



# Pemodelan Teks Tweet pada Isu Pelecehan Seksual Berbasis Analisis Sentimen dan Leksikon Emosi

Tresna Maulana Fahrudin<sup>1</sup>, Allan Ruhui Fatmah Sari<sup>2</sup>, Adhisa Shilfadianis Iffadah<sup>3</sup>,  
Chrysilla Citra Windyadari<sup>4</sup>, Gema Khusnul Ma'rifah<sup>5</sup>

<sup>1,2,3,4,5</sup>Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur  
[tresna.maulana.ds@upnjatim.ac.id](mailto:tresna.maulana.ds@upnjatim.ac.id), <sup>1</sup>[21083010007](https://orcid.org/0000-0002-1083-0100), <sup>2</sup>[21083010016](https://orcid.org/0000-0002-1083-0100), <sup>3</sup>[21083010023](https://orcid.org/0000-0002-1083-0100), <sup>4</sup>[21083010034](https://orcid.org/0000-0002-1083-0100), <sup>5</sup>[@student.upnjatim.ac.id](mailto:@student.upnjatim.ac.id)

Corresponding author email: [tresna.maulana.ds@upnjatim.ac.id](mailto:tresna.maulana.ds@upnjatim.ac.id)

## Abstract:

Cases of sexual harassment are still often found in Indonesia, both against female and male victims. The number of cases of sexual harassment in Indonesia based on the SIMFONI PPPA data recorded as many as 12,948 cases, including 1,994 male victims and 11,974 female victims, as of January 1 2022 until the data has been verified in the current month. Incidents of sexual harassment are widely reported through the mass media, television, radio, and information is increasingly spreading through social media. This study tries to analyze the public's perception of the issue of sexual harassment from the perspective of social media, one of which is Twitter. This study aims to analyze the public's perception of the issue of sexual harassment that is currently happening based on tweet text data obtained from Twitter using sentiment analysis and emotional lexicon. The results showed that the accuracy of the classification model built using the Naïve Bayes method was 74% with the percentage of sentiment proportion 72% negative tweets, 18.8% positive tweets, and 9.2% neutral tweets. This shows that there are still more negative opinions regarding the issue of sexual harassment than positive and neutral opinions. The results of emotional analysis using the NRC Emotion Lexicon produce a histogram of various tweet emotions such as fear, anger, anticipation, trust, and other emotions. The Word Cloud visualization displays several frequency words that often appear such as 'harassed', 'sexual', 'victim', 'woman', 'law', 'action', and others.

**Keywords:** Tweet Text Modeling, Social Media, Sexual Harassment, Sentiment Analysis, Emotion Lexicon

## Abstrak:

Kasus pelecehan seksual saat ini masih sering ditemui di Indonesia baik terhadap korban perempuan maupun laki-laki. Jumlah kasus pelecehan seksual di Indonesia berdasarkan data SIMFONI PPPA mencatat sebanyak 12.948 kasus yang diantaranya 1.994 korban laki-laki dan 11.974 korban perempuan yang terhitung sejak tanggal 1 Januari 2022 hingga data yang sudah terverifikasi pada bulan berjalan. Kejadian pelecehan seksual cukup banyak diberitakan melalui media massa, televisi, radio, dan informasi semakin cepat menyebar melalui media sosial. Penelitian ini mencoba untuk menganalisis persepsi masyarakat terhadap isu pelecehan seksual dari perspektif media sosial, salah satunya Twitter. Penelitian ini bertujuan untuk menganalisis persepsi masyarakat terhadap isu pelecehan seksual yang sedang terjadi berdasarkan data teks tweet yang didapatkan dari Twitter dengan menggunakan analisis sentimen dan leksikon emosi. Hasil penelitian menunjukkan akurasi dari model klasifikasi yang dibangun menggunakan metode Naïve Bayes sebesar 74% dengan persentase proporsi sentimen 72% tweet negatif, 18.8% tweet positif, dan 9.2% tweet netral. Hal tersebut menunjukkan bahwa opini negatif terkait isu kekerasan pelecehan seksual masih lebih banyak dibandingkan opini positif maupun opini netral. Hasil analisis emosi menggunakan NRC Emotion Lexicon menghasilkan histogram berbagai emosi tweet seperti ketakutan, kemarahan, kewaspadaan, kepercayaan, dan emosi lainnya. Visualisasi Word Cloud menampilkan beberapa frekuensi kata yang sering muncul seperti 'leceh', 'seksual', 'korban', 'wanita', 'perempuan', 'hukum', 'tindak', dan lainnya.

