**PROPOSAL PRA SKRIPSI**

**ANALISA TOPIK DATA MEDIA SOSIAL TWITTER TERHADAP**

**PEJABAT PUBLIK MENGGUNAKAN LATENT SEMANTIC ANALYSIS**

**(Studi Kasus : Gubernur Jawa Tengah)**



**Disusun Oleh :**

**WIDYA SULISTYANI**

**Nomor Mahasiswa : 165410086**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER AKAKOM YOGYAKARTA**

**2020**

## HALAMAN PERSETUJUAN

|  |  |  |
| --- | --- | --- |
| **Judul** | **:** | **Analisa Topik Data Media Sosial Twitter Terhadap Pejabat Publik Menggunakan Latent Semantic Analysis** |
| **Nama** | **:** | **Widya Sulistyani** |
| **Nomor Mhs** | **:** | **165410086** |
| **Program Studi** | **:** | **Teknik Infomatika** |
| **Jenjang** | **:** | **Strata Satu (S1)** |
| **Tahun** | **:** | **2020** |

**Telah diperiksa dan disetujui**

**Yogyakarta, 29 Mei 2020**

**Mengetahui**

**Dosen Pembimbing**

[**Dini Fakta Sari, S.T., M.T.**](https://scholar.google.co.id/citations?user=RQCxqBQAAAAJ&hl=id)

## DAFTAR ISI

[HALAMAN PERSETUJUAN i](#_Toc42845665)

[DAFTAR ISI ii](#_Toc42845666)

[DAFTAR ISI iii](#_Toc42845667)

[DAFTAR GAMBAR iii](#_Toc42845668)

[DAFTAR TABEL iii](#_Toc42845669)

[BAB I PENDAHULUAN 1](#_Toc42845670)

[1.1 Latar Belakang Masalah 1](#_Toc42845671)

[1.2 Rumusan Masalah 2](#_Toc42845672)

[1.3 Ruang Lingkup 3](#_Toc42845673)

[1.4 Tujuan Penelitian 3](#_Toc42845674)

[1.5 Manfaat Penelitian 3](#_Toc42845675)

[BAB 2. TINJAUAN PUSTAKA DAN DASAR TEORI 4](#_Toc42845676)

[1.6 Tinjauan Pustaka 4](#_Toc42845677)

[1.7 Dasar Teori 9](#_Toc42845678)

[1. Singular Value Decomposition 13](#_Toc42845679)

[Singular Value Decomposition (SVD) adalah salah satu teknik reduksi dimensi yang bermanfaat untuk memperkecil nilai kompleksitas dalam pemrosesan term-document matrix. SVD merupakan teorema aljabar linier yang menyebutkan bahwa persegi panjang dari term-document matrix dapat dipecah/didekomposisikan menjadi tiga matriks, yaitu : – Matriks ortogonal U – Matriks diagonal S – Transpose dari matriks ortogonal V Yang dirumuskan dengan : 13](#_Toc42845680)

[2. Cosine Similarity Measurement 14](#_Toc42845681)

[Cosine similarity digunakan untuk menghitung nilai kosinus sudut antara vektor dokumen dengan vektor kueri. Semakin kecil sudut yang dihasilkan, maka tingkat kemiripan esai semakin tinggi. Formula dari cosine similarity adalah sebagai berikut: 14](#_Toc42845682)

[BAB 3. METODE PENELITIAN 15](#_Toc42845683)

[1.8 Bahan/Data 15](#_Toc42845684)

[1.9 Peralatan 15](#_Toc42845685)

[1.10 Prosedur dan Pengumpulan Data 15](#_Toc42845686)

[1.11 Analisis dan Rancangan Sistem 15](#_Toc42845687)

# DAFTAR ISI

# DAFTAR GAMBAR

# DAFTAR TABEL

## 

## BAB I PENDAHULUAN

### Latar Belakang Masalah

Media sosial merupakan sebuah *platform* yang penggunanya dapat berbagi pandangan, ide, perasaan, dan emosi mengenai suatu konteks peristiwa atau hal yang sedang terjadi saat itu. Media sosial semakin populer seiring dengan semakin mudahnya akses internet pada laptop, komputer, dan perangkat *mobile*. Banyaknya pengguna media sosial yang aktif menyuarakan pendapat mereka, menjadikan media sosial sebagai sumber data utama dalam bidang *social data mining* (Matwin 2013).

