

# Rekomendasi Topik Khotbah Jumat Berdasarkan Twitter Trending Topics Menggunakan Vector Space Model

*by* Indah Sawitri, Ardiansyah; Ramonasari , Ardiansyah;

---

**Submission date:** 21-Jun-2019 09:09AM (UTC+0700)

**Submission ID:** 1145703683

**File name:** SNATi\_2019\_paper\_59.pdf (770.3K)

**Word count:** 3181

**Character count:** 19826

# Rekomendasi Topik Khotbah Jumat Berdasarkan Twitter Trending Topics Menggunakan Vector Space Model

22

Indah Sawitri Ramonasari, Ardiansyah

Program Studi Teknik Informatika Fakultas Teknologi Industri

Universitas Ahmad Dahlan

Kota Yogyakarta

indahsramonasari.19@gmail.com, ardiansyah@tif.uad.ac.id

13

**Abstrak**—Topik khotbah Jumat yang ideal adalah yang sesuai dengan isu yang sedang berkembang di masyarakat. Namun, pada kenyataannya topik khotbah Jumat hanya dipilih berdasarkan materi yang dikuasai oleh khatib saja. Penelitian ini bertujuan untuk membangun sistem rekomendasi materi khotbah Jumat berdasarkan *trending topics* di Twitter dengan menerapkan metode *Vector Space Model*. Hasil penelitian menunjukkan bahwa sistem rekomendasi materi khotbah sudah mampu berfungsi untuk memberikan rekomendasi topik dan materi khotbah Jumat. Sistem memiliki fitur rekomendasi berdasarkan *trending topics* dan *input query* sebagai alternatif ketika pengguna tidak menemukan rekomendasi sesuai dengan *trending topics*. Topik yang direkomendasikan oleh sistem ini telah sesuai dengan isu yang sedang berkembang di masyarakat. Evaluasi sistem dilakukan dengan menggunakan 14 *query* dan 360 dokumen. Hasil evaluasi menunjukkan nilai akurasi *mean average precision* (MAP) untuk sistem rekomendasi khotbah Jumat adalah sebesar 0.906 atau 91%. Nilai MAP tersebut dinyatakan sangat baik atau masuk pada klasifikasi *excellent recommendation* berdasarkan *diagnostic test*. Hasil penelitian ini selanjutnya bisa digunakan oleh para khotib di seluruh Indonesia dalam menemukan topik dan materi khotbah Jumat yang sesuai dengan isu terkini di masyarakat.

**Kata kunci**—sistem rekomendasi; *vector space model*; *cosine similarity*; *TF-IDF*; *trending*

## 4 I. PENDAHULUAN

Khotbah yang baik dan efektif sangat ditentukan oleh khatibnya, mulai dari pemilihan topik, penyajian materi, penyusunan naskah dan gaya bahasa atau segi pemanfaatan waktu dan penampilan khatib [1].

Beberapa temuan hasil dari wawancara kepada khatib Jumat yaitu Drs. Sutijan didapatkan bahwa pemilihan topik khotbah Jumat kadang ditentukan hanya berdasarkan materi khotbah yang dikuasai khatib saja. Khatib jarang menyampaikan topik khotbah yang berhubungan dengan isu yang sedang hangat di masyarakat karena keterbatasan waktu dalam penentuan topik. Beliau memaparkan bahwa sebenarnya telah tersedia kumpulan dokumen khotbah Jumat. Akan tetapi, khatib masih merasa kesulitan dalam mengelola dokumen khotbah Jumat tersebut. Tidak jauh beda dengan yang dipaparkan oleh Dr. Okrisal Eka Putra, Lc., M.Ag. (Pengurus

MIUMI DIY dan PP Muhammadiyah), bahwa kumpulan dokumen khotbah Jumat masih belum dikelompokkan berdasarkan topik tertentu sehingga mempersulit proses penentuan topik.

Dalam memilih tema khotbah, seorang khatib harus mengambil tema yang bersifat konsumtif yakni tema khotbah yang disampaikan harus betul-betul dirasakan sesuai kebutuhan jemaah yang mendesak [2]. Tantangan lain bagi khatib adalah betapa sulitnya membuat jemaah tetap tekun dan antusias mendengarkan uraiannya sepanjang khotbah berlangsung. Tak jarang terdapat khotbah yang kurang diperhatikan dan ditinggal tidur oleh jemaahnya [1]. Fakta ini hendaknya segera dicari penyebab dan solusinya, sehingga khotbah yang disampaikan bisa secara optimal diterima dan disimak oleh jemaah.

