

## PENERAPAN ALGORITMA K-NEAREST NEIGHBOR PADA PENGKLASIFIKASIAN DOKUMEN BERITA ONLINE

Andi Yulia Muniar<sup>1</sup>, Pasnur<sup>2</sup>, Kiki Ria Lestari<sup>3</sup>

Program Studi Sistem Informasi, STMIK AKBA

Jl. P. Kemerdekaan Km. 9 No. 75 Makassar, telp/fax : 0411-588371

email: andiyulia@akba.ac.id<sup>1</sup>, pasnur@akba.ac.id<sup>2</sup>, kiki14@mhs.akba.ac.id<sup>3</sup>

### ABSTRAK

Jumlah dokumen berita online meningkat dengan pesat. Dokumen-dokumen tersebut perlu dikelompokkan agar memudahkan menemukan berita pada topik terkait. Pengelompokan secara manual menggunakan bantuan manusia tidak efisien, terutama dengan kondisi jumlah dokumen yang sangat banyak. Pada penelitian ini diusulkan metode pengelompokan atau pengklasifikasian dokumen berita secara otomatis. Penelitian ini bertujuan untuk membuat suatu sistem yang mampu mengklasifikasikan dokumen berita *online* secara otomatis menggunakan metode *k-Nearest Neighbour (KNN)* dan diterapkan menggunakan bahasa pemrograman *PHP Hypertext Preprocessor (PHP)* serta database *MySQL*. Hasil pengujian pada dokumen berita yang dikumpulkan dari situs berita online *TribunNews* menunjukkan nilai akurasi rata-rata 89,9% pada nilai  $k=7$ . Hasil penelitian menunjukkan bahwa metode *KNN* memiliki tingkat akurasi yang baik dalam melakukan pengklasifikasian dokumen berita secara otomatis.

**Kata Kunci:** Dokumen Berita, Klasifikasi, k-Nearest Neighbor

### ABSTRACT

*The number of online news documents is increasing rapidly. These documents need to be classified to make it easier to find news on related topics. Manual classification by using human assistance is not efficient, especially with the large number of documents. This research proposes an automatic classification method for news documents. This study aims to create a system capable of automatically classifying online news documents using the k-Nearest Neighbor (KNN) method and implemented using the PHP Hypertext Preprocessor (PHP) programming language and the MySQL database. The experimental results on news documents collected from the online news site TribunNews showed an average accuracy value of 89.9% at a value of  $k = 7$ . The results show that the KNN method has a good level of accuracy in automatic news documents classification.*

**Keywords:** news documents, classification, k-nearest neighbor

### 1. Pendahuluan

Teknologi informasi merupakan salah satu hal yang tidak akan lepas dari kehidupan manusia. Tanpa adanya teknologi, manusia akan kesulitan dalam berkomunikasi dan menyampaikan informasi. Perlu adanya sistem yang secara otomatis yang dapat

mengelompokkan berita sesuai kategori berita dengan menggunakan *text mining*. Pada umumnya, berita yang disampaikan dalam website terdiri dari beberapa kategori seperti berita politik, olahraga, ekonomi, kesehatan, dan lain-lain (sebagai contoh pada website *kompas.com*, *detik.com*, dan *vivanews.com*). Sejauh ini, mengelompokkan berita dalam beberapa kategori tersebut dilakukan oleh editor

secara manual. Prosesnya, sebelum diunggah harus terlebih dahulu diketahui isi berita secara keseluruhan, Hal ini membutuhkan ketelitian dan waktu yang tidak sebentar dalam sistem pengelompokan. Oleh karena itu, perlu adanya sistem yang secara otomatis dapat mengelompokkan berita sesuai dengan kategori berita dengan menggunakan *text mining*. Dokumen berita akan diproses menggunakan *teks preprocessing* dan kemudian akan dilakukan proses klasifikasi menggunakan metode *k-Nearest Neighbour* berdasarkan nilai *k* terdekat.

## 2. Kajian Literatur

### 2.1 Dokumen Berita

Dokumen merupakan sumber data yang digunakan untuk melengkapi penelitian, baik berupa sumber tertulis, film, gambar (foto), dan karya-karya monumental, yang semuanya itu memberikan informasi bagi proses penelitian (Gharuty, 2009). Menurut Kamus Besar Bahasa Indonesia (KBBI), Berita adalah cerita atau keterangan mengenai kejadian atau peristiwa yang hangat; kabar, Laporan, Pemberitahuan; pengumuman. Berita memiliki aliran yang dinamis dimana informasi yang terkandung didalamnya memungkinkan sebuah informasi baru yang tidak ada dalam dokumen sebelumnya (Widodo, 2016).

