

Implementasi Metode Word2Vec Dan TextRank Dalam Aplikasi Mobile Peringkat Berita Olahraga

Nathaniel Andrew ¹⁾ Viny Christanti Mawardi ²⁾ Novario Jaya Perdana ³⁾

^{1) 2) 3)} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta

¹⁾email: nathaniel..535200054@stu.untar.ac.id, ²⁾email: viny@fti.untar.ac.id, ³⁾email: novariojp@fti.untar.ac.id

ABSTRACT

Dalam era informasi digital saat ini, volume besar dokumen berita olahraga online memunculkan kebutuhan akan alat peringkat otomatis yang efisien. Peningkatan daya tampung informasi ini memunculkan tantangan bagi pembaca untuk mendapatkan pemahaman cepat dan efisien dari berita olahraga terkini. Oleh karena itu, pengembangan aplikasi peringkat otomatis menjadi topik yang menarik untuk menyediakan ringkasan yang informatif dan ringkas dari berita atau artikel yang panjang. Metode yang akan di gunakan menggunakan metode word2vec dan textrank dan hasil kesimpulan yang di dapat menghasilkan ringkasan berita olahraga yang terbilang cukup dapat dibaca dan dimengerti.

Key words

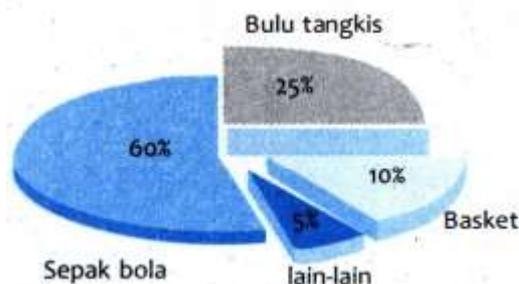
Android, Word2Vec, TextRank, Peringkat, Olahraga

1. Pendahuluan

Dalam era informasi digital saat ini, volume besar dokumen berita olahraga online memunculkan kebutuhan akan alat peringkat otomatis yang efisien. Peningkatan daya tampung informasi ini memunculkan tantangan bagi pembaca untuk mendapatkan pemahaman cepat dan akurat dari berita olahraga terkini. Oleh karena itu, pengembangan aplikasi peringkat otomatis menjadi penting untuk menyediakan ringkasan yang informatif dan ringkas dari berita atau artikel yang panjang [1]. Olahraga adalah suatu aktivitas manusia yang mencakup berbagai bentuk latihan fisik, kompetisi, dan rekreasi yang dilakukan secara sadar dan terstruktur dengan tujuan untuk meningkatkan kondisi fisik, mental, dan sosial individu. Aktivitas olahraga melibatkan gerakan fisik yang berkisar dari yang sederhana hingga kompleks, sering kali melibatkan interaksi dengan orang lain atau lingkungan sekitar, serta mengikuti aturan atau pedoman tertentu yang mengatur jalannya kompetisi atau latihan.

Olahraga bukan hanya tentang gerakan fisik semata, tetapi juga tentang pengembangan keterampilan, dan nilai-nilai positif dalam kehidupan sehari-hari [2]. Olahraga juga merupakan sebuah hiburan atau Entertainment bagi Masyarakat dunia. Di Indonesia sendiri, banyak masyarakat yang gemar dengan cabang olahraga secara global dan selalu mengikuti perkembangan berita cabang olahraga seperti sepak bola, bulutangkis dan basket agar mereka tetap selalu

mengetahui berita perkembangan cabang olahraga yang Masyarakat minati atau senang. Berikut adalah data statistik tahun 2021-2023 dari beberapa cabang olahraga yang memiliki penggemar dari penduduk asia yang khusus nya ada negara Indonesia. Berita olahraga merupakan salah satu cara agar Masyarakat pecinta cabang olahraga tertentu tetap terus mengikuti perkembangan dunia olahraga yang mereka gemari mulai dari cabang olahraga tertentu atau klub maupun individu atlet yang disukai. Berikut diagram data dari penggemar olahraga di negara asia.



