

Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace

Charles Joergensen E Munthe, Nelly Astuti Hasibuan, Hukendik Hutabarat

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma

Jalan Sisingamangaraja No. 338, Medan, Sumatera Utara, Indonesia

Email: ¹charlesjoer123@gmail.com

Abstrak-Pada perkembangan zaman yang semakin canggih dan modern seperti saat ini, Marketplace online adalah salah satu layanan e-commerce atau layanan virtual jual beli yang menjadi tempat bertemunya antara penjual dan pembeli secara online untuk melakukan proses transaksi jual beli. Menentukan Algoritma Text mining merupakan proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Proses dalam text mining meliputi proses tokenisasi, stemming dan filtering. Metode pengumpulan data dengan metode kepustakaan. Tahapan pengembangan aplikasi meliputi perancangan proses, perancangan tabel, implementasi dan pengujian sistem. Adapun pengerjaan dalam tugas proyek akhir ini menggunakan studi kasus pada sebuah marketplace. Yang mana tugas proyek akhir ini menggunakan algoritma Text Mining dan algoritma TF-RF dalam menentukan promo produk pada marketplace tersebut.

Kata Kunci: Text Mining; TF-RF; Marketplace

Abstract-In the development of an increasingly sophisticated and modern era like today, the online marketplace is one of the e-commerce services or virtual buying and selling services which is a meeting place between sellers and buyers online to process buying and selling transactions. Determining the Algorithm Text mining is a process of mining data in the form of text where the data source is usually obtained from documents and the goal is to find words that can represent the contents of the document so that an analysis of the connectivity between documents can be carried out. Processes in text mining include tokenization, stemming and filtering processes. The method of collecting data is the library method. The stages of application development include process design, table design, implementation and system testing. The work in this final project uses a case study on a marketplace. In which this final project uses the Text Mining algorithm and the TF-RF algorithm in determining product promotions in the marketplace.

Keywords: Text Mining; TF-RF; Marketplace

1. PENDAHULUAN

Seiring dengan semakin berkembangnya teknologi, pertumbuhan *E-Commerce* di Indonesia memang bisa dikatakan sangat pesat, bahkan menjadi salah satu negara dengan pertumbuhan *marketplace* terpesat. Menurut ICD, pertumbuhan *e-commerce* Indonesia merupakan yang terbesar di kawasan Asia Tenggara. Awal mulanya marketplace berada di Amerika, *marketplace* mulai populer pada tahun 1995 dengan adanya *eBay* dan *Amazon*. Di China, *marketplace* mulai ramai setelah Jack Ma mendirikan *Alibaba* yang kini menjadi *marketplace* raksasa. Sementara di Indonesia, awal mula toko online dimulai pada tahun 1999 dengan didirikannya forum jual beli Kaskus. Namun di masa awal berdirinya forum jual beli online, kebanyakan orang hanya menggunakan *platform* tersebut untuk menunjukkan produknya. Sementara proses transaksinya tetap dilakukan secara offline. Beberapa tahun setelahnya, Tokobagus.com yang kini menjadi *olx* berdiri. Dengan mulai berkembangnya ekosistem-ekosistem digital di Indonesia. Tren marketplace terus berkembang seiring perkembangan internet dan teknologi digital di Indonesia. Pada tahun 2009, Tokopedia berdiri dan mengalami perkembangan yang sangat pesat. Marketplace pun menjadi semakin populer sebagai tempat bagi masyarakat untuk berbelanja berbagai kebutuhan. Pada tahun 2010 sampai sekarang, marketplace baru bermunculan diantaranya adalah Bukalapak, Lazada, Shoppe, dll turut bersaing diantara para marketplace terdahulu.

Penyelesaian masalah penelitian ini mengenai masalah yang terjadi pada penjualan produk dengan melakukan promo produk / promosi produk handphone namun hasilnya tidak sesuai dengan yang apa yang diharapkan dikarenakan tidak tepat sasaran.