### 2.2 Klasifikasi

Klasifikasi merupakan suatu metode untuk mengelompokkan sebuah objek ke dalam kelompok atau kelas tertentu. Klasifikasi dokumen adalah bidang penelitian dalam perolehan informasi dengan mengembangkan metode untuk menentukan atau mengkategorikan dokumen ke dalam satu atau lebih kelompok yang sebelumnya telah diakui secara otomatis berdasarkan isi dokumen.

Klasifikasi dokumen bertujuan untuk mengklasifikasikan dokumen tidak terstruktur ke dalam kelompok yang menggambarkan isi dari dokumen. Dokumen dapat berupa teks dokumen seperti artikel berita (Efendi & Malik, 2012).

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning (fase training)*, dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi (Rohman, 2012).

### 2.3 *k-Nearest Neighbor (kNN)*

*k-Nearest Neighbor (kNN)* adalah suatu metode yang menggunakan algoritma *supervised* dimana data testing yang baru diklasifikasikan berdasarkan mayoritas kelas pada *kNN* (Fauzi, 2017). Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasarkan *atribut* dan data *training*. Pengklasifikasian tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Prinsip dari *kNN* adalah menemukan *k* objek dari data *training* yang paling dekat dengan data *testing*. Algoritma *kNN* sangat sederhana, bekerja berdasarkan pada jarak terdekat dari data testing dengan data *training* untuk menentukan *k*-tetangga terdekat (*kNN*), kemudian diambil mayoritas dari *kNN* untuk dijadikan prediksi dari data *testing*.

*KNN* memiliki beberapa kelebihan yaitu ketangguhan terhadap *training data* yang memiliki banyak *noise* dan efektif apabila *training data*-nya besar. Sedangkan, kelemahan *KNN* adalah *KNN* perlu menentukan nilai dari parameter *k* (jumlah dari tetangga terdekat), *training* berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang

harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap *query instance* pada keseluruhan *training sample*.

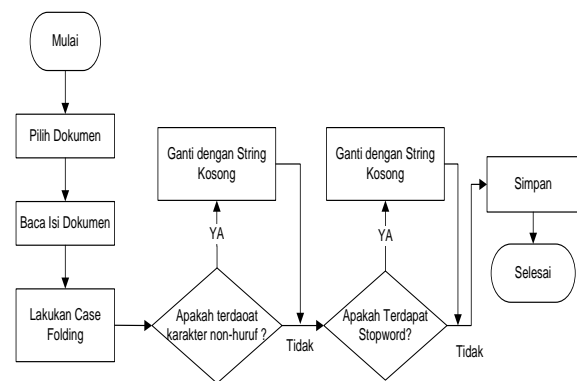
Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk testing data (yang klasifikasinya tidak diketahui). Jarak dari vektor baru yang ini terhadap seluruh vektor *training sample* dihitung dan sejumlah k buah yang paling dekat diambil. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi menjadi lebih baik. cara kerja dari algoritma kNN dimulai dengan Input data testing kemudian menetapkan nilai k setelah itu menghitung jarak kedekatan menggunakan rumus *cosine similarity*, selanjutnya mengurutkan jarak dan memilih kelas mayoritas sebagai hasil akhir dari klasifikasi dokumen berita. Menghitung jarak kedekatan menggunakan *Cosine Similarity* Karena dalam memperhitungkan jarak selain pada data uji dan data sampel, sudut yang dibentuk antar 2 data tersebut juga diperhitungkan dalam penentuan data yang termasuk dalam *Nearest Neighbors*. Saat nilai *similarity* ditemukan, maka akan ditentukan data yang merupakan *Nearest Neighbours* dari data uji sebanyak nilai K. Nilai K yang dimaksud adalah banyaknya tetangga terdekat atau data yang memiliki kemiripan tertinggi di antara semua data yang ada.

