



# SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,  
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



## Informasi Pelaksanaan :

SNESTIK II - Surabaya, 26 Maret 2022

Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

## Informasi Artikel:

DOI : 10.31284/p.snestik.2022.2712

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya  
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043

Email : [snestik@itats.ac.id](mailto:snestik@itats.ac.id)

## Analisa Sentimen Mahasiswa Terhadap Minat Berorganisasi Dengan Metode *Naïve Bayes*

Aldio Nur Samsi<sup>1</sup>, Yulian Findawati<sup>2</sup>, Irwan A. Kautsar<sup>3</sup>

Universitas Muhammadiyah Sidoarjo<sup>1,2,3</sup>

*e-mail: aldio.nursamsi97@gmail.com*

### ABSTRACT

*Student organization is a collection of several students who have goals to be achieved and to accommodate student interests and talents. Student organizations vary, ranging from the Department of Student Association and the Student Executive Board. However, the number of Student Organizations is not matched with interest of students to join. And this research to analyze student sentiment by organizational interest using the naive Bayes method and using TF-IDF word weighting. Methods to perform the test are Confusion Matrix, Precision and Recall. By obtaining 3 document classifications, namely negative classification, positive classification and neutral classification. The document is shared for 2 subsets, namely training data and test data. Then it produces a confusion matrix with 23 data classified as positive, 59 data classified as negative and 51 data classified as neutral correctly. With values obtained positive precision of 60.5%, negative precision 64.8% and neutral precision 72.8%. Then for a positive recall value of 49.9%, a negative recall of 78.6% and a neutral recall of 68%.*

**Keywords:** *College student; Confusion Matrix; Naive Bayes; Organization; Precision; Recall; TF-IDF.*

### ABSTRAK

Organisasi mahasiswa ialah kumpulan dari beberapa mahasiswa yang memiliki tujuan yang akan dicapai serta untuk mawadahi minat dan bakat mahasiswa. Organisasi mahasiswa bermacam-macam, mulai dari Himpunan Mahasiswa Jurusan dan Badan Eksekutif Mahasiswa. Namun banyaknya Organisasi Mahasiswa tidak diimbangi dengan minat mahasiswa untuk bergabung. Dan penelitian ini bertujuan untuk melakukan analisa sentimen mahasiswa terhadap minat berorganisasi dengan metode *Naïve Bayes* dan menggunakan pembobotan kata TF-IDF. Metode untuk melakukan pengujian adalah *Confusion Matrix*, *Precision* dan *Recall*. Dengan memperoleh 3 klasifikasi dokumen yaitu klasifikasi negatif, klasifikasi positif dan

klasifikasi netral. Dokumen tersebut dibedakan 2 subset, yaitu data *training* dan data tes. Maka menghasilkan *confusion matrix* dengan 23 data terklasifikasi positif, 59 data terklasifikasi negatif dan 51 data terklasifikasi netral dengan benar. Dengan nilai yang diperoleh *Precision* positif sebesar 60,5%, *precision* negatif 64,8% dan *precision* netral 72,8%. Lalu untuk nilai *recall* positif sebesar 49,9%, *recall* negatif 78,6% dan *recall* netral sebesar 68%.

**Kata kunci:** *Confusion Matrix*; Mahasiswa; *Naïve Bayes*; Organisasi; *Precision*; *Recall*; TF-IDF.

## PENDAHULUAN

Organisasi mahasiswa ialah fasilitas untuk mahasiswa menambah ilmu pengetahuan, dan untuk membentuk karakter pada mahasiswa[1]. Namun banyak mahasiswa yang tidak memiliki minat untuk berorganisasi. Minat berorganisasi adalah suatu rasa keinginan pada mahasiswa untuk mengikuti kegiatan berorganisasi. Hal ini akan membuat mahasiswa untuk memberikan perhatian, keyakinan, ketertarikan dan keinginan untuk berorganisasi[2].

Penulis akan melakukan penelitian untuk mengetahui pendapat atau komentar dari mahasiswa tentang minat berorganisasi. Komentar tersebut akan diolah untuk menjadikan informasi berupa data teks. Penulis akan menggunakan metode *Naïve Bayes* untuk pengolahan data tersebut. Metode ini sebelumnya sudah digunakan oleh Fransiska Vina Sari pada tahun 2019 untuk menganalisa pelanggan toko online JD.ID dengan basis konversi ikon emosi. Dengan menghasilkan tingkat akurasi mencapai 96,44% , apabila dijumlahkan dengan TF-IDF maka akan meningkat menjadi 98% [3].

Penulis menggunakan pembobotan TF-IDF karena metode pembobotan ini memiliki perhitungan cukup mudah dan hasil yang akurat serta efisien. Berdasarkan latar belakang ini penulis ingin melakukan penelitian tentang “Analisa Sentimen Mahasiswa Terhadap Minat Berorganisasi Dengan Metode *Naïve Bayes*”. Penelitian ini menghasilkan 3 klasifikasi pada pendapat yaitu klasifikasi negatif, klasifikasi netral dan klasifikasi positif.

