

***APPLICATION PROGRAMMING INTERFACE (API) DETEKSI
BAHASA MENGGUNAKAN METODE NAIVE BAYES***

SKRIPSI

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh:

ARINI SHOFA IMANIA

NIM D1041201055



FAKULTAS TEKNIK
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini.

Nama : Arini Shofa Imania

NIM : D1041201055

Menyatakan bahwa dalam skripsi yang berjudul “*APPLICATION PROGRAMMING INTERFACE (API) DETEKSI BAHASA MENGGUNAKAN METODE NAIVE BAYES*” tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan Saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya sanggup menerima konsekuensi akademis dan hukum di kemudian hari apabila pernyataan yang dibuat ini tidak benar.

Pontianak, 31 Januari 2025

Arini Shofa Imania

NIM D1041201055



KEMENTERIAN PENDIDIKAN TINGGI, SAINS,
DAN TEKNOLOGI
UNIVERSITAS TANJUNGPURA
FAKULTAS TEKNIK

Jalan Prof. Dr. H. Hadari Nawawi Pontianak 78124
Telepon (0561) 740186, WA: +6282152280907
Email : ft@untan.ac.id Website : <http://teknik.untan.ac.id>

HALAMAN PENGESAHAN

APPLICATION PROGRAMMING INTERFACE (API) DETEKSI BAHASA
MENGUNAKAN METODE NAIVE BAYES

Program Studi Sarjana Informatika
Jurusan Informatika


Oleh:

Arini Shofa Imania
NIM D1041201055

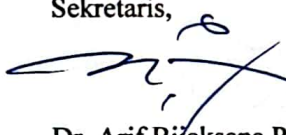
Telah dipertahankan di depan Penguji Skripsi pada tanggal 31 Januari 2025 dan
diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana.

Susunan Penguji Skripsi:

Ketua,


Prof. Dr. Herry Sujaini, S.T., M.T.
NIP 196806291997021001

Sekretaris,


Dr. Arif Bijaksana Putra Negara, S.T., M.T.
NIP 197208081998021002

Penguji Utama,


Rina Septiriana, ST, M.Cs
NIP 198709232020122001

Penguji Pendamping,


Khairul Hafidh, S.T., M.Kom.
NIP 199202232022031003



Dr.-Ing. R. Slamet Widodo, M.T., IPM
NIP 196712231992031002

HALAMAN PERSEMBAHAN

Segala puji syukur kepada Allah SWT. atas segala nikmat serta limpahan karunia-Nya telah menguatkan dan menyertai segala kelancaran dalam menuntaskan Tugas akhir ini hingga selesai sebagai salah satu syarat mendapatkan gelar keserjanaan.

Tugas akhir ini saya persembahkan untuk kedua orang tua tercinta Ayah dan Mama. Terima kasih atas kasih sayang yang tulus, doa, nasihat yang bijaksana, dan pengorbanan yang tak terhingga. Ayah dan Mama adalah kekuatan dari setiap langkah yang saya ambil dan semoga karya ini dapat menjadi sedikit dari rasa terima kasih saya untuk segala yang telah kalian lakukan.

Segala hormat serta terima kasih saya sampaikan kepada Bapak Prof. Dr. Herry Sujaini, S.T., M.T. dan Bapak Dr. Arif Bijaksana Putra Negara, S.T., M.T. selaku dosen pembimbing tugas akhir, serta Bapak Dr. Ir. Yus Sholva, S.T., M.T. selaku dosen pembimbing akademik saya. Terima kasih atas segala ilmu, dan arahan yang berharga diberikan selama menjalani tugas akhir ini.

Untuk teman-teman saya, Diva, Galuh, Reza, dan Anisa, yang senantiasa menemani dari awal perkuliahan hingga akhir masa tugas akhir ini. Terima kasih saya sampaikan sebanyak-banyaknya atas segala kenangan perkuliahan yang tak akan saya lupakan.

Untuk Abang, Ayuk Yaya, Ayuk Caca, Kakak Utin, Abang Randy, dan Faril, saudara yang menemani dikala susah dan senang. Tak pernah berucap sayang, maaf, dan terimakasih tak menjadikan hal-hal yang telah kita lalui dari saya lahir hingga sekarang tidak berarti.

Teruntuk Yafa kesayang Ateu, yang selalu lucu dimanapun dan kapanpun.

