

KLASIFIKASI BERITA OLAHRAGA PADA PORTAL BERITA ONLINE DENGAN METODE K-NEAREST NEIGHBOUR (KNN) DAN LEVENSHTTEIN DISTANCE

Rizal, S.Si., M.IT¹, Hafizh Al Kautsar Aidilof, S.T., M.Kom², Wahyu Kurniawan³

Teknik Informatika Universitas Malikussaleh Lhokseumawe

Jl. Batam Kampus Bukit Indah Gedung Teknik Informatika, Aceh Utara, 141
Indonesia

³Mahasiswa Teknik Informatika Fakultas Teknik Universitas Malikussaleh

^{1 2}Dosen Teknik Informatika Fakultas Teknik Universitas Malikussaleh

e-mail: wahyukurniawan1011@gmail.com

Abstrak — Dalam perkembangannya, pemberitaan mengenai olahraga pada portal berita online berbasis web semakin berkembang. Minat masyarakat untuk membaca berita olahraga online semakin meningkat. Sehingga pertumbuhan berita olahraga pada portal berita online semakin hari semakin berkembang. K-Nearest Neighbour (KNN) adalah metode klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Pada teori informasi dan ilmu komputer, Levenshtein Distance merupakan matriks untuk mengukur nilai jumlah perbedaan antara 2 string yaitu string sumber (s) dan string target (t). Pada penelitian ini menggunakan data sebanyak 1000 data latih dan 60 data uji. Untuk melakukan pengujian terhadap sistem maka yang akan diinput adalah berita uji, dan tiap berita uji akan diuji berdasarkan kemiripan tiap kalimatnya dengan data latih. Hasil pengujian data uji dengan 1000 data latih dengan jumlah minimal kalimat yang masuk ke dalam kelas olahraga sebesar 0%, 50%, 90% dari jumlah seluruh kalimat data yang diuji menghasilkan akurasi sebesar 100%, 48,3% dan 13,3%.

Kata kunci : *Berita Olahraga, Portal Berita Online, KNN, Levenshtein Distance.*

PENDAHULUAN

Berita adalah suatu tulisan yang mengandung informasi dan fakta tentang suatu hal. Olahraga adalah topik yang sangat menarik dijadikan berita serta memiliki banyak pembaca yang selalu antusias dalam mengakses informasinya. Dalam perkembangannya, pemberitaan mengenai olahraga pada portal berita online berbasis web semakin berkembang. Hal ini tidak terlepas dari berkembangnya teknologi informasi dan komunikasi yang begitu pesat. Menurut (Firmansyah, 2018), portal berita olahraga online seperti bola.com, bolasport.com, bolalob.com menempati urutan teratas situs portal berita olahraga online yang paling banyak dikunjungi.

Pengklasifikasian berita dilakukan oleh pengelola dari portal berita online yang dibantu dengan editor konten berita. Namun, dalam mengklasifikasi berita ke dalam kategori-kategori tersebut untuk saat ini masih dilakukan secara manual, artinya dalam mengunggah berita pengunggah harus terlebih dahulu mengetahui isi dari berita yang akan diunggah secara keseluruhan untuk selanjutnya dimasukkan ke dalam kategori yang tepat (Kurniawan, Effendi, & Sitompul, 2012). Bila jumlah berita semakin banyak, maka dapat semakin merepotkan pengelola dan editor portal berita dalam mengklasifikasikan berita, karena harus membaca isi berita satu per satu secara keseluruhan terlebih dahulu. Masalah lain yang muncul adalah bila terdapat berita dengan kategori yang berbeda namun memiliki kemiripan dalam kata dan kalimat.

Hal ini tentu dapat membuat pembaca kesulitan dan membutuhkan waktu yang tidak sedikit untuk mencari kategori berita yang ingin dibacanya jika berita belum diklasifikasikan, sehingga perlu adanya sistem yang dapat secara otomatis mengklasifikasikan berita menggunakan salah satu cabang *data mining*, yaitu *text mining*.

LANDASAN TEORI

Text Mining

Menurut (Kurniawan et al., 2012) text mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi. Dimana, text mining merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. Selain digunakan untuk klasifikasi, text mining juga digunakan untuk menangani masalah clustering, information extration, dan information retrieval.

