



Article

Algoritma Cocke-Younger-Kasami untuk Sistem Deteksi Kesalahan Pola Kalimat Bahasa Sunda

Oktavia Putri Handayani^{1*}, Sony Kartika Wibisono², Khoirun Nisa³

¹⁻³Informatika, Universitas Harapan Bangsa, Purwokerto, Indonesia

* Correspondence: oktaviaputri378@gmail.com

Abstract: *This study develops a sentence error detection system for the Sundanese language using the Cocke-Younger-Kasami (CYK) algorithm, implemented through a website platform. The system aims to efficiently and accurately detect errors in the structure of Sundanese sentences. The CYK algorithm is used to analyze sentences based on a predefined context-free grammar (CFG). The system is built using HTML, PHP, and CSS to create the user interface and process the data. Based on testing results, the system successfully identifies both valid and invalid sentence patterns with an accuracy of 94.44%. This research is expected to contribute to the preservation and understanding of the Sundanese language, as well as make it easier for people to correct errors in their everyday communication.*

Keywords: Error detection; sentence patterns; sundanese language; CYK algorithm; web-based system.

Received: 05 January 2025

Revised: 20 February 2025

Accepted: 22 February 2025

Published: 24 February 2025



Copyright: © 2025 by the authors.

License Universitas Harapan Bangsa, Purwokerto, Indonesia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Pendahuluan

Bahasa merupakan alat komunikasi utama yang digunakan manusia untuk menyampaikan ide, gagasan, dan informasi. Bahasa memiliki sifat yang arbitrer, produktif, dan dinamis (Syahrin Effendi, 2012). Selain itu, bahasa juga berfungsi sebagai identitas budaya dan warisan tradisional yang sangat berharga (Meko et al., 2011). Indonesia memiliki lebih dari 700 bahasa daerah, termasuk Bahasa Sunda, yang merupakan salah satu bahasa daerah dengan jumlah penutur terbesar di Indonesia (Maksum et al., 2024). Bahasa Sunda memiliki keunikan pada pola kalimat, tata bahasa, dan struktur yang khas, sehingga keberadaannya menjadi bagian penting dari identitas budaya masyarakat Sunda.

Namun, di tengah modernisasi dan globalisasi, minat masyarakat terhadap penggunaan Bahasa Sunda mulai menurun. Hal ini tidak hanya berdampak pada berkurangnya pengguna aktif, tetapi juga pada meningkatnya kesalahan dalam penggunaan tata bahasa, baik dalam penulisan maupun pengucapan (Fatimah Djaja Sudarma et al., 2018). Kesalahan ini sering kali terjadi karena kurangnya pemahaman terhadap kaidah kebahasaan, khususnya pola kalimat yang menjadi inti dari struktur tata bahasa (Annyzaq Pitri Nasution, 2024).

Dalam perkembangan teknologi digital, pengembangan sistem berbasis komputer menjadi solusi untuk mendukung pelestarian dan pemahaman bahasa daerah. Salah satu metode yang dapat digunakan untuk mendeteksi kesalahan pola kalimat adalah Algoritma *Cocke-Younger-Kasami* (CYK), yang dikenal efektif dalam menganalisis tata bahasa bebas konteks (*context-free grammar*) (Hopcroft et al., 2007). Algoritma CYK menggunakan pendekatan parsing bottom-up dengan pemrograman dinamis untuk menganalisis struktur kalimat secara efisien (Jurafsky & Martin, 2000).

Penelitian sebelumnya menunjukkan bahwa Algoritma CYK berhasil diterapkan untuk deteksi kesalahan struktur kalimat dalam berbagai bahasa. Misalnya, Prabowo, dkk. menggunakan algoritma ini untuk mendeteksi kesalahan tata bahasa dan memberikan rekomendasi perbaikan, dengan akurasi mencapai 96%. Sementara itu, Sulianto dan Herwati menggunakan algoritma CYK untuk menganalisis kalimat sesuai kaidah EYD, yang hasilnya menunjukkan efisiensi tinggi dalam identifikasi elemen kalimat (Prabowo et al., 2020).