#### 2.4 Teks mining

*Teks mining* secara umum adalah teori tentang pengolahan koleksi dokumen dalam jumlah besar yang ada dari waktu ke waktu dengan menggunakan beberapa analisis, tujuan pengolahan teks adalah mengetahui dan mengekstrak informasi yang berguna dari sumber data dengan identifikasi dan eksplorasi pola menarik. Dalam kasus

*text mining*, sumber data yang dipergunakan adalah kumpulan atau koleksi dokumen tidak terstruktur dan memerlukan adanya pengelompokan untuk diketahui informasi sejenis (Lestari, 2014).

Seperti *data mining*, aplikasi *text mining* pada suatu studi kasus, harus dilakukan sesuai prosedur analisis. Langkah awal sebelum suatu data teks dianalisis menggunakan metode *dalam text mining* adalah melakukan *preprocessing teks* (*case folding, tokenizing, filtering, stemming*). Selanjutnya, setelah didapatkan data yang siap diolah, analisis *text mining* dapat dilakukan.



Gambar 1. Tahapan Persiapan dokumen Berita (preprocessing)

#### 2.5 Pembobotan Kata (*Term Weighting*)

Hal yang perlu diperhatikan dalam pencarian informasi dari koleksi dokumen yang *heterogen* adalah pembobotan *term*. *Term* dapat berupa kata, *frase* atau unit hasil *indexing* lainnya dalam suatu dokumen yang dapat digunakan untuk mengetahui konteks dari dokumen tersebut. Karena setiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, maka untuk setiap kata tersebut diberikan sebuah indikator, yaitu *term weight*. *Term*

*weighting* atau pembobotan *term* sangat dipengaruhi oleh hal-hal berikut ini :

1. *Term Frequency* (tf), yaitu faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu *term* (tf tinggi) dalam dokumen, maka akan semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.
2. *Inverse Document Frequency* (idf) , yaitu pengurangan dominansi *term* yang sering muncul di berbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul di berbagai dokumen, dapat dianggap sebagai *term* umum (*common term*) sehingga tidak penting nilainya. Sebaliknya faktor kejarangmunculan kata (*term scarcity*) dalam koleksi dokumen harus diperhatikan dalam pemberian bobot. 'Kata yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting (*uncommon tems*) disbanding dengan kata yang muncul pada banyak dokumen. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*inverse document frequency*).

Metode TF-IDF merupakan metode pembobotan *term* yang banyak digunakan sebagai metode pembanding terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot *term* *t* dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency*.

## 2.5 Cosine Similarity

*Cosine similarity* berfungsi untuk membandingkan kemiripan antar dokumen, dalam hal ini yang dibandingkan adalah *query* dengan dokumen latih . Dalam menghitung *cosine similarity* pertama yaitu melakukan perkalian skalar antara *query* dengan dokumen kemudian dijumlahkan, setelah itu melakukan perkalian antara panjang dokumen dengan panjang *query* yang telah dikuadratkan, setelah itu dihitung akar pangkat dua. Selanjutnya hasil perkalian skalar tersebut di bagi dengan hasil perkalian panjang dokumen dan *query*. Rumus dapat dilihat sebagai berikut :

$$\text{cosSim}(d_j, q_k) = \frac{\sum_{i=1}^n (td_{ij} \times tq_{ik})}{\sqrt{\sum_{i=1}^n td_{ij}^2 \times \sum_{i=1}^n tq_{ik}^2}}$$

di mana :