Metode *text mining* merupakan pengembangan dari metode data mining yang dapat diterapkan untuk mengatasi masalah tersebut. Algoritma-algoritma dalam text mining dibuat untuk dapat mengenali data yang sifatnya semi terstruktur misalnya data-data, abstrak maupun isi dari dokumen-dokumen. *Text mining* adalah salah satu penambangan informasi yang berguna dari data – data yang berupa tulisan, 2 dokumen atau text dalam bentuk klasifikasi, penentuan maupun clustering. Text mining masih merupakan bagian dari data mining dimana akan memproses data – data atau text – text serta dokumen – dokumen yang bisa jadi dalam jumlah sangat besar. Untuk memproses data yang sangat besar tentulah akan memakan sumber daya yang tidak sedikit kaitanya dengan pengolahan data tersebut. Disinilah diperukanya sebuah pemrosesan awal atau preprocessing data text tersebut sebelum data tersebut di lakukan proses text mining sesuai algoritma yang akan diterapkan. Untuk dapat melakukan penambangan informasi atau text mining dalam menentukan promo produk pada marketplace maka perlu dilakukan beberapa tahapan yang harus dilakukan untuk mengolah sumber data baik dan yang terstruktur, perlu dilakukan proses awal atau di sebut sebagai preprocessing data promo yang bermaksud mengolah data awal yang masih bermacam – macam untuk dijadikan sebuah data teratur yang dapat dikenai atau diterapkan di metode text mining ini.[1]

Metode TF-RF yang berasal dari TF-IDF, TF-RF adalah pengukuran yang paling sederhana dalam metode pembobotan. Pada metode ini, masing-masing *term* di asumsikan mempunyai proporsi kepentingan sesuai jumlah terjadinya (munculnya) dalam teks (dokumen). *Term frequency* dapat memperbaiki nilai *recall* pada informasi *retrieval*, tetapi tidak selalu memperbaiki nilai *precision*. Hal ini disebabkan *term* yang *frequent* cenderung muncul di banyak teks, sehingga *term-term* tersebut memiliki kekuatan pembeda (keunikan) yang kecil. Sedangkan *term frequency* (TF) adalah pengukuran yang paling sederhana dalam metode pembobotan. Pada metode ini, masing-masing *term* di asumsikan mempunyai proporsi kepentingan sesuai jumlah terjadinya (munculnya) dalam teks (dokumen). *Term frequency* dapat memperbaiki nilai *recall* pada information retrieval, tetapi tidak selalu memperbaiki nilai *precision*. [2]

Pada penelitian-penelitian sebelumnya penulis menggunakan algoritma *text mining* dalam menentukan pencarian promo produk yang dilakukan oleh bambang kurniawan dengan judul Penerapan Algoritma *Text Mining* Untuk Menentukan Promo Produk Pada Supermarket Transmart menyimpulkan bahwa penentuan data promo produk secara tepat dan proses semakin akurat jika data yang digunakan dalam pembelanjaan memiliki data yang rinci dan menjadikan data promo tepat sasaran.

2. METODOLOGI PENELITIAN

2.1 Text Mining

Text Mining merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi, dimana *text mining* merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. Dan juga *text mining* dapat dikatakan sebagai penerapan konsep dari teknik data mining untuk mencari pola dalam teks, bertujuan untuk mencari informasi yang bermanfaat dengan tujuan tertentu.

Berdasarkan ketidak teraturan struktur data teks, maka proses *text mining* memerlukan beberapa tahap awal yang pada intinya mempersiapkan agar teks dapat diubah menjadi lebih terstruktur (Budi Susanto, Teknik Informatika Yogyakarta). Dalam penerapan *text mining*, proses *text mining* dibagi menjadi 3 tahap utama, yaitu proses awal terhadap teks (*text preprocessing*), transformasi teks (*text transformation*), dan penemuan pola (*pattern discovery*). [3]

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

TF-IDF merupakan gabungan dari metode Term Frequency (TF) dengan Inverse Document Frequency (IDF) yang dihasilkan dari perkalian metode TF dengan metode IDF. Metode TF-IDF ini memberikan nilai bobot yang tinggi kepada term yang sering muncul pada suatu dokumen, tetapi jarang muncul dalam kumpulan dokumen. Persamaan metode TF-IDF sebagai berikut :

$$TF * IDF (d,t) = TF(d,t) * \log \frac{N}{df(t)} \quad (1)$$

Dimana :

$TF * IDF (d, t)$: Pembobotan TF-IDF

$TF (d, t)$: Frekuensi munculnya *term t* pada dokumen *d*.

N : Jumlah dari semua kumpulan dokumen.

$df(t)$: Jumlah dari dokumen yang mengandung *term t*.

