

# IMPLEMENTASI DAN ANALISIS ORIENTATION DETECTION (OPINI POSITIF DAN NEGATIF) PADA ONLINE PRODUCT REVIEW DENGAN NAIVE BAYES DAN FEATURE SELECTION

Wina Permana Sari<sup>1</sup>, Warih Maharani<sup>2</sup>, Intan Nurma Yulita<sup>3</sup>

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

---

## Abstrak

Dewasa ini semakin canggihnya teknologi dunia maya terutama internet, membuat semua hal menjadi mudah. Termasuk dalam memberikan opini untuk siapa saja dan apa saja. Misalnya saja opini untuk online product dalam suatu online shop. Bagi perusahaan profit yang berpusat pada pelanggan (customer), setiap opini dari mereka sangatlah penting karena dengan mengetahui opini tersebut maka akan dapat membantu perusahaan untuk memberikan pertimbangan dalam mengambil keputusan. Keputusan mengenai apakah suatu produk akan ditambah lagi produksinya atau tidak. Untuk mengetahui produk mana yang memiliki tingkat yang paling tinggi (best seller) maka perlu diketahui apakah opini tersebut merupakan opini positif maupun negatif. Untuk mengatasi hal tersebut maka dalam Tugas Akhir ini membahas mengenai klasifikasi opini positif dan negatif dalam online product. Untuk melakukan klasifikasi tersebut digunakan metode naïve bayes. Dan untuk menanggulangi jumlah dataset yang besar (memiliki jumlah fitur/kata yang banyak) sehingga meningkatkan akurasi naïve bayes maka naïve bayes tersebut dikombinasikan dengan feature selection. Dalam hal ini teknik feature selection yang digunakan yaitu Chi-Square. Dari hasil pengujian tersebut, maka terbukti bahwa feature selection yang dikombinasikan dengan Naïve Bayes menghasilkan performansi system yang lebih baik dibandingkan dengan tidak menggunakan feature selection untuk dataset yang memiliki fitur kata yang banyak.

Kata Kunci : opini, online product, naïve bayes, feature selection, chi-Square

---

## Abstract

In now days, there are many enterprenership that are take a chalenge to improve their bisnis. In order to have a good salary they are focus in customer reviews. There two kinds of reviews, the positive and negative. To help the bisnisman get the right choice, in this final project i used naive bayes to clasify the oipni and feature selection to improve the naive bayes performance. In for the last, in my assesment there are improve that the performance naive bayes bayes and feature selection is better than did'nt use the feature selection.

Keywords : opinion, online product, naïve bayes, feature selection, chi-Square

---

# 1. PENDAHULUAN

## 1.1 Latar belakang masalah

Dewasa ini tingkat kebutuhan manusia semakin tinggi seiring dengan meningkatnya teknologi. Semakin banyak pula perusahaan yang berkompetisi untuk membuat inovasi-inovasi baru agar tidak ketinggalan jaman dan produknya kian eksis dan pesat tingkat produksinya. Selain memasarkan produknya melalui outlet-outlet cabang, saat ini banyak juga yang sudah memanfaatkan teknologi internet dalam pemasarannya, hal ini dapat dilihat dari banyaknya online shop yang berdiri, seperti *amazon.com* merupakan salah satu online shop yang terpercaya dan cukup memiliki citra perusahaan yang bagus. Dalam *amazon.com* ini, customer dapat memberikan opini mereka terhadap produk yang ditawarkan oleh *amazon.com* sehingga opini ini dapat menjadi acuan bagi customer lain untuk membeli produk tersebut yang dikenal sebagai *online product reviews (opinion)*. Namun terkadang untuk mengetahui tanggapan pasar (pelanggan) akan suatu produk menjadi tantangan tersendiri bagi masing-masing perusahaan. Customer merupakan ladang keuntungan bagi perusahaan penyedia produk maupun jasa. Pendapat atau opini mereka pun menjadi satu hal penting bagi perusahaan. Jika perusahaan mendapat tanggapan/opini positif tentunya suatu perusahaan tersebut akan semakin meningkatkan jumlah produksi untuk dapat mengambil keuntungan yang lebih besar lagi. Pertimbangan inilah yang sangat penting bagi perusahaan sehingga tidak salah dalam mengambil keputusan selanjutnya. Karena jika hal ini tidak dapat ditangani tentunya akan merugikan bagi perusahaan untuk mulai memasarkan produknya.

Untuk menanggulangi jumlah data pada *online product reviews* (khususnya untuk opini positif dan negatif) dan meningkatkan performansi pada data yang sedikit, maka digunakan metode *Naïve Bayes Classifier (Nb)*. *Naïve Bayes Classifier (Nb)* merupakan sebuah pengklasifikasian atau pengkategorisasian probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (*independent*) yang tinggi[2], Sehingga waktu komputasinya relative lebih singkat dibanding dengan metode klasifikasi yang lainnya.

Permasalahan yang sering muncul pada *opinion mining* yaitu tingginya dimensi data (banyaknya fitur / kata dalam satu opini). Hal ini tentu mengganggu efektifitas dari klasifikasi itu sendiri. Oleh karena itu, untuk mengurangi tingginya dimensi data harus dilakukan pemilihan beberapa atribut yang dapat memberikan pengaruh terhadap hasil klasifikasi, yaitu *feature selection*[1]. Terdapat banyak metode dalam *feature selection* seperti, *Chi-Square*, *Chi-WSS* varian dari *Chi-Square*, *information gain*, *weight of evidence* dan lain lain. Keunggulan dari *Chi-Square* dibanding dengan metode lainnya adalah dapat meningkatkan performansi dari suatu kategori teks dilihat dari segi akurasi[1].

Berdasarkan pertimbangan di atas, tugas akhir ini akan meneliti mengenai *orientation detection* (khususnya opini positif dan negatif) dengan *Naïve Bayes Classifier (Nb)* dengan *Feature selection* dengan *Chi-Square*.

## 1.2 Perumusan masalah

1. Bagaimana mengimplemetasikan *Naïve Bayes Classifier (Nb)* dan *feature selection* pada opinion mining?
2. Bagaimana pengaruh *feature selection* terhadap hasil akurasi klasifikasi opini positif dan negatif

## 1.3 Batasan Masalah

1. Dataset yang digunakan merupakan opini yang memiliki format .txt (dataset yang berasal dari online product review dari *amazon.com*)
2. Opini yang akan diproses hanya yang berbahasa Inggris dengan kaedah yang benar
3. Tidak menangani opini yang tidak berkaitan dengan topik
4. Klasifikasi yang dibuat hanya terbatas pada opini positif dan negatif

## 1.4 Tujuan

Dalam masalah *orientation detection* banyak dilakukan dengan berbagai metode klasifikasi, dimana klasifikasi opinion positif dan negatif ini merupakan salah satu dari klasifikasi teks. Adapun dalam konteks klasifikasi teks sebuah opini dipandang sebagai sebuah dokumen dan proses *selection* dilakukan untuk mendapatkan fitur (kata) yang berpengaruh terhadap sebuah opini. Dan fitur inilah yang membantu menentukan sebuah opini tersebut positif atau negatif.

Tujuan dari tugas akhir ini yaitu :

1. Menganalisa implementasi *Naïve Bayes Classifier (Nb)* dan *feature selection* dalam klasifikasi opini
2. Menganalisa pengaruh *feature selection* pada *Naïve Bayes Classifier (Nb)* dari hasil akurasi

## 1.5 Metodologi penyelesaian masalah

Metode yang digunakan dalam tugas akhir ini adalah sebagai berikut:

### a. Studi literature

Tahap awal dari penelitian ini adalah melakukan studi literature. Tahap ini bertujuan untuk memperoleh berbagai informasi. Studi Literatur dilakukan untuk mencari teori mengenai *orientation detection*, algoritma *Naïve Bayes Classifier (Nb)* dan *feature selection*. Pencarian teori dilakukan pada text book dan jurnal-jurnal ilmiah.

### b. Pengumpulan data

Pada tahap ini dilakukan pengumpulan dataset yang akan digunakan dalam penentuan *klasifikasi orientation detection* dengan menggunakan algoritma *Naïve Bayes Classifier (Nb)*. Dataset tersebut berasal dari dataset dari sebuah *online product reviews* yang terpercaya dan akurat (*amazon.com*)

**c. Analisis dan perancangan**

Pada tahap ini dilakukan analisis dan perancangan perangkat lunak dengan menggunakan pendekatan *Object Oriented*.

**d. Implementasi**

Tahap implementasi perangkat lunak dilakukan dengan menggunakan Netbeans sesuai dengan spesifikasi dan perancangan yang telah ditentukan.

**e. Pengujian dan analisis hasil**

Pada tahap ini dilakukan pengujian terhadap performansi algoritma *Naïve Bayes Classifier (Nb)* dan *feature selection*. Kemudian dilakukan analisis terhadap kelebihan dan keterbatasan metode tersebut pada kasus orientation detection (opini negatif dan positif).

**f. Penyusunan dokumentasi**

Tahap akhir dari penelitian adalah penyusunan dokumentasi. Dokumentasi ditulis dalam bentuk buku Tugas Akhir dan berisi dasar teori, tahapan prosen penelitian, serta hasil penelitian menggunakan yang telah dilakukan.



## 5. PENUTUP

### 5.1 Kesimpulan

Berdasarkan pengujian dan analisis yang telah dilakukan, dapat diberikan beberapa kesimpulan sebagai berikut:

- a. *Naïve Bayes (Nb)* saja tanpa *feature selection* akan menghasilkan yang lebih baik daripada *Naïve Bayes (Nb)* dengan *feature selection* ketika dataset berukuran kecil (jumlah fitur/katanya sedikit)
- b. *Naïve Bayes (Nb)* saja dengan *feature selection* akan menghasilkan yang lebih baik daripada *Naïve Bayes (Nb)* tanpa *feature selection* ketika dataset berukuran besar (jumlah fitur/katanya banyak)
- c. Karakteristik data dan pembagian dataset sangat mempengaruhi kinerja dari *Naïve Bayes (Nb)* dan *feature Selection*

### 5.2 Saran

Adapun saran untuk penelitian lebih lanjut berkaitan dengan Opinion Mining maupun Feature Selection sebagai berikut:

- a. Untuk metode klasifikasi opini positif dan negatif dapat digunakan metode lain selain *naïve bayes*, misalnya teknik SVM, KNN, Algoritma genetika, dll
- b. Untuk metode feature selection dapat digunakan metode lain selain Chi-Square misalnya Information gain, Chi-WSS, dll

## DAFTAR PUSTAKA

- [1] Abraham Ranjit, 2009, "Effective Discretization and Hybrid Feature Selection Using Naïve Bayesian Classifier for Medical Data Mining", Dr. MGR University, Chennai, India [Paper Referensi]
- [2] Andria, Johanes, 2006, "JST back propagation dan feature selection dalam email spam filtering", Universitas Indonesia
- [3] Ian H. Witten and Eibe Frank, 2005, "Data Mining : Practical Machine learning Tools and Techniques 2<sup>nd</sup> edition", San Fransisco : Morgan Kaufmann Publisher
- [4] Meesad Phayung, Boonrawd Pudsadee, Nuijian Vatine, 2011, "A chi-square Test for Word Importance Diffrentiation in Text Classification", Bangkok, King Mongkut's University of Technology North Bangkok
- [5] Pang-Ning Tan, 2006, "Data Mining", Michigan State University, Michael Steinbach, University of Minnesota Vipin Kumar, University of Minnesota
- [6] Thabtah Fadi, Ali Mohammad, Zamzeer Mannam, Hadi Musa Wa'el, 2009 , "Naïve Bayesian Based on Chi-square to categorize Arabic Data ", AL-Isra Private Univercity,
- [7] Wenqian Shang, Houkuan Huang, Haibin Zhu, YOUNMIN LIN, Youli Qu, and Zhihai Wang. 2007. A Novel Feature Selection Algorithm for Text Categorization. [www.npissing.edu/haibin\\_zhu](http://www.npissing.edu/haibin_zhu) [7 April 2010]
- [8] [www.en.wikipedia.org/wiki/Naive\\_bayes\\_classifier](http://www.en.wikipedia.org/wiki/Naive_bayes_classifier) [26 April 2008]
- [9] Yang C Christopher, 2009, "Classifying Web Review Opinions for Consumer Product Analysis", Philadelphia, Drexel University.[Paper Utama]