

IMPLEMENTASI INFORMATION RETRIVALS UNTUK MENINGKATKAN PEMASARAN PRODUK

Hasan Basri¹, Richardius Eko Indrajit²

^{1,2} Program Studi Magister Ilmu Komputer

¹ STMIK Nusa Mandiri Jakarta, ²ABFI Institute Perbanas
Jakarta, Indonesia

hasanbasrisukses@gmail.com, indrajit@post.harvard.edu

Abstract— *Marketing is a promotional activities of products that aim to increase sales. Effective and efficient marketing should produce a comparison between the costs incurred and obtaining sales income. Companies that are already great, sometimes not yet able to achieve appropriate reciprocal results from marketing the product. That is because the less Interestingly information available in a brochure. This article describes how to implement an information retrieval algorithm is an effective and efficient marketing practitioners to help improve the performance of their sales. As for the algorithms used are Naïve Bayes and Vector Space Model. Naïve bayes Algorithm serves as a search engine text using the classification in order to be effective and efficient. Vector Space Model is a word search algorithm using vector space by looking at how big of a level of similarity on a document. The study is expected to assist the consumer in determining the best product from a brochure.*

Intisari—Brosur marketing adalah alat yang digunakan oleh seorang developer perumahan untuk menawarkan produknya. Namun kendala yang dihadapi pada saat ini brosur marketing tidak begitu efektif digunakan karena hasil pencapaian penjualan dari produk tidak memenuhi target. Hal tersebut dibuktikan dengan hasil observasi di beberapa perumahan yang ada di kota karawang. Data yang didapat dari hasil observasi diklasifikasi dengan metode naïve bayes untuk melihat sejauh mana peluang yang didapatkan dari pemanfaatan brosur itu sendiri. Berdasarkan permasalahan itu maka perlu digunakan metode yang lain. *information retrivals* merupakan suatu mesin pencarian materil koleksi dalam jumlah besar yang tersimpan dalam komputer. Elemen penting yang mendasar dari sebuah system temu kembali informasi adalah inverted index yang dapat menyimpan informasi semua kata yang ada dalam sekumpulan dokumen yang digunakan untuk membangun sistem temu kembali informasi tersebut. metode yang digunakan dalam kesempatan ini adalah Vektor Space Model. Metode atau algoritma vector space model merupakan sebuah algoritma pencarian

informasi dalam ruang lingkup dokumen yang berjumlah besar menggunakan ruang vector. Dengan metode Vektor Space Model ini diharapkan dapat membantu pencarian informasi secara cepat dan tepat dalam pemilihan perumahan.

Kata Kunci: *Marketing, Naïve Bayes, Vector Space Model, Algoritma, Brosur.*

PENDAHULUAN

Kenyataan yang terjadi saat ini adalah konsumen kebingungan dalam memilih perumahan yang sesuai dengan kebutuhan untuk keluarganya. Permasalahan tersebut terlihat semakin lengkap dengan banyaknya *marketing developer* perumahan yang menawarkan dengan harga yang menarik dan bonus yang menggiurkan (Sailendra & Latar, 2014).

Pada umumnya *marketing developer* perumahan menawarkan produk secara face to face kepada konsumen. *Marketing developer* menawarkan dengan cara menarik perhatian konsumen yang didukung dengan sebuah brosur marketing. Konsumen bisa melihat secara tiga dimensi bagaimana tampilan rumah yang terdapat dalam katalog promosi perumahan (Zuniargoprabowo et al., 2015)

Kemajuan teknologi yang bisa dirasakan sampai saat ini memungkinkan semua individu memperoleh informasi secara cepat. Transformasi informasi yang begitu cepat tidak menjamin bahwa kualitas informasi tersebut memiliki tingkat keakuratan informasi yang tinggi. Salah satu contohnya adalah informasi pencarian perumahan yang sering dilakukan kebanyakan orang pada umumnya pada mesin pencari, seperti *Google, MozillaFirefox* dan mesin pencari lainnya.

