

Pemanfaatan Natural Language Processing Untuk Pengecekan Ejaan Sesuai KBBI

Adventus Michael Bala Ledjap¹, Febriyan Putri Rochmawati², Dea Ayu Eka Marsanda³,
Anggraini Puspita Sari*

^{1,2,3,4} Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya
Email : 22081010084@student.upnjatim.ac.id, 22081010074@student.upnjatim.ac.id,
22081010336@student.upnjatim.ac.id, anggraini.puspita.if@upnjatim.ac.id
Corresponding author : anggraini.puspita.if@upnjatim.ac.id

Abstract

Spelling mistakes in the Indonesian language often occur due to lack of knowledge of spelling rules, the influence of foreign languages, and the use of informal language. This research aims to develop a Natural Language Processing (NLP) based spelling checking system that refers to the Dictionary of Indonesian Language (KBBI). The system is expected to be able to perform automatic spelling verification and correction with high accuracy. The method used is qualitative with an analytic descriptive approach, involving data collection from the KBBI and processing using Python. The research stages include collection, data cleaning, tokenization, and data analysis. The system was trained using a large Indonesian text dataset. The results showed the system achieved a spelling check accuracy of 81.64%, more accurate, easy to use, flexible, and adaptive than conventional systems. The number of words in the text affects the checking time, with longer text taking longer.

Keywords: *Natural Language Processing (NLP), Kamus Besar Bahasa Indonesia (KBBI), Text Analysis.*

Abstrak

Kesalahan ejaan dalam bahasa Indonesia sering terjadi akibat kurangnya pengetahuan aturan ejaan, pengaruh bahasa asing, dan penggunaan bahasa informal. Penelitian ini bertujuan mengembangkan sistem pengecekan ejaan berbasis Natural Language Processing (NLP) yang mengacu pada Kamus Besar Bahasa Indonesia (KBBI). Sistem diharapkan mampu melakukan verifikasi dan koreksi ejaan otomatis dengan akurasi tinggi. Metode yang digunakan adalah kualitatif dengan pendekatan deskriptif analitik, melibatkan pengumpulan data dari KBBI dan pemrosesan menggunakan Python. Tahapan penelitian mencakup pengumpulan, pembersihan data, tokenisasi, dan analisis data. Sistem dilatih menggunakan dataset teks bahasa Indonesia yang besar. Hasil penelitian menunjukkan sistem mencapai akurasi pengecekan ejaan sebesar 81,64%, lebih akurat, mudah digunakan, fleksibel, dan adaptif dibandingkan sistem konvensional. Jumlah kata dalam teks mempengaruhi waktu pengecekan, dengan teks yang lebih panjang membutuhkan waktu lebih lama.

Kata Kunci: *Natural Language Processing (NLP), Kamus Besar Bahasa Indonesia (KBBI), Analisis Teks.*

PENDAHULUAN

Penggunaan bahasa yang baik dan benar menjadi salah satu kunci untuk mencapai komunikasi yang efektif dan efisien. Melalui bahasa, kita dapat memahami budaya dan pola pikir suatu masyarakat. Menurut Keraf (1997: 1), bahasa adalah alat komunikasi antara anggota masyarakat yang berupa simbol bunyi yang dihasilkan oleh alat ucap manusia. Namun, dalam praktiknya, masih banyak orang yang melakukan kesalahan ejaan dalam menulis bahasa Indonesia. Hal ini dapat disebabkan oleh berbagai faktor, seperti kurangnya pengetahuan tentang aturan ejaan, kurangnya latihan, atau pengaruh bahasa asing. Kesalahan ejaan dapat berakibat fatal, terutama dalam dokumen resmi atau publikasi ilmiah, kesalahan ejaan juga dapat membuat informasi menjadi sulit dipahami, bahkan dapat mengubah makna kalimat.

Variasi pada Bahasa Indonesia memiliki banyak kata yang memiliki bunyi dan penulisan yang mirip, yang dapat menyebabkan kesalahan ejaan. Pengaruh bahasa asing dalam bahasa Indonesia juga dapat menyebabkan kesalahan ejaan, terutama dalam penggunaan kata-kata serapan. Menurut Firmansyah (2015), Bahasa Indonesia adalah salah satu mata pelajaran yang diajarkan sejak jenjang pendidikan dasar dan sangat diperlukan dalam kehidupan sehari-hari. Masalah umum yang sering ditemui pada tahap preprocessing teks adalah penggunaan bahasa informal dan kesalahan pengetikan pada teks. Masyarakat kadang-kadang menggunakan kata yang tidak baku atau bahasa gaul dalam opini mereka. Selain itu, kesalahan ejaan atau yang biasa disebut typo juga sering muncul dalam kalimat. Jika kata-kata tersebut tetap diproses, analisis yang dihasilkan akan menjadi tidak akurat.

