion sebesar 60.7%. * Feature Selection dari sumber data pelatihan dan fesfer yang berbeda menghasilkan precision kurang dari 60%, karena jangkauan topic   pembicaraan yang cukup berbeda.  Penggunaan metadata (hasil klasifikasi) dalam proses pencarian Oapit meningkatkan tingkat recall. Saran untuk pengembangan dan perbaikan sistem ini adalah: r' '   * Perlu adanya perbaikan struktur data untuk mempercepat proses text mining, karena semakin banyak jumlah data, semakin lama pula proses yang dibutuhkannya. * Pengembangan sistem, yaitu sistem dapat menambahkan kategori baru.   Pengembangan pencarian dengan kata kunci berupa frasa atau dengan penggunaan Boolean Operator (misal: OR, AND). |
| **Kelebihan** | tingkat ketepptan , kumpulan dokumen yang relevan,kumpulan dokumen yang ditemukan |