Gambar 1. Statistik Penggemar Olahraga Di Negara Asia

Masyarakat Indonesia umumnya masih membaca berita olahraga website berita dan berganti-ganti portal website berita lainnya untuk membaca berita cabang olahraga lainnya. Rancangan yang akan dibuat yaitu aplikasi berita olahraga berbasis android yang berisikan berbagai macam cabang olahraga di dalamnya. Cabang olahraga yang nantinya akan tersedia dalam aplikasi seperti Sepakbola, Bulutangkis, Basket dan lainnya. Rancangan ini juga memiliki tujuan untuk meringkas artikel berita olahraga dari beberapa website ke dalam sebuah aplikasi berbasis android. Secara keseluruhan, aplikasi peringkat otomatis diharapkan dapat membantu meringkas berita olahraga dengan harapan yang sesuai [3]. Harapan dari perancangan aplikasi peringkat ini dapat menyajikan informasi dengan cara yang mudah dipahami dan dan dibaca. Dengan menggunakan teknologi peringkasan otomatis, pembaca nantinya dapat dengan mudah memperoleh inti dari isi sebuah artikel olahraga tanpa harus membaca keseluruhan artikel yang terlalu panjang [4].

Dengan latar belakang masalah tersebut, rancangan aplikasi yang akan dibuat yaitu meringkas berita dari

berbagai cabang olahraga sebesar 50% dan kemudian hasil ringkasan tersebut akan tersedia nanti pada rancangan aplikasi yang akan dibuat. Ada beberapa metode peringkasan yang biasa digunakan seperti PageRank, BERT, LSTM, BERTSUM dan LSA. Untuk merealisasikan pembuatan rancangan aplikasi, akan menggunakan metode TextRank untuk melakukan ekstraksi ringkasan.

TextRank adalah metode dalam pemrosesan bahasa alami yang digunakan untuk ekstraksi informasi dan ringkasan teks. TextRank bekerja dengan cara memandang teks sebagai graf, di mana setiap kalimat atau entitas teks seperti kata atau frasa direpresentasikan sebagai simpul dalam graf. Hubungan antar kalimat atau entitas dinyatakan sebagai edge atau garis sambung yang menghubungkan simpul-simpul tersebut. Kemudian, metode TextRank digunakan untuk menghitung skor kepentingan ranking dari setiap simpul dalam graf berdasarkan struktur hubungannya [5]. Sebelum Metode TextRank berjalan, setiap kata dalam teks berita yang sudah dikumpulkan melalui scraping data akan diberikan bobot menggunakan Word2Vec agar nantinya dapat diolah hasilnya ke dalam proses metode TextRank. Word2Vec adalah sebuah teknik dalam pemrosesan bahasa alami Natural Language Processing yang digunakan untuk menghasilkan representasi vektor kata dari teks. Ide dasarnya adalah untuk mengonversi kata-kata menjadi nilai vektor dalam ruang berdimensi sehingga kata-kata yang memiliki makna serupa memiliki representasi vektor yang mendekati satu sama lain [6]. Cara kerja Word2Vec melibatkan pembelajaran mesin untuk memahami konteks kata berdasarkan hubungannya dengan kata-kata di sekitarnya dalam teks. Word2Vec telah menjadi alat yang sangat berguna dalam banyak aplikasi NLP, termasuk klasterisasi dokumen, penerjemahan mesin, dan analisis sentimen, karena representasi vektor kata yang dihasilkannya memungkinkan model untuk lebih baik memahami dan menangkap makna dari teks [7].

