e-ISSN: 2460-870X | p-ISSN: 2085-1367

Analisis Sentimen Terhadap Film Sri Asih Dengan CNN, KNN, dan Logistic Regression

Sentiment Analysis of Film Sri Asih with CNN, KNN, and Logistic Regression

Ester Caroline Dwi Wijaya[1], Hendry[2]
[1,2]Teknik Informatika, Universitas Kristen Satya Wacana
[1,2]Jl. Diponegoro, Salatiga, Indonesia
[1]estercaroline43@gmail.com, [2]hendry@uksw.edu

ABSTRAK

Perkembangan teknologi menjadi dampak yang besar di berbagai sektor, salah satunya adalah di industri perfilman. Hal ini tentunya menjadi tantangan bagi para produser film untuk membuat karya yang sesuai dengan selera masyarakat. Film hangat yang menjadi perhatian masyarakat adalah film Sri Asih. Dengan adanya media sosial Twitter, penulis dapat menggunakannya sebagai pengamatan opini masyarakat untuk menganalis sentimen masyarakat terhadap film Sri Asih, dan melakukan observasi emosi pada sentimen tersebut. Dalam proses klasifikasi tersebut penelitian ini dilakukan dengan tiga metode yaitu Convolutional Neural Network (CNN) dengan hasil akurasi 78%, K-Nearest Neighbour (KNN) dengan hasil akurasi 61%, dan Logistic Regression dengan hasil akurasi 81%. Dengan menggunakan tiga metode yang berbeda, peneliti dapat menemukan bahwa algoritma baru yang kompleks dan membutuhkan resource yang besar, belum tentu dapat menyelesaikan permasalahan pada penelitian ini. Oleh karena itu dapat disimpulkan bahwa Logistic Regression yang merupakan algoritma yang sederhana, masih relevan dan mampu menyelesaikan permasalahan yang ada pada penelitian ini. Dengan hasil akurasi dari algoritma Logistic Regression, dapat digunakan untuk memenuhi tujuan penelitian ini yaitu memberikan informasi dengan baik mengenai opini masyarakat terhadap film Sri Asih, serta menjadi dasar evaluasi bagi produser film dalam mengembangkan kualitas film Sri Asih menjadi lebih baik.

Kata kunci : Analisis Sentimen, Twitter, CNN, KNN, Logistic Regression

ABSTRACT

Technological developments have had a big impact on various sectors, one of which is the film industry. This is of course a challenge for film producers to make works that suit people's tastes. The latest film that has become the public's attention is Sri Asih. With the existence of Twitter social media, the writer can use it as an observation of public opinion to analyze public sentiment towards the Sri Asih film and make emotional observations on this sentiment. In the classification process, this research was carried out using three methods, namely Convolutional Neural Network (CNN) with an accuracy of 78%, K-Nearest Neighbor (KNN) with an accuracy of 61%, and Logistic Regression with an accuracy of 81%. By using three different methods, researchers can find that new algorithms that are complex and require large resources, may not necessarily be able to solve the problems in this study. Therefore it can be concluded that Logistic Regression which is a simple algorithm, is still relevant and able to solve the problems that exist in this study. With the results of the accuracy of the Logistic Regression algorithm, it can be used to fulfill the objectives of this study, which is to provide good information about public opinion about the Sri Asih film, as well as being the basis for evaluation for film producers in developing the quality of the Sri Asih film for the better.

Keywords : Sentiment Analysis, Twitter, CNN, KNN, Logistic Regression

Info Artikel :

Disubmit:10 Mei *Direview:01 Juni* *Diterima :11 September*

Copyright © 2022 – CSRID Journal. All rights reserved.

1. PENDAHULUAN

Pesatnya perkembangan dunia teknologi memiliki pengaruh terhadap segala aktivitas manusia, baik di sektor pendidikan, perdagangan, komunikasi, dan lain sebagainya. Begitupun dengan perkembangan didunia perfilman yang semakin menjadi sorotan masyarakat, Film merupakan suatu

komunikasi yang berbentuk massa visual. Undang Undang Perfilman Nomor 33 tahun 2009 Ketentuan-Ketentuan Umum Ayat 1 menyatakan bahwa film merupakan salah satu media komunikasi massa yang merupakan karya seni budaya uang dan pranata sosial yang dibentuk dan dibuat berdasarkan sinematografi dengan suara atau tanpa suara yang dapat dipertunjukan [1]. Selain itu, film juga merupakan karya manusia yang memiliki estetika yang tinggi dan media yang berguna sebagai sarana untuk menyebarkan pesan kepada publik. Jadi, dapat dikatakan bahwa produser memproduksi film berguna untuk menyampaikan pesan melalui film yang dihasilkan [2]. Namun, menurut Ketua Gabungan Pengusaha Biskop Indonesia tidak sedikit masyarakat yang beranggapan bahwa film Indonesia tidak bermutu sehingga mengakibatkan daya tarik masyarakat dalam menonton film menjadi berkurang [3]. Hal ini tentunya dibutuhkan pemahaman mengenai opini masyarakat dalam pengalamannya menonton suatu film. Oleh karena itu, diperlukanlah penelitian analisis sentimen untuk mengetahui opini masyarakat terhadap film Indonesia supaya menjadi informasi yang berharga bagi masyarakat maupun produser film untuk mengevaluasi film Indonesia menjadi lebih berkualitas.

Terdapat beberapa penelitian sebelumnya yang melakukan analisis sentimen terhadap film Indonesia. Salah satunya adalah penelitian yang dilakukan oleh Adri Gabriel Sooai yang bertujuan untuk membantu masyarakat Indonesia dalam memperoleh informasi film dan membaca *review* film dari *website* yang dibuat oleh peneliti. Melalui data analisis sentimen yang diambil dari *user* lokal Twitter, Adri merepresentasikan sentimen tersebut kedalam *website review* film. Penelitian ini menyoroti analisis sentimen positif dan negatif dengan metode *Rule Based* dan *Naïve Bayes Classifier*. Dengan menggunakan metode tersebut penelitian ini menemukan hasil bahwa *Rule Based System* memiliki akurasi yang lebih besar yaitu 82,64% Sementara 74,09% untuk *Naïve Bayes Classifier* [4].

Penelitian yang lainnya adalah penelitian yang dilakukan oleh Arifin Kurniawan dkk. Arifin mengangkat isu analisis sentimen opini film dengan menggunakan *Naïve Bayes* dan *Lexicon Based Features*. Berdasarkan evaluasi dan data yang diambil dari forum KASKUS, penelitian ini menemukan bahwa masyarakat lebih cenderung memberikan hasil positif terhadap film Indonesia dengan hasil akurasi sebesar 80%. Penelitian ini menunjukkan pentingnya analisis sentimen untuk mengetahui umpan balik masyarakat terhadap suatu film [5].

