

PENCARIAN PRODUK E-COMMERCE BERDASARKAN JUDUL DAN DESKRIPSI MENGGUNAKAN *VECTOR SPACE MODEL*

Ayuni Asistiyasari

ayuni.yas@bsi.ac.id

Prodi Sistem Informasi, Fakultas Teknologi Informatika,
Universitas Bina Sarana Informatika

ABSTRACT

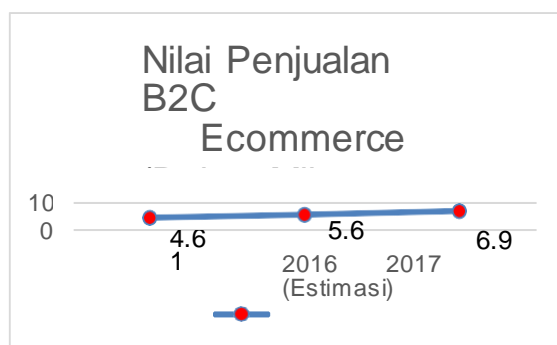
E-Commerce in Indonesia has increased every year, especially B2C sales and this is directly proportional to the increasing number of e-commerce sites in Indonesia. One of the things that this site offers is the ease of finding the product you want to find. However, in several tests conducted, it turns out that some Indonesian e-commerce sites are unable to search for products based on the description keyword. Therefore, the writer tries to do this research using the VSM algorithm and produces an accuracy of 60%, a precision of 72.73% and a recall of 80%.

Keywords: *E-Commerce, Vector Space Model, Text Mining, CRISP-DM*

1. PENDAHULUAN

Penjualan *E-Commerce* di Indonesia diprediksi meningkat setiap tahunnya seperti yang ditunjukkan pada gambar 1.1. Bahkan pemerintah secara khusus menerbitkan *roadmap e-commerce* sebagai peta agar Indonesia menjadi pemimpin

ekonomi digital di wilayah Asia Tenggara pada 2020. Namun hal itu tentunya tidak akan tercapai jika tidak ditunjang dengan kualitas dari platform *e-commerce* yang ada di Indonesia. Salah satu aspek kualitas dari platform adalah kualitas dari sistem pencari yang ada didalam platform tersebut.



Sumber : validnews.co(2020)

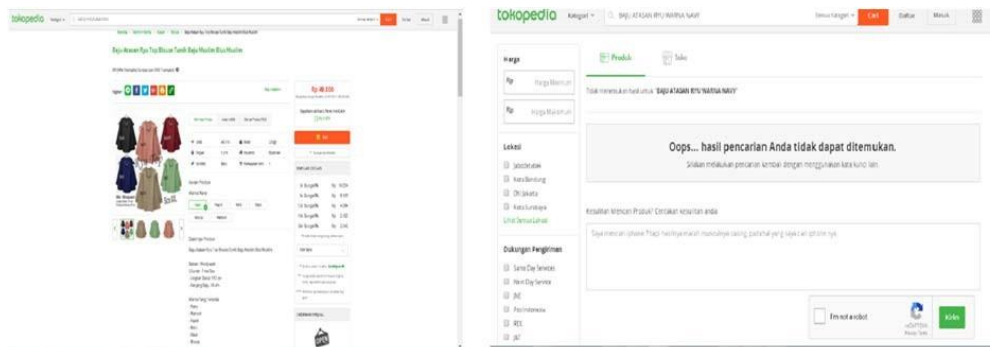
Gambar 1.1 Estimasi Penjualan B2C Ecommerce

Salah satu kekurangan dalam *e-commerce* Indonesia adalah bagaimana kebanyakan dari mereka hanya

mempertimbangkan judulnya saja sebagai target dari pencarian. Sehingga banyak kemungkinan suatu produk ditambah

keterangan tambahan yang terdapat dalam deskripsi tidak akan muncul dalam pencarian tersebut. Jangan berharap dapat

menemukan produk yang terpampang pada deskripsi tambahan.



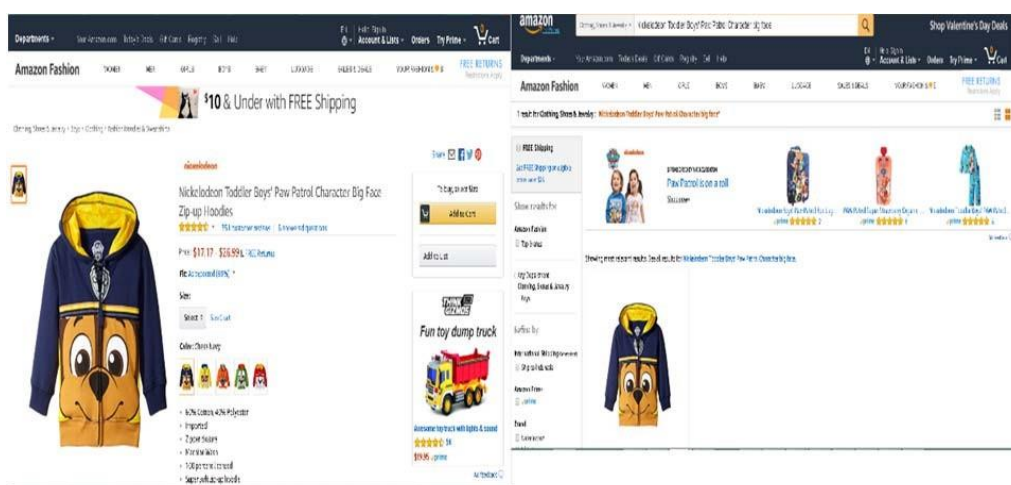
Sumber: Tokopedia.com (2019)