2. Dasar Teori

2.1 Rancangan Sistem

Sistem yang akan dirancang atau dibuat merupakan peringkasan text dari hasil scrapping data teks artikel atau website-website informasi berita olahraga. Pengambilan data teks atau scrapping data akan dilakukan menggunakan bahasa pemrograman Python yang menggunakan code editor third party Google Colab. Data teks yang akan diambil nantinya hanya data yang berhubungan dengan informasi berita olahraga. Kemudian setelah semua data teks sudah terkumpul dari proses scrapping data menggunakan Python, data teks tersebut akan dilakukan tahapan pre-processing untuk merapikan data-data teks yang nantinya akan lebih mudah diolah lagi menggunakan metode Word2Vec dan metode TextRank. berupa teks kosakata.

2.2 Olahraga

Olahraga merupakan aktivitas fisik yang dilakukan dengan tujuan meningkatkan kebugaran fisik, kesehatan, dan keterampilan atletis. Aktivitas ini melibatkan gerakan tubuh yang dapat membantu pembentukan otot, peningkatan daya tahan kardiovaskular, serta pengembangan keterampilan motorik. Olahraga dapat dilakukan secara individu atau dalam bentuk tim, dan melibatkan berbagai jenis disiplin, mulai dari olahraga rekreasi hingga olahraga kompetitif. Kebugaran Fisik Olahraga secara konsisten dapat meningkatkan kebugaran fisik, termasuk kekuatan otot, daya tahan, kelenturan, dan keseimbangan tubuh. Aktivitas fisik juga membantu menjaga berat badan yang sehat dan mencegah penyakit terkait gaya hidup tidak sehat [2]. Kesehatan Mental Olahraga memiliki dampak positif pada kesehatan mental. Aktivitas fisik merangsang pelepasan endorfin, hormon yang dapat meningkatkan suasana hati dan mengurangi stres serta kecemasan. Olahraga juga dapat menjadi saluran untuk melepaskan energi negatif dan meningkatkan kesejahteraan psikologis.

Banyak olahraga dilakukan dalam kelompok atau tim, menciptakan kesempatan untuk berinteraksi sosial. Ini dapat memperkuat hubungan interpersonal, mengembangkan keterampilan komunikasi, dan membangun rasa kebersamaan. Pengembangan Keterampilan Berbagai olahraga melibatkan pengembangan keterampilan teknis, taktis, dan fisik. Misalnya, olahraga seperti sepak bola, bola basket, atau tenis melibatkan latihan keterampilan khusus yang dapat meningkatkan koordinasi dan keahlian atletik. Fair Play dan Etika Olahraga sering kali memiliki aturan dan etika tertentu yang harus diikuti oleh para pesertanya. Prinsip-prinsip fair play, menghargai lawan, dan menghormati aturan merupakan nilai-nilai inti dalam banyak kompetisi olahraga. Kesempatan Karier bagi beberapa individu yang berbakat, olahraga dapat menjadi jalur karier. Profesionalisme dalam olahraga menciptakan peluang bagi atlet untuk bersaing di tingkat nasional dan internasional, dan menjadi atlet profesional. Olahraga memiliki dampak positif yang luas pada kesehatan dan kesejahteraan manusia, tidak hanya fisik tetapi juga mental dan sosial. Olahraga dapat diadaptasi sesuai dengan preferensi individu, termasuk pilihan olahraga rekreasi, olahraga kompetitif, atau aktivitas fisik lainnya yang mendukung gaya hidup sehat [8]. Berita olahraga memberikan informasi tentang peristiwa atau kejadian terkini di dunia olahraga. Hal ini dapat mencakup hasil pertandingan, prestasi atlet, berita transfer, kontroversi, cedera, dan banyak aspek olahraga lainnya. Konteks Kompetisi: Berita olahraga sering kali memberikan konteks kompetitif tentang performa tim atau atlet dalam pertandingan atau kompetisi tertentu. Hal ini dapat mencakup statistik, skor, analisis taktis, dan liputan pertandingan. Selain informasi dasar, berita olahraga sering kali menyertakan komentar dan analisis dari pakar atau pakar olahraga. Mereka dapat memberikan informasi tentang kinerja tim atau atlet, mengevaluasi strategi, dan

