

# ANALISIS PENDAPAT PUBLIK TERHADAP *PUBLIC FIGURE* DENGAN MENGGUNAKAN METODE NAIVE BAYES

Januar Mansur<sup>1)</sup> Viny Christanti M.<sup>2)</sup> Tri Sutrisno<sup>3)</sup>

<sup>1,2,3)</sup> Teknik Informatika Universitas Tarumangara  
Jl. Let. Jend. S. Parman No.1, Jakarta 11440 Indonesia  
januarmansur@yahoo.com<sup>1)</sup>, viny@untar.ac.id<sup>2)</sup>, tris@fti.untar.ac.id<sup>3)</sup>

## ABSTRACT

*Public Figure is a talented idol figure. For example, such as sportsmen, music players or film actors. Because it is the focus of the community, the behavior of a public figure is interesting to comment on. This is supported by the use of social media by Indonesian people such as Twitter. On the public figure page, existing comments can be analyzed to find out the sentiments of the people who are classified as positive, negative, and neutral. For a small amount of data, the comment classification process can be done manually, but if the data is too much it requires a system equipped with classification methods, so classification can be done quickly. The application of this classification will use a diagram feature, when the user types the public figure name the desired one will output the classification with a diagram. In determining the classification results will use the Naïve Bayes classification method. This application uses 6 training data classes of 2104 data with results of the system produced an accuracy of 96.47% and 6 training data balance of classes of 750 data with the result of the system produced an accuracy of 87.97%.*

## Key words

*Analisis sentimen, Naive Bayes, Public Figure, Twitter*

## 1. Pendahuluan

Perkembangan dunia teknologi informasi dan komunikasi yang pesat tidak terlepas dari penyedia layanan web yang menyediakan informasi yang beragam. Informasi yang menyebabkan penambahan data yang kebanyakan berupa data teks dapat dijadikan sumber yang sangat potensial untuk digali lebih dalam. Salah satu contohnya adalah data text yang diambil dari twitter.

Twitter adalah layanan jejaring sosial dan microblog daring yang memungkinkan penggunaanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan tweet. Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di

internet, dan dijuluki dengan pesan singkat dari internet . Di twitter, pengguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menulis kicauan melalui antarmuka situs web, pesan singkat (SMS) atau melalui berbagai aplikasi untuk perangkat seluler [1].

Pendapat publik adalah pendapat, ide atau pikiran untuk menjelaskan kecenderungan atau preferensi tertentu terhadap perspektif dan ideologi akan tetapi bersifat tidak objektif karena belum mendapatkan pemastian atau pengujian, dapat pula merupakan sebuah pernyataan tentang sesuatu yang berlaku pada masa depan dan kebenaran atau kesalahannya serta tidak dapat langsung ditentukan misalnya menurut pembuktian melalui induksi [2].

Analisis Sentimen adalah riset komputasional dari opini, sentiment dan emosi yang diekspresikan secara tekstual. Untuk melakukan analisis sentimen ada beberapa metode yang dapat digunakan dan salah satu nya adalah metode Naïve Bayes. Naive Bayes classifier merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan dengan ciri utamanya adalah asumsi yang sangat kuat (naif) akan ketergantungan dari masing-masing kondisi/kejadian.[3]

Penelitian dalam bidang analisis sentimen ini banyak dilakukan karena topik ini sangat menarik untuk dibahas. Beberapa penelitian yang pernah dibuat adalah penelitian opinion mining untuk ulasan produk dengan klasifikasi naïve bayes oleh Albert Jeremy dengan akurasi sebesar 77% [4], *Opinion Extraction of Public Figure Based on Sentiment Analysis from twitter* oleh Nur hayatin dengan akurasi sebesar 76.67% [5].

## 2. Dasar Teori

### 2.1 *Public Figure*

*Public Figure* adalah kata yang berasal dari bahasa Inggris yang dalam terjemahannya memiliki arti sosok atau tokoh yang dikenal secara luas oleh masyarakat umum. Namun di Indonesia sendiri, penggunaan kata sosok atau tokoh masyarakat lebih cenderung dipakai untuk menggambarkan atau menyebut pejabat, pemuka adat, kepala suku atau orang-orang dari instansi pemerintahan. Sedangkan kata *public figure* sendiri

cenderung digunakan oleh masyarakat sebagai sebutan kepada artis, penyanyi, pesinetron dan mereka yang sering muncul di layar kaca[6].