Berdasarkan data kuisioner awal terhadap jemaah shalat Jumat dari berbagai masjid, 52,9% dari 34 responden masih merasa bahwa topik yang disampaikan jarang berhubungan dengan isu yang sedang hangat di masyarakat. Seyogyanya, dari isu yang sedang berkembang tersebut sebenarnya dapat diambil sebagai topik khotbah Jumat sehingga jemaah dapat mengetahui bagaimana cara untuk menyikapi isu tersebut. Topik khotbah yang berhubungan dengan isu yang sedang populer juga dapat mempengaruhi semangat jemaah dalam mendengarkan khotbah.

Disisi lain, dokumen tertulis baik di majalah maupun situs online banyak memuat isi dokumen khotbah Jumat. Arsip dokumen khotbah Jumat dikeluarkan secara online oleh organisasi kemasyarakatan seperti Muhammadiyah yang telah melakukan *posting* dokumen khotbah sebanyak 38 dokumen, Nahdlatul Ulama sebanyak 15 dokumen, dan Yufid sebanyak 307 dokumen. Perkembangan data teks khotbah Jumat yang mencapai jumlah besar tentu berkaitan erat dengan *text mining*. Dengan mengaplikasikan proses-proses dalam *text mining*, maka akan diperoleh pola-pola data, tren, dan ekstraksi dari pengetahuan-pengetahuan yang potensial dari data teks [3]. Dengan *text mining*, pemrosesan dokumen khotbah Jumat untuk pemilihan topik khotbah yang kontekstual dapat diwujudkan.

Beberapa penelitian telah menerapkan *text mining* untuk klasifikasi berita *online* [4] dan sistem penentuan dosen penguji skripsi [5]. Penelitian klasifikasi berita *online* dilakukan menggunakan metode *Single Pass Clustering*, sedangkan penentuan dosen penguji skripsi menggunakan metode *Vector Space Model*. Akan tetapi, dari semua sistem rekomendasi dan metode yang digunakan, belum pernah dibangun sistem rekomendasi khotbah Jumat, sehingga belum diketahui bagaimana penerapan metode-metode tersebut secara nyata pada sistem rekomendasi khotbah Jumat.

Berdasarkan uraian masalah yang telah dipaparkan, maka penelitian ini akan mengembangkan sistem rekomendasi khotbah Jumat berdasarkan *trending topics Twitter* dengan *Vector Space Model*. *Trending topics* diperoleh melalui proses *crawling* terhadap *trending topics* terkini yang sedang terjadi di Indonesia. Penelitian ini bertujuan untuk membangun sistem rekomendasi khotbah Jumat berdasarkan *trending topics* dengan menerapkan metode *Vector Space Model* sehingga diperoleh topik khotbah sesuai dengan isu yang sedang hangat di masyarakat.

## II. KAJIAN PUSTAKA

### A. Information Retrieval

*Information Retrieval* merupakan suatu sistem yang menyimpan informasi dan menemukan kembali informasi tersebut. Secara konsep ada beberapa dokumen atau kumpulan *record* yang berisi informasi yang diorganisasikan ke dalam sebuah media penyimpanan untuk tujuan mempermudah ditemukan kembali. Dokumen yang tersimpan tersebut dapat berupa kumpulan *record* inform bibliografi maupun data lainnya [6]. *Information retrieval* menemukan bahan (biasanya dokumen) bersifat tidak terstruktur (biasanya teks) yang memenuhi kebutuhan informasi dari dalam koleksi besar (biasanya disimpan di komputer) [7].

### B. Vector Space Model

*Vector Space Model* merupakan suatu metode yang digunakan untuk mengukur tingkat kedekatan atau kesamaan (similarity) term dengan cara pembobotan pada term [8]. Perhitungan kemiripan antara vektor dokumen dan vektor query dilihat dari sudut yang paling kecil. Sudut yang dibentuk oleh dua buah vektor dapat dihitung dengan melakukan inner product [9].

Proses penting yang dilakukan dalam penerapan *Vector Space Model* antara lain:

#### 1. Term Frequency-Inverse Document Frequency

*Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan teknik dalam memberikan bobot hubungan suatu term terhadap sebuah dokumen [10]. Proses pembobotan kata dijabarkan pada persamaan (1).

$$TF-IDF(t_i, d_j) = tf(t_i, d_j) * \log(N \div N(t_i)) \quad (1)$$

Keterangan :

TF-IDF ( $t_i, d_j$ ) = bobot term  $t_i$  terhadap dokumen  $d_j$

tf ( $t_i, d_j$ ) = jumlah kemunculan term  $t_i$  dalam dokumen  $d_j$

N = jumlah semua dokumen

$N(t_i)$  = jumlah dokumen yang mengandung term  $t_i$

### 2. Cosine Similarity

*Cosine similarity* merupakan ukuran kesamaan antara dua vektor n dimensi dengan mencari kosinus sudut antara mereka, *cosine similarity* sering digunakan untuk membandingkan dokumen dalam penambangan teks [10]. Perhitungan *cosine similarity* ditunjukkan dalam persamaan 2.

$$\text{Cos}(x, y) = (x * y) \div (||x|| * ||y||) \quad (2)$$

Keterangan :

$||x||$  = Panjang vektor x

$||y||$  = Panjang vektor y

*Cosine similarity* antara vektor *query* dan vektor dokumen dapat digunakan sebagai ukuran skor dokumen untuk *query* tersebut, nilai yang dihasilkan kemudian dapat digunakan untuk memilih penilaian atas dokumen untuk *query* [7].

