PENERAPAN *NAÏVE BAYES* UNTUK KLASIFIKASI PENYEDERHANAAN TEKS BACAAN ANAK DISLEKSIA

LAPORAN TUGAS AKHIR

Proposal ini disusun guna memenuhi salah satu syarat untuk menyelesaikan program studi Teknik Informatika S-1 pada Fakultas Teknologi Industri Universitas Islam Sultan Agung

Disusun Oleh:

Nama : Ika Rahmawati

NIM : 32602100055

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

SEMARANG

2025

APPLICATION OF NAÏVE BAYES FOR SIMPLIFIED CLASSIFICATION OF READING TEXTS FOR DYSLEXIC CHILDREN

FINAL PROJECT

This report was prepared to fulfill one of the requirements for completing the Undergraduate Informatics Engineering Study Progem at the Faculty of Industrial Technology, Sultan Agung Islamic University.

Arranged by:

Ika Rahmawati

32602100055

MAJORING OF INFORMATICS ENGINEERING

FACULTY OF INDUSTRIAL TECHNOLOGY

SULTAN AGUNG ISLAMIC UNIVERSITY

2025

LEMBAR PENGESAHAN

TUGAS AKHIR

PENERAPAN NAÏVE BAYES UNTUK KLASIFIKASI PENYEDERHANAAN TEKS BACAAN ANAK DISLEKSIA

IKA RAHMAWATI
NIM 32602100055

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 30 Juli 2025

TIM PENGUJI UJIAN SARJANA :

Dedy Kurniadi, ST.M.Kom 20-08-2025
NIK. 210615048
(Ketua Penguji)

Bagus Satrio Waluyo Poetro, S.Kom, M.Cs 21-08-2025
NIK. 210616051
(Anggota Penguji)

Mustofa ST, MM., M.Kom 22-08-2025
NIK. 2106100040
(Pembimbing)

Semarang, 30 Juli 2025
Mengetahui,
Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch. Taufik, S.T., MIT
NIK. 210604034

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini:

Nama	: Ika Rahmawati
NIM	: 32602100055
Judul Tugas Akhir	: Penerapan *Naïve Bayes* untuk Klasifikasi Penyederhanaan Teks Bacaan Anak Disleksia

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apbila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 30 Juli 2025

Yang Menyatakan,

Ika Rahmawati

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama	: Ika Rahmawati
NIM	: 32602100055
Program Studi	: Teknik Informatika
Fakultas	: Teknologi industri
Alamat Asal	: Kab. Semarang, Jawa Tengah

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : "Penerapan *Naïve Bayes* untuk Klasifikasi Penyederhanaan Teks Bacaan Anak Disleksia" Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 30 Juli 2025

Yang menyatakan,

Ika Rahmawati

KATA PENGANTAR

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan Tugas Akhir ini dengan judul "Penerapan *Naïve Bayes* untuk Klasifikasi Penyederhanaan Teks Bacaan Anak Disleksia". Tugas Akhir ini disusun sebagai salah satu syarat kelulusan dalam menempuh studi serta untuk memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika, Fakultas teknologi Industri, Universitas Islam Sultan Agung Semarang.

Dalam penyusunan Tugas Akhir ini, saya mendapatkan banyak bantuan, baik dalam aspek materi maupun teknis dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, saya ingin mengucapkan terima kasih kepada :

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Hj. Novi Marlyana, S.T., M.T., IPU., ASEAN Eng
3. Dosen Pembimbing Bapak Mustafa, ST, MM., M.Kom yang telah meluangkan waktu, membimbing, dan memberi ilmu.
4. Orang tua penulis Bapak Sutarno dan Ibu Naila Tunnajah yang selalu memberikan restu, dukungan, serta Doa dalam menyelesaikan Tugas Akhir ini.
5. Ariani selaku sahabat penulis yang senantiasa menemani penulis dalam keaadan sulit dan senang, dan memberikan dukungan kepada penulis.

Penulis menyadari bahwa dalam penyusunan laporan ini masih terdapat banyak kekurangan, untuk itu penulis mengharap kritik dan saran dari pembaca untuk sempurnanya laporan ini. Semoga dengan ditulisnya laporan ini dapat menjadi sumber ilmu bagi setiap pembaca.

Semarang, 21 Juli 2025

Ika Rahmawati

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Anak-anak dengan disleksia sering mengalami kesulitan dalam memahami teks bacaan yang kompleks, sehingga diperlukan sistem yang mampu menyederhanakan teks sesuai dengan kebutuhan mereka. Penelitian ini bertujuan untuk mengklasifikasikan tipe penyederhanaan teks bacaan bagi anak disleksia ke dalam dua kategori utama, yaitu leksikal dan sintaksis, dengan menerapkan algoritma *Multinomial Naïve Bayes*. Sebelum klasifikasi dilakukan, teks terlebih dahulu disederhanakan menggunakan metode berbasis aturan, yaitu teknik *Phrase Matcher* untuk mengganti kata atau frasa sulit dengan padanan yang lebih mudah dipahami, serta proses *stemming* untuk mengubah kata berimbuhan menjadi bentuk dasar. Teks yang telah disederhanakan kemudian diubah menjadi representasi numerik menggunakan vektorisasi TF-IDF, dan diklasifikasikan ke dalam tipe penyederhanaan menggunakan model *Multinomial Naïve Bayes*. Hasil evaluasi menunjukkan bahwa model mampu mengklasifikasikan dengan akurasi sebesar 90.09%. Pendekatan ini diharapkan dapat membantu guru atau pendidik dalam menyediakan materi bacaan yang lebih mudah dipahami secara efektif dan inklusif bagi anak-anak dengan disleksia.

Kata kunci: Naïve Bayes, klasifikasi, Penyederhanaan Teks, Anak Disleksia, Phrase Matcher, Stemming.

ABSTRACT

Children with dyslexia often have difficulty understanding complex reading texts, so a system that can simplify texts according to their needs is necessary. This research aims to classify the types of text simplification for children with dyslexia into two main categories, namely lexical and syntactic, by applying the Multinomial Naïve Bayes algorithm. Before classification is performed, the text is first simplified using rule-based methods, specifically the Phrase Matcher technique to replace difficult words or phrases with more easily understood equivalents, and the stemming process to convert inflected words into their base forms. The simplified text is then converted into a numerical representation using TF-IDF vectorization and classified into types of simplification using the Multinomial Naïve Bayes model. Evaluation results show that the model is capable of classifying with an accuracy of 90.09%. This approach is expected to assist teachers or educators in providing reading materials that are more easily understood in an effective and inclusive manner for children with dyslexia.

Keywords: Naïve Bayes, classification, Text Simplification, Dyslexic Child, Phrase Matcher, Stemming.

BAB I
PENDAHULUAN

1.1 Latar Belakang

Salah satu masalah belajar yang dihadapi anak adalah disleksia, yang merupakan ketidakmampuan mereka untuk membaca. Disleksia, menurut *National Institute of Neurological Disorders and Stroke*, adalah masalah belajar berbasis neurologi yang mempengaruhi kemampuan seseorang untuk membaca dan berbahasa. Karena membaca merupakan bagian penting dari proses belajar, ketidakmampuan membaca dapat menghambat proses belajar. Seseorang yang menderita disleksia mengalami kesulitan dalam belajar mengeja, membaca, dan menulis(Yuliana Putri dkk., 2024a). faktor genetic menjadi penyebab utama, namun faktor exsternal seperti kurangnya dukungan dari orang tua dan sistem Pendidikan yang tidak memadai juga berkontribusi terhadap kesulitan belajar anak. Anak-anak disleksia sering mengalami masalah dengan daya ingat jangka pendek dan berkonsentrasi. yang mengakibatkan pemahaman teks menjadi lebih buruk.

Mengingat tantangan yang dihadapi oleh anak-anak disleksia penyederhanaan teks bacaan menjadi sangat penting. Inovasi baru dengan sistem yang mudah dan praktis serta pelayanan yang lebih cepat dihasilkan oleh kemajuan dan modernisasi teknologi informasi dan komunikasi. Dengan kemajuan teknologi, banyak anak yang mengalami gangguan belajar, yang berdampak negatif. pada prestasi akademik mereka. Pendekatan otomatis untuk menyelesaikan masalah gangguan belajar mulai menjadi mungkin berkat kemajuan teknologi, terutama dalam bidang *Natural Language Processing* (NLP). Membangun sistem klasifikasi teks yang dapat digunakan untuk membagi teks yang dibaca berdasarkan tingkat kesederhanaannya adalah salah satu cara yang dapat digunakan. Dengan demikian, proses pengenalan teks yang sesuai untuk siswa disleksia dapat dilakukan secara lebih efektif, akurat, dan konsisten.

Salah satu pendekatan yang digunakan untuk menyelesaikan masalah ini yaitu dengan menerapkan algoritma *Naïve Bayes* untuk klasifikasi penyederhanaan teks bacaan. Algoritma *Naïve Bayes* dikenal sebagai metode probabilistic yang sederhana namun efektif. Dapat diterapkan untuk mengidentifikasi elemen dalam teks yang membuat sulit dipahami. Melalui metode ini memungkinkan teks secara otomatis di ubah menjadi versi yang lebih sederhana tanpa menghilangkan makna utamanya. Dengan menggunakan metode probalistik, algoritma ini dapat memprediksi elemen-elemen bahasa yang paling relevan untuk klasifikasi penyederhanaan teks. Dalam konteks klasifikasi penyederhanaan teks untuk anak disleksia, algoritma *Naïve Bayes* dapat membantu mengidentifikasi kata atau frasa yang sulit.

Namun, teknologi masih jarang digunakan untuk kebutuhan inklusif seperti menyederhanakan teks untuk anak disleksia. Pendekatan ini tidak hanya membantu anak-anak disleksia secara langsung, tetapi juga membantu guru, orang tua, dan pendamping dalam menyediakan bahan Pendidikan yang lebih inklusif. Adanya teknologi yang dapat menyederhanakan teks secara otomatis akan memungkinkan pengembangan Pendidikan berbasis teknologi untuk anak berkebutuhan khusus untuk berkembang seirirng dengan upaya pemerintah dalam meningkatkan akses ke Pendidikan inklusif Indonesia. Dalam beberapa penelitian sebelumnya salah satu di antaranya (Liana & Sinaga, 2021)dalam penelitian yang berjudul “Sistem Pakar Diagnosa Penyakit Disleksia pada Anak Dengan Metode *Naïve Bayes* Berbasis Web” menunjukkan bahwa metode *Naïve Bayes* berbasis web telah berhasil digunakan untuk membangun sistem pakar diagnose disleksia pada anak yang memberikan hasil diagnosis dengan klasifikasi tertinggi.

Ketidakmampuan belajar yang disebabkan oleh gangguan neurologis yang menyebabkan kesulitan membaca dan menulis dikenal sebagai disleksia. Metode multisensori, bercerita, dan membaca nyaring adalah cara guru dapat membantu anak disleksia meningkatkan keterampilan membaca mereka. Di antara 2.470 artikel ilmiah yang digunakan dalam penelitian selama sepuluh tahun terakhir, data yang terkait dengan topik penelitian dikumpulkan; dari jumlah tersebut, 7 artikel telah ditemukan sesuai dengan tema penelitian. Pendidikan umumnya adalah upaya

yang dilakukan secara sadar dan terencana untuk menciptakan keadaan belajar dan sistem evaluasi untuk anak dan atau peserta didik dengan tujuan meningkatkan pengetahuan spiritual, pengendalian diri, potensi kecerdasan, akhlak, dan keterampilan. Pendidikan adalah proses pembelajaran sekelompok orang tentang pengetahuan, kemampuan, kebiasaan yang diwariskan melalui Pelajaran dari satu generasi ke generasi berikutnya. Tujuannya adalah untuk menggunakan Pendidikan sebagai alat utama untuk menyukseskan Pembangunan nasional. Pembelajaran individual, juga dikenal sebagai program Pendidikan khusus, adalah salah satu jenis Pendidikan yang dirancang khusus untuk anak-anak yang memiliki kebutuhan khusus.

