

## PERBANDINGAN KINERJA ALGORITMA SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR TERHADAP ANALISIS SENTIMEN KEBIJAKAN NEW NORMAL

Didin Muhidin<sup>1</sup>, Arief Wibowo<sup>2</sup>  
Magister Ilmu Komputer, Universitas Budi Luhur  
dmuhidin49@gmail.com<sup>1</sup>, arief.wibowo@budiluhur.ac.id<sup>2</sup>

*Submitted July 19, 2020; Revised November 17, 2020; Accepted November 18, 2020*

### Abstrak

Twitter merupakan salah satu situs *microblogging* populer di kalangan pengguna internet sehingga banyak masyarakat menggunakan twitter untuk menyampaikan sentimen positif dan negatifnya terhadap kebijakan new normal. Masa pandemi memunculkan banyak sentimen masyarakat terhadap kebijakan adaptasi kebiasaan baru atau *new normal*. Penelitian ini bertujuan untuk mengklasifikasi twit sentimen ke dalam kelas positif dan negatif. Algoritma klasifikasi yang digunakan adalah k-NN dan SVM. Hasil pengujian menunjukkan bahwa algoritma k-NN lebih baik dibandingkan SVM dalam menyelesaikan kasus sentimen ini dengan menghasilkan akurasi sebesar 72,96%.

**Kata Kunci :** Teks Mining, Analisis Sentimen, SVM, k-NN, New Normal.

### Abstract

*Twitter is one of the popular microblogging sites among internet users, so that many people use Twitter to convey their positive and negative sentiments towards the new normal policy. The pandemic period raises much public sentiment towards the policy of adapting to the new normal. This study aims to classify sentiment tweets into positive and negative classes. The classification algorithms used are k-NN and SVM. The test results show that the k-NN algorithm is better than SVM in solving this sentiment case with an accuracy of 72.96%.*

*Keywords: Mining Text, Sentiment Analysis, SVM, k-NN, New Normal.*

### 1. PENDAHULUAN

Perkembangan teknologi internet di Indonesia mencapai telah mencapai 63 juta orang di tahun 2013, meningkat pada tahun 2014 menjadi 83,7 juta, sedangkan pengguna twitter di Indonesia pada tahun 2014 berjumlah 19,5 juta pengguna (kominfo.go.id). Perkembangan pesat ini menjadikan twitter sebagai platform media sosial yang dominan. Twitter merupakan salah satu situs *microblogging* yang telah menjadi alat komunikasi yang populer di kalangan pengguna internet, sehingga akan terdapat banyak data yang dapat dimanfaatkan. Salah satu manfaat yang bisa dianalisis dari penggunaan twitter adalah sentimen pengguna twitter terhadap kebijakan-kebijakan yang di keluarkan oleh pemerintah karena pengguna internet

akan menuliskan pendapatnya di pesan twitter.

Pada tanggal 2 Maret 2020, di Indonesia pertama kali diumumkan kasus korban virus covid 19 dengan jumlah 1.155 kasus (www.kompas.com). Pada akhir Maret 2020, pemerintah pusat menghimbau kepada pemerintah daerah untuk memberlakukan *lockdown* di daerah masing-masing sehingga banyak perusahaan, sekolah dan tempat umum lainnya di tutup untuk menghindari penyebaran covid-19. Efek dari pemberlakuan *lockdown* ini adalah banyaknya aktivitas masyarakat yang terganggu seperti sektor perdagangan, pekerjaan, pendidikan bahkan sampai dengan faktor ekonomi.

Pada 1 Juni 2020, pemerintah memberlakukan *new normal* untuk empat

provinsi di Indonesia. Di satu sisi masyarakat memang perlu untuk beraktifitas normal kembali namun di sisi lain korban covid-19 terus bertambah.