**Kata kunci:** Pemodelan Teks Tweet, Media Sosial, Pelecehan Seksual, Analisis Sentimen, Leksikon Emosi

## I. PENDAHULUAN

Akhir-akhir ini sering ditemukan kasus pelecehan seksual di Indonesia yang dilakukan oleh oknum tertentu. Hal ini dibuktikan dengan jumlah kasus pelecehan seksual di Indonesia berdasarkan data Sistem Informasi Online Perlindungan Perempuan dan Anak Kementerian PPPA (SIMFONI PPPA) mencatat sebanyak 12.948 kasus, diantaranya 1.994 korban laki-laki dan 11.974 korban perempuan yang terhitung sejak tanggal 1 Januari 2022 hingga data yang sudah terverifikasi pada bulan berjalan [1]. Pengertian pelecehan seksual sendiri menurut Adita Miranti dan Yudi Sudiana, yaitu merujuk pada tingkah laku atau tindakan yang dilakukan secara verbal atau nonverbal yang diantaranya sebagai suatu



keadaan yang tidak dapat diterima, baik secara fisik, lisan atau isyarat seksual dan pernyataan-pernyataan yang bersifat menghina atau pernyataan seksual yang bersifat membedakan antara perempuan dan laki-laki, serta membuat seseorang merasa terancam, dipermalukan, dibodohi, dilecehkan dan dilemahkan kondisi keamanannya [2]. Pelecehan seksual banyak menimpa pada orang yang dianggap lemah tak berdaya atau pada orang yang tidak memiliki kekuasaan.

Kasus pelecehan seksual saat ini masih sering terjadi pada perempuan yang disebabkan oleh sistem tata nilai yang memandang perempuan adalah sebagai kaum marginal dan memiliki status lebih rendah dari pada laki-laki. Walaupun, belakangan ini terjadi kasus pelecehan seksual terhadap laki-laki. Namun, masih banyak kasus terjadi pada perempuan. Selain itu, kentalnya sistem patriarki di masyarakat Indonesia masih sangat terasa di sekitar kita. Sistem patriarki merupakan suatu sikap yang mengutamakan laki-laki sebagai pemegang kekuasaan atau mendominasi dalam sistem sosial di masyarakat.

Pada tahun 2021, terdapat kasus pelecehan seksual terhadap 13 santri di salah satu pondok pesantren daerah Bandung, Jawa Barat. Pelaku merupakan seorang ustaz yang mengelola pondok pesantren tersebut. Ia melangsungkan aksinya sudah dilakukan sejak tahun 2016 dan baru terkuak pada bulan Mei tahun 2021 lalu. Selama itu, sebanyak 8 korban sudah melahirkan 9 bayi dan rentang usia korban adalah 14-20 tahun. Korban selalu dibujuk akan dibiayai selama di pesantren, sekolah gratis menjadi polisi wanita hingga dibiayai kuliah. Akhirnya, setelah melewati persidangan, pelaku divonis hukuman mati dan membayar restitusi sebesar 300 juta rupiah oleh Pengadilan Tinggi Bandung [3].

Kemudian pada tahun yang sama, terbongkar kasus pelecehan seksual terhadap pegawai yang bekerja di Komisi Penyiaran Indonesia (KPI). Korban merupakan seorang lelaki dan pelakunya adalah beberapa seniornya di tempat kerjanya. Korban diterima di KPI pada tahun 2011 dan kerap kali diperlakukan tidak baik oleh seniornya hingga tahun 2014. Kemudian, pada tahun 2015-2020, korban mendapatkan kekerasan dan pelecehan, mengalami gangguan mental dan berusaha melaporkan pada Komnas HAM dan Kepolisian. Namun, pengaduan yang dilakukan oleh korban mendapatkan hasil yang tidak diinginkan dan diarahkan untuk menyelesaikan kasus secara internal. Hingga pada tahun 2021, korban tidak tahan dan membuat pengakuan di media sosial dan kasus tersebut baru diterima dan diproses pada tanggal 1 September 2021 [4].

