

## **KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KENAIKAN HARGA TIKET PESAWAT PADA TWITTER MENGGUNAKAN NAÏVE BAYES**

**AGUSTINA MARDEKA RAYA<sup>1</sup>, FITRI NURBAITI<sup>2</sup>, DETIN SOFIA<sup>3</sup>**

Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jl. Ciledug Raya, RT.10/RW.2, Petukangan Utara, Kec. Pesanggrahan, Kota Jakarta Selatan, DKI Jakarta 12260 e-mail : merdekaraya.amr@gmail.com, nurbaiti288@gmail.com, [sofia.detin@gmail.com](mailto:sofia.detin@gmail.com)

### **ABSTRAK**

Analisa sentiment atau opinion mining merupakan teknik yang digunakan untuk mengolah data yang didapatkan dari media sosial. Kenaikan harga tiket pesawat terjadi setelah adanya kecelakaan dari salah satu maskapai domestik. Imbas dari kecelakaan tersebut menyebabkan semua tiket maskapai penerbangan Indonesia mengalami kenaikan. Banyak masyarakat yang menyayangkan dengan kenaikan tiket pesawat tersebut. Penelitian ini akan memanfaatkan data tweet pada twitter untuk melihat persepsi masyarakat terhadap kenaikan harga tiket pesawat dengan menggunakan metode naïve bayes dan KNN dengan mengklasifikasikan sentimen sentiment positif, negative dan netral. Tingkat akurasi klasifikasi menggunakan naïve bayes sebesar 90.70% sedangkan tingkat akurasi dengan metode KNN sebesar 62.79%

*Kata Kunci* : Analisa Sentimen, Klasifikasi, Naïve bayes, Tiket Pesawat.

### **I.PENDAHULUAN**

Teknologi informasi saat ini tidak bisa lepas dari kehidupan sehari-hari. Banyak kemudahan yang diberikan untuk baik untuk mencari informasi, berkomunikasi atau kemudahannya lainnya. Pada era teknologi informasi saat ini, banyak informasi yang mengalir melalui jaringan. Web dengan berbagai konten meliputi informasi objektif maupun opini subjektif.

Analisa sentiment merupakan bagian dari text mining. Text mining merupakan proses pengambilan informasi data tekstual yang memiliki kualitas tinggi serta dapat mengetahui permasalahan dalam teks dari sebuah topik tertentu. Media sosial twitter menyediakan layanan tweet dimana user dapat menulis semua keluhan mereka atau pendapat mereka mengenai masalah dan topik tertentu. Pendapat tersebut merupakan pandangan positif maupun negative user terhadap suatu topik. Pendapat-pendapat user pada twitter tersebut dapat dijadikan penelitian untuk melihat informasi.

Menurut (Adisasmita, 2010), Bangsa yang maju adalah bangsa yang memiliki sistem transformasi yang berkualitas. Transformasi yang diperlukan untuk menunjang sektor lain untuk berkembang guna mendorong perkembangan daerah tertinggal.

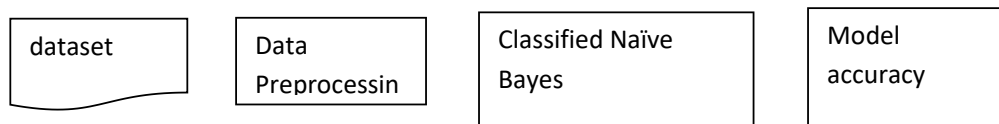
Transportasi udara atau pesawat merupakan transportasi yang dapat menempuh jarak yang jauh dengan waktu yang relative singkat. Karena alasan efisiensi waktu tersebut tingkat penggunaan transportasi udara di Indonesia mengalami peningkatan setiap tahunnya. Selama Januari-Desember 2016 jumlah penumpang domestik mencapai 80,4 juta orang atau naik 16,97 persen dan jumlah penumpang internasional mencapai 14,8 juta orang atau naik 8,16 persen dibanding periode yang sama tahun 2015. (bps.go.id) November 2018 hingga juni 2019 harga tiket pesawat naik rata-rata sebesar 40% hingga 120%. Kenaikan harga tiket pesawat terjadi pasca kecelakaan salah satu maskapai penerbangan domestic. Banyak masyarakat yang mengeluh atas naiknya harga tiket pesawat. Penelitian ini bertujuan untuk melihat persepsi masyarakat terhadap kenaikan harga tiket pesawat melalui media sosial twitter.

Hasil penelitian yang dilakukan oleh Fauziah Afshoh (2017) berjudul analisa sentimen menggunakan naïve bayes untuk melihat persepsi masyarakat terhadap kenaikan harga jual rokok pada media sosial twitter, menghasilkan nilai precision 53% pada data opini negatif dan nilai recall 57% pada data opini negatif. Untuk nilai akurasi menggunakan Lexion Based menghasilkan nilai persentase sebesar 81% pada data opini netral. Pengujian menggunakan metode naïve bayes classifier nilai precision sebesar 76% dengan opini positif, dan nilai recall 93% dengan opini positif, dengan nilai accuracy saat pengujian sebesar 88% dengan opini netral. Dapat dilihat bahwa kinerja naïve bayes classifier dalam melakukan klasifikasi terbukti lebih baik dari pada lexion based. Dengan hasil presentasi tweet 5% netral, 53% positif dan 41% negatif.

