

## Pengembangan Model Klasifikasi Teks Berbasis CNN untuk Deteksi Wuku Kurantil

Dika Aurelya Aleandra Taroreh <sup>1</sup>, Didik Dwi Prasetya <sup>2</sup>

1. Universitas Negeri Malang 1, Indonesia 1 | [dika.aurelya.2205356@students.um.ac.id](mailto:dika.aurelya.2205356@students.um.ac.id) 1
2. Universitas Negeri Malang 1, Indonesia 2 | [didikdwi@um.ac.id](mailto:didikdwi@um.ac.id) 2

### Abstrak

Penelitian ini bertujuan mengembangkan model klasifikasi teks berbasis *Convolutional Neural Network* (CNN) untuk mendeteksi keterkaitan kalimat dalam bahasa Jawa dengan Wuku Kurantil, bagian dari sistem kalender tradisional Pawukon yang terekam dalam naskah Primbon. Permasalahan utama yang diangkat adalah minimnya pemahaman generasi muda terhadap sistem penanggalan tradisional akibat keterbatasan akses dan format digitalisasi konten budaya lokal. Penelitian menggunakan pendekatan *Conceive, Design, Implement, Operate* (CDIO) untuk membangun sistem klasifikasi otomatis. Data diperoleh dari situs sastra.org, kemudian diproses melalui tahapan pembersihan data, normalisasi, dan penghapusan *stopwords*, diikuti vektorisasi menggunakan Word2Vec. Arsitektur CNN dirancang dengan parameter optimal dan dilatih menggunakan *train-test split* (80:20). Hasil evaluasi model menunjukkan akurasi sebesar 80%, dengan *precision* dan *recall* yang tinggi pada kelas mayoritas (bukan Wuku Kurantil), namun nihil pada kelas minoritas (Wuku Kurantil), mengindikasikan adanya ketimpangan kinerja akibat distribusi data yang tidak seimbang. Untuk mengatasi hal tersebut, direkomendasikan penerapan *oversampling*, *class weighting*, atau penggunaan algoritma yang lebih adaptif seperti *Random Forest* atau *ensemble models*. Meskipun demikian, model yang dikembangkan telah mampu menunjukkan potensi dalam digitalisasi budaya berbasis teks Jawa. Studi ini memberikan kontribusi terhadap pelestarian warisan budaya lokal melalui pemanfaatan teknologi *Natural Language Processing* (NLP), serta membuka peluang pengembangan lebih lanjut di bidang *digital humanities* dan linguistik komputasional.

### Kata Kunci

NLP, CNN, Klasifikasi Teks, Wuku

## 1. Pendahuluan

Dalam khazanah budaya Jawa terdapat warisan intelektual yang dikenal luas di Nusantara yakni Primbon Jawa. Salah satu aspek penting dari primbon adalah sistem penanggalan tradisional yang dikenal sebagai Pawukon (Sidhartani et al., 2023). Sistem ini merupakan siklus waktu sepanjang 210 hari, terbagi dalam 30 bagian yang disebut wuku. Salah satu wuku tersebut adalah Wuku Kurantil, yang dipercaya membawa makna spiritual dan karakter tersendiri dalam kepercayaan masyarakat Jawa. Pawukon digunakan oleh masyarakat Jawa sebagai panduan untuk menentukan hari-hari baik dalam berbagai kegiatan, seperti perkawinan, khitanan, masa bercocok tanam, hingga pembangunan rumah. Dalam masyarakat Jawa, penentuan hari baik diyakini memiliki peran penting untuk menghindarkan kesialan dan mendukung kelancaran suatu acara. Namun, seiring perkembangan zaman, kekayaan budaya ini semakin jarang dipahami oleh generasi muda karena terbatasnya media modern yang menjelaskan konsep tersebut dengan cara yang relevan dan mudah diakses.

Hal ini menjadi sebuah tantangan bagi pelestarian budaya lokal, termasuk wuku, umumnya terdokumentasi dalam teks-teks tradisional seperti Primbon Jawa. Teks-teks tradisional seperti Primbon Jawa yang memuat informasi mengenai wuku umumnya belum terdigitalisasi secara sistematis dan belum tersedia sistem otomatis yang mampu mengklasifikasikan konten berdasarkan wuku, termasuk Wuku Kurantil. Dilansir dari [tribunnews.com](http://tribunnews.com) Kriesinar menyatakan bahwa pengetahuan tentang Pakuwon serta Primbon Jawa nyaris terkubur oleh derasnya alur kebarat-baratan di kalangan masyarakat Jawa. Fenomena tersebut menunjukkan bahwa kebudayaan ini memiliki potensi terancam kehilangan relevansi di tengah kehidupan modern.

