

Analisis Sentimen Pada Media Sosial Twitter Menggunakan *Support Vector Machine (SVM)* Terhadap Kata Kunci #TheLinkinJKT

1st AsrianaFakultas Rekayasa Industri
Universitas Telkom

Bandung, Indonesia

asrianayaya@student.telkomuniversity.
com2nd Oktariani Nurul PratiwiFakultas Rekayasa Industri
Universitas Telkom

Bandung, Indonesia

onurulp@telkomuniversity.ac.id

3rd Hanif FakhurrojaFakultas Rekayasa Industri
Universitas Telkom

Bandung, Indonesia

haniff@telkomuniversity.ac.id

Abstrak—Pada penelitian ini dilakukan sentimen analisis terhadap *tweetss* terkait #TheLinkinJKT mengenai segi “Sistem Ticketing” dan “Kepuasan Event” dengan kategori sentimen label positif dan negatif. Sentimen analisis dilakukan menggunakan pendekatan *Machine Learning* dengan algoritma *Support Vector Machine (SVM)*. Terdapat 1.331 jumlah data keseluruhan pada “Sistem Ticketing” dengan 650 *tweetss* dengan label positif dan 705 *tweetss* dengan label negatif. Sedangkan pada “Kepuasan Event” terdapat 4.400 jumlah data keseluruhan dengan 1.995 dengan label positif dan 2.402 dengan label negatif, pelabelan data dilakukan secara manual. Berdasarkan hasil analisis menggunakan SVM, akurasi tertinggi pada “Sistem Ticketing” dengan rasio split 70:30 sebesar 0,737 atau 73% dan akurasi tertinggi pada “Kepuasan Event” dengan rasio split 80:20 sebesar 0,743 atau 74%. Adapun hasil evaluasi kinerja model menggunakan *K-Fold Cross Validation* pada “Sistem Ticketing” dengan hasil *accuracy* sebesar 0,74, *precision* sebesar 0,71, *recall* sebesar 0,71 dan *f1-score* sebesar 0,71, sedangkan pada “Kepuasan Event” hasil *accuracy* sebesar 0,74, *precision* sebesar 0,82, *recall* sebesar 0,57, dan *f1-score* sebesar 0,68. Hasil dari analisis ini diharapkan dapat memberikan digunakan sebagai *feedback* bagi pihak terkait untuk meningkatkan kualitas layanan yang diberikan dari sisi sistem *ticketing* dan kepuasan konsumen terhadap *event* yang diselenggarakan oleh pihak promotor Dyandra Global Edutainment untuk acara-acara berikutnya.

Kata kunci— *Analisis Sentimen, Support Vector Machine, Natural Language Processing*

I. PENDAHULUAN

Perkembangan teknologi informasi di Indonesia berkembang sangat pesat, hal ini memberikan dampak pada setiap pengguna atau masyarakat dalam mendapatkan berbagai informasi baik dari media sosial, media cetak atau media digital. Perkembangan tersebut memberikan kebebasan masyarakat untuk mengeluarkan pendapatnya dan menjadi penyebar informasi melalui media sosial yang dimiliki. Media sosial menjadi salah satu platform bagi pengguna untuk saling berbagi informasi, pendapat, dan perasaan mereka secara online.[1]

Twitter merupakan sosial media yang populer Indonesia yang dimana pada saat ini, Twitter merupakan

wadah untuk mengekspresikan pendapat mereka secara bebas. Dengan demikian, twitter banyak diminati dan memberikan pengaruh dalam penelitian terutama dengan menggunakan metode sentimen analisis. [2] Sentimen analisis adalah suatu proses yang dapat menganalisis data text seperti pada *tweets*, komentar, ulasan dan lainnya. Sentimen analisis mampu mengelompokkan polaritas dari suatu teks dalam kalimat untuk mengetahui apakah opini dari kalimat tersebut termasuk positif atau negatif. Penelitian ini bertujuan untuk melakukan analisis sentimen pada *tweets* yang berkaitan dengan “sistem *ticketing*” dan “kepuasan *event*” berdasarkan *hashtag* #TheLinkinJKT menggunakan metode *Support Vector Machine (SVM)*, untuk memberikan informasi mengenai sentimen yang dominan pada *tweets* tersebut. *Support Vector Machine (SVM)* merupakan salah satu *classifier* yang dapat menghasilkan hasil yang paling baik dengan memiliki tingkat akurasi paling tinggi dibandingkan dengan *classifier* lainnya [2]. SVM kerap diterapkan pada banyak kasus dengan akurasi tinggi pada kategorisasinya [3]. Maka dari itu, dalam penelitian ini akan menggunakan SVM sebagai klasifikasi pengkategorian penelitian. Dengan menggunakan SVM, diharapkan dapat diketahui sentimen pengguna Twitter yang berkaitan dengan “sistem *ticketing*” dan “kepuasan *event*” berdasarkan *hashtag* #TheLinkinJKT. [1] Dengan menggunakan *hashtag* atau “#” sebagai kata kunci untuk pencarian suatu informasi di Twitter, dapat memudahkan pengguna untuk mencari suatu informasi. Pada penelitian kali ini, membahas tentang penggunaan *hashtag* #TheLinkinJKT yang menjadi populer di kalangan NCTzen, yaitu sebutan bagi penggemar *boyband* NCT 127 dari Korea Selatan. *Boyband* ini telah sukses mengadakan *event* konser tur dunia yang bertepatan NCT 127 2ND TOUR, NEO CITY: THE LINK – JAKARTA, khususnya untuk penggemar di Indonesia. Acara tur konser tersebut dilaksanakan pada tanggal 4-5 November 2022 di Indonesia Convention Exhibition (ICE) BSD, Tangerang. Acara konser ini diselenggarakan oleh pihak promotor Dyandra Global Edutainment. *Hashtag* #TheLinkinJKT digunakan oleh penggemar

