



**KLASIFIKASI KATA MENGGUNAKAN METODE NAÏVE BAYES  
UNTUK DOKUMEN BERBAHASA INDONESIA BERTEMA  
PENDIDIKAN**

**SKRIPSI**

**TITA SUHARYATI**

**1110511019**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S1 TEKNIK INFORMATIKA**

**2015**



**KLASIFIKASI KATA MENGGUNAKAN METODE NAÏVE BAYES  
UNTUK DOKUMEN BERBAHASA INDONESIA BERTEMA  
PENDIDIKAN**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Komputer**

**TITA SUHARYATI**

**1110511019**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S1 TEKNIK INFORMATIKA**

**2015**

## PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Tita Suharyati

NRP : 111051109

Tanggal : 31 Juli 2015

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 31 Juli 2015

Yang Menyatakan



(Tita Suharyati)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI  
UNTUK KEPENTINGAN AKADEMIS**

---

Sebagai citivas akademik universitas pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini :

Nama : Tita Suharyati  
NRP : 1110511019  
Fakultas : Ilmu Komputer  
Program Studi : Teknik Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royaltis Non eksklusif (Non-exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul :

**Klasifikasi Kata Menggunakan Metode Naïve Bayes Untuk Dokumen Berbahasa Indonesia Bertema Pendidikan.**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat : Jakarta

Pada Tanggal : 31 Juli 2015

Yang Menyatakan



(Tita Suharyati)

## PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Tita Suharyati

NRP : 1110511019

Program Studi : S1 Teknik Informatika

Judul Skripsi : Klasifikasi Kata Menggunakan Metode Naïve Bayes Untuk Dokumen Berbahasa Indonesia Bertema Pendidikan

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Strata 1 Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta.



Titin Pramiyati, S.Kom, M.Si

Ketua Penguji



Theresia Wati, S.Kom, M.T.I

Penguji I



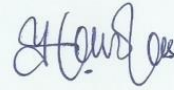
Iin Ernawati, S.Kom, M.Si

Penguji II (Pembimbing)



Dr. Nidjo Sandjojo, MSc

Dekan



Yuni Widiastiwi, S.Kom, M.Si

Ka. Prodi

Ditetapkan Di : Jakarta

Tanggal Ujian : 31 Juli 2015

# **KLASIFIKASI KATA MENGGUNAKAN METODE NAÏVE BAYES UNTUK DOKUMEN BERBAHASA INDONESIA BERTEMA PENDIDIKAN**

**Tita Suharyati**

## **Abstrak**

Semakin berkembangnya teknologi menuntut agar pekerjaan dapat diselesaikan dengan cepat dan meminimalkan campur tangan dari manusia. Semakin banyak pekerjaan yang dikerjakan oleh suatu teknologi maka semakin mempermudah manusia dalam menyelesaikan pekerjaannya. Kategorisasi teks yang sering dijumpai masih banyak yang manual sehingga sistem hanya dapat menentukan kategori teks tersebut berdasarkan yang telah pengguna tetapkan sebelumnya. Oleh karena itu, dibutuhkan teknik khusus untuk mengidentifikasi kata yang relevan terhadap topik dalam sebuah dokumen berbahasa Indonesia. Dalam hal ini membahas tentang klasifikasi kata menggunakan metode Naïve Bayes. Tujuan dari penelitian ini untuk mengetahui metode Naïve Bayes dalam melakukan klasifikasi dokumen teks berbahasa Indonesia bertema pendidikan & Melakukan pengklasifikasian kata berdasarkan dokumen berbahasa Indonesia bertema pendidikan. Hal ini terlihat dari hasil perhitungan, yaitu kata yang paling banyak muncul adalah kata “program” dengan hasil probabilitas 13.4% dan kata yang paling sedikit muncul adalah kata “keilmuan” dengan hasil probabilitas 0.3% dengan jumlah dokumen yang diuji sebanyak 35 dokumen.

**Kata Kunci** : Klasifikasi, Naïve Bayes, Probabilitas

# **WORD CLASSIFICATION USING NAIVE BAYES METHOD FOR INDONESIAN LANGUAGE DOCUMENT FOR EDUCATION THEMED**

**Tita Suharyati**

## **Abstract**

Growing technology demanded that the job can be completed quickly and minimize the interference of man. The more work done by a technology the more ease in completing his work man. Text categorization often encountered many manual so that the system can only determine the category of the text based on the user who has set previously. Therefore, it takes a special technique to identify words that are relevant to the topic in a document language, Indonesia. In this case discusses the classification of the word Naïve Bayes method. The purpose of this research is to know the methods of Naïve