Permasalahan yang dihadapi oleh seseorang dalam mencari informasi perumahan berdasarkan brosur yang menarik.

Oleh Karena itu perlu suatu metode pencarian informasi yang tepat untuk menemukan sebuah informasi tersebut. Untuk memperoleh informasi yang tepat, maka

diperlukan sebuah metode pengklasifikasian informasi.

Metode klasifikasi merupakan sebuah proses model yang berfungsi untuk membedakan konsep atau model data, dengan tujuan dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui (Sulaehani, 2016).

Pada kesempatan ini algoritma naïve bayes dapat digunakan untuk melihat sejauh mana peluang yang dimiliki oleh developer marketing dalam menggunakan brosur .SVM dan Bayesian Classifier. (Herliana, 2016). Untuk meningkatkan hasil pencarian informasi perumahan pada brosur secara online dapat diterapkan suatu algoritma *vector space model*.

BAHAN DAN METODE

Populasi Data

Populasi data yang dibutuhkan untuk meningkatkan pemasaran produk perumahan adalah dengan mengetahui angka penjualan perumahan setiap bulan yang kemudian

dibandingkan dengan penjualan perumahan pada bulan-bulan berikutnya.

Angka penjualan perumahan tersebut tentunya dipengaruhi suatu media promosi yaitu brosur marketing.

Selain angka penjualan perumahan yang dipengaruhi oleh brosur marketing, untuk mengetahui sejauh mana tanggapan konsumen atau calon konsumen dalam memilih perumahan dapat dilakukan secara pengisian kuisisioner kepada konsumen untuk melihat kepuasan dalam memilih perumahan.

Format Data

Format atau karakteristik data yang dibutuhkan adalah dalam bentuk statistik penjualan perumahan/bulan, dari penjualan setiap bulan dapat terlihat keefektifan penggunaan brosur. Sehingga peluang penjualan perumahan dapat diprediksi hasilnya. Berikut adalah hasil klasifikasi dari penelitian terkait pengaruh brosur marketing pada 10 populasi perumahan menggunakan algoritma naïve bayes.

Tabel 1. Klasifikasi Brosur Marketing

NO	HARGA	LOKASI	SPESIFIKASI	PROMO	PEMBAYARAN	KEPUTUSAN PEMBELI
1	MURAH	KURANG STRATEGIS	TIPE B	DISKON RENDAH	KREDIT	YA
2	MURAH	KURANG STRATEGIS	TIPE B	DISKON RENDAH	KREDIT	TIDAK
3	MAHAL	STRATEGIS	TIPE A	DISKON SEDANG	KREDIT	YA
4	MURAH	STRATEGIS	TIPE B	DISKON SEDANG	KREDIT	YA
5	SEDANG	STRATEGIS	TIPE C	DISKON TINGGI	KREDIT	YA
6	SEDANG	KURANG STRATEGIS	TIPE B	DISKON SEDANG	KREDIT	YA
7	MAHAL	KURANG STRATEGIS	TIPE B	DISKON RENDAH	KREDIT	TIDAK
8	MAHAL	STRATEGIS	TIPE A	DISKON TINGGI	KREDIT	YA
9	MURAH	KURANG STRATEGIS	TIPE C	DISKON RENDAH	KREDIT	TIDAK
10	MURAH	KURANG STRATEGIS	TIPE C	DISKON RENDAH	KREDIT	YA

Sumber: Basri & Indrajit (2017)