Dengan latar belakang, ini penelitian ini bertujuan untuk membantu anak-anak disleksia lebih baik dalam memahami teks, meningkatkan keterampilan membaca mereka secara bertahap, dan menumbuhkan rasa percaya diri. Perkembangan Pendidikan inklusif semakin memperhatikan kebutuhan belajar individu.

1.2 Perumusan Masalah

Berdasarkan uraian pada bagian latar belakang, maka rumusan masalah dalam penelitian ini dapat disusun sebagai berikut:

1. Bagaimana algoritma *Naïve Bayes* dapat membantu anak-anak dengan disleksia membaca teks bacaan dengan mengklasifikasikan kata atau frasa yang sulit dan menggantinya dengan padanan yang lebih mudah.
2. Bagaimana akurasi hasil klasifikasi tipe penyederhanaan teks menggunakan algoritma *Naïve Bayes* dalam membedakan antara penyederhanaan leksikal atau sinaksis.

1.3 Pembatasan Masalah

Batasan masalah dalam penelitian ini ditentukan sebagai berikut:

1. Berdasarkan fitur linguistik teks, penelitian ini hanya membagi teks bacaan anak ke dalam dua kategori: "sintaksis" dan "leksikal".
2. Penelitian ini hanya menggunakan teks bacaan berbahasa Indonesia yang ditujukan untuk anak-anak usia sekolah dasar. Teks dalam bahasa asing atau untuk usia lain tidak dimasukkan.

1.4 Tujuan Tugas Akhir

Adapun tujuan dari penulisan tugas akhir ini adalah sebagai berikut:

1. Menggunakan algoritma *Naïve Bayes* untuk membuat model klasifikasi teks yang dibaca anak berdasarkan seberapa sederhana teks tersebut.
2. Mengkategorikan teks bacaan anak ke dalam kategori "sintaksis" dan "leksikal".

1.5 Manfaat

Adapun manfaat dari penulisan tugas akhir ini adalah sebagai berikut:

1. Sistem klasifikasi otomatis berbasis algoritma *Naïve Bayes* mempermudah proses penyaringan dan kurasi teks bacaan, mengurangi beban kerja manual bagi guru dan penyusun materi pembelajaran untuk anak disleksia.
2. Menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan teks dan berkontribusi pada pengembangan teknik dan teknologi bantu pembelajaran inklusif, terutama untuk anak dengan masalah membaca seperti disleksia.
3. Menyediakan solusi teknologi yang membantu anak disleksia mendapatkan akses ke materi bacaan yang mudah dipahami, mengurangi kesenjangan pembelajaran, mendukung upaya untuk meningkatkan kualitas pendidikan inklusif.

1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan oleh penulis dalam sebuah pembuatan laporan tugas akhir ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Pada Bab 1, penulis mengutarakan urgensi dari penelitian yang diangkat, mulai dari penulisan latar belakang, membuat rumusan masalah, membatasi permasalahan yang dibahas, serta tujuan dan manfaat yang diperoleh, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Pada bab 2, penulis membuat dasar teori yang digunakan, serta rujukan dari penelitian terdahulu yang akan digunakan dalam perancangan sistem, dan membantu penulis untuk memahami teori yang berhubungan dengan *Naïve Bayes* selama proses penelitian.

BAB III : METODE PENELITIAN

Pada bab 3, penulis mengungkapkan proses dan tahapan penelitian yang dimulai dari mendapatkan dataset hingga proses pemodelan topik data yang ada.

BAB IV : HASIL DAN ANALISIS PENELITIAN

Pada bab 4, penulis mengungkapkan hasil penelitian, dimulai dengan hasil akhir sistem, klasifikasi data uji, dan akurasi dari sistem.

BAB V : KESIMPULAN DAN SARAN

Pada bab 5, penulis memaparkan kesimpulan dari hasil proses penelitan mulai dari awal sampai akhir.

BAB II
TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Tinjauan penemuan penelitian sebelumnya sangat penting untuk mendukung penelitian yang akan datang. Disatu sisi, tinjauan tersebut memberikan bahan yang berguna untuk membandingkan penelitian yang telah dilakukan sebelumnya, membantu mengidentifikasi keunggulan dalam kelemahan. Selain itu, tinjauan tersebut juga membantu memperkuat argument yang dibuat dalam penelitian baru, yang memungkinkan penelitu untuk mendapatkan. Berikut adalah beberapa penelitian yang serupa:

Adapun penelitian sebelumnya (Silvana dkk., 2020)Melakukan penelitian menggunakan metode *Naïve Bayes* saat menerapkan sistem pakar untuk menilai masalah perkembangan anak. Dukumen ini membahas pembuatan mesin inferensi yang dirancang untuk mendiagnosis gangguan perkembangan pada anak melalui sistem berbasis aturan yang memanfaatkan inferensi chaining maju. Ini menyoroti identifikasi kondisi seperti hiperaktivitas dan disleksia berdasarkan gejala tertentu, menunjukkan akurasi diagnostic yang lebih tinggi diantara di antara para ahli dibandingkan dengan non ahli. Selain itu, dokumen ini mengeksplorasi penerapan metode *Naïve Bayes* untuk deteksi dini gangguan perkembangan, menekankan pentingnya antarmuka yang ramah pengguna untuk input gejala dan diagnosis pada akhirnya bertujuan untuk membantu orang tua dalam mengenali gangguan ini dengan lebih efektif.

Kemudian penelitian (Agastya & Aripin, 2020)Melakukan penelitian pemetaan emosi dominan dalam kalimat bahasa Indonesia majemuk menggunakan multinomial *Naïve Bayes*. Studi berfokus pada pemetaan emosi dominan dalam kalimat majemuk bahasa Indonesia menggunakan model Multinomial *Naïve Bayes*, dengan tujuan untuk mengklasifikasikan kalimat kedalam enam emosi dasar: kebahagiaan, kesedihan, kemarahan, ketakutan, jijik, dan kejutan. Penelitian ini menyoroti pentingnya pemetaan emosi dalam bidang seperti produksi film animasi dan interaksi manusia-komputer, terutama untuk menghasilkan ekspresi wajah yang

sesuai pada karakter virtual. Metodelogi yang digunakan melibatkan klasifikasi teks dan persamaan batas dominan untuk mengidentifikasi emosi, dengan hasil yang menunjukkan bahwa pendekatan ini dapat secara efektif memetakan beberapa emosi dominan dalam kalimat majemuk, mengatasi keterbatasan metode tradisional yang biasanya hanya menghasilkan satu kelas emosi.

Kemudian dari hasil penelitian (Surayya & Mubarok, 2021) Pengaruh aplikasi marbel membaca pada kemampuan membaca anak yang didiagnosis dengan disleksia. Karena ketidakmampuan mereka untuk membaca dan menghafal abjad, anak disleksia akan sangat tertinggal dibandingkan dengan anak di usianya, sehingga pembelajaran harus lebih menekankan pada proses belajar mereka dalam mengeja, menulis, dan membaca. Untuk menyelesaikan masalah ini, penelitian eksperimen yang menggunakan aplikasi pembelajaran harus dilakukan untuk membantu anak disleksia menghafal abjad. Tujuan dari penelitian ini adalah untuk membantu Aplikasi *Marbel Reading* membantu anak disleksia menghafal abjad dengan baik. Uji ekspresimen dilakukan oleh peneliti setelah menentukan bahan, alat, dan strategi penelitian, menentukan tujuan penelitian, dan menyiapkan peralatan dan sarana yang diperlukan

Kemudian dari penelitian (Oktamarina dkk., 2022)Melakukan penelitian gangguan gejala disleksia pada anak usia dini untuk mengetahui tentang gangguan gejala disleksia pada anak usia dini. Metode penelitiannya adalah jenis penelitian yang dipakai adalah penelitian kualitatif dengan metode studi kasus. Studi kasus yaitu jenis penelitiannya kualitatif yang memiliki kemampuan untuk menyelidiki proses, mengumpulkan data, dan mendapatkan informasi lebih lanjut. Namun kasus yang diteliti adalah perkembangan kognitif anak dengan disleksia. Data yang dikumpulkan melalui pengamatan anak dan wawancara dengan orang tua, dan pendidik yang menangani anak-anak dengan disleksia akan kesulitan menemukan dan mengubah kata-kata yang tepat menjadi huruf atau kalimat. Hasil penelitian menunjukkan disleksia tergolong sebagai gangguan saraf pada bagian otak, yang memproses bahasa. Terbukti bahwa mendeteksi dan menangani penderita sejak usia dini meningkatkan kemampuan mereka dalam membaca.

Kemudian penelitian (Sinaga, 2023)Melakukan penelitian Penerapan algoritma *Naïve Bayes* dalam pemrosesan bahasa alamiyah. Eksplorasi teoretis dan praktis terhadap algoritma *Naïve Bayes* dalam pemrosesan bahasa alamiah menyoroti keefektifan model ini dalam meningkatkan pemahaman dan analisis teks. Algoritma *Naïve Bayes*, sebagai pengklasifikasi probabilistik, memberikan dasar yang kuat untuk mengatasi tugas klasifikasi teks, terutama dalam analisis sentimen. Model ini menggunakan pendekatan berbasis teorema Bayes dengan mengasumsikan independensi fitur, menjadikannya sederhana namun kuat. Metodologi penelitian mencakup langkah-langkah terstruktur mulai dari impor pustaka *Python* hingga evaluasi model *Naive Bayes*, dengan feature engineering, penskalaan fitur, dan optimasi analisis sentimen sebagai Langkah-langkah kunci untuk meningkatkan kinerja. Studi kasus tentang analisis sentimen dengan *Naïve Bayes* memberikan gambaran konkret tentang aplikasinya dalam memprediksi sentimen dari teks, dengan penggunaan leksikon sentimen, *Naïve Bayes* biner, dan penanganan negasi memberikan peningkatan yang signifikan terhadap generalisasi model.

Kemudian dari penelitian (Pebdika dkk., 2023)Melakukan penelitian penelitian klasifikasi yang mengidentifikasi kandidat penerima pip menggunakan teknik *Naïve Bayes*. Dokumen ini membahas implementasi program Indonesia pintar (PIP), yang bertujuan untuk meningkatkan akses ke pendidikan bagi anak-anak yang tidak mau bersekolah, terutama melalui penerapan Teknik data mining seperti algoritma *Naïve Bayes*. Dokumen ini menyoroti efektivitas algoritma ini dalam mengklasifikasikan calon penerima bantuan Pendidikan, dengan menunjukkan tingkat akurasi yang tinggi sebesar 88,89% dalam memprediksi kelayakan. Penelitian ini menekankan pentingnya pendidikan bagi siswa yang kurang mampu dan peran inisiatif pemerintah dalam mengatasi ketimpangan Pendidikan.