## 2.2 Analisis Sentimen

Informasi tekstual secara umum dapat dibagi menjadi informasi fakta dan opini[7]. Fakta adalah ekspresi obyektif terhadap suatu benda, kejadian dan kepunyaan benda tersebut. Opini biasanya berupa ekspresi subyektif yang menggambarkan sentimen, penilaian, atau perasaan seseorang terhadap suatu benda, kejadian atau kepunyaan dari benda tersebut. Menjelaskan bahwa analisis sentimen adalah bagian dari pekerjaan yang meninjau segala sesuatu berhubungan dengan pendapat komputasi, sentimen dan subjektivitas teks. Ditambahkan oleh bahwa analisis sentimen adalah alat untuk memproses koleksi hasil pencarian yang bertujuan dengan mencari atribut suatu produk (kualitas, fitur, dll) dan proses memperoleh hasil pendapatnya.

Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapat yang dikemukakan dalam dokumen bersifat positif, negatif atau netral. Penelitian mengenai analisis sentimen telah berkembang sejak tahun 2003 dan merupakan bagian dari text mining yang merupakan penelitian komputasi berdasarkan sentimen, emoticon, pendapat, komentar dan setiap ekspresi yang diungkapkan oleh teks. Analisis sentimen difokuskan untuk review klasifikasi berdasarkan polaritas. Berdasarkan klasifikasi, analisis sentimen dibagi menjadi dua kelompok utama. Yaitu dokumen klasifikasi ke pendapat atau fakta, atau dikenal sebagai klasifikasi subjektivitas (*subjectivity classification*) dan dokumen klasifikasi ke dalam positif atau negatif, atau dikenal sebagai analisis sentimen. Hal ini adalah proses yang penting untuk menentukan dokumen yang memiliki opini dan dokumen yang menyimpulkan opini bernilai positif, negatif maupun netral.

## 2.3 Preprocessing

*Preprocessing* adalah tahap proses awal *text mining* terhadap teks untuk mempersiapkan teks menjadi data yang dapat diolah lebih lanjut[8]. Sekumpulan karakter yang bersambung (teks) harus dipecah-pecah menjadi unsur yang lebih berarti. Suatu dokumen dapat dipecah menjadi bab, sub-bab, paragraf, kalimat, kata dan bahkan suku kata. Adapun tahapan *preprocessing* yaitu: *case folding* untuk mengubah semua huruf dokumen menjadi huruf kecil, *filtering* untuk mengambil kata penting dengan membuang *stopword*, dan *stemming* untuk mendapat kata dasar.

## 2.4 Visualisasi IR

Visualisasi merupakan metode komputasi yang mengubah data simbolik menjadi data geometric,

memungkinkan peneliti untuk mengobservasi hasil penelitian dalam tampilan yang lebih kaya informasi. Visualisasi merupakan proses mengubah data, informasi, dan pengetahuan menjadi presentasi grafik untuk mendukung pekerjaan-pekerjaan seperti analisis, memprediksi, pencarian pola, dan lain-lain[9].

Berikut adalah beberapa visualisasi yang umum pada IR:

### 1. Histogram

Histogram merupakan representasi grafik dari distribusi frekuensi dalam bentuk diagram batang. Histogram memiliki variasi tampilan tampilan diagram, beberapa diagram diantaranya adalah 2 dimensi dan 3 dimensi. Tampilan ini terdiri dari kumpulan persegi panjang dengan panjang yang berbeda-beda. Semakin besar frekuensi suatu kelas, ukuran persegi panjangnya akan semakin tinggi.

### 2. Pie Chart

*Pie Chart* merupakan grafik yang paling dasar dan mudah dikenali. Tampilan ini biasa digunakan untuk menampilkan persentase kelas dalam rupa grafik berbentuk pie. *Pie Chart* memiliki informasi yang mudah dipahami karena ukuran potongan grafik bergantung pada presentase yang ada.

## 2.4 Klasifikasi Data

Klasifikasi merupakan salah satu teknik atau metode dalam data mining dan metode klasifikasi termasuk dalam jenis *supervised learning*. *Supervised learning* adalah *meachine learning* yang membutuhkan label sebagai tujuan dari pelatihan data (*data training*).

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data ke dalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu.