Kesenjangan ini dapat diatasi dengan berbagai cara, salah satunya dengan memanfaatkan teknologi pemrosesan bahasa

alami (NLP). NLP adalah bidang ilmu komputer yang mempelajari interaksi antara komputer dan bahasa manusia, tujuan dari Natural Language Processing (NLP) adalah merancang serta membangun perangkat lunak yang mampu menganalisis, dan menghasilkan bahasa manusia secara alami. Oleh karena itu, saat berinteraksi dengan komputer, pengguna merasa seperti sedang berkomunikasi dengan manusia lainnya (Mishra & Kaur, 2013). Terdapat banyak teori dan penelitian mengenai pengecekan ejaan dengan NLP, salah satu metode yang populer adalah dengan menggunakan model statistik. Model statistik ini dilatih dengan menggunakan data teks yang besar, dan kemudian digunakan untuk memprediksi ejaan yang benar untuk kata-kata baru. Metode lain yang juga populer adalah dengan menggunakan kamus elektronik. Kamus elektronik ini berisi daftar kata-kata yang benar dan ejaannya. NLP dapat digunakan untuk mencari kata-kata dalam kamus elektronik dan untuk mencocokkannya dengan kata-kata yang digunakan dalam teks.

Penelitian tentang pemanfaatan NLP untuk pengecekan ejaan sesuai KBBI memiliki nilai baru yang signifikan. Penelitian ini dapat membantu meningkatkan akses informasi bagi masyarakat luas, karena dengan ejaan yang benar, informasi akan lebih mudah dipahami. Selain itu, penelitian ini juga dapat membantu mengembangkan aplikasi dan perangkat lunak yang dapat membantu orang-orang untuk menggunakan bahasa Indonesia dengan baik dan benar. Manusia secara alami mampu memahami sintaks linguistik, tata bahasa, dan hubungan spasial yang tersirat, namun komputer menghadapi tantangan besar dalam memproses permintaan dalam bahasa alami (Allen, 1995). Pemanfaatan NLP untuk pengecekan ejaan sesuai KBBI memiliki potensi yang besar untuk meningkatkan kualitas bahasa Indonesia yang digunakan

dalam berbagai aspek kehidupan. Penelitian tentang hal ini masih terus berkembang, dan diharapkan dapat menghasilkan inovasi-

inovasi baru yang bermanfaat bagi masyarakat luas.

METODE

Penelitian ini menggunakan metode kualitatif dengan pendekatan deskriptif analitik. Pendekatan ini dipilih untuk mengembangkan metode pengecekan ejaan bahasa Indonesia berbasis Natural Language Processing (NLP) yang mengacu pada Kamus Besar Bahasa Indonesia (KBBI). Melalui pendekatan ini, diharapkan dapat dihasilkan sebuah sistem yang mampu melakukan verifikasi dan koreksi ejaan secara otomatis dengan tingkat akurasi yang tinggi. Implementasi NLP dalam penelitian ini akan mencakup analisis dan pemrosesan teks untuk mengidentifikasi kesalahan ejaan, serta memberikan saran perbaikan yang sesuai berdasarkan aturan ejaan yang tercantum dalam KBBI.

1. Bahasa pemrograman

Python merupakan bahasa pemrograman serbaguna yang dirancang dengan tujuan untuk menghasilkan kode yang mudah dibaca. Python dikenal sebagai bahasa yang memadukan kapabilitas dan kinerja dengan sintaksis yang sangat jelas, serta dilengkapi dengan pustaka standar yang luas dan lengkap. Penelitian ini menggunakan bahasa pemrograman Python sebagai alat utama untuk mengembangkan algoritma dan model yang diperlukan.

2. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menganalisis mendalam terhadap kamus elektronik, khususnya Kamus Besar Bahasa Indonesia (KBBI). Proses ini melibatkan ekstraksi data ejaan, makna kata, dan penggunaan kata yang relevan dari KBBI. Data yang diperoleh sangat penting dalam membangun sistem pengecekan ejaan yang akurat dan sesuai

dengan standar bahasa Indonesia. Melalui ekstraksi dan analisis data dari KBBI, penelitian ini bertujuan untuk mengembangkan algoritma yang mampu mendeteksi dan memperbaiki kesalahan ejaan secara efektif, memastikan kesesuaian dengan aturan dan konvensi yang ditetapkan dalam bahasa Indonesia.

