

ANALISIS SENTIMEN DENGAN ALGORITMA *SUPPORT VECTOR MACHINE* TERHADAP PENYAKIT *HEPATITIS* AKUT MISTERIUS

Januar Adiputra¹, Deni Mahdiana^{2*}

^{1,2}Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹januarptr@gmail.com, ²deni.mahdiana@budiluhur.ac.id

(* : coresponding author)

Abstrak—Setelah meredanya pandemi *Covid-19*, saat ini sudah mulai muncul penyakit *hepatitis* akut yang dianggap misterius oleh masyarakat. Penyakit *hepatitis* akut sampai saat ini masih belum diketahui penyebabnya, penyakit ini pertama kali ditemukan di Inggris Raya pada tanggal 05 April 2022. Penyakit ini berbeda dari penyakit *hepatitis* biasanya, karena menurut WHO (*World Health Organization*) penyebab dari penyakit *hepatitis* akut bukan dari virus *hepatitis* A, B, C, D, dan E yang biasanya. Di Indonesia sendiri, terdapat tiga kasus suspek *hepatitis* akut pada anak yang meninggal setelah pengobatan, terkait dengan *hepatitis* akut yang terjadi dalam waktu dua minggu hingga 30 April 2022. Oleh karena itu pada penelitian ini akan membutuhkan sebuah analisa sentimen terhadap opini masyarakat berdasarkan data *twitter* terkait penyakit *hepatitis* akut misterius. Permasalahan yang diangkat pada penelitian ini adalah mengukur seberapa besar tingkat akurasi yang akan dihasilkan pada penerapan algoritma *Support Vector Machine* dalam mengklasifikasikan sentimen dari sebuah *tweet* terkait penyakit *hepatitis* akut misterius. Pada penelitian ini akan melakukan analisa sentimen menggunakan algoritma *Support Vector Machine* (SVM) dan menerapkan metodologi *Cross-Industry Standard Process for Data Mining* (CRIPS-DM). Metodologi CRIPS-DM meliputi tahapan mulai dari pemahaman bisnis, pemahaman data, persiapan data, pemodelan, serta evaluasi. Data bersih yang berhasil dikumpulkan berjumlah 2058 data yang diambil dari media sosial *twitter* pada 24 Mei 2022 hingga 30 Mei 2022 dengan total sentimen positif 1590 dan total sentimen negatif 468 dengan menggunakan *Azure Machine*. Penerapan model *Support Vector Machine* pada penelitian ini menghasilkan nilai *Accuracy* sebesar 81.06%, *Precision* 81.88%, dan *Recall* 97.19% menggunakan pengujian *split data* pada perbandingan 80:20 untuk data *training* dan data *testing*.

Kata Kunci: *text mining*, analisis sentimen, *SVM*, *hepatitis* akut misterius

Abstract—After the easing of the *Covid-19* pandemic, an acute *hepatitis* disease has begun to appear which is considered mysterious by the public. Until now, the cause of acute *hepatitis* is still unknown, this disease was first discovered in Great Britain on April 5, 2022. This disease is different from usual *hepatitis*, because according to WHO (*World Health Organization*) the cause of acute *hepatitis* is not viral *hepatitis*. The usual A, B, C, D, and E. In Indonesia alone, there are three suspected cases of acute *hepatitis* in children who died after treatment, related to acute *hepatitis* that occurred within two weeks until 30 April 2022. Therefore, this study will require a sentiment analysis of public opinion based on related *twitter* data. mysterious acute *hepatitis*. The problem raised in this study is to measure how much accuracy will be generated in the application of the *Support Vector Machine* algorithm in classifying the sentiment of a *tweet* related to mysterious acute *hepatitis*. In this study, sentiment analysis will be carried out using the *Support Vector Machine* (SVM) algorithm and applying the *Cross-Industry Standard Process for Data Mining* (CRIPS-DM) methodology. The CRIPS-DM methodology includes stages ranging from business understanding, data understanding, data preparation, modeling, and evaluation. The net data collected was 2058 data taken from social media *twitter* on May 24, 2022 to May 30, 2022 with a total of 1590 positive sentiments and 468 total negative sentiments using *Azure Machine*. The application of the *Support Vector Machine* model in this study resulted in an *Accuracy* value of 81.06%, *Precision* 81.88%, and *Recall* 97.19% using data split testing at a ratio of 80:20 for training data and testing data.