KATA PENGANTAR

Allhamdulillah. Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat serta nikmat-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir dengan judul “*Application Programming Interface (API) Deteksi Bahasa Menggunakan Metode Naive Bayes*” sebagai salah satu persyaratan untuk bisa menyelesaikan Program Sarjana (S1) pada Fakultas Teknik, Jurusan Informatika, Universitas Tanjungpura.

Penulis menyadari banyak pihak yang telah memberikan bantuan, masukan serta bimbingan selama menyelesaikan studi dan tugas akhir ini. Pada kesempatan ini, secara khusus penulis ingin menyampaikan ucapan terima kasih kepada Bapak Prof. Dr. Herry Sujaini, S.T, M.T. selaku dosen Pembimbing Pertama dan Bapak Dr. Arif Bijaksana Putra Negara, S.T., M.T. selaku dosen Pembimbing Kedua, serta Bapak Dr. Ir. Yus Sholva, S.T., M.T. selaku Dosen Pembimbing Akademik serta seluruh dosen Program Studi Informatika atas ilmu-ilmu yang diberikan selama masa perkuliahan. Dalam pengerjaan skripsi ini, penulis menyadari bahwa Skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi penyempurnaan penulisan ini dimasa yang akan datang. Semoga hasil penelitian ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan serta dapat menjadi referensi bagi penelitian-penelitian selanjutnya.

Pontianak, 31 Januari 2025

Penulis,

Arini Shofa Imania

ABSTRAK

Penelitian ini mengembangkan model deteksi bahasa daerah menggunakan pendekatan klasifikasi dengan algoritma Multinomial Naive Bayes (MNB) untuk mengenali bahasa Melayu Pontianak, Jawa, dan Dayak. Proses pelatihan dilakukan dengan menggunakan 10-Fold Cross Validation pada tiga skenario dataset yang berbeda: dataset tidak seimbang, dataset seimbang minimum, dan dataset seimbang maksimum (dengan teknik SMOTE). Hasil evaluasi model menunjukkan performa yang sangat baik, dengan precision, recall, dan F1-score yang mencapai lebih dari 97% pada semua skenario dataset yang diuji. Peningkatan signifikan terlihat pada deteksi bahasa minoritas, terutama pada skenario dataset seimbang, yang menunjukkan kemampuan model dalam mengenali kelas yang lebih jarang. *Application Programming Interface* (API) berbasis Flask dikembangkan untuk memproses input teks dan memberikan prediksi bahasa beserta probabilitasnya dalam format JSON, yang memudahkan integrasi dengan aplikasi eksternal. Pengujian black-box API menunjukkan bahwa API ini dapat mendeteksi bahasa daerah dengan akurasi yang tinggi, meskipun terdapat keterbatasan pada teks yang tidak termasuk dalam cakupan pelatihan, seperti bahasa yang tidak terdata dalam dataset yang digunakan. Hasil ini membuktikan fleksibilitas API dalam memberikan prediksi dan dapat menjadi dasar pengembangan lebih lanjut untuk mendeteksi bahasa daerah lainnya

Kata kunci: API, Multinomial Naive Bayes, Deteksi Bahasa, Bahasa Daerah.

ABSTRACT

This study develops a regional language detection model using a classification approach with the Multinomial Naive Bayes (MNB) algorithm to recognize Pontianak Malay, Javanese, and Dayak languages. The training process was carried out using 10-Fold Cross Validation on three different dataset scenarios: imbalanced dataset, minimum balanced dataset, and maximum balanced dataset (with the SMOTE technique). The model evaluation results showed very good performance, with precision, recall, and F1-score reaching more than 97% in all tested dataset scenarios. Significant improvements were seen in minority language detection, especially in the balanced dataset scenario, which shows the model's ability to recognize rarer classes. A Flask-based Application Programming Interface (API) was developed to process text input and provide language predictions along with their probabilities in JSON format, which facilitates integration with external applications. Black-box testing of the API showed that this API can detect regional languages with high accuracy, although there are limitations in texts that are not included in the training scope, such as languages that are not recorded in the dataset used. These results prove the flexibility of the API in providing predictions and can be the basis for further development to detect other regional languages.

Keywords: API, Multinomial Naive Bayes, Language Detection, Regional Languages.