Algoritma CYK yang diimplementasikan dalam penelitian ini diharapkan dapat meningkatkan kualitas penggunaan Bahasa Sunda dan mendukung pelestarian bahasa serta budaya daerah. Sistem ini dirancang menggunakan PHP sebagai bahasa pemrograman utama untuk logika aplikasi, dan CSS untuk desain antarmuka pengguna (Fitrina Annisa Mustada, 2021). Sistem ini diharapkan dapat membantu pengguna memahami tata bahasa Sunda secara lebih baik, serta mendukung pelestarian bahasa daerah dengan memanfaatkan teknologi modern (Sulianto & Herawati, 2019). Dengan ini, masyarakat dapat meminimalkan kesalahan dalam penulisan Bahasa Sunda sekaligus meningkatkan kesadaran akan pentingnya melestarikan warisan budaya.

Metode Penelitian

Desain dan Komponen Sistem

Penelitian ini bertujuan untuk mengembangkan sistem berbasis website yang mampu mendeteksi kesalahan pola kalimat dalam Bahasa Sunda menggunakan algoritma *Cocke-Younger-Kasami* (CYK). Sistem ini dikembangkan menggunakan PHP untuk pemrosesan server, HTML dan CSS untuk antarmuka pengguna. Komponen utama yang membangun sistem ini meliputi:

1. Index.php berisi struktur utama halaman web menggunakan HTML, formulir untuk menerima input teks dari pengguna, dan logika pemrosesan menggunakan algoritma CYK untuk memvalidasi pola kalimat berdasarkan tata bahasa Sunda yang telah didefinisikan dalam variabel *\$grammar*. Algoritma ini bekerja dengan memeriksa susunan kata menggunakan aturan tata bahasa dalam bentuk CNF dan menampilkan hasil validasi apakah kalimat tersebut sesuai atau tidak.
2. Styles.css berfungsi untuk mengatur tampilan halaman, termasuk desain latar belakang dengan gambar, tata letak responsif, serta elemen visual seperti *form*, tombol, dan hasil validasi agar terlihat menarik dan terorganisir.

Algoritma CYK

Algoritma *Cocke-Younger-Kasami* (CYK) adalah metode parsing yang digunakan untuk menganalisis sintaksis kalimat berdasarkan tata bahasa bebas konteks (*Context-Free Grammar*, CFG) yang dinyatakan dalam *Chomsky Normal Form* (CNF). Algoritma ini merupakan metode *parsing bottom-up*, yang artinya mulai memeriksa bagian terkecil dari kalimat dan secara bertahap membangun struktur sintaksis yang lebih besar. CYK menggunakan tabel dua dimensi untuk menyimpan informasi tentang kemungkinan non-terminal yang dapat menghasilkan substrings dari kalimat input. Algoritma ini bekerja dengan langkah-langkah sebagai berikut:

1. Inisialisasi Tabel

Tabel dua dimensi P berukuran $(n+1) \times (n+1)$, di mana n adalah jumlah kata dalam kalimat input. Setiap elemen $P[i, j]$ dalam tabel menyimpan himpunan non-terminal yang dapat menghasilkan substring dari posisi i hingga j dalam kalimat. Diagonal tabel, $P[i, i + 1]$, diisi dengan *non-terminal* yang sesuai dengan kata terminal pada posisi i dari kalimat *input* menggunakan aturan *unary*.

2. Pengisian Tabel

Setelah tabel diinisialisasi, tabel diisi menggunakan aturan biner yang ada dalam tata bahasa. Untuk setiap *substring* yang lebih panjang dari satu kata, algoritma memeriksa semua kemungkinan pembagian substring menjadi dua bagian. Jika ada *non-terminal* yang dapat menghasilkan kedua bagian tersebut, *non-terminal* yang sesuai ditambahkan ke tabel.

3. Validasi Kalimat

Setelah tabel terisi, kalimat dianggap valid jika simbol awal tata bahasa (*start symbol*) ada di dalam entri tabel yang mewakili keseluruhan kalimat, yaitu pada $P[0, n]$, yang mewakili keseluruhan kalimat tersebut. Jika simbol awal ditemukan, kalimat tersebut valid sesuai dengan aturan tata bahasa yang digunakan.