3. Pra-pemrosesan Data Teks

Pra-pemrosesan data teks melibatkan beberapa tahap penting untuk mempersiapkan data sebelum digunakan dalam pemodelan bahasa :

- a. Pengumpulan dan Pembersihan Data
Data teks dikumpulkan dari berbagai sumber dan dibersihkan untuk menghapus elemen yang tidak diperlukan.
- b. Tokenisasi
Metode ini melibatkan pemecahan teks menjadi unit-unit kecil yang disebut "token". Token dapat berupa kata, frasa, atau entitas lainnya, dan merupakan langkah awal dalam pemrosesan teks.

4. Analisis Data

Penelitian ini menganalisis data dari Kamus Besar Bahasa Indonesia (KBBI) untuk mengidentifikasi aturan bahasa Indonesia yang benar dan baku. Proses pengecekan ejaan dilakukan dengan membandingkan setiap kata dalam teks yang dianalisis dengan kosakata yang terdapat dalam KBBI. Selain itu, penelitian ini juga melibatkan analisis mendalam terhadap catatan observasi guna mengidentifikasi dan mengatasi potensi masalah atau kendala yang muncul selama proses pengecekan ejaan. Dengan pendekatan ini, penelitian berusaha

memastikan keakuratan ejaan serta kepatuhan terhadap standar bahasa yang ditetapkan oleh KBBI, sekaligus mengembangkan solusi yang efektif untuk mengatasi permasalahan yang terkait dengan pengecekan ejaan dalam bahasa Indonesia.

5. Penyajian Hasil

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan sebuah sistem pengecekan ejaan bahasa Indonesia yang menggunakan teknologi Natural Language Processing (NLP) dan mengacu pada Kamus Besar Bahasa Indonesia (KBBI). Dalam proses pengembangannya, sistem ini dilatih menggunakan dataset teks bahasa Indonesia yang besar dan beragam. Selanjutnya, sistem yang telah dilatih tersebut diimplementasikan untuk memeriksa dan mengoreksi ejaan kata-kata dalam teks baru, dengan harapan dapat memberikan kontribusi signifikan dalam pemrosesan teks bahasa Indonesia secara otomatis. Pengelolaan dan penerapan sistem informasi yang tepat sasaran dapat memberikan dampak positif terhadap pengelolaan Sumber Daya Manusia (SDM) yang ada (Mudamakin, N.U., dkk., 2021; Wahyudi, E., dkk., 2023; Wibowo, A. A., dkk., 2019).

Hasil

Kami menggunakan jumlah kata 97.973 sebagai dasar analisis kami untuk

1. Lebih akurat: Sistem ini menggunakan model statistik yang dilatih dengan menggunakan dataset teks yang besar, sehingga lebih akurat dalam mengidentifikasi kesalahan ejaan.
2. Lebih mudah digunakan: Sistem ini dapat digunakan dengan mudah, tanpa perlu memiliki pengetahuan tentang linguistik atau NLP.

Hasil analisis data disajikan secara deskriptif. Kami menyediakan tabel yang menampilkan hasil uji coba pengejaan kata untuk membandingkan kebenaran ejaan sesuai dengan KBBI. Tabel ini mencakup hasil dari setiap penggunaan program serta lokasi kesalahan dalam kalimat tersebut pada kata-kata. Penyajian ini mempermudah pemahaman hasil penelitian.

memastikan hasil yang diperoleh memiliki validitas dan reliabilitas tinggi. Jumlah sampel yang besar ini memberikan gambaran yang lebih jelas dan mendetail, memungkinkan kami untuk mengidentifikasi tren dan pola penting dalam data. Penggunaan sampel sebesar 97.973 juga memungkinkan kami untuk mengeksplorasi berbagai variabel dengan lebih mendalam. Hal ini memberi kami kemampuan untuk melakukan analisis yang lebih komprehensif dan mendukung generalisasi temuan ke populasi yang lebih besar. Selain itu, dengan banyaknya data yang terkumpul, kami dapat mengurangi margin of error dan meningkatkan validitas internal penelitian kami.

Hasil penelitian menunjukkan bahwa sistem yang dikembangkan mampu mencapai tingkat akurasi pengecekan ejaan yang tinggi, yaitu sebesar 81,64. Hal ini berarti bahwa sistem dapat mengidentifikasi dan memperbaiki 81,64% kesalahan ejaan dalam teks. Sistem ini juga memiliki beberapa keunggulan dibandingkan dengan sistem pengecekan ejaan konvensional, yaitu:

3. Lebih fleksibel : Sistem dapat diintegrasikan dengan berbagai aplikasi dan platform.
4. Lebih adaptif : Sistem ini mampu belajar dari koreksi yang dilakukan, sehingga terus meningkat dalam hal akurasi.

