

# Prediksi Sentimen Pada Sosial Media Twitter Mengenai Produk Smartphone Menggunakan Algoritma K-NN Classification

Akbar Nugraha, Yulison Herry Chrisnanto, Rezky Yuniarti\*

Jurusan Informatika, Univ. Jendral Achmad Yani, Jl. Terusan Jendral Sudirman Kota Cimahi, Indonesia

Email: <sup>1</sup>akbnugrah4@gmail.com, <sup>2\*</sup>y.chrisnanto@gmail.com.

## Abstrak

Twitter merupakan salah satu media sosial yang memungkinkan penggunaanya untuk melakukan microblogging dan mem-publish status (tweet) dengan mudah. Salah satu aspek yang membantu perkembangan twitter adalah karena adanya pemasaran dan peningkatan produk, dalam sebuah laporan menyatakan ada sekitar 126 juta pengguna aktif setiap harinya. Sebagian dari tweet yang dibuat berisi sebuah opini mengenai produk ataupun sebuah iklan untuk pemasaran. Dengan jumlah data yang besar membuat kumpulan tweet dapat dimanfaatkan untuk melakukan analisis sentiment terhadap suatu produk untuk melihat respon dari konsumen. Penelitian ini menemukan hubungan kata kunci dan pola dari opini mengenai smartphone menggunakan metode K-nearest neighbor(K-NN). Beberapa penelitian terdahulu melakukan analisis sentiment dengan menggunakan metode Naïve Bayes classifier dengan hasil akurasi pada pemobotan tekstual sebanyak 68,52%, dan pada pemobotan non tekstual sebanyak 75,93% dari 11 kali pengujian. Penelitian ini akan membuat system untuk mengidentifikasi sentiment yang disampaikan masyarakat pada media sosial twitter mengenai produk smartphone. Luaran dari penelitian adalah system yang mampu mengklasifikasikan tweet, apakah tweet tersebut masuk kedalam opini negative, positif atau netral yang akan diimplementasikan dalam perangkat lunak.

**Kata Kunci:** Twitter, Analisis Sentimen, Klasifikasi, K-Nearest Neighbor, Vector Space Model, Smartphone

## 1. PENDAHULUAN

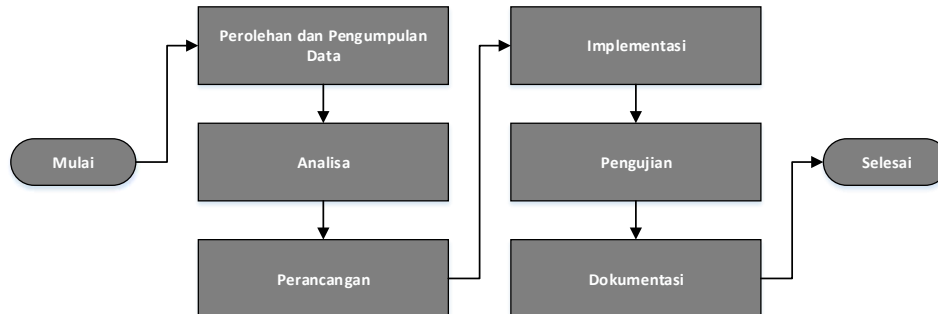
Media sosial merupakan sebuah media untuk bersosialisasi satu sama lain dan dilakukan secara online yang memungkinkan manusia untuk saling berinteraksi tanpa dibatasi ruang dan waktu. Kemudahan orang untuk menyampaikan pendapat kini dapat disampaikan lewat media sosial sehingga diterbitkan Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) untuk membatasi publik untuk memberikan kritik. *Twitter* merupakan salah satu media sosial yang digunakan untuk menyampaikan pesan-pesan singkat yang berupa *tweet* yang tidak jarang membuat sesuatu menjadi lebih heboh dengan fitur *Trending topic* yang dimiliki oleh *Twitter*. Indonesia menduduki posisi ketiga sebagai Negara dengan pengguna *Twitter* terbanyak, hal ini terlihat dari jumlah pengguna twitter di Indonesia yang mencapai lima puluh juta pada pertengahan tahun 2015 [1]. Kata yang disampaikan dalam *Twitter* tidak disampaikan dalam bentuk yang terstruktur, sehingga menimbulkan kesulitan dalam mencerna informasi [2], dan kebiasaan masyarakat mengutarakan pendapatnya melalui *Twitter*, dapat menjadi acuan untuk mengetahui sentiment masyarakat terhadap berbagai topik seperti tokoh publik [3], tempat, situasi, dan lainnya. Sebagai contoh dalam pemilihan kepala daerah (Pilkada) sentiment ini dapat membantu dalam menilai keberhasilan kampanye dan banyaknya dukungan [4].

