

BAB I

PENDAHULUAN

Bab ini mencakup latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup, dan sistematika penulisan sebagai landasan laporan skripsi.

1.1 Latar Belakang

Artificial Intelligence (AI) telah mengalami kemajuan pesat dalam beberapa tahun terakhir, membawa dampak signifikan di berbagai sektor, termasuk pendidikan. Salah satu manifestasi utama dari penerapan AI dalam pendidikan adalah pengembangan dan penggunaan *generative AI*. Penggunaan *generative AI* dalam pendidikan membawa manfaat seperti peningkatan keterlibatan siswa, personalisasi pembelajaran, dan aksesibilitas sepanjang waktu (Ali dkk., 2024; Kasneci dkk., 2023). *Generative AI* memungkinkan pengalaman belajar yang lebih menarik dan fleksibel bagi siswa, menjadikannya alat yang potensial untuk mendukung sistem pendidikan modern.

Maraknya penggunaan *generative AI* oleh pelajar juga menimbulkan kekhawatiran terkait plagiarisme, keadilan, dan integritas akademik. *Generative AI* mampu menghasilkan teks yang berpotensi digunakan pelajar sebagai karya mereka sendiri, sebagaimana dilaporkan oleh Cotton dkk. (2024). Data dari Kompas.com mengungkap bahwa antara 30% hingga 83% kasus plagiarisme terdeteksi pada periode Agustus 2019 hingga Oktober 2020 (Sahrani, 2020), sementara survei menunjukkan bahwa pelajar menyadari risiko integritas akademik saat menggunakan *generative AI* (Prokhorova dkk., 2024). Kekhawatiran ini diperkuat oleh Peraturan Menteri Pendidikan RI Nomor 17 Tahun 2010 tentang pencegahan plagiat, yang merujuk pada Undang-Undang Hak Cipta Nomor 19 Tahun 2002 terkait sanksi bagi pelaku plagiarisme (Menteri Pendidikan Nasional, 2010).

Penelitian yang ada telah mengembangkan model klasifikasi teks untuk mengatasi kekhawatiran tersebut dengan membedakan teks buatan manusia dan *generative AI*. Penelitian terdahulu oleh Islam dkk. (2023) menggunakan teknik ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma

classical Machine Learning (ML) seperti *Logistic Regression* (LR), *Support Vector Machine* (SVM), *Decision Tree*, *K-Nearest Neighbor* (KNN), *Random Forest* (RF), *AdaBoost*, *Bagging Classifier*, *Gradient Boosting*, *Extremely Randomized Trees*, serta algoritma *deep learning* seperti *Multi-layer Perceptron* (MLP), *Long Short-Term Memory* (LSTM) (Islam dkk., 2023). Hasil penelitian menunjukkan bahwa *Extremely Randomized Trees* menghasilkan akurasi terbaik sebesar 77% yang mana hasil tersebut mengungguli MLP dan LSTM

Penelitian lain oleh Vora dkk. (2023) dengan ekstraksi fitur berbasis *knowledge graph*, serta algoritma *Random Forest*, SVM, dan *Bidirectional Encoder Representations from Transformers* (BERT) untuk *classifier*-nya. Hasil menunjukkan bahwa BERT mampu mencapai akurasi sebesar 89,99%. Meskipun penggunaan *deep learning* yang canggih seperti BERT menunjukkan potensi yang baik, namun performa optimalnya belum sepenuhnya tercapai. Hal ini disebabkan oleh kebutuhan yang besar akan data, sumber daya komputasi, serta waktu pelatihan yang lama (Li dkk., 2022). Sebaliknya, algoritma *classical ML*, meskipun lebih efisien secara komputasi, sering kali kurang efektif dalam menangkap pola bahasa yang kompleks (Pearl, 2018).

Arsitektur *One Dimensional Convolutional Neural Network* (CNN) menawarkan keseimbangan antara efisiensi komputasi dan kemampuan mengenali pola kompleks, menjadikannya alternatif yang efektif untuk klasifikasi teks dibandingkan *classical machine learning*, BERT, atau LSTM (Romero dkk., 2022). Penelitian terdahulu oleh Fesseha dkk. (2021) menggunakan *static word embedding* untuk ekstraksi fitur dan model CNN sebagai *classifier*-nya menunjukkan performa unggul untuk klasifikasi teks dengan akurasi 93,41% (Fesseha dkk., 2021) dan bahkan 100% untuk deteksi berita palsu menggunakan teknik ekstraksi fitur TF-IDF (Kamel & Waleed, 2024). Penelitian yang ada tersebut menunjukkan potensi CNN untuk memberikan hasil yang unggul tanpa perlu menggabungkan arsitektur lain yang lebih kompleks.