Organisasi Penyimpanan Informasi

Syntax Pencarian Informasi

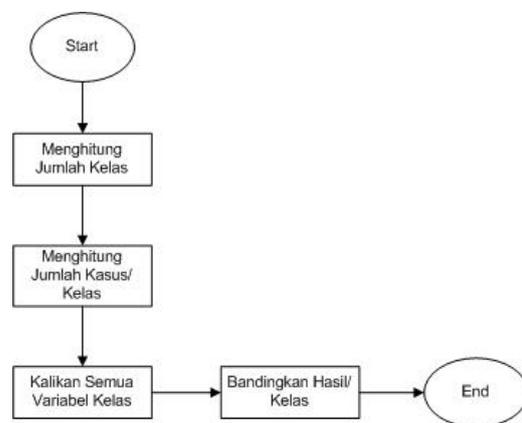
Syntax pencarian yang digunakan adalah menggunakan syntax SQL. Pencarian SQL dengan menggunakan logika “Or” data pencarian perumahan akan semakin sering muncul. Selain menggunakan logika “Or” dalam syntax pencarian juga dikombinasikan dengan menggunakan algoritma Vektor Space Model untuk membantu mempercepat dalam pencarian informasi.

Algoritma Naïve Bayes

Algoritma yang digunakan dalam penelitian ini adalah menggunakan algoritma naïve bayes. Metode bayes ini merupakan metode yang baik dalam mesin pembelajaran berdasarkan data training, dengan menggunakan probabilitas bersyarat sebagai syaratnya. Algoritma Naive Bayes juga salah satu algoritma

yang terdapat pada teknik klasifikasi (Informatika, Malikussaleh, & Utara, 2014).

Adapun algoritma naïve bayes sebagai berikut:

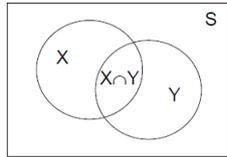


Sumber: Basri & Indrajit (2017)

Gambar 1. Flowchart Algoritma Naïve Bayes

Dari flowchart algoritma dapat dicari sebuah probabilitas bersyarat dari sekumpulan data yang ada. Untuk probabilitas syarat dapat dinyatakan sebagai berikut:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \dots\dots\dots(1)$$



Sumber: Basri & Indrajit (2017)
 Gambar 2. Notasi Algoritma Naïve Bayes

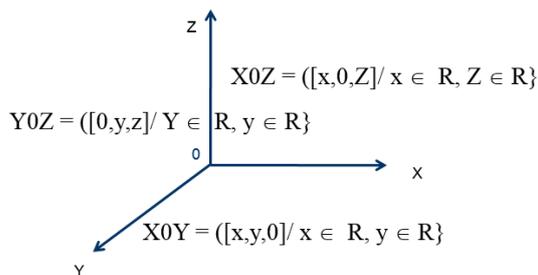
Probabilitas X di dalam Y adalah probabilitas interseksi X dan Y dari probabilitas Y, atau dengan kata lain $P(X|Y)$ adalah prosentase banyaknya X didalam Y.

Secara garis besar untuk metode algoritma naïve bayes dibagi menjadi 2 tahapan, diantaranya:

1. Tahap Pembelajaran (Learning Phase)
 Pada tahap ini perlu proses klasifikasi data berdasarkan kelas yang telah ditentukan.
 Tahap Pengujian (Testing Phase)
 Pada tahap ini dilakukan proses perbandingan antara kelas yang telah ditemukan setiap kelasnya, sehingga dapat diketahui tingkat akurasi dari kelas tersebut.

Vektor Space Model

Vector space model adalah suatu algoritma dari sistem temu balik informasi yang dapat dimodelkan dengan menggunakan ruang vektor. Pendekatan ini memudahkan untuk pencarian informasi yang terdapat pada suatu dokumen terhadap n kata yang berbeda. Untuk selanjutnya kata-kata yang dicari akan membentuk ruang vector yang memiliki dimensi sebesar n. Model Ruang Vektor ini juga dapat digunakan untuk memodelkan seberapa besar relevansi kata kunci atau query yang diberikan dengan kata pada sebuah dokumen yang ada.