Klasifikasi yang digunakan untuk proses penyelesaian masalah sentimen analisis adalah dengan menggunakan *text processing*, dengan menggunakan sumber data yang berasal dari twitter API (*Application Programming Interfaces*). Proses pengambilan data komentar yang berada di website menggunakan teknik *scrapping*. Teknik *scrapping* merupakan metode ekstraksi data teks yang berada pada halaman website twitter API. Setelah data berhasil didapatkan, maka data tersebut akan diproses dengan tahap *stemming* dan akan langsung masuk ke proses *labeling* kata untuk menjadi data training.

Klasifikasi terdiri dari dua tahap atau langkah proses. Tahap pertama adalah *learning* (fase training), yaitu dimana algoritma klasifikasi dibuat untuk menganalisis data training atau data latih lalu direpresentasikan kedalam bentuk rule atau aturan klasifikasi. Sedangkan proses yang kedua adalah proses klasifikasi, yaitu data tes yang digunakan untuk memperkirakan akurasi dari aturan

klasifikasi yang telah terbentuk. Pada proses klasifikasi, dipengaruhi oleh empat komponen:

1. Class Label  
Variable Dependent yang bertipe kategorikal yang mempresentasikan label yang terdapat pada objek.
2. Data Training  
Suatu set data yang berisi nilai dari kedua komponen diatas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
3. Data Testing  
Berisi data baru yang akan diklasifikasikan oleh model yang telah terbuat dan akurasi klasifikasi dievaluasi.

## 2.5 Naïve Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [10]. Naive Bayes merupakan metode yang digunakan dalam mengklasifikasi dokumen. Metode klasifikasi Naive berperan dalam menyederhanakan pengklasifikasian dokumen dengan bekerja melalui dasar asumsi bahwa atribut-atribut yang digunakan bersifat conditionally independent antara satu dan yang lainnya, dengan kata lain dalam setiap kategori, setiap kata independent satu sama lain.

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

Persamaan dari teorema Bayes adalah:

Teorema Bayes berawal dari persamaan 1, yaitu:

$$P(A|B) = \frac{P(B \cap A)}{P(B)} \quad \dots (1)$$

Dengan  $P(A|B)$  artinya peluang A jika diketahui keadaan B. Kemudian dari persamaan 1 didapatkan persamaan 2.

$$P(B \cap A) = P(B|A) \cdot P(A) \quad \dots (2)$$

Sehingga didapatkan teorema bayes seperti persamaan yang ditunjukkan pada persamaan 3.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad \dots (3)$$

Pada pengklasifikasian teks, perhitungan persamaan dapat didefinisikan:

$$P(v_j) = \frac{docs_j}{|D|} \quad \dots (4)$$

$$P(w_k|v_j) = \frac{n_k+1}{n+|kata|} \quad \dots (5)$$

Keterangan:

1.  $Docs_j$  : kumpulan dokumen yang memiliki kategori  $v_j$ .
2.  $|D|$  : jumlah dokumen yang digunakan dalam pelatihan (kumpulan data latih).
3.  $n$  : jumlah total kata yang terdapat di dalam kata tekstual yang memiliki nilai fungsi target yang sesuai.
4.  $n_k$  : jumlah kemunculan kata  $w_k$  pada semua data tekstual yang memiliki nilai fungsi target yang sesuai.
5.  $|kata|$ : jumlah kata yang berbeda yang muncul dalam seluruh data tekstual yang digunakan.

Selanjutnya menggunakan rumus *argument* maksimum untuk menemukan nilai maksimum dari hasilnya akan mencapai proses dari klasifikasi tersebut.

$$V_{map} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i|V_j) \quad \dots (6)$$

$V_{map}$  adalah nilai probabilitas hasil perhitungan Naive Bayes. Untuk nilai fungsi target yang bersangkutan. Frekuensi kemunculan kata menjadi dasar perhitungan nilai dari  $P(v_j)$  dan  $P(a_i | v_j)$ .