2.2 Dasar Teori

2.2.1 Disleksia

Disleksia, berasal dari kata Yunani "dys" yang berarti "sulit dalam" dan "lex" yang berarti "kata", didefinisikan sebagai ketidakmampuan untuk memahami informasi melalui proses pembelajaran. Sebagai sindrom, disleksia menyebabkan kesulitan dalam mempelajari dan mengintegrasikan kompenen kata dan kalimat serta waktu, arah, dan masa. Anak-anak yang didiagnosis dengan disleksia mengalami kesulitan membaca karena kelainan syaraf yang terjadi pada otak mereka. Jenis kesulitan belajar membaca bervariasi, tetapi semuanya menunjukkan bahwa ada gangguan dalam fungsi otak. Berpikir linear dan berbicara, anak-anak dengan disleksia memiliki bagian otak kiri yang lebih kecil daripada manusia normal. Inilah yang membedakan pemrosesan data dan kemampuan berbahasa(Indriastuti, 2015).

Seseorang yang menderita disleksia menunjukkan kesulitan yang signifikan dalam bidang berbahasa, seperti mengeja, membaca, dan menulis. Kebanyakan orang menganggap disleksia sebagai kondisi di mana anak-anak mengalami kesulitan belajar membaca, kesulitan menulis jika mereka menulis dengan huruf yang terbalik, atau kesulitan mengeja jika ada banyak huruf yang hilang. Karena gangguan disleksia, seorang anak mengalami kesulitan membaca, menulis, dan memahami kata dan kalimat yang diterimanya. Gangguan ini menyebabkan nilai rendah dan prestasi rendah di sekolah(Darmayanti dkk., 2023).

Disleksia adalah suatu bentuk kesulitan belajar spesifik yang ditandai dengan kesulitan dalam membaca susunan kalimat atau menulis, serta kesulitan Orang dengan disleksia mungkin mengalami kesulitan dalam membedakan huruf tertentu, seperti b dan d, M dan N, u dan w, p dan f, c dan e, a dan o, R dan P, dan p dan d. Mereka juga mungkin mengalami kesulitan dalam menghubungkan bunyi dengan huruf, dan memahami teks tertulis dengan baik. Akibatnya, sangat penting untuk memahami situasi ini dan menemukan strategi pembelajaran yang tepat untuk membantu mereka meningkatkan kemampuan membaca mereka(Rahmawati & Muhroji, 2024).

Menurut *National Institute of Neurological Disorders and Stroke* (NINDS, 2011), disleksia didefinisikan sebagai ketidakmampuan belajar yang disebabkan oleh masalah neurologis yang secara khusus mempengaruhi kemampuan seseorang untuk membaca dan menulis. Rowan juga menganggap disleksia sebagai masalah ekspresi diri secara tertulis, baik dalam membaca maupun mengeja, serta keterampilan membaca dan menulis yang buruk. Disleksia adalah gangguan yang menyebabkan kesulitan dalam belajar membaca, mengeja, dan menulis, seringkali kurang dari yang diharapkan meskipun mereka memiliki kecerdasan normal(Yuliana Putri dkk., 2024).

Disleksia merupakan gangguan neurologis yang mengganggu kemampuan seseorang untuk membaca, mengeja, dan memahami kata. Gangguan ini tidak sebanding dengan tingkat kecerdasan atau pendidikan mereka(Siti Alyunita Mega Lestari dkk., 2024). Anak-anak dengan dyslexia sering menghadapi masalah di sekolah dan dalam kehidupan sehari-hari, terutama dalam mengenali huruf, menghubungkan suara dengan huruf, dan memahami urutan kata. Mereka juga mungkin mengalami kesulitan mengeja kata, membaca lancar, dan memahami teks, yang seringkali membuat mereka frustrasi dan menghindari kegiatan membaca. Hasil definisi menunjukkan bahwa anak-anak dengan gangguan disleksia tidak dapat diklasifikasikan sebagai anak keterbelakangan mental. Sebelum menandai seorang anak sebagai kelompok risiko disleksia, seseorang harus memastikan bahwa tingkat kecerdasan kognitifnya berada dalam rentang normal atau di atas rata-rata.

Didasarkan pada beberapa pengertian di atas, disleksia adalah gangguan belajar yang ditandai dengan kesulitan membaca, menulis, atau mengeja. Penderita akan menghadapi kesulitan dalam mengidentifikasi kata-kata yang mereka ucapkan dan mengubahnya menjadi huruf atau kalimat. Disleksia adalah gangguan saraf pada bagian otak yang bertanggung jawab untuk memproses bahasa. Kondisi ini dapat terjadi pada anak-anak atau orang dewasa. Tingkat kecerdasan penderitanya tidak dipengaruhi oleh disleksia, meskipun penyakit ini menyebabkan kesulitan belajar. Oleh karena itu, pendekatan pembelajaran khusus diperlukan, yang mencakup pendekatan pengajaran yang interaktif, penyederhanaan bahasa, dan

dukungan emosional untuk menjaga semangat siswa dan memahami materi dengan cara yang sesuai dengan kemampuan mereka.

2.2.2 Penyederhanaan Teks

Penyederhanaan teks merupakan proses mengubah teks asli menjadi bentuk yang lebih sederhana tanpa menghilangkan makna inti dari teks sebelumnya. Proses ini bertujuan untuk meningkatkan aksesibilitas dan pemahaman bagi pembaca seperti anak-anak pelajar dengan kesulitan belajar atau sering juga di sebut anak disleksia. Tujuan penyederhanaan teks sangat membantu anak-anak disleksia, yang pertama untuk mempermudah pemahaman, meningkatkan aksesibilitas, dan mengurangi beban kognitif. Berdasarkan masalah tersebut maka diusulkan menggunakan *Pharase Matcer* yang dapat membantu menyederhanakan sebuah kalimat dalam suatu dokumen.

Dalam penelitian ini, proses penyederhanaan teks dilakukan menggunakan pendekatan *rule-based* yang terdiri dari dua tahapan utama, yaitu *phrase matcher* dan *stemming*:

1. *Phrase Matcher*

   Phrase Matcher adalah salah satu fitur dalam pustaka *spaCy* yang berfungsi untuk mendeteksi kata atau frasa tertentu dalam teks yang telah ditentukan sebelumnya. Teknik ini digunakan untuk mengganti kata atau frasa sulit dengan padanan kata yang lebih sederhana, berdasarkan daftar kata yang telah ditentukan melalui pemetaan manual. Misalnya, kata “konstitusi” dapat diganti menjadi “aturan”.

2. *Stemming*

   Stemming merupakan proses untuk mendapatkan kata dasar dari kata asli pada suatu kalimat. Dalam kasus tertentu, kata asli dapat mengandung imbuhan yang terpisah, seperti "makan", "dimakan", dan "memakan", yang masing-masing memiliki kata dasar yang sama, "makan"(Kusuma Wardana dkk., 2019).

2.2.3 Klasifikasi

Klasifikasi adalah suatu proses pembelajaran yang bertujuan untuk menentukan setiap himpunan karakteristik dari sebuah objek. Dalam beberapa

kasus, klasifikasi juga disebut sebagai pengelompokkan suatu data. Pengelompokkan adalah penempatan suatu objek ke dalam kelas yang memiliki fungsi yang sama dengannya. Penempatan suatu objek ke dalam kelas dengan tujuan yang sama dikenal sebagai kelompokan. Sebuah dataset yang memiliki setiap jenis data, yaitu berbentuk nomial atau biner, dapat digambarkan dengan penggunaan klasifikasi. Dalam klasifikasi data yang diawasi, data akan dibagi menjadi dua data pelatihan dan data pengujian. Data pelatihan akan dianalisis dengan menggunakan algoritma klasifikasi(Marito Putry & Nurina Sari, 2022).

Klasifikasi teks merupakan proses mengelompokkan dokumen atau teks ke dalam kategori tertentu berdasarkan isi atau ciri-ciri tertentu. Dalam penelitian ini, klasifikasi dilakukan untuk membedakan jenis penyederhanaan teks menjadi dua kategori, leksikal dan sintaksis(Wibowo, 2020).

a. Penyederhanaan Leksikal

   Berfokus pada pengganti kata yang kompleks dengan padanan yang lebih sederhana.

b. Penyederhanaan Sintaksis

   Mengatur ulang struktur kalimat agar lebih sederhana, seperti menghapus kaluso subordinat atau pengurangan elemen yang tidak esensial.

Metode klasifikasi menempatkan benda dalam kelompok tertentu berdasarkan karakteristik yang dimilikinya. Dalam prosesnya, klasifikasi dapat dilakukan dengan berbagai cara, seperti secara manual atau dengan bantuan teknologi. Klasifikasi manual berarti klasifikasi dilakukan oleh manusia tanpa bantuan algoritma cerdas komputer. Meskipun klasifikasi menggunakan teknologi, ada banyak algoritma, termasuk *Naïve Bayes, Support Vector Machine*, Decission Tree, Fuzzy, dan Jaringan Saraf Tiruan. Klasifikasi adalah pembagian objek ke dalam kelompok tertentu berdasarkan kelompoknya, yang biasanya disebut sebagai kelas. Semua ahli setuju bahwa klasifikasi adalah proses menempatkan sesuatu ke dalam kelompok berdasarkan karakteristiknya. Proses klasifikasi dapat dilakukan secara teknologi atau manual oleh manusia menggunakan berbagai algoritma seperti Jaringan Saraf Tiruan, *Naïve Bayes*, *Support Vector Machine, Decision Tree*,

dan Fuzzy. Metode klasifikasi digunakan untuk mengkategorikan barang-barang ke dalam kelompok atau kelas tertentu. Dengan kata lain, klasifikasi membantu pemahaman dan pengorganisasian data atau objek berdasarkan persamaan atau perbedaan atributnya, baik secara manual maupun dengan bantuan komputer(Oktaviyani dkk., 2024).

2.2.4 Padanan Kata

Padanan kata digunakan untuk membantu menyederhanakan teks, terutama untuk anak-anak disleksia, untuk menggantikan kata-kata yang dianggap sulit, jarang digunakan, atau teknis dengan kata-kata yang lebih sederhana, umum, dan mudah dipahami. Tujuan dari proses ini adalah untuk meningkatkan keterbacaan dan pemahaman teks tanpa mengubah makna utama kalimat. Karena kata padanan tersebut digunakan lebih sering dalam kehidupan sehari-hari dan menjadi lebih sederhana untuk dipahami oleh anak-anak dengan gangguan membaca, kata-kata seperti "implementasi" dapat diganti dengan "pelaksanaan" atau "aturan" dapat diganti dengan "aturan". Proses simplifikasi bahasa sangat bergantung pada penggunaan padanan kata, terutama ketika menggunakan metode pembelajaran mesin seperti *Naïve Bayes*, yang dapat menemukan kata yang sulit dan menyarankan kata pengganti yang lebih sederhana(Bernard dkk., 2018).

Pemilihan padanan kata tidak hanya mempertimbangkan kesamaan makna, tetapi juga panjang kata, keterbacaan, dan familiaritas kata dengan pembaca sasaran—anak-anak dengan disleksia. Dengan demikian, padanan kata yang dipilih idealnya harus memiliki suku kata yang lebih sedikit dan tidak mengandung elemen bahasa akademik atau teknis. Selain itu, kamus sederhana atau kamus sinonim yang disesuaikan dengan usia dan kemampuan kognitif anak dapat membantu dalam memilih padanan kata.

Padanan kata berfungsi sebagai output dalam sistem penyederhanaan teks berbasis algoritma seperti *Naïve Bayes*, menggantikan kata-kata yang telah diklasifikasikan sebagai kompleks. Oleh karena itu, padanan kata sangat penting untuk memastikan bahwa teks yang telah disederhanakan tetap memiliki makna aslinya, tetapi membuatnya lebih mudah dipahami oleh pembaca yang mengalami kesulitan memahami teks konvensional, seperti anak-anak dengan disleksia.