$\text{ccosSim}(d_j, q_k)$  : tingkat kesamaan dokumen dengan query tertentu

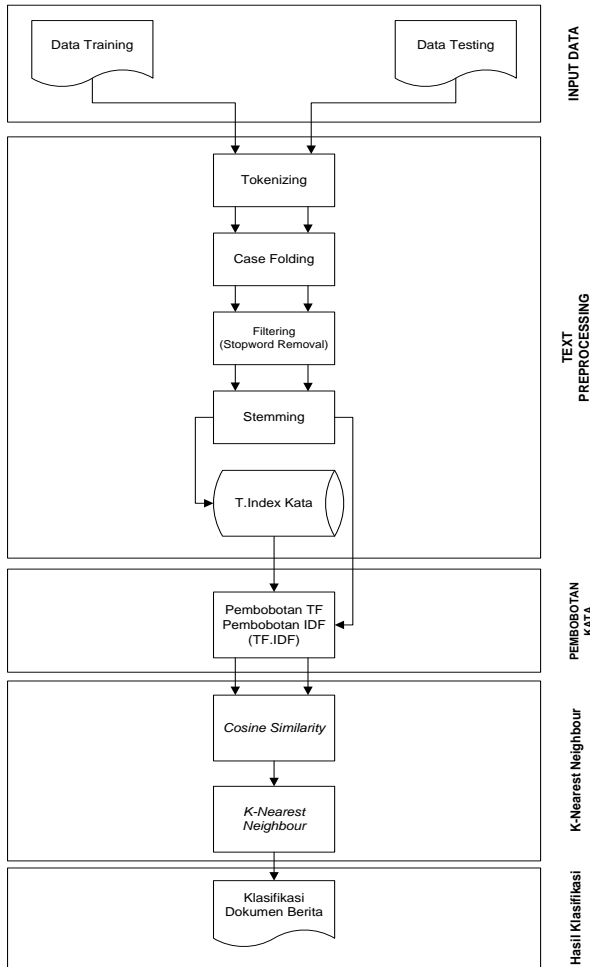
$td_{ij}$  : term ke-I vector untuk dokumen ke-j

$tq_{ik}$  : term ke-I vector untuk dokumen ke-k

$n$  : jumlah *term* yang unik dalam data set

## 3. Metode Penelitian

Pengolahan data diawali dengan mengumpulkan berita yang didapatkan dari situs *Tribunnews.com* kemudian mengubahnya menjadi dokumen *plaint text* (TXT). Dokumen uji coba yang digunakan berjumlah 60 data yang merupakan kumpulan dari beberapa kategori seperti : Sport, Kesehatan, dan Techno dengan 6 data testing (uji) dan sebanyak 54 data training (latih) yang akan di proses dalam tahap *preprocessing* (*case folding, tokenizing, filtering, stemming*).

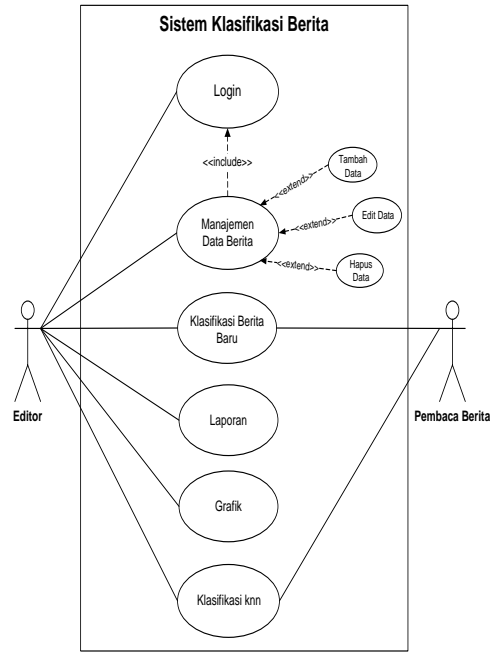


Gambar 2. Blok Diagram sebagai gambaran umum sistem

### 3.2 Use Case

Use case diagram dapat dilihat pada gambar 3, dimana terdapat 2 aktor yang terlibat dalam sistem ini, yaitu selaku *admin* (*editor*) dan *user* (*pembaca berita*).

*Admin* adalah aktor yang memiliki hak akses dalam mengelola keseluruhan sistem, termasuk didalamnya manajemen data berita meliputi penambahan, mengedit atau menghapus berita. *Admin* memiliki hak untuk dapat melihat dan mengetahui hasil klasifikasi berita serta hak untuk mengakses pengujian terhadap metode yang digunakan. *User* adalah pengguna sistem yang memiliki hak untuk melihat klasifikasi berita, melakukan klasifikasi berita baru



