

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

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

Abstrak—Topik khotbah Jumat yang ideal adalah yang sesuai dengan isu atau topik yang sedang terjadi 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* 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 atau topik yang sedang terjadi 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 khatib 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*

I. PENDAHULUAN

Khotbah yang ideal tentu bergantung pada khatibnya, mulai dari pengambilan topik sampai dengan penyampaian khotbah [1]. Berdasarkan studi pendahuluan dengan melakukan wawancara dengan khatib di Yogyakarta diketahui bahwa pemilihan topik khotbah Jumat ditentukan hanya berdasarkan materi khotbah yang dikuasai khatib saja. Khatib jarang menyampaikan topik khotbah yang berhubungan dengan isu atau topik yang sedang terjadi di masyarakat, dikarenakan keterbatasan waktu dalam penentuan topik yang sesuai. Sebenarnya, kumpulan dokumen khotbah Jumat telah tersedia, akan tetapi khatib masih merasa kesulitan dalam mengelola dokumen khotbah Jumat tersebut. Di sisi lain, kumpulan dokumen khotbah Jumat masih belum dikelompokkan berdasarkan kategori tertentu sehingga mempersulit proses penentuan topik khotbah Jumat yang akan dibawakan.

*Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2019
Yogyakarta, 03 Agustus 2019*

Dalam memilih tema khotbah, seorang khatib harus mengambil tema yang bersifat konsumtif yakni tema khotbah yang disampaikan harus betul-betul dirasakan sebagai kebutuhan jemaah yang mendesak [2]. Mempertahankan antusias jemaah dalam mendengarkan uraian khotbah juga menjadi tantangan lain bagi khatib. Seringkali terjadi, jemaah kurang memperhatikan khotbah bahkan sampai tertidur [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 atau topik yang sedang terjadi di masyarakat. Seyogyanya, dari isu atau topik yang sedang terjadi 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.

Di sisi 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 kontesktual 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

ISSN: 1907 – 5022

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 atau topik yang sedang terjadi 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* informasi bibliografi maupun data lainnya [6].

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 [7]. 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 [8]. Proses pembobotan kata dijabarkan pada persamaan (1).

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

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 algoritma untuk mengukur kemiripan antara dua vektor n dimensi dengan mencari kosinus sudut antara dua vektor tersebut, *cosine similarity* sering digunakan untuk membandingkan dokumen dalam penambangan teks [8]. Proses 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

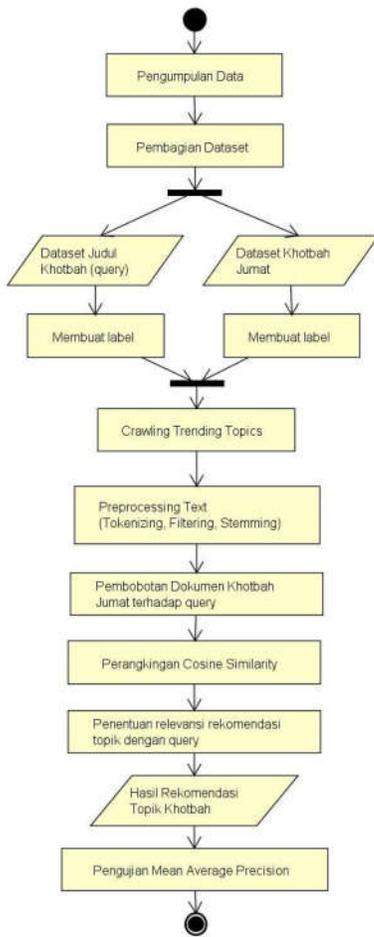
Cosine similarity antara vektor *query* dan vektor dokumen dapat digunakan sebagai ukuran skor dokumen untuk *query* tersebut, nilai yang dihasilkan kemudian dapat digunakan untuk memilih penilaian atas dokumen untuk *query* [9].

C. Text Mining

Text mining adalah proses analisis data dalam bentuk teks. Sumber data *text mining* diperoleh dari dokumen [10]. Tahapan *text mining* meliputi pengumpulan data dan ekstraksi fitur [10]. 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 hasil rekomendasi topik khotbah Jumat. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

IV. IMPLEMENTASI DAN EVALUASI

D. Hasil Pengumpulan Data

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 Muhammadiyah¹, Nahdlatul ‘Ulama², maupun Yufid³ yang membagikan teks khotbah kemudian menyimpannya satu per satu ke dalam file excel. Atribut yang dicari terdiri dari isi khotbah dan *link* khotbah tersebut didapatkan. Hasil pengumpulan dataset khotbah Jumat yang didapatkan adalah sebanyak 360 data.

