



Rekomendasi Destinasi Wisata di Indonesia Menggunakan Metode Item2Vec

Aisha Kusuma Putri, Suliadi Suliadi*

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ARTICLE INFO

Article history :

Received : 10/2/2023

Revised : 26/06/2023

Published : 13/7/2023



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Volume : 3

No. : 1

Halaman : 11-18

Terbitan : Juli 2023

ABSTRAK

Item2Vec merupakan sistem rekomendasi item yang hanya menggunakan data rating pengguna dalam algoritmanya. Item yang direkomendasikan kepada pengguna yaitu item yang memiliki kemiripan dengan item yang disukai sebelumnya. Dalam algoritmanya, Item2Vec dibangun dengan collaborative filtering (CF) dan Word2Vec Skip Gram with Negative Sampling (SGNS). Melalui Word2Vec SGNS, direpresentasikannya item ke dalam vektor angka melalui proses neural network yaitu feedforward, meminimumkan fungsi objektif dan backpropagation. Melalui CF, vektor setiap item dihitung

menggunakan cosine similarity sebagai nilai kemiripan antar item. Hasil nilai cosine similarity tertinggi diartikan sebagai item yang paling direkomendasikan untuk setiap item dan juga sebaliknya. Item2Vec memiliki akurasi lebih besar dan kinerja lebih efektif dibandingkan sistem rekomendasi yang dibangun oleh Singular Value Decomposition (SVD) dengan CF. Tujuan dari penelitian ini adalah memberikan rekomendasi destinasi wisata di Indonesia, dengan menggunakan metode Item2Vec. Metode ini mempermudah wisatawan dalam memilih destinasi wisata dan juga mempermudah biro perjalanan dalam menawarkan destinasi wisata untuk keperluan promosi. Metode Item2Vec menghasilkan rekomendasi destinasi wisata di Indonesia dengan akurasi sebesar 61% dan termasuk ke dalam performa cukup baik.

Kata Kunci : Collaborative Filtering, Destinasi Wisata; Item2Vec.

ABSTRACT

Item2Vec is an item recommendation method that only uses user rating data in its algorithm. The items recommended to users are items that are similar to items previously liked. In its algorithm, Item2Vec is built with collaborative filtering (CF) and Word2Vec Skip Gram with Negative Sampling (SGNS). Through Word2Vec SGNS, it represents items into a vector of numbers through a neural network process, that is feedforward, that minimizing the objective function and backpropagation. Through CF, the vector for each item will be calculated using cosine similarity as the similarity value between items. The highest cosine similarity value results are interpreted as the most recommended items for each item and otherwise. Item2Vec has greater accuracy and more effective performance than recommendation systems built by Singular Value Decomposition (SVD) with CF. The objective of this research is to provide recommendation for tourist destination in Indonesia, by using Item2Vec method. This method makes it easier for tourists to choose tourist destinations and also makes it easier for travel agents to offer tourist destinations for promotion purposes. The Item2Vec method produces recommendations for tourist destinations in Indonesia with an accuracy of 61% and is considered to be performing quite well.

Keywords : Collaborative Filtering; Tourist Destination; Item2Vec.

© 2023 Jurnal Riset Statistika Unisba Press. All rights reserved.

A. Pendahuluan

Machine learning adalah bagian dari kecerdasan buatan yang mengolah data besar untuk digunakan dalam segala aspek kehidupan. *Machine learning* memiliki banyak metode yang dirancang untuk mencari pola keputusan yang cepat dan akurat, salah satunya yaitu sistem rekomendasi [1]. Dalam sistem rekomendasi, dicarinya pola keinginan seseorang dalam memilih atau membeli suatu *item*. Sistem rekomendasi memiliki beragam jenis, salah satunya yang populer yaitu *collaborative filtering* atau CF [2]. CF merupakan sistem rekomendasi berdasarkan *rating* dari pengguna. *Rating* merupakan informasi yang cukup akurat karena sesuai dengan opini pribadi terkait kesukaan pengguna terhadap suatu *item*. CF digunakan dalam sistem rekomendasi Item2Vec dengan memanfaatkan *rating* suka dan tidak suka dari pengguna. Selain dibangun oleh CF, Item2Vec menggunakan Word2Vec *Skip Gram with Negative Sampling* (SGNS) dalam merepresentasikan *item* menjadi vektor angka. Algoritma Item2Vec identik dengan algoritma Word2Vec SGNS yaitu menggunakan proses *neural network* untuk menghasilkan vektor angka bagi setiap *item* [3]. Vektor angka kemudian dihitung menggunakan nilai *cosine similarity* pada proses CF.

