



Implementasi Tunelling pada Perancangan Sistem Peringkasan dan Klasifikasi Berita Otomatis menggunakan *Textrank* dan *KNN*

Tunelling Implementation on Design of Sports News Summarization and Classification System using Textrank and KNN

Falahah^{1*}, Ari Fajar Santoso², Abdullah Fajar³^{1,2,3}Program Studi Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom¹falahah@telkomuniversity.ac.id**Abstract**

News summarization is very important in the news analysis process. However, in the summarization process, there are often obstacles such as the large number of news stories and the need for news classification. This research aims to build a simple web-based system that can be used to summarize and classify news which will be very useful in the news analysis process. The proposed summarization method is *Textrank*, and the news classification method that will be used is *KNN*. This system is expected to provide an automatic summarization function to make it easier to analyze news content. The data that will be used as the basis for classification modeling is sports news in 3 months, and the classification that will be used to determine whether the news includes sports news in three branches, namely football, rackets or basketball. Testing of the summarization model using *textrank* was carried out by applying *ROUGE-1* and *ROUGE-2*, with results of 0.79 and 0.67. Meanwhile, testing the classification model using *KNN* with $k=3$ and $k=5$ is 0.9866 and 0.9666 so $k=3$ will be used. This system will be built using the web scrapping library, *textrank*, *stopword* from *PySastrawi*, *scikit-learn* for the classification module using the *KNN* algorithm, and *ngrok* for publishing web-based applications. By using *ngrok*, we can expose the application through internet with a temporary public url without hosting required

Keywords: news, summarization, classification, *textrank*, *KNN*, *ngrok*.

Abstrak

Peringkasan berita merupakan hal yang sangat penting pada proses analisis berita. Namun dalam proses peringkasan tersebut, seringkali terdapat kendala seperti banyaknya jumlah berita dan perlunya klasifikasi berita. Penelitian ini bertujuan untuk merancang sistem sederhana berbasis web yang dapat digunakan untuk melakukan peringkasan dan klasifikasi berita yang akan sangat berguna dalam proses analisis berita. Metode peringkasan yang diusulkan adalah *Textrank*, dan metode pengklasifikasian berita yang akan digunakan adalah *KNN*. Sistem ini diharapkan dapat menyediakan fungsi peringkasan otomatis sehingga memudahkan dalam menganalisis isi berita. Data yang akan digunakan sebagai dasar pemodelan klasifikasi adalah berita olahraga selama tiga bulan, dan klasifikasi dilakukan untuk menentukan apakah berita tersebut termasuk berita olahraga pada tiga cabang yaitu sepakbola, raket atau basket. Pengujian pada model peringkasan menggunakan *textrank*, dilakukan dengan menerapkan *ROUGE-1* dan *ROUGE-2*, dengan hasil 0.79 dan 0.67. Sedangkan pengujian pada model klasifikasi menggunakan *KNN* dengan $k=3$ dan $k=5$ adalah 0.9866 dan 0.9666 sehingga akan digunakan $k=3$. Sistem ini akan dibangun memanfaatkan pustaka *web scrapping*, *textrank*, *stopword* dari *PySastrawi*, *scikit-learn* untuk modul klasifikasi menggunakan algoritma *KNN*, dan *ngrok* untuk publikasi aplikasi berbasis web. Pemanfaatan *ngrok* memungkinkan kita untuk mempublikasikan aplikasi di internet dengan menggunakan url publik sementara tanpa harus melakukan hosting.

Kata kunci: berita, peringkasan, klasifikasi, *textrank*, *KNN*, *ngrok*.

1. Pendahuluan

Analisis berita merupakan pekerjaan penting pada kegiatan jurnalistik. Informasi yang dipublikasikan di berbagai media, termasuk media massa mainstream, tak cukup hanya dibaca. Praktisi komunikasi dituntut

mampu untuk melakukan analisis berita, sehingga hasilnya dapat menjadi strategi komunikasi pengemasan informasi yang ingin disampaikan kepada masyarakat

Analisis isi dipakai untuk menganalisis isi media baik cetak maupun elektronik. Analisis dilakukan secara ilmiah untuk mempelajari dan menarik kesimpulan atas suatu fenomena, dengan memanfaatkan dokumen pada isi media cetak maupun elektronik. Lewat analisis isi, peneliti dapat mempelajari gambaran isi, karakteristik pesan, dan perkembangan dari suatu isi. Selain itu, dapat dilihat juga mana yang diuntungkan atau dirugikan, pihak mana yang menindas maupun tertindas.

Bersamaan dengan massifnya jumlah berita saat ini, maka pekerjaan analisis berita tidak dapat berlangsung dengan cepat, tanpa terlebih dahulu mempersiapkan konten berita yang akan dianalisis, serta kategori yang harus dipenuhi oleh konten berita tersebut. Guna membantu proses analisis ini, maka berita-berita tersebut harus dipilih dengan cermat, disarikan isinya, kemudian diklasifikasikan. Dengan cara ini, penganalisis berita dapat lebih cepat memilih berita yang tepat sesuai kebutuhan.

Kendala pada proses pemilihan berita ini adalah bagaimana cara mengumpulkan berita dari sumber media daring dengan cepat, mensarikan isinya dan mengklasifikasikannya. Hal ini terkait dengan bidang ilmu analisis text dan pemrosesan bahasa natural (*natural language processing*), yang sangat terkait erat dengan pustaka bahasa yang akan digunakan. Terkait dengan berita dalam bahasa Indonesia, masih banyak ditemui kendala dalam pemilihan pustaka bahasa yang dapat dengan mudah digunakan pada perangkat lunak ekstraksi dan klasifikasi data. Penelitian ini akan mengusulkan langkah-langkah untuk menyelesaikan masalah di atas dengan menerapkan teknik web scrapping, text summarization dan text classification.