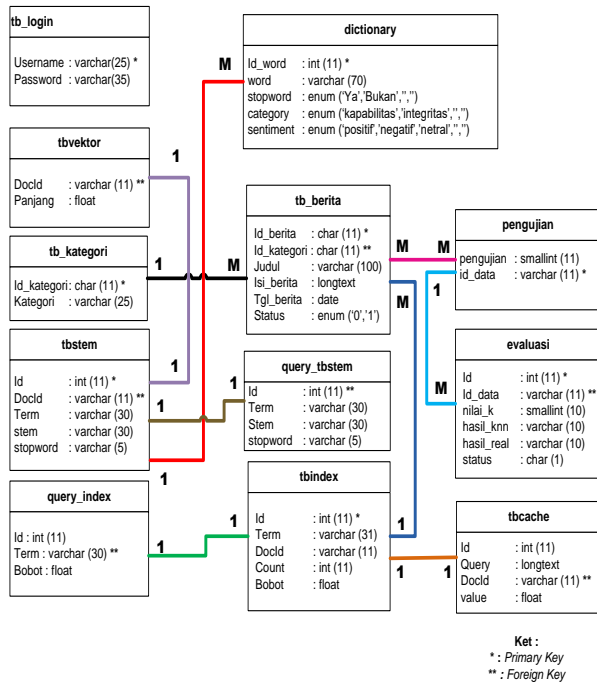
Gambar 3. Use Case Diagram

### 3.3 Perancangan Basis Data

Perancangan basis data ini untuk menggambarkan data yang digunakan dan direlasikan dalam pengaplikasian sistem klasifikasi ini. Terdapat beberapa bagian data base yang terhubung yaitu database login, database kategori, database berita, database vektor, database stemming, database index, database cach, database query table stemming, query index, table dictionary, table evaluasi, table pengujian, dan database evaluasi. Perancangan ini menunjukkan adanya editor dan pembaca berita yang dihungkan oleh table klasifikasi dokumen berita. Untuk melihat hubungan antar masing – masing entitas pada sistem ini dapat dilihat pada gambar 4.

## 4. Hasil dan Pembahasan

Hasil penelitian dari penerapan KNN dalam pengklasifikasian dokumen berita online. Aktifitas akan dimulai oleh admin (*editor*) terlebih dahulu dengan mengakses menu Login seperti yang terlihat pada gambar 5.



Gambar 4. Relasi Tabel Database

Gambar 5. Form Login Admin (editor)

Setelah Admin login maka akan dihadapkan pada tampilan halaman utama yang terdiri dari 3 menu utama yaitu menu Dashboard, manajemen data, dan klasifikasi KNN yang dapat dilihat pada Gambar 6.

Gambar 6. Form Halaman Utama

Dalam menu manajemen data terbagi menjadi dua sub menu, kategori berita dan dokumen berita dalam sub menu dokumen berita ini berisi dua sub-sub menu yaitu, menu berita yang berisi keseluruhan data dokumen berita yang digunakan dan sub-sub menu data pengujian yang menampilkan ID dari berita yang akan menjadi berita latih (*training*) dan ID berita yang digunakan sebagai berita test (*testing*). Seperti yang terlihat pada gambar 7.

Gambar 7. Form Manajemen data

selanjutnya pada menu Klasifikasi KNN terdapat empat sub menu berupa, KNN, Klasifikasi Berita Baru, Laporan, Grafik. Menu ini akan menampilkan 2 tabel utama yang akan menunjukkan jarak dari semua data yang dekat dengan data testing serta tabel jarak dengan semua data yang dekat dengan data testing yang sudah diurutkan sesuai dengan jarak yang paling dekat dan akan menampilkan hasil klasifikasi berita baru seperti yang terlihat pada Gambar 8 dan Gambar 9.

Tabel Jarak dengan semua data yang dekat dengan data testing yang sudah diurutkan jarak terdekat:

No	ID Berita	Judul	Jarak	Kategori
1	BR0031	Valentino Rossi Diecebut Tengah Tembosi pada Rivalnya, Marc Marquez	0.151853	SPORT
2	BR0002	Firman Abdul Kholik Pasukan Kemenangan Indonesia Asas Korset, Tim Thomas Juara Grup D	0.047721	SPORT
3	BR0013	Rasailan 10 Manfaat Luar Biasa untuk Badan dan Kebiasaan Baik Berguna Kali Setiap Hari	0.044591	KESEHATAN
4	BR0004	Zinedine Zidane mundur, bagei 'tedakan boni' yang mengguncangkan Real Madrid	0.038999	SPORT
5	BR0003	Tim Vidi Putra Indonesia Juara Piala Lemuelposbank 2018 di Vietnam	0.032699	SPORT
6	BR0037	Simall In Dua Syarat dan Rafael Nadal untuk Menjadi Petenis Elite	0.024676	SPORT
7	BR0008	Ada Bahaya Mengintai Jika Tudor Sambil Menyajikan Kipas Angin, Berikut Sebabnya	0.020891	KESEHATAN
8	BR0054	Indonesia Raih Kemenangan dalam Epson Regional App Challenge	0.019977	TECHNO
9	BR0022	Usa 20-an Rawan Kena Gigi Senekat, Sebenamya Apa Penyebabnya?	0.018138	KESEHATAN

