

# ANALISA SISTEM KLASIFIKASI JUDUL SKRIPSI MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Dwi Puspitasari<sup>1</sup>, Pramana Yoga Saputra<sup>2</sup>, Ilham Agung Prakoso<sup>3</sup>

Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang

<sup>1</sup>dwipuspitasari@polinema.ac.id., <sup>2</sup>pramanay@polinema.ac.id, <sup>3</sup>ilhamagungp6196@gmail.com

---

## Abstrak

Skripsi adalah hasil karya tulis ilmiah yang berbasis pada penelitian dengan mengulas permasalahan beserta solusi yang dihasilkan pada disiplin ilmu tertentu. Hal ini juga didukung oleh rujukan pustaka terdahulu, beserta temuan dari hasil pengembangan eksperimen, pengamatan dengan arahan dosen pembimbing. Pada Jurusan Teknik Informatika Politeknik Negeri Malang terdapat banyak judul skripsi dalam berbagai bidang yang telah dibuat, tetapi judul tersebut belum terklasifikasikan kedalam bidangnya masing-masing. Berangkat dari masalah yang ada, dibuatlah sistem klasifikasi judul skripsi dengan mengambil studi kasus di jurusan Teknologi Informasi Politeknik Negeri Malang. Pada sistem ini data judul skripsi akan dikategorikan kedalam daftar bidang keahlian Teknik Informatika, sehingga setiap judul akan memiliki label sesuai dengan kategori bidang keahlian yang ada. Mengacu pada *Association for Computing Machinery (ACM)* terdapat 24 kategori bidang keahlian diantaranya : *Networking, Information Management, Web and Mobile System, Platform Technology, System Integration, Software Fundamental, Cyber Security, Global Professional Practice, System Administration and Maintenance, Applied Network, Data Scalability, Integrated System, IoT, Software development Management, User Experience Desain, Virtual System and Services, Social Responsibility, Mobile Application, Cloud Computing, Information Retrieval, Digital Image Processing, CyberSecurity Emerging Challenges, Geographical Information Service dan Game Development*. Pengkategorian judul skripsi sangat diperlukan untuk mengetahui persebaran judul skripsi pada masing-masing kategori. Penelitian ini bertujuan untuk mengkategorikan judul skripsi dibidang informatika kedalam 24 kategori menggunakan Naïve Bayes Classifier, karena menurut penelitian sebelumnya metode tersebut memberikan nilai akurasi yang cenderung baik. Pengujian algoritma menggunakan data training pada 50 judul, 100 judul dan 150 judul menghasilkan nilai akurasi sebesar 52%, 56% dan 58%. Berdasarkan penelitian tersebut dapat disimpulkan bahwa semakin banyak data training yang digunakan, maka nilai akurasi akan semakin meningkat.

**Kata Kunci :** klasifikasi, persebaran judul skripsi, *Naïve Bayes Classifier*

---

## 1. Pendahuluan

Perkembangan data teks saat ini telah mencapai jumlah yang cukup besar. Hal tersebut disebabkan oleh berkembangnya dunia teknologi informasi yang terdiri dari data teks di dalamnya. Saat ini, berbagai macam media online seperti blog, situs berita online, dan jejaring sosial menjadi sumber data teks yang sangat potensial untuk digali lebih dalam. Namun, data berbentuk teks memiliki karakteristik yang tidak terstruktur dan banyak memuat noise. Oleh karena itu, text mining memiliki peran penting dalam pengolahan data teks. Diantara proses yang dapat dilakukan dalam text mining adalah klasifikasi teks.

Klasifikasi teks dapat didefinisikan sebagai proses untuk menentukan suatu dokumen teks ke dalam suatu kelas tertentu. Dengan mengaplikasikan proses- proses dalam text mining, maka akan diperoleh pola-pola data, tren, dan ekstraksi dari pengetahuan-pengetahuan yang potensial dari data teks. Salah satu bentuk data teks yang dapat diolah yaitu data teks judul skripsi.

Skripsi adalah istilah yang digunakan di Indonesia untuk mengilustrasikan suatu karya tulis ilmiah berupa paparan tulisan hasil penelitian sarjana S1 yang membahas suatu permasalahan atau fenomena dalam bidang ilmu tertentu berdasarkan hasil kajian pustaka yang diteliti oleh para ahli, hasil penelitian lapangan atau hasil pengembangan atau eksperimen. Skripsi merupakan karya tulis ilmiah berdasarkan hasil penelitian lapangan atau studi kepustakaan yang disusun mahasiswa sesuai dengan bidang studinya sebagai tugas akhir dalam studi formalnya di Perguruan Tinggi (Munslich Mansnur, 2009;

## 2. Landasan Teori

### 2.1 Text Mining

Text Mining adalah proses yang mencoba mengekstrak informasi berguna dari teks natural language. Hal itu bisa diartikan sebagai proses menganalisa teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu Lokesh & Parul (2013). *Text mining* memiliki tujuan dan menggunakan

proses yang sama dengan *data mining*, namun memiliki *input* yang berbeda.

### 2.2 Naive Bayes Classifier

*Naive Bayes Classifier* merupakan salah satu metoda machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan vocabulary, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sedapat mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen.

### 2.3 Skripsi

Skripsi adalah istilah yang digunakan di Indonesia untuk mengilustrasikan suatu karya tulis ilmiah berupa paparan tulisan hasil penelitian sarjana S1 yang membahas suatu permasalahan / fenomena dalam bidang ilmu tertentu dengan menggunakan kaidah-kaidah yang berlaku. Skripsi merupakan karya tulis ilmiah berdasarkan hasil penelitian lapangan dan atau studi kepustakaan yang disusun mahasiswa sesuai dengan bisang studinya sebagai tugas akhir dalam studiformalnya di Perguruan Tinggi. Skripsi adalah karya ilmiah yang ditulis mahasiswa program S-1 yang membahas topik atau bidang tertentu berdasarkan hasil kajian pustaka yang diteliti oleh para ahli, hasil penelitian lapangan atau hasil pengembangan atau eksperimen (Munslich Mansnur, 2009: 4).

## 3. Metodologi

### 3.1. Metode Pengumpulan Data

Metode pengumpulan data digunakan untuk mengumpulkan data- data pendukung yang dibutuhkan dalam proses pembuatan aplikasi. Beberapa cara yang dapat digunakan adalah melalui studi literatur, yaitu dengan mengumpulkan dan mempelajari beberapa referensi dari berbagai sumber yang berkaitan dengan judul penelitian yang dilakukan.

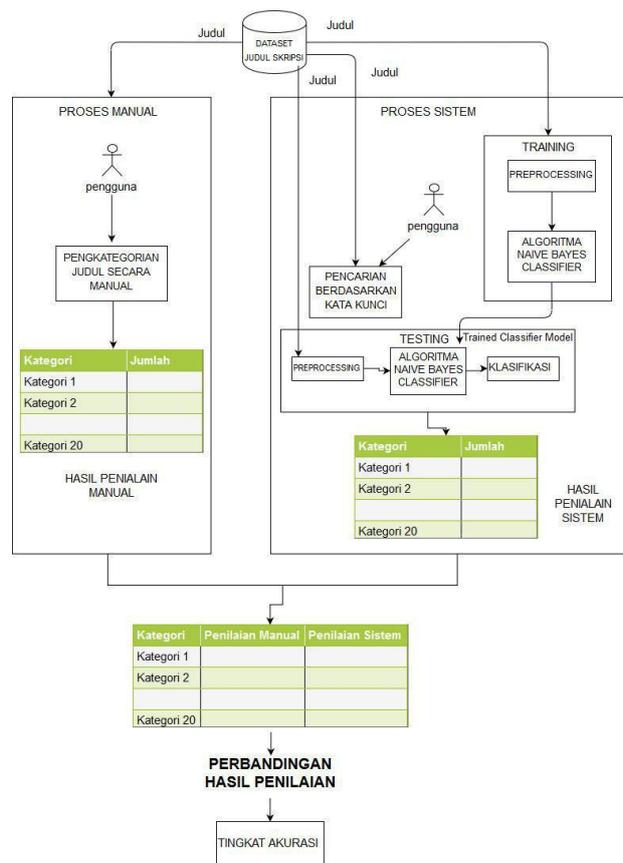
Data-data yang dibutuhkan dalam penelitian ini adalah sebagai berikut :

- Dataset judul skripsi pada jurusan teknologi informasi tahun 2014 sampai tahun 2017 yang diambil dari website jurnal informatika polinema(JIP), website jurnal seminar informatika aplikatif polinema (SIAP) dan juga dosen panitia skripsi

- Daftar kategori bidang keahlian jurusan teknologi informasi mengacupada Association for Computing Machinery (ACM) Daftar kata stopword Bahasa Indonesia, stopwords adalah kata-kata yang diabaikan dalam pemrosesan data, contoh: ada.apa.awal,bagi, dll.

