

Implementasi Sistem Automatic Text Summarization Berbasis Fitur dan Metode Jaringan Syaraf Tiruan Propagasi Balik

Muhammad Sulthan Syaddad¹, Mohammad Syafrullah²

^{1,2}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Indonesia

¹sulthansyaddad331@gmail.com, ²mohammad.syafrullah@budiluhur.ac.id

Article Info

Article history:

Received Aug 27, 2023

Revised Apr 02, 2024

Accepted Apr 04, 2024

Keywords:

Text Summarization

Neural Network

Backpropagation

Feature-Based

ABSTRACT

In the era of Industry 4.0, information has become a primary necessity for society today as it enables people to know about various current events worldwide. With the rapid development of information technology and the internet, there has been an abundance of documents available that we can search according to our needs. Text Summarization Machines have the function of presenting essential information from the original documents in a shorter format while still preserving the main content and helping users understand the information from lengthy documents faster. In this case, the method used is the Text Summarization Feature-Based approach, utilizing the Backpropagation Artificial Neural Network algorithm for sentence prediction calculations. The Backpropagation Artificial Neural Network algorithm seeks the most optimal weights during its process. In the testing process with five document samples, the final result obtained was a text summary model that could predict the overall number of labels correctly. However, it struggled in predicting which ones should be labeled as "true" and which ones should be labeled as "false".

Copyright © 2024 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Muhammad Sulthan Syaddad,
Teknik Informatika, Fakultas Teknologi Informasi,
Universitas Budi Luhur,
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260, DKI Jakarta.
Email: sulthansyaddad331@gmail.com

1. PENDAHULUAN

Pada era industri 4.0 informasi menjadi kebutuhan utama masyarakat saat ini karena memungkinkan masyarakat untuk mengetahui berbagai peristiwa terkini di seluruh dunia. Dengan berkembangnya teknologi informasi dan teknologi internet yang berkembang semakin pesat, telah memungkinkan munculnya dokumen-dokumen yang tersedia secara melimpah yang dapat kita cari sesuai kebutuhan kita. Namun, terdapat kesulitan untuk membaca dokumen secara cepat dan dapat mengerti apa saja informasi yang tersimpan dalam dokumen tersebut. Untuk Membaca dan memahami seluruh konten sebuah berita membutuhkan waktu yang lebih lama, sedangkan dengan menggunakan ringkasan teks saja, kita dapat memahami informasi yang terdapat dalam dokumen dengan lebih efisien. Oleh karena itu, ringkasan teks memiliki peran yang sangat penting dalam memahami isi dokumen untuk memperoleh suatu informasi yang dibutuhkan secara cepat. Dengan adanya ringkasan, pembaca dapat dengan mudah dan cepat memahami isi dokumen tanpa harus membaca keseluruhan dokumen tersebut. Hal ini membantu menghemat waktu pembaca dengan menghindari membaca suatu dokumen yang tidak relevan dengan informasi yang diharapkan oleh pembaca, terutama karena jumlah dokumen yang tersedia di internet sangat banyak saat ini. Mesin Peringkasan Teks hadir sebagai alat yang cukup penting dan efisien dalam membantu mengurangi waktu yang diperlukan untuk membaca dokumen yang tebal.

Mesin Peringkasan Teks memiliki fungsi untuk menyajikan informasi penting dari dokumen asli dalam format yang lebih singkat, tetapi tetap mempertahankan konten utama dan membantu pengguna memahami informasi dari dokumen yang panjang secara lebih cepat [1] [2] [3]. Mesin ini berfokus pada dua masalah utama, yaitu pemilihan bagian terpenting dari teks dan pembuatan ringkasan yang koheren. Proses ini berbeda dengan cara manusia membuat ringkasan, karena manusia dapat menangkap makna mendalam dan tema dokumen dengan lebih baik, suatu hal yang sulit diimplementasikan dalam mesin. Terdapat dua Teknik dari mesin peringkasan teks yaitu ekstraksi dan abstraksi. Teknik ekstraksi bekerja dengan cara memilih kalimat paling penting dan informatif dari sebuah dokumen. Sedangkan Teknik abstraksi mengambil ide utama dari sebuah dokumen dan membuat kalimat baru dari ide tersebut [4].

Selama masa pengembangannya, Mesin Peringkasan Teks menemui beberapa kendala yang membuat mesin ini memiliki batasannya tersendiri. Pada model Encoder-Decoder, hasil pengukuran interaksi antara tahap encoder dan decoder dengan nilai yang besar mengidentifikasi bahwa masukan ke i telah digunakan dan sulit untuk digunakan kembali untuk mencetak keluaran ke t . Kendala yang selanjutnya dihadapi yaitu hasil dari membandingkan abstrak yang dihasilkan dari sistem dan yang dihasilkan secara manual hanya memiliki sedikit kalimat saja dengan tingkat kemiripannya tertinggi. Ini dapat terjadi karena penulis biasanya memiliki susunan kalimatnya sendiri yang umumnya memiliki kata yang tidak ada pada badan teks. Jumlah kalimat yang dihasilkan oleh mesin ini juga berpengaruh terhadap seberapa banyaknya informasi penting yang dapat dicakup dari teks aslinya.