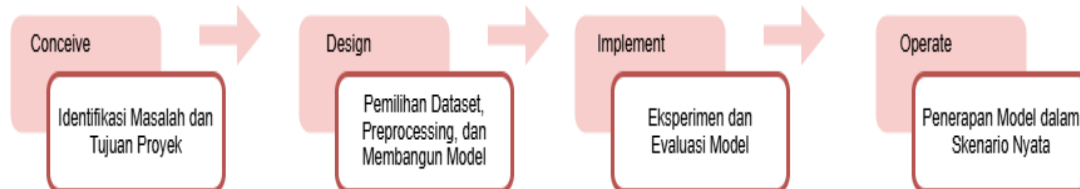
Sebagai upaya pelestarian budaya, perkembangan teknologi *Natural Language Processing* (NLP) yang merupakan cabang dari *Machine learning* (ML) telah memberikan kemampuan bagi komputer untuk memahami, memproses, dan menganalisis bahasa manusia (Sumarna & Riadi, n.d.). Dengan menggunakan teknologi NLP yang dipadukan dengan *Convolutional Neural Networks* (CNN) dapat digunakan untuk mengenali pola linguistik dalam teks tradisional. CNN efektif dalam mengekstraksi ciri penting dari teks dan dapat dioptimalkan untuk mengklasifikasikan wuku berdasarkan data dari teks sastra Jawa. Melalui pelatihan pada korpus lokal yang diperoleh dari sumber digital seperti [sastra.org](http://sastra.org), model CNN dipadukan dengan pendekatan NLP untuk membentuk sistem klasifikasi berbasis teks. Sistem ini dirancang untuk menerima masukan berupa kalimat dan menghasilkan keluaran berupa prediksi apakah teks tersebut berkaitan dengan Wuku Kurantil.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi wuku yang bersifat akurat, efisien, dan kontekstual, serta berkontribusi dalam digitalisasi warisan budaya Jawa melalui pendekatan komputasional. Dengan membangun sistem klasifikasi otomatis yang mampu mengenali struktur bahasa tradisional secara kontekstual, studi ini tidak hanya mendukung pelestarian nilai-nilai budaya lokal, tetapi juga memperluas cakupan penerapan teknologi NLP dalam pemrosesan bahasa daerah. Hasil penelitian ini diharapkan dapat menjadi fondasi bagi

pengembangan lebih lanjut dalam digital humanities dan aplikasi linguistik komputasional berbasis kearifan lokal (Dharma et al., 2022; Fan et al., 2021).

## 2. Metodologi Penelitian

Penelitian ini menerapkan kerangka CDIO (*Conceive, Design, Implement, Operate*) seperti Gambar 1 sebagai pendekatan sistematis dalam pengembangan model klasifikasi teks berbasis CNN. CDIO dikenal sebagai metode rekayasa berorientasi proyek yang mencakup seluruh siklus pengembangan produk dan sistem secara berkelanjutan, serta relevan dalam konteks pendidikan teknik (Muludi et al., 2023). Penerapan metode CDIO diharapkan dapat memberikan alur pembelajaran yang berskema dan terstruktur sehingga dapat terlihat penerapannya pada kurikulum (Kurniadi et al., 2021).



Gambar 1. Alur Metode CDIO

### 1) *Conceive*

Tahap *conceive* merupakan proses awal untuk mengidentifikasi kebutuhan dan perumusan masalah (Wesley & Gunawan, 2024). Dalam konteks ini, penurunan pemahaman terhadap sistem kalender Pawukon, khususnya Wuku Kurantil, menjadi isu utama yang diangkat. Informasi mengenai wuku umumnya tersebar dalam teks-teks tradisional berbahasa Jawa, seperti Primbon, yang bersifat kompleks dan sulit dipahami tanpa latar belakang budaya. Menyikapi hal tersebut, penelitian ini bertujuan mengembangkan model klasifikasi otomatis berbasis CNN yang dapat mengidentifikasi apakah suatu kalimat berkaitan dengan Wuku Kurantil atau tidak. Pengembangan model ini diharapkan menjadi kontribusi nyata dalam digitalisasi pengetahuan budaya lokal serta memperluas penerapan NLP pada bahasa daerah.

