

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, manfaat dan tujuan, ruang lingkup, dan sistematika penulisan pada skripsi yang berjudul Implementasi MobileNetV3 dengan SimAM, *Cosine Classifier*, serta MixUp dan CutMix untuk Meningkatkan Performa Klasifikasi Ekspresi Wajah.

1.1 Latar Belakang

Ekspresi wajah merupakan salah satu bentuk komunikasi non-verbal yang berperan penting dalam menyampaikan emosi dan perasaan manusia. Dalam interaksi sehari-hari, ekspresi wajah dapat mencerminkan berbagai kondisi emosional, seperti kepuasan, kebingungan, maupun ketidakpuasan terhadap suatu layanan. Oleh karena itu, kemampuan untuk mengenali ekspresi wajah secara otomatis menjadi hal yang penting, khususnya dalam pengembangan sistem berbasis kecerdasan buatan (Ekman, 1992).

Dalam bidang *Human-Computer Interaction* (HCI), interaksi antara manusia dan komputer diharapkan dapat berlangsung secara lebih alami, adaptif, dan responsif terhadap kondisi pengguna. Salah satu pendekatan yang dapat digunakan untuk mendukung hal tersebut adalah dengan memanfaatkan teknologi *Facial Expression Recognition* (FER), yaitu sistem yang mampu mengidentifikasi ekspresi wajah manusia secara otomatis. Teknologi ini menjadikan sistem tidak hanya merespons masukan eksplisit dari pengguna, tetapi juga dapat memahami kondisi emosional pengguna berdasarkan ekspresi wajah yang ditampilkan (Picard, 1997).

Pada konteks layanan pelanggan (*customer service*), pemahaman terhadap kondisi emosional pelanggan merupakan faktor penting dalam meningkatkan kualitas pelayanan. Namun, metode konvensional seperti survei kepuasan pelanggan memiliki keterbatasan karena bersifat subjektif dan tidak dilakukan secara *real-time*, serta bergantung pada partisipasi pelanggan. Hal ini menyebabkan penyedia layanan mengalami kesulitan dalam memperoleh informasi mengenai kondisi emosional pelanggan secara langsung selama proses interaksi berlangsung (Kotler & Keller, 2016).

Perkembangan teknologi *deep learning*, khususnya *Convolutional Neural Network* (CNN), telah memberikan kontribusi signifikan dalam tugas klasifikasi citra, termasuk pengenalan ekspresi wajah. CNN memiliki kemampuan untuk mengekstraksi fitur secara otomatis dari citra, sehingga menghasilkan performa yang lebih baik dibandingkan metode konvensional (Krizhevsky dkk., 2012).

Salah satu arsitektur CNN yang banyak digunakan adalah MobileNetV3, yang dikenal sebagai model *lightweight* dengan efisiensi komputasi yang tinggi, jumlah parameter yang relatif kecil, serta tetap mampu menghasilkan performa yang baik pada perangkat dengan keterbatasan sumber daya. Hal ini menjadikan MobileNetV3 sesuai untuk implementasi sistem *real-time* berbasis kamera. Namun demikian, sebagai model yang ringan, MobileNetV3 memiliki keterbatasan dalam menangkap pola fitur yang kompleks, terutama pada kondisi citra dengan variasi tinggi seperti perubahan pencahayaan, pose, dan *occlusion*, sehingga dapat memengaruhi performa klasifikasi (Howard dkk., 2019).

Keterbatasan tersebut menunjukkan bahwa pengenalan ekspresi wajah masih dihadapkan pada berbagai tantangan, seperti variasi pencahayaan, perubahan sudut pandang (pose), *occlusion*, serta ketidakseimbangan distribusi data antar kelas. Kondisi ini dapat menyulitkan model dalam mengekstraksi fitur secara optimal dan berdampak pada penurunan performa klasifikasi. Selain itu, ketidakseimbangan jumlah data pada setiap kelas dapat menyebabkan model cenderung bias terhadap kelas mayoritas dan kurang akurat dalam mengenali kelas minoritas (Li & Deng, 2020). Oleh karena itu, diperlukan pendekatan tambahan untuk meningkatkan kinerja model dalam menghadapi berbagai tantangan tersebut.