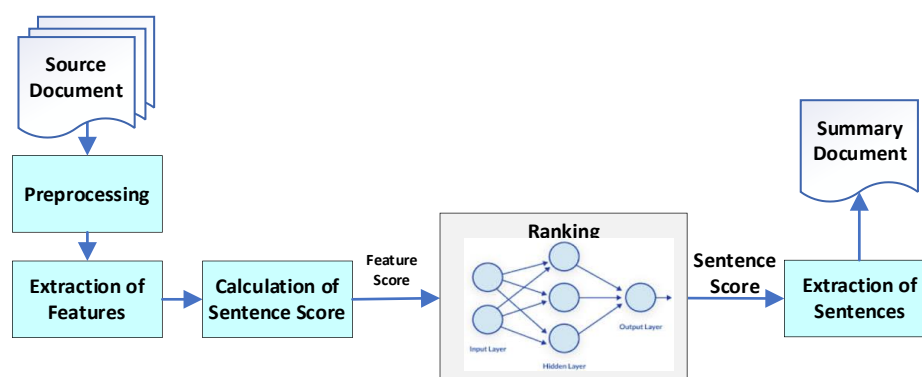
Penelitian sebelumnya yang dilakukan oleh Logan Lebanoff [5] menjelaskan bahwa Mesin Peringkasan Teks yang dibuatnya menggunakan metode abstraksi dengan algoritma PG-MMR mendemonstrasikan hasil yang kompetitif. Lalu pada penelitian C Slamet [6] menjelaskan bahwa model VSM yang digunakan untuk meringkas jurnal menghasilkan ringkasan yang dimana hasilnya menunjukkan perbandingan antara ringkasan menggunakan sistem dan manual sampai pada keluaran empat kalimat dengan tingkat kemiripan tertinggi. Selanjutnya, pada penelitian Ladda Suanmali [7] menjelaskan penerapan metode ROUGE dan algoritma fuzzy logic dapat mengekstrak ringkasan se-informatif mungkin seperti keseluruhan teks pada dokumen dengan cakupan informasi yang lebih baik.

Membuat Mesin Peringkasan Teks otomatis dilakukan dengan cara pemberian bobot menggunakan beberapa fitur untuk setiap kalimat dan mencari kata yang mirip dengan judul dokumen dari kata yang terdapat didalam kalimat. Dalam penelitian ini ringkasan yang dihasilkan merupakan kumpulan kalimat dari hasil perhitungan dengan metode Jaringan Syaraf Tiruan Propagasi Balik. Metode ini bertujuan untuk mengambil sebagian kalimat dari keseluruhan teks dokumen yang penting, ditandai dengan bobot tinggi dari hasil perhitungan dengan metode Jaringan Syaraf Tiruan Propagasi Balik. Keluaran yang dihasilkan akan menampilkan sebuah ringkasan dokumen yang diharapkan sesuai dengan kebutuhan pembaca.

2. METODE PENELITIAN

2.1 Penerapan Metode

Pada proses penerapan dalam membangun sistem Text Summarization, topik ini akan menggunakan basis fitur (Feature based) dan jaringan syaraf tiruan propagasi balik untuk memberi nilai atau skor dengan perhitungan matematis masing-masing fitur untuk setiap kata dan kalimat yang ada di dalam dataset yang telah ditentukan, yang bertujuan untuk mencari kalimat mana saja yang memiliki relevansi tertinggi terhadap topik utama dokumen asal. Pada penelitian yang dilakukan ini, terdapat beberapa tahap yang akan dibuat dan bisa dilihat pada gambar 1 dibawah ini. Gambar 1 merupakan suatu proses bagaimana data atau dokumen yang dimasukkan ke dalam mesin Text Summarization diolah dan kemudian dapat menghasilkan sebuah ringkasan teks baru.



Gambar 1. Flowchart Tahapan Proses Sistem Text Summarization

a. Pengumpulan Dataset

Berdasarkan studi literatur yang telah dipelajari sebelumnya, hal paling utama yang harus dilakukan adalah mengumpulkan dataset yang akan diolah untuk menghasilkan informasi baru. Data yang diperoleh untuk penelitian ini didapat dari data publik yaitu dataset IndoSum yang sudah memiliki gold standard.

b. Transformasi dan Preprocessing Dataset

Setelah menyaring data apa saja yang bisa digunakan, tahap selanjutnya adalah melakukan *preprocessing* data dengan melakukan transformasi data dari format json ke dalam format *worksheet*. Tahap ini dilakukan dengan cara manual yang ada pada Microsoft excel dikarenakan format data sulit untuk diterapkan. Sehingga, pada tahap ini penulis hanya menggunakan beberapa attribute yang diperlukan. Kemudian attribut tersebut selanjutnya akan dibersihkan melalui tahapan *preprocessing* dengan cara dilakukan penghilangan karakter maupun term yang tidak penting. Selanjutnya, data yang sudah dibersihkan diubah ke bentuk numerik sehingga dapat dilakukan perhitungan matematis [8] sebelum akhirnya siap untuk kebutuhan yang diperlukan dalam tahap pengolahan.

