

**PENERAPAN METODE BERT UNTUK KLASIFIKASI POLA
PENULISAN PUISI PENYAIR INDONESIA**
(Studi Kasus : Balai Bahasa Provinsi Kalimantan Barat)

NAVILA NATASYAHRANI
NIM H1101211027

SKRIPSI



PROGRAM STUDI SISTEM INFORMASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

**PENERAPAN METODE BERT UNTUK KLASIFIKASI POLA
PENULISAN PUISI PENYAIR INDONESIA**
(Studi Kasus : Balai Bahasa Provinsi Kalimantan Barat)

NAVILA NATASYAHRANI
NIM H1101211027

Skripsi

Skripsi Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi



PROGRAM STUDI SISTEM INFORMASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir : Penerapan Metode BERT untuk Klasifikasi Pola
Penulisan Puisi Penyair Indonesia (Studi Kasus:
Balai Bahasa Provinsi Kalimantan Barat)

Nama Mahasiswa : Navila Natasyahrani
NIM : H1101211027
Jurusan/Program Studi : Sistem Informasi
Tanggal Lulus : 31 Januari 2025
SK Pembimbing : 3431/UN22.8/TD.06/2024/1 November 2024
SK Penguji : 227/UN22.8/TD.06/2025/24 Januari 2025

Pembimbing I

Dosen Pembimbing

Dian Prawira, S.T., M.Eng.
NIP 198411132015041001

Pembimbing II

Ibnur Rusi, S.Kom., M.M.
NIP 198907282019031008

Dosen Penguji

Ketua Penguji

Renny Puspita Sari, S.T., M.T.
NIP 198704182015042001

Anggota Penguji

Syahru Rahmayuda, S.Kom., M.Kom.
NIDK.8884370018

Pimpinan Sidang
(merangkap anggota penguji)

Dian Prawira, S.T., M.Eng.
NIP 198411132015041001

Sekretaris Sidang
(merangkap anggota penguji)

Ibnur Rusi, S.Kom., M.M.
NIP 198907282019031008

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura


Prof. Dr. Gusrizal, S.Si., M.Si.
NIP 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Navila Natasyahrani
NIM : H1101211027
Program Studi/ Jurusan : Sistem Informasi/ Sistem Informasi
Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 31 Januari 2025



Navila Natasyahrani

H1101211027

PENERAPAN METODE BERT UNTUK KLASIFIKASI POLA PENULISAN PUISI PENYAIR INDONESIA

(Studi Kasus : Balai Bahasa Provinsi Kalimantan Barat)

Abstrak

Klasifikasi terhadap pola penulisan puisi seorang penyair sering kali didasarkan pada persepsi dan pengetahuan subjektif pembaca yang dipengaruhi oleh latar belakang budaya, pengalaman, dan preferensi pribadi. Hal ini juga menjadi perhatian Balai Bahasa Provinsi Kalimantan Barat, sebagai lembaga pelindungan dan pemasyarakatan bahasa dan sastra. Dalam konteks tersebut, penelitian ini menganalisis 390 puisi dari enam penyair yaitu Chairil Anwar, Joko Pinurbo, Sapardi Djoko Damono, M Aan Mansyur, Tia Ragat, dan Pradono menggunakan metode BERT. *Bidirectional Encoder Representations from Transformers* (BERT) merupakan model deep learning berbasis *Transformer* yang memproses teks secara bidirectional untuk memahami konteks secara mendalam. BERT digunakan untuk analisis mendalam, menghasilkan *accuracy* sebesar 79% dengan pembagian data 80:20 dan 87% dengan pembagian data 90:10 dengan hasil prediksi puisi yang diuji adalah Sapardi Djoko Damono. Selain itu, analisis stilistika dilakukan untuk mengevaluasi struktur, tema, dan nada emosional dalam puisi, serta menghasilkan hasil analisis puisi yang diuji dan skor kemiripan sebesar 96.44%. Kombinasi metode BERT dan stilistika memberikan hasil analisis yang lebih mendalam terhadap pola penulisan puisi setiap penyair. Penelitian ini diharapkan dapat memberikan kontribusi dalam memahami pola penulisan dan karakteristik gaya bahasa penyair Indonesia secara lebih objektif dan mendalam.

