

ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI REVIEW FILM PADA IMDB**Gempita Buana Putri^{1*}, Abdul Khoir Diaz² Tata Sutabri³**^{1,2,3}Magister Teknik Informatika, Universitas Bina Darma, Palembang, Sumatera Selatan
email: gempitabuanaputri@gmail.com^{1*}, akd7646@gmail.com², tata.sutabri@binadarma.ac.id³**ABSTRAK**

Perkembangan teknologi informasi saat ini sudah berkembang dengan pesat. Internet Movie Database (IMDb) adalah situs web yang menyediakan informasi mengenai film dari seluruh dunia, termasuk orang-orang yang terlibat di dalamnya mulai dari aktor/aktris, sutradara, penulis, penata rias, soundtrack, nominasi yang pernah didapat, dan review dari para pengunjung. Ini mempengaruhi bagaimana cara pengguna mendapatkan informasi, dimana informasi dan hiburan dengan mudahnya kita akses lewat media digital. Permasalahan yang terjadi ialah bagaimana melakukan klasifikasi terhadap data sentimen atau *review* film yang berjumlah 1000 *record data*. Tujuan penelitian ini ialah melakukan proses klasifikasi terhadap data *review film* yang telah dikumpulkan oleh peneliti dengan menggunakan algoritma *support vector machine*. Berdasarkan 5 Kali uji coba terhadap data testing dan data training, didapat bahwa pengujian dengan sampel data 80:20 memiliki akurasi tertinggi yaitu sebesar 0.687 atau 68.7% dengan menggunakan metode *support vector machine* pada penelitian Algoritma *Machine Learning* Untuk Klasifikasi Review Film Pada IMDB.

Kata Kunci : *Algoritma, Machine Learning, Klasifikasi, SVM, Review*

Abstract: *The development of information technology is currently growing rapidly. The Internet Movie Database (IMDb) is a website that provides information about films from around the world, including the people involved in them, from actors/actresses, directors, writers, make-up artists, soundtracks, teenagers who have been found, and reviews from critics. visitors. This affects how users get information, where we can easily access information and entertainment through digital media. The problem that occurs is how to classify sentiment data or film reviews totaling 1000 data records. The purpose of this research is to carry out a classification process for film review data that has been collected by researchers using a support vector machine algorithm. Based on 5 trials of data testing and training data, it was found that testing with sample data 80.20 had the highest accuracy of 0.687 or 68.7% using the support vector machine method in Machine Learning Algorithm research for Classification of Film Reviews on IMDB.*

Keywords : *Algorithm, Machine Learning, Classification, SVM, Review***PENDAHULUAN**

Perkembangan teknologi informasi saat ini sudah berkembang dengan pesat[1]. Perkembangan teknologi internet saat ini membawa dampak perubahan pada beberapa sektor kehidupan manusia, seperti hanya pada bidang pendidikan, perdagangan, pemerintahan hingga komunikasi social. Di era sekarang ini pertumbuhan internet dan media sosial telah berkembang pesat sehingga menyediakan berbagai informasi opini dari orang lain[2].

Internet Movie Database (IMDb) adalah situs web yang menyediakan informasi mengenai film dari seluruh dunia, termasuk orang-orang yang terlibat di dalamnya mulai dari aktor/aktris, sutradara, penulis, penata rias, soundtrack, nominasi yang pernah didapat, dan review dari para pengunjung. Ini mempengaruhi bagaimana cara pengguna mendapatkan informasi, dimana informasi dan hiburan dengan mudahnya kita akses lewat media digital. Selain itu, pengguna juga dapat berinteraksi satu sama lain antar sesama pengunjung dan memberikan tanggapan mengenai suatu film. Pengguna IMDb di Indonesia mencapai 1.3% dari seluruh dunia yaitu sekitar 1.3% dari 57 juta visitors. Hal ini membuat Indonesia berada diperingkat 12 menurut statistik dari *hypestat*[3]. IMDb merupakan sebuah portal media informasi sudah lebih dari cukup untuk membantu responden dalam memiliki informasi berkaitan dengan film tanpa harus mengecek portal sejenis. hasil penelitian menunjukkan bahwa rating IMDb berpengaruh signifikan terhadap keputusan untuk menonton sebuah film, disamping itu pengguna juga dapat membaca ulasan-ulasan yang menyertai ranting tersebut[4].