Berbagai penelitian sebelumnya telah mengusulkan beberapa pendekatan untuk mengatasi permasalahan tersebut. Salah satunya adalah penggunaan *attention mechanism* seperti SimAM yang mampu membantu model dalam memfokuskan perhatian pada bagian penting dari citra tanpa menambah jumlah parameter secara signifikan (Yang dkk., 2021). Selain itu, teknik data *augmentation* seperti MixUp dan CutMix terbukti dapat meningkatkan variasi data *train* serta mengurangi *overfitting* (Zhang dkk., 2018; Yun dkk., 2019). Penggunaan *cosine classifier* juga diketahui dapat meningkatkan kemampuan model dalam membedakan antar kelas, khususnya pada kondisi data yang tidak seimbang (Wang dkk., 2018).

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan yang menggabungkan arsitektur MobileNetV3 dengan SimAM, teknik data *augmentation* (MixUp dan CutMix), serta *cosine classifier* untuk meningkatkan performa klasifikasi ekspresi wajah. Sistem yang dikembangkan diharapkan dapat diimplementasikan dalam layanan pelanggan berbasis kamera untuk mendeteksi emosi secara *real-time*, sehingga mendukung peningkatan kualitas layanan yang lebih adaptif dan responsif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa model MobileNetV3-Large dalam klasifikasi ekspresi wajah?
2. Bagaimana pengaruh penerapan SimAM, *cosine classifier*, serta teknik augmentasi MixUp dan CutMix terhadap performa model MobileNetV3-Large?
3. Bagaimana kombinasi metode yang menghasilkan performa terbaik dalam klasifikasi ekspresi wajah?

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah memperoleh model klasifikasi ekspresi wajah berbasis MobileNetV3-Large dengan performa yang baik melalui penerapan SimAM, *cosine classifier*, serta teknik augmentasi MixUp dan CutMix. Selain itu, penelitian ini juga menganalisis pengaruh penerapan metode-metode tersebut terhadap performa model.

Manfaat dari penelitian ini adalah memberikan kontribusi dalam pengembangan model klasifikasi ekspresi wajah yang ringan dan memiliki performa yang baik, serta dapat menjadi referensi bagi penelitian selanjutnya di bidang klasifikasi ekspresi wajah dan HCI.

1.4 Ruang Lingkup

Agar penelitian lebih terarah dan tidak menyimpang dari tujuan, maka ruang lingkup penelitian ini adalah sebagai berikut:

1. *Dataset* yang digunakan adalah *dataset* FERPlus.
2. Klasifikasi dilakukan pada kategori emosi yang terdapat dalam *dataset* FERPlus.
3. Penelitian ini tidak menerapkan metode penanganan *class imbalance* pada proses pelatihan model.

4. Penelitian ini tidak melakukan proses *hyperparameter tuning* secara khusus dan konfigurasi parameter ditentukan secara manual.

1.5 Sistematika Penulisan

Untuk memberikan gambaran yang urut, runtut, dan jelas mengenai pembahasan penyusunan laporan penelitian ini, maka dibuatlah sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang yang menjelaskan alasan dilakukannya penelitian, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan laporan.

BAB II TINJAUAN PUSTAKA

Bab ini memuat teori-teori pendukung seperti *facial expression recognition*, *deep learning*, CNN, MobileNetV3, SimAM, *data augmentation*, *cosine classifier*, serta metrik evaluasi, dan penelitian terdahulu yang relevan. Bab ini juga membahas konsep dan metode yang digunakan sebagai dasar dalam perancangan serta pengembangan model pada penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan penelitian secara sistematis, meliputi proses pengumpulan data, penggunaan *dataset*, perancangan model, serta metode pelatihan dan pengujian yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil eksperimen dari proses pelatihan dan pengujian model, serta analisis performa model untuk mengetahui pengaruh metode yang diterapkan. Selain itu, dilakukan perbandingan antar variasi model berdasarkan nilai *accuracy*, *precision*, *recall*, *f1-score*, dan *confusion matrix* untuk memperoleh model dengan performa terbaik.

BAB V PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian berdasarkan rumusan masalah serta saran yang dapat digunakan untuk pengembangan penelitian selanjutnya.