Kata Kunci : BERT, Stilistika, Puisi, Klasifikasi, Balai Bahasa Provinsi Kalimantan Barat

PENERAPAN METODE BERT UNTUK KLASIFIKASI POLA PENULISAN PUISI PENYAIR INDONESIA

(Studi Kasus : Balai Bahasa Provinsi Kalimantan Barat)

Abstract

The classification of a poet's poetic writing pattern is often based on the reader's subjective perception and knowledge influenced by cultural background, experience, and personal preference. This is also a concern of the Language Center of West Kalimantan Province, as an institution for the protection and popularization of language and literature. In this context, this study analyzed 390 poems from six poets namely Chairil Anwar, Joko Pinurbo, Sapardi Djoko Damono, M Aan Mansyur, Tia Ragat, and Pradono using the BERT method. Bidirectional Encoder Representations from Transformers (BERT) is a Transformer-based deep learning model that processes text bidirectionally to understand context in depth. BERT is used for deep analysis, resulting in an accuracy of 79% with a data division of 80:20 and 87% with a data division of 90:10 with the prediction results of the tested poem is Sapardi Djoko Damono. In addition, stylistic analysis was conducted to evaluate the structure, theme, and emotional tone in the poems, and resulted in the analysis of the tested poems and a similarity score of 96.44%. The combination of BERT and stylistic methods provides a more in-depth analysis of the poetry writing patterns of each poet. This research is expected to contribute to understanding the writing patterns and stylistic characteristics of Indonesian poets more objectively and deeply.

Keywords : BERT, Stylistics, Poetry, Classification, Balai Bahasa Provinsi Kalimantan Barat

PRAKATA

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya, sehingga saat ini penulis berhasil menyelesaikan skripsi yang berjudul “Penerapan Metode BERT untuk Klasifikasi Pola Penulisan Puisi Penyair Indonesia (Studi Kasus : Balai Bahasa Provinsi Kalimantan Barat). Dalam penyelesaian penulisan skripsi ini, penulis banyak mendapatkan bimbingan dan bantuan dari berbagai pihak. Oleh sebab itu, penulis mengucapkan terima kasih kepada:

1. Orang tua dan keluarga yang senantiasa selalu memberikan dukungan baik secara materil maupun moril.
2. Bapak Ibnur Rusi, S.Kom., M.M. selaku dosen pembimbing akademik dan pembimbing II yang telah memberikan bimbingan, masukan dan saran dari awal perkuliahan hingga akhir penulisan skripsi ini.
3. Bapak Dian Prawira, S.T., M.Eng. selaku dosen pembimbing I yang telah memberikan bimbingan, masukan dan saran dalam penulisan skripsi ini.
4. Ibu Renny Puspita Sari S.T., M.T. selaku Dosen Penguji I dan Ketua Jurusan Sistem Informasi yang telah memberikan masukan dan saran.
5. Bapak Syahru Rahmayuda, S.Kom., M.Kom. selaku Dosen penguji II yang telah memberikan masukan dan saran.
6. Bapak dan Ibu serta staf Program Studi Sistem Informasi yang telah membantu mengurus administrasi selama perkuliahan.
7. Dera Elvany Riyadi dan Ulfy Luthvia yang selalu menghibur, mendukung, dan menemani penulis ketika pusing menghadapi hidup.
8. Nanda Permata Lutfina yang selalu menemani mengerjakan tugas hingga malam hari selama pembuatan skripsi.
9. Teman-teman dekat saya di Sistem Infomasi angkatan 2021 yang selalu membantu selama perkuliahan dari awal hingga akhir.
10. Teman-teman Eternity 2021 yang telah memberikan dukungan dan selamat, semoga bisa cepat lulus juga.