Dengan banyaknya film maupun review pada IMDb menyebabkan user atau visitor harus memeriksa review satu per satu untuk mengetahui film rating tersebut. *Text mining*, mengacu pada proses mengambil informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya diperoleh melalui peramalan pola dan kecenderungan melalui sarana seperti pembelajaran pola statistik. Proses *textmining* yang khas meliputi kategorisasi teks, *text clustering*, ekstraksi konsep/entitas, produksi taksonomi granular, *sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas[5].

Opini adalah pendapat pribadi seseorang tentang suatu topik atau masalah yang sedang dihadapi, dan opini publik saat ini merupakan faktor penting dalam perkembangan perusahaan atau kebijakan yang harus ditempuh oleh pemerintah. Data digital ini dapat berupa opini publik tentang suatu topik yang sedang tren di masyarakat, atau peringkat, komentar, dan opini konsumen tentang produk atau layanan yang mereka beli dari produk atau layanan tersebut[6]. Tujuan sentiment analysis ialah untuk membantu user secara cepat dan efisien dalam menentukan film yang akan ditonton berdasarkan sentiment hasil review. Dalam proses sentiment analysis dapat digunakan metode *text classification*[7]

Tahapan analisis sentimen terdiri dari pengumpulan data, pemrosesan awal, transformasi data, seleksi fitur dan tahapan klasifikasi menggunakan metode naïve bayes. Teknik pengumpulan data menggunakan observasi, interview, dan dokumentasi. Pemrosesan awal menerapkan tahapan *text mining* yakni *casefolding*, *non alpha numeric removal*, *stop words removal*, dan *stemming*[8].

Untuk mendukung proses klasifikasi terhadap data *review* film dibutuhkan sebuah metode *machine learning* yang dapat mengolah dataset. *Support Vector Machine* merupakan sistem pembelajaran yang menggunakan hipotesis berupa fungsi-fungsi linear dalam sebuah fitur yang berdimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran yang didasarkan dengan teori optimasi[9]. *Support Vector Machine (SVM)* merupakan salah satu metode yang sudah banyak diterapkan untuk berbagai jenis penelitian dibidang data dan *text mining* karena telah mampu menunjukkan perfoma yang lebih baik[10].

Maka pada penelitian ini akan di implementasikan metode tersebut dan dicari hasil akurasi dengan model *Support Vector Machine (SVM)* dan menggunakan alat bantu *Python3* serta hasil klasifikasi tersebut dapat membantu dalam menentukan dan mengelompokkan jenis *review* dengan menggunakan salah satu algoritma *machine learning*.

2. TINJAUAN PUSTAKA

2.1 Analisa

Analisis adalah sebuah kegiatan untuk mencari pola ataupun metode berasumsi yang berhubungan dengan pengamatan dengan cara analitis kepada suatu buat memastikan bagian, jalinan antar bagian, dan ikatan dengan totalitas. Analisa merupakan sesuatu upaya buat menguraikan sesuatu permasalahan jadi bagian-bagian(*decomposition*) alhasil lapisan wujud sesuatu yang dijabarkan itu nampak dengan nyata alhasil dapat dimengerti permasalahannya[11]

2.2 Analisa Sentimen

Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral. Saat ini, peneliti secara luas menggunakan analisis sentimen sebagai salah satu cabang penelitian dalam ilmu komputer. Jejaring sosial, seperti Twitter, umumnya digunakan dalam analisis sentimen untuk menentukan persepsi publik[12].

2.3 Big Data

Big Data adalah data yang melebihi proses kapasitas dari kovensi sistem database yang ada. Data terlalu besar dan terlalu cepat atau tidak sesuai dengan struktur arsitektur database yang ada. Untuk mendapatkan nilai dari data, maka harus memilih jalan alternatif untuk memprosesnya[13].