Adapun tugas khusus dari text mining antara lain yaitu pengkategorisasian teks (text categorization) dan pengelompokan teks (text clustering). Dalam proses text mining ini, terdapat proses text preprocessing, yaitu proses pemrosesan data teks sebelum melakukan proses mining yang sebenarnya. Ide awal pembuatan text mining adalah untuk menemukan pola-pola informasi yang dapat digali dari suatu teks yang tidak terstruktur (Asiyah & Fithriasari, 2016). Dengan demikian, text mining mengacu juga kepada istilah text data mining atau penemuan pengetahuan dari basis data teks. Saat ini, text mining telah mendapat perhatian dalam berbagai bidang, antara lain dibidang keamanan, biomedis, pengembangan perangkat lunak dan aplikasi, media online, pemasaran, dan akademik. Seperti halnya dalam data mining, aplikasi text mining pada suatu studi kasus, harus dilakukan sesuai prosedur analisis. Langkah awal sebelum suatu data teks dianalisis menggunakan metode-metode dalam text mining adalah melakukan pre processing teks. Selanjutnya, setelah didapatkan data yang siap diolah, analisis text mining dapat dilakukan.

Klasifikasi Teks

Menurut (Nurhadi, 2015) klasifikasi teks merupakan bagian penting dari text mining. Klasifikasi teks berdasarkan para ahli, bagaimana mengklasifikasikan dokumen didalam beberapa kategori-kategori (dalam Kaur, 2013). Klasifikasi atau kategorisasi teks adalah proses penempatan suatu dokumen ke suatu kategori atau kelas sesuai dengan karakteristik dari dokumen tersebut. Dalam text mining, klasifikasi mengacu kepada aktifitas menganalisis atau mempelajari himpunan dokumen teks preclassified untuk memperoleh suatu model

atau fungsi yang dapat digunakan untuk mengelompokkan dokumen teks lain yang belum diketahui kelasnya ke dalam satu atau lebih kelas pre-defined (dalam Sebastiani, 2002).

Algoritma K-Nearest Neighbour (KNN)

K-Nearest Neighbour (KNN) menurut (Dzikrulloh & Setiawan, 2017)(dalam Tri, 2010) adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training sample. Diberikan suatu titik query, selanjutnya akan ditemukan sejumlah K objek atau titik training yang paling dekat dengan titik query. Nilai prediksi dari query akan ditentukan berdasarkan klasifikasi tetangga. Sebelum melakukan perhitungan dengan metode K-Nearest Neighbour, terlebih dahulu harus menentukan data latih dan data uji. Kemudian akan dilakukan proses perhitungan untuk mencari jarak menggunakan Metode Manhattan. Persamaan Manhattan adalah sebagai berikut :

$$d(a, b) = \sum_{i=0}^n |X_i - Y_i|$$

dimana: $d(a, b)$ = jarak Manhattan

x = data 1

y = data 2

i = fitur ke - i

n = jumlah fitur

Algoritma Levenshtein Distance

Pada teori informasi dan ilmu komputer, Levenshtein Distance merupakan matriks untuk mengukur nilai jumlah perbedaan antara 2 string yaitu string sumber (s) dan string target (t). Nilai Levenshtein Distance antara dua kata merupakan nilai minimum dari pengeditan single-character (yaitu insertion, deletion maupun substitution) membutuhkan perubahan pada salah satu kata. Levenshtein Distance antara dua string ditentukan berdasarkan jumlah minimum pengeditan yang diperlukan untuk melakukan transformasi dari satu bentuk string ke bentuk string yang lain. Notasi yang digunakan untuk Levenshtein Distance adalah $LD(s, t)$ dengan s yaitu sumber dan t adalah target. Misalnya, jika source string (s) adalah "tahun" dan target string (t) adalah "tahun" maka nilai Levenshtein Distance adalah 1, dalam hal ini berarti dibutuhkan sebuah operasi yaitu substitution untuk mengubah source string (s) menjadi sama dengan target string (t) (Fahma, Cholissodin, & Perdana, 2018).

Operasi-operasi pada algoritma Levenshtein distance, terdapat tiga macam operasi yang dapat dilakukan (Pratama & Pamungkas, 2016) yaitu:

1. Operasi Penyisipan Karakter (Insertion) Operasi penyisipan karakter berarti menyisipkan karakter ke dalam suatu string. Contohnya string 'disrit' menjadi string 'diskrit', dilakukan penyisipan karakter 'k' di akhir string. Penyisipan karakter tidak hanya dilakukan di tengah string, namun bisa disisipkan diawal maupun disisipkan diakhir string. Ilustrasi:
String 1 d i s k r i t
String 2 d i s - r i t
Insertion k
2. Operasi Penghapusan Karakter (Deletion) Operasi penghapusan karakter dilakukan untuk menghilangkan karakter dari suatu string. Contohnya string 'matematikan' karakter terakhir dihilangkan sehingga menjadi string 'matematika'.