Penulis menyadari bahwa Tugas Akhir ini masih memiliki berbagai kekurangan karena keterbatasan dari penulis. Oleh karena itu, penulis memohon maaf apabila terdapat hal-hal yang kurang berkenan dalam penulisan ini. Kritik dan saran yang membangun sangat diharapkan untuk perbaikan di masa mendatang. Semoga Tugas Akhir ini dapat memberikan manfaat bagi penulis maupun pihak-pihak lain yang membutuhkan.

Pontianak, 16 Januari 2025

Navila Natasyahrani

H1101211027

DAFTAR ISI

Abstrak	iv
Abstract	v
PRAKATA.....	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
DAFTAR LISTING CODE.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI	7
2.1 Dasar Teori	7
2.1.1. <i>Text Mining</i>	7
2.1.2. <i>Text Preprocessing</i>	7
2.1.3. Klasifikasi	8
2.1.4. <i>Bidirectional Encoder Representations from Transformers (BERT)</i>	8
2.1.5. Balai Bahasa Provinsi Kalimantan Barat	9
2.1.6. <i>Confusion Matrix</i>	9
2.1.7. Puisi.....	11
2.1.8. Struktur Stilistika	11
2.1.9 <i>Cosine Similarity</i>	11
2.2 Kajian Penelitian	12
BAB III METODOLOGI PENELITIAN.....	15
3.1 Bahan Penelitian.....	15
3.2 Alat Penelitian	15
3.3 Kerangka Kerja Penelitian	15
3.4 Metode Penelitian.....	19

3.5	Data Penelitian	21
3.5.1	Objek Penelitian	21
3.5.2	Populasi dan Sampel	21
3.5.3	Jenis dan Sumber Data	21
3.5.4	Metode Pengumpulan Data	22
BAB IV	HASIL DAN PEMBAHASAN.....	23
4.1	Studi Literatur	23
4.2	Pengumpulan Data	23
4.3	<i>Install dan Import Packages</i>	25
4.4	<i>Load Dataset</i>	26
4.5	<i>Bidirectional Encoder Representations from Transformers (BERT)</i>	27
4.5.1	<i>Data Split</i>	27
4.5.2	<i>Data Preprocessing</i>	28
4.5.3	Konfigurasi Model	30
4.5.4	Arsitektur Model	31
4.5.5	<i>Model Training</i>	32
4.6	Stilistika	34
4.7	Analisis dan Prediksi Puisi.....	42
4.6	Analisis Komparatif	46
4.7	Visualisasi	58
4.7.1	Visualisasi Skor Kemiripan.....	58
4.7.2	Visualisasi Puisi	59
4.7.3	Visualisasi Frekuensi Kata	61
4.8	Evaluasi	66
4.9	Pengujian Prediksi Puisi.....	82
4.10	Visualisasi <i>Similarity Score</i> Penyair.....	87
BAB V	PENUTUP.....	89
5.1	Kesimpulan	89
5.2	Saran.....	90
REFERENSI		91
LAMPIRAN		94