Penelitian yang berjudul sistem klasifikasi opini pengguna maskapai penerbangan di Indonesia pada jejaring sosial twitter menggunakan metode KNearest Neighbour. Data sampling pada sistem ini menggunakan 3.324 opinion tweet yang diproses dengan algoritma WIDF dalam pembobotan dan metode KNearest Neighbour dalam tahap klasifikasi opini. Hasil pengujian memperoleh range 3 pada algoritma Cosine Similarity menjadi nilai akurasi yang tinggi dengan nilai akurasi sebesar 86.67% opini positif dan 93.34% pada opini negative.

### III. METODOLOGI PENELITIAN

Tahapan metodologi penelitian pada penelitian ini adalah sebagai berikut :



Gambar 1. Model yang diusulkan

### 1. Dataset

Data yang digunakan pada penelitian ini diambil dari tweets pengguna twitter dengan menggunakan API twitter.

### 2. Data Preprocessing

Data yang didapat dari twitter di lakukan preprocessing untuk menghilangkan kata-kata yang dapat mengganggu proses klasifikasi.

### 3. Classified Naïve Bayes

Metode yang digunakan adalah metode Naïve bayes untuk pengklasifikasi data. Naïve bayes dipilih karena sangat sederhana, efisien dan merupakan teknik *machine learning* yang populer untuk klasifikasi teks. Dan memiliki performa yang baik masyarakat. Opini masyarakat dapat dilihat dari kata-kata yang digunakan dan dari *emoticon* yang dicantumkan pada tweet tersebut.

## II. HASIL DAN PEMBAHASAN

### A. *Data Set*

Tweet yang diambil dibatasi dari bulan November 2018 sampai Juni 2019 dimana pada range bulan tersebut merupakan puncak kenaikan harga tiket pesawat. Data tweets yang didapatkan dari twitter dengan menggunakan API twitter adalah sebanyak 651 tweets.

### B. *Data Preprocessing*

Data preprocessing merupakan tahap awal untuk megolah data teks menjadi analisa sentimen. Pada penelitian ini preprocessing yang dilakukan adalah

1. Stopword, yaitu menghilangkan kata tertentu yang tidak diperlukan dalam proses klasifikasi.
2. Menghapus URL
3. Melakukan Regexp, yaitu menghilangkan karakter selain huruf dan angka.

### C. *Data sebelum proses Preprocessing*

Data yang belum dilakukan proses preprocessing masih terdapat *noise* yang dapat mengganggu proses klasifikasi

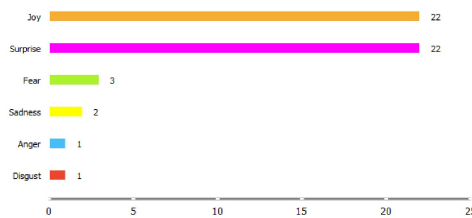
**D.** *Data setelah proses preprocessing*

Data yang sudah dilakukan proses preprocessing sudah siap untuk di lakukan klasifikasi.

**E.** *Classified Naïve Bayes*

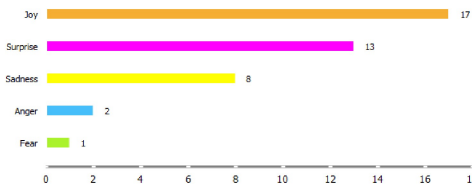
Klasifikasi adalah proses untuk menentukan sebuah kalimat sebagai anggota kelas positif, negative atau kelas netral. ini menggunakan data latih 90 tweets dan data uji sebanyak 45 tweet. Dari tweet tersebut dapat dilihat berbagai opini

**1) Klasifikasi tweet profiler pada data positif**



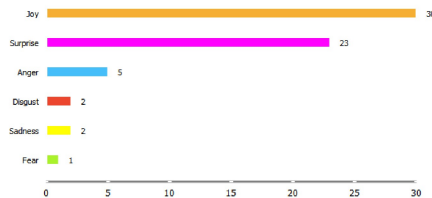
Gambar 2. Klasifikasi tweet profiler pada data positif

**2) Klasifikasi tweet profiler pada data negatif**



Gambar 3. Klasifikasi tweet profiler pada data negatif

**3) Klasifikasi tweet profiler pada data netral**



Gambar 4. Klasifikasi tweet profiler data netral

4) Hasil Presentase Tweet

Bedasarkan data tweet, penelitian yang dilakukan dapat diketahui bahwa nilai sentiment yang paling banyakterbentuk adalah sentiment neutral sebesar 62%.

