

ANALISIS SENTIMEN OPINI PUBLIK MENGENAI SARANA DAN TRANSPORTASI MUDIK TAHUN 2019 PADA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES, NEURAL NETWORK, KNN DAN SVM

Melisa Winda Pertiwi

Sistem Informasi
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
<http://www.nusamandiri.ac.id>
melisa.mwp@nusamandiri.ac.id

Abstract— Various moments are immortalized in social networking, one of them Twitter. Twitter is one of the social media that allows users to interact, share information, or even to express feelings and opinions as well as in expressing opinions about "Mudik". Every tweets not all contain the same value, therefore there needs to be a sentiment analysis and tweet classification that discusses about to know the amount of public sentiment to mudik year 2019. This will be a concern for "facilities transportation for goinghome" next year. One of method used is classification method using k-NN, SVM, Naïve Bayes and Neural Network. The results of this study can be seen that positive sentiments appear more than the value of other sentiments for Mudik 2019, next time negative's opinion appear more than positive Mudik 2019. and k-NN algorithm has a higher accuracy, it is accuracy=90.76 and AUC=0.939, SVM generated accuracy=89.03% and AUC=0.500, Naïve Bayes generated accuracy=78.16% and AUC=0.567 then Neural Network generated accuracy=52.73% and AUC=0.000.

Keywords: Accuracy, KNN, Naïve Bayes, Neural Network, Sentiment Analysis, Twitter, SVM.

Abstrak— Berbagai momen diabadikan dalam jejaring sosial, salah satunya Twitter. Twitter adalah salah satu media sosial yang memungkinkan pengguna untuk berinteraksi berbagai informasi atau bahkan mengekspresikan perasaan dan pendapat serta dalam mengekspresikan pendapat tentang "Mudik". Setiap tweets tidak semua mengandung nilai yang sama. Oleh karena itu perlu ada analisis sentiment dan klasifikasi tweets yang membahas tentang jumlah sentiment public mengenai Mudik tahun 2019. Ini akan menjadi perhatian untuk transportasi fasilitas kebijakan untuk mudik tahun berikutnya. Beberapa metode yang digunakan adalah metode klasifikasi menggunakan k-NN, SVM, Naïve Bayes dan Neural Network. Hasil penelitian ini dapat dilihat bahwa sentiment positif muncul lebih pada Mudik 2019 sedangkan untuk sentiment negatif muncul lebih banyak hari-hari berikutnya Mudik 2019. Algoritma k-NN memiliki akurasi yang lebih tinggi, akurasi= 90,76% dan AUC= 0,939; akurasi yang dihasilkan SVM= 89,03% dan AUC= 0,5; akurasi Naïve Bayes= 78,16% dan AUC= 0,567 kemudian yang dihasilkan Neural Network= 52,73% dan AUC=0,0.

Kata kunci: Akurasi, Analisa Sentimen, KNN, Mudik, Naïve Bayes, Neural Network, Twitter, SVM

PENDAHULUAN

Hari Raya merupakan salahsatu hari besar bagi masyarakat Indonesia. Memanfaatkan momen hari raya sebagai momen untuk berkumpul dengan keluarga dan itu sudah menjadi suatu tradisi. Pada Hari Raya tersebut, hampir seluruh instansi kantor pemerintah, perusahaan, sekolah, toserba dan lainnya diliburkan terkecuali bagi mereka yang memiliki tugas pada hari tersebut. Dikenal dengan istilah "mudik" salahsatu hal yang sering menjadi wacana dalam beberapa media ketika libur hari raya maupun libur panjang lainnya. Hal yang berhubungan dengan mudik hari raya adalah anggapan mengenai sarana dan transportasi saat

momen tersebut terjadi. Berbagai momen mudik diabadikan di jejaring sosial, salahsatunya Twitter. Twitter adalah layanan jejaring sosial dan mikroblog yang memungkinkan penggunaanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan/tweet.

Setiap twit tidak selalu mengandung nilai yang sama, untuk itulah diperlukan suatu analisis sentiment dan klasifikasi negatif, positif dan netral dari twit yang membahas tentang mudik hari raya sepanjang sepanjang tahun 2019. *Sentiment analysis* atau *opinion mining* adalah studi komputasional dari opinion - opini orang, sentiment dan emosi melalui entitas dan atribut

yang dimiliki yang diekspresikan dalam bentuk teks

Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral (Liu, 2012)..