Teknik peringkasan berita menerapkan konsep peringkasan teks yang dapat dibagi menjadi dua kategori yaitu: ringkasan ekstratif dan ringkasan abstraktif. Ringkasan ekstratif adalah teknik yang akan memilih himpunan bagian kalimat dari teks asal sebagai pembentuk hasil rangkuman. Sedangkan ringkasan abstratif melibatkan reorganisasi bahasa pada teks asal, dan jika diperlukan, penambahan kata-kata baru atau frase pada hasil peringkasan. Oleh karena itu, proses peringkasan ekstratif relatif lebih sederhana dibandingkan dengan peringkasan abstraktif. Hasil peringkasan ekstratif juga dapat dievaluasi menggunakan metrik ROUGE [1].

Penelitian terkait peringkasan teks sudah banyak dilakukan, yang menerapkan berbagai algoritma antara lain: Bert [2], Textrank [3], Fuzzy Logic [4], LSA [5], TF-IDF [6], dan lain-lain. Evaluasi hasil peringkasan dapat menggunakan metrik ROUGE atau bleu [7]. Dari sekian banyak algoritma tersebut, Textrank merupakan algoritma yang lebih sering digunakan karena sifat kompleksitas komputasionalnya relatif rendah dibandingkan yang lain, sehingga proses eksekusinya

lebih cepat dibandingkan BERT. Beberapa riset sebelumnya menunjukkan bahwa algoritma Textrank memberikan hasil kinerja yang lebih baik dibandingkan dengan algoritma lainnya [8] - [12].

Pada prinsipnya, TextRank adalah algoritma supervised yang dapat digunakan untuk meringkas teks bahasa natural, yang dapat mengambil bagian terpenting dari sumber teks dan menggabung-gabungkannya hingga membentuk rangkuman. Textrank menggunakan metode berbasis graf yang mengasumsikan kata dan kalimat dianggap sebagai verteks graf [13]. Algoritma pemeringkatan berbasis graf sendiri merupakan sebuah metode untuk mengetahui node (titik) yang relevan pada sebuah graf dengan cara mendapatkan informasi global dari seluruh jaringan node, secara rekursif.

Misal, $G = (V,E)$ sebagai graf dengan V adalah sekumpulan simpul dan E adalah sekumpulan sisi, dimana E merupakan sebagian dari $V \times V$. Sebuah sisi memiliki bobot yang dapat ditentukan nilainya dengan rumus kesamaan (similar formula). Pada simpul V_i , terdapat $In(V_i)$, simpul-simpul yang menunjuk ke simpul tersebut (predecessor) dan $Out(V_i)$, sekumpulan simpul-simpul yang ditunjuk oleh V_i (successor). Skor total pada simpul V_i didefinisikan seperti pada Rumus 1 [14]:

$$S(V_i) = (1 - d) + d * \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j) \quad (1)$$

dimana d adalah damping factor yang dapat bernilai diantara 0 - 1. Nilai d berperan dalam mengintegrasikan probabilitas perpindahan dari satu simpul ke satu simpul acak pada graf. Pada konteks web surfing, algoritma ini menerapkan "random surfer model", dimana pengguna secara acak menekan link dengan kemungkinan d dan berpindah ke halaman yang sama sekali baru dengan kemungkinan $1-d$. Faktor d biasanya bernilai 0,85. Setiap iterasi diperoleh nilai simpul baru. Error diketahui dengan mengurangi hasil iterasi sekarang dengan hasil iterasi sebelumnya dengan Rumus 2:

$$S^{k+1}(V_i) - S^k(V_i) \quad (2)$$

Setelah menjalankan algoritma, skor dihubungkan pada tiap simpul, yang mewakili "pentingnya" simpul dalam suatu graf. Nilai akhir yang diperoleh setelah TextRank selesai tidak dipengaruhi oleh pilihan nilai awal melainkan jumlah iterasi menuju ambang bataslah yang berpengaruh.

Metode yang diterapkan merupakan graf berbobot maka perlu diketahui bobot tiap sisi yang menghubungkan simpul-simpul. Maka rumus TextRank seperti terlihat pada Rumus 3:

$$WS(V_i) = (1 - d) + d * \sum_{j \in In(V_i)} \frac{w_{ji}}{V_k \in Out(V_j)} WS(V_j) \quad (3)$$

$WS(V_i)$ adalah Bobot simpul, V_i & d adalah damping factor (0.85), V_k adalah Himpunan simpul tetangga, V_j

WS(Vi) adalah Bobot simpul, Vj Vj adalah Himpunan simpul tetangga, Vi wji & wjk adalah bobot sisi antara simpul j&i dan j&k secara berurutan