Implementasi Algoritma CYK dalam Sistem

Implementasi algoritma *Cocke-Younger-Kasami* (CYK) pada sistem deteksi kesalahan pola kalimat bertujuan untuk menganalisis struktur sintaksis kalimat secara sistematis berdasarkan tata bahasa bebas konteks (CFG). Algoritma CYK dipilih karena keandalannya dalam memproses parsing bottom-up dengan menggunakan pendekatan pemrograman dinamis. Dalam sistem ini, algoritma diterapkan untuk memverifikasi kesesuaian susunan kata pada kalimat input dengan aturan tata bahasa yang telah didefinisikan. Berikut adalah tahapnya:

1. Input dan Pemrosesan

Pada tahap Input dan Pemrosesan, pengguna memulai dengan memasukkan kalimat melalui antarmuka web yang disediakan oleh sistem, yang diolah oleh file `index.php`. Kalimat yang dimasukkan kemudian dipecah menjadi kata-kata individu menggunakan fungsi `explode()` dalam PHP. Setelah itu, setiap kata dicocokkan dengan aturan tata bahasa yang telah didefinisikan dalam sistem, seperti aturan yang ada dalam variabel `$grammar`. Pola kalimat disusun berdasarkan urutan kelas kata yang telah dipetakan, seperti subjek, predikat, objek, atau lainnya, sehingga membentuk representasi struktural dari kalimat tersebut.

2. Validasi Pola Kalimat

Pada tahap Validasi Pola Kalimat, sistem menginisialisasi tabel dua dimensi untuk menyimpan *non-terminal* yang dapat menghasilkan *substring* dari kalimat input. Diagonal tabel diisi dengan *non-terminal* yang sesuai dengan kata terminal yang ada pada kalimat tersebut. Kemudian, tabel diisi menggunakan aturan produksi biner dari tata bahasa yang berlaku. Sebagai contoh, jika sebuah *non-terminal* S dapat menghasilkan kombinasi dua bagian kalimat, tabel akan mencatat informasi tersebut. Pada akhirnya, sistem memeriksa apakah simbol awal tata bahasa, seperti S , terdapat pada entri tabel yang mewakili keseluruhan kalimat. Jika simbol tersebut ada, kalimat dianggap valid sesuai dengan aturan tata bahasa yang telah ditentukan.

Pengujian dan Evaluasi

pengujian dilakukan untuk memastikan bahwa sistem dapat mendeteksi kesalahan dalam kalimat dengan benar. Pengujian positif dilakukan dengan menggunakan kalimat yang benar dan sesuai dengan aturan tata bahasa untuk memastikan sistem dapat memvalidasi kalimat yang valid. Sebaliknya, pengujian negatif dilakukan dengan memasukkan kalimat yang salah, misalnya dengan mengacak urutan kata, untuk memastikan bahwa sistem dapat mendeteksi kesalahan dan menandainya sebagai kalimat yang tidak valid. Evaluasi sistem dilakukan dengan

memperhatikan dua aspek utama, yaitu akurasi dalam mendeteksi kalimat yang benar dan salah serta waktu pemrosesan yang efisien. Pengujian positif memastikan kalimat yang sesuai dengan aturan dapat divalidasi dengan benar, sementara pengujian negatif memastikan kesalahan dapat terdeteksi. Dengan demikian, metode pengujian ini memberikan gambaran yang jelas mengenai kinerja dan efektivitas sistem dalam mendeteksi kesalahan pola kalimat. Tahap ini menggunakan *confusion matrix* yang memiliki dua kelas yaitu sesuai harapan dan tidak sesuai harapan. Dimana TP, TN, FN dan FP merupakan singkatan dari *true positives*, *true negatives*, *false positives*, dan *false negatives*. Akurasi, Presisi, recall, dan f1 score digunakan sebagai metrik pada evaluasi performa dari klasifikasi. Metrik tersebut diformulasikan pada persamaan (1), (2), (3) dan (4).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