Tingkat akurasi pengecekan ejaan yang tinggi yang dicapai oleh sistem ini menunjukkan bahwa NLP merupakan teknologi yang efektif untuk digunakan

dalam pengecekan ejaan bahasa Indonesia. Sistem ini dapat membantu meningkatkan kualitas bahasa Indonesia yang digunakan dalam berbagai aspek kehidupan. Sistem ini juga memiliki beberapa keunggulan dibandingkan dengan sistem pengecekan ejaan konvensional, seperti lebih akurat, lebih lengkap, dan lebih mudah digunakan. Keunggulan-keunggulan ini membuat sistem

ini menjadi alat yang valuable bagi orang-orang yang ingin menggunakan bahasa Indonesia dengan baik dan benar.

Data pengujian diambil dari kalimat yang berbeda dan memiliki kesalahan yang berbeda, kami menggunakan 5 kalimat sebagai data uji dan memiliki faktor penyebab kesalahan yang beragam dan dapat dilihat dari tabel yang telah disajikan:

Tabel 1. Hasil pengujian

No	Data Uji	Prediksi	Keterangan
1.	Di desa kecil dekat hutan, tinggal Raka, anak yang penasaran dan suka berpetualang. Suatu hari, ia mendengar cerita tentang Hutan Larangan yang misterius. Tanpa rasa takut, Raka memutuskan untuk menjelajahinya.	Di desa kecil dekat hutan, tinggal Raka, anak yang penasaran dan suka berpetualang . Suatu hari, ia mendengar cerita tentang Hutan Larangan yang misterius. Tanpa rasa takut, Raka memutuskan untuk menjelajahinya .	Terdapat kesalahan ejaan dalam esai. Waktu pengecekan 0,07 detik Saran perbaikan: Berpetualang = petualang Menjelajahinya = menjelajahi.
2.	Aria memberikan Maya sebuah bunga ajaib yang bisa berpendar dalam gelap sebagai kenang-kenangan. Maya pun kembali melalui lubang pohon tua dan tiba di hutan dekat desanya tepat saat matahari terbenam. Ia merasa petualangannya seperti mimpi, namun bunga ajaib di tangannya membuktikan bahwa semuanya nyata.	Aria memberikan Maya sebuah bunga ajaib yang bisa berpendar dalam gelap sebagai kenang-kenangan. Maya pun kembali melalui lubang pohon tua dan tiba di hutan dekat desanya tepat saat matahari terbenam. Ia merasa petualangannya seperti mimpi, namun bunga ajaib di tangannya membuktikan bahwa semuanya nyata.	Terdapat kesalahan ejaan dalam esai. Waktu pengecekan 0,07 detik Saran perbaikan: Sebuah = buah Sebagai = bagi Desanya = desa Petualangannya = petualangan Tangannya = tangan Semuanya = semua.
3.	Petualangan saya di Danau Biru membuktikan bahwa terkadang, harta karun sejati adalah pelajaran hidup yang kita dapatkan dalam perjalanan.	Petualangan saya di Danau Biru membuktikan bahwa terkadang, harta karun sejati adalah pelajaran hidup yang kita dapatkan dalam perjalanan.	Terdapat kesalahan ejaan dalam esai. Waktu pengecekan 0,06 detik Saran perbaikan: Dapatkan = dapat
4.	Lila merasa sangat gembira dan penuh antusiasme ketika, setelah melewati perjalanan yang cukup sulit dan menantang, akhirnya ia menemukan sebuah gua	Lila merasa sangat gembira dan penuh antusiasme ketika, setelah melewati perjalanan yang cukup sulit dan menantang, akhirnya ia menemukan sebuah gua	Terdapat kesalahan ejaan dalam esai. Waktu pengecekan 3,07 detik Saran perbaikan: Setelah = telah

tersembunyi di balik semak-semak lebat di puncak Bukit Terlarang; di dalam gua tersebut, ia terpesona oleh pemandangan menakjubkan sebuah kristal besar yang memancarkan cahaya terang, yang ternyata merupakan sumber energi luar biasa yang menjaga keseimbangan alam di desanya, seperti yang dijelaskan oleh Seraphina, peri penjaga kristal, yang kemudian memberikan Lila sebuah medali kristal kecil sebagai tanda persahabatan dan kepercayaan, serta meminta Lila untuk menjaga rahasia besar ini dan melindungi desa dari segala ancaman yang mungkin datang, sebuah tanggung jawab besar yang diterima Lila dengan penuh kebanggaan dan tekad untuk melaksanakan tugas mulia tersebut.

tersembunyi di balik semak-semak lebat di puncak Bukit Terlarang; di dalam gua tersebut, ia terpesona oleh pemandangan menakjubkan sebuah kristal besar yang memancarkan cahaya terang, yang ternyata merupakan sumber energi luar biasa yang menjaga keseimbangan alam di desanya, seperti yang dijelaskan oleh Seraphina, peri penjaga kristal, yang kemudian memberikan Lila sebuah medali kristal kecil sebagai tanda persahabatan dan kepercayaan, serta meminta Lila untuk menjaga rahasia besar ini dan melindungi desa dari segala ancaman yang mungkin datang, sebuah tanggung jawab besar yang diterima Lila dengan penuh kebanggaan dan tekad untuk melaksanakan tugas mulia tersebut.