2.3 Term Frequency-Relevance Frequency (TF-RF)

TF-RF adalah metode gabungan antara TF dan RF dengan tujuan untuk mendapatkan performansi yang lebih baik dari metode pembobotan kata lainnya. Metode ini mempertimbangkan relevansi dokumen dilihat dari frekuensi kemunculan *term* di kategori yang berkaitan. Persamaan metode TF-RF sebagai berikut :

$$TF * RF (t,c) = TF(d,t) * \log_2 \left(2 + \frac{a}{\max(1,c)} \right) \quad (2)$$

Dimana :

$TF * RF (d, t)$: Pembobotan TF-RF.

$TF (d, t)$: Frekuensi munculnya *term t* pada dokumen *d*.

c : Kelas kategori.

t : *term*.

3. HASIL DAN PEMBAHASAN

3.1 Analisa

Dalam penelitian ini data yang digunakan adalah data yang berhubungan dengan URL pada marketplace tersebut. Pada penelitian ini data url marketplace akan di proses untuk menghasilkan promo produk pada marketplace dengan menggunakan Text Mining dan TF-RF.

Untuk menyelesaikan masalah diatas penulis menerapkan algoritma *Text Mining* yang merupakan suatu teknik data mining untuk mencari pola dalam penentuan promo produk pada marketplace. Proses penganalisaan teks guna menemukan informasi yang bermanfaat untuk tujuan tertentu. Algoritma text mining memiliki tahapan case folding, tokenizing, filtering, stemming, spelling normalization, dan analyzing hasil dari tahapan-tahapan adalah kata-kata root atau kata-kata akar dari suatu kalimat.

informasi yang didapatkan mengenai pola suatu barang yang memiliki tingkat penjualan tertinggi. dapat dimanfaatkan untuk penambahan stok barang tersebut. Selain itu juga, dari data transaksi dapat dimanfaatkan keterkaitan antar barang dari keranjang belanja pelanggan, dari situ dapat digunakan untuk penataan barang agar bisa lebih efisien serta dapat menarik minat pelanggan. Aplikasi yang biasanya digunakan untuk menganalisis data transaksi dari keranjang belanja para konsumen adalah market basket analysis

Untuk menerapkan pengetahuan awal kepada promo produk yang akan ditentukan kategorinya penulis menerapkan algoritma *Term Frquency – Relevance Frequency* (TF-RF), hasil dari penerapan algoritma tersebut berupa bobot keterkaitan data target atau promo produk yang belum ditentukan kategorinya dengan pengetahuan awal atau disebut dengan data target.

3.1.1 Penerapan Algoritma *Text Mining* dan TF-RF

Tahapan awal yang dilakukan pada penelitian ini adalah menentukan data awal sebagai pengetahuan awal, dalam penelitian tersebut peneliti telah mengambil data produk – produk beserta kategorinya sebanyak 10 produk yang dapat dilihat pada tabel 1.

Tabel 1.Data Awal Promo Produk

No.	Nama Produk	Kategori
1.	Laptop Asus	Komputer
2.	Kabel LAN RJ 45	Jaringan
3.	Adobe Photoshop CC 2020	Software
4.	Printer	Elektronik Komputer
5.	Matlab Book	Buku
6.	Blender	Elektronik Dapur
7.	Digital Printing	Elektronik Kantor
8.	Handphone Samsung	Gadget
9.	Ikan Salmon	Ikan
10.	Baju Kemeja	Fashion Pria

Setelah data awal tentukan maka dilakukan pengolahan teks sesuai dengan algoritma text mining, adapun tahapan dari algoritma text mining adalah sebagai berikut :

1. Case Folding

Case Folding adalah tahapan selanjutnya setelah tahapan *cleaning* dimana sistem mengubah semua huruf dalam teks menjadi huruf kecil. Karena tidak semua kata dalam teks menggunakan huruf kapital oleh karena itu peran *case folding* dibutuhkan untuk mengkonversi keseluruhan teks menjadi sebuah bentuk standar sehingga memudahkan untuk melanjutkan ke tahapan *tokenizing* selanjutnya. Hasil dari tahapan *case folding* dapat dilihat pada tabel 2. berikut :