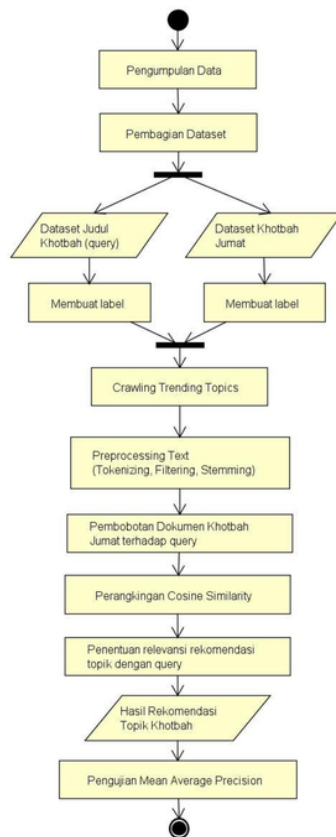
### C. Text Mining

*Text mining* merupakan proses analisis dalam data yang berupa teks di mana sumber data didapatkan dari dokumen [11]. *Text mining* digunakan untuk mendeskripsikan teknik dari *data mining* yang secara otomatis menemukan sesuatu hal yang berguna atau sebuah pengetahuan baru dari sebuah teks yang tidak terstruktur [12].

Tahapan analisa pada *text mining* yaitu melakukan pengumpulan data kemudian melakukan ekstraksi terhadap fitur yang akan digunakan [11]. Dengan mengaplikasikan proses-proses dalam *text mining*, maka akan diperoleh pola-pola data, tren, dan ekstraksi dari pengetahuan-pengetahuan yang potensial dari data teks [3].

## III. METODOLOGI PENELITIAN

Tahapan penelitian yang telah diuraikan mulai dari proses pengumpulan data, pembagian dataset, *crawling trending topics* pada *Twitter*, *text preprocessing*, pembobotan dokumen khotbah Jumat terhadap *query*, perhitungan nilai cosine similarity, penentuan relevansi rekomendasi topik dengan *query*, perolehan hasil rekomendasi, hingga proses evaluasi sil rekomendasi topik khotbah Jumat. Tahapan penelitian tersebut ditunjukkan pada gambar 1.



Gambar 1. Tahapan penelitian

#### IV. IMPLEMENTASI DAN EVALUASI

##### A. Hasil Pengumpulan Data

Pada penelitian ini dataset penelitian yang digunakan sebagai korpus adalah kumpulan dokumen khotbah Jumat. Setelah ditelusuri, tidak ada organisasi yang menyediakan dokumen khotbah Jumat dalam bentuk dataset. Pengumpulan dataset dilakukan sendiri dengan cara menelusuri situs Muhammadiyah<sup>1</sup>, Nahdlatul Ulama<sup>2</sup>, maupun Yufid<sup>3</sup> yang membagikan teks khotbah kemudian menyalinnya satu per satu ke dalam file excel. Atribut yang dicari terdiri dari isi khotbah dan link di mana khotbah tersebut didapatkan. Hasil pengumpulan dataset khotbah Jumat yang didapatkan adalah sebanyak 360 data.

##### B. Crawling

Pada penelitian ini data yang digunakan sebagai query adalah data trending topics pada Twitter. Dalam proses

<sup>1</sup> [www.suaramuhammadiyah.id/category/khutbah](http://www.suaramuhammadiyah.id/category/khutbah)

<sup>2</sup> [www.nu.or.id/post/9/khutbah](http://www.nu.or.id/post/9/khutbah)

<sup>3</sup> [www.khotbahjumat.com/category/artikel-khutbah-jumat-2](http://www.khotbahjumat.com/category/artikel-khutbah-jumat-2)

memperoleh trending topics, API key yang telah didapatkan dari Twitter akan digunakan bersamaan dengan library tweepy untuk menelusuri dan membantu dalam mengumpulkan atau mengunduh data trending topics dari database Twitter. Data hasil crawling tersebut selanjutnya diolah dan diproses menjadi query agar menghasilkan rekomendasi yang tepat.