untuk mengekspresikan antusiasme menyambut kedatangan NCT 127 ke Indonesia dan mengikuti konser yang telah diadakan. Penggunaan *hashtag* #TheLinkinJKT ini menjadi *trending* 1 pada *platform* Twitter, dimulai dari penjualan tiket hingga berakhirnya konser. Penelitian ini bertujuan untuk melakukan analisis sentimen pada *tweets* yang berkaitan dengan “Sistem *Ticketing*” dan “Kepuasan *Event*” berdasarkan *hashtag* #TheLinkinJKT dengan menggunakan metode *Support Vector Machine* (SVM). Hasil dari analisis ini diharapkan dapat memberikan informasi mengenai sentimen yang dominan pada *tweets* tersebut, sehingga dapat digunakan sebagai *feedback* bagi pihak terkait untuk meningkatkan kualitas layanan yang diberikan dari sisi sistem *ticketing* dan kepuasan konsumen terhadap *event* yang diselenggarakan oleh pihak promotor Dyandra Global Edutainment untuk acara-acara berikutnya, sehingga kejadian seperti acara konser yang dibubarkan seperti yang terjadi pada konser NCT 127 di Indonesia tidak terulang kembali dan dapat berlangsung dengan tertib.

II. KAJIAN TEORI

A. Twitter

Twitter adalah *platform* media sosial populer yang memungkinkan orang berkomunikasi melalui *tweets* dan teks pendek. Panjang maksimum *tweets* dibatasi hingga 140 karakter [2]. Batasan ini mengakibatkan penggunaan bahasa informal dan kurang formal secara konsisten, karena pengguna hanya perlu mengungkapkan perasaannya dalam satu atau dua kalimat. Banyak penelitian telah dilakukan di bidang ini dalam beberapa tahun terakhir [3]. Twitter adalah salah satu media sosial yang banyak digunakan oleh pengguna untuk berbagi informasi dan pendapat mereka secara online. Twitter memiliki fitur *hashtag* atau “#” yang dapat digunakan oleh pengguna untuk mengumpulkan *tweets* yang berkaitan dengan topik tertentu [1].

B. Analisis Sentimen

Analisis sentimen adalah proses yang digunakan untuk menganalisis data teks seperti *tweets*, komentar, ulasan dan lainnya untuk mengelompokkan polaritas dari suatu teks dalam kalimat. Analisis sentimen dapat memberikan informasi mengenai persepsi atau pandangan pengguna terhadap suatu topik tertentu [4]. sehingga dapat digunakan untuk memberikan *feedback* bagi pihak terkait untuk meningkatkan kualitas layanan yang diberikan.

C. Natural Language Processing (NLP)

Natural language processing (NLP) merupakan pengolahan bahasa natural yang dapat digunakan untuk mengatasi hambatan pengenalan Bahasa percakapan sehari-hari oleh sistem komputer [1]. Sentimen analisis merupakan bagian dari *Natural Language Processing* [2]. Penerapan NLP memiliki potensi untuk memberikan bantuan dalam beragam aspek, terutama dalam analisis dan ekstraksi informasi dari data yang tidak terstruktur seperti pada Twitter [3]. Terdapat dua teknik utama dalam memahami NLP yaitu analisis sintaksis dan analisis semantik. Kedua teknik ini berfungsi untuk memvalidasi struktur bahasa. Analisis sintaksis berfokus pada tata bahasa, sementara analisis semantik bertujuan untuk menafsirkan makna suatu kalimat [4].

D. Data Preprocessing

Data preprocessing adalah suatu tahap dalam pengolahan data yang bertujuan untuk mengubah data mentah menjadi data yang siap untuk dianalisis. Tahapan ini dilakukan agar bersih dari noise dan lebih terstruktur sehingga bisa diolah lebih lanjut [8]. Adapun tahapan *data preprocessing* memiliki beberapa proses, yaitu *data labeling*, *data cleaning*, *case folding*, *tokenizing*, *stopwords removal*, normalisasi kata dan *stemming*.

Pada tahapan-tahapan yang dilakukan di atas dapat disesuaikan dengan kebutuhan masing-masing proses. Tahapan dapat berbeda antara satu dengan yang lain, berikut merupakan tahapan yang dilakukan:

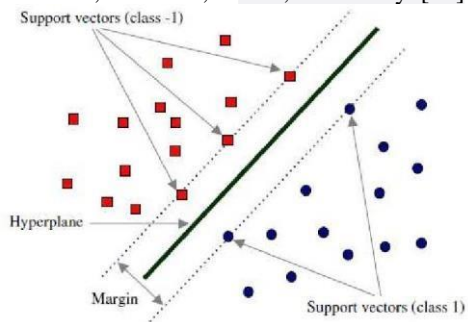
1. *Data labeling*, proses *labeling* melibatkan pemberian label positif atau negatif pada *tweets* atau teks berdasarkan kata-kata yang terdapat didalamnya. Langkah ini dilakukan dengan membandingkan kata-kata dalam *tweets* atau teks dengan daftar kata-kata yang ada, baik untuk kategori positif maupun negatif. [5]
2. *Data Cleaning*, merupakan proses pembersihan teks seperti data teks yang tidak diperlukan atau tidak memiliki arti yang bersumber dari data teks yang didapat [6]. Dalam proses data cleaning ini dapat dilakukan beberapa proses yang dapat disesuaikan dengan kebutuhan data cleaning pada masing-masing data:
 - a. Penghapusan *mention* (@) atau penyebutan *username* pada *tweets*
 - b. Penghapusan karakter tunggal atau *single char* agar menghilangkan karakter yang tidak relevan.
 - c. Penghapusan tanda baca atau *punctuation* agar menghapus tanda baca pada *tweets*.
 - d. Penghapusan *hashtag* # agar menghilangkan *hashtag* (#) atau tagar pada *tweets*.
 - e. Penghapusan URL agar menghilangkan tautan pada *tweets*.
 - f. Penghapusan *whitespace* agar menghilangkan garis baru pada *tweets*.
 - g. Penghapusan angka agar menghilangkan angka pada *tweets*.
3. *Case Folding*, adalah proses mengubah huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*) atau sebaliknya. *Case folding* dilakukan untuk menyeimbangkan bentuk huruf dalam teks, sehingga tidak terdapat perbedaan dalam pemrosesan teks.
4. *Tokenizing*, adalah proses pemecahan suatu teks menjadi unit-unit kecil yang disebut token.

Token dapat berupa kata, karakter, atau frase dan tanda baca yang terdapat dalam teks [10].

5. *Stopwords Removal*, adalah proses menghapus kata-kata yang tidak memiliki makna atau tidak berguna dalam teks. *Stopwords* biasanya adalah kata-kata yang umum atau baku seperti kata yang, ke, di, dll. Penghilangan *stopwords* dilakukan untuk mengurangi jumlah token yang tidak bermakna dan meningkatkan akurasi dari hasil analisis teks [10].
6. Normalisasi Kata, jika ditemukan karakter yang diulang secara berurutan maka akan diubah menjadi satu karakter seperti contoh karakter berulang seperti kata “okeee” akan menjadi “oke”, “baikkk” menjadi “baik”. Pada normalisasi juga bisa menghilangkan kata yang mengandung satu karakter seperti kaya “y” atau “g” yang kemudian kata slang tersebut bisa diubah sesuai [11] [15]
7. *Stemming*, adalah proses mengubah kata dasar (base word) dari sebuah kata yang memiliki bentuk yang berbeda namun memiliki arti yang sama. *Stemming* dilakukan untuk mengurangi jumlah token yang tidak bermakna dan meningkatkan akurasi dari hasil analisis teks [10].

E. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu metode dalam klasifikasi data yang dapat menghasilkan hasil akurasi yang paling baik [12]. SVM dapat diterapkan pada banyak kasus dengan akurasi tinggi dalam mengklasifikasikan data. SVM bekerja dengan memilih garis yang paling tepat untuk memisahkan data menjadi dua kelas dalam mengidentifikasi *hyperplane*. SVM dapat digunakan untuk mengklasifikasikan teks, seperti *tweets*, komentar, ulasan, dan lainnya [13].



GAMBAR 1.
Hyperplane SVM

Untuk mendapatkan garis *hyperplane* yang paling optimal untuk memisahkan data ke dalam dua buah kelas tersebut, dapat menggunakan perhitungan *margin hyperplane* sehingga mendapatkan titik maksimal. Untuk mendapatkan *hyperplane* pada SVM, dapat menggunakan rumus berikut:

$$1 = w \cdot x + b$$

Persamaan (1)

$$-1 = w \cdot x + b$$

Persamaan (2)

Kelas positif memenuhi persamaan $1 = w \cdot x + b$ dan kelas negatif memenuhi persamaan $-1 = w \cdot x + b$. Luas antara dua persamaan paralel merupakan *margin* antara dua *hyperplane* yang dinyatakan dalam persamaan [3].

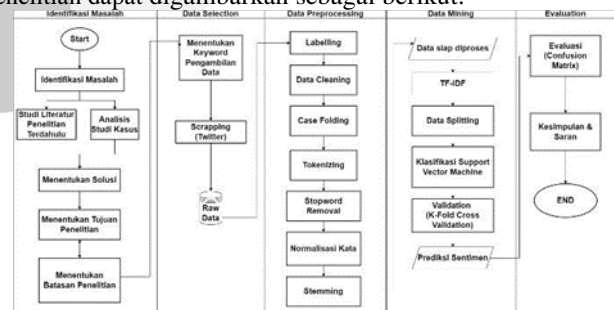
F. Evaluation

Dalam penelitian ini menggunakan metode evaluasi *K-Fold Cross Validation* untuk mengevaluasi kinerja pemodelan dan *Confusion Matrix* untuk menilai akurasi dari suatu model klasifikasi. Pada proses ini, dilakukan dengan membagi data menjadi lipatan yang sama atau K subset, dengan salah satu subset menjadi data latih dan data uji. Dalam penelitian ini, menggunakan dengan salah satu subset menjadi data latih dan data uji. Dalam penelitian ini, menggunakan modul *KFold_cross_val_score* dari modul *sklearn.model_selection* dalam library *sklearn* dengan menggunakan parameter *k_folds* = 5 yang menunjukkan K=5 dan menggunakan *random_state*=42.