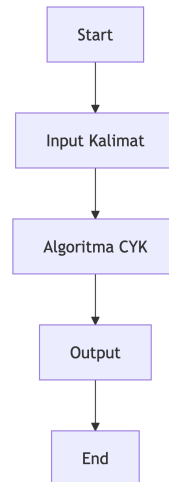
$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$\text{F1 - Score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

Hasil dan Pembahasan

Sistem deteksi kesalahan pola kalimat Bahasa Sunda berbasis algoritma *Cocke-Younger-Kasami* (CYK) telah berhasil dikembangkan dan diimplementasikan dengan menggunakan beberapa komponen penting. Sistem ini dibuat menggunakan HTML, PHP, dan CSS. HTML berfungsi untuk menyusun struktur halaman website dengan menggunakan tag-tag untuk menata setiap elemen yang ada di dalam halaman. HTML adalah bahasa standar untuk membangun struktur dokumen web, yang dapat dikelola dan disesuaikan oleh penggunaannya melalui W3C (*World Wide Web Consortium*) (Canggih Ajika Pamungkas, 2020). PHP adalah bahasa pemrograman yang digunakan sebagai interpreter untuk menerjemahkan kode sumber menjadi instruksi yang dipahami oleh mesin saat kode tersebut dieksekusi. PHP adalah bahasa pemrograman dengan lisensi sumber terbuka, yang memungkinkan pengguna untuk menyesuaikan dan mengembangkan fungsinya sesuai dengan kebutuhan. Sedangkan CSS digunakan untuk menentukan tampilan elemen-elemen HTML di halaman web, memberikan desain visual yang lebih menarik dan konsisten dengan memisahkan struktur dan desain dari konten (Satria Utama et al., 2023). Sistem deteksi ini dibangun untuk mendeteksi kesalahan pola kalimat Bahasa Sunda secara struktural, menggunakan algoritma CYK untuk melakukan analisis pola kalimat.

Diagram alur sistem deteksi kesalahan pola kalimat Bahasa Sunda yang dikembangkan dalam penelitian ini disajikan pada Gambar 1.

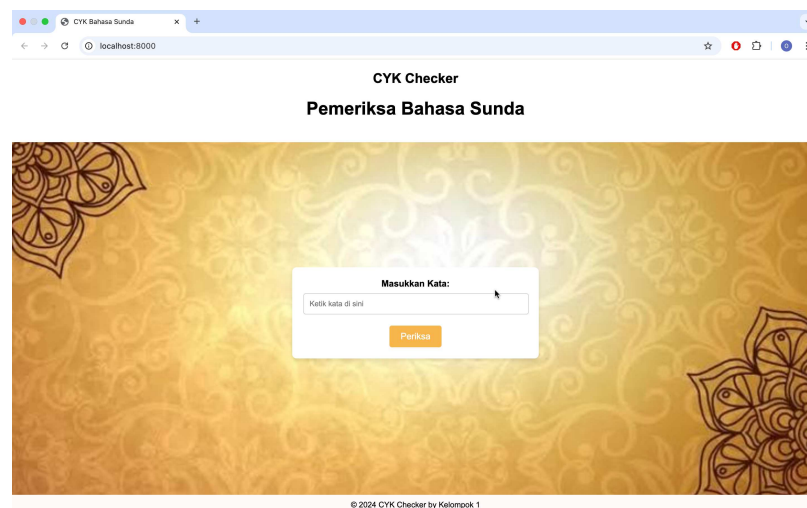


Gambar 1. Diagram Alur Sistem Deteksi Kesalahan Pola Kalimat Bahasa Sunda

Berdasarkan Diagram Alur Sistem Deteksi Kesalahan Pola Kalimat Bahasa Sunda pada Gambar 1, pengembangan sistem dilakukan menggunakan HTML, PHP, dan CSS untuk menghasilkan website yang terdiri dari halaman utama dan halaman deteksi kalimat. Berikut adalah pembahasan mengenai dua halaman utama dalam sistem ini:

1. Halaman Utama

Halaman utama menyediakan antarmuka pengguna yang sederhana dan intuitif untuk memasukkan kalimat Bahasa Sunda yang ingin dianalisis. Pengguna dapat mengetikkan kalimat di kolom yang disediakan dan sistem akan melakukan deteksi pola kalimat apakah sesuai dengan aturan yang telah ditentukan. Tampilan antarmuka halaman utama dapat dilihat pada Gambar 2.

