

## Searching Information Mountain on the Java Island to help improve Tourism with the Vector Space Model

Yuri Yuliani<sup>1</sup>, Muhamad Azhar<sup>2</sup>, Ikhwanul Hakim<sup>3</sup>, Sfenrianto<sup>4</sup>, Achmad Rifai<sup>5</sup>

<sup>1,2,3</sup>Master of Computer Science Postgraduate Program, STMIK Nusa Mandiri, Jakarta, Indonesia,

<sup>1</sup>14002311@nusamandiri.ac.id, <sup>2</sup>muhammadazhar88@gmail.com, <sup>3</sup>ikhwanulhakim@gmail.com

<sup>4</sup>Information System management Departemen, BINUS Graduate Program, Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia, sfenrianto@binus.edu

<sup>5</sup>Technical Information, STMIK Nusa Mandiri, Jakarta, Indonesia, achmad.acf@nusamandiri.ac.id

### ABSTRACT

Mountain expeditions and climbers are not hindered by advances in digital technology at the moment, seen increasing mountain climbing on the island of Java. This is due to the abundance of biodiversity in the mountains of the island of Java such as Mount Gede Pangrango which has many nail trees, at the foot of Mount Rinjani which is surrounded by butterfly species, Mount Merbabu with 70 species of birds, natural savanna with an area of  $\pm 300$  Ha and has the highest mountain, Mount Semeru with 3,676 meters above sea level. The potential and natural wealth can attract tourists to travel there so that it can be a factor to develop the economy in the area. This research helps tourism to search information mountain in Java Island by using the Vector Space Model (VSM) algorithm. The algorithm using to data preparation, tokenizing, filtering, stemming, calculation, and finally document ranking framework. Generated 14 sample data from search results on Google with the query "mountain on the island of Java". VSM algorithm runs well enough to produce the highest ranking of 3.716189934.

**Key words :** Searching Information, Mountain Java, Tourism, VSM.

### 1. INTRODUCTION

The passion among nature lovers to be able to see the beauty on the top of the mountain does not disappear with technological advancements in the current digital era. Seen from the expedition and mountaineering that is increasingly carried out by the climbers in the mountains of Indonesia, especially mountain climbing on the island of Java [1].

Java Island is one of the islands in Indonesia which has many volcanoes, high mountains and has high biodiversity mountains such as Semeru Mountain, (3,676 mdpl) [2], Gunung Gede Pangrango which has many nail trees [3], then

at the foot of Mount Rinjani which is abundant with butterfly species [4], Merbabu mountain with 70 species of birds [5], baluran mountain which has a natural savanna, namely bekol savanna with an area of  $\pm 300$  Ha which is the largest savanna on the island of Java [6].

With the advantages possessed by the mountains on the island of Java can make many natural connoisseurs who hunt for the beauty that is owned by every mountain on this island of Java. Because the potential or superior natural wealth in an area can be a factor in regional economic development [7].

This study is intended to determine the extent to which tourists know the attractions on the largest island in Indonesia. By using the Vector Space Model (VSM) is expected to help tourism on the island of Java in promoting tourist attractions, by helping to improve the information [8] to increase the number of tours, especially climbers.

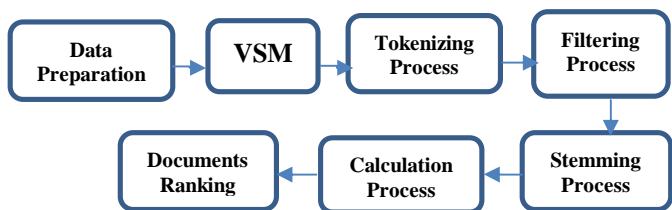
### 2. RESEARCH FRAMEWORK

#### 2.1 Information Retrieval

Formation retrieval is a science to get information in the form of text collected from documents based on queries. The information presented is a list of documents based on queries that have been assessed based on similarities using the similarity function [9]. The ranking results obtained from documents that have relevance to the query, for the level of relevance is a subjective thing that is influenced by factor topics, time, sources of information and user objectives [10].

