

# PERANCANGAN ALAT BANTU BRAND MONITORING DI SITUS JEJARING SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER DAN MUTUAL INFORMATION

Apriadhie Purnama Ramadhan<sup>1</sup>, Z.k. Abdurahman Baizal<sup>2, 3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

---

## Abstrak

Brand Monitoring pada Microblogging seperti twitter adalah salah satu cara yang dapat digunakan guna mengetahui pendapat masyarakat terhadap sebuah barang atau jasa yang diberikan oleh perusahaan. Brand monitoring dapat dilakukan dengan cara melihat opini positif maupun negatif yang diberikan masyarakat. Permasalahan pengklasifikasian pendapat ke dalam opini negatif dan positif dapat diselesaikan dengan menggunakan algoritma Naive Bayes Classifier.

Naive Bayes Classifier adalah salah satu contoh algoritma pengklasifikasian. Dengan menggabungkan feature selection Mutual Information diharapkan algoritma ini dapat menghasilkan akurasi yang baik. Pada Naive Bayes Classifier terjadi dua proses yaitu proses latihan dan tes. Proses latihan dilakukan terhadap sebagian data sample, dari pelatihan ini kan didapatkan sebuah model probabilitas, yang nantinya model ini akan digunakan oleh data pada proses tes.

Pada tugas akhir ini dilakukan perbandingan nilai akurasi diantara kedua metode yaitu dengan menggunakan MI dan tanpa menggunakan MI. Hasil percobaan menunjukkan bahwa metode dengan menggunakan Mutual Information memiliki akurasi yang lebih baik dengan nilai rata-rata akurasi 84,5%.

Kata Kunci : brand monitoring, feature selection, mutual information, naive bayes, klasifikasi, opini.

---

## Abstract

Brand Monitoring on Microblogging like twitter is one way that can be used to determine public opinion against the goods or services supplied by the company. Brand monitoring can be done by looking at the positive and negative opinions given by the society. Problems classification opinion into positive and negative opinions can be solved by using a Naive Bayes classifier algorithm Multinomial.

Naive bayes Classifier is one example of classification algorithms. By combining feature selection Mutual Information, this algorithm is expected to produce a good accuracy. There are two processes in Naive bayes Classifier happened, they are the process of training and testing. The process of training is applied to some of the data samples, from the training, a probability model can be defined, which is later, the model will be used by the data on the test.

In this final threshold value comparison on the comparison method by using Mutual Information and without the use of Mutual Information. And showed that the method using the Mutual Information has better accuracy with value 84,5%.

Keywords : brand monitoring, feature selection, mutual information, naive bayes, classification, opinions.

---

# 1 PENDAHULUAN

## 1.1 Latar Belakang Masalah

Kemudahan masyarakat mengakses internet berdampak pada meningkatnya pengguna media sosial khususnya *twitter*. *Twitter* adalah salah satu contoh media sosial yang biasa digunakan untuk ‘berbagi’ pengalaman atau pendapat seseorang terhadap apa yang mereka rasakan. Tidak jarang sebuah akun *twitter* memberikan pendapatnya tentang sebuah produk atau *brand*, baik berupa pendapat positif maupun negatif. Hal ini memberikan peluang kepada perusahaan maupun *costumer* untuk mengetahui ‘tren’ yang sedang terjadi pada sebuah produk maupun pelayanan. Bagi perusahaan dengan mengetahui opini masyarakat tentang produk mereka, perusahaan dapat mengatur strategi yang tepat agar produk mereka tetap baik bahkan menjadi lebih baik. Sedangkan bagi pelanggan yang ingin membeli sebuah produk tertentu, mereka dapat melihat bagaimana performansi produk tersebut secara ‘*real*’ dari pelanggan lain yang sudah menggunakan produk tersebut.

Pada tugas akhir ini penulis akan mengambil salah satu *provider* telekomunikasi di Indonesia yaitu PT Indosat Tbk sebagai contoh kasus *opinion mining* di *twitter*. PT Indosat Tbk adalah sebuah perusahaan yang bergerak dilayanan penyedia layanan telekomunikasi. Produk layanan yang ditawarkan Indosat sangat beragam dan banyak dikenal masyarakat luas. Dalam penelitian ini penulis akan melihat bagaimana tanggapan masyarakat terhadap layanan indosat khususnya di *twitter*.