Data trending topics Twitter yang berhasil di crawling memiliki dua bentuk kalimat berbeda. Pertama yang berbentuk tagar di mana kalimat trending topics yang terhubung tanpa spasi, kalimat trending topics yang berbentuk seperti pada contoh tersebut harus melalui proses pemecahan terlebih dahulu sebelum dilakukan proses preprocessing agar menjadi sebuah query. Bentuk kedua, kalimat trending topics yang bukan merupakan tagar dari pengguna, bentuk ini tidak perlu melalui proses pemecahan terlebih dahulu.

Library untuk melakukan proses pemecahan atau penambahan spasi pada kalimat terbatas hanya untuk kalimat berbahasa Inggris saja, maka dibuat fungsi dengan memanfaatkan file KataDasar.txt yang berisi sekumpulan kata dasar bahasa Indonesia dan bahasa Inggris. Kata dasar berfungsi untuk membantu proses pengecekan kata pada kalimat tanpa spasi sebelum dipecah.

##### C. Text Preprocessing

Pada tahapan text preprocessing, digunakan library nltk dan sastrawi untuk mengubah suatu dataset dokumen yang terdiri dari judul dan isi dokumen khotbah Jumat ke dalam bentuk kata dasar. Selain dokumen, data trending topics juga harus melalui tahapan preprocessing sebelum digunakan sebagai query. Proses yang dilakukan pada tahapan ini terdiri dari tokenizing, filtering dan stemming. Contoh hasil pemecahan kalimat menjadi kata dan memperoleh bentuk kata dasarnya dapat dilihat pada tabel 1.

TABEL 1. CONTOH PREPROCESSING TRENDING TOPICS

Trending Topics	Query
MalamJumatReceh	malam jumat receh
SilentMajorityWinElection	silent majority win election
BANGTAN MI CASA	bangtan mi casa
Memetwit	meme twit
Kasep	kasep

TABEL 2. CONTOH PREPROCESSING DOKUMEN KHOTBAH

Dokumen	Preprocessed Dokumen
8 muslim Pemakmur Bumi, Jama'ah shalat Jum'at yang dirahmati Allah.	muslim makmur bumi jamaah shalat jumat rahmat allah

Tabel 2 diperoleh melalui proses tokenizing, filtering dan stemming. Hasil dari preprocessing dokumen khotbah selanjutnya akan di simpan ke dalam file excel baru dengan nama preprocessed-dataset.xlsx. Dataset yang digunakan terdiri dari tiga atribut yaitu dokumen, link di mana dokumen khotbah dapat diunduh dan hasil preprocessing dokumen. Sampel dataset ditunjukkan pada tabel 3.

TABEL 3. SAMPEL DATASET

Dokumen	Link	Preprocessed
Muslim Pemakmur	<a href="http://www">http://www</a>	muslim



Bumi, 8 jamaah shalat Jumat yang dirahmati Allah. Pada hari ini tidak ada alasan bagi kita untuk tidak mengucapkan syukur kepada Allah SwT.	.suaramuhammadiyah.id/2017/02/16/muslim-pemakmurbumi/	makmur bumi jamaah shalat jumat rahmat allah hari alas ucap syukur allah swt
---	---	--

#### D. TF-IDF dan Cosine Similarity

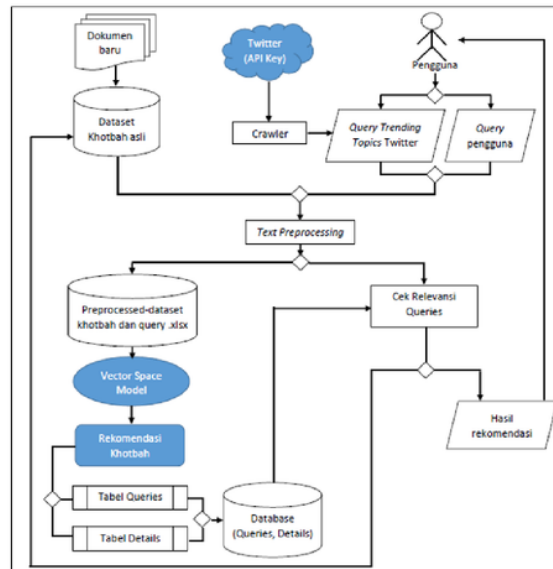
Pembobotan dokumen dilakukan dengan menggunakan library TF-IDF dari *scikit learn python*. Pembobotan dilakukan untuk menentukan bobot term pada setiap dokumen yang diperoleh dari *query* dan dataset. Setelah proses pembobotan TF-IDF, kemudian tiap term pada *query* dan dokumen k<sub>2</sub> diubah ke dalam bentuk *vector* untuk perhitungan *cosine similarity*. Hasil perhitungan *cosine similarity* akan menentukan nilai kemiripan antara *query* yang di masukan dengan dokumen khotbah yang tersedia. Sampel hasil perhitungan nilai *cosine similarity* ditunjukkan pada gambar 2.

