Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 4, Nomer: 2, Desember 2023

DETEKSI KATA SERAPAN TERHADAP DOKUMEN MENGGUNAKAN PENDEKATAN DEEP LEARNING

Windi Halimardani [1], Edy Rahman Syahputra [2], Husni Lubis[3]

[1] *Universitas Harapan Medan*
Jl. H.M. Jhoni No.70 C

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sebuah metode deteksi kata serapan dalam dokumen teks menggunakan pendekatan Deep Learning. Kata serapan adalah kata-kata yang berasal dari bahasa asing dan telah diadopsi ke dalam bahasa lokal. Metode ini memiliki potensi untuk mengidentifikasi kata serapan dengan akurasi yang tinggi, bahkan dalam konteks dokumen yang besar dan beragam. Pendekatan Deep Learning akan digunakan dalam analisis teks untuk mengambil fitur-fitur yang relevan dan kompleks dari kata-kata dalam dokumen. Model Deep Learning yang akan dibangun dapat memahami konteks penggunaan kata serapan dalam bahasa lokal, serta dapat membedakannya dari kata-kata asli bahasa tersebut. Selain memberikan solusi untuk tugas deteksi kata serapan, penelitian ini juga akan menggali potensi penerapan Deep Learning dalam pemrosesan teks dan linguistik komputasional. Hasil dari penelitian ini diharapkan dapat membantu dalam memahami lebih baik aspek-aspek bahasa yang berkaitan dengan kata serapan, serta dapat berguna dalam aplikasi yang berkaitan dengan analisis teks seperti terjemahan otomatis, analisis sentimen, dan banyak lagi.

Kata Kunci: Serapan, Deteksi, Deep learning

ABSTRACT

This research aims to develop a method for detecting loanwords in text documents using a Deep Learning approach. Loan words are words that originate from a foreign language and have been adopted into the local language. This method has the potential to identify loanwords with high accuracy, even in the context of large and diverse documents. Deep Learning approaches will be used in text analysis to extract relevant and complex features from words in documents. The Deep Learning model that will be built can understand the context of the use of loan words in local languages and can differentiate them from native words in that language. Apart from providing a solution to the task of loan word detection, this research will also explore the potential application of Deep Learning in text processing and computational linguistics. It is hoped that the results of this research will help in better understanding aspects of language related to loanwords, and can be useful in applications related to text analysis such as automatic translation, sentiment analysis, and many more

Keywords: Absorption, Detection, Deep learning

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 4, Nomer: 2, Desember 2023

I. PENDAHULUAN

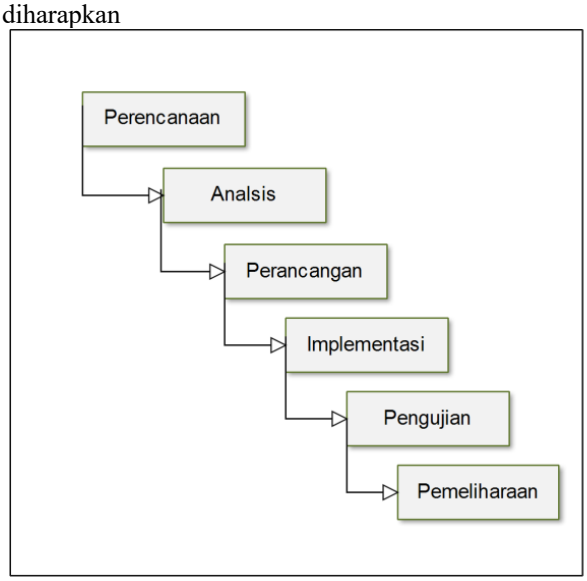
Deteksi kata serapan dalam dokumen menggunakan pendekatan deep learning adalah topik yang menarik dalam bidang pemrosesan bahasa alami (natural language processing)(He et al., 2023). Dalam beberapa tahun terakhir, deep learning telah menjadi metode yang sangat populer untuk memahami dan memproses teks dengan akurasi yang tinggi (Nurhikmat, 2018). Serapan merupakan proses penyerapan kata asing yang digunakan karena memiliki makna sama dalam Bahasa Indonesia. Namun kata tersebut telah mengalami perubahan dalam ejaan, pengucapan, dan penulisan sesuai kaidah Bahasa Indonesia [1]. Namun, dalam dokumen tertulis, deteksi serapan dapat menjadi lebih mudah dengan menggunakan teknik-teknik deep learning[2].