Keywords: *text mining*, sentiment analysis, *SVM*, mysterious acute *hepatitis*

1. PENDAHULUAN

Belum lama setelah meredanya pandemi *Covid-19*, saat ini sudah mulai muncul penyakit *hepatitis* akut yang dianggap misterius oleh masyarakat. Penyakit *hepatitis* akut sampai saat ini masih belum diketahui penyebabnya, penyakit ini pertama kali ditemukan di Inggris Raya pada tanggal 05 April 2022. Menurut WHO (*World Health Organization*), *hepatitis* akut tidak disebabkan oleh virus *hepatitis* A, B, C, D, dan E yang umum, sehingga penyakitnya berbeda dengan *hepatitis* biasa. Gejala awal *hepatitis* akut adalah gangguan saluran cerna seperti mual, muntah, sakit perut, dan diare. Di Indonesia sendiri, terdapat tiga kasus suspek *hepatitis* akut pada anak yang meninggal setelah pengobatan, terkait dengan *hepatitis* akut yang terjadi dalam waktu dua minggu hingga 30 April 2022. WHO menetapkan penyakit *hepatitis* akut sebagai Kejadian Luar Biasa (KLB), sehingga kejadian terkait penyakit ini menjadi banyak perbincangan di media sosial salah satunya yaitu *twitter*. Banyak

masyarakat yang memberikan pendapat dan opini melalui *twitter* yang membahas terkait penyakit *hepatitis* akut. Opini tersebut bisa berupa opini positif dan juga opini negatif, tergantung bagaimana masyarakat menanggapi terkait penyakit tersebut. Karena penyakit *hepatitis* akut ini termasuk ke dalam kriteria penyakit yang berbahaya dan menular, sehingga kementerian kesehatan RI menghimbau masyarakat agar tetap waspada dan tetap mematuhi protokol kesehatan meskipun pada saat ini sudah ada kebijakan pelonggaran masker terkait *Covid-19*. Oleh karena itu dibutuhkan sebuah analisa sentimen terhadap opini masyarakat dari sebuah *tweet* di *twitter* terkait penyakit *hepatitis* akut yang sekarang sedang banyak diperbincangkan dimana-mana.

Analisis sentimen adalah sebuah cara yang digunakan untuk mengolah komentar yang diberikan oleh pemesan atau pelanggan melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi [1]. Analisis sentimen, juga dikenal sebagai penambangan opini, berguna dalam manajemen bahasa alami, linguistik komputasi, dan penambangan teks atau biasa disebut *text mining*. Tujuan dari analisis sentimen adalah untuk menentukan perilaku dan pendapat penulis tentang topik tertentu. Tindakan tersebut dapat menunjukkan penilaian serta alasan dan kondisi untuk kecenderungan [2].

Text mining merupakan penerapan konsep dan teknik data mining untuk mencari pola dalam teks, yaitu proses menganalisis teks guna menyarikan informasi yang bermanfaat untuk tujuan tertentu [3]. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen atau mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen [4].

Metodologi pada penelitian ini akan menggunakan *Cross-Industry Standard Process for Data Mining* atau yang dikenal dengan CRISP-DM. CRISP-DM adalah metodologi yang menyediakan proses standar untuk *data mining* yang dapat diterapkan pada strategi pemecahan masalah umum dalam bisnis dan penelitian. Dengan setiap *fase* terstruktur dan terdefinisi dengan baik, CRISP-DM adalah metodologi penambangan data yang lengkap dan terdokumentasi dengan baik yang mudah digunakan bahkan untuk pemula [5].

Dalam penelitian ini menggunakan algoritma *Support Vector Machine* (SVM). *Support Vector Machine* (SVM), yang dikembangkan oleh Boser dan Guyon di Vapnik pada tahun 1992, merupakan teknik klasifikasi yang bertujuan untuk menemukan *hyperplane* dengan margin terbesar. *Hyperplane* yang baik diperoleh dengan memaksimalkan nilai *margin*. *Margin* adalah jarak titik antara *hyperplane* dan *support vector* yang paling dekat dengan *hyperplane* [6]. Keistimewaan dari *Support Vector Machine* berasal dari kemampuan untuk menerapkan pemisahan *linear* pada *input* data *non linear* berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi *kernel* yang diperlukan. *Support Vector Machine* (SVM) telah menjadi metode klasifikasi dan *regresi* yang populer untuk masalah *linear* dan *nonlinear*. Banyak peneliti telah melaporkan bahwa *Support Vector Machine* metode yang paling akurat untuk teks klasifikasi [7].