Perkembangan smartphone semakin beragam tidak hanya dari penampilan saja tetapi juga dari segi fitur dan spesifikasi. Pasar smartphone juga telah berkembang pesat, tidak hanya dalam penjualan konvensional tetapi telah merambah di toko online. Sehingga percakapan di media sosial tentunya akan mengandung sentiment berupa opini dari masyarakat mengenai pendapat mereka tentang produk Smartphone yang mereka gunakan ataupun yang akan mereka beli, baik itu berupa sentiment positif ataupun negatif [5] [6]. Karena tidak semua Smartphone memiliki kualitas yang baik untuk mendukung kebutuhan konsumen dan itu harus diperhatikan oleh konsumen. Sebelum konsumen memutuskan untuk membeli smartphone, mereka harus mengetahui detail spesifikasi dan fungsi smartphone, hal ini dapat dipelajari dari kesaksian dan pendapat atau hasil peninjauan pengguna ponsel cerdas. Saat ini konsumen yang menulis opini dan pengalaman online meningkat. Jika konsumen membaca seluruh ulasan itu bisa menghabiskan banyak waktu. Selain itu jika dibaca tanpa evaluasi yang benar akan menimbulkan kesalahan dalam menilai. Klasifikasi teks adalah bagian dari *text mining* [7], dan merupakan suatu metode untuk memprediksi kategori kelas dari suatu data. Beberapa metode klasifikasi cukup banyak digunakan untuk melakukan klasifikasi teks seperti metode *Naïve Bayes Classifier*, *K-Nearest Neighbor* [8], *Support Vektor Machines*, hingga *Neural Network* [9]. Penelitian ini menggunakan metode K-NN dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasarkan attribute dan data latih. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari K obyek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru dan jarak dihitung berdasarkan jarak *euclidian*. Kelebihan dari algoritma ini adalah memiliki kemampuan yang dapat menangani data yang sangat besar dan rumit.

Penelitian ini berkaitan dengan tahapan klasifikasi [10], dengan metode *Double Propagation* (DP) untuk kasus data *tweet* yang telah dilakukan [6], mampu mengubah parameter menjadi 7 parameter sentiment yaitu sangat positif, positif, agak positif, netral, agak negatif, negative, sangat negative. Dengan data sebanyak 128 *tweet* didapatkan hasil akurasi 23,43%. Kemampuan klasifikasi metode K-NN yang mampu menangani banyak data diharapkan dapat menaikkan tingkat akurasi dan efektivitas dalam sistem prediksi sentiment *twitter* yang akan dibangun.

## 2. METODOLOGI PENELITIAN

Dalam melakukan sebuah penelitian, dibutuhkan suatu acuan pelaksanaan yang dinamakan metodologi penelitian. Metodologi penelitian terdiri dari beberapa tahapan kerangka kerja penelitian secara terstruktur / sistematis mulai dari tahap awal penelitian hingga mendapatkan hasil yang ingin dicapai. Berikut adalah gambaran tahapan yang dilakukan dalam penelitian yang akan disajikan pada Gambar 1 berikut.



**Gambar 1.** Metode Penelitian

### 2.1 Perolehan dan Pengumpulan Data.

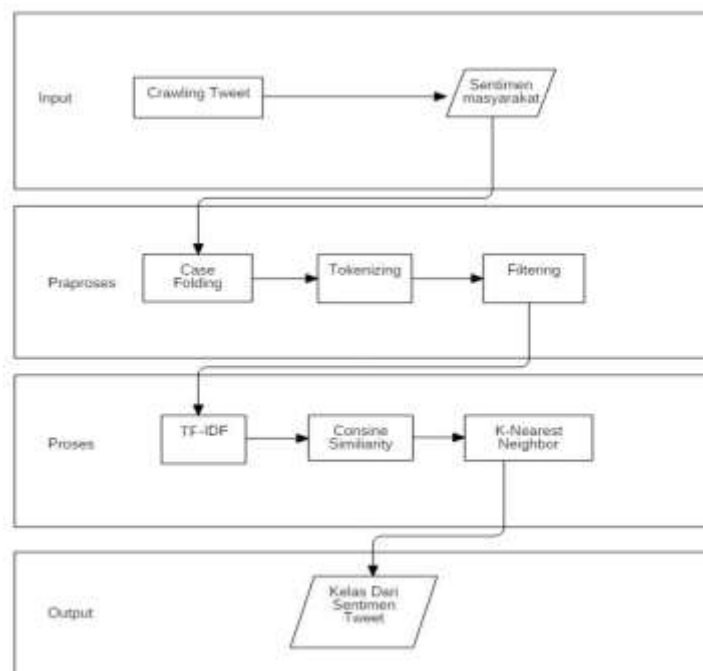
Data yang akan digunakan berasal dari data tweet semua pengguna twitter yang dipilih berdasarkan kebutuhan mengenai sentiment smartphone. Data yang digunakan sebanyak 125 data tweet untuk dijadikan sebagai data training dan data testing.

