

Sistem Rekomendasi Event Online Menggunakan Metode Content Based Filtering

Fahmi Nurfalalah¹, Asriyanik², Agung Pambudi³

^{1,2,3}Universitas Muhammadiyah Sukabumi

Jl. R. Syamsudin, S.H. No. 50, Cikole, Kec. Cikole, Kota Sukabumi, Jawa Barat 43113, (0266) 218345,

e-mail: fahminurfalalah9@ummi.ac.id¹, asriyanik263@ummi.ac.id², agungpambd@ummi.ac.id³

ARTICLE INFO

Article history:

Received 30 Mei 2022

Received in revised form 2 Juni 2022

Accepted 4 Juni 2022

Available online 2 Desember 2022

ABSTRACT

Covid-19 has spread widely with various variants, disrupting human social activities, the activities in question include the implementation of events, but there are other alternatives by utilizing technology, events can still be held online. However, the number of events that are held online makes it difficult for someone to find the event they want because there are too many events on the website so someone experiences information overload. Therefore, the authors conduct research with the goal to make it easier for someone to find the event they want by implementing a recommendation system on the website. In this study, the authors make a recommendation system using the Content-Based Filtering method in recommending an online event. The results of this study obtained an accuracy of 86% of the performance of the model carried out by SVM.

Keywords: *event online*, sistem rekomendasi, *content-based filtering*, *cosine similarity*, SVM

1. Pendahuluan

Covid-19 membuat semua kegiatan menjadi terganggu, khususnya kegiatan sosial seperti *event*. *Event* menurut Faizal Noer Hakim [1] merupakan sebuah kegiatan yang terencana dan tersusun untuk suatu tujuan tertentu, dengan menghadirkan sejumlah peserta yang mencapai target bahkan lebih. *Event* dibagi menjadi publik dan pribadi, pada *event* publik misalnya seperti webinar, *workshop*, lomba/kejuaraan, hiburan, edukasi, pengetahuan dan lain-lain. Namun demikian, adanya pandemi tidak membuat *event* terganggu, karena ada alternatif lain yang menunjang *event* tersebut dapat terus berjalan. Dengan memanfaatkan internet semua orang yang ikut *event* tersebut tidak perlu repot-repot datang ke suatu tempat dan saling bertatap muka secara langsung, cukup secara virtual menggunakan aplikasi konferensi *video* seperti Google Meet, Zoom, dan lainnya.

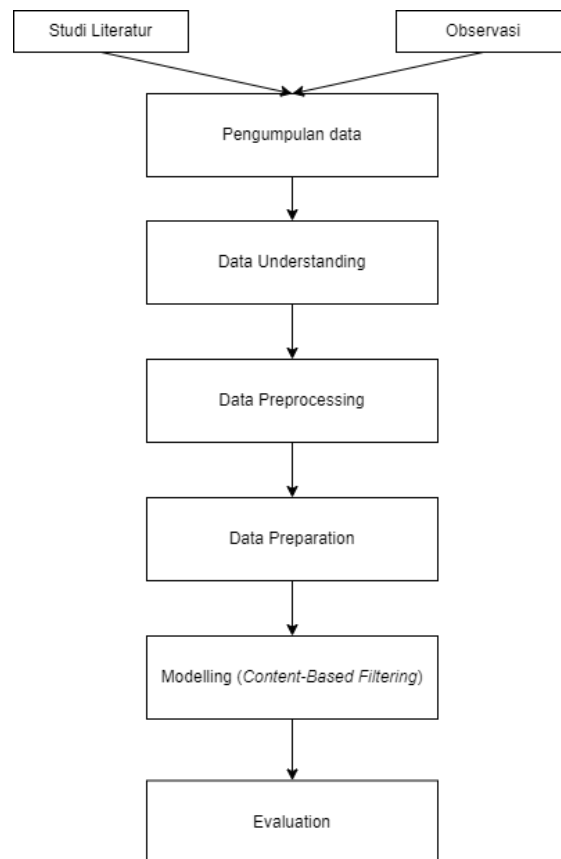
Dengan begitu, banyak *event organizer* yang menyelenggarakan sebuah *event* sehingga banyaknya informasi terkait *event online* ini yang membuat seseorang mungkin akan merasa stres karena mendapatkan informasi yang terlalu banyak (*overload information*) sehingga membuat seseorang tidak mendapatkan informasi yang ingin ia dapatkan secara tepat dan cepat. Menurut Prof. Dr. Anis Eliyana [2] menyebutkan bahwa semakin banyak informasi yang

didapatkan seseorang dari internet, memungkinkan seseorang tersebut akan mengalami kelelahan sosial media bahkan sampai mengalami stres.