Dengan penjelasan pada pendahuluan mengenai penelitian ini, identifikasi masalah yang akan dibahas dalam penelitian ini adalah seberapa besar akurasi yang dapat diperoleh ketika menerapkan algoritma *support vector machine* dalam mengklasifikasikan sentimen *tweet* tentang *hepatitis* akut misterius pada data di *Twitter*.

Maka tujuan dari penelitian ini adalah melakukan analisa sentimen masyarakat berdasarkan opini *tweet* menggunakan algoritma *Support Vector Machine*, agar mengetahui seberapa besar akurasi yang akan dihasilkan dengan algoritma *Support Vector Machine* dalam menentukan sentimen dari sebuah *tweet* terkait penyakit *hepatitis* akut di *twitter*.

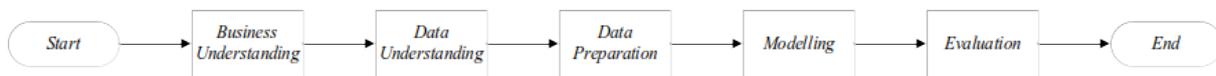
Pada penelitian ini mengacu pada beberapa referensi jurnal mengenai penelitian yang tidak jauh berbeda dengan masalah yang dibahas yaitu mengenai analisis sentimen berdasarkan data *twitter*. Berdasarkan penelitian yang telah dilakukan oleh (Imelya Susianti, Sri Soerya Ningsih, M. Al Haris, Tiani Wahyu Utami, 2020) dengan judul "Analisis sentimen pada *twitter* terkait *new normal* dengan metode *naïve bayes classifier*". Pada tahun 2020, dunia diguncang oleh merebaknya virus berbahaya, yaitu virus corona. Virus ini pertama kali berasal dari Wuhan, China dan telah menyebar ke negara lain, termasuk Indonesia. Mulai dari *social distancing* hingga *lockdown* hingga *new normal*, Indonesia mengambil berbagai inisiatif untuk mengantisipasi virus tersebut. Begitu banyak reaksi dan reaksi dari masyarakat terkait *new normal*. Terutama di media sosial, *Twitter*. Berdasarkan 3000 data *review* yang diperoleh dari aplikasi *Twitter*, 2.409 *review* masuk dalam kelas sentimen positif dan 591 *review* masuk dalam kelas sentimen negatif. Hasil penerapan metode *Naive Bayes classifier* untuk mengklasifikasikan data *review* pada *new normal* pada 80% perbandingan antara data *latih* dan data *uji*: 20% mencapai hasil klasifikasi sentimen dengan akurasi 87%. Tentunya penelitian ini tidak lepas dari kekurangan dan kelemahan. Berdasarkan visualisasi *wordcloud* beberapa kata yang paling banyak muncul diantaranya kata "new", "normal", "harga", "era", "hidup" [8].

Berdasarkan penelitian yang dilakukan oleh Teguh Ansyor Lorosae, Burhanudin Dwi Prakoso, Saifudin, Kusri (Seminar Nasional Teknologi Informasi dan Multimedia 2018) dengan judul "Analisis Sentimen Berdasarkan Opini Masyarakat Pada *Twitter* Menggunakan *Naive Bayes*". Menangkap opini publik mengenai kebutuhan dan keinginan pengguna jasa ekspedisi bukanlah hal yang mudah. Ini memiliki banyak aspek, salah satunya adalah jejaring sosial *Twitter*. Respon masyarakat terhadap layanan jasa pesan antar melalui media sosial *Twitter* sangat positif. Banyak opini publik dikomunikasikan melalui *tweet* media sosial. Dari gambaran publik

tentang layanan yang diberikan, layanan pengiriman tidak dapat sepenuhnya memahami berbagai layanan yang diberikan. Kita membutuhkan sebuah sistem yang dapat mengubah informasi opini publik menjadi kesimpulan yang emosional. Penelitian ini terdiri dari beberapa tahapan seperti identifikasi data input, preprocessing, klasifikasi Naive Bayesian, dan output. Data input berupa data training dan test yang digunakan diperoleh dari fan page akun twitter perusahaan jasa ekspedisi TIKI berupa opini masyarakat tentang layanan tersebut. Berdasarkan 11 analisis dan hasil pengujian yang dilakukan pada 100 data pengujian, metode Naive Bayes Classifier (NBC) dengan preprocessing menghasilkan akurasi rata-rata 84%. Dengan menggunakan Naive Bayes Classifier (NBC) dengan sistem ini, kita dapat menghasilkan klasifikasi mood positif dan negatif [9].