#### 2.2 VSM Model

Vector Space Model introduces tf-IDF weighting as the term weight scheme. Each weight that has a frequency is called a (tf) factor that measures the frequency of occurrences in a document or text and the frequency document factor (IDF) measures the inverse of the number of documents containing a query or document [11]. With a framework like the following:



**Figure 1:** Research Framework

The data taken is secondary data taken on a mountain search on the island of Java on Google. Then the data will be processed using the Vector Space Model (SVM) to calculate the number of values obtained based on keywords by the process of tokenizing, filtering, stemming, calculation and ranking.

Then proceed with the tokenizing process. Tokenizing itself is a process of breaking down word by word in a sentence or document, to get weight on each word [12], which is calculated using Term Frequency (Tf) and Document Frequency (Df). After the tokenizing process, it is followed by a filtering process which functions to solve words and discard words that are not important for classification [13].

The next stage is the process of the stem which is the process of producing basic words by removing affixes to a word [14], numbers, and who have punctuation. Then do the calculation process and end with ranking on the document

### 3. DESCUSSION

#### 3.1 Research Data Preparation

Research data obtained from the results of a search on google with the query "gunung di pulau jawa". Then take 14 sample titles from search results on Google, including:

- [D.1] Terpikat Pesona Gunung Ciremai, Gunung Tertinggi di Jawa Barat.
- [D.2] Menyaksikan Sunrice Gunung Semeru, Cantik Bikin Rindu Mendaki
- [D.3] Semeru dan Empat Gunung Mempesona di Malang
- [D.4] Menggapai Puncak Semeru, Atap Pulau Jawa
- [D.5] Gunung Slamet, Gunung Tertinggi Kedua di Pulau Jawa dengan Tinggi 3428 Mdpl
- [D.6] Menyaksikan Sunrise Gunung di Jawa, Cantik Bikin Rindu Mendaki
- [D.7] Gunung Agrapura, Trek Pendakian Terpanjang di Jawa
- [D.8] Legenda Gunung Semeru Sebagai Paku Bumi Pulau Jawa
- [D.9] Gunung Slamet Pusat Spiritualitas di Jawa
- [D.10] Keindahan Gunung Bromo, Semeru, Prau dan gede-Pangrango di Pulau Jawa ini Bikin Susah Move-on
- [D.11] Tertutup Awan, Gunung Sumbing Disebut

Nampak Mengerikan

- [D.12] Letusan Gunung Merapi yang Konon Mengubah Sejarah Jawa
- [D.13] Gunung Api Nglanggeran, Kebumen, Kaldera pasir Tengger merupakan gunung api Purba di Pulau Jawa
- [D.14] Gunung Papandayan merupakan gunung yang ada sabana di jawa

#### 3.2 Tokenizing Process

After the data is collected, a tokenizing process is carried out, with the following results:

**Table 1:** Tokenizing Process

terpikat	Pesona	gunung	Ceremai
gunung	Tertinggi	di	Jawa
Barat	menyaksikan	sunrise	Gunung
Semeru	Cantik	bikin	Rindu
Pendaki	Semeru	dan	Empat
Gunung	mempesona	di	Malang
Menggapai	Puncak	semeru	Atap
Pulau	Jawa	gunung	Slamet
Gunung	tertinggi	kedua	Di
Pulau	Jawa	dengan	Tinggi
3428	Mdpl	menyaksikan	Sunrise
Gunung	Di	jawa	Cantik
Bikin	Rindu	mendaki	Gunung
Argapura	Trek	pendakian	Terpanjang
Di	Jawa	legenda	Gunung
Semeru	Sebagai	paku	Bumi
Pulau	Jawa	gunung	Slamet
Pusat	spiritualitas	di	Jawa
Keindahan	Gunung	bromo	Semeru
Prau	Dan	gede	Pangrango
Di	Pulau	jawa	Ini
Bikin	Susah	move	On
Tertutup	Awan	gunung	Sumbing
Disebut	Nampak	mengerikan	Letusan
Gunung	Merapi	yang	Konon
Mengubah	Sejarah	jawa	Gunung
Api	nglanggeran	kebumen	kalder
Pasir	Tengger	merupakan	Gunung
Api	Purba	di	Pulau
Jawa	Gunung	papandayan	Merupakan
Gunung	Yang	ada	Sabana
Di	Jawa		