Deep learning adalah sub-bidang dari machine learning yang menggunakan jaringan saraf tiruan (artificial neural networks) yang memiliki banyak lapisan (layers) untuk mempelajari pola-pola yang kompleks dalam data. Dalam konteks deteksi kata serapan, deep learning dapat membantu dalam mempelajari representasi kata-kata yang mencerminkan serapan, sehingga dapat mengidentifikasi kata-kata yang digunakan secara ironis. Untuk melatih model deep learning dalam deteksi kata serapan, diperlukan kumpulan data yang besar yang berisi dokumen-dokumen dengan kata serapa [3]. Proses pelatihan melibatkan memasukkan teks dokumen ke dalam model deep learning dan mengoptimasi parameter-parameter model untuk menghasilkan prediksi yang akurat. Semakin besar dan representatif kumpulan data pelatihan, semakin baik kinerja model dalam deteksi serapan(Syarifuddin, 2020). Salah satu pendekatan deep learning yang dapat digunakan dalam deteksi kata serapan adalah jaringan saraf konvolusional (convolutional neural network) [4]. CNN memiliki kemampuan untuk mempelajari fitur-fitur lokal dalam teks, seperti kombinasi kata-kata yang sering muncul dalam konteks serapan. Dengan melatih CNN pada kumpulan data yang cukup besar [5].

Selain itu, ada beberapa faktor yang perlu diperhatikan dalam deteksi kata serapan menggunakan deep learning. Pertama, kualitas dan representativitas kumpulan data pelatihan sangat penting pendekatan. Kumpulan data harus mencakup variasi yang cukup dalam gaya bahasa, topik, dan jenis dokumen.

II.METODE PENELITIAN

Model pengembangan sistem yang digunakan penulis dalam penelitian ini yaitu menggunakan model SDLC (*System Development Life Cyle*) pengembangan atau rekayasa sistem informasi (*software engineering*). SDLC digunakan untuk membangun suatu sistem informasi agar dapat berjalan sesuai dengan apa yang diharapkan



Gambar 1. Kerangka Kerja Pengembangan Sistem Informasi (SDLC)

Keterangan:

1. Perencanaan
   Tahap awal dari pengembangan sistem, tahap ini bertujuan untuk mengidentifikasi sistem yang akan dikembangkan,
2. Analisis
   Kebutuhan perangkat lunak tahap analisis dilakukan secara intensif untuk menspesifikasikan kebutuhan sistem agar dapat di pahami sistem seperti apa yang dibutuhkan user
3. Perancangan
   Proses yang fokus pada perancangan program termasuk struktur data, arsitektur sistem, dan representasi antarmuka. Tahap ini mentranlasi kebutuhan sistem dri tahap analisis kebutuhan ke representasi perancangan agar dapat diimplementasikan menjadi program pada tahap selanjutnya.
4. Implementasi
   Pada implementasi sistem akan di tampilkan hasil dari perancangan antar muka yang ada pada penelitian ini untuk dapat digunakan oleh pengguna
5. Pengujian
   Tahap pengujian fokus pada sistem dari segi logika dan fungsional dan memastikan bahwa semua bagian sudah diuji. Hal ini dilakukan untuk meminimalisis kesalahan dan memastikan keluaran yang dihasilkan sesuai dengan yang diinginkan.

