

BAB I

PENDAHULUAN

Bab pendahuluan ini menyajikan dasar pemikiran penelitian tentang analisis pola emosi melalui sinyal sEMG wajah menggunakan teknologi Apache Spark. Pembahasan dimulai dari latar belakang masalah hingga urgensi pengembangan sistem identifikasi pola emosi yang lebih akurat serta mampu mengolah data dalam jumlah besar.

1.1 Latar Belakang

Emosi merupakan konstruksi psikologis yang kompleks, menggabungkan aspek kognitif, fisiologis, dan sosial, serta berperan signifikan dalam perilaku manusia, termasuk dalam pengambilan keputusan [1]. Untuk memahami emosi lebih mendalam melalui ekspresi wajah, berbagai metode deteksi emosi telah dikembangkan. Metode berbasis gambar atau video telah banyak digunakan dalam berbagai aplikasi, seperti sistem mobil pintar yang menyesuaikan kecepatan dan pencahayaan berdasarkan ekspresi pengemudi, deteksi perilaku mencurigakan di area publik, serta personalisasi konten *digital* dalam bidang pendidikan dan kustomisasi *website* [2], [3]. Metode konvensional ini memiliki keterbatasan karena faktor-faktor seperti variasi individu, kondisi pencahayaan lingkungan, dan kemampuan individu dalam mengontrol ekspresi mereka [4]. Oleh karena itu, *surface electromyography* (sEMG) hadir sebagai pendekatan alternatif yang objektif dalam mengukur aktivitas otot wajah terkait dengan ekspresi emosi [5], [6]. sEMG memungkinkan deteksi ekspresi emosi yang halus, bahkan yang tidak terlihat secara visual, sehingga membuka peluang untuk pengembangan studi kognisi emosional yang lebih akurat [3].

Sejalan dengan perkembangan era *digital* yang ditandai dengan pertumbuhan data visual yang eksponensial, kebutuhan akan algoritma untuk menginterpretasikan pola ekspresi wajah manusia semakin meningkat, terutama dalam konteks interaksi manusia-komputer dan pelayanan kesehatan mental [7]. Namun, peningkatan volume data sEMG serta kompleksitas sinyal yang dihasilkan

menimbulkan tantangan dalam hal pemrosesan dan analisis data. Sistem konvensional tidak mampu menangani beban komputasi yang besar untuk analisis data sEMG secara efisien dalam jumlah besar, sehingga mengakibatkan keterlambatan respons dalam identifikasi pola [8].

Sinyal sEMG wajah telah menunjukkan potensi besar dalam mengungkap emosi melalui aktivitas otot wajah. Identifikasi pola emosi menggunakan sEMG telah terbukti efektif sebagai alat deteksi emosi dengan akurasi mencapai 69-91% dalam kondisi alami saat subjek terpengaruh situasi yang memicu perasaan tertentu. Akurasi tersebut meningkat signifikan hingga 91-97% ketika individu secara sengaja mengekspresikan emosi seperti tersenyum [9]. Sistem identifikasi pola emosi berbasis sEMG dengan akurasi klasifikasi hingga 88,1% setelah 300 iterasi juga telah berhasil dikembangkan meskipun hanya menggunakan *dataset* yang relatif kecil dari 10 partisipan. Temuan ini menunjukkan potensi signifikan sinyal sEMG untuk pengembangan sistem pengenalan emosi lebih akurat dengan *dataset* yang lebih besar [10]. Keunggulan sEMG dalam mendeteksi ekspresi mikro, yakni ekspresi tidak sadar dan sulit dideteksi secara visual telah terbukti melalui penelitian sebelumnya [11]. Selain itu, kemampuan sEMG untuk merekam aktivitas otot wajah tanpa prosedur invasif memberikan peluang deteksi aktivitas otot halus yang berkorelasi dengan emosi seseorang [12].