Kosakata suatu bahasa dapat menunjukkan kemajuan peradaban bangsa pemiliknya karena kosakata merupakan alat untuk pengungkapan seni, ilmu, dan teknologi(Sugono & Qodratillah, 2013). Seiring perkembangan sejarah Indonesia, kosa kata terus berkembang. Kemajuan ini didorong oleh kemajuan teknologi informasi yang dapat melampaui ruang dan waktu. Kamus membantu orang memahami arti kata. Orang yang berbicara memiliki ide, tetapi mereka tidak menemukan kata yang tepat untuk mengkomunikasikannya. Ini adalah di mana tesaurus diperlukan. Pusat Bahasa sekarang menyediakan Tesaurus bahasa Indonesia, yang disusun berdasarkan penelitian tentang berbagai area penggunaan bahasa.

Tesaurus ini menyediakan kumpulan kata dengan makna yang sama atau hampir sama. Tesaurus adalah kumpulan kata yang memiliki arti yang terkait satu sama lain. Tesaurus pada dasarnya berfungsi untuk mengubah ide menjadi kata atau sebaliknya. Berdasarkan abjad, sinonim lema terdiri dari lema yang memiliki makna yang sama yang terhubung ke kata-kata dasar, kata-kata turunan, dan frasa atau grup kata. Kata-kata yang bersinonim dapat berasal dari bahasa baku, bahasa konvensional, bahasa modern, atau bahasa kuno. Tesaurus ini tidak mencantumkan label ragam lain selain kiasan dan percakapan. Ini dilakukan untuk memastikan bahwa kata-kata dapat digunakan lagi saat berbicara.

Oleh karena itu, padanan kata membantu penyederhanaan teks tidak hanya sebagai pengganti semata, tetapi juga membantu anak-anak dengan disleksia memahami dan menggunakan bacaan dengan lebih mudah. Metode ini memungkinkan mereka untuk mendapatkan informasi yang sama tanpa terbebani oleh bahasa yang kompleks. Dikombinasikan dengan teknik seperti *Naïve Bayes*, proses ini menjadi lebih sistematis, terukur, dan dapat diotomatisasi. Ini memungkinkan pengembangan alat bantu baca berbasis teknologi yang lebih inklusif dan sesuai dengan kebutuhan pembaca yang memiliki keterbatasan kognitif atau bahasa.

Berikut beberapa contoh padanan kata untuk membantu peneliti menyederhanakan teks (Sugono & Qodratillah, 2013):

Tabel 2. 1 Contoh Padanan kata

Kata	Persamaan
Kanda	Kakak, mas, abang, uda
Abu-abu	Samar, kelabu, samar-samar
Acak	Awur, random, sembarang
Terakhir	Berhenti, beres, selesai
Alibi	Alasan, dalih
Beralih	Beranjak, berasak
Mengamati	Memandangi, memeriksa
Awut-awutan	Memberantakkan, porak-porak
Awas	Cermat, telik
Cantik	Geulis, elok, indah, rupawan
Ayunan	Buaian, limbai, sandungan
Azab	Hukuman, kesengsaraan, siksaan
Atau	Alias, ataupun, maupun
Rajin	Tekun, giat, aktif
Mudah	Gampang, enteng, ringan
Sulit	Rumit, susah, pelik
Bodoh	Dungu, tolol, bebal
Sabar	Tabah, tahan uji, tenang
Bohong	Duta, palsu, tipu
Jujur	Tulus, Amanah, lurus
Sedih	Duka, pilu, tipu
Bahagia	Senang, gembira, riang
Takut	Cemas, gentar, ngeri
Lambat	Perlahan, pelan, lelet
Cepat	Lekas, gesit, tangkas
Miskin	Melarat, sengsara
Kaya	Berada, melarat, sengsara
Pandai	Pinter, cerdas, bijak
Marah	Murka, geram, gusar

Kecil	Mini, mungil, sempit
Baik	Bagus, positif, mulia
Cepat	Lekas, kilat, tangkas
Mengajar	Membimbing, mendidik
Belajar	Menimba ilmu, studi, mempelajari
Guru	Pengajar, dosen
Siswa	Murid, pelajar, anak didik
Ujian	Tes, evaluasi, asesmen
Prestasi	Keberhasilan, pencapaian, hasil
perpustakaan	Pustaka, pusat baca

2.2.5 Algoritma *Naïve Bayes*

Naïve Bayes adalah pengklasifikasi probabilitas sederhana yang menghasilkan sekumpulan kemungkinan dengan menjumlahkan frekuansi yang menggabungkan nilai dari dataset tertentu. Algoritma ini mengambil semua karakteristik independen atau tidak saling ketergantungan yang diberikan oleh nilai variable kelas, dan menggunakan teorema Bayes untuk melakukannya. Nilai penyederhanaan, yang didasarkan pada *Naïve Bayes*, menyatakan bahwa jika nilai output diberikan kepada atribut, nilai-nilai tersebut secara kondisional tidak berhubungan satu sama lain. Keuntungan dari Metode *Naïve Bayes* hanya membutuhkan jumlah data pelatihan yang relatif kecil untuk menentukan perkiraan parameter yang diperlukan untuk proses pengklasifikasian(Susilawati & Iqbal, 2025).

$$\mathrm{P}\,(\mathrm{H} \mid \mathrm{X}) = \left(\frac{\mathrm{P}\,(\mathrm{H}|X)\;P(H)}{\mathrm{P}(\mathrm{X})}\right)$$

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probalitas hipotesis H berdasar kondisi X

P(H) = Probabilitas hipotesis H (prior probability)

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probalitas X

Salah satu penelitian sebelumnya tentang *Naïve Bayes* adalah mendeskripsikan *Naïve Bayes* sebagai algoritma klasifikasi probabilistik yang sangat efesien dalam melakukan analisis sentiment pada teks. Algoritma ini berasaskan pada teorema Bayes, suatu metode untuk memperkirakan kemungkinan terjadinya suatu kejadian berdasarkan data yang ada (Fadhilah Az-Haari dkk., 2024)

Sederhananya, metode klasifikasi *Naïve Bayes* menghitung semua kemungkinan dengan menggabungkan jumlah kombinasi dan kerapatan nilai dari suatu dataset yang sudah di peroleh(Darmawan dkk., 2023). Model algoritma klasifikasi *Naïve Bayes* sudah sering digunakan untuk melakukan klasifikasi data menggunakan perhitungan probabilistic. Sentimen komputasi positif dan negatif dapat dihasilkan dari perhitungan probabilitas untuk setiap kata dalam dokumen atau dataset pada pembobotan kata TF-IDF.

Naïve Bayes merupakan metode yang tidak memiliki aturan dan menggunakan cabang matematika yang disebut teori probabilitas untuk mendapatkan peluang setinggi mungkin dengan melihat frekuensi atau jumlah kemunculan setiap klasifikasi dalam data pelatihan. Dalam pengembangan basis datanya, *Naïve Bayes* melibatkan jenis pembelajaran mesin yang membutuhkan sampel sebagai data label pelatihan dan menggunakan metode ini untuk mendapatkan data label pelatihan. dibagi menjadi dua bagian, yaitu Klasifikasi dan regresi: Variabel diklasifikasikan menjadi kategori, misalnya panas atau dingin, sakit atau tidak sakit(Khoirul dkk., 2023).Berdasarkan teorema Bayes, yang diusulkan oleh ilmuwan inggris Thomas Bayes, algoritma *Naïve Bayes* adalah metode klasifikasi probabilistic sederhana yang menganggap bahwa setiap kondisi dan peristiwa memiliki tingkat kemandirian yang tinggi. Kemampuannya untuk membuat parameter klasifikasi dengan menggunakan data pelatihan adalah salah satu keunggulan algoritma *Naïve Bayes* yang relatif kecil, meskipun metode operasinya sederhana, kinerjanya Salah satu keunggulan algoritma *Naïve Bayes*, sebanding dengan algoritma lain seperti pohon keputusan dan jaringan saraf C, adalah kemampuannya untuk menghasilkan parameter klasifikasi hanya dengan

menggunakan sejumlah kecil data pelatihan; ini membuatnya pilihan yang lebih baik untuk kumpulan data yang besar(Gaja Rismanda dkk., 2023).

Metode ini digunakan untuk memprediksi kelas data berdasarkan probabilitasnya dengan menghitung kemungkinan bahwa data akan termasuk dalam kelas tertentu. *Naïve Bayes* terbukti sangat efektif dalam banyak aplikasi, terutama dalam pengolahan teks seperti klasifikasi spam, analisis sentimen, dan penyederhanaan bahasa. Namun, asumsi independensi fitur ini jarang benar sepenuhnya. Kemampuannya untuk bekerja dengan cepat pada dataset besar dan memberikan hasil yang akurat dengan persyaratan komputasi yang relatif rendah adalah keunggulan utama *Naïve Bayes*. Algoritma ini mampu secara otomatis menentukan kategori yang paling mungkin untuk data baru dengan menghitung probabilitas masing-masing kelas berdasarkan fitur yang ada.

Algoritma *Naïve Bayes* untuk klasifikasi didasarkan pada Teorema Bayes dan menganggap secara naive (sederhana) bahwa setiap sifat atau ciri dalam data tidak bergantung satu sama lain terhadap kelas yang diprediksi. Terlepas dari kenyataan bahwa asumsi ini jarang benar dalam kehidupan nyata, algoritma ini masih sangat efektif dalam berbagai tugas klasifikasi teks, seperti analisis sentimen, deteksi spam, dan kategorisasi dokumen. Untuk melakukan klasifikasi dibantu menggunakan model *Multinomial Naïve Bayes*, yang merupakan pengembangan dari algoritma Bayes.

2.2.6 *Multinomial Naïve Bayes*

Model *Multinomial Naïve Bayes*, yang merupakan pengembangan dari algoritma Bayes, digunakan untuk mengklasifikasikan teks atau dokumen. Rumus klasifikator Multinomial *Naïve Bayes* menghitung kedua jumlah kemunculan kata untuk menentukan kelas dokumen(Ramadhan dkk., 2023).

$$p\ (c,d) = \frac{N_c}{N} \text{x P}(t_{1,}\ \text{c})\text{x} \ldots \text{x P}(t_{n,}c)$$

Keterangan:

$p\ (c,d)$: Probabilitas suatu dokumen termasuk kelas c

N_c : Jumlah kelas c pada seluruh dokumen

N : Jumlah seluruh dokumen

t_n : Kata dokumen d ke-n

$p\ (t_{n,}c)$:Probabilitas kata ke-n dengan diketahui kelas c

Rumus probabilitas kata ke-n yang digunakan dengan pembobotan kata TF-IDF dapat dilihat pada keterangan dibawah ini.

$$P\ (t_{n,}c) = \frac{Wct + 1}{(\sum W'€VW'ct + B'}$$

Keterangan:

$W\ ct$: Nilai pembobotan tfidf ayau W dari term t di kategori c

$W'€VW'ct$:Jumlah total W dari keseluruhan term pada kategori c

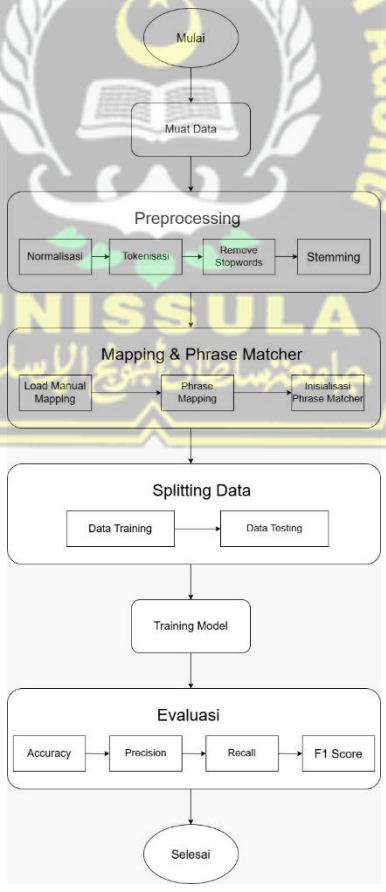
B' : Jumlah W kata unik (nilai IDF tidak dikali dengan tf) pada seluruh dokumen

Salah satu metode pembelajaran probabilistic yang digunakan dalam pengolahan *Natural Language Processing* (NLP) adalah algoritma *multinomial Naïve Bayes* yang didasarkan pada teorema Bayes(Yuyun dkk., 2021). Algoritma ini bergantung pada konsep "frekuensi", yaitu berapa kali kata tersebut muncul dalam sebuah dokumen. Model ini menjelaskan dua fakta: apakah kata tersebut muncul atau tidak, dan frekuensi kemunculannya dalam dokumen. Multinomial *Naïve Bayes* adalah model *Naïve Bayes* yang paling umum digunakan dalam klasifikasi teks.

