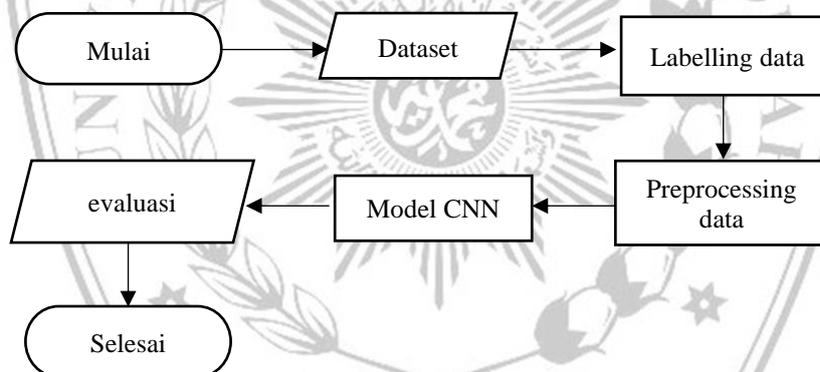


## BAB III

### METODOLOGI PENELITIAN

Pada bab Penelitian ini menggunakan model penelitian dengan Convolutional Neural Network (CNN ) merupakan sebuah penelitian atau studi yang dilakukan untuk menganalisis sentimen atau pendapat pengguna terhadap fitur musik dalam aplikasi Instagram di Google Play Store. Metode yang digunakan dalam penelitian ini adalah Convolutional Neural Network, yang merupakan sebuah jenis model deep learning yang terinspirasi oleh cara kerja visual cortex dalam otak manusia. Penelitian ini dilakukan Untuk identifikasi sentiment negatif dan positif pada ulasan yang mengandung fitur musik aplikasi instagram pada google playstore. Berikut beberapa tahapan yang dilakukan pada penelitian ini dijelaskan pada gambar 1.



Gambar 2 Alur penelitian

Gambar 2. Menunjukkan diagram langkah penelitian yang digunakan untuk melakukan analisis sentimen ulasan aplikasi intagram pada playstore. Langkah penelitian yang pertama adalah memulai dengan dataset yang diambil menggunakan teknik web screpping. Kedua melakukan langkah pemberian label pada data, data yang sudah di ambil maka disini data akan dilakukan pemberian label sentimen negatif dan positif. Sebelum pemberian label, dataset juga dilakukan preprocessing awalan, contohnya menghapus angka atau emoticon, selanjutnya dilakukan

tahap preprocessing lanjutan atau pemrosesan data. Ketiga, melakukan pemrosesan data, pada tahap ini akan dibagi menjadi beberapa tahap cleaning data. Keempat melakukan tahap modelling data menggunakan CNN (Convolutional Neural Network), tahap ini akan dilatih menggunakan data train yang telah diberi label sentimen negatif dan positif. Kelima melakukan prediksi dengan testing data menggunakan metode CNN akan dilatih dengan data train dan di evaluasi menggunakan data test.

### **3.1 Mengumpulkan Data Ulasan**

Data Ulasan diambil dengan menggunakan Teknik Web Screpping. Library yang digunakan adalah google play scraper berbasis python, data ulasan yang diambil dari google playstore pada aplikasi Instagram. Google Play Scraper merupakan suatu api yang digunakan untuk dilakukan ekstraksi data dari Google Play Store. Setelah itu, data disimpan dalam bentuk csv yang terkumpul data teks yang terkait fitur musik yang dimiliki oleh aplikasi instagram sebanyak 2.260 ulasan.

### **3.2 Labelling Data**

Data ulasan aplikasi instagram pada Google Play Store hasil screpping yang telah didapat, Tahap selanjutnya membuat labelling secara manual. Pemberian label yang dilakukan ini agar data dapat terfilter yang sesuai terkait fitur musik aplikasi instagram. Setelah itu, data dapat dilakukan klasifikasi sentimen pada dataset teks. Proses ini mempermudah klasifikasi sentimen meskipun digunakan secara manual akan tetapi dapat mengetahui setiap detailnya ulasan yang mengandung fitur musik aplikasi instagram. Proses pemberian label ini memungkinkan untuk klasifikasi teks dalam kategori positif dan negatif.