**Tabel 1.** Sample Data Tweet

No.	Tweet	Sentimen
1	Xiaomi Redmi 5 Plus mantap buat game bro..! @mryanto	Positif
2	Handphone Samsung spesifikasi kecil harga mahal	Negatif
...	...	...
123	Iphone X Resmi Rilis di Indonesia	Netral
124	Vivo udah murah spesifikasi tinggi dapet skin mobile legend lagi	Positif
125	Xiaomi performanya turun terus seiring pemakaian	Negatif

### 2.2 Analisa dan Perancangan Sistem

Sistem yang dibangun merupakan sistem yang dapat mengklasifikasikan sebuah tweet yang berisi opini mengenai smartphone. Terdapat dua proses pada sistem yaitu proses data latih dan data uji. Dimana data uji akan dibandingkan dengan data latih yang ada, sehingga menghasilkan rekomendasi informasi. Gambaran umum dari sistem yang akan dibuat dapat dilihat pada Gambar 1.



**Gambar 2.** Perancangan Sistem

Keterangan :

Dalam perancangan sistem ini terdapat 4 tahapan yaitu :

1. Input. Terdapat satu data masukan dari hasil crawling yang merupakan data uji dimana tweet akan diambil berdasarkan keyword yang dimasukkan lalu tweet akan dibandingkan dengan sentiment masyarakat yang merupakan data latih dan dilanjutkan ke praproses.
2. Praproses. Dalam tahap ini terdapat beberapa langkah yang dilakukan yaitu :
  - a. Case Folding: Dimana data latih dan data uji diubah semuanya menjadi huruf kecil dan karakter selain huruf ‘a’ sampai dengan ‘z’ akan dianggap sebagai delimiter.
  - b. Tokenizing : adalah tahap pemotongan kalimat menjadi kata.
  - c. Filtering : adalah tahap mengambil kata-kata penting dari hasil token menggunakan stopwords.
3. Proses. Dalam tahap ini terdapat beberapa langkah yang dilakukan yaitu :
  - a. TF-IDF : adalah proses dimana tweet hasil pencarian atau data uji akan dihitung bobotnya bersamaan dengan data latih yang ada berdasarkan jumlah kata yang sama.
  - b. Cosine Similarity : adalah proses penghitungan jarak kemiripan antara data uji dan data latih.
  - c. K-Nearest Neighbor : adalah proses klasifikasi dimana dari beberapa jarak yang ada akan dihitung mana tetangga yang lebih dekat dan lebih banyak sejumlah nilai K.
4. Output. Dalam tahap ini hasil akhir sudah diketahui, bahwa dari data tweet atau data uji mendekati kelas yang mana dari 15 kelas yang ada.

## 2.2 Implementasi

### 2.3.1 Text Mining

Text mining merupakan salah satu cabang ilmu data mining yang menganalisis suatu data berupa teks. Text mining adalah suatu langkah analisis teks yang dilakukan otomatis oleh komputer untuk menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen. Ide awal pembuatan text mining adalah untuk menemukan pola-pola informasi yang dapat digali dari teks yang tidak terstruktur [2]. Saat ini, text mining telah mendapat perhatian dalam berbagai bidang, antara lain dibidang keamanan, biomedis, pengembangan perangkat lunak dan aplikasi, media online, pemasaran, akademik hingga politik [1]. Seperti halnya dalam data mining, aplikasi text mining pada suatu studi kasus, harus dilakukan sesuai prosedur analisis. Langkah awal sebelum suatu data teks dianalisis menggunakan metode dalam text mining adalah melakukan pre-processing teks [7]. Sehingga, setelah didapatkan data yang siap diolah, analisis text mining dapat dilakukan.