DAFTAR GAMBAR

Gambar 2. 1 Pre-training dan Fine-Tuning model BERT	8
Gambar 3. 1 Kerangka Kerja IS Research	16
Gambar 3. 2 Flowchart Penelitian.....	20
Gambar 4. 1 Rata-Rata Kata per Puisi	60
Gambar 4. 2 Rata-Rata Baris per Puisi	61
Gambar 4. 3 Lima Teratas Kata-Kata Umum Chairil Anwar.....	63
Gambar 4. 4 Lima Teratas Kata-Kata Umum Joko Pinurbo	63
Gambar 4. 5 Lima Teratas Kata-Kata Umum Sapardi Djoko Damono	64
Gambar 4. 6 Lima Teratas Kata-Kata Umum M Aan Mansyur	65
Gambar 4. 7 Lima Teratas Kata-Kata Umum Tia Ragat.....	65
Gambar 4. 8 Lima Teratas Kata-Kata Umum Pradono	66
Gambar 4. 9 Loss Across Epoch Skenario 1	76
Gambar 4. 10 Accuracy Across Epochs Skenario 1	76
Gambar 4. 11 Confusion Matrix Skenario 1	77
Gambar 4. 12 Loss Across Epochs Skenario 2	79
Gambar 4. 13 Accuracy Across Epochs Skenario 2.....	80
Gambar 4. 14 Confusion Matrix Skenario 2	80
Gambar 4. 15 Hasil Analisis Puisi Testing.....	84
Gambar 4. 16 Hasil Prediksi Puisi Testing.....	85
Gambar 4. 17 Hasil Perbandingan dengan Penyair Lain	86
Gambar 4. 18 Similarity Scores Between Poem and Poet Profiles.....	87

DAFTAR TABEL

Tabel 2. 1 Tabel Confusion Matrix.....	10
Tabel 2. 2 Kajian Penelitian	12
Tabel 4. 1 Contoh Dataset Puisi	23
Tabel 4. 2 Contoh Text Cleaning.....	28
Tabel 4. 3 Contoh Encoding Labels	29
Tabel 4. 4 Hasil Evaluasi Skenario 1 (80 20)	74
Tabel 4. 4 Hasil Evaluasi Skenario 1 (80 20) (Lanjutan)	75
Tabel 4. 5 Hasil Evaluasi Skenario 2 (90 10)	78

DAFTAR LISTING CODE

Listing Code 4. 1 Install dan Import Packages	25
Listing Code 4. 2 Download NLTK	26
Listing Code 4. 3 Load Data	26
Listing Code 4. 4 Data Split.....	27
Listing Code 4. 5 Data Preprocessing (Text Clening).....	28
Listing Code 4. 6 Encoding Labels	29
Listing Code 4. 7 Tokenization	29
Listing Code 4. 8 Model Configuration	30
Listing Code 4. 9 Arsitektur Model.....	31
Listing Code 4. 10 Model Training	34
Listing Code 4. 11 Theme Keywords.....	35
Listing Code 4. 12 Stanza Analysis	36
Listing Code 4. 13 Syllable and Word per Line Analysis	37
Listing Code 4. 14 Structur Analyzer.....	38
Listing Code 4. 15 Theme Analyzer	39
Listing Code 4. 16 Emotional Tone Analyzer	40
Listing Code 4. 17 Style Markers Analyzer	41
Listing Code 4. 18 Prediction	43
Listing Code 4. 19 Explanation.....	45
Listing Code 4. 20 Inisialisasi Class PoetComparativeAnalysis	46
Listing Code 4. 21 Poet Profiles	48
Listing Code 4. 22 Vektor Fitur.....	50
Listing Code 4. 23 Perhitungan Similarity.....	51
Listing Code 4. 24 Ekstrak Statistik Puisi.....	52
Listing Code 4. 25 Analyze Style Match	53
Listing Code 4. 26 Compare Themes.....	54
Listing Code 4. 27 Calculate Confidence	55
Listing Code 4. 28 Analyze Poem with Comparison	57
Listing Code 4. 29 Visualisasi Skor Kemiripan	59
Listing Code 4. 30 Poetry Visualization	60
Listing Code 4. 31 Word Frequency Visualization	62
Listing Code 4. 32 Evaluasi Model.....	68
Listing Code 4. 33 Visualisasi Training Metrik	69
Listing Code 4. 34 Predict New Poem	72
Listing Code 4. 35 Main Script.....	74
Listing Code 4. 36 Pengujian Prediksi Puisi.....	84