E. Crawling

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

¹ www.suaramuhammadiyah.id/category/khutbah

² www.nu.or.id/post/9/khutbah

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 yaitu 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.

F. 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*. Tahapan *text preprocessing* yang dilakukan meliputi tokenisasi, *filtering* dan *stemming*. Contoh hasil pemecahan kalimat menjadi kata dan memperoleh bentuk kata dasarnya ditunjukkan pada Tabel 1.

TABEL I. CONTOH *PREPROCESSING* *TRENDING TOPICS*

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

TABEL II. CONTOH *PREPROCESSING* DOKUMEN KHOTBAH

Dokumen	Preprocessed Dokumen
Muslim Pemakmur Bumi, Jemaah shalat Jumat yang dirahmati Allah.	muslim makmur bumi jemaah shalat jumat rahmat allah

Tabel 2 diperoleh melalui proses *tokenizing*, *filtering* dan *stemming*. Hasil dari *preprocessing* dokumen khotbah selanjutnya akan disimpan ke dalam file excel baru dengan nama *preprocessed-dataset.xlsx*. Dataset yang digunakan terdiri dari tiga atribut yaitu dokumen, *link*, dan *preprocessed*. *Preprocessed* pada dataset menunjukkan hasil *preprocessing*

³ www.khotbahjumat.com/category/artikel-khutbah-jumat-2

dokumen. *Link* pada dataset menunjukkan alamat tautan dokumen agar bisa diunduh. Sampel dataset ditunjukkan pada Tabel 3.

TABEL III. SAMPEL DATASET

Dokumen	Link	Preprocessed
Muslim Pemakmur Bumi, Jemaah shalat Jumat yang dirahmati Allah Pada hari ini tidak ada alasan bagi kita untuk tidak mengucapkan syukur kepada Allah SwT.	http://www.suaramuhammadiyah.id/2017/02/16/muslim-pemakmur-bumi/	muslim makmur bumi jemaah shalat jumat rahmat allah hari alas ucap syukur allah swt

G. TF-IDF dan Cosine Similarity

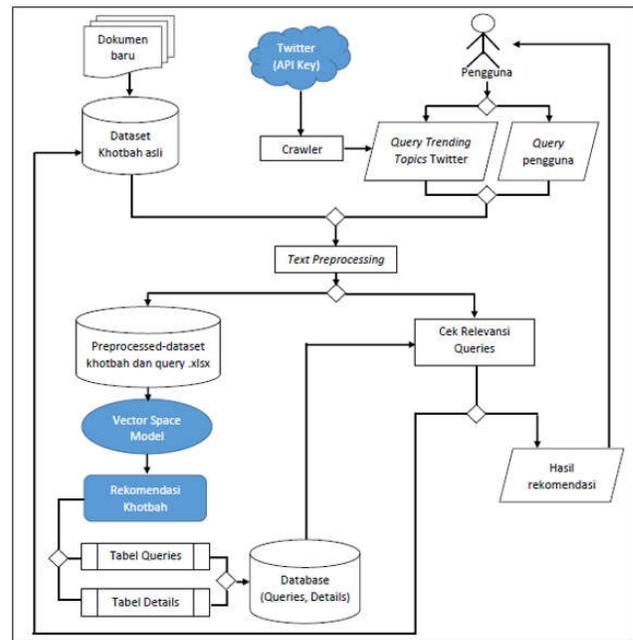
Pembobotan dokumen dilakukan dengan menggunakan *library* TF-IDF dari *scikit learn* python. Pembobotan dilakukan untuk menentukan bobot term pada setiap dokumen yang diperoleh dari *query* dan dataset. Setelah proses pembobotan TF-IDF, kemudian tiap term pada *query* dan dokumen khotbah diubah ke dalam bentuk *vector* untuk menghitung nilai *cosine similarity*. Nilai *cosine similarity* menentukan nilai kemiripan antara *query* yang dimasukkan 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

H. 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 crawling*, selanjutnya akan melewati tahap *preprocessing*. Hasil dari *preprocessing query* akan digunakan untuk melakukan pengecekan kemiripan *query* pada database *history*. Jika hasil *preprocessing query* tersedia di database *history*, maka sistem akan langsung memberikan rekomendasi. Jika hasil *preprocessing query* tidak tersedia di database *history*, 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