Dengan hanya menggunakan informasi *rating*, Item2Vec dinilai lebih efektif dibandingkan sistem rekomendasi yang dibangun oleh *Singular Value Decomposition* (SVD) dan CF [3]. Sistem rekomendasi SVD dengan CF selain menggunakan *rating*, sistem tersebut memerlukan informasi karakteristik pengguna seperti usia dan pekerjaan. Selain itu, Item2Vec juga cukup baik dalam nilai akurasinya yaitu sebesar 68% dibandingkan dengan sistem rekomendasi SVD dengan CF yaitu sebesar 54% [3].

Item2Vec dapat digunakan dalam bidang apapun selama tersedianya data *rating* dari pengguna, salah satunya yaitu bidang pariwisata. Pariwisata merupakan salah satu sumber perekonomian Indonesia sehingga banyak wisatawan lokal dan asing berdatangan. Namun, adanya pandemi *Covid-19* selama dua tahun membuat pariwisata menurun sebesar 80% [4]. Pada tahun 2022 kondisi global sudah membaik dari pandemi *Covid-19*, sehingga Kementerian Pariwisata dan Ekonomi Kreatif berupaya menghidupkan kembali pariwisata Indonesia dengan melakukan promosi [5]. Promosi akan berhasil ketika seseorang tertarik dengan sesuatu yang ditawarkan. Item2Vec dapat diterapkan untuk membuat promosi berhasil yaitu dengan memberikan rekomendasi destinasi wisata yang sesuai dengan keinginan wisatawan. Rekomendasi yang diberikan yaitu destinasi wisata yang mirip dengan destinasi wisata yang wisatawan sukai sebelumnya. Selain itu, rekomendasi ini bermanfaat bagi biro perjalanan dalam menstrategikan destinasi wisata yang akan ditawarkan kepada wisatawan [6]. Berdasarkan latar belakang yang telah diuraikan, tujuan dalam penelitian ini yaitu memberikan rekomendasi destinasi wisata di Indonesia dengan menggunakan metode Item2Vec dan memberikan rekomendasi yang memiliki akurasi cukup baik.

B. Metode Penelitian

Item2Vec

Item2Vec merupakan sistem rekomendasi *item* dengan cara merepresentasikan *item* ke dalam vektor angka yang kemudian dicari hubungannya. Item2Vec dibangun dengan *collaborative filtering* (CF) dan metode Word2Vec *Skip-Gram with Negative Sampling* (SGNS). Word2Vec SGNS dibangun dengan proses *neural network* yang terdiri atas *feedforward*, meminimumkan fungsi objektif dan *backpropagation* [7]. Word2Vec SGNS berperan sebagai merepresentasikan *item* ke dalam bentuk angka. Sedangkan CF berperan membuat rekomendasi dengan menghitung nilai *cosine similarity* antar *item* yang sebelumnya telah diubah menjadi bentuk angka. Rekomendasi diseleksi menjadi K *item* berdasarkan nilai *cosine similarity* terbesar sampai terkecil (*Top-K Recommendation*). Proses Item2Vec diantaranya yaitu tahapan pra analisis (*split data* dan persiapan data), kemudian tahapan analisis (*neural network* dan *collaborative filtering*), dan terakhir yaitu tahapan evaluasi model [8].

Tahapan Pra Analisis Item2Vec

1) Melakukan Split Data. Data dibagi menjadi data *training* dan *testing* dengan proporsi sebesar 80:20. Proporsi tersebut dipilih karena merupakan proporsi yang umum digunakan pada *neural network* dan menurut [9] mampu menghasilkan nilai akurasi yang paling tinggi dibandingkan proporsi lainnya pada penerapan *neural network*. Data *training* digunakan pada tahapan analisis, sedangkan data *testing* digunakan pada

evaluasi model. 2) Mempersiapkan Data. Data dibagi menjadi daftar positif dan negatif. Daftar positif berisi *item* yang disukai oleh pengguna (*item* positif), sedangkan daftar negatif berisi *item* yang tidak disukai oleh pengguna (*item* negatif). Penentuan *item* positif dan negatif sesuai dengan kriteria peneliti. Untuk skala penilaian 1 sampai 5, daftar positif terdiri atas *item* yang diberi nilai 4 dan 5, sedangkan daftar negatif terdiri atas *item* yang diberi nilai 1, 2 dan 3 [10].