Terdapat korelasi signifikan antara aktivitas otot wajah tertentu dengan pola emosi dasar manusia, khususnya pada otot *Zygomaticus major* yang menunjukkan peningkatan nilai sinyal sEMG amplitudo saat individu tersenyum dan otot *Corrugator supercilii* yang mengalami penurunan dalam kondisi mengerutkan dahi. Meskipun demikian, penelitian tersebut terbatas pada *dataset* yang kecil dan analisis yang relatif sederhana, sehingga sulit untuk menggeneralisasi hasil dan mendeteksi pola emosi yang lebih kompleks, serta hasil *clustering* yang diperoleh belum sepenuhnya sesuai dengan kategori video emosi [13]. Oleh karena itu, dibutuhkan pendekatan yang lebih terukur untuk mengolah data yang semakin besar untuk meningkatkan akurasi pengenalan emosi berbasis sinyal sEMG [14]. Apache Spark menawarkan solusi yang menjanjikan untuk menangani data dengan

kompleksitas analisis [15], [16]. Kemampuannya dalam melakukan paralelisasi tugas dan manajemen memori yang optimal sesuai untuk menangani data sinyal sEMG yang *high-dimensional* dan *time-series* [17]–[19]. Selain itu, Spark menggunakan arsitektur *in-memory computing* yang memungkinkan pemrosesan data 10-100 kali lebih cepat dibandingkan *on-disk* Hadoop *MapReduce* [20]. Berdasarkan hal tersebut, penelitian ini memanfaatkan Apache Spark untuk mengidentifikasi pola emosi melalui teknik *clustering*, yang memungkinkan pemrosesan dan pengelompokan data sinyal sEMG dalam jumlah besar secara efisien untuk menghasilkan pola emosi yang lebih akurat [21].

Orisinalitas penelitian ini dapat ditinjau dari tiga aspek utama. Aspek pertama adalah mengadopsi pendekatan *big data* yang memanfaatkan keseluruhan *dataset* yang tersedia, sehingga memungkinkan identifikasi pola emosi yang lebih komprehensif dan akurat dibandingkan penelitian sebelumnya. Aspek kedua berkaitan dengan inovasi teknologi, yakni penggunaan Apache Spark sebagai *platform* komputasi pemrosesan data sEMG yang merupakan pendekatan baru dan belum banyak dieksplorasi dalam konteks pengenalan emosi [22]. Aspek ketiga berfokus pada perluasan cakupan penelitian dengan mengidentifikasi spektrum emosi yang lebih kompleks, mencakup variasi keadaan afektif yang lebih mendalam dibandingkan penelitian-penelitian relevan sebelumnya [8], [10], [13].

Berdasarkan ketiga aspek orisinalitas tersebut, pengembangan sistem identifikasi emosi berbasis sEMG yang didukung komputasi Spark memiliki urgensi tinggi, terutama mengingat semakin meningkatnya kebutuhan akan sistem yang dapat memahami dan merespons emosi pengguna secara akurat. Dalam konteks kesehatan mental, implementasi sistem ini dapat memberikan terobosan dalam diagnosis dini gangguan mental dan pemantauan kesehatan emosional berkelanjutan [2], [9]. Di bidang teknologi asistif, sistem ini dapat membantu pengembangan *interface* adaptif yang lebih responsif untuk penyandang disabilitas [4]. Lebih lanjut, keberhasilan implementasi sistem ini juga dapat membuka jalan bagi pengembangan aplikasi-aplikasi inovatif dalam bidang pendidikan adaptif, keamanan, dan pelayanan pelanggan [2], [3].