Penelitian terkait peringkasan berita sudah banyak dilakukan dengan menggunakan beberapa algoritma misalnya *Lexical chain* dan *word sense disambiguation* [15], TF-IDF [6], *Maximum Marginal Relevance* [16], dan *Cross Latent Semantic Analysis* [17]. Sedangkan penelitian terkait klasifikasi berita secara otomatis juga sudah dilakukan dengan menerapkan berbagai algoritma seperti *Naïve Bayes*[18], ANN[19], SVM [20] dan KNN[21]. *K-Nearest Neighbors* (KNN) merupakan salah satu algoritma yang relatif lebih sering digunakan [22]. Beberapa penelitian sebelumnya menunjukkan bahwa akurasi KNN dianggap lebih baik dibandingkan teknik klasifikasi lainnya, seperti SVM, regresi logistik dan *random forest* (RF) [23] [24]. Pada umumnya penelitian terkait peringkasan berita belum dikombinasikan dengan teknik klasifikasi berita, serta hasil akhir penelitian berupa penerapan algoritma peringkasan atau klasifikasi pada studi kasus berita beserta hasil pengujian tingkat akurasi, namun belum ditemui hasil akhir berupa implementasi kedua model tersebut dalam bentuk aplikasi. Hal ini karena perangkat yang digunakan untuk menerapkan algoritma tersebut pada umumnya menggunakan pustaka dari Python, baik yang berjalan pada *Google colab* maupun yang terpasang di lokal.

Di lain pihak, peluang untuk mempublikasikan model menjadi satu aplikasi, meskipun model tersebut diimplementasikan pada *Google colab*, dimungkinkan dengan adanya fasilitas tunneling dari ngrok. Pada prinsipnya, ngrok menyediakan fitur yang memungkinkan kita untuk mempublikasikan server lokal melalui tunnels (terowongan) sehingga dapat diakses melalui URL publik [25]. Dengan adanya fitur ini, kita dapat mempublikasikan aplikasi tanpa harus memiliki akun di webserver publik. Ngrok juga dapat diintegrasikan dengan model machine learning yang sudah disiapkan di *google colab* [26].

Penelitian ini bertujuan untuk menyelesaikan masalah terkait kebutuhan awal untuk membantu analisis berita, yaitu bagaimana menyediakan isi berita secara singkat dan cepat, dengan mengusulkan rancangan sistem peringkasan dan klasifikasi berita secara otomatis. Sistem ini akan dibangun dengan memanfaatkan teknik *web scrapping*, algoritma peringkasan teks dan klasifikasi teks, serta pustaka-pustaka pendukung pada python yang dapat berjalan pada *google colab*, termasuk fitur publikasi berupa aplikasi web yang berjalan pada url publik menggunakan fitur *tunneling* dari ngrok. Pertanyaan yang akan dijawab pada penelitian ini adalah: Bagaimana melakukan web scrapping pada situs berita online?; Bagaimana melakukan peringkasan berita?; Bagaimana membangun model klasifikasi berita?; Bagaimana

merancang sistem yang memuat model peringkasan dan klasifikasi, yang berjalan pada lingkungan cloud.

Perbedaan penelitian ini dari penelitian terkait sebelumnya adalah penggabungan antara proses peringkasan dan klasifikasi. Sedangkan kontribusi yang dihasilkan dari penelitian ini adalah tersedianya rancangan sistem peringkasan berita otomatis yang dapat berjalan pada lingkungan cloud, sehingga dapat dengan mudah dibagi-pakaikan ke pihak yang memerlukan.

2. Metode Penelitian

Pada penelitian ini, digunakan metode penelitian *Canonical Action Research* (CAR) yang merupakan salah satu bentuk metode *Action Research*. Metode *Action Research* (AR) sendiri banyak digunakan pada bidang pendidikan dan mengacu pada usaha untuk meningkatkan kualitas organisasi dan kinerjanya. Tetapi, metode AR juga sudah banyak digunakan dalam pembangunan software [27] [28] dan diterapkan pada beberapa pembangunan aplikasi di kalangan industri [29].

Metode CAR, yang merupakan salah satu varian dari metode AR, dilakukan melalui lima tahapan seperti yang terlihat pada gambar 1. Perbedaan mendasar atas metode CAR terhadap metode AR adalah proses penelitian berlangsung secara iteratif, kolaboratif dan terstruktur [30].

Tahapan-tahapan pada metode CAR adalah:

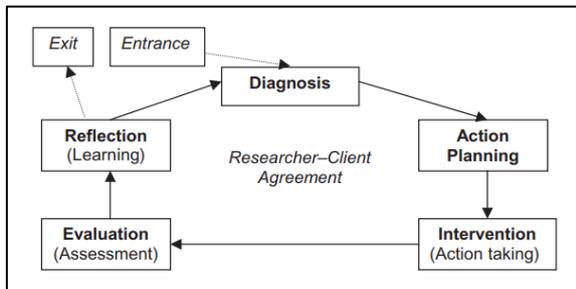
Diagnosis, meliputi eksplorasi terhadap ruang lingkup penelitian, identifikasi permasalahan, kebutuhan dan keinginan stakeholder secara keseluruhan. Pada fase ini perlu ditetapkan tema penelitian yang mencerminkan permasalahan dan ruang lingkup pengetahuan yang akan dicakup.

Planning, merupakan fase perencanaan terhadap tindakan yang akan dilakukan, berdasarkan kondisi dan ruang lingkup yang sudah diketahui pada fase sebelumnya. Pada fase ini ditentukan batasan masalah, asumsi-asumsi, definisi-definisi serta kendala yang harus dipertimbangkan. Hasilnya berupa hipotesa atau usulan solusi yang harus diikuti dengan landasan teoritis ilmiah yang relevan.

Intervention, merupakan tahapan implementasi rencana.

Evaluation, merupakan tahapan untuk menganalisis efek atas tindakan yang dilakukan, yang disesuaikan dengan latar belakang teori yang diacu, ketika menyusun atau merencanakan tindakan.

Reflection, merupakan tahap diseminasi pengetahuan atau pelajaran yang diperoleh dari hasil implementasi dan evaluasi, kepada berbagai pihak di organisasi.



