RANKING INDEX BERITA NEW NORMAL DENGAN METODE INFORMATION RETRIEVAL MENGGUNAKAN VECTOR SPACE MODEL

Nandang Suwela
Informatika, Universitas Indraprasta PGRI
suwela.nandang@gmail.com

Submitted June 16, 2020; Revised July 25, 2020; Accepted August 1, 2020

Abstrak

Setelah selama tiga bulan melaksanakan Pembatasan Sosial Berskala Besar (PSBB), Indonesia sedang bersiap untuk memasuki tatanan kehidupan baru atau disebut dengan *new normal*, suatu hal baru yang terpaksa harus dijalani oleh semua penduduk di berbagai negara termasuk Indonesia. Tatanan *new normal* menjadi hal yang menarik untuk dibicarakan sehingga banyak masyarakat yang mencari berita tentang hal tersebut untuk mengetahui lebih jauh apa, bagaimana *new normal* dan kapan persisnya *new normal* diberlakukan. Tujuan dari penelitian ini adalah untuk menghitung keakuratan pencarian berita dengan cara menghitung *score rank* sehingga kata kunci yang digunakan bisa mendapatkan hasil seperti yang diinginkan. Penelitian ini menggunakan metode eksperimen dengan mengumpulkan berita-berita yang berada di dunia maya. Salah satu cara untuk mendapatkan berita yang sesuai adalah dengan membuat *Ranking*. Ranking adalah salah satu cabang dari Information Retrieval System dan salah satu model yang dapat digunakan untuk mengatasi permasalahan tersebut adalah *Vector Space Model*. Dataset yang digunakan adalah berita-berita *corona* yang melalui tahapan *tokenizing*, *filtering* dan *stemming* pada proses *preprocessing*. Hasil penelitian membuktikan bahwa *Vector Space Model* (*VSM*) dapat digunakan dan dapat membantu untuk mencari serta menghitung *score rank* dari sebuah berita.
Kata Kunci : *New Normal, Vector Space Model, Relevan,* PSBB

Abstract

*After three months of large scale social distancing major (*PSBB*), Indonesia is getting ready to enter a new order or commonly known as new normal, a new thing that many countries including Indonesia are forced to deal with. The new normal has become a rather interesting thing to discuss that many people are browsing for news orbiting around such topic regarding of what it is, how it is done and when exactly it's implemented. The initial of this research is to calculate the accuracy of news browsing by computing the score rank so that the keywords that we use can acquire the results as we wanted. This research is using an experiment method in which gathering the news in the virtual world. .One of the ways to acquire the right documents is to make a ranking. Ranking is one of the branches from Information Retrieval System and one of the models that we can use to deal with such problem is the Vector Space Model. The dataset that we use is the Corona-related news through tokenizing, filtering, and stemming on preprocessing. The research results in proving that Vector Space Model (VSM) can be used and can help finding and computing the score rank of the news.*
Keywords : *New Normal, Vector Space Model, Relevan,* PSBB.

1. PENDAHULUAN

Pandemi corona yang sedang terjadi di seluruh dunia, ternyata berakibat sangat luas, karena bukan hanya berakibat pada bidang kesehatan tetapi juga berpengaruh pada bidang ekonomi secara luas. Kebijakan yang dilakukan pemerintah adalah untuk menekan penularan yang terjadi di masyarakat, mulai dari penutupan kantor, sekolah, pasar, mal, tempat rekreasi sampai pada penutupan rumah ibadah. Ternyata penutupan tempat-tempat tersebut