DAFTAR ISI

HALAMAN PERNYATAAN	ii
HALAMAN PERSEMBAHAN	iv
Kata Pengantar	v
Abstrak	vi
Abstract	vii
Daftar Isi	viii
Daftar Tabel	x
Daftar Gambar	xi
Daftar Lampiran	xii
BAB I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Pembatasan Masalah	5
1.5 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Kajian Terkait	7
2.2 Landasan Teori.....	8
2.2.1 <i>Application Programming Interface</i>	9
2.2.2 <i>JavaScript Object Notation</i>	9
2.2.3 Deteksi Bahasa	10
2.2.4 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	10
2.2.5 Klasifikasi	10
2.2.6 Naïve Bayes	11
2.2.7 <i>Confusion Matrix</i>	13
2.2.8 Mesin Penerjemah.....	15
2.2.9 Bahasa Daerah	15

BAB III METODOLOGI PENELITIAN	17
3.1 Studi Literatur	17
3.2 Perangkat Penelitian.....	17
3.3 Pengumpulan Data	18
3.4 <i>Preprocessing Data</i>	21
3.2.1 <i>Data Cleaning</i>	21
3.2.2 <i>Case Folding</i>	22
3.2.3 Tokenisasi	22
3.2.4 <i>TfidfVectorizer</i>	23
3.5 Pengembangan Model Deteksi Bahasa.....	24
3.6 Pengembangan API Deteksi Bahasa	27
3.7 Pengujian dan Evaluasi	30
BAB IV HASIL DAN ANALISIS	33
4.1 Pengumpulan Data	33
4.2 <i>Preprocessing Data</i>	36
4.3 Pengembangan dan Pelatihan Model Naive Bayes.....	38
4.3.1 Pembagian Dataset.....	39
4.3.2 Pelatihan Model	40
4.3.3 Penyimpanan Model	41
4.4 Implementasi API Deteksi Bahasa.....	42
4.5 Evaluasi dan Analisis	46
4.5.1 Evaluasi Model	46
4.5.2 Evaluasi Fungsionalitas Sistem.....	55
BAB V KESIMPULAN DAN SARAN	63
5.1 Kesimpulan	63
5.2 Saran	64
Daftar Pustaka.....	65

DAFTAR TABEL

Tabel 2. 1 Contoh <i>Confusion Matrix</i>	14
Tabel 3. 1 Spesifikasi <i>Software</i> dan <i>Hardware</i>	17
Tabel 3. 2 Contoh Dataset Sumber Melayu Pontianak	18
Tabel 3. 3 Contoh Dataset Sumber Dayak	18
Tabel 3. 4 Contoh Dataset Sumber Jawa	19
Tabel 3. 5 Distribusi Jumlah Kalimat	20
Tabel 3. 6 Skenario Pengujian Model	31
Tabel 3. 7 Skenario Pengujian API	32
Tabel 4. 1 Contoh Dataset Melayu Pontianak	33
Tabel 4. 2 Contoh Dataset Jawa	34
Tabel 4. 3 Contoh Dataset Dayak	34
Tabel 4. 4 Statistik Korpus	35
Tabel 4. 5 Kosakata Similiar	35
Tabel 4. 6 Contoh Hasil <i>Preprocessing</i>	37
Tabel 4. 7 Contoh Hasil <i>Preprocessing</i>	38
Tabel 4. 8 <i>Confusion Matrix</i> Skenario 1	47
Tabel 4. 9 <i>Confusion Matrix</i> skenario 2	49
Tabel 4. 10 <i>Confusion Matrix</i> Skenario 3	50
Tabel 4. 11 Hasil dan Analisis Pengujian Model	53
Tabel 4. 12 Hasil Skenario Pengujian API	58

DAFTAR GAMBAR

Gambar 3. 1 Diagram Distribusi Dataset	20
Gambar 3. 2 <i>Data Cleaning</i>	22
Gambar 3. 3 Case Folding	22
Gambar 3. 4 Tokenisasi	23
Gambar 3. 5 TF-IDF	23
Gambar 3. 6 <i>Flowchart</i> Pengembangan Model.	25
Gambar 3. 7 <i>Flowchart</i> Pengembangan API	27
Gambar 4. 1 Arsitektur Sistem.....	42
Gambar 4. 2 <i>Request</i> JSON.	43
Gambar 4. 3 <i>Flask</i>	43
Gambar 4. 4 <i>Preprocessing</i> API	43
Gambar 4. 5 Prediksi API	44
Gambar 4. 6 Input Teks Yang Dikirim	45
Gambar 4. 7 Respon JSON Yang Diterima Dari API.....	45
Gambar 4. 8 <i>Classification Report Table</i> skenario 1	46
Gambar 4. 9 <i>Confusion Matrix</i> skenario 1	47
Gambar 4. 10 <i>Classification Report Table</i> skenario 2	48
Gambar 4. 11 <i>Confusion matrix</i> skenario 2	48
Gambar 4. 12 <i>Classification Report Table</i> skenario 3	50
Gambar 4. 13 <i>Confusion Matrix</i> skenario 3	50
Gambar 4. 14 Hasil Akurasi Model Untuk Skenario 1	51
Gambar 4. 15 Hasil Akurasi Model Untuk Skenario 2	52
Gambar 4. 16 Hasil Akurasi Model Skenario 3 (SMOTE).....	53
Gambar 4. 17 Hasil <i>Test Case</i>	57