Teks yang akan digunakan untuk proses *text mining*, pada umumnya memiliki beberapa karakteristik diantaranya adalah memiliki dimensi yang tinggi, terdapat *noise* pada data, dan terdapat struktur teks yang tidak baik. Cara yang digunakan dalam mempelajari suatu data teks, adalah dengan terlebih dahulu menentukan fitur-fitur yang mewakili setiap kata untuk setiap fitur yang ada pada dokumen. Sebelum menentukan fitur-fitur yang mewakili, diperlukan tahap *preprocessing*. Tahapan *preprocessing* [9] dalam *text mining* secara umum adalah *case folding*, *tokenizing*, *filtering*, *stemming*, *tagging*, dan *analysing*.

### 2.3.2 Term Frequency-Inverse Document Frequency (TF-IDF)

*TF-IDF* adalah metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini akan menghitung nilai *term frequency* (TF) dan *inverse document frequency* (IDF) pada setiap *token* (kata) pada setiap dokumen atau tweet. Metode ini digunakan untuk menghitung bobot setiap token *t* pada dokumen *d* dengan rumus:

$$W_{dt} = tf_{dt} * IDF_t \dots \dots \dots (1)$$

Keterangan:

- d : dokumen ke-d
- t : kata ke-t dari kata kunci
- W : bobot dokumen ke-d terhadap kata ke-t
- tf : banyaknya kata yang dicari pada sebuah dokumen
- IDF : Inversed Document Frequency

Kemudian dihitung nilai IDF yang akan dilakukan dengan Persamaan 2 berikut:

$$IDF(t) = \log(D/df) \dots \dots \dots (2)$$

Keterangan:

- D : Total dokumen
- df : banyak dokumen yang mengandung kata yang dicari.

Tahap selanjutnya setelah bobot (W) masing masing tweet diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai W, semakin besar tingkat similiaritas tweet terhadap kata kunci dan sebaliknya.

### 2.3.3 Cosine Similarity

Cosine Similarity adalah perhitungan kemiripan antar dua vector n dimensi dimana pencarian menggunakan cosinus dari sudut antara kedua vector dan digunakan untuk membandingkan dokumen dalam text mining. Rumus untuk Cosine similarity dapat dilihat pada persamaan berikut :

$$\text{Similarity}(x, y) = \cos \theta = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \dots\dots\dots(3)$$

Dimana :

- $\sum_{i=1}^n x_i y_i$  = jumlah bobot kata dokumen x terhadap dokumen y.
- $\sqrt{\sum_{i=1}^n x_i^2}$  = akar dari jumlah bobot dokumen x.
- $\sqrt{\sum_{i=1}^n y_i^2}$  = akar dari jumlah bobot dokumen y.

### 2.3.4 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data *k* objek dalam data (*training*) yang paling dekat (mirip) dengan objek pada data baru (*testing*). *Nearest neighbor* melakukan pendekatan untuk mencari kemiripan dengan menghitung nilai kedekatan antara kasus baru dengan kasus lama. Nilai kemiripan tersebut diperoleh dari proses perhitungan dengan mendefinisikan bobot dari fitur atau atribut yang dijadikan parameter [8]. Dekat atau jauhnya tetangga dapat dihitung berdasarkan jarak *Euclidean* dengan rumus persamaan 4 berikut:

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \dots\dots\dots(3)$$

Keterangan :

- d* : jarak.
- a* : data uji.
- b* : *sample data*.
- i* : variable data.
- n* : dimensi Data

## 3. ANALISA DAN PEMBAHASAN

Dalam pengambilan data, data yang akan digunakan berasal dari tweet pengguna media social twitter yang diambil berdasarkan kata kunci, jadi untuk mengambil data diperlukan sebuah *keyword* yang akan memunculkan data yang berhubungan dengan *keyword* tersebut. Dalam penelitian ini data yang dapat diproses hanya mengenai produk smartphone karena data latih yang disediakan merupakan tweet mengenai pendapat masyarakat tentang *Smartphone* dalam Bahasa Indonesia dan apabila diluar bahasan tersebut program tidak akan efektif.