c. Perhitungan Fitur

Pada tahap ini data yang sudah melalui tahap *preprocessing* akan dilakukan perhitungan matematis berdasarkan 6 fitur yang akan diterapkan pada penelitian ini. 6 fitur ini diantaranya yaitu *Title Feature*, *Sentence Length Feature*, *Sentence Position Feature*, *Sentence to Sentence Similarity Feature*, *Thematic Word Feature* dan *Numerical Word Feature* yang akan dibahas pada bagian 2.2.

d. Proses Perhitungan Prediksi Jaringan Syaraf Tiruan Propagasi Balik

Tahap selanjutnya adalah menerapkan proses perhitungan prediksi menggunakan algoritme jaringan syaraf tiruan propagasi balik. Menentukan prediksi ini merupakan hal yang sangat penting karena ini berdampak pada ringkasan teks yang dihasilkan oleh Text Summarization. Kalimat yang memiliki nilai dengan semakin mendekati angka '1' maka berpeluang besar dipilih untuk menjadi kalimat dalam hasil ringkasan teks.

e. Ekstraksi Ringkasan Teks

Tahap terakhir yaitu mengekstraksi kalimat-kalimat yang memiliki nilai akhir tertinggi dan menjadikannya ringkasan teks terbaru. Ini dapat dilakukan dengan cara mengurutkan skor nilai akhir dari kalimat-kalimat yang sudah diproses sebelumnya. Selanjutnya dipilih berapa banyak kalimat yang akan dijadikan ringkasan teks yang baru. Terakhir menampilkan keluaran ringkasan teks yang sudah dibuat ke tampilan layar. Berisi penjelasan tentang tahapan penelitian yang menggambarkan urutan logis untuk mendapatkan hasil penelitian sesuai dengan harapan dan gambaran sistem. Jika ada gambar dan tabel, itu harus disajikan dengan nama tabel dan gambar yang disertai dengan nomor urut.

2.2 Fitur Kalimat

Setelah dokumen selesai melewati tahap *preprocessing*, tiap kalimat dalam dokumen diubah menjadi atribut vektor yang mencerminkan berbagai fitur. Fitur-fitur ini adalah atribut yang berusaha mewakili data sesuai dengan tugasnya. Pada penelitian ini penulis akan berfokus pada tujuh fitur untuk setiap kalimat. Setiap fitur diberi nilai antara '0' sampai '1' [7]. Terdapat enam fitur, diantaranya adalah:

a. Title Feature

Kata-kata yang ada dalam kalimat dan juga ada dalam judul dokumen memberikan skor yang cukup tinggi. Ini dapat ditentukan dengan menghitung jumlah kata yang cocok antara kata yang ada di kalimat dan kata yang ada pada judul. Penulis menghitung skor fitur ini dengan membagi jumlah kata yang muncul baik dalam kalimat maupun judul dengan total kata dalam judul.

$$S_{F1(s)} = \frac{\Sigma \text{ kata yang muncul pada } S}{\Sigma \text{ kata pada judul}} \quad (1)$$

b. Sentence Length

Fitur ini berguna dalam menyaring kalimat pendek seperti jadwal dan nama penulis yang sering ditemukan dalam artikel berita. Kalimat-kalimat pendek ini tidak diharapkan muncul pada ringkasan. Penulis menggunakan panjang suatu kalimat dengan membagi jumlah kata yang muncul pada kalimat dengan jumlah kata yang muncul pada kalimat terpanjang dari sebuah dokumen.

$$S_{F2(s)} = \frac{\Sigma \text{ kata yang muncul pada } S}{\Sigma \text{ kata yang muncul pada kalimat terpanjang}} \quad (2)$$

c. Sentence Position

Baik itu pada 5 kalimat pertama di paragraf, posisi kalimat dalam teks memberi indikasi tentang seberapa penting kalimat tersebut. Fitur ini melibatkan beberapa elemen seperti posisi dari sebuah kalimat dalam dokumen, paragraf, dan lainnya. Kalimat pertama dianggap sebagai kalimat dengan peringkat tertinggi. Dengan menggunakan skor dari fitur ini, penulis dapat menentukan bahwa lima kalimat pertama dalam paragraf tersebut merupakan kalimat yang paling penting. Rumus untuk menghitung skor fitur ini dapat dirumuskan sebagai berikut.

$$S_{F4(S)} = \begin{array}{l} 5/5 \text{ untuk kalimat 1, } 4/5 \text{ untuk kalimat 2,} \\ 3/5 \text{ untuk kalimat 3, } 2/5 \text{ untuk kalimat 4,} \\ 1/5 \text{ untuk kalimat 5, } 0/5 \text{ untuk kalimat lainnya} \end{array} \quad (3)$$

d. Sentence to Sentence Similarity

Fitur ini menghitung tingkat kesamaan antar kalimat. Untuk setiap kalimat S , tingkat kesamaan antara S dan kalimat lain dihitung menggunakan *cosine similarity* dengan rentang nilai berkisar antara '0' dan '1' [9]. Bobot term w_i dan w_j dari term t hingga term ke n dalam kalimat S_i dan S_j direpresentasikan sebagai vektor. Tingkat kesamaan antara setiap pasangan kalimat dapat dihitung menggunakan rumus similarity sebagai berikut [10].