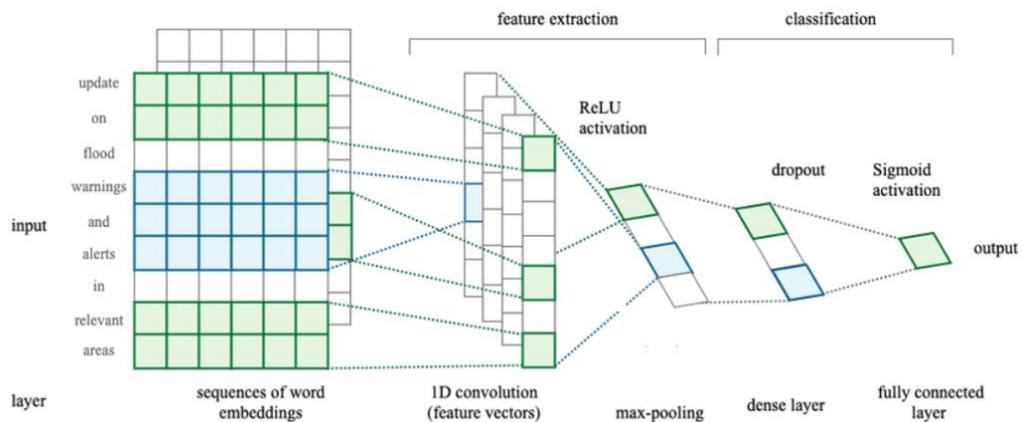
### **3.3 Preprocessing**

Pada Dasarnya, Dataset Yang Diperoleh belum terstruktur. Oleh Karena Itu, Sebelum Dataset Dimasukan Kedalam Model, Data Terlebih Dahulu Melalui Tahap Preprocessing Data. Tahapan Ini Dilakukan Untuk Membersihkan Data Dari Noise Dan Mengubah Data Menjadi Data Yang Terstruktur. Tahap Preprocessing Meliputi :

- a) Case Folding ialah Tahapan Untuk Menyeragamkan Seluruh Teks Yang Akan Dimasukan Kedalam Model Menjadi Huruf Kecil Semua (Lowercase)[21].
- b) Tokenize Yaitu Proses Pemecahan Kata Pada Kalimat Ulasan, Pemisahan Kata Dalam Kalimat Umumnya Menggunakan Karakter[16].
- c) Remove mentions yaitu suatu tahap preprocessing untuk dilakukan menghapus username, seperti @username dalam teks ulasan yang termasuk[22].
- d) Remove hashtags yaitu tahap preprocessing dilakukan tahap menghilangkan tanda pagar, seperti # dalam suatu teks ulasan[22].
- e) Remove punctuation yaitu tahap preprocessing untuk menghilangkan tanda baca pada suatu teks ulasan[22].
- f) Remove links yaitu tahap preprocessing untuk menghilangkan tautan atau URLs dalam suatu teks ulasan, seperti http[22].
- g) Stopword Removal Yaitu Proses Menghilangkan Kata Yang Termasuk Kedalam Kategori Stopword. Stopword Adalah Kata Yang Sering Muncul Namun Dianggap Tidak Memiliki Arti[8].
- h) Lemmatization yaitu suatu tahap preprocessing yang digunakan untuk mengubah kata menjadi bentuk dasarnya, seperti kata benda menjadi kata benda bentuk dasarnya[23].

### **3.4 Model CNN (Convolutional Neural Network)**

Arsitektur CNN yang digunakan dalam penelitian ini memiliki lima lapisan: convolutional layer, embedding layer, dropout layer, polling layer, dense layer. Berikut ilustrasi arsitektur CNN dapat dilihat pada Gambar 3.



Gambar 3 Arsitektur CNN[24]

Pada gambar 3 menunjukkan layer-layer yang digunakan dalam penelitian ini. Layer pertama embedding ,kedua model memiliki vektor-vektor dimensi 100, yang artinya setiap kata di representasikan oleh vektor 100-dimensi untuk memahami teks secara mendalam tentang teks ulasan yang diolah. Layer kedua convolutional, model satu menggunakan filters yang memiliki ukuran kernel 3 dan 4 yang artinya setiap filter melihat,semisal tiga kata secara bersamaan dalam teks ulasan, sedangkan model dua menggunakan ukuran 4 kernel. Layer ketiga dropout, kedua model menggunakan layer ini dengan rate 0,5 untuk mengurangi overfitting pada model. Layer keempat pooling, kedua model digunakan max pooling dengan ukuran pool 2 yang artinya setiap pool mengambil nilai maksimum dari dua nilai berturut. Layer kelima dense, kedua model memiliki 128 neuron untuk membantu mengekstraksi fitur-fitur di proses sebelumnya. Penelitian ini juga menggunakan K-Fold Cross Validation untuk membantu memberikan performa model yang lebih akurat dengan membagi dataset menjadi k subset atau fold dengan angka 5 serta menggunakan 0,0001 untuk angka learning rate.

