

Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM

Yuliana Romadhoni^{1*)}, Khadijah Fahmi Hayati Holle²

^{1,2}Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim, Malang
^{1,2}Jl. Gajayana No.50, Dinoyo, Kec. Lowokwaru, Kota Malang, Jawa Timur 65144
email: ¹18650012@student.uin-malang.ac.id, ²khadijah.holle@uin-malang.ac.id

Abstract – Research on Sentiment Analysis of public responses to PERMENDIKBUD No. 30 on Twitter social media can use Machine Learning and Deep Learning models. This study uses 2 methods derived from the two models, namely the Naïve Bayes method and the Long Short-Term Memory method. Data collection by crawling data using the Twitter API which uses keywords in the form of "permendikbud30" and "Sexual violence on campus". contains "Negative" and "Positive" However, the dataset that has been preprocessed is reduced to 471 data. After preprocessing is done, then the weighting process is carried out using the TF-IDF method and continued with the calculation method. The results of this study indicate that the LSTM method gets a higher performance value, namely the Accuracy value of 77%, Precision of 84%, Recall of 75%, and F1-Score of 80%. testing the Naïve Bayes method obtained results of 76% accuracy, 75% precision, 75% recall value and 75% F1-Score.

Kata Kunci – Sentiment Analysis, PERMENDIKBUD No.30, Naïve Bayes, Long Short-Term Memory

Abstrak – Penelitian Analisis Sentimen tanggapan masyarakat terhadap PERMENDIKBUD No.30 pada media sosial twitter dapat menggunakan model *Machine Learning* dan *Deep Learning*. Pada penelitian ini menggunakan 2 metode yang berasal dari kedua model tersebut yaitu metode Naïve Bayes dan metode Long Short-Term Memory. Pengumpulan data dengan cara *crawling* data menggunakan API Twitter yang mana menggunakan *keyword* berupa "permendikbud30" dan "Kekerasan seksual di kampus" hasil *crawling* pada 2 *keyword* tersebut terdapat data sebanyak 2765 teks *tweet*, yang mana data ini akan dilakukan proses labeling dengan mengisi label pada teks berisi "Negatif" dan "Positif" Namun dataset yang telah dilakukan *Preprocessing* berkurang hingga menjadi 471 data. Setelah dilakukan *Preprocessing*, selanjutnya dilakukan proses pembobotan dengan menggunakan metode TF-IDF dan dilanjutkan dengan perhitungan metode. Hasil penelitian ini menyatakan bahwa metode LSTM mendapatkan nilai performa lebih tinggi yaitu nilai Akurasi sebesar 77%, Presisi sebesar 84%, *Recall* sebesar 75%, dan *F1-Score* sebesar 80%. pengujian metode Naïve Bayes didapatkan hasil Akurasi sebesar 76%, Presisi sebesar 75%, Nilai *Recall* sebesar 75% dan *F1-Score* sebesar 75%.

Kata Kunci – Analisis Sentimen, PERMENDIKBUD No.30, Naïve Bayes, Long Short-Term Memory

*) penulis korespondensi: Yuliana Romadhoni
Email: 18650012@student.uin-malang.ac.id

I. PENDAHULUAN

Beragam kekerasan seksual, kekerasan seksual ini terjadi pada semua subjek hukum khususnya anak-anak dan perempuan. Namun telah dilakukannya penghapusan kekerasan seksual yang telah diatur pada RUU merupakan elaborasi dari kewajiban suatu negara dalam menegakkan hukum dan mengurangi adanya kekerasan seksual yang banyak terjadi pada perempuan dan anak-anak [1]. Kekerasan seksual terjadi bukan hanya di lingkungan pelacuran, pasar, transportasi umum, di jalan atau lingkungan yang memungkinkan orang lawan jenis saling berinteraksi, namun kekerasan seksual sering dijumpai pada lingkungan kantor, lingkungan keluarga bahkan lingkungan Pendidikan [2]. Ramainya perbincangan adanya kekerasan seksual pada perguruan tinggi, baik perguruan tinggi keagamaan (PTKI) maupun perguruan tinggi umum. Maka terdapat buku yang disusun oleh pihak PTKI yang membahas tentang pedoman pencegahan dan penanganan kasus kekerasan seksual di PTKI yang bertujuan untuk menjunjung tinggi nilai kemanusiaan dan norma kesusilaan [3]. Hal ini perlu dijunjung tinggi untuk meminimalisir adanya kekerasan seksual di Indonesia khususnya pada perguruan tinggi. Namun pada 14 November 2021 sedang ramai diperbincangkan mengenai isi Permendikbud No 30 tahun 2021. Beberapa kalangan menilai isi Permendikbud Ristek ini melegalkan seks bebas dikarenakan terdapat isi yang menimbulkan arti melegalkan perbuatan asusila dan seks bebas yang berbasis persetujuan [4]. Pembahasan ini sedang ramai diperbincangkan pada Media Sosial, banyaknya korban yang bercerita dan mengungkapkan tentang adanya perlakuan kekerasan seksual yang telah menimpa dirinya. Media sosial merupakan salah satu media online yang dapat digunakan untuk mempermudah manusia yang pada dasarnya makhluk sosial untuk melakukan interaksi berbagai hal. [5] Media sosial yang ramai membicarakan Permendikbud No 30 salah satunya yaitu Twitter. Terdapat Pro dan Kontra pada pembahasan Permendikbud No.30, maka dari itu dilakukan penelitian ini untuk mengukur nilai akurasi dari klasifikasi menggunakan metode LSTM pada pembahasan Permendikbud No 30. Penelitian ini melakukan klasifikasi sentimen ini yang bertujuan agar pemerintah menegakkan dan lebih memperkuat aturan pada undang-undang dengan memperhatikan tanggapan masyarakat pada media sosial, maka dari itu penelitian ini membutuhkan metode yang digunakan pada pengukuran