[ 0.416, 0.194, 0.010, 0.037, 0.111, 0.023, 0.028, 0.005, 0.457, 0.045, 0.080, 0.304, 0.051, 0.005, 0.018, 0.345, 0.153, 0.038, 0.232, 0.025, 0.018, 0.021, 0.142, 0.123, 0.073, 0.106, 0.003, 0.021, 0.073, 0.277, 0.018, 0.058, 0.289, 0.052, 0.012, dst...]
--

Gambar 2. Sampel hasil perhitungan nilai *cosine similarity* dalam bentuk list

#### E. Perancangan

Adapun desain arsitektur yang diusulkan ditunjukkan seperti pada gambar 3. Bagian yang memiliki peranan penting pada arsitektur adalah pengguna. Pengguna berperan untuk memilih bagaimana cara mendapatkan rekomendasi, apakah dengan cara memasukkan *query* atau memperoleh rekomendasi berdasarkan *trending topics*. *Query* yang diperoleh baik dari proses *input* maupun *crawling*, selanjutnya akan melewati tahap *preprocessing*. Hasil dari *preprocessing query* akan digunakan untuk melakukan pengecekan kemiripan *query* pada database *history*. Jika tersedia maka sistem akan langsung memberikan rekomendasi, jika tidak tersedia maka sistem akan memproses pembobotan *query* dengan dokumen *preprocessed-dataset* menggunakan *vector space model*. Hasil rekomendasi yang menggunakan *query* baru akan disimpan pada database *history* yang terdiri dari dua tabel yaitu *queries* untuk menyimpan *query* dan tabel *details* untuk menyimpan hasil rekomendasi.



Gambar 3. Arsitektur sistem rekomendasi yang diusulkan

#### F. Implementasi

Hasil implementasi halaman rekomendasi dilakukan dengan membuat tiga menu untuk rekomendasi *input query*, rekomendasi *trending topics* dan unggah dataset. Rekomendasi dengan inputan *query* akan menampilkan data hasil rekomendasi berdasarkan *input query* dari pengguna. Sampel hasil implementasi rekomendasi inputan *query* dapat dilihat pada tabel 4 dengan *query* "pemimpin yang baik".

TABEL 4. SAMPEL REKOMENDASI BERDASARKAN INPUTAN QUERY

Dokumen	Link	Score
Document 216	<a href="https://khotbahjumat.com/2694-khotbah-jumat-memilih-pemimpin-yang-baik.html">https://khotbahjumat.com/2694-khotbah-jumat-memilih-pemimpin-yang-baik.html</a>	0.496
Document 107	<a href="https://khotbahjumat.com/3132-keshalehan-rakyat-adalah-pilar-kekuatan-negara.html">https://khotbahjumat.com/3132-keshalehan-rakyat-adalah-pilar-kekuatan-negara.html</a>	0.495
Document 55	<a href="https://khotbahjumat.com/4007-yang-terbaik.html">https://khotbahjumat.com/4007-yang-terbaik.html</a>	0.33
Document 194	<a href="https://khotbahjumat.com/2817-siapapun-presidennya-tetap-harus-ditaati.html">https://khotbahjumat.com/2817-siapapun-presidennya-tetap-harus-ditaati.html</a>	0.329
Document 358	<a href="http://www.suaramuhammadiyah.id/2017/01/26/negeri-baldatun-thoyibatun-warabun-ghofur/">http://www.suaramuhammadiyah.id/2017/01/26/negeri-baldatun-thoyibatun-warabun-ghofur/</a>	0.282

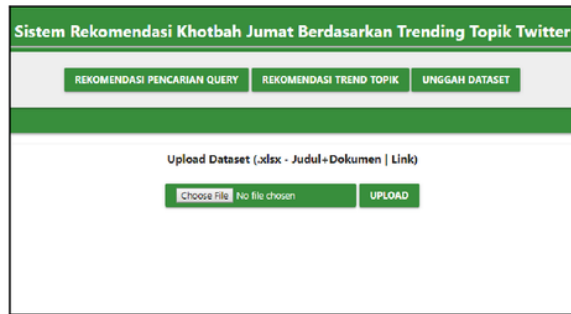
Hasil implementasi rekomendasi berdasarkan *trending topics* menampilkan lima *trending topics* teratas dan masing-masing rekomendasinya. Sampel hasil implementasi rekomendasi *trending topics* Twitter dapat dilihat pada tabel 5 dengan *trending topics* "dingin".