Berdasarkan penelitian yang telah dilakukan oleh (Rachman & Pramana, 2020) dengan judul “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter “. Wabah penyakit baru yang biasa disebut *coronavirus* (2019-nCoV), atau COVID-19, resmi ditetapkan sebagai pandemi global oleh Organisasi Kesehatan Dunia (WHO) pada 11 Maret 2020 (1). Pusat penyebaran virus Virus yang terletak di Wuhan, China pada akhir tahun 2019, kini telah menyebar ke seluruh dunia internasional, menyebabkan lebih dari 41,5 juta kasus dan lebih dari 1,1 juta kematian setiap tahunnya pada tahun 2020. Pada penelitian ini, nilai sentimen untuk setiap tweet diperoleh dari penghitungan jumlah kata positif & negatif yang terdapat pada suatu tweet. Secara rata-rata, diperoleh nilai sentimen untuk keseluruhan respon masyarakat terhadap vaksin COVID-19 yaitu sebesar 0,055. Hasil tersebut mengindikasikan masyarakat cenderung memberikan respon yang bersentimen positif dibandingkan dengan respon yang bersentimen negatif, meskipun nilai rata-rata yang dihasilkan sangat mendekati nilai 0 yang mengindikasikan banyaknya respon masyarakat yang tidak bersentimen (netral). Hasil tersebut sejalan dengan jumlah tweets berdasarkan jenis sentimen yang diperoleh. Pada Tabel 1 di atas terlihat bahwa masyarakat lebih banyak memberikan respon yang bersentimen positif dibandingkan dengan respon yang bersentimen negatif yaitu sebesar 29,6% banding 23,6%. Hal tersebut dapat diartikan sebagai masyarakat lebih banyak memberikan respon positif terhadap wacana vaksinasi yang dikeluarkan oleh pemerintah dibandingkan dengan respon negatifnya [10].

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Pada Gambar 1 merupakan alur dari tahapan penelitian, yaitu mulai dari pemahaman bisnis, pemahaman data, persiapan data, pemodelan, hingga evaluasi berdasarkan topik penelitian yang dibahas yaitu penyakit hepatitis akut misterius.

2.1. Business Understanding (Pemahaman Bisnis)

Pada tahapan ini peneliti akan mencoba untuk memahami permasalahan yang akan diangkat yaitu bagaimana sentimen masyarakat terhadap penyakit *hepatitis* akut yang dianggap misterius dan baru masuk ke Indonesia akhir-akhir ini, dengan mempelajari informasi di artikel atau berita di internet. Kemudian peneliti mencari informasi lebih di media sosial *twitter* yaitu *tweet* masyarakat bagaimana menanggapi terkait masalah penyakit *hepatitis* misterius tersebut dengan menggunakan *keyword* yang sudah ditentukan.

2.2. Data Understanding (Pemahaman Data)

Pada tahapan ini peneliti akan mencoba memahami proses pengambilan data tweet di twitter yang sesuai dengan topik yang dibahas dengan menggunakan keyword seperti hepatitis misterius dengan menggunakan software rapid miner. Sebelum mengambil data tweet, peneliti di haruskan membuat koneksi agar terhubung dengan API twitter agar proses pengambilan data dapat di lakukan. Dengan jumlah data tweet yang di ambil sejumlah 4495, mulai dari tanggal 24 Mei 2022 hingga 30 Mei 2022. Kemudian dari hasil pengambilan data, di lakukan proses remove duplicate agar menghilangkan data tweet yang sama dengan menggunakan software rapid miner. Setelah proses remove duplicate sudah di lakukan lalu mendapatkan hasil 1000 data. Hasil pemrosesan tersebut di gunakan untuk ke tahapan selanjutnya yaitu preprocessing data.

2.3. Data Preparation (Persiapan Data)

Pada tahapan ini peneliti akan melakukan preprocessing data. Sebelum melakukan preprocessing data, hasil data tweet yang sudah di remove duplicate, di lakukan pemberian label sentimen dengan menggunakan azure machine yang terdapat di software Microsoft excel. Hasil data dari pemberian label sentimen tersebut baru di

lakukan preprocessing data. Selanjutnya data yang sudah melewati tahapan preprocessing akan melakukan pemodelan berdasarkan algoritma yang sudah ditentukan.