performa. Beberapa metode yang dapat digunakan untuk mengukur performa yaitu metode *Naïve bayes* dan juga metode *Long Short-Term Memory* (LSTM) yang mana metode ini telah memiliki kelebihan pada masing-masing fungsinya. Penelitian yang menggunakan metode Metode *Long Short-Term Memory* ini menyatakan bahwa metode tersebut banyak diusulkan pada permasalahan klasifikasi dan mendapatkan nilai akurasi tinggi sebesar 95,38% [6]. Selain metode *Deep Learning* terdapat metode *Machine Learning* yang banyak direkomendasikan dan dilakukan penelitiannya yaitu metode *Naïve Bayes*. Metode ini telah dilakukan pada penelitian [7] telah melakukan penelitian dan menyatakan bahwa rata-rata performansi yang dilakukan menggunakan metode *naïve bayes* nilainya stabil dan mendapat nilai akurasi sebesar 85,5%, presisi 83%, recall 86% dan f1-score 84%.

II. PENELITIAN YANG TERKAIT

Banyaknya penelitian *text mining* mengakibatkan berkembangnya model yang dapat digunakan untuk melakukan perhitungan pada teks mining berikut. Pada penelitian [8] melakukan penelitian analisis sentimen menggunakan kalimat berbahasa Arab dan menunjukkan bahwa pendekatan yang diusulkan mengungguli metode dasar dan relatif kompetitif untuk pekerjaan terkait lainnya. Selain itu, menerapkan *bidirectional LSTM* dengan *embedding feature* telah secara signifikan meningkatkan hasil pendekatan untuk ekstraksi ekspresi target opini aspek (OTE) (meningkat sebesar 39% dibandingkan hasil dasar). Penelitian ini juga menggunakan fitur berbasis *attention-based* dengan menghubungkan aspek *embedding* ke *corresponding word embedding* untuk klasifikasi polaritas sentimen (AB-LSTM PC) telah signifikan meningkatkan hasil pendekatan dalam identifikasi polaritas sentimen aspek sebesar 6%. Penelitian [9] mengidentifikasi setiap data yang berupa kata tweet penelitian tersebut memberikan arti ke dalamnya. Pekerjaan fitur digabungkan dengan kata-kata tweet, *word2vec*, *stopword* dan diintegrasikan ke dalam teknik *deep learning* model jaringan saraf *Convolution* dan *Long short Term Memory*, algoritma ini dapat mengidentifikasi pola jumlah *stop word* dengan strateginya sendiri. Kedua model tersebut dilatih dengan baik dan diterapkan untuk dataset IMDB yang berisi 50.000 review film. Dengan sejumlah besar data twitter diproses untuk memprediksi *tweet* sentimenal untuk klasifikasi. Dengan metodologi yang diusulkan, sampel yang dikumpulkan secara eksperimental dari lingkungan waktu nyata dapat dibedakan dengan baik dan kemandirian sistem ditingkatkan. Hasil algoritma *Deep Learning* bertujuan untuk menilai *tweet review* dan juga mampu mengidentifikasi review film dengan akurasi pengujian sebesar 87,74% untuk CNN dan nilai akurasi menggunakan LSTM sebesar 88,02%. Penelitian [10] menyatakan bahwa telah melakukan Ekstraksi fitur berbasis model skip gram telah digunakan untuk penyisipan kata. Lewati-gram Word ke representasi Vektor membutuhkan lebih sedikit ruang memori dan menghasilkan akurasi yang terus menerus lebih tinggi jika dibandingkan dengan representasi Word ke vektor lainnya. LSTM dapat mencapai hasil yang baik bahkan tanpa optimasi karena kemampuan generalisasi yang baik. Untuk lebih meningkatkan kinerja LSTM, digunakan algoritma APSO untuk optimasi parameter bobot LSTM. Kontribusi APSO dalam memilih parameter