Gambar 8. Form klasifikasi KNN

Query

Momen kehamilannya membuat seluruh wanita di dunia ini merasa sempurna dan bahagia. Hal itu pula yang dirasakan dokter Resia Broto Asmoro, ia mengaku senang saat menjalani proses kehamilan anak pertamanya maupun kedua. Wanita yang masih terlihat cantik meski telah memiliki anak itu kemudian menyebutkan sejumlah masalah yang pernah ia alami saat hamil Ania dan baby Yoda. Perubahan hormonal hingga stretch mark, merupakan beberapa hal yang membuatnya sempat khawatir. Baik itu saat hamil muda, hingga dirinya hendak melahirkan kedua anaknya. "Tentu sebabnya itu hamil harus menjaga ketebalan kulitnya," tegas Resia. Namun, permasalahan baru ternyata muncul saat dirinya memasuki usia kehamilan "tujuh". Bibir tubuhnya yang semakin besar, membuat kakinya selalu merasa pegal karena harus menopang badannya itu. "Kalau sudah hamil besar, beda lagi khasiatnya, kali pegal banget karena menopang tubuh yang makin berat kan badan ya," kata Resia. Setelah masa kehamilannya berakhir, Resia pun kini tengah menikmati momen indahannya menjadi seorang ibu muda yang sibuk mengurus kedua buah hatinya.

Hasil Klasifikasi dengan KNN adalah: KESEHATAN

Perhitungan Proses Klasifikasi Dengan KNN

1. Tabel Jarak dengan semua data yang dekat dengan data testing

No	ID Berita	Judul	Jarak	Kategori
1	BR0001	Sirkuit Jerez Target Maverick Vinales Benihnya	0.0386073	SPORT
2	BR0002	Firman Abadi Kholik Pastikan Kemenangan Indonesia Atas Korser Tim Thomas Juara Grup B	0.0247154	SPORT
3	BR0003	Tin Vui Putra Indonesia Juara Pala Leningpostbani 2018 di Vietnam	0.0114541	SPORT
4	BR0004	Zinedine Zidane mundur, bage 'tedakan boni' yang mengguncangkan Real Madrid	0.0091103	SPORT

2. Tabel Jarak dengan semua data yang dekat dengan data testing yang sudah diurutkan jarak terdekat

No	ID Berita	Judul	Jarak	Kategori
1	BR0002	Usia 20-an Rawan Kera-Gigi Sensitif, Sebenarnya Apa Penyebabnya?	0.162792	KESEHATAN
2	BR0013	Rasakan 10 Manfaat Luar Biasa untuk Badan dan Kebiasaan Baik Berjalan Kali Setiap Hari	0.14341	KESEHATAN
3	BR0021	Berjuta Manfaat Daun Kelor Buat Ibu Hamil dan Menyusui	0.100807	KESEHATAN
4	BR0011	Manfaat Kunyit bagi Kesehatan	0.101665	KESEHATAN

Gambar 9. Form Hasil Klasifikasi Berita Baru

Pada menu laporan seperti yang terlihat pada Gambar 10 berisi tampilan keseluruhan dari hasil klasifikasi, terdapat empat button utama yakni 1) Hasil Klasifikasi, laporan yang berisi informasi hasil dari 10 pengujian, kategori sebenarnya, serta kategori dari hasil klasifikasi yg telah dilakukan oleh sistem. 2) Evaluasi K, Laporan mendetail dari nilai *precision*, *recall*, *F1-Measure* serta *akurasi* di sepuluh kali pengujian sesuai dengan inputan nilai k yang telah dipilih. 3) Rekap Per K-Pengujian, laporan ini akan menyatukan keseluruhan nilai dari 10 pengujian dimulai dari k=1,3,5,7,9. 4) Rekap Per K, jenis laporan ini hanya akan menampilkan nilai per nilai k pengujian tanpa menampilkan detail dari isi 10 nilai pengujian seperti dilaporan rekap per k-pengujian sebelumnya.