Auziah Mumtaz, 2025

ANALISIS BIG DATA SINYAL FACIAL SURFACE ELECTROMYOGRAPHY MENGGUNAKAN APACHE SPARK PADA KOMPUTER TERDISTRIBUSI UNTUK PENINGKATAN AKURASI IDENTIFIKASI POLA EMOSI

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Penelitian ini akan mengembangkan dan mengevaluasi proses *clustering* untuk analisis pola aktivitas otot wajah menggunakan *dataset* EmgDataVR yang dikumpulkan oleh EmteqLabs melalui sistem emteqPRO [23]. *Dataset* ini mencakup sinyal sEMG dari area otot wajah yang direkam dari 38 partisipan saat menonton 25 video dengan stimulus emosional dengan kategori label negatif, netral, atau positif. Implementasi *framework* Apache Spark akan diimplementasikan secara terdistribusi untuk mengatasi keterbatasan komputasi yang ditemui pada penelitian sebelumnya, dimana analisis terbatas pada 0,196% dari total *dataset* karena keterbatasan komputasi memori pada Google Colab [13]. Dengan memanfaatkan kemampuan *distributed computing* Spark, penelitian ini akan menganalisis keseluruhan *dataset* 800.000 baris data per partisipan (38 partisipan, total $\pm 30,4$ juta baris dengan kapasitas $\pm 13,9$ GB) untuk mengidentifikasi pola emosi yang lebih kompleks. Hasil penelitian diharapkan dapat memberikan pemahaman komprehensif tentang pola aktivitas otot wajah dalam mengekspresikan emosi melalui analisis *dataset* yang lebih menyeluruh.

1.2 Rumusan Masalah

Penelitian ini mengidentifikasi permasalahan utama dalam pemrosesan sinyal sEMG wajah menggunakan teknologi *big data*. Berdasarkan latar belakang masalah yang telah diuraikan, peneliti merumuskan beberapa permasalahan penelitian sebagai berikut:

1. Bagaimana pengembangan teknik analisis *big data* untuk pemrosesan sinyal elektromiografi wajah menggunakan Apache Spark pada komputasi terdistribusi untuk mengatasi keterbatasan komputasi pada studi pendahuluan?
2. Bagaimana efektivitas penggunaan *big data* dalam meningkatkan kualitas hasil identifikasi pola emosi berdasarkan *dataset* EmgDataVR dari EmteqLabs?
3. Bagaimana perbandingan performa dan skalabilitas komputasi terdistribusi menggunakan Apache Spark dalam menganalisis keseluruhan *dataset* EmgDataVR dibandingkan dengan keterbatasan pemrosesan pada sistem komputasi tunggal?

Auziah Mumtaz, 2025

ANALISIS BIG DATA SINYAL FACIAL SURFACE ELECTROMYOGRAPHY MENGGUNAKAN APACHE SPARK PADA KOMPUTER TERDISTRIBUSI UNTUK PENINGKATAN AKURASI IDENTIFIKASI POLA EMOSI

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

1.3 Batasan Penelitian

Untuk memastikan penelitian yang terarah dan terfokus, diperlukan penetapan batasan-batasan yang jelas mencakup aspek teknis dan spesifikasi *dataset*. Berikut batasan-batasan untuk memfokuskan penelitian ini:

1. Bahasa pemrograman PySpark digunakan pada *cluster* Apache Spark yang terdistribusi di beberapa server untuk pemrosesan data sinyal sEMG dan implementasi *clustering* bertujuan mengatasi keterbatasan komputasi pada penelitian sebelumnya.
2. *Library* MLlib digunakan sebagai pustaka *machine learning* terintegrasi dalam Apache Spark yang menyediakan algoritma *clustering* terdistribusi untuk pemrosesan data sinyal sEMG dalam skala besar.
3. *Dataset* yang digunakan adalah EmgDataVR dari Emteq Labs yang mencakup data sinyal sEMG dari 38 responden sehat (14 wanita, 24 pria) dengan rentang usia 18 - 68 tahun ($\bar{x} = 33,4 \pm 13,6$ tahun) dengan persetujuan etik dari Bournemouth University; data direkam menggunakan sistem emteqPRO dengan Pico Neo 2 Eye VR *headset* selama total durasi $\pm 8,5$ jam, dimana setiap responden menghasilkan ± 800.000 baris data dengan 31 kolom (~ 300 MB per responden), menghasilkan total *dataset* sebesar 30.327.000 baris dengan kapasitas $\pm 13,9$ GB; *dataset* mencakup tiga tipe sinyal sEMG (contact, filtered, dan amplitude) yang diukur pada 4 area otot wajah (*Frontalis* untuk gerakan alis, *Zygomaticus major* untuk tersenyum, *Orbicularis oculi* untuk penutupan mata, dan *Corrugator supercilii* untuk pengerutan dahi) di 7 titik pengukuran; stimulus berupa 25 video terstandarisasi yang diambil dari pustaka video publik Samson *et al.*, studi Gnacek *et.al*, dan 7 video positif baru (ditambahkan karena masalah kualitas pada video yang tersedia), terdiri dari: 6 video negatif berdurasi total 93 detik ($\bar{x} = 0,248$ detik), 10 video netral berdurasi total 157 detik ($\bar{x} = 0,257$ detik), dan 9 video positif berdurasi total 78 detik ($\bar{x} = 0,125$ detik), dengan interval *self-report* 10 detik antar video.

4. Infrastruktur komputasi terdiri dari komputer terdistribusi dengan konfigurasi 1 *master node* dan 9 *worker nodes* (masing-masing 8,00 GB RAM dan 50 GB harddisk) melalui Python API (PySpark) untuk mendukung *distributed computing* dalam pemrosesan *big data* sinyal sEMG.
5. Sistem komputasi terdistribusi dibangun pada ekosistem Hadoop, dengan konfigurasi Apache Spark sebagai kerangka kerja pemrosesan paralel dan HDFS (Hadoop Distributed File System) sebagai sistem penyimpanan data secara terdistribusi.
6. Validasi hasil *clustering* dibandingkan dengan label emosi dari stimulus video dan pengujian performa komputasi dibandingkan penelitian sebelumnya yang menggunakan Google Colab untuk mengukur peningkatan akurasi dan efisiensi.

1.4 Tujuan Penelitian

Penelitian ini diarahkan untuk mengatasi permasalahan dalam analisis sinyal elektromiografi permukaan (sEMG) wajah dengan mengembangkan metode analisis berbasis *big data*. Berdasarkan rumusan masalah yang telah diuraikan, tujuan dari penelitian ini adalah untuk:

1. Mengembangkan teknik analisis *big data* untuk pemrosesan *dataset* skala besar sinyal elektromiografi wajah menggunakan Apache Spark pada komputasi terdistribusi untuk mengatasi keterbatasan komputasi pada studi pendahuluan.
2. Menganalisis efektivitas penggunaan *big data* dalam meningkatkan kualitas hasil identifikasi pola emosi berdasarkan *dataset* EmgDataVR dari EmteqLabs.
3. Mengevaluasi performa dan skalabilitas komputasi terdistribusi menggunakan Apache Spark dalam menganalisis *dataset* EmgDataVR dibandingkan dengan keterbatasan pemrosesan pada komputasi tunggal.

1.5 Manfaat Penelitian

Penelitian ini memberikan kontribusi dalam pengembangan metode analisis sinyal elektromiografi permukaan pada wajah dengan mengembangkan metode analisis berbasis *big data*. Berikut beberapa manfaat dari penelitian ini:

1. Bagi Peneliti

Penelitian ini memberikan manfaat bagi pengembangan kapasitas peneliti. Melalui penelitian ini, peneliti dapat mengembangkan wawasan mengenai sinyal biofisiologis berbasis *big data*.

1. Memberikan kontribusi ilmiah dalam pengembangan metode analisis sinyal biofisiologis untuk pengenalan pola emosi manusia.
2. Memperluas wawasan dan pengetahuan peneliti mengenai metode analisis *big data* untuk sinyal biofisiologis, khususnya sEMG wajah.
3. Memberikan pengalaman praktis pengembangan sistem komputasi terdistribusi Apache Spark untuk pemrosesan data skala besar.