Tahapan Analisis Item2Vec

Menurut Barkan & Koenigstein [3], tahapan Item2Vec sama dengan Word2Vec SGNS. Di bawah ini merupakan tahapan Word2Vec SGNS menurut Rong [7] yang digunakan pada tahapan Item2Vec.

Feedforward

Feedforward adalah proses *neural network* untuk menghubungkan *input layer*, *hidden layer* dan *output layer* melalui nilai bobot. 1) Menentukan *item* target positif ($x_k^{(+)}$) dan *item* konteks positif ($x_j^{(+)}$). *Item* target positif ($x_k^{(+)}$) adalah *item* yang menjadi fokus dalam pencarian rekomendasi, sedangkan *item* konteks positif ($x_j^{(+)}$) adalah *item* yang berada dalam daftar positif dengan mengeluarkan *item* target positif. Banyaknya *item* konteks disebut sebagai *window size*, yang ditentukan berdasarkan banyaknya *item* dalam daftar positif. Pada Item2Vec, tidak masalah jika urutan *item* dalam satu set *item* diacak karena dibuangnya informasi spasial sehingga urutan atau waktu pemberian *rating* pada *item* tidak bermakna[3]. 2) Menentukan sampel *item* negatif. Dalam menentukan *item* target dan konteks untuk daftar negatif diperlukannya menghitung sampel menggunakan persamaan di bawah ini.

$$P_n(x_j)^{3/4} = \left(\frac{fрек(x_j)}{fрек(x_t)} \right)^{3/4} \tag{1}$$

Di mana $fрек(x_j)$ adalah frekuensi munculnya *item* j dalam seluruh data, $fрек(x_t)$ adalah frekuensi seluruh data, sedangkan $P_n(x_j)^{3/4}$ adalah peluang kemunculan *item* j dalam data dengan $j = 1, 2, \dots, n$. Untuk nilai $P_n(x_j)^{3/4}$ yang paling tinggi diartikan dengan paling sering muncul pada seluruh data dan cenderung dipilih. Adapun banyaknya *item* negatif yang dipilih yaitu sebanyak *item* dalam daftar positif dikurangi satu. Namun, jika *item* dalam daftar negatif lebih sedikit dibandingkan *item* dalam daftar positif maka semua *item* negatif diambil. *Item* negatif yang terpilih dinotasikan dengan $X_{ns}^{u,(-)} = \{x_1^{u,(-)}, x_2^{u,(-)}, \dots, x_n^{u,(-)}\}$. (3) Menentukan *item* target negatif ($x_k^{(-)}$) dan *item* konteks negatif ($x_j^{(-)}$). *Item* target negatif ($x_k^{(-)}$) adalah salah satu *item* yang dipilih dari *item* konteks positif, sedangkan *item* konteks negatif ($x_j^{(-)}$) adalah *item* yang berada dalam daftar sampel negatif dengan mengeluarkan *item* target negatif. (4) Menentukan matriks bobot pertama, $W = \{w_{ki}\}$. Matriks bobot pertama ditentukan untuk daftar positif dan daftar negatif yang dinotasikan dengan $W_u^{(+)}$ dan $W_u^{(-)}$, u menunjukkan pengguna ke- u . Matriks ini berukuran $V \times N$, di mana V adalah banyaknya *item* dalam daftar positif atau negatif. Sedangkan N adalah representasi *item* ke dalam bentuk vektor, pada Item2Vec digunakan N sebesar 100 untuk data berukuran 2500 sampai 20000 [3]. Angka pada matriks bobot diperoleh dari angka *random* 0 sampai 1 pada *feedforward* pertama [11]. 5) Menentukan matriks bobot kedua, $W' = \{w'_{ij}\}$. Matriks bobot kedua ditentukan untuk daftar positif dan daftar negatif yang dinotasikan dengan $W_u'^{(+)}$ dan $W_u'^{-}$, u menunjukkan pengguna ke- u . Ukuran matriks kedua merupakan kebalikan dari ukuran matriks pertama yaitu $N \times V$. Angka pada matriks bobot kedua juga diperoleh dari angka *random* 0 sampai 1 pada *feedforward* pertama [11]. 6) Lakukan tahapan 7 sampai 9 untuk setiap *item* target positif dan *item* target negatif. 7) Membuat *input layer* untuk *item* target k (x_k). *Input layer* pada Item2Vec berupa *one hot encoded* yang merupakan vektor sepanjang V , di mana V adalah banyaknya $x_{pos}^{u,(+)}$ atau *item* positif untuk wisatawan yang bersangkutan. *One hot encoded* pada *input layer* positif dan negatif terdiri atas *item* target yang diberi nilai 1 dan *item* konteks yang diberi nilai 0.