Gambar 1. Tahapan Canonical Action Research [30]

Pada penelitian ini metoda CAR diterapkan hingga langkah 3 yaitu *intervention*. Hal ini karena cakupan tahapan perancangan tidak meliputi evaluasi sistem yang sudah berjalan ataupun penyebarluasan hasil kepada berbagai pihak terkait.

3. Hasil dan Pembahasan

Pada fase diagnosis, ditetapkan identifikasi permasalahan dan ruang lingkup penelitian. Ruang lingkup permasalahan yang akan dicakup meliputi penyediaan solusi untuk mengatasi masalah ekstraksi konten berita, peringkasan dan klasifikasi isi berita, dengan menerapkan teknik dan algoritma sebagai berikut: Ekstraksi konten berita akan menggunakan teknik web scrapping; Peringkasan isi berita akan menggunakan metode Textrank; Klasifikasi isi berita akan menggunakan metode KNN.

Dikarenakan penelitian ini berfokus pada kasus peringkasan berita olahraga dari media berbahasa Indonesia, maka beberapa batasan yang diterapkan antara lain: Ruang lingkup berita dibatasi pada berita olahraga, yang mencakup tiga kategori yaitu sepakbola, raket dan basket; Dataset berita diambil dari media sport.detik.com; Durasi pengambilan berita antara Juli 2023 hingga November 2023; Proses yang akan dicakup pada sistem meliputi ekstraksi data, peringkasan dan klasifikasi data teks.

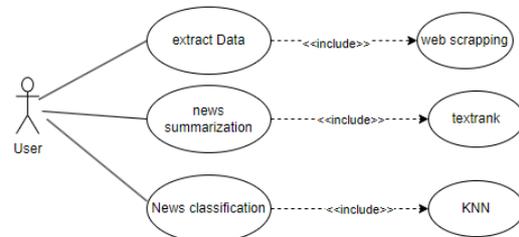
Adapun asumsi-asumsi yang digunakan untuk menyusun solusi atas permasalahan ini adalah: Sistem harus berjalan pada lingkungan komputasi awan dan dapat digunakan oleh siapa saja tanpa perlu proses instalasi yang rumit; Pengguna harus memahami struktur dokumen situs berita untuk memudahkan proses web-scrapping; Pengelompokan berita mengikuti pengkategorian berita yang digunakan pada situs sumber berita.

Berdasarkan asumsi dan batasan masalah di atas, maka sistem diusulkan untuk berjalan pada arsitektur komputasi awan dengan memanfaatkan fitur dari *google colab*, dipadukan dengan *ngrok* sebagai sarana publikasi aplikasi berbasis web.

Tahapan *intervention* adalah realisasi dari rencana implementasi, yaitu perancangan aplikasi dan

pembangunannya pada platform teknologi yang dipilih. Tahapan perancangan aplikasi meliputi: Identifikasi kebutuhan pengguna; Spesifikasi fungsional; mplementasi spesifikasi fungsional; Pengujian fungsional; dan perancangan antarmuka.

Secara garis besar, fungsional sistem dapat dilihat pada usecase diagram seperti pada gambar 2.



Gambar 2. Fungsional Sistem Peringkasan dan Klasifikasi Berita

Berdasarkan identifikasi kebutuhan pengguna, maka sistem yang akan dibangun meliputi spesifikasi serta rancangan aplikasi seperti diuraikan pada bagian berikut:

Extract Data, dilakukan dengan menerapkan teknik *web scrapping*. *Web scrapping* dilakukan dengan cara mempelajari struktur halaman web situs berita target dan memodifikasi blok program untuk membaca konten.

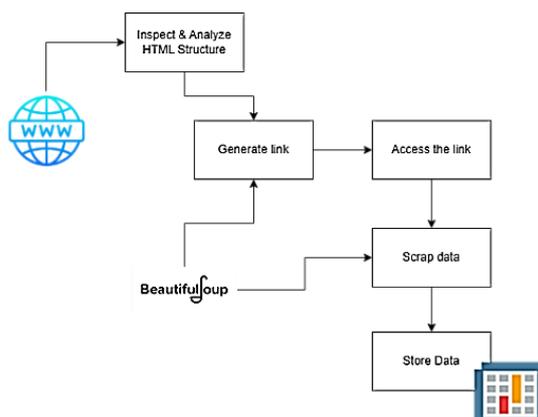
Peringkasan teks, merupakan fitur untuk membaca data konten yang dihasilkan dari proses *web scrapping* dan membuat peringkasan dengan menerapkan metode *textrank*.

Klasifikasi teks, merupakan fitur untuk mengelompokkan konten yang sudah dirangkum menjadi tiga kategori yaitu raket, sepakbola dan basket. Metode yang digunakan adalah KNN.

Desain sistem di atas kemudian diimplementasikan secara teknis dengan memanfaatkan beberapa pustaka pada *Python* dan mengikuti prosedur sebagai berikut:

Fungsi Web Scrapping, yang dibangun berupa fungsi yang menggunakan beberapa pustaka sebagai berikut: *Request-html*, merupakan pustaka *Python* berupa modul *web-scrapping* yang dapat mengeksekusi *HTTP request* dan *rendering Javascript*; dan *Bs4 (Beautiful soup)*, merupakan pustaka *Python* untuk mendapatkan struktur data, dan berinteraksi melalui HTML.

Secara garis besar, proses web scrapping dapat dilihat seperti pada Gambar 3. Hasilnya adalah diperoleh text yang memuat isi berita, yang disimpan dalam tabel dataframe dengan struktur: link, title, judul, kategori, content.



Gambar 3. Proses Web Scrapping

Fungsi Peringkasan Text yang dibangun menggunakan putaka Python: (a) nltk dan nltk.tokenized, (b) Glove, untuk *glove vector of words*, utk proses word embedding, dan (c) *Pysastrawi*, untuk *stopwords* berbahasa Indonesia. Proses yang dilakukan sebagai berikut: *tokenized; word embedding; text preprocessing (remove stopwords); extract word vector; similarity matrix (menggunakan cosine similarity)*;Menerapkan *pagerank*.

Fungsi Klasifikasi berita yang dilakukan setelah diperoleh konten berita yang sudah diringkas. Model klasifikasi yang dipilih adalah KNN, dan tahapan-tahapannya adalah: Memberikan label pada dataset berita hasil peringkasan, menjadi tiga kategori yaitu sepakbola, raket, dan basket; Membagi dua dataset menjadi data latih dan data uji; *Preprocessing text* seperti melakukan *tokenizing*, membuang *stopwords* dan membersihkan tanda baca; dan menerapkan model KNN menggunakan pustaka *scikit-learn*.

Pengujian teknis terhadap model yang diusulkan sudah dilakukan pada penelitian sebelumnya [31], sehingga pada artikel ini akan ditampilkan beberapa contoh hasil pengujian yaitu pengujian model peringkasan berita dan pengujian model klasifikasi berita.

Pengujian fungsional terhadap hasil peringkasan berita menunjukkan bahwa prosedur peringkasan berita menggunakan metode *textrank* sudah bekerja dengan baik. Tabel 1 menampilkan contoh hasil metode *textrank* yang dapat meringkas berita menjadi hanya memuat kata-kata penting dari isi berita tersebut.

Pengujian terhadap model klasifikasi berita yang sudah dilakukan pada penelitian sebelumnya [31], menunjukkan bahwa akurasi untuk model klasifikasi dengan metode kNN untuk k =3, adalah 0.98666, dan untuk k=5 tingkat akurasinya adalah 0.9666, sehingga model yang akan diterapkan adalah dengan k=3.

Selain itu, pengujian juga dilakukan terhadap model prediksi, yang berupa penerapan model tersebut pada teks berita yang belum diberi label untuk mengetahui apakah model mampu mengklasifikasi berita dengan

tepat. Sebagai contoh, konten berita akan diambil dari situs Kompas dan CNN. Tabel 2 menampilkan beberapa contoh hasil prediksi kategori berita olahraga.

Tabel 1. Contoh Hasil Peringkasan Berita

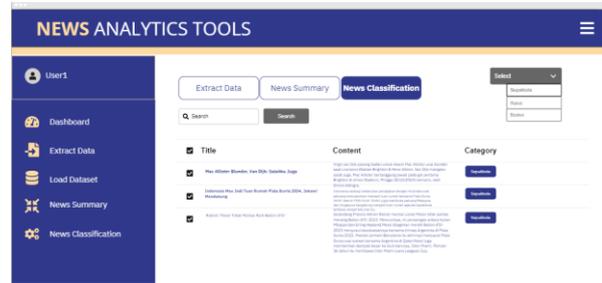
Attribute	Data
Link	https://sport.detik.com/sepakbola/bola-dunia/d-6975897/rabiot-messi-tidak-pantas-raih-ballon-dor
Title	Rabiot: Messi Tidak Pantas Raih Ballon d'Or
Category	Sepakbola
Content	Gelandang Prancis Adrien Rabiot menilai Lionel Messi tidak pantas menang Ballon d'Or 2023. Menurutnya, ini persaingan antara Kylian Mbappe dan Erling Haaland.Messi dijagokan meraih Ballon d'Or 2023 menyusul kesuksesannya bersama timnas Argentina di Piala Dunia 2022. Mantan pemain Barcelona itu akhirnya menjuarai Piala Dunia usai sukses bersama Argentina di Qatar.Messi juga memberikan dampak besar ke klub barunya, Inter Miami. Pemain 36 tahun itu membawa Inter Miami juara Leagues Cup. Namun, Messi dianggap gagal dan dapat kritik saat masih di Paris Saint-Germain. Meski juara Ligue 1, PSG gagal di Liga Champions."Saya terus mendengar kalau Messi akan melakukannya. Tapi dalam kaitannya dengan olahraga, seharusnya antara Kylian Mbappe dan Erling Haaland," ujar Rabiot kepada RMC Sport."Sulit untuk memilih satu dari yang lain, tergantung apa dasarnya. Tidak semua orang akan setuju, tapi ini akan antara mereka berdua," katanya. Kylian Mbappe memang berkaitan dengan Lionel Messi di level klub karena pernah satu tim di PSG. Namun di level individu, Mbappe juga bersinar di Piala Dunia Dunia 2022 dengan meraih Golden Boot.Sementara itu, Erling Haaland sukses besar bersama Manchester City musim lalu. Striker Norwegia itu mencetak total 52 gol di semua kompetisi dan ikut mengantarkan The Citizens meraih treble.
Summary	Menurutnya, ini persaingan antara Kylian Mbappe dan Erling Haaland.Messi dijagokan meraih Ballon d'Or 2023 menyusul kesuksesannya bersama timnas Argentina di Piala Dunia 2022. Mantan pemain Barcelona itu akhirnya menjuarai Piala Dunia usai sukses bersama Argentina di Qatar.Messi juga memberikan dampak besar ke klub barunya, Inter Miami. Tidak semua orang akan setuju, tapi ini akan antara mereka berdua," katanya. Kylian Mbappe memang berkaitan dengan Lionel Messi di level klub karena pernah satu tim di PSG. Namun di level individu, Mbappe juga bersinar di Piala Dunia Dunia 2022 dengan meraih Golden Boot.Sementara itu, Erling Haaland sukses besar bersama Manchester City musim lalu.

Tabel 2. Contoh Hasil Prediksi Klasifikasi Berita Olahraga

No	Sumber	Konten Berita	Prediksi Kategori
1	Kompas	Tunggal putra Denmark, Viktor Axelsen, melempar raket seuaai juara Indonesia Open 2023. Viktor Axelsen juara Indonesia Open 2023 setelah mengalahkan wakil tuan rumah, Anthony Sinisuka Ginting, pada laga final di Istora Senayan, Jakarta, Minggu (18/6/2023). Axelsen meraih kemenangan straight game atau dua gim langsung dengan skor 21-14, 21-13.	Raket

No	Sumber	Konten Berita	Prediksi Kategori
2	CNN	<p>Axelsen lantas melakukan selebrasi usai memastikan gelar juara. Peraih medali emas Olimpiade Tokyo 2020 itu memberikan dua raket kepada penonton. Raket pertama ia berikan kepada seorang penggemar yang berada di tribune belakang lapangannya.</p> <p>Indonesia harus tampil di babak playoff karena menjadi juru kunci Grup G Kualifikasi Piala Dunia 2022. Indonesia termasuk empat tim yang performanya paling buruk dalam kualifikasi tersebut. Indonesia akan bertindak sebagai tuan rumah terlebih dahulu pada 7 September 2021. Sedangkan pertandingan leg kedua akan berlangsung pada 12 Oktober 2021 di Taiwan.</p> <p>Sebanyak 24 tim yang dibagi menjadi enam grup akan tampil di babak kualifikasi Piala Asia 2023. Sebanyak 11 tiket ke putaran final Piala Asia 2023, yang akan berlangsung di China pada 16 Juni hingga 16 Juli 2023, diperebutkan.</p>	Sepak bola

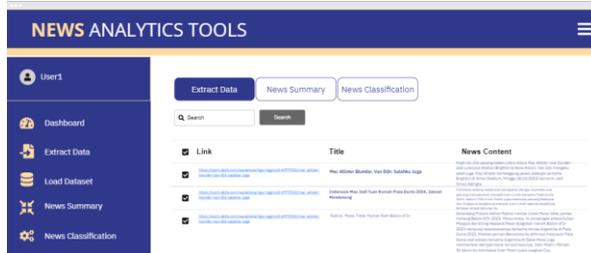
Konten berita yang sudah diringkas kemudian dijadikan masukan untuk membangun model klasifikasi. Gambar 6 menampilkan hasil klasifikasi berita untuk data-data yang sudah dimasukkan.



Gambar 6. Hasil Klasifikasi Berita

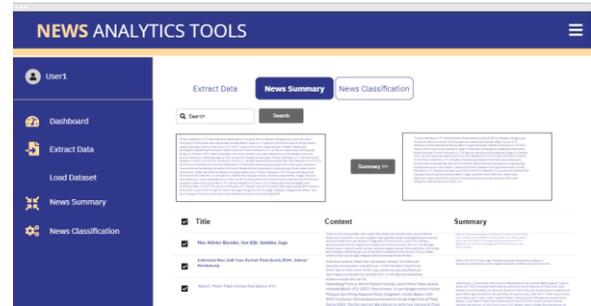
Setelah diperoleh model klasifikasi yang kemudian disimpan pada sistem, model ini dapat digunakan untuk memprediksi klasifikasi konten berita yang baru atau dari sumber berita yang berbeda, baik dengan menyimpan konten berita dalam satu dataset khusus maupun secara langsung dengan cara menyalin konten berita pada area teks dan memanggil fungsi klasifikasi sesuai model yang sudah dibangun sebelumnya (Gambar 7).

Setelah diperoleh model prediksi yang baik, selanjutnya model tersebut diterapkan pada antarmuka berbasis web, dengan memanfaatkan pustaka ngrok untuk Python pada google colab. Gambar 4 menampilkan antarmuka untuk hasil ekstrak data dari berbagai sumber berita.

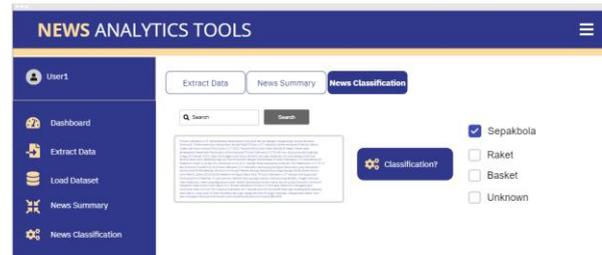


Gambar 4. Antarmuka Ekstrak Data dari Sumber Berita

Setelah diperoleh konten berita, selanjutnya dapat dilakukan proses peringkasan, baik dilakukan secara kolektif ataupun untuk satu blok artikel yang disesuaikan dengan kebutuhan pengguna (Gambar 5).



Gambar 5. Antarmuka Proses Peringkasan Berita



Gambar 7. Prediksi Klasifikasi Berita

Tahapan selanjutnya adalah publikasi web menggunakan fitur tunneling dari ngrok, dengan langkah-langkah sebagai berikut: Menyimpan model untuk peringkasan dan klasifikasi berita; Instalasi pustaka *pyngrok* pada *google colab*; Menyiapkan *ngrok autotoken*; Menyiapkan platform aplikasi berbasis web; Menjalankan aplikasi web (*mode external*), dan catat port yang digunakan; Memanggil *ngrok* dari *Python* dan memulai *tunelling/connection*, catat *public-url* yang dihasilkan.