Pada operasi ini dilakukan penghapusan karakter 'n'.
Ilustrasi:

```
String 1 m A t E m a t i k a -
String 2 m A t E m a t i k a n
Deletion      n
```

3. Operasi Penukaran Karakter (Subtitution) Operasi penukaran karakter merupakan operasi menukar sebuah karakter dengan karakter lain. Contohnya penulis menuliskan string 'gimpunan' menjadi 'himpunan'. Dalam kasus ini karakter 'g' yang terdapat pada awal string, diganti dengan huruf 'h'.
Ilustrasi:

```
String 1 H i M p u n a n
String 2 G i M p u n a n
Subtitution H
```

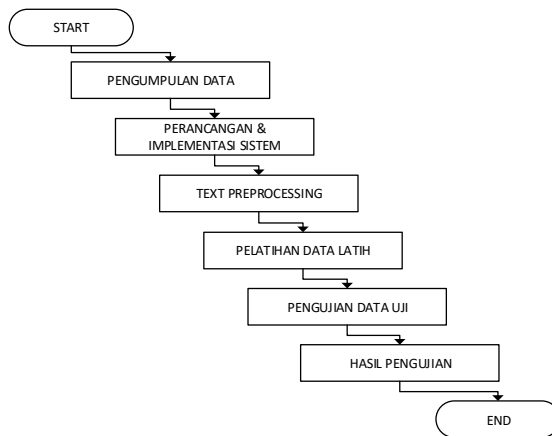
Codeigniter

CodeIgniter adalah sumber kode terbuka yang merupakan salah satu kerangka kerja dari bahasa pemrograman PHP. Kerangka kerja adalah kumpulan instruksi yang dikumpulkan dalam kelas, fungsi-fungsi tertentu yang dibuat untuk memudahkan developer dalam membangun program, yang bisa menghemat waktu dalam pengembangan aplikasi, serta menghindari penulisan sintaks yang berulang-ulang. CodeIgniter adalah kerangka kerja PHP yang ringan, cepat, lengkap dokumentasi penggunaannya, dan bersifat open source. CodeIgniter menerapkan konsep MVC (Model, View, Controller) dalam penulisan kodenya, yang membantu mempercepat pengembangan web oleh developer.

METODOLOGI PENELITIAN

Langkah Penelitian

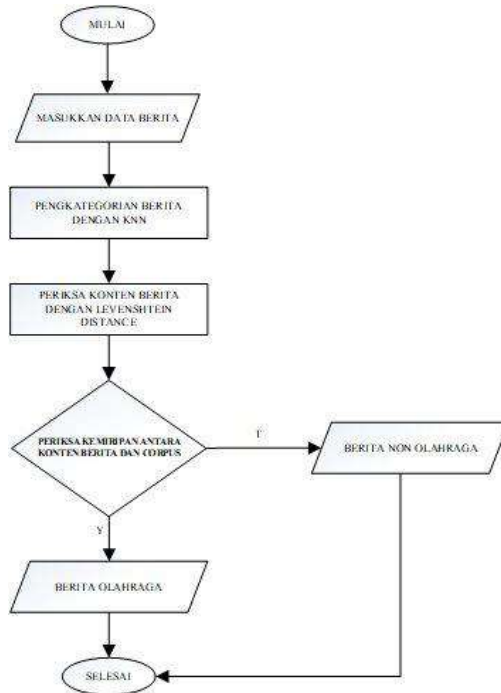
Langkah penelitian yang penulis lakukan dalam proses penelitian tugas akhir ini adalah dengan menggunakan metode waterfall, dimana langkah penelitian diawali dengan tahap pengumpulan data, hingga ke pengujian sistem.



Gambar 1 Langkah-langkah Penelitian

Skema Sistem

Skema klasifikasi berita olahraga pada portal berita online dengan metode K-Nearest Neighbor (KNN) dan algoritma Levenshtein Distance dapat dilihat pada gambar berikut ini.



