

BAB III

METODOLOGI PENELITIAN

3.1. Objek Penelitian

Objek penelitian pada penelitian ini adalah dataset yang digunakan pada penelitian sebelumnya [8]. Tabel 3.1 di bawah menunjukkan dataset terdiri 2 kolom yaitu kolom *tweet* dan label. Dataset ini memiliki 4401 tweet dengan 5 label yang berbeda yaitu *happy*, *love*, *fear*, *sadness*, dan *anger*. Terdapat 1101 label *anger*, 1017 label *happy*, 997 label *sadness*, 649 label *fear*, dan 637 label *love*. *Username*, *URL*, dan nomor-nomor sensitif seperti nomor telepon dan nomor invoice telah diubah kedalam format tertentu. Proses anotasi atau pelabelan dataset dilakukan menggunakan kamus emosi Indonesia sebagai pedoman [31].

Tabel 3.1. Contoh Data *Twitter Emotion Dataset*

No	Label	Tweet
1	Anger	Ku juga pengen ngamuk bacain komen2 netizen yg mahabenaar. Ya Lord. Ga abis pikir. Komen2 macem gini yg bikin kasus pelecehan seksual akan tenggelam begitu saja. Tidak ada hukuman untuk pelaku. Lebih sedihnya, rerata yg komen jahat adalah perempuan. Mbaq2, kesehatanmu loh! [URL]
2	Fear	Tadi lagi asik ngobrol sama temen-temen SMA terus terbersit di pikiran sendiri "gimana ya aku nanti di alam kubur teman-temanku ini kan hanya fana" serem banget pikiranku ini
3	Happy	Kepingin gudeg mbarek Bu hj. Amad Foto dari google, sengaja, biar teman-teman jg membayangkannya. Berbagi itu indah.
4	Love	kan kupeluk engkau erat2 hingga tak ada seorang pun yang bisa merebut mu dari pelukan ku happy anniversary syg #6nov #31months
5	Sadness	Turut berduka cita buat kawan a cak. Semoga. Keluarga yg di tinggalkan a. Bisa di beri kesabaran & meridoi kepergian alm.. #sajete #wani

3.2. Metode Penelitian

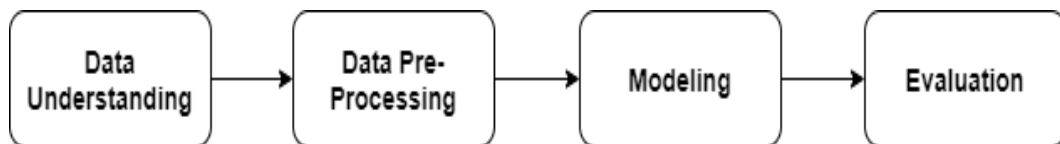
3.2.1. Variabel Independen

Variabel independen adalah variabel yang mempengaruhi variabel lain. Dalam dataset *twitter_emotion_dataset*, yang merupakan variabel independen adalah *tweet*.

3.2.2. Variabel Dependen

Variabel dependen adalah variabel yang dipengaruhi oleh variabel lain. Dalam dataset *twitter_emotion_dataset*, yang merupakan variabel dependen adalah label.

3.3. Kerangka Berpikir



Gambar 3.1. Kerangka berpikir

3.3.1 *Data Understanding*

Data Understanding ditujukan untuk mengenal dan menganalisa dataset yang akan digunakan.

3.3.2 *Data Pre-Processing*

Data pre-processing adalah tahap dimana dataset akan disiapkan terlebih dahulu sebelum diteruskan untuk diproses oleh model yang akan

digunakan. Terdapat beberapa tahap yang dilakukan pada data *pre-processing*, yaitu:

a. *Data normalization*

Pada tahap ini, dilakukan beberapa hal yaitu:

- *Case folding* dimana data *tweet* akan diubah menjadi huruf kecil (a-z) semua.
- Menghapus [*USERNAME*], [*URL*], [*SENSITIVE-NO*], angka, *white space* dan tanda baca.
- Merubah kata singkatan atau kata salah eja menjadi kata yang benar menggunakan Kamus *Typography* Indonesia yang telah ada di penelitian sebelumnya [32] [33], *open source* Kamus *Typography* Indonesia dan Kamus *Typography* Indonesia yang dibuat secara manual.

b. *Stemming*

Pada tahap ini, data *tweet* yang telah dinormalisasi akan diubah menjadi kata dasar dengan tujuan untuk meningkatkan probabilitas kata yang mirip dengan kata dasar yang mirip dihitung atau dianggap satu kata.

c. *Tokenizing*

Pada tahap ini, data *tweet* akan di bagi menjadi kata dan masing-masing kata akan dijadikan token dimana nantinya setiap kalimat atau *tweet* yang ada pada dataset akan diubah kedalam

bentuk token untuk memudahkan model menganalisa arti dari suatu kalimat atau *text* berdasarkan urutan katanya [34].

d. *One Hot Encoding*

One hot encoding adalah salah satu teknik yang dapat digunakan untuk merubah data kategorikal menjadi data integer [35]. Pada tahap ini, akan dilakukan *one hot encoding* terhadap label dari dataset yang digunakan.

e. *Split data*

Pada tahap ini, data akan dibagi menjadi 2 yaitu data *training* dan data *validation*.

3.3.3 *Modeling*

Modeling adalah tahap dimana model akan dibangun dan dilatih dengan dataset yang telah disiapkan untuk memenuhi tujuan penelitian. Model akan dibuat menggunakan salah satu algoritma *deep learning* yaitu *CNN* dengan *layer LSTM*. Algoritma ini digunakan karena berdasarkan penelitian terdahulu, algoritma *CNN* dengan *layer LSTM* dapat memberikan akurasi dan performa yang baik untuk mengklasifikasi teks.

3.3.4 *Evaluation*

Evaluation adalah tahap dimana model yang dibangun akan dievaluasi performanya. Evaluasi dilakukan dengan membandingkan performa model yang telah dibangun dengan *baseline* pada penelitian yaitu model yang telah dikembangkan pada penelitian sebelumnya. Performa

yang akan dibandingkan dalam penelitian ini adalah akurasi model dimana akurasi dari model yang dikembangkan dalam penelitian sebelumnya adalah sebesar 69.73% [8].