Text mining (penambangan teks) adalah penambangan yang dilakukan oleh computer untuk mendapatkan sesuatu yang baru dalam bentuk sebuah informasi, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda. Tujuan utama dari Text Mining adalah menganalisa informasi untuk menemukan pola-pola (Simpson & Demner-Fushman, 2012).

Hal ini sejalan dengan data mining yang bertujuan untuk mengekstrak pola-pola menarik dari data, bedanya untuk text mining data ini berbentuk teks sedangkan untuk data mining data ini berbentuk angka (Aggarwal & Zhai, 2013). Data mining berasumsi bahwa data sudah dalam bentuk terstruktur sedangkan pada text mining data dalam bentuk tidak terstruktur dari koleksi dokumen (corpus) harus diolah terlebih dahulu (pre processing) menjadi bentuk terstruktur (Simpson & Demner-Fushman, 2012).

Sentiment Analysis atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif (Liu, 2012).

Sentiment analysis adalah kegiatan melakukan analisa terhadap pendapat, opini, sikap atau emosi seseorang mengenai suatu produk, topik atau permasalahan tertentu sehingga bisa diketahui hal tersebut masuk kedalam sentimen positif, negatif atau netral.

Sentiment analysis dapat dibedakan berdasarkan sumber datanya, beberapa level yang sering digunakan dalam penelitian sentiment analysis adalah sentiment analysis pada level dokumen dan pada level kalimat (Fink, Chou, Kopecky, & Llorens, 2011). Berdasarkan level sumber datanya sentiment analysis terbagi ke dalam 2 kelompok besar yaitu: Coarse-grained Sentiment Analysis dan Fine-grained *Sentiment Analysis*.

Penelitian ini berfokus untuk mengklasifikasi terhadap twit yang mengandung

nilai positif dan negatif mengenai mudik hari raya menggunakan algoritma k-Nearest Neighbor (k-NN), *Support Vector Machine* (SVM), Naïve Bayes dan *Neural Network*. Hasil penelitian tersebut akan menjadi gambaran mengenai Mudik 2019 apakah opini publik lebih cenderung ke opini positif/negatif/netral dan membandingkan hasil akurasi yang lebih tinggi dari algoritma yang digunakan yaitu k-Nearest Neighbor (k-NN), *Support Vector Machine* (SVM), Naïve Bayes dan *Neural Network*.

BAHAN DAN METODE

1. Pengumpulan Data

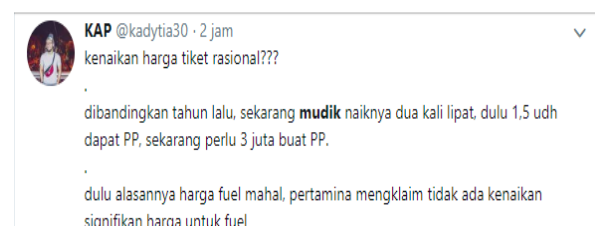
Data yang digunakan pada penelitian ini berasal dari tweet yang berisi opini dengan kata kunci yang berhubungan yaitu: "mudik", "sarana", "transportasi" dan "macet". Data tersebut diperoleh dari situs jejaring sosial Twitter, www.twitter.com. Terdapat banyak opini mengenai "Sarana dan Transportasi Mudik" yang diakses pada bulan Mei 2019. Data diambil secara random sebanyak 347 tweets.

Berikut ini contoh data opini positif dan negatif mengenai "Sarana dan Transportasi Mudik Tahun 2019" yang diambil pada *website* resmi www.twiiter.com:



Sumber: (Twitter, 2019)

Gambar1 Contoh Opini Positif Mengenai Sarana dan Transportasi Mudik 2019



Sumber: (Twitter, 2019):