2.4. Modelling (Pemodelan)

Pada tahapan ini peneliti akan melakukan pencarian model terhadap dataset tweet yang sudah di preprocessing. Dataset tersebut akan dibagi menjadi dua dataset, yaitu dataset training dan dataset testing yang dimana nantinya akan melalui dua tahapan pada penelitian ini yaitu split data dan cross fold validation. Pada tahapan split data, dataset akan dibagi dengan perbandingan data training dan data testing yaitu 60:40, 70:30, dan 80:20. Kemudian pada tahapan cross fold validation peneliti akan membagi secara acak ke dalam 10 bagian (number of fold=10). Pada tahapan pemodelan ini, peneliti akan menggunakan algoritma *Support Vector Machine* (SVM) yang pada sebelumnya sudah dilakukan perbandingan model klasifikasi dengan menggunakan algoritma *Naïve Bayes* (NB) dan *K-Nearest Neighbors* (KNN).

Karena pada penelitian ini menggunakan algoritma *support vector machine*, berikut merupakan tahapan dalam perhitungannya.

- Menentukan kata yang sering muncul dari tiap dokumen atau tweet yang digunakan.
- Menentukan inialisasi awal untuk nilai $\alpha=0.5, C=1, \lambda=0.5, \text{gamma}=0.5$ dan $\text{epsilon}=0.001$.
- Menghitung matriks dengan rumus (1):

$$D_{ij} = y_i y_j (K(x_i \times x_j) + \lambda^2) \quad (1)$$

Keterangan :

D_{ij} : Elemen matriks data ke-ij

y_i : Kelas atau label data ke-i

y_j : Kelas atau label data ke-j

$(K(x_i \times x_j))$: Fungsi Kernel

λ : Turunan batas teoritis

- Pada data ke-n = 123, ... n menggunakan persamaan (2), (3), (4) seperti berikut ini.

$$E_i = \sum_{j=1}^n \alpha_j D_{ij} \quad (2)$$

$$\delta \alpha_i = \min \left\{ \max \left[\gamma (1 - E_i), -\alpha_i \right], C - \alpha_i \right\} \quad (3)$$

$$\alpha_i = \alpha_i + \delta \alpha_i \quad (4)$$

Keterangan :

E_i : Nilai error data ke-i

γ : Tingkat pembelajaran

$\max(i) D_{ij}$: Nilai maksimum diagonal matriks hessian

- Mencari nilai bias (b) dengan menggunakan persamaan (5) sebagai berikut.

$$b = -\frac{1}{2} w \cdot x^+ + w \cdot x^- \quad (5)$$

- Pengujian pada dokumen atau data yang diuji

- Perhitungan keputusan

Jika hasil perhitungan keputusan lebih dari sama dengan 0 maka nilai sign $h(x)$ adalah +1, maka termasuk kelas positif dan jika hasil perhitungan keputusan kurang dari 0 maka nilai sign $h(x)$ nilai -1, maka termasuk kelas negatif. Perhitungan keputusan dengan menggunakan persamaan (6).

$$h(x) = w \cdot x + b \quad (6)$$

atau

$$h(x) = \sum_{i=1}^m (\alpha_i - y_i) K(x_i, x) + b \quad (7)$$

2.5. Evaluation (Evaluasi)

Pada tahapan ini peneliti akan melakukan evaluasi terhadap model yang sudah dibuat dengan mengukur performa terbaik atau paling tinggi menggunakan confusion matrix terhadap algoritma Support Vector Machine (SVM).

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Dalam menganalisa dan membahas usaha ini, peneliti berusaha untuk memahami dan mempelajari permasalahan yang dibahas selama penelitian ini. Pertanyaan yang diangkat dan diperdebatkan adalah seberapa akurat analisis sentimen dari data Twitter dalam bentuk tweet yang membahas kemunculan hepatitis akut misterius baru-baru ini dengan menggunakan algoritma mesin vektor pendukung. Karena penelitian ini adalah analisis sentimen berbasis opini publik, peneliti mengumpulkan data dari Twitter, media sosial yang populer dan banyak digunakan, untuk menentukan seberapa relevan opini pengguna Twitter dengan masalah yang sedang dibahas. Dengan kata lain, hepatitis akut yang misterius di Indonesia.

3.2. Data Understanding

Dalam menganalisis dan membahas pemahaman data, peneliti menggunakan sebuah software atau alat bernama Rapid Miner untuk mengumpulkan data yang mereka dapatkan dari salah satu media sosial yaitu Twitter. Berdasarkan masalah yang akan dibahas, peneliti mengidentifikasi beberapa kata kunci untuk pengumpulan data yang terkait dengan masalah dan memahami data yang akan dikumpulkan dan digunakan.