Tabel 2. Tahap *Case Folding*

No.	Nama Produk	Kategori
1.	laptop asus	Komputer
2.	kabel lan rj 45	Jaringan
3.	adobe photoshop cc 2020	Software
4.	Printer	Elektronik Komputer
5.	matlab book	Buku
6.	blender	Elektronik Dapur
7.	digital printing	Elektronik Kantor
8.	handphone samsung	Gadget
9.	ikan salmon	Ikan
10.	baju kemeja	Fashion Pria

2. Tokenizing

Tokenizing adalah tahap pemotongan string input berdasarkan tiap kalimat yang menyusunnya. Tokenisasi secara garis besar memecah kalimat dalam suatu teks ke dalam bentuk sekumpulan satuan kalimat untuk dihitung hasilnya nanti. Proses tokenisasi akan semakin sulit dikarenakan harus memperhatikan struktur bahasa. Tahapan *Tokenizing* dapat dilihat seperti pada tabel 3. berikut :

Tabel 3. Tahapan *Tokenizing*

No.	Nama Produk	Kategori
1.	laptop asus	Komputer
2.	kabel lan rj 45	Jaringan
3.	adobe photoshop cc 2020	Software
4.	Printer	Elektronik Komputer
5.	buku matlab	Buku
6.	blender	Elektronik Dapur
7.	digital printing	Elektronik Kantor
8.	handphone samsung	Gadget
9.	ikan salmon	Ikan
10.	baju kemeja	Fashion Pria

3. *Filtering*

Pada tahapan ini akan dilakukan penghapusan kata yang tidak penting, pada tahapan tersebut untuk memudahkan penghapusan kata tidak penting penulis memanfaatkan library stopwords dari tala. Hasil dari tahapan tersebut dapat dilihat pada tabel 4.

Tabel 4. Tahapan *Filtering*

No.	Nama Produk	Kategori
1.	Laptop	Komputer
2.	kabel lan	Jaringan
3.	photoshop	Software
4.	Print	Elektronik Komputer
5.	matlab	Buku
6.	Blend	Elektronik Dapur
7.	dig printing	Elektronik Kantor
8.	handphone	Gadget
9.	salmon	Ikan
10.	kemeja	Fashion Pria

4. *Stemming*

Stemming merupakan suatu proses yang terdapat dalam sistem IR yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Hasil dari tahapan tersebut dapat dilihat pada tabel 5.

Tabel 5. Tahapan *Stemming*

No.	Nama Produk	Kategori
1.	laptop asus	Komputer
2.	kabel lan rj 45	Jaringan
3.	adobe photoshop cc 2020	Software
4.	Printer	Elektronik Komputer
5.	buku matlab	Buku
6.	blender	Elektronik Dapur
7.	digital printing	Elektronik Kantor
8.	handphone samsung	Gadget
9.	ikan salmon	Ikan
10.	baju kemeja	Fashion Pria

5. *Analyzing*

Analyzing Merupakan tahap penentuan seberapa jauh keterhubungan antar kata-kata dokumen yang ada. Untuk memudahkan analisa keterhubungan antar kata-kata pada setiap nama produk yang telah diketahui kategorinya nama produk yang belum diketahui kategorinya, maka penulis menerapkan algoritma *Term Frequency – Relevance Frequency* (TF-RF). Algoritma tersebut akan membantu penulis dalam menganalisa hubungan antara masing-masing kata dari setiap nama produk.

Pada penelitian tersebut penulis telah menentukan nama produk yang akan ditentukan kategorinya yaitu “**adobe photoshop cc 2020**” Untuk memudahkan proses perbandingan antara nama produk yang sudah diketahui kategorinya dengan nama produk yang belum diketahui kategorinya maka penulis melakukan tahapan algoritma text mining terhadap nama produk yang belum diketahui kategorinya tersebut. Hasil dari penerapan algoritma text mining terhadap nama produk yang belum diketahui kategorinya tersebut adalah “**Handphone Samsung**”. Setelah nama produk diterapkan algoritma text mining maka selanjutnya melakukan perbandingan atas kedua judul tersebut dengan

menerapkan algoritma *Term Frequency – Relevance Frequency* (TF-RF). Hasil dari algoritma tersebut dapat dilihat pada tabel 3.6.