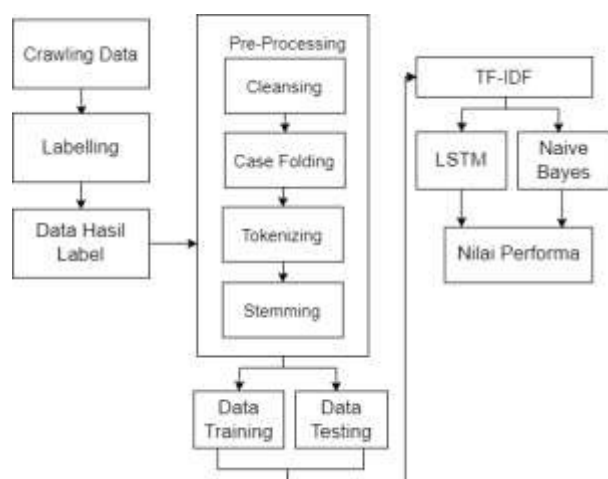
bobot untuk jaringan saraf tiruan LSTM meningkatkan akurasi dan menurunkan kompleksitas komputasi. Untuk analisis eksperimental, empat jenis dataset telah digunakan. Evaluasi kinerja metode yang diusulkan telah dilakukan dengan menggunakan metrik yang berbeda seperti akurasi, *recall*, presisi, dan *F-measure*. Metodologi yang diusulkan mencapai akurasi maksimum 96,8% untuk kumpulan data Amazon, 97,8% untuk kumpulan data penasihat perjalanan, 93,2% untuk kumpulan data demonetisasi, dan 95,2%. Penelitian [11] menggunakan pembobotan fitur dan kalibrasi Laplace dan melakukan algoritma klasifikasi *Naïve Bayes* yang ditingkatkan. Melalui simulasi numerik, ditemukan bahwa ketika ukuran sampel besar, akurasi algoritma klasifikasi *Naïve Bayes* yang ditingkatkan lebih dari 99%, dan sangat stabil; ketika atribut sampel kurang dari 400 dan jumlah kategori kurang dari 24, akurasi algoritma klasifikasi *Naïve Bayes* yang ditingkatkan lebih dari 95%. Melalui penelitian empiris, ditemukan bahwa algoritma klasifikasi *Naïve Bayes* yang ditingkatkan dapat sangat meningkatkan dan tingkat analisis diskriminasi yang benar dari 49,5 menjadi 92%. Melalui analisis sentimen dengan algoritma klasifikasi *Naïve Bayes* yang ditingkatkan memiliki akurasi yang lebih tinggi. Penelitian [12] melakukan perbandingan antara metode LSTM dengan metode *Naïve bayes* dan menyatakan bahwa Metode Long Short-Term Memory memiliki akurasi

yang lebih baik daripada metode *Naïve Bayes* dengan akurasi 72,85%, presisi 73%, recall 72%, dan f-measure 72% dibandingkan hasil metode *Naïve Bayes* dengan akurasi 67,88%, 69% presisi, 68% recall, dan 68% f-measure. Metode LSTM 2 layer memiliki akurasi yang lebih baik daripada LSTM 1 layer yang menghasilkan akurasi 72,85%, presisi 73%, recall 72%, dan f-measure 72% sedangkan layer LSTM 1 menghasilkan akurasi 71,20%, presisi 71%, recall 71%, dan 70% f-measure. Artinya menambahkan lapisan pada metode LSTM dapat meningkatkan hasil akurasi.

III. METODE PENELITIAN

A. Analisis Sentimen

Analisis sentimen merupakan penelitian berupa *text mining* yang memiliki tujuan sebagai analisis pendapat, evaluasi, sikap, sentimen seseorang yang berupa teks untuk ditinjau pada suatu objek apakah kalimat tersebut termasuk kedalam kalimat positif atau kedalam kalimat negatif. Ada banyak penelitian yang menerapkan analisis sentimen pada sebuah review, bahkan Kaggle dan SemEval 2017 mengadakan kompetisi untuk mengidentifikasi metode terbaik untuk klasifikasi sentimen. Metode machine learning seperti *Naïve Bayes*, Maximum Entropy (ME), dan Support Vector Machine (SVM) sering digunakan dalam menemukan model dan fitur yang sesuai dengan target masalah. SVM dan ME merupakan model yang kompleks sehingga waktu pelatihan lebih lama, sedangkan *Naïve Bayes* merupakan model yang sederhana dan cepat [12]. Selain metode *naïve bayes* terdapat metode LSTM yang biasa digunakan untuk klasifikasi teks. Setelah mengetahui metode yang akan digunakan penelitian ini merancang desain sistem sebagai berikut.