### 3.2. Metode Pengolahan Data

Data judul Skripsi yang digunakan yaitu data pada tahun 2014 hingga tahun 2017 yang didapat dari website jurnal informatika polinema(JIP) dan website jurnal seminar informatika aplikatif polinema (SIAP). Data tersebut perlu dilakukan pengolahan agar menjadi data yang mudah digunakan dalam proses klasifikasi. Data judul akan mengalami penyeleksian kata- kata sehingga judul menjadi lebih ringkas. Beberapa komponen dari judul akan dihapus untuk diseleksi. Proses ini dapat disebut dengan preprocessing. Setelah proses preprocessing selesai maka data akan diproses menggunakan algoritma Naive Bayes Classifier agar menghasilkan nilai peluang pada masing masing kategori. Nilai ini yang akan menjadi masukan untuk proses klasifikasi pada tahap testing.



### 3.3. Metode Pengujian Sistem

#### a. Pengujian Perangkat Lunak

Pengujian perangkat lunak merupakan proses untuk mengetahui apakah sistem berjalan dengan baik dan tidak ada kerusakan. Metode pengujian yang digunakan adalah *black box*. *Black box* testing adalah pengujian yang dilakukan hanya mengamati hasil eksekusi melalui data uji dan memeriksa fungsional dari perangkat lunak.

#### b. Performa Algoritma

Untuk mengetahui performa algoritma yang digunakan yaitu *Naive Bayes Classifier* maka perlu dilakukan pengujian tingkat akurasi dan presisi. Tahap yang dilakukan dalam analisis tingkat akurasi dan presisi sebagai berikut:

- Merekap klasifikasi opini yang dilakukan melalui *testing* menggunakan aplikasi
- Merekap klasifikasi opini yang dilakukan melalui *testing* secara manual
- Melakukan perbandingan antara hasil klasifikasi opini melalui *testing* menggunakan aplikasi dengan *testing* secara manual
- Melakukan penghitungan hasil yang sama dan hasil yang berbeda dari kedua *testing* yang dilakukan
- Melakukan penghitungan tingkat akurasi algoritma *Naive Bayes Classifier* dengan rumus sebagai berikut:

$$Akurasi = \frac{\sum \text{data benar}}{n \text{ dokumen}} \times 100 \quad (1)$$

- Melakukan penghitungan hasil yang positif dan negatif
- Melakukan penghitungan tingkat presisi algoritma *Naive Bayes Classifier* dengan rumus sebagai berikut :

$$Presisi = \frac{\sum \text{data positif atau negatif}}{n \text{ dokumen positif atau negatif}} \times 100 \quad (2)$$

### 4. Kesimpulan dan Saran

Berdasarkan hasil penelitian dan pengujian tingkat akurasi yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Algoritma *Naive Bayes Classifier* dapat digunakan untuk mengklasifikasikan judul skripsi
2. Pengujian dengan menggunakan jumlah data *training* yang berbeda menghasilkan nilai akurasi yang berbeda dengan kecenderungan akurasi yang semakin meningkat. Hasil akurasi klasifikasi algoritma *Naive Bayes Classifier* menggunakan *training* 50 judul, 100 judul, 150 judul didapat akurasi sistem sebesar 52%, 56 %, dan 58%.

#### Daftar Pustaka:

- Hamza, Amir. 2012 “Klasifikasi Teks Dengan Naive Bayes Classifier (Nbc) Untuk Pengelompokan Teks Berita Dan Abstract Akademis” Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III.
- Lokesh, Kumar & Parul, Kalra Bhatia. 2013. “Text Mining: Concepts, Process And Applications”. Journal of Global Research in Computer Science
- Nathaniel Chandra, Daniel., Indrawan, Gede., Sukajaya, Nyoman, Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naive Bayes Dengan Fitur N-Gram. Jurnal Ilmiah Teknologi dan Informasia ASIA (JITIKA) Vol.10, No.1, ISSN: 0852-730X. Singaraja. Universitas Pendidikan Ganesha
- Sukma, Alifian., Puji Santoso, Bagus., Ramadhan, Dian., Wiraswari, Ni Made Ayu Karina., Sari, Tiara Ratna., “Klasifikasi Dokumen Bahasa Jawa Menggunakan Metode N-Gram” Jurusan Sistem Informasi, Fakultas Sains dan Teknologi Universitas Airlangga, Surabaya
- Vicky Indriyono, Bonifacius., Ema Utami. “Klasifikasi Jenis Buku Berdasarkan Judul dan Sinopsis Menggunakan Metode NBC” .Program Pascasarjana Magister Teknik Informatika. STMIK AMIKOM, Yogyakarta