### 2) *Design*

Tahap desain merupakan fase strategis dalam pengembangan sistem klasifikasi, dengan fokus pada perumusan solusi teknis terhadap permasalahan yang telah diidentifikasi (Gori et al., 2024). Proses ini mencakup analisis kebutuhan data, perencanaan pengolahan informasi, serta perancangan arsitektur model klasifikasi. Dataset yang digunakan bersumber dari situs sastra.org, yang memuat naskah-naskah digital berbahasa Jawa seperti Primbon dan sistem kalender Pawukon. Meskipun tersedia dalam bentuk digital, seleksi dan validasi tetap dilakukan

untuk memastikan data merepresentasikan konsep yang relevan, khususnya Wuku Kurantil sebagai objek klasifikasi.

Setelah pengumpulan data, dilakukan tahap pra-pemrosesan berupa normalisasi teks, pembersihan karakter non-alfabet, penghapusan *stopwords*, pelabelan, dan integrasi data (Martantoh & Yanih, 2022). Data yang telah dibersihkan kemudian direpresentasikan dalam bentuk vektor menggunakan Word2Vec, untuk menangkap relasi semantik antar kata dalam korpus berbahasa Jawa. Model klasifikasi dikembangkan menggunakan algoritma CNN yang dirancang secara optimal dengan mempertimbangkan parameter seperti jumlah filter, ukuran kernel, dan metode pooling. Evaluasi performa model dilakukan melalui metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Model dengan kinerja terbaik disimpan sebagai sistem klasifikasi otomatis untuk mendeteksi keberadaan referensi terhadap Wuku Kurantil dalam input teks secara efisien dan akurat.

### 3) Implement

Tahap implementasi merupakan proses realisasi rancangan solusi ke dalam bentuk program yang dijalankan menggunakan bahasa Python (Widhiyasana et al., 2021). Pada tahap ini, eksperimen dilakukan dengan mengembangkan kode program sesuai dengan strategi teknis yang telah ditentukan. Implementasi bertujuan untuk menguji efektivitas model dalam menyelesaikan permasalahan klasifikasi berdasarkan data yang telah dipersiapkan sebelumnya. Adapun beberapa proses yang perlu diikuti dalam tahap ini antara lain:

#### a. Pemilihan Dataset

Dataset yang digunakan diperoleh dari situs *sastra.org*, mencakup naskah *Primbon Jawa* dan kalender *Pawukon* dalam format digital. Dataset terdiri dari 49 instance dengan dua atribut utama: "Katrangan" sebagai prediktor dan "Title" sebagai label awal. Untuk menyederhanakan klasifikasi dan meningkatkan akurasi model, label "Title" yang semula terdiri dari 30 kelas dikategorikan ulang menjadi dua kelas biner: *is\_wuku\_kurantil* dan *is\_not\_wuku\_kurantil*, yang direpresentasikan secara numerik (1 dan 0). Hasil akhir klasifikasi terdiri dari 20 instance berlabel "1" dan 29 instance berlabel "0", yang selanjutnya digunakan dalam pelatihan dan evaluasi model klasifikasi. Daftar atribut dalam dataset dapat dilihat dalam Tabel 1. Proporsi jumlah instance masing-masing kelas dalam dataset dapat dilihat dalam Tabel 2.

**Tabel 1.** Daftar atribut dalam dataset

No	Nama atribut	Penjelasan atribut	Tipe data	Rentang nilai / enumerasi
1	Katrangan	Atribut prediktor	nominal	Katrangan berupa kalimat yang terkandung dalam wuku
2	<i>is_wuku_kurantil</i>	Atribut kelas	numerik	(1), (0)