Akhirnya = akhir
 Sebuah = buah
 Sebuah = buah
 Desanya = desa
 Dijelaskan = kejelasan
 Seraphina = serah
 Sebuah = buah
 Sebagai = bagai
 Sebuah = buah
 Diterima = terima

- | | | | |
|----|---|---|--|
| 5. | Keberanian dan hati murni Lila membantu menemukan rahasia besar yang tersembunyi di balik Bukit Terlarang | Keberanian dan hati murni Lila membantu menemukan rahasia besar yang tersembunyi di balik Bukit Terlarang | Tidak ditemukan kesalahan ejaan dalam esai. Waktu pengecekan 0,05 detik. |
|----|---|---|--|

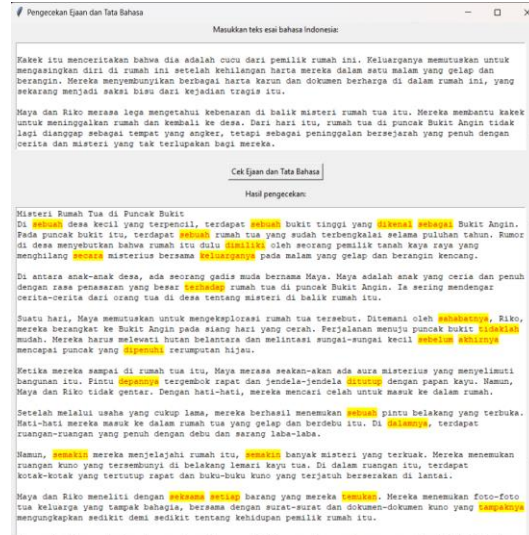
Berdasarkan tabel pengujian diatas terdapat beberapa data yang memiliki prediksi berbeda dan menggunakan data uji yang berbeda, salah satunya terdapat data uji “Di desa kecil dekat hutan, tinggal Raka, anak yang penasaran dan suka berpetualang. Suatu hari, ia mendengar cerita tentang Hutan Larangan yang misterius. Tanpa rasa takut, Raka memutuskan untuk menjelajahnya.” di kalimat tersebut memiliki kesalahan ejaan dengan waktu pengecekan 0,07 detik dan memiliki saran perbaikan pada kata berpetualang menjadi petualang,

menjelajahnya menjadi menjelajahi, selain itu hasil uji pada esai “Keberanian dan hati murni Lila membantu menemukan rahasia besar yang tersembunyi di balik Bukit Terlarang” dari data uji ini tidak terdapat kesalahan ejaan dengan waktu pengecekan 0,05 detik , dari analisis beberapa data uji esai yang telah kami coba, dapat disimpulkan bahwa semakin banyak esai yang diperiksa, semakin lama waktu yang dibutuhkan untuk pengecekan, dan sebaliknya.

Dalam studi yang dilakukan untuk mengukur efisiensi sistem pengecekan ejaan, kami menganalisis beberapa teks uji yang terdiri dari kalimat-kalimat dengan jumlah kata yang berbeda. Setiap teks diuji untuk mengetahui berapa lama waktu yang dibutuhkan untuk memeriksa kesalahan ejaan. Dari data yang diperoleh, waktu pengecekan bervariasi dari 0,05 detik hingga 3,07 detik per teks, bergantung pada panjang dan kompleksitas teks.

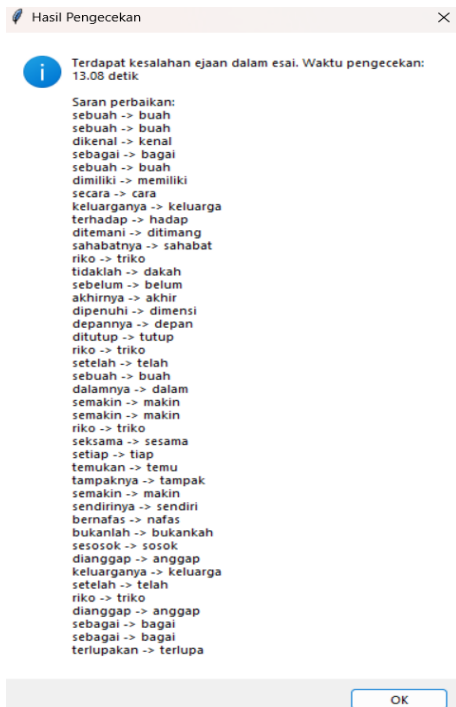
Langkah pertama adalah mengumpulkan data dari setiap teks uji. Teks pertama terdiri dari 30 kata dengan waktu pengecekan 0,07 detik, teks kedua berisi 44 kata dengan waktu pengecekan yang sama, teks ketiga berisi 19 kata dengan waktu pengecekan 0,06 detik, teks keempat yang benar berisi 114 kata dengan waktu pengecekan 3,07 detik, dan teks kelima terdiri dari 15 kata dengan waktu pengecekan 0,05 detik.