### 3.1 Pra-Proses

Sebagai contoh data uji terdapat sebuah tweet yang belum diketahui kelasnya dengan kode "DU" yaitu : "Kamera dari Iphone X gila keren BANGET!" data yang digunakan merupakan data hasil pencarian tweet dari *keyword* "Kamera Iphone x."

#### 3.1.1 Case Folding

Data uji yang dituliskan akan dimasukan pada tahap case folding untuk merubah text data uji menjadi lower case dan menghilangkan demiliter yang akan menghasilkan text seperti pada tabel 2 berikut :

**Tabel 2.** Case Folding

Tweet Asli	Tweet Hasil Case Folding
Kamera dari Iphone gila keren BANGET!	kamera dari iphone gila keren banget

#### 3.1.2 Tokenizing

Tokenizing adalah proses dimana tweet yang didapat setelah hasil case folding dipisahkan berdasarkan "Spasi" atau dipisah menjadi kata sebagai contoh dapat dilihat pada tabel 3:

**Tabel 3.** Tokenizing

Tweet	Token
kamera dari iphone gila keren banget	Kamera Dari Iphone Gila

Keren  
Banget

### 3.1.3 Filtering

Filtering adalah proses untuk mengambil kata-kata yang penting dari hasil tokenizing atau pembuangan kata yang tidak penting seperti kata yang tidak deskriptif seperti halnya “yang”, “dan”, “di”, “dari” dan lainnya. Seperti yang ditunjukkan pada Tabel 3.3.

**Tabel 4.** Filtering

Hasil Tokenizing	Hasil Filter
kamera	kamera
dari	-
iphone	iphone
gila	gila
keren	keren
banget	-

## 3.2 Siklus Algoritma

Siklus ini menggambarkan proses Algoritma untuk klasifikasi teks menggunakan K-NN dan TF-IDF setelah pra-proses sampai dengan output hasil klasifikasi teks tweet yang diambil.

### 3.2.1 Pemetaan Data Uji Dalam Wordnet Data Latih

Setelah pra-proses selesai, teks dari hasil *Crawling* akan dipetakan dalam kumpulan wordnet data latih. Pada uji coba ini digunakan dokumen data latih sebanyak 10 tweet dengan kelas yang berbeda beda. Karena keterbatasan halaman maka hasil dari pemetaan akan ditulis secara singkat, berikut adalah hasil pada Tabel 3.4 :

**Tabel 5.** Pemetaan Wordnet

Term	DU	DL0	DL1	DL2	DL3	DL4	DL5	...	DL9	DF
kamera	1	1	1	0	1	1	0	...	1	6
iphone	1	1	0	0	0	0	1	...	0	5
keren	1	1	1	0	1	1	0	...	0	7
xiaomi	0	0	0	1	1	0	1	...	0	4
asus	0	0	1	0	0	0	1	...	0	3
modelnya	0	0	1	0	0	0	0	...	0	2
begitu	0	0	1	0	0	0	0	...	0	1
saja	0	0	1	0	0	0	0	...	0	1
...	...	...	...	...	...	...	...	...	...	...
bagus	0	0	0	1	0	0	1	...	0	3
samsung	0	0	0	0	0	1	1	...	1	4
mereka	0	0	0	0	0	0	1	...	0	2
Total term	8	9	11	8	7	6	8	...	5	-

### 3.2.2 Pembobotan TF dan IDF

Setelah kata dipetakan dalam wordnet selanjutnya dihitung bobot setiap kata dengan menggunakan persamaan 1 untuk bobot Term dan persamaan 2 untuk bobot IDF, berikut pada Tabel 3.5 adalah hasil perhitungannya :

**Tabel 6.** Pembobotan TF dan IDF

Term	DL0	DL1	DL2	DL3	DL4	DL5	...	DL9	IDF
kamera	0.221849	0.221849	0	0.221849	0.221849	0	...	0.221849	0.221849
iphone	0.30103	0.30103	0	0	0	1	...	0	0.30103
keren	0.154902	0.154902	0	0.309804	0.154902	0	...	0	0.154902
xiaomi	0	0	0.39794	0.39794	0	1	...	0	0.39794
asus	0	0	0	0	0	1	...	0	0.522879
modelnya	0	0	0	0	0	0	...	0	0.69897
begitu	0	0	0	0	0	0	...	0	1
saja	0	0	0	0	0	0	...	0	1
...	...	...	...	...	...	...	...	...	...
bagus	0	0	0.522879	0	0	1	...	0	0.522879
samsung	0	0	0	0	0.39794	1	...	1	0.39794
mereka	0	0	0	0	0	1	...	0	0.69897

