

**Analisis Cuitan Masyarakat Indonesia dalam Media Sosial Twitter terhadap
Permendikbud dan Kekerasan Seksual di Perguruan Tinggi Menggunakan
Pemrosesan Bahasa Alami**

Maria Khelli¹

Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika,
Institut Teknologi Bandung, Jl. Ganesha no. 10 Bandung, Indonesia, 40132

¹13520115@std.stei.itb.ac.id

Abstrak

Sosial media kerap dipakai untuk menganalisis pendapat masyarakat atau konsumen terhadap suatu produk atau kejadian. Media sosial yang sering dipakai adalah Twitter karena Twitter merupakan media sosial berbasis teks. Peraturan penggunaan Twitter pun tidak seketat media sosial lain. Maka dari itu, data dari Twitter cocok untuk menganalisis suatu isu karena cakupan pendapat bisa lebih luas, baik positif maupun negatif. Isu yang akhir-akhir ini ada di Indonesia adalah isu Permendikbud dan kekerasan seksual di perguruan tinggi. Dalam kesempatan ini, penulis ingin menganalisis pendapat masyarakat Indonesia terhadap isu tersebut melalui platform Twitter dengan menggunakan teknik pemrosesan bahasa alami. Didapatkan bahwa sebagian besar masyarakat setuju atas Permendikbud yang baru saja diresmikan ini.

Kata kunci: Twitter, pemrosesan bahasa alami, kekerasan seksual, perguruan tinggi, permendikbud

Pendahuluan

Twitter merupakan salah satu platform media sosial yang sering digunakan untuk mengunggah tulisan-tulisan singkat. Tidak seperti media sosial lain, Twitter lebih berbasis pada tulisan (teks) daripada gambar atau foto seperti Instagram dan Facebook. Karena itu, data dari Twitter cocok untuk dianalisis dengan menggunakan kecerdasan buatan, secara spesifik, pemrosesan bahasa alami.

Selain itu, Twitter juga merupakan platform yang lebih bebas daripada platform lainnya. Misalnya, satu unggahan dengan konteks yang sama bisa saja tidak dapat diunggah di platform lain, tetapi bisa diunggah pada platform Twitter. Kasus yang sering terjadi adalah pengguna Twitter memublikasikan video peristiwa yang seharusnya disensor di platform lain.

Peraturan platform yang kurang membatasi ini menjadikan Twitter sebagai “sasaran” masyarakat untuk berkeluh-kesah secara daring. Dengan demikian, kita bisa menganalisis pandangan atau pendapat masyarakat, khususnya masyarakat Indonesia, tentang sebuah peristiwa terbaru.

Akhir-akhir ini, kasus yang sering dibahas adalah kebijakan Menteri Kebudayaan, Riset, dan Teknologi (Mendikbud Ristek), Nadiem Anwar Makarim, tentang kekerasan seksual di perguruan tinggi. Kemudian, banyak kasus yang sudah terjadi maupun yang baru saja terjadi muncul ke permukaan setelah Permendikbud Ristek Nomor 30 Tahun 2021 ditetapkan secara resmi. Beberapa di antaranya adalah kasus dari Universitas Riau dan Universitas Sriwijaya.

Berdasarkan data dari Komisi Nasional (Komnas) Perempuan via Program Mata Najwa, kekerasan seksual dan diskriminasi berdasarkan jenjang pendidikan terjadi paling banyak pada lingkungan universitas.



Grafik kekerasan seksual dan diskriminasi berdasarkan jenjang pendidikan

Pada Studium Generale keenam, menurut Ketua Komnas Perempuan, Andy Yentriyani mengatakan bahwa relasi kuasa dapat menjadikan kelompok tertentu menjadi lebih rentan mengalami kekerasan. Dalam lingkungan universitas sendiri, ada beberapa relasi kuasa, yaitu antarmahasiswa (senior dan junior) dan antara mahasiswa dan dosen.

Menurut Mendikbud Ristek Nadiem dalam sinar bersama Cinta Laura dan Deddy Corbuzier, relasi kuasa dapat menimbulkan adanya *dominating of power* dari pelaku kepada korban. Dominasi ini diperparah dengan tidak adanya peraturan dan penegakan hukum yang jelas. Bahkan, korban justru kerap dilaporkan balik atas pencemaran nama baik atau, dalam beberapa kasus, atas pelanggaran UU ITE. Dominasi dan ketidakpastian inilah yang dimanfaatkan oleh predator seksual saat “memangsa” korban.

Sebelum adanya Permendikbud Nomor 30 Tahun 2021 tentang Pencegahan dan Penanganan Kekerasan Seksual, korban hanya memiliki sedikit ruang untuk bersuara atau melaporkan. Pada beberapa kasus, laporan memang telah diajukan, tetapi prosesnya terhambat karena kekurangan bukti aduan. Namun, dengan adanya Permendikbud ini, akan ada satuan tugas yang khusus menangani, menyelidiki, dan mendampingi korban dalam kasus kekerasan seksual di perguruan tinggi (Pasal 1).

Di sisi lain, meski tujuan dari Permendikbud adalah baik, terdapat pihak yang kurang setuju terhadap beberapa bagian isi dari Permendikbud ini. Maka dari itu, penulis ingin menganalisis cuitan masyarakat Indonesia dalam platform Twitter menggunakan kecerdasan buatan (pemrosesan bahasa alami) untuk mendapatkan—secara garis besar—pendapat masyarakat mengenai peraturan ini.

Metodologi

Dalam makalah ini, untuk keperluan pendidikan, *disclaimer*: penulis mengambil data dari situs Twitter menggunakan robot secara otomatis—yang sebetulnya melanggar kode etik penggunaan Twitter. Penulis menggunakan bahasa pemrograman Python untuk mengumpulkan dan menganalisis data. Untuk spesifikasi *libraries* yang digunakan, penulis akan merincikannya sebagai berikut.

1. Twitter Web Scrapping: selenium (*automated test software*) dan MS Edge Driver (browser).
2. Exploratory Data Analysis: pandas, NumPy, regex, wordcloud, matplotlib, seaborn, sklearn, collections.

Penambangan Data Cuitan

Pada tahap pengambilan data menggunakan robot, penulis mengekstraksi nama akun, username, waktu unggah, dan isi cuitan dari setiap blok *tweet*. Setiap elemen yang diekstrak diambil dari struktur elemen HTML dari situs Twitter. Beberapa komponennya adalah *div*, *article*, dan *span*.

Blok *tweet* ini diambil setelah kita melakukan kueri dengan kata kunci tertentu. Pada makalah ini, kueri pencarian yang dimasukkan adalah “kampus seksual,” artinya setiap cuitan yang muncul harus mengandung dua kata ini. Kueri pencarian juga tidak harus memiliki arti semantik yang benar.

Untuk memudahkan, penulis menggunakan metode pencarian elemen secara relatif pada suatu elemen HTML, bukan mencari berdasarkan *identifier* (ID) ataupun nama kelas. Hal ini juga sekaligus menghindari ada cuitan yang terlewat karena bisa saja ada nama kelas yang berbeda untuk beberapa blok *tweet*—yang dapat diambil program hanya yang dispesifikasikan.

Setelah mendapatkan data, browser akan ditutup dan data yang telah didapatkan akan dipindahkan pada struktur data bertipe array yang akan dikonversi menjadi tipe DataFrame dari *library* pandas. Kemudian, dokumen akan disimpan dengan ekstensi .csv (*comma-separated-values*).

Pembersihan dan Pemrosesan Data Cuitan

Agar analisis dapat dilakukan lebih menyeluruh, data harus dibersihkan dan diproses terlebih dahulu. Pembersihan data dilakukan dengan membuang kolom-kolom yang tidak diperlukan dan membuang baris-baris yang tidak memiliki makna, misalnya baris data yang tidak memiliki isi cuitan dan baris yang memiliki duplikat. Kemudian, data waktu unggah diproses agar mengikuti waktu lokal. Zona waktu penulis adalah WIB sehingga dilakukan konversi dari UTC ke WIB.

Lalu, pemrosesan dilakukan sebagian besar pada kolom isi cuitan (*tweet*). Pada tiap isi cuitan di kolom ini, dilakukan prosedur berikut.

1. Mengubah karakter baris baru “\n” menjadi spasi.
2. Membuat seluruh karakter menjadi huruf kecil.
3. Membuang tautan, *mention*, angka, dan tanda baca.

Setelah itu, karena isi cuitan banyak kata-kata yang tidak seragam dan sering kali terjadi saltik, penulis melakukan pemetaan. Hal ini dilakukan karena komputer menganggap kata *gw*, *gue*, *gua*, dan *w* berbeda. Padahal, menurut pemahaman manusia, keempat kata itu memiliki makna yang sama.

Jadi, misalkan pada contoh tersebut, jika kalimat, “Gw barusan pulang dari kampus,” akan diubah menjadi “Gue barusan pulang dari kampus.” Kata-kata yang nonbaku tidak terlalu penting di sini karena penulis merasa bahasa nonbaku tidak memberi hambatan saat analisis.

Terakhir, penulis membuang kata-kata yang tidak akan memberikan wawasan apa pun saat analisis, misalnya kata-kata seperti *yang*, *ini*, dan *itu*. Hal ini dilakukan untuk menghilangkan *noise* atau *clutter* pada data sehingga lebih fokus pada kata yang penting saat dianalisis.

Berikut tampilan set data 5 teratas saat ini.

account_name	username	post_date	tweet
hy friends kinda la bcs sibok ya miskah	@cupidnyamrkyck	2021-12-03 19:03:58+00:00	aneh pelaku kekerasan seksual dicoret namanya ...
eri4tul	@erisul	2021-12-03 18:58:39+00:00	kakak unsri unsri tak tau valid pelecehan seks...
Avogadro	@galaktosaaa	2021-12-03 18:45:08+00:00	ditengah banyaknya pelecehan seksual dterlepas...
KPM Unpad	@kpmunpad	2021-12-03 18:34:05+00:00	webinar series pembinaan mental kebangsaan mar...
tiara,	@yaelahbraay	2021-12-03 18:17:10+00:00	lihat berita kampus gue dongkoi ya internalny...

Gambar cuplikan set data hasil pembersihan dan pemrosesan

Data dan Analisis

Pada makalah ini, data yang diambil merupakan data dari 10 November 2021, sehari setelah Mendikbud Ristek Nadiem Makarim berdiskusi dalam Program Mata Najwa, sampai dengan 3 Desember 2021, hari ketika yudisium korban pelecehan seksual Universitas Sriwijaya ditangguhkan.

```
# Tweet terlawas: 10 November 2021
# Bertepatan sehari setelah Nadiem Makarín ada di Mata Najwa
df.post_date.tail(1)

1674 2021-11-10 16:35:41+00:00
Name: post_date, dtype: datetime64[ns, UTC]

# Tweet terbaru: 03 Desember 2021
# Kasus terbaru Universitas Sriwijaya
df.post_date.head(1)

0 2021-12-03 19:03:58+00:00
Name: post_date, dtype: datetime64[ns, UTC]
```

Gambar waktu awal (tail) dan akhir (head)

Total data cuitan yang dikumpulkan berjumlah 1.626 data cuitan (setelah dibersihkan) dengan total karakter sebanyak 272.809 karakter dan total kata 33.173 kata.

Analisis *Word Cloud*

Word cloud (awan kata) merupakan representasi grafik frekuensi kata dalam sebuah dokumen. Interpretasinya adalah semakin banyak frekuensi—semakin banyak disebutkan—berarti semakin besar ukurannya dalam grafik. Dari data yang didapat, grafik awan katanya adalah sebagai berikut.



Grafik awan kata hasil penambangan

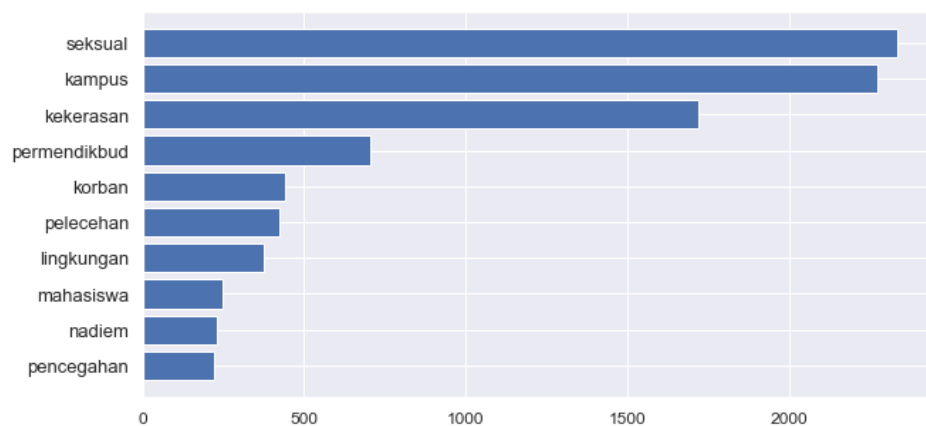
Dari grafik awan kata yang ditampilkan, kita melihat bahwa 5 kata dengan ukuran huruf yang cukup besar adalah *kekerasan seksual*, *seksual kampus*, *pelecehan seksual*, dan *permendikbud*. Untuk frasa *seksual kampus* karena mirip dengan *kampus seksual* yang merupakan kueri pencarian, maka bisa dikatakan wajar muncul sebagai lima teratas.

Namun, frasa lain, seperti *kekerasan seksual*, *pelecehan seksual*, dan *permendikbud*, berbeda dengan kueri pencarian. Komputer juga tidak mengetahui secara eksplisit bahwa ketiga hal tersebut terhubung satu sama lain. Maka dari itu, didapatkan bahwa topik “seksual” di “kampus” (kueri pencarian) berhubungan erat dengan kekerasan seksual, pelecehan seksual, dan permendikbud berdasarkan analisis grafik awan kata.

Analisis N-Gram

N-gram merupakan analisis untuk menunjukkan kata tunggal atau beberapa kontigu yang sering disebut. Dalam keseluruhan cuitan yang didapatkan, berikut data n-gram 10 teratasnya.

1. Unigram

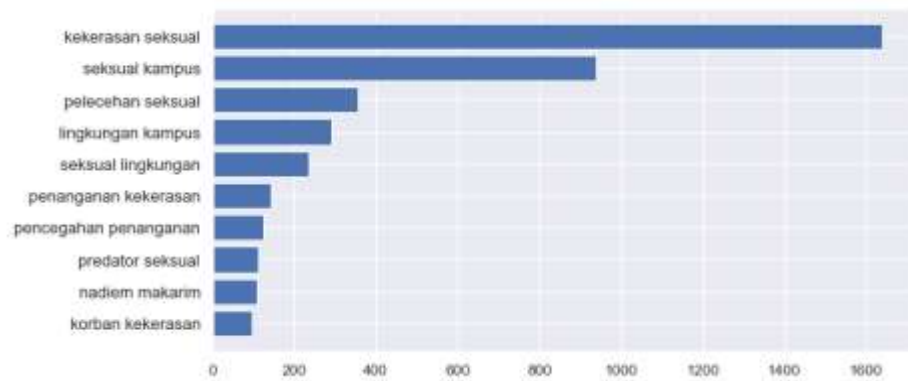


Grafik 10 unigram dengan frekuensi tertinggi

Dari grafik unigram, kita dapat melihat bahwa selain kata seksual dan kampus (kueri pencarian), terdapat kata seperti *permendikbud*, *korban*, dan *pencegahan*. Hal ini bisa jadi karena permendikbud mengambil peran penting dalam pencegahan kekerasan seksual di kampus. Salah satu cuitan tentang ini adalah “Topik utama Permendikbud Ristek 30/2021 adalah

pencegahan kekerasan seksual di lingkungan kampus. Peraturan itu bukan berarti tindakan asusila di luar kampus dibenarkan,” oleh akun bernama Regina Bonafida.

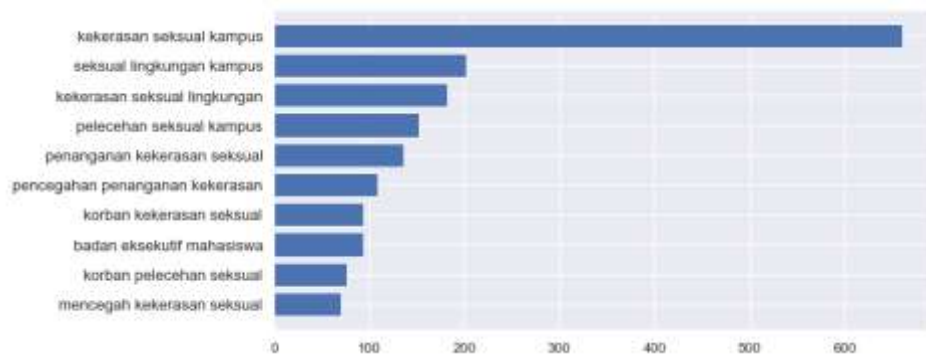
2. Bigram



Grafik 10 bigram dengan frekuensi tertinggi

Data yang ditunjukkan grafik bigram tidak berbeda jauh dengan grafik awan kata yang sudah dianalisis sebelumnya. Hanya saja, yang menarik adalah frasa *kekerasan seksual* disebut lebih dari 1600 kali, frekuensinya lebih tinggi daripada frasa *seksual kampus* yang dikueri.

3. Trigram



Grafik 10 trigram dengan frekuensi tertinggi

Hasilnya berkaitan dengan grafik N-gram sebelumnya, tetapi ada yang menarik, yaitu frasa “badan eksekutif mahasiswa” yang berarti ada peran BEM dalam kejadian kekerasan seksual di kampus. Berdasarkan data yang ada, sekitar 30 cuitan mengenai topik ini berasal dari akun resmi BEM/KM dari sejumlah universitas.

Kesimpulan

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa sebagian besar masyarakat mendukung upaya pencegahan dan perlindungan korban kekerasan seksual. Salah satu upaya yang dimaksud adalah adanya Permendikbud Nomor 30 Tahun 2021. Menurut Ketua Komnas Perempuan, pada kuliah umum keenam, permendikbud ini menambah rujukan regulasi. Dengan demikian, peraturan menjadi lebih jelas sehingga korban kekerasan seksual memiliki pendamping dan pelindung saat melakukan pelaporan atas pelaku. Selain itu, permendikbud ini juga memberikan usaha pemulihan fisik dan psikologis kepada korban.

Di sisi lain, adanya dukungan dan suara dari mahasiswa yang diwakilkan secara resmi melalui BEM/KM dapat menjadikan lingkungan kampus lebih aman dan terbebas dari predator seksual. Sebagai individu, kita sebaiknya mendukung korban dan tidak mengucilkan mereka. Kita juga harus mengubah persepsi-persepsi yang dapat memperburuk keadaan korban.

Daftar Pustaka

- Budi, M. O. (n.d.). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. *ALW3: 3rd Workshop on Abusive Language Online*, 46-57.
- Cotra, A. K. (2020, Juni 2). *Analysis on Tweets Using Python and TWINT*. Retrieved from <https://towardsdatascience.com/>
- Ibrohim, M. O., & Budi, I. (2019). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. *In ALW3: 3rd Workshop on Abusive Language Online*, 46-57.
- Institut Teknologi Bandung. (2021, Oktober 25). *SG KU-4078 : Andy Yentriyani*. Retrieved from https://www.youtube.com/watch?v=b9Kz8Kbe6Ws&ab_channel=InstitutTeknologiBandung
- Komisi Nasional Perempuan. (2021, Maret 5). *Catahu 2021: Catatan Tahunan Kekerasan Terhadap Perempuan Tahun 2020*. Retrieved from <https://komnasperempuan.go.id/uploadedFiles/1466.1614933645.pdf>
- Makarim, N., & Laura, C. (2021, November 16). Nadiem Makarim x Cinta Laura - Deddy Corbuzier Podcast. (D. Corbuzier, Interviewer)