III. METODE

Metodologi yang digunakan pada penelitian ini diadopsi dari *Design Science in Information System Research* [7] Metode ini dipilih karena dapat membantu peneliti dalam memecahkan masalah yang kompleks, membantu mengidentifikasi dan mengurangi celah antara teknologi dan bisnis, membantu mengembangkan produk/layanan berbasis teknologi yang inovatif, membantu mengidentifikasi, menganalisis dan menerapkan solusi untuk meningkatkan kinerja serta mengurangi biaya, serta dapat membantu mengembangkan teknik, alat dan metodologi guna meningkatkan efisiensi penelitian yang dilakukan.

Sistematika penyelesaian masalah dalam penelitian ini dengan menggunakan *Knowledge Discovery in Database*. *Dataset* yang digunakan dianalisis untuk mengetahui pola yang teridentifikasi pada dataset. Adapun tahapan dari sistematika penyelesaiannya yaitu: identifikasi masalah, data selection, data preprocessing, data transformation, data mining, dan evaluasi. Sebelum memulai tahapan proses data selection, akan dilakukan identifikasi masalah dan studi literatur terlebih dahulu sehingga solusi yang dihasilkan dapat sesuai atau relevan dengan masalah terkait. Tahapan-tahapan penelitian dapat digambarkan sebagai berikut:



GAMBAR 2.
Sistematika Penelitian

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

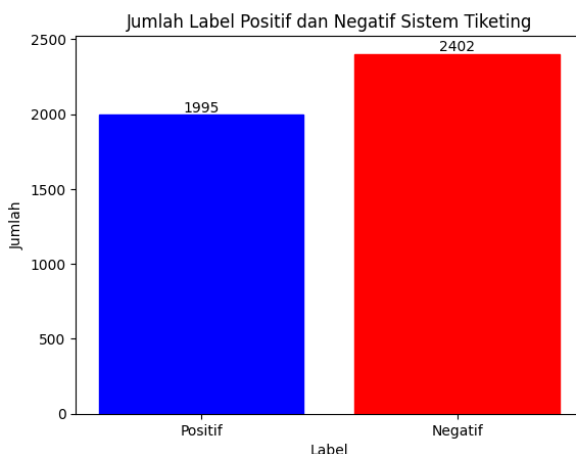
Dataset adalah proses atau tahapan pengambilan data atau *crawling data* dari internet yang dapat dilakukan secara otomatis. Data yang didapat akan digunakan untuk penelitian. *Dataset* diambil menggunakan *snsrape* pada Python. Data yang diambil merupakan data dari Twitter. Pengambilan data menggunakan kata kunci “#TheLinkinJKT”. data diambil memiliki rentan waktu yaitu antara 12 September 2022 sampai 6 November 2022. Data yang akan digunakan pada pembuatan model *Machine Learning* dalam mempengaruhi kualitas model agar dapat menghasilkan model yang baik, maka dibutuhkan adanya persiapan *dataset*. Setelah dilakukan *crawling data*, pelabelan secara manual dan label sentimen yang dipakai pada penelitian ini adalah label positif dan label negatif, data dengan label netral yang didapat tidak digunakan karena tidak relevan dengan penelitian. Sehingga diperoleh total 1.331 data untuk sistem *ticketing* dan 4.400 data untuk kepuasan *event* yang akan digunakan untuk penelitian.

B. Data Labeling

Pelabelan data merupakan proses dimana dataset yang telah diambil atau dilakukan *crawling data* diberi label. Pada tahap ini, pelabelan dilakukan secara manual terhadap 1.331 data pada sistem *ticketing* dan 4.400 data pada kepuasan *event* yang akan digunakan untuk pemodelan.

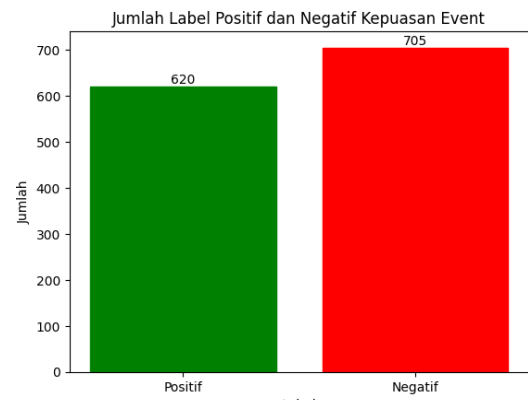
Pelabelan ini bertujuan untuk meningkatkan akurasi klasifikasi. Adapun teknik pelabelan yang dipakai untuk penelitian ini yaitu berdasarkan konteks. Semua *tweetss* diberi label positif dan label negatif berdasarkan konteks dari *tweetss* tersebut.

Pada awalnya penulis mengidentifikasi label netral dimana teks tidak menunjukkan dukungan ataupun menjelekkan terhadap aspek sistem *ticketing* dan kepuasan *event*. Pada akhirnya, label netral tidak digunakan dalam pemrosesan karena label netral dianggap tidak relevan karena tidak dapat memberikan informasi yang jelas sehingga terdapat kalimat yang sulit dianalisis dengan jelas.



GAMBAR II3.

Persebaran Label Sistem *Ticketing*



GAMBAR 4.