Sumber: Basri & Indrajit (2017)

Gambar 3. Ruang Vektor Model

Elemen mendasar yang terpenting dari sistem temu informasi adalah inverted index yang menyimpan informasi semua kata yang terdapat pada sekumpulan dokumen. Kumpulan seluruh kata yang ada pada dokumen tersebut sebagai kamus. Kata yang dimasukkan kedalam kamus diperoleh melalui proses tokenization yang dilakukan pada dokumen. Tokenization menerima inputan berupa sekumpulan karakter secara berurutan, setelah karakter tersebut berurutan selanjutnya masuk kedalam tahapan membagi menjadi beberapa bagian yang lebih kecil (Token) dengan menghilangkan karakter spesial seperti tanda baca. Token yang merupakan kumpulan karakter yang mempunyai makna secara semantik umumnya dianggap sama dengan kata (term) walaupun sebenarnya tidak selalu memiliki arti. Adapun proses tokenization dapat dilihat sebagai berikut:

Dokumen	Soft Marketing Adalah Kemampuan Memasarkan Produk Secara Personal....		
Token	Soft	Marketing	Adalah
	Kemampuan		Memasarkan
	Produk	Secara	Personal
Term	Soft	Marketing	Adalah
	Kemampuan		Memasar
	Produk	Secara	Personal

Sumber: Basri & Indrajit (2017)
 Gambar 4. Proses Tokenization Dokumen

Proses Indexing

Indexing subsystem merupakan sebuah proses subsystem yang merepresentasikan beberapa dokumen kedalam bentuk tertentu untuk mempermudah dan mempercepat proses pencarian data dan temu kembali dokumen yang relevan. Pengembangan index dari beberapa dokumen merupakan point penting pada tahapan preprocessing di dalam information retrieval. Kualitas index akan mempengaruhi efektifitas dan efiseinsi sistem information retrieval. Index dokumen merupakan himpunan term yang menunjukkan isi atau topik yang terdapat pada sebuah dokumen. Index dapat memberikan perbedaan suatu dokumen dari dokumen lain yang berada di dalam koleksi. Ukuran index yang kecil dapat memberikan hasil yang kurang optimal dan dapat mengakibatkan beberapa item

ysng relevan terabaikan. Index yang besar memungkinkan ditemukan oleh banyak dokumen yang relevan tetapi dapat juga menaikkan jumlah dokumen yang tidak relevan serta mengakibatkan kecepatan pencarian menurun.

Pembuatan inverted index harus melibatkan konsep linguistic processing yang bertujuan untuk mengekstrak term-term penting dari dokumen yang direpresentasikan sebagai bag-of-word. Ekstraksi term ini biasanya melibatkan dua operasi utama sebagai berikut:

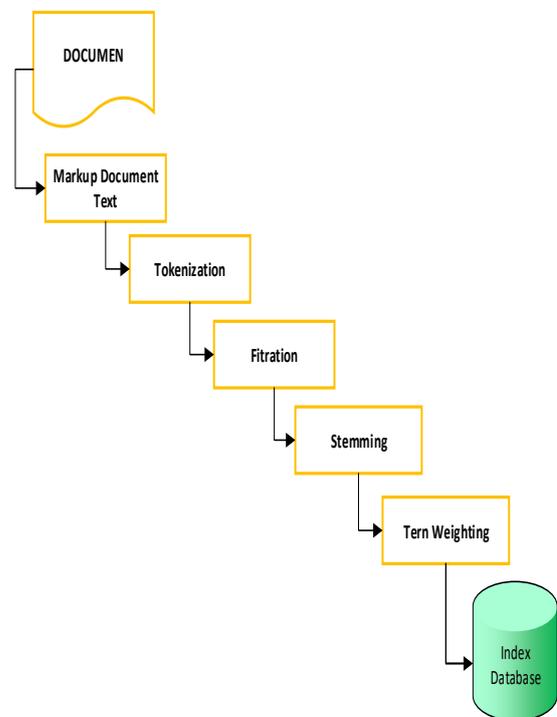
1. Penghapusan *stop-word* diartikan sebagai term yang tidak berhubungan dengan subyek utama dari database meskipun kata tersebut sering kali terdapat di dalam sebuah dokumen. Contohnya adalah dalam Bahasa Indonesia : juga, dari, dia, kami, aku, itu, atau, pada, dengan, dan, maka;

