



**ANALISIS SEMANTIK MASYARAKAT INDONESIA
TERHADAP KEBIJAKAN PEMERINTAH
DENGAN MENGGUNAKAN METODE
BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS**

Disusun Oleh:

Tim NamaTimnyaApa

Anggota:

Setyo Nugroho

Cindy Alifia Putri

Nurliah Awaliah

Asal Universitas:

Telkom University



DAFTAR ISI

DAFTAR GAMBAR	2
DAFTAR TABEL	3
1. LATAR BELAKANG	4
1.1 Tujuan dan Manfaat	5
1.2 Batasan Masalah	6
2. METODE	7
2.1 Pre-train BERT	7
2.2 Fine-tuning BERT	8
3. DESAIN DAN IMPLEMENTASI	9
3.1 Gambaran Sistem	9
3.2 Training	10
3.3 Implementasi	13
4. ANALISIS	18
4.1. Kebijakan Pemerintah terhadap Upaya Pemindahan Ibu Kota	18
4.2. Kebijakan Pemerintah terhadap Undang-Undang KPK Baru	20
4.3. Kebijakan Pemerintah terhadap Revisi Undang-Undang KUHP	23
5. KESIMPULAN	26
6. DOKUMENTASI	27
REFERENSI	30

DAFTAR GAMBAR

Gambar 2.1 Langkah-Langkah yang Digunakan pada Metode BERT	7
Gambar 2.2 Langkah Pre-Train	8
Gambar 3.1 Flowchart Training	9
Gambar 3.2 Flowchart Implementasi	10
Gambar 3.3 Tweet berbahasa Inggris	11
Gambar 3.4 <i>Tweet Berbahasa Indonesia</i>	11
Gambar 3.5 Arsitektur Jaringan pada Metode BERT	12
Gambar 3.6 Contoh Hasil Prediksi Data Real	16
Gambar 3.7 Visualisasi dari Presentasi Pro dan Kontra terhadap Kebijakan RUU KPK	16
Gambar 4.1 Visualisasi dari Pro dan Kontra terhadap Kebijakan Pemindahan Ibukota	19
Gambar 4.2 Visualisasi Persentase Pro dan Kontra terhadap Kebijakan UU KPK terbaru menggunakan Dataset dari Youtube	21
Gambar 4.3 Visualisasi Persentase Pro dan Kontra terhadap Kebijakan UU KPK terbaru menggunakan Dataset dari Berita Kompas	22
Gambar 4.4 <i>Visualisasi Persentase Pro dan Kontra terhadap Kebijakan UU KPK terbaru menggunakan Dataset dari Berita CNN</i>	23
Gambar 4.5 <i>Visualisasi Persentase Pro dan Kontra terhadap Kebijakan RUU KUHP</i>	25
Gambar 6.1 Load Model	27
Gambar 6.2 Crawling Dataset Kebijakan Pindah Ibukota	27
Gambar 6.3 Crawling Dataset Kebijakan RUU KPK	28
Gambar 6.4 Contoh Dataset dan Analisis Kebijakan Pindah Ibukota	28
Gambar 6.5 Crawling Dataset Kebijakan RUU KUHP	29

DAFTAR TABEL

Tabel 3.1 Hasil Training	13
Tabel 4.1 Crawling Twitter pada Kebijakan Pemerintah terhadap Upaya Pemindahan Ibu Kota	18
Tabel 4.2 Crawling Twitter pada Kebijakan Pemerintah terhadap Undang-Undang KPK Baru	20
Tabel 4.3 Crawling Twitter pada Kebijakan Pemerintah terhadap Revisi Undang Undang KUHP	24

1. LATAR BELAKANG

Undang-Undang memiliki peran yang sangat penting terhadap Negara Republik Indonesia. Setiap pasal-pasal yang ada di Undang-Undang memiliki fungsi-fungsi yang sangat mendasar guna menjalankan sebuah sistem dari Negara Indonesia. Fungsi-fungsi mendasar tersebut salah satunya adalah sebagai alat kontrol yang tegas dan konkret [1]. Apabila suatu negara tidak memiliki alat kontrol, tentu saja akan banyak peraturan-peraturan dan banyak pula hukum-hukum negara yang dilanggar, Hak Asasi Manusia yang tidak terpenuhi, penyalahgunaan kekuasaan dan akhirnya terjadi perpecahan pada rakyatnya.

Pada saat ini, Warga Negara Republik Indonesia sedang ramai perihal Undang-Undang KPK yang baru saja disahkan dan juga terkait UU KUHP. Banyak terdapat pro dan kontra terkait UU KPK yang baru ataupun UU KUHP tersebut. Bahkan banyak mahasiswa dan masyarakat yang sampai mengadakan demonstrasi dan menyampaikan berbagai aspirasinya baik secara langsung di gedung DPR maupun melalui perantara sosial media. Beberapa dosen dan juga tokoh-tokoh negara pun banyak yang menolak tetapi ada juga beberapa yang menyetujui Undang-Undang KPK yang baru tersebut.

Terdapat beberapa pasal yang dianggap menjadikan Warga Negara Republik Indonesia tidak setuju terhadap pengesahan Undang-Undang KPK, diantaranya adalah dikutip dari tempo.co [2] pada Pasal 21 ayat (4) dan (6) UU KPK yang lama, "pimpinan KPK merupakan penanggung jawab tertinggi yang berwenang menerbitkan surat perintah penyelidikan, penyidikan, penahanan, penuntutan, dan penangkapan". Sedangkan dalam UU baru kewenangan pimpinan sebagai penanggung jawab tertinggi, penyidik, dan penuntut umum dihapus. Di UU yang baru, hampir semua kewenangan pimpinan KPK diambil alih oleh dewan pengawas. Selain itu, terdapat pasal lainnya yang banyak mendapatkan perhatian lebih dari Warga Negara Indonesia yaitu Pasal 47 UU KPK yang baru, kewenangan menggeledah dan menyita harus melalui izin dewan pengawas. Pasal 12B mengatur penyadapan juga harus melalui izin tertulis dewan pengawas. Jangka waktu penyadapan dibatasi hanya selama 1x6 bulan dan dapat diperpanjang 1x6 bulan.

Dari sekian banyaknya masyarakat yang setuju dan menolak pengesahan Undang-Undang KPK yang baru dan beberapa permasalahan lainnya terkait dengan kebijakan pemerintah, dapat dilakukan *sentiment analysis* dengan menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT) agar dapat diketahui semantik dari setiap masyarakat. *Sentiment analysis* adalah suatu proses untuk mengidentifikasi pendapat ataupun emosi yang memiliki makna negatif atau positif dan selanjutnya dapat

dilakukan analisis terhadap hasil identifikasi tersebut [3]. Informasi semantik dari setiap masyarakat dapat digunakan untuk mengetahui berapa besar persentase masyarakat yang pro dan juga persentase kontra. Informasi tersebut, dapat digunakan sebagai pertimbangan terkait pengesahan kebijakan-kebijakan pemerintah selanjutnya supaya tidak terjadi demonstrasi-demonstrasi oleh mahasiswa dan juga mendapatkan banyak persetujuan rakyat.

Perlunya penggunaan metode BERT dikarenakan data yang akan dianalisa berjumlah sangat besar dan akan terus membesar setiap harinya. Akibat data yang terus menerus bertambah, sangat perlu dilakukan pemodelan dengan menggunakan *deep learning* yang dinilai efektif untuk mendapatkan model dari data yang sangat besar, sehingga selanjutnya model tersebut dapat digunakan untuk memproses sentiment analysis masyarakat

Metode BERT adalah model representasi bahasa baru yang menghasilkan model *pre-train* representasi *bidirectional* dari teks yang tidak berlabel dengan bersama-sama mengkondisikan dari kedua konteks di semua *layer*. Model ini memiliki 2 *steps* penting yaitu *pre-train* dan *fine-tuning*. Pada tahap *pre-train*, program dilatih untuk mengetahui tata bahasa, struktur bahasa, dan gaya bahasa dari artikel wikipedia. Sedangkan *fine-tuning* adalah tahap dimana setelah didapatkan model *pre-train*, model tersebut akan ditambah dengan beberapa layer agar didapatkan output sesuai dengan kasus yang ingin dipecahkan. Pada tahap *fine-tuning* hanya dilakukan penambahan layer saja, untuk dilakukan *training* lanjutan dari model *pre-train* sebelumnya.

1.1 Tujuan dan Manfaat

Penelitian yang dilakukan bertujuan untuk membangun *sentiment analysis* terhadap persentase pro dan kontra masyarakat yang dapat digunakan pemerintah dalam mempertimbangkan kebijakan-kebijakan baru yang akan diterapkan. Dengan menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT) tujuannya adalah bahwa hasil analisis yang didapatkan bisa merepresentasikan *sentiment* masyarakat.

Berdasarkan hasil persentase *sentiment* yang didapatkan, akan diketahui jumlah masyarakat yang setuju dan tidak setuju, sehingga diharapkan dapat memberikan wawasan lebih kepada pemerintah dalam memutuskan suatu kebijakan baru untuk Indonesia. Pemerintah sebagai alat pertimbangan pengambilan keputusan terhadap kebijakan yang akan ditetapkan tentu saja membutuhkan opini masyarakat untuk melakukan perbaikan Negara Indonesia menjadi lebih baik lagi.

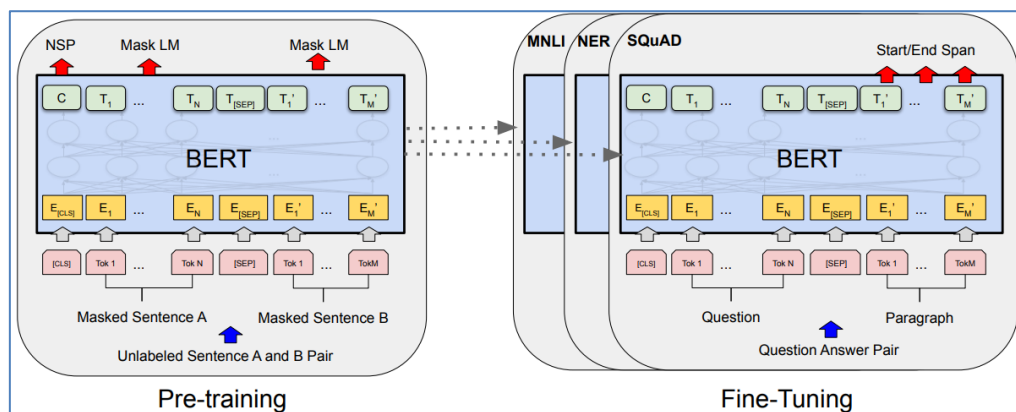
1.2 Batasan Masalah

Dalam makalah ini kami membatasi masalah yang kami angkat sebagai berikut:

1. Diutamakan text berbahasa Indonesia.
2. Karakter maksimal yang diproses adalah 128 karakter.
3. Text bukan merupakan sarkasme.
4. Untuk data dari Twitter, maksimal data satu minggu yang lalu dari waktu menjalankan program yang dapat diproses.

2. METODE

Metode yang dipakai dalam penelitian ini adalah menggunakan metode BERT (*Bidirectional Encoder Transformers for Language Understanding*). Metode ini dipopulerkan oleh Google pada 11 Oktober 2018 tahun lalu dan revisi terakhir dilakukan pada 24 Mei 2019. BERT adalah model representasi bahasa baru yang menghasilkan model *pre-train* representasi *bidirectional* dari teks yang tidak berlabel dengan bersama-sama mengkondisikan dari kedua konteks di semua *layer*. Sehingga, model BERT yang telah *ditrain*, dapat disesuaikan dengan hanya menambahkan satu *layer output* saja.

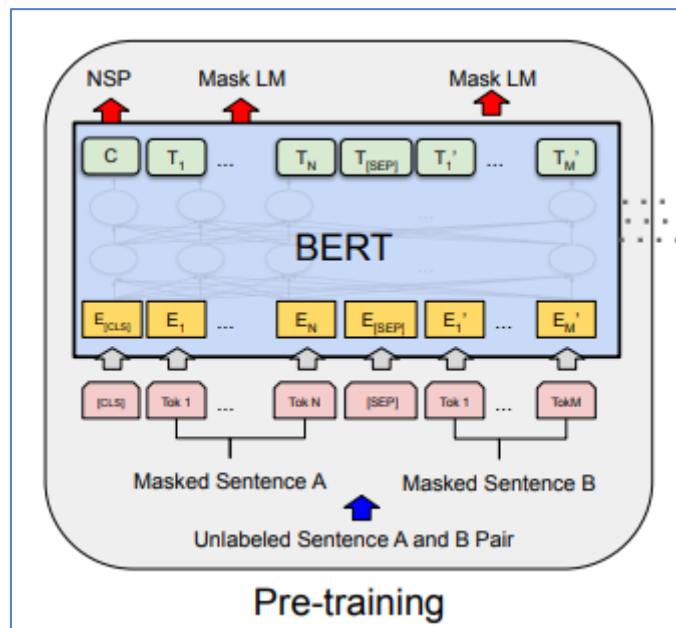


Gambar 2.1 Langkah-Langkah yang Digunakan pada Metode BERT

Seperti yang terlihat pada **Gambar 2.1** yang diambil dari seminar nlp di stanford [4], terdapat 2 *steps* penting yang digunakan dalam metode ini, yaitu *pre-train* dan *fine-tuning*. Dalam step *pre-train*, dilakukan *training* model terhadap data yang tidak berlabel, sedangkan dalam step *fine-tuning* model BERT diinisialisasi dengan parameter *pre-train*, dan semua parameter tersebut disesuaikan dengan menggunakan data yang berlabel. Sehingga dalam implementasinya, metode ini akan mengambil *input* dari dua segmen [5].

2.1 Pre-train BERT

Untuk membuat representasi bidirectional yang baik dan akurat, maka perlu dilakukan 'Masked Language Model' (MLM) yaitu proses dimana beberapa persen dari total token *input* diberi 'mask' atau ditutup secara *random* dan kemudian dicari prediksi terhadap token *input* yang sudah diberi 'mask' tersebut.



Gambar 2.2 Langkah Pre-Train

Dalam **Gambar 2.2** [4] terdapat [CLS] yang merupakan token pertama dari setiap urutan yang dinamakan sebagai *Special Classification Token* ([CLS]) dan token terakhir merepresentasikan tugas klasifikasi. Simbol [SEP] pada gambar merupakan token khusus. Simbol C adalah vektor tersembunyi terakhir dari token [CLS], sedangkan simbol T_i adalah vektor tersembunyi terakhir untuk token *input* ke *i*.

Pada step ini, vektor tersembunyi terakhir yang terkait dengan token 'mask' dimasukkan kedalam *softmax output* melalui *vocabulary*, seperti pada *Language Modelling* yang sederhana. Dalam hal ini, digunakan token 'mask' sebanyak 15% dari seluruh token dan kemudian dicari prediksi terhadap token 'mask' tersebut sehingga didapatkan model *bidirectional pre-train* [6].

2.2 Fine-tuning BERT

Model *pre-train* yang dihasilkan memudahkan BERT untuk memodelkan banyak *task* dengan cukup mengganti *input* dan *output* yang sesuai. BERT menggunakan mekanisme *self-attention* untuk aplikasi yang melibatkan pasangan teks. Pada setiap *task*, *fine-tuning* cukup dilakukan dengan mencocokkan *input* dan *output* spesifik dan menyetel semua parameter secara *end-to-end*. Dalam melakukan *fine-tuning* untuk makalah ini, digunakan 2 *dense layer* untuk melakukan prediksi.

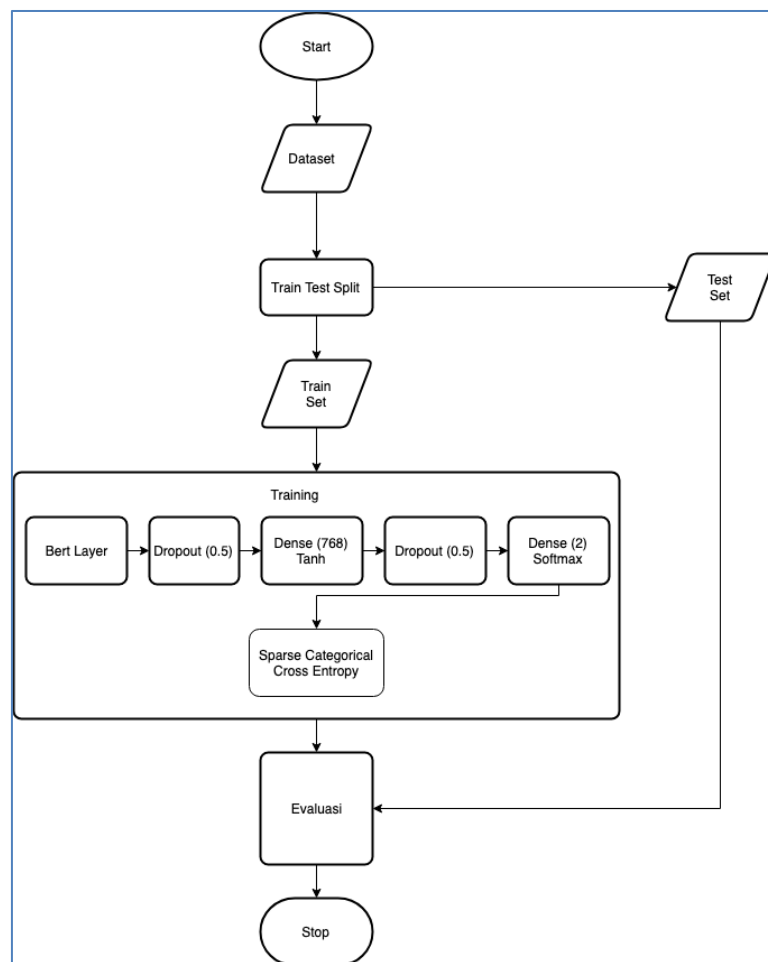
3. DESAIN DAN IMPLEMENTASI

3.1 Gambaran Sistem

Sistem yang kami bangun memiliki dua bagian, yaitu *Training* dan Implementasi.

3.1.1 Training

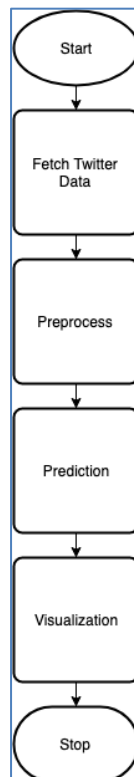
Proses *training* melewati beberapa proses penting yang harus dilakukan seperti, membuka dataset, *split* dataset, *training*, dan evaluasi.



Gambar 3.1 Flowchart Training

3.1.2 Implementasi

Untuk implementasinya sendiri memerlukan beberapa proses seperti, fetch data dari Twitter, *preprocess* data yang diambil, melakukan prediksi, dan visualisasi.



Gambar 3.2 Flowchart Implementasi

3.2 Training

Disini kami melakukan proses *fine-tuning* dengan model BERT yang kami pilih untuk kasus *sentiment analysis*. Kami menggunakan *framework Tensorflow 2.0* dengan *library bert-for-tf2* <https://github.com/kpe/bert-for-tf2> . Kode sumber yang kami gunakan untuk *training* dapat dilihat di <http://bit.ly/namatimnyaapa-training>.

3.2.1 Dataset

Dataset yang kami gunakan berasal dari <http://ridi.staff.ugm.ac.id/2019/03/06/indonesia-sentiment-analysis-dataset/>. Dataset tersebut dikembangkan oleh UGM dan berlisensi CC BY-NC (Common Creative Non Commercial). Dataset tersebut juga sudah melalui proses data *cleaning* sehingga cocok untuk proses *training*. Sebelum *training* dataset

tersebut kami *split* dengan rasio 0.2 untuk *test set* nya. Untuk ukuran dataset tersebut adalah 5472 baris.

3.3.2 Model BERT

Model BERT yang kami gunakan adalah *BERT-Base, Multilingual Cased (multi_cased_L-12_H-768_A-12)*. Model tersebut adalah model yang mendukung *Multilingual* (Multi Bahasa) dan bersumber dari <https://github.com/google-research/bert/blob/master/multilingual.md>. Model *Multilingual* sendiri mendukung 104 bahasa termasuk Bahasa Indonesia dan Bahasa Inggris. Model tersebut kami pilih karena sesuai dengan kasus yang kami angkat dimana *tweet* yang *diposting* dapat berupa Bahasa Inggris dan Indonesia.



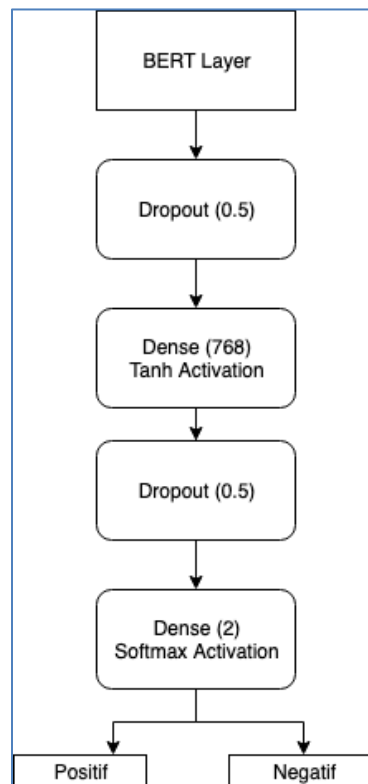
Gambar 3.3 Tweet berbahasa Inggris



Gambar 3.4 Tweet Berbahasa Indonesia

3.2.3 Arsitektur Jaringan

BERT sendiri adalah arsitektur *pre-training* untuk *general language understanding* sehingga tidak dapat langsung digunakan untuk melakukan kasus tertentu, dalam kasus kami adalah *Sentiment Analysis*. Oleh karena itu *output* dari BERT harus diproses lagi dengan menambahkan *layer* baru untuk klasifikasi.



Gambar 3.5 Arsitektur Jaringan pada Metode BERT

Disini kami menambahkan *layer dropout* dengan *rate* 0.5 pada pada keluaran *layer* BERT, lalu kami menambahkan *layer Dense* sebanyak 768 *neuron* dengan *Tanh activation*, disambung *dropout* 0.5 dan *Dense 2 neuron* dengan *activation softmax*. *Dropout rate* 0.5 kami dapatkan setelah mencoba training tanpa dropout dan *dropout rate* yang kecil justru membuat hasilnya menjadi *overfitting*. Dengan *rate* 0.5 kami berhasil mengurangi *overfitting* yang dihasilkan.

Hasil dari arsitektur tersebut adalah probabilitas untuk positif dan negatif. Untuk *loss function* kami menggunakan *Sparse Categorical Cross Entropy*. Kami menggunakan *loss function* tersebut karena *output* yang dihasilkan adalah *categorical numerik* (0 atau 1).

3.2.4 Hyperparameter

Dalam proses *training* kami menggunakan *Hyperparameter* sebagai berikut:

- Model BERT : multi_cased_L-12_H-768_A-12
- Max Sequence Length : 128
- Loss function : Sparse Categorical Cross Entropy
- Optimizer : Adam [7]
- Epoch : 50
- Validation Split : 0.1
- Learning Rate : Kami menggunakan dua strategi *learning rate* yaitu *Linear Warm Up* dengan *learning rate maximum* 1e-5 dan *Polynomial Decay* yang menurunkan *learning rate* dari *learning rate maximum* menuju ke 1e-7. Proses *Linear Warm Up* sendiri dilakukan dengan 20 epoch.

3.2.5 Hasil Training

Dengan penjelasan kami diatas, kami kemudian melakukan *training* dengan *Environment Google Colab* dan menggunakan GPU. Proses *training* memakan waktu selama 1 jam 31 menit 32 detik dan kami mendapatkan hasil *training* dengan akurasi sebagai berikut:

Tabel 3.1 Hasil Training

Data Set	Jumlah Data	Akurasi
Train Set	4383	87%
Test Set	1096	74%

Hasil *weight* dari model yang sudah *ditraining* dapat *download* pada *link* berikut: <http://bit.ly/namatimnyapa-model-weight>.

3.3 Implementasi

Setelah melakukan proses *training* seperti yang dijelaskan diatas, proses selanjutnya adalah implementasi model yang didapatkan untuk data *real* dari *Twitter*. Kode sumber untuk implementasi sendiri dapat dilihat pada <http://bit.ly/namatimnyaapa-implementasi>.

3.3.1 Load Model

Dari hasil *training* diatas, kami mendapatkan *weight* yang disimpan format H5/HDF5. Disini kami membuat kembali model prediksi dan memuat *weight* yang sudah didapatkan dari hasil *training*.

3.3.2 Mengambil data dari Twitter

Disini kami menggunakan library *Tweepy* (<https://www.tweepy.org/>) untuk mengambil data dari Twitter. Data yang kami ambil berjumlah 500 data dengan *filter* yang mengecualikan *retweet*. Sehingga data yang didapat adalah *tweet original*.

3.3.3 Preprocessing

Data yang didapat dari Twitter sangat tidak beraturan dan tidak dapat langsung diprediksi. Oleh karena itu kami melakukan *preprocessing* terlebih dahulu. Kami melakukan *preprocessing* sebagai berikut:

1. HTML Decoding

Hasil dari API masih memiliki kode HTML *special entities* seperti & yang berarti simbol & (dan). Oleh karena itu kita perlu melakukan *decode* untuk mengkonversinya ke dalam bentuk aslinya.

Contoh: "Mahasiswa &mb seluruh masyarakat meminta agar ruu kpk tidak di sahkan"

Menjadi: "Mahasiswa & seluruh masyarakat meminta agar ruu kpk tidak di sahkan"

2. Menghapus @mention

Meskipun terkadang *mention* dapat memiliki informasi, namun informasi tersebut tidak bermanfaat untuk model *sentiment analysis* yang dibangun. Oleh karena itu kita perlu menghapus *mention* tersebut.

Contoh: "@DenResched @kurawa Prof hukum siapa yg diundang utk merumuskan berbagai macam RUU ini ?"

Menjadi: "Prof hukum siapa yg diundang utk merumuskan berbagai macam RUU ini ?"

3. Menghapus Link URL

Seperti *mention*, meskipun URL dapat memiliki informasi namun URL tidak memberikan manfaat untuk *sentiment analysis*.

Contoh: “Dedi Mulyadi: Ada Kekacauan Berpikir dalam RUU KUHP <https://t.co/Ujucgecpnl>”

Menjadi: “Dedi Mulyadi: Ada Kekacauan Berpikir dalam RUU KUHP”

4. Menghapus *Hashtag*

Disini kami hanya menghapus simbol (#) hashtag nya saja karena hashtag memberikan informasi yang berguna untuk *sentiment analysis*.

Contoh: “Mantappp, DPR Provinsi Jawa timur Mau menampung aspirasi mahasiswa dan menolak revisi RUU KUHP dan revisi UU KPK...
👍 #SurabayaMenggugat #MAHASISWABERSATU”

Menjadi: “Mantappp, DPR Provinsi Jawa timur Mau menampung aspirasi mahasiswa dan menolak revisi RUU KUHP dan revisi UU KPK...
👍 SurabayaMenggugat MAHASISWABERSATU”

5. Menghapus semua karakter selain huruf

Kami kemudian menghapus semua karakter selain huruf untuk menghasilkan data yang lebih mudah diproses oleh sistem. Dalam proses ini angka, emoji, dan simbol-simbol dihilangkan.

Contoh: “Mantappp, DPR Provinsi Jawa timur Mau menampung aspirasi mahasiswa dan menolak revisi RUU KUHP dan revisi UU KPK...
👍 SurabayaMenggugat MAHASISWABERSATU”

Menjadi: “Mantappp, DPR Provinsi Jawa timur Mau menampung aspirasi mahasiswa dan menolak revisi RUU KUHP dan revisi UU KPK SurabayaMenggugat MAHASISWABERSATU”

6. Menghapus *whitespace* berlebih

Hasil dari proses diatas terkadang menghasilkan adanya *whitespace* yang berlebih. Diproses ini kita menghilangkan *whitespace* tersebut.

7. Mengambil 128 karakter pertama

Karena kami menggunakan *max sequence length* sebesar 128, model hanya dapat menerima kalimat dengan maksimum karakter sebanyak 128 sehingga kita perlu memotong kalimat tersebut. Disini memotong text nya dari depan atau *head-only* [7].

3.3.4 Prediksi

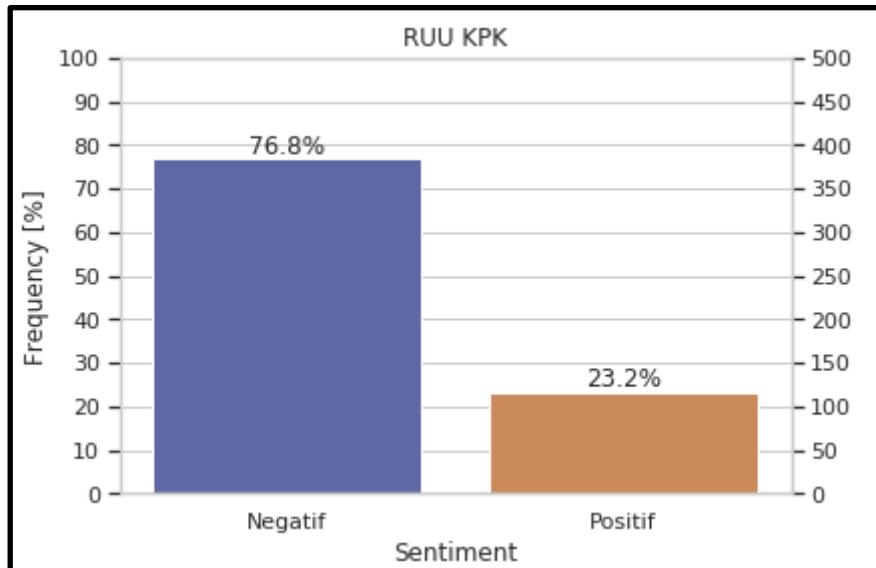
Proses selanjutnya adalah prediksi, disini kami memasukkan data *tweet* yang sudah kami *pre-process* ke dalam model sehingga mendapat probabilitas untuk sentimen negatif atau positifnya.

```
text: bpjs naik subsidi listrik dikurangi kpk dilemahkan rkuhp dan ruu lain dianggap aneh ibukota pindah ngabisin t
hasil: Negatif
text: zery tapi satu hal lagi ketika semua org menghina bang mereka lupa siapa yang mengesahkan ruu kpk tersebut d
hasil: Negatif
text: dapet pesen dari mami lewat telfon kamu boleh ikut demo ikut protes kalo kamu udah baca udah paham semua isi
hasil: Negatif
text: senator abdul aziz tolak ruu kpk
hasil: Positif
text: kriting klo soal ruu udh ditunda s d periode berikutnya dievaluasi ulang yg udh sah ruu kpk tinggal ini so d
hasil: Negatif
text: senator abdul aziz tolak ruu kpk
hasil: Positif
text: lagi pula wajar klo menuntut jokowi mundur udh banyak boongnya terutama soal ruu kpk mana ada lembaga anti re
hasil: Negatif
text: zery hahahahaha pantau terusss winnnn urg juga kurang setuju masalah kpk ini harusnya di diskusikan lagi untuk
hasil: Positif
text: saya mendukung gerakan tolak ruu kpk saya mendukung gerakan tolak ruu kuhp tapi jika melihat kondisi sudah se
hasil: Positif
text: hanya bapak ini kayanya yg punya dendam kesumat sama kpk undang nya diperlemah aj demo nya berhari apalagi di
hasil: Negatif
```

Gambar 3.6 Contoh Hasil Prediksi Data Real

3.3.5 Visualisasi

Kami kemudian melakukan visualisasi pada hasil prediksi. Kami menggunakan *library matplotlib* dan *seaborn* untuk visualisasi. Contoh hasil visualisasi dapat dilihat pada gambar berikut.



Gambar 3.7 Visualisasi dari Presentasi Pro dan Kontra terhadap Kebijakan RUU KPK

Visualisasi yang dihasilkan mencerminkan bahwa tingkat persentase pro dan kontra terkait kebijakan RUU KPK memiliki rentang yang sangat jauh yaitu

1:3 dimana 1 untuk representasi masyarakat yang setuju terhadap UU KPK yang baru.

4. ANALISIS

Analisis dilakukan dengan melakukan perhitungan semantik Masyarakat Indonesia terhadap kebijakan-kebijakan pemerintah yang telah ditetapkan atau bahkan kebijakan yang masih dalam proses untuk diimplementasi.

4.1. Kebijakan Pemerintah terhadap Upaya Pemindahan Ibu Kota

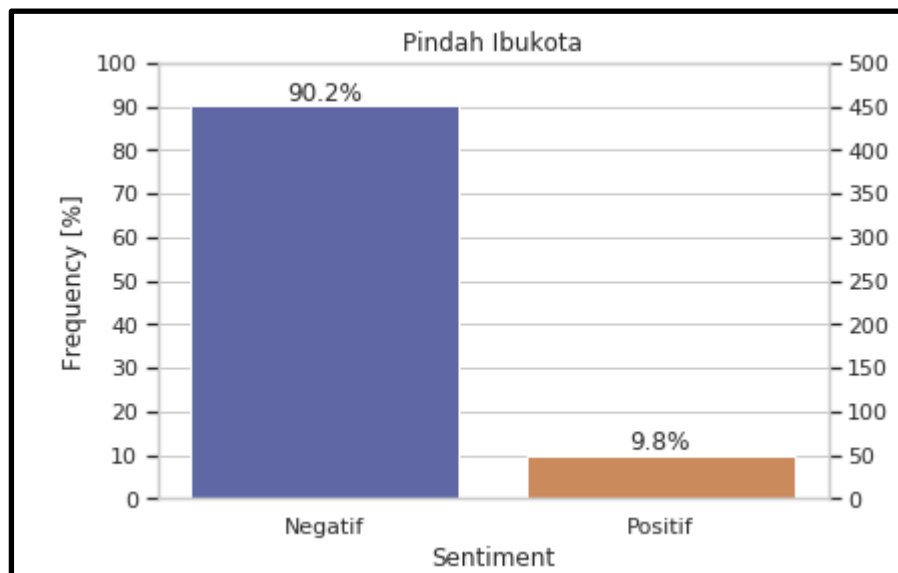
Dalam proses pemindahan ibu kota baru banyak juga terdapat pro dan kontra masyarakat terhadap kebijakan tersebut. Maka, dilakukan analisis terhadap sentimen masyarakat dengan tujuan didapatkan persentase pro dan kontra sehingga didapatkan kesimpulan apakah masyarakat lebih condong positif terhadap kebijakan tersebut atau sebaliknya. Analisis dilakukan terhadap dataset yang merupakan hasil *crawling* dari twitter dengan menggunakan *keyword* 'Pindah Ibukota'. Sehingga didapatkan dataset dan didapatkan hasil analisis terhadap sentimen tiap *tweet* masyarakat seperti berikut ini:

Tabel 4.1 Crawling Twitter pada Kebijakan Pemerintah terhadap Upaya Pemindahan Ibu Kota

No	Komentar	Sentiment
1	didu maksa amat pindah ibukota	Negatif
2	sy sdh tau dari awal klo pindah ibukota bkl butuh danah bsar knp g di gunakan tuk mengirangi ini sj jd negara g di jdikan boneka	Negatif
3	ingin segera cepat ibukota pindah ke kalimantan biar kalau mau demo harus mikir dlu karna nyebrang pulau	Negatif
4	aku malah dr kmrn mikirnya gini pas ada wacana ibukota pindah ke kalimantan itu yg aku takutin cuman kan kalimantan hutannya luar	Negatif
5	kalau bgtu model mahasiswa disumbar kalimantan lebih baik gak usah pindah ibukota bapak presiden terlalu mereka itu sayabersamaj	Negatif
6	untung ibukota negara segera pindah	Negatif

7	buruan pindah deh ibukota ke kalimantan kata anis ow pantes pak de ngotot bgt pengen pindah biar gak didemo jd mahasiwa kudu ngo	Negatif
8	kecuali gua pegawai kemenkumham gua akan bersuara tapi ibukota ttp pindah yaa	Negatif
9	segitu parahkah jakarta sehingga harus pindah ibukota pindah ibukota dgn menjual aset negara kenapa uang itu digunakan utk menst	Negatif
10	pindah ibukota biar santuy korupsinya ya iya	Negatif

Dari dataset tersebut terlihat bahwa kalimat-kalimat yang digunakan adalah murni kalimat yang didapatkan dari hasil *crawling* twitter dan belum memiliki label *class* positif atau negatif. Setelah dilakukan proses sub bab desain dan implementasi agar dataset mentah tersebut dapat diolah, maka didapatkan *sentiment analysis* seperti yang dilihat pada hasil res atau hasil *sentiment* dari dataset tersebut. Apabila dilakukan pengelompokan res negatif dan positif dari tweetan masyarakat, maka hasil persentase yang didapatkan adalah sebagai berikut:



Gambar 4.1 Visualisasi dari Pro dan Kontra terhadap Kebijakan Pemindahan Ibukota

Dari **Gambar 4.1**, terlihat bahwa masyarakat cenderung memiliki sentimen negatif terhadap kebijakan pemerintah untuk mengganti ibu kota

negara. Hal tersebut, dapat terlihat dari persentase yang dihasilkan memiliki *range* yang jauh berbeda yaitu 90.2% dan 9.8% dari total 500 dataset yang digunakan.

4.2. Kebijakan Pemerintah terhadap Undang-Undang KPK Baru

Baru-baru ini, masyarakat sedang dibuat ricuh dengan kebijakan pemerintah dalam mengganti Undang-Undang KPK. Beberapa dosen bahkan mengizinkan mahasiswanya untuk ikut melakukan demonstrasi dan meliburkan kelasnya pada tanggal 24 September 2019. Aksi demonstrasi menuntut pencabutan dan penolakan terkait dengan Undang-Undang KPK yang baru. Oleh karena itu, dibuatlah analisis terhadap sentimen masyarakat dalam upaya mengetahui persentase pro dan kontra masyarakat terkait dengan kebijakan tersebut. Dataset diambil dalam dua sisi, yaitu pertama dari twitter dan yang kedua adalah dari youtube 'mata najwa', komentar berita CNN, dan komentar berita kompas.com.

Pada dataset yang pertama yaitu *crawling twitter* dengan kata kunci 'RUU KUHP' dan didapatkan 500 komentar masyarakat. Setelah dilakukan tahap pemrosesan agar dataset yang didapatkan dapat diolah, maka dilakukan pemrosesan seperti pada sub bab desain dan implementasi, sehingga analisis yang didapatkan adalah sebagai berikut :

Tabel 4.2 Crawling Twitter pada Kebijakan Pemerintah terhadap Undang-Undang KPK Baru

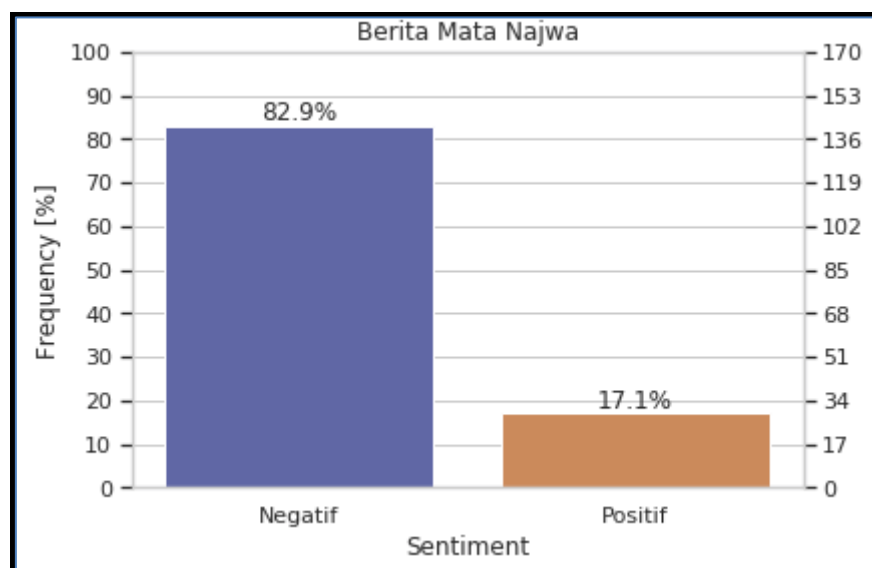
No	Komentar	Sentiment
1	bentrokan antara mahasiswa yang melakukan aksi menolak ruu kuhp dan pengesahan uu kpk dengan polisi membuat kawasan senayan jaka	Negatif
2	revocar la ley de kpk y la ley de recurcos naturales pembatalan uu kpk dan uu sda aprobar la ley de violencia sexual y la ley de	Positif
3	ruu kpk sudh disahkan dan ruu khup sudh ditunda oleh presiden terus mahasiswa yang turun kejalan dalam rangka apa jika kalian ka	Positif

4	bpjs naik subsidi listrik dikurangi kpk dilemahkan rkuhp dan ruu lain dianggap aneh ibukota pindah ngabisin uang dll gak milih o	Negatif
---	--	---------

Terdapat 500 *row* atau data komentar masyarakat terkait dengan diubahnya Undang-Undang KPK tersebut. Dari 4 data pertama yang diambil, dapat diketahui bahwa rata-rata masyarakat memiliki sentimen yang imbang antara negatif dan positif. Dari **Gambar 3.7** didapatkan kesimpulan bahwa tidak semua masyarakat kontra terhadap kebijakan tersebut, akan tetapi persentase pro dan kontra masyarakat masih sangat jauh yaitu 3:1, perbandingan 3 untuk masyarakat yang memiliki sentimen negatif.

Pada dataset yang kedua yaitu dataset yang didapatkan dengan cara mengambil komentar masyarakat pada kolom komentar youtube 'KPK: Kiamat Pemberantasan Korupsi - Laode Syarief: Ini *Darkest* Moment KPK (Part 1) | Mata Najwa' dan juga dari artikel berita kompas.com dan CNN. Total berita yang didapatkan adalah sebanyak 170 *row* komentar.

Terdapat 170 *row* komentar masyarakat terkait dengan diubahnya Undang-Undang KPK tersebut. Dari 10 data pertama yang diambil, dapat diketahui bahwa semuanya memiliki sentimen negatif. Dari total data didapatkan persentase negatif dan positif adalah sebagai berikut:

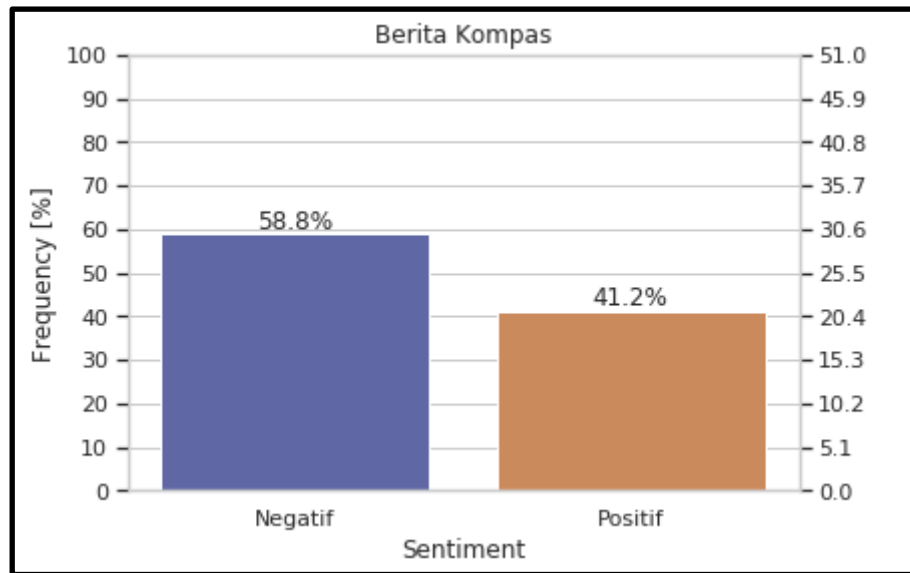


Gambar 4.2 Visualisasi Persentase Pro dan Kontra terhadap Kebijakan UU KPK terbaru menggunakan Dataset dari Youtube

Dari **Gambar 4.2** didapatkan kesimpulan bahwa masyarakat yang kontra terhadap kebijakan pemerintah dari dataset youtube 'mata najwa' adalah

sebesar 82.9% sedangkan persentase masyarakat yang pro adalah sebesar 17.1%.

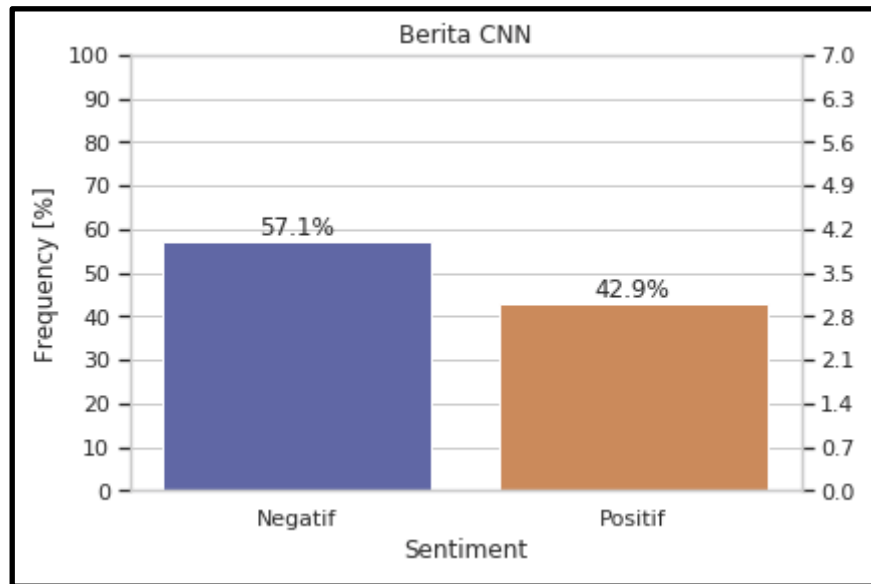
Selanjutnya adalah dataset yang didapatkan dari crawling komentar pada berita Kompas.com dan didapatkan sebanyak 51 komentar, berikut adalah visualisasi persentase pro dan kontra yang didapatkan:



Gambar 4.3 Visualisasi Persentase Pro dan Kontra terhadap Kebijakan UU KPK terbaru menggunakan Dataset dari Berita Kompas

Dari **Gambar 4.3** didapatkan kesimpulan bahwa masyarakat yang kontra terhadap kebijakan pemerintah dari komentar-komentar pada berita Kompas.com adalah sebesar 58.8% sedangkan persentase masyarakat yang pro adalah sebesar 41.2%.

Dataset terakhir untuk kasus *sentiment analysis* terhadap UU KPK adalah dataset yang didapatkan dari *crawling* komentar pada berita Kompas.com dan didapatkan sebanyak 7 komentar, berikut adalah visualisasi persentase pro dan kontra yang didapatkan:



Gambar 4.4 Visualisasi Persentase Pro dan Kontra terhadap Kebijakan UU KPK terbaru menggunakan Dataset dari Berita CNN

Dari **Gambar 4.4** didapatkan kesimpulan bahwa masyarakat yang kontra terhadap kebijakan pemerintah dari komentar-komentar pada berita CNN adalah sebesar 57.1% sedangkan persentase masyarakat yang pro adalah sebesar 42.9%.

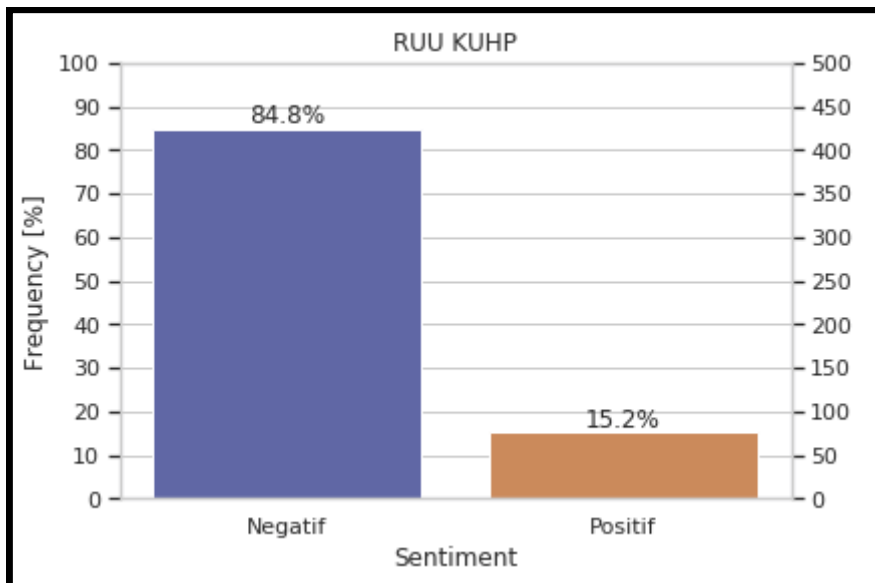
4.3. Kebijakan Pemerintah terhadap Revisi Undang-Undang KUHP

Pengesahan RUU KUHP ditunda, hal tersebut dikarenakan banyak sekali aksi-aksi demonstrasi kepada DPR RI yang meminta agar pengesahan terhadap RUU KUHP ditunda. Banyak mahasiswa-mahasiswa yang menuntut demonstrasi ke gedung-gedung DPR pada tanggal 24 September 2019. Karena kekacauan yang diakibatkan sampai begitu besar dan akhirnya memaksa DPR untuk menunda pengesahan RUU KUHP. Untuk melihat seberapa besar masyarakat menolak adanya RUU KUHP, maka dilakukan analisis sentimen masyarakat terhadap adanya Revisi Undang-Undang KUHP. Dataset yang diambil adalah *crawling* twitter dengan menggunakan 500 komentar masyarakat dari *tweet*-an twitter. Dataset harus diolah terlebih dahulu agar dapat dilakukan analisis terhadap sentimennya. Analisis dilakukan seperti pada langkah-langkah sebelumnya, setelah dilakukan tahap *preprocessing* dan analisis, maka didapatkan *sentiment* untuk setiap komennya seperti tabel berikut ini:

Tabel 4.3 Crawling Twitter pada Kebijakan Pemerintah terhadap Revisi Undang Undang KUHP

No.	Komentar	Sentiment
1	para pelajar tersebut mengatakan tuntutan nya agar ruu kuhp tidak perlu dilakukan oleh anggota dpr emak emak cari anaknya yang de	Negatif
2	agar apa ruu kuhp di sahkan agar indonesia makin lemah di mata negara lain entah apa yang merasuki dpr mensahkan ruu itu faedahn	Negatif
3	tetep aja kalo ngata ngatain dpr kena juga mamah wkwwkwk harusnya orang orang yang bikin ruu kuhp goblog biar ga durhaka wkwwkwk	Negatif
4	dulu rebutan sekarang	Negatif
5	hotman paris sebut ruu kuhp aneh	Negatif
6	cek polisi terluka saat amankan demo mahasiswa tolak ruu kuhp mereka luka karena	Negatif
7	cek polisi terluka saat amankan demo mahasiswa tolak ruu kuhp mereka luka karena	Negatif
8	terpilihlah wacana ruu kuhp dkk itu hasile rakyat panik akhirnya ketidakstabilan negara terjadi akibatnya saham turun rupiah ju	Negatif
9	koko yazid di tv dan media pers lainnya mahasiswa lantang mengatakan gerakan mereka itu bertujuan menolak penetapan ruu kpk dan	Negatif

Setelah didapatkan sentimen untuk setiap komentar masyarakat, ternyata diketahui pada 9 tweet pertama, seluruhnya memiliki *sentiment* yang negatif. Apabila dilakukan perhitungan persentase untuk 500 orang, maka didapatkanlah **Gambar 4.2** berikut ini:



Gambar 4.5 Visualisasi Persentase Pro dan Kontra terhadap Kebijakan RUU KUHP

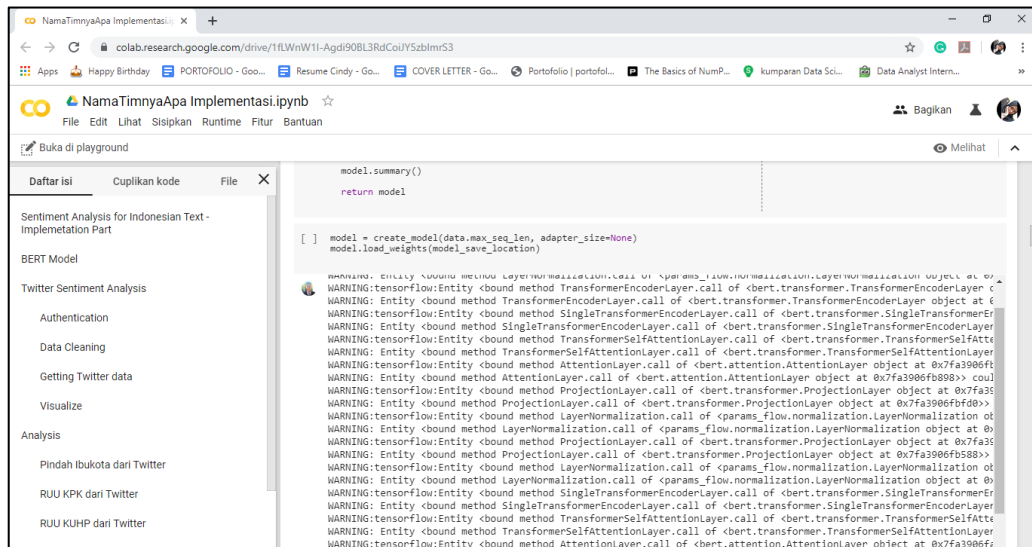
Dari **Gambar 4.5** didapatkan kesimpulan bahwa memang sebagian besar Masyarakat Indonesia menentang adanya Revisi terhadap Undang-Undang KUHP karena dari visualisasi **Gambar 4.2** menunjukkan bahwa 84.8% dari 500 masyarakat menyatakan kontra terhadap kebijakan pemerintah dalam upaya RUU KUHP.

5. KESIMPULAN

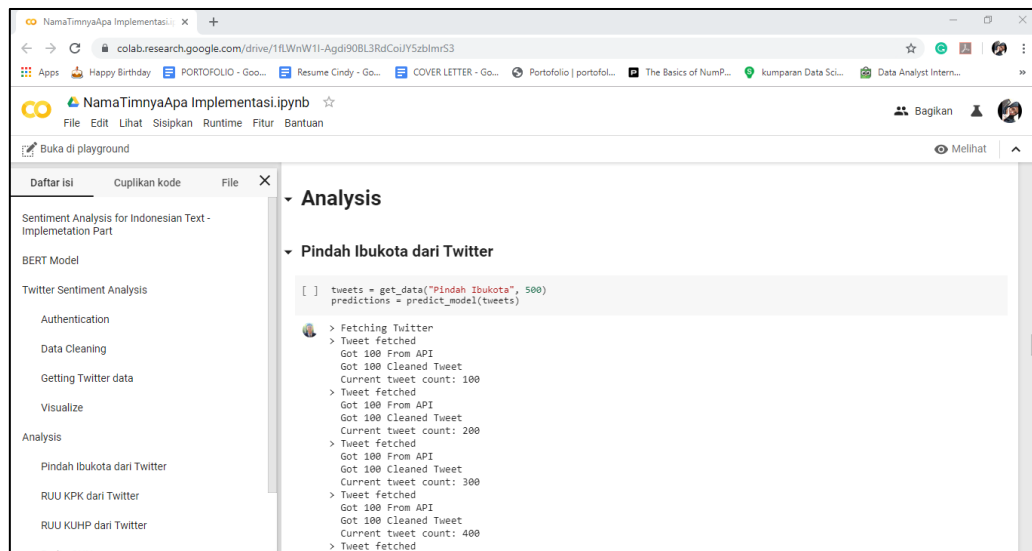
Pembelajaran dengan konteks bahasa atau yang sering disebut dengan *Natural Language Processing* dengan metode *Bidirectional Encoder Representations from Transformers* terbukti dapat diimplementasikan untuk pengaplikasian *sentiment analysis* baik menggunakan dataset dari twitter ataupun menggunakan berbagai dataset seperti komentar-komentar berita ataupun youtube. Hal tersebut, dapat dilihat dari *sentiment analysis* pada setiap *row* (data) komentar telah memberikan *output* yang terklasifikasi dengan benar. Kelemahan menggunakan metode ini adalah untuk komentar yang bersifat sarkasme masih tidak dapat terklasifikasi dengan benar.

Dari hasil *sentiment analysis* yang sudah dilakukan, manfaatnya adalah agar dapat dilakukan analisis dan pertimbangan bagi pemerintah untuk menetapkan suatu kebijakan baru agar mendapat wawasan lebih tentang persetujuan rakyat terhadap kebijakan-kebijakan baru tersebut.

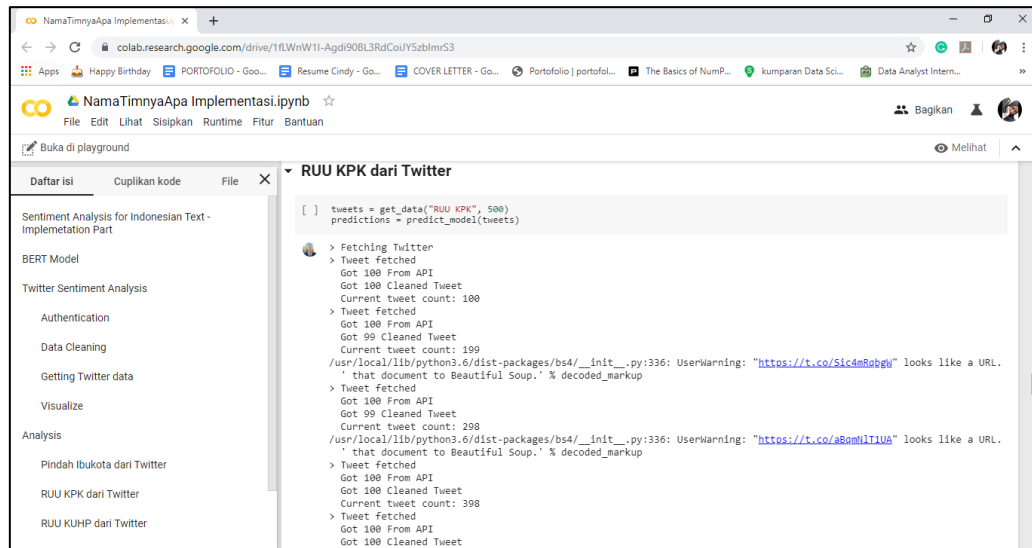
6. DOKUMENTASI



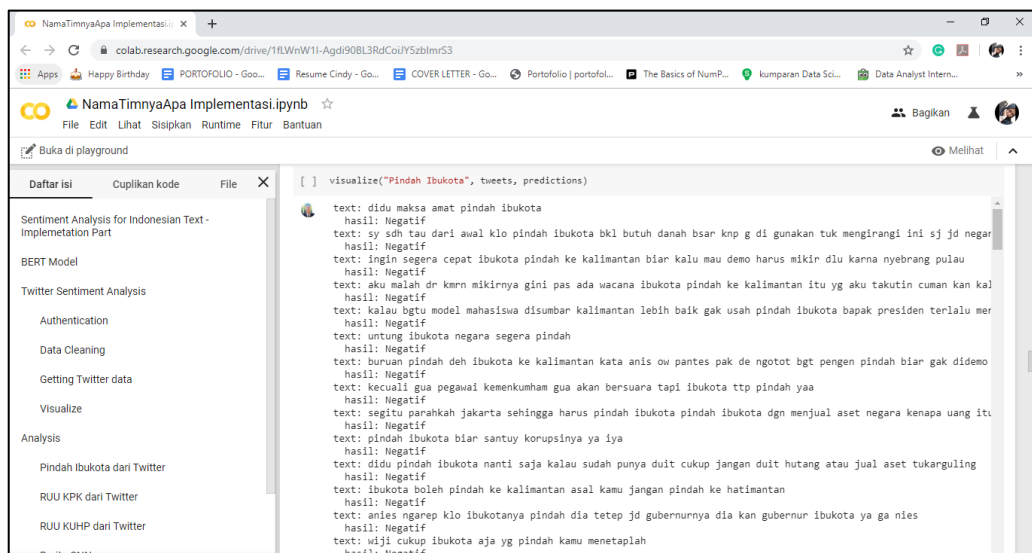
Gambar 6.1 Load Model



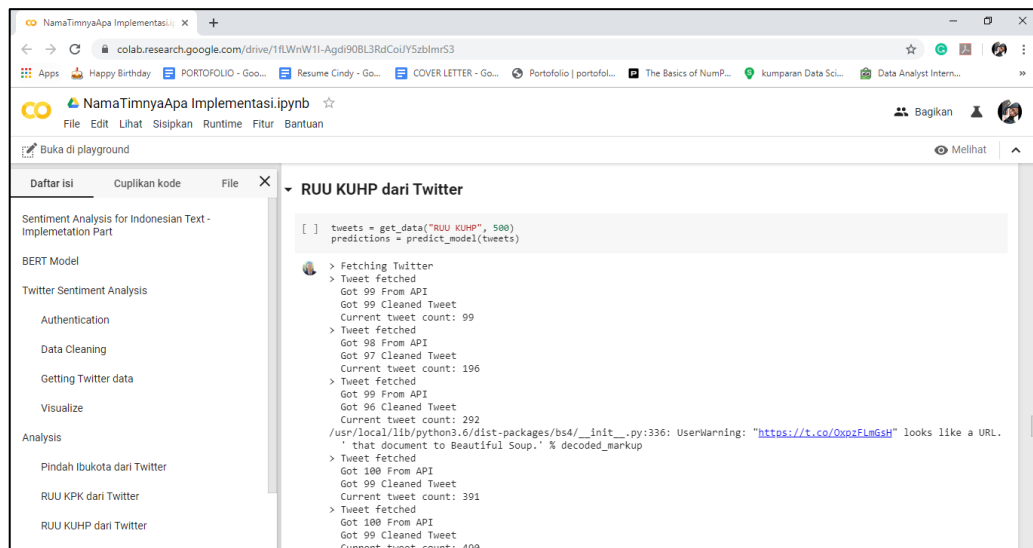
Gambar 6.2 Crawling Dataset Kebijakan Pindah Ibukota



Gambar 6.3 Crawling Dataset Kebijakan RUU KPK



Gambar 6.4 Contoh Dataset dan Analisis Kebijakan Pindah Ibukota



Gambar 6.5 Crawling Dataset Kebijakan RUU KUHP

REFERENSI

- [1] J. Simamora, "Tafsir Makna Negara Hukum Dalam Perspektif Undang-Undang Dasar Negara Republik Indonesia Tahun 1945," *J. Din. Huk.*, vol. 14, no. 3, pp. 547–561, 2014.
- [2] "kpk-sebut-penetapan-tersangka-imam-nahrawi-tak-bermotif-politik @nasional.tempo.co." .
- [3] T. I. Jain and D. Nemade, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Int. J. Comput. Appl.*, vol. 7, no. 5, pp. 12–21, 2010.
- [4] J. Devlin, "BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding Pre-training in NLP • Word embeddings are the basis of deep learning for NLP," *Bert-Ppt*.
- [5] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019.
- [6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," no. Mlm, 2018.
- [7] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to Fine-Tune BERT for Text Classification?," no. 3, 2019.