Dari banyaknya kejadian pelecehan seksual yang diberitakan melalui media massa, televisi, radio, dan beberapa media lainnya. Penelitian ini mencoba untuk menganalisis persepsi masyarakat terhadap isu pelecehan seksual dari perspektif media sosial, salah satunya Twitter. Pengguna twitter saat ini cukup banyak yang aktif dan berpartisipasi menyampaikan kejadian pelecehan seksual yang *up to date*, misalnya pengguna masyarakat yang meresahkan pernyataan bahwasannya pelecehan seksual terjadi karena pakaian wanita yang terlalu terbuka, sering terjadinya unsur pemaksaan terhadap korban pelecehan seksual untuk *speak up* tanpa mempertimbangkan keadaan kesehatan mental korban. Selain itu, menyampaikan bahwasannya edukasi juga harus diberikan kepada orang tua dan lain-lain. Penelitian ini bertujuan untuk menganalisis persepsi masyarakat terhadap isu pelecehan seksual yang sedang terjadi berdasarkan data teks tweet yang didapatkan di Twitter dengan menggunakan analisis sentimen dan leksikon emosi.

## II. PENELITIAN TERKAIT

Penelitian mengenai analisis pengaruh emoticon dan sarkasme terhadap sentimen bahasa Indonesia pernah dilakukan oleh Debby Alita dari Universitas Teknokrat Indonesia, Sigit Priyanta dan Nur Rokhman dari Universitas Gajah Mada. Penelitian ini bertujuan untuk mengetahui sentimen publik mengandung sarkasme atau tidak terhadap pelayanan suatu produk, perusahaan dan pemerintah. Peneliti menggunakan metode Naïve Bayes dan Support Vector Machine (SVM) dalam proses analisis sentimennya sehingga menghasilkan data yang berlabel positif, negatif, serta netral. Selanjutnya dilakukan pendeteksian sarkasme pada data berlabel positif dengan menggunakan perbandingan metode



Random Forest Classifier, Naïve Bayes dan SVM. Dari hasil penelitian didapatkan metode terbaik dalam analisis sentimen sarkasme dengan emoticon adalah metode SVM dengan nilai 69,74% [5].

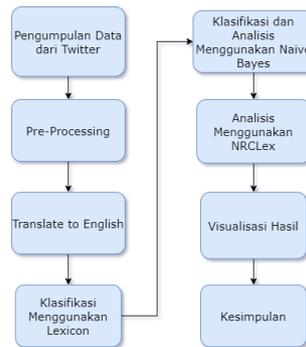
Penelitian pada tahun 2020, terdapat penelitian tentang analisis sentimen respons masyarakat Indonesia terhadap *influencer* yang semakin banyak di media sosial. Penelitian ini dilakukan oleh dua mahasiswa dari Universitas Indonesia yakni Syafi Muhammad Tauhid dan Yova Ruldeviyani. Penelitian tersebut bertujuan untuk mengidentifikasi tanggapan pengguna twitter terhadap tweet *influencer* Fiersa Besari dan Keanu. Peneliti menggunakan beberapa metode diantaranya Naïve Bayes, Support Vector Machine (SVM), Decision Tree, dan Logistic Regression. Dataset dibagi menjadi 80% data *training*, 10% data *testing*, dan 10% data *cross validation*. Dataset diberi label sentimen positif dan sentimen negatif. Dari hasil penelitian model yang memiliki performa nilai F-Score terbaik adalah model Naïve Bayes dengan nilai 81.56% untuk *influencer* Fiersa Besari dan 70.68% untuk *influencer* Keanu. Selain itu, perbandingan hasil akurasi data *testing* dengan data *cross validation* hanya selisih 5% yang berarti tidak terjadi *overfitting* pada model yang digunakan [6].

Penelitian selanjutnya pada tahun 2021 yang dilakukan oleh Azka Al Azkiya dan Utami Syafitri dari Insitut Pertanian Bogor, menganalisis perbandingan opini pengguna twitter terkait isu kesetaraan gender dari tahun 2018-2021 dan menganalisis nilai akurasi sentimen tersebut. Dalam analisisnya peneliti menggunakan metode *Lexicon-Based Classification* dengan bantuan kamus AFINN-111 versi Bahasa Indonesia sebagai penentu polaritas sentimen data teks. Hasil penelitian menunjukkan bahwa setelah dilakukan lima kali pengulangan analisis akurasi didapatkan rata-rata nilai akurasi yang cukup baik yaitu sebesar 66.08% [7].

Penelitian lainnya mengenai analisis sentimen deteksi depresi terhadap pengguna twitter dengan menggunakan metode K-Nearest Neighbor pernah dilakukan oleh Arianti Primadhani Tirtopangarsa dan Warih Maharani dari Universitas Telkom Indonesia. Penelitian ini berdasarkan cuitan para pengguna twitter dan untuk mengetahui performa model. Dalam analisisnya peneliti membagi dataset menjadi dua tipe yaitu dataset label yang terdapat fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) sebelum dilakukan klasifikasi sentimen menggunakan KNN dan dataset *unlabel* sebagai hasil prediksi oleh model KNN. Dari hasil penelitian didapatkan nilai akurasi untuk tipe dataset label sebesar 78.18%, sedangkan untuk dataset *unlabel* nilai akurasinya hanya sebesar 64.78% [8].

