

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Seiring terus berkembangnya ilmu pengetahuan dan teknologi yang membawa kebaharuan dan kenyamanan bagi manusia dalam berbagai aspek termasuk di bidang teknologi itu sendiri. Teknologi yang saat ini berkembang sangat cepat, membawa perubahan yang signifikan, yang dahulunya hanya sebatas untuk bertukar informasi akan tetapi saat ini di era digital, teknologi membuat pekerjaan manusia menjadi lebih mudah dan efisien, tentunya dengan kemajuan teknologi itu salah satunya kecerdasan buatan atau sebuah kecerdasan yang menjadikan program dapat merasakan, berpikir, bertindak, dan beradaptasi sehingga memudahkan pekerjaan manusia.

Berbicara mengenai kecerdasan buatan tentunya ada kaitannya dengan *Machine Learning* dan *Deep Learning*, *Machine Learning* yaitu algoritma yang terus menerus belajar dari data dan meningkatkan kinerjanya dari waktu ke waktu sedangkan *Deep Learning* bagian dari *Machine Learning* berupa *neural network* yang belajar dari data dalam jumlah yang besar (Hidayatullah, 2021). Teknologi yang semakin maju ini tentunya dijadikan sebagai alat bantu kerja penerapannya seperti pengenalan wajah dan pengenalan emosi berupa ekspresi Wajah, penerapan umumnya diterapkan di banyak bidang dalam hal ini pada keamanan dan pembayaran di restoran atau *cafe* dengan mesin *Point of Sale* (Sutisnawati et al., 2021).

Selain itu dalam bidang pendidikan yaitu saat pembelajaran, pelaksanaan pembelajaran antara pendidik dan peserta didik terjadi hubungan yang dinamis, berhubungan dengan pedagogi, sesuai dengan metode terkait perubahan perilaku dan emosi peserta didik, tentunya pembelajaran yang berlangsung akan mempengaruhi hasil belajarnya, dengan begitu tentunya perlu adanya penilaian emosi peserta didik dan identifikasi perubahan emosi selama sesi pembelajaran, sehingga dapat dijadikan evaluasi dalam pembelajaran, akan tetapi banyaknya peserta didik atau jumlahnya yang banyak, pendidik susah dalam mengidentifikasi satu-satu (Mukhopadhyay et al., 2020).

Saat pembelajaran berlangsung, interaksi antara pendidik dan peserta didik mengalami hubungan yang memungkinkan bertatap muka secara langsung maupun tidak, akan tetapi yang menjadi permasalahan, bagaimana untuk mengetahui antusias dari peserta didik, apakah benar-benar siap, sudah memahami materi dan hal lain terkait pelaksanaan pembelajaran, pada kondisi ini pendidik sulit untuk mengetahui antusias dalam pembelajaran, ketika jumlahnya 1 orang masih bisa melihat antusias peserta didik, lain halnya dalam kasus peserta didik banyak dalam satu ruangan, bagaimana bisa melihat antusias peserta didik khususnya dari ekspresi wajahnya.

Permasalahan muncul juga selain di pendidikan yaitu perusahaan, pada perusahaan, stres kerja memiliki dampak yang dapat mengakibatkan penurunan kualitas dan kuantitas pekerja dalam melaksanakan pekerjaannya di lapangan. Dampak dari stres kerja antara lain adalah meningkatnya tingkat absensi, ketidakpuasan kerja, sikap yang menunda nunda pekerjaan, kecelakaan kerja yang tinggi (Farid et al., 2019).

Dari hasil kutipan tersebut muncul dari sisi karyawan meningkatkan produksi dan kemajuan dalam perusahaan tersebut, tentunya perlu adanya *training* dan pengelolaan *stress* yang berasal dari karyawan, apakah karyawan tersebut *antusiasme* dan siap dalam bekerja, dalam kesiapan karyawan dan *antusiasme* karyawan tersebut akan meningkatkan produksi dan menghindari *accident* dari hal tersebut dilakukan pendeteksian ekspresi wajah dalam jumlah besar dalam satu waktu sehingga hasilnya bisa dijadikan evaluasi berupa *training* dan edukasi terhadap karyawan untuk memajukan perusahaan dan meningkatkan produksi serta menghindari *accident* atau kecelakaan kerja. Pada permasalahan tersebut tentunya erat kaitannya dengan ekspresi wajah dengan berupa deteksi emosi manusia.

Ekspresi emosi merupakan usaha yang dilakukan oleh seseorang untuk mengkomunikasikan status perasaan (emosi) sebagai respon terhadap situasi tertentu baik internal maupun eksternal yang terlihat dari perubahan biologis, fisiologis dan serangkaian kecenderungan tindakan (sikap dan tingkah laku) berorientasi pada tujuan (Kusdiananggalih et al., 2021).

Wajah adalah sebuah objek untuk mengidentifikasi seseorang dari ekspresi yang dapat menunjukkan aktualisasi diri atau emosinya. Ekspresi atau mimik wajah

merupakan bentuk komunikasi nonverbal yang mana dalam ekspresi tersebut dapat memberikan keadaan emosi dari seseorang kepada orang yang mengamatinya (Yin & Liu, 2018).

Interaksi terhadap makhluk hidup salah satunya manusia tentunya berkaitan dengan wajah yang dimunculkan melalui ekspresi wajah. Ekspresi wajah adalah cara paling umum untuk menyampaikan suasana hati atau perasaan, bukan hanya untuk manusia tetapi juga untuk banyak *organisme* hidup lainnya (V & Bahel, 2021). Pendeteksian ekspresi emosi manusia berdasarkan ekspresi wajah dalam interaksi dapat memberikan informasi dalam hal mengetahui perasaan seseorang dengan mudah apabila dilakukan kepada satu orang, lain halnya dengan banyaknya orang sehingga sulit dalam mengenalinya, maka diperlukan sistem yang membantu dalam menangani hal seperti itu, sehingga ekspresi emosi manusia berdasarkan ekspresi wajah dengan sistem secara otomatis akan memudahkan dalam pendeteksian tersebut meskipun dalam jumlah yang banyak, Pendeteksian ekspresi wajah secara otomatis erat kaitannya dengan *Deep Learning* terutama domain *Computer Vision*.

Deep Learning adalah jenis *Machine Learning* yang terinspirasi oleh fungsi sel otak manusia yang disebut neuron. Dari Inspirasi tersebut terdapat konsep yang dinamakan *Neural Network* (NN). Namun pada *deep learning*, *neural network*-nya lebih dari dua lapisan. Selain itu terdapat beberapa hal yang menjadi evolusi *neural network* pada *deep learning* diantaranya lebih banyak neuron dari jaringan sebelumnya, cara yang lebih kompleks untuk menghubungkan lapisan atau neuron, dan ekstraksi fitur otomatis (Hidayatullah, 2021).

Kemajuan di bidang *Deep Learning* sangat pesat termasuk dalam domain *Computer Vision*. *Computer Vision* sendiri adalah persepsi visual dengan menggabungkan kamera, komputasi tepi atau berbasis *cloud*, perangkat lunak, dan kecerdasan buatan (AI) sehingga sistem dapat melihat dan mengidentifikasi objek. *Computer Vision* menggunakan pembelajaran mendalam untuk membentuk jaringan saraf yang memandu sistem untuk pemrosesan dan analisis. Model *Computer Vision* yang terlatih dapat mengenali objek, mendeteksi dan mengenali orang, dan bahkan melacak gerakan. Contoh dalam *Computer Vision* seperti sistem keamanan serta pemantauan, termasuk dalam hal yang berkaitan dengan penguatan

ketepatan sistem pada mengenali paras atau ekspresi pada wajah manusia, sebagai akibatnya wajah masih menjadi objek yang terus menerus dikembangkan peneliti, dalam pendeteksi emosi wajah berdasarkan ekspresi wajah ini menggunakan *Deep Learning*, dalam *Deep Learning* mampu mempelajari data dalam jumlah besar yang ada dan cara pembelajaran yang berasal dari data kemudian menghasilkan prediksi di masa yang akan datang.

Penelitian dari Azhari dan Fitriyani (2020) mengemukakan bahwa Pembelajaran mendalam menggunakan metode *Convolutional Neural Network* untuk mendeteksi emosi berdasarkan ekspresi wajah secara *real-time* menghasilkan hasil yang memuaskan, sementara di lain hal itu membutuhkan sumber daya perangkat keras yang sangat besar untuk memprediksi ambang batas emosi (Azhari & Fitriyani, 2020). Dataset yang digunakan dalam penelitiannya tersebut bersumber dari kaggle FER 2013. Hasil dari pengkombinasian antara metode CNN, *Viola-Jones* dan *Haar Cascade* dengan tingkat keakuratan yang terbilang cukup bagus antara rentang 67-83%.

Penelitian yang dilakukan Omar, dkk (2021) mengemukakan bahwa pada penelitian yang dilakukannya, model CNN *training* menggunakan *library tensorflow-gpu version 2.3.0*. Dataset dari *training* yang digunakan adalah dataset FER-2013 yang didapatkan dari kaggle. Hasil yang didapatkan, bahwa model berhasil untuk melakukan klasifikasi ekspresi wajah manusia. Hal ini didukung dengan akurasi sebesar 73,56%, dengan menggunakan algoritma *Haar Cascade*. Model CNN yang di *training* mencapai akurasi sebesar 73.56% (Omar et al., 2021). Selain itu, sistem juga berhasil untuk melakukan pengenalan berbagai ekspresi wajah partisipan terhadap input gambar yang berupa gambar *screenshot* dari *virtual meeting zoom*. Grafik kurva dari *training* dan *validation* juga menunjukkan bahwa model CNN yang terdapat pada penelitian ini tidak terjadi *overfitting*.

Penelitian yang dilakukan Seandrio, dkk (2021) dalam penelitiannya mengemukakan bahwa dalam penelitian yang dilakukan, beberapa lapisan pada situasi yang berbelit-belit tidak menghasilkan jaminan yang baik dan menghasilkan hasil yang buruk untuk hasil yang diinginkan. Arsitektur jaringan saraf *convolutional* menggunakan model 16 lapisan *Visual Geometry Group (VGG)*, yang mampu menganalisis data emosi wajah yang diklasifikasikan. Hal ini dapat

ditunjukkan oleh *Confusion Matrix* pengujian nilai, yang menunjukkan akurasi rata-rata sekitar 70,47%, rata-rata presisi sekitar 70,86%, dan rata-rata recall sekitar 69,57% (Seandrio et al., 2021). Pengujian dilakukan dengan menggunakan sekitar 3589 dataset uji data, yang terdiri dari 491 data marah, 55 data jijik, 528 data takut, 879 data senang, 594 data sedih, 416 data kaget, dan 626 data biasa. Hasil penelitian yang kurang akurat dipengaruhi oleh faktor lain termasuk goyangan kamera dan konflik emosi atau kemiripan emosi satu sama lain.

Penelitian yang dilakukan oleh Pratama, dkk (2019) mengemukakan bahwa dalam hasil pengujian secara otomatis menggunakan berbagai pola variasi dalam kumpulan data gambar Aberdeen, pengujian hingga 60 gambar dari 30 orang berbeda yang digunakan dalam Sistem Pencarian Pengenalan Wajah menggunakan *Multitasking Cascade Convolutional Neural Network* (MTCNN) menghasilkan kesimpulan, beberapa parameter citra yang mempengaruhi seperti variasi kecerahan (gelap, terang, sedang), variasi posisi objek (antara lain dapat diambil dekat dengan wajah, dari jarak dengan kemiringan yang berbeda), kemudian posisi grip (sedikit miring, ekspresi wajah dalam gambar objek yang sangat berbeda (marah, sedih, kesal, tersenyum, datar, cemberut), sistem pencarian pengenalan wajah menggunakan MTCNN (*Multitask Cascaded Convolutional Neural Network*) berhasil melakukan pengenalan wajah 100% (Pratama et al., 2019). Oleh karena itu, nilai positif sejati sama dengan jumlah data yang diuji dan nilai negatif sebenarnya adalah nol.

Berdasarkan penelitian terdahulu terkait *Multi Task Convolutional Neural Network* (MTCNN) dan *Visual Geometry Group* (VGG)-16, penulis tertarik untuk melakukan penggabungan kedua hal tersebut, dikarenakan akurasi yang didapatkan pada penelitian tersebut kurang memuaskan, penulis mencoba melakukan implementasi deteksi emosi manusia berdasarkan ekspresi wajah, yang mana dalam pendeteksian wajah menggunakan Algoritma *Multi Task Convolutional Neural Network* (MTCNN). Algoritma ini dipilih karena setelah melihat hasil penelitian-penelitian terdahulu ternyata algoritma ini lebih efektif untuk digunakan dalam kasus pendeteksian wajah, dan Arsitektur *Visual Geometry Group* (VGG) yang menggunakan 16 lapisan yang mampu menghasilkan model dengan baik, kemudian Pendeteksian dengan Algoritma MTCNN berhasil melakukan pengenalan wajah

100% yang berdasar dari penelitian terdahulu yang telah dilakukan, maka dari itu, dengan menggunakan Arsitektur VGG 16 dan pengenalan wajah menggunakan Algoritma MTCNN terbukti dapat menangani data emosi wajah berdasarkan ekspresi yang diklasifikasi dan akan menghasilkan akurasi atau performa model yang baik, sehingga pendeteksian secara otomatis dalam bentuk sistem dengan lebih dari satu orang dapat dilakukan dengan mudah.

Pada penelitian yang dilakukan penulis menggunakan *Framework AI Project Life Cycle*, karena langkahnya yang menyesuaikan dari penggunaannya. Dataset yang akan digunakan menggunakan data FER 2013. Dataset yang digunakan dilakukan eksplorasi dengan melakukan augmentasi gambar, pembagian dataset dan *modelling* menggunakan Arsitektur *Visual Geometry Group (VGG) 16* untuk ekspresi wajah dan melakukan evaluasi model jika model baik dan bagus performanya dilakukan *deployment* dengan menggunakan *framework streamlit* dalam bentuk *Graphical User Interface (GUI)* dengan input berupa gambar dan video, menggunakan *framework streamlit* karena *framework* ini adalah alat yang memudahkan dalam menampung model pembelajaran mesin untuk membuat presentasi yang elegan dan dinamis untuk data dan pengelolaannya juga mudah karena telah tersedia dokumentasi bawaannya (Rajan, 2021). Selain itu batasan masalah dalam penelitian ini hanya membahas mengenai implementasi algoritma *Multi Task Convolutional Neural Network (MTCNN)* dan arsitektur *Visual Geometry Group (VGG)-16* untuk deteksi emosi manusia berdasarkan ekspresi wajah, hanya di implementasinya saja, kemudian dalam pendeteksiannya hasil yang diperoleh terkait ekspresi wajah tidak perlu mendapatkan validasi dari orang yang dideteksi.

1.2 Rumusan Masalah Penelitian

Berdasarkan latar belakang masalah yang telah disampaikan adapun rumusan masalah penelitian ini sebagai berikut:

1. Bagaimana implementasi Algoritma *Multi Task Convolutional Neural Network (MTCNN)* dan Arsitektur *Visual Geometry Group (VGG) 16* untuk Deteksi Emosi berdasarkan ekspresi wajah?

2. Bagaimana hasil pengujian dari Algoritma *Multi Task Convolutional Neural Network* (MTCNN) dan Arsitektur *Visual Geometry Group* (VGG)-16 untuk Deteksi Emosi berdasarkan ekspresi wajah?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini antara lain sebagai berikut:

1. Untuk mengetahui implementasi Algoritma *Multi Task Convolutional Neural Network* (MTCNN) dan Arsitektur *Visual Geometry Group* (VGG)-16 untuk Deteksi Emosi berdasarkan ekspresi wajah;
2. Untuk mengetahui hasil pengujian dari Algoritma *Multi Task Convolutional Neural Network* (MTCNN) dan Arsitektur *Visual Geometry Group* (VGG)-16.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini antara lain sebagai berikut:

1. Menerapkan Arsitektur *Visual Geometry Group* (VGG)-16 dengan pendeteksian wajah menggunakan Algoritma *Multi Task Convolutional Neural Network* (MTCNN);
2. Mengetahui hasil pengujian Arsitektur *Visual Geometry Group* (VGG)-16 dan Algoritma *Multi Task Convolutional Neural Network* (MTCNN) untuk mendeteksi Emosi berdasarkan ekspresi wajah;
3. Menghasilkan sistem deteksi emosi manusia berdasarkan ekspresi wajah pada gambar dan video;
4. Manfaat bagi peneliti lain, sebagai acuan atau bahan kajian lebih lanjut penelitian selanjutnya dengan kesamaan topik atau yang menggunakan algoritma dan arsitektur yang sama pula;
5. Manfaat bagi penulis, memperbaharui *knowledge* khususnya dalam pemahaman yang bersinggungan dengan Algoritma *Multi Task Convolutional Neural Network* (MTCNN) untuk pengenalan wajah dan arsitektur *Visual Geometry Group* (VGG)-16;
6. Manfaat bagi pembaca, memperbaharui informasi atau *knowledge* tentang Algoritma *Multi Task Convolutional Neural Network* (MTCNN) dan Arsitektur *Visual Geometry Group* (VGG)-16 serta penerapannya.

1.5 Struktur Organisasi Skripsi

Sistematika penulisan skripsi ini terdiri dari lima bab, diawali dengan pendahuluan dan diakhiri dengan simpulan, implikasi, dan rekomendasi. Dalam penelitian ini penulis menggunakan susunan ini sesuai dengan Pedoman Penulisan Karya Ilmiah Universitas Pendidikan Indonesia 2019. Adapun bagian isi secara rinci sebagai berikut:

1. BAB I PENDAHULUAN

Pada bab I dalam penelitian ini terdiri dari latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan struktur organisasi skripsi.

2. BAB II KAJIAN PUSTAKA

Pada bab II dalam penelitian ini terdiri dari penelitian terdahulu dan kajian teori mulai dari Ekspresi wajah, *Convolutional Neural Network*, Arsitektur *Visual Geometry Group (VGG)-16*, Algoritma *Multi Task Convolutional Neural Network (MTCNN)*, *Confusion Matrix*, Kurva ROC dan AUC, Bahasa pemrograman *Python*, dan *Framework Streamlit*.

3. BAB III METODE PENELITIAN

Pada bab III dalam penelitian ini terdiri dari metode dan desain penelitian dengan menggunakan *Framework AI Project Life Cycle* yang terdiri dari *Problem Scoping, Data Acquisition, Data Exploration, Modelling, Evaluation*, dan *Deployment* serta prosedur perancangan.

4. BAB IV HASIL DAN PEMBAHASAN

Pada bab IV dalam penelitian ini terdiri dari temuan berupa hasil penelitian, *hardware* dan *software* pendukung, *Problem Scoping, Data Acquisition, Data Exploration, Modelling, Evaluation*, kemudian *Deployment* menggunakan *framework streamlit*, selain itu melakukan hasil pengujian meliputi pengujian ekspresi wajah dengan arsitektur VGG-16 dan pengujian pendeteksian wajah dengan Algoritma MTCNN, selanjutnya pengujian pendeteksian deteksi emosi manusia berupa ekspresi wajah dalam bentuk *Graphical User Interface* secara individu dan kelompok dengan data berupa foto dan video, hasil video akan ditampilkan data hasil pengujian dalam bentuk *excel* berbentuk *log*.

5. BAB V SIMPULAN, IMPLIKASI, DAN REKOMENDASI

Pada bab V dalam penelitian ini berisi tentang kesimpulan isi dari keseluruhan bab-bab sebelumnya, implikasi, dan rekomendasi berupa saran-saran dari hasil yang diperoleh dan diharapkan dapat bermanfaat dalam pengembangan dan pemanfaatannya.