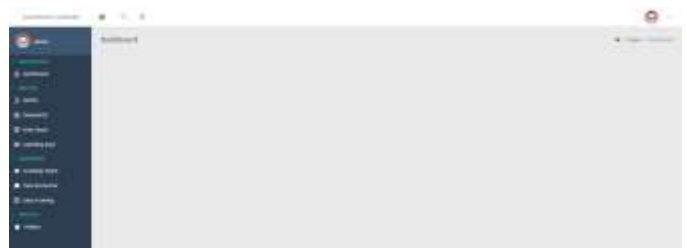
## 2.6 Akurasi

Akurasi dibutuhkan untuk menentukan performa dari system. *Confusion Matrix* adalah alat test untuk menganalisis sistem klasifikasi. Evaluasi yang dilakukan dengan menggunakan tabel *confusion matrix* untuk melakukan perbandingan antara kategori actual dan prediksi. Variabel dalam *confusion matrix* memiliki arti sebagai berikut TP (*True Positive*), FP (*False positive*), FN (*False Negative*), TN (*True Negative*).

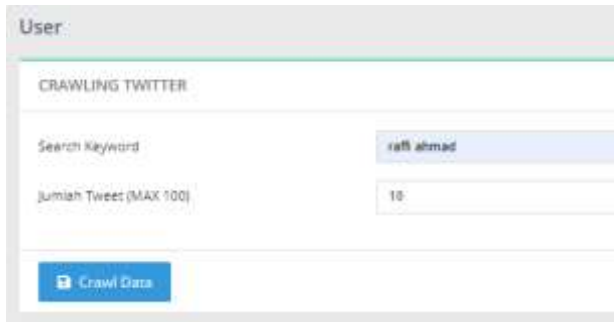
$$AKURASI = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad \dots (7)$$

## 3. Pembuatan

Pembahasan modul interface akan membahas tentang modul-modul yang berada di dalam program. Di dalam program terdapat 10 modul yaitu modul dashboard, modul admin, modul stopwords, modul kata dasar, modul labelling kata, modul crawling tweet admin, modul data komentar, modul data training, modul analisa, modul user analisa. Berikut akan membahas tentang modul-modul utama yang akan dipakai:



Gambar 1. Halaman utama analisis *public figure*



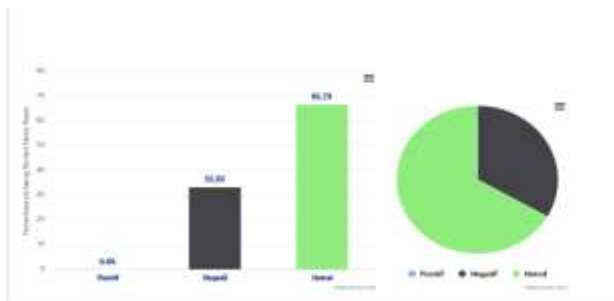
Gambar 2. Ketik nama *public figure*

Modul yang berada di Gambar 1 membahas tentang cara mengambil data testing dari twitter, pengguna hanya perlu memasukan nama *public figure* dan data test yang ingin diambil.



Gambar 3. Hasil yang didapat

Modul yang berada di gambar 2 membahas setelah proses pengambilan data testing, maka akan keluar klasifikasi total data positif, negatif, dan netral dari hasil klasifikasi. Hasil dari hasil analisis rafi ahmad terdapat 6 data testing yang masuk dalam pengklasifikasian 4 data netral dan 2 data negatif sehingga sentimen akan masuk ke dalam sentiment netral.



Gambar 4. Hasil Grafik

Dari hasil grafik didapatkan bahwa rafi ahmad mempunyai sentiment netral sebesar 66.67% dan negatif sebesar 33.3% dari 6 data tweet testing.

#### 4.1 Pengujian

Tabel 1 Jumlah Data tidak *balance* Label Per-Kelas ke -1

Kelas	Jumlah
Positif	648
Negatif	435
Netral	1021
Total	2104

Tabel 1 menunjukkan bahwa terdapat 648 data positif, 435 data negatif, dan 1021 data netral. Pengujian ini akan termasuk dalam pengujian data tidak *balance*.

Tabel 2 Jumlah Data *balance* Label Per-Kelas ke -2

Kelas	Jumlah
Positif	250
Negatif	250
Netral	250
Total	750

Tabel 2 menunjukkan bahwa terdapat 250 data positif, 250 data negatif, dan 250 data netral. Pengujian ini akan termasuk dalam pengujian data *balance*.

Untuk Pengujian *confusion matrix* akan menggunakan semua data dan akan diuji sebanyak enam kali untuk data training dengan 2104 data dan enam kali dengan data 750 data. Untuk tabel pengujian *Confusion Matrix* dapat dilihat pada Tabel 3.