### 3.2.3 Vector Space Model

Setelah mendapatkan bobot dari TF-IDF, selanjutnya masuk dalam proses Vector Space Model (VSM) untuk mencari sudut antara kedua vector dengan menghitung dua sudut antara bobot data latih dan data uji. Berikut adalah hasil perkalian dari data Uji dan Data latih dan VSM:

**Tabel 7.** Vektor Space Model (1)

Term	DU*DL <sub>i</sub>							
	DL0	DL1	DL2	DL3	DL4	DL5	...	DL9
kamera	0.049	0.049	0	0.049	0.049	0	...	0.049
iphone	0.09	0.09	0	0	0	0.09	...	0
keren	0.023	0.023	0	0.047	0.023	0	...	0
xiaomi	0	0	0	0	0	0	...	0
asus	0	0	0	0	0	0	...	0
modelnya	0	0	0	0	0	0	...	0
begitu	0	0	0	0	0	0	...	0
saja	0	0	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...
bagus	0	0	0	0	0	0	...	0
samsung	0	0	0	0	0	0	...	0
mereka	0	0	0	0	0	0	...	0
TOTAL	0.327	0.327	0.048	0.261	0.237	0.254	...	0.213

Setelah mendapatkan hasil dari perkalian skalar maka akan dilanjutkan dengan menghitung Panjang vector untuk menyelesaikan rumus VSM. Berikut adalah perhitungan Panjang vector dapat dilihat pada Tabel 3.7

**Tabel 8.** Vektor Space Model (2)

Term	DU	Panjang Vektor							
		DL0	DL1	DL2	DL3	DL4	DL5	...	DL9
kamera	0.049217	0.049217	0.049217	0	0.049217	0.049217	0	...	0.049217
iphone	0.090619	0.090619	0.090619	0	0	0	0.090619	...	0
keren	0.023995	0.023995	0.023995	0	0.095978	0.023995	0	...	0
xiaomi	0	0	0	0.158356	0.158356	0	0.158356	...	0
asus	0	0	0	0	0	0	0.273402	...	0
modelnya	0	0	0	0	0	0	0	...	0
begitu	0	0	0	0	0	0	0	...	0
saja	0	0	0	0	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...
bagus	0	0	0	0.273402	0	0	0.273402	...	0
samsung	0	0	0	0	0	0.158356	0.158356	...	0.158356
mereka	0	0	0	0	0	0	0.488559	...	0
TOTAL	0.259831	0.259831	0.283831	0.503758	0.375552	0.303568	1.538695	...	0.207573

Selanjutnya setelah bobot didapatkan hitung bobot yang sudah didapat menggunakan Persamaan 3, berikut adalah nilai yang didapat dari hasil perhitungan akar bobot TF-IDF dan Consine Similarity :

**Tabel 9.** Vektor Space Model (3)

	DU	DL0	DL1	DL2	DL3	DL4	DL5	...	DL9
$\sqrt{\sum TF-IDF}$	0.509736	0.509736	0.532757	0.1059	0.612823	0.55097	1.240441	..	0.551602
Cosine	-	1.231572	1.178353	0.741001	0.832326	0.854552	0.395384	...	0.746876

### 3.2.4 K-Nearest Neighbor

Setelah mendapatkan jarak kemiripan dari hasil perhitungan sebelumnya, maka perhitungan dilanjutkan menggunakan K-Nearest Neighbor untuk menentukan kelas dari data uji yang belum diketahui kelasnya. Setelah bobot cosine didapat maka akan diurutkan bobot kemiripan dari bobot tertinggi hingga paling rendah, berikut adalah hasil bobot beserta kelas dari setiap data latih dapat dilihat pada tabel 10:

**Tabel 10.** Bobot dan Kelas Data Latih terhadap Data Uji

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
1.	DL0	1.231572	Positif

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
2.	DL1	1.178353	Positif
3.	DL2	0.741001	Negatif
4.	DL3	0.832326	Netral
5.	DL4	0.854552	Positif
6.	DL5	0.395384	Netral
7.	DL6	0.457286	Negatif
8.	DL7	0.682685	Negatif
9.	DL8	0.477291	Positif
10.	DL9	0.746876	Netral

Selanjutnya setelah data diketahui nilai kemiripan dengan dokumen uji nilai akan diurutkan mulai dari terbesar ke terkecil, berikut adalah hasil pengurutan nilai :