Persebaran Label Kepuasan *Event*

Persebaran label dari sentimen positif dan sentimen negatif yang berjumlah sebanyak 620 positif dan 705 negatif dari dataset sistem *ticketing*. Sedangkan adalah persebaran label dari sentimen positif dan sentimen negatif yang berjumlah sebanyak 1.995 positif dan 2.402 negatif dari dataset kepuasan *event*.

C. Akurasi Algoritma SVM

TABEL 1
Tabel Hasil Akurasi SVM setiap rasio

Dataset	Akurasi		
	80:20	70:30	60:40
Sistem <i>Ticketing</i>	0,722	0,737	0,724
Kepuasan <i>Event</i>	0,743	0,731	0,734

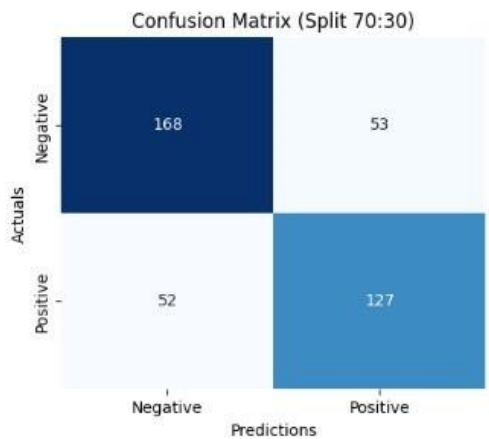
Berdasarkan tabel di atas, akurasi dari *dataset* sistem *ticketing* tertinggi didapatkan dari rasio *data splitting* 70:30 dengan akurasi sebesar 0,737 atau 73%, sedangkan akurasi tertinggi dari *dataset* kepuasan *event* didapatkan dari rasio *data splitting* 80:20 dengan akurasi sebesar 0,743 atau 74%.

D. Evaluasi Model

Dalam penelitian ini menggunakan dua evaluasi model yaitu dengan *K-Fold Cross Validation* untuk mengevaluasi model dan menggunakan *Confusion Matrix*.

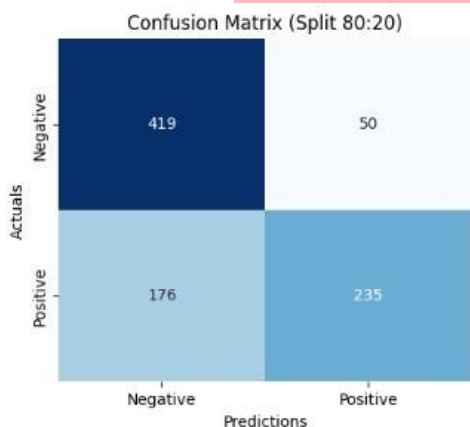
Didapatkan hasil dari akurasi *dataset* tertinggi pada “Sistem *Ticketing*” dengan rasio *data split* terbesar yaitu rasio 70:30 dengan akurasi sebesar 0,737 atau 73% dan akurasi tertinggi dari *dataset* “Kepuasan *Event*” didapatkan dari rasio *data split* 80:20 dengan akurasi terbesar yaitu 0,743 atau 74%. Kemudian dilakukan evaluasi model dengan menggunakan *K-Fold Cross Validation* dan *Confusion Matrix*, didapatkan dari hasil rasio tertinggi 70:30 pada “Sistem *Ticketing*” dengan hasil *confusion matrix* TP sebesar 127, TN sebesar 168, FP sebesar 53 dan FN sebesar 52, dengan evaluasi kinerja menggunakan *K-Fold Cross Validation* memiliki akurasi sebesar 0,74, *precision* sebesar 0,71, *recall* sebesar 0,71, dan *f1-score* sebesar 0,71. Pada “Kepuasan *Event*” didapatkan dari hasil data split tertinggi yaitu dengan rasio 80:20 dengan hasil *confusion matrix* TP sebesar 235,

TN sebesar 419, FP sebesar 50 dan FN sebesar 176, dengan evaluasi kinerja menggunakan *K-Fold Cross Validation* dengan akurasi sebesar 0,74, *precision* sebesar 0,82, *recall* sebesar 0,57, dan *f1-score* sebesar 0,68.



GAMBAR 5.

Confusion Matrix Split 70:30 pada Sistem Ticketing



GAMBAR 6.

Confusion Matrix Split 80:20 Kepuasan Event

TABEL 2.

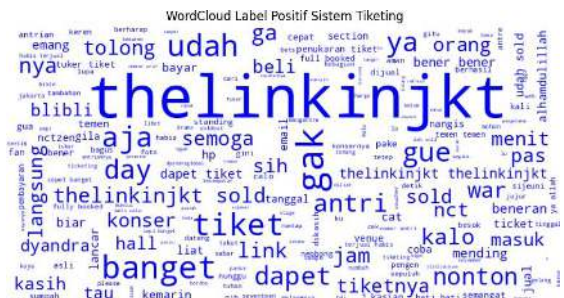
Tabel Hasil *Classification Report* pada Sistem *Ticketing*

Rasio	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
80:20	0,72	0,67	0,72	0,69
70:30	0,74	0,71	0,71	0,71
60:40	0,72	0,67	0,72	0,73

TABEL 3

Tabel Hasil *Classification Report* pada Kepuasan Event

Rasio	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
80:20	0,74	0,82	0,57	0,68
70:30	0,73	0,79	0,57	0,72
60:40	0,73	0,79	0,56	0,66



GAMBAR 7.