**Tabel 2.** Proporsi jumlah instance masing-masing kelas dalam dataset

Label Kelas	Jumlah Instance	Persentase
Kelas "1"	20	40,82%
Kelas "0"	29	59.18%

## b. Preprocessing

Proses ini meliputi beberapa langkah, seperti pembersihan data (*Cleaning Data*) untuk menghapus unsur-unsur yang tidak penting atau gangguan dalam teks. *Cleaning Data* juga bertujuan untuk meningkatkan kualitas data dengan imputasi nilai yang hilang dan penghapusan outlier (Aisah et al., 2023). Selain itu, pre-processing juga dilakukan dengan menggunakan cara *stopword removal* untuk menyederhanakan dan meningkatkan kualitas data. Pembangunan data operasional biasanya berkualitas buruk dan karenanya, preprocessing data sering kali diperlukan untuk memastikan keandalan data analisis dengan menggunakan berbagai teknik. Dengan tahapan ini, kumpulan data yang digunakan menjadi lebih siap dan cocok untuk analisis lebih lanjut, serta mempermudah model dalam menemukan pola yang ada.

## c. *Data Vectorization and Standardization* Menggunakan Word2Vec

Pada tahap ini, data teks hasil pra-pemrosesan dikonversi ke bentuk numerik melalui proses vektorisasi dan standarisasi. Vektorisasi mengubah susunan kata dalam teks menjadi vektor angka, sementara standarisasi memastikan keseragaman format data agar sesuai dengan kebutuhan pemodelan. Standar yang digunakan mengacu pada hasil kajian sebelumnya, sehingga data yang digunakan selaras dengan karakteristik algoritma klasifikasi yang akan diterapkan (Intan Af et al., 2021).

Proses ini menggunakan model Word2Vec untuk menghasilkan representasi vektor kata yang mencerminkan hubungan semantik antar kata dalam korpus. Setiap kata dalam data diubah menjadi vektor numerik berdasarkan model Word2Vec yang telah dilatih. Kata yang tidak terdapat dalam model akan diabaikan, dan padding bernilai nol ditambahkan untuk menyamakan panjang vektor antar dokumen sesuai batas maksimal yang ditentukan. Pendekatan ini memungkinkan model pembelajaran mesin memahami konteks kata dalam teks berbahasa Jawa secara efektif.

## d. Klasifikasi dengan Algoritma CNN

Proses pelatihan model CNN menggunakan pembagian dataset menjadi data latih (80%) dan data uji (20%) melalui metode *train-test split*. *Split data* adalah teknik yang digunakan untuk mempartisi dataset adalah salah satu dari beberapa aspek yang mempengaruhi seberapa baik kinerja model klasifikasi pada algoritma pembelajaran mesin (Nurhopipah & Hasanah, 2020). Data training digunakan untuk menghasilkan model yang sesuai dari proses pelatihan algoritma. Sedangkan data testing digunakan untuk pengujian dan penentuan kinerja model yang diperoleh dari hasil pengujian algoritma. Pembagian ini dilakukan guna menghindari *overfitting*, yaitu situasi di mana model bekerja sangat baik pada data latih tetapi kurang efektif dalam melakukan prediksi

pada data baru. Proses validasi sangat penting untuk dilakukan, tujuannya agar setiap data memiliki peluang sebagai pelatihan data dan pengujian data.

Dalam membangun model CNN diperlukan parameter-parameter yang sesuai dengan kebutuhan. Parameter-parameter yang digunakan dalam membangun arsitektur CNN dirancang untuk mengoptimalkan proses ekstraksi fitur dan klasifikasi data sekuensial. Filter dan ukuran kernel pada lapisan convolutional menentukan kemampuan model dalam menangkap pola lokal, sementara padding dan strides memastikan dimensi data tetap sesuai untuk pemrosesan lebih lanjut. GlobalMaxPooling digunakan untuk mereduksi dimensi data tanpa kehilangan informasi penting, sehingga meningkatkan efisiensi komputasi. Jumlah node pada lapisan dense serta dropout probability dipilih untuk menyeimbangkan kapasitas pembelajaran dan mencegah overfitting. Fungsi aktivasi relu pada lapisan hidden memperkenalkan non-linearitas, sedangkan sigmoid pada lapisan output menghasilkan probabilitas untuk tugas klasifikasi biner. Tabel 3 memperlihatkan parameter yang digunakan dalam model CNN.