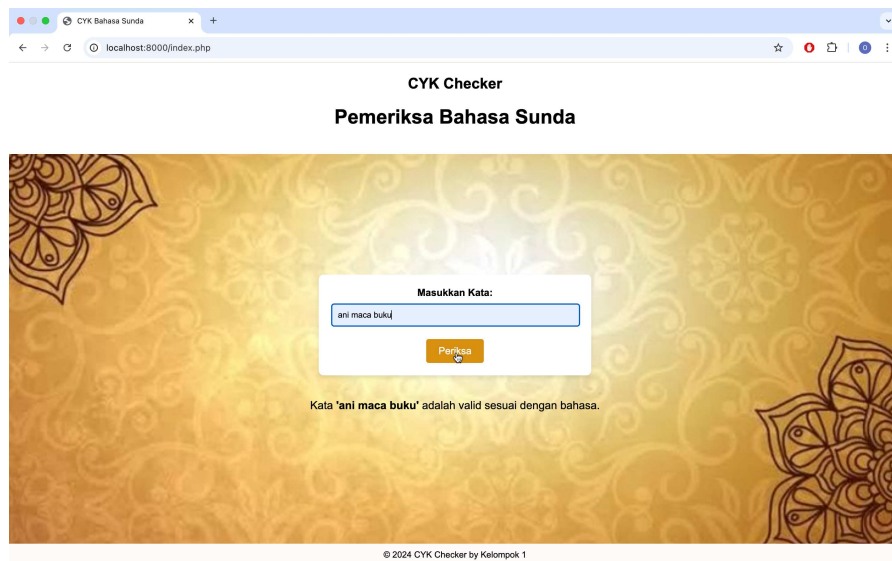


Gambar 2. Halaman Utama Website

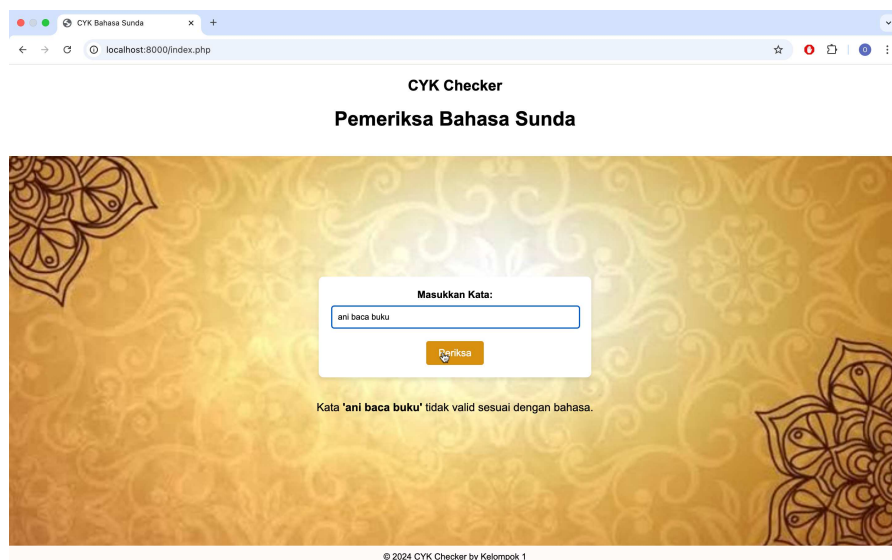
2. Halaman Deteksi Bahasa

Setelah kalimat dimasukkan, halaman deteksi akan menampilkan hasil analisis menggunakan algoritma CYK. Algoritma ini menganalisis struktur kalimat berdasarkan aturan yang didefinisikan dalam tata bahasa yang ada di dalam variabel *\$grammar*. Jika kalimat yang dimasukkan sesuai dengan aturan yang

ditentukan, sistem akan menampilkan hasil “valid”. Sebaliknya, jika kalimat tidak sesuai dengan pola yang telah ditentukan, hasil yang ditampilkan adalah “tidak valid”.