2.4 Informasi

Informasi adalah merupakan hasil dari pengolahan data menjadi bentuk yang lebih berguna bagi yang menerimanya yang menggambarkan suatu kejadian – kejadian nyata yang dapat digunakan sebagai alat bantu untuk mengambil suatu keputusan[14]. Syarat-syarat informasi berupa Kerahasiaan, Otentifikasi, Integritas, Tidak dapat dibantahkan serta mudah diakses oleh siapa saja bagi yang butuh terhadap data tersebut. informasi adalah data yang telah diklasifikasikan atau diolah atau diinterpretasikan untuk digunakan dalam proses pengambilan keputusan[15][16]

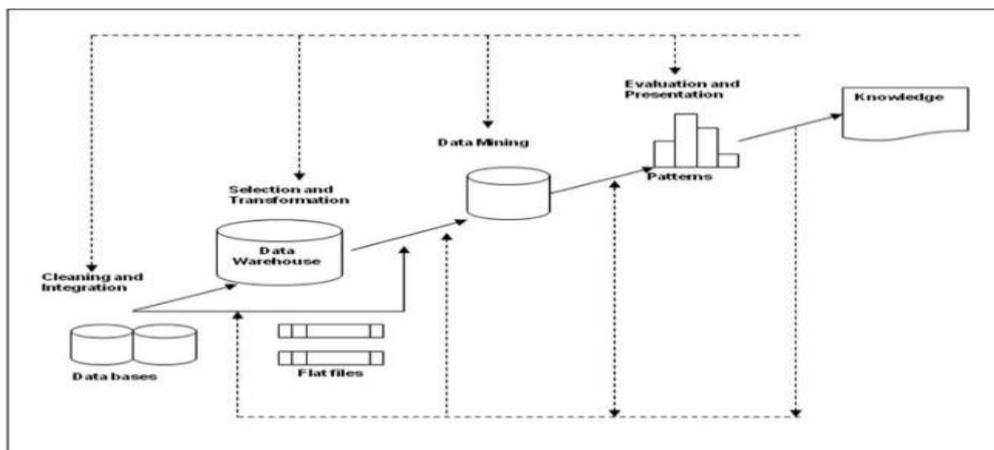
2.5 Data

Data merupakan objek mentah, yang belum diolah dan akan diolah. Sedangkan, informasi adalah data yang telah diolah dan sifatnya menjadi data lain yang bermanfaat. Pembelajaran digital adalah produk revolusi industry 4.0[17].

2.6 Data Mining

Data mining, sering juga disebut *Knowledge Discovery In Database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan

keputusan di masa depan. Saat ini istilah pengenalan pola (pattern recognition) jarang digunakan karena ia termasuk bagian dari data mining[18]. *Knowledge Discovery in Database* adalah proses untuk menggali dan menganalisis sejumlah data dan mengestrak informasi dan pengetahuan yang berguna. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan untuk keperluan pengambilan keputusan”. Proses dalam KDD adalah proses yang digambarkan pada dan terdiri dari rangkaian proses sebagai berikut[19]:



Gambar 1. Tahapan KDD

Data mining merupakan salah satu langkah dari proses Knowledge Discovery from Data atau lebih dikenal dengan singkatan KDD”. Berikut langkah langkah dari KDD (Jiawei and Micheline, 2006): Tahap-tahap data mining ada 6 yaitu:

1. Pembersihan data (data cleaning)
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi data (data integration)
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. .
3. Seleksi Data (Data Selection)
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.
4. Transformasi data (Data Transformation)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses mining,
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi pola (pattern evaluation),
Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.
7. Presentasi pengetahuan (knowledge presentation),
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining

3. METODE PENELITIAN

3.1 Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah observasi dan dokumentasi. Berikut penjelasannya:

1. Wawancara kerja (interviews). Komunikasi dua arah untuk mengambil data dari responden. Wawancara (interviews) meliputi wawancara pribadi (*individual interview*), wawancara intersepsi (*interception interview*), dan wawancara telepon (*telephone interview*). Proses wawancara dilakukan dengan beberapa stakeholder terkait salah satunya pihak kecamatan yang menangani bidang pencatatan kependudukan, peneliti melakukan wawancara bagaimana prosedur yang dilakukan saat ini dalam melakukan evaluasi ataupun proses pendataan data yang diperoleh seperti pihak-pihak terkait dalam penelitian ini.
2. Observasi (pengamatan). Suatu teknik atau pendekatan yang memperoleh data primer dengan cara mengamati secara langsung objek data. Pendekatan observasional dapat dibagi menjadi observasi behavioral (observasi perilaku) dan observasi non-perilaku (nonbehavioral observasi). Hasil proses observasi yaitu data-data yang digunakan pada penelitian ini salah satu sampel data yaitu data review film pada IMDb.
3. Studi Sastra (studi sastra). Melakukan penelitian dengan menelaah buku, jurnal dan literatur lain yang berhubungan dengan judul penelitian. Salah satunya dengan menggunakan jurnal mengenai penelitian terkait seperti penelitian visualisasi atau analisa dengan menggunakan visualisasi.