I. 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 IV. SAMPEL REKOMENDASI BERDASARKAN INPUTAN QUERY

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

TABEL V. SAMPEL REKOMENDASI BERDASARKAN *TRENDING TOPICS*

Dokumen	Link	Score
Document 327	http://www.suaramuhammadiyah.id/2016/01/21/muslim-dan-budaya-utama/	0.03
Document 180	https://khotbahjumat.com/2864-adab-seorang-mukmin-kepada-sahabat-nabi.html	0.016
Document 4	https://khotbahjumat.com/5223-pemimpin-yang-mmenyesatkan.html	0.006
Document 1	https://khotbahjumat.com/5254-kekeliruan-dalam-masalah-syarat-sah-shalat.html	0
Document 2	https://khotbahjumat.com/5251-allah-maha-mengetahui-apa-yang-kita-kerjakan.html	0

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. Implementasi halaman dari unggah dataset khotbah Jumat ditunjukkan pada Gambar 4.



Gambar 4. Implementasi halaman unggah dataset khotbah

J. Evaluasi

Evaluasi dilakukan menggunakan teknik evaluasi *Information Retrieval* untuk yang menghasilkan ranking [9]. 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 dijumlahkan 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 evaluasi MAP ditunjukkan pada Tabel 6.

TABEL VI. SAMPEL EVALUASI *MEAN AVERAGE PRECISION*

Query	Q	R	NR	Avg. Precision
malam jumat receh	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
Total				12.717
Mean Average Precision				0.906

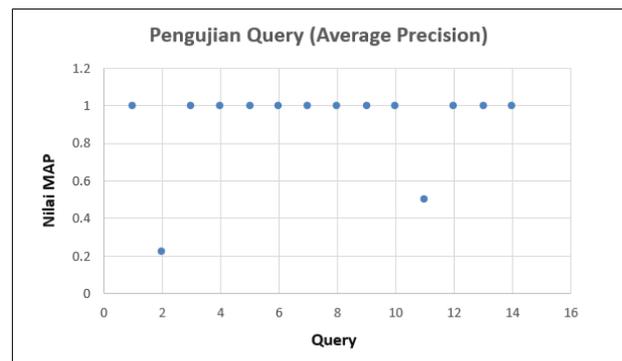
Keterangan :

Q = Indeks *query*

R = Jumlah dokumen yang relevan

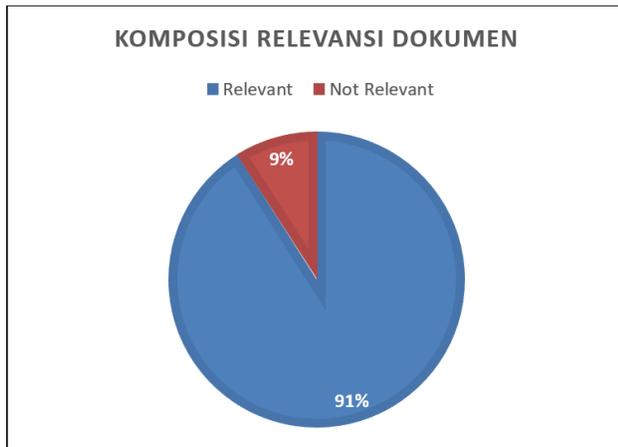
NR= Jumlah dokumen yang tidak relevan

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 pada 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

Dari hasil penelitian yang telah dilakukan, diperoleh kesimpulan 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 atau topik yang sedang terjadi 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 Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2019 Yogyakarta, 03 Agustus 2019

para Khotib Di Kota Samarinda)," *FENOMENA*, vol. V, no. 2, pp. 229–240, 2013.

[2] E. J. Thaib, "Analisis Minat Jamaah Masjid Terhadap Penyampaian Khotbah 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. 1–2, 2005.

[4] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "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. 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] 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.

[8] K. P. N. V Satya and J. V. R. Murthy, "Clustering based on cosine similarity measure," no. 3, pp. 508–512, 2012.

[9] P. R. and H. S. Manning Christopher D., "manning2008-Introduction to information retrieval.pdf." 2008.

[10] R. Feldman and J. Sanger, *The text mining handbook*. New York: Cambridge University Press, 2007.