

IMPLEMENTASI *TEXT MINING* PADA MESIN PENCARIAN *TWITTER* UNTUK MENGANALISIS TOPIK – TOPIK TERKAIT “KPK dan JOKOWI”

¹Imam Adiyana, ²Dr. RB. Fajriyah Hakim, S.Si., M.Si.
^{1,2}Program Studi Statistika, Universitas Islam Indonesia,
¹imam02adiyana@gmail.com, ²fajriyahakim@yahoo.com

ABSTRAK

Sebagai salah satu media jejaring sosial yang keberadaannya masih diminati oleh masyarakat luas sampai saat ini, twitter sering digunakan untuk berbagi kegiatan seperti menjalin relasi bisnis, bertukar informasi ataupun tempat untuk mencurahkan isi hati. Pada twitter pengguna dapat menemukan berbagai macam informasi, bahkan ketika pengguna tidak mengetahui akun dari topik yang dicari pengguna dapat menggunakan bantuan mesin pencarian twitter untuk mencari informasi terkait topik yang dicari. Banyaknya kegiatan yang dapat dilakukan menggunakan twitter, maka tidak dipungkiri twitter akan menghasilkan kumpulan data yang besar. Hal tersebut ditunjukkan dari hasil survei bahwa setiap detiknya pada saat kehidupan normal rata-rata jumlah tweet yang ada dalam twitter adalah 600 tweet. Perlu adanya suatu penanganan menggunakan metode khusus untuk mengalisis data pada twitter sehingga tidak terdapat suatu kondisi yang disebut “Rich of Data but Poor of Information”. Makalah ini membahas mengenai penerapan metode text mining untuk data tweet terkait topik KPK dan topik Jokowi, dimana didapatkan beberapa informasi yang bermanfaat seperti keseringan penggunaan kata-kata menurut aturan asosiasi yang menyertai kata KPK adalah kata Polri dan Laporan, serta kata Jokowi adalah kata Widodo, Menghadiri, Izin, Pintu, Satu, Investor, Urus, Presiden, Nilai, Aktif, Bahaya, Manuver, Menang, Mulai, Relawan, dan Sejumlah. Dalam makalah ini juga dibahas kesamaan topik utama yang dibahas mengenai topik KPK dan Jokowi pada tweet yaitu topik utama KPK dan Polri.

Keyword : *Teks Mining, Twitter, Mesin Pencarian*

I. PENDAHULUAN

Sebagai makhluk sosial, manusia tidak lepas dari kebutuhan dasar untuk bersosialisasi. Sosialisasi secara umum adalah proses belajar individu untuk mengenal dan menghayati norma-norma serta nilai-nilai sosial sehingga terjadi pembentukan sikap untuk berperilaku sesuai dengan tuntutan atau perilaku masyarakatnya. Salah satu cara bersosialisasi dapat dilakukan melalui komunikasi verbal maupun non verbal dan secara langsung ataupun tidak langsung [3]. Melalui komunikasi antar individu dapat bertukar kabar atau berita yang menghasilkan suatu informasi.

Di era modernisasi seperti sekarang ini, sosialisasi antar individu dapat dilakukan dengan komunikasi tidak langsung yaitu melalui media sosial. Media sosial atau sering disebut situs jejaring sosial (*social network sites*) adalah suatu alat (situs media *online*) yang dapat digunakan untuk melakukan komunikasi tanpa adanya interaksi langsung antar individu. Menurut Andreas Kaplan dan Michael Haenlein, mendefinisikan media sosial sebagai “sebuah kelompok aplikasi berbasis internet yang membangun di atas dasar ideologi dan teknologi Web 2.0, dan yang memungkinkan penciptaan dan pertukaran “user-generated

content”. Terdapat banyak jenis media sosial yang berkembang sampai saat ini, salah satunya adalah situs jejaring sosial (*social network sites*) *Twitter*.