## METODE

### Pengumpulan data

Proses ini dilakukan menggunakan proses *crawling* data berupa komentar pada aplikasi *Twitter* dengan menggunakan RStudio dengan kata kunci mahasiswa, organisasi, berorganisasi, universitas, kampus. Data yang diperlukan adalah dokumen berformat csv. Untuk dokumen yang dimasukkan memiliki 2 kolom dan sudah diklasifikasi secara manual yaitu terdiri dari klasifikasi positif, negatif dan netral. Setelah itu dilakukan tahap *preprocessing* atau menyeleksi data lalu dipilah menjadi 2 bagian, yakni data *training* dan data tes.

### *Preprocessing*

Untuk mencegah data yang belum siap untuk diproses semacam data yang masih ditemukan tanda baca serta data yang belum menjadi kata baku[4]. Tahap *preprocessing* terdiri dari 4 bagian, yang pertama *Case Folding* adalah proses mengubah kata yang terdapat huruf besar diubah ke huruf kecil serta menghilangkan tanda baca. *Tokenizing* adalah proses pemenggalan kalimat menjadi string atau kata. *Filtering* ialah proses pengambilan kata yang dianggap penting dan akan digunakan untuk proses selanjutnya. *Stemming* yaitu merubah kata berimbuhan menjadi kata baku.

## Pembobotan kata

### Perhitungan TF-IDF

Metode TF-IDF ialah untuk memberi bobot setiap kata pada dokumen. TF-IDF merupakan sebuah penilaian statistik untuk mengevaluasi penting atau tidak pada sebuah kata di dalam sebuah dokumen atau sebuah kalimat [5]. Berikut adalah rumus TF-IDF yaitu:

$$W_{t,d} = t_{ft,d} \times id_{ft} = t_{ft,d} \times \log \frac{N}{d_{ft}} \quad (1)$$

Keterangan:

$W_{t,d}$  = nilai TF-IDF  
 $t_{ft,d}$  = kekerapan kata  
 $id_{ft}$  = kebalikan kekerapan dokumen setiap kata  
 $d_{ft}$  = kekerapan dokumen setiap kata  
 $N$  = jumlah dokumen

### Naïve Bayes Classifier

Naïve Bayes yakni metode pembelajaran klasifikasi yang mengabdikan aturan Bayes dengan taksiran jika atributnya independen bersyarat yang diberikan kelas[6]. Berikut persamaan umum dalam metode naïve bayes:

$$P(C_j|W_i) = \frac{P(C_j) \times P(W_i|C_j)}{P(W_i)} \quad (2)$$

Keterangan:

$P(C_j|W_i)$  = munculnya kesempatan bagian “j” ketika ada pengulangan kata “i”  
 $P(C_j)$  = kesempatan kehadiran dokumen pada bagian “j”  
 $P(W_i|C_j)$  = kesempatan sebuah kata “i” yang ada pada bagian “j”  
 $P(W_i)$  = kesempatan kehadiran sebuah kata  
 $i$  = indikator kata yang dimulai dari 1 sampai kata ke-k  
 $j$  = indikator bagian yang dimulai dari 1 sampai dengan bagian ke-n

### Probabilitas Kelas

Tahap ini merupakan Multinomial Naïve Bayes yaitu proses akhir dari proses data latih. Semua kata pada setiap kategori akan mendapat nilai probabilitas. Nilai tersebut akan dijadikan acuan untuk menghitung proses data uji.

$$P(w|c) = \frac{\text{count}(w,c) + 1}{\text{count}(c) + |V|} \quad (3)$$

Keterangan:

$\text{count}(w,c)$  = adanya kata “w” dari kategori “c”  
 $\text{count}(c)$  = adanya seluruh kata kategori “c”  
 $|V|$  = term unik atau fitur

## HASIL DAN PEMBAHASAN

### Text Preprocessing

Pada tahap ini digunakan untuk menghilangkan kata berimbuhan dan tanda baca. Serta menjadikan kata yang terdapat imbuhan menjadi kata dasar atau kata baku.

Tabel 1. *Pre-processing*

Data	Setelah di preprocess	Klasifikasi
Pengalaman berorganisasi gue apa ya???	alam organisasi	Netral
Memperluas jaringan yang penting untuk kelancaran karir kamu di masa depan..	luas jaringan karir	Positif

Sumber: Dokumen yang diambil dari komentar Twitter

Pada kata “pengalaman berorganisasi gue apa ya???” tentukan klasifikasinya, lalu diubah menjadi kata”alam” karena menghilangkan kata (gue, apa, ya) serta menghilangkan tanda baca “???”. Lalu untuk kata “pengalaman” dan “berorganisasi” diubah menjadi kata “alam” dan “organisasi”.