Gambar 2 Skema Sistem

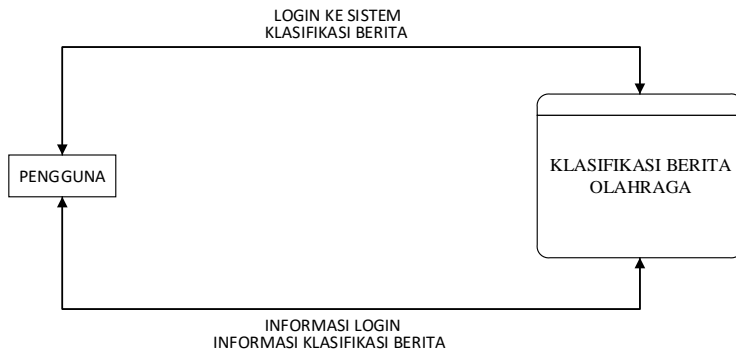
Skema sistem yang bakal berjalan diatas dimulai dari start hingga stop, di mana masing-masing tahapan akan dijelaskan sebagai berikut.

- Mulai**
Mulai adalah proses inialisasi permulaan sistem.
- Masukkan Data Berita**
Masukkan data berita adalah proses menginput data berita yang akan diklasifikasikan oleh sistem.
- Pengkategorian Berita dengan KNN**
Pada tahap ini, berita yang telah diinput akan diklasifikasikan oleh algoritma KNN, Sebelumnya data berita sudah mengalami tahapan teks preprocessing.
- Periksa Konten Berita dengan Levenshtein Distance**

- Pada tahap ini, berita yang telah diinput akan diperiksa tiap kalimatnya dengan algoritma Levenshtein Distance
- e. Periksa Kemiripan antara Konten Berita dan Corpus
Pada tahap ini, data berita yang telah diklasifikasikan dengan metode KNN akan di hitung jarak kemiripan kalimat kontennya dengan corpus (bank data) berita olahraga menggunakan metode Levenshtein Distance.
 - f. Berita Olahraga
Jika nilai jarak kemiripan kalimat konten berita telah memenuhi kriteria nilai minimal yang telah ditetapkan, maka berita yang diklasifikasikan akan masuk ke kategori berita olahraga.
 - g. Berita Non Olahraga
Jika nilai jarak kemiripan kalimat konten berita belum memenuhi kriteria nilai minimal yang telah ditetapkan, maka berita yang diklasifikasikan akan masuk ke kategori berita non olahraga.
 - h. Selesai
Tahap ini adalah akhir dari sistem pengklasifikasian berita olahraga.