**Tabel 3.** Parameter-parameter untuk membangun model algoritma CNN

Parameter	Nilai
Filters	4
Kernel_size	3
Padding	'same'
Strides	1
Activation (Conv1D)	relu
Input_shape	72 (max_length), 300 (vector_size)
Pooling	GlobalMaxPooling1D
Hidden_layer_1_nodes	8
Hidden_layer_2_nodes	4
Dropout_prob	0.2
Output_layer	Aktivasi dengan Sigmoid
Num_Epochs	100
Batch_size	49 (jumlah data)
Filters	4

Model dilatih selama 100 epoch, yang berarti seluruh dataset latih diproses sebanyak 100 kali. Satu epoch berarti sebuah algoritma *machine learning* telah 'belajar' dari data training secara keseluruhan [29]. Jumlah epoch yang cukup memberikan model kesempatan untuk mempelajari pola dalam data secara menyeluruh. Namun, diperlukan keseimbangan untuk mencegah overfitting, di mana model terlalu terfokus pada data latih sehingga performanya menurun pada data uji. Selain itu, ukuran batch yang digunakan adalah 49 yaitu sesuai dengan jumlah data, yang mengacu pada jumlah sampel yang diproses dalam satu iterasi. Batch size merupakan jumlah sampel data pelatihan yang digunakan pada setiap iterasi (Islam Riau et al., 2022). Pemilihan batch size ini bertujuan untuk mempercepat proses pelatihan sekaligus menjaga

kestabilan performa model. Proses optimasi menggunakan algoritma Adam dengan fungsi *loss binary crossentropy*, memastikan pembaruan bobot yang efisien selama proses pelatihan.

#### e. Evaluasi Model

Pada tahap ini, dilakukan evaluasi untuk menilai performa model dalam mengklasifikasikan dataset menjadi dua kelas, yaitu 1 untuk Wuku Kurantil dan 0 untuk bukan Wuku Kurantil. Evaluasi dilakukan dengan menganalisis *confusion matrix*, yang mencakup penghitungan akurasi serta metrik tambahan seperti *precision*, *recall*, dan *F1-score* untuk memberikan gambaran menyeluruh tentang performa model. Pada tahap *confusion matrix* akan menghasilkan nilai *accuracy*, *precision*, *recall*, dan *F1-Score* (Buda et al., 2018). Akurasi adalah proporsi dari total prediksi true dari semua data. *Precision* adalah ukuran ketepatan dari hasil suatu model. *Recall* adalah ukuran kelengkapan dari sebuah model. *F-measure* merupakan harmonic mean dari *precision* dan *recall*. Langkah ini bertujuan untuk memastikan bahwa model yang dibangun tidak hanya akurat, tetapi juga dapat diandalkan dalam menangani distribusi kelas yang tidak seimbang.

Matriks ini mencakup empat komponen utama: *True Positive* (TP), yaitu jumlah instance yang benar-benar positif dan diklasifikasikan sebagai positif; *True Negative* (TN), yaitu jumlah instance yang benar-benar negatif dan diklasifikasikan sebagai negatif; *False Positive* (FP), yaitu jumlah instance yang sebenarnya negatif tetapi salah diklasifikasikan sebagai positif; dan *False Negative* (FN), yaitu jumlah instance yang sebenarnya positif tetapi salah diklasifikasikan sebagai negatif. Berdasarkan komponen-komponen ini, kita dapat menghitung matrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Pers. (1) – Pers. (4) merupakan persamaan yang digunakan untuk menilai performa model klasifikasi.

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \quad \text{Pers. (1)}$$

$$Presisi = \frac{TP}{TP+FP} \quad \text{Pers. (2)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{Pers. (3)}$$

$$F - 1 \text{ Score} = 2x \frac{(Recall*Precision)}{(Recall+Precision)} \quad \text{Pers. (4)}$$

#### 4) Operate

Tahap operate dalam kerangka CDIO mencakup implementasi dan pengujian sistem klasifikasi yang telah dirancang. Pada tahap ini, model CNN yang telah dilatih diintegrasikan ke dalam program klasifikasi berbasis teks, yang berfungsi menerima input kalimat dan menghasilkan output berupa prediksi keterkaitan dengan Wuku Kurantil. Program ini dirancang agar dapat beroperasi secara mandiri, sebagai wujud nyata penerapan model dalam konteks penggunaan praktis.