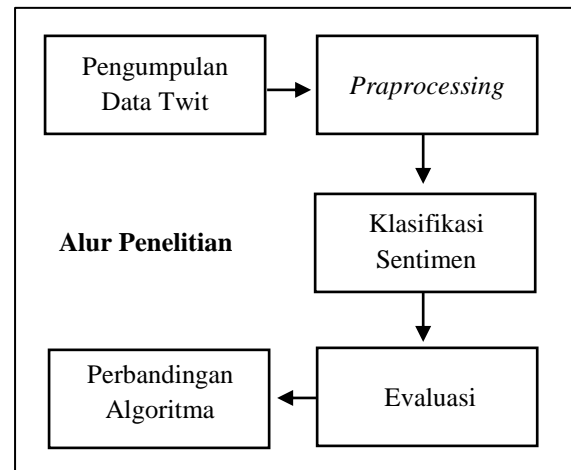
Berdasarkan uraian di atas, penulis bermaksud menganalisis sentimen masyarakat terhadap kebijakan *new normal* yang ditetapkan oleh pemerintah dengan cara menganalisis data yang di dapat dari twitter. Analisis sentimen merupakan metode untuk proses ekstraksi data berupa opini atau sentimen, yang prosesnya memaknai data tekstual untuk mengetahui sentimen yang termuat [1]. Dengan demikian maka analisis sentimen terdiri dari rangkaian metode untuk menentukan polaritas dari kumpulan teks dan membedakannya ke dalam nilai positif, negatif, atau netral [2].

Adapun tahapan yang akan dilakukan penulis yaitu dengan mengambil data secara langsung dari twitter dengan keyword pencarian *New Normal* dan #NewNormal. Selanjutnya, data yang sudah terkumpul akan di olah menggunakan *Text Preprocessing* dengan tujuan agar data yang didapat akan lebih terstruktur dan lebih mudah dilakukan pengolahan data.

Tahapan selanjutnya yaitu Klasifikasi sentimen menggunakan algoritma *k-nearest neighbors* (KNN) dan algoritma *support vector machine* (SVM) untuk menentukan sentimen positif atau negatif. Kemudian yang selanjutnya adalah evaluasi menggunakan *k-fold cross validation* untuk mengukur validasi dalam pengujian metode yang digunakan. Dan tahapan yang terakhir adalah membandingkan akurasi algoritma SVM dengan K-NN.

## 2. METODE PENELITIAN

Alur kerja penelitian ini adalah sebagai berikut:



Gambar 1. Alur Kerja Penelitian

Untuk penjelasan alur kerja penelitian secara keseluruhan dapat dijelaskan sebagai berikut.

### Pengumpulan Data

Data dikumpulkan merupakan data twit yang diambil dari twitter dengan kata kunci “New Normal” dan #NewNormal. Data yang di ambil adalah data dari tanggal 01 Juni 2020 sampai 03 Juni 2020 dengan jumlah data yang terkumpul sebanyak 222 data twit.

### Preprocessing Data

*Preprocessing* pada Data Mining merupakan proses mengubah data agar sesuai dengan format yang dibutuhkan pada proses berikutnya. Tahap ini dilakukan dengan cara menganalisis dan melakukan penggalian informasi, hingga pengolahan dan pengatuarn informasi untuk mengetahui ada tidaknya bentuk tekstual yang terstruktur maupun tidak terstruktur [2].

Data yang terkumpul dari twitter seringkali kurang berkualitas, seperti terdapat singkatan-singkatan, kata kurang baku, salah input dan tidak konsisten sehingga perlu dilakukan *preprocessing* data terlebih dahulu.

Data teks akan dibersihkan menggunakan teks *preprocessing* yang disediakan oleh data *framework* yang terdiri dari

*transformation remove url, tokenization, Indonesian stemming, transformation not (negative) dan stopword removal.*

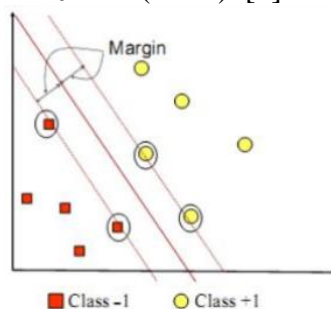
### Pemrosesan Teks

Setelah data dikumpulkan dan dilakukan *preprocessing data*, selanjutnya data akan di klasifikasi dengan algoritma *Support Vector Machine (SVM)* dan algoritma *K-Nearest Neighbor (k-NN)*.

