

Implementasi Doc2Vec untuk rekomendasi penginapan di Bali

Nyoman Purnama*¹

^{1,2}STMIK PRIMAKARA; Jalan Tukad Badung 135

³Jurusan Sistem Informasi, STMIK PRIMAKARA, Denpasar

e-mail: *¹ purnama@primakara.ac.id

Abstrak

Dengan semakin menjamurnya hunian berupa villa di Bali, memberikan pilihan yang beragam bagi para wisatawan yang akan menginap di Bali. Berbagai macam pilihan yang ada seringkali membingungkan wisatawan untuk memilih penginapan sesuai dengan kriteria yang diinginkan. Informasi yang tersedia di internet pun sangat beragam dengan jumlah tak terbatas. Apalagi dengan layanan pemesanan online yang juga banyak bertebaran di internet, wisatawan seringkali harus mengunjungi situs-situs ini satu persatu. Rekomendasi penginapan secara otomatis sangat diperlukan untuk membantu wisatawan dalam mempersempit pilihannya dalam mencari penginapan. Maka pada penelitian ini dibangun sebuah system rekomendasi menggunakan algoritma doc2Vec. Doc2Vec merupakan algoritma pembelajaran tidak terawasi yang akan mengubah dokumen menjadi vektor. Sistem rekomendasi ini dikembangkan menggunakan Bahasa python. Dataset yang digunakan merupakan deskripsi kumpulan villa yang tersebar diseluruh Bali yang berasal dari salah satu agen penyewaan villa. Dari hasil penelitian yang telah dilaksanakan terbukti algoritma doc2vec memberikan hasil rekomendasi terbaik berdasarkan kueri yang diminta oleh pengguna. Dimana metode PV-DM memberikan hasil yang lebih baik jika dibandingkan metode PV-DBOW dengan tingkat kemiripan rata-rata diatas 60%.

Keywords: villa, word embedding, doc2vec, rekomendasi

Abstract

With the proliferation of occupancy in the form of villas in Bali, providing a variety of choices for tourists who will stay in Bali. The wide range of available options often confuses tourists to choose lodging according to the desired criteria. The information available on the internet is also very diverse with an unlimited amount. Especially with online ordering services that are also widely scattered on the internet, tourists often have to visit these sites one by one. Automatic lodging recommendations are needed to assist tourists in narrowing their choices in finding lodging. So in this study a recommendation system was built using the doc2Vec algorithm. Doc2Vec is an unsupervised learning algorithm that will convert documents into vectors. This recommendation system was developed using the Python language. The dataset used is a description of a collection of villas scattered throughout Bali from one of the villa rental agents. From the results of the research that has been carried out, it is proven that the doc2vec algorithm provides the best recommendation results based on the queries requested by the user. Where the PV-DM method gives better results when compared to the PV-DBOW method with an average similarity level of above 60%.

Keywords—Villa, recommendation, word embedding, doc2vec

1. PENDAHULUAN

Bali merupakan salah satu destinasi wisata yang terbaik di Indonesia. Dengan jumlah wisatawan yang datang setiap tahunnya semakin meningkat sebelum adanya pandemi[1]. Menginap di sebuah fasilitas penginapan biasa dilakukan ketika Anda tengah bepergian ke Bali. Beberapa fasilitas penginapan yang sangat menjamur di Bali diantaranya adalah villa. Villa merupakan sebuah fasilitas tempat tinggal sementara yang biasanya terletak di daerah yang berhawa sejuk, misalnya seperti di pegunungan, dataran tinggi, maupun pinggir pantai. Villa umumnya ditinggali pada akhir pekan atau musim liburan sebagai sarana penunjang liburan. Villa juga menyediakan berbagai fasilitas hiburan seperti kolam renang taman bunga yang indah, area bermain anak, danau atau kolam pemancingan, fasilitas olahraga, serta berbagai sarana rekreasi lainnya. Villa

juga dilengkapi dengan peralatan dan fasilitas penunjang kegiatan sehari-hari layaknya di rumah Anda sendiri[2].