Tabel 3 Tabel Pengujian data tidak *balance*

L	Data Training	Data testing	Proporsi Pembagian Data
1	1683	421	80% / 20%
2	1472	632	70% / 30%
3	1262	842	60% / 40%
4	2104	421	100 / 20%
5	2104	632	100 / 30%
6	2104	842	100 / 40%

Dari tabel pengujian dapat dilihat bahwa terdapat 6 data pengujian. Tabel 3 untuk data pengujian tidak *balance*, dengan pembagian proporsi data yang berbeda-beda untuk menemukan akurasi yang paling baik. Pada data satu sampai ketiga akan memisah data *training* dan data *testing*, untuk data ke-empat sampai ke-enam data akan ditraining 100% dan data *testing* akan diambil random dari data *training*.

Tabel 4 Tabel Pengujian data *balance*

L	Data Training	Data testing	Proporsi Pembagian Data
1	600	150	80% / 20%
2	525	225	70% / 30%
3	450	300	60% / 40%
4	750	150	100 / 20%
5	750	225	100 / 30%
6	750	300	100 / 40%

Dari tabel pengujian dapat dilihat bahwa terdapat 6 data pengujian. Tabel 3 untuk data pengujian *balance*, dengan pembagian proporsi data yang berbeda-beda untuk menemukan akurasi yang paling baik. Pada data satu sampai ketiga akan memisah data *training* dan data *testing*, untuk data keempat sampai ke-enam data akan ditraining 100% dan data *testing* akan diambil random dari data *training*.

### 4.2 Hasil Pembahasan Pengujian

Tabel 5 Tabel Hasil Pengujian *Confusion Matrix* data tidak *balance*

Pengujian Ke	Hasil Akurasi
1	57%
2	58.22%
3	57.24%
4	86.93%
5	87.97%
6	87.66%

Dari ke-enam hasil pengujian yang dilakukan didapatkan hasil akurasi terbesar berada di percobaan kelima dengan akurasi 87.97% dan akurasi terkecil berada pada percobaan pertama dengan hasil akurasi 57%.

Tabel 6 Tabel Hasil Pengujian *Confusion Matrix* data *balance*

Pengujian Ke	Hasil Akurasi
1	46%
2	57.88%
3	59%
4	94.67
5	95.11
6	96.47%

Dari ke-enam hasil pengujian yang dilakukan didapatkan hasil akurasi terbesar berada di percobaan kelima dengan akurasi 96.47% dan akurasi terkecil berada pada percobaan pertama dengan hasil akurasi 46%.

Tabel 7 Hasil Pengujian *Confusion Matrix* ke-5 Data Tidak *Balance*

		Kelas Prediksi			
		L1	L2	L3	Jumlah
Kelas sebenarnya	L1	134	40	9	183
	L2	0	263	24	287
	L3	0	3	159	162
					632

Keterangan :  
 L1 : positif  
 L2 : negatif  
 L3 : netral

Tabel 7 menampilkan hasil perhitungan *confusion matrix* ke-lima dengan data tidak *balance* dengan pembagian proporsi data 2104 data training dan 632 data testing. Di kolom L1 yang merupakan data positif tidak terjadi kesalahan, *confusion matrix* berhasil mengklasifikasikan 134 data positif, 40 data positif yang masuk ke negatif, dan 9 data positif yang masuk ke netral. Di kolom L2 terdapat 0 data negatif yang masuk ke positif, 263 data negatif, dan 24 data negatif yang masuk ke netral. Di kolom L3 terdapat 0 data netral yang masuk ke positif, 3 data netral masuk ke negatif, dan 159 data netral.

Tabel 8 Hasil Pengujian *Confusion Matrix* ke-6 data *balance*

		Kelas Prediksi			
		L1	L2	L3	Jumlah
Kelas sebenarnya	L1	97	0	0	97
	L2	5	99	1	105
	L3	8	2	88	98
					300

Keterangan :  
 L1 : positif  
 L2 : negatif  
 L3 : netral

Tabel 8 menampilkan hasil perhitungan *confusion matrix* ke-lima dengan data *balance* dengan pembagian proporsi data 750 data training dan 300 data testing. Di kolom L1 terdapat 97 data positif dan tidak ada kesalahan dalam perhitungan data L1. Di kolom L2 terdapat 5 data negatif yang masuk ke dalam positif, 99 data negatif yang masuk ke dalam positif, dan 1 data negatif yang masuk ke netral. Di kolom L3 terdapat 8 data netral yang masuk ke dalam positif, 2 data netral yang masuk ke negatif, dan 88 data netral yang masuk ke netral.