Twitter merupakan situs jejaring sosial yang keberadaannya masih diminiati oleh masyarakat sampai saat ini. *Twitter* adalah jejaring sosial berupa blog ukuran kecil yang didirikan oleh Jack Dorsey pada bulan Maret 2006. Melalui *Twitter* pengguna dapat mengirim dan membaca pesan, berbagi informasi, menjalin relasi bisnis, menuangkan isi hati dan pikiran dalam bentuk tulisan (sering disebut *tweet*), dengan kapasitas kata yang bisa diunggah dan ditampilkan pada *timeline* *penggunatwitter* mencapai 140 karakter. Sama halnya dengan situs jejaring sosial lain dalam *Twitter* disediakan suatu mesin pencarian (*search engine*) yang berguna untuk mempermudah pengguna dalam menemukan informasi menggunakan kata kunci. Melalui *search engine* pengguna dapat menemukan lebih banyak informasi yang dibutuhkan terkait topik yang ingin dicari, yaitu lebih dari satu akun yang ada di *twitter*.

Twitter sebagai hasil dari perkembangan teknologi informasi memungkinkan setiap waktu untuk menghasilkan kumpulan data yang banyak, dimana setiap detik pada saat kehidupan normal rata-rata jumlah *tweet* yang ada dalam *twitter* adalah 600 *tweet* [4]. Hal tersebut tidak berlaku jika suatu waktu terjadi peristiwa-peristiwa tertentu yang menyebabkan peningkatan atau penurunan rata-rata jumlah *tweet* perdetiknya. Dengan adanya kumpulan data yang terus meningkat setiap waktunya yaitu berupa data *tweet* perlu dilakukan suatu penanganan menggunakan metode khusus untuk mengalisis data pada *twitter* sehingga menghasilkan suatu informasi yang bermanfaat dan mengurangi kondisi yang biasa disebut “rich of data but poor of information”.

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual, dimana *data mining* memiliki fungsi umum untuk membentuk *assosiation*, *sequence*, *clustering*, *classification*, *regretion*, *forecasting*, dan *solution* [1]. Dalam *data mining* terdapat beberapa metode yang dapat digunakan untuk melakukan analisis data, salah satunya adalah *Teks Mining*. *Teks Mining* didefinisikan sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam *data mining* dimana salah satu fungsinya adalah kategorisasi [2].

Dalam makalah ini akan dibahas mengenai bagaimana cara pengolahan data yang banyak di *twitter* menggunakan metode *teks mining* dan *word cloud*, yang bertujuan untuk menganalisis kumpulan data *tweet* dalam *search engine twitter* terkait topik berita KPK dan Jokowi serta meng*cluster*kan kata-kata yang terkait dengan KPK dan Jokowi dalam suatu *cluster* tertentu menggunakan metode *k-means*. Bagian pertama dalam makalah ini akan membahas alasan perlunya dilakukan analisis data *tweet* pada *twitter*. Bagian kedua akan membahas mengenai metodologi penelitian yang digunakan. Bagian ketiga akan membahas mengenai hasil hasil dan analisis dari penelitian ini, dan bagian terakhir akan membahas kesimpulan serta ringkasan makalah.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan data-data teks pada media sosial *twitter* yang berkaitan dengan kata kunci “kpk”. Tujuan akhir penelitian ini adalah dapat mendeskripsikan topik

utama dan kata - kata yang melekat pada “KPK dan Jokowi”, serta mengelompokkan topik - topik lain yang saling berkaitan.

Populasi

Populasi dari penelitian ini adalah berita/dokumen (teks) pada sosial media *twitter* yang termasuk dalam kategori *microblogging* tentang “kpk” dan “jokowi”. *Tweet-tweet* pengguna pada *twitter* dapat dijadikan sebagai topik-topik pada media *online*.

Sampel dan Teknik Pengambilan Sampel

Sampel yang diambil adalah data teks dari *twitter* yang diambil dengan permintaan kepada sistem sebanyak 500 *tweet* pada periode waktu tertentu.

Sumber Data dan Metode Pengumpulan Data

Data yang digunakan adalah data primer yang dikumpulkan dari media sosial *twitter*. Data diambil dengan cara *download* dan mengumpulkan data *tweet* mengenai KPK dan Jokowi dari hasil pencarian *search engine twitter* (mesin pencarian *twitter*) dan menyimpannya menggunakan suatu program dengan *software* “R”.