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada era yang terus berubah dan berkembang ini, karya sastra senantiasa menghadirkan pengalaman membaca yang menyenangkan. Meskipun ada pengaruh teknologi, karya sastra dapat diadaptasi untuk dinikmati oleh semua orang. Bahkan, karya sastra kuno pun tetap abadi dan relevan sepanjang waktu. Sastra telah mengubah perspektif individu terhadap kehidupan, mengajarkan cara menghadapi masalah, mengubah pola pikir, dan memahami moralitas zaman sekarang (Simaremare et al., 2023). Salah satu karya sastra yang biasanya digunakan sebagai sarana dalam menyampaikan pesan moral yaitu puisi.

Puisi adalah jenis karya sastra yang telah eksis sejak zaman dulu, disampaikan secara soliter dengan penggunaan kata-kata yang indah dan sarat makna (Hasminur et al., 2023). Puisi dibuat oleh seorang penyair untuk mengekspresikan perasaan melalui rangkaian kata-kata, menghasilkan ekspresi yang unik dan mendalam (Tiadilona et al., 2023). Setiap penyair memiliki sudut pandang masing-masing yang dapat diolah menjadi sebuah karya sastra yang memiliki nilai estetika. Selain penggunaan bahasa yang indah, puisi juga menggabungkan ide dan struktur makna yang ingin disampaikan oleh penyair (Dirman, 2022). Peran penting puisi dalam melestarikan budaya dan bahasa inilah yang menjadikan lembaga seperti Balai Bahasa Provinsi Kalimantan Barat begitu relevan.

Balai Bahasa Provinsi Kalimantan Barat adalah sebuah lembaga yang beroperasi di bawah otoritas dan bertanggung jawab kepada Badan Pengembangan dan Pembinaan Bahasa. Lembaga ini memiliki peran utama dalam menjaga penggunaan bahasa dan sastra Indonesia di wilayah kerjanya, serta melestarikan bahasa daerah yang kaya akan nilai-nilai budaya (Laporan Kinerja Balai Bahasa Kalimantan Barat, 2019). Salah satu bentuk sastra yang dijaga oleh Balai Bahasa adalah puisi, yang sering kali menjadi medium ekspresi budaya dan identitas lokal.

Dalam hal ini, Balai Bahasa tidak hanya melestarikan warisan budaya, tetapi juga mendukung para penyair lokal untuk terus berkarya dan mengekspresikan diri melalui puisi, memperkaya identitas budaya Kalimantan Barat di tengah dinamika masyarakat modern.

Puisi bagian dari warisan budaya tradisional yang dikenal memiliki makna mendalam, sentimen sederhana, dan gaya yang memukau. Namun, selama ini penilaian terhadap gaya puisi didasarkan pada persepsi dan pengetahuan pembaca yang subjektif, tanpa adanya aturan kuantitatif atau formal yang jelas. (Wang & Ni, 2022). Hal ini menimbulkan tantangan dalam mengidentifikasi pola penulisan dan karakteristik dari setiap penyair, terutama ketika karya-karya tersebut dibandingkan satu sama lain. Kurangnya pendekatan yang sistematis dan terstandarisasi dalam analisis puisi membuat interpretasi pola penulisan menjadi tidak konsisten. Masalah ini memerlukan pengklasifikasian pola penulisan puisi secara objektif dan terstruktur. Klasifikasi bertujuan untuk mengidentifikasi ciri khas setiap penyair berdasarkan pola penulisannya. Dalam melakukan klasifikasi penting memperhatikan metode yang digunakan serta kaitannya dengan studi kasus yang sedang dijalankan. Salah satu pendekatan yang dapat diterapkan dengan mengklasifikasikan pola penulisan puisi menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT). BERT adalah model pemahaman bahasa alami berbasis arsitektur transformator yang menggunakan *transfer learning* untuk membantu mesin memahami konteks frasa (Bello et al., 2023).