Dari penelitian sebelumnya, belum terdapat penelitian yang membahas secara luas mengenai emosi yang terkandung dalam setiap sentimen masyarakat. Oleh karena itu penulis akan melakukan observasi mengenai emosi yang terdapat dalam sentimen tersebut. Dengan melakukan observasi emosi, data dari hasil sentimen analisis akan memberikan informasi yang lebih jelas, apakah sentimen tersebut mengandung emosi *joy* (bahagia), *anger* (marah), *surprise* (terkejut), *sadness* (sedih), *fear* (takut), atau *disgust* (menjijikan). Selain itu, penulis menggunakan tiga kelas dalam penelitian ini untuk mendapatkan pemahaman yang lebih komprehensif tentang sentimen atau opini yang dianalisis. Klasifikasi kelas netral tetap dilakukan karena tidak semua teks memiliki sentimen positif dan negatif, kelas netral dapat digunakan untuk menganalisis opini yang cenderung ragu-ragu (tidak positif dan tidak negatif). Jika banyak opini yang bersifat netral, artinya opini tersebut tidak mengekspresikan emosi yang kuat. Namun, opini netral dapat menyesuaikan konteks dalam penelitian, dan ketika diobservasi dengan isu-isu yang relevan maka dapat di representasikan sebagai opini positif atau negatif. Penulis akan mengklasifikasikan opini positif, netral, dan negatif dengan menggunakan metode *Convolutional Neural Network* (CNN), *K-Nearest Neighbour* (KNN), dan *Logistic Regression*. Ketiga metode ini memiliki cara dan keunikan masing-masing dalam pendekatan analisis sentimen, sehingga akan dilakukan observasi metode mana yang mampu menyelesaikan permasalahan dari penelitian ini.

Dalam penelitian ini, penulis akan fokus pada film Indonesia yang mendapat sorotan masyarakat, film tersebut adalah film “Sri Asih”. Sri Asih merupakan film pertama *superhero* wanita yang mengangkat cerita asli karya anak bangsa. Sri Asih rilis pada 17 November 2022 di seluruh layar lebar Indonesia. Sri Asih adalah film yang diadaptasi dari sebuah komik karya anak nusantara berjudul “Sri Asih”. Film ini diproduksi oleh Upi Avianto bersama Joko Anwar yang menceritakan

e-ISSN: 2460-870X | p-ISSN: 2085-1367

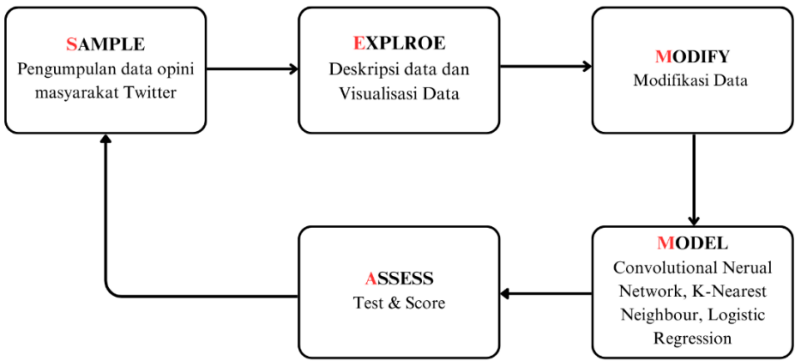
tentang Alana yaitu *superhero* wanita yang memiliki sifat marah yang menggebu, namun ia berusaha untuk mengendalikan amarahnya dan menjadi kebaikan untuk kehidupan dan sekitarnya.

Dengan genre dan cerita yang tergolong baru dan *fresh*, film ini tentunya menarik perhatian masyarakat Indonesia dan menuai banyak opini dari masyarakat. Salah satu cara masyarakat dalam menyampaikan opininya adalah dengan melalui media sosial Twitter. Twitter merupakan media sosial yang menjadi media komunikasi bagi hampir seluruh masyarakat di dunia. Berdasarkan laporan finansial dari Twitter pada kuartal ke-3 di tahun 2019, Twitter di klaim menjadi salah satu media sosial yang memiliki pengguna aktif harian yang paling banyak [6]. Oleh karena itu, penulis akan memanfaatkan Twitter sebagai penelitian terhadap film Sri Asih, dengan melakukan observasi apakah masyarakat cenderung beropini positif, netral, atau negatif [7].

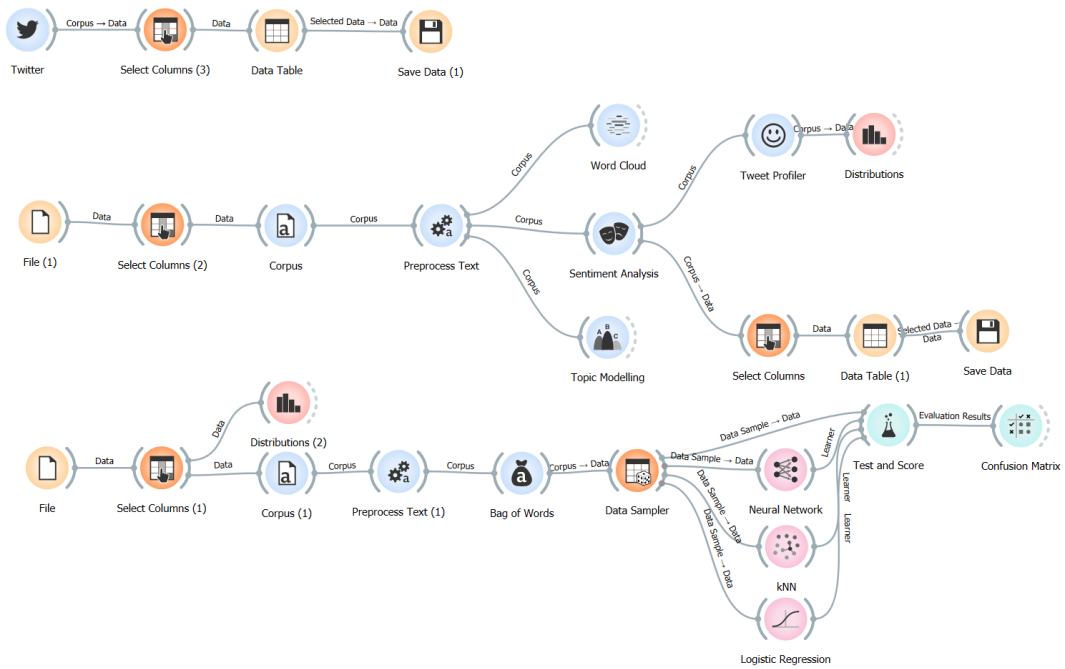
Dalam penelitian ini penulis menggunakan aplikasi Orange Data Mining untuk membantu setiap proses yang dilakukan, dengan harapan dapat mengetahui kecenderungan opini masyarakat sesuai klasifikasi positif, negatif, ataupun netral. Aplikasi Orange Data Mining merupakan aplikasi *open source* yang digunakan untuk menganalisis dan visualisasi data eksploratif yang dikembangkan di Universitas Ljubljana, Slovenia. Orange memiliki GNU General Public License yang dipublikasikan oleh Free Software Foundation [8]. Orange menyediakan berbagai fitur eksperimen, pembuatan model, eksplorasi, dan sistem rekomendasi [9]. Pada penelitian kali ini, aplikasi Orange memiliki fitur-fitur yang dibutuhkan antara lain *crawling* data Twitter, analisis sentimen, berbagai model algoritma, dan Tweet Profiler. Hasil dari penelitian ini diharapkan dapat memenuhi tujuan penelitian ini yaitu untuk memberikan informasi yang baik mengenai opini masyarakat terhadap film Sri Asih, Selain itu, penelitian ini juga bertujuan sebagai dasar dan evaluasi bagi produser film Indonesia dalam mengambil keputusan, reformasi strategi, dan mengembangkan produksi dan pemasaran film menjadi lebih baik sesuai dengan opini yang telah diberikan oleh masyarakat.