DAFTAR LAMPIRAN

Lampiran A71

BAB I PENDAHULUAN

1.1 Latar Belakang

Saat ini, aplikasi penerjemahan teks telah menjadi alat yang berguna dalam memfasilitasi komunikasi lintas budaya. Terkhusus Indonesia yang memiliki lebih dari 700 bahasa daerah yang beragam (Kemendikbudristek, 2022). Menurut Setia, kesuksesan suatu proses penerjemahan sangat tergantung pada tujuan di balik terjemahan tersebut, yang akan tercermin dalam hasilnya dan sesuai dengan kebutuhan orang yang memerlukannya (Setia, 2007). Identifikasi bahasa asli suatu teks menjadi salah satu landasan utama untuk menyediakan penerjemahan yang relevan. Pada umumnya, deteksi bahasa atau *language detection* yang biasa juga disebut identifikasi bahasa merujuk pada kemampuan sistem untuk mengenali bahasa yang digunakan dalam suatu teks atau dokumen dengan proses menentukan bahasa asli dari teks tersebut (Ahmed et al., 2004).

Menurut Jauhiainen et al., meskipun deteksi bahasa dapat melibatkan bahasa, dialek, atau variasi bahasa, dari sudut pandang komputasi, algoritma dan fitur yang digunakan untuk membedakan antara bahasa, variasi bahasa, dan dialek seringkali sama (Jauhiainen et al., 2019). Hal ini mencerminkan upaya untuk mengembangkan metode yang umum dan dapat diterapkan secara fleksibel untuk mengatasi berbagai tantangan dalam deteksi bahasa. Dalam penelitian terkait yang menggunakan strategi yang berbeda dalam mendeteksi bahasa untuk kompetisi TweetLID-2014 dengan menggunakan *Ranking-based* dan Naive Bayes. Hasilnya menunjukkan bahwa *ranking dictionaries* berperforma lebih baik dengan korpus pelatihan kecil yang distribusinya serupa dengan dataset uji, sementara algoritma Naive Bayes meningkatkan skor dengan model yang besar, bahkan jika data tidak seimbang terkait dataset uji. Dalam evaluasi final, klasifikasi Naive Bayes mendapatkan posisi pertama diantara sistem-sistem tanpa batasan dilatih dengan sumber eksternal yang berpartisipasi dalam kompetisi (Gamallo et al., 2014).

Naive Bayes adalah suatu algoritma pembelajaran yang sederhana dengan asumsi kuat bahwa atribut bersifat kondisional independen jika diberikan kelasnya. Meskipun asumsi independensi ini sering tidak terpenuhi dalam praktik, Naive

Bayes tetap mampu memberikan tingkat akurasi klasifikasi yang bersaing (Webb, 2016). Untuk deteksi bahasa, Naive Bayes dapat diimplementasikan dengan membangun model probabilitas untuk setiap bahasa yang ingin dideteksi. Model ini akan didasarkan pada frekuensi kemunculan kata-kata dalam teks pelatihan yang diberikan untuk setiap bahasa. Penghitungan menggunakan Teorema Bayes dilakukan dengan membandingkan kemungkinan terjadinya suatu peristiwa dengan kemungkinan tidak terjadinya peristiwa tersebut (Candra & Mirwanto, 2018).