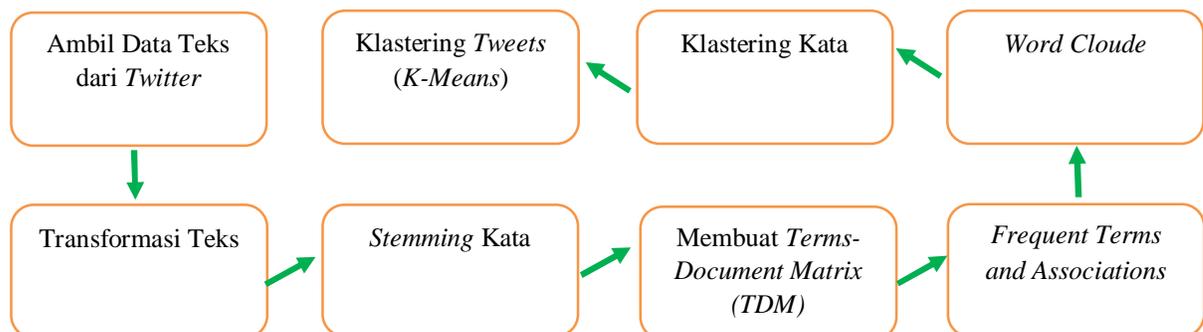
Tahapan Analisis Data

Secara umum, tahapan melakukan analisis *text mining* dapat digambarkan dalam diagram alir berikut :



Pengolahan Data

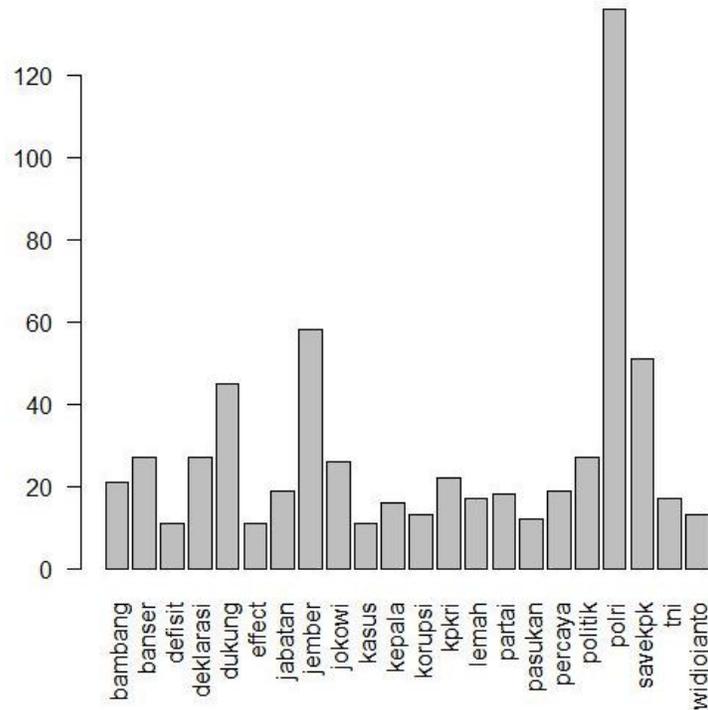
Pengolahan data teks yang didapatkan dari *tweet twitter* diolah menggunakan *Software R* dengan tahapan ditunjukkan pada diagram alir berikut :