Sistem rekomendasi menjadi salah satu solusi yang dapat dimanfaatkan sehingga pengguna tidak lagi mengalami *overload information* karena sistem rekomendasi mampu merekomendasikan suatu *item* yang relevan dan sesuai dengan yang dicari oleh pengguna. Ada beberapa penelitian terkait topik ini, seperti dalam penelitian Muhtar Yusuf [3], menerapkan sistem rekomendasi untuk seminar menggunakan algoritma TF-IDF dan *Cosine Similarity*, dan juga penelitian dari Michael Wijaya Arifin [4] yang menerapkan sistem rekomendasi artikel berita menggunakan metode *content-based filtering*. Hasilnya sistem dapat merekomendasikan artikel berita dengan akurat namun tingkat *precision* yang ada pada sistem berkurang.

Oleh karena itu, adanya masalah tersebut maka diperlukan membangun sebuah sistem rekomendasi dengan tujuan dapat membantu seseorang yang sedang mencari *event online* secara tepat dan cepat sesuai preferensinya dengan akurasi yang bagus menggunakan *content-based filtering* dan algoritma SVM sebagai evaluasi performa dari model yang dibuat.

2. Metode Penelitian



Gambar 1. Tahapan Metode Penelitian

Penelitian ini dilakukan selama 4 bulan, dengan melakukan pengumpulan data *event online*, proses *preprocessing* dan *modeling* menggunakan metode *content based filtering* dan algoritma SVM sebagai evaluasi model.

2.1. Pengumpulan Data

Pengumpulan data dilakukan dengan cara *scraping* data. *Scraping* merupakan teknik pengikisan data dari *website* dan disajikan ke dalam bentuk yang terstruktur dengan menggunakan alat *scraper* pada ekstensi *browser* [5]. Data yang didapatkan sebanyak 1180 data *event online* dari kumpulan *website* penyedia *event*.

2.2. Data Understanding

Dilakukan analisis data untuk pengecekan nama atribut yang ada pada *dataset* dan dilakukan visualisasi data yang telah dikumpulkan sebelumnya. Dalam hal ini dilakukan analisis data *event online*.

2.3. Data Preprocessing dan Preparation

Dilakukan pembersihan data, yaitu dengan melakukan pengecekan nilai kosong, proses *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*. Proses *case folding* merupakan proses merubah huruf kapital menjadi huruf kecil, *tokenizing* merupakan proses memecah satu dokumen atau kalimat menjadi kumpulan perkata dengan spasi sebagai pemisahannya, serta menghilangkan tanda baca dan spesial karakter, *stopwords removal* merupakan tahap menghilangkan kata tidak penting seperti 'dari', 'saya', 'dan', dan sebagainya, dan *stemming* merupakan proses menghilangkan kata imbuhan -an, me-, ke-, ber-, dan lain-lain [3].

2.4. Modeling

Membuat model dengan metode *content based filtering*, *content based filtering* merupakan metode sistem rekomendasi berbentuk dokumen dengan merekomendasikan kemiripan antar dokumen [6]. data hasil *preprocessing* dilakukan pembobotan dengan TF-IDF yang mengeluarkan hasil berupa vektor matriks dengan mencari kemunculan kata pada suatu dokumen. Rumus yang digunakan TF-IDF adalah sebagai berikut [3][7].

$$IDF = \log \frac{N}{df} + 1 \quad (1)$$

Keterangan:

N = jumlah dokumen yang ada,

df = jumlah kemunculan kata pada semua dokumen

$$W = tf \times idf \quad (2)$$

Keterangan:

W = bobot (*weight*) TF-IDF

tf = kemunculan kata dalam dokumen

idf = hasil invers pada rumus (1) di atas

Hasil dari TF-IDF akan dicari kesamaan antar dokumen menggunakan *cosine similarity*. Jika hasil dari *cosine similarity* tinggi antara dokumen A dengan dokumen B yang ada maka tingkat kemiripan semakin tinggi. Dengan rumus yang digunakan untuk *cosine similarity* sebagai berikut [3][8].

$$\cos(A,B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n (A_i)^2 \cdot \sum_{i=1}^n (B_i)^2}} \quad (3)$$

Keterangan:

A_i = jumlah bobot *term* ke-i dalam vektor untuk dokumen ke-A

B_i = jumlah *term* ke-i dalam vektor untuk dokumen ke-B

2.5. Evaluasi

Evaluasi model dilakukan untuk menguji performa model yang sudah dibuat. Pengujian performa dilakukan menggunakan algoritma SVM. SVM sederhananya memiliki konsep mencari *hyperplane* terbaik untuk memisahkan dua buah kelas atau lebih [9], dimana hasil dari prediksi SVM ini akan dihitung akurasi, presisi, *recall*, dan *f1-score* menggunakan *confusion matrix*. Berikut rumus yang digunakan *confusion matrix* [9][10].

$$Accuracy = \frac{\sum TP}{\sum Support} \tag{5}$$

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F1-score = \frac{2 \cdot \frac{precision \cdot recall}{precision + recall}}{\tag{8}}$$

Keterangan:

TP = *True Positive* (data aktual positif, dan data prediksi juga positif)

TN = *True Negative* (data aktual negatif, dan data prediksi juga negatif)

FP = *False Positive* (data aktual negatif, tapi data prediksi positif)

FN = *False Negative* (data aktual positif, tapi data prediksi negatif)

3. Hasil dan Analisa

3.1. Pengumpulan Data

Data hasil *scraping* dari *website* penyedia *event online*, diperoleh data sebanyak 1180 data dengan sampel dataset sebagai berikut.

ID Event	Nama Event	Jenis Event	Tahun Event	Tanggal Pelaksanaan Event	Waktu Dimulai	Tempat	Deskripsi Event	Tags	Upload Gambar/Poster Event
event001	Moderasi Islam Berken Webinar		2021	15/12/2021	08:30:00	Youtube	Moderasi Islam E#Webinar #Keislaman		https://drive.google.com/uc?id=1V...
event002	Program Bangkit 2022 Beasiswa		2021	31/12/2021		Full Onli	Bangkit 2022 aka #beasiswa #bangki		https://drive.google.com/uc?id=1t...
event003	Kewirausahaan Membu Webinar		2020	13/08/2020	13:30:00	Zoom C	*BEM UNIVERSE #webinar #kewirauc		https://drive.google.com/uc?id=1g...
event004	Kelas Gratis Wirausaha Workshop		2020	06/08/2020	14:00:00	Live Yo	*Kelas Gratis Wira#kelasOnlineGratis		https://drive.google.com/uc?id=1Q...
event005	ITech 5.0 How To Build Webinar		2021	05/12/2021	08:30:00	Zoom M	ITECH 5.0 by TI #informatika #teknai		https://drive.google.com/uc?id=1xf...
event006	SOCIAL MEDIA ANAL Webinar		2021	28/11/2021	13:00:00	Zoom M	WEBINAR SOCI #webinar #informat		https://drive.google.com/uc?id=1e...
event007	IT Series Vol.3 Design Webinar		2021	23/10/2021	13:00:00	Zoom M	IT Webinar Seri #webinar #informat		https://drive.google.com/uc?id=1Y...
event008	Series Memulai Karir S Webinar		2021	29/08/2021	13:00:00	Zoom M	Seri Webinar Prc #webinar #informat		https://drive.google.com/uc?id=1aj...
event009	How To Build Career in Webinar		2021	03/07/2021	19:00:00	Zoom M	Ayo ikuti shanng #informatika #infor		https://drive.google.com/uc?id=1TI...
event010	React Native Navigatio Webinar		2021	26/02/2021	13:00:00	Zoom M	Webinar Kolabor #webinar #bsaa #i		https://drive.google.com/uc?id=1a...