Jumlah villa yang semakin menjamur di pulau Bali, sering menyulitkan para wisatawan untuk memilih villa yang tepat sesuai kriteria yang diinginkan. Diperlukan banyak kriteria bagi para wisatawan untuk memilih villa yang sesuai. Seringkali para agen penginapan akan membantu para wisatawan untuk memberikan alternatif terbaik sesuai dengan pilihan awal yang diinginkan oleh wisatawan[2]. Disinilah system rekomendasi sangat dibutuhkan, system rekomendasi akan memberikan rekomendasi villa dengan memperhitungkan nilai untuk poin-poin penting pada saat melakukan penilaian. Rekomendasi memberikan nilai lebih jika dibandingkan dengan sistem pencarian biasa dimana hasil yang diberikan tidak terduga oleh pengguna tetapi ada[3]. Rekomendasi ini diperoleh berdasarkan deskripsi dari masing-masing villa. Agen penyewaan villa menggunakan deskripsi villa untuk memberikan keterangan dari setiap villa yang akan disewakan.

Salah satu contoh algoritma rekomendasi yang bisa digunakan yakni Doc2Vec. Doc2vec merupakan pengembangan dari algoritma word2Vec, dimana word2vec menangkap pengertian kata dengan kata yang muncul disekelilingnya[4]. Doc2vec menambahkan konteks didalamnya[5]. Doc2vec menggunakan algoritma neural untuk menjaga konteks dan semantic dalam suatu kata. Doc2vec juga merupakan Teknik content based filtering yang menggunakan ketersediaan konten sebuah item sebagai basis dalam pemberian rekomendasi[6]. Teknik lain untuk memperoleh rekomendasi sudah pernah dilakukan pada penelitian sebelumnya dengan bantuan system pendukung keputusan dan metode fuzzy mamdani[7]. Adapun tujuan dari penelitian adalah mengimplementasikan algoritma doc2Vec untuk merekomendasikan villa berdasarkan deskripsi yang ada dengan mencari kemiripan dari permintaan pengguna dengan dataset yang ada. Bagi para wisatawan hasil rekomendasi ini memberikan kemudahan untuk mendapatkan villa yang diinginkan. Sedangkan bagi agen, sistem rekomendasi akan mempermudah dalam memberikan daftar villa sesuai keinginan wisatawan.

2. METODE PENELITIAN

Adapun metode yang digunakan dalam penelitian ini yakni pertama kali akan dilakukan proses penyiapan data, pre-processing, membuat model dan menghitung kemiripan menggunakan cosine similarity.

2.1 Dataset

Dalam penelitian ini digunakan dataset berupa data villa yang didapatkan dengan metode web scrapping dari salah satu website agen penyewaan villa di Bali, dengan jumlah dataset sebesar 890 data villa. Bahasa yang digunakan pada website ini yaitu Bahasa Inggris. Proses web scrapping ini dilakukan dengan bantuan plugin dari salah satu browser internet. Dataset yang dihasilkan terdiri dari beberapa kolom yang merupakan atribut dari masing-masing villa yakni nama villa, jumlah bedroom, deskripsi villa, deskripsi bedroom, deskripsi outdoor, deskripsi indoor, lokasi, lokasi gambar. Data utama yang akan digunakan dalam proses pelatihan yakni kolom deskripsi baik itu dari bedroom, outdoor dan indoor. Dengan panjang data dari masing-masing kolom bervariasi. Dataset hasil scrapping diperoleh dalam bentuk format csv, format ini kemudian dirubah terlebih dahulu menjadi format json yang akan diolah selanjutnya pada tahapan pre processing.