Proses klasifikasi diawali dengan tahapan preprocessing, seperti pembersihan karakter tidak relevan, penghapusan *stopwords*, dan normalisasi teks. Selanjutnya, teks dikonversi menjadi bentuk numerik melalui teknik vectorization menggunakan Word2Vec. Representasi ini diproses oleh model CNN untuk menghasilkan prediksi probabilistik. Apabila nilai probabilitas melebihi ambang batas tertentu, sistem akan mengklasifikasikan teks sebagai termasuk Wuku Kurantil; jika tidak, teks dikategorikan sebagai bukan bagian dari wuku tersebut. Dengan pendekatan ini, sistem mampu melakukan klasifikasi secara otomatis, efisien, dan konsisten terhadap input teks berbahasa Jawa.

### 3. Hasil dan Pembahasan

Dalam penelitian ini, digunakan algoritma CNN dalam membangun model yang dirancang sebagai sistem klasifikasi otomatis yang menerima input berupa kalimat berbahasa Jawa dan menghasilkan output berupa prediksi biner, yaitu apakah kalimat tersebut mengandung referensi terhadap Wuku Kurantil atau tidak. Sistem ini bekerja melalui tahapan preprocessing teks, vektorisasi menggunakan Word2Vec, dan klasifikasi menggunakan CNN yang telah dilatih.

Evaluasi model dilakukan menggunakan confusion matrix dan metrik turunan seperti precision, recall, dan F1-score, dengan hasil sebagaimana dalam Tabel 4.

**Tabel 4.** Evaluasi Model Menggunakan Confusion Matrix

Kelas	Precision	Recall	F1-Score	Support
0 (bukan Kurantil)	80%	100%	88,89%	8
1 (Kurantil)	0%	0%	0%	2
<b>Akurasi total</b>	80%	80%	80%	10

Hasil ini menunjukkan bahwa model memiliki performa tinggi dalam mengklasifikasikan teks yang tidak mengandung Wuku Kurantil, namun gagal sepenuhnya dalam mengenali teks yang termasuk kelas minoritas (Wuku Kurantil). *Recall* sebesar 100% pada kelas mayoritas (kelas 0) mengindikasikan bahwa semua data non-Kurantil dapat diprediksi dengan benar, sedangkan precision sebesar 80% menunjukkan bahwa sebagian besar prediksi tersebut juga akurat. Sebaliknya, metrik performa pada kelas 1 (Wuku Kurantil) bernilai nol, menandakan model tidak mampu mengenali pola dari kelas minoritas secara efektif. Ketimpangan performa ini disebabkan oleh class imbalance, yaitu ketidakseimbangan distribusi data antar kelas yang membuat model bias terhadap kelas mayoritas. Meskipun akurasi keseluruhan cukup tinggi, metrik tersebut kurang representatif dalam konteks distribusi data yang tidak seimbang. Oleh karena itu, *precision*, *recall*, dan *F1-score* menjadi metrik yang lebih tepat untuk menilai kualitas model dalam skenario ini.

Untuk meningkatkan performa klasifikasi terhadap kelas minoritas, perlu dilakukan penyesuaian pada data atau arsitektur model. Strategi yang disarankan meliputi oversampling pada kelas minoritas (misalnya dengan metode SMOTE), undersampling pada kelas mayoritas,

atau penggunaan class weighting pada fungsi loss untuk memberikan bobot lebih pada kelas yang jarang. Alternatif lainnya adalah menggunakan algoritma yang lebih adaptif terhadap data tidak seimbang, seperti Random Forest atau ensemble models. Dengan penerapan pendekatan-pendekatan tersebut, diharapkan model dapat lebih adil dan akurat dalam melakukan klasifikasi dua arah, sehingga kinerjanya meningkat secara menyeluruh. Luaran dari penelitian ini adalah model klasifikasi teks berbasis CNN yang dapat digunakan secara mandiri sebagai komponen sistem digitalisasi budaya berbasis NLP untuk teks Jawa.