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 4, Nomer: 2, Desember 2023

III. HASIL DAN PEMBAHASAN

1. Analisa data
   Analisis data merupakan langkah penting dalam mengembangkan sistem deteksi kata serapan berbasis deep learning. Langkah-langkah ini bertujuan untuk memahami sifat data, mengeksplorasi pola, dan mengidentifikasi cara terbaik untuk mempersiapkan data sebelum digunakan untuk melatih model. Berikut ini data kata serapan yang akan digunakan yang sterdapat pada tabel 1.

Tabel 1. Data kata serapan

No	Istilah Asing	Bahasa Indonesia
1	Account	Akun
2	Actor	Aktor
3	Admin	Administrator
4	Analysis	Analisis
5	Application	Aplikasi
6	Artifact	Artifak
7	Attachment	Lampiran
8	Bandwidth	Lebar Pita
9	Basis Path	Jalur Dasar
10	Blackbox	Kotak Hitam

2. Perhitungan algoritma
   Berikut ini perhitungan algoritma dalam melakukan deteksi kata serapan.

- Data

Dengan melihat data dari file excel yang diberikan, maka dapat disimpulkan bahwa yang mempengaruhi deteksi kata serapan yaitu label kata serapan. Pada penelitian ini akan menggunakann 5 data pertama yang akan dihitung yang telah dibuat untuk melakukan validasi terkait apakah data kalimat tersebut mengandung kata serapan. Berikut merupakan tampilan datanya pada tabel 2 :

Tabel 2. Dataset

Opini	Kata serapan	Kata tidak serapan	Kata netral	Cluster
1	2	4	3	C1= serapan
2	2	3	1	C2= tidak serapan
3	1	4	2	C3= Netral

Perhitungan Jarak Data Dengan Centroid Kemudian akan dihitung jarak dari setiap data ke setiap pusat cluster yang ada dengan rumus Euclidean distance. Sehingga ditemukan jarak terdekat dari seiap data ke centroid.

- Data gopay
  Akan dilakukan perubahan kata dari bentuk teks menjadi numerik seperti pada gambar berikut ini

Tabel 3. Data Opini

Opini	Kata serapan	Kata tidak serapan	Kata netral
1	4	2	1
2	2	2	1
3	1	1	1
4	2	3	2
5	4	2	2

Perhitungan jarak data dengan centroid menggunakan rumus Euclidean distance:

$$d(ai,bj) = \sqrt{\sum(ai - bj)^2}$$

dimana:
ai: data kriteria,
bj: centroid pada cluster ke-j

Jarak data opini 1 dengan centroid 1:

$d(1,1) = \sqrt{(4-2)^2 + (2-4)^2 + (1-3)^2} = 3.4$

$d(2,1) = \sqrt{(2-2)^2 + (2-4)^2 + (1-3)^2} = 2.8$

$d(3,1) = \sqrt{(1-2)^2 + (1-4)^2 + (1-3)^2} = 3,7$

$d(4,1) = \sqrt{(2-2)^2 + (3-4)^2 + (2-3)^2} = 1.4$

$d(5,1) = \sqrt{(4-2)^2 + (2-4)^2 + (2-3)^2} = 3$

Jarak data opini 1 dengan centroid 2 :

$d(1,2) = \sqrt{(4-2)^2 + (2-3)^2 + (1-1)^2} = 2,2$

$d(2,2) = \sqrt{(2-2)^2 + (2-3)^2 + (1-1)^2} = 1$

$d(3,2) = \sqrt{(1-2)^2 + (1-3)^2 + (1-1)^2} = 2.2$

$d(4,2) = \sqrt{(2-2)^2 + (3-3)^2 + (2-1)^2} = 1$

$d(5,2) = \sqrt{(4-2)^2 + (2-3)^2 + (2-1)^2} = 2.4$

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 4, Nomer: 2, Desember 2023

Jarak data opini 1 dengan centroid 3 :