Sebagai contoh, berikut skrip untuk memanggil aplikasi dan menampilkan public url dari ngrok. Pada contoh ini, port yang digunakan adalah 8000, dan contoh public url yang dihasilkan adalah: <http://02cb-33-82-95-132.ngrok.io>

```
ngrok_tunnel = ngrok.connect(8000)
print('Public URL:', ngrok_tunnel.public_url)
nest_asyncio.apply()
uvicorn.run(app, port=8000)
```

Hasil eksekusi

INFO: Started server process [61]
 INFO: Waiting for application startup.

INFO: Application startup complete.
INFO: Uvicorn running on <http://127.0.0.1:8000>
(Press CTRL+C to quit)
Public URL: <http://02cb-33-82-95-132.ngrok.io>

Dalam mengembangkan sistem peringkasan berita ini, masih terdapat kendala yaitu: Model klasifikasi yang dibangun, masih dibatasi untuk tiga kategori, sehingga masih perlu ditambahkan kemampuan untuk klasifikasi berdasarkan kategori lainnya; Kemampuan peringkasan perlu ditambah dengan penetapan keyword dan penentuan topik, agar lebih memudahkan praktisi jurnalistik melakukan analisis berita; dan *hosting* yang dilakukan masih bersifat sementara (*temporary*), sehingga perlu disiapkan platform yang lebih permanen untuk hosting agar sistem ini dapat dimanfaatkan secara luas.

4. Kesimpulan

Berdasarkan hasil penelitian yang sudah diuraikan sebelumnya, dapat disimpulkan bahwa sistem peringkasan dan klasifikasi berita olahraga yang dibangun dengan menggunakan metode textrank untuk peringkasan berita dan KNN untuk klasifikasi berita sudah dapat membantu menjawab beberapa permasalahan seperti kemampuan untuk melakukan web scrapping untuk mendapatkan konten berita secara otomatis, fasilitas peringkasan konten berita dan penentuan kategori berita olahraga pada tiga kategori yaitu sepakbola, raket dan basket. Sistem juga dapat berjalan pada lingkungan komputasi awan karena dijalankan pada google colab yang dapat dioperasikan oleh siapa saja menggunakan akun Google masing-masing. Sedangkan antarmuka berbasis web dapat difasilitasi dengan memanfaatkan fasilitas dari ngrok. Pengembangan lebih lanjut dapat diarahkan dengan menambahkan beberapa fitur pada fungsi peringkasan seperti ekstrak keywords, penentuan topik, serta penambahan jenis kategori pada fitur klasifikasi berita