Twitter merupakan salah satu media sosial populer di dunia sejak pertama kali dipublikasikan pada tahun 2006 (Singh 2015). Indonesia menggunakan Twitter sejak awal dipublikasikan dan termasuk dalam pengguna Twitter yang paling produktif. Terdapat sekitar 29 juta pengguna Twitter di Indonesia. Pada tahun 2014, Indonesia menempati peringkat kelima sebagai negara dengan *tweet* terbanyak (Carley *et al* 2015).

Banyaknya pengguna Twitter di Indonesia menjadikan twitter sebagai alat pendukung … demokrasi di Indonesia. Ini dapat dibuktikan dengan banyaknya pejabat pemerintah yang menggunakan twitter sebagai alat komunikasi dengan masyarakat, salah satunya gubernur jawa tengah. Dengan adanya twitter, pejabat dapat menyampaikan berbagai informasi, dan masyarakat dapat dengan cepat menerima informasi, merespon, bahkan menyampaikan aspirasi maupun membahas berbagai keresahan maupun isu yang berkembang di masyarakat.

Dengan demikian masyarakat dapat mengikuti perkembangan terkini suatu isu dan bisa memberikan opininya. Sehingga twitter bisa menjadi sumber data terhadap suatu isu yang berkembang.

Cuitan pada twitter atau biasa disebut tweet umumnya terdiri dari kalimat panjang dari banyak akun sehingga diperlukan sebuah sistem untuk melakukan analisis topik yang sedang ramai diperbincangkan, tanpa harus membaca keseluruhan tweet yang ada. Salah satu bidang yang mampu mengatasi masalah ini adalah pemodelan topik.

Analisis topik dilakukan untuk memberikan gambaran umum yang mewakili isi dari topik utama. Begitu banyak *tweet* yang dapat dianalisis untuk memperoleh topik utama. Diperlukan suatu metode yang cepat dan efisien yaitu dengan metode pemodelan topik. Pemodelan topik dapat digambarkan sebagai metode untuk menemukan kelompok kata (topik) dari kumpulan dokumen yang dapat merepresentasikan dengan baik informasi yang ada dalam kumpulan dokumen tersebut (Nair 2016). Salah satu model dari pemodelan topik adalah model *Latent Semantic Analysis.*

### Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah :

#### Bagaimana membuat sistem yang mampu melakukan pengambilan data twitter.

#### Bagaimana memproses t*weets* untuk memperoleh informasi berupa topik apa saja yang dibicarakan dan menampilkan trending topik menggunakan pemodelam topik Latent Semantic Analysis.

### Ruang Lingkup

Ruang lingkup dari penelitian ini adalah sebagai berikut :

#### Data *tweet* yang digunakan diperoleh dari proses *crawling* pada akun @ganjarpranowo selaku gubernur jawa tengah.

#### Penelitian dilakukan 3 bulan yaitu 20 oktober 2019 sampai 28 januari 2020.

#### Korpus yang digunakan adalah korpus berbahasa Indonesia.

#### Pemodelan topik yang digunakan adalah *Latent Semantic Analysis* (LSA).

### Tujuan Penelitian

Penelitian ini diharapkan dapat memberikan pengetahuan tentang bagaimana cara melakukan eksplorasi opini, memberikan pengetahuan tentang tanggapan atau aspirasi pengguna jejaring sosial twitter terhadap tokoh publik.

### Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Bagi pihak @ganjarpranowo selaku pejabat publik, penelitian ini dapat membantu mengefisienkan waktu untuk mengetahui, isu yang berkembang, tanggapan atau aspirasi masyarakat mengenai konten twitter maupun kebijakan yang dilakukan saat menjabat, sehingga dapat menjadi bahan evaluasi bagi @ganjarpranowo dalam pembuatan konten maupun dalam pembuatan kebijakan jika diperlukan.
2. Bagi peneliti lain, penelitian ini dapat menjadi bahan referensi untuk penelitian pemodelan topik, utamanya *latent semantic analysis*.

## BAB 2. TINJAUAN PUSTAKA DAN DASAR TEORI

### Tinjauan Pustaka

Untuk membedakan penelitian ini dengan penelitian sebelumnya maka penulis melakukan studi pustaka terhadap peneliti peneliti terdahulu.

Ardytha Luthfiarta, Junta Zeniarja, Abu Salam(2013). Peneliti melakukan peringkasan dokumen otomatis dengan menggunakan algoritma *Latent Semantic Analysis* (LSA). Berdasarkan percobaan-percobaan yang telah dilakukan dapat disimpulkan bahwa Peringkas Dokumen Otomatis dengan *Latent Semantic Analysis* (LSA) pada Proses Clustering Dokumen Teks Berbahasa Indonesia dapat meningkatkan kinerja clustering dokumen lebih baik dari pada Peringkas Dokumen Otomatis dengan Metode Fitur dan Proses Clustering Dokumen Standar, mengalami peningkatan dari tingkat akurasi 65,92 % untuk proses clustering standar menjadi 71,04% untuk proses clustering dokumen menggunakan peringkas dokumen otomatis dengan *Latent Semantic Analysis* (LSA).

Dary Mochamad Irfan (2016). Peneliti melakukan uji efektifitas penggunaan metode latent semantic analysis yang merupakan proses analisis semantik pada short text al-Quran untuk pencarian nilai kesamaan makna ayatnya sesuai makna yang terkandung dari tiap ayat. Mengingat LSA bersifat dinamis, dimana tidak ada cara yang pasti untuk menentukan berapa dimensi yang harus digunakan untuk setiap kasus, maka perlu diketahui bagaimana pengaruh pemilihan dimensi pada nilai similarity. Berdasarkan hasil pengujian, term-document matrix tanpa LSA memiliki akurasi yang tinggi yaitu 85% tetapi berdasarkan pengukuran F-measure, efektifitasnya sangat rendah yaitu 0% karena tidak bisa mendeteksi ayat yang maknanya serupa. Dibandingkan dengan pengukuran dengan LSA walaupun akurasi tertinggi dari salah satu variabel 71% tapi nilai F-measure mencapai 40% yang dimiliki oleh variabel LSA yang lain. Selain itu dimensi yang digunakan berpengaruh pada hasil. Jika terlalu rendah maka nilai kesamaan akan terlalu seragam, bias dan hasilnya tidak akan optimal.

Jaka Eka Sembodo, Abdurahman Baizal, Erwin Budi Setiawan (2016) membangun sistem crawling data twitter secara otomatis menggunakan *Application Programming Interface* (API) twitter dengan bahasa pemograman PHP dan basisdata MySQL. Pencarian dapat dilakukan dengan by user dan by keyword. Batasan maksimum data user yang diunduh maksimum 200 tweet dan maksimum data tweet yang diunduh maksimum 100 tweet. Ekstrasi fitur berdasarkan indeks twitter untuk data user meliputi: id user, nama dana kun user, total tweet, total followers, total status, total likes, bio profile, website dan lokasi. Sedangkan untuk ekstraksi fitur untuk data tweet meliputi: Id tweet, teks tweet, retweet, url, hashtag, mention, sumber tweet, total like, total retweet dan waktu tweet.