Gambar 3. Hasil Deteksi “Valid”



Gambar 4. Hasil Deteksi “Tidak Valid”

3. Hasil Pengujian Kalimat Bahasa Sunda

Pengujian terhadap sistem deteksi kesalahan pola kalimat Bahasa Sunda dilakukan dengan menggunakan beberapa kalimat uji yang bervariasi. Tujuan dari pengujian ini adalah untuk memastikan bahwa sistem dapat mendeteksi apakah susunan kata dalam kalimat tersebut sesuai dengan pola yang benar atau tidak. Tabel 1 menunjukkan hasil deteksi sistem terhadap beberapa kalimat yang diuji.

Tabel 1. Hasil Pengujian Kalimat Bahasa Sunda

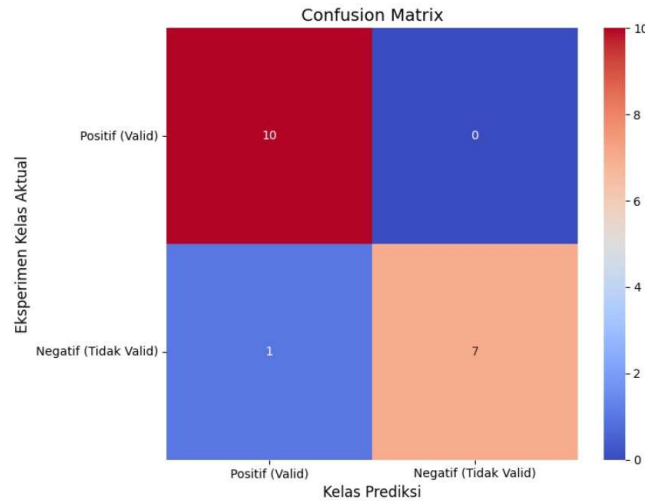
Kalimat	Pola Kalimat	Hasil Deteksi Sistem	Hasil Pengujian
"ani maca buku"	Benar	Valid	Sesuai Harapan
"budi dahar roti"	Benar	Valid	Sesuai Harapan
"sari nulis air"	Benar	Valid	Sesuai Harapan
"kucing nginum peuyeum"	Benar	Valid	Sesuai Harapan
"anjing lila piring"	Benar	Valid	Sesuai Harapan
"guru ngajual balon"	Benar	Valid	Sesuai Harapan
"ani maca gancang"	Benar	Valid	Sesuai Harapan
"budi dahar pelan"	Benar	Valid	Sesuai Harapan
"sari nulis gancang"	Benar	Valid	Sesuai Harapan
"kucing nginum pelan"	Benar	Valid	Sesuai Harapan
"guru ngajual pelan"	Salah	Valid	Tidak Sesuai Harapan
"budi kucing"	Salah	Tidak Valid	Tidak Sesuai Harapan
"ani roti"	Salah	Tidak Valid	Tidak Sesuai Harapan
"sari beureum"	Salah	Tidak Valid	Tidak Sesuai Harapan
"kucing dahar"	Salah	Tidak Valid	Tidak Sesuai Harapan
"anjing maca"	Salah	Tidak Valid	Tidak Sesuai Harapan
"maca buku"	Salah	Tidak Valid	Tidak Sesuai Harapan
"dahar air"	Salah	Tidak Valid	Tidak Sesuai Harapan

Berdasarkan hasil pengujian sistem didapatkan tabel untuk menghitung *confusion matrix* berupa TN, FN, FP, dan TP tertuang pada tabel 2.