Gambar2 Contoh Opini Negatif Mengenai Sarana dan Transportasi Mudik 2019

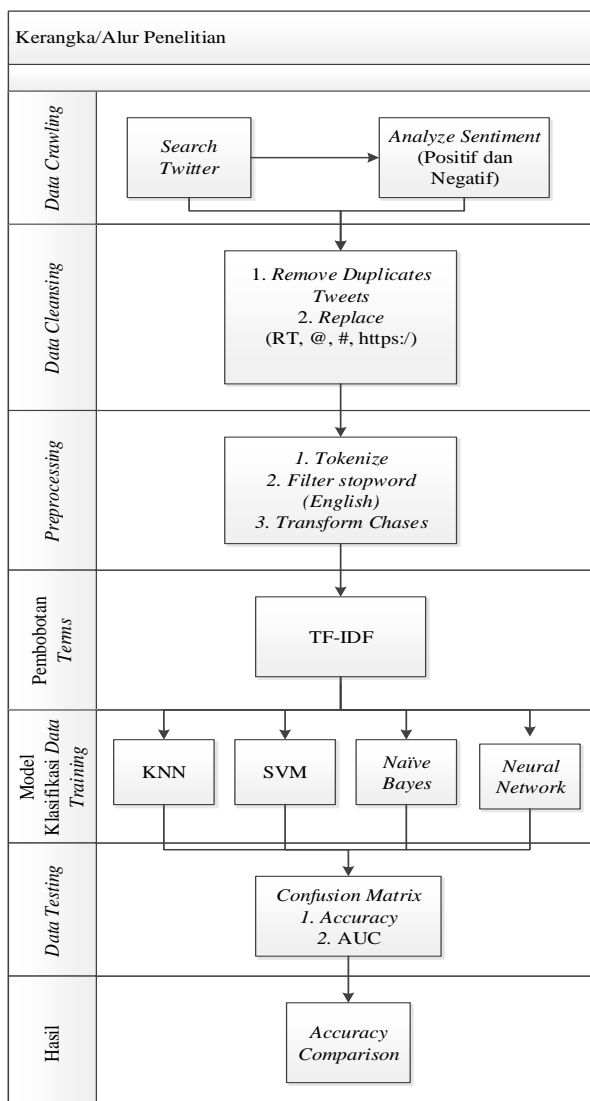
2. Pengolahan Data Awal

Pada tahap ini dilakukan klasifikasi sentimen (kategori: positif dan negatif) lalu dilanjutkan dengan tahapan data cleansing. Tahapan tersebut mengubah beberapa karakter pada tweet. Contohnya “RT” diganti dengan “retweet”, “@” diganti dengan “at” dll. Selain itu ada juga atribut – atribut data yang tidak relevan. Data tersebut dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining.

3. Metode yang diusulkan

Metode yang diusulkan yaitu metode klasifikasi dengan model yang menggunakan empat algoritma k-Nearest Neighbor (k-NN) dan Support Vector Machine (SVM), Naive Bayes dan Neural Network.

Hasilnya berupa perbandingan akurasi yang baik diantara kedua algoritma tersebut.



Sumber : Hasil Penelitian (Pertwi, 2019)
Gambar1. Kerangka Alur Penelitian

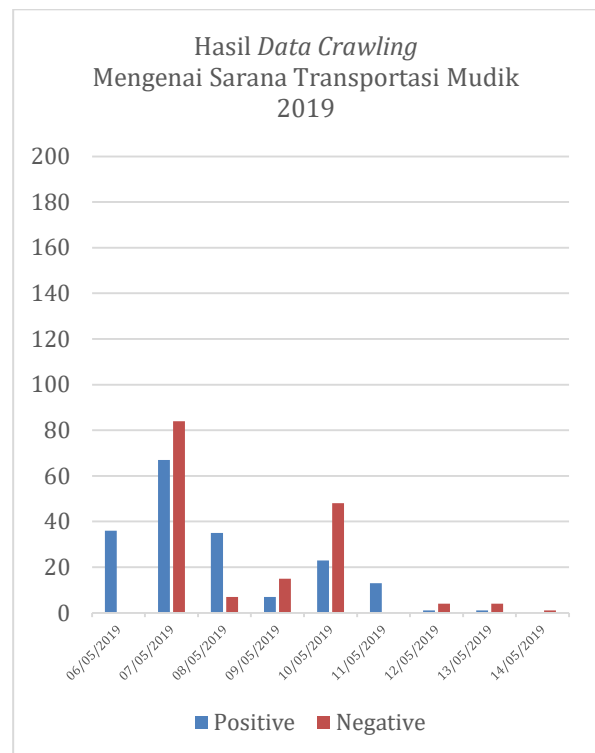
HASIL DAN PEMBAHASAN

Data training yang digunakan pada saat pengujian data, diambil dari www.twitter.com. Pengujian data dilakukan dengan menggunakan opini publik tentang tahun baru dengan kata kunci Mudik. Data diambil secara acak menggunakan operator search twitter pada RapidMiner dengan sample 347 data. Kemudian di dapat 164 opini negatif dan 183 opini positif. Kemudian dilakukan testing dan training dataset sehingga didapat accuracy dan AUC. Berikut akan dijelaskan lebih rinci mengenai hasil penelitian yang diperoleh