TABEL 5. SAMPEL REKOMENDASI BERDASARKAN TRENDING TOPICS

Dokumen	Link	Score
Document 54	<a href="https://khotbahjumat.com/4103-matahari-adalah-bukti-kekuasaan-allah.html">https://khotbahjumat.com/4103-matahari-adalah-bukti-kekuasaan-allah.html</a>	0.034
Document 328	<a href="http://www.suaramuhammadiyah.id/2016/01/28/mengambil-ibrah-dari-dua-maulid/">http://www.suaramuhammadiyah.id/2016/01/28/mengambil-ibrah-dari-dua-maulid/</a>	0.023
Document	<a href="https://khotbahjumat.com/3159-baktimu-">https://khotbahjumat.com/3159-baktimu-</a>	0.013

100	pada-dua-orang-tua.html	
Document 298	https://khotbahjumat.com/1448-persiapan-menghadapi-kematian.html	0.013
Document 73	https://khotbahjumat.com/3394-dialah-teladan-bagi-pemuda.html	0.011

Hasil implementasi halaman unggah dataset terdiri dari *input file* kemudian *upload*. Dataset yang diunggah harus dalam bentuk excel dengan format .xlsx dan terdiri dari tiga atribut yaitu dokumen dan *link*. Dataset yang berhasil diunggah akan melalui tahap *text preprocessing*, kemudian disimpan sebagai dataset. 25 implementasi halaman dari unggah dataset khotbah Jumat dapat dilihat pada gambar 4.



Gambar 4. Hasil unggah dataset khotbah

#### G. Evaluasi

Evaluasi dilakukan menggunakan teknik evaluasi *Information Retrieval* untuk yang menghasilkan ranking [7]. Setiap rekomendasi akan diberi label berdasarkan skor *cosine similarity*. Rekomendasi yang relevan akan diberi label angka satu, sedangkan yang tidak relevan akan diberi label angka nol. Rekomendasi yang bernilai lebih dari nol akan dianggap relevan, sedangkan jika bernilai sama dengan nol maka rekomendasi dianggap tidak relevan. Label akan digunakan untuk menghitung atau mengukur nilai *precision* dari data rekomendasi yang terambil. Perhitungan nilai *precision* dilakukan pada lima belas rekomendasi teratas pada setiap *query* yang dimasukkan. Dari hasil tersebut akan dihitung nilai *average precision*. Kemudian skor *average precision* setiap *query* akan dijumlah dan dibagi dengan jumlah *query* untuk mendapatkan nilai *mean average precision* (MAP). Pengujian dilakukan menggunakan 14 *query* dari *trending topics Twitter* dan 360 dokumen dataset. Sampel pengujian dapat dilihat pada tabel 6.

TABEL 6. SAMPEL PENGUJIAN MEAN AVERAGE PRECISION

Query	Q	R	NR	Avg. Precision
malam jumat rekeh	Q1	15	0	1
meme twit	Q3	1	14	0.221
siapa presiden	Q8	15	0	1
jaga tps kawal hasil suara	Q9	15	0	1
jangan curang indonesia	Q10	15	0	1
indonesia memilih	Q11	15	0	1
sambut ramadhan	Q12	15	0	1

ramadhan mubarak	Q13	15	0	1
puasa	Q14	15	0	1
jumat berkah	Q15	15	0	1
kuis sahur pandit fandom	Q16	3	12	0.496
ulin yusron kebal hukum	Q17	15	0	1
ini sahur tiket com	Q18	15	0	1
kita menang tebal	Q19	15	0	1
<b>Total</b>				12.717
<b>Mean Average Precision</b>				0.908

Keterangan :

Q = Indeks query

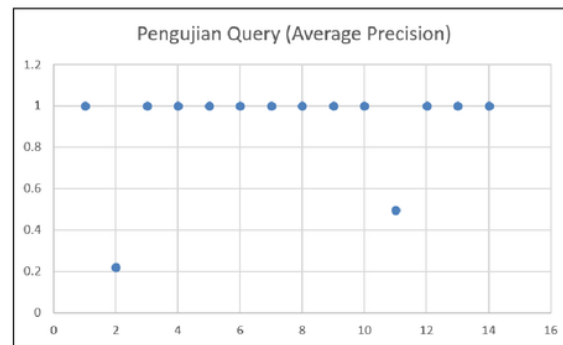
R = Banyaknya/jumlah dokumen yang relevan

NR= Banyaknya/jumlah dokumen yang tidak relevan

Pengujian dilakukan dengan formula perhitungan seperti pada persamaan 3.