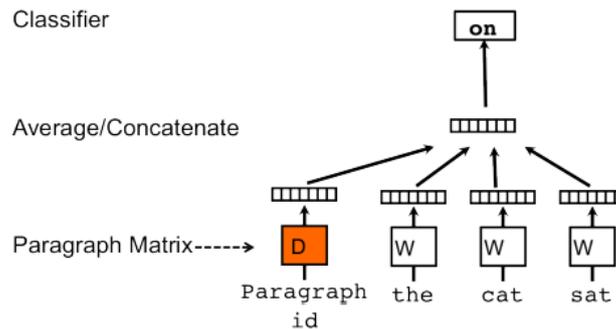
2.2 Pre Processing

Langkah selanjutnya setelah dataset berhasil dikumpulkan yakni proses pengolahan dataset yang berupa teks kolom deskripsi dari masing-masing villa. Langkah pertama sebelum data digunakan pada proses doc2vec yakni membangun "tagging" yang akan membuat pemetaan dari masing masing kalimat dengan index yang unik. Preprocessing juga melibatkan proses pengolahan kata dengan text mining seperti tokenizing, Stop words removal dan stemming[8], sebelum kata ini diproses menjadi vektor dengan algoritma Doc2vec. Tokenizing akan memisahkan deskripsi villa menjadi per kata dan menjadi sebuah tag dari masing-masing dokumen. Stop words removal merupakan Teknik untuk menghilangkan kata stop words atau kata yang tidak memiliki arti dalam pengolahan nantinya. Stemming akan mengembalikan tag kedalam bentuk dasarnya.

2.3 Doc2Vec

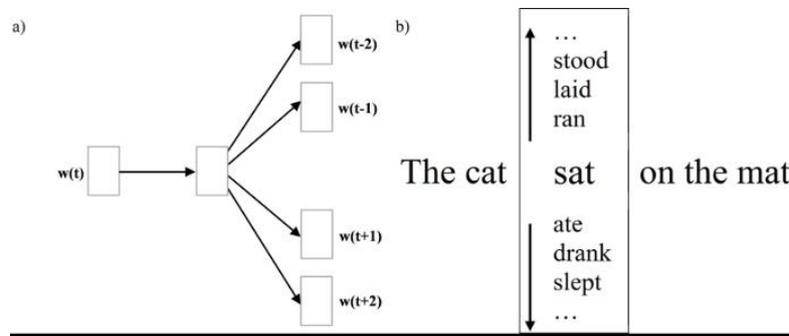
Pembuatan model merupakan bagian utama pada proses pembuatan system rekomendasi. Pada penelitian ini menggunakan doc2Vec sebagai metode untuk mentransformasikan dokumen kedalam vektor. Doc2Vec merupakan algoritma yang berbasis word2vec yang menggambarkan sebuah kata kedalam vektor

yang bisa menyesuaikan dengan kalimat dan paragraf serta dokumen[9]. Doc2vec juga menggunakan neural network untuk menjaga semantic dan konteks dari suatu kata. Dan berdasarkan penelitian, memberikan peningkatan 30% terhadap klasifikasi teks. Hal ini sangat sesuai dengan dataset yang digunakan berupa deskripsi villa yang memerlukan semantic dan konteks dalam proses mencari kemiripan. Seperti pada algoritma word2vec, doc2vec memiliki 2 model berbeda yang bisa digunakan yakni Distributed Memory version of Paragraph Vector (PV-DM) dan Words version of Paragraph Vector (PV-DBOW). PV-DM memiliki kesamaan pada model Continuous Bag of Words pada model word2Vec tetapi terdapat sebuah vektor tambahan yang ditambahkan untuk menggambarkan keunikan dari sebuah dokumen. Vektor bernama "W" memegang representasi numerik dan mewakili konsep sebuah kata, Demikian pula, vektor bernama 'D' adalah vektor dokumen yang memegang representasi numerik dan mewakili konsep dokumen.



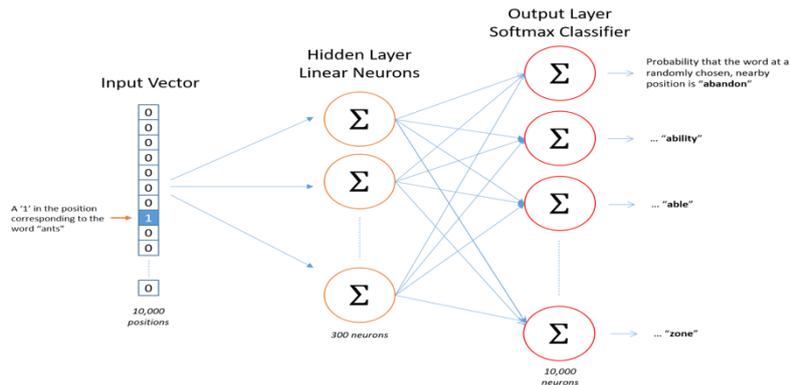
Gambar 1. Distributed Memory version of Paragraph Vector (PV-DM)