2. METODE PENELITIAN

Langkah-langkah proses penelitian ini ditunjukan pada Gambar 1. dan Gambar 2. Setiap proses yang dilewati harus berhasil dilakukan karena proses demi proses saling terhubung untuk dapat melakukan proses selanjutnya.



Gambar 1. Sistem Arsitektur



Gambar 2. Desain Alur Proses pada Aplikasi Orange

A. Sample

Dalam tahap ini dilakukan proses pengumpulan data. Data yang digunakan pada penelitian ini merupakan data opini masyarakat yang diambil langsung dari media sosial Twitter. Pengambilan data ini memerlukan API *key* Twitter yang sebelumnya diperoleh dari Twitter *Developer*. *Application Programming Interface* (API) merupakan salah satu cara suatu sistem komputer dapat berkomunikasi satu sama lain supaya mereka dapat meminta dan menerima informasi yang dibutuhkan. *Platform* API Twitter ini menyediakan akses ke dalam data Twitter publik yang dibagikan oleh pengguna ke seluruh dunia. Maka dari itu penulis dapat mengelola data Twitter dengan API *key* Twitter ini.

Dalam aplikasi Orange untuk menggunakan API *key* tersebut, penulis menggunakan *widget* Twitter pada aplikasi, data dapat diambil dengan memasukan kata kunci dari topik yang ingin dianalisis yaitu "sri asih" dan "bumi langit". Peneliti mengambil data sejumlah 8362 data unggahan Twitter pada 7 Desember 2022 sampai dengan 12 Desember 2022. Data tersebut kemudian disimpan dalam bentuk *file* excel untuk selanjutnya dapat dilakukan *pre-processing*.

B. Explore

Pada tahap ini dilakukan pemeriksaan dataset untuk mengetahui gambaran besar data yang dipakai, dengan menganalisis secara deskriptif dan visual untuk memahami karakteristik dataset. Dalam tahap eksplorasi dilakukan pemahaman data, menganalisis atribut data yang penting, dan data anomali.

C. Modify

Dalam tahap modifikasi dilakukan *pre-processing* data dan *Bag of Words*. *Pre-processing* data merupakan tahap dalam mengolah data teks untuk menghilangkan data-data yang tidak diperlukan. Sedangkan *Bag of Words* merupakan pembobotan kata dalam bahasa Indonesia.Hal ini diperlukan agar data yang digunakan dapat dilakukan analisis sentimen [10]. Adapun tahap yang dilakukan pada penelitian ini adalah sebagai berikut :

a) *Transformation* merupakan tahap membersihkan dan mengubah data input. Dalam *transformation* terdapat beberapa hal yang dilakukan yaitu *Lowercase* mengubah semua data teks menjadi huruf kecil. *Remove accents* yaitu menghapus karakter karakter aksen seperti

ä, ü, ě, dan lain sebagainya. *Parse html* yaitu mendeteksi *syntax* atau *tag* html yang terdapat didata menjadi teks biasa. *Remove urls* akan menghapus *urls* yang terdapat di data *input*.

b) *Tokenization* adalah proses untuk memecah teks menjadi suatu kata.
c) *Filtering* merupakan proses untuk menghapus kata yang tidak dibutuhkan setelah proses *tokenizing* sebelumnya. Dalam proses *filtering* kata akan di *stopword* berdasarkan kamus jenis kata yang dibuat sendiri berdasarkan dari data yang digunakan. Selain itu, dilakukan juga penghapusan tanda baca dan angka dengan menggunakan fitur regexp dan *numbers* yang bisa menghapus angka dan tanda baca seperti berikut "\.|,|:|;|!|\?|\(|\)|\||\+|\'|\"|‘|’|“|”|\'|\’|…|\-|–|—|\$|&|*|>|<|\/|\[|\]".
d) *Normalization* adalah proses *stemming* yaitu mengembalikan bentukan kata menjadi kata dasar dengan menghilangkan imbuhan dan menjadi kata yang berdiri sendiri.