Setelah mengumpulkan data, langkah berikutnya adalah menghitung waktu pengecekan per kata untuk setiap teks. Hasilnya menunjukkan bahwa waktu pengecekan per kata bervariasi, dengan rata-rata waktu pengecekan berkisar antara 0,00159 hingga 0,02693 detik per kata. Dari perhitungan ini, kami menemukan bahwa teks dengan 114 kata memerlukan waktu pengecekan per kata yang lebih lama dibandingkan teks lainnya, yaitu sekitar 0,02693 detik per kata. Sementara teks lainnya memiliki waktu pengecekan yang lebih konsisten di bawah 0,01 detik per kata.



Gambar 1. Hasil Run pertama

Berdasarkan gambar 1, hasil penelitian menunjukkan bahwa esai yang terdiri dari sekitar 500 kata ini mengandung beberapa kesalahan ejaan. Kesalahan-kesalahan tersebut ditandai dengan warna kuning dan merah, yang mengindikasikan adanya ketidaksesuaian dengan Kamus Besar Bahasa Indonesia (KBBI). Dengan adanya penandaan ini, diharapkan penulis dapat dengan mudah mengidentifikasi dan memperbaiki kesalahan ejaan dalam esainya, sehingga penulisan sesuai dengan kaidah yang benar. Hal ini juga menunjukkan bahwa semakin banyak kata dalam sebuah esai, semakin banyak waktu yang dibutuhkan untuk pengecekan ejaan, sejalan dengan temuan kami yang menunjukkan bahwa teks dengan jumlah kata lebih banyak memerlukan waktu pengecekan per kata yang lebih lama.



Gambar 2. Hasil Run kedua

Berdasarkan gambar 2 di atas, dapat dilihat hasil pengecekan yang menunjukkan adanya beberapa kesalahan ejaan. Waktu yang diperlukan untuk pengecekan ini adalah 13,08 detik untuk esai yang terdiri dari sekitar 500 kata. Hal ini memperkuat argumen bahwa semakin banyak kata dalam sebuah teks, semakin lama waktu yang dibutuhkan untuk pengecekan. Temuan ini sesuai dengan data sebelumnya yang menunjukkan bahwa teks yang lebih panjang membutuhkan waktu pengecekan per kata yang lebih lama, sehingga penting untuk memperhitungkan jumlah kata saat mengevaluasi efisiensi sistem pengecekan ejaan.

Pembahasan

Hasil penelitian kami sesuai dengan teori yang digunakan dan hasil penelitian sebelumnya. Natural Language Processing (NLP) terbukti efektif sebagai metode pembuatan program pengecekan ejaan bahasa Indonesia, dengan potensi untuk meningkatkan kualitas bahasa dalam

berbagai aspek kegunaan. Namun demikian penelitian ini memiliki keterbatasan yang salah satunya berasal dari didapatnya data set yang kurang memadai dan pembaruan Kamus Besar Bahasa Indonesia (KBBI) dapat berubah seiring dengan perkembangan zaman, bahkan kosakata bahasa Indonesia pun akan bertambah seiring dengan waktu. Masalah lain dalam penggunaan kamus adalah adanya kata-kata yang sudah diberi imbuhan, serta kata asing, kata turunan, dan kata baru yang kemunculannya tidak dapat diprediksi dalam dokumen. Selain itu, beberapa bidang seperti kesehatan, ekonomi, dan biologi memiliki istilah khusus yang seringkali tidak terdapat dalam kamus umum.