Gambar 2. Sampel *Dataset* Penelitian Hasil Pengumpulan Data

3.2. Data Understanding

Proses memahami data dan menganalisa atribut apa saja yang ada pada *dataset* beserta fungsinya. Berikut atribut yang akan digunakan untuk proses *modeling*.

Tabel 1 Atribut Pada *Dataset* Penelitian

No	Atribut/Variabel	Keterangan
1	ID <i>Event</i>	Sebagai identitas unik data <i>event online</i>
2	Nama <i>Event</i>	Sebagai nama dari <i>event</i> yang diselenggarakan
3	Jenis <i>Event</i>	Sebagai jenis <i>event</i> yang akan diselenggarakan
4	Tahun <i>Event</i>	Menunjukkan tahun <i>event</i> yang akan dilaksanakan
5	Tanggal Pelaksanaan <i>Event</i>	Tanggal pelaksanaan/tanggal pendaftaran pada <i>event</i> yang dilaksanakan
6	Waktu Dimulai <i>Event</i>	Pukul dilaksanakannya <i>event</i>
7	Tempat pelaksanaan	Tempat pelaksanaan <i>event</i> (dalam penelitian ini semua

	<i>Event</i>	<i>event</i> dilaksanakan online)
8	Deskripsi <i>Event</i>	Penjelasan detail tentang <i>event</i> yang akan dilaksanakan
9	Tags	Tags <i>event</i> (diawali dengan simbol '#')
10	Upload Gambar/Poster <i>Event</i>	Sebagai poster <i>event</i> (berisi link url poster)

3.3. Data Preprocessing dan Data Preparation

Pada tahapan preprocessing, dilakukan proses pembersihan data dilanjutkan dengan menggabungkan atribut 'Nama *Event*', 'Tags' menjadi satu atribut khusus bernama '*bag of words*'. Hasilnya seperti gambar berikut.

Tabel 2 Hasil *Preprocessing* Data

No	Bag of words
1	moderasi islam maju bingkai indonesia webinar keislaman universitasmuhammadiyahsukabumi islam dakwah
2	program bangkit registration beasiswa bangkit viral universitasmuhammadiyahsukabumi
3	kewirausahaan bangun ekonomi kreatif adaptasi biasa webinar kewirausahaan bemummi ekonomikreatif
4	kelas gratis wirausaha google womenwill indonesia bem ummi kelasonlinegratis kelasonline bemummi workshop womenwill wirausaha kewirausahaan
5	itech build start business sucessfully social media analyst informatika teknikinformatika informatikaummi hmifummi webinar technology teknologi engineering
6	social media analysis machine learning webinar informatika informatikaummi machinelearning technology
7	series vol design thinking webinar informatika informatikaummi technology softskill designthinking
8	series karir lulus webinar informatika informatikaummi technology softskill
9	build career startup company informatika informatikaummi sukabumi kuliah ummi universitasmuhammadiyahsukabumi sbmptn webinar technology
10	react native navigation local storage webinar bisaai tampil ummi webinarseries series idg event mobileapps reactnative react native android appandroid navigation localstorage technology informatika

3.4. Modeling

Pada tahap ini dilakukan proses *modeling* dengan melakukan pembobotan menggunakan TF-IDF dan *cosine similarity*. Mengambil sampel dokumen 1 dan dokumen 2. Berikut perhitungannya.