Model yang dihasilkan dari penelitian terdahulu dalam klasifikasi teks buatan manusia dan AI sejauh ini belum mencapai performa optimal. Masih ada peluang untuk mengeksplorasi model lebih lanjut guna meningkatkan akurasi. Oleh karena itu, penelitian ini dilakukan untuk mengembangkan model klasifikasi teks buatan manusia

dan AI berbasis CNN. Model ini dirancang agar dapat mengungguli metode *classical machine learning*, tetapi tetap lebih sederhana dibandingkan arsitektur yang lebih kompleks seperti BERT atau LSTM. Dalam penelitian ini, CNN dilakukan pengaturan *hyperparameter* untuk mendapatkan performa yang optimal. Hipotesis yang diajukan adalah pengaturan *hyperparameter* dapat meningkatkan performa model.

1.2 Rumusan Masalah

Rumusan masalah yang diajukan dalam penelitian ini adalah bagaimana kombinasi *hyperparameter* terbaik pada model CNN untuk mengklasifikasikan teks buatan manusia versus *generative AI*?

1.3 Tujuan

Tujuan dari penelitian ini adalah mendapatkan kombinasi *hyperparameter* terbaik pada model CNN untuk mengklasifikasikan teks buatan manusia versus *generative AI*.

1.4 Manfaat

Manfaat yang diperoleh dari penelitian ini adalah model klasifikasi teks berbasis CNN dapat membantu pendidik dalam membedakan jawaban pelajar apakah teks hasil buatan manusia atau *generative AI*. Penelitian ini juga menghasilkan *dataset* baru teks berbahasa Indonesia, yang dapat menjadi sumber penting bagi penelitian lanjutan di bidang pengolahan bahasa alami.

1.5 Ruang Lingkup

Ruang lingkup bertujuan untuk menetapkan batasan agar pembahasan tetap fokus pada tujuan utama. Batasan-batasan dalam laporan skripsi ini adalah sebagai berikut:

1. Data yang digunakan adalah jawaban kuis mata kuliah *Internet of Things* (IoT) pada mahasiswa Departemen Informatika Universitas Diponegoro tahun ajaran 2023/2024 Semester 2 mengenai arsitektur komputer dan perbedaan mikrokontroler dengan komputer.

2. Rentang waktu pengumpulan data dari kuis tersebut adalah tanggal 29 April hingga 20 Juni 2024.
3. Rentang waktu pengumpulan data dari berbagai *platform generative AI* adalah tanggal 8 hingga 10 Oktober 2024.
4. Data terdiri atas 819 buah, yang mencakup teks buatan mahasiswa *generative AI* dalam bahasa Indonesia.
5. Arsitektur CNN dibangun dengan menggunakan *library TensorFlow*.
6. *Hyperparameter* yang ditentukan adalah *learning rate*, *batch size*, dan *optimizer*.
7. Pelatihan dan pengujian model dilakukan di lingkungan Jupyter Notebook pada *platform web Google Colaboratory*.

1.6 Sistematika Penulisan

Sistematika penulisan memberikan gambaran yang jelas dan terstruktur mengenai laporan skripsi ini. Berikut adalah urutan penulisan skripsi ini.

BAB I PENDAHULUAN

Bab ini menyajikan informasi mengenai beberapa aspek penting, termasuk latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup, dan sistematika penulisan dari laporan skripsi mengenai Klasifikasi Teks Buatan Manusia Versus *Generative AI Menggunakan Convolutional Neural Network* dengan Pengaturan *Hyperparameter*.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan informasi mengenai teori yang berhubungan dengan masalah yang dibahas pada skripsi ini. Bab ini memiliki beberapa subbab mengenai studi literatur, AI, pengolahan bahasa alami, *text preprocessing*, *K-Fold Cross Validation*, ekstraksi fitur, klasifikasi teks, *deep learning*, *hyperparameter*, CNN, *early stopping*, Adam *optimizer*, metrik evaluasi, *tools* dan *library*.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan informasi mengenai metodologi yang digunakan pada penelitian. Metode yang dipaparkan mencakup tahapan

pengumpulan data, *preprocessing*, pembagian data, pelatihan model, pengaturan *hyperparameter*, klasifikasi, dan evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan informasi mengenai lingkungan dan perangkat yang digunakan untuk skripsi ini, serta menyajikan hasil dan analisis dari hasil penelitian optimalisasi *hyperparameter* pada klasifikasi teks buatan manusia dan AI menggunakan metode CNN.

BAB V PENUTUP

Bab ini berisi kesimpulan dari seluruh bab yang telah dibahas dan saran sebagai bahan masukan untuk pengembangan atau penelitian selanjutnya.