$$v_k^{(+)} = \begin{pmatrix} x_k^{(+)} \\ \vdots \\ x_j^{(+)} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}; v_k^{(-)} = \begin{pmatrix} x_k^{(-)} \\ \vdots \\ x_j^{(-)} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} \tag{2}$$

Di mana $v_k^{(+)}$ adalah vektor *one hot encoded* untuk *item* target positif, $v_k^{(-)}$ adalah vektor *one hot encoded* untuk *item* target negatif, $x_k^{(+)}$ adalah *item* target positif, $x_k^{(-)}$ adalah *item* target negatif, $x_j^{(+)}$ adalah *item* konteks positif dan $x_j^{(-)}$ adalah *item* konteks negatif. 8) Menghitung *hidden layer*. *Hidden layer* (h_k) merupakan vektor berdimensi N dan merupakan representasi dari *item* target. Proses dari *input layer* melewati matriks bobot pertama yang kemudian diperolehnya vektor *hidden layer*. Berikut ini merupakan perhitungan *hidden layer* untuk daftar positif dan negatif.

$$h_k^{(+)} = (W_u^{(+)})^T \cdot v_k^{(+)} ; h_k^{(-)} = (W_u^{(-)})^T \cdot v_k^{(-)} \quad (3)$$

Di mana $W_u^{(+)}$ adalah matriks bobot pertama untuk daftar positif pada pengguna ke- u dan $W_u^{(-)}$ adalah matriks bobot pertama untuk daftar negatif pada pengguna ke- u . 9) Menghitung *output layer*. Dari proses *hidden layer* menuju ke *output layer* digunakannya matriks bobot kedua $W' = \{w'_{ij}\}$. *Output layer* dihitung menggunakan fungsi aktivasi *negative sampling* di bawah ini [3].

$$p(x_j|x_k) = y_{j,k} = \left(\frac{1}{1 + \exp(-v'_{wj}^{(+)} h_k^{(+)})} \right) \prod_{j=1}^{X_{ns}} \frac{1}{1 + \exp(v'_{wj}^{(-)} h_k^{(-)})} \quad (3)$$

Di mana $y_{j,k}$ adalah *output* untuk *item* target positif ke- k pada *item* konteks positif ke- j , $v'_{wj}^{(+)}$ adalah baris ke- j dari matriks *transpose* $W_u'^{(+)}$, $v'_{wj}^{(-)}$ adalah baris ke- j dari matriks *transpose* $W_u'^{-}$, $h_k^{(+)}$ adalah *hidden layer* untuk *item* target positif, $h_k^{(-)}$ adalah *hidden layer* untuk *item* target negatif dan X_{ns} adalah *item* negatif.

Meminimumkan Fungsi Objektif

Setelah diperolehnya nilai $p(x_j|x_k)$ antara *item* target positif dengan *item* konteks positif pada *output layer*, maka selanjutnya meminimumkan fungsi objektif (E) antara *item* target positif ($x_k^{(+)}$) dan *item* konteks positif ($x_j^{(+)}$) dengan menggunakan persamaan di bawah ini.

$$E = -\frac{1}{J} \sum_{j=1}^J \sum_{-j \leq j \leq j, j \neq 0} \log p(x_j|x_k) \quad (4)$$

Jika $E >$ toleransi kesalahan (α) maka lanjut ke tahapan *backpropagation*, namun jika $E \leq$ toleransi kesalahan (α) maka lanjut ke membuat rekomendasi. Toleransi kesalahan sebesar 0,001 umum digunakan pada proses *neural network* yang mampu dicapai dengan iterasi sebanyak 100 [12].