memprediksi hasil di masa depan. Berita olahraga sering kali juga memberikan konteks yang lebih luas mengenai dampak acara olahraga tertentu. Misalnya, bagaimana kemenangan atau kekalahan suatu tim dapat memengaruhi posisinya di peringkat, atau bagaimana prestasi seorang atlet dapat memengaruhi karier dan reputasinya. Hiburan dan interaksi penggemar: Berita olahraga tidak hanya memberikan informasi, namun juga berfungsi sebagai sumber daya hiburan dan keterlibatan untuk penggemar olahraga. Berita-berita ini dapat menimbulkan emosi, menimbulkan perdebatan dan mendorong semangat kompetitif di kalangan penggemar tim atau atlet tersebut. Oleh karena itu, berita olahraga tidak hanya memberikan update mengenai dunia olahraga, tetapi juga berperan penting dalam membentuk budaya dan identitas penggemar dan komunitas olahraga yang lebih luas.

2.3 Natural Language Processing

Natural Language Processing merupakan sub bidang pembelajaran kecerdasan buatan atau AI yang memiliki karakteristik untuk membuat sebuah kecerdasan buatan yang dapat memahami bahasa manusia dengan komputer untuk tujuan tertentu. NLP sendiri sebenarnya tidak memiliki metode algoritma perhitungan sendiri melainkan mengambil metode algoritma perhitungan dari sub bidang AI lainnya. NLP sudah banyak sekali diterapkan dan dapat digunakan oleh manusia untuk saat ini seperti contohnya yang semua orang tau dan gunakan yaitu google. Meskipun NLP sudah sudah maju dengan sangat cepat, NLP sendiri masih bisa berkembang lebih jauh lagi untuk membuka peluang-peluang baru dalam berbagai macam bidang dengan kemajuan teknologi saat ini [3]. Data teks merupakan sebuah data berskala besar atau kecil yang berisikan teks-teks string atau karakter yang bisa berupa dokumen, artikel, pesan, dan lainnya. Biasanya data ini digunakan dalam pembelajaran Natural Language Processing yang nantinya dapat diolah menggunakan tahap pre-processing dan metode perhitungan untuk mencapai tujuan tertentu.

2.4 Scrapping

Scrapping data merupakan sebuah Teknik dalam bidang Natural language Processing untuk mengekstraksi informasi data-data dari sebuah web ataupun aplikasi secara otomatis. Data informasi yang bisa di ekstrak seperti data teks, data suara, data gambar, dan lainnya. Scrapping dapat dilakukan secara manual satu per-satu tetapi akan memakan waktu yang sangat lama [9]. Menggunakan kode pemrograman akan sangat membantu untuk melakukan scrapping data seperti menggunakan Python untuk melakukan scrapping data dan terdapat tools yang dapat melakukan scraping secara otomatis menggunakan Web Scraper dan UI Path. Ini akan memakan waktu yang jauh lebih sedikit dibanding dengan cara manual karena komputer yang akan mengambilkan data yang kita inginkan. Biasanya scrapping data dilakukan untuk melakukan penelitian

data, pengumpulan data, analisa data sesuai dengan kebutuhan masing-masing.

2.5 Preprocessing

Pre-processing merupakan Langkah awal penting setelah melakukan scrapping data yang sudah dilakukan. Tujuan dari melakukan pre-processing untuk mengolah data mentah atau data kotor menjadi lebih bersih, mengelompokkan data, dan lainnya untuk menjadi lebih mudah di analisa dan di proses pada suatu algoritma metode perhitungan. Tahapan pre-procesing yang bisa dilakukan:

1. Tokenizing
Tokenizing merupakan tahapan yang akan memecah teks menjadi bagian-bagian yang lebih kecil yang disebut Token. Token dapat berupa data teks, symbol, angka yang memiliki makna tersendiri.
2. Case Folding
Case Folding merupakan tahapan yang dapat mengubah huruf pada data teks menjadi huruf besar atau kecil sesuai dengan kebutuhan.
3. Punctuation Removal
Punctuation Removal adalah tahap penghapusan karakter tanda baca yang tidak memiliki relevan sama sekali pada sebuah rangkaian data teks seperti titik '.', koma ',', tanda seru '!', tanda tanya '?'.