#### 4. Kesimpulan

Pada tahap implementasi, eksperimen dilakukan dengan mengembangkan model klasifikasi berbasis CNN untuk memprediksi teks terkait Wuku Kurantil, menggunakan dataset dari situs sastra.org. Dataset yang terdiri dari 49 instance diproses dengan langkah preprocessing, termasuk pembersihan data dan penghilangan stopwords. Data kemudian dianalisis menggunakan Word2Vec untuk vektorisasi dan standarisasi. Model CNN dipilih karena kinerjanya yang superior dibandingkan algoritma lain seperti Naïve Bayes, SVM, dan ANN, dengan akurasi mencapai 80%. Data latih dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, dengan hasil konvolusi menggunakan kernel dan stride untuk mengidentifikasi pola dalam data teks. Evaluasi model menggunakan confusion matrix menunjukkan akurasi 80%, tetapi dengan ketimpangan performa antara kelas mayoritas (bukan wuku kurantil) dan minoritas (wuku kurantil), dengan hasil 0% pada kelas minoritas. Untuk mengatasi masalah ini, solusi seperti oversampling atau penggunaan algoritma adaptif terhadap ketidakseimbangan kelas disarankan. Hasil implementasi chatbot untuk klasifikasi wuku kurantil mampu memberikan respons yang baik, meskipun perbaikan lebih lanjut diperlukan untuk meningkatkan akurasi pada kelas minoritas.

#### Daftar Rujukan

- Aisah, I. S., Irawan, B., & Suprpti, T. (2023). Algoritma Support Vector Machine (SVM) untuk Analisis Sentimen Ulasan Aplikasi AI Qur'an Digital. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).
- Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- Dharma, E. M., Lumban Gaol, F., Leslie, H., Warnars, H. S., & Soewito, B. (2022). The Accuracy Comparison Among Word2vec, Glove, and Fasttext Towards Convolution Neural Network (Cnn) Text Classification. *Journal of Theoretical and Applied Information Technology*, 31(2). [www.jatit.org](http://www.jatit.org)
- Fan, C., Chen, M., Wang, X., Wang, J., & Huang, B. (2021). A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational

- Data. In *Frontiers in Energy Research* (Vol. 9). Frontiers Media S.A. <https://doi.org/10.3389/fenrg.2021.652801>
- Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Intan Af, D., Febbi Handayani, S., Wijayatun Pratiwi, R., Teknik Informatika, J., Harapan Bersama, P., Mataram No, J., & Lor, P. (2021). *Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen*. 6(3).
- Islam Riau, U., Kaharuddin Nasution, J., Marpoyan, P., Marpoyan Damai, K., & Pekanbaru, K. (2022). Standarisasi Kualitas Riset di Sekolah Dasar Zaka Hadikusuma Ramadan. *Journal on Education*, 05(02), 1932–1939.
- Kurniadi, B. W., Prasetyo, H., Ahmad, G. L., Aditya Wibisono, B., & Sandya Prasvita, D. (2021). Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah. In *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*.
- Martantoh, E., & Yanih, N. (2022). Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP MySQL Implementation of Naive Bayes Method for Classification of Student's Personality Characteristics at MTS Darussa'adah School Using PHP Mysql. In *JTSI* (Vol. 3, Issue 2).
- Muludi, K., Naufal Humam, M., Shofiana, D. A., Syarif, A., Komputer, J. I., & Lampung, U. (2023). Perbandingan Kinerja CNN dan Naïve Bayes pada Analisis Sentimen Performa Manchester United di Twitter. In *Journal Information Engineering and Educational Technology* (Vol. 07).
- Nurhopipah, A., & Hasanah, U. (2020). Dataset Splitting Techniques Comparison For Face Classification on CCTV Images. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(4), 341. <https://doi.org/10.22146/ijccs.58092>
- Sidhartani, S., Sugihartono, R. A., & Suyanto, S. (2023). Ekspresi Budaya Jawa pada Figur Dewa dalam Ilustrasi Pawukon. *Jurnal Desain*, 11(1), 171. <https://doi.org/10.30998/jd.v11i1.17376>
- Sumarna, A. D., & Riadi, S. (n.d.). *Simposium Nasional Akuntansi Vokasi (SNAV) XII 14-17 Mei Tahun 2024 Politeknik Negeri Jember Project Based Learning CDIO: Sebuah Metode Integrated Learning Untuk Program Studi Akuntansi*.
- Wesley, R., & Gunawan, R. (2024). Literatur Review: Metode Deep Learning untuk Analisis Teks. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 5).
- Widhiyasana, Y., Semiawan, T., Gibran, I., Mudzakir, A., & Noor, M. R. (2021). Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia (Convolutional Long Short-Term Memory Implementation for Indonesian News Classification). In *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 10, Issue 4).