

Laporan Akhir Tugas Akhir Analitika Media Sosial

“What Netizens Think About Donald Trump: Sentiment Analysis”

1. MOTIVASI

1.1. Deskripsi Pekerjaan

Peneliti melakukan penelitian *sentiment analysis* atas opini *netizen* terhadap Donald Trump secara keseluruhan. Penelitian dilakukan dengan meneliti data Twitter sebagai *dataset*. Kami menggunakan *tweepy* untuk melakukan pengumpulan data yang akan digunakan sebagai *dataset*. Selain itu, kami menggunakan Gephi sebagai *tools* untuk mendapatkan analisis banyak kata yang muncul yang berhubungan dengan Trump. Selanjutnya, untuk *sentiment analysis*, kami gunakan Natural Language Toolkit dari Python.

1.2. Latar Belakang

Alasan mengapa tim peneliti memilih opini *netizen* terhadap Donald Trump adalah karena betapa kontroversial keputusan-keputusannya dan dirinya sendiri yang menyebabkan banyaknya pro dan kontra yang terjadi di khalayak umum, sehingga kami rasa akan sangat menarik untuk diteliti. Oleh karena itu, kami selaku tim peneliti tertarik untuk mengetahui tentang opini khalayak umum terhadap Donald Trump di sosial media, untuk penelitian ini kami memilih sosial media Twitter.

1.3. Manfaat

Manfaat dari penelitian ini adalah untuk mengetahui pendapat khalayak umum terhadap sosok Donald Trump. Pendapat khalayak yang ingin diketahui bukan hanya dari pendapat negatif, tetapi juga pendapat positif terhadap Donald Trump itu sendiri. Selain itu, penelitian ini juga memiliki manfaat untuk mengetahui kejadian apa yang sedang terjadi yang berkaitan Donald Trump pada saat *dataset* sedang diambil.

2. METODOLOGI

2.1. Pengumpulan Data

Data dikumpulkan dengan *crawling* tweets menggunakan Tweepy dengan kata kunci ‘Trump’ dan ‘Donald Trump’. *Crawling* tweets dilakukan pada tanggal

10 Mei 2017 dari jam 22.00 WIB sampai 24.00 WIB. Dari hasil *crawling* tersebut, didapatkan *dataset* sebanyak kurang lebih 70.000 tweets. Hasil *crawling* dikumpulkan dalam bentuk JSON yang kemudian difilter menjadi file dengan ekstensi *.tweets* dan *.txt*. File *.tweets* ini merupakan file yang nantinya menjadi dasar untuk pembuatan file *.csv* untuk analisis graf dengan Gephi.

2.2. Preprocessing

Preprocessing dilakukan untuk menghilangkan atau setidaknya mengurangi *noise* dari *dataset* kami. Ini dilakukan agar hasil analisis yang kami lakukan selanjutnya lebih efektif dan efisien.

- Proses pertama yaitu proses penghapusan *stop words*. Kami menggunakan *dictionary stop words* yang kami dapatkan dari Scele dan GitHub (<https://github.com/stanfordnlp/CoreNLP/blob/master/data/edu/stanford/nlp/patterns/surface/stopwords.txt>). Kami lalu menggabungkan kedua *dictionary* tersebut untuk hasil *filtering stop words* yang lebih maksimal.
- Selanjutnya dilakukan Lemmatization untuk mengubah suatu kata menjadi kata dasarnya.
- Setelah itu, kami melakukan proses *POS tagging* untuk mendapatkan label atau jenis dari setiap kata yang muncul dalam tweets. Kata-kata yang telah memiliki tag akan digunakan pada saat *sentiment analysis*.
- Tahap preprocessing selanjutnya yaitu, kami mengambil kata-kata yang telah di *tag* dengan menggunakan *POS tagger*. Kata yang diambil yakni kata-kata dengan *tag adjective* (JJ).
- Selanjutnya, kami menggunakan *dictionary* tentang kata-kata yang tergolong ke dalam kata positif dan kata-kata yang tergolong ke dalam kata negatif.

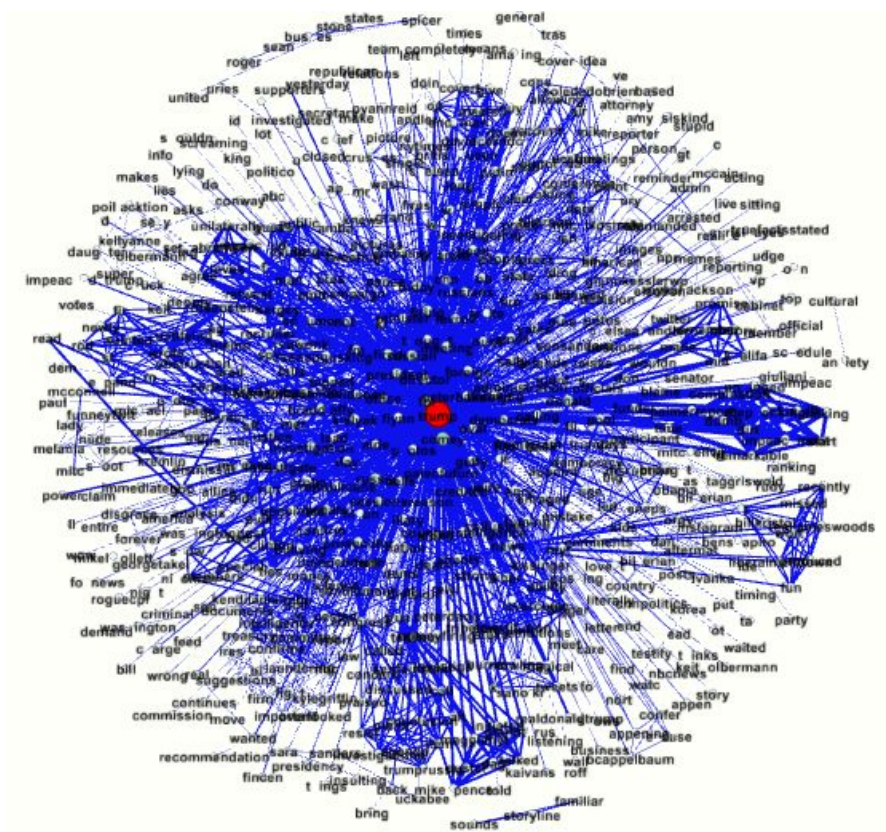
2.3. Pekerjaan utama yang dilakukan

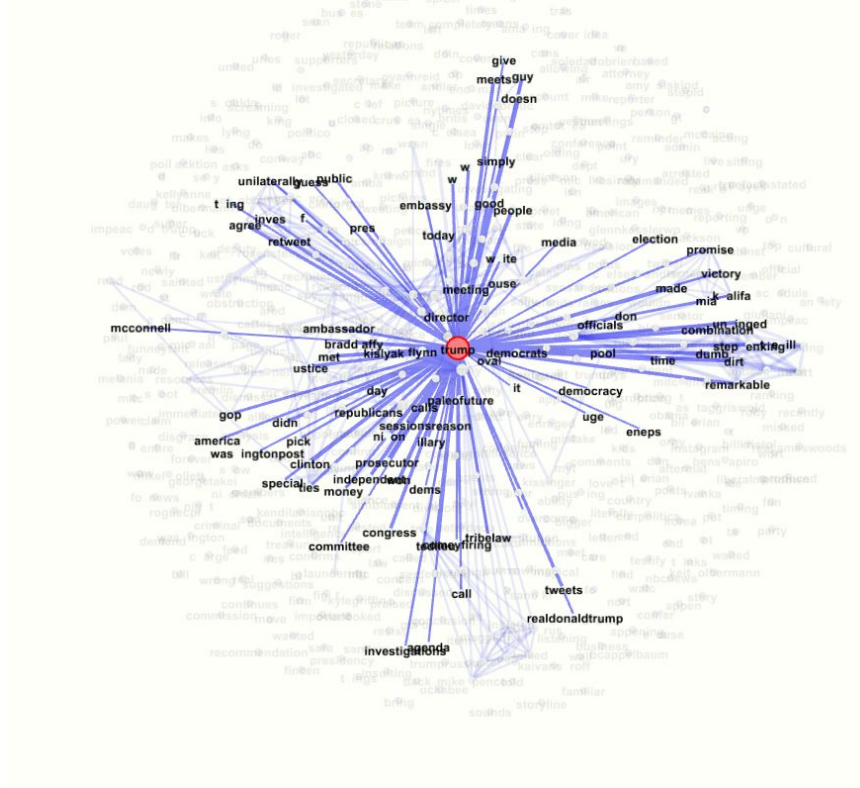
Kami akan melakukan *sentiment analysis* terhadap tweets yang telah kami dapatkan di dalam *dataset*. Nantinya, akan ditentukan apakah suatu tweet mempunyai nilai positif atau negatif. Sebuah tweet memiliki nilai positif apabila dia memiliki kata positif atau dalam arti mendukung Donald Trump. Sebaliknya, sebuah tweet memiliki nilai negatif apabila tweet tersebut mengandung kata yang negatif atau menjatuhkan Donald Trump.

Setelah dilakukan *sentiment analysis*, hasil analisis tersebut akan divisualisasikan dalam bentuk grafik. Hal ini dilakukan agar pembaca hasil analisis menjadi lebih mudah.

3. HASIL ANALISIS

Tim peneliti telah mendapatkan *dataset* berupa *tweet* mengenai Donald Trump yang diperoleh dari Twitter. *Dataset* yang didapatkan berupa *tweet* khalayak umum tentang Donald Trump dalam Bahasa Inggris. Berikut merupakan hasil visualisasi grafik dari *dataset* yang telah dikumpulkan.





3.1. Peristiwa yang terjadi ketika proses pengambilan *dataset*

Berikut merupakan 10 kata yang paling sering muncul di dalam dataset beserta banyak kemunculannya:

'comey' (141), 'russia' (89), 'russian' (81), 'meeting' (55), 'fbi' (54), 'investigation' (54), 'fired' (52), 'president' (47), 'firing' (47), dan 'today' (44).

Dari 10 kata yang paling sering muncul tersebut, ternyata dapat dilihat bahwa saat pengambilan data dari Twitter, banyak *netizen* yang *men-tweet* kata terkait dengan **Rusia, FBI, pertemuan, investigasi, presiden, dan pemecatan**. Dari kata-kata tersebut, dapat diambil salah satu kesimpulan bahwa adanya **pertemuan antara presiden dan Rusia yang menyebabkan adanya investigasi dan pemecatan**. Ternyata, memang benar pada saat data diambil, terjadi **pemecatan Kepala Direktur FBI** yaitu **James Comey** setelah FBI menginvestigasi adanya hubungan/pertemuan antara **Presiden Donald Trump dengan Rusia**.

Selanjutnya untuk perkembangan dari sentiment analysis, kode untuk *stop-words removal* dan *lemmatization* telah selesai. Untuk perkembangan kode *POS Tagging* telah berjalan 50% dari keseluruhan. Kendala yang dihadapi adalah *syntax* Python dalam *array comprehension*.

3.2. *Sentiment Analysis* tentang sifat Donald Trump dengan *tweet* yang dilakukan oleh *netizen* melalui Twitter

Selanjutnya tim peneliti ingin mengetahui *sentiment analysis* dari pendapat *netizen* terhadap Donald Trump. Pada tahap ini tim peneliti melakukan pengumpulan data sifat positif maupun negatif Donald Trump dari pandangan *netizen*. Selanjutnya data tersebut diurutkan berdasarkan jumlah kemunculan terbanyak dalam *tweet* yang dilakukan oleh *netizen* untuk mengetahui lima sifat positif serta lima sifat negatif Donald Trump menurut pendapat *netizen*.

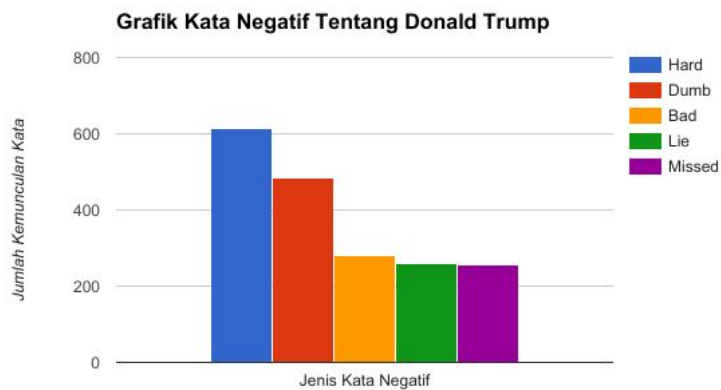
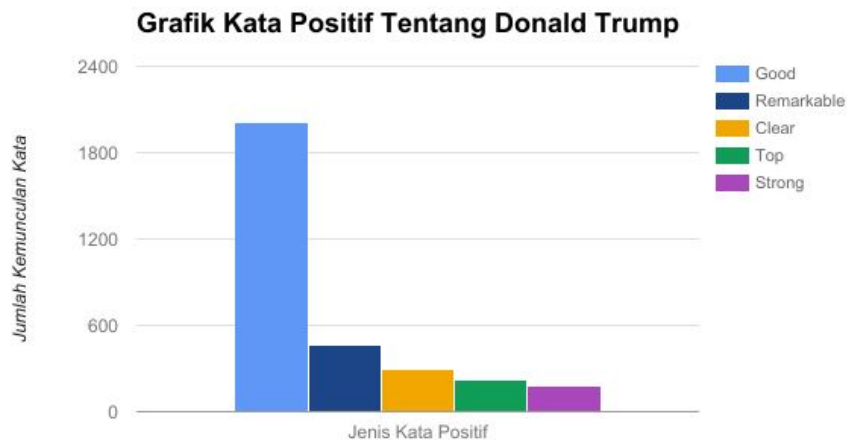
Dengan menggunakan *dataset* yang ada dilakukan *POS tagging* untuk mendapatkan kata sifat dari *tweet* yang dilakukan oleh *netizen*. Selanjutnya dari hasil *POS tagging* tersebut diambil seluruh kata sifat yang ada dari *POS tags*. Kata sifat tersebut dikumpulkan dalam sebuah *array* yang sudah disediakan.

Sebenarnya kami ingin melakukan *dependency parsing* juga untuk dapat mengerjakan *aspect-based sentiment analysis* namun dikarenakan setelah dicari di Internet, ternyata sulit untuk dilakukan dan kami belum terlalu mengerti bagaimana cara memakainya. Oleh karena itu, kami belum dapat melakukan baik *dependency parsing* maupun *aspect-based sentiment analysis*.

Kemudian akan dilakukan analisis terhadap *array* yang berisi kata sifat tersebut. Analisis ini dilakukan untuk mengetahui apakah elemen kata sifat dalam *array* merupakan kata sifat positif atau kata sifat negatif. Proses analisis dilakukan dengan membandingkan elemen dalam *array* dengan *negative-word.txt* dan *positive-word.txt* (*file* terlampir). Elemen dalam *array* merupakan kata sifat positif apabila kata tersebut sama dengan kata sifat yang terdapat dalam *positive-word.txt*, begitu juga sebaliknya. Kemudian kata sifat positif maupun negatif tersebut diurutkan berdasarkan banyaknya kemunculan dalam *tweet netizen*.

Setelah program dijalankan, ternyata ditemukan kata positif sebanyak 6591, kata negatif sebanyak 5757 kata, dan kata tidak terkategori 34086 kata. Setelah diteliti lebih lanjut ternyata kata-kata tidak terkategori tersebut kebanyakan bukan kata sifat sehingga tidak dapat dikategorisasi menjadi kata sifat positif atau negatif.

Dari hasil analisis tersebut didapatkan lima teratas sifat positif serta lima teratas sifat negatif Donald Trump melalui *tweet* yang dilakukan oleh *netizen*. Lima kata positif tentang Donald Trump adalah ‘good’(2013), ‘remarkable’ (468), ‘clear’ (293), ‘top’ (222), dan ‘strong’ (181). Selanjutnya lima kata negatif tentang Donald Trump adalah ‘hard’ (613), ‘dumb’ (484), ‘bad’ (281), ‘lie’ (258), dan ‘missed’ (256). Berikut ini visualisasi terkait kata positif dan negatif untuk Donald Trump:



4. KESIMPULAN

Setelah dilakukan percobaan, kami mendapatkan dua kesimpulan:

1. Banyak kata di dalam tweets yang berhubungan dengan Donald Trump merupakan kata-kata positif. Ini dapat disimpulkan bahwa banyak tweet yang mengandung sentimen positif untuk Trump.
2. Pada saat dataset diambil, event yang sedang marak dibicarakan adalah tentang pemecatan Direktur FBI yaitu James Comey oleh Donald Trump dikarenakan menginvestigasi keterkaitannya dengan pihak Rusia.
3. POS Tagger yang telah ada masih kurang bagus karena masih banyaknya kata yang non-sifat masuk ke dalam kategori *adjective*.