Kemudian terdapat penelitian lainnya yang menggunakan metode Naïve Bayes, Support Vector Machine (SVM), dan Logistic Regression oleh Garry, dkk dari Insitut Teknologi Sepuluh November. Penelitian ini bertujuan untuk mengetahui sentimen publik melalui media sosial twitter terhadap Rancangan Undang-Undang Penghapusan Kekerasan Seksual (RUU PKS) serta menentukan model algoritma terbaik dalam analisis sentimen. Label dalam penelitian ini meliputi label sentimen positif yang menunjukkan urgensi pengesahan RUU PKS, label sentimen netral yang menunjukkan kecenderungan masyarakat mendukung kebijakan, dan label sentimen negatif yang menunjukkan penolakan terhadap RUU PKS. Dari hasil penelitian ini didapatkan publik mendukung urgensi pengesahan RUU PKS dan model algoritma terbaik diperoleh algoritma Naïve Bayes dengan akurasi sebesar 65% [9].

### III. METODE PENELITIAN



Gambar 1. Diagram Alir Penelitian

Diagram alir penelitian ini seperti yang ditunjukkan pada Gambar 1 dimulai dari pengumpulan data dari Twitter, *preprocessing*, *translate* tweet ke Bahasa Inggris, *lexicon-based sentiment analysis*, klasifikasi menggunakan Naïve Bayes, NRC Emotion Lexicon, dan visualisasi.

#### 3.1. Pengumpulan Data

Pengambilan data pada penelitian ini melalui media sosial Twitter dengan bantuan website ‘netlytic.org’ untuk mengumpulkan data sebanyak 17.681 tweet berdasarkan *keyword*. *Keyword* yang digunakan yaitu “kesetaraan gender” dan “pelecehan seksual”.

#### 3.2. Preprocessing

Data yang dikumpulkan berbentuk data yang tidak terstruktur (*unstructured dataset*) sehingga membutuhkan teknik *preprocessing*. *Preprocessing* dilakukan melalui beberapa tahapan yaitu *remove punctuation*, *repeated character*, *negative expression replacer*, *stop word*, *stemming*, dan *lemmatization*.

##### 3.2.1. Case Folding

Tahap ini bertujuan untuk menyeragamkan karakter huruf dalam sebuah kalimat menjadi karakter huruf kecil keseluruhan, misalnya seperti kata “MaKan” menjadi “makan”.

##### 3.2.2. Remove Punctuation

Tanda baca atau karakter unik seperti (,!.()-[]{};:'" \ / < > ? @ # % ^ & \* \_ ~ dan sebagainya) pada data teks tweet akan dihapus karena tanda baca tidak dibutuhkan dalam token.

##### 3.2.3. Repeated Character

Tahap ini menghapus karakter yang berulang menjadi makna kata yang sebenarnya sehingga dapat diketahui oleh komputer, misalnya kata “cantikk” akan dirubah menjadi “cantik”.

##### 3.2.4. Tokenization

Tahap ini memisahkan teks menjadi potongan kata yang disebut sebagai token untuk kemudian digunakan dalam analisis. *Tokenize* dapat diimplementasikan pada paragraf maupun kalimat, misalnya kalimat “rumahku terletak jauh dari sini” akan diubah menjadi [‘rumahku’, ‘terletak’, ‘jauh’, ‘dari’, ‘sini’].



### 3.2.5. Stopword

Tahap ini menghapus kata yang tidak memiliki makna atau nilai penting. Dalam hal ini *stopword* membantu menghapus kata yang tidak bermakna seperti kata imbuhan “di”, “yang”, “dalam”, “adalah”, dan sebagainya.

### 3.2.6. Stemming

Proses mengubah kata yang memiliki imbuhan *prefix* dan *suffix* ke dalam bentuk dasarnya, misalnya dalam teks terdapat kata “mencari”, maka akan di-*stemming* menjadi “cari”.

### 3.2.7. Lemmatization

*Lemmatization* merupakan akar kata. Gabungan dari akar kata akan membentuk kata dasar. Akar kata ini juga biasanya berhubungan dengan bahasa lain, misalnya akar kata “pra” memiliki terjemahan “sebelum”, lalu dari akar kata “pra” dapat membentuk kata dasar seperti “pranikah” sehingga *lemmatization* ini mengubah suatu kata ke bentuk akar katanya.