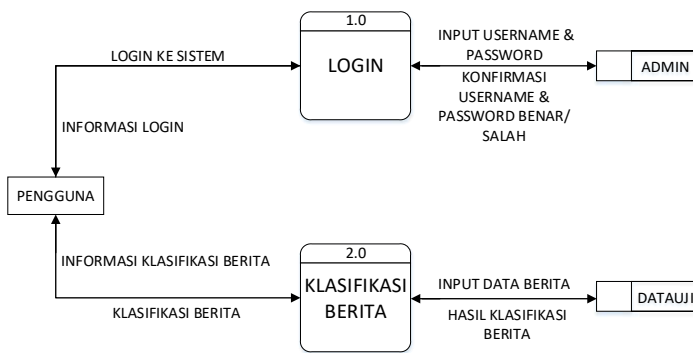
HASIL & PEMBAHASAN

Diagram Context



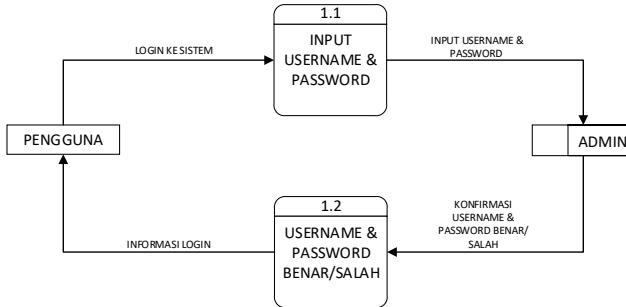
Gambar 3 Diagram Context

Data Flow Diagram Level 1



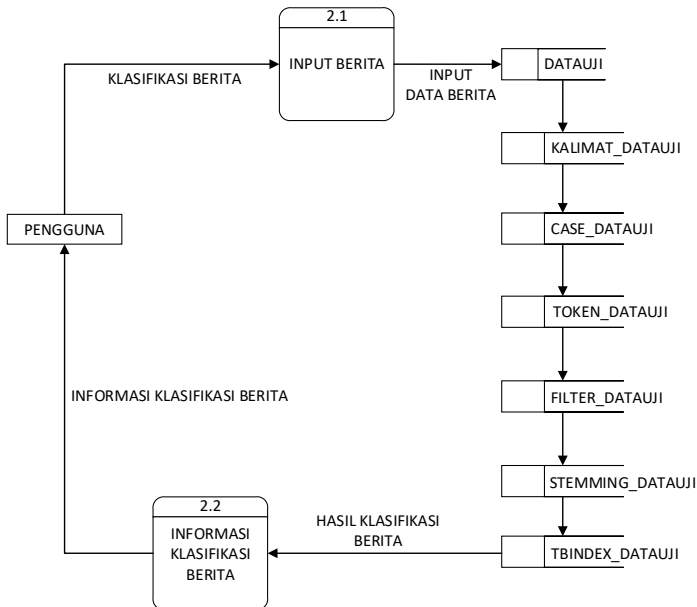
Gambar 4 DFD Level 1

Data Flow Diagram Level 2.0 Proses 1.0 Login



Gambar 5 DFD Level 2 Proses 1.0

Data Flow Diagram Level 2.0 Proses 2.0 Klasifikasi Berita

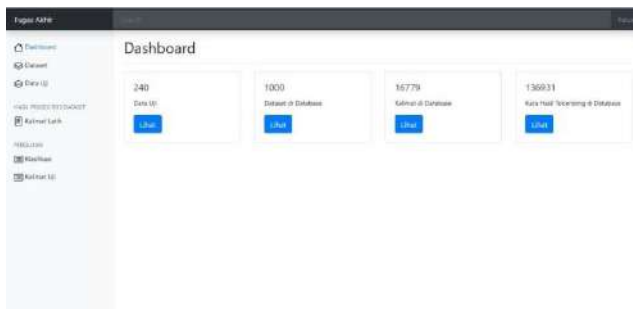


Gambar 6 DFD Level 2 Proses 2.0

Implementasi Antar Muka



Gambar 7 Halaman Login



Gambar 8 Halaman Beranda



Gambar 9 Halaman Klasifikasi Berita

Detail Hasil Pengujian

Judul: Hui Priadi PHS Penangin (re Alam di Polesia Olimpiade)

Jumlah Kalimat: 0

Jumlah Kalimat Olahraga: 0

Klasifikasi: Senja Non Olahraga

| Kalimat | ID | Kemiripan | Klasifikasi | Bobot |
|---|----|-----------|--------------------|-------|
| jelante trady beas sabbu bergabung pelatnas renang kualifikasi olimpiade tokyo | 42 | 50.5062 | Senja Non Olahraga | 0,331 |
| mengikuti olimpiade pra jabatan pra kongres | 29 | 48.9589 | Senja Non Olahraga | 0,523 |
| peraih medali emas games malaya masuk cibai gelombang atlet atlet cabang olahraga | 41 | 56.8621 | Senja Non Olahraga | 1,340 |
| peraihhan jekon puslalat kemah non jelante | 27 | 44.8768 | Senja Non Olahraga | 5,117 |
| beres pra jabatan binaan kemper | 21 | 51.7423 | Senja Non Olahraga | 5,478 |
| menjadi tahap pertamanya akan berangkat aktualisasi | 33 | 44.0678 | Senja Non Olahraga | 6,22 |

Gambar 10 Halaman Detail Hasil Klasifikasi

Pengujian Data Uji

Pada pengujian, data yang diuji merupakan berita dengan kategori yang sama dengan data latih. Dalam penelitian ini, pengujian data akan dilakukan berdasarkan kalimat dalam tiap berita. Untuk menghitung kemiripan tiap kalimat uji dengan kalimat latih akan digunakan penghitungan similaritas Levenshtein Distance dan pengklasifikasiannya menggunakan metode KNN. Tingkat similaritas atau kemiripan minimal levenshtein yang diterapkan pada sistem pada proses pengujian adalah sebesar 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, dan 90%, 100%. Sehingga jika data uji memenuhi nilai kemiripan tersebut maka akan masuk ke dalam kelas olahraga, dan jika kurang dari nilai kemiripan yang telah ditetapkan maka akan masuk ke dalam kelas non olahraga.