Dalam penelitian ini, teknologi deteksi bahasa menjadi langkah awal yang untuk memastikan relevansi hasil terjemahan. Bahasa daerah yang menjadi fokus dalam penelitian ini adalah Melayu Pontianak, Jawa, dan Dayak. Melayu Pontianak bahasa Melayu yang digunakan di Kalimantan Barat dengan ciri khas tersendiri yang berfungsi sebagai alat komunikasi terutama digunakan oleh masyarakat di ibu kota Pontianak, namun, di luar wilayah ibu kota Pontianak, bahasa ini tidak lagi digunakan, karena setiap daerah kabupaten memiliki bahasa yang berbeda (E. P. D. Putra & Hartati, 2016). Bahasa Dayak merupakan salah satu bahasa daerah yang berada di Indonesia. Bahasa Dayak melambangkan kelompok bahasa dari masyarakat adat Kalimantan yang memiliki karakteristik unik, baik dalam kosakata maupun struktur bahasanya. Dayak juga merupakan kelompok masyarakat yang besar dari sisi jumlah, persebaran, dan perannya salah satunya di Kalimantan Barat, dimana komunitas ini membentuk 30 hingga 40 persen dari penduduk Kalbar dengan 151 subsuku dan 100 sub subsuku (Yusriadi, 2018). Sebagai bahasa daerah yang eksistensinya masih dipakai dalam berkomunikasi, dihargai, dipelihara oleh masyarakat, dan negara karena bahasa Dayak merupakan bagian dari khazanah dan budaya Indonesia (Istianingrum, 2015). Bahasa Jawa adalah bagian dari kebudayaan dan identitas suatu bangsa di Indonesia, namun, di era modern penggunaan bahasa Jawa mulai memudar seiring dengan perkembangan zaman (Nadhiroh & Setyawan, 2021). Sebagai salah satu bahasa daerah, bahasa Jawa memiliki peran sebagai simbol identitas daerah, simbol kebanggaan masyarakat daerah, dan sarana komunikasi dalam lingkungan keluarga dan masyarakat di daerah tersebut (Bhakti, 2020). Ketiga bahasa ini dipilih untuk merepresentasikan keberagaman linguistik Indonesia sekaligus menjadi tantangan dalam

pengembangan model deteksi bahasa akibat kemiripan kosakata antarbahasa dimana terdapat di cakupan pulau yang sama yaitu Melayu Pontianak dan Dayak, serta bahasa yang pola kosakatanya berbeda yaitu Jawa dan distribusi data yang tidak merata.

Penelitian ini menggunakan pendekatan klasifikasi untuk mendeteksi bahasa berdasarkan teks input. Klasifikasi merupakan bentuk dasar dari analisis data yang mempunyai teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada (Azis et al., 2020). Algoritma yang digunakan adalah Multinomial Naive Bayes, salah satu metode pembelajaran mesin berbasis probabilitas yang efisien untuk tugas klasifikasi teks. Multinomial Naïve Bayes menggunakan pendekatan probabilitas dan termasuk dalam teknik *supervised learning* untuk kasus klasifikasi teks (Satria Mahendra, 2024). Algoritma ini bekerja dengan menghitung probabilitas setiap kata dalam suatu kelas bahasa, sehingga memungkinkan prediksi berdasarkan pola distribusi kata yang dipelajari dari dataset pelatihan. Multinomial Naive Bayes juga dikenal karena kesederhanaannya dalam implementasi dan kinerjanya yang kompetitif, terutama pada data teks. Untuk meningkatkan akurasi pada bahasa minoritas yaitu Melayu Pontianak dan Jawa teknik optimasi seperti *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan untuk mengatasi masalah distribusi data yang tidak seimbang.

Penelitian ini berfokus pada pengembangan Application Programming Interface (API) deteksi bahasa yang dibangun menggunakan Flask. API ini dirancang untuk memuat model deteksi bahasa dan memproses input teks. Tujuan utama dari pembuatan API ini adalah agar dapat diimplementasikan dan diintegrasikan dengan berbagai aplikasi eksternal tanpa memerlukan perubahan pada struktur model atau dataset, sehingga API dapat diakses secara independen oleh aplikasi apa pun yang memerlukan layanan deteksi bahasa. Dengan menggunakan API, pengguna dapat dengan mudah mengakses layanan deteksi bahasa hanya dengan mengirimkan request, yang memungkinkan proses deteksi bahasa dilakukan secara otomatis tanpa intervensi manual (Baharuddin et al., 2022). API ini memiliki arsitektur berbasis HTTP, yang memungkinkannya untuk

menerima input teks, mengolah teks tersebut menggunakan model Multinomial Naive Bayes, dan menyajikan hasil klasifikasi dalam format JSON. Format JSON ini sangat memudahkan integrasi dengan sistem lain yang memerlukan informasi hasil deteksi bahasa.

Model klasifikasi Multinomial Naive Bayes yang digunakan dalam API ini diharapkan mampu menghasilkan prediksi bahasa yang akurat sesuai dengan input teks yang diberikan. Dengan pengembangan API berbasis Flask ini, diharapkan proses deteksi bahasa mudah diakses oleh berbagai aplikasi yang memerlukan layanan tersebut.