2. Bagi Program Studi

Program studi memperoleh manfaat signifikan dari hasil penelitian ini untuk pengembangan kurikulum dan keilmuan. Penelitian ini menyumbangkan landasan teoretis dan praktis yang dapat diintegrasikan ke dalam materi pembelajaran.

1. Memperkaya literatur tentang pendekatan komputasi terdistribusi dalam pengolahan data biofisiologis skala besar.
2. Menjadi landasan teoretis untuk pengembangan model pengenalan emosi yang lebih akurat melalui pemrosesan data kompleks.
3. Hasil penelitian dapat digunakan sebagai materi pembelajaran dan pengembangan kurikulum dalam bidang analisis *big data*, sinyal biofisiologis, dan komputasi terdistribusi.

3. Bagi Masyarakat

Penelitian ini memberikan dampak sosial yang positif bagi masyarakat luas melalui pengembangan teknologi yang lebih adaptif. Hasil penelitian

berpotensi meningkatkan kualitas interaksi manusia-komputer dalam kehidupan sehari-hari.

1. Memperluas wawasan dan berkontribusi pada pengembangan teknologi yang lebih adaptif dan responsif terhadap kondisi emosional pengguna.
2. Berpotensi meningkatkan kualitas interaksi manusia-komputer dalam berbagai aplikasi sehari-hari.

4. Bagi Industri

Sektor industri dapat memanfaatkan hasil penelitian ini untuk mengembangkan produk dan layanan yang lebih inovatif. Teknologi yang dikembangkan memiliki aplikasi praktis dalam berbagai bidang industri yang membutuhkan interaksi manusia-komputer yang responsif.

1. Menciptakan peluang inovasi dalam pengembangan produk dan layanan yang dapat menginterpretasi dan merespons emosi pengguna.
2. Berpotensi meningkatkan pengalaman pengguna dan efektivitas sistem dalam berbagai aplikasi industri.

1.6 Struktur Penulisan Skripsi

Penelitian ini disusun dengan mengacu pada Pedoman Penulisan Karya Ilmiah Universitas Pendidikan Indonesia (UPI) edisi 2024. Sesuai dengan pedoman tersebut, struktur skripsi terdiri dari:

1. Bab I: Pendahuluan

Bab ini menguraikan latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, serta batasan penelitian. Latar belakang penelitian menjelaskan topik dan isu yang dibahas secara sistematis dan relevan dengan permasalahan yang diteliti.

2. Bab II: Tinjauan Pustaka

Bab ini memuat uraian teori dan penelitian sebelumnya yang relevan sebagai dasar untuk mendukung penelitian. Tinjauan pustaka mencakup kerangka teori dan konsep yang menjadi landasan penelitian, serta posisi teoritis peneliti terkait masalah yang dikaji.

3. Bab III: Metode Penelitian

Bab ini menjelaskan pendekatan dan metode yang diterapkan dalam penelitian secara sistematis, jenis dan pendekatan penelitian, tahapan alur penelitian, perancangan arsitektur sistem, alur kerja, algoritma utama yang diterapkan, serta alat dan bahan yang digunakan. Metode penelitian disusun secara terstruktur untuk menjawab rumusan masalah dan mencapai tujuan penelitian yang telah ditetapkan.

4. Bab IV: Hasil dan Pembahasan

Bab ini menyajikan temuan atau hasil penelitian dalam format teks, tabel, atau grafik, disertai interpretasi dan pembahasan mendalam. Pada bagian ini, hasil penelitian dikaitkan dengan teori atau penelitian sebelumnya yang relevan untuk menghasilkan analisis yang komprehensif.

5. Bab V: Simpulan dan Saran

Bab ini merangkum hasil penelitian secara singkat serta menjawab rumusan masalah. Bagian ini juga menyajikan rekomendasi untuk penelitian lanjutan atau implikasi praktis dari temuan penelitian.