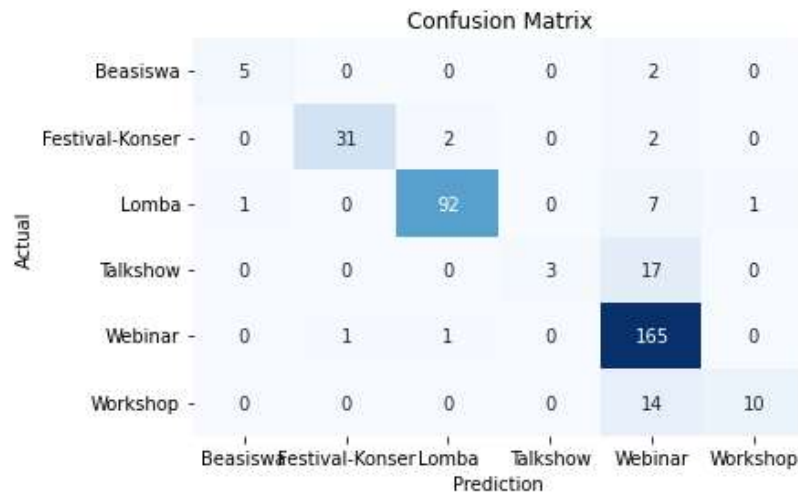
Tabel 3. Perhitungan TF-IDF dan *Cosine Similarity* (Sampel D1 dan D2)

Terms	tf		df	idf	W (tf.idf)		Aij.Biq
	Query (D1)	D2			D1	D2	
moderasi	1	0	1	1,693	1,693	0	0
islam	2	0	1	1,693	3,386	0	0
maju	1	0	1	1,693	1,693	0	0

bingkai	1	0	1	1,693	1,693	0	0
indonesia	1	0	2	1	1	0	0
program	0	1	1	1,693	0	1,693	0
bangkit	0	2	1	1,693	0	3,386	0
registration	0	1	1	1,693	0	1,693	0
webinar	1	0	2	1	1	0	0
keislaman	1	0	1	1,693	1,693	0	0
universitasmuham madiyahsukabumi	1	1	2	1	1	1	1
dakwah	1	0	1	1,693	1,693	0	0
beasiswa	0	1	1	1,693	0	1,693	0
viral	0	1	1	1,693	0	1,693	0
					28,796	23,93	1
					5,366	4,892	26,25
						<i>Cosine Sim</i>	0,038

3.5. Evaluasi

Proses evaluasi memisahkan data untuk *train* sebesar 70% dan data *test* sebesar 30% dengan atribut ‘Jenis *Event*’ sebagai labelnya. Hasil evaluasi dari SVM menghasilkan sebuah *confusion matrix* sebagai berikut.



Gambar 3. *Confusion Matrix* dari Hasil Klasifikasi SVM

Dengan *confusion matrix* dapat dicari *accuracy*, *precision*, *recall*, dan *f1-score* nya. Berikut hasil perhitungannya.

$$Accuracy = \frac{5+31+92+3+165+10}{354} \tag{5}$$

$$= 0.86$$

$$Precision \text{ label Beasiswa} = \frac{5}{(5+0+1+0+0+0)} \tag{6}$$

$$= \frac{5}{6} = 0.83$$

$$Recall \text{ label Beasiswa} = \frac{5}{(5+0+0+0+2+0)} \tag{7}$$

$$\begin{aligned}
 F1\text{-score label Beasiswa} &= \frac{5}{7} = 0.71 \\
 &= 2 \frac{(0.83+0.71)}{(0.83+0.71)} \\
 &= 2*0.382 = 0.77
 \end{aligned}$$

Hasil lengkap perhitungan *accuracy*, *precision*, *recall*, *f1-score* ada pada gambar berikut.