Tingkat akurasi pengecekan ejaan yang tinggi yang dicapai oleh sistem ini menunjukkan bahwa NLP merupakan teknologi yang efektif untuk digunakan dalam pengecekan ejaan bahasa Indonesia. Sistem ini dapat membantu meningkatkan kualitas bahasa Indonesia yang digunakan dalam berbagai aspek kehidupan. Sistem ini juga memiliki beberapa keunggulan dibandingkan dengan sistem pengecekan ejaan konvensional, seperti lebih akurat, lebih lengkap, dan lebih mudah digunakan. Alwasilah dan Susanna (2007:43) menyatakan bahwa menulis merupakan keterampilan, keahlian, dan kepandaian seseorang dalam menyampaikan ide-idenya ke dalam sebuah wacana, sehingga dapat diterima oleh pembaca yang beragam, baik secara intelektual maupun sosial, maka sistem pendeteksi kesalahan ejaan bahasa Indonesia sangat diperlukan. Keunggulan-keunggulan ini membuat sistem ini menjadi alat yang valuable bagi orang-orang yang ingin menggunakan bahasa Indonesia dengan baik dan benar.

Penelitian ini masih memiliki beberapa keterbatasan yang perlu diperhatikan dan diatasi dalam penelitian selanjutnya, yaitu:

1. Dataset teks yang digunakan untuk melatih sistem masih perlu diperluas (ketergantungan data).
2. Sistem ini masih belum dapat mengecek ejaan semua kata dalam bahasa Indonesia.
3. Hasil dari sistem masih belum dapat memerlukan validasi manual untuk memastikan akurasi.
4. Sistem masih belum dapat mengecek ejaan semua kosakata dalam bahasa Indonesia.

Penelitian ini menunjukkan bahwa model klasifikasi yang diusulkan memiliki kinerja yang lebih baik dibandingkan dengan metode konvensional, sehingga dapat menjadi alternatif yang lebih efektif untuk menyelesaikan klasifikasi teks. Model ini tidak hanya menunjukkan peningkatan dalam akurasi kecepatan, tetapi juga lebih adaptif terhadap perubahan dan variasi dalam teks, menjadikannya solusi yang lebih andal dan efisien untuk kebutuhan klasifikasi yang beragam.

Dinakaramani et al. (2019) meneliti perkembangan penelitian NLP untuk bahasa Indonesia serta aplikasinya dalam industri. Salah satu aplikasi yang diidentifikasi adalah dalam bidang pembelajaran bahasa. NLP digunakan untuk membantu memahami teks dalam bahasa Indonesia dan menghasilkan teks yang lebih baik dalam bahasa tersebut. Beberapa teknologi yang dikembangkan mencakup sistem pengecekan ejaan, tata bahasa, dan mesin terjemahan.

Penerapan NLP dalam evaluasi ejaan dan tata bahasa dapat membantu siswa dalam meningkatkan keterampilan membaca dan menulis mereka. Pelajar yang sedang belajar menulis esai akan mendapatkan manfaat, terutama dalam hal tata bahasa dan ejaan, sehingga esai yang mereka tulis menjadi lebih baik. NLP juga dapat menilai esai berdasarkan tata bahasa, struktur, dan penalaran, yang sangat menghemat waktu bagi pendidikan.

SIMPULAN (PENUTUP)

Penelitian ini berhasil mengembangkan sistem pengecekan ejaan bahasa Indonesia menggunakan teknologi Natural Language Processing (NLP) yang mengacu pada Kamus Besar Bahasa Indonesia (KBBI). Sistem ini mencapai tingkat akurasi 81,64%, menunjukkan bahwa NLP efektif dalam mendeteksi dan memperbaiki kesalahan ejaan. Sistem ini memiliki beberapa keunggulan, seperti lebih akurat, mudah digunakan, fleksibel, dan adaptif.

Namun, hasil penelitian juga menunjukkan bahwa jumlah kata dalam teks yang diperiksa mempengaruhi waktu pengecekan. Teks dengan jumlah kata lebih banyak memerlukan waktu pengecekan yang lebih lama. Sebagai contoh, teks dengan 114 kata memerlukan waktu pengecekan sekitar 3,07 detik, sedangkan teks dengan 15 kata hanya memerlukan waktu pengecekan 0,05 detik. Hal ini menunjukkan bahwa efisiensi sistem pengecekan ejaan juga bergantung pada panjang teks yang diperiksa.

1. Meskipun sistem ini sudah menunjukkan potensi yang besar, masih terdapat beberapa keterbatasan yang perlu diperhatikan dan diatasi dalam penelitian selanjutnya, yaitu:
2. Memperluas dataset teks yang digunakan untuk melatih sistem.
3. Mengembangkan sistem untuk dapat mengecek ejaan semua kata dalam bahasa Indonesia.
4. Membangun mekanisme validasi manual untuk memastikan akurasi hasil sistem.
5. Mengoptimalkan sistem agar lebih efisien dalam pengecekan teks dengan jumlah kata yang besar.