2. Stemming diartikan sebagai kata-kata yang muncul di dalam dokumen sering memiliki varian morfologik. Oleh karenanya, setiap kata yang bukan *stop-word* direduksi ke dalam bentuk *stemmed word (term)* yang cocok. Kemudian kata tersebut di term untuk mendapatkan bentuk akarnya dengan menghilangkan awalan atau akhiran, maka dengan cara seperti inilah diperoleh kelompok kata yang mempunyai makna serupa tetapi berbeda wujud sintaktis antara kata satu dengan kata yang lainnya. Kelompok tersebut dapat direpresentasikan oleh satu kata tertentu. Sebagai contoh kata "diselamatkan", "Menyelamatkan", "selamatkan" adalah satu kelompok dan dapat diwakili oleh satu kata umum yaitu : selamat. Terdapat 5 langkah pembangunan inverted index, diantaranya:

- a. Penghapusan format dan markup dari dalam dokumen. Tahapan ini menghapus semua tag markup dan format khusus dari dokumen, terutama pada dokumen yang mempunyai banyak tag dan format seperti dokumen (x)HTML.
- b. Pemisahan rangkaian kata (*tokenization*). *Tokenization* adalah tahap yang dapat memisahkan serangkaian kata dalam sebuah kalimat, paragraph atau halaman menjadi *token* atau potongan kata tunggal atau *termed word*. Pada tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua token ke dalam bentuk huruf kecil (*lower case*)
- c. Penyaringan (*filtration*). Pada tahapan ini ditentukan term mana yang akan digunakan untuk merepresentasikan dokumen sehingga dapat mendeskripsikan isi dari sebuah dokumen serta dapat membedakan dokumen tersebut dari dokumen lain di dalam kumpulan dokumen. Term yang sering

dipakai tidak dapat digunakan untuk tujuan ini, hal ini disebabkan karena 2 hal. Pertama, karena jumlah dokumen yang relevan terhadap query kemungkinan besar merupakan bagian kecil dari koleksi. Term yang efektif dalam proses pemisahan dokumen yang relevan dari dokumen yang tidak relevan kemungkinan besar adalah term yang muncul pada sedikit dokumen. Kedua, term yang muncul dalam banyak dokumen tidak mencerminkan definisi dari topik atau dari sub-topik dokumen. Oleh karena itu term yang sering digunakan dapat disebut juga sebagai *stop-word* dan dihapus.

- d. Konversi term ke bentuk dasar (*stemming*). *Stemming* adalah proses konversi term kedalam bentuk umumnya. Dokumen juga dapat diekspansi dengan cara mencari sinonim bagi term-term tertentu di dalamnya. Sinonim adalah kata-kata yang memiliki pengertian yang sama tetapi berbeda dari sudut pandang morfologis
- e. Pemberian bobot terhadap term (*weighting*). Setiap term diberikan bobot sesuai dengan skema pembobotan yang dipilih, apakah pembobotan local, global atau kombinasi keduanya. Banyak berbagai macam aplikasi yang menerapkan pembobotan dengan mengkombinasikan berupa perkalian bobot secara *local term frequency* dan *global inverse document* atau dikenal dengan istilah TF-IDF.

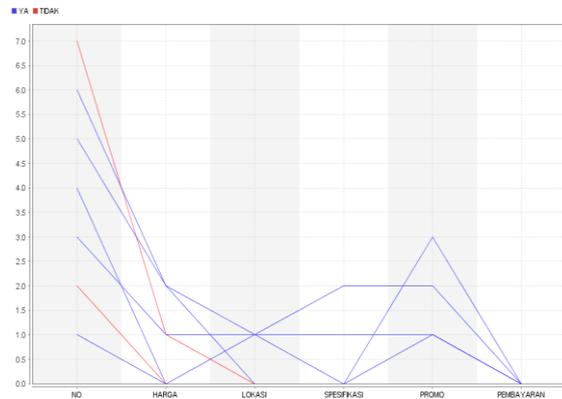


Sumber: Basri & Indrajit (2017)

Gambar 5. Proses Tokenization.