Melalui penelitian ini, dapat diketahui bahwa metode BERT dapat menjadi alat yang efektif untuk mengidentifikasi pola gaya penulisan puisi yang berbeda-beda antara penyair-penyair Indonesia. Dengan mengklasifikasikan puisi berdasarkan gaya penulisannya, maka dapat mengungkapkan keunikan dan ciri khas setiap penyair dalam konteks sastra Indonesia (Sulistiyo & Syihabuddin, 2022). *Transfer learning* menggunakan BERT akan memberikan wawasan mendalam tentang penggunaan bahasa, struktur kalimat, dan tema-tema yang muncul dalam karya-karya puisi. Penelitian ini diharapkan dapat mengembangkan integrasi antara ilmu pengetahuan dan teknologi informasi untuk mengapresiasi karya penyair Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah tersebut, beberapa rumusan masalah telah dirumuskan sebagai dasar untuk penelitian ini, yaitu :

1. Bagaimana menentukan karakteristik penulisan dari setiap penyair secara objektif dan konsisten?
2. Bagaimana cara mempelajari dataset puisi untuk mengidentifikasi pola penulisan yang khas dari masing-masing penyair secara efektif?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini meliputi hal-hal berikut :

1. Menerapkan klasifikasi pola penulisan puisi setiap penyair secara terukur sehingga memudahkan proses identifikasi berdasarkan ciri khas dalam karyanya.
2. Menerapkan model BERT untuk mempelajari dataset puisi secara mendalam sehingga dapat digunakan untuk memprediksi penyair berdasarkan pola penulisan yang khas.

1.4 Batasan Masalah

Batasan masalah yang ditetapkan dalam penelitian ini meliputi :

1. Penelitian ini berfokus pada klasifikasi pola penulisan penyair Indonesia
2. Metode yang digunakan adalah *Bidirectional Encoder Representations from Transformers* (BERT)
3. Melakukan *labeling data* menggunakan nama penyair puisi
4. Puisi yang digunakan hanya yang Bahasa Indonesia
5. Data diambil dari enam penyair Indonesia, yaitu Chairil Anwar, Joko Pinurbo, Sapardi Djoko Damono, M Aan Mansyur, Tia Ragat, dan Pradono.
6. Setiap penyair diwakili oleh 65 puisi dalam penelitian ini.
7. Menggunakan bahasa pemrograman *Python*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini antara lain adalah :

1. Bagi Peneliti

Manfaat yang diperoleh peneliti yaitu pemahaman mendalam tentang text mining dengan penerapannya pada metode *Bidirectional Encoder Representations from Transformers* (BERT). Penelitian ini dapat mengetahui gaya penulisan penyair Indonesia yang memiliki ciri khas berbeda.

2. Bagi Balai Bahasa Provinsi Kalimantan Barat

Manfaat yang diperoleh Balai Bahasa Provinsi Kalimantan Barat yaitu memperoleh wawasan mengenai pola penulisan puisi dari penyair-penyair lokal yang teridentifikasi dalam dataset dimiliki. Ini bisa menjadi kontribusi penting dalam mendokumentasikan dan mengarsipkan pola penulisan yang khas dari penyair-penyair Indonesia, terutama yang berada di wilayah Kalimantan Barat.

3. Bagi Bidang Ilmu Sistem Informasi

Manfaat bagi bidang ilmu sistem informasi adalah penelitian ini dapat menjadi referensi yang berguna untuk studi selanjutnya yang terkait dengan klasifikasi menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT).