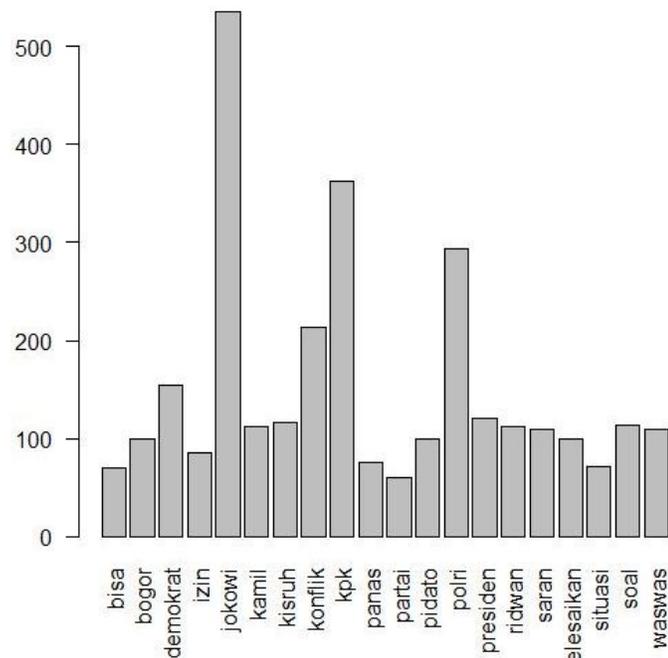
III. HASIL DAN PEMBAHASAN

Term of FrequencyKPK dan Jokowi

Berikut ini merupakan gambar *term of frequency* topik-topik terkait KPK dan Jokowi



Gambar 3.1. Topik-topik terkait KPK



Gambar 3.2. Topik-topik terkait Jokowi

Berdasarkan *barchart* (diagram batang) pada gambar diatas dapat diketahui *term of frequency* dari kata-kata (topik) lain yang sering muncul bersamaan dengan kata KPK dan

Jokowi. *Term of frequency* adalah kata-kata yang sering muncul dari data teks yang dianalisis, yang ditampilkan dalam bentuk diagram batang dimana topik utama terkait kata kunci ditampilkan dengan gambar batang yang lebih panjang dibanding lainnya.

Gambar 3.1. *barchart* menunjukkan kata-kata yang menjadi topik utama dan sering muncul saat pengguna *twitter* mencari informasi terkait KPK pada *search engine twitter* adalah kata Polri, Jember, savekpk (Save KPK), dan Dukung (dukungan). Sedangkan kata-kata lain seperti Bambang, Banser, Defisit, Deklarasi, *Effect* (efek), Jabatan, Jokowi, Kasus, Kepala, Korupsi, kpkri (KPK RI), Lemah, Partai, Pasukan, Percaya, Politik, TNI, dan Widjojanto merupakan topik menarik lain yang cukup sering digunakan terkait pencarian topik KPK.

Gambar 3.2. *barchart* menunjukkan kata-kata yang menjadi topik utama dan sering muncul saat pengguna *twitter* mencari informasi terkait Jokowi pada *search engine twitter* adalah kata Jokowi, KPK, Polri, Konflik, dan Demokrat. Sedangkan kata-kata lain seperti Bisa, Bogor, Izin, Kamil, Kisruh, Panas, Partai, Pidato, Presiden, Ridwan, Saran, Selesaikan, Situasi, Soal, waswas (Was – was) merupakan topik menarik lain yang cukup sering digunakan terkait pencarian topik Jokowi.

Dari gambar 3.1. dan gambar 3.2. diatas dapat dilihat bahwa terdapat kesamaan topik utama ketika pengguna *twitter* mencari informasi terkait KPK dan Jokowi, yaitu topik utama Polri dan KPK. Hal tersebut menunjukkan bahwa terdapat hubungan antara topik KPK dan Jokowi, dimana ketika pengguna *twitter* membuat *tweet* (kicauan) menggunakan kata KPK atau Jokowi pengguna juga sering menggunakan kata Polri dan KPK.

Word Cloud KPK dan Jokowi

Dibawah ini disajikan *word cloud* untuk topik terkait KPK dan Jokowi

Word cloud (awan kata) adalah kumpulan kata-kata yang paling banyak muncul dalam data teks yang dianalisis. Kata-kata tersebut terkumpul seperti sebuah gumpalan awan yang berisi kata-kata sehingga disebut awan kata. Intensitas keseringan kata yang digunakan, ditunjukkan dengan ukuran huruf pada kata. Semakin besar huruf dari kata yang terdapat di awan kata menunjukkan semakin sering kata tersebut muncul. Tampilan awan kata ini lebih menarik serta cepat untuk menemukan kata-kata yang sering muncul, akan tetapi awan kata memiliki kekurangan yaitu tidak dapat menunjukkan frekuensi kata-kata muncul dalam suatu teks yang dianalisis.