2.6 Word2Vec

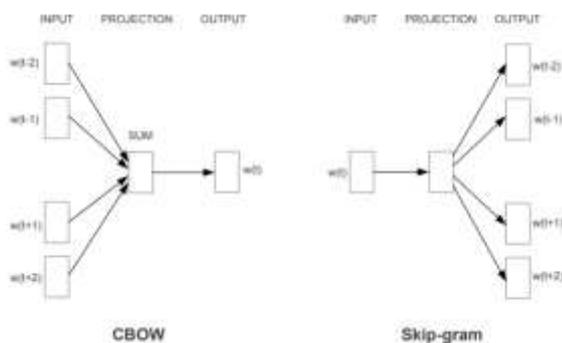
Word2Vec adalah sebuah teknik dalam pemrosesan bahasa alami Natural Language Processing atau NLP yang digunakan untuk menghasilkan representasi vektor kata dari teks. Word2Vec bertujuan untuk menyematkan kata-kata ke dalam ruang vektor, di mana kata-kata yang memiliki makna serupa ditempatkan lebih dekat satu sama lain dalam ruang vektor tersebut. Pendekatan Word2Vec memanfaatkan ide distribusi kata-kata, yang menyatakan bahwa kata-kata yang sering muncul bersama-sama dalam konteks yang sama cenderung memiliki makna yang mirip atau berkaitan. Word2Vec telah menjadi komponen penting dalam banyak aplikasi NLP, seperti pemahaman bahasa alami, penerjemahan mesin, klasifikasi teks, dan analisis sentimen. Representasi vektor kata yang dihasilkan oleh Word2Vec membantu mesin untuk memahami makna kata-kata dalam konteks dan meningkatkan kinerja model dalam tugas-tugas berbasis teks [10]. Berikut cara kerja metode Word2Vec:

1. Pelatihan model
Word2Vec memerlukan pelatihan model untuk mendapatkan hasil yang maksimal ketika melakukan proses pengolahan data untuk mencapai hasil yang maksimal.
2. Membangun kosa kata
Teks yang akan di proses akan dipecah menjadi sebuah token vector berupa nilai atau angka yang

direpresentasikan menjadi sebuah dimensi untuk membuat kosa kata.

3. Membuat konteks
Word2vec akan membuat sebuah konteks kata yang sering muncul dan akan memperbarui nilai vector pada kata berdasarkan kecenderungan kemunculan kata dalam konteks.
4. Menghasilkan Nilai Vector
Kalimat-kalimat atau teks yang sudah di proses sebelumnya akan menampilkan hasil nilai vector atau bobot representasi kata untuk diolah lebih lanjut menggunakan metode lain.

Word2Vec memiliki 2 arsitektur utama yaitu Skip-Gram dan Continuous Bags Of Word (CBOW). Keduanya memiliki tujuan yang sama tetapi dengan proses yang berbeda. Proses arsitektur Skip-gram adalah memprediksi kata-kata yang mungkin muncul di sekitarnya. Untuk proses arsitektur Continuous Bags Of Words (CBOW) adalah memprediksi kata target dalam sebuah dataset. Berikut gambar di bawah ini merupakan arsitektur pada word2vec.



Gambar 2. Arsitektur Pada Word2Vec

Rancangan yang akan dibuat akan menggunakan arsitektur Skip-gram dikarenakan lebih optimal dengan dataset yang berjumlah besar. Berikut rumus dari Word2Vec arsitektur Skip-gram :

$$P(c | t) = \frac{\sum \exp(v_c \cdot v_t)}{\sum \exp(v_w \cdot v_t)}$$

Penjelasan:

1. Σ : Simbol sigma menandakan penjumlahan.
2. exp: Fungsi eksponensial.
3. v_c : Vektor representasi kata konteks (c) dengan dimensi n.
4. v_t : Vektor representasi kata target (t) dengan dimensi n.