Gambar. 1 Desain Sistem

B. Preprocessing

Data yang digunakan pada teks mining menggunakan data yang belum diproses atau bisa disebut dengan data mentah, maka sebelum melakukan proses analisis sentimen, data harus melewati proses *Preprocessing* terlebih dahulu, *Preprocessing* yang digunakan pada penelitian ini terdapat 4 proses, yaitu sebagai berikut :

a. Cleansing Data

Proses ini dilakukan untuk membersihkan hal yang tidak dibutuhkan seperti angka, simbol-simbol atau tanda baca, emoticon, alamat url, mention twitter dll. Dilakukan *cleansing data* agar data lebih bersih dari simbol dan kata.

b. Case Folding

Proses *Case Folding* digunakan untuk merubah seluruh huruf kapital menjadi huruf kecil atau lower case. Hal ini dilakukan tanpa terkecuali pada awal kalimat, penamaan orang, kota dll.

c. Tokenizing

Proses *tokenizing* ini proses yang dibutuhkan pada proses TF-IDF dikarenakan proses ini merubah kalimat menjadi perkata, yang mana kata pada proses *tokenizing* ini akan diberikan bobot. Proses ini dilakukan dengan adanya spasi.

d. Stemming

Proses *stemming* ini proses untuk merubah setiap kata yang memiliki imbuhan menjadi kata dasar, *stemming* dilakukan pada setiap kata yang memiliki imbuhan, imbuhan pada awal kata (*prefixes*), imbuhan pada tengah kata (*infixes*) maupun imbuhan diakhir kata (*suffixes*).

C. TF-IDF

Penelitian analisis sentimen ini sebelum melakukan perhitungan menggunakan metode deep learning ataupun metode *machine learning* terlebih dahulu harus melakukan pembobotan kata, agar setiap kata memiliki bobot dan dapat dilakukan perhitungan. *Term Frequency-Inverse Document Frequency* (TF-IDF) ini merupakan algoritma yang dapat melakukan pembobotan kata. Pada algoritma ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap dokumen yang telah ditentukan pada korpus. Pengertian secara mudah TF-IDF ini untuk mengetahui seberapa sering kata pada suatu dokumen muncul [13]. Terdapat rumus yang biasa digunakan yaitu :

$$Tf(w) = \frac{\text{frekuensi muncul kata } w \text{ di dokumen } d}{\text{Total kata dokumen } d} \quad (1)$$

$$IDF(w) = \log_e \frac{\text{Jumlah total dokumen}}{\text{total dokumen pada kata } w} \quad (2)$$

D. Naïve Bayes

Setelah dilakukan proses TF-IDF selanjutnya melakukan perhitungan salah satunya menggunakan metode *Naïve bayes*. *Naïve bayes* merupakan metode *machine learning* yang mempunyai model dalam bentuk peluang atau probabilitas. *Naïve bayes* ini mempunyai kemampuan mengklasifikasikan serupa dengan *Decision Tree* dan *Neural Network* yang telah terbukti menghasilkan nilai akurasi dan kecepatan yang cukup tinggi [7]. Berikut adalah persamaan pada algoritma *Naïve Bayes*.

$$P(C) = \frac{N_c}{N} \quad (3)$$

Keterangan :

N_c = Banyak dokumen pada kelas c

N = Jumlah dokumen

Persamaan (3) digunakan untuk menghitung nilai yang ada dari probabilitas *prior* [14]. Selain untuk menentukan probabilitas *prior*, ada selanjutnya menghitung *Posterior*.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad 4$$

Keterangan :

X: Data pada kelas yang belum diketahui

C: Hipotesa data yaitu suatu kelas yang spesifik

$P(C|X)$: Probabilitas *Posterior*, merupakan probabilitas hipotesis C kondisi X

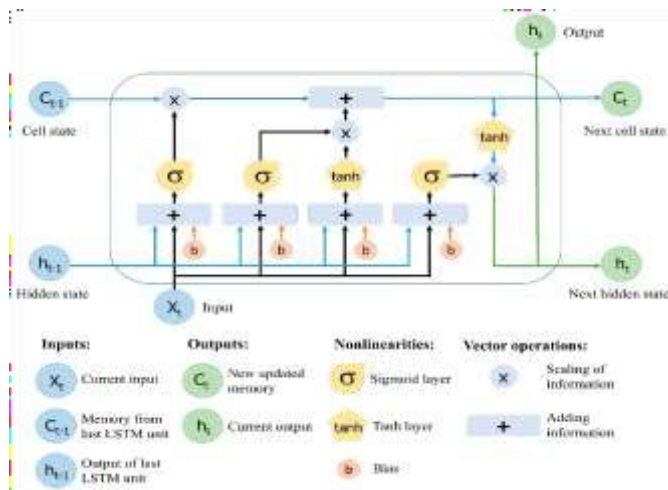
$P(C)$: Probabilitas *Prior*, probabilitas hipotesa C

$P(X|C)$: *Likelihood* atau *Conditional Probability*, probabilitas hipotesa X pada kondisi C

$P(X)$: Evidence, adalah peluang (*probability*) munculnya sebuah kata.

E. Long Short-Term Memory

Metode *Long Short-Term Memory* ini merupakan metode deep learning yang dapat digunakan pada bidang *Natural Language Processing* (NLP) diantaranya yaitu penerjemahan teks, pengenalan ucapan dan dapat digunakan untuk klasifikasi gambar maupun teks. Telah banyak yang melakukan penelitian menggunakan metode LSTM ini dan menyatakan bahwa metode ini mempunyai hasil yang lebih baik dibandingkan dengan metode konvensional, dan metode ini dapat dikatakan cocok dalam penerapan analisis sentimen [12]. Berikut adalah struktur jaringan dan juga rumus perhitungan dengan metode LSTM.