WordCloud Label Positif Sistem Ticketing

Pada Gambar 7 di atas merupakan hasil wordcloud sentimen label positif pada sistem *ticketing*. Untuk menggambarkan mengenai sistem *ticketing* berdasarkan data yang memiliki sentimen positif terdapat beberapa kata yang sering muncul diantaranya seperti kata “*thelikinjkt*”, “*tiket*”, “*banget*”, “*blibli*”, dll. Berdasarkan kata yang sering muncul dengan sentimen label positif di atas, menggambarkan bahwa customer banyak menuliskan *tweets* dengan kata kunci “*thelinkinjkt*” untuk menandakan bahwa *tweetss* tersebut mengenai *event* konser NCT 127 yang diselenggarakan. Customer menuliskan pada label sentimen positif ini menyukai sistem *ticketing* pada aplikasi dan website *blibli* yang berjalan dengan lancar ketika *ticketing* berlangsung sehingga pada beberapa customer dengan cepat bisa langsung membeli tiket konser nct 127.



GAMBAR 8.

WordCloud Label Negatif Sistem *Ticketing*

Pada Gambar 8 di atas merupakan hasil wordcloud sentimen label negatif pada sistem *ticketing*. Untuk menggambarkan mengenai sistem *ticketing* berdasarkan data yang memiliki sentimen negatif terdapat beberapa kata yang sering muncul diantaranya seperti kata “thelikinjkt”, “gak”, “jual”, “sold”, dll. Berdasarkan kata yang sering muncul dengan sentimen label negatif di atas, menggambarkan bahwa customer banyak menuliskan *tweets* dengan kata kunci “thelinkinjkt” untuk menandakan bahwa *tweetss* tersebut mengenai *event* konser NCT 127 yang diselenggarakan. Customer menuliskan pada label sentimen negatif ini bahwa tiket dengan cepat terjual atau sold. Terdapat pula kata seperti “calo” yang menandakan bahwa *customer* kecewa karena penggemar tidak kebagian tiket sedangkan tiket banyak berhasil dibeli oleh calo dan langsung dijual dengan harganya yang tinggi.



GAMBAR 9.

WordCloud Label Positif Kepuasan *Event*

Pada Gambar 9 di atas merupakan hasil wordcloud sentimen label positif pada kepuasan *event*. Untuk menggambarkan mengenai kepuasan *event* berdasarkan data yang memiliki sentimen positif terdapat beberapa kata yang sering muncul diantaranya seperti kata “thelikinjt”, “konser”, “senang”, “tolong”, dll. Berdasarkan kata yang sering muncul dengan sentimen label positif di atas, menggambarkan bahwa customer banyak menuliskan *tweets* dengan kata kunci “thelinkinjt” untuk menandakan bahwa *tweetss* tersebut mengenai *event* konser NCT 127 yang diselenggarakan. Pengguna menuliskan pada label sentimen positif ini menuliskan *tweets* tentang para penggemar yang datang ke konser NCT 127 untuk bersenang-senang menikmati konser yang diselenggarakan dan menuliskan cuitan agar berhati-hati untuk tidak rusuh dan dorong-dorongan ketika konser berlangsung dan agar penggemar yang datang untuk menonton konser NCT 127 dapat tertib.



GAMBAR 10.

WordCloud Label Negatif Kepuasan *Event*

Pada Gambar 10 di atas merupakan hasil wordcloud sentimen label negatif pada kepuasan *event*. Untuk menggambarkan mengenai kepuasan *event* berdasarkan data yang memiliki sentimen negatif terdapat beberapa kata yang sering muncul diantaranya seperti kata “thelikinjt”, “gak”, “dorong”, “rusuh”, dll. Berdasarkan kata yang sering muncul dengan sentimen label negatif di atas, menggambarkan bahwa pengguna banyak menuliskan *tweets* dengan kata kunci “thelinkinjt” untuk menandakan bahwa *tweetss* tersebut mengenai *event* konser NCT 127 yang diselenggarakan. Pengguna menuliskan pada label sentimen negatif ini menuliskan *tweets* tentang para penggemar yang datang ke konser NCT 127 yang kecewa terhadap konser dibubarkan pada tanggal 4 November 2022. Terjadi dorong-dorongan sehingga beberapa penggemar merasa malu dan tidak bisa bersenang-senang menikmati konser yang diselenggarakan.

N-gram Label Positif Sistem Tiketing:

```
0      udah sold out
1      dah sold out
2      tucker tiket day
3      sold out aja
4      terjual habis cepat
5      sold out guys
6      penukaran tiket day
7      pengen banget nonton
8      langsung full booked
9      hp harapan orang
10     harapan orang tua
11     web nya udah
12     web nya lancar
13     war tiket dapet
14     udah dapet tiket
15     tucker tiket tanggal
16     tucker tiket gak
17     tujuh day hall
18     tiket tanggal hall
19     tiket habis terjual
Name: bigram/trigram, dtype: object
```

N-gram Label Negatif Sistem Tiketing:

```
0      gak niat nonton
1      udah sold out
2      jual harga mahal
3      email pembayaran berhasil
4      udah full booked
5      jual tiket harga
6      war tiket dijual
7      war tiket gak
8      udah sold aja
9      udah dapet email
10     udah berhasil bayar
11     tiket fully booked
12     pembayaran berhasil pas
13     nonton gak war
14     nonton gak dapet
15     kasian benaran nonton
16     gak nomor urut
17     gak nomor antri
18     dapet email pembayaran
19     war tiket jual
Name: bigram/trigram, dtype: object
```

GAMBAR 11.