Daftar Rujukan

- [1] C. Zhu, "Applications and future of machine reading comprehension," in *Machine Reading Comprehension*, Elsevier, 2021, pp. 185–207.
- [2] D. Miller, "Leveraging BERT for Extractive Text Summarization on Lectures," Jun. 2019, [Online]. Available: <http://arxiv.org/abs/1906.04165>.
- [3] N. Zhou, W. Shi, R. Liang, and N. Zhong, "TextRank Keyword Extraction Algorithm Using Word Vector Clustering Based on Rough Data-Deduction," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–19, Jan. 2022, doi: 10.1155/2022/5649994.
- [4] S. Kemahduta, "Automatic Text Summarization dengan kategorisasi pada berita online mengenai tokoh masyarakat indonesia dengan metode Fuzzy Logic," Universitas Sebelas Maret, 2019.
- [5] H. Gupta and M. Patel, "Method Of Text Summarization Using Lsa And Sentence Based Topic Modelling With Bert," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Mar. 2021, pp. 511–517, doi: 10.1109/ICAIS50930.2021.9395976.
- [6] K. U. Manjari, S. Rousha, D. Sumanth, and J. Sirisha Devi, "Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, Jun. 2020, pp. 648–652, doi: 10.1109/ICOEI48184.2020.9142938.
- [7] P. Modaresi, P. Gross, S. Sefidrodi, M. Eckhof, and S. Conrad, "On (Commercial) Benefits of Automatic Text Summarization Systems in the News Domain: A Case of Media Monitoring and Media Response Analysis," Jan. 2017, [Online]. Available: <http://arxiv.org/abs/1701.00728>.
- [8] K. S. Thakkar, R. V. Dharaskar, and M. B. Chandak, "Graph-Based Algorithms for Text Summarization," in *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, Nov. 2010, pp. 516–519, doi: 10.1109/ICETET.2010.104.
- [9] A. Abdurrohman, "Evaluasi Algoritma Textrank pada Peringkasan Teks Berbahasa Indonesia," Universitas Sumatera Utara, 2018.
- [10] Y. Marsyah and S. H. Wijaya, "Perbandingan Kinerja Algoritme TextRank dengan Algoritme LexRank pada Peringkasan Dokumen Bahasa Indonesia," IPB University, 2013.
- [11] S. R. K. Harinatha, B. T. Tasara, and N. N. Qomariyah, "Evaluating Extractive Summarization Techniques on News Articles," in *2021 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Jul. 2021, pp. 88–94, doi: 10.1109/ISITIA52817.2021.9502230.
- [12] M. Zhang, X. Li, S. Yue, and L. Yang, "An Empirical Study of TextRank for Keyword Extraction," *IEEE Access*, vol. 8, pp. 178849–178858, 2020, doi: 10.1109/ACCESS.2020.3027567.
- [13] S. Mishra, M. Kuznetsov, G. Srivastava, and M. Sviridenko, "VisualTextRank: Unsupervised Graph-based Content Extraction for Automating Ad Text to Image Search," Aug. 2021, doi: 10.1145/1122445.1122456.
- [14] Y. Chen and Q. Song, "News Text Summarization Method based on BART-TextRank Model," in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Mar. 2021, pp. 2005–2010, doi: 10.1109/IAEAC50856.2021.9390683.
- [15] N. A. Maghfiroh, G. W. Wicaksono, and C. S. K. Aditya, "Peringkasan Berita Online Corona Virus dengan Metode Lexical Chain dan Word Sense Disambiguation," *Komputika J. Sist. Komput.*, vol. 10, no. 2, pp. 145–151, Aug. 2021, doi: 10.34010/komputika.v10i2.4499.
- [16] S. Tuhpatussania and L. M. Nurkholis, "Automatic Text Summarization Artikel Berita Menggunakan Metode Maximum Marginal Relevance," *EXPLORE*, vol. 12, no. 2, p. 18, Jul. 2022, doi: 10.35200/explore.v12i2.543.
- [17] G. W. Wicaksono, M. N. M. Hakim, N. Hayatin, N. P. Hidayah, and T. I. Sari, "Text Summarization on Verdicts of Industrial Relations Disputes Using the Cross-Latent Semantic Analysis and Long Short-Term Memory," *JOIV Int. J. Informatics Vis.*, vol. 7, no. 3, pp. 847–853, Sep. 2023, doi: 10.30630/joiv.7.3.2052.
- [18] B. Imran, M. N. Karim, and N. I. Ningsih, "KLASIFIKASI BERITA HOAX TERKAIT PEMILIHAN UMUM PRESIDEN REPUBLIK INDONESIA TAHUN 2024 MENGGUNAKAN NAÏVE BAYES DAN SVM," *Din. Rekayasa*, vol. 20, no. 1, pp. 1–9, Jan. 2024, doi: 10.20884/1.dinarek.2024.20.1.27.
- [19] A. S. Ridwan, Y. H. Chrisnanto, and R. Ilyas, "KLASIFIKASI KALIMAT PADA BERITA OLAAHRAGA SECARA OTOMATIS MENGGUNAKAN METODE ARTIFICIAL NEURAL NETWORK," *J. Komput. dan Inform.*, vol. 9, no. 1, pp. 88–97, Apr. 2021, doi: 10.35508/jicon.v9i1.3708.
- [20] G. Elisabeth, Rahma Salsa Bilah, S. N. Ardini, N. Agustina, and D. A. Rismayadi, "KLASIFIKASI BERITA PALSU KENAIKAN HARGA BAHAN BAKAR MINYAK (BBM) MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 2, pp. 104–109, Dec. 2023, doi: 10.53580/naratif.v5i2.188.
- [21] R. M. Juwita, E. Haerani, S. K. Gusti, and S. Ramadhani,

- “Klasifikasi Berita Menggunakan Metode K-Nearest Neighbor,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 2, pp. 259–268, Apr. 2022, doi: 10.32672/jnkti.v5i2.4192.
- [22] J. Ahmed and M. Ahmed, “ONLINE NEWS CLASSIFICATION USING MACHINE LEARNING TECHNIQUES,” *IJUM Eng. J.*, vol. 22, no. 2, pp. 210–225, Jul. 2021, doi: 10.31436/iiumej.v22i2.1662.
- [23] Nur Ghaniaviyanto Ramadhan, “Indonesian Online News Topics Classification using Word2Vec and K-Nearest Neighbor,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1083–1089, Dec. 2021, doi: 10.29207/resti.v5i6.3547.
- [24] K. Munawaroh and A. Alamsyah, “Performance Comparison of SVM, Naïve Bayes, and KNN Algorithms for Analysis of Public Opinion Sentiment Against COVID-19 Vaccination on Twitter,” *J. Adv. Inf. Syst. Technol.*, vol. 4, no. 2, pp. 113–125, Mar. 2023, doi: 10.15294/jaist.v4i2.59493.
- [25] Ngrok, “What is ngrok.” <https://ngrok.com/docs/what-is-ngrok/>.
- [26] A. RS, “Quickly share ML WebApps from Google Colab using ngrok for Free,” *Toward Data Science*, 2020. <https://towardsdatascience.com/quickly-share-ml-webapps-from-google-colab-using-ngrok-for-free-ae899ca2661a> (accessed Apr. 04, 2024).
- [27] G. . Santos, P.S.M, Travassos, “Action Research Can Swing the Balance in Experimental Software Engineering,” *Adv. Comput.*, vol. 83, pp. 205–276.
- [28] M. Staron, *Action Research in Software Engineering*. Cham: Springer International Publishing, 2020.
- [29] S. M, “Action Research in Software Engineering: Metrics’ Research Perspective,” in *Theory and Practice of Computer Science. SOFSEM 2019. Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2019.
- [30] N. Davidson, R.M, Martinsons, M.G., Kock, “Systems Journal: Principles of Canonical Action Research,” *J. Inf.*, vol. 14, pp. 65–86, 2004.
- [31] Falahah, “Summarization and Classification of Sports News using Textrank and KNN,” *Int. J. Syst. Eng. Inf. Technol.*, vol. 3, no. 1, pp. 23–29, 2024, doi: 10.29207/joseit.v3i1.5706.