Salah satu metode yang bisa digunakan untuk implementasi *brand monitoring* adalah *sentiment analysis*. *Sentiment analysis* adalah pengolahan data berupa teks dan kemudian menganalisa isi dari dokumen tersebut. Dalam pengerjaannya penulis akan menggunakan algoritma *Naïve Bayes*. *Multinomial Naïve Bayes* sangat sering digunakan dalam pengklasifikasian teks karena sangat cepat dan mudah untuk diimplementasikan [9] dan dipadukan dengan *feature selection Mutual Information*, karena menurut penelitian sebelumnya[4] gabungan kedua metode ini merupakan metode yang paling baik untuk kasus *sentiment analysis* dengan nilai akurasi mencapai 84%.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan diatas, terdapat beberapa hal yang menjadi masalah dalam pembuatan tugas akhir ini, seperti:

1. Bagaimana cara mengimplementasikan *Naïve Bayes Classifier* dan *Mutual Information* untuk mengkategorikan opini.
2. Bagaimana cara memvisualisasikan hasil dari klasifikasi.
3. Bagaimana mengetahui performansi *Naïve Bayes Classifier* untuk memberikan hasil optimum pada klasifikasi.

Dalam menyelesaikan tugas akhir ini ada beberapa batasan atau ruang lingkup, yaitu:

1. Bahasa yang digunakan adalah bahasa Indonesia
2. Permasalahan ini dibatasi hanya untuk layanan Indosat.
3. Data yang dianalisis berbentuk xml.
4. Data yang akan dianalisis secara *offline*.
5. Tidak menangani opini yang bermakna ambigu.

### 1.3 Tujuan

Adapun beberapa hal yang ingin dicapai dari pengerjaan tugas akhir ini, seperti:

1. Mengimplementasikan *Naïve Bayes Classifier* dan *Mutual Information* pada kasus *opinion mining*.
2. Mengklasifikasikan opini positif, negatif, dan netral serta memvisualkannya sebagai fungsionalitas tambahan dari aplikasi.
3. Menganalisis performansi akurasi *Naïve Bayes Classifier* dan *Mutual Information* pada kasus *opinion mining* dengan parameter *recall*, *precision*, dan akurasi.

### 1.4 Metodologi Penyelesaian Masalah

Metode yang akan digunakan dalam menyelesaikan tugas akhir ini sebagai berikut:

#### a. Studi Literatur

Tahap ini adalah tahap paling awal yang dilakukan. Pekerjaan pada tahap ini adalah pencarian sumber-sumber tulisan (referensi) mengenai *Opinion mining*, metode yang dapat dipakai untuk itu, cara pengambilan data secara *offline*, dan segala hal yang berhubungan dengan tugas akhir ini. Hal ini dilakukan untuk mendukung pemahaman untuk melanjutkan ke tahap-tahap berikutnya.

#### b. Analisis Kebutuhan dan Perancangan Sistem

Setelah melakukan studi literatur dan menguasai materi, tahap selanjutnya adalah menganalisis kebutuhan aplikasi yang nantinya akan dibuat dan bagaimana rancangannya. Analisis kebutuhan ini juga harus mengacu pada literatur-literatur yang ada. Setelah analisis kebutuhan, dilakukan perancangan sistem. Perancangan sistem ini harus dikerjakan secara mendetil, yaitu terdiri dari diagram alur sistem yang akan dibuat beserta penjelasannya, *usecase diagram* dan *scenario*, UML, dan ERD.

#### c. Implementasi Sistem

Setelah analisis kebutuhan dan perancangan sistemnya sudah lengkap, maka tahap selanjutnya adalah mengimplementasikannya dengan menggunakan tools java.