Sedangkan versi skip gram dari word2vec adalah Words version of Paragraph Vector (PV-DBOW). Model skip gram akan mencoba untuk memprediksi konteks sekelilingnya dengan target kata yang diberikan. Pada gambar 2 merupakan gambaran dari skip gram.



Gambar 2. Words version of Paragraph Vector (PV-DBOW)

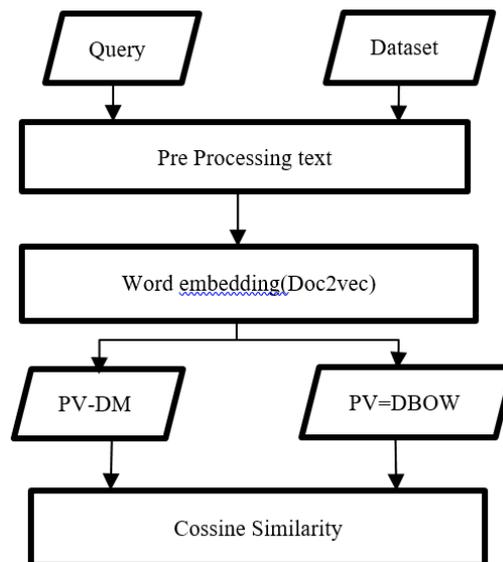
Pada penelitian ini digunakan model PV-DBOW yang merupakan ekstensi dari skip gram karena berdasarkan penelitian [9], memiliki kecepatan komputasi yang lebih cepat karena mengkonsumsi memori yang lebih rendah. Pada penelitian menggunakan skip gram, dimana pada saat pertama kali penulis akan memasukkan kata dan dapat beberapa sebagai balasannya. Doc2Vec sama seperti word2vec yang merupakan jaringan syaraf tiruan yang memiliki beberapa lapisan diproyeksikan didalam jaringan ini[8]. Pada gambar 3 merupakan model skip gram yang digunakan pada penelitian ini. Terdapat 3 layer pada jaringan neural ini yaitu yang pertama input layer yang akan menerima kata yang ingin dicari. Input layer ini berupa matriks berisi semua kata-kata dalam sebuah korpus. Pada lapisan tersembunyi terdapat matriks yang berisi bobot untuk setiap kata pada neuron. Layer output akan berisi perkalian antara matriks input dan hidden layer berupa representasi vektor. Pada layer output, nilai matriks hasil perkalian tadi diproses menggunakan *softmax regression classifier*, dimana setiap neuron emngembalikan nilai 0 dan 1.



Gambar 3. Model skip gram[8]

2.4 Alur penelitian

Alur proses system rekomendasi ini dimulai dengan pre processing dataset seperti yang telah dijelaskan sebelumnya. Setelah itu dilakukan proses representasi kata menjadi vektor menggunakan metode doc2vec. Disini kita akan menggunakan 2 arsitektur doc2vec yakni PV-DM dan PV-DBOW. Kedua arsitektur ini akan dicari perbandingannya dengan jumlah dataset yang berbeda yakni 100,200,300 dan 400. Pengujian dilakukan dengan mencari nilai kesamaan menggunakan cosine simmilarity antara query dan dataset yang digunakan. Berikut pada gambar 4 merupakan alur proses rekomendasi secara garis besar.



Gambar 4. Alur Penelitian

Pada penelitian ini tahapan pre proses teks yang dilakukan yakni dengan mengolah permintaan pengguna dan deskripsi villa dengan beberapa tahapan yakni tokenisasi, stop words dan stemming. Proses ini dilakukan sebelum teks ini dirubah menjadi nilai numerik yang dapat diolah dan juga supaya data teks tersebut lebih bersih dari kata-kata yang diperlukan.