### 5. Kesimpulan

Kesimpulan yang dapat diperoleh berdasarkan pembuatan dan pengujian dari aplikasi ini adalah sebagai berikut:

1. Pembuatan skripsi menghasilkan website analisis yang digunakan untuk menganalisis *public figure* berdasarkan data dari twitter dengan menampilkan visualisasi berupa histogram dan *pie chart*.
2. Persentase klasifikasi analisis yang paling efektif diantara enam pengujian data tidak *balance* ada pada pengujian ke-lima dengan akurasi sebesar 87.97% dengan data latih sebesar 750 dan data testing sebesar 300. Diantara enam pengujian data *balance* akurasi terbesar berada di pengujian ke 6 dengan akurasi 96.47%.
3. Data testing dari twitter tidak sesuai dengan jumlah permintaan sehingga data yang diterima tidak sesuai dengan data yang diterima.

Saran yang dapat ditambahkan untuk pengembangan aplikasi ini adalah perlunya ditambahkan penjelasan tentang visualisasi IR, menggunakan data latih yang lebih bervariasi agar dapat menghasilkan klasifikasi yang lebih akurat terhadap sentiment *public figure*.

## REFERENSI

- [1] D'Monte, Leslie. "Swine flu's tweet tweet causes online flutter." *Business Standard* 29 (2009).
- [2] Rish, Irina. "An empirical study of the naive Bayes classifier." *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. No. 22. 2001.
- [3] Hayatin, Nur, Mustika Mentari, and Abidatul Izzah. "Opinion Extraction of Public Figure Based on Sentiment Analysis from Twitter." *IPTEK The Journal of Engineering* 1.1 (2014).
- [4] Jeremy, Albert, Viny Christanti, and Bagus Mulyawan. "Opinion Mining Untuk Ulasan Produk Dengan Klasifikasi Naive Bayes." *Jurnal Ilmu Komputer dan Sistem Informasi* 6.1 (2018): 9.
- [5] Kuku, Widyatmoko. "Dramaturgi Kalangan *Public Figure*." <http://www.kabarindonesia.com/berita.php?pil=20&jd=Dramaturgi+Kalangan+Public+Figure&dn=20091021103217>, 15 Juli 2019.
- [6] Liu, Bing. "Opinion mining and sentiment analysis." *Web Data Mining*. Springer, Berlin, Heidelberg, 2011. 459-526.
- [7] Aditya, Bayu Rima. "Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining." *Jurnal Infotel* 7.2 (2015): 93-100.
- [8] Raharja, Yosoa Putra. "Rancang Bangun Sistem Rekomendasi Beasiswa Menggunakan Algoritma Klasifikasi C4. 5 Pada Universitas Dian Nuswantoro." *Universitas Dian Nuswantoro, Semarang, skripsi* (2014).
- [9] Nagel, Wolfgang E., et al. "VAMPIR: Visualization and analysis of MPI resources." (1996).
- [10] Patil, Tina R., and S. S. Sherekar. "Performance analysis of Naive Bayes and J48 classification algorithm for data classification." *International journal of computer science and applications* 6.2 (2013): 256-261.

**Januar Mansur**, seorang mahasiswa pada program studi Fakultas Teknologi Informasi di Universitas Tarumanagara

**Viny Christanti Mawardi**, memperoleh gelar S.Kom. dari Universitas Tarumanagara, Indonesia tahun 2004. Kemudian, tahun 2008 memperoleh gelar M.Kom. Dari Universitas Indonesia, Indonesia. Saat ini sebagai staf pengajar program studi Teknik Informatika, Universitas Tarumanagara.

**Tri Sutrisno**, memperoleh gelar S.Si dari Universitas Diponegoro, Indonesia tahun 2011. Kemudian tahun 2015 memperoleh M.Sc. dari Universitas Gadjah Mada, Indonesia. Saat ini sebagai Staf Pengajar Program studi Teknik Informatika, Universitas Tarumanagara