Gambar 1.2

Pencarian Query Judul dan Deskripsi Tidak Ditemukan

Berbeda dengan *e-commerce* ternama luar negeri seperti Amazon yang menerapkan target pencarian dengan

mempertimbangkan judul dan deskripsi suatu barang seperti yang tampak pada gambar 1.3 di bawah ini.



Sumber: Amazon.com (2019)

Gambar 1.3

Pencarian Query Judul dan Deskripsi Ditemukan

Sebenarnya banyak penelitian-penelitian di Indonesia terkait sistem pencarian dengan berbagai *algoritma*. Di antaranya Leman. D dan Andesa. K (2015) mengemukakan bahwa dengan “*Algoritma Vector Space Model* dapat diterapkan dalam sistem pencarian buku”. Sedangkan pada penelitian lainnya Irmawati (2017) mengemukakan bahwa “*Metode Vector Space Model*” mampu menghasilkan rata-

rata *recall* hampir 100% dan rata-rata *precision* sekitar 73,6%. Pada penelitian lainnya, Prasetya (2017) mengemukakan bahwa metode *K-nn* digabungkan dengan *Content Based* menghasilkan rata-rata nilai *precision* dan *recall* yang lebih tinggi dibandingkan menggunakan *collaborative filtering* dan *hybrid* pada ke-10 sebesar 0.0080 dan 0.148.

Namun dari penelitan di atas ternyata masih memiliki nilai *accuracy*, *precision* dan *recall* yang rendah, dan yang terpenting ternyata pada penelitian tersebut hanya mempertimbangkan suatu judul saja, tidak keseluruhan seperti deskripsi atau pun keterangan lainnya.

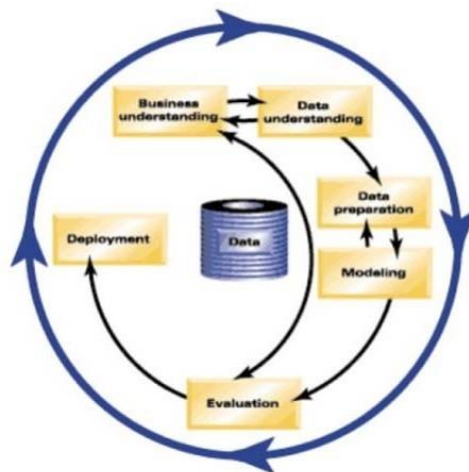
Pada penelitian-penelitian berbahasa Inggris tidak hanya terdapat penelitian yang hanya mempertimbangkan judulnya saja sebagai target pencarian. Namun juga terdapat penelitian-penelitian yang mempertimbangkan keseluruhan. Baik dari judul maupun keterangan lainnya. Pada penelitian yang mempertimbangkan keseluruhan baik judul dan keterangan lainnya Sawant dkk. (2017) mengungkapkan “Pencarian produk *e-fashion* yang menggunakan pemecahan text yang dalam mampu menghasilkan nilai *f1* 92%”. Akan tetapi apabila model tersebut langsung diterapkan terhadap *e-commerce* berbahasa Indonesia tentunya

akan sulit. Hal ini disebabkan karena pada penelitian tersebut mereka menggunakan *wordembedding* yang memang dikhususkan mengenali *vocabulary* bahasa Inggris.

Oleh sebab itu penulis mencoba melakukan penelitian dengan mempertimbangkan keseluruhan judul maupun suatu deskripsi barang menggunakan *Vector Space model* untuk mendapatkan label suatu barang.

2. METODE

Menurut Larose dalam (Kusrini& Luthfi. 2009:8) Data mining adalah sebuah proses, yang mana dalam melakukan prosesnya harus sesuai dengan prosedur dari proses tersebut, yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri dari keseluruhan proses, *preprocessing* data, pembentukan model, model evaluasi, dan tahap akhir penyebaran model.



Gambar 2.1 Proses CRISP-DM (Kusrini & Luthfi: 2009:8)

1. Business/Research Understanding Phase (Fase Pemahaman Bisnis)

Teknik pencarian produk pada toko *online* sangat penting. Seperti yang digambarkan sebelumnya, bagaimana teknik

pencarian sangat memengaruhi hasil dari pencarian. Apabila teknik pencariannya baik maka akan memungkinkan produk yang ingin dibeli dapat ditemukan dengan cepat dan tepat. Sebaliknya, apabila

teknik pencariannya kurang baik maka produk yang ingin dibeli akan sulit ditemukan.