Gambar 5. Alur pemrosesan teks

3. HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai implementasi metode doc2vec dengan menggunakan library gensim yang terdapat pada python. Kemudian dilanjutkan dengan pengujian dengan menggunakan arsitektur berbeda dan dataset yang berubah-ubah. Penelitian ini bertujuan untuk mengimplementasikan metode doc2vec menggunakan bahasa python. Dimana dataset yang digunakan adalah deskripsi villa yang diperoleh menggunakan Teknik web scrapping. Web yang akan discrapping merupakan website booking agent villa di Bali berbahasa Inggris. Dataset yang digunakan berupa deskripsi dari villa-villa yang ada di beberapa daerah wisata di Bali yakni Seminyak, Canggu, Ubud, Uluwatu, Nusa dua dan Kerobokan. Proses pelatihan dan pengujian dilakukan menggunakan Bahasa python. Dimana Bahasa ini merupakan Bahasa tingkat tinggi yang memiliki banyak library. Sebagian besar library diinstalasi dengan menggunakan PIP. Pip merupakan package manager yang mengatur paket instalasi di python. Pada penelitian ini digunakan library gensim yang didalamnya terdapat fungsi feature extraction menggunakan doc2vec. Berikut contoh hasil web scrapping dari website agen villa di Bali dengan menggunakan library Python:

-scraper-order	web-scraper-start-url	villa-link	villa-link-href	nama_villa	jumlah_bedroom	villa_deskripsi	bedrooms_deskripsi	outdoor_deskripsi
595486-443	https://www.villa-bali.com/en/find	More details	https://www.villa-bali.com/en/villa/canggu/sun...	Sungai Jungle Villa I	3 bedrooms	A place like Sungai Jungle Villa I is not some...	Sungai Jungle Villa I is home for 3 finely app...	Sungai Jungle Villa I boasts an exclusive use ...
595765-552	https://www.villa-bali.com/en/find	More details	https://www.villa-bali.com/en/villa/uluwatu/mo...	Morabito Art Cliff	9 bedrooms	"Wow" is the best word to describe what everyo...	Santorini: Massalia	Every one of these gorgeous lofts has a terrac...
595709-530	https://www.villa-bali.com/en/find	More details	https://www.villa-bali.com/en/villa/nusa-dua/v...	Villa Vinla Penthouse	4 bedrooms	The penthouse suite with 4 bedrooms at Villa V...	The penthouse has 4 bedrooms and all of them a...	The penthouse villa has a private swimming poo...
595342-388	https://www.villa-bali.com/en/find	More details	https://www.villa-bali.com/en/villa/canggu/vil...	Villa Sally	4 bedrooms	Villa Sally is a luxury retreat nestled in a q...	Villa Sally houses two master bedrooms on grou...	Stepping stones over ponds with rippling fount...
596348-783	https://www.villa-bali.com/en/find	More details	https://www.villa-bali.com/en/villa/seminyak/v...	Villa Ibiza	3 bedrooms	Villa Ibiza is a beautiful 3-bedroom luxury vi...	Villa Ibiza houses 3 bedroom suites, each feat...	Wooden-carved doors open up to the garden of t...

Gambar 5. Hasil web scrapping

Dari hasil web scrapping ini diambil semua kolom deskripsi dari masing-masing villa. Selanjutnya dilakukan pembuatan model doc2vec. Proses pembuatan model doc2vec ini dinamakan dengan proses pelatihan/training. Pada proses pelatihan ini akan menghasilkan matriks antar kata dengan dimensi m x d. Dimana m merupakan jumlah kata yang ada, sedangkan d adalah jumlah fitur/dimensi vektor. Setelah didapatkan vektor antara permintaan oleh pengguna dan data deskripsi villa, kemudian dilakukan perhitungan kemiripan dua vektor tersebut. Permintaan oleh pengguna merupakan sebuah kumpulan kata-kata berupa keinginan pengguna akan penginapan villa yang diinginkan. Permintaan ini pada umumnya berisi lokasi tujuan, jumlah kamar tidur dan beberapa permintaan lainnya yang bervariasi. Mengingat banyaknya variasi dan atribut dari sebuah villa, dengan adanya rekomendasi ini maka akan mempersempit hasil pencarian berdasarkan kemiripan deskripsi dengan permintaan pengguna. Hasil kemiripan ini akan dihitung nilai kesalahan/errornya. Data yang digunakan pada saat pelatihan dibagi menjadi 100, 200 dan 300. Selain perbedaan jumlah data latih yang digunakan, dua arsitektur doc2vec juga akan dievaluasi performanya dengan nilai cosine similarity-nya. Nilai cosine similarity dihitung berdasarkan kesamaan nilai antara query dengan dataset yang digunakan. Pada penelitian ini menggunakan fungsi most_similar dari library Gensim python.

```
# Calculate the cosine similarity and return top 20 matches
score = model.docvecs.most_similar([model.infer_vector(test_docs)], topn=20)
```