Tabel 1. Hasil Data Crawling

Created-at	Positive	Negative
5/6/2019	36	0
5/7/2019	67	84
5/8/2019	35	7
5/9/2019	7	15
5/10/2019	23	48
5/11/2019	13	0
5/12/2019	1	4
5/13/2019	1	4
5/14/2019	0	1
	183	164

Sumber: (Twitter, 2019)



Sumber: Hasil Penelitian (Pertwi, 2019)
Gambar3. Grafik Hasil Data Crawling Twitter

Mengenai Sarana Transportasi Mudik Tahun 2019

1. Data Cleansing

Tabel 2. Teks Sebelum dan Sesudah Tahapan *Data Cleansing*

Sebelum tahapan <i>data cleansing</i>	@fthon666 @AutonetMagz @TMCPoldaMetro nagrek macet
Sesudah tahapan <i>data cleansing</i>	At fthon666 at AutonetMagz at TMCPoldaMetro nagrek macet

Sumber: Hasil Penelitian (Pertiwi, 2019)

2. Tokenize

Tabel 3. Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Tokenize*

Teks dilakukan <i>tokenize</i> sebelum proses	Jalur pinggir kali Ketanggungan - Mergasari ramai lancar, aspal basah habia hujan ☺
Teks setelah dilakukan proses <i>tokenize</i>	jalur pinggir kali ketanggungan - mergasari ramai lancar, aspal basah habia hujan

Sumber: Hasil Penelitian (Pertiwi, 2019)

3. Filter Stopwords (Indonesia)

Tabel 4. Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Stopwords Removal*

Teks dilakukan <i>stopwords removal</i> sebelum proses	H-1 Lebaran, Stasiun Pasar Senen msh dipadati Penumpang. Hari ini sebanyak 26rb org diberangkatkan dg tujuan Jogja terbanyak
Teks setelah dilakukan proses <i>stopwords removal</i>	h-1 lebaran, stasiun pasar senen dipadati penumpang. hari ini sebanyak 26rb diberangkatkan tujuan jogja terbanyak

Sumber: (Pertiwi, 2019)

4. Transform Cases

Tabel 5. Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Transform Cases*

Teks dilakukan <i>transform cases</i> sebelum proses	Indonesia memiliki tradisi unik dalam menyambut hari raya idul fitri. MUDIK!
Teks setelah dilakukan proses <i>transform cases</i>	indonesia memiliki tradisi unik dalam menyambut hari raya idul fitri. mudik

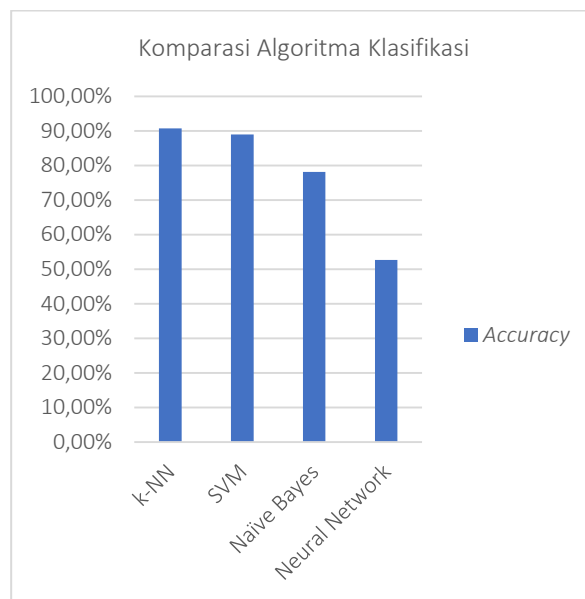
Sumber: (Pertiwi, 2019)