D. Model

a) Data Training & Data Testing

Pada tahap ini akan dilakukan pembagian dataset. Dalam klasifikasi data umumnya dibagi menjadi dua bagian antara lain data *training* dan data *testing*. Proporsi untuk melakukan data *training* biasanya lebih besar dibanding data *testing*, namun biasanya proporsi yang digunakan adalah 70:30, 80:20, atau 90:10 [11]. Oleh karena itu, dengan menggunakan *widget* Data Sampler penulis membagi data menjadi 90% data *training* dan 10% data *testing* untuk mendapatkan hasil yang terbaik.

b) Analisis Sentimen

Dikarenakan data yang diambil merupakan data yang menggunakan bahasa Indonesia maka pada tahap ini dilakukan proses analisis sentimen dengan metode *multilingual sentiment* bahasa Indonesia. Pada analisis sentimen akan dilakukan proses penentuan sentimen negatif, sentimen positif, dan sentimen netral [12]. Dalam proses analisis sentimen akan dihasilkan data yang sudah dikategorikan menjadi opini yang bersifat negatif dengan nilai dibawah 0, positif nilai diatas 0, dan netral nilai sama dengan 0. Data yang dihasilkan akan disimpan dalam bentuk *file* excel dan diberi label positif, negatif, dan netral.

c) Pengujian Algoritma

Pada tahap ini akan dilakukan pengujian model dari data sentimen yang sudah diperoleh. Data sentimen yang telah dibagi menjadi data *training* dan data *testing* dilakukan pengujian dengan menggunakan metode *Nearest Neighbour (KNN), Neural Network*, dan *Logistic Regression*.

1) Convolutional Neural Network

Convolutional Neural Network (CNN) adalah metode perkembangan dari *multiplayer perceptron* (MLP). MLP didesain untuk mengolah data dua dimensi. Pada awalnya metode *Convolutional Neural Network* (CNN) digunakan pada bidang *image processing* atau *computer vision*. Namun, pada tahun 2015 metode CNN telah digunakan dalam bidang *natural language processing* (NLP), penelitian tersebut dilakukan untuk mengklasifikasi teks [13].

CNN adalah jenis *neural network* yang mengolah data dalam dimensi tinggi dan terdapat *neuron* yang saling terhubung satu sama lain. CNN terdiri dari beberapa lapisan yang saling terhubung untuk memproses data yang akan di *input*. Lapisan tersebut adalah *convolutional, pooling,* dan *fully connected*. Lapisan *convolutional* dan *pooling* bertugas untuk mempelajari data yang diinputkan, lalu lapisan *fully connected* bertugas untuk mengklasifikasikan data yang telah di *input* [14].

2) K-Nearest Neighbour (KNN)

K-Nearest Neighbour (KNN) adalah suatu metode yang digunakan untuk klasifikasi terhadap objek yang berdasarkan data training memiliki jarak terdekat dengan objek

tersebut [15]. *K-Nearest Neighbour* merupakan metode yang termasuk *instance-based* learning. Nilai k adalah bilangan bulat positif yang ditentukan oleh pengguna dengan pertimbangan semakin kecil nilai k akan semakin rentan dengan pengaruh noise/outlier, sedangkan semakin besar nilai k maka pengaruh noise/outlier akan semakin kecil namun batasan antar kelas akan semakin kabur [16]. Rumus *Euclidean Distance* adalah rumus yang digunakan dalam perhitungan *K-Nearest Neighbour*.

$$d\,(x,y) = \sqrt{(\sum_{i=1}^{n} (xi - yi)^2)} \quad (1)$$

3) Logistic Regression

Logistic regression merupakan salah satu teknik regresi yang memisahkan dataset menjadi dua kelompok. Namun, terdapat metode generalisasi dari *logistic regression* antara lain *multinomial logistic regression* yang dapat memprediksi data lebih dari dua kelompok. Dengan tujuan untuk melakukan analisis data kategori, menggunakan *logistic regression* merupakan model yang sederhana untuk dapat menjelaskan hubungan antar *output* variabel respon (Y) dan variabel penjelas (X) [6]. Rumus *logistic regression* adalah sebagai berikut:

$$Akurasi = \sum_{i=1}^{m} \frac{(y_{pred}^{(i)} == y_{true}^{(i)}}{m} \quad (3)$$

Keterangan :
m = jumlah data *test*

E. Assess

Tahap akhir adalah dengan melakukan evaluasi model yang telah dibangun dengan *widget Test & Score.* Dari tahap ini akan mengukur kinerja model dengan AUC, *accuracy, precision, F1*, dan *Recall.*

3. HASIL DAN PEMBAHASAN

A. Sample dan Explore

Dalam tahap *sample* dan *explore* penulis berhasil memperoleh data sebanyak 8362 data dalam periode waktu 9 Januari 2023 pukul 17.04 WIB sampai dengan 22.38 WIB. Data ini diambil dengan menggunakan *widget* Twitter dengan menginput *API key Twitter* dan *keyword* "sri asih" dan "bumi langit".

Data yang telah diambil kemudian dianalisis dan diambil atribut data yang akan digunakan dalam proses selanjutnya, yaitu atribut *content*. Kemudian disimpan dalam bentuk file excel menggunakan *widget Save Data*.

Tabel 1. Contoh Data Twitter

Content
Sri Asih tampil memukau di babak pertama. Plotnya padet. Cinematografinya keren ba...
cocologiku pagi ini. cover theme phantom ngingetin sama jeng sri asih sumpah ga boong g…
Di film ini Randy Korompis rasanya sangat jorjoran dengan CGI karna terlibat sebagai sutr…
Terus karena Teduh mulai ngeh sama temen-teman bapak ibunya siapa aja, dia juga ngeh…
Wah, apa aja sih keistimewaan dan asal-usul dari Sri Asih ini? Apa lagi fakta-fakta menarik…

e-ISSN: 2460-870X | p-ISSN: 2085-1367

B. Modify

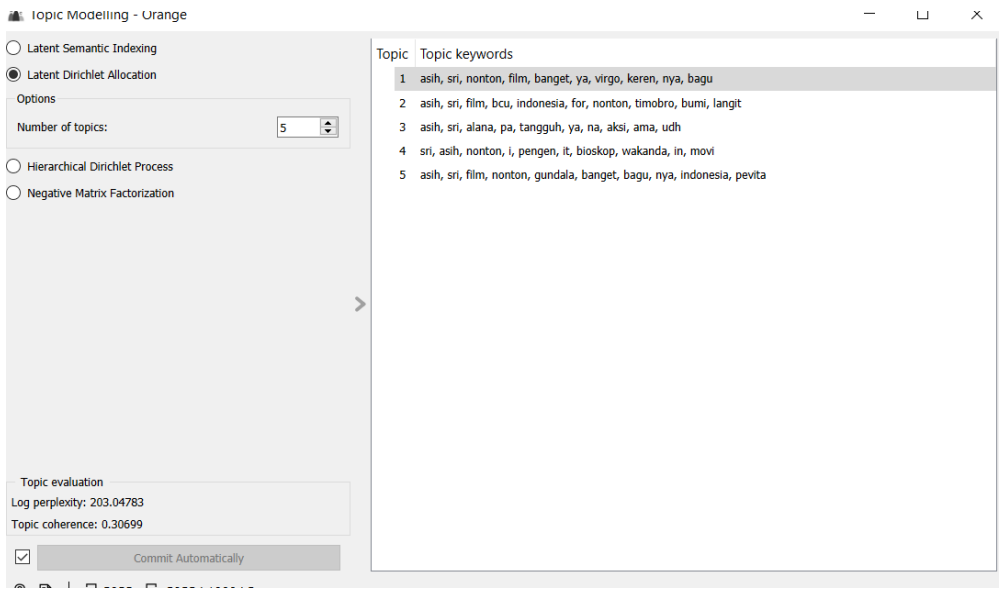
Data yang telah diperoleh dan dipilih atribut pentingnya kemudian dilakukan pengolahan data tersebut dilakukan *pre-processing* data agar komponen-komponen yang tidak diperlukan dapat dihapus dari data input. *Pre-processing* yang dilakukan antara lain menghapus tanda baca, kata imbuhan, memecah kalimat menjadi kata demi kata, dan lain sebagainya