1.6 Sistematika Penulisan

Struktur penulisan laporan tugas akhir mengikuti panduan format Sistem Informasi. Terdapat lima bab utama, dan setiap bab terdiri dari beberapa sub bab yang saling terkait. Berikut adalah sistematika penulisan:

BAB I PENDAHULUAN

Bab I Pendahuluan memberikan gambaran awal mengenai penelitian yang dilakukan, dimulai dari latar belakang yang menjelaskan terkait topik penelitian serta alasan dilaksanakannya studi ini. Selanjutnya, bagian rumusan masalah merumuskan pertanyaan-pertanyaan penelitian yang menjadi fokus utama studi. Bagian tujuan penelitian menguraikan hasil yang ingin dicapai untuk menjawab rumusan masalah tersebut. Pada batasan masalah menjelaskan ruang lingkup penelitian untuk memastikan fokus studi tetap terarah dan sesuai tujuan. Terakhir, bagian manfaat penelitian memaparkan kontribusi penelitian terhadap Balai Bahasa Provinsi Kalimantan Barat dan pengembangan ilmu pengetahuan. Bab ini secara keseluruhan memberikan dasar yang jelas dan terstruktur untuk memahami arah penelitian..

BAB II LANDASAN TEORI

Bab II Landasan Teori membahas dasar-dasar konseptual penelitian, termasuk *text mining*, *text processing*, dan klasifikasi untuk proses pengelompokan data berdasarkan pola tertentu. Selain itu, bagian ini menjelaskan metode utama yang digunakan yaitu *Bidirectional Encoder Representations from Transformers* (BERT), yang menjadi inti dalam analisis pola penulisan puisi serta alat implementasinya seperti *Google Colaboratory*. Bagian ini juga menjelaskan tentang puisi dan memaparkan konteks penelitian termasuk informasi Balai Bahasa Provinsi Kalimantan Barat sebagai penyedia dataset. Konsep puisi seperti struktur stilistika untuk analisis karakteristik penulisan dan *Cosine Similarity* digunakan untuk mengukur kemiripan teks. Di bagian akhir, kajian penelitian mengulas penelitian sebelumnya untuk memperkuat landasan teori dan posisi penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab III Metodologi Penelitian menjelaskan secara rinci pendekatan dan langkah-langkah yang digunakan dalam penelitian ini. Bagian bahan penelitian menguraikan data dan material yang digunakan untuk mendukung analisis, sementara alat penelitian membahas perangkat lunak dan alat yang digunakan dalam proses penelitian. Kerangka kerja penelitian menggambarkan alur dan

tahapan penelitian secara sistematis untuk mencapai tujuan penelitian. Pada bagian metode penelitian dijelaskan tahapan-tahapan yang dilakukan pada penelitian. Selanjutnya, bagian data penelitian menguraikan detail seperti objek penelitian, populasi dan sampel, jenis dan sumber data, dan metode pengumpulan data yang mencakup teknik pengambilan dan pengolahan data. Bab ini bertujuan memberikan panduan yang terstruktur tentang cara penelitian dilakukan sehingga dapat dipahami oleh peneliti lain.

BAB IV HASIL DAN PEMBAHASAN

Bab ini Bab IV Hasil dan Pembahasan menyajikan hasil penelitian secara terstruktur, dimulai dari studi literatur sebagai dasar teoretis dan diikuti oleh proses teknis seperti *data collection*, *install* dan *import packages*, serta *load dataset*. Penjelasan mendalam diberikan pada penerapan model BERT untuk analisis pola penulisan puisi, *stylistic analysis* untuk identifikasi karakteristik, dan *Prediction of Poetry* untuk memprediksi penyair berdasarkan pola penulisan. Bab ini juga mencakup *comparative analysis* untuk membandingkan hasil, *visualizations* untuk menampilkan data, *evaluation* untuk menilai kinerja model, dan pengujian dan analisis hasil prediksi puisi untuk mengukur dan menginterpretasikan keakuratan model terhadap dataset.

BAB V PENUTUP

Bab ini menyajikan kesimpulan dari penelitian yang telah dilakukan serta memberikan saran yang berkaitan dengan pembahasan yang telah dijelaskan pada bab-bab sebelumnya. Selain itu, disampaikan saran pengembangan penelitian di masa mendatang.