Gambar. 2 Struktur Jaringan LSTM [15]

Pada LSTM terdapat 4 *gates units* yaitu : *Input gate*, *Forget Gate*, *Cell gates*, dan *Output Gates*[15]. *Forget Gate* informasi yang ada pada input data akan diolah dan dipilah data mana yang layak untuk disimpan atau tidak layak disimpan pada *memory cells*. Terdapat persamaan fungsi *forget gate* yaitu [13]

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (5)$$

Gates pada *Input gates* sebanyak 2 *gates* yang akan dilaksanakan, yang pertama menentukan nilai yang diperbarui dengan fungsi aktivasi sigmoid. Dan nilai aktivasi tanh akan membuat vector nilai terbaru yang akan disimpan di *memory cell*. Persamaannya yaitu [13] :

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (6)$$

$$C'_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (7)$$

Cell gates akan berfungsi sebagai pengganti *memory cell* yang lama dengan *memory cell* yang baru, nilai ini berasal dari gabungan *forget gates* dengan *input gates*. Berikut persamaan dari *Cell gates*[13].

$$C_t = (f_t * C_{t-1} + i_t * C'_t) \quad (8)$$

Langkah terakhir yang akan dilakukan yaitu *output gates*, yang akan melakukan 2 langkah, Langkah pertama menentukan nilai yang ada pada *memory cell* yang akan dikeluarkan menggunakan fungsi aktivasi sigmoid, yang dilanjutkan penempatan nilai *memory cell* dengan menggunakan fungsi tanh. Langkah terakhir kedua *gates* akan dilakukan perkalian sehingga mendapatkan hasil nilai output. Dengan persamaan sebagai berikut [13].

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (9)$$

$$h_t = o_t * \tanh(C_t) \quad (10)$$

Keterangan Rumus:

- f_t = Forget gate
- i_t = Input gate
- c_t = Cell gates
- o_t = Output gate
- h_t = Hidden State
- \check{c}_t = Cell Aktivasi
- b = Bias
- W = Bobot

F. Confussion Matrix

Tujuan penelitian ini untuk mendapatkan performa pada analisis sentimen, maka dibutuhkan *confussion matrix* untuk mengukur performa, *confussion matrix* memiliki persamaan sebagai berikut.

$$\text{Akurasi \%} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (11)$$

$$\text{Presisi \%} = \frac{TP}{TP + FP} \times 100\% \quad (12)$$

$$\text{Recal \%} = \frac{TP}{TP + FN} \times 100\% \quad (13)$$

$$F1 - \text{Score \%} = \frac{2 \times \text{Presisi} + \text{Recall}}{\text{Presisi} + \text{Recall}} \times 100\% \quad (14)$$

Keterangan :

TP = True Positive yaitu jumlah data yang memiliki nilai positif dan diprediksi positif.

FP = False Positive yaitu data yang bernilai negatif tetapi diprediksi positif.

FN = False Negative yaitu jumlah data positif tetapi diprediksi negatif.

TN = True Negative yaitu data yang bernilai negatif dan diprediksi negatif.

IV. HASIL DAN PEMBAHASAN

Penelitian analisis sentimen tersebut telah melakukan pengumpulan data dengan cara *crawling* data menggunakan API Twitter yang mana menggunakan *keyword* berupa “permendikbud30” dan “Kekerasan seksual di kampus” hasil *crawling* pada 2 *keyword* tersebut terdapat data sebanyak 2765 teks *tweet*, yang mana data ini akan dilakukan proses labeling dengan mengisi label pada teks berisi “Negatif” dan “Positif” oleh ahli Bahasa yang mana dapat melakukan pelabelan secara benar dan baik secara manual. Berikut adalah data yang telah dilakukan *crawling* dan juga labeling manual.

Tabel 1 Pengumpulan Data

| Teks | Label |
|--|---------|
| b'Bahas kekerasan seksual sama hima kampus di zoom, pematerynya orang bem fakultas. Gak jelas, kenapa bahas ke kesetaraan gender. Pelecehan seksual gak akan sama dengan kesetaraan gender, kalau enggak ada solusi yg bagus. Skip. Berasa baca buku dongeng tentang feminisme.' | Negatif |
| b'[PERNYATAAN SIKAP BERSAMA DUKUNG IMPLEMENTASI PERMENDIKBUD RISTEK NO 30 TAHUN 2021 DI UGM]\n\nKekerasan seksual di lingkungan Universitas Gadjah Mada adalah permasalahan yang masih kerap terjadi. Hal ini berarti kampus belum dapat menjadi ruang aman untuk seluruh civitas akademika https://t.co/b2B8Aewqc5' | Negatif |