Backpropagation

Backpropagation adalah proses pembelajaran yang memungkinkan *neural network* menghasilkan *output* yang diinginkan [13]. Pada proses ini dilakukan perhitungan ulang nilai bobot berdasarkan nilai *output* yang dihasilkan pada *feedforward*. 1) *Update* matriks bobot kedua, $W' = \{w'_{ij}\}$. *Update* bobot dimulai dari bobot kedua (W') dengan melakukan *update* pada vektor bobot untuk *item* target positif (v'_{wk}) dan *item* konteks positif j (v'_{wj}), yang mana vektor ini didapat dari baris j matriks W'^T . *Update* dihitung menggunakan persamaan di bawah ini [7].

$$v'_{wj}^{(+)(baru)} = v'_{wj}^{(+)(lama)} - \eta \left(\left(\frac{1}{1 + \exp(-v'_{wj}^{(+)} h_k^{(+)})} \right) - t_j \right) h_k^{(+)} \quad (5)$$

Di mana $\eta > 0$ adalah *learning rate*, sedangkan $t_j = 1$ ketika *item* j merupakan *item* positif, jika tidak maka $t_j = 0$. Dalam *backpropagation*, *learning rate* sebesar 0,5 mampu memberikan hasil mendekati dengan toleransi kesalahan yang ditentukan [14]. 2) *Update* matriks bobot pertama, $W = \{w_{ki}\}$. Setelah *update* matriks bobot kedua kemudian dilanjutkan dengan *update* matriks bobot pertama (W). Matriks bobot pertama (W) hanya melakukan *update* pada vektor bobot untuk *item* target k (v_{wk}) dan yang lainnya tetap, yang mana vektor ini didapat dari baris k dari matriks bobot pertama (W). *Update* matriks bobot kedua dihitung menggunakan persamaan di bawah ini [7].

$$v_{wk}^{(+)(baru)} = v_{wk}^{(+)(lama)} - \eta \left(\sum_{w_j \in \{w_j\} \cup X_{ns}} \left(\left(\frac{1}{1 + \exp \left(- \left(v_{w_j}^{(+)} \right)^T h_k^{(+)} \right)} \right) - t_j \right) v_{w_j}^{\prime(+)} \right)^T \quad (6)$$

Di mana $w_j \in \{w_j\} \cup X_{ns}$ yang berarti bahwa w_j terdiri atas *item* konteks dan *item* negatif. Matriks W' dan W yang baru, digunakan dalam proses *feedforward* selanjutnya sampai menghasilkan E yang sesuai dengan kriteria peneliti. Jika E sudah sesuai, maka matriks bobot kedua (W') merupakan representasi akhir sebuah *item* dan dapat dilakukan ke proses *collaborative filtering* untuk dihitung nilai *cosine similarity* antar *item*. 3) Kembali ke tahapan menghitung *hidden layer*.

Membuat Rekomendasi

Rekomendasi sebanyak K untuk setiap item target positif ($x_k^{(+)}$) (Top-K Recommendation) diperoleh dari pengurutan nilai cosine similarity terbesar sampai terkecil. Nilai cosine similarity antara item target positif ($x_k^{(+)}$) dan item konteks positif ($x_j^{(+)}$) dihitung dengan mengkorelasikan bobot antar item. Maka dari itu matriks W' merupakan representasi untuk item k dan item j [3]. Masing-masing kolom matriks $W' = \{v'_{wj}, \dots, v'_{wk}, \dots, v'_{wj}\}$ adalah wakil untuk setiap item, sehingga dihitungnya $v'_{wj} = \{w_{ji}, \dots, w_{jN}\}$ dan $v'_{wk} = \{w_{ki}, \dots, w_{kN}\}$ menggunakan persamaan di bawah ini.

$$sim(j, k) = \frac{\sum_{i=1}^N w_{ji} w_{ki}}{\sqrt{\sum_{i=1}^N w_{ji}^2} \sqrt{\sum_{i=1}^N w_{ki}^2}} \quad (7)$$

Di mana $sim(j, k)$ adalah nilai cosine similarity antara item target positif ($x_k^{(+)}$) dengan item konteks positif ($x_j^{(+)}$).