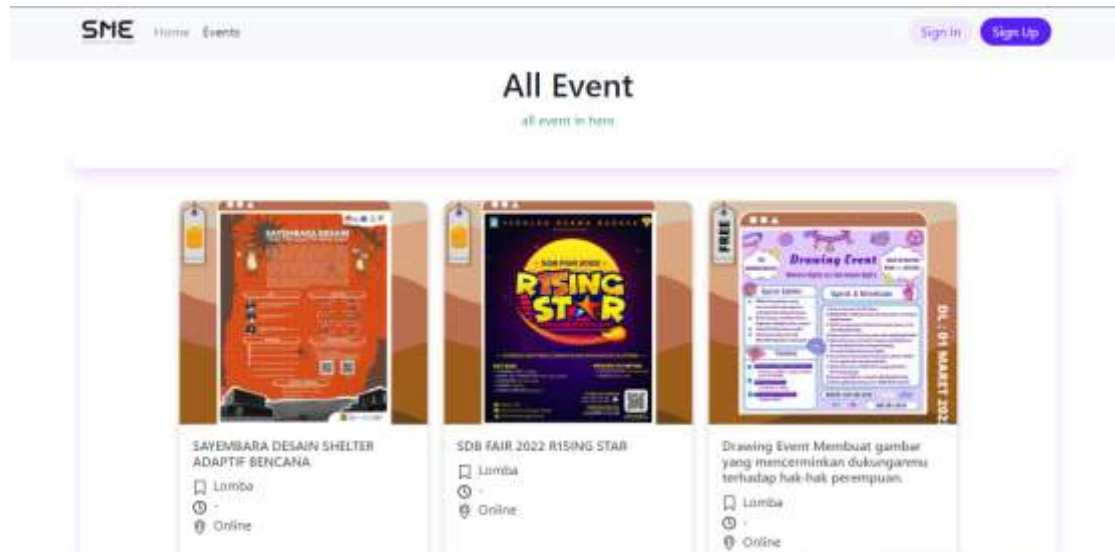
	precision	recall	f1-score	support
Beasiswa	0.83	0.71	0.77	7
Festival-Konser	0.97	0.89	0.93	35
Lomba	0.97	0.91	0.94	101
Talkshow	1.00	0.15	0.26	20
Webinar	0.80	0.99	0.88	167
Workshop	0.91	0.42	0.57	24
accuracy			0.86	354
macro avg	0.91	0.68	0.72	354
weighted avg	0.88	0.86	0.84	354

Gambar 4. Hasil Evaluasi Performa Model

3.6. Implementasi Program

1. Tampilan Halaman *Events*

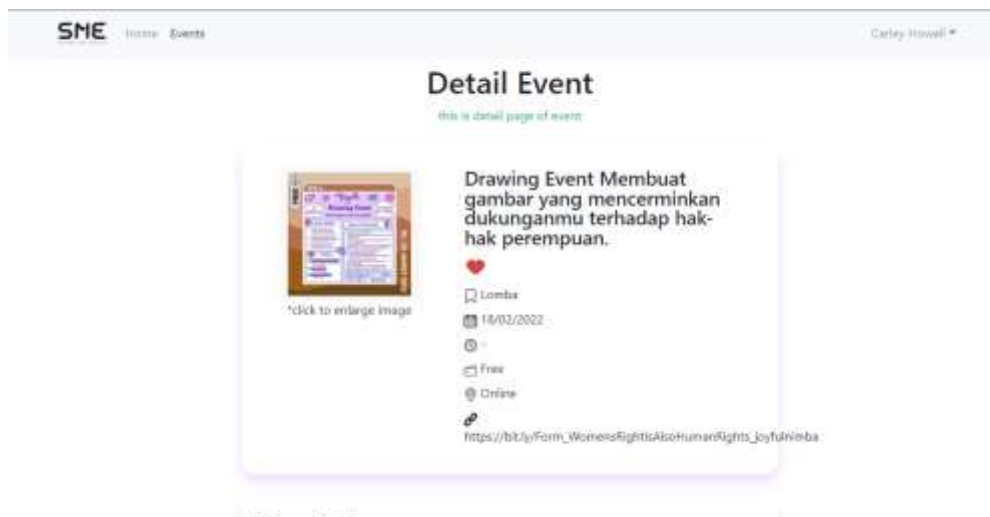
Pada halaman ini menampilkan semua *event online* sebanyak 1180 data yang ada pada *dataset*.