Untuk menentukan tiap berita masuk ke dalam kelas olahraga maka jumlah minimal kalimatnya yang masuk ke dalam kelas olahraga yang diterapkan pada sistem pada proses pengujian adalah sebesar 0%, 50%, 90% dari keseluruhan jumlah kalimat. Dimana penghitungan dari kedua metode ini akan menghasilkan nilai kemiripan dan klasifikasi dari data berita yang diujikan.

Untuk menentukan tiap berita masuk ke dalam kelas olahraga maka jumlah minimal kalimatnya yang masuk ke dalam kelas olahraga yang diterapkan pada sistem pada proses pengujian adalah sebesar 0%, 50%, 90% dari keseluruhan jumlah kalimat.

Dimana penghitungan dari kedua metode ini akan menghasilkan nilai kemiripan dan klasifikasi dari data berita yang diujikan.

Dibawah ini merupakan contoh perhitungan manual pengklasifikasian kalimat dari berita yang diuji. Persamaan yang digunakan untuk menentukan suatu berita masuk ke dalam kelas olahraga adalah sebagai berikut:

$$\text{Klasifikasi} = \frac{\sum JK_{OA}}{\sum JK_A} \geq X_{parameter}$$

dimana:

Klasifikasi = Kelas Berita

$\sum JK_{OA}$ = Jumlah Kalimat Olahraga Berita Uji

$\sum JK_A$ = Jumlah Kalimat Berita Uji

$X_{parameter}$ = Jumlah Minimal Kalimat Olahraga yang Diterapkan sistem

Kalimat Uji:

- a. jakarta ketua pbsi agung firman sampurna mengubah program pelatnas bulutangkis

Kalimat Latih Olahraga:

- a. jakarta ketua pp pbsi terpilih agung firman sampurna membangun sport industri bulutangkis
- b. jakarta juara dunia junior muhammad zohri insiden soal bendera merah putih diperpanjang
- c. timnas basket putra filipina sempat memutuskan mundur asian games sepuluh pemain filipina
- d. menpora zainudin amali barang barang bekas asian games pekan olahraga nasional pon papua
- e. seperti anthony ginting jonatan christie menjalani latihan pelatnas pbsi cipayung

Kalimat Latih Non Olahraga:

- a. tujuan menjaga keberlanjutan kegiatan usaha mendukung menjalankan kebijakan peraturan pemerintah tutupnya
- b. jakarta restoran mengambil kebijakan menerjunkan karyawannya berjualan pinggir jalan ngemper

- c. pengelola pizza hut prinsip mendukung kebijakan pemerintah mengatasi pandemi covid
- d. jakarta pasangan pengantin bandung kedatangan tamu diundang numpang makan
- e. kunyit berkhasiat perkuat daya tahan tubuh menyembuhkan peradangan

Pengukuran Similaritas Levenshtein Distance

Kalimat uji : jakarta ketua pbsi agung firman sampurna mengubah program pelatnas bulutangkis

Tabel 1 Pengukuran Similaritas Levenshtein Distance

| Kalimat Latih | Klasifikasi | Tingkat Kemiripan |
|---|---------------------|--------------------------|
| jakarta ketua pp pbsi terpilih agung firman sampurna membangun sport industri bulutangkis | Berita Olahraga | 64,37 |
| jakarta juara dunia junior muhammad zohri insiden soal bendera merah putih diperpanjang | Berita Olahraga | 33,34 |
| timnas basket putra filipina sempat memutuskan mundur asian games sepuluh pemain filipina | Berita Olahraga | 28,19 |
| menpora zainudin amali barang barang bekas asian games pekan olahraga nasional pon papua | Berita Olahraga | 28,41 |
| seperti anthony ginting jonatan christie menjalani latihan pelatnas pbsi cipayung | Berita Olahraga | 34,67 |
| tujuan menjaga keberlanjutan kegiatan usaha mendukung menjalankan kebijakan peraturan pemerintah tutupnya | Berita Non Olahraga | 31,43 |
| jakarta restoran mengambil kebijakan menerjunkan karyawannya berjualan pinggir jalan ngemper | Berita Non Olahraga | 31,52 |

| | | |
|--|---------------------|-------|
| pengelola pizza hut prinsip mendukung kebijakan pemerintah mengatasi pandemi covid | Berita Non Olahraga | 23,17 |
| jakarta pasangan pengantin bandung kedatangan tamu diundang numpang makan | Berita Non Olahraga | 32,05 |
| kunyit berkhasiat perkuat daya tahan tubuh menyembuhkan peradangan | Berita Non Olahraga | 26,92 |

Klasifikasi K-Nearest Neighbour (KNN)

Setelah didapatkan nilai kemiripan antara data kalimat uji dengan data kalimat latih, selanjutnya nilai kemiripan diurutkan dari terbesar hingga ke yang terkecil.