### 3.5 Skenario Percobaan

Pada penelitian ini menggunakan model 1D-CNN untuk pengolahan data menggunakan dataset teks. Sebelum melakukan tahap modelling, dilakukan terlebih dahulu penyeimbangan data dengan teknik Smote. Lalu

melakukan pembagian data atau splitting yaitu data train 80% dan data test 20%. Peneliti membandingkan menggunakan 2 percobaan yang masing-masing memiliki 2 model untuk dijadikan skenario eksperimen. Terdapat 6 layer yang digunakan oleh ke dua percobaan pada masing-masing model, yaitu Embedding, Conv1D, MaxPooling1D, GlobalMaxPooling1D, Dense, dan Dropout.

Perbedaan kedua percobaan model tersebut dilihat dari hyperparameter dari Conv1D. Percobaan pertama skenario model pertama memiliki filters 32,16,8. Sedangkan, percobaan pertama Skenario model kedua memiliki jumlah filters 256,128,32. Penggunaan optimizer pada percobaan pertama menggunakan Adam. Pada percobaan pertama ini menggunakan K-fold Cross Validation dengan random state 32. Perbedaan lain dari percobaan pertama menggunakan jumlah epoch 50 dan jumlah batch 20. Kedua model dalam percobaan ini menghasilkan akurasi 93% dan nilai loss untuk model pertama 24% dan 34% untuk model kedua. Hasil tersebut dipengaruhi oleh tingkat penggunaan jumlah epoch, ukuran batch, dan keseimbangan data. Percobaan ini dilakukan dengan tujuan untuk mengukur performa model dalam hyperparameter yang digunakan.

Perbedaan percobaan kedua skenario model pertama memiliki jumlah filters 64,32,16 dan untuk model kedua memiliki jumlah filters 128,64,32. Penggunaan random state K-fold Cross Validation sejumlah 42. Selain itu, perbedaan lain juga ada pada penggunaan 80 epoch dan 32 batch size pada percobaan kedua ini. Hasil dari percobaan ini untuk nilai akurasi model pertama 97% dan model kedua 95%, sedangkan untuk nilai loss model pertama 20% dan 21% untuk nilai loss model kedua. Hasil percobaan kedua ini dipengaruhi oleh penggunaan jumlah random state pada k-fold sejumlah 42, jumlah batch, epoch, dan keseimbangan data.

### 3.6 Evaluasi

Penelitian ini setelah diimplementasikan pemodelan dari kedua percobaan tersebut, maka dilakukan tahap evaluasi dengan menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengukur kinerja kedua model dalam mengklasifikasikan sentimen ulasan. Setelah model CNN dievaluasi, model dapat digunakan untuk identifikasi sentiment negatif dan positif pada ulasan yang mengandung fitur musik aplikasi instagram pada google playstore. Hasil perbedaan antara kedua percobaan model arsitektur yang digunakan, dapat mengetahui hasil performa data penelitian yang menggunakan CNN. Hasil performa ditunjukkan dalam bentuk diagram plot loss, plot akurasi, classification report, dan confusion matrix. Pada confusion matrix yang menunjukkan pola kesalahan klasifikasi yang dibuat oleh model pertama dan kedua pada masing-masing kelas. Berikut ulasan yang salah diprediksi oleh sistem.

Ulasan	Terprediksi oleh sistem	Label asli
Fitur musik instagram bermasalah kebanyakan musik terkenal tersedia fitur musik jelek	Positif	Negatif
Suka dengan instagram fitur memudahkan fitur musik semoga mempengaruhi kedepannya	Negatif	Positif
Fitur musik bahasa indonesianya tolong perbaiki aplikasinya	Negatif	Positif

Tabel 2 Prediksi