N-Gram Sistem *Ticketing*

Berdasarkan Gambar 11 di atas, menunjukkan bahwa kata atau kalimat yang sering muncul sebagai representasi sentimen positif dari sistem *ticketing* yaitu tiket terjual habis dengan cepat yang menunjukkan bahwa minat penggemar NCT 127 terhadap konser yang diselenggarakan sangat tinggi. Kemudian terdapat kalimat web nya lancar yang menunjukkan bahwa customer puas terhadap website dan aplikasi yang digunakan sebagai partner *ticketing* untuk pembelian tiket konser NCT 127. Kemudian pada representasi sentimen negatif dari sistem *ticketing* yaitu terdapat kalimat “jual harga mahal” yang merujuk kepada calo yang membeli tiket konser NCT 127 dan menjual tiket dengan harga mahal, kemudian terdapat kalimat “gak nomor urut” yang merujuk kepada tidak adanya nomor urut atau *queue number* yang diberikan kepada *customer* yang membeli tiket konser NCT 127. Terdapat pula kalimat “pembayaran berhasil pas” yang merujuk kepada ketika *customer* ingin melakukan pembayaran namun ketika diklik pembayaran tidak bisa dilakukan karena kesalahan sistem atau *system error* pada aplikasi.


```

N-gram Label Positif Kepuasan Event:
0   selamat bersenang senang
1   day berjalan lancar
2   buang sampah sembarangan
3   terima kasih sijeuni
4   hati selamat bersenang
5   bersenang senang nonton
6   nonton konser nct
7   nonton day tertib
8   hati hati selamat
9   ga dorong dorongan
10  nonton day tolong
11  tolong banget tertib
12  tolong banget mah
13  tolong banget day
14  rusuh dorong dorongan
15  konser selamat bersenang
16  day tolong banget
17  sijeuni udah tertib
18  konser berjalan lancar
19  day tolong tertib
Name: bigram/trigram, dtype: object

N-gram Label Negatif Kepuasan Event:
0   buang sampah sembarangan
1   nonton konser nct
2   selamat bersenang senang
3   ga dorong dorongan
4   dorong dorong an
5   tertib dorong dorongan
6   nonton day tolong
7   malu malu in banget
8   dorong dorongan ga
9   sedih banget liat
10  rusuh dorong dorongan
11  kejadian kaya gini
12  kejadian itaewon kemarin
13  dorong dorongan tolong
14  tolong hati hati
15  tolong banget day
16  orang pingsan konser
17  malu malu in
18  hati hati tolong
19  day tolong tertib
Name: bigram/trigram, dtype: object

```

GAMBAR 12.
N-Gram Kepuasan Event

Berdasarkan Gambar 12 di atas, menunjukkan bahwa kata atau kalimat yang sering muncul sebagai representasi sentimen positif dari kepuasan event yaitu selamat bersenang-senang yang menunjukkan bahwa cuitan terkait memberikan selamat agar penggemar yang datang ke konser NCT 127 dapat bersenang-senang menikmati konser yang diselenggarakan. Kemudian terdapat kalimat tolong banget tertib yang menunjukkan bahwa cuitan tersebut mengingatkan agar tertib ketika konser berlangsung. Kemudian pada representasi sentimen label negatif yaitu terdapat kalimat “buang sampah sembarangan” dimana menunjukkan bahwa kekecewaan terhadap para penggemar yang datang ke event konser NCT 127 dan membuang sampah sembarangan di venue konser. Terdapat pula kalimat “malu-maluin” yang menunjukkan bahwa karena kejadian konser pada tanggal 4 November diberhentikan banyak penggemar merasa malu dan kecewa karena kejadian tersebut.

V. KESIMPULAN

Adapun kesimpulan dari penelitian ini adalah sebagai berikut:

1. Pada penelitian ini menggunakan data dari opini masyarakat yang merupakan pengguna media sosial Twitter yang melakukan *tweetss* terkait event konser NCT 127 di Indonesia dengan kata kunci #TheLinkinJKT. Berdasarkan hasil analisis sentimen menggunakan SVM pada media sosial Twitter terkait “Sistem Ticketing” dan “Kepuasan Event” dengan kata kunci #TheLinkinJKT diperoleh jumlah keseluruhan data untuk “Sistem

Ticketing” yaitu 1.331 data *tweetss* dengan dilakukan *labeling* secara manual diperoleh hasil label positif berjumlah 620 dan hasil label negatif berjumlah 705 data *tweetss*. Pada jumlah keseluruhan data “Kepuasan Event” diperoleh 4.400 data *tweetss*, dengan dilakukan *labeling* secara manual diperoleh hasil label positif berjumlah 1.995 dan jumlah label negatif berjumlah 2.402.