1.2 Perumusan Masalah

Dalam penelitian ini, pengembangan deteksi bahasa menjadi fokus utama, namun, terdapat beberapa hal yang menjadi fokus utama. Berdasarkan hal tersebut, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana merancang model deteksi bahasa menggunakan pendekatan klasifikasi berbasis Multinomial Naive?
2. Bagaimana membangun API dengan *Flask* yang dapat memuat model deteksi bahasa secara efisien, memproses *input* teks, dan memberikan prediksi yang akurat dalam format JSON untuk memastikan keandalannya dalam mendeteksi bahasa daerah?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan deteksi bahasa daerah dengan Multinomial Naive Bayes dalam mendukung penerjemahan teks multibahasa. Penelitian ini dirancang dengan beberapa tujuan, sebagai berikut.

1. Merancang model deteksi bahasa menggunakan pendekatan klasifikasi berbasis Multinomial Naive Bayes.
2. Membangun API dengan *Flask* yang dapat memuat model deteksi bahasa agar dapat memproses *input* teks, dan memberikan prediksi bahasa secara akurat dalam format JSON.

1.4 Pembatasan Masalah

Agar penelitian terfokus pada pengembangan deteksi bahasa daerah dengan Multinomial Naive Bayes dalam mendukung penerjemahan teks multibahasa. serta memastikan penelitian memiliki fokus yang jelas, maka penulis membuat batasan sebagai berikut.

1. Penelitian ini akan berfokus pada data teks dari bahasa daerah di Indonesia, yaitu bahasa Melayu Pontianak, Jawa, dan Dayak.
2. Pengembangan model deteksi bahasa dengan menggunakan Multinomial Naïve Bayes untuk mengoptimalkan hasil klasifikasi yang dirancang untuk mengenali tiga bahasa daerah Indonesia.
3. Pengembangan *Application Programming Interface* (API) dirancang untuk dapat diakses secara independen oleh *platform* yang memerlukan layanan deteksi bahasa. Dengan arsitektur berbasis HTTP dan penyajian hasil dalam format JSON, melakukan *preprocessing* data, dan menyediakan hasil prediksi bahasa dengan transparansi melalui probabilitas keyakinan model.
4. Proses pengujian dan evaluasi kinerja model deteksi bahasa akan difokuskan pada pengukuran akurasi sebagai indikator utama keberhasilan sistem serta kemampuan model untuk beradaptasi terhadap perbedaan karakteristik bahasa daerah yang diuji.

1.5 Sistematika Penulisan

Dalam penelitian ini, penulis membagi sistematika penelitian menjadi lima bab. Sebelum bab pertama pada bagian awal, peneliti mencantumkan halaman judul, halaman pernyataan, lembar pengesahan, halaman persembahan, kata pengantar, abstrak, daftar isi, daftar tabel, dan daftar gambar.

BAB I PENDAHULUAN

Bab pendahuluan berisi latar belakang, serta sistem pendeteksiannya, sedangkan inti dari permasalahan akan digambarkan dalam perumusan masalah,

pembatasan masalah menjelaskan batasan-batasan dari sistem yang akan dibuat supaya tidak keluar dari ketentuan-ketentuan yang telah ditetapkan, tujuan penelitian berupa harapan dari hasil yang akan dicapai dari rancang bangun sistem tersebut.

BAB II TINJAUAN PUSTAKA

Tinjauan pustaka berisi teori-teori yang berhubungan dengan pembuatan aplikasi ini, meliputi deteksi bahasa, naïve bayes, bahasa daerah, dan teori-teori penunjang lainnya yang berkaitan dengan sistem tersebut.

BAB III METODOLOGI PENELITIAN

Metodologi penelitian mengemukakan metode penelitian yang dilakukan dalam perancangan dan implementasi, gambaran umum sistem dengan metode naïve bayes berupa API pendeteksian bahasa untuk mesin penerjemah teks multibahasa daerah Indonesia.

BAB IV HASIL DAN PEMBAHASAN

Hasil dan pembahasan memaparkan dari hasil-hasil tahapan penelitian, mulai dari analisis hingga hasil testing dan implementasinya untuk API deteksi multibahasa daerah Indonesia.

BAB V PENUTUP

Bab penutup berisi kesimpulan jawaban dari masalah yang diteliti, saran untuk pengembangan sistem atau untuk penelitian selanjutnya serta rekomendasi bagi masyarakat sekitar maupun peneliti lain.