Query yang akan digunakan berbahasa Inggris, yang berhubungan dengan permintaan penyewaan villa. Query yang dimasukkan kedalam system akan ditokenisasi dan dilakukan proses pengolahan kata. Berikut contoh permintaan dari pengguna yang digunakan pada penelitian ini :

Tabel 1. Contoh query yang digunakan

Contoh Query
"canggu villa for rent great view"
"I would like to rent a villa in seminyak 2 bedroom great pool"
I would like to enquire about the Vacation Villas in Ubud
Any villa for rent in Kerobokan, for big family

A. Pengujian dengan arsitektur PV-DM

Pengujian pertama dilakukan dengan model PV-DM(Distributed Memory version of Paragraph Vector). Dimana pada arsitektur ini sebagai memori yang mengingat apa yang hilang dari konteks saat ini (topik paragraf). Model ini akan mencoba memprediksi pusat dari kata dari sekumpulan kata, dengan mempertimbangkan input kata konteks dan id dari sebuah paragraf. Pada library genism model ini bisa diterapkan dengan merubah nilai pada saat melakukan inisialisasi model dengan merubah nilai parameter dm dengan nilai 1. Kemudian dilakukan perubahan data latih yakni 100, 200, 300 dan 400. Berikut merupakan hasil pengamatan dengan menggunakan model PV-DM.

Tabel 2. Pengujian cosine similarity pada model PV-DM

Jumlah data latih	Cossine simmiliarity	Akurasi
100	63,9%	35%
200	72.5%	65%
300	72%	80%
400	67%	45%

Setelah dilakukan pengujian dengan data latih yang berbeda pada metode PV-DM diperoleh nilai cosine similarity rata-rata diatas 60%. Hal ini menandakan system mampu memberikan rekomendasi berdasarkan tingkat kemiripan antara kueri dan dataset villa. Dimana dengan metode PV-DM dengan semakin besar jumlah data latih tidak mempengaruhi nilai cosine similarity, hal ini dipengaruhi tingkat variasi dataset yang digunakan. Dimana dengan semakin besar variasi tingkat variasi data dengan kueri yang diberikan, bisa mengakibatkan berkurangnya kemampuan Doc2vec dalam melakukan proses pelatihan. Nilai kemiripan dan akurasi terbaik diperoleh pada saat dataset yang digunakan sejumlah 300 data latih.

B. Pengujian dengan arsitektur PV-DBOW

Pengujian kedua dilakukan dengan model DBOW, dimana dengan model ini sedikit berbeda dengan model PV-DM. Model DBOW mengabaikan konteks data pada input dan memaksakan model untuk memprediksi kata secara acak dari data output. Pada library genism model ini bisa diterapkan dengan merubah nilai pada saat melakukan inisialisasi model dengan merubah nilai parameter dm dengan nilai 0. Kemudian dilakukan perubahan data latih yakni 100, 200, 300 dan 400. Berikut merupakan hasil pengamatan dengan menggunakan model PV-DBOW.