3.2 Support Vector Machine

SVM merupakan singkatan dari Support Vector Machine, SVM bekerja dengan membagi dua kelompok kelas data menggunakan fungsi linear dalam sebuah ruang fitur berdimensi tinggi dengan proses menemukan garis pemisah (hyperplane) terbaik sehingga dapat menemukan ukuran margin yang maksimal antara ruang input dengan ruang ciri menggunakan kaidah kernel[10].

SVM sering digunakan dalam berbagai masalah termasuk pengenalan pola, bioinformatika dan kategori teks dengan menguraikan hyperplane sebagai set input kedalam ruang fitur yang terdiri dari dua kelas tetapi kemudian dioptimasi kembali sehingga dapat digunakan kedalam bentuk lebih dari dua kelas[20].

4. HASIL DAN PEMBAHASAN

Spesifikasi kebutuhan dilakukan dengan melakukan sorting dan pengelolaan data yang akan digunakan pada penelitian ini dan melakukan observasi dan analisa kasus review film pada IMDb dengan menggunakan visualisasi Data itu sendiri. Pada tahap ini dikumpulkan kebutuhan dari user serta tujuan dari sistem.

Proses pertama setelah menentukan atribut adalah pemanggilan data dengan menggunakan bahasa pemrograman *python3* pada alat bantu *notebook google collabatory*. Langkah pertama adalah melakukan pemanggilan sampel data yang sudah dimasukkan kedalam file *ekstensi csv*, seperti pada Gambar 2 berikut

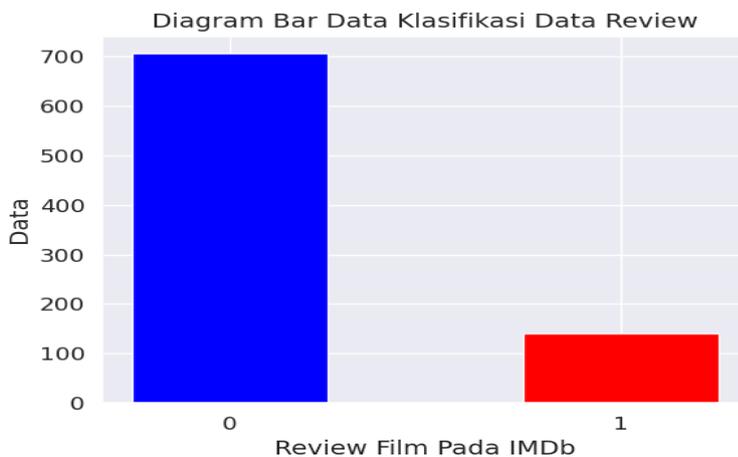
```
[1] # Packages for visuals
import matplotlib.pyplot as plt
import seaborn as sns; sns.set(font_scale=1.2)

# Load Data
review = pd.read_csv(r"text2.csv")
review.head(10)
```

	text	label
0	I always wrote this series off as being a comp...	0
1	1st watched 12/7/2002 - 3 out of 10(Dir-Steve ...	0
2	This movie was so poorly written and directed ...	0
3	The most interesting thing about Miryang (Secr...	1
4	when i first read about "berlin am meer" i did...	0
5	I saw this film on September 1st, 2005 in Indi...	1
6	I saw a screening of this movie last night. I ...	0
7	William Hurt may not be an American matinee id...	1
8	IT IS A PIFCF OF CRAP! not funny at all durin	0

Gambar 2. Proses Penarikan Data

Pada langkah berikutnya melakukan proses pemanggilan operator untuk membuat visual grafik dan pada langkah ini peneliti langsung menggunakan grafik bar untuk menyusun data berdasarkan atribut seperti Gambar 3 berikut.