$$Sim(S_i, S_j) = \frac{\sum_{t=1}^n w_{it} \times w_{jt}}{\sqrt{\sum_{t=1}^n w_{it}^2} \times \sqrt{\sum_{t=1}^n w_{jt}^2}} \quad (4)$$

Skor fitur ini untuk kalimat S diperoleh dengan menghitung rasio antara tingkat kesamaan kalimat S dengan setiap kalimat lainnya dalam ringkasan, kemudian dibagi dengan nilai maksimum tingkat kesamaan dalam ringkasan.

$$S_{F5(S)} = \frac{\sum Sim(S_i, S_j)}{Max(\sum Sim(S_i, S_j))} \quad (5)$$

e. Thematic Word

Jumlah kata tematik dalam sebuah kalimat menjadi fitur penting karena term yang sering muncul dalam dokumen memiliki kemungkinan berkaitan dengan topik dokumen tersebut. Jumlah kata tematik mengindikasikan bahwa kata-kata tersebut memiliki hubungan dengan tingkat kemungkinan yang tinggi dengan topik dokumen. Penulis menggunakan 10 kata dengan frekuensi tertinggi sebagai pertimbangan kata-kata tematik. Skor fitur ini diungkapkan sebagai rasio antara jumlah kata tematik yang muncul dalam kalimat dibagi dengan nilai maksimum dari jumlah kata tematik dalam ringkasan kalimat.

$$S_{F7(S)} = \frac{Jumlah \ kata \ thematic \ pada \ S}{Max(Jumlah \ kata \ thematic)} \quad (6)$$

f. Numerical Data

Jumlah data numerik dalam sebuah kalimat menjadi fitur penting, karena kalimat yang mengandung data numerik memiliki kemungkinan tinggi untuk dimasukkan dalam ringkasan dokumen. Skor untuk fitur ini dirumuskan dengan membagi jumlah data numerik yang muncul dalam kalimat dengan panjang kalimat.

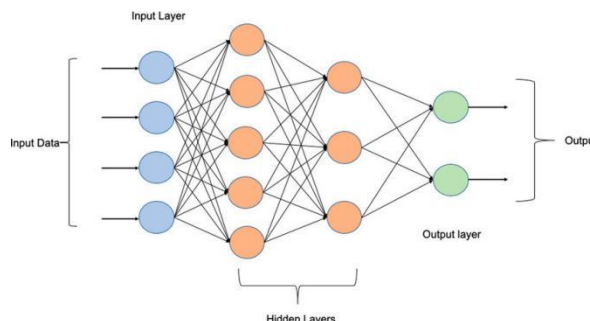
$$S_{F8(S)} = \frac{Jumlah \ data \ numerik \ pada \ S}{Panjang \ Kalimat \ (S)} \quad (7)$$

2.3 Neural Network Backpropagation: Multi-Layer Perceptron Regresor

Jaringan syaraf tiruan (Artificial Neural Network) merupakan sebuah penerapan sistem komputasi yang memiliki dasar analogi pemodelan dari jaringan syaraf biologis otak manusia dan karena itu banyak terminologi dipinjam dari *neuroscience* [11]. Jaringan syaraf tiruan terhubung untuk tiap prosesor sederhana melalui koneksi berbobot. Pendekatan dari jaringan syaraf tiruan dalam pengenalan pola, didasarkan pada cara pembelajaran yang dipakai untuk mendapatkan output dari jaringan. Pada pendekatan kali ini sistem akan dilatih sebelumnya untuk mendapatkan output yang diinginkan.

a. *Backpropagation (Propagasi Balik)*

Backpropagation merupakan algoritma jaringan syaraf tiruan yang bekerja dengan cara memberikan bobot pada jaringan secara acak, yang kemudian bobot akan disesuaikan saat melakukan perhitungan pengkoreksian pada saat propagasi balik (ke arah lapisan masukan). Propagasi balik terjadi setelah jaringan menghasilkan keluaran yang mengandung error[11].



Gambar 1. Arsitektur Backpropagation

Backpropagation meliputi 2 fase, yaitu Forward-Phase (ke arah lapisan keluaran) yang bertujuan untuk menghasilkan keluaran, dan Backward-Phase (ke arah lapisan masukan) yang bertujuan untuk melakukan pengkoreksian bobot. Proses ini dilakukan secara berulang kali sampai ditemukan nilai error terkecil untuk semua jaringan[12]. Arsitektur pada metode Backpropagation dapat dilihat pada Gambar 1. Setelah backpropagation selesai menghasilkan keluaran, Hasil pelatihan tersebut akan disimpan dan kemudian digunakan untuk menghitung nilai output dari sampel baru.