Tabel 2. *Confusion Matrix*

Eksperimen Kelas Aktual	Kelas Prediksi	
	Positif(Valid)	Negatif(Tidak Valid)
Positif(Valid)	10	0
Negatif(Tidak Valid)	1	7

Evaluasi model klasifikasi merupakan tahapan penting dalam mengukur keandalan dan akurasi algoritma terhadap data yang diolah. Pada penelitian ini, performa model dinilai menggunakan *confusion matrix*, sebagaimana ditampilkan pada Gambar 5.



Gambar 5. Performa Model Menggunakan Confusion Matrix

Hasil analisis menunjukkan bahwa model mampu memprediksi 10 data dalam kelas aktual "Positif (Valid)" dengan benar sebagai "Positif (Valid)", menghasilkan nilai TP = 10. Model juga berhasil mengklasifikasikan dengan benar sebanyak 7 data dalam kelas aktual "Negatif (Tidak Valid)" sebagai "Negatif (Tidak Valid)" (TN = 7). Namun, terdapat 1 data dari kelas "Negatif (Tidak Valid)" yang salah diprediksi sebagai "Positif (Valid)" (FP = 1), sementara tidak ada data yang salah diklasifikasikan dari "Positif (Valid)" ke "Negatif (Tidak Valid)" (FN = 0). Data tersebut digunakan untuk menghitung akurasi, presisi, recall dan *f1-score*.

1. Akurasi

$$\text{Akurasi} = \frac{10 + 7}{10 + 7 + 1 + 0} \times 100\%$$

$$\text{Akurasi} = \frac{17}{18} \times 100\%$$

$$\text{Akurasi} = 94.44\%$$

2. Presisi

$$\text{Presisi} = \frac{10}{10 + 1} \times 100\%$$

$$\text{Presisi} = \frac{10}{11} \times 100\%$$

$$\text{Presisi} = 90.91\%$$

3. Recall

$$\text{Recall} = \frac{10}{10 + 0} \times 100\%$$

$$\text{Recall} = \frac{10}{10} \times 100\%$$

$$\text{Recall} = 100\%$$

4. F1-Score

$$\text{F1 - Score} = \frac{2 \times 90.91 \times 100}{90.91 + 100}$$

$$\text{F1 - Score} = \frac{18182}{190.91}$$

$$\text{F1 - Score} = 95.24\%$$

Berdasarkan hasil *Confusion Matrix* yang diperoleh, evaluasi performa model menunjukkan akurasi sebesar 94,44%. Nilai ini diperoleh dengan membandingkan jumlah prediksi benar (*True Positive* dan *True Negative*) terhadap total data. Selanjutnya, presisi model dihitung sebesar 90,91%, yang menunjukkan kemampuan model dalam memprediksi kelas Positif (Valid) secara benar dari seluruh prediksi Positif yang dibuat. Recall, yang mengukur kemampuan model dalam mengidentifikasi seluruh data Positif yang sebenarnya, tercatat sebesar 100%, menandakan bahwa semua data Positif (Valid) berhasil teridentifikasi oleh model.

Kombinasi antara presisi dan recall kemudian diukur menggunakan metrik *F1-Score*, yang dihasilkan sebesar 95,24%. Nilai ini menggambarkan keseimbangan antara presisi dan *recall* yang cukup baik, menunjukkan bahwa model mampu memberikan performa yang optimal dalam klasifikasi data. Berdasarkan hasil evaluasi ini, model memiliki performa yang sangat baik dan konsisten dalam mengklasifikasikan data dengan tingkat kesalahan yang rendah. Hasil ini memberikan kepercayaan bahwa model dapat digunakan untuk tugas klasifikasi serupa dengan tingkat akurasi yang tinggi.

Kesimpulan

Penelitian ini berhasil mengembangkan sistem deteksi kesalahan pola kalimat Bahasa Sunda berbasis algoritma *Cocke-Younger-Kasami* (CYK) yang diimplementasikan dalam platform website. Sistem ini mampu mendeteksi pola kalimat yang valid dan tidak valid dengan tingkat akurasi sebesar 94,44%, presisi 90,91%, recall 100%, dan *F1-score* 95,24%. Hasil ini menunjukkan bahwa algoritma CYK efektif dalam menganalisis struktur kalimat sesuai tata bahasa bebas konteks (CFG) yang telah didefinisikan, memberikan solusi cepat dan akurat dalam mendeteksi kesalahan sintaksis pada kalimat Bahasa Sunda.