#### d. Pengujian Sistem

Setelah sistem yang dibuat telah selesai, selanjutnya dilakukan uji sistem. Pada tahap ini dilakukan sedemikian cara untuk mencari kesalahan sistem. Pertama, pencocokkan sistem yang telah selesai dengan rancangannya. Kedua, dilakukan pencari bug-bug lainnya. Jika terdapat kesalahan, maka secepatnya harus diperbaiki. Selain itu, karena sistem ini tidak secara langsung menuju kebutuhan akhir, maka setiap selesai pengujian sistem awal akan segera menjadi acuan untuk mengembangkan sistem tersebut menjadi sistem selanjutnya.

e. Analisis Hasil Pengujian

Setelah sistem akhir yang layak dipakai sudah dapat dipastikan benar dan berjalan dengan baik, dilakukan analisis hasil pengujian. Analisis ini dilihat dari akurasi dan bagaimana improvisasi sistem untuk mendapatkan hasil yang lebih baik.

f. Penyusunan Laporan

Membuat laporan tertulis atas semua kegiatan yang telah dilakukan selama menyelesaikan tugas akhir.



## 5 KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

1. Algoritma *Naïve BayesClassifier* dapat diimplementasikan untuk mengkategorisasikan *tweet* berbahasa Indonesia.
2. Algoritma *Naïve BayesClassifier* memberikan akurasi yang lebih baik saat dikombinasikan dengan metode *mutual information*.
3. Pengurangan *term* yang berlebihan dapat mengurangi nilai akurasi.

### 5.2 Saran

Berdasarkan percobaan yang dilakukan di atas, sebaiknya perlu diperhatikan beberapa poin sebagai berikut:

1. Proses pengambilan data tes dapat dilakukan secara online dan otomatis, sehingga tidak ada lagi proses persiapan data secara manual.
2. Penggunaan alur data yang lebih baik sehingga waktu proses train tidak terlalu lama.
3. Implementasi *Naive Bayes Classifier* dapat menggunakan teknik *feature selection* yang lain seperti *indonesian parser*, *Indonesian Morphological Analyzer*, dll.

## 6 DAFTAR PUSTAKA

- [1.] B. Liu, "Sentiment analysis: A multifaceted problem," IEEE. Intelligent System, vol. 25, no. 3, pp. 76–80, 2010.
- [2.] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval, vol.2, nos.1-2, pp.1-135, 2008.
- [3.] Fayyad, U. Piatesky-Shapiro, G. Smyth, P. 1996. From *Data mining to Knowledge Discovery in Database*, *AI Magazine*, 37-54.
- [4.] Go, A., Huang, L., & Bhayani, R. (2009). Twitter Sentiment Analysis. Final Project Report, Stanford University, Department of Computer Science.
- [5.] Hasugian, J. 2006. Penggunaan Bahasa Alamiah dan Kosa Kata Terkendali dalam Sistem Temu Balik Informasi Berbasis Teks. *Jurnal Studi dan Informasi* (Vol. 2, No.2). Fakultas Sastra: Universitas Sumatra Utara.
- [6.] Larose, Daniel T. 2006. Data Mining Methods and Models. Canada: Jhon Wiley & Sons, Inc.
- [7.] Manning, Cristoper. 2009. An Introduction to Information Retrieval, Cambridge UP: Cambridge.
- [8.] Mitchell, Tom M. (1997). Machine Learning, International Editions. McGraww-Hill, Inc. Singapore.
- [9.] Rennie, J.D.M, Shih, L. Teevan, J., karger, D.R. 2003. *Tackling the poor assumptions of naive Bayes text classifier*, In : Proceeding of Twentieth International Conference on Machine Learning. Washinton DC.
- [10.] Trisedya, Bayu D. *Klasifikasi Dokumen Menggunakan Algoritma Naive Bayes dengan Penambahan Parameter Probabilitas Parent Category*. Fakultas Ilmu Komputer Univrsitas Indonesia: Depok.
- [11.] Payam Refaeilzadeh, Lei Tang, Huan Liu (2008), "Cross-Validation", Arizona State University. [dikutip 5-32013]. Tersedia di: [http://www.cse.iitb.ac.in/~tarung/smt/papers\\_ppt/ency-cross-validation.pdf](http://www.cse.iitb.ac.in/~tarung/smt/papers_ppt/ency-cross-validation.pdf)