Peneliti mengumpulkan data di Twitter selama tujuh hari. Dari 24 Mei 2022 hingga 30 Mei 2022, dua kata kunci akan digunakan untuk dua perayapan atau dua pengambilan. Berikut adalah beberapa kata kunci yang diidentifikasi berdasarkan masalah yang dibahas.

Tabel 1. Keyword Crawling Data

TANGGAL	KEYWORD
24 Mei 2022 – 30 Mei 2022	<i>Hepatitis Akut</i> <i>Hepatitis Misterius</i>

Pada Tabel 1 merupakan *keyword* dari *crawling* data yang dilakukan. *Crawling* data dilakukan dengan mengambil data *tweet* dari tanggal 24 Mei 2022 sampai 30 Mei 2022 dengan kata kunci *hepatitis* akut dan *hepatitis* misterius.

3.3. Data Preparation

3.3.1. Pemberian Label Sentimen

Pada tahapan pemberian ini maka data atau tweet harus diberikan label sentimen terlebih dahulu sebelum memulai tahapan selanjutnya yaitu preprocessing data. Salah satu software atau tools yang dapat digunakan untuk melakukan pemberian label dalam analisis sentimen adalah Microsoft Excel dengan get ads-in Azure Machine. Azure machine di excel ini menggunakan MPQA Subjectivity Lexicon. Pemberian label tersebut adalah berupa sentimen positif atau sentimen negatif dari setiap tweet yang ada pada data. Sentimen yang digunakan pada penelitian ini hanya sentimen positif dan sentimen negatif. Berikut ini adalah contoh hasil dari pemberian label sentimen menggunakan azure machine, seperti pada Tabel 2.

Tabel 2. Hasil Pemberian Label Sentimen

TEXT	SENTIMEN	SCORE
yuk kenali dan waspadai mengenai penyakit #HepatitisMisterius agar bisa menvega penularan sejak dini. https://t.co/yz3mosYeKD	<i>Positive</i>	0.63372129201889
Hoaks penyebaran hepatitis akut	<i>Negative</i>	0.448214620351791
G20 Indonesia jaya https://t.co/KXtOWi6Bta		

Pada Tabel 2 merupakan hasil pemberian label sentimen dengan menggunakan Azure Machine, menghasilkan sentimen positive dan sentimen negative dengan masing-masing score yang berbeda.

3.3.2. Preprocessing Data

Pada tahapan preprocessing data akan melalui beberapa alur proses. Alur proses pada tahapan preprocessing data terdiri dari yaitu proses cleaning, proses casefolding, proses tokenizing, proses filtering stopword, dan proses stemming. Dari dataset text yang sudah dicrawling dari twitter, kemudian akan melewati proses preprocessing data. Berikut adalah hasil dari preprocessing data dari awal sampai akhir.

Tabel 3. Hasil Preprocessing Data

Sebelum Preprocessing	Hasil Preprocessing
Kemenkes mencatat tambahan kasus baru diduga hepatitis akut misterius Total menjadi kasus orang meninggal dua di antaranya probable	kemenkes catat tambah duga hepatitis akut misterius total orang tinggal probable
Kasus Dugaan Hepatitis Akut Misterius di RI Bertambah Jadi	duga hepatitis akut misterius ri tambah
Pentingnya Edukasi Masyarakat Cegah Hepatitis Akut Misterius Jangan Sampai Anak-Anak Terpapar	edukasi masyarakat cegah hepatitis akut misterius anakanak papar

Pada Tabel 3 merupakan hasil dari *preprocessing* data yang terdiri dari proses *cleaning*, proses *casefolding*, proses *tokenizing*, proses *filtering stopword*, hingga proses *stemming*.

3.3.3. Pembagian Data Training dan Data Testing

Setelah tahapan *preprocessing* data selesai, maka hasil dari data *preprocessing* tersebut akan dibagi menjadi data *training* dan data *testing*. Data *training* merupakan data latih yang akan digunakan untuk proses latihan pada model klasifikasi. Sedangkan data *testing* merupakan data yang akan digunakan untuk proses uji *rule* klasifikasi. Pada penelitian ini untuk data *training* dan data *testing* akan dibagi menjadi beberapa perbandingan, yaitu perbandingan 60:40, 70:30, dan 80:20. Untuk metode pembagiannya ke dalam perbandingan tersebut, menggunakan metode *stratified sampling*. Metode tersebut merupakan metode pemilihan sampel yang dilakukan secara acak dan terstruktur pada masing-masing kelompok.