2.7 TextRank

TextRank adalah metode ekstraksi ringkasan yang digunakan dalam Natural Language Processing untuk mengidentifikasi dan mengekstraksi kalimat-kalimat penting dari sebuah teks tanpa mengubah kalimat awal dari teks yang akan diringkas. Berbeda dengan abstraksi, abstraksi sendiri memiliki tujuan untuk mengambil ini dari

kalimat awal dan mengubah kata-katanya tetapi tidak mengubah makna nya. Tujuan utama dari TextRank adalah menyederhanakan teks dengan mempertahankan informasi yang paling relevan dan signifikan. Metode TextRank sering digunakan dalam tugas-tugas ekstraksi ringkasan otomatis, identifikasi kata kunci, dan pemahaman dokumen dengan memanfaatkan struktur hubungan antara kata-kata untuk mengekstrak informasi penting dari teks [5]. Metode ini juga memiliki langkah-langkah untuk mencapai hasil akhirnya, berikut adalah langkahnya :

1. Pembentukan Graf Kata
teks awal akan dipecah menjadi kata-kata dan kalimat. Kemudian, dibentuk menjadi graf yang diwakili simpul atau node, korelasi antara kata-kata ini ditentukan oleh seberapa sering kata-kata tersebut muncul bersamaan dalam dokumen teks.
2. Memberi Bobot Graf
Setiap sisi antara dua simpul dalam graf akan diberi bobot berdasarkan seberapa kuat hubungan antara kata-kata tersebut dan bobot dapat dihitung.
3. Menghitung Nilai Penting
Setiap simpul akan diberi skor berdasarkan pengaruhnya dalam graf. Skor ini dapat dihitung Node-node yang lebih penting atau memiliki lebih banyak keterkaitan akan memiliki skor yang lebih tinggi.
4. Melakukan seleksi Kalimat
Kalimat-kalimat yang mewakili simpul dengan skor tertinggi akan dipilih sebagai bagian dari ringkasan. Kalimat-kalimat akan diurutkan berdasarkan skor dan diambil dalam urutan tertentu untuk membentuk ringkasan teks.

Dengan mengimplementasikan metode TextRank, akan membuat pengguna mempermudah dan mempercepat dalam memahami makna dari sebuah kalimat teks [19]. Berikut merupakan rumus dari TextRank :

$$PR(u) = (1 - d) * \sum (PR(v) / |N(v)|) + d * (1 / |V|)$$

Penjelasan:

1. PR(u) adalah nilai PageRank untuk kalimat u.
 2. N(u) adalah himpunan kalimat yang terhubung langsung ke kalimat u.
 3. PR(v) adalah nilai PageRank untuk kalimat v yang terhubung ke kalimat u.
 4. |N(v)| adalah jumlah kalimat yang terhubung langsung ke kalimat v.
 5. |V| adalah total jumlah kalimat dalam dokumen.
- D adalah faktor damping (dalam contoh kali ini menggunakan 0,85 sebagai contoh).

3. Hasil Percobaan

Setelah rancangan sistem sudah dibuat, akan dilakukan metode pengujian tahap – tahap setiap langkah pada program. Untuk menguji kinerja rancangan sistem

dan metode, akan di ambil data proses dari awal artikel yang belum di ringkas sampai ke tahap artikel berhasil di ringkas menggunakan metode. Untuk melakukan pengujian pada hasil ringkasan dan tampilan antar muka, akan dilakukan pengisian kuisioner untuk mendapatkan responden orang lain tentang rancangan sistem yang dibuat. Pembuatan kuisioner juga memiliki tujuan untuk melakukan evaluasi keberhasilan rancangan sistem apakah rancangan yang dibuat sudah baik dalam pembuatannya atau tidak. Untuk evaluasi dan hasil pengujian akan difokuskan pada hasil ringkasan apakah dapat dimengerti dan tampilan antarmuka apakah sudah terbilang cukup baik. Hasil dari pengujian dan evaluasi akan di jelaskan lebih lanjut pada bab dan sub – bab berikut ini.