Gambar 5. Halaman Semua *Event Online*

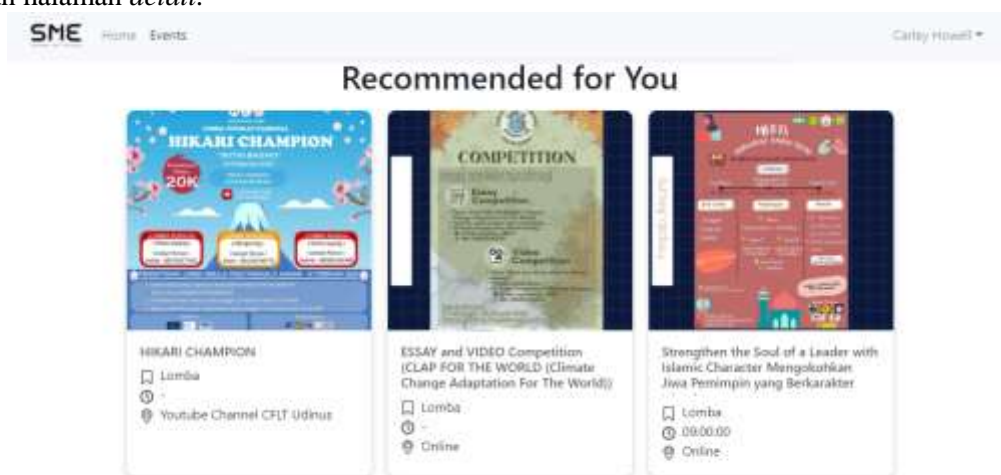
2. Tampilan Halaman *Detail Event* dan Rekomendasi *Event*

Halaman *detail event* menampilkan penjelasan *event* lebih lengkap mulai dari nama sampai deskripsi *event*.



Gambar 6. Halaman Detail *Event Online*

Jika menekan tombol ikon berbentuk hati, nantinya akan muncul *event* yang sama terkait dengan *event* yang ditambah ke favorit tadi dan hasil rekomendasi muncul di bawah tampilan halaman *detail*.



Gambar 7. Tampilan Rekomendasi *Event Online* dari Halaman *Detail Event*

4. Kesimpulan

Berdasarkan hasil penelitian terdapat beberapa kesimpulan yang diperoleh, diantaranya.

1. Penerapan sistem rekomendasi *event online* dapat membantu seseorang dalam mencari *event* yang relevan dan sesuai dengan preferensinya.
2. Sistem rekomendasi dapat menjadi sarana promosi bagi penyelenggara *event* karena sistem akan merekomendasikan *event* walaupun *event* tersebut baru diunggah.
3. Sistem rekomendasi dapat merekomendasikan *event online* secara baik, dengan bukti hasil evaluasi performa model didapatkan hasil *accuracy* sebesar 86%.

Referensi

- [1] Hakim FN. PENGALAMAN PENGUNJUNG PADA EVENT SEHARI BERSAMA COKLAT KITA 2019. Bandung, SEKOLAH TINGGI PARIWISATA NHI, 2019.

- [2] Eliyana A. *Kelebihan Informasi dan Komunikasi terhadap Social Media Exhaustion dan Prestasi Kerja*. Unair News, 2021.
- [3] Yusuf M, Cherid A. Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar. *Jurnal Ilmu Fakultas Ilmu Komputer*. 2020;9:8–16.
- [4] Arifin MW. Rancang Bangun Sistem Rekomendasi Artikel Berita Menggunakan Teknik Content-Based Filtering dan TF-IDF. Tangerang. Universitas Multimedia Nusantara, 2020.
- [5] Alkaberi WS. Web Scraper Application for Extracting Scientific Journals Data. *ICFNDS*. 2021;15:220–4.
- [6] Adi PS. Sistem Rekomendasi Nilai Mata Kuliah menggunakan Metode Content-Based Filtering. *Seminar Nasional Informatika 2014*;2010:A.90-A.94.
- [7] Herwijayanti B, Ratnawati DE, Muflikhah L. Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Pengembangan Teknologi Informasi dan Ilmu Komputer* 2018;2:306–12.
- [8] Rismayani, Ummareng C, F N. Sistem Rekomendasi Pencarian Jodoh Syariah Menggunakan Algoritma Cosine Similarity Berbasis Android. *Journal Teknologi Informasi dan Komunikasi* 2018;7:88–97.
- [9] Fibrianda MF. ANALISIS PERBANDINGAN AKURASI DETEKSI SERANGAN PADA JARINGAN KOMPUTER DENGAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM). Malang. Universitas Brawijaya, 2018.
- [10] Anggraeny MS. *Confusion Matrix*. Binus SOCS. 2020.