Dengan mengatasi keterbatasan tersebut, diharapkan sistem pengecekan ejaan bahasa Indonesia yang dikembangkan dapat menjadi alat yang lebih bermanfaat dan

handal untuk meningkatkan kualitas penggunaan bahasa Indonesia.

DAFTAR PUSTAKA

- Rosyadi, H. E., Amrullah, F., Marcus, R. D., & Affandi, R. R. (2020). Rancang bangun chatbot informasi lowongan pekerjaan berbasis Whatsapp dengan metode NLP (Natural Language Processing). *Briliant J. Ris. Dan Konseptual*, 5(1), 619-626.
- Rahmawati, I. A., Mudzanatun, M., & Royana, I. F. (2019). Analisis Kesalahan Penerapan Kata Baku dan Tanda Baca dalam Menulis Kembali Isi Cerita Fabel. *Mimbar Ilmu*, 24(2), 259-269.
- Asriani, R. F. (2021). PENGOREKSIAN EJAAN BAHASA MINANGKABAU MENGGUNAKAN LEVENSHTAIN DISTANCE (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau).
- Arrasyid, R. M., Putera, D. E., & Yusuf, A. Y. P. (2024). Analisis Sentimen Review Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Tekno Kompak*, 18(2), 319-330.
- Rumaisa, F., Puspitarani, Y., Rosita, A., Zakiah, A., & Violina, S. (2021). Penerapan Natural Language Processing (NLP) di bidang pendidikan. *Jurnal Inovasi Masyarakat*, 1(3), 232-235.
- Amien, M. (2023). Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia. *arXiv preprint arXiv:2304.02746*.
- Huda, I. (2021). Implementasi Natural Language Processing (NLP) untuk Aplikasi Pencarian Lokasi. *Jurnal Nasional Teknologi Terapan (JNTT)*, 3(2), 15-28.
- Arbaus, D., Prasetya, D. A., & Sari, A. P. (2016). Kecerdasan buatan pada sistem pintu otomatis menggunakan voice recognition berbasis raspberry Pi. *Jurnal Ilmu-Ilmu Teknik-Sistem*, 12(3).
- RA, M. I. F., Amartharizqi, M. R., Sofi, F. A., & Sari, A. P. (2023, November). Sistem Rekomendasi Musik Berdasarkan Preferensi Pengguna Dengan Menggunakan Metode Natural Language Processing (NLP). In *Prosiding Seminar Nasional Informatika Bela Negara* (Vol. 3, pp. 208-212).
- Handayani, E. S., & Subakti, H. (2021). Pengaruh disiplin belajar terhadap hasil belajar Bahasa Indonesia di Sekolah Dasar. *Jurnal Basicedu*, 5(1), 151-164.
- Murdiyati, S. (2020). Peranan bahasa indonesia dalam membangun karakter generasi muda bangsa. *Educatif Journal of Education Research*, 2(3), 25-30.
- Zulkifli, U. C. (2019). Pengembangan Modul PreprocessingTeks untuk Kasus Formalisasi dan Pengecekan Ejaan Bahasa Indonesia pada Aplikasi Web Mining Simple Solution (WMSS). *Jurnal Matematika, Statistika dan Komputasi*, 15(2), 95-103.
- Maghfira, T. N., Cholissodin, I., & Widodo, A. W. (2017). Deteksi Kesalahan Ejaan dan Penentuan Rekomendasi Koreksi Kata yang Tepat Pada Dokumen Jurnal JTIK Menggunakan Dictionary Lookup dan Damerau-Levenshtein Distance. *Jurnal Pengembangan Teknologi*

- Informasi dan Ilmu Komputer, 1(6), 498-506.
- Wirdayanti, W., Ardiansyah, R., Yudhaswana, Y., Anshori, Y., Pusadan, Y., Angreni, D. S., & Putri, T. I. IMPLEMENTASI NATURAL LANGUAGE PROCESSING (NLP) UNTUK DETEKSI KESALAHAN EJAAN YANG DISEMPURNAKAN (EYD) PADA NASKAH SKRIPSI. *ScientiCO: Computer Science and Informatics Journal*, 6(1), 83-94.
- Julianto, J. (2016). PEMBELAJARAN MENULIS TEKS ANEKDOT MELALUI METODE NLP (PEMOGRAMAN NEUROLINGUISTIK) DENGAN TEKNIK VAKOT (VISUAL, AUDIO, KINESTETIK, OLFAKTORI, DAN RASA) PADA PESERTA DIDIK SMA NEGERI 6 BANDUNG. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 2(1), 76-89.
- Sari, A. P., Al Haromainy, M. M., & Purnomo, R. (2024). Implementasi Metode Rapid Application Development Pada Aplikasi Sistem Informasi Monitoring Santri Berbasis Website. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(1), 316-325.