2. Data Understanding Phase (Fase Pemahaman Data)

Data yang akan digunakan dalam penelitian ini diambil dari beberapa toko *online* dengan total 480 *record* dari pencarian yang tidak dapat ditemukan menggunakan pencarian berdasarkan judul dan deskripsi. Setiap *record* data terbagi menjadi 2, yaitu judul dan deskripsi.

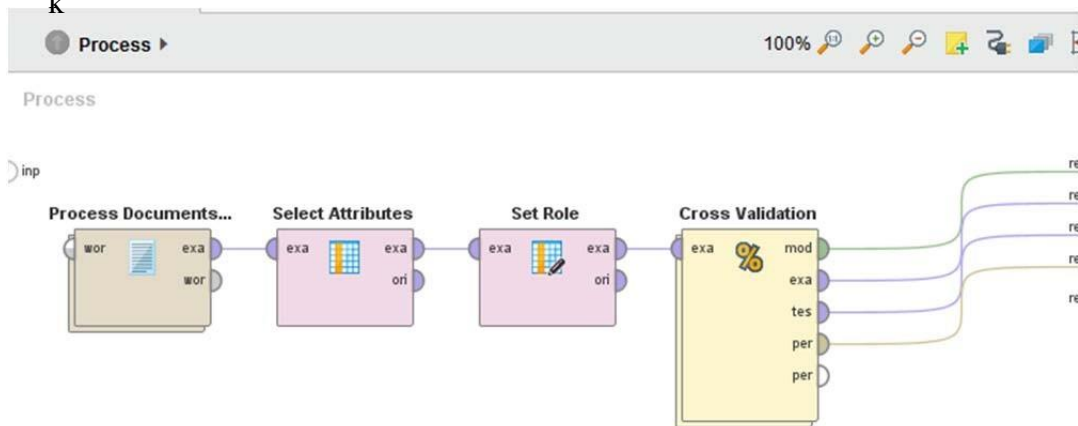
3. Data Preparation Phase (Fase Pengolahan Data)

Dari data yang diambil sebanyak 480 *record* tersebut k

i menjadi 2 yaitu data training sebanyak 240 *record* dan data *testing* sebanyak 240 *record*. Kemudian dilakukan *text preprocessing* yaitu *case folding*, *tokenization*, *filtering*, *stopremoval* dan *stemming* pada data training serta pembentukan label tambahan.

4. Modeling Phase (Fase Pemodelan)

Data yang sudah di *preprocessing* kemudian dimasukan kedalam model yang akan digunakan. Pemodel yang digunakan dalam penelitian ini yaitu menggunakan *Algoritma Vector Space Model* menggunakan rapid miner.



g

Gambar 3.1 Modeling VSM menggunakan Rapid Miner

5. Evaluation Phase (Fase Evaluasi)

Untuk evaluasi yang dilakukan pada penelitian ini adalah menggunakan *confusion matrix*.

6. Deployment Phase (Fase Penyebaran)

Deployment yang dilakukan pada penelitian ini sendiri yaitu dengan pembuatan aplikasi

pencarian produk berbasis *website*.

3. HASIL DAN PEMBAHASAN

Setelah melakukan *preprocessing* maka peneliti langsung memasukan data ke dalam *rapid miner* untuk melihat akurasi, presisi dan *recall* seperti tampilan di bawah ini.

accuracy: 60.00% +/- 43.59% (mikro: 64.29%)

	true negatif	true positif	class precision
pred. negatif	10	20	33.33%
pred. positif	30	80	72.73%
class recall	25.00%	80.00%	

Gambar 3.2 Hasil Akurasi

4. KESIMPULAN

Dari hasil perhitungan yang sudah dilakukan bahwa *Algoritma Vector Space Model* mampu menghasilkan akurasi sebesar 60%, presisi sebesar 72,73% dan

recall sebesar 80 % dari *record* yang tidak mampu ditemukan oleh pencarian sebelumnya, yang artinya mampu meningkatkan pencarian yang lebih optimal dibandingkan model pencarian berdasarkan judul saja.

DAFTAR PUSTAKA

- Aprilla, D, Baskoro, Donny Aji, Ambarwati, Lia, & Wicaksana, I Wayan Simri. (2013). *Belajar Data Mining dengan Rapid Miner*. Jakarta.
- Irmawati. (2017). *Sistem Temu Kembali Informasi Pada Dokumen Dengan Metode Vector Space Model*. Jurnal Ilmiah FIFO. Vol IX, No. 1
- Kusrini & Luthfi. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Larose, Daniel. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons. Inc.
- Leman, D dan Andesa, K. (2015). *Implementasi Vector Space Model Untuk Meningkatkan Kualitas Pada Sistem Pencarian Buku Perpustakaan*. Medan: Seminar Nasional Informatika. Vol 1, No. 1.
- Prasetya Chandra Saha Dewa. (2017). *Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor*. Jurnal Teknologi Informasi dan Ilmu Komputer. Vol. 4, No. 3
- Sawant, Uma, Gabale & Subramanian. (2017). *E-fashion Product Discovery Via Deep Text Parsing*. Australia: International Word Web Conference Committee.