### Perhitungan TF-IDF

Tahap ini (1) berfungsi untuk pembobotan pada tiap kata dengan cara mengkalikan TF dengan IDF, maka mendapatkan hasil seperti pada gambar berikut:

Kata	TF	DF	TF-IDF
beri	1	1	1.62325
wewenang	1	1	1.62325
tiap	1	2	1.32222
kelompok	1	1	1.62325

Gambar 1. Hasil dari TF-IDF

Sumber: Dokumen pribadi dari perhitungan aplikasi

Pada Gambar 1 kata “beri” memiliki nilai TF sebesar 1 dan nilai DF ada 1 dengan jumlah TF-IDF sebesar 1,62325. Hasil perhitungan tersebut menggunakan jumlah total dokumen sebanyak 140 data.

### Perhitungan Probabilitas Kelas

Di bagian Tahap (3) akan mendapatkan hasil seperti berikut:

Jumlah Seluruh Term Unik = 270  
 Jumlah Term Pada Kategori Positif = 185  
 Jumlah Term Pada Kategori Negatif = 102  
 Jumlah Term Pada Kategori Netral = 122

Kata	Probabilitas Positif	Probabilitas Negatif	Probabilitas Netral
beri	0.00577	0.00269	0.00255
wewenang	0.00577	0.00269	0.00255
tiap	0.00801	0.00269	0.00255
kelompok	0.00577	0.00269	0.00255

Gambar 3. Perhitungan Probabilitas Kelas

Sumber: Dokumen pribadi dari perhitungan aplikasi

### Confusion Matrix

Digunakan untuk pengukuran nilai kebenaran (*true*) dan nilai kesalahan (*false*) pada data tes ke dalam matrix. Berikut tampilan *Confusion Matrix* :

Prediksi	Kategori		
	Positif	Negatif	Netral
Positif	23	20	17
Negatif	10	6	4
Netral	7	6	2

Gambar 4. Confusion Matrix

Sumber: Dokumen pribadi dari perhitungan aplikasi

Dari Gambar 4 menghasilkan *confusion matrix* pada kategori *true* positif sebanyak 23, klasifikasi *true* negatif sebanyak 6 kata dan klasifikasi *true* netral sebanyak 2 kata.

### Precision dan Recall

*Precision* untuk perbandingan antara kebenaran positif dengan banyaknya data yang diprediksi positif, lalu *Recall* untuk perbandingan antara kebenaran positif dengan banyaknya data yang sebenarnya positif. Berikut tampilan untuk hasil *precision* dan *recall*:

	Precision	Recall
Positif	38.33333333333333	57.5
Negatif	30	18.75
Netral	13.33333333333333	8.695652173913
Akurasi	8.9337175792507	

Gambar 5. Hasil dari *Precision* dan Hasil dari *Recall*

Sumber: Dokumen pribadi dari perhitungan aplikasi

Pada Gambar 5 *precision* positif sebesar 38,33%, *precision* negatif 30% dan *precision* netral 13,33%. Lalu untuk nilai *recall* positif sebesar 57,5%, *recall* negatif 18,75% dan *recall* netral sebesar 8,69%. Sehingga menghasilkan akurasi sebesar 8,93%.

## KESIMPULAN

Pada penelitian ini dengan mengambil dari komentar pada *Twitter* maka mendapatkan 140 data yang terbagi menjadi 40 data training dan 100 data tes sehingga menjadi 3 kategori yaitu kategori positif, kategori negatif dan netral. Dari data tersebut terdapat *confusion matrix*, *precision* dan *recall*, sehingga menghasilkan akurasi sebesar 8,93%.

## DAFTAR PUSTAKA

- [1] S. Rivaldi, "PROGRAM STUDI PENDIDIKAN EKONOMI JURUSAN PENDIDIKAN ILMU PENGETAHUAN SOSIAL FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN UNIVERSITAS TANJUNGPURA PONTIANAK 2013," p. 9.
- [2] . N. M. K. U., . Dr. K. A., and . Dr. D. G. H. D., "Hubungan Motivasi Berprestasi dan Minat Berorganisasi Terhadap Indeks Prestasi Belajar Mahasiswa Pada Jurusan Pendidikan Teknik Informatika," *Kumpul. Artik. Mhs. Pendidik. Tek. Inform. KARMAPATI*, vol. 6, no. 2, p. 267, May 2017, doi: 10.23887/karmapati.v6i2.10112.

- 
- [3] F. V. Sari and A. Wibowo, “ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI,” vol. 10, no. 2, p. 6, 2019.
- [4] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, Aug. 2019, doi: 10.52985/insyst.v1i1.36.
- [5] V. Amrizal, “PENERAPAN METODE TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DAN COSINE SIMILARITY PADA SISTEM TEMU KEMBALI INFORMASI UNTUK MENGETAHUI SYARAH HADITS BERBASIS WEB (STUDI KASUS: HADITS SHAHIH BUKHARI-MUSLIM),” *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, Nov. 2018, doi: 10.15408/jti.v11i2.8623.
- [6] G. I. Webb, “Naïve Bayes,” in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2016, pp. 1–2. doi: 10.1007/978-1-4899-7502-7\_581-1.