Keke Putri Utami (2017) melakukan pemodelan topik menggunakan *Latent Dirichlet Allocation* pada lima lokasi *tweet* di Kota Bogor dan rentang waktu tertentu, dan berhasil membentuk topik dengan informasi atau deskripsi topik untuk setiap lokasi *tweet*. Deskripsi topik untuk setiap lokasi *tweet* menunjukkan topik-topik tersebut sedang banyak dibicarakan oleh pengguna Twitter pada rentang waktu yang ditentukan. Jumlah topik yang ditentukan untuk data csv setiap lokasi *tweet* telah dapat menghasilkan kumpulan kata yang membentuk topik dengan baik. Informasi yang mewakili isi topik dapat dimanfaatkan oleh pembaca atau *stakeholder* terkait dalam memahami setiap perkembangan isu terkini.

Nurina Savanti Widya Gotami, Indriati, Ratih Kartika Dewi (2018).Peneliti melakukan peringkasan teks otomatis secara ekstraktif pada artikel berita kesehatan berbahasa Indonesia dengan menggunakan LSA dapat diterapkan dengan cara LSA sebagai algoritme untuk mendapat kalimat kalimat yang memiliki keterkaitan kata dengan pendekatan secara semantik dengan menggunakan SVD sebagai fitur penghilang redudansi atau *noise* pada kata tertentu. Serta penggunaan *Cross method* LSA sebagai pengekstrasi ringkasan yang akan dipilih dari artikel berita kesehatan dalam data teks dokumen.

Tinjauan keenam ditulis oleh Septian Narsa Putra(2018). Peneliti membuat sistem yang dapat melakukan klasifikasi terhadap berita yang diupload pada akun twitter Divis Humas Polri ke dalam tiga katagori yaitu berita kegiatan polisi, komentar masyarakat dan layanan masyarakat selama empat tahun ke belakang. Setelah berita diklasifikasi kemudian dicari sentimen dari setiap topik. Metode yang digunakan adalah Naive Bayes Classifier.

Widya Sulistyani (2020)

Perbandingan dari penelitian - penelitian diatas dapat dilihat pada table 2.1

**TABEL 2.1 PERBANDINGAN PENELITIAN**

|  |  |  |  |
| --- | --- | --- | --- |
| Penulis | Objek | Metode | Hasil |
| Ardytha Luthfiarta 1, Junta Zeniarja 2, Abu Salam 3 (2013) | Dokumen Teks Berbahasa Indonesia | *Latent Semantic Analysis (LSA)* | Sistem untuk proses c*lustering d*okumen teks berbahasa Indonesia. |
| Dary Mochamad Irfan (2016) | Al-Quran | *Latent Semantic Analysis (LSA)* | Sistem untuk uji efektifitas penggunaan metode latent semantic analysis yang merupakan proses analisis semantik pada short text al-Quran untuk pencarian nilai kesamaan makna ayatnya sesuai makna yang terkandung dari tiap ayat. |
| Jaka Eka Sembodo 1,  Abdurahman Baizal 2,  Erwin Budi Setiawan 3 (2016) | Data Twitter | *Application Programming Interface* (API) twitter | Membangun sistem crawling data twitter secara otomatis menggunakan *Application Programming Interface* (API) twitter dengan bahasa pemograman  PHP dan basisdata MySQL. |
| Keke Putri Utami (2017) | Data Twitter | *Latent Dirichlet Allocation* | Pemodelan topik LDA pada 5 lokasi tweet di kota bogor membentuk informasi topik setiap lokasi tweet. Deskripsi topik untuk setiap lokasi *tweet*  menunjukkan topik-topik tersebut sedang banyak dibicarakan oleh pengguna  Twitter pada rentang waktu yang ditentukan |
| Nurina Savanti Widya Gotami 1, Indriati 2, Ratih Kartika Dewi 3 (2018) | Artikel berita kesehatan dalam data teks dokumen | *Latent Semantic Analysis (LSA)* | Sistem peringkasan teks otomatis secara ekstraktif pada artikel kesehatan berbahasa Indonesia. |
| Septian Narsa Putra (2018) | Data Twitter | Naive Bayes Classifier | Sistem klasifikasi berita pada akun twitter. |
| Widya Sulistyani (2020) | Data Twitter | *Latent Semantic Analysis (LSA)* | Sistem ….. |