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi analisis, hasil implementasi ataupun pengujian serta pembahasan dari topik penelitian, yang bisa dibuat terlebih dahulu metodologi penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan, gambar, tabel dan lainnya.

3.1 Implementasi Metode

a. Pengumpulan Dataset

Untuk melakukan implementasi metode ke dalam sistem aplikasi yang akan dibuat, tahap awal yang harus dilakukan yaitu mengumpulkan dataset. Dataset yang akan digunakan adalah data dari IndoSum yang ditransformasikan ke dalam worksheet dan sudah memiliki *gold standard* untuk menjadi standar acuan.

b. Seleksi Dataset

Setelah dataset berhasil didapatkan, maka tahap selanjutnya yaitu menyeleksi dataset. Penyeleksian dataset ini bertujuan untuk memilih data mana saja yang akan dipakai untuk proses *Text Summarization*. Pada Gambar 2 merupakan sampel Dataset IndoSum yang pada penelitian ini penulis akan menggunakan data Paragraphs dan Gold_Standard yang ditransformasikan ke bentuk worksheet.

Tabel 2. Sampel Perhitungan Tiap-Tiap Fitur

Title_Feature	Sentence_Length_Feature	Sentence_Position_Feature	Sentence_To_Similarity_Feature	Thematic_Word_Feature	Numerical_Data_Feature
0.625	0.8461538461538461	1.0	0.2617801047120419	0.2	0.0
0.5	0.9230769230769231	0.9444444444444444	0.5314685314685315	0.4	0.0
0.375	1.0	0.8888888888888888	0.39892183288409705	0.3	0.0
0.0	0.15384615384615385	0.8333333333333333	0.2530612244897959	0.1	0.0
0.0	0.5384615384615384	0.7777777777777778	0.3058823529411765	0.0	0.0
0.0	0.7692307692307693	0.7222222222222222	0.16216216216216217	0.0	0.02222222222222223
0.125	0.5384615384615384	0.6666666666666666	0.17757009345794392	0.0	0.0

e. Perhitungan Prediksi dengan Jaringan Syaraf Tiruan Propagasi Balik

Proses perhitungan prediksi menggunakan jaringan syaraf tiruan propagasi balik ini bertujuan untuk memberikan label prediksi yang ada pada masing-masing kalimat agar dapat mengetahui kalimat mana saja yang akan masuk ke hasil akhir ringkasan teks berdasarkan pengujian pada dataset Indosum. Pada penelitian ini penulis menggunakan 10.000 iterasi, learning rate bersifat adaptive, dan hidden layer sebanyak 100 x 100 node. Sebagai contoh, dapat dilihat pada Gambar 3 dibawah ini penulis akan menggunakan jaringan syaraf tiruan propagasi balik pada beberapa kalimat:

Gold Label: 1	Hasil Prediksi: 1	Ketua MPR Zulkifli Hasan menyesalkan kisruh yang terjadi antara pelaku sarana transportasi
Gold Label: 1	Hasil Prediksi: 1	Zulkifli menyarankan adanya pertemuan bersama antara pemerintah, pelaku transportasi online dan
Gold Label: 1	Hasil Prediksi: 1	Zulkifli menyarankan adanya pertemuan bersama antara pemerintah, pelaku transportasi online dan
Gold Label: 0	Hasil Prediksi: 0	"Tidak hanya tarif tapi apa saja harus diatur.
Gold Label: 0	Hasil Prediksi: 0	Dipanggil keduanya untuk berbicara masing-masing, musyawarah, duduk bareng kemudian dibuat atur
Gold Label: 0	Hasil Prediksi: 0	Kan bisa saling melengkapi, negara lain bisa masa kita enggak bisa," ucap Zulkifli di Gedung DP
Gold Label: 0	Hasil Prediksi: 0	Baca juga: Setya Novanto: Jangan Sampai Kisruh Taksi dan Ojek Online Jadi Besar
Gold Label: 0	Hasil Prediksi: 0	Ketua Umum PAN menambahkan bahwa hal ini harus diatur karena menyangkut mata pencaharian dari m
Gold Label: 0	Hasil Prediksi: 0	Menurut dia, titik temu masih bisa dicari.
Gold Label: 0	Hasil Prediksi: 0	Sebab, transportasi online tradisional masih memiliki pasar sendiri dan transportasi tradisional
Gold Label: 0	Hasil Prediksi: 0	"Yang tradisional tentu akan ada pasar sendiri, yang online punya pasar sendiri.
Gold Label: 0	Hasil Prediksi: 0	Nah itulah segera ditata kalau tidak tentu terjadi bentrokan-bentrokan.
Gold Label: 0	Hasil Prediksi: 0	Itu namanya urusan perut, urusan rezeki tentu kalau tidak ditata dengan baik akan menimbulkan k
Gold Label: 0	Hasil Prediksi: 0	Zulkifli menilai bahwa tidak mungkin bila keduanya bertarung bebas.
Gold Label: 0	Hasil Prediksi: 0	Diperlukan adanya aturan yang membahas hal tersebut.
Gold Label: 0	Hasil Prediksi: 0	"Itu mesti diatur kan kalau semua tarung bebas kan enggak mungkin.
Gold Label: 0	Hasil Prediksi: 0	Kalau semua dibiarkan tarung bebas atau disebut pasar bebas itu tentu nanti yang tradisional ak
Gold Label: 0	Hasil Prediksi: 0	Baca juga: Sebelum 1 April Sudah Ada Ketentuan Tarif Taksi Online di Jakarta.
Gold Label: 1	Hasil Prediksi: 1	Suara.com - Cerita sekuel terbaru James Bond bocor.
Gold Label: 1	Hasil Prediksi: 1	Menurut sumber yang terlibat dalam produksi film ini, agen rahasia 007 berhenti menjadi mata-ma
Gold Label: 0	Hasil Prediksi: 0	"Bond berhenti menjadi agen rahasia karena jatuh cinta dan menikah dengan perempuan yang dicint
Gold Label: 0	Hasil Prediksi: 0	Dalam film tersebut, Bond diduga menikahi Madeleine Swann yang diperankan oleh Lea Seydoux.
Gold Label: 0	Hasil Prediksi: 0	Lea diketahui bermain sebagai gadis Bond di sekuel Spectre pada 2015 silam.
Gold Label: 1	Hasil Prediksi: 1	Jika benar, ini merupakan satu-satunya sekuel yang bercerita pernikahan James Bond sejak 1969.
Gold Label: 1	Hasil Prediksi: 1	Sebelumnya, di sekuel On Her Majesty, James Bond menikahi Tracy Draco yang diperankan Diana Rig
Gold Label: 1	Hasil Prediksi: 0	Namun di film itu Draco terbunuh.