Dalam tahap ini dilakukan juga persiapan dengan menggunakan *widget Bag of Words.* Dimana widget ini memberikan bobot pada setiap kata yang ada di *dataset*. Dari data hasil *pre-pro-processing* dan *Bag of Words* dapat dilihat topik yang sering muncul dan berkaitan dengan "sri asih" dan "bumi langit", untuk melihat topik tersebut digunakan *widget Word Cloud*. Fungsi pada *Word Cloud* adalah menampilkan kata berdasarkan frekuensi kemunculan kata tersebut. Semakin besar frekuensi kemunculannya maka semakin besar ukuran dari kata tersebut. Hasil tampilan *Word Cloud* dalam penelitian ini dapat dilihat di Gambar 3. Dari Gambar 3. Dapat dilihat bahwa kata yang sering muncul adalah sri, asih, film, gundala, nonton, keren, banget.



Gambar 3. Tampilan Word Cloud

Untuk lebih memperkuat analisis topik apa saja yang berkaitan dengan film Sri Asih maka digunakan *widget Topic Modelling*. *Topic Modelling* akan menampilkan kata yang paling sering mencul atau yang sering digunakan. Hasil dari *widget Topic Modelling* dapat dilihat dari Gambar 4. Dari Gambar 4 menunjukkan 5 topik teratas yang sering digunakan oleh pengguna Twitter dalam unggahannya, dalam setiap topik memiliki beberapa kata yang divisualisasikan dalam tampilan *word cloud*.



Gambar 4. Tampilan Topic Modelling

C. *Model*

a) *Pembagian Dataset dan Analisis Sentimen*

Selanjutnya dalam tahap ini akan dihasilkan data analisis sentimen yang telah diklasifikasikan menjadi tiga kelas, yaitu positif, netral, dan negatif. Data yang dipakai adalah data yang telah dibagi manjadi data *training* dan data *testing* dengan proporsi 90:10. Contoh dari hasil analisis sentimen adalah sebagai berikut :

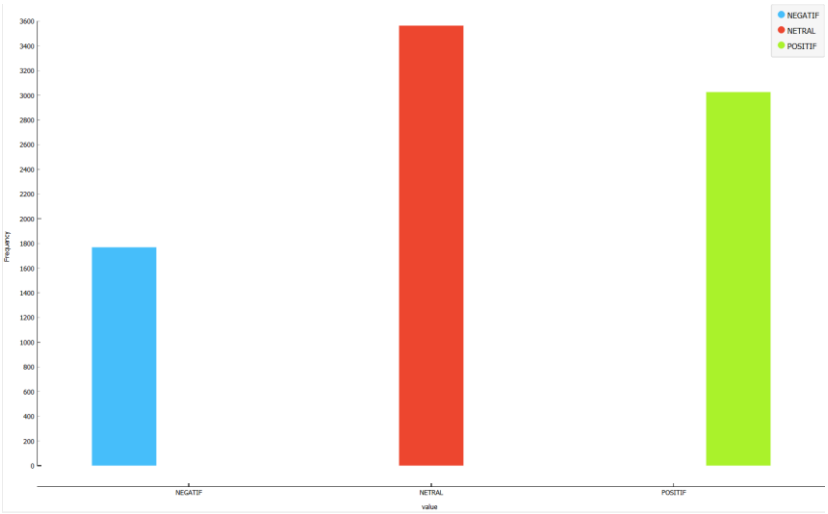
Tabel 1. Contoh Hasil Analisis Sentimen

Positif	Gue percaya Film Indonesia kedepannya arahnya akan bagaimana. Gue juga nemuin titik terang untuk masa depan perfilman Indonesia akhir akhir ini, dan diperkuat dengan film Sri Asih. Oh yaa, buat Ka Pevita, lu juga keren banget. Totalitas banget.. Gue Respect banget sama aktingya!
Netral	film sri asih jadi pioneer baru standar film indonesia dalam genre superhero dan kualitas CGI/VFX, secara story sebenarnya biasa aja dan banyak hal hal yang belum dijelaskan, banyak juga karakter dan villain yang dihilangkan dan dilupakan begitu aja.
Negatif	Plotnya kurang banget. Background cerita alana juga berantakan. Konflik kurang dalam. Dialog kaku. adegan berantem yang keliatan banget koreo-nya. Sri asih cantik banget tapi habis dihantam sana sini masa tetap badai rambut lurus semampai. Building character development nya kurang.
Positif	@bicaraboxoffice Padahal bagus kok, nikmati filmnya, jangan dibanding2ke, adegan aksinya juga keren, sampe terharu dan bangga lihat Indonesia punya jagoan cewek sehebat @sriasihofficial aku sih jatuh hati sama Sri Asih.
Netral	Overall keren banget walaupun masih ga terlalu menuhin ekspektasi ku Sri Asih nya bakal keluarin kekuatan yang lebih keren lagi (mungkin di film selanjutnya kali ya) tapi cukup mengobati rasa kangen setelah nonton Gundala terus ditambah bakal muncul Godam
Negatif	Gimana mau bagus filmnya kalo gak ada rasa "tertarik" sama characternya? Padahal pengemasan awal udah lumayan meyakinkan tapi gue ga dpt ketertarikan emosional dari sri asih karena karakternya lempeng. Justru malah lebih tertarik sm sahabatnya sri asih drpd main chara