Berdasarkan gambar 3.3. dapat diketahui kata (topik) yang sering muncul dalam *word cloud* bersamaan dengan kata KPK adalah kata KPKPolri, Kisruh, Oegroseno, Jimly, Bahas, dan Istana. Sedangkan kata lain yang cukup sering muncul bersamaan dengan kata KPK adalah kata Prasaja, Pejabat, Ketua, Laporan, Wakil, dan negarara (Negarawan). Gambar 3.4. menunjukkan kata (topik) yang sering muncul dalam *word cloud* bersamaan dengan kata Jokowi adalah kata Polri, Konflik dan KPK. Kata lain yang cukup sering muncul bersamaan dengan kata Jokowi adalah kata Selesaikan, Bogor, Demokrat, Presiden, Was-was, Ridwan, Soal, Saran, Kamil, Kisruh, dan Pidato. Sama halnya dengan penjelasan *barchart* pada gambar 3.1. dan 3.2. diatas bahwa pada *word cloud* untuk KPK dan Jokowi juga terdapat kesamaan topik yang sering muncul bersamaan dengan kata KPK dan Jokowi yaitu kata KPK dan Polri. Kesamaan topik utama yang sering muncul pada topik KPK dan Jokowi bisa jadi merupakan kicauan hasil opini atau pemberitaan yang sedang hangat dibicarakan dan diberitakan oleh media mengenai konflik antara KPK dan Polri yang menyangkut juga dengan Presiden Indonesia Joko Widodo.

Asosiasi Kata

Pencarian asosiasi kata-kata yang paling sering keluar selain dengan *barchart* dan *word cloud*, juga dapat dilihat berdasarkan peluang munculnya kata tersebut, sebagaimana tersaji pada keluaran berikut ini. :

```
> findAssocs(myTdm, 'kpk', 0.3)
      kpk
polri 0.50
lapor 0.35
```

Gambar 3.5. Peluang Kata yang Berasosiasi dengan Kata “KPK”

Berdasarkan gambar 3.5. diatas, dengan nilai korelasi tidak kurang dari 0.30 maka dapat diketahui kata-kata yang berasosiasi dengan kata KPK adalah kata Polri dan Laporan, dimana kata Polri lebih sering muncul bersama kata KPK dibanding dengan kata Laporan jika dilihat dari nilai asosiasi.

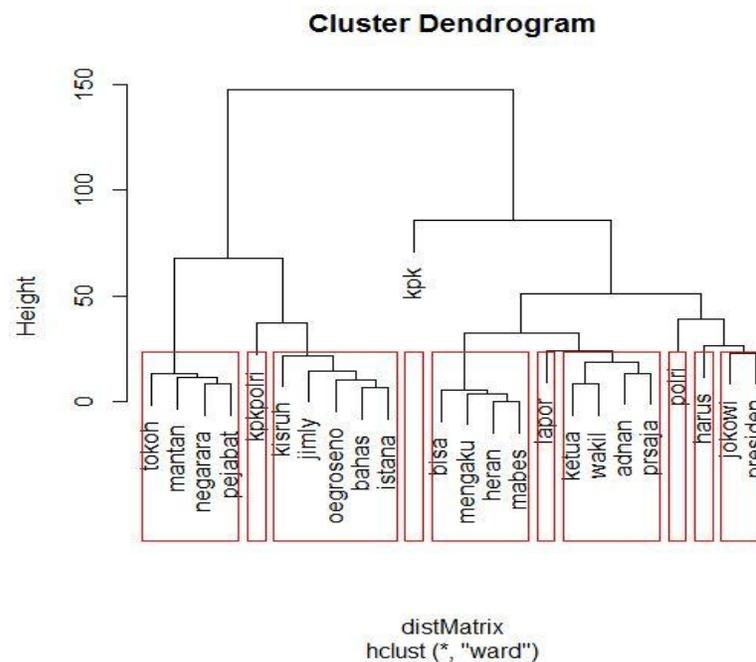
```
> findAssocs(myTdm, 'jokowi', 0.3)
      jokowi
widodo      0.67      nilai      0.36
menghadiri  0.62      aktif      0.35
izin        0.61      bahaya     0.35
pintu       0.61      manuver   0.35
satu        0.61      menang    0.35
investor    0.60      mulai     0.35
urus        0.60      relawan   0.35
presiden    0.50      sejumlah 0.35
```