Gambar 3. Grafik Review film pada IMDb Berdasarkan Tahun

Pada grafik diatas menampilkan informasi grafik berdasarkan jenis sentimen review. Berikut ini merupakan proses dari pembagian dataset data dari review film yang berjumlah 1000 record data pada dataset *review film pada IMDb* yang dapat dilihat pada Gambar 4.

```
[9] # Split into train and test data

train_X, test_X, train_Y, test_Y = model_selection.train_test_split(review['text'], review['label'], test_size = 0.2, random_state = 3)
# random_state = 0 menyatakan tidak ada pengacakan pada data yang di split yang artinya urutannya masih sama

df_train90 = pd.DataFrame()
df_train90['text'] = train_X
df_train90['label'] = train_Y

df_test10 = pd.DataFrame()
df_test10['text'] = test_X
df_test10['label'] = test_Y
```

Gambar 4 Dataset Review film pada IMDb

Pada proses selanjutnya ialah proses pengolahan dataset menjadi data testing dan data training sebelum diolah dengan algoritma *svm* dari dataset *review film pada IMDb* yang dapat dilihat pada Gambar 5.

```
[12] # TF-IDF
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vect_9010 = TfidfVectorizer(max_features = 10)
tfidf_vect_9010.fit(review['text'])
train_X_tfidf_9010 = tfidf_vect_9010.transform(df_train90['text'])
test_X_tfidf_9010 = tfidf_vect_9010.transform(df_test10['text'])

# Proses Pelatihan
from sklearn.svm import SVC

model = SVC(kernel='linear')
model.fit(train_X_tfidf_9010,train_Y)
```

SVC
SVC(kernel='linear')

Gambar 5 Proses Pelatihan SVM

Proses berikutnya adalah pengujian terhadap dataset untuk menghasilkan akurasi dari 1000 *record data* review film yang sudah dibagi menjadi data *training* dan data *testing* yang dapat dilihat pada Gambar 6 berikut.

```
[14] # Proses Pengujian

from sklearn.metrics import accuracy_score

predictions_SVM_9010 = model.predict(test_X_tfidf_9010)
test_prediction_9010 = pd.DataFrame()
test_prediction_9010['text'] = test_X
test_prediction_9010['label'] = predictions_SVM_9010
SVM_accuracy_9010 = accuracy_score(predictions_SVM_9010, test_Y)*100
SVM_accuracy_9010 = round(SVM_accuracy_9010,1)
```

Gambar 6 Proses Pengujian Dataset

Berikut ini merupakan hasil dari pengujian yang dilakukan peneliti sebanyak 5 kali uji dengan membagi data *training* dan data *testing*.

```
SVM_accuracy_9010
68.7

from sklearn.metrics import classification_report

print ("\nHere is the classification report:")
print (classification_report(test_Y, predictions_SVM_9010))
```

Here is the classification report:

	precision	recall	f1-score	support
0	0.64	0.63	0.64	279
1	0.54	0.55	0.55	221
accuracy			0.60	500
macro avg	0.59	0.59	0.59	500
weighted avg	0.60	0.60	0.60	500

Gambar 7 Hasil Pengujian Dataset

Berdasarkan hasil penelitian maka peneliti dengan menggunakan *review film pada IMDb* dan dengan menggunakan metode *support vector machine* didapati bahwa dengan menggunakan 5 kali pengujian dari pembagian data training dan data testing secara *random* dan 1000 *record data*, menghasilkan akurasi sebesar 68.7% pada penelitian Algoritma *Machine Learning* Untuk Klasifikasi *Review Film* Pada IMDb.

KESIMPULAN DAN SARAN

Berdasarkan 5 Kali uji coba terhadap data testing dan data training, didapat bahwa pengujian dengan sampel data 80:20 memiliki akurasi tertinggi yaitu sebesar 0.687 atau 68.7% dengan menggunakan metode *support vector machine* pada penelitian Algoritma *Machine Learning* Untuk Klasifikasi *Review Film* Pada IMDb. Hasil tersebut sudah dapat dikatakan baik karena memiliki akurasi sebesar 68.7%. keterbatasan penelitian ini masih menggunakan satu metode serta metode *support vector machine* masih bisa dioptimalkan lagi untuk mendapatkan akurasi yang lebih baik.