Analisis sentimen pada penelitian ini menggunakan metode *Multilingual Sentiment* dengan basis bahasa Indonesia. Dengan *widget distribution* hasil analisis sentimen dapat divisualisasikan dalam bentuk diagram batang seperti Gambar 5. Dalam Gambar 5. Menunjukan bahwa dari 8362 data sebanyak 1770 (21,17%) masuk dalam kategori Negatif, sebanyak 3565 (42,63%) masuk dalam kategori Netral, dan sebanyak 3027 (36,20%) masuk dalam kategori Positif. Dilihat dari hasil visualisasi analisis sentimen tersebut terlihat bahwa lebih banyak masyarakat yang beranggapan netral terhadap rilisnya film Sri Asih di Indonesia. Hal ini menunjukkan bahwa masih banyak keraguan masyarakat terhadap kualitas film Sri Asih. Didukung dengan opini *user* Twitter yang menganggap bahwa film Sri Asih merupakan permulaan film *superhero* yang cukup

e-ISSN: 2460-870X | p-ISSN: 2085-1367

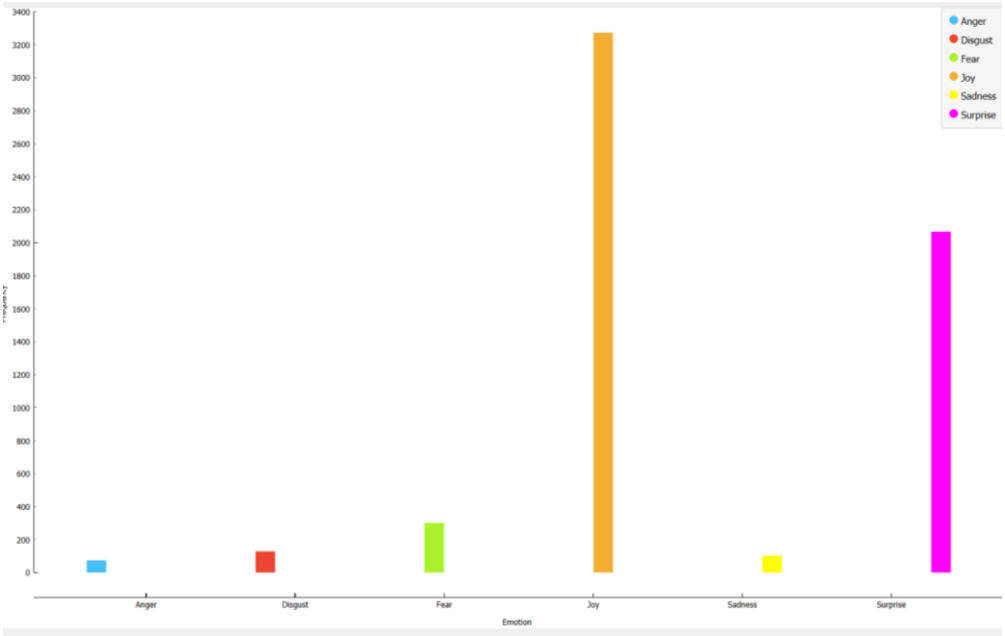
di Indonesia namun juga menyebutkan kelemahan-kelemahan dari film Sri Asih. Tidak sedikit juga masyarakat Indonesia beranggapan positif dengan mengatakan film Sri Asih dapat menjadi film *superhero* terbaik di Indonesia. Selain itu dalam kategori negatif, masih ada beberapa masyarakat yang menganggap film Sri Asih perlu banyak perbaikan dan peningkatan, serta menjadi film yang kurang layak ditonton.



Gambar 5. Hasil Analisis Sentimen

b) Tweet Profiler

Dari hasil analisis sentimen tersebut dapat dilihat emosi tiap opini para pengguna. Penelitian emosi dalam analisis sentimen dapat menjelaskan informasi sentimen menjadi lebih kompeherensif. Untuk memperoleh analisis emosi dari data input dapat menggunakan widget *Tweet Profiler*. *Widget* ini berfungsi menampilkan emosi berdasarkan kata-kata yang termuat di data input. Emosi yang ditampilkan antara lain *joy* (bahagia), *anger* (marah), *surprise* (terkejut), *sadness* (sedih), *fear* (takut), dan *disgust* (menjijikan). Hasil dari analisis emosi dapat dilihat dari Gambar 6. yang divisualisasikan dengan *widget distribution*. Dari hasil tersebut dapat dilihat bahwa film Sri Asih lebih banyak memberikan dampak emosi *joy* (kebahagiaan) terhadap para pengguna Twitter.



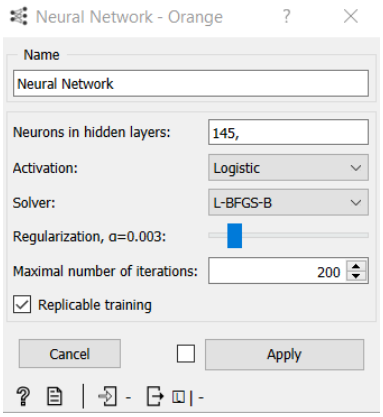
Gambar 6. Hasil Analisis Emotion

c) Cara Kerja Algoritma

Selanjutnya dilakukan pengujian algoritma analisis sentimen dengan beberapa algoritma yaitu, *K-Nearest Neighbour*, *Convolutional Neural Network*, dan *Logistic Regression*. Penggunaan dari ketiga metode ini bertujuan untuk mengetahui metode manakah yang dapat menyelesaikan permasalahan pada penelitian ini dengan baik.

1) Convolutional Neural Network (CNN)

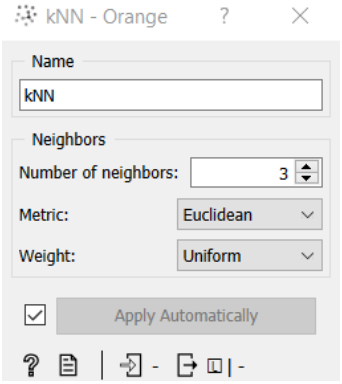
Pendekatan CNN dalam melakukan analisis sentimen yaitu dengan memperhatikan fitur-fitur spasial dalam teks. Dengan menggunakan *widget neural network* pada aplikasi Orange, peneliti dapat mengatur parameter untuk mendapatkan hasil yang terbaik.



Gambar 7. Parameter CNN yang digunakan

2) K-Nearest Neighbour (KNN)

Dalam pendekatannya untuk analisis sentimen KNN mempertimbangkan kemiripan teks dengan tetangga terdekat. Dalam penelitian ini penulis dibantu dengan *widget* KNN pada Orange dan mengatur parameter KNN untuk mendapatkan hasil yang terbaik. Parameter tersebut dapat dilihat pada Gambar 8.