Gambar 3. Sampel Perhitungan Prediksi dengan Jaringan Syaraf Tiruan Propagasi Balik

f. Pengurutan Data Berdasarkan Urutan Posisi Kalimat

Proses pengurutan data berdasarkan urutan posisi kalimat ini bertujuan untuk mengurutkan kalimat dengan nilai skor akhir yang didapat dari masing-masing kalimat pada dataset Indosum secara *descending* (besar ke kecil). Sebagai contoh, dapat dilihat pada Gambar 4 dibawah ini penulis akan menggunakan beberapa sampel untuk menampilkan penerapan pengurutan data berdasarkan nilai skor akhir yang didapat pada beberapa kalimat:

Gold Label: 1	Hasil Prediksi: [0.92071449]	Ketua MPR Zulkifli Hasan menyesalkan kisruh yang terjadi antara pelaku sar
Gold Label: 1	Hasil Prediksi: [0.94155659]	Zulkifli menyarankan adanya pertemuan bersama antara pemerintah, pelaku tr
Gold Label: 1	Hasil Prediksi: [0.74803845]	Zulkifli menyarankan adanya pertemuan bersama antara pemerintah, pelaku tr
Gold Label: 1	Hasil Prediksi: [0.92940558]	Suara.com - Cerita sekuel terbaru James Bond bocor.
Gold Label: 1	Hasil Prediksi: [0.871451]	Menurut sumber yang terlibat dalam produksi film ini, agen rahasia 007 berhe
Gold Label: 1	Hasil Prediksi: [0.84531485]	Jika benar, ini merupakan satu-satunya sekuel yang bercerita pernikahan Ja
Gold Label: 1	Hasil Prediksi: [0.90027468]	Sebelumnya, di sekuel On Her Majesty, James Bond menikahi Tracy Draco yang
Gold Label: 1	Hasil Prediksi: [0.82765202]	Menteri Pertanian Andi Amran Sulaiman mengatakan bahwa saat ini pemerintah
Gold Label: 1	Hasil Prediksi: [0.88243471]	Hal tersebut disampaikannya saat melepas ekspor perdana produk daging ayan
Gold Label: 1	Hasil Prediksi: [0.58232967]	Charoen Pokphand Indonesia ke Papua New Guinea, di Kawasan Industri Moderr
Gold Label: 1	Hasil Prediksi: [0.73489698]	Kata Amran, produksi ayam ras nasional di Indonesia saat ini mengalami sur
Gold Label: 1	Hasil Prediksi: [1.07141915]	SPANYOL \u2013 Barcelona berhasil memboyong kemenangan besar di laga Clasi
Gold Label: 1	Hasil Prediksi: [0.95175652]	Gol yang diciptakan di laga yang digelar di kandang Madrid, Santiago Berr
Gold Label: 1	Hasil Prediksi: [0.63888139]	Kemenangan ini semakin membuat Barcelona semakin jauh di puncak klasemen c
Gold Label: 1	Hasil Prediksi: [1.18134991]	Wamena (ANTARA News) - Pemerintah Kabupaten Jayawijaya, Papua, mendatangka
Gold Label: 1	Hasil Prediksi: [0.8937741]	Kepala Dinas Kebudayaan dan Pariwisata (Disbudpar) Jayawijaya Alpius Wetipic