Tahapan Evaluasi Model Item2Vec

Pada tahapan ini menggunakan data *testing* sebanyak 20%. Berdasarkan tahapan sebelumnya, model Item2Vec menghasilkan sejumlah rekomendasi pada setiap *item* di setiap *item* memiliki *Top-K Recommendation*. K adalah jumlah rekomendasi yang dibuat untuk setiap *item*. Kinerja *Top-K Recommendation* dievaluasi dengan perhitungan akurasi. Dalam evaluasi model digunakannya daftar positif pengguna, adapun langkahnya yaitu [15]: 1) Siapkan data *testing* yang berisikan daftar positif untuk setiap pengguna. 2) Daftar positif setiap pengguna dinotasikan dengan $X_{pos}^{ru,(+)} = \{x_1^{ru,(+)}, x_2^{ru,(+)}, \dots, x_m^{ru,(+)}\}$ dan m adalah notasi untuk *item* terakhir pada daftar positif. 3) Dengan model Item2Vec dilihatnya *Top-K Recommendation* dari *item* k pada setiap pengguna. 4) Jika *item* j terdapat pada *Top-K Recommendation* dari *item* k maka $hit = 1$, yang berarti model Item2Vec dapat merekomendasikan *item* yang relevan. 5) Jika *item* j tidak terdapat pada *Top-K Recommendation* dari *item* k maka $hit = 0$, yang berarti model Item2Vec tidak dapat merekomendasikan *item* yang relevan. 6) Dari langkah 4 dan 5 diperoleh banyaknya *hit* bernilai 1 dan 0. Kemudian *hit* bernilai 1 dijumlahkan dan dinotasikan dengan $\#hits$, $\#hits$ adalah jumlah *item* rekomendasi yang relevan dengan model Item2Vec. 7) Menghitung akurasi atau *accuracy* dengan persamaan di bawah ini.

$$accuracy = \frac{\#hits}{test\ size} \quad (8)$$

Di mana $\#hits$ adalah banyaknya *hit* bernilai 1 dan *test size* adalah banyaknya *item* dalam data *testing*. Pada **Tabel 1.** terdapat kriteria performa model berdasarkan nilai akurasi [16].

Tabel 1. Kriteria Performa Model

Akurasi (%)	Performa
≥ 90	SangatBaik
$\geq 70 - < 90$	Baik
$\geq 60 - < 70$	Cukup Baik
< 60	Buruk

Sumber: Stephen (2022)

Data

Data yang digunakan merupakan data 10.000 *rating* dari 300 wisatawan terhadap 437 destinasi wisata di Indonesia. Pada data, setiap wisatawan memberikan *rating* sebanyak 21 sampai 53 destinasi wisata. Data ini dihimpun oleh akun *Kaggle* bernama A Prabowo pada Mei 2021 dan diunggah pada Juli 2021. Dalam pengolahannya dibantu oleh *Software Jupyter Notebook* versi 3.6.5 dengan modul *word2vec*. Variabel yang digunakan dalam penelitian ini adalah: 1) *Id* : nomor untuk setiap wisatawan lokal. *Id* ini terdiri dari wisatawan ke-1 sampai dengan wisatawan ke-300. 2) *Kunjungan Destinasi* : destinasi wisata di Indonesia yang dikunjungi oleh para wisatawan ke-1 sampai dengan wisatawan ke-300. 3) *Rating* : penilaian dengan skala 1 (sangat tidak suka) sampai 5 (sangat suka) yang diberikan oleh setiap wisatawan terhadap destinasi wisata yang telah dikunjunginya. 4) *No* : nomor untuk setiap destinasi wisata di Indonesia. 5) *Nama Destinasi* : nama destinasi wisata di Indonesia. 6) *Kategori* : kategori dari destinasi wisata. 7) *Kota* : kota asal dari setiap destinasi wisata.

C. Hasil dan Pembahasan

Hasil Analisis Item2Vec

Berdasarkan model *Item2Vec*, dihasilkan *Top-5 Recommendation* atau 5 rekomendasi untuk 437 destinasi wisata di Indonesia. Pada bagian ini hanya dilampirkan rekomendasi untuk destinasi wisata yang paling banyak diberikan *rating* positif oleh wisatawan. Sehingga, pada **Tabel 2.** terlampir *Top-5 Recommendation* untuk *item* 416, 300, 322, 402, 157 dan 167.