| | |
|---|---------|
| b'@adstrophille Mmf moots kalau ini lewat. Sender, ini bisa dilaporkan ya. Ini termasuk dalam pelecehan seksual. Sekarang juga sudah ada payung hukumnya ttg kekerasan seksual di lingkungan kampus, Permendikbud no 30 th 2021. Pls jangan dibiarin.' | Positif |
| b'@bertanyar! Ya brarti permendikbud 30 bagus buat ngurusin kasus kekerasan seksual di kampus. Buktinya para korban banyak yang berani speak up, lapor ke pihak kampus/berwenang.' | Positif |
| b'@catchmeupid Akhirnya, semoga dengan ini kekerasan seksual di kampus bisa diberantas habis agar tak ada lagi porn fugitive di kampus-kampus Indonesia!' | Positif |
| b'@Ji_Star99 Kalo yang ui bisa dipidanakan bakal jadi bom kasus kekerasan seksual dikampus lain nggak sih wkwk di kampus lain kek iceberg soalnya' | Negatif |
| b'@glrhn Indeed. Bahkan media seperti tempo jadi makin berani untuk buat reportase soal kekerasan seksual di kampus. Which is goood is' | Positif |
| b'@ghastpiracy This is why we need satgas kekerasan seksual di kampus, biar clear alasannya kalo ada kejadian kaya begini' | Positif |
| b'@geloraco Harusnya tak perlu permendikbud untuk kejahatan kekerasan seksual, sudah bisa dipidana perkosaan. Justru dengan permendukbud yg baru, Guru Besar itu boleh berhubungan seks di kampus jika suka sama suka.' | Negatif |
| b'@collegemenfess Bukannya peraturan ini dibuat mengkhusus untuk kekerasan seksual ya? Gak ada kampus yg melegalkan seks bebas bahkan di norma kita aja gak ada. Bukan berarti diterbitkan peraturan ini mhswh malah bisa se enaknyanya. Udah gede. Udah tau mana yang boleh, mana yang salah.' | Positif |

Setelah melakukan labeling secara manual maka dilakukan *Preprocessing* pada tahap ini penulis menggunakan beberapa library untuk membantu penelitian ini agar lebih mudah yaitu library Sastrawi, library Porter, library Keras, library Scikit-learn. tahapan pada preprocessing sesuai dengan pada Gambar.1 Desain Sistem yaitu:

1. Cleansing Data

Tabel 2 Cleansing Data

| | |
|------------------------|--|
| Sebelum Cleansing Data | b'Bahas kekerasan seksual sama hima kampus di zoom, pamerterinya orang bem fakultas. Gak jelas, kenapa bahas ke kesetaraan gender. Pelecehan seksual gak akan sama dengan sama kesetaraan gender, kalau enggak ada solusi yg bagus. Skip. Berasa baca buku dongeng tentang feminisme.' |
|------------------------|--|

| | |
|------------------------|---|
| Sesudah Cleansing Data | Bahas kekerasan seksual sama hima kampus di zoom pamerterinya orang bem fakultas Gak jelas kenapa bahas ke kesetaraan gender Pelecehan seksual gak akan sama dengan sama kesetaraan gender kalau enggak ada solusi yg bagus Skip Berasa baca buku dongeng tentang feminisme |
|------------------------|---|

2. Case Folding

Tabel 3 Case Folding

| | |
|----------------------|---|
| Sebelum Case Folding | Bahas kekerasan seksual sama hima kampus di zoom pamerterinya orang bem fakultas Gak jelas kenapa nyerempet ke kesetaraan gender Pelecehan seksual gak akan sama dengan sama kesetaraan gender kalau enggak ada solusi yg bagus Skip Berasa baca buku dongeng tentang feminisme |
| Sesudah Case Folding | bahas kekerasan seksual sama hima kampus di zoom pamerterinya orang bem fakultas gak jelas kenapa nyerempet ke kesetaraan gender pelecehan seksual gak akan sama dengan sama kesetaraan gender kalau enggak ada solusi yg bagus skip berasa baca buku dongeng tentang feminisme |

3. Tokenizing

Tabel 4 Tokenizing

| | |
|--------------------|--|
| Sebelum Tokenizing | bahas kekerasan seksual sama hima kampus di zoom pamerterinya orang bem fakultas gak jelas kenapa nyerempet ke kesetaraan gender pelecehan seksual gak akan sama dengan sama kesetaraan gender kalau enggak ada solusi yg bagus skip berasa baca buku dongeng tentang feminisme |
| Sesudah Tokenizing | “bahas” “kekerasan” “seksual” “sama” “hima” “kampus” “di” “zoom” “pamerterinya” “orang” “bem” “fakultas” “gak” “jelas” “kenapa” “nyerempet” “ke” “kesetaraan” “gender” “pelecehan” “seksual” “gak” “akan” “sama dengan” “sama” “kesetaraan” “gender” “kalua” “enggak” “ada” “solusi” “yg” “bagus” “skip” “berasa” “baca” |