Tabel 6. Hasil Perhitungan Term Frequency

Q	Tf										rf	$\frac{R}{rf}$	RF	RF+1
	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10				
Asus	1	0	0	1	0	0	0	1	1	0	4	2,5	0,397	1,397
Rj45	0	1	0	0	0	0	0	0	1	0	2	5	0,698	1,698
Photoshop	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Print	0	0	1	1	1	0	1	0	1	0	5	2	0,301	1,301
Blend	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Digital	0	0	1	0	1	0	0	1	0	1	4	2,5	0,397	1,397
Hp	0	0	0	0	0	0	0	0	0	0	0	0	0	0
salmon	2	0	1	0	0	0	0	0	0	0	2	5	0,698	1,698
kemeja	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabel 7. Hasil Pembobotan TF-RF

W=tf* (RF+1)										
r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	
1,397	0	0	1397	0	0	0	1397	1,397	0	
0	1,698	0	0	0	0	0	0	1,698	0	
0	0	0	0	0	0	0	0	0	0	
0	0	1,301	1,301	1,301	0	1,301	0	1,301	0	
0	0	0	0	0	0	0	0	0	0	
0	0	1,397	0	1,397	0	0	1,397	0	1,397	
0	0	0	0	0	0	0	0	0	0	
3,396	0	1,698	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	
sum(d1)	sum(d2)	sum(d3)	sum(d4)	sum(d5)	sum(d6)	sum(d7)	sum(d8)	sum(d9)	sum(d10)	
4,793	1,698	5,919	2,698	2,698	0	1,301	2,794	4,396	1,397	

Dari hasil perhitungan dengan algoritma TF-RF, menunjukkan bahwa pada nama produk ke-3 adalah nilai nilai bobot yang paling besar yaitu 5.919, berdasarkan hal tersebut maka nama produk “**Adobe Photoshop CC 2020**” dan kategorinya adalah “**Software**”.

4. KESIMPULAN

Setelah dilakukan pengujian dengan menggunakan metode *Text Mining* dan TF-RF dalam penentuan promo produk, penulis dapat mengambil kesimpulan yang diharapkan dapat bermanfaat bagi banyak bidang terutama dalam bidang pendidikan. Adapun simpulan tersebut adalah Pengujian metode *Text Mining* dan TF-RF dalam menentukan promo produk bisa dapat dijalankan sesuai dengan yang diinginkan. Algoritma *Text Mining* dan TF-RF dapat diterapkan dalam program from produk untuk menentukan promo pada produk-produk yg telah ditentukan

REFERENCES

- [1] S. M. Holida, T. Alawiyah, and H. Sutisna, “Penerapan Animasi Interaktif Dalam Pengenalan Aksara Sunda,” *J. Inform.*, vol. 1, no. 2, pp. 111–122, 2016, doi: 10.31311/ji.v1i2.39.
- [2] Haniperdana, (November 20, 2012). Operators of genetic algoritmn [online]. Available: <https://technos21.wordpress.com/2012/11/20/pengertian-hardware-software-dan-brainware>
- [3] Even, Yahir dan Zohar. 2002. Introduction to text mining. Automeated Learning Group National Center For Supercomputing Applications. University of Illions.
- [4] A. T. J. H, “Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining,” *Inform. UPGRIS*, vol. 1, pp. 1–9, 2015.
- [5] A. Deolika, K. Kusri, and E. T. Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [6] Lesa James (06 Desember 2016). Mengenal Definisi Promo [online] Available : <http://hargamurah.org/amp/mengenal-definisi-promo-dan-pentingnya-melakukan-promo-terbaru/>
- [7] Helmy Ashari, (7 April 2019). Apa itu Marketplace Online? [online] Available : <https://idwebhost.com/blog/apa-itu-marketplace-online/>
- [8] R. A.S and M. Shalahudin, *Modul Pembelajaran: Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek)*, Modula. Bandung, 2011.
- [9] Mohtasham, “UML Tutorial [online].” Available : <https://www.tutorialspoint.com/uml/index.htm>

- [10] H. Maulana, “Analisis Dan Perancangan Sistem Replikasi Database Mysql Dengan Menggunakan Vmware Pada Sistem Operasi Open Source,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 1, pp. 32–37, 2016, doi: 10.30743/infotekjar.v1i1.37.
- [11] Rahmat Priyanto, *Langsung Bisa Visual Basic.Net 2008*. 2009.
- [12] M. P. Simatupang and D. P. Utomo, “Analisa Testimonial Dengan Menggunakan Algoritma Text Mining Dan Term Frequency-Inverse Document Frequence (Tf-Idf) Pada Toko Allmeeart,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 808–814, 2019.