HASIL DAN PEMBAHASAN

Implementasi Naïve Bayes



Sumber: Basri & Indrajit (2017)

Gambar 6. Grafik Pengujian Menggunakan Naïve Bayes

Dari tabel grafik diatas menjelaskan bahwa konsumen akan memilih perumahan yang dibutuhkan dengan ketentuan Harga, Lokasi, Spesifikasi, Promo, serta cara Pembayaran yang ditawarkan oleh *Marketing Developer* Perumahan. Konsumen tidak akan membeli

rumah apabila lokasi perumahan tidak strategis. Hal ini dipengaruhi dengan kenyamanan para konsumen dalam memilih rumah. Secara garis besar konsumen akan tetap membeli rumah sesuai dengan batas kemampuan para konsumen dalam memilihnya.

Adapun hasil akurasi pengujian data training dan data testing dari pengaruh brosur marketing.

Tabel 2. Hasil Akurasi Brosur Marketing.

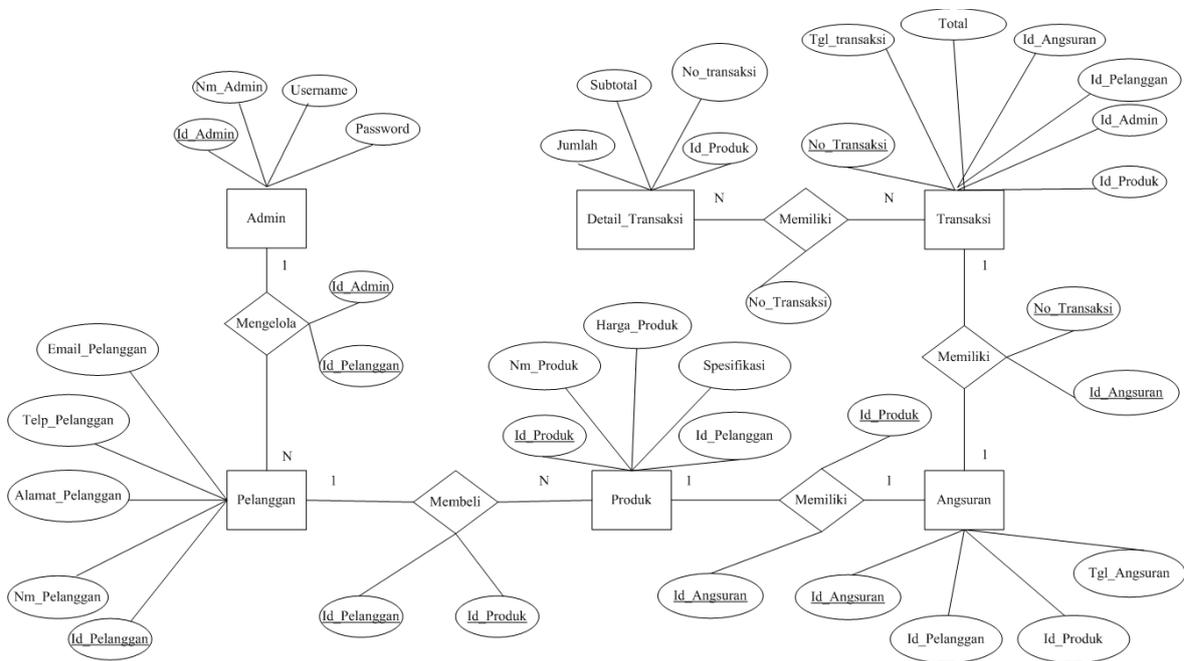
Model	Accuracy
Naïve Bayes	85.71%

Sumber: Basri & Indrajit (2017)

Rancangan Penerapan Vektor Space Model

Rancangan penerapan vektor space model dalam penelitian ini digunakan untuk mempermudah pencarian informasi tentang brosur marketing perumahan. Informasi perumahan yang tersimpan dalam sebuah database akan dapat ditemukan dengan cepat dan tepat. Berikut beberapa langkah-langkah yang harus dilakukan untuk membuat sistem pencarian informasi dengan menggunakan vector space model.