Tabel 2. *Top-5 Recommendation* untuk Destinasi dengan *Rating* Positif Terbanyak

<i>Item</i>	<i>Top-5 Recommendation</i>	<i>Kategori</i>	<i>Kota</i>
No : 416 Nama : Keraton Surabaya Kategori : Budaya Kota : Surabaya	Perkebunan Teh Malabar	Cagar Alam	Bandung
	Bendung Lepen	Taman Hiburan	Yogyakarta
	Grojogan Watu Purbo Bangunrejo	Taman Hiburan	Yogyakarta
	Taman Harmoni Keputih	Cagar Alam	Surabaya
	Taman Barunawati	Taman Hiburan	Surabaya
No : 300 Nama : Sanghyang Heuleut Kategori : Cagar Alam Kota : Bandung	Monumen Palagan Ambarawa	Budaya	Semarang
	Taman Film	Budaya	Bandung
	Water Park Bandung Indah	Taman Hiburan	Bandung
	Taman Hutan Raya Ir. H. Juanda	Cagar Alam	Bandung
	Monumen Sanapati	Budaya	Yogyakarta
No : 322 Nama : Bukit Jamur Kategori : Cagar Alam Kota : Bandung	Wisata Alam Wana Wisata Penggaron	Cagar Alam	Semarang
	Taman Impian Jaya Ancol	Taman Hiburan	Jakarta
	Kampung Tematik Jawi	Budaya	Semarang
	Jembatan Kota Intan	Budaya	Jakarta
	Masjid Agung Trans Studio Bandung	Tempat Ibadah	Bandung
No : 402 Nama : Food Junction Grand Pakuwon Kategori : Pusat Perbelanjaan Kota : Surabaya	Taman Legenda Keong Emas	Taman Hiburan	Jakarta
	Kawasan Punclut	Taman Hiburan	Bandung
	Desa Wisata Gamplong	Taman Hiburan	Yogyakarta
	Kampung Pelangi	Taman Hiburan	Semarang
	Taman Agrowisata Cilangkap	Taman Hiburan	Jakarta
No : 157 Nama : Pantai Baron Kategori : Bahari Kota : Yogyakarta	Monumen Perjuangan Rakyat Jawa Barat	Budaya	Bandung
	Curug Cimahi	Cagar Alam	Bandung
	Curug Bugbrug	Cagar Alam	Bandung
	Puncak Segoro	Cagar Alam	Yogyakarta
	Candi Donotirto	Budaya	Yogyakarta
No : 167 Nama : Geoforest Watu Payung Turunan Kategori : Cagar Alam Kota : Yogyakarta	Monumen Bandung Lautan Api	Budaya	Bandung
	Bandros City Tour	Budaya	Bandung
	Kawasan Wisata Sosrowijayan	Pusat Perbelanjaan	Yogyakarta
	Monumen Sanapati	Budaya	Yogyakarta
	Babakan Siliwangi City Forest Path Bandung	Cagar Alam	Bandung

Sumber: Pengolahan Software Jupyter Notebook

Berdasarkan **Tabel 2.** diperoleh bahwa, ketika wisatawan telah pergi dan menyukai destinasi wisata nomor 416 atau Keraton Surabaya, maka wisatawan tersebut akan diberikan rekomendasi pertama yaitu Perkebunan Teh Malabar, rekomendasi kedua yaitu Bendung Lepen, rekomendasi ketiga yaitu Grojogan Watu

Purbo Bangunrejo, rekomendasi keempat yaitu Taman Harmoni Keputih dan rekomendasi kelima yaitu Taman Barunawati. Berdasarkan hasil tersebut tidak ada pola kecocokan antara kategori dan Kota antara *item* yang direkomendasikan. Oleh karena itu, rekomendasi ini hanya berdasarkan pendapat dari wisatawan terkait perjalanan sebelum dan sesudah.