| | |
|--|---|
| | “buku” “dongeng” “tentang” “feminisme” |
|--|---|

4. Stemming

Tabel 5 Stemming

| | |
|------------------|--|
| Sebelum Stemming | “bahas” “kekerasan” “seksual” “sama” “hima” “kampus” “di” “zoom” “pematerinya” “orang” “bem” “fakultas” “gak” “jelas” “kenapa” “nyerempet” “ke” “kesetaraan” “gender” “pelecehan” “seksual” “gak” “akan” “sama” “dengan” “sama” “kesetaraan” “gender” “kalua” “enggak” “ada” “solusi” “yg” “bagus” “skip” “berasa” “baca” “buku” “dongeng” “tentang” “feminisme” |
| Setelah Stemming | “bahas” “keras” “seksual” “sama” “hima” “kampus” “di” “zoom” “materi” “orang” “bem” “fakultas” “gak” “jelas” “kenapa” “nyerempet” “ke” “setara” “gender” “leceh” “seksual” “gak” “akan” “sama” “dengan” “sama” “setara” “gender” “kalua” “enggak” “ada” “solusi” “yg” “bagus” “skip” “berasa” “baca” “buku” “dongeng” “tentang” “feminisme” |

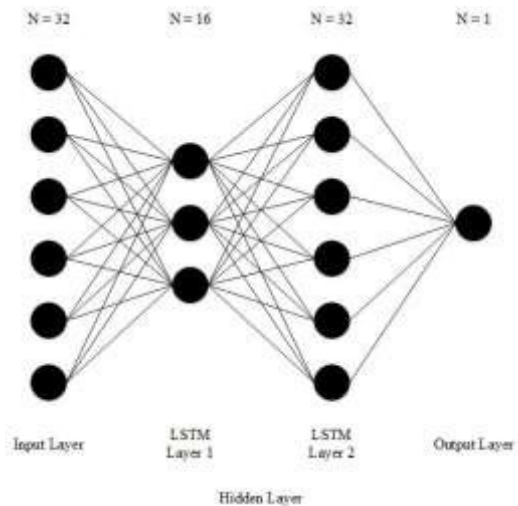
Setelah dilakukan *Preprocessing* maka data akan berkurang dan menjadi 471 teks tweet. Setelah melakukan *Preprocessing* maka data dapat dilakukan proses TF-IDF atau pembobotan. Setelah melakukan pembobotan lalu dilakukan perhitungan performa dengan metode *Naïve Bayes* dan *Long Short Term Memory*.

A. *Naïve Bayes*

Metode ini menggunakan data yang sama dengan pengujian metode LSTM, yaitu sebesar 470 data tweet, dengan presentase pembagian data latih dan data uji sebesar 80% dan 20% yang mana data latih sebanyak 376 data dan data uji sebanyak 94 data. Data tersebut juga telah melalui proses *preprocessing* yang sama sedangkan ekstraksi ciri yang digunakan adalah TF (Term-Frequency). Selanjutnya model tersebut diuji dengan data testing. Hasil pengujian metode *Naïve Bayes* didapatkan hasil Akurasi sebesar 76%, Presisi sebesar 75%, Nilai *Recall* sebesar 75% dan *F1-Score* sebesar 75%.

B. *Long Short-term Memory*

Data yang diolah pada metode ini sama persis dengan data yang digunakan pada *Naïve Bayes* presentase pembagian pada data latih dan uji sama-sama sebesar 80% dan 20%. Pada LSTM terdapat arsitektur LSTM yang diilustrasikan pada gambar berikut.



Pada Gambar 3. Menyatakan bahwa *Input Layer* didapatkan nilai sebesar $n=32$ dan terdapat 2 *Hidden Layer* yang berupa LSTM Layer 1 dan LSTM Layer 2 yang mana LSTM Layer 1 bernilai $n=16$ dan LSTM Layer 2 bernilai $n=32$ dan *Output Layer* bernilai $n=1$. Setelah mengukur dengan metode LSTM ini maka didapatkan nilai Akurasi sebesar 77%, Presisi sebesar 84%, *Recall* sebesar 75%, dan *F1-Score* sebesar 80%.

C. Perbandingan Metode *Naïve Bayes* dan *Long Short-Term Memory*

Data dan Tahapan *Processing* pada penelitian analisis sentimen yang telah dilakukan ini telah diberikan data dan proses yang sama, dengan menggunakan 2 metode dalam menentukan hasil performa maka dapat dilakukan perbandingan pada 2 metode tersebut, berikut hasil perbandingan menggunakan metode *Naïve Bayes* dan LSTM.

Tabel 6 Perbandingan Metode

| Metode | Waktu | Akurasi | Presisi | Recall | F1-Score |
|-------------|--------|---------|---------|--------|----------|
| Naïve Bayes | 01'01' | 76% | 75% | 75% | 75% |
| LSTM | 02'05' | 78% | 84% | 78% | 81% |

Nilai performa perbandingan metode tersebut menyatakan bahwa model *Deep Learning* lebih tinggi dibandingkan model *Naïve Bayes* dikarenakan model *Deep Learning* lebih modern dibandingkan model *Machine Learning*. Namun pada *Deep Learning* membutuhkan data yang besar. Metode LSTM juga membutuhkan waktu lebih lama dibandingkan metode *Naïve Bayes*, dikarenakan pada LSTM membutuhkan pengolahan epoch dimana jumlah epoch mempengaruhi tingginya nilai akurasi.