berdampak serius, salah satunya adalah imbas yang dirasakan oleh dunia pendidikan. Penutupan sekolah bersamaan dengan akan diselenggarakannya UN (Ujian Nasional) untuk tingkat dasar dan menengah serta ujian masuk untuk perguruan tinggi. Penutupan sekolah barakibat ditiadakannya UN ditingkat pendidikan dasar dan menengah sehingga seleksi masuk kejenjang berikutnya dilakukan dengan besaran umur calon siswa, hal ini menuai protes dari orang tua murid karena dirasa tidak adil. Pemerintah tetap dengan keputusannya karena salah satu tujuannya adalah ingin menghilangkan sekolah-sekolah favorit. Cara belajarpun ikut berubah, dari pembelajaran tatap muka menjadi pembelajaran jarak jauh melalui internet (daring). Banyak seminar dan rapat yang juga dilakukan secara daring. Selain penutupan di berbagai tempat, usaha pemerintah juga dibarengi dengan anjuran untuk menggunakan masker dan juga menjaga jarak jika harus bertemu dengan orang lain. Pengaruh pada bidang ekonomi terjadi karena diberlakukan kebijakan Pembatasan Sosial Berskala Besar (PSBB) sebagai upaya untuk menghentikan atau paling tidak menghambat penularan yang terus terjadi dengan cepat. Kebijakan PSBB berakibat terhentinya semua kegiatan masyarakat dan secara otomatis menghentikan putaran roda perekonomian. Setelah lebih kurang 3 bulan menerapkan PSBB, dan diklaim bahwa tingkat penyebaran terus menurun, pemerintah berencana untuk melonggarkan pembatasan dan akan masuk pada fase tatanan kehidupan baru atau yang disebut *new normal*. Fase ini menarik perhatian masyarakat dan banyak dari mereka yang berusaha mencari informasi untuk mengetahui lebih jauh apa dan bagaimana kehidupan pada fase *new normal*. Dengan diterapkannya fase baru ini, maka akan merubah kebiasaan hidup masyarakat. Banyak berita tersebar di dunia maya tentang hal ini, dan masalah mulai muncul ketika pencarian yang dilakukan tidak mendapat berita yang diharapkan, artinya berita yang diperoleh kurang relevan dengan berita yang diharapkan muncul dan kata kunci yang digunakan terkadang tidak mendapatkan hasil seperti yang diinginkan. Seringkali pada web, dimana kita mencari suatu informasi tertentu, banyak hal yang penting justru terlewatkan, malah yang tidak penting banyak terserap.[7]

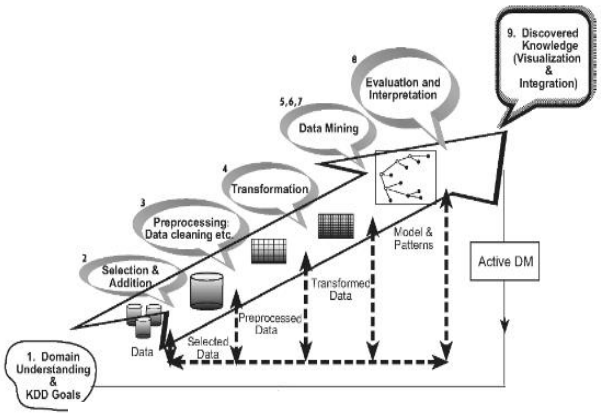
Pengertian Data *Mining*

Data *mining* merupakan teknologi yang kompleks dan berakar pada banyak disiplin ilmu: matematika, statistik, ilmu komputer, fisika, teknik, biologi, dll., serta beragam aplikasi yang berbeda domain : bisnis, perawatan kesehatan, sains dan teknik, dll. Pada dasarnya data *mining* dapat dilihat sebagai ilmu mengeksplorasi dataset besar untuk mendapatkan informasi yang tersirat, yang tidak dikenal sebelumnya dan berpotensi mempunyai manfaat.[1]

Text Mining merupakan suatu bidang khusus dari Data *Mining*. *Text mining* dapat didefinisikan sebagai suatu proses menggali informasi dimana seseorang user berinteraksi dengan sekumpulan dokumen menggunakan *tool analisis* yang merupakan komponen-komponen dalam data *mining*.[8]

Banyak yang menggunakan data *mining* sebagai istilah popular dari *Knowledge Discovery in Database* (KDD). Data *mining* merupakan inti dari proses *Knowledge Discovery in Database* (KDD). [2]

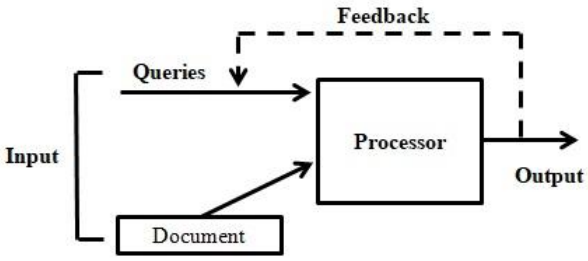
KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah *data set* yang besar dan kompleks. [2]