Nilai sentiment negative sebesar 22%. Sentiment negative dapat diartikan bahwa masyarakat tidak menyetujui dengan kenaikan harga tiket pesawat. Dan sisanya 16% untuk sentiment positif. yang artinya masyarakat tidak mempermasalahkan dengan adanya kenaikan harga tiket pesawat.

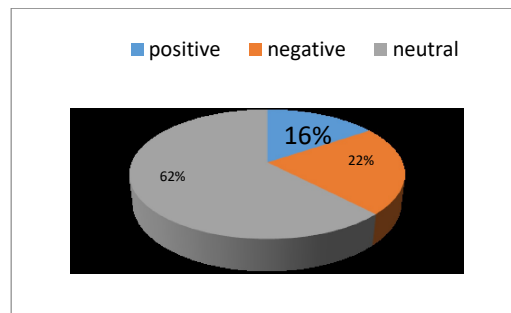


Diagram 1. Hasil Analisa Sentimen.

F. Evaluasi pengujian

Pengujian pada penelitian ini menggunakan Confusion matriks untuk menunjukkan seberapa baik model yang terbentuk. Accuracy menggunakan algoritma *Naïve bayes* sebesar 90,70%. Yang artinya dengan menggunakan algoritma *Naïve bayes* tingkat akurasi lebih besar dibandingkan menggunakan *K-Nearest Naighbour* yang digunakan pada penelitian sebelumnya. Akurasi *K-Nearest Naighbour* hanya sebesar 62,79%.

1) Hasil data latih dengan algoritma *Naïve Bayes*

accuracy: 100.00%				
	true neutral	true positive	true negative	class precision
pred. neutral	66	0	0	100.00%
pred. positive	0	11	0	100.00%
pred. negative	0	0	13	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 5. Data latih dengan algoritma *Naïve bayes*

2) Hasil Data Uji dengan Algoritma Naïve Bayes

accuracy: 90.70%

	true neutral	true negative	true positive	class precision
pred. neutral	27	2	2	87.10%
pred. negative	0	7	0	100.00%
pred. positive	0	0	5	100.00%
class recall	100.00%	77.78%	71.43%	

Gambar 6. Data uji dengan algoritma Naïve bayes

3) Hasil data latih dengan algoritma K-Nearest Neighbour

accuracy: 73.33%

	true neutral	true positive	true negative	class precision
pred. neutral	66	11	13	73.33%
pred. positive	0	0	0	0.00%
pred. negative	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

Gambar 7. Data latih dengan algoritma K-Nearest Neighbour

4) Hasil data latih dengan algoritma K-Nearest Neighbour

accuracy: 62.79%

	true neutral	true negative	true positive	class precision
pred. neutral	27	9	7	62.79%
pred. negative	0	0	0	0.00%
pred. positive	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

Gambar 8. Data uji dengan algoritma K-Nearest Neighbour

5) Pengujian model dengan confusion matriks

Tabel 1. Perbandingan Algoritma Naïve Bayes dengan Algoritma K-Nearest Neighbour.

Sentimen	Analisis	Naïve Bayes	KNN
Positive	7	5	0
Negative	10	7	0
Neutral	28	33	45
Accuracy		90,70%	62,79%

#### iv.KESIMPULAN

Berdasarkan hasil analisa tersebut, dapat disimpulkan bahwa Tweet kenaikan harga tiket pesawat dapat di klasifikasi dengan baik ke dalam bentuk positive dan negative. 62% masyarakat menganggap kenaikan harga tiket pesawat sebagai hal yang biasa saja, 16% memberikan tanggapan positive dan 22% memberikan tanggapa negative. Akurasi data latih Naïve Bayes yaitu sebesar 100% dan *K- Nearest Neighbor Classifier* sebesar 73,33 %. Untuk akurasi data uji Naïve Bayes yaitu sebesar 90,70% dan *K- Nearest Neighbor Classifier* sebesar 62,79 %. Yang artinya algoritma naïve bayes menghasilkan akurasi yang lebih tinggi di bandingkan KNN.

#### DAFTAR PUSTAKA

- Lee, M. (2010). M ULTICLASS S ENTIMENT A NALYSIS WITH RESTAURANT REVIEWS.
- Ye, Q., Zhang, Z., & Law, R. (2009). Expert Systems with Applications Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems With Applications*, 36(3), 6527–6535. K. Elissa, “Title of paper if known,” unpublished.
- Suprpto, Aji (2017) Sistem Klasifikasi Opini Pengguna Maskapai Penerbangan di Indonesia Pda Jejaring Sosial Twitter menggunakan Metode *K-Nearest Neighbour*. Bps.go.id
- Afshoh, Fauziah (2017) Analisa Sentimen menggunakan Naïve Bayes untk meliohat persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter. Skripsi. Informatika. Universitas Surakarta Muhammadiyah.
- witten, I.H (2005). Text Mining. Practical Handbook of internet computing, edited by M.P. Singh, pp. 14-1 -14-22. Chapman & Hall/CRC Press, Boca Raton, Florida.