5. Hasil Perbandingan Komparasi Algoritma k-NN, SVM, Naive Bayes dan *Neural Network*

Tabel 6. Komparasi *Accuracy* dan AUC Algoritma Klasifikasi

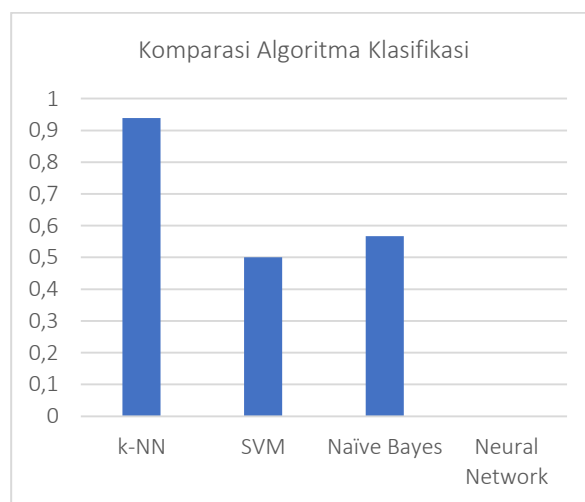
Algoritma	<i>Accuracy</i>	AUC
k-NN	90.76%	0.939
SVM	89.03%	0.500
<i>Naive Bayes</i>	78.16%	0.567
<i>Neural Network</i>	52.73%	0.000

Sumber: (Pertiwi, 2019)



Sumber: Hasil Penelitian (Pertiwi, 2019)

Gambar 4. Diagram Komparasi Nilai *Accuracy* Algoritma Klasifikasi



Sumber: (Pertiwi, 2019)

Gambar 5. Diagram Komparasi Nilai AUC Algoritma Klasifikasi

Hasilnya menunjukkan bahwa algoritma k-NN lebih tinggi dibandingkan algoritma SVM, Naïve Bayes dan Neural Network terhadap dataset yang digunakan dan untuk perbandingannya. Dalam penelitian ini, accuracy k-NN 90.76% dan nilai AUC=0.939 terhadap nilai k=5, SVM lebih rendah dari k-NN mendapatkan accuracy=89.03 dan AUC=0.500, algoritma Naïve Bayes menghasilkan akurasi lebih rendah dibanding k-NN dan SVM yaitu nilai accuracy=78.16% dan nilai AUC=0.657 sedangkan akurasi paling rendah terhadap dataset yaitu pada algoritma Neural Network yang mendapatkan nilai accuracy=52.73% dan AUC=0.000.

KESIMPULAN

Berdasarkan hasil penelitian dapat ditarik beberapa kesimpulan bahwa penelitian mengenai Sarana dan Transportasi Mudik setiap tahunnya dapat mengalami perubahan karena opini setiap orang akan berbeda, dan penelitian ini dapat digunakan sebagai acuan untuk Mudik tahun selanjutnya khususnya pada Sarana dan Transportasi maupun penelitian terkait mengenai "Mudik" lainnya.

Hasil sentimen publik yang diambil dengan kata kunci Mudik Hari Raya 2019 menunjukkan bahwa pengguna Twitter lebih banyak memberikan opini positif. Opini tersebut selama mudik berlangsung semakin banyak yang bernilai positif namun setelah beberapa hari jumlah opini negatif meningkat. Ini membuktikan bahwa pengguna Twitter menunjukkan respon yang positif terhadap Mudik 2019 khususnya pada mudik hari raya.

Algoritma k-NN akurasi lebih tinggi dibandingkan algoritma SVM, Naïve Bayes dan Neural Network untuk melakukan klasifikasi sentimen analisis teks berbahasa Inggris dengan nilai accuracy=90.76% dan AUC=0.939.

REFERENSI

- Aggarwal, C. C., & Zhai, C. X. (2013). Mining text data. *Mining Text Data*, 9781461432, 1-522. <https://doi.org/10.1007/978-1-4614-3223-4>
- Fink, C. R., Chou, D. S., Kopecky, J. J., & Llorens, A. J. (2011). Coarse- and fine-grained sentiment analysis of social media text. *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, 30(1), 22-30.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human*

Language Technologies, 5(1), 1-167. <https://doi.org/https://doi.org/10.2200/S00416ED1V01Y201204HLT016>

Pertiwi, M. W. (2019). *Laporan Akhir Penelitian*. STMIK Nusa Mandiri.

Simpson, M. S., & Demner-Fushman, D. (2012). *Biomedical Text Mining: A Survey of Recent Progress*. (C. C. Aggarwal & C. Zhai, Eds.), *Mining Text Data*. New York: Springer. <https://doi.org/10.1007/978-1-4614-3223-4>

Twitter. (2019). Mudik. Retrieved from <https://twitter.com/search?q=mudik&src=typd>