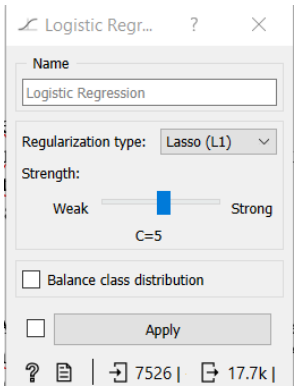


Gambar. 8 Parameter KNN yang digunakan

3) Logistic Regression

Logistic Regression merupakan algoritma yang sederhana dibanding dengan kedua metode sebelumnya, dalam melakukan pendekatan analisis sentimen Logistic regression memodelkan hubungan linear antara fitur dan sentimen. Dengan menggunakan *widget Logistic Regression*, penulis dapat mengatur parameter yang diperlukan untuk mendapatkan hasil terbaik.

e-ISSN: 2460-870X | p-ISSN: 2085-1367



Gambar 9. Parameter Logistic Regression yang digunakan

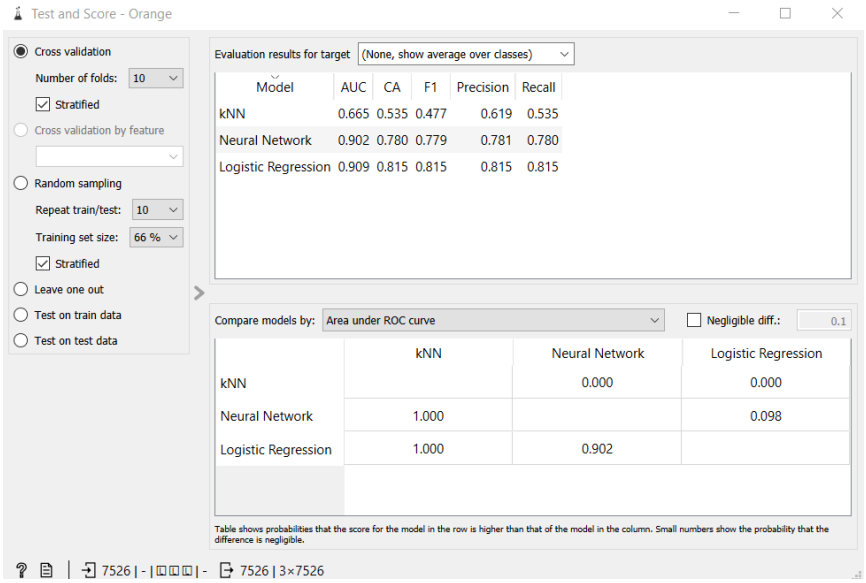
D. Assess

Hasil pembangunan model akan dievaluasi untuk mengukur kinerja dari masing-masing algoritma. Untuk melakukan evaluasi ini penulis menggunakan *widget Test & Score* dan *Confussion Matrix.*

Dari *widget Test & Score* diperoleh AUC (*Area Under ROC Curve*) yang dapat dengan mudah membandingkan model satu dengan yang lainnya, CA (*Classification Accuracy*) adalah akurasi dataset, F1 (F1 *Score*) yaitu acuan perbandingan *precision* dan *recall*, *Precision* digunakan untuk dapat mengetahui ketepatan data, dan *Recall* digunakan untuk mengetahui tingkat keberhasilan sistem dalam menemukan kembali data tersebut. Pada penelitian ini *Test & Score* menggunakan *Cross Validation* dengan nilai *k-fold 10*. Maka diperoleh hasil dari pengujian algoritma *K-Nearest Neighbour*, *Convolutional Neural Network*, dan *Logistic Regression* sebagai berikut :

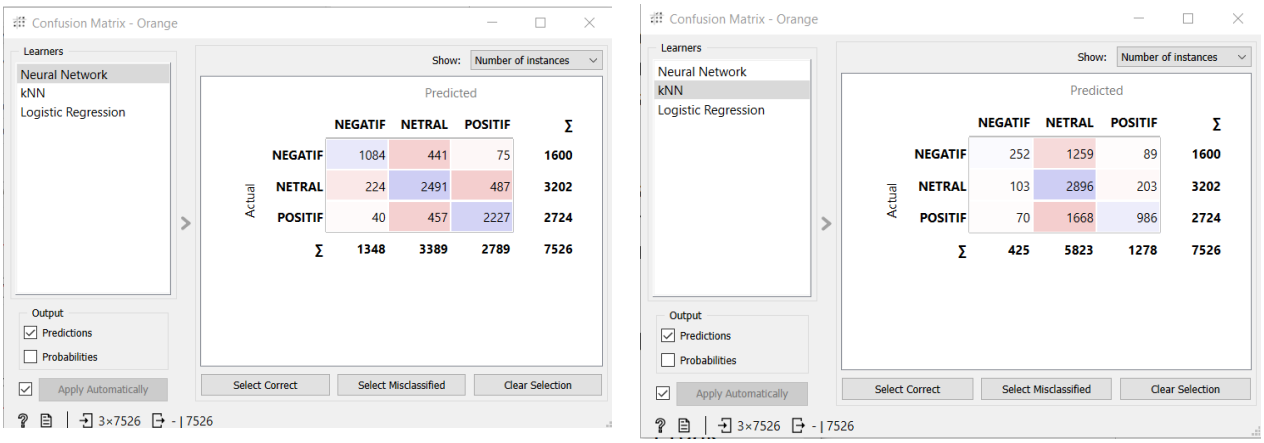
Tabel 2. Hasil Test & Score dalam Bentuk Persen

Model	AUC	CA	F1	Precision	Recall
kNN	66,5%	53,5%	47,7%	61,9%	53,5%
Neural Network	90,2%	78,0%	77,9%	78,1%	78,0%
Logistic Regression	90,9%	81,5%	81,5%	81,5%	81,5%