### Dasar Teori

#### Twitter

Twitter merupakan salah satu media sosial dengan layanan *microbloging* yang terkenal dan memungkinkan para penggunanya untuk menulis sesuatu atau yang biasa disebut tweet. Twitter digunakan untuk mengutarakan opini publik maupun berita resmi dari suatu instansi atau dari akun kantor berita. Twitter dibangun oleh Jack Dorsey pada tahun 2006 dengan alamat http://www.twitter.com, jika seseorang ingin menggunakan twiiter seseorang harus terlebih dahulu memiliki akun, untuk registrasinya dapat dilakukan pada alamat tersebut.

Pengguna dapat menulis pesan berdasarkan topik dengan tanda #(tagar). Sedangkan untuk menyebut atau membalas pesan dari pengguna lain bisa menggunakan tanda @(diikuti nama akun yang akan dibalas).

#### Crawling data merupakan tahap dalam penelitian yang bertujuan untuk mengumpulkan atau mengunduh data dari suatu database (3, 4). Pengumpulan data dari penelitian ini yaitu data yang diunduh dari server twitter berupa user dan tweet beserta atribut-atributnya.

#### API Twitter

API Twitter atau Application Programming Interface (API) twitter adalah suatu program atau aplikasi yang disediakan oleh twitter untuk mempermudah developer lain dalam mengakses informasi yang ada di website twitter (3, 4). Pendaftaran sebagai developer aplikasi twitter untuk menggunakan API twitter dapat dilakukan di lama *https://dev.twitter.com*. Setelah mendaftar developer akan mendapatkan consumer key, consumer access, access token dan access token secret yang akan digunakan sebagai syarat otentifikasi dari aplikasi yang akan kita bangun. Tujuan dari otentifikasi adalah untuk hak akses developer dalam mengunduh data yang ada di twitter.

#### Text Mining

Text Mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. Penambangan teks memiliki tujuan dan menggunakan proses yang sama dengan penambangan data, namun memiliki masukan yang berbeda. Masukan untuk penambangan teks adalah data yang tidak (atau kurang) terstruktur, sedangkan masukan untuk penambangan data adalah data yang terstruktur (Ronen Feldman, 2006). *Text mining* dapat diterapkan untuk pelacakan topik, katagorisasi dan analisis sentimen.

Dalam proses penambangan teks terdapat proses pengubahan data menjadi data yang terstruktur sesuai dengan kebutuhan, proses ini disebut *Text Prepocessing*, dengan tahapannya adalah Case Folding, Tokenizing, Filtering dan Analyzing. *Case Folding* adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil (lowercase). Hanya huruf “a” sampai dengan “z” yang diterima dan karakter selain huruf dihilangkan. Tahapan *Tokenizing* adalah proses pemotongan kalimat (*string*) berdasarkan kata peyusunnya. Kemudian tahap *filtering*, merupakan tahapan mengambil kata - kata penting dari hasil token dan membuang sebagian kata tertentu. Kemudian tahap *filtering*, merupakan tahapan mengambil kata - kata penting dari hasil token dan membuang sebagian kata tertentu. Yang terakhir adalah tahap *analyzing* yaitu tahap untuk mencari seberapa jauh keterhubungan antar kata - kata setiap dokumen.

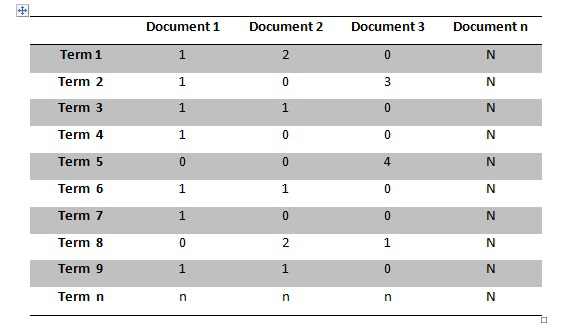
#### Pemodelan Topik

#### Latent Semantic Analysis

Latent Semantic Analysis (LSA) merupakan sebuah metode yang memanfaatkan model statistik matematis untuk menganalisa struktur semantik suatu teks. Berikut merupakan langkah langkah menggunakan metode Latent Semantic Analysis.