BAB III
METODE PENELITIAN

3.1 Metode Penelitian

Penelitian ini memanfaatkan pendekatan klasifikasi otomatis untuk menyederhanakan teks bacaan yang ditujukan bagi anak-anak disleksia. Algoritma yang digunakan dalam penelitian ini adalah *Multinomial Naïve Bayes*, yang digunakan untuk mengklasifikasikan tipe penyederhanaan teks kedalam dua kategori yaitu leksikal dan sintaksis. Proses penyederhanaan dilakukan dengan mengganti kata atau frasa yang sulit menjadi bentuk yang lebih mudah dipahami, berdasarkan padanan kata yang telah disusun secara manual maupun dengan pendekatan *Phrase Matching*. Langkah-langkah yang harus dilakukan dalam penelitian ini adalah sebagai berikut:



Gambar 3. 1 Flowchart Penelitian

3.1.1 Pengumpulan Data

Pada tahap awal penelitian ini, proses pengumpulan data dilakukan terhadap teks bacaan anak yang akan dianalisis tingkat kesederhanaannya. 500 kalimat teks yang digunakan berasal dari berbagai sumber, termasuk buku bacaan anak sekolah dasar, artikel pendidikan anak, dan konten pembelajaran dasar dari situs web edukatif yang tersedia secara online. Sumber-sumber ini dipilih karena memiliki gaya bahasa yang relevan untuk anak-anak, memiliki variasi dalam struktur kalimat dan tingkat kompleksitas kosakata yang berbeda. Setelah data dikumpulkan, setiap teks diberi label manual sebagai "Sintaksis" atau "Leksikal". Kriteria yang digunakan untuk pelabelan ini adalah kesederhanaan bahasa. Kriteria ini termasuk penggunaan kosakata umum, struktur kalimat yang sederhana, dan kemudahan pembacaan bagi anak-anak, terutama mereka yang mengalami disleksia. Proses ini sangat penting untuk memastikan bahwa data yang digunakan untuk pelatihan dan pengujian model klasifikasi dengan algoritma *Naïve Bayes* tepat dan memenuhi kebutuhan pengguna akhir.

3.1.2 Preprocessing Teks

Preprocessing teks adalah langkah penting dalam *Natural Language Processing* (NLP). Tujuan dari langkah ini adalah untuk membersihkan dan menyiapkan data teks sehingga model pembelajaran mesin dapat lebih mudah menganalisisnya. Beberapa tahapan utama preprocessing dalam penelitian ini termasuk setword, tokenisasi, dan stemming.

1. Tokenisasi

   Tokenisasi, proses membagi teks yang dibaca menjadi bagian yang lebih kecil, seperti kalimat, adalah langkah pertama dalam praproses data. Ini dilakukan untuk membuat analisis dan pemrosesan data lebih mudah.

2. Stemming

   Stemming dilakukan untuk mengubah kata menjadi dasar atau akar katanya. Misalnya “memahami” menjadi “paham.

3. Stopword Removal

   Tahapan preprocessing ini menghasilkan data teks yang lebih bersih dan terstruktur. Ini memungkinkan model klasifikasi untuk lebih baik mendeteksi dan menyederhanakan teks bacaan secara lebih akurat.

3.1.3 Pemberian Label

Label berfungsi sebagai penanda kategori untuk setiap teks yang digunakan, sehingga tahap pemberian label merupakan langkah penting dalam membangun model klasifikasi. Dalam penelitian ini, setiap teks yang dibaca setelah proses preprocessing akan diberi label "Sintaksis" atau "Leksikal". Teks yang diberi label "Sederhana" memiliki struktur kalimat yang mudah dipahami, kosakata umum, dan sesuai dengan kemampuan pemahaman anak disleksia. Namun, label "Tidak Sederhana" digunakan untuk teks yang memiliki struktur kalimat yang kompleks, penggunaan kata-kata yang sulit atau teknis, dan makna yang cenderung ambigu atau sulit dipahami. Kriteria kebahasaan dan kesesuaian tingkat keterbacaan digunakan untuk melakukan proses pelabelan ini secara manual. Tahap ini akan menghasilkan data latih dan data uji yang akan digunakan untuk mempelajari pola bahasa yang membedakan teks sederhana dan tidak sederhana.

3.1.4 Pembagian Data

Data kemudian dibagi menjadi dua bagian utama setelah proses pelabelan selesai: data latihan (training data) dan data uji (testing data). Tujuan dari bagian ini adalah untuk membedakan data yang digunakan untuk membangun model (melatih algoritma) dari data yang digunakan untuk menguji atau mengevaluasi kinerja model yang dilatih. Pembagian data ini biasanya dilakukan secara rasional; misalnya, dua puluh persen untuk data uji dan delapan puluh persen untuk data latih. Algoritma *Naïve Bayes* dapat membedakan teks sederhana dan tidak sederhana dengan menggunakan data latih untuk mengenalkan pola dan karakteristik teks. Sementara itu, data uji digunakan untuk mengevaluasi kemampuan model untuk mengklasifikasikan data baru yang belum pernah dilihat sebelumnya. Pembagian data yang tepat sangat penting untuk memastikan bahwa model tidak hanya berfungsi dengan baik pada data pelatihan tetapi juga dapat generalisasi dengan baik pada data di luar pelatihan.

3.1.5 Penerapan Phrase Matcher

Setelah proses klasifikasi dilakukan menggunakan algoritma *Naïve Bayes*, sistem akan menentukan tipe penyederhanaan dari suatu teks, yaitu leksikal atau sintaksis. Untuk menyederhanakan teks berdasarkan hasil klasifikasi tersebut, penelitian ini menerapkan teknik *Phrase Matcher* dari libary SpaCy. *Phrase Matcher* digunakan untuk mencocokkan kata atau frasa sulit dalam teks dengan daftar padanan kata yang telah disiapkan sebelumnya, baik dari dataset maupun file manual mapping. Pola-pola frasa tersebut ditambahkan ke *Phrase Matcher* dan digunakan untuk mendeteksi bagian-bagian teks yang perlu disederhanakan. Jika ditemukan kecocokan, sistem secara otomatis mengganti kata atau frasa tersebut dengan bentuk yang lebih mudah dipahami oleh anak disleksia. Pada tipe penyederhanaan leksikal, *Phrase Matcher* digunakan untuk menggantikan kata sulit satu per satu berdasarkan padanan yang tersedia. Sedangkan pada tipe sintaksis, sistem akan terlebih dahulu memeriksa apakah terdapat hasil penyederhanaan langsung dari mapping berbasis kalimat. Jika tidak ditemukan, *Phrase Matcher* tetap digunakan sebagai alternatif untuk menyederhanakan bagian tertentu dari teks. Dengan cara ini, *Phrase Matcher* berperan sebagai komponen kunci dalam menghasilkan teks sederhana yang sesuai dengan tipe penyederhanaan, sekaligus menjaga keterbacaan dan makna utama dari teks asli.

3.1.6 Training Model *Naïve Bayes*

Pada tahap pelatihan (training) model, algoritma *Naïve Bayes,* tepatnya jenis *Multinomial Naïve Bayes,* digunakan untuk membangun model klasifikasi dari data latihan yang telah melalui proses pra-pemrosesan. Semua teks pada data latih diubah menjadi representasi numerik menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), yang mengukur seberapa penting suatu kata dalam dokumen dan korpus. *Multinomial Naïve Bayes* merupakan varian dari *Naïve Bayes* yang dirancang khusus untuk menangani data dalam bentuk frekuensi atau bobot kata, sehingga sangat cocok untuk klasifikasi teks. Model ini bekerja dengan menghitung kemungkinan suatu kata muncul dalam masing-masing kelas label, yaitu sintaksis dan leksikal, berdasarkan distribusi kata pada data latih. Proses pelatihan ini menggunakan prinsip Teorema Bayes, di mana sistem menghitung

probabilitas suatu teks masuk ke salah satu kelas dengan asumsi bahwa tiap fitur (dalam hal ini kata-kata) dianggap saling independen. Asumsi inilah yang membuat metode ini disebut sebagai “Naïve”. Hasil dari proses ini adalah sebuah model klasifikasi yang mampu mengenali pola kemunculan kata dan kemudian memprediksi tipe penyederhanaan dari suatu teks baru, apakah termasuk kategori leksikal atau sintaksis.

3.1.7 Evaluasi Model

Dalam penelitian ini, evaluasi model dilakukan untuk mengukur kinerja algoritma klasifikasi yang digunakan. Empat metrik utama digunakan: akurasi, *precision, recall, dan F1-score*. Akurasi mengukur persentase prediksi yang benar dari keseluruhan data yang diuji, memberikan gambaran umum seberapa baik model mengenali data secara keseluruhan. Precision mengukur ketepatan prediksi positif, yaitu seberapa banyak dari hasil prediksi positif yang benar atau relevan. Sementara recall menilai kemampuan model dalam menemukan semua data positif yang sebenarnya akurat. F1-Score untuk menganalisis kualitas prediksi. Dengan menggunakan keempat metrik ini, evaluasi model menjadi lebih komprehensif baik dari segi ketepatan maupun kelengkapan hasil prediksi. Selanjutnya untuk *Confusion Matrix* untuk memahami kesalahan dan kinerja model secara detail. Dengan demikian, menjadi lebih mudah untuk menilai keandalan model pada tugas klasifikasi teks bacaan yang diberikan kepada anak-anak yang didiagnosis dengan disleksia.

BAB IV
HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada bagian ini, peneliti membahas mengenai yang di dapatkan setelah melalui tahap metode penelitian sebagai berikut:

4.1.1 Pengumpulan Data

Proses pengumpulan dataset pada penelitian ini dilakukan secara manual karena belum tersedia dataset yang sesuai dengan kebutuhan penelitian. Dataset ini disusun dengan cara mereview beberapa buku anak, dan majalah anak. Kemudian memilih beberapa kalimat yang sulit bagi anak-anak untuk memahami disleksia yang berusia 10-12 tahun.