Sistem ini memberikan manfaat praktis dalam membantu masyarakat memeriksa dan memperbaiki kesalahan kalimat dengan mudah, serta mendukung pelestarian bahasa daerah, khususnya Bahasa Sunda, di era digital. Meski telah menunjukkan performa yang memuaskan, penelitian lebih lanjut diperlukan untuk meningkatkan akurasi dan cakupan sistem. Salah satu langkah yang dapat dilakukan adalah memperluas database kosakata dan aturan tata bahasa, serta mengintegrasikan algoritma parsing yang lebih canggih untuk menangani struktur kalimat yang lebih kompleks.

Dengan pengembangan yang berkelanjutan, sistem ini berpotensi menjadi alat bantu penting dalam pendidikan bahasa dan pelestarian budaya Sunda, sekaligus berkontribusi terhadap kemajuan teknologi bahasa di Indonesia.

Sistem ini tidak hanya membantu pelestarian budaya tetapi juga memperkuat pengembangan teknologi lokal yang relevan di masa depan.

Daftar Pustaka

- Annyzaq Pitri Nasution 1 Melyani Sari Sitepu 2 Dewi Kesuma Nasution3. (2024). Analisis Kesalahan Berbahasa dan Membaca Di Muslim Satun Wittaya School. *Jurnal Review Pendidikan Dan Pengajaran*.
- Canggih Ajika Pamungkas, M. K. (2020). Pengantar Pemrograman Web Dengan Php Dan Html.
- Prabowo, B., Cahya Rustamadji, H., & Fauziah, Y. (2020). Algoritma Cocke Younger Kasami Untuk Deteksi Struktur Kalimat Dan Merekomendasikanya Menggunakan Algoritma Damerau Levenshtein Distance. In *TELEMATIKA* (Vol. 17, Issue 2).
- Fatimah Djaja Sudarma, T., Citraresmana, E., Indira, D., Muhtadin, T., & Hera Meganova Lyra, dan. (2018). Upaya Pemertahanan Bahasa-Budaya Sunda Di Tengah Pengaruh Globalisasi.
- Fitrina Annisa Mustada, N. I. U. O. N. P. (2021). Implementasi Frontend Untuk Sistem Deteksi Defect Produksi Kain Tekstil Pt. Gracia Mega Karya Menggunakan Metode User Centered Design (Ucd).
- Hopcroft, J. E. ., Motwani, Rajeev., & Ullman, J. D. . (2007). Introduction to automata theory, languages, and computation. Pearson/Addison Wesley.
- Jurafsky, Dan., & Martin, J. H. . (2000). Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall.
- Maksum, M. S., Arifin, T., Rohidin, R., Prasetya, M. A. B., & Anshori, I. F. (2024). Optimalisasi Algoritma Terjemahan Bahasa Dengan Model Transformer: Pendekatan Statistical Machine Learning. *Infotech Journal*, 10(2), 282–287. <https://doi.org/10.31949/infotech.v10i2.11132>
- Meko, A., Ketut, M., Nyoman, A. I., Mulyadi, S., & Elfiando, Z. (2011). Bahasa Ibu Sebagai Pilar Jati Diri Bangsa Yang Majemuk.
- Satria Utama, B., Kowi Mahmud, R., Cahyo Syahputra, Z., & Haryo Sulaksono, D. (2023). Pembuatan Website Responsif Berbasis Bahasa Pemrograman HTML dan CSS di PT. Hore Indonesia. *Prosiding Seminar Implementasi Teknologi Informasi Dan Komunikasi*, 2(1). <https://doi.org/10.31284/p.semtik.2023-1.3991>
- Sulianto, T. T., & Herawati, R. (2019). The Implementation Of Cyk Algorithm To Separate Sentence-Forming Elements In Indonesian.
- Syahrin Effendi, O. M. (2012). Linguistik Sebagai Ilmu Bahasa.