Algoritma *K-Nearest Neighbor (KNN)* bekerja dengan cara mengklasifikasi suatu objek, atas dasar nilai  $k$  pada data latih yang jaraknya paling dekat dengan obyek yang dianalisis. Syarat nilai  $k$  adalah harus lebih kecil dari jumlah data latih, selain itu nilai  $k$  bersifat ganjil dan lebih dari satu. Kedekatan jarak data dengan data latih selanjutnya akan diklasifikasi untuk dihitung menggunakan metode *cosine similiarity* [7].

Metode *Cosine similiarity* digunakan untuk melihat nilai kemiripan isi antar dokumen. *Cosine similiarity* bekerja dengan menghitung ukuran interpretasi kedekatan jarak berdasarkan kemiripan suatu dokumen [7].

*Support Vector Machine (SVM)* pertama kali diperkenalkan sebagai rangkaian harmonis konsep- konsep unggulan dalam bidang *pattern recognition* Feldman oleh Vapnik. SVM merupakan algoritma *machine learning* yang bertujuan untuk menemukan *hyperplane* terbaik dalam memisahkan dua buah *class* pada *input space* yang bekerja atas prinsip *Structural Risk Minimization (SRM)* [8].



**Gambar 2. Hyperlane Terbaik Yang Memisahkan Kedua Kelas**

### Evaluasi

Evaluasi yang dilakukan untuk mengukur tingkat akurasi pada penelitian ini menggunakan *k fold cross validation*.

### Perbandingan Algoritma

Setelah tahap evaluasi, tahap selanjutnya adalah membandingkan nilai akurasi antara algoritma  $k$ -NN dengan SVM dengan tujuan untuk memilih akurasi algoritma yang lebih baik pada penyelesaian kasus ini.

## 3. HASIL DAN PEMBAHASAN

### *Knowledge Discovery Goals*

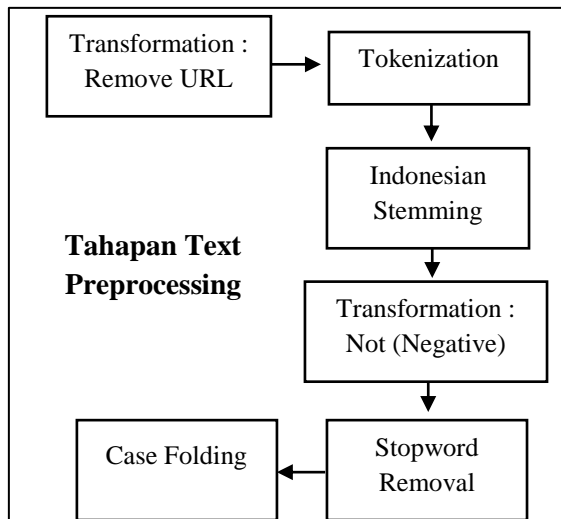
Penelitian ini bertujuan untuk mengklasifikasi tweet tentang *new normal* kedalam kelas yang sudah ditentukan, yaitu positif dan negatif. Proses klasifikasi ini dilakukan dengan menggunakan algoritma  $k$ -NN dan SVM.

### *Data Integration*

Data mentah yang akan dijadikan data uji adalah data tweet masyarakat terhadap kebijakan *new normal* yang sudah ditetapkan oleh pemerintah. Data uji yang di dapat terdiri dari atribut teks dan kelas dengan jumlah data sebanyak 222 data ulasan yang akan diberikan label secara manual. Seluruh data yang sudah diberikan label akan disimpan dalam bentuk file microsoft excel.

### *Data Pre-processing*

*Pre-proceesing* pada *text mining* adalah rangkain proses mengubah data sesuai dengan format yang dibutuhkan. Data dibersihkan menggunakan metode Gata Framework (Windu Gata, 2018) dan fitur *text preprocessing* yang ada pada RapidMiner. Tahapan data *preprocessing* yang dilakukan dapat dilihat pada gambar berikut:



**Gambar 3. Tahapan Text Preprocessing**

1. *Transformation Remove URL*, yaitu proses yang digunakan untuk membersihkan data twit dari *link-link* yang terdapat dalam data tersebut.

**Tabel 1. Perbandingan Proses Transformation URL**

Proses	Teks
Sebelum dilakukan <i>transformation URL</i>	Halo teman2 berdomisili di DKI.  Sbtr lg PSBB berakhir tp blm ada info berlanjut atau tidak. Namun apa kita memahami resiko COVID19? Apa kita siap masuki "New Normal"??  Yuk isi survey @laporcovid agar kita pny gambaran apa yg ada di pikiran pend DKI?  <a href="http://s.id/resikocovid">http://s.id/resikocovid</a>
Sesudah dilakukan <i>case folding</i>	halo teman2 berdomisili di dki. sbtr lg psbb berakhir tp blm ada info berlanjut atau tidak. namun apa kita memahami resiko covid19? apa kita siap masuki new normal? yuk isi survey @laporcovid agar kita pny gambaran apa yg ada di pikiran pend dki?

2. *Tokenization*, merupakan pemotongan kalimat menjadi kata. Metode ini akan memotong kata berdasarkan spasi serta dan kemudian mengeliminasi karakter simbol-simbol atau tanda baca. Tahap *tokenizing* membagi urutan karakter

menjadi kalimat, selanjutnya kalimat diubah menjadi *token* [4]. Proses ini bertujuan untuk memecah kalimat menjadi satuan kata. Dan karakter selain teks akan dihilangkan.

**Tabel 2. Perbandingan Proses Tokenization**

Proses	Teks
Sebelum dilakukan <i>tokenization</i>	Sultan: Tak perlu buru-buru terapkan new normal #Yogyakarta #COVID19indonesia
Sesudah dilakukan <i>tokenization</i>	Sultan Tak perlu buru buru terapkan new normal Yogyakarta COVID indonesia

3. *Indonesian stemming*, yaitu mengembalikan kata pada kata dasarnya atau menghilangkan imbuhan-imbuhan. *Stemming* merupakan tahapan pada *text preprocessing* yang bertujuan untuk mengembalikan suatu kata bentuk akar katanya atau menghilangkan imbuhan (awalan dan akhiran) yang terdapat pada kata tersebut [6].

**Tabel 3. Perbandingan Proses Indonesian Stemming**

Proses	Teks
Sebelum dilakukan <i>Indonesian stemming</i>	New normal new normal. Negara lain udah beradaptasi itu kapan mau maju sih ini !!
Sesudah dilakukan <i>Indonesian stemming</i>	New normal new normal. Negara lain udah adaptasi itu kapan mau maju sih ini !!

4. *Transform not (negative)*, yaitu tahapan yang digunakan untuk menghilangkan kata yang ambigu seperti tidak mengecewakan, tidak lama yang mengandung arti positif. Transformasi data merupakan tahap mengubah data dari suatu skala ke skala yang lain [2].

*Transformation* juga dilakukan untuk menghilangkan kata yang abigu [5].

**Tabel 4. Perbandingan Proses Transform Not (Negative)**

Proses	Teks
Sebelum dilakukan <i>Transform not (negative)</i>	New normal New normal kencet, pola pikir kau yak belum normal hmmm
Sesudah dilakukan <i>Transform not (negative)</i>	New normal New normal kencet, pola pikir kau yak belum_normal hmmm

5. *Stopword removal*, merupakan tahap penghapusan kata-kata yang terkandung pada daftar kata *stopword*. Kata-kata *stopwords* dapat menambah dimensi data pada proses klasifikasi [6]. *Stopword removal* juga digunakan untuk menghilangkan kata yang tidak relevan atau yang tidak bermakna [5].

**Tabel 5. Perbandingan Proses Stopword Removal**

Proses	Teks
Sebelum dilakukan <i>stopword removal</i>	Banyak yang bingung apa itu New Normal. Atiqah Hasiholan juga bingung. semoga kamu gak bingung. bingung New Normal.
Sesudah dilakukan <i>stopword removal</i>	Atiqah Hasiholan bingung. Semoga tidak_bingung.

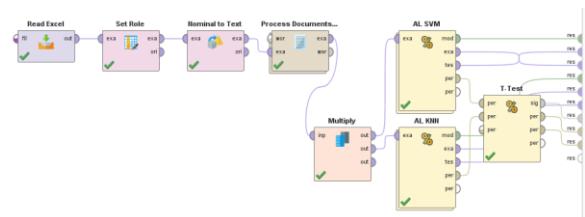
6. *Case folding*, merupakan proses yang dilakukan untuk menyeragamkan karakter pada data di dalam *text preprocessing*. Proses *case folding* adalah proses mengubah seluruh huruf pada data menjadi huruf kecil [4]. Pada proses ini akan mengubah semua huruf yang ada pada data twit menjadi huruf kecil (*non capital*).

**Tabel 6. Perbandingan Proses Case Folding**

Proses	Teks
Sebelum dilakukan <i>Case folding</i>	New Normal = Bunuh Diri Massal New Normal = Seleksi Alam New Normal = Nyobain mati Tau lah tau, serah dah serah.
Sesudah dilakukan <i>case folding</i>	new normal = bunuh diri massal new normal = seleksi alam new normal = nyobain mati tau lah tau, serah dah serah.

### Data Mining

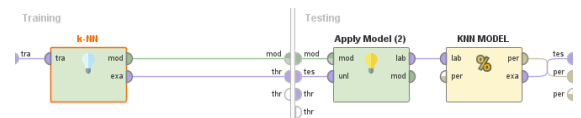
Pada tahapan ini, algoritma yang digunakan yaitu algoritma k-NN dan SVM yang diimplementasikan pada tool Rapidminer 9.7 dengan desain yang diajukan sebagai berikut:



**Gambar 4. Desain Model k-NN dan SVM**

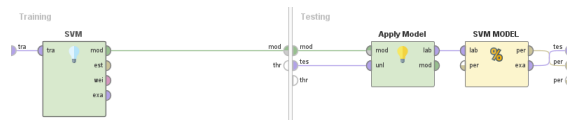
### Evaluasi

Setelah model *data mining* didapatkan, proses selanjutnya adalah membandingkan nilai akurasi antara algoritma k-NN dengan SVM menggunakan *10-fold cross validation*. Desain model pengujian algoritma k-NN pada data twit yang digunakan dapat dilihat pada gambar 5.



**Gambar 5. Desain Model k-NN**

Desain model pengujian algoritma SVM pada data twit yang digunakan dapat dilihat pada gambar 6.



**Gambar 6. Desain Model SVM**

Gambar 5 dan 6 adalah penjelasan detail dari gambar 4. Data yang digunakan untuk pengujian validasi terdiri dari tweet positif sebanyak 99 data dan tweet negatif sebanyak 123 data. Jumlah  $k$  yang digunakan pada model  $k$ -NN adalah 5.

Tahap selanjutnya adalah menghitung akurasi dari *confusion matrix*. Hasil perhitungan model algoritma SVM dapat dilihat pada tabel 7.

**Tabel 7. Confused Matrix Algoritma SVM**

Accuracy: 71.19% +/- 5.89%

	true Negatif	true Positif	class precision
pred. Negatif	122	63	65.95%
pred. Positif	1	36	97.30%
class recall	99.19%	36.36%	

Akurasi yang diperoleh sebesar 72,96%. Data kelas negatif yang sesuai dengan prediksi adalah 88 data. Data negatif yang diprediksi ke dalam kelas positif ada sebanyak 25 data. Data positif yang diprediksi secara benar ada sebanyak 74 data dan data positif diprediksi menjadi kelas negatif ada sebanyak 35 data.

Terakhir, hasil perhitungan model  $k$ -NN dapat dilihat pada tabel 8.

**Tabel 8. Confused Matrix Algoritma  $k$ -NN**

Accuracy: 72.96% +/- 11.68%

	true Negatif	true Positif	class precision
pred. Negatif	88	25	77.88%
pred. Positif	35	74	67.89%
class recall	71.54%	74.75%	

Akurasi yang diperoleh sebesar 71,19%. Data kelas negatif yang sesuai dengan prediksi adalah 122 data. Data negatif yang diprediksi ke dalam kelas positif ada sebanyak 63 data. Data positif yang diprediksi secara benar ada sebanyak 36 data dan data positif diprediksi menjadi kelas negatif ada sebanyak 1 data

Dari tabel 7 dan 8, akurasi yang dimiliki algoritma  $k$ -NN sebesar 72,96% lebih baik dibandingkan dengan kinerja algoritma SVM sebesar 71,19%.

#### 4. SIMPULAN

Dari seluruh tahap penelitian yang telah dilakukan, dapat disimpulkan sebagai berikut:

1. Algoritma *K-Nearest Neighbor* ( $k$ -NN) dan algoritma *Support Vector Machine* (SVM) mampu menganalisis sentimen masyarakat tentang kebijakan *new normal*. Hasil klasifikasi dari analisis ini berupa kelas sentimen positif dan sentimen negatif.
2. Hasil dari pengujian ini menunjukkan bahwa akurasi yang dimiliki oleh algoritma SVM sebesar 71,19% dan algoritma  $k$ -NN dengan nilai  $k=5$  telah mendapatkan akurasi sebesar 72,96%. Pengujian tersebut menunjukkan bahwa kinerja algoritma  $k$ -NN terbukti lebih baik dibandingkan algoritma SVM dalam pengujian kasus ini.

#### Saran

Disarankan pada peneliti di masa mendatang dapat melakukan dengan pengembangan pada analisis bentuk kata-kata yang ambigu, serta pada akronim atau singkatan serta frasa asing yang diduga akan meningkatkan akurasi dari proses klasifikasi.

### DAFTAR PUSTAKA

- [1] F. V. Sari, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," 2019, vol. 10, no. 2, pp. 681–686.
- [2] A. Taufik, "Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel," in *Jurnal Teknik Komputer AMIK BSI (JTK)*, 2018, vol. IV, no. 2, pp. 69–74, doi: 10.31294/jtk.v4i2.3461.
- [3] N. L. Ratniasih, M. Sudarma, and N. Gunantara, "Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat," in *Majalah Ilmiah Teknologi Elektro*, 2017, vol. 16, no. 3, p. 13, doi: 10.24843/mite.2017.v16i03p03.
- [4] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," in *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2017, vol. 1, no. 1, p. 19, doi: 10.29207/resti.v1i1.11.
- [5] I. Saputra and D. Rosiyadi, "Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp," 2019, vol. 12, no. 2, pp. 101–111.
- [6] E. Z. Yonathan Sari Mahardhika, "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes," in *Prosiding SINTAK 2018*, 2018, no. 2015, pp. 409–413.
- [7] M. Rivki and A. M. Bachtiar, "Jurnal Sistem Informasi ( Journal of Information Systems ). 1 / 13 ( 2017 ), 31-37 DOI: <http://dx.doi.org/10.21609/jsi.v13i1.500>," 2017, vol. 13, pp. 31–37.
- [8] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine," in *Jurnal Informatika Polinema*, 2018, vol. 4, no. 2, p. 93, doi: 10.33795/jip.v4i2.152.