**3.3 Filtering Process**

After getting the words that have been processed at the tokenizing stage, then the filtering stage is then carried out with the results of the data. Table 2 shown filtering process in the study.

**Table 2: Filtering Process**

ada	api	argapura	atap
awan	barat	bikin	bromo
bumi	cantik	ceremai	dan
dengan	di	disebut	empat
gede	gunung	ini	jawa
kalder	kebumen	kedua	keindahan
konon	legenda	letusan	malang
mdpl	mempesona	mendaki	mengerikan
menggapai	mengubah	menyaksikan	menyaksikan
merapi	merupakan	move	nampak
nglanggeran	on	paku	pangrango
papandayan	pasir	pendaki	pendakian
pesona	prau	pulau	puncak
purba	pusat	rindu	sabana
sebagai	sejarah	semeru	slamet
spiritualitas	sumbing	sunrise	susah
tengger	terpanjang	terpikat	tertinggi
tertutup	Tinggi	trek	Yang

**3.4 Stemming Process**

From the data generated at the filtering stage of 72 words, it will proceed to the stemming stage which produces 61 words as follows:

**Table 3: Stemming Precess**

ada	api	argapura	atap
awan	bagai	barat	bikin
bromo	bumi	cantik	ceremai
daki	dua	empat	gapai
gede	gunung	indah	jawa
kalder	kebumen	konon	legenda
letus	malang	mdpl	merapi
nampak	negeri	nglanggeran	paku
pangrango	panjang	papandayan	pasir
pendaki	pesona	pikat	prau
pulau	puncak	purba	pusat
rindu	rupa	sabana	saksi
sebut	sejarah	semeru	slamet
spiritualitas	sumbing	sunrise	susah
tengger	tinggi	trek	tutup
ubah			

**3.5 Calculation Process**

Look for the number of words in a document (tf) and

documents containing the searched word (df) (see figure 2)

Token	tf														
	Q	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
ada	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
api	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
argapura	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
atap	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
awan	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
bagai	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
barat	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
bikin	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0
bromo	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
bumi	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
cantik	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
ceremai	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
daki	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
dua	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
empat	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
gapai	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
gede	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
gunung	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
indah	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
jawa	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
kalder	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
kebumen	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
konon	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
legenda	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
letus	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
malang	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
mdpl	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
merapi	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
nampak	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
ngeri	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
nglanggeran	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
paku	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
pangrango	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
panjang	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
papandayan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
pasir	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
pendaki	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
pesona	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
pikat	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
prau	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
pulau	1	0	0	0	1	1	0	0	1	0	1	0	0	1	0
puncak	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
purba	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
pusat	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
rindu	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
rupa	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
sabana	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
saksi	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
sebut	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
sejarah	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
semeru	0	0	1	1	1	0	0	0	1	0	1	0	0	0	0
slamet	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
spiritualitas	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
sumbing	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
sunrise	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
susah	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
tengger	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
tinggi	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
trek	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
tutup	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
ubah	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

**Figure 2: Grading Per-Word Value**







## 5. CONCLUSION

The algorithm used in this study shows good results with 14 document sample data. The ranking process uses the stages of the VSM (Vector Space Model) algorithm with the highest value of 3.716189934 which is owned by document 9 from the ranking which produces 13 ranks. The results of the algorithm used in the case of mountain data search on the island of Java can be proven to run well.

This research development still needs to be done which consists of data samples and can be developed using other methods that can get better results in this case.