Gambar 3.6. Peluang Kata yang Berasosiasi dengan Kata “Jokowi”

Gambar 3.6. menunjukkan hubungan kata lain dengan kata Jokowi, dimana dengan nilai korelasi tidak kurang dari 0.30 maka dapat diketahui kata-kata yang berasosiasi dengan kata Jokowi adalah kata Widodo, Menghadiri, Izin, Pintu, Satu, Investor, Urus, Presiden, Nilai, Aktif, Bahaya, Manuver, Menang, Mulai, Relawan, dan Sejumlah.

Clustering Word

Pengelompokan kata digunakan untuk melihat kata-kata yang memiliki karakteristik dan hubungan yang sama dan sering muncul dalam suatu kicauan secara bersamaan. Berikut ini merupakan hasil pengelompokan kata untuk topik KPK dan Jokowi.



Gambar 3.7. Dendrogram Topik “KPK”

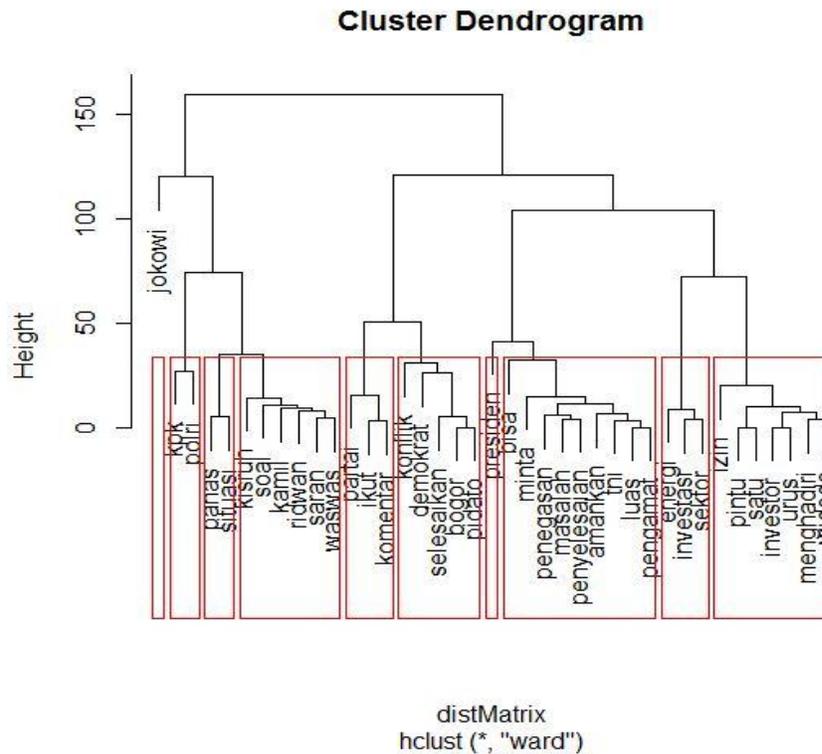
```
> (groups <- cutree(fit, k=10))
  adnan   bahas   bisa   harus   heran   istana   jimly   jokowi
    1     2     3     4     3     2     2     5
  ketua  kisruh  kpk   kpkpolri  lapor   mabes   mantan  mengaku
    1     2     6     7     8     3     9     3
  negarara oegroseno pejabat  polri  presiden  prsaja  tokoh   wakil
    9     2     9     10    5     1     9     1
```