d(1,3) = $\sqrt{(4-1)^2 + (2-4)^2 + (1-2)^2} = 3.7$

d(2,3) = $\sqrt{(2-1)^2 + (2-4)^2 + (1-2)^2} = 2.4$

d(3,3) = $\sqrt{(1-1)^2 + (1-4)^2 + (1-2)^2} =$ 3.1

d(4,3) = $\sqrt{(2-1)^2 + (3-4)^2 + (2-2)^2} =$ 1,3

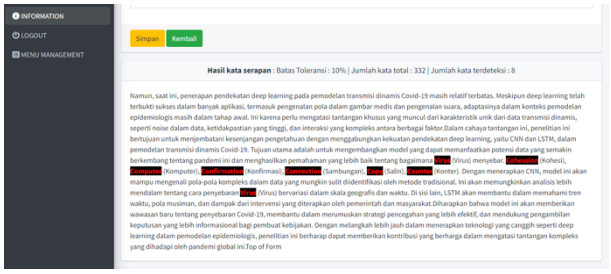
d(5,3) = $\sqrt{(4-1)^2 + (2-4)^2 + (2-2)^2} =$ 3,6

kemudian dilakukan perhitungan sebagai berikut

Tabel 4.12 Data hasil perhitungan

No	Kata serapan	Kata tidak serapan	Kata netral	C1	C2	C3	Cluster
1	4	2	1	3.4	2.1	3.7	C1
2	2	2	1	2.8	1	2.4	C1
3	1	1	1	3.7	2.2	3.1	C1
4	2	3	2	1.4	1	1.3	C1
5	4	2	2	3	2.4	3.6	C3

Berdasarkan hasil tersebut bahwa dalam perhitungan cluster pada data 5 opini menghasilkan label kata serapan berjumlah 4 data dan label netral 1 data sehingga dapat disimpulan. Selanjutnya di uji melalui website yang telah di rancang di halaman deteksi kata serapan bertujuan melakukan deteksi kata serapan berdasarkan data dokumen yang diupload, proses deteksi kata serapan akan menggunakan algoritma CNN yang dapat dilihat pada gambar 4.7 berikut ini.



Gambar 2. Halaman menu deteksi kata serapan

IV. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah penulis lakukan maka dapat disimpulkan bahwa :

1. Serapan adalah kata yang berasal dari bahasa asing dan telah diadopsi ke dalam bahasa lokal.
2. Penelitian ini menghasilkan sistem untuk mengidentifikasi kata-kata serapan dalam suatu konteks tertentu yang terdapat pada dokumen dengan format doc
3. Penelitian ini memiliki implikasi praktis dalam bidang pemrosesan bahasa alami dan linguistik komputasional. Deteksi kata serapan adalah aspek penting dalam memahami dan menganalisis bahasa dalam konteks yang lebih luas.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah membantu penelitian ini sampai selesai.

REFERENSI

[1] S. R. Dewi, “Deep Learning Object Detection Pada Video Menggunakan Tensorflow Dan Convolutional Neural Network,” 2018.

[2] A. I. S. Azis, V. Suhartono, and H. Himawan, “Model Multi-Class SVM menggunakan strategi 1V1 Untuk Klasifikasi Wall-Following Robot Navigation Data,” *J. Cyberku*, vol. 13, no. 2, p. 8, 2017.

[3] A. K. N. Bany, “Analisis Sentimen Dan Deteksi Emosi Dengan Pendekatan Lexicon pada Judul Berita Media Online Mengenai Covid-19 di Indonesia.” Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.

[4] M. A. Rahman, H. Budianto, and E. I. Setiawan, “Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network,” *INSYST J. Intell. Syst. Comput.*, vol. 1, no. 2, pp. 50–57, 2019.

[5] F. A. Wijaya, “Studi empiris analisis sentimen kenaikan cukai rokok Indonesia di Tahun 2019 menggunakan data twitter dan ensemble learning.” Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah ….