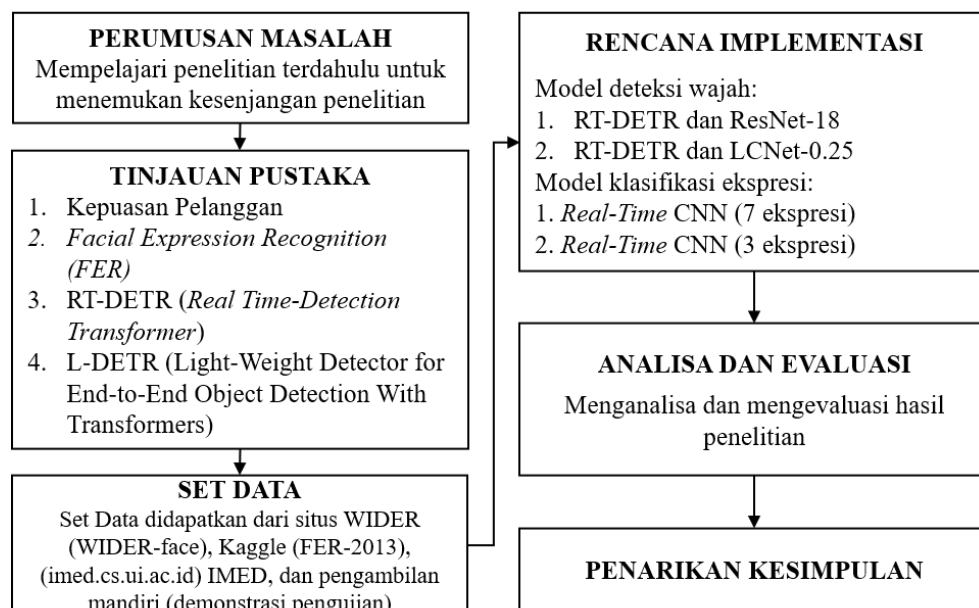


BAB III METODE PENELITIAN

Bab ini menjelaskan semua tahapan pembangunan sistem pengenalan ekspresi wajah untuk kepuasan pelanggan. Diawali dengan pengembangan model deteksi wajah menggunakan metode RT-DETR (*Real Time-Detection Transformer*), ResNet-18, dan LNet (*Lightweight CPU Convolutional Neural Network*) dengan set data WIDER-face. Kemudian, pengembangan model klasifikasi wajah menggunakan metode Real-Time CNN dengan set data FER-2013 (*Facial Expression Recognition-2013*) serta IMED (*The Indonesian Mixed Emotion Dataset*). Hingga terakhir, yaitu pengujian integrasi model deteksi wajah dan klasifikasi ekspresi menggunakan set data demonstrasi pengujian. Metodologi penelitian terdiri dari desain penelitian, rancangan model dan eksperimen, serta lingkungan komputasi eksperimen.

3.1. Desain Penelitian

Penelitian yang dilakukan melibatkan tujuh tahap pengerjaan. Tahapan tersebut terdiri dari perumusan masalah, tinjauan pustaka, pengumpulan data, rancangan model, implementasi, evaluasi, serta penarikan kesimpulan. Seluruh tahapan penelitian tersebut diilustrasikan pada gambar 3.1.



Gambar 3.1 Desain penelitian

Penjelasan lebih lanjut mengenai masing-masing tahapan adalah sebagai berikut:

3.1.1. Rumusan Masalah

Pada perumusan masalah, dilakukan proses identifikasi masalah dengan mempelajari penelitian terdahulu untuk menemukan kesenjangan penelitian yang dapat dijadikan basis dari penelitian yang dilakukan. Dari perumusan masalah ditemukan basis penelitian seperti, latar belakang, masalah utama, tujuan penelitian, dan metode yang akan digunakan.

3.1.2. Tinjauan Pustaka

Setelah perumusan masalah selesai dilakukan merupakan tinjauan pustaka yang berfungsi untuk menguraikan hubungan berbagai teori yang melandasi atau menjadi argumen penelitian. Tinjauan pustaka juga berisi penjelasan tentang variabel-variabel yang diteliti.

3.1.3. Pengumpulan Data

Penelitian ini membutuhkan data untuk melatih model agar dapat mendeteksi ekspresi wajah. Data yang digunakan, yaitu set data WIDER-face, FER-2013, IMED, dan set data demonstrasi pengujian.

3.1.4. Analisa dan Evaluasi

Tahap ini menganalisa serta mengevaluasi hasil eksperimen yang telah dilakukan. Diharapkan dari eksperimen yang telah dilakukan dapat menjawab rumusan masalah yang telah disusun.

3.1.5. Rancangan Implementasi

Pada tahap ini dibangun model sesuai rancangan serta eksperimen penelitian. Pembangunan model deteksi wajah melibatkan RT-DETR dengan ResNet-18 serta LCNet-0.25 dan model klasifikasi ekspresi melibatkan Real-Time CNN menggunakan 7 ekspresi dan 3 ekspresi.

Kemudian, eksperimen dilakukan berdasarkan rancangan model yang telah disusun. Dimulai dari praproses, *training*, pemilihan bobot model terbaik, dan *testing*. Selanjutnya, dilakukan eksperimen pada beberapa jenis model yang terdiri dari varian kombinasi metode yang telah dipilih pada rumusan masalah.

3.1.6. Penarikan Kesimpulan

Tahap terakhir dari penelitian adalah penarikan kesimpulan yang dilakukan dengan cara membandingkan hasil analisa dan evaluasi eksperimen dengan rumusan serta tujuan penelitian. Dari kesimpulan yang telah ditarik, terbentuk juga saran yang akan membantu dalam pengarahannya penelitian selanjutnya.

3.2. Set Data

Pada penelitian ini, digunakan empat set data yang berbeda pada model deteksi serta klasifikasi wajah. Set data yang digunakan, yaitu WIDER-face (S. Yang dkk., 2016), FER-2013 (Goodfellow dkk., 2013), IMED (Liliana dkk., 2018) serta set data demonstrasi pengujian.

3.2.1. WIDER-face



Gambar 3.2 Sampel set data WIDER-face

Set data WIDER-face yang dikembangkan oleh S. Yang dkk. (2016) didapatkan dengan mengunduhnya secara langsung pada situs web terkait dan digunakan untuk pengembangan model deteksi wajah, dengan metode RT-DETR (*Real Time-Detection Transformer*), ResNet-18, serta LCNet (*Lightweight CPU Convolutional Neural Network*). Evaluasi set data WIDER dilakukan dengan mengirimkan hasil prediksi model ke pengurus situs WIDER. Hasil evaluasi akan didapatkan dalam sebuah file MatLab, yang berisi nilai *precision* dan *recall*. Set data ini terdiri dari variasi kondisi seperti kemiringan, pose, ekspresi, halangan, serta pencahayaan. WIDER-face terbagi menjadi tiga bagian dengan persentase

training (40%), *validation* (10%), dan *testing* (50%). Gambar 3.2 merupakan sampel data dari set data WIDER-face.

Kemudian, set data ini digunakan karena memiliki anotasi *bounding box* yang mana sesuai dengan spesifikasi dari metode RT-DETR yang mendeteksi wajah dengan *bounding box*. Dengan kondisi foto pada set data, yaitu *wild* atau tidak mengikuti ketentuan/pose tertentu, menjadikan set data ini sesuai dengan kebutuhan sistem yang nantinya digunakan melalui CCTV yang mendeteksi wajah melalui sudut pandang atas.

3.2.2. FER-2013 (*Facial Expression Recognition-2013*)

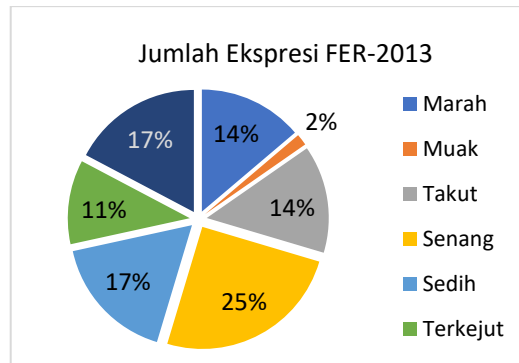


Gambar 3.3 Sampel set data FER-2013

Set data FER-2013 (*Facial Expression Recognition 2013*) yang dikembangkan oleh Goodfellow dkk. (2013) didapatkan dengan registrasi pada tantangan yang diadakan oleh Kaggle dan digunakan untuk pengembangan model klasifikasi ekspresi. Set data dibuat dengan menggunakan fitur pencarian citra pada Google dengan memasang nama ekspresi dengan gender, umur, atau etnis. Selanjutnya foto yang telah didapatkan dari Google diproses lebih lanjut agar tidak ada set data duplikat dan salah label. Hasil akhir dari set data adalah data berukuran 48x48 piksel dan terkonversi menjadi *grayscale*. FER-2013 terbagi menjadi tiga bagian dengan persentase *training* (80%), *validation* (10%), dan *testing* (10%). Gambar 3.3 merupakan contoh dari set data FER-2013.

Penelitian ini menggunakan set data FER-2013 yang juga pernah digunakan pada penelitian FER terdahulu (Meena dkk., 2023). Kemudian, kondisi foto pada set data merupakan *wild* atau tidak mengikuti ketentuan/pose tertentu. Kedua alasan tersebut menjadikan set data ini sesuai dengan tujuan pengembangan sistem, yaitu

pengimplementasian pengenalan ekspresi wajah melalui CCTV sudut pandang atas. Selanjutnya foto pada set data juga telah di-*crop* mengikuti batasan wajah yang menjadikan tahap praproses lebih mudah untuk dilakukan.



Gambar 3.4 Bagan jumlah ekspresi FER-2013

Keseluruhan set data FER-2013 berjumlah 35.887 data dengan rincian sebagai yaitu, 4953 foto untuk ekspresi marah, 547 foto untuk ekspresi muak, 5121 foto untuk ekspresi takut, 8989 foto untuk ekspresi senang, 6077 foto untuk ekspresi sedih, 4002 foto untuk ekspresi terkejut, dan 6198 foto untuk ekspresi netral. Berdasarkan jumlah data tersebut, kelas muak mengalami *imbalance* dengan persentase yang hanya sebesar 1.52%. Gambar 3.4 merupakan bagan yang merepresentasikan jumlah ekspresi pada set data FER-2013.

3.2.3. IMED (*Indonesian Mixed Emotion Dataset*)

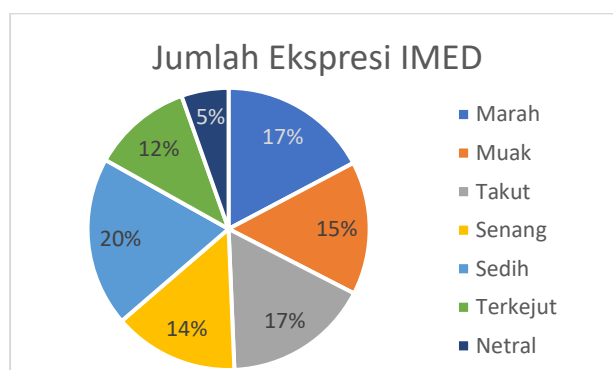


Gambar 3.5 Sampel set data IMED

Set data IMED yang dibuat oleh Liliana dkk. (2018) didapatkan dengan registrasi di situs IMED dan mengajukan permintaan akses kepada pemilik dataset dan digunakan untuk pengembangan model klasifikasi ekspresi. Pembuatan data

dilakukan dengan mengikuti prosedur yang telah ditentukan, dengan partisipan berjumlah 15 orang yang terdiri dari 9 perempuan dan 6 laki-laki serta berbagai macam suku, yaitu 40% Jawa, 20% Sunda, 13.3% Padang, 13.3% Melayu, 6.7% Batak, dan 6.7% Manado. Terdapat 19 jenis ekspresi yang terdiri dari 7 ekspresi dasar dan 12 ekspresi campuran. Penelitian ini hanya menggunakan 7 ekspresi dasar, yaitu netral, senang, sedih, muak, marah, takut, dan terkejut. Gambar 3.5 merupakan contoh dari set data IMED.

Berdasarkan studi mengenai bias yang terjadi pada klasifikasi gender dan ras oleh Buolamwini & Gebru (2018), model klasifikasi komersil umumnya memiliki performa terbaik pada subjek pria berkulit putih serta terburuk pada perempuan berkulit gelap. Oleh karena itu, untuk mengantisipasi hal tersebut, digunakan IMED agar model bisa memiliki performa lebih baik pada wajah Indonesia karena keunikan etnis subjeknya yang memiliki karakteristik wajah orang Indonesia.



Gambar 3.6 Bagan jumlah ekspresi IMED

Keseluruhan set data IMED (7 ekspresi dasar) berjumlah 5956 data dengan rincian sebagai yaitu, 1028 foto untuk ekspresi marah, 911 foto untuk ekspresi muak, 1000 foto untuk ekspresi takut, 854 foto untuk ekspresi senang, 1158 foto untuk ekspresi sedih, 683 foto untuk ekspresi terkejut, dan 322 foto untuk ekspresi netral. Berdasarkan jumlah data tersebut, kelas netral mengalami *imbalance* dengan persentase yang hanya sebesar 5.4%. Gambar 3.6 merupakan bagan yang merepresentasikan jumlah ekspresi pada set data.

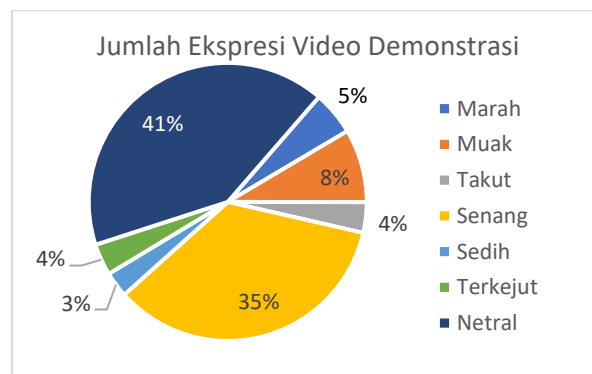
3.2.4. Set Data Demonstrasi Pengujian



Gambar 3.7 Sampel set data demonstrasi pengujian

Set data ini digunakan untuk pengujian model integrasi deteksi wajah dan klasifikasi wajah. Video diambil secara mandiri dengan subjek mahasiswa program studi Ilmu Komputer Universitas Pendidikan Indonesia. Semua subjek (aktor) memiliki karakteristik wajah orang Indonesia, yaitu kulit berwarna sawo matang hingga kuning kecoklatan, iris mata berwarna coklat kehitaman, rambut cenderung berwarna hitam hingga kecoklatan dengan tekstur lurus hingga ikal bergelombang, serta hidung lebar, tetapi tidak terlalu menonjol.

Agar wajah subjek dapat dievaluasi oleh sistem, kamera diletakkan di sebelah kiri kasir dengan posisi membelakangi cahaya. Subjek juga diharuskan untuk melepas kacamata, agar tidak terdapat pantulan cahaya dan wajah dapat terlihat jelas. Pada akhirnya, video berhasil diambil dengan durasi sepanjang 6 menit 44 detik, format 60 FPS, dan resolusi 1920 x 1080. Namun, karena terdapat subjek dengan wajah yang tidak terlihat jelas, maka video diedit dan menjadi berdurasi 4 menit 42 detik. Gambar 3.7 adalah sampel frame set data video demonstrasi pengujian.

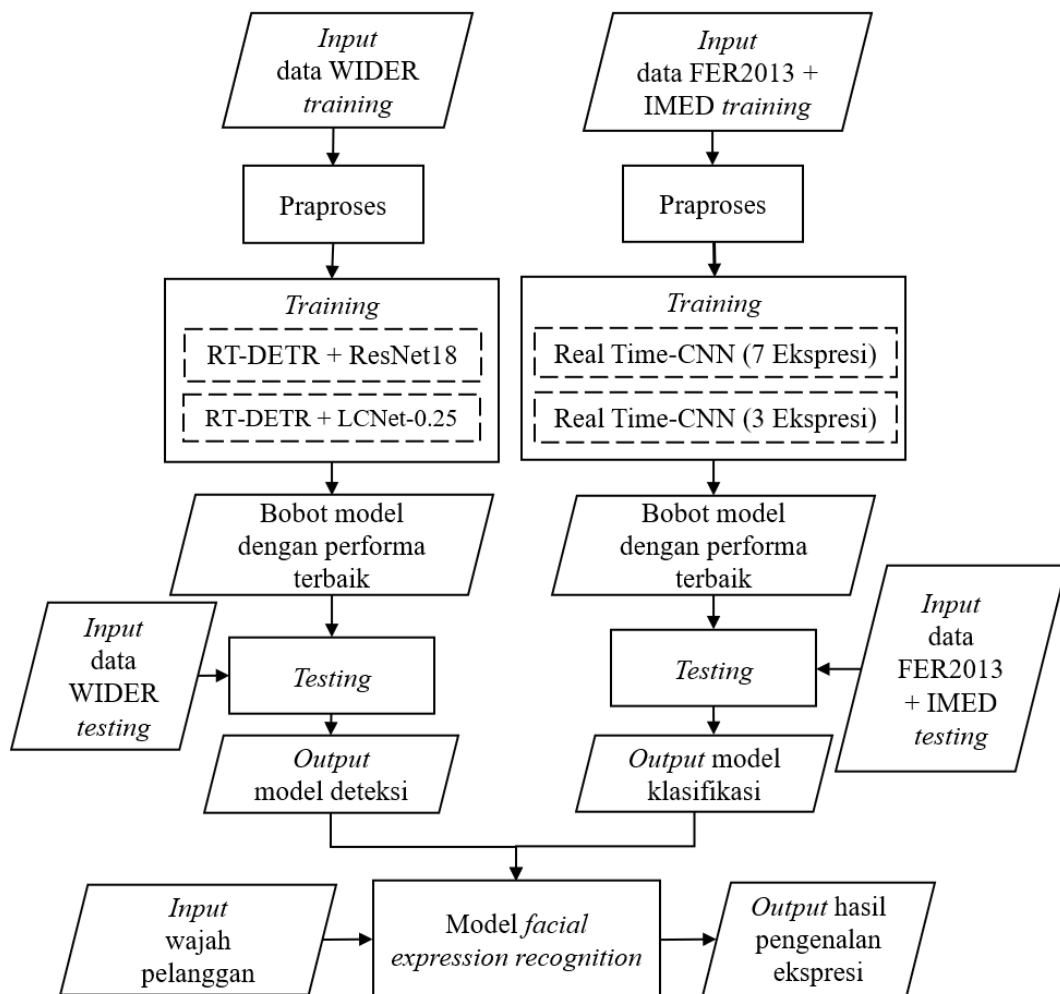


Gambar 3.8 Persentase jenis ekspresi

Pada video ini subjek diarahkan untuk berperilaku sebagai pelanggan pada minimarket, setiap subjek bergiliran menampilkan ekspresi netral, senang, sedih, muak, marah, takut, dan terkejut secara berurutan ketika sedang melakukan transaksi dengan kasir. Sayangnya karena subjek tidak memiliki keterampilan untuk berekspresi seperti aktor, ekspresi yang dipraktikkan subjek tidak serumit yang telah diarahkan. Pada gambar 3.8, dapat diketahui pembagian jenis ekspresi yang mampu dilakukan subjek.

Walaupun set data diambil dalam bentuk video, tetapi ketika digunakan untuk pengujian model pengenalan ekspresi wajah, video akan dikonversi menjadi *frames* atau citra (*still image*).

3.3. Rancangan Implementasi



Gambar 3.9 Rancangan implementasi

Pada tahap ini dibangun model sesuai pada rancangan gambar 3.9 yang melibatkan RT-DETR dengan ResNet-18 dan LCNet-0.25, serta Real-Time CNN dengan 7 ekspresi dan 3 ekspresi. Selanjutnya model diimplementasikan menggunakan set data WIDER-face, FER-2013 dan IMED, serta set data demonstrasi pengujian. Eksperimen terdiri dari dua tahap, yaitu tahap deteksi wajah dan klasifikasi ekspresi.

Walaupun metode utama dari penelitian ini, yaitu RT-DETR, memiliki kemampuan untuk deteksi dan klasifikasi secara langsung (*end-to-end*), tetapi set data yang digunakan memiliki keterbatasan. Keterbatasan set data tersebut yaitu tidak adanya set data deteksi wajah yang telah melibatkan jenis kelas ekspresi. Hal tersebut mengakibatkan harus menggunakan dua jenis set data untuk deteksi wajah dan klasifikasi ekspresi secara terpisah.

3.3.1. Model Deteksi Wajah

Model deteksi wajah dikembangkan menggunakan set data WIDER-face. Sebelum melalui tahap *training*, set data WIDER-face melalui tahap praproses yang berfungsi agar data anotasi sesuai dengan format COCO. Hal tersebut bertujuan agar set data dapat diterima oleh RT-DETR. Selanjutnya, pada tahap *training*, dilakukan eksperimen pada tiga kombinasi model, dengan rincian sebagai berikut:

1. RT-DETR dan *backbone* ResNet-18
2. RT-DETR dan *backbone* LCNet-0.25

Setelah *training* selesai dilakukan pada seluruh kombinasi model, selanjutnya dipilih bobot model terbaik berdasarkan performa, yaitu *epoch* terbaik berdasarkan besar IoU dan *average precision* (AP). Kemudian, dengan bobot model terbaik, dilakukan *testing* menggunakan set data WIDER-face. Evaluasi model dilakukan dengan metrik *average precision* (AP), FPS, dan ukuran model (parameter). AP yang digunakan memiliki *threshold* IoU 0.50 hingga 0.95 karena merupakan metrik evaluasi utama yang digunakan pada evaluasi COCO (Lin dkk., 2014).

3.3.2. Model Klasifikasi Ekspresi

Secara umum, pengimplementasian model dimulai dari *input* set data FER-2013 dan IMED *training* dan *testing* yang telah melewati praproses. Tahap praproses dilakukan agar set data IMED sesuai dengan format FER-2013. Format tersebut, yaitu *grayscale*, rasio 1:1, ukuran 48 x 48 piksel, serta dalam format *array* piksel yang disimpan dalam CSV.

Setelah set data IMED selesai dipraproses, maka data ini digabung dengan set data FER-2013. Memasuki tahap *training*, set data digunakan sebagai input untuk melatih model Real-Time CNN. Selanjutnya, model melalui tahap *testing* untuk menguji performa model klasifikasi ekspresi. Evaluasi model dilakukan dengan metrik *precision*, *recall*, dan *F1-score*.

Berdasarkan gambar 3.6, set data FER-2013 dan IMED, mengalami *imbalance* pada kelas netral. Selain pengembangan menggunakan 7 jenis ekspresi wajah (Ekman, 1970), untuk mempermudah pengevaluasian ekspresi wajah, model klasifikasi wajah juga akan dikembangkan menggunakan 3 jenis emosi saja, yaitu positif, negatif, dan netral. Pengkategorian 3 jenis ekspresi wajah dilakukan berdasarkan penelitian Robinson (2008) dengan rincian yang dapat dilihat pada tabel berikut. Setelah model klasifikasi ekspresi dikembangkan berdasarkan 2 tipe jenis ekspresi, yaitu 7 jenis ekspresi dan 3 jenis ekspresi, selanjutnya model dipilih berdasarkan performa terbaik.

Tabel 3.1 Rincian kategori 7 ekspresi dan 3 ekspresi

7 Ekspresi (Ekman, 1970)	3 Ekspresi (Robinson, 2008)
Senang	Positif
Kaget	
Takut	Negatif
Marah	
Sedih	
Muak	
Netral	Netral

3.3.3. Integrasi Model Deteksi Wajah dan Klasifikasi Ekspresi

Setelah model deteksi wajah dan klasifikasi ekspresi selesai dikembangkan, kedua model tersebut diintegrasikan agar terbentuk menjadi sebuah sistem

pengenalan ekspresi wajah. Kemudian, eksperimen dilanjutkan untuk melakukan evaluasi pengenalan ekspresi wajah pada set data demonstrasi pengujian yang telah melewati tahap praproses.

Supaya set data dapat digunakan sebagai data pengujian, video harus diubah menjadi *frame* dengan jumlah *frame* berdasarkan besar FPS yang dapat diproses oleh model integrasi. Selain itu, perlu ada anotasi *ground truth* berupa *bounding box* serta label kelas ekspresi. *Bounding box* dianotasi melalui model yang telah teruji untuk deteksi wajah, yaitu YuNet (Wu dkk., 2023) yang mencapai 81.1% mAP pada set data WIDER. Kemudian, anotasi label kelas ekspresi dilakukan secara manual oleh manusia berdasarkan sistem FACS (*Facial Action Coding System*) (Ekman & Friesen, 1978). Dilakukan juga, penentuan batas ROI agar model tidak memproses wajah pelanggan lain selain yang sedang bertransaksi.

Terakhir, model memproses *input* wajah pelanggan dan menghasilkan *output* berupa hasil pengenalan ekspresi wajah pelanggan. Setelah model selesai memprediksi posisi dan ekspresi wajah. Selanjutnya hasil akan dievaluasi untuk mendapatkan nilai FPS dan *average precision*.

3.4. Analisa dan Evaluasi

Tahap ini berfungsi untuk mengevaluasi kinerja model berdasarkan hasil eksperimen yang telah dilakukan. Berikut merupakan rincian dari seluruh metrik evaluasi yang digunakan:

1) *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

2) *Precision*

$$Precision = \frac{TP}{TP+FP}$$

3) *Recall*

$$Recall = \frac{TP}{TP+FN}$$

4) *F1-score*

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

5) *Average precision (AP)*

$$AP = \sum_{k=1}^{k=n-1} [Recall_{k+1} - Recall_k] p_{interp}(r_{i+1})$$

6) *Framerate per second (FPS)*

7) Ukuran model atau jumlah parameter

3.5. Lingkungan Komputasi

Penelitian ini dilakukan pada beberapa perangkat keras dan perangkat lunak tertentu. Rincian spesifikasi dari seluruh perangkat yang dilakukan adalah sebagai berikut:

Tabel 3.2 Spesifikasi perangkat

Laptop pribadi	Komputer Laboratorium Kecerdasan Buatan
<i>Perangkat keras</i>	
• RAM 20GB • SSD 477GB • 8 CPU AMD Ryzen 5 3500U • GPU AMD Radeon(TM) Vega 8 Graphics	• RAM 64GB • SSD 1TB • 16 CPU Intel(R) Core(TM) i7-6900K • GPU Nvidia GeForce GTX 1080 Ti
<i>Perangkat lunak</i>	
• Windows 10 • Google Colab • Visual Studio Code • Google Chrome • Microsoft Edge	• Ubuntu 20.04 LTS • Jupyter Hub