Gambar 4. Sampel Pengurutan Data Berdasarkan Urutan Posisi Kalimat

g. Pembuatan Hasil Ringkasan Teks

Pada proses terakhir ini bertujuan untuk mengambil dan menyusun beberapa kalimat dengan nilai skor tertinggi pada dataset Indosum menjadi ringkasan teks yang baru. Sebagai contoh, dapat dilihat pada Gambar 5 dibawah ini penulis akan menggunakan contoh dokumen untuk melakukan pembuatan ringkasan teks dari beberapa kalimat dengan nilai skor tertinggi:

```

=====
=====HASIL RINGKASAN TEKS=====
=====

Kalimat ke 1:

SPANYOL \u2013 Barcelona berhasil memboyong kemenangan besar di laga Clasico 2017 dengan skor 3 - 0 atas Real Madrid, Sabtu ( 23 / 12 / 2017 ).

Kalimat ke 2:

Gol yang diciptakan di laga yang digelar di kandang Madrid, Santiago Bernabeu, \u00a0diciptakan Luis Suarez, Lionel Messi dan Aleix Vidal di babak kedua.

Kalimat ke 3:

Kemenangan atas Madrid ini membuat Barcelona mencatat tak pernah mengalami kekalahan sepanjang paruh musim ini.

Kalimat ke 4:

Kekalahan terakhir Barcelona terjadi di bulan Agustus lalu, saat Barcelona sebagai juara Copa del Rey menghadapi Real Madrid, juara La Liga musim 2016

Kalimat ke 5:

Kala itu, Barcelona ditaklukkan 2 - 0 oleh Real Madrid.

```

Gambar 5. Contoh Hasil Ringkasan Teks

3.2 Pengujian

Metode pengujian ini menggunakan metode pengujian perhitungan hasil tingkat Presisi, *Recall*, *F-Measure*, dan Akurasi. Tujuan dari pengujian ini adalah untuk menentukan seberapa tinggi ketepatan hasil ringkasan teks dari Text Summarization yang telah dibuat dengan Gold Standard yang sudah ada sebagai acuan standard tingkat ketepatan hasil. Pada Tabel 3 menampilkan hasil perhitungan pengujian pada data latih dan pada Tabel 4 menampilkan hasil perhitungan pengujian pada data uji.

$$Precision = (TP)/(TP + FP) \quad (8)$$

$$Recall = (TP)/(TP + FN) \quad (9)$$

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$F - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (11)$$

Tabel 3. Hasil Pengujian *Precision*, *Recall* dan *F-Measure* Pada Data Latih

Data Latih								
Doc	TP	TN	FP	FN	precision	recall	Akurasi	f-measure
1	3	15	0	0	1	1	1	1
2	4	5	0	1	1	0,8	0,9	0,888889
3	4	19	0	0	1	1	1	1
4	3	39	0	0	1	1	1	1
5	2	1	0	0	1	1	1	1
6	1	15	0	1	1	0,5	0,941176	0,666667
7	3	14	1	0	0,75	1	0,944444	0,857143
8	4	12	0	0	1	1	1	1
9	4	18	0	0	1	1	1	1
10	3	14	0	2	1	0,6	0,894737	0,75
11	3	28	0	0	1	1	1	1
12	3	10	0	0	1	1	1	1
13	1	13	0	1	1	0,5	0,933333	0,666667
14	3	16	2	0	0,6	1	0,904762	0,75
15	3	10	0	1	1	0,75	0,928571	0,857143
Total	44	229	3	6	0,93617	0,88	0,963135	0,907216

Tabel 4. Hasil Pengujian *Precision*, *Recall* dan *F-Measure* Pada Data Uji

Data Uji								
Doc	TP	TN	FP	FN	precision	recall	Akurasi	f-measure
1	1	6	0	2	1	0,333333	0,78	0,5
2	2	10	0	4	1	0,333333	0,75	0,5
3	2	11	1	0	0,666667	1	0,93	0,8
4	2	6	0	3	1	0,4	0,73	0,571429
5	1	2	0	4	1	0,2	0,43	0,333333
Total	8	35	1	13	0,888889	0,380952	0,72	0,533333

Berdasarkan hasil pengukuran pada presisi, recall, akurasi dan f-measure didapatkan analisis yaitu akurasi yang didapat menggunakan data uji memperlihatkan bahwa dari 5 dokumen yang diuji didapatkan ketepatan prediksi sebesar 72% dari total keseluruhan label prediksi. Presisi yang didapat sebesar 88.89% menunjukkan tingginya ketepatan hasil memprediksi jumlah label “benar” terhadap jumlah label “benar” sesungguhnya dari keseluruhan data uji. Hasil dari recall sebesar 38.1% pada data uji menunjukkan cukup rendahnya ketepatan posisi label “benar” yang sesungguhnya diprediksi secara tepat. Namun pada perhitungan f-measure didapatkan hasil sebesar 53.3% yang dimana hasil ini menunjukkan seberapa besar ketepatan rata-rata antara hubungan hasil dari presisi dan recall. Hasil ini dapat diartikan bahwa model dapat memprediksi jumlah keseluruhan label secara benar namun kurang baik dalam memprediksi mana yang harus berlabel “benar” dan mana yang harus berlabel “salah”.