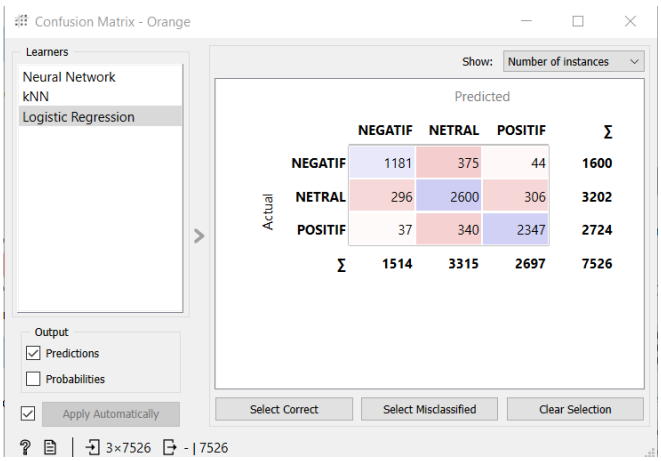


Gambar 10. Hasil Evaluasi Model dengan Test & Score

Lalu untuk melihat hasil evaluasi kinerja algoritma dalam melakukan prediksi analisis sentimen dapat menggunakan *widget Confussion Matrix.* Dalam penelitian ini, penulis mendapatkan hasil sebagai berikut :



Gambar 11. Confussion Matrix KNN



Gambar 12. Confussion Matrix Logistic Regression

Dari evaluasi dengan *widget Test & Score* dan *Confusiion Matrix* pada masing-masing algoritma, *Logistic Regression* memiliki hasil akurasi yang lebih tinggi yaitu diatas 81%. Hal ini menjadi penemuan bahwa dalam penelitian ini tidak diperlukan algoritma baru yang kompleks dan membutuhkan *resource* besar. Namun, algoritma yang sederhana seperti *Logistic Regression* sudah cukup dinyatakan mampu menyelesaikan permasalahan pada penelitian ini.

4. KESIMPULAN

Dari penelitian ini, dapat disimpulkan bahwa algoritma yang sederhana seperti *Logistic Regression* mampu menyelesaikan permasalahan penelitian ini dengan akurasi diatas 81%. Dengan meneliti emosi dari analisis sentimen ini, dihasilkan bahwa banyak masyarakat yang bahagia (*joy)*, namun cenderung belum puas dengan kualitas film Sri Asih.

Saran untuk penelitian selanjutnya diharapkan data yang digunakan dapat lebih banyak dari penelitian sebelumnya, supaya dapat meningkatkan pemberian informasi dan memperoleh akurasi algoritma yang lebih baik dan akurat. Serta disarankan untuk dapat melakukan penelitian dengan menggunakan metode yang berbeda, agar dapat lebih mengetahui perbandingan dengan metode-metode yang digunakan sebelumnya.

REFERENSI

[1] N. Yuwita, 2018, “Representasi Nasionalisme Dalam Film Rudy Habibie (Study Analisis Semiotika Charles Sanders Pierce),” J. Herit., vol. 6, no. 1, pp. 1689–1699.

[2] R. S. M. Permana, L. Puspitasari, and S. S. Indriani, 2019, "Industri film Indonesia dalam perspektif sineas Komunitas Film Sumatera Utara," ProTVF, vol. 3, no. 2, pp. 185, doi: 10.24198/ptvf.v3i2.23667.

[3] M. Y. Rizky and Y. Stellarosa, 2019, "Preferensi Penonton Terhadap Film Indonesia," Commun. J. Commun. Stud., vol. 4, no. 1, pp. 15, doi: 10.37535/101004120172.

[4] A. G. Sooai and M. Laniwati, 2021, "Ekstraksi Ulasan Sentimen Film dari Twitter dengan Naïve Bayes pada Situs Web Media Sosial Penggemar Film," J. Intell. Syst. Comput., vol. 3, no. 1, pp. 49–54, doi: 10.52985/insyst.v3i1.186.

[5] Arifin Kurniawan, Indriati Indriati, and Sigit Adinugroho, 2019, "Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 3, no. 9, pp. 8335–8342.

[6] K. Kelvin, J. Banjarnahor, E. I., and M. NK Nababan, 2022, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA), vol. 5, no. 2, pp. 47–52, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365.

[7] B. Pamungkas, M. E. Purbaya, and D. J. A.K, 2021, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," J. Informatics, Inf. Syst. Softw. Eng. Appl., vol. 3, no. 2, pp. 10–20.

[8] F. Tscherne, N. Wilke, B. Schachenhofer, K. Roux, and G. Tavlaridis, 2016, "The thermo lignum ecological insect pest eradication process: The effects on gilded and painted wooden objects," Int. J. Conserv. Sci., vol. 7, no. 1, pp. 295–300.

[9] S. Indah Nurhafida and F. Sembiring, 2021, "Analisis Text Clustering Masyarakat Di Twiter Mengenai Mcdonald'Sxbts Menggunakan Orange Data Mining," SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform., vol. 1, no.1, pp. 28–35.

[10] A. J. Putri, A. S. Syafira, M. E. Purbaya, and D. Purnomo, 2022, "Analisis Sentimen E-Commerce Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine," J. TRINISTIK J. Tek. Ind. Bisnis Digit. dan Tek. Logistik, vol. 1, no. 1, pp. 16–21, doi: 10.20895/trinistik.v1i1.447.

[11] P. Alam Jusia, F. Muhammad Irfan, and Kurniabudi, 2019, "Clustering Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means," J. IKRA-ITH Inform., vol. 3, no. 3, pp. 75.

[12] M. P. Agustina and H. Hendry, 2021, "Sentimen Masyarakat Terkait Perpindahan Ibukota Via Model Random Forest dan Logistic Regression," Aiti, vol. 18, no. 2, pp. 111–124, doi: 10.24246/aiti.v18i2.111-124.

[13] D. H. Qudsi, J. H. Lubis, K. U. Syaliman, and N. F. Najwa, 2021, "Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan Sentiment Analysis in the Student ' S Reviews of College Department Performance Using," J. Teknol. Inf. dan Ilmu Komput., vol. 8, no. 5, pp. 1067–1076, doi: 10.25126/jtiik.202184842.

[14] A. R. Maulana and N. Rochmawati, 2020, "Opinion Mining Terhadap Pemberitaan Corona di Instagram menggunakan Convolutional Neural Network," J. Informatics Comput. Sci., vol. 2, no. 1, pp. 53–59, doi: 10.26740/jinacs.v2n01.

[15] J. A. Josen Limbong, I. Sembiring, K. Dwi Hartomo, 2019, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based With Naïve Bayes and K-Nearest Neighbor Meth," J. Teknol. Inf. dan Ilmu Komput., vol. 9, no. 2, pp. 347–356, doi: 10.25126/jtiik.202294960.

[16] N. R. Wardani, S. Saepudin, and C. Warman, 2022, "Sentimen Analisis Kegiatan Trading Pada Aplikasi Twitter dengan Algoritma SVM , KNN Dan Random Forrest," J. S. Komputer, vol. 6, no.2, pp. 863–870.