Tabel 2 Pengurutan Hasil Pengukuran Kemiripan

| Kalimat Latih | Klasifikasi | Tingkat Kemiripan |
|---|---------------------|-------------------|
| jakarta ketua pp pbsi terpilih agung firman sempurna membangun sport industri bulutangkis | Berita Olahraga | 64,37 |
| seperti anthony ginting jonatan christie menjalani latihan pelatnas pbsi cipayung | Berita Olahraga | 34,67 |
| jakarta juara dunia junior muhammad zohri insiden soal bendera merah putih diperpanjang | Berita Olahraga | 33,34 |
| jakarta pasangan pengantin bandung kedatangan tamu diundang numpang makan | Berita Olahraga | 32,05 |
| jakarta restoran mengambil kebijakan menerjunkan karyawannya berjualan pinggir jalan ngemper | Berita Olahraga | 31,52 |
| tujuan menjaga keberlanjutan kegiatan usaha mendukung menjalankan kebijakan peraturan pemerintah tutupnya | Berita Non Olahraga | 31,43 |

| | | |
|---|------------------------|-------|
| menpora zainudin amali barang barang bekas asian games pekan olahraga nasional pon papua | Berita Non Olahraga | 28,41 |
| timnas basket putra filipina sempat memutuskan mundur asian games sepuluh pemain filipina | Berita Non Olahraga | 28,19 |
| kunyit berkhasiat perkuat daya tahan tubuh menyembuhkan peradangan | Berita Non Olahraga | 26,92 |
| pengelola pizza hut prinsip mendukung kebijakan pemerintah mengatasi pandemi covid | Berita Non Olahraga | 23,17 |

Selanjutnya dipilih K sebanyak K = 1 yaitu nilai kemiripan terbesar dari hasil pengurutan. Sehingga didapatkan kalimat latih yang paling mirip dengan kalimat uji adalah kalimat latih jakarta ketua pp pbsi terpilih agung firman sampurna membangun sport industri bulutangkis dengan nilai kemiripan sebesar 64,37 % yang merupakan kalimat dalam kelas olahraga.

Berdasarkan nilai K-1 tersebut, ternyata nilai kemiripan yang dihasilkan telah sesuai dengan nilai kemiripan minimal yang telah ditetapkan jika nilai similaritas minimalnya sebesar 60%, karena $64,37 \geq 60$. Maka kalimat uji jakarta ketua pbsi agung firman sampurna mengubah program pelatnas bulutangkis masuk ke kelas olahraga.

Akurasi Pengujian

Tabel 3 Akurasi Pengujian

| Similaritas Levenshtein Distance | Jumlah Minimal Kalimat Olahraga | | |
|-------------------------------------|---------------------------------|---------|---------|
| | 0% | 50% | 90% |
| 0% | 100% | 48,3% % | 11,7% % |
| 10% | 100% | 48,3% | 11,7% |
| 20% | 100% | 48,3% | 11,7% |
| 30% | 100% | 48,3% | 13,3% |
| 40% | 100% | 46,7% | 8,3% |

| | | | |
|------|-------|------|----|
| 50% | 73,3% | 5% | 0% |
| 60% | 26,7% | 1,7% | 0% |
| 70% | 10% | 1,7% | 0% |
| 80% | 5% | 0% | 0% |
| 90% | 0% | 0% | 0% |
| 100% | 3,3% | 0% | 0% |

KESIMPULAN

Berdasarkan pengujian berita uji terhadap data latih oleh sistem yang telah penulis lakukan seperti yang telah dipaparkan diatas, maka kesimpulan hasil pengujian yang dapat ditarik adalah sebagai berikut:

1. Jumlah data latih berpengaruh terhadap tingkat similaritas levenshtein distance data uji.
2. Semakin rendah nilai similaritas Levenshtein Distance yang diterapkan ke dalam sistem, maka tingkat kemiripan data uji terhadap data latih semakin jauh dan akurasi hasil pengujian semakin tinggi dari hasil pengujian nilai similaritas Levenshtein Distance sebesar 40%
3. Semakin tinggi nilai similaritas Levenshtein Distance yang diterapkan ke dalam sistem, maka tingkat kemiripan data uji terhadap data latih semakin dekat dan akurasi hasil pengujian semakin rendah dari hasil pengujian nilai similaritas Levenshtein Distance lebih dari 40%.
4. Semakin rendah jumlah minimal kalimat olahraga yang diterapkan dalam sistem pada pengujian berita uji, maka semakin banyak kalimat berita uji yang masuk ke dalam kelas olahraga dan persentase