Sumber : Maimon. 2010

Gambar 1. Proses KDD

Salah satu cara untuk mendapatkan kembali informasi yang terdapat dalam dokumen yang sesuai dengan kebutuhan pengguna adalah dengan melakukan *Ranking*. *Ranking* merupakan salah satu cabang ilmu dari *information retrieval*.
Ranking merupakan bagian penting dari masalah pencarian informasi, seperti pengambilan dokumen, penyaringan informasi, penempatan iklan *online*, dan lain-lain.[10]
Information Retrieval akan menemukan



Sumber : C.J. Rijsbergen, 1979

Gambar 2. Diagram Alur Information Retrieval

Tujuan Information Retrieval adalah untuk mengambil semua dokumen yang *relevant* dan dalam waktu yang bersamaan mengambil sesedikit mungkin dokumen yang tidak *relevant*. Salah satu model yang dapat digunakan adalah *VSM (Vector Space Model)* yaitu suatu metode untuk melihat tingkat kedekatan atau kesamaan *(similarity)* term dengan cara melakukan pembobotan *term* menggunakan metode pembobotan TF-IDF. Dokumen dan kata kunci dipandang sebagai sebuah vector yang memiliki jarak dan arah. Relevansi sebuah dokumen kesebuah kata kunci didasarkan pada similaritas diantara *vector* dokumen dan *vector* kata kunci.[4]. Sistem memiliki dua pekerjaan yaitu, yaitu melakukan *pre-processing* terhadap database dan kemudian menerapkan metode tertentu untuk menghitung kedekatan relevansi atau similarity antara dokumen di dalam database yang telah di*preprocess* dengan *query* pengguna.[9]

2. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen adapun tujuannya adalah untuk mengukur akurasi algoritma VSM. Penelitian ini juga bertujuan untuk menghitung *score rank* dari sebuah dokumen. Tahapan dalam penelitian dimulai dengan mengumpulkan dokumen berupa berita-berita yang banyak terdapat didunia maya. Dokumen tersebut akan melalui proses tokenisasi, filtering dan stemming.

Pengumpulan Data

Data adalah hal yang sangat penting dalam suatu penelitian, jumlah dan variasi data dapat mempengaruhi hasil dari penelitian tersebut. Salah satu metode pengumpulan data adalah dengan metode literasi yaitu cara pengumpulan data yang menggunakan data sekunder. Dalam penelitian ini data yang digunakan adalah data sekunder yang diambil dari kumpulan berita yang beredar di media internet.

Preprosesing

Tahap data *preparation* merupakan tahap dengan proses penyiapan data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam

penelitian. Dalam text mining tahapan awal yang akan dilakukan adalah tahap *preprocessing*.[6]
Preprosesing yaitu tahapan proses yang bertujuan untuk membentuk sebuah database yang bersumber dari kumpulan data sekunder yang tidak terstruktur hingga siap untuk diolah. Tahapan ini terbagi menjadi 3, yaitu :

1. Tokenisasi
   Tokensasi adalah proses pembacaan dokumen serta memisahkan kalimatnya menjadi *token* (kata tunggal). Kemudian proses ini juga menghilangkan *special character* seperti tanda baca serta merubah semua hurufnya menjadi huruf kecil. Tujuan dari tokenisasi adalah untuk mendapatkan kata yang unik dari sebuah dokumen sehingga dapat dihitung frekuensi dan mendapatkan bobot (*weight*) dari kata tersebut.

   Contoh Tokenisasi :
   "Tidak!! Itu tidak boleh dikerjakan."

   Proses Tokenisasi :
   a. Hilangkan *special character*,
      Tidak Itu tidak boleh dikerjakan

   b. Pisahkan menjadi kata tunggal dan unik,
      Tidak, itu, boleh, dikerjakan

   c. Ubah menjadi huruf kecil,
      tidak, itu, boleh, dikerjakan

   Sehingga hasil dari proses ini mengubah kalimat tersebut menjadi 4 kata, yaitu :
   tidak, itu, boleh, dikerjakan

2. Filtering
   Pada tahap ini dilakukan penghilangan kata penghubung atau kata yang sering muncul (*Stopword Removal*).
   Hasil tokenisasi diatas adalah kata-kata:
   tidak, itu, boleh, dikerjakan
   *Stopword*nya adalah kata "itu", sehingga hasil filtering menjadi :
   tidak, boleh, dikerjakan

3. Stemming
   Algortima yang digunakan adalah *Porter Stemmer* yaitu proses pencarian kata dasar. Dalam bahasa Indonesia biasanya menggunakan kaidah prefix + kata dasar + suffix.