Tabel 3. Pengujian cosine similarity pada model PV-DBOW

Jumlah data latih	Cossine simmiliarity	Akurasi
100	51%	30%
200	45%	40%
300	64%	30%
400	70%	35%

Setelah dilakukan pengujian dengan data latih yang berbeda pada metode PV-DBOW diperoleh nilai cosine similarity rata-rata diatas 50%. Nilai kemiripan yang diperoleh lebih kecil dibandingkan dengan metode PV-DM. Namun dengan metode PV-DBOW proses pelatihan dilakukan lebih cepat dibandingkan dengan

metode PV-DM. Hal ini sesuai dengan penelitian yang telah dilakukan oleh Abhinav Roy[10], dimana metode PV-DM lebih akurat dibandingkan dengan metode DBOW.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, pembuatan rekomendasi dengan menggunakan algoritma doc2vec memberikan hasil rekomendasi yang cukup baik dengan tingkat kemiripan diatas 50%. Pada penelitian dengan 2 model yang berbeda yakni PV-DM dan PV-DBOW diperoleh kesimpulan bahwa metode PV-DM memiliki tingkat kemiripan yang lebih baik dibandingkan PV-DBOW. Secara garis besar vektor yang dihasilkan oleh doc2vec mampu mewakili sebuah dokumen dengan baik. Dimana dataset yang digunakan saat pelatihan dengan doc2vec memiliki peranan penting yang akan menentukan hasil rekomendasi yang diberikan.

5. SARAN

Adapun saran dari penelitian ini kedepannya adalah pembuatan deployment hasil model yang telah dikerjakan dalam bentuk web, sehingga bisa diakses dengan lebih mudah. Serta penambahan jumlah dataset untuk memperbaiki hasil rekomendasi yang diberikan, meliputi tambahan villa di daerah wisata lainnya di Bali.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung dalam pembuatan penelitian ini.

DAFTAR PUSTAKA

- [1] A. Andriani, "Sistem Rekomendasi Promosi Hotel Pada Wisatawan mancanegara Berbasis Data Mining," *Semnasteknomedia Online*, vol. 2, no. 1, pp. 2-02–37, 2014, [Online]. Available: <https://www.ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1087>.
- [2] G. A. Rahmadanu, E. Santoso, and Sutrisno, "Implementasi Naïve Bayes dan Weighted Product Dalam Memberi Rekomendasi Hotel Terbaik Saat Berwisata Di Bali," vol. 3, no. 2, pp. 1617–1624, 2019.
- [3] A. N. Laili, P. P. Adikara, and S. Adinugroho, "Rekomendasi Film Berdasarkan Sinopsis Menggunakan Metode Word2Vec," vol. 3, no. 6, pp. 6035–6043, 2019.
- [4] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, and Z. Abidin, "Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks," *J. Tekno Kompak*, vol. 14, no. 2, p. 74, 2020, doi: 10.33365/jtk.v14i2.732.
- [5] M. Marcińczuk Michał and Gniewkowski, T. Walkowiak, and M. Błedkowski, "Text Document Clustering: {W}ordnet vs. {TF}-{IDF} vs. Word Embeddings," *Proc. 11th Glob. Wordnet Conf.*, pp. 207–214, 2021, [Online]. Available: <https://www.aclweb.org/anthology/2021.gwc-1.24>.
- [6] A. E. Wijaya and D. Alfian, "Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering," *J. Comput. Bisnis*, vol. 12, no. 1, pp. 11–27, 2018.
- [7] Y. P. Lestari, T. Komputer, S. Informasi, L. Keja, and K. Sama, "Implementasi Logika Fuzzy Mamdani," no. x, pp. 106–115, 2012.
- [8] I. Syarif, R. Asmara, and B. Dewangkara, "Rekomendasi Kendaraan Roda 4 Berdasarkan Tweet Customer Menggunakan Word2Vec," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 1, p. 58, 2020, doi: 10.35314/isi.v5i1.1096.
- [9] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," *31st Int. Conf. Mach. Learn. ICML 2014*, vol. 4, pp. 2931–2939, 2014.
- [10] "Getting started with Doc2Vec. Hands-on guide for building your own... | by Abhinav Roy | Heartbeat." <https://heartbeat.comet.ml/getting-started-with-doc2vec-2645e3e9f137> (accessed Oct. 04, 2021).