3.3 Evaluasi

Evaluasi adalah tahap terakhir dalam pengujian system aplikasi untuk mengetahui apa saja kelebihan dan kekurangan yang terdapat pada sistem aplikasi yang telah dibuat. Setelah melakukan pengujian, didapatkan beberapa kelebihan dan kekurangan pada program aplikasi yang akan dijelaskan dibawah ini:

a. Kelebihan Sistem Aplikasi *Text Summarization*

Sistem aplikasi ini memiliki beberapa kelebihan, yaitu:

1. Memiliki tingkat ketepatan yang tinggi jika dibandingkan antara kalimat yang sudah diprediksi akan masuk sebagai hasil ringkasan teks dengan hasil ringkasan teks yang sudah dijadikan gold standard.
2. Hasil yang didapat dari perhitungan ketepatan prediksi diartikan bahwa model dapat memprediksi jumlah keseluruhan label secara benar namun kurang baik dalam memprediksi mana yang harus berlabel “benar” dan mana yang harus berlabel “salah”.

b. Kekurangan Sistem Aplikasi Text Summarization

Selain memiliki kelebihan, sistem aplikasi ini juga memiliki beberapa kekurangan, yaitu:

1. Sistem hanya bisa menerima masukkan satu dokumen saja.
2. Sistem tidak membentuk kalimat baru untuk setiap kalimat yang ada di dalam ringkasan teks.

4. PENUTUP

Berdasarkan hasil evaluasi dari penelitian, setelah menggunakan dataset IndoSum yang telah ditransformasikan menjadi bentuk text didapatkan hasil pengukuran ketepatan dengan rata-rata sebesar 0.889 untuk Presisi, 0.381 untuk Recall, 0.533 untuk F-measure dan untuk akurasi sebesar 0.72. Hasil yang didapat dari perhitungan ketepatan prediksi diartikan bahwa model dapat memprediksi jumlah keseluruhan label secara benar namun kurang baik dalam memprediksi mana yang harus berlabel “benar” dan mana yang harus berlabel “salah”. Pada penelitian di masa depan dapat dilakukan penerapan Text Summarization menggunakan basis Fitur dan metode Jaringan Syaraf Tiruan Propagasi Balik menggunakan pendekatan abstraktif dengan masukan multi dokumen.

DAFTAR PUSTAKA

- [1] D. W. Brata and A. Hetami, “Perancangan Information Retrieval (IR) Untuk Pencarian Ide Pokok Teks Artikel Berbahasa Inggris Dengan Pembobotan Vector Space Model,” *J. Ilm. Teknol. Inf. Asia*, vol. 9, no. 1, pp. 53–59, 2015.
- [2] Waseemullah *et al.*, “A Novel Approach for Semantic Extractive Text Summarization,” *MDPL*, pp. 1–14, 2022.
- [3] D. Delvin, D. Arisandi, and T. Sutrisno, “APLIKASI PERINGKASAN DOKUMEN MENGGUNAKAN METODE MAXIMUM MARGINAL RELEVANCE (MMR),” *J. Ilmu Komput. dan Sist. Inf.*, vol. 10, no. 1, pp. 62–68, Mar. 2022.
- [4] P. M. Sabuna and D. B. Setyohadi, “Summarizing Indonesian text automatically by using sentence scoring and decision tree,” *Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017*, vol. 2018-Janua, no. February, pp. 1–6, 2018.
- [5] L. Lebanoff, K. Song, and F. Liu, “Adapting the neural encoder-decoder framework from single to multi-document summarization,” *Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 4131–4141, 2018.
- [6] C. Slamet, A. R. Atmadja, D. S. Maylawati, R. S. Lestari, W. Darmalaksana, and M. A. Ramdhani, “Automated Text Summarization for Indonesian Article Using Vector Space Model,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 288, no. 1, 2018.
- [7] L. Suanmali, N. Salim, and M. S. Binwahlan, “Fuzzy Logic Based Method for Improving Text Summarization,” vol. 2, no. 1, 2009.
- [8] D. P. Ismi and F. Ardianto, “Peringkasan Ekstraktif Teks Bahasa Indonesia dengan Pendekatan Unsupervised Menggunakan Metode Clustering,” *Cybernetics*, vol. 3, no. 02, p. 90, 2020.
- [9] A. J. C. Trappey and C. V. Trappey, “An R and D knowledge management method for patent document summarization,” *Ind. Manag. Data Syst.*, vol. 108, no. 2, pp. 245–257, 2008.
- [10] A. Hernández-castañeda, R. A. García-hernández, Y. Ledeneva, and C. E. Millán-hernández, “Extractive Automatic Text Summarization Based on Lexical-Semantic Keywords,” *IEEE Access*, vol. 8, p. <https://creativecommons.org/licenses/by/4.0/>, 2020.
- [11] A. D. Dongare, R. R. Kharde, and A. D. Kachare, “Introduction to Artificial Neural Network,” 2008.
- [12] P. Bhasin, “Back Propagation Algorithm: An Artificial Neural Network Approach.”