Sehingga hasil akhir dari preprosesing pada contoh diatas adalah kata :
tidak, boleh, kerja

Inverted Index
Inverted index adalah salah satu proses untuk mengideksan sebuah koleksi teks yang digunakan untuk mempercepat proses pencarian. Dalam dokumen *inverted index* didapat dari proses *preprocessing* yaitu setelah proses *tokenization, stopword dan stemming* dilakukan.[5]
Hasil dari preprosesing kemudian dibuatkan index, yang bertujuan untuk mempermudah dan mempercepat pencarian kata dari sebuah text dokumen. Tahapan berikutnya adalah mencari kata didalam dokumen kemudian dihitung jumlahnya.

Contoh :
Kata-1 (T1) ada didalam dokumen-1 (D1), dokumen-2 (D2) dan dokumen-3 (D3), sedangkan kata-2 (T2) ada didalam dokumen-1 (D1) dan dokumen-3 (D3), maka *inverted index* yang dihasilkan adalah :
T1 → D1, D2, D3
T2 → D1, D3 (3)

***Term Frequency-Inverse Document Frequency* (TF-IDF)**
TF-IDF adalah algoritma yang digunakan untuk menghitung *Weight* (bobot) dari sebuah dokumen yang relevan dengan kata kunci yang digunakan.
TF (*Term Frequency*) adalah jumlah kata (*Term*) yang sama yang muncul dalam sebuah dokumen.

IDF (*Inverse Document Frequency*) adalah pengukuran jumlah frekuensi suatu kata (*Term*) dalam sekumpulan dokumen.
TF-IDF dituliskan dengan notasi :

$idf = \log(D/df)$ (1)

dimana :
D = jumlah dokumen
df = Frekuensi *Term* dalam dokumen

Perhitungan *Weight* (bobot) menggunakan notasi :

$Wd,f = tfd,t * IDF$ (2)

dimana :
d = Dokumen ke-d dari dokumen yang ada di basis data
t = Kata (Term) ke-t dari kata kunci
tf = Banyak kata yang dicari dalam dokumen
W= Bobot dokumen ke-d terhadap kata kunci ke-t

3. HASIL DAN PEMBAHASAN

Dalam pembahasan hasil penelitian diperlukan adanya pemahaman awal tentang tujuan dari penelitian. Tujuan dari penelitian ini adalah mencari penghitungan ranking index berita new normal, sehingga dapat diketahui berita mana yang lebih relevan dengan kata kunci yang digunakan.

Dataset yang digunakan adalah dokumen – dokumen berita tentang corona, covid-19 dan *new normal* yang mempunyai kata kunci yang sama.

Preprosesing

Dari kumpulan berita tersebut kemudian dilakukan proses tokenisasi, dengan menghapus *special character*, mencari kata tunggal yang unik dan mengubahnya menjadi huruf kecil. Proses dilanjutkan dengan melakukan *filtering* dengan menjalankan proses *stopword removal* yaitu menghilangkan kata penghubung serta kata yang sering keluar. Proses selanjutnya adalah stemming yaitu mencari kata dasar dari kalimat yang ada, sehingga menghasilkan data sebagai berikut :

Tabel 1. Preprosesing

Token	*tf*					
	Q	D1	D2	D3	D4	D5
masyarakat	0	1	0	0	0	0
bingung	0	1	0	0	0	0
mudik	0	1	0	0	0	0
kantong	0	1	0	0	0	0
sikm	0	1	0	0	0	0
jakarta	0	1	0	1	1	0
new	0	1	0	1	1	0
publik	0	1	0	0	0	0
update	0	0	1	0	0	0
kabar	0	0	1	0	0	0
baik	0	0	1	0	0	0
tangan	0	0	1	0	0	0
virus	0	0	1	0	0	1
corona	0	0	1	0	1	1
indonesia	0	0	1	0	0	1
anies	0	0	0	1	0	0
baswedan	0	0	0	1	0	0
terap	0	0	0	1	0	0
psbl	0	0	0	1	0	0
usai	0	0	0	1	0	0
psbb	0	0	0	1	0	0
khusus	0	0	0	1	0	0
zona	0	0	0	1	0	0
merah	0	0	0	1	0	0
hadap	0	0	0	0	1	0
kasus	0	0	0	0	1	1
tentu	0	0	0	0	1	0
provinsi	0	0	0	0	0	1
aktif	0	1	0	0	0	1
kecil	0	0	0	0	0	1
diy	0	0	0	0	0	1
urut	0	0	0	0	0	1
normal	1	1	0	1	1	0