Tahap hasil pengujian yang akan di ambil datanya difokuskan pada waktu lamanya proses eksekusi program pada setiap Langkah, dan hasil responen kuisioner yang sudah dibuat dan diberikan kepada orang lain untuk di isi. Untuk dataset berhasil terkumpul sebanyak 1524 data dan jumlah responden yang sudah mengisi kuisioner mendapat 33 responden dari 10 pertanyaan. Berikut sub – bab untuk membahas hasilnya.

Proses eksekusi untuk melakukan scrapping pada sumber berita memiliki waktu yang bervariasi tergantung dari sumber berita yang di scrapping. Sumber Kompas memakan waktu selama 20-27 detik sebanyak 15-20 data yang berhasil di scrapping. Sumber berita CNN memakan waktu selama 12–15 detik sebanyak 10–11 data yang berhasil di scrapping. Sumber berita iNews memakan waktu selama 10–15 detik sebanyak 10–11 data yang berhasil di scrapping. Sumber berita Detik memakan waktu selama 12–15 detik sebanyak 10–11 data yang berhasil di scrapping. Sumber berita Sindonews memakan waktu 10–15 detik sebanyak 10–11 data yang berhasil di scrapping.

Proses eksekusi data pada bagian preprocessing tergantung pada banyaknya dataset yang akan di bersihkan datanya. Preprocessing terdiri dari casefolding, punctuation, dan tokenize. Untuk pengujian kali ini skema yang digunakan adalah dengan 1 buah artikel data dan seluruh dataset yang terdiri dari 1524 data yang akan di uji proses eksekusinya secara keseluruhan. Untuk hanya 1 buah artikel data yang terdiri dari 7–10 kalimat memakan waktu sekitar 4 detik. Untuk seluruh dataset memakan waktu sekita 1 menit 38 detik.

Proses klasifikasi bertujuan untuk memberikan tanda pada setiap baris data berita artikel bahwa data tersebut merupakan jenis kategori olahraga apa dan berasal dari sumber yang mana dan klasifikasi ini menggunakan kode program. Untuk skema pengujian sama seperti pada pre – processing yaitu dengan 1 buah data dan seluruh dataset. Untuk 1 buah data memakan waktu sekitar 5 detik. Untuk seluruh dataset memakan waktu sekitar 3 menit 23 detik.

Pengujian proses eksekusi Metode Word2Vec dilakukan setelah proses pre–processing dan klasifikasi berita. Untuk 1 buah data yang di eksekusi memakan waktu sekitar 6 detik. Untuk keseluruhan dataset yang di

eksekusi dalam metode Word2Vec memakan waktu sekitar 4 menit 10 detik.

Tahap terakhir untuk menghasilkan ringkasan artikel dari data mentah sampai data artikel yang diringkas adalah dengan metode TextRank. Proses metode TextRank diberikan parameter untuk mengambil 60 kata dari 5 kalimat yang ada di dataset artikel dan menghasilkan ringkasan. Hasil pengujian metode TextRank untuk 1 buah data memakan waktu sekitar 12 detik. Untuk keseluruhan dataset dengan jumlah 1524 yang sudah Melawati tahap pre-processing, klasifikasi, dan metode Word2Vec memakan waktu sekitar 7 menit 31 detik.

4. Kesimpulan

Kesimpulan dari uraian di atas adalah bahwa pengujian dan evaluasi kinerja sistem peringkasan artikel menunjukkan hasil yang hanya bisa dibilang memadai. Waktu eksekusi setiap langkah dalam proses, mulai dari scrapping, preprocessing, klasifikasi, hingga metode Word2Vec dan TextRank, berada dalam batas yang dapat diterima. Evaluasi melalui kuisioner yang melibatkan 33 responden menunjukkan bahwa sistem ini memiliki potensi yang baik dalam hal pemahaman hasil ringkasan dan kualitas tampilan antarmuka. Meskipun demikian, feedback dari kuisioner akan sangat berguna untuk lebih meningkatkan sistem di masa depan. Secara keseluruhan, sistem ini berhasil menjalankan tugasnya dengan baik, namun masih ada ruang untuk perbaikan berdasarkan evaluasi pengguna.