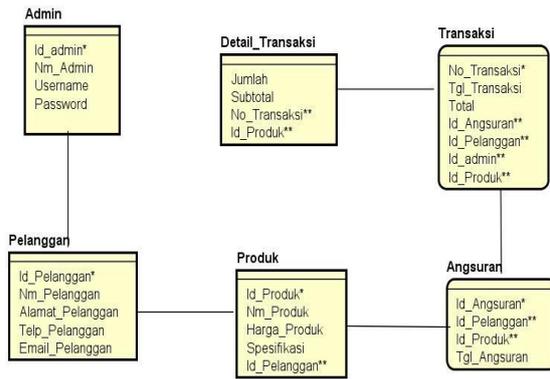
1. ERD



Sumber: Basri & Indrajit (2017)

Gambar 7. Entity Relationship Diagram

2. LRS



Sumber: Basri & Indrajit (2017)
Gambar 8. Logical Relationship Structure.

KESIMPULAN

Keunggulan metode algoritma naïve bayes dengan menggunakan pendekatan backward elimination sangat efektif karena hasil akurasi dapat mencapai lebih dari 80 %. Begitu juga dengan algoritma Vektor Space Model yang dapat mencari informasi yang cepat dengan konsep mencari kemiripan informasi dari sebuah dokumen sampai dengan hasil akhir dari informasi yang dicari. Dari hasil peluang yang didapatkan terkait pengaruh brosur marketing untuk meningkatkan penjualan brosur ternyata hasilnya kurang optimal. Hal tersebut dapat terlihat dari hasil pengujian data menggunakan algoritma Naïve bayes dimana dijelaskan bahwa konsumen akan lebih memilih tempat tinggal yang lebih strategis dibandingkan dengan aspek yang lainnya. Jika diperhatikan ada aspek harga dan spesifikasi rumah yang dipilih. Grafik yang terlihat untuk aspek Harga dan Spesifikasi rumah tidak begitu berpengaruh besar karena harga rumah akan menyesuaikan dari rumah yang dipilih.

Untuk mendapatkan hasil yang optimal maka perlu dipastikan bahwa pada saat menggunakan algoritma naïve bayes harus

memperhatikan keidenpendenan atribut. Asumsi keidenpendenan atribut akan menghilangkan kebutuhan banyaknya jumlah data training dari pencarian perkalian kartesius seluruh atribut yang dibutuhkan untuk mengklasifikasin suatu data. Sebagai bahan penelitian lanjutan penulis menyarankan kepada para peneliti untuk menggunakan algoritma Decision Tree Sebagai pengambilan keputusan dan untuk pencarian informasi menggunakan algoritma *Weight Tree Similarity*.

REFERENSI

Basri, H., & Indrajit, (2017). Laporan Akhir Penelitian Mandiri. Jakarta: STMIK Nusa Mandiri Pascasarjana

Herliana, A. (2016). Optimasi Klasifikasi Sel Tunggal Pap Smear Menggunakan Correlation Based Features Selection (CFS) Berbasis C4 . 5 Dan Naive BAYES, 3 (September), 148-155.

Informatika, T., Malikussaleh, U., & Utara, A. (2014). PENERAPAN ALGORITMA NAIVE BAYES, 8(1), 884-898.

Sailendra, S., & Latar, P. (2014). Strategi pemasaran barang dan jasa perusahaan melalui media iklan, 2.

Sulaehani, R. (2016). Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive, 8 (Desember), 182-189.

Zuniargoprabowo, A., Satoto, K. I., Martono, K. T., Studi, P., Komputer, S., Teknik, F., ... Reality, A. (2015). Perancangan Dan Implementasi Augmented Reality Sebagai, 3 (1), 161-170.