Tahapan pengumpulan data dilakukan sebagai berikut:

a. Mengumpulkan beberapa buku anak, majalah anak, atau juga bisa dari *E-Book*
b. Mereview buku anak dan majalah anak, proses pengumpulan dimulai dengan mereview kalimat yang sulit dipahami. Setiap kalimat memuat pemahaman sulit dan solusinya dicatat untuk membuat dataset.

Tabel 4. 1 Contoh Pengumpulan Data

NO	Teks Asli	Kata Sulit
1	Benar, Kanda. Semoga ia menjadi putra yang baik	Kanda, ia, putra
2	Ia sadar bahwa tingkah lakunya kemarin tidak seperti anak yang mandiri	Tingkah laku, mandiri
3	Buku cerita itu penuh dengan petualangan yang menarik	petualangan

4	Dia merencanakan untuk pergi berlibur ke tempat yang jauh	Merencanakan, berlibur
5	iklim di samarinda ini cukup panas, pada dasarnya beriklim tropis	iklim; pada dasarnya; beriklim tropis

4.1.2 Preprocessing Teks

Beberapa tahapan utama preprocessing dalam penelitian ini termasuk normalisasi, tokenisasi, stopword removal, dan stemming. menunjukkan langkah-langkah awal pengolahan teks dengan Python, terutama untuk teks berbahasa Indonesia. Normalisasi, Stopword remover, Steming, dan Tokenizer adalah empat komponen penting yang digunakan. Pertama, objek stopword_factory digunakan untuk mengakses daftar kata umum, atau stopwords, seperti "yang", "di", "ke", dan sebagainya, yang biasanya dihapus karena tidak penting. Meskipun demikian, kata-kata seperti "itu" dan "ini" disimpan karena dianggap penting untuk menjaga makna kalimat tetap jelas. Oleh karena itu, objek stemmer dibuat untuk mengubah kata ke bentuk aslinya, seperti "berlari" menjadi "berlari", sehingga analisis teks menjadi lebih konsisten. Terakhir, tokenizer digunakan untuk memecah teks menjadi kata-kata terpisah dengan menggunakan pola tertentu, hanya mengambil kata yang terdiri dari huruf dan angka. Proses ini membantu mempersiapkan teks sebelum dianalisis lebih lanjut.

4.1.3 *Mapping* dan *Phrase Matcher*

Pada tahap ini, sistem melakukan proses pemetaan (mapping) dan deteksi otomatis terhadap kata-kata sulit dalam teks bacaan, khususnya pada tipe penyederhanaan leksikal. Tujuannya adalah untuk membuat padanan kata yang lebih mudah dipahami oleh anak-anak dengan disleksia dengan mengganti kata atau frasa yang sulit dengan padanan kata yang lebih mudah dipahami. Proses ini dilakukan secara sistematis, dimulai dengan pemuatan data padanan kata dan kemudian memanfaatkan PhraseMatcher dari spaCy untuk pencocokan otomatis. Berikut ini merupakan penjelasan dari alur kerja Mapping dan *Phrase Matcher*:

a. Load Manual Mapping

Terlebih dahulu, sistem memuat file padanan kata yang disusun secara manual dalam file manual_mapping.txt. File ini berisi pasangan kata yang sulit dan kata pengganti yang lebih mudah dipahami, seperti "kanda: kakak" dan "putra: anak". Data dibaca satu per satu sebelum dimasukkan ke dalam struktur dictionary mapping_manual, yang akan digunakan dalam proses penyederhanaan.

b. Phrase Mapping

Sistem menggabungkan data manual_mapping dan dataset untuk membangun struktur phrase_mapping setelah data dimuat. Sistem membaca seluruh baris data yang memiliki tipe penyederhanaan leksikal. Setelah normalisasi dan tokenisasi teks asli dan teks sederhana dari masing-masing baris, sistem membandingkan kata-kata yang sulit dengan padanannya. Jika ada perbedaan dalam bentuk atau posisi kata, pasangan tersebut dimasukkan ke dalam phrase_mapping. Selain itu, sistem juga menambahkan data dari manual_mapping ke phrase_mapping jika belum ada. Ini meningkatkan jumlah pasangan kata sederhana dan sulit.

c. Inisialisasi Phrase Matcher

Pada tahap terakhir, librari spaCy memulai PhraseMatcher. Ini dilakukan dengan mengubah semua kata yang ada dalam phrase_mapping menjadi objek dokumen spaCy (nlp.make_doc), dan kemudian menambahkan pola pencocokan ke dalam PhraseMatcher dengan label "KataSulit". PhraseMatcher ini akan digunakan untuk mencocokkan kata sulit yang muncul dalam teks masukan dan menggantinya dengan kata yang lebih sederhana sesuai dengan phrase_mapping

4.1.4 Vektorisasi TF-IDF

Setelah data teks dibersihkan Selanjutnya, teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk mengubah data teks menjadi data numerik. TfidfVectorizer dari scikit-learn digunakan untuk menyelesaikan tugas

ini. Itu dikonfigurasi untuk memproses satu katagram dengan parameter ngram_range=(1,1). Berdasarkan bobot TF-IDF-nya, hanya 250 kata paling penting yang dapat dihitung berdasarkan jumlah fitur tertinggi. Vektor yang dihasilkan dari transformasi ini menggambarkan setiap dokumen sebagai matriks berdimensi [jumlah dokumen x jumlah fitur], dengan bobot penting kata terhadap dokumen dan seluruh korpus ditunjukkan oleh setiap elemen. Selanjutnya, hasil vektorisasi disimpan dalam variabel X_tfidf. Variabel ini akan digunakan sebagai input utama untuk pelatihan model klasifikas dan tahap pemisahan data (splitting).

4.1.5 Splitting Data

Setelah melalui tahapan preprocessing dan vektorisasi menggunakan TF-IDF, langkah selanjutnya dalam penelitian ini adalah melakukan pembagian data (splitting data) ke dalam dua kelompok utama, yaitu data training dan data testing. Tujuan dari pembagian ini adalah untuk memastikan bahwa model klasifikasi Multinomial *Naïve Bayes* yang dibangun dapat dilatih pada sebagian data dan diuji performanya pada data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi menjadi lebih akurat dan tidak biasa. Pembagian data dilakukan dengan menggunakan fungsi train_test_split dari library Scikit-learn, di mana data dibagi dengan perbandingan 70% untuk pelatihan (training) dan 30% untuk pengujian (testing). Dalam proses ini, digunakan juga parameter stratify untuk memastikan distribusi label antara dua kelas penyederhanaan, yaitu *leksikal* dan *sintaksis*, tetap proporsional di kedua subset data. Dengan ini, potensi ketidakseimbangan data yang bisa mempengaruhi akurasi model dapat diminimalisasi. Dari hasil pemisahan ini diperoleh empat buah variabel utama, yaitu X_train dan y_train sebagai input dan label data pelatihan, serta X_test dan y_test sebagai input dan label data pengujian. Data X_train dan X_test berisi representasi teks yang telah dikonversi menjadi bentuk vektor numerik menggunakan metode TF-IDF, sedangkan y_train dan y_test berisi label klasifikasi berupa 0 (untuk sintaksis) dan 1 (untuk leksikal).

4.1.6 Training Model

Setelah data dipecah menjadi dua bagian, yaitu data latih (X_train, y_train) dan data uji (X_test, y_test), tahap selanjutnya adalah melakukan pelatihan model dengan algoritma *Multinomial Naïve Bayes.* Model ini digunakan karena sesuai

untuk klasifikasi teks yang mewakili data dalam bentuk frekuensi kata atau skor TF-IDF. Dalam implementasinya, model dilatih dengan memanggil metode.fit() menggunakan data latih. Parameter alpha diset sebesar 0.05 untuk menerapkan teknik *Laplace smoothing*, yaitu metode untuk menangani kemungkinan nilai probabilitas nol jika suatu fitur tidak muncul pada salah satu kelas. Setelah proses pelatihan selesai, model kemudian digunakan untuk melakukan prediksi terhadap data pelatihan dan data pengujian melalui metode. predict(). Hasil prediksi ini kemudian dibandingkan dengan label sebenarnya untuk mengukur sejauh mana kemampuan model dalam mengklasifikasikan jenis penyederhanaan teks. Prediksi dikategorikan ke dalam dua label yaitu 0 untuk sintaksis dan 1 untuk leksikal, sesuai dengan anotasi label yang telah disiapkan pada dataset. Tahap ini menjadi penting untuk mengetahui apakah model mampu mengenali pola teks yang telah disederhanakan dan dapat membedakan antara perubahan pada tingkat kata (leksikal) dan perubahan pada struktur kalimat (sintaksis).

4.1.7 Evaluasi

Setelah proses pelatihan model selesai, langkah berikutnya dalam penelitian ini adalah melakukan evaluasi terhadap kinerja model klasifikasi yang telah dibangun. Tujuan dari evaluasi ini adalah untuk mengukur kemampuan model *Multinomial Naïve Bayes* untuk mengenali pola penyederhanaan teks berdasarkan dua kategori, yaitu leksikal dan sintaksis, secara akurat dan andal. Dengan melakukan evaluasi ini, peneliti dapat menilai apakah model sudah cukup baik untuk digunakan untuk membantu anak-anak dengan disleksia memahami teks dengan lebih mudah. Untuk menguji kinerja model, data yang telah diproses melalui TF-IDF selanjutnya dibagi menjadi dua bagian: data latih (training data) dan data uji (testing data) menggunakan metode train_test_split. Model dibuat dengan data latih, dan data uji digunakan untuk menilai seberapa baik model dapat beradaptasi dengan data yang belum pernah dilihat sebelumnya. Model Multinomial *Naïve Bayes* kemudian dilatih pada data latih dan digunakan untuk memprediksi hasil klasifikasi pada data uji. Evaluasi dilakukan menggunakan beberapa metrik standar, yaitu: Akurasi*, Precision, Recall, F1-Score*. Selain metrik

numerik, digunakan pula *confusion matrix* untuk menggambarkan secara visual jumlah prediksi benar dan salah yang dilakukan oleh model terhadap kedua kelas.

Berdasarkan hasil prediksi yang dilakukan pada data uji, diperoleh beberapa metrik evaluasi yang menggambarkan performa dari model klasifikasi. Pada tabel 4.1 menyajikan nilai akurasi, precision, recall, dan f1-score dari model *Multinomial Naïve Bayes* terhadap dua kelas, yaitu sintaksis dan leksikal. Nilai-nilai ini menjadi dasar untuk menilai sejauh mana model dapat mengenali dan membedakan tipe penyederhanaan teks secara efektif dan konsisten.

Tabel 4. 2 Hasil Precision, Recall, dan F1-Score

	precision	recall	F1-Score	support
sintaksis	0.88	0.91	0.90	443
leksikal	0.92	0.88	0.90	475
accuracy			0.90	918
Macro avg	0.90	0.90	0.90	918
Weighted avg	0.90	0.90	0.90	918

Tabel 4.2 menunjukkan hasil evaluasi performa model klasifikasi yang ditampilkan dalam tabel di atas, dapat disimpulkan bahwa algoritma *Multinomial Naïve Bayes* menunjukkan kinerja yang sangat baik dalam membedakan antara tipe penyederhanaan sintaksis dan leksikal. Nilai *precision*, *recall*, dan *F1-score* untuk kedua kelas berada pada angka 0.88 hingga 0.92, yang mencerminkan tingkat keakuratan model dalam melakukan prediksi. Akurasi keseluruhan model mencapai 90%, yang berarti 9 dari 10 prediksi yang dihasilkan sesuai dengan label sebenarnya. Nilai rata-rata makro dan rata-rata tertimbang (weighted average) juga menunjukkan angka yang sama, yaitu 0.90, yang mengindikasikan bahwa model bekerja secara konsisten dan seimbang meskipun jumlah data antar kelas tidak sama persis. Hal ini menegaskan bahwa metode klasifikasi otomatis yang digunakan dalam penelitian ini dapat diandalkan sebagai dasar untuk proses penyederhanaan teks untuk anak-anak yang didiagnosis dengan disleksia.

Selain itu, Sebagai bagian dari tahapan evaluasi model klasifikasi, penggunaan *confusion matrix* atau matriks kebingungan menjadi salah satu metode yang penting untuk memberikan gambaran menyeluruh terhadap kinerja prediktif

dari algoritma yang diterapkan. Dalam penelitian ini, *confusion matrix* digunakan untuk mengevaluasi hasil klasifikasi model *Multinomial Naïve Bayes* terhadap dua kelas target, yaitu tipe penyederhanaan *sintaksis* dan *leksikal*. Melalui *confusion matrix*, peneliti dapat mengetahui seberapa banyak data yang berhasil diklasifikasikan secara tepat oleh model, serta seberapa besar kesalahan klasifikasi yang terjadi. Selain menunjukkan akurasi keseluruhan, evaluasi ini menunjukkan pola kesalahan klasifikasi antar kelas secara lebih mendalam. Oleh karena itu, evaluasi ini dapat digunakan sebagai dasar untuk analisis lebih lanjut terhadap kekuatan dan kelemahan model. Confusion matrix digambarkan dalam skema warna biru dibawah ini

Gambar 4. 1 Confusion Matrix

Gambar 4.1 diketahui bahwa model berhasil mengklasifikasikan sebanyak 405 data dengan tipe sintaksis secara benar, dan 418 data bertipe leksikal juga diklasifikasikan dengan benar. Hal ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengenali kedua jenis penyederhanaan. Namun demikian, masih terdapat 38 data bertipe sintaksis yang salah diklasifikasikan sebagai leksikal, serta 57 data bertipe leksikal yang keliru diprediksi sebagai sintaksis. Ada kemiripan pola antar teks dari dua kelas tersebut

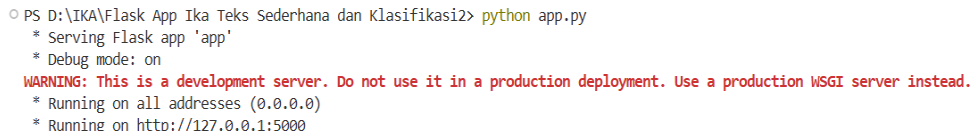
atau batasan representasi fitur yang dihasilkan dari proses vektorisasi TF-IDF yang dapat menyebabkan kesalahan prediksi ini. Meskipun demikian, prediksi yang tepat lebih banyak dibandingkan prediksi yang salah, sehingga dapat disimpulkan bahwa model secara keseluruhan bekerja dengan baik dalam klasifikasi. Dalam konteks klasifikasi teks dengan dua label, tingkat kesalahan yang muncul juga masih berada dalam batas yang wajar. Oleh karena itu, hasil ini mendukung bahwa algoritma *Multinomial Naïve Bayes* cukup efektif dalam melakukan klasifikasi tipe penyederhanaan teks untuk membantu anak-anak dengan disleksia memahami bacaan dengan lebih baik.

4.1.8 Implementasi

Tahap implementasi merupakan langkah akhir dalam pengembangan sistem klasifikasi penyederhanaan teks, yang bertujuan untuk menguji penerapan model dalam skenario nyata. Implementasi ini dilakukan dengan memanfaatkan fungsi simplify_text() untuk memproses sejumlah kalimat baru di luar data pelatihan. Sistem menggunakan model *Multinomial Naïve Bayes* yang telah dilatih sebelumnya untuk memprediksi jenis penyederhanaan teks setelah menerimanya. Berdasarkan prediksi tersebut, sistem akan melakukan proses penyederhanaan sesuai dengan jenisnya. Hasil dari proses ini mencakup teks asli, hasil prediksi, teks sederhana, kata sulit yang terdeteksi, dan padanan kata yang digunakan sebagai penyederhana. Semua hasil implementasi disimpan dalam file hasil_penyederhanaan.csv sebagai dokumentasi dan evaluasi akhir. Tahapan ini menunjukkan bagaimana sistem yang dikembangkan dapat diterapkan dan memberikan manfaat nyata, terutama dalam membantu anak-anak dengan disleksia memahami bacaan lebih mudah.

Setelah proses pelatihan dan evaluasi model klasifikasi teks berhasil dilakukan, tahap selanjutnya adalah menjalankan aplikasi agar dapat digunakan secara langsung oleh pengguna. Untuk itu, eksekusi dilakukan melalui terminal dengan menjalankan perintah python app.py, yang akan memulai server lokal berbasis framework Flask. Saat server berjalan, program dapat diakses melalui browser melalui alamat URL berikut: http://127.0.0.1:5000. Tampilan ini menunjukkan bahwa sistem telah diuji secara langsung; pengguna dapat mencoba

fitur klasifikasi dan penyederhanaan teks secara instan melalui antarmuka aplikasi, meskipun masih dalam mode pengembangan (development mode).



Gambar 4. 2 Tampilan Terminal

Gambar 4.2 menunjukkan proses eksekusi aplikasi web berbasis Flask melalui terminal menggunakan perintah python app.py. Perintah ini digunakan untuk menjalankan file utama yang berisi kode backend aplikasi klasifikasi dan penyederhanaan teks. Setelah perintah dijalankan, terminal menampilkan beberapa informasi penting. Pertama, sistem menampilkan bahwa aplikasi Flask berhasil dilayani dengan nama 'app'. Selain itu, mode debug diaktifkan (Debug mode: on), yang memungkinkan pengembang untuk melihat pesan kesalahan secara langsung jika terjadi kendala selama pengujian. Terdapat pula peringatan dari Flask bahwa server yang digunakan adalah development server, bukan server untuk keperluan produksi. Hal ini merupakan hal wajar dalam tahap pengembangan aplikasi. Flask kemudian memulai server lokal pada alamat http://127.0.0.1:5000, yang dapat diakses melalui browser untuk melihat dan mencoba langsung fitur-fitur aplikasi yang telah dikembangkan. Alamat tersebut menunjukkan bahwa server berjalan pada mode lokal (localhost), sehingga hanya bisa diakses dari perangkat pengembang. Gambar ini mengilustrasikan keberhasilan inisialisasi server lokal sebagai bagian dari proses implementasi sistem klasifikasi teks.

Setelah server lokal berhasil dijalankan melalui terminal, langkah selanjutnya dalam proses implementasi adalah mengakses aplikasi melalui peramban web. Akses ini dilakukan dengan membuka alamat http://127.0.0.1:5000, yang merupakan endpoint default dari server Flask yang sedang aktif. Pada tahap ini, antarmuka pengguna (*user interface*) dari sistem klasifikasi dan penyederhanaan teks ditampilkan secara visual, memungkinkan pengguna untuk langsung mencoba fungsionalitas yang telah dikembangkan. Antarmuka ini berfungsi sebagai jembatan antara pengguna dan sistem klasifikasi yang dibangun

menggunakan pembelajaran mesin. Selain itu, itu merupakan representasi nyata dari hasil implementasi penelitian dalam bentuk aplikasi interaktif berbasis web.

Gambar 4. 3 Tampilan Halaman Awal

Gambar 4.3 menampilkan tampilan antarmuka utama dari aplikasi berbasis web yang telah dikembangkan. Aplikasi ini dirancang dengan pendekatan antarmuka yang sederhana dan ramah pengguna, sehingga mudah diakses dan digunakan oleh siapa saja, termasuk orang tua, guru, maupun pendamping anak dengan disleksia. Pada bagian atas halaman, ditampilkan informasi identitas berupa nama pengembang, NIM, serta judul tugas akhir, yang mencerminkan konteks penelitian yang dilakukan. Di bagian tengah halaman, terdapat formulir input dengan label "Masukkan Teks" yang memungkinkan pengguna untuk menyalin atau mengetikkan kalimat yang ingin disederhanakan. Setelah memasukkan teks, pengguna dapat mengklik tombol "Sederhanakan" untuk memproses teks tersebut. Tombol ini akan memicu backend aplikasi untuk menjalankan model klasifikasi dan penyederhanaan teks berdasarkan algoritma *Naïve Bayes* yang telah dilatih sebelumnya. Tampilan antarmuka ini tidak hanya mendemonstrasikan hasil implementasi sistem klasifikasi, tetapi juga menunjukkan penerapan nyata dari penelitian dalam bentuk aplikasi yang fungsional dan interaktif. Dengan adanya aplikasi ini, proses klasifikasi dan penyederhanaan teks menjadi lebih mudah, cepat, dan dapat digunakan secara langsung oleh pengguna akhir yang membutuhkan dukungan dalam memahami teks bacaan.

Setelah antarmuka utama berhasil dimuat, pengguna dapat langsung menguji sistem dengan memasukkan teks bacaan ke dalam kolom input yang tersedia. Pada tahap ini, pengguna diberikan kebebasan untuk mencoba berbagai

bentuk kalimat, baik yang kompleks maupun yang mengandung kata sulit. Tombol "Sederhanakan" di bagian bawah berfungsi untuk mengirimkan teks ke dalam sistem agar diproses menggunakan model klasifikasi dan teknik penyederhanaan yang telah dilatih sebelumnya. Proses ini mencerminkan implementasi nyata dari tujuan penelitian, yaitu memberikan kemudahan akses terhadap bacaan yang ramah bagi anak-anak dengan disleksia. Tampilan ini menjadi titik interaksi antara pengguna dan sistem, sekaligus memperlihatkan bagaimana model dapat diintegrasikan dalam aplikasi berbasis web yang sederhana namun fungsional.

Gambar 4. 4 Tampilan Input Teks

Gambar 4.4 menunjukkan antarmuka awal dari aplikasi sistem klasifikasi dan penyederhanaan teks yang dirancang khusus untuk membantu anak-anak dengan disleksia. Pada tampilan ini, terdapat kolom input tempat pengguna dapat memasukkan kalimat atau paragraf yang ingin diproses. Desain antarmuka dirancang dengan sederhana dan intuitif agar mudah digunakan oleh berbagai kalangan, baik guru, orang tua, maupun tenaga pendidik lainnya yang mendampingi anak dengan kebutuhan khusus. Fokus utama dari tampilan ini adalah area input teks yang jelas terlihat di tengah halaman, dengan label "Masukkan Teks:" sebagai petunjuk bagi pengguna. Setelah pengguna mengetikkan teks ke dalam kolom tersebut, mereka dapat langsung menekan tombol "Sederhanakan" untuk memulai proses klasifikasi dan penyederhanaan. Proses ini akan berjalan menggunakan model yang telah dilatih sebelumnya dengan algoritma *Naïve Bayes*, yang akan menentukan apakah teks termasuk kategori leksikal atau sintaksis, kemudian menyederhanakannya sesuai kategori tersebut. Contoh input dalam gambar yaitu "Buku cerita itu penuh dengan petualangan yang menarik" menunjukkan bahwa sistem dapat menerima kalimat dalam bahasa Indonesia yang kompleks atau

mengandung kata-kata tidak umum, untuk kemudian dianalisis dan disesuaikan tingkat keterbacaannya.

Setelah proses input teks dilakukan, sistem akan menjalankan tahap penyederhanaan dan klasifikasi secara otomatis. Pada tahap ini, pengguna dapat melihat hasil lengkap dari proses yang telah dijalankan oleh sistem. Hasil tersebut mencakup berbagai informasi penting yang disusun secara sistematis untuk memudahkan pemahaman pengguna terhadap perubahan yang dilakukan pada teks. Karena menunjukkan cara sistem menyederhanakan dan mengklasifikasikan kalimat yang dimasukkan, tampilan ini sangat penting untuk proses implementasi.

Gambar 4. 5 Tampilan Output Sistem

Gambar 4.5 menunjukkan tampilan hasil dari proses penyederhanaan teks yang dilakukan oleh sistem. Setelah pengguna memasukkan kalimat ke dalam kolom input dan menekan tombol “Sederhanakan,” sistem akan secara otomatis menampilkan hasil penyederhanaan teks berdasarkan model klasifikasi yang telah dilatih sebelumnya. Pada tampilan ini terlihat bahwa teks asli yang dimasukkan pengguna diubah menjadi bentuk teks yang lebih sederhana agar lebih mudah

dipahami, khususnya oleh anak-anak dengan disleksia. Selain menampilkan teks hasil penyederhanaan, sistem juga menampilkan versi highlight yang menunjukkan secara visual bagian-bagian teks yang telah mengalami perubahan. Hal ini bertujuan agar pengguna dapat membandingkan secara langsung antara teks asli dan hasil penyederhanaannya. Prediksi kategori teks, apakah termasuk dalam klasifikasi leksikal atau sintaksis, juga ditampilkan untuk memberikan informasi tambahan kepada pengguna. Tampilan ini juga dilengkapi dengan daftar kata sulit yang terdeteksi dalam teks, beserta padanan katanya yang lebih sederhana. Dengan adanya penjelasan tersebut, pengguna bisa mengetahui bagian mana dari teks yang dianggap kompleks dan bagaimana sistem menyederhanakannya. Hal ini menjadi bukti bahwa sistem tidak hanya mampu melakukan klasifikasi, tetapi juga memberikan hasil penyederhanaan yang sesuai dan relevan.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Adapun tujuan dari penelitian ini adalah untuk membangun model klasifikasi penyederhanaan teks bacaan anak disleksia menggunakan algoritma *Naive Bayes*. Model ini tidak hanya menentukan apakah suatu teks telah disederhanakan atau belum, tetapi juga mampu mengkategorikan jenis penyederhanaan yang dilakukan, yaitu leksikal dan sintaksis. Penyederhanaan leksikal melibatkan penggantian kata-kata sulit dengan padanan yang lebih mudah dipahami, sedangkan penyederhanaan sintaksis berkaitan dengan perubahan struktur kalimat agar lebih sederhana. Dengan demikian, sistem ini diharapkan dapat membantu anak-anak dengan disleksia dalam memahami teks bacaan secara lebih efektif, serta menjadi acuan bagi pendidik dalam memilih atau menyusun materi ajar yang sesuai.

Model yang telah dibangun kemudian dievaluasi menggunakan data uji sebanyak 918 sampel, dan menghasilkan akurasi sebesar 90%. Selain itu, nilai precision dan recall masing-masing berkisar antara 88% hingga 92% pada kedua kelas, yang menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat kesalahan yang rendah dan ketepatan yang tinggi. Nilai f1-score rata-rata sebesar 0.90 juga membuktikan bahwa model memiliki performa yang seimbang dalam mengenali jenis penyederhanaan teks. Dengan hasil evaluasi tersebut, dapat disimpulkan bahwa algoritma *Naive Bayes* cukup efektif dan layak diterapkan dalam sistem klasifikasi penyederhanaan teks untuk mendukung pembelajaran anak-anak dengan disleksia.

5.2 Saran

Berdasarkan hasil yang diperoleh dari penelitian ini, terdapat beberapa hal yang dapat dijadikan saran untuk pengembangan lebih lanjut. Pertama, jumlah dataset yang digunakan masih tergolong terbatas. Maka dari itu, disarankan untuk menambah jumlah dan variasi teks bacaan, baik dari tingkat kesulitan maupun jenis bahasa yang digunakan, misalnya dengan menambahkan teks berbahasa Inggris, Jawa, atau bahasa lain. Hal ini bertujuan agar model memiliki cakupan klasifikasi yang lebih luas dan mampu mengenali berbagai pola penyederhanaan teks dengan lebih akurat.

Selain itu, proses penyederhanaan teks dapat dikembangkan dengan membandingkan metode dan model lain di luar Naïve Bayes, seperti algoritma berbasis deep learning atau sequence-to-sequence. Dengan perbandingan ini, dapat diperoleh gambaran mengenai metode mana yang paling optimal dalam mendukung pemahaman bacaan bagi anak disleksia. Ke depannya, sistem dapat diperluas tidak hanya terbatas pada klasifikasi jenis penyederhanaan saja, tetapi juga dengan memberikan hasil teks yang telah disederhanakan secara otomatis. Fitur interaktif, seperti opsi tingkat kesederhanaan yang bisa dipilih oleh pengguna, juga dapat menjadi pengembangan menarik agar sistem lebih fleksibel, adaptif, dan mudah digunakan sesuai kebutuhan pembaca.

DAFTAR PUSTAKA

Agastya, W., & Aripin. (2020). Pemetaan Emosi Dominan Pada Kalimat Majemuk Bahasa Indonesia Menggunakan Multinomial Naïve Bayes (Mapping Dominant Emotion In Indonesian Compound Sentences Using Multinomial Naïve Bayes). *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* |, *9*(2).

Bernard, M., Asror, I., & Sardi, I. L. (2018). *Penyederhanaan Kalimat Dalam Dokumen Menggunakan Metode A Noisy-Channel*. *5*(2).

Darmawan, G., Alam, S., Imam Sulistyo, M., Studi Teknik Informatika, P., Tinggi Teknologi Wastukancana Purwakarta, S., & Artikel, R. (2023). *Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes Info Artikel Abstrak*. *2*(3), 100–108. Https://Doi.Org/10.55123

Darmayanti, N., Hayati, N., Rohali, A., & Marpaung, Z. E. (2023). Pendidikan Dan Bimbingan Anak Berkebutuhan Khusus (Abk) Disleksia. *El-Mujtama: Jurnal Pengabdian Masyarakat*, *4*(2), 854–862. Https://Doi.Org/10.47467/Elmujtama.V4i2.4431

Fadhilah Az-Haari, N., Juardi, D., & Jamaludin, A. (2024). Analisis Sentimen Terhadap Boikot Brand Pro-Israel Pada Twitter Menggunakan Metode Naïve Bayes (Studi Kasus: Starbucks). Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Nomor 3).

Gaja Rismanda, M. Y., Maulana, I., & Komarudin, O. (2023). Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan Playstore Menggunakan Algoritma Naive Bayes. *Jurnal Mahasiswa Teknik Informatika*, *7*(4).

Indriastuti, F. (2015). *Pengembangan Buku Audio Untuk Mengatasi Kesulitan Belajar Anak Disleksia Development Of Audio Book For Learning Dyslexic*. *3*(2), 91–106. Www.Learningally.Org

Khoirul, M., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Nomor 1).

Kusuma Wardana, H., Swanita, I., & Yohanes, B. W. (2019). Sistem Pemeriksa Pola Kalimat Bahasa Indonesia Berbasis Algoritme Left-Corner Parsing Dengan Stemming. Dalam *Jnteti* (Vol. 8, Nomor 3).

Liana, C. F., & Sinaga, B. (2021). Sistem Pakar Diagnosa Penyakit Dyslexia Pada Anak Dengan Metode Naive Bayes Berbasis Web. *Jurnal Ilmu Komputer Dan Bisnis*, *12*(2a), 173–183. Https://Doi.Org/10.47927/Jikb.V12i2a.219

Marito Putry, N., & Nurina Sari, B. (2022). Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus. *Jurnal Sains Dan Manajemen*, *10*(1).

Oktamarina, L., Rosalina, E., Septiani Utami, L., Dzakiyyah, C., Fitri Kurnia Duati, S., Puspa Sari, R., & Sales Julita, M. (2022). *Gangguan Gejala Disleksia Pada Anak Usia Dini*. *2*.

Oktaviyani, A., Heryati, A., & Alie, M. F. (2024). *Penerapan Metode Naive Bayes Untuk Klasifikasi Kategori Olah Pangan (Studi Kasus Dinas Kesehatan Kota Palembang)*. *2*(1), 30–38.

Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Nomor 1).

Rahmawati, E., & Muhroji, M. (2024). Pengaruh Media Puzzle Huruf Untuk Meningkatkan Kemampuan Membaca Anak Disleksia. *Ideguru: Jurnal Karya Ilmiah Guru*, *9*(3), 1408–1413. Https://Doi.Org/10.51169/Ideguru.V9i3.1103

Ramadhan, F. A., Sitorus, S. H., & Rismawan, T. (2023). Penerapan Metode Multinomial Naïve Bayes Untuk Klasifikasi Judul Berita Clickbait Dengan Term Frequency - Inverse Document Frequency. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, *11*(1), 70. Https://Doi.Org/10.26418/Justin.V11i1.57452

Silvana, M., Akbar, R., & Syahnum, A. (2020). Pemanfaatan Metode Naïve Bayes Dalam Implementasi Sistem Pakar Untuk Menganalisis Gangguan Perkembangan Anak. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *6*(2), 74–81. Https://Doi.Org/10.25077/Teknosi.V6i2.2020.74-81

Sinaga, T. (2023). *Penerapan Algoritma Naive Bayes Dalam Pemrosesan Bahasa Alamiah*. Https://Www.Researchgate.Net/Publication/376720851

Siti Alyunita Mega Lestari, Akim M.H. Pardede, & Magdalena Simanjuntak. (2024). Prediksi Disleksia Pada Anak Menggunakan Metode Naive Bayes. *Jurnal Kajian Dan Penelitian Umum*, *2*(5), 37–51. Https://Doi.Org/10.47861/Jkpu-Nalanda.V2i5.1287

Sugono, D., & Qodratillah, M. T. (2013). *Tesaurus Bahasa Indonesia Pusat Bahasa*.

Surayya, S., & Mubarok, H. (2021). *Pengaruh Aplikasi Marbel Membaca Terhadap Kemampuan Membaca Anak Disleksia*. *6*(2).

Susilawati, S., & Iqbal, M. (2025). Penerapan Metode Naïve Bayes Untuk Mengidentifikasi Sentimen Pengguna Pada Ulasan Aplikasi Reelshort Di Google Play Store. *Simkom*, *10*(1), 49–59. Https://Doi.Org/10.51717/Simkom.V10i1.686

Wibowo, M. S. (2020). *Penyederhanaan Leksikal Dan Sintaksis Untuk Teks Bahasa Indonesia*.

Yuliana Putri, D., Siti Lathifah, A., Mukholis Aji Prasetyo, C., & Suparmi, S. (2024a). Peran Guru Dalam Meningkatkan Keterampilan Membaca Anak Disleksia. *Wahana Karya Ilmiah Pendidikan*, *8*(01), 26–36. Https://Doi.Org/10.35706/Wkip.V8i01.11578

Yuliana Putri, D., Siti Lathifah, A., Mukholis Aji Prasetyo, C., & Suparmi, S. (2024b). Peran Guru Dalam Meningkatkan Keterampilan Membaca Anak Disleksia. *Wahana Karya Ilmiah Pendidikan*, *8*(01), 26–36. Https://Doi.Org/10.35706/Wkip.V8i01.11578

Yuyun, Nurul Hidayah, & Supriadi Sahibu. (2021). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 820–826. Https://Doi.Org/10.29207/Resti.V5i4.3146