## 3.3. Analisis Sentimen

Klasifikasi sentimen menggunakan metode Naïve Bayes Classifier dalam menentukan positif serta negatifnya suatu opini. Metode ini digunakan karena memiliki keunggulan yaitu sederhana dalam proses komputasinya, tetapi memiliki akurasi yang tinggi. Metode ini memanfaatkan konsep probabilitas dalam perhitungan.

Konsep yang digunakan yaitu teorema peluang bersyarat di mana peluang kejadian A bersyarat B ditentukan peluang A dan peluang B bersyarat A [10] seperti persamaan (1) berikut:

$$P(A|B) = P(A) \cdot P(B|A) \quad (1)$$

Kemudian persamaan (1) tersebut dapat dikembangkan menjadi persamaan (2) berikut:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2)$$

Persamaan ini digunakan untuk klasifikasi dokumen dengan menghitung probabilitasnya. Klasifikasi bertujuan untuk menentukan kategori dari sebuah dokumen. Maka persamaan dapat diubah menjadi persamaan (3) sebagai berikut:

$$P(C_i|D) = \frac{P(D|C_i) \cdot P(C_i)}{P(D)} \quad (3)$$

Pada metode ini setiap variabel berdiri bebas (*independent variable*), tidak ada keterkaitan satu variabel dengan variabel lainnya sehingga suatu dokumen diasumsikan sebagai kumpulan dari token atau kata. Selain itu, juga tidak memperhatikan urutan kemunculan kata pada dokumen. Perhitungan probabilitasnya adalah sebagai hasil perkalian dari probabilitas kemunculan kata-kata dalam dokumen. Persamaan (4) pada metode Naïve Bayes Classifier sebagai berikut:

$$P(C_i) = \frac{f^D(C_i)}{|D|} \quad (4)$$

Keterangan:

$P(C_i)$  : Probabilitas dari suatu kategori dokumen  
 $f^D(C_i)$  : Frekuensi dokumen yang memiliki kategori  $C_i$   
 $|D|$  : Jumlah seluruh dokumen data latih

$$P(W_{kj}|C_i) = \frac{f(W_{kj}|C_i)+1}{f(C_i)+|W|} \quad (5)$$

Keterangan:

$P(W_{kj}|C_i)$  : Probabilitas kemunculan kata  $W_{kj}$  pada suatu dokumen dengan kategori kelas  $C_i$   
 $W_{kj}$  : Frekuensi kata  $ke-k$  pada setiap kategori



“netral” jika *compound score* bernilai sama dengan nol. *Lexicon-based sentiment analysis* ini menggunakan pustaka berbahasa Inggris sehingga perlu diterjemahkan terlebih dahulu teks pada tahap sebelumnya. Berikut Gambar 3 menunjukkan hasil analisis sentimen pada tahap *lexicon-based sentiment analysis*.

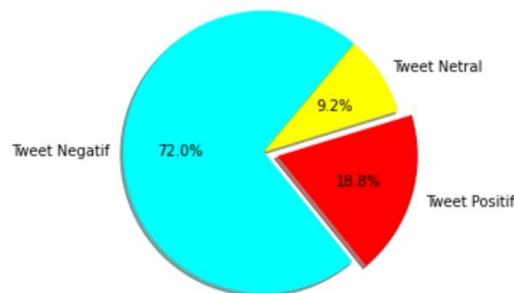
index	Description	Compound_Score	Sentiments
0	0 ma wariiii the act of sexually harassing behav...	-0.0516	Negatif
1	1 management of the park mall solo clarifies reg...	-0.3818	Negatif
2	2 I forgot the subject of this sexual abuse, it'...	-0.6369	Negatif
3	3 fes, let's create a free safe space, let's hel...	0.8750	Positif
4	4 so basically i have to change the subject of m...	0.0000	Netral
5	5 netizen sexual harassment center finished	-0.5423	Negatif
6	6 the son of the kiai, the son of habib, the cle...	-0.6486	Negatif
7	7 sexual harassment	-0.5423	Negatif
8	8 try digitizing it and teach a girl to be sexua...	-0.5423	Negatif
9	9 Gibran suspects that a member of the JKT Mall ...	-0.7096	Negatif

**Gambar 3.** Analisis Sentimen pada Dataset Tweet Pelecehan Seksual

#### 4.2. Tahap Naïve Bayes Classifier

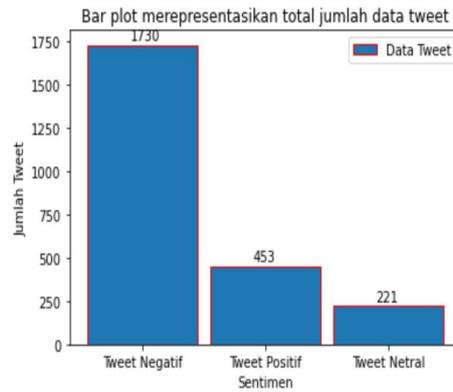
Setelah melakukan *lexicon-based sentiment analysis* yang menghasilkan nilai *compound score* dan data positif, negatif, dan netral, selanjutnya adalah tahap membuat model klasifikasi menggunakan metode Naive Bayes. Pada tahap ini juga akan dilakukan polarisasi data yakni mengubah label data positif menjadi angka 1, label data negatif menjadi angka -1, dan label data netral menjadi angka 0 sehingga didapatkan sebanyak 453 data opini positif, 1.730 data opini negatif, dan 221 data opini netral. Dari data yang telah dikelompokkan tersebut dapat divisualisasikan seperti yang ditunjukkan pada Gambar 4 dalam bentuk *pie chart* dan Gambar 5 dalam bentuk histogram:

Representasi Data Tweet berdasarkan sentimen kedalam pie chart



**Gambar 4.** Pie Chart Tweet Opini Pelecehan Seksual

Pada Gambar 4 disajikan dalam bentuk *pie chart* dengan persentase kategori tweet positif sebesar 18.8%, tweet netral sebesar 9.2%, dan tweet negatif sebesar 72%, sedangkan pada Gambar 5 disajikan dalam bentuk histogram jumlah tweet berdasarkan kategori tweet positif sebesar 453, tweet netral sebesar 221 dan tweet negatif sebesar 1.730.



Gambar 5. Histogram Tweet Opini Pelecehan Seksual

Selanjutnya adalah mengukur keakuratan model klasifikasi sentimen tweet dan didapatkan hasil akurasi sebesar 74% seperti yang ditunjukkan pada Gambar 6.

	precision	recall	f1-score	support
-1	0.74	1.00	0.85	351
0	1.00	0.06	0.12	49
1	0.00	0.00	0.00	81
accuracy			0.74	481
macro avg	0.58	0.35	0.32	481
weighted avg	0.64	0.74	0.63	481

Gambar 6. Akurasi Pada Klasifikasi Data Tweet Pelecehan Seksual

#### 4.3. Tahap NRC Emotion Lexicon (EmoLex)

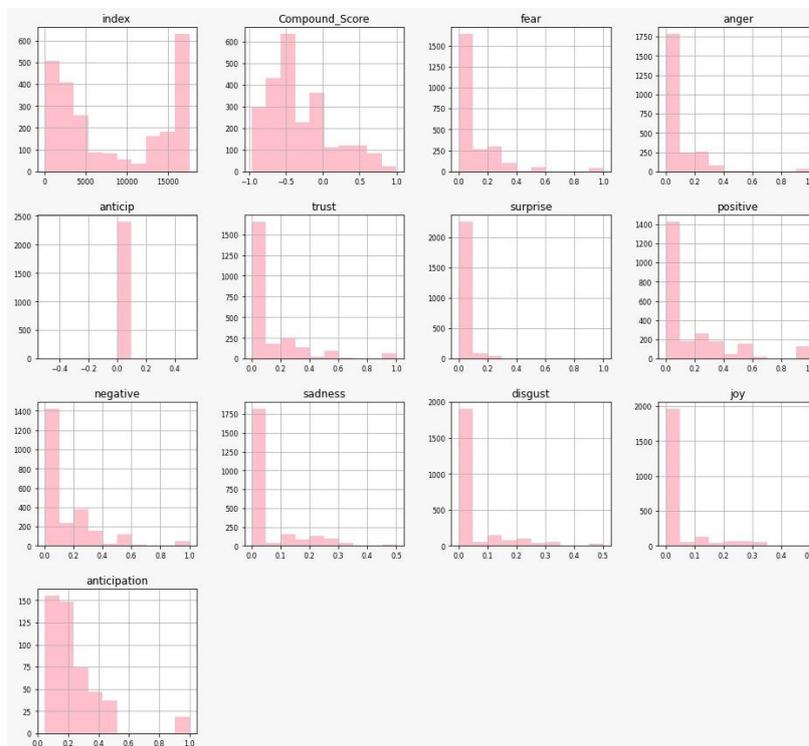
Tahapan selanjutnya adalah mengimplementasikan NRC Emotion Lexicon yang digunakan untuk mengkategorikan label emosi apa saja yang terdapat pada kalimat tersebut. Pelabelan emosi tersebut dilakukan secara otomatis berdasarkan pustaka NRCLex. Tiap label emosi pada teks tweet memiliki nilai skor masing-masing berdasarkan pustaka NRCLex yang di mana akan disimpulkan kecenderungan label emosi, apakah dikategorikan takut, marah, percaya, kaget, sedih, dan label lainnya. Hasil labelisasi emosi menggunakan NRC Emotion Lexicon pada data tweet pelecehan seksual ditunjukkan pada Gambar 7.

```
1 #@title Teks judul default { display-mode: "code" }
2 df = pd.read_csv('hasilnrc_snd_pelecehan_seksual.csv')
3 df.head()
```

index	Description	Compound_Score	Sentiments	fear	anger	anticip	trust	surprise	positive	negative	sadness	disgust	joy	anticipation	emotions
0	ma warilil the act of sexually harassing behav...	-0.0516	Negatif	0.000000	1.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	NaN	[('anger', 1.0)]
1	management of the park mall solo clarifies reg...	-0.3818	Negatif	0.000000	0.000000	0.0	0.500000	0.0	0.500000	0.000000	0.000000	0.000000	0.000000	NaN	[('trust', 0.5), ('positive', 0.5)]
2	I forgot the subject of this sexual abuse, if...	-0.6369	Negatif	0.166667	0.166667	0.0	0.000000	0.0	0.000000	0.333333	0.166667	0.166667	0.000000	NaN	[('negative', 0.3333333333333333)]
3	fes, let's create a free safe space, let's hel...	0.8750	Positif	0.000000	0.000000	0.0	0.285714	0.0	0.428571	0.000000	0.000000	0.000000	0.285714	NaN	[('positive', 0.42857142857142855)]
4	so basically i have to change the subject of	0.0000	Netral	0.500000	0.000000	0.0	0.000000	0.0	0.000000	0.500000	0.000000	0.000000	0.000000	NaN	[('fear', 0.5), ('negative', 0.5)]

Gambar 7. Labelisasi Emosi menggunakan NRC Emotion Lexicon pada Data Tweet Pelecehan Seksual

Labelisasi emosi data tweet ini juga dapat divisualisasikan dalam bentuk histogram yang menunjukkan bagaimana intensitas atau kadar pada masing-masing label emosi. Berikut Gambar 8 menunjukkan histogram NRC Emotion Lexicon pada data tweet pelecehan seksual.



Gambar 8. Histogram NRC Emotion Lexicon Data Tweet Pelecehan Seksual

Berdasarkan histogram pada Gambar 8 didapatkan hasil analisis data tweet pelecehan seksual dengan penilaian negatif sebanyak 1.500 tweet berdasarkan *compound score* pada *axis-x* mulai dari -1 hingga -0.5, hal ini menunjukkan bahwa banyak opini dari para pengguna twitter kontra terhadap pelecehan seksual. Misalnya, terdapat tweet “Mau anak Kiai kek, mau anak Habib, anak Ustadz klo ngelakuin tindak pidana yah tetep harus dihukum. Apalagi ngelakuin pelecehan seksual tuh”. Sebanyak 380 tweet mengomentari kasus pelecehan seksual ini secara netral atau tidak memihak pihak manapun, terdapat juga 300 tweet dengan penilaian positif pada *axis-x* mulai dari 0.5 hingga 1.0 yang di mana terdapat pengguna twitter yang tidak setuju dengan orang yang membenarkan pelaku pelecehan seksual.





## V. KESIMPULAN

Analisis sentimen dan leksikon emosi pada isu pelecehan seksual menjadi hal yang menarik untuk diteliti. Penelitian ini menerapkan algoritma *Lexicon-based* dan Naïve Bayes untuk klasifikasi sentimen dan dilanjutkan dengan NRCLex untuk mendeteksi emosi dari data tweet. Tujuannya untuk mengetahui respons masyarakat mengenai permasalahan pelecehan seksual dan mengetahui kecenderungan emosi yang sering ditunjukkan pada media sosial khususnya pada tweet baik sentimen positif, negatif, netral maupun emosi takut, jijik, marah, sedih, dan emosi lainnya. Penelitian ini menggunakan data tweet dengan kata kunci berbahasa Indonesia terkait permasalahan pelecehan seksual dengan rentang pengambilan data pada tanggal 4 Juli hingga 10 Juli 2022. Selanjutnya, data diolah melalui tahapan *preprocessing* mulai dari pembersihan data yang redundan, tanda baca, simbol dan karakter. Selain itu, juga melakukan *case folding*, *tokenization*, *stopword*, *stemming*, dan *lemmatization*. Dalam analisis data pelecehan seksual, penelitian ini menerapkan metode *lexicon-based sentiment analysis* yang menghasilkan nilai *compound score* dari sentimen. Akurasi dari model klasifikasi yang dibangun menggunakan metode Naïve Bayes sebesar 74% dengan persentase proporsi sentimen 72% tweet negatif, 18.8% tweet positif, dan 9.2% tweet netral. Hal tersebut menunjukkan bahwa opini negatif terkait isu kekerasan pelecehan seksual masih lebih banyak dibandingkan opini positif maupun opini netral. Hasil analisis emosi menggunakan NRC Emotion Lexicon juga menghasilkan histogram berbagai emosi tweet seperti ketakutan, kemarahan, kewaspadaan, kepercayaan, dan emosi lainnya. Visualisasi Word Cloud juga menampilkan beberapa frekuensi kata yang sering muncul seperti '*leceh*', '*seksual*', '*korban*', '*wanita*', '*perempuan*', '*hukum*', '*tindak*', dan lainnya. Hasil dari penelitian ini diharapkan menjadi kontribusi pengetahuan terutama bagi perempuan agar selalu waspada terkait tindakan pelecehan seksual yang saat ini marak terjadi di Indonesia.

## REFERENSI

- [1] Kementerian Perlindungan Perempuan dan Anak, "SIMFONI PPA," Kementerian Perlindungan Perempuan dan Anak, 2022. [Online]. Available: <https://kekerasan.kemennppa.go.id/ringkasan>. [Accessed 10 Agustus 2022].
- [2] . A. Miranti dan Y. Sudiana, "Pelecehan Seksual Pada Laki-laki Dan Perspektif Masyarakat Terhadap Maskulinitas (Analisis Wacana Kritis Norman Fairclough)," *Bricolage : Jurnal Magister Ilmu Komunikasi*, vol. 7, no. 2, pp. 261-276, 2021.
- [3] I. Masitoh, M. Muhajir, H. Hasbullah, T. Fachmi dan A. Adriadi, "Pencegahan Kekerasan Seksual Remaja Pada Model Pembelajaran Fiqih Kontekstual: Penelitian Di Mts Al-fathaniyah Serang Banten," *Geneologi PAI Jurnal Pendidikan Agama Islam*, vol. 9, no. 1, pp. 517-528, 2022.
- [4] L. . L. Suryani dan H. Setiawan, "Analisis Framing Berita Mengenai Kasus Pelecehan Seksual Pada Media Online Suara.Com Dan Tribun News," *Jurnal Pendidikan Tambusai*, vol. 6, no. 1, pp. 3685-3693, 2022.
- [5] D. Alita, S. Priyanta dan N. Rokhman, "Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter," *Information Systems Engineering and Business Intelligence*, vol. 5, no. 2, pp. 100-109, 2019.
- [6] S. M. Tauhid dan Y. Ruldeviyani, "Sentiment Analysis of Indonesians Response to Influencer in Social Media," dalam *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, 2020.
- [7] A. Al Azkiya, I. P. Vega, M. Iqbal, Z. N. Fatimah dan U. D. Syafitri, "Kata Netizen tentang Kesetaraan Gender dalam Sentimen Warganet Twitter," *Martabat: Jurnal Perempuan dan Anak*, vol. 5, no. 2, pp. 434-458, 2021.
- [8] A. P. Tirtopangarsa dan W. Maharani, "Sentiment Analysis of Depression Detection on Twitter Social Media Users Using the K-Nearest Neighbor Method," dalam *Seminar Nasional Informatika (SEMNASIF) 2021*, Yogyakarta, 2021.
- [9] G. D. Hamidi, . F. A. Bestari dan . A. Situmorang, "Sentiment Analysis on the Ratification of Penghapusan Kekerasan Seksual Bill on Twitter," *JUTISI: Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 3, pp. 655-665, 2021.



- [10] L. Oktasari, Y. H. Chrisnanto dan R. Yuniarti, “Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naïve Bayes Classifier,” dalam *Prosiding Seminar Nasional Sains dan Teknologi ke-7 2016*, Semarang, 2016.