Hasil Evaluasi Model Item2Vec

Dengan menggunakan *software* Jupyter Notebook diperoleh nilai akurasi sebesar 0,6085 atau 61%. Sebagaimana pada Tabel 2., model Item2Vec yang diterapkan pada data rating wisatawan terhadap destinasi wisata di Indonesia memiliki akurasi cukup baik. Akurasi tersebut juga berada pada rentang yang sama dengan model Item2Vec oleh Barkan & Koenigstein [3] yaitu sekitar 60%. Walaupun memiliki akurasi yang tidak terlalu besar, tetapi mampu memberikan rekomendasi yang cukup baik dan mampu mengungguli sistem rekomendasi *Singular Value Decomposition* (SVD) dengan CF yang memiliki akurasi sebesar 58,4%.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, dapat disimpulkan bahwa diperoleh 5 hasil rekomendasi atau *Top-5 Recommendation* untuk destinasi wisata di Indonesia. Wisatawan yang telah berpergian dari suatu destinasi wisata dapat diberikan rekomendasi sebanyak 5 destinasi wisata untuk dapat didatangi selanjutnya. *Top-5 Recommendation* destinasi wisata di Indonesia dengan menggunakan metode Item2Vec memiliki akurasi yang cukup baik yaitu sebesar 61%.

Daftar Pustaka

- [1] A. Firdaus, "Aplikasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Omicron Covid-19," *Jurnal Riset Statistika*, pp. 85–92, Dec. 2022, doi: 10.29313/jrs.v2i2.1148.
- [2] A. Susanto, "Implementasi Sistem Rekomendasi Menggunakan Pendekatan Collaborative Filtering dan Algoritma Slope One Pada Sistem Informasi Penjualan," *Universitas Teknologi Yogyakarta*, 2020.
- [3] O. Barkan and N. Koenigstein, "Item2Vec: Neural Item Embedding for Collaborative Filtering," *Vietri Sul Mare: IEEE*, pp. 1–6, 2016, doi: 10.1109/MLSP.2016.7738886.
- [4] A. A. Faqir, "Sandiaga Uno: Industri Pariwisata Alami Kontraksi Hebat, Kunjungan Turun 80 Persen," *merdeka*, Sep. 27, 2021.
- [5] B. Nugraha, "Sandiaga Uno: Perlu Gerak Cepat Pulihkan Pariwisata Indonesia," *Viva*, Nov. 26, 2021. <https://www.viva.co.id/berita/nasional/1426670-sandiaga-uno-perlu-gerak-cepat-pulihkan-pariwisata-indonesia>
- [6] N. Yuniar, "Paket menarik ditawarkan demi tingkatkan pariwisata domestik," *Antaranews*, Jan. 05, 2022. <https://www.antaranews.com/berita/2624693/paket-menarik-ditawarkan-demi-tingkatkan-pariwisata-domestik>
- [7] X. Rong, "word2vec Parameter Learning Explained," Nov. 2014.
- [8] M. N. Zain, "Algoritma Artificial Neural Network dalam Klasifikasi Chest X-Rays Pasien COVID-19," *Jurnal Riset Statistika*, pp. 137–144, Dec. 2022, doi: 10.29313/jrs.v2i2.1426.
- [9] P. D. Inas Azizah, "Penerapan Probabilistic Neural Network pada Klasifikasi Berat Bayi Baru Lahir," *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 152–159, Feb. 2022, doi: 10.29313/jrs.v1i2.524.
- [10] V. Kuzmin, "Item2Vec-based Approach to a Recommender System," University of Tartu, 2017.
- [11] A. Mnih and Y. W. Teh, "A fast and simple algorithm for training neural probabilistic language models," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 8, pp. 1751–1758, 2012.

- [12] A. Irianti, P. H. Rantelinggi, A. Taufik, N. Zulkarnaim, and S. Cokrowibowo, "Implementation of Backpropagation Artificial Neural Network For Food Price Prediction in Majene Central Market," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 3, pp. 681–688, 2022, doi: 10.20884/1.jutif.2022.3.3.226.
- [13] J. Lalis, B. Gerardo, and Y. Byun, "An adaptive stopping criterion for backpropagation learning in feedforward neural network," *International Journal of Multimedia and Ubiquitous Engineering*, pp. 149–156, 2014, doi: 10.14257/ijmue.2014.9.8.13.
- [14] N. Nisa, A. Trihasto, and D. Pravitasari, "Optimasi Daya Keluaran Permanent Magnet Shynchronous Generator pada Pembangkit Listrik Tenaga Bayu," *Journal of Electrical Engineering, Computer, and Information Technology*, vol. 02, no. 01, 2021.
- [15] B. A. Yilma, "Recommendation based on sequence (ITEM2VEC)," *Trento: Università degli Studi di Trento*, 2016.
- [16] S. Allwright, "What is a good accuracy score in machine learning?," *stephenallwright.com*, May 14, 2022.