Tabel 4. Perbandingan Algoritma Data Training dan Data Testing

ALGORTIMA	PERBANDINGAN
<i>Naïve Bayes, Support Vector Machine, K-Nearest Neighbors</i>	60:40
<i>Naïve Bayes, Support Vector Machine, K-Nearest Neighbors</i>	70:30
<i>Naïve Bayes, Support Vector Machine, K-Nearest Neighbors</i>	80:20

Pada Tabel 4 merupakan perbandingan algoritma dari data training dan data testing. Perbandingan algoritma terdiri dari algoritma *Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbors*. Untuk perbandingan data training dan data testing yaitu perbandingan 60:40, 70:30, dan 80:20.

3.4. Modelling

3.4.1. Modelling Menggunakan Split Data

Pada *modelling* menggunakan *split* data, data akan dibagi menjadi 3 model perbandingan yang terdiri dari 60:40, 70:30, dan 80:20 dengan menggunakan algoritma *Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbors*. Perbandingan data tersebut terdiri dari masing-masing data *training* dan data *testing*.

3.4.2. Modelling Menggunakan Cross Validation

Pada *modelling* menggunakan *cross validation* maka data *training* dan data *testing* akan di *split* secara acak menjadi 10 bagian (*number of folds* = 10). Sepuluh bagian tersebut merupakan dataset yang sudah diacak, kemudian hasil dari pengacakan sebanyak 10 bagian akan diambil dengan hasil terbesar.

3.5. Hasil Pengujian Modelling

Setelah melakukan *modelling* dengan 2 jenis yaitu *modelling* menggunakan *split* data dan *modelling* menggunakan *cross validation*, berikut adalah hasil dari kedua *modelling* tersebut.

3.5.1. Hasil Pengujian Model Menggunakan *Split Data*

Hasil pengujian *modelling* menggunakan *split data* dengan menggunakan perbandingan dari data *training* dan data *testing* dapat dilihat pada tabel berikut.

Tabel 5. Perbandingan Hasil *Accuracy Split Data*

PERBANDINGAN	ALGORITMA		
	<i>Naïve Bayes</i>	<i>Support Vector Machine</i>	<i>K-Nearest Neighbors</i>
60:40	76.33%	79.00%	78.88%
70:30	77.50%	79.28%	78.15%
80:20	78.64%	81.06%	79.61%

3.5.2. Hasil Pengujian Model Menggunakan *Cross Validation*

Hasil akurasi dari pengujian *modelling* menggunakan *cross validation* terhadap *dataset* yang sudah di *preprocessing* dapat dilihat pada tabel berikut.

Tabel 6. Hasil *Accuracy Cross Validation*

ALGORITMA	ACCURACY
<i>Naïve Bayes</i>	77.60%
<i>Support Vector Machine</i>	79.43%
<i>K-Nearest Neighbors</i>	77.91%

Berdasarkan pada tabel 5 dan tabel 6 yang berisi hasil performa akurasi model klasifikasi pada algoritma *Support Vector Machine* dengan 2 cara *modelling* yaitu *modelling split data* dan *modelling cross validation*, maka peneliti menyimpulkan bahwa pemodelan dengan *split data* perbandingan 80:20 menghasilkan akurasi sebesar 81.06% sedangkan untuk pemodelan *cross validation* menghasilkan akurasi sebesar 79.43%, sehingga akurasi tertinggi yang dihasilkan dalam penelitian ini adalah dengan pemodelan *split data* perbandingan 80:20 sebesar 81.06%.

3.6. Evaluation

Penelitian ini akan melakukan pengujian sebagai evaluasi dengan menggunakan *confussion matrix* yang sudah diperoleh dari tahapan *modelling* dengan menggunakan algoritma *Support Vector Machine*, sehingga menghasilkan akurasi terbesar dengan menggunakan pemodelan *split data* pada perbandingan 80:20. Sehingga *confussion matrix* yang ditampilkan pada tabel 4.14 adalah hasil dari evaluasi pengukuran klasifikasi menggunakan algoritma *support vector machine* dengan *dataset* yang berjumlah 2058 terdapat 1646 data *training* dan 412 data *testing*.

Tabel 7. Hasil *Accuracy Cross Validation*

	True Positive	True Negative	Class Precision
Pred. Positive	312	69	81.88%
Pre. Negative	9	22	70.96%
Class Recall	97.19%	24.17%	

Perhitungan *confussion matrix* dari tabel 7 adalah sebagai berikut :

- $Precision = \frac{TP}{TP+FP} = \frac{312}{312+69} = 0.8188 = 81.88\%$
- $Recall = \frac{TP}{TP+FN} = \frac{312}{312+9} = 0.9719 = 97.19\%$
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{22+312}{22+312+69+9} = 0.8106 = 81.06\%$

Berdasarkan hasil perhitungan pada tabel 4.14 dapat dilihat bahwa hasil dari pengolahan data sentimen terkait *hepatitis* akut misterius dengan menggunakan algoritma *support vector machine* menghasilkan *precision* sebesar 81.88%, *recall* sebesar 97.19%, dan *accuracy* sebesar 81.06%.