Sumber : data diolah. 2020

Dimana :

token = kata hasil preprosesing

tf = frekuensi dari setiap token di semua dokumen

***Term Frequency-Inverse Document Frequency* (TF-IDF)**

Setelah diketahui frekuensi tiap *term* dari semua dokumen, kemudian dihitung IDF (*Inverse Document Frequency*), dengan hasil sebagai berikut :

Tabel 2. Perhitungan *idf*

df	D/*df*	*idf* = log (D/*df*)
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
3	1.666667	0.22184875
3	1.666667	0.22184875
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
2	2.5	0.397940009
3	1.666667	0.22184875
2	2.5	0.397940009
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
2	2.5	0.397940009
1	5	0.698970004
1	5	0.698970004
2	2.5	0.397940009
1	5	0.698970004
1	5	0.698970004
1	5	0.698970004
3	1.666667	0.22184875

Sumber : data diolah, 2020

Weight

Perhitungan bobot (*Weight*) dilakukan dengan cara mengalikan *Term Frequency* (*tf*) dengan Inverse Document Frequency (*idf*), sehingga didapat hasil sebagai berikut :

Tabel 3. Perhitungan *Weight*

Weight					
Q	D1	D2	D3	D4	D5
0	0.69897	0	0	0	0
0	0.69897	0	0	0	0
0	0.69897	0	0	0	0
0	0.69897	0	0	0	0
0	0.69897	0	0	0	0
0	0.221849	0	0.221849	0.221849	0
0	0.221849	0	0.221849	0.221849	0
0	0.69897	0	0	0	0
0	0	0.69897	0	0	0
0	0	0.69897	0	0	0
0	0	0.69897	0	0	0
0	0	0.69897	0	0	0
0	0	0.39794	0	0	0.39794
0	0	0.221849	0	0.221849	0.221849
0	0	0.39794	0	0	0.39794
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0.69897	0	0
0	0	0	0	0.69897	0
0	0	0	0	0.39794	0.39794
0	0	0	0	0.69897	0
0	0	0	0	0	0.69897
0	0.39794	0	0	0	0.39794
0	0	0	0	0	0.69897
0	0	0	0	0	0.69897
0	0	0	0	0	0.69897
0.221849	0.221849	0	0.221849	0.221849	0

Sumber : data diolah, 2020

Akar Pangkat 2

Sebelum melakukan penghitungsn *score ranking*, terlebih dahulu dilakukan penghitungan akar pangkat dua dari masing-masing *term* dari semua dokumen yang ada, sehingga menghasilkan data sebagai berikut :

Tabel 3. Perhitungan Akar Pangkat Dua

	Q^2	D1^2	D2^2	D3^2	D4^2	D5^2
	0	0.488559	0	0	0	0
	0	0.488559	0	0	0	0
	0	0.488559	0	0	0	0
	0	0.488559	0	0	0	0
	0	0.488559	0	0	0	0
	0	0.049217	0	0.049217	0.049217	0
	0	0.049217	0	0.049217	0.049217	0
	0	0.488559	0	0	0	0
	0	0	0.488559	0	0	0
	0	0	0.488559	0	0	0
	0	0	0.488559	0	0	0
	0	0	0.488559	0	0	0
	0	0	0.158356	0	0	0.15835625
	0	0	0.049217	0	0.049217	0.04921687
	0	0	0.158356	0	0	0.15835625
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0.488559	0	0
	0	0	0	0	0.488559	0
	0	0	0	0	0.158356	0.15835625
	0	0	0	0	0.488559	0
	0	0	0	0	0	0.48855907
	0	0.158356	0	0	0	0.15835625
	0	0	0	0	0	0.48855907
	0	0	0	0	0	0.48855907
	0	0	0	0	0	0.48855907
	0.049217	0.049217	0	0.049217	0.049217	0
Q	0.049217	Sqrt(Di)				
Sqrt(Q)	0.221849					
		1.799267	1.523209	2.131826	1.154271	1.62384671

Sumber : data diolah, 2020

The Score Ranking

Setelah menjalankan semua perhitungan diatas, maka tahap terakhir adalah menghitung *score ranking* untuk mengetahui dokumen mana yang paling relevan dengan kata kunci yang telah ditentukan. Perhitungan dilakukan dengan cara mengalikan *query* dengan masing-masing bobot *term*.