## REFERENCES

1. Yuliyana, I. A., Februariyanti, H., & Santoso, D. B. **Weather Underground Untuk Layanan Cuaca Pada Sistem Pendakian Gunung.** *Prosiding Sintak*, 83, 2018.
2. Wulandari, A., & Rahmawati, R. D. **Tingkat Ploidii Paku Sayur (Diplazium Esculentum) Pada Ketinggian Yang Berbeda Di Gunung Semeru.** *Edubiotik*, 59, 2018.  
<https://doi.org/10.33503/ebio.v3i02.98>
3. Maulidia, A., Sedayu, A., Sakti, D. P., Puspita, E. D., Kusumaningtiyas, F., Ristanto, R. H., Et Al. **Keanekaragaman Tanaman Paku (Pteridophyta) Di Jalur Ciwalen Taman Nasional Gunung Gede Pangrango, Jawa Barat.** *Biosfer, J.Bio. & Pend.Bio.*, 29, 2018.
4. Sumiati, Idrus, A. A., & Ilhamdi, L. **Keanekaragaman Kupu-Kupu (Subordo Rhopalocera) Di Kawasan Hutan Jeruk Manis.** *Prosiding Seminar Nasional Pendidikan Biologi*, 399-400, 2018.
5. Aditya, Nugroho, G. D., Jauhar, M. F., & Sunarto. **Keanekaragaman Burung Diurnal Dan Potensi Burung Sebagai Objek Daya Tarik Avitourism Di Taman Nasional Gunung Merbabu, Jawa Tengah.** *PROS SEM NAS MASY BIODIV INDON*, 362-263. 2019
6. Nugroho, H. T. **Pengembangan Taman Nasional Baluran Sebagai Destinasiwisata Unggulan Berbasis Alam Di Situbondo Jawa Timur.** *Jurnal Ilmiah Domestic Case Study*. 2019
7. Destiningsih, R., Acha, A., & Septiani, Y. (2019). **Analisis Potensi Wilayah Provinsi Jawa Tengah (Studi Kasus : Tahun 2010-2016).** *Jurnal REP (Riset Ekonomi Pembangunan)*, 74, 2019.  
<https://doi.org/10.31002/rep.v4i1.1343>
8. Lakshmi, D., Rani, K. P., & Reddy, M. P. (2019). **A Comparative Study of Navigation Techniques and Information.** *International Journal of Advanced Trends in Computer Science and Engineering*, 10.
9. Prasetyo, V. R., Hartanto, B., & Mulyono, A. A. **Penentuan Pembimbing Tugas Akhir Mahasiswa Jurusan Teknik Informatika Universitas Surabaya Dengan Metode Dice Coefficient.** *TEKNIKA*, 45, 2019.
10. Fauziah, S., Sulistyowati, D. N., & Asra, T. **Optimasi Algoritma Vector Space Model Dengan Algoritma K-Nearest Neighbour Pada Pencarian Judul Artikel Jurnal.** *Jurnal Pilar Nusa Mandiri*, 22, 2019
11. Heristian, S., Kautsar, H. A., & Sayfulloh, A. **Rancang Bangun Information Retrieval System (IRS) Kamus Bahasa-Sunda.Com Dengan Metode Vector Space Model (VSM).** *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 66, 2019.
12. Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). **Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine.** *Komputek*, 58, 2019.  
<https://doi.org/10.24269/jkt.v3i2.270>
13. Afandi, M. I., Adiwijaya, & Astuti, W. **Klasifikasi Multilabel Pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan Mutual Information Dan Support Vector Machine.** *E-Proceeding Of Engineering*, 9103, 2019
14. Subagyo, I., Yulianto, L. D., Permadi, W., Dewantara, A. W., & Hartanto, A. D. **Sentiment Analisis Review Film Di IMDB Menggunakan Algoritma SVM.** *Jurnal Sistem Informasi Dan Teknologi Informasi*, 48, 2019