**Tabel 11.** Pengurutan Nilai Kedekatan

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
1	DL0	1.231572	Positif
2	DL1	1.178353	Positif
5	DL4	0.854552	Positif
4	DL3	0.832326	Netral
10	DL9	0.746876	Netral
3	DL2	0.741001	Negatif
8	DL7	0.682685	Negatif
9	DL8	0.477291	Positif
7	DL6	0.457286	Negatif
6	DL5	0.395384	Netral

### 3.2.5 Perangkingan Nilai K

Untuk mengetahui hasil akhir dari dokumen uji selanjutnya adalah mencari dokumen yang paling dekat dengan data uji dengan nilai K= 1, 3, 5, 7, dan 9. Berikut adalah hasil perangkingan nilai berdasarkan nilai K:

**Tabel 12.** Nilai K=1

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
1	DL0	1.231572	<b>Positif</b>

**Tabel 13.** Nilai K=3

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
1	DL0	1.231572	<b>Positif</b>
2	DL1	1.178353	<b>Positif</b>
5	DL4	0.854552	<b>Positif</b>

**Tabel 14.** Nilai K=5

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
1	DL0	1.231572	<b>Positif</b>
2	DL1	1.178353	<b>Positif</b>
5	DL4	0.854552	<b>Positif</b>
4	DL3	0.832326	Netral
10	DL9	0.746876	Netral

**Tabel 15.** Nilai K=7

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
1	DL0	1.231572	<b>Positif</b>
2	DL1	1.178353	<b>Positif</b>
5	DL4	0.854552	<b>Positif</b>
4	DL3	0.832326	Netral
10	DL9	0.746876	Netral
3	DL2	0.741001	Negatif
8	DL7	0.682685	Negatif

**Tabel 16.** Nilai K=9

No	Data Latih	Nilai Kedekatan Dengan Dokumen Uji	Kelas
----	------------	------------------------------------	-------

1	DL0	1.231572	<b>Positif</b>
2	DL1	1.178353	<b>Positif</b>
5	DL4	0.854552	<b>Positif</b>
4	DL3	0.832326	Netral
10	DL9	0.746876	Netral
3	DL2	0.741001	Negatif
8	DL7	0.682685	Negatif
9	DL8	0.477291	<b>Positif</b>
7	DL6	0.457286	Negatif

#### 4. KESIMPULAN

Berdasarkan data uji berupa tweet, dengan menerapkan Teknik VSM dimana tweet berupa teks yang diubah menjadi kumpulan angka lalu dihitung nilai kedekatan setiap kelas dengan menggunakan metode K-NN serta menguji jumlah kedekatan dengan nilai K yang beragam dapat diperoleh kelas terbanyak pada setiap pengujian menunjukkan bahwa tweet yang diuji dapat diidentifikasi sebagai kelompok kelas tertentu, dalam hal ini tweet yang digunakan sebagai data uji pada penelitian ini mengarah pada kelas Positif.

#### REFERENCES

- [1] R. T. Lestari, R. S. Perdana and M. A. Fauzi, "Analisis Sentimen Tentang Opini Pilkada DKI 2017 pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan Pembobotan Emoji," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 12, pp. 1718-1724, 2017.
- [2] Hamzah, "Klasifikasi Teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis," in *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, Yogyakarta, 2014.
- [3] Sunni and D. H. Widyantoro, "Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik," *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, vol. 1, no. 2, pp. 200-206, 2015.
- [4] G. A. Buntoro, T. B. Adji and A. E. Purnamasari, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," in *CITEE 2014*, Yogyakarta, 2014.
- [5] V. Bijalwan, V. Kumar, P. Kumari and J. Pascual, "KNN Based Machine Learning Approach for Text and Document Mining," *International Journal of Database Theory and Application*, vol. 7, pp. 61-70, 2014.
- [6] [www.teorikomputer.com/pengertian-mysql-beserta-kelebihan-dan.html](http://www.teorikomputer.com/pengertian-mysql-beserta-kelebihan-dan.html), 10 2015.
- [7] <https://informatikalogi.com/term-weighting-tf-idf>.
- [8] Surya Prasath, Abu Alfeilat, Lasassmeh, Hassanat, & Tarawneh, Distance and Similiarity Measures Effect on The Performance of K-Nearest Neighbor Classifier, 2019
- [9] <https://cmry.github.io/notes/euclidean-v-cosine>