Gambar 3.8. Pembagian Nomor *Cluster* Kata untuk Topik “KPK”

Pada gambar 3.7 dapat dilihat pembagian kelompok kata berdasarkan keamatan hubungan antar kata, dimana kata-kata yang berada dalam satu kelompok memiliki karakteristik yang sama dan biasanya akan selalu muncul bersamaan pada suatu *tweet*. Gambar 3.8 menunjukkan pembagian nomor kelompok kata-kata. Dari hasil *cluster* diketahui bahwa kedua topik (KPK dan Jokowi) dikelompokkan kedalam 10 kelompok kata.

Berdasarkan gambar 3.7 dan 3.8 didapatkan pembagian kelompok kata untuk topik “KPK” sebagai berikut :

- Kelompok 1 : kata Adnan, Ketua, PR Saja, dan Wakil
- Kelompok 2 : kata Bahas, Istana, Jimly, Kisruh, dan Oegroseno
- Kelompok 3 : kata Bisa, Heran, Mengaku, dan Mabes
- Kelompok 4 : kata Harus
- Kelompok 5 : kata Jokowi dan Presiden
- Kelompok 6 : kata KPK
- Kelompok 7 : kata Kpk Polri
- Kelompok 8 : kata Laporan
- Kelompok 9 : kata Mantan, Negarawan (negarara), Pejabat, dan Tokoh
- Kelompok 10 : kata Polri



Gambar 3.9. Dendrogram Topik “Jokowi”

```

> (groups <- cutree(fit, k=10))
  amankan      bisa      bogor      demokrat      energi      ikut
    1         1         2         2         3         4
  investasi  investor      izin      jokowi      kamil      kisruh
    3         5         5         6         7         7
  komentar   konflik      kpk      luas      masalah  menghadiri
    4         2         8         1         1         5
  minta      panas      partai  penegasan  pengamat  penyelesaian
    1         9         4         1         1         1
  pidato     pintu      polri   presiden  ridwan     saran
    2         5         8        10         7         7
  satu       sektor   selesaikan  situasi  soal       tni
    5         3         2         9         7         1
  urus       waswas   widodo
    5         7         5

```

Gambar 3.10. Pembagian Nomor *Cluster* Kata untuk Topik “Jokowi”

Berdasarkan gambar 3.9 dan 3.10 didapatkan pembagian kelompok kata untuk topik “Jokowi” sebagai berikut :

- Kelompok 1 : kata Amankan, Bisa, Luas, Masalah, Minta, Penegasan, Pengamat, TNI, dan Penyelesaian.
- Kelompok 2 : kata Bogor, Demokrat, Konflik, Pidato, dan Selesaikan
- Kelompok 3 : kata Energi, Investasi, dan Sektor
- Kelompok 4 : kata Ikut, Komentar, dan Partai
- Kelompok 5 : kata Investor, Izin, Menghadiri, Pintu, Satu, Urus, dan Widodo
- Kelompok 6 : kata Jokowi
- Kelompok 7 : kata Kamil, Kisruh, Ridwan, Saran, Soal, dan Was-was
- Kelompok 8 : kata KPK dan Polri
- Kelompok 9 : kata Panas dan Situasi
- Kelompok 10 : kata Presiden

Clustering Tweet

Clustering tweet digunakan untuk melihat kelompok *tweet* berdasarkan topik-topik yang berbeda tetapi memiliki karakteristik yang sama. Berikut ini merupakan hasil pengelompokan *tweet* menggunakan metode K-Means untuk topik KPK dan Jokowi.

```

cluster 1: bahas istana kisruh
cluster 2: kpkpolri presiden harus
cluster 3: adnan prsaja ketua
cluster 4: kpk kisruh polri
cluster 5: kpk polri lapor
cluster 6: mantan kpk harus
cluster 7: adnan ketua kpk
cluster 8: kpk polri jokowi