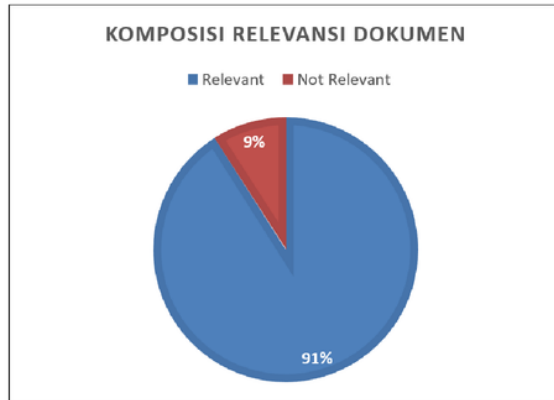
$$\begin{aligned} \text{MAP} &= (1+0.221+1+1+1+1+\dots+0.496+1+1+1) / 14 \\ &= 12.717 / 14 \\ &= 0.906 \end{aligned}$$

Berdasarkan klasifikasi keakuratan pengujian *diagnostic test* maka skor 0.9 sampai 1.00 masuk pada klasifikasi *excellent recommendation* atau dianggap sebagai rekomendasi yang sangat baik. Hasil perhitungan nilai MAP yang diperoleh pada penelitian ini adalah 0.906 atau 91% jika dibulatkan dalam bentuk persen. Hal ini menunjukkan bahwa rekomendasi yang diberikan pada sistem ini masuk pada klasifikasi sangat baik.



Gambar 5. Scatter Hasil Average Precision

Gambar 5 menunjukkan hasil nilai average precision berada pada kisaran 0 sampai 1. Semakin nilai average precision mendekati angka satu artinya semakin banyak juga rekomendasi relevan yang diberikan, begitu juga sebaliknya. Scatter tersebut menunjukkan bahwa nilai average precision sebagian besar berada di angka satu atau mendekati satu.



Gambar 6. Komposisi Relevansi Dokumen

Gambar 6 merepresentasikan hasil rekomendasi yang relevan dengan yang tidak relevan dari setiap *query trending topics*. Hasil rekomendasi relevan adalah 91% sedangkan rekomendasi yang tidak relevan adalah 9%.

## V. KESIMPULAN

Berdasarkan hasil dan pembahasan dari penelitian ini, maka dapat disimpulkan bahwa :

1. Telah dibangun sistem rekomendasi khotbah Jumat berdasarkan *trending topics Twitter* yang mampu memberikan informasi yang bermakna untuk menentukan kecocokan antara *trending topics* dengan dokumen khotbah, sehingga dapat membantu khatib dalam menemukan rekomendasi khotbah yang sesuai dengan isu yang sedang hangat di masyarakat.
2. Akurasi yang diperoleh melalui evaluasi sistem rekomendasi khotbah Jumat menggunakan metode MAP adalah sebesar 0.906 atau 91% jika dibulatkan dalam bentuk persen. Evaluasi melibatkan *query* sebanyak 14 dan dokumen sebanyak 360. Nilai MAP tersebut dinyatakan sangat baik atau masuk pada klasifikasi *excellent recommendation* pada *diagnostic test*.
3. Penerapan *Vector Space Model* pada penelitian ini telah menghasilkan rekomendasi yang baik dengan keakuratan yang tinggi. Namun, bagian rekomendasi berdasarkan *trending topics Twitter* seringkali memberikan rekomendasi yang tidak relevan. Kasus tersebut terjadi karena tidak semua *trending topics* dapat di-*preprocessing* menjadi *query*. Hal ini terjadi karena *trending topics* tersebut berbentuk *slang* atau memang tidak berkaitan dengan dokumen khotbah Jumat. Faktor lain dikarenakan jumlah dataset khotbah Jumat yang terbatas yaitu sebanyak 360 dokumen.

## REFERENSI

- [1] M. Tahir, "Khotbah Jum'at Di Kota Samarinda (Analisis Kesiapan para Khotib Di Kota Samarinda),"

- [2] *MENA*, vol. V, no. 2, pp. 229–240, 2013.
- [2] E. J. Thaib, "Analisis Minat Jamaah Masjid Terhadap Penyampaian Khutbah Jumat Di Kecamatan Talaga Jaya Kabupaten Gorontalo," *J. Madani*, vol. 4, pp. 75–94, 2014.
- [3] A. Kao and S. Poteet, "Text mining and natural language processing – introduction for the special issue," *SIGKDD Explor. Newsl*, vol. 7, no. 1, pp. 1–18, 2005.
- [4] B. Herwijayanti, D. E. Ratnawati, and L. Muflikah, "Klasifikasi berita online dengan menggunakan pembobotan TF-IDF dan cosine similarity," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018.
- [5] R. A. Siregar, F. Amelia Sinaga, and R. Arianto, "Aplikasi penentuan dosen penguji skripsi menggunakan metode TF-IDF dan vector space model," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 2, pp. 171–186, 2017.
- [6] G. Salton, *Automatic text preprocessing*. New York: Addison-Wesley Publishing Company, 1989.
- [7] P. R. and H. S. Manning Christopher D., "manning2008-Introduction to information retrieval.pdf," 2008.
- [8] W. S. J. Saputra and F. Muttaqin, "Pengenalan karakter pada proses digitalisasi dokumen menggunakan cosine similarity," *Semin. Nas. Tek. Inform.*, no. September, pp. 51–56, 2013.
- [9] O. Karmayasa and I. B. Mahendra, "Implementasi vector space model dan beberapa notasi metode term frequency inverse document frequency (TF- IDF) pada sistem temu kembali informasi," *J. Elektron. Ilmu Comput. Univ. Udayana*, vol. 1, 2012.
- [10] K. P. N. V Satya and J. V. R. Murthy, "Clustering based on cosine similarity measure," no. 3, pp. 508–514, 2012.
- [11] R. Feldman and J. Sanger, *The text mining handbook*. New York: Cambridge University Press, 2007.
- [12] A. Tan, "Text mining: The state of the art and the challenges concept-based," *Proc Pacific Asia Conf Knowl. Discov. Data Min. PAKDD '99 Work. Knowl. Discov. from Adv. Databases*, 1999.