V. KESIMPULAN

Kesimpulan pada penelitian Analisis Sentimen yang dapat membandingkan 2 metode menyatakan bahwa:

- Hasil 2 metode tidak sama meskipun tahapan

preprocessing dan jumlah data. Metode LSTM memiliki nilai performa yang lebih besar daripada metode Naïve Bayes. Dengan nilai Akurasi sebesar 77%, Presisi sebesar 84%, Recall sebesar 75%, dan F1-Score sebesar 80%. Naïve Bayes mendapatkan nilai Akurasi sebesar 76%, Presisi sebesar 75%, Nilai Recall sebesar 75% dan F1-Score sebesar 75%.

- Waktu yang digunakan akan berbeda pada 2 metode tersebut. LSTM lebih lama dibandingkan Naïve Bayes.

Namun pada penelitian tersebut masih banyak kekurangan yaitu.

- Data yang digunakan pada penelitian terbilang kecil, dikarenakan metode *deep learning* membutuhkan data yang besar.
- Membandingkan metode sesama *deep learning* (*Convolution Neural Network, Neural Network*) atau sesama metode *machine learning* (*Support Vector Machine, K-Nearest Neighbor*).

UCAPAN TERIMA KASIH

Terima kasih kepada pihak yang telah membantu dalam melakukan labeling manual, dan terima kasih kepada dosen pengampu mata kuliah *Natural Language Processing* yaitu Ibu Khadijah Fahmi Holle yang telah membantu dan memberikan materi agar terselesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] A. Purwanti and M. Hardiyanti, "Strategi Penyelesaian Tindak Kekerasan Seksual Terhadap Perempuan Dan Anak Melalui Ruu Kekerasan Seksual" vol. jilid 47 N, pp. 138–148, 2018.
- [2] Y. K. Huda, P. Studi, I. Hukum, F. Hukum, and U. M. Surakarta, "(Studi Putusan Nomor Perkara : 76 / PID / 2015 / PT . MDN dan Putusan Nomor Perkara : 145 / Pid . B / 2016 / PT . PBR)," 2020, [Online]. Available: <https://core.ac.uk/download/pdf/289186285.pdf>.
- [3] Komnas Perempuan, *Pedoman Pencegahan & Penanganan Kasus Kekerasan Seksual Di Perguruan Tinggi Keagamaan Islam (Ptiki)*. 2020.
- [4] "Isi Permendikbud Ristek Nomor 30 Tahun 2021 yang Tuai Pro Kontra." <https://www.kompas.com/tren/read/2021/11/14/140000465/isi-permendikbud-ristek-nomor-30-tahun-2021-yang-tuai-pro-kontra> (accessed Dec. 09, 2021).
- [5] Anang Sugeng Cahyono, "Pengaruh media sosial terhadap perubahan sosial masyarakat di Indonesia," *J. Ilmu Sos. Ilmu Polit. diterbitkan oleh Fak. Ilmu Sos. Polit. Univ. Tulungagung*, vol. 9, no. 1, pp. 140–157, 2016, Accessed: Aug. 02, 2021. [Online]. Available: <https://journal.unita.ac.id/index.php/publiciana/article/view/79>.
- [6] Winda Kurnia Sari, "Klasifikasi Teks Menggunakan Long Short-Term Memory Dengan Fitur Word Embedding Glove Dan Word2Vec," 2020.
- [7] A. Nisa, E. Darwiyanto, and I. Asror, "Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 8650–8659, 2019.
- [8] M. Al-Smadi, B. Talafha, M. Al-Ayyoub, and Y. Jararweh, "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 8, pp. 2163–2175, 2019, doi: 10.1007/s13042-018-0799-4.
- [9] U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu, and G. Karthick, "Sentimen Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," *Wirel. Pers. Commun.*, no. 0123456789, 2021, doi: 10.1007/s11277-021-08580-3.
- [10] J. Shobana and M. Murali, "An efficient sentiment analysis methodology based on long short-term memory networks," *Complex Intell. Syst.*, vol. 7, no. 5, pp. 2485–2501, 2021, doi: 10.1007/s40747-021-00436-4.
- [11] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00742-6.
- [12] M. A. Nurrohmat and A. SN, "Sentimen Analysis of Novel Review Using Long Short-Term Memory Method," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 3, p. 209, 2019, doi: 10.22146/ijccs.41236.
- [13] I. L. Rais and J. Jondri, "Klasifikasi Data Kuesioner dengan Metode Recurrent Neural Network," *eProceedings Eng.*, vol. 7, no. 1, pp. 2817–2826, 2020.
- [14] P. Aliandu, "Sentimen Analysis to Determine Accommodation, Shopping and Culinary Location on Foursquare in Kupang City," 2015, doi: 10.1016/j.procs.2015.12.144.
- [15] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of Long Short-Term Memory (LSTM) neural network for flood forecasting," *Water (Switzerland)*, vol. 11, no. 7, 2019, doi: 10.3390/w11071387.