#### 4.1 Pengujian Akurasi Algoritma KNN

Pengujian keakuratan dalam melakukan klasifikasi dokumen berita menggunakan artikel dalam bentuk file teks.



Gambar 10. Form Menu Laporan

Didalam pengujian ini menggunakan 60 dokumen berita dengan tiga kategori dari media elektronik yaitu, sport, kesehatan dan techno. Dokumen berita yang digunakan terdiri dari dua bagian, diantaranya dokumen pelatihan dan dokumen pengujian. Dokumen pelatihan berperan sebagai data contoh yang akan digunakan dalam proses pelatihan. Sedangkan dokumen pengujian digunakan sebagai data pengujian untuk melihat tingkat akurasi. Dalam pengujian ini, digunakan sebanyak 54 dokumen pelatihan dan 6 dokumen pengujian, hal ini dilakukan setiap pengujian. Rata-rata jumlah kata yang terkandung didalam dokumen berita yang digunakan sebanyak  $\pm 295$  kata. Pada setiap proses klasifikasi ataupun pelatihan, semua dokumen yang digunakan harus melewati proses *text mining* terlebih dahulu, yaitu proses *case folding* (mengubah teks menjadi karakter, membuang karakter selain a-z, menghilangkan symbol, tanda baca dan numerik) *tokenizing* (pemecahan kata), *filtering* (penyaringan kata), *stemming* (penghilangan imbuhan menjadikan kata dasar). Dalam melakukan pengujian terhadap hasil keluaran sistem menggunakan empat

jenis pengujian *precision*, *recall*, *f-measure*, dan *accuracy*. Hasil akurasi dari masing-masing nilai k dapat dilihat grafik yang ditampilkan pada Gambar 10



Gambar 10. Grafik akurasi KNN

Hasil grafik akurasi knn didapatkan bahwa pada  $k=1$  sebesar 88.8,  $k=3$  sebesar 87.6,  $k=5$  sebesar 88.8,  $k=7$  sebesar 89.9, sedangkan  $k=9$  sebesar 85.5. maka dari hasil tingkat akurasi tertinggi didapatkan dari nilai  $k=7$  yakni 89.9.

## 5. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dalam proses klasifikasi dokumen  $k=7$  merupakan nilai dengan tingkat akurasi terbaik sebesar 89.9% dengan menggunakan 60 dokumen yang terdiri dari 6 data uji dan 54 data latih dari 3 kategori yaitu, sport, kesehatan dan techno. Dari 60 dokumen uji hasil klasifikasi yang dilakukan oleh sistem bahwa sebanyak 44 dokumen uji memiliki nilai yang benar yang artinya hasil klasifikasi dan kategori kelas sebenarnya telah sesuai sedangkan 16 dokumen sisanya diklasifikasikan dalam kategori yang berbeda, hal ini dikarenakan kurangnya kata-kata yang penting yang mencerminkan suatu dokumen dan juga terdapat kata-kata yang dominan ke kategori lain, inilah

yang menjadi penyebab kesalahan dalam pengklasifikasian.

## Daftar Pustaka

- Efendi, R., Malik, R. F., & U Mila Sari, J. ; (2012). *Klasifikasi Dokumen Berbahasa Indonesia Menggunakan Naive Bayes Classifier*. *Research in Computer Science and Applications*, I(1), 7–13.
- Fauzi, A. M. 2017. Text Mining. <http://malifauzi.lecture.ub.ac.id/2017/09/text-mining-20172018/>, diakses 4 April 2020.
- Gharuty, F. (2009). Studi Dokumen Dalam Penelitian Kualitatif. <https://adzelgar.wordpress.com/2009/02/02/studi-dokumen-dalam-penelitian-kualitatif/> diakses tanggal 20 April 2019
- Lestari, M. E. I. (2014). Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) untuk Mendeteksi Penyakit Jantung. *Faktor Exacta*, 7(4), 366–371. Jakarta Selatan Universitas Indraprasta
- Rohman, A. (2012). Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa. Semarang : Fakultas Teknik Universitas Pandanaran Semarang
- Widodo, P., et al. (2016). Klasifikasi Kategori Dokumen Berita Berbahasa Indonesia Dengan Metode Kategorisasi Multi-Label Berbasis Domain Specific Ontology, II(2), 126–137. Surabaya : Institut Teknologi Sepuluh Nopember (ITS)