# Rekomendasi Topik Khotbah Jumat Berdasarkan Twitter Trending Topics Menggunakan Vector Space Model

## ORIGINALITY REPORT

23%

SIMILARITY INDEX

20%

INTERNET SOURCES

7%

PUBLICATIONS

14%

STUDENT PAPERS

## PRIMARY SOURCES

1

[media.neliti.com](http://media.neliti.com)

Internet Source

4%

2

Submitted to Universitas Brawijaya

Student Paper

3%

3

[ejournals.umn.ac.id](http://ejournals.umn.ac.id)

Internet Source

2%

4

[journal.iain-samarinda.ac.id](http://journal.iain-samarinda.ac.id)

Internet Source

2%

5

[ejournal.undip.ac.id](http://ejournal.undip.ac.id)

Internet Source

1%

6

[jutei.ukdw.ac.id](http://jutei.ukdw.ac.id)

Internet Source

1%

7

Alfirna Rizqi Lahitani, Adhistya Erna Permanasari, Noor Akhmad Setiawan. "Cosine similarity to determine similarity measure: Study case in online essay assessment", 2016 4th International Conference on Cyber and IT Service Management, 2016

1%



---

8	<a href="http://www.suaramuhammadiyah.id">www.suaramuhammadiyah.id</a> Internet Source	1%
9	<a href="http://eprints.uns.ac.id">eprints.uns.ac.id</a> Internet Source	1%
10	Tao Jiang. "", IEEE Transactions on Knowledge and Data Engineering, 2/2007 Publication	1%
11	<a href="http://tubagusisnaeni.blogspot.com">tubagusisnaeni.blogspot.com</a> Internet Source	1%
12	<a href="http://digilib.uinsby.ac.id">digilib.uinsby.ac.id</a> Internet Source	1%
13	<a href="http://aieluwuk.blogspot.com">aieluwuk.blogspot.com</a> Internet Source	<1%
14	<a href="http://ejournal.poltektegal.ac.id">ejournal.poltektegal.ac.id</a> Internet Source	<1%
15	<a href="http://jurnaldampak.ft.unand.ac.id">jurnaldampak.ft.unand.ac.id</a> Internet Source	<1%
16	<a href="http://ilkom.unnes.ac.id">ilkom.unnes.ac.id</a> Internet Source	<1%
17	<a href="http://eprints.upnjatim.ac.id">eprints.upnjatim.ac.id</a> Internet Source	<1%
18	Nurkholis, Meila Zuwari Negara, Guruh Fajar Shidik, A. Zainul Fanani, Muljono, Edi	<1%

---

Noersasongko. "Sentiment Analysis of Indonesian News Using Deep Learning (Case Study: TVKU Broadcast)", 2018 International Seminar on Application for Technology of Information and Communication, 2018

Publication

---

19 [hardianti1411.blogspot.com](http://hardianti1411.blogspot.com) <1 %  
Internet Source

---

20 [ml.scribd.com](http://ml.scribd.com) <1 %  
Internet Source

---

21 Submitted to Politeknik Negeri Bandung <1 %  
Student Paper

---

22 [journal.uad.ac.id](http://journal.uad.ac.id) <1 %  
Internet Source

---

23 Submitted to Binus University International <1 %  
Student Paper

---

24 [internationalscienceindex.org](http://internationalscienceindex.org) <1 %  
Internet Source

---

25 [www.scribd.com](http://www.scribd.com) <1 %  
Internet Source

---

26 [es.scribd.com](http://es.scribd.com) <1 %  
Internet Source

---

27 [de.scribd.com](http://de.scribd.com) <1 %  
Internet Source

---

28

[digilib.uin-suka.ac.id](http://digilib.uin-suka.ac.id)

Internet Source

<1%

---

29

[www.devieriana.com](http://www.devieriana.com)

Internet Source

<1%

---

Exclude quotes Off

Exclude matches Off

Exclude bibliography On