4. KESIMPULAN

Berdasarkan hasil yang didapat untuk pengujian dan analisis pada penelitian ini, maka dapat disimpulkan terkait penelitian ini mengenai analisis sentimen terhadap penyakit *hepatitis* akut misterius berdasarkan data dari sosial media *twitter* yaitu dengan jumlah *dataset* yang digunakan sebanyak 2058 data dengan sentimen positif sejumlah 1590 data dan sentimen negatif sejumlah 468 data yang pengujian modelnya menggunakan algoritma *Support Vector Machine* dengan perbandingan 80:20 pada *modelling split* data, maka nilai hasil *accuracy* yang didapatkan sebesar 81.06%, selanjutnya dengan jumlah sentimen positif lebih banyak dibanding jumlah sentimen negatif, maka kebanyakan opini masyarakat di media sosial *twitter* terkait penyakit ini merupakan respon yang baik dan positif untuk pihak terkait yaitu pihak pemerintah dan pihak kesehatan, sehingga pihak-pihak tersebut bisa mengambil keputusan untuk kedepannya dalam menangani penyakit *hepatitis* akut misterius yang dibahas pada penelitian ini.

DAFTAR PUSTAKA

- [1] E. M. Sipayung, H. Maharani, And I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *J. Sist. Inf.*, Vol. 8, No. 1, Pp. 2355–4614, 2016, [Online]. Available: [Http://Ejournal.Unsri.Ac.Id/Index.Php/Jsi/Index](http://Ejournal.Unsri.Ac.Id/Index.Php/Jsi/Index).
- [2] W. A. Luqyana, *Instagram Dengan Metode Klasifikasi Support Vector Machine*. 2018.
- [3] A. N. Roifa, "Text Mining Dengan Metode Naive Bayes Classifier Untuk Mengklasifikasikan Berita Berdasarkan Konten," 2018, [Online]. Available: [Https://Repository.Its.Ac.Id/51007/](https://Repository.Its.Ac.Id/51007/).
- [4] D. Wijaya, "Aplikasi Text Mining Untuk Pencarian Dokumen Tugas Akhir Berdasarkan Semantic Relatedness Menggunakan Machine Readable Dictionary," Pp. 5–14, 2017.
- [5] V. R. Hananto, "Analisis Penentuan Metode Data Mining Untuk," Pp. 1–11, 2017.
- [6] I. M. Yulietha, S. Al Faraby, F. Teknik, And U. Telkom, "Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine Sentiment Classification Of Movie Reviews," Vol. 4, No. 3, Pp. 4740–4750, 2017.
- [7] E. Indrayuni, "Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization," *J. Evolusi* Vol. 4 Nomor 2 - 2016, Vol. 4, No. 2, Pp. 20–27, 2016.
- [8] T. W. U. Imelya Susianti, Sri Soerya Ningsih, M. Al Haris, "Analisis Sentimen Pada Twitter Terkait New Normal Dengan Metode Naive Bayes Classifier," *Pros. Semin. Edusainstech Fmipa Unimus*, Pp. 354–363, 2020, [Online]. Available: [Https://Prosiding.Unimus.Ac.Id/Index.Php/Edusaintek/Article/View/576/578](https://Prosiding.Unimus.Ac.Id/Index.Php/Edusaintek/Article/View/576/578).
- [9] T. A. Lorosae And B. D. Prakoso, "Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan Naive Bayes," *Semin. Nas. Teknol. Inf. Dan Multimed.* 2018 Univ. Amikom Yogyakarta, 10 Februari 2018, Pp. 25–30, 2018.
- [10] F. F. Rachman And S. Pramana, "Analisis Sentimen Pro Dan Kontra Masyarakat Indonesia Tentang Vaksin Covid-19 Pada Media Sosial Twitter," *Heal. Inf. Manag. J.*, Vol. 8, No. 2, Pp. 100–109, 2020, [Online]. Available: [Https://Inohim.Esaunggul.Ac.Id/Index.Php/Ino/Article/View/223/175](https://Inohim.Esaunggul.Ac.Id/Index.Php/Ino/Article/View/223/175).