Hasil perhitungan tersebut dapat dilihat pada tabel 4.

Tabel 4. Perhitungan *Score Rank*

	Q*D1	Q*D2	Q*D3	Q*D4	Q*D5
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0	0	0	0	0
	0.002422	0	0.002422	0.002422	0
	Sum(Q*Di)				
	0.002422	0	0.002422	0.002422	0
Score Rank	0.006068	0	0.005122	0.009459	0

Penelitian saat ini hanya menggunakan 5 dokumen berita, sehingga terlihat hasil akhir pada Table.4 ada 5 *score rank*, yaitu:

D-1 memiliki *score rank* 0.006068
D-2 memiliki *score rank* 0
D-3 memiliki *score rank* 0.005122
D-4 memiliki *score rank* 0.009459
D-5 memiliki *score rank* 0

Dari hasil tersebut dapat diketahui dokumen mana yang paling relevan dan dokumen mana yang tidak relevan dengan kata kunci yang digunakan. Sehingga jika diurutkan dokumen mulai dari yang paling relevan adalah :
Dokumen 4 (D-4)
Dokumen 1 (D-1)
Dokumen 3 (D-3)

Sedangkan dokumen 2 (D-2) dan dokumen 5 (D-5) mempunyai score ranking 0, sehingga dapat dipastikan dokumen 2 dan dokumen 5 tidak relevan dengan kata kunci yang digunakan.

4. SIMPULAN

Setelah dilakukan perhitungan dan analisa pada penelitian ini, dapat diambil beberapa kesimpulan sebagai berikut :

1. Bahwa Vector Space Model (VSM) dapat digunakan dan dapat membantu untuk mencari serta menghitung score rank dari sebuah dokumen.
2. Dengan diketahuinya score rank tersebut dapat diketahui relevansi antara kata kunci yang digunakan dengan dokumen yang dicari.
3. Penelitian ini adalah penelitian awal, yang dikemudian hari dapat dikembangkan dengan penelitian lain dengan *scope* yang lebih besar.

DAFTAR PUSTAKA

[1] F. Gorunescu, *Data Mining Concepts, Models and Techniques*, 2011.

[2] O.Maimon dan L. Lokarch, *Data Mining and Knowledge Discovery Handbook*, Edisi ke-2. 2010.

[3] Manning, Christopher D., Raghavan, Prabhakar,. Schutze. *Introduction to Information Retrieval*. Cambridge University Press, New York, USA, 2008

[4] F. Amin, Implementasi *Search Engine (*Mesin Pencari*)* Menggunakan *MetodeVector Space Model.*2013

[5] N. Annisa, Warnia Nengsih, Ananda, Implementasi Algoritma Vector Space Model Dalam Pencarian *E-Book*, 2015, https://www.researchgate.net/publication/313477433

[6] A. Yudha, Y. Nuryaman, I. Nuddin, A. Andhikawati, Ernawati, N. Suwela, Sentiment Analysis pandangan masyarakat terhadap tarif tol Trans-Jawa menggunakan *Support Vector Machine* dan *Particle Swarm Optimization*, *The 10th University Research Colloqium 2019* Sekolah Tinggi Ilmu Kesehatan Muhammadiyah Gombong, 2019

[7] D. Susandi, U. Sholahuin, Pemanfaatan *Vector Space* Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi *Similarity Cosine* untuk Pembobotan IDF dan WIDF pada *Prototipe* Sistem Klasifikasi Teks Bahasa Indonesia, *Jurnal ProTekInfo* Vol. 3 No. 1 September 2016

[8] B. Amburika, Y. H. Chrisnanto, W. Uriawan, Teknik *Vector Space Model* (VSM) Dalam Penentuan Penanganan Dampak *Game Online* Pada Anak, *Prosiding SNST ke-7* Fakultas Teknik Universitas Wahid Hasyim Semarang, 2016.

[9] Anna, A. Hedini, Implementasi *Vector Space Model* Pada Sistem Pencarian Mesin Karaoke, *Jurnal Evolusi Volume* 6 No 1 - 2018.

[10] A. A. Abdillah, I. B. Muktyas, Implementasi *Vector Space* Model untuk Pencarian Dokumen, *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika* 2013.