2. Pada penelitian ini dilakukan proses atau tahapan data *scraping* untuk mengolah data mentah menjadi data yang siap untuk dilakukan pemodelan dengan melakukan tahapan *data cleaning*, *case folding*, *tokenizing*, *stopword removal*, normalisasi kata, dan *stemming*. Kemudian data yang telah dilakukan tahapan *data scraping* siap diproses untuk dilakukan tahapan selanjutnya yaitu *data transformation* dengan menggunakan TF-IDF. Berdasarkan hasil analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) dilakukan data *splitting* dengan melakukan pengujian rasio untuk data latih dan data uji sebesar 80:20, 70:30, dan 60:40. Didapatkan hasil dari akurasi *dataset* tertinggi pada “Sistem Ticketing” dengan rasio data *split* terbesar yaitu rasio 70:30 dengan akurasi sebesar 0,737 atau 73% dan akurasi tertinggi dari *dataset* “Kepuasan Event” didapatkan dari rasio data *split* 80:20 dengan akurasi terbesar yaitu 0,743 atau 74%. Kemudian dilakukan evaluasi model dengan menggunakan *K-Fold Cross Validation* dan *Confusion Matrix*, didapatkan dari hasil rasio tertinggi 70:30 pada “Sistem Ticketing” dengan hasil *confusion matrix* TP sebesar 127, TN sebesar 168, FP sebesar 53 dan FN sebesar 52, dengan evaluasi kinerja menggunakan *K-Fold Cross Validation* memiliki akurasi sebesar 0,74, *precision* sebesar 0,71, *recall* sebesar 0,71, dan *f1-score* sebesar 0,71. Pada “Kepuasan Event” didapatkan dari hasil *data split* tertinggi yaitu dengan rasio 80:20 dengan hasil *confusion matrix* TP sebesar 235, TN sebesar 419, FP sebesar 50 dan FN sebesar 176, dengan evaluasi kinerja menggunakan *K-Fold Cross Validation* dengan akurasi sebesar 0,74, *precision* sebesar 0,82, *recall* sebesar 0,57, dan *f1-score* sebesar 0,68.

Kesimpulan ini menunjukkan bahwa metode SVM dapat digunakan untuk melakukan analisis sentimen pada data *tweetss* terkait “Sistem Ticketing” dan “Kepuasan Event” pada kata kunci #TheLinkinJKT pada media sosial Twitter dengan tingkat akurasi yang cukup baik. Pada data “Kepuasan Event”, akurasi yang didapatkan lebih tinggi dibandingkan dengan data “Sistem Ticketing”, tetapi *recall* pada data “Kepuasan Event” lebih rendah, menunjukkan bahwa ada beberapa *tweetss* positif yang tidak terdeteksi dengan baik. Evaluasi ini memberikan pandangan tentang seberapa baik model dapat mengklasifikasikan sentimen *tweetss* yang diberikan dan dapat membantu dalam memahami tanggapan pengguna terkait topik “Sistem Ticketing” dan “Kepuasan Event” pada kata kunci #TheLinkinJKT di Twitter.

REFERENSI

- [1] Elik Hari Muktafin, "Analisis Sentimen pada Ulasan Pembelian Produk di *Marketplace* Shopee Menggunakan Pendekatan Natural Language Processing," *Jurnal Eksplora Informatika*, 2020.
- [2] Ardianne Luthfika Fairuz, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *Jurnal Dinda*, vol. 1, 2021.
- [3] Fitrah Rumaisa, "Penerapan *Natural Language Processing* (NLP) Di Bidang Pendidikan," *Jurnal Inovasi Masyarakat*, vol. 1, pp. 232-235, 2021.
- [4] Vincentius Riandaru Prasetyo, "Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya," 2021.
- [5] S. Kevin Yudi, "*Sentiment Analysis of Transjakarta Based on Twitter using Convolutional Neural Network*," *Advances in Science, Technology and Engineering Systems Journal*, vol. 4, pp. 281-286, 2019.
- [6] Mohammad Reza Faisal, "Text Mining Untuk Pemula," *Scripta Cendekia*, 2022.
- [7] A. R. Hevner, "*Design Science In Information Systems Research*," *MIS Quarterly*, vol. 28, 2004.
- [8] S. B. Somiya Rani, "Sentiment Analysis on Twitter Data using Support Vector Machine," *International Journal of Scientific Research in Computer Science Applications and Management Studies*, vol. 7, no. 3, 2018.
- [9] T. Jo, "Text Mining Concepts, Implementation, and Big Data Challenge," vol. 45, 2022.
- [10] Annisa Raudya Wibowo, "Analisis Sentimen Hashtag "DIRUMAHAJA" Saat Pandemi COVID-19 Di Indonesia Menggunakan NLP," *Jurnal Informatika dan Sistem Informasi*, vol. 1, 2020.
- [11] Agus Sasmito Aribowo, "Cross-domain sentiment analysis model on Indonesian YouTube comment," *International Journal of Advances in Intelligent Informatics*, vol. 1, pp. 12-25, 2021.
- [12] Ikhwanul Hakim, "Sentimen Analisis Stay Home menggunakan metode klasifikasi Naive Bayes, Support Vector Machine, dan k-Nearest Neighbor," *Paradigma – Jurnal Informatika dan Komputer*, vol. 22, 2020.
- [13] Deni Gunawan, "Komparasi Algoritma Support Vector Machine Dan Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023," *Jurnal Teknik Komputer AMIK BSI*, vol. VI, 2020.
- [14] Abdy Yoga Syantara, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Terhadap Kata Kunci "#Asiangames2018"," vol. 3, pp. 493-500, 2021.
- [15] Septian Andre Jeremy, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *Journal Of Intelligent Systems and Computation*, 2019.