- akurasi pengujian semakin besar dari hasil pengujian jumlah minimal kalimat olahraga sebesar 50%.
5. Semakin tinggi jumlah minimal kalimat olahraga yang diterapkan dalam sistem pada pengujian berita uji, maka semakin sedikit kalimat berita uji yang masuk ke dalam kelas olahraga dan persentase akurasi pengujian semakin kecil dari hasil pengujian jumlah minimal kalimat olahraga lebih dari 50%.
 6. Semakin dekat jarak similaritas Levenshtein Distance kalimat uji dengan kalimat latih akan membuat klasifikasi semakin jauh.

DAFTAR PUSTAKA

- Asiyah, S., & Fithriasari, K. (2016). Klasifikasi Berita Online Menggunakan Metode Support Vector Machine Dan K-Nearest Neighbor. *Jurnal Sains Dan Seni ITS*, 5(2). <https://doi.org/10.12962/j23373520.v5i2.16643>
- Bagus, I., & Surya, K. (n.d.). *Implementasi Algoritma Levenshtein Pada Sistem Pencarian Judul Skripsi / Tugas Akhir*. 46–53.
- Chandra, D. N., Indrawan, G., & Sukajaya, I. N. (2016). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram. *Jurnal Ilmiah Teknologi Informasi Asia*, 10(1), 11–19.
- Dzikrulloh, N. N., & Setiawan, B. D. (2017). Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri). *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(5), 378–385.
- Efendi, Z., & Mustakim, M. (2017). Text Mining Classification sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi. *Seminar Nasional Teknologi Informasi Komunikasi Dan Industri*, 0(0), 235–242. Retrieved from <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/3273>

- Fahma, A. I., Cholissodin, I., & Perdana, R. S. (2018). Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 53–62.
- Firmansyah, N. (2018). 5 Situs Olahraga Populer di Indonesia. Retrieved November 22, 2019, from www.starnage.com website: <https://starnage.com/5-situs-olahraga-populer-di-indonesia/>
- informatikalogi.com. (2017). Algoritma K-Nearest Neighbor (K-NN). Retrieved November 23, 2019, from www.informatikalogi.com website: <https://informatikalogi.com/algoritma-k-nn-k-nearest-neighbor/>
- Kurniawan, B., Effendi, S., & Sitompul, O. S. (2012). Klasifikasi Konten Berita Dengan Metode Text Mining. *Jurnal Dunia Teknologi Informasi*, 1(1), 14–19. Retrieved from <http://download.portalgaruda.org/article.php?article=58993&val=4123>
- Mahmudy, W. F., & Widodo, A. W. (2014). Klasifikasi Artikel Berita Menggunakan Naive Bayes Classifier yang Dimodifikasi. *Tekno*, 21, 1–10.
- Nurhadi, A. (2015). Klasifikasi Konten Berita Digital Bahasa Indonesia Menggunakan Support Vector Machines (SVM) Berbasis Particle Swarm Optimization (PSO). *Jurnal Bianglala Informatika*, 3(2), 1–9.
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 269. <https://doi.org/10.25126/jtiik.201853810>
- Pratama, B. P., & Pamungkas, S. A. (2016). Analisis Kinerja Algoritma Levenshtein Distance dalam Mendeteksi Kemiripan Dokumen Teks. *Jurnal "LOG!K@,"* 6(2), 131–143.
- Triawati, C. (2009). Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia. *Universitas Telkom*.
- Iqbal, Agami. IMPLEMENTASI APLIKASI *e-BOARDING HOUSE* DI KOTA LHOKESEMAWE MENGGUNAKAN ALGORITMA

LEVENSHTEIN DISTANCE. TGA. Universitas Malikussaleh Lhokseumawe. 2016.

Fauzi, Muhammad. IDENTIFIKASI KEMIRIPAN KARYA ILMIAH DENGAN ALGORITMA K-NEAREST NEIGHBOUR. TGA. Universitas Malikussaleh Lhokseumawe. 2014.