DAFTAR PUSTAKA

- [1] T. Sutabri, T. Sugiharto, R. A. Krisdiawan, and M. A. Azis, "Pengembangan Sistem Informasi Monitoring Progres Proyek Properti Berbasis Website Pada PT Peruri Properti," *J. Teknol. Inform. dan Komput.*, vol. 8, no. 2, pp. 17–29, 2022.
- [2] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [3] R. I. Pristiyanti, M. A. Fauzi, and L. Muflikhah, "Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 3, pp. 1179–1186, 2018, [Online]. Available: <http://j-ptiik.uib.ac.id>
- [4] A. S. Pangestu, H. M. Jumhur, and M. Hum, "Pengaruh Rating Imdb Terhadap Keputusan Menonton Film Game of Thrones Di Kota Bandung Tahun 2018 Impact of Imdb Rating on the Decision To Watch Game of Thrones Film in Bandung City in 2018," vol. 5, no. 3, pp. 3253–3258, 2018.
- [5] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [6] F. Septianingrum and A. S. Y. Irawan, "Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 799, 2021, doi: 10.30865/mib.v5i3.2983.
- [7] G. Cahyani, W. Widayani, S. D. Anggita, Y. Pristyanto, I. Ikmah, and A. Sidauruk, "Klasifikasi Data Review IMDb Berdasarkan Analisis Sentimen Menggunakan Algoritma Support Vector Machine," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1418, 2022, doi: 10.30865/mib.v6i3.4023.
- [8] E. Kartika and J. Gondohanindijo, "Rancang Bangun Model Sentimen Analisis Review Produk Pada Toko Online Menggunakan Naive Bayes," *Semin. Nas. Has. Penelit. dan Pengabd. Kpd. Masy.*, pp. 201–212, 2020.
- [9] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM," *Jdmsi*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: <https://t.co/NfhmfMjtXw>
- [10] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, p. 86, 2020, doi: 10.33365/jtk.v14i2.792.
- [11] P. M. Purba, A. C. Amandha, R. H. Purnama, and A. Ikhwan, "Analisis Keamanan Website Prodi Sistem Informasi Uinsu Menggunakan Metode Application Scanning," *J. Inform. Teknol. dan Sains*, vol. 4, no. 4, pp. 325–329, 2022.
- [12] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [13] A. Pujiyanto, A. Mulyati, and R. Novaria, "Pemanfaatan Big Data Dan Perlindungan Privasi Konsumen Di Era Ekonomi Digital," *Maj. Ilm. Bijak*, vol. 15, no. 2, pp. 127–137, 2018, doi: 10.31334/bijak.v15i2.201.
- [14] Y. Djahir and D. Pratita, *Sistem Informasi Manajemen*, vol. 1, no. 1. Yogyakarta: Deepublish, 2014. [Online]. Available: <https://rudiatko.files.wordpress.com/2008/10/kuliahsim-1-2.pdf>
- [15] T. Sutabri, *Analisis Sitem Informasi*, vol. 53, no. 9. 2014.
- [16] T. Sutabri, *Konsep Sistem Informasi*. Yogyakarta: Andi, 2012.
- [17] N. Nana and E. Surahman, "Pengembangan Inovasi Pembelajaran Digital Menggunakan Model Blended POE2WE di Era Revolusi Industri 4.0," *Pros. SNFA (Seminar Nas. Fis. dan Apl.)*, vol. 4, p. 82, 2019, doi: 10.20961/prosidingsnfa.v4i0.35915.
- [18] D. Novianti, "Implementasi Algoritma Naive Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. J. Komput. dan Inform. Univ. Bina Sarana Inform.*, vol. 21, no. 2, pp. 143–148, 2019, doi: 10.31294/p.v20i2.
- [19] E. T. L. Kusriani, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [20] D. Alita and Y. Fernando, "Multiclass SVM Algorithm for Sarcasm Text in Twitter," *JATISI (Jurnal Tek. Inform. dan Sist. Informatika)*, vol. 8, no. 1, pp. 118–128, 2021, doi: 10.35957/jatisi.v8i1.646.