1. Term Document Matric

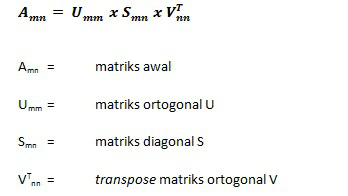
Setelah melalui stopwords removal dan stemming, matriks term-document dibangun dengan menempatkan kata hasil proses stemming (term) ke dalam baris. Matriks ini disebut term-document matrix. Setiap baris mewakili sebuah kata yang unik, sedangkan setiap kolom mewakili konteks dari mana kata-kata tersebut diambil. Konteks yang dimaksud bisa berupa kalimat, paragraf, atau seluruh bagian dari teks.  
Di bawah ini merupakan contoh term-document matrix:



Pada tabel di atas, baris pertama mewakili stemmed term (term 1, term 2, dst), dan bagian kolom mewakili konteks, yaitu teks. Nilai yang terletak pada setiap cell pada tabel menunjukkan berapa kali sebuah term muncul dalam sebuah dokumen. Contohnya, term 1 muncul 1 kali pada dokumen ke-1, dan muncul 2 kali pada dokumen ke-2, namun term 1 tidak muncul pada dokumen 3, dan seterusnya.

## Singular Value Decomposition

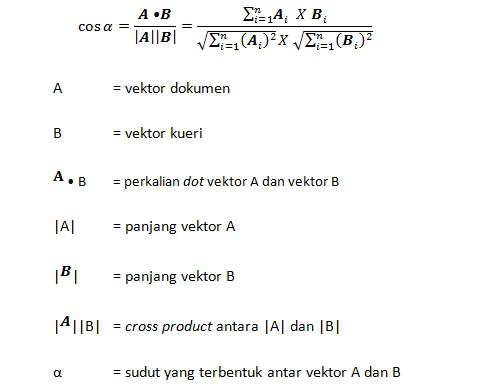
## Singular Value Decomposition (SVD) adalah salah satu teknik reduksi dimensi yang bermanfaat untuk memperkecil nilai kompleksitas dalam pemrosesan term-document matrix. SVD merupakan teorema aljabar linier yang menyebutkan bahwa persegi panjang dari term-document matrix dapat dipecah/didekomposisikan menjadi tiga matriks, yaitu : – Matriks ortogonal U – Matriks diagonal S – Transpose dari matriks ortogonal V Yang dirumuskan dengan :



Hasil dari proses SVD adalah vektor yang akan digunakan untuk menghitung similaritasnya dengan pendekatan cosine similarity.

## Cosine Similarity Measurement

## Cosine similarity digunakan untuk menghitung nilai kosinus sudut antara vektor dokumen dengan vektor kueri. Semakin kecil sudut yang dihasilkan, maka tingkat kemiripan esai semakin tinggi. Formula dari cosine similarity adalah sebagai berikut:



Dari hasil cosine similarity, akan didapatkan nilai yang akan dibandingkan dengan penilaian manusia untuk diuji selisih nilainya.

## BAB 3. METODE PENELITIAN

### Bahan/Data

### Peralatan

### Prosedur dan Pengumpulan Data

### Analisis dan Rancangan Sistem

**DAFTAR PUSTAKA**

Penulisan daftar pustaka berisi:

nama pengarang (urut alpabet), tahun, judul (italic), penerbit, kota penerbit

Bila berasal dari web site: nama pengarang (urut alpabet), tahun, judul (italic), url atau web address, tangggal dan jam akses.