REFERENSI

- [1] Idhafi, Zaky, Surya Agustian, and Febi Yanto. "Peringkasan teks otomatis pada artikel berbahasa Indonesia menggunakan metode maximum marginal relevance." *Jurnal CoSciTech (Computer Science and Information Technology)* 4, no. 3 (2023): 609-618.
- [2] Candra, Oki, Tri Prasetyo, and Ahmad Rahmadani. "Pembentukan Karakter Melalui Olahraga." (2023).
- [3] Yuliska, Yuliska, and Khairul Umam Syaliman. "Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia." *IT Journal Research and Development* 5, no. 1 (2020): 19-31.
- [4] Abidin, Aa Zezen Zaenal, and Enung Nurjanah. "Sistem Peringkasan Teks Otomatis Multi Dokumen Kliping Artikel Berita Gempa Menggunakan Metode Tf-idf." *Jurnal Teknologi Informasi dan Komunikasi* 15, no. 1 (2020): 52-69.
- [5] Khaqiqi, Ryan Fajar. "Rancang Bangun Aplikasi Baca Berita Online dengan Peringkasan Teks Otomatis Menggunakan Algoritma Textrank." PhD diss., Universitas Islam Sultan Agung, 2021.
- [6] Susanto, Edy, Viny Christanti Mawardi, and Manatap Dolok Lauro. "Aplikasi Clustering Berita Dengan Metode K Means Dan Peringkasan Berita Dengan Metode Maximum Marginal Relevance." *Jurnal Ilmu Komputer dan Sistem Informasi* 9, no. 1 (2021): 62-68.

- [7] Firdaus, Ilmi Rasyidah. “Peringkasan artikel online berbasis Algoritma TextRank dan Similarity.” PhD diss., Universitas Islam Negeri Maulana Malik Ibrahim, 2023.
- [8] Djoko, Sri Wahyuni, Chita Widia, I. Putu Agus Dharma Hita, Billy Aditya Pratama, Pauzan Pauzan, Farah Fauziyah Radhiyatulqalbi Ahmad, Marina Indriasari et al. “Anatomi & Fisiologi Olahraga.” (2023).
- [9] Mulyani, Asri, Dede Kurniadi, and Iqbal Lukmanul Hakim. “Web Scraping pada Web Media Digital untuk Membangun Aplikasi Android.” *Jurnal Algoritma* 18.1 (2021): 313-322.
- [10] Naufal, Hildan Fawwaz, and Erwin Budi Setiawan. “Ekspansi Fitur Pada Analisis Sentimen Twitter Dengan Pendekatan Metode Word2Vec.” *eProceedings of Engineering* 8, no. 5 (2021).

Nathaniel Andrew, mahasiswa pada program studi Fakultas Teknologi Informasi di Universitas Tarumanagara.

Viny Christanti Mawardi S.Kom., M.Kom. memperoleh gelar S.Kom. dari Universitas Tarumanagara tahun 2004. Kemudian memperoleh gelar M.Kom. dari Universitas Indonesia tahun 2008. Saat ini aktif sebagai Dosen Tetap Fakultas Teknologi Informasi Universitas Tarumanagara.

Novario Jaya Perdana S.Kom., M.T. memperoleh gelar S.Kom. dari Institut Teknologi Sepuluh Nopember tahun 2011. Kemudian memperoleh gelar M.T. dari Universitas Indonesia pada tahun 2016. Saat ini aktif sebagai Dosen Tetap Fakultas Teknologi Informasi Universitas Tarumanagara.