```

Gambar 3.11. *Cluster Tweet* untuk Topik “KPK”

```

cluster 1: demokrat bogor jokowi
cluster 2: jokowi bisa kpk
cluster 3: izin jokowi pintu
cluster 4: jokowi sektor investasi
cluster 5: polri jokowi kpk
cluster 6: kpk polri jokowi
cluster 7: jokowi presiden energi
cluster 8: kpk tni amankan

```

Gambar 3.12. *Cluster Tweet* untuk Topik “Jokowi”

Gambar 3.11. dan 3.12. diatas adalah hasil *cluster tweet* untuk masing-masing topik (KPK dan Jokowi), dimana jika diperhatikan terdapat kesamaan *cluster tweet* untuk topik KPK dengan topik Jokowi yaitu *cluster 8* untuk topik KPK dan *cluster 6* untuk topik Jokowi. Hal tersebut menunjukkan bahwa baik untuk topik KPK maupun Jokowi memiliki hubungan yang dapat dilihat dari topik utama (ditunjukkan dengan kata pertama) pada masing-masing *cluster tweet*. Jika diperhatikan untuk kata-kata pada topik Jokowi untuk *cluster 5* dan *6* adalah sama, tapi yang membedakan disini adalah topik utama dari masing-masing *cluster* yaitu topik utama Polri untuk *cluster 5* dan topik utama KPK untuk *cluster ke-6*.

IV. SIMPULAN

Dari hasil pembahasan diatas mengenai aplikasi *teks mining* untuk penanganan data besar hasil pencarian topik-topik terkait pada *search engine twitter* dengan studi kasus topik KPK dan Jokowi maka dapat ditarik kesimpulan yang menghasilkan sebuah informasi bahwa : terdapat kesamaan topik utama yang digunakan pada *tweet-tweet* ketika pengguna *twitter* mencari informasi seputar topik KPK dan Jokowi pada *search engine twitter* yaitu topik utama KPK dan Polri. Berdasarkan ukuran asosiasi kata dengan nilai korelasi tidak kurang dari 0.30, kata-kata yang berasosiasi dengan kata KPK adalah kata Polri dan Laporan. Sedangkan kata-kata yang berasosiasi dengan Jokowi dimana nilai korelasi tidak kurang dari 0.30 adalah kata Widodo, Menghadiri, Izin, Pintu, Satu, Investor, Urus, Presiden, Nilai, Aktif, Bahaya, Manuver, Menang, Mulai, Relawan, dan Sejumlah. Hasil pengelompokan kata berdasarkan tingkat keseringan kata-kata muncul bersamaan didapatkan 10 kelompok untuk kedua topik (KPK dan Jokowi). Terdapat kesamaan topik yang dibahas pada pengelompokan *tweet* untuk topik KPK dan Jokowi yaitu *tweet* yang membahas mengenai topik "kpk, polri, dan jokowi". Kesamaan-kesamaan penggunaan kata pada *tweet* untuk dua topik yang berbeda menunjukkan adanya hubungan dari kedua topik tersebut, dimana ditunjukkan dari hasil analisis mengenai data *tweet* topik KPK dan data *tweet* topik Jokowi yang memiliki keseringan dan kesamaan penggunaan kata pada *tweet*.

DAFTAR PUSTAKA

- [1] Pramudiono, I. 2007. “Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data”. <http://www.ilmukomputer.org/wpcontent/uploads/2006/08/iko-data-mining.zip>(diakses pada 11 januari 2015)
- [2] Triawati, Candra. 2009. “Metode Pembobotan *Statistical Concept Based* untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia”. <http://digilib.tes.telkomuni>

www.ums.ac.id

<http://www.ums.ac.id/index.php?view=article&catid=20:informatika&id=590:textmining&tmpl=component&print=1&page=> (diakses pada 11 januari 2015)

- [3] Sasrawan, Hedi. 2013. “Pengertian Sosialisasi”.
<http://hedisasrawan.blogspot.com/2013/01/pengertian-sosialisasi-artikel-lengkap.html> (diakses pada 11 januari 2015)
- [4] <http://media.kompasiana.com/new-media/2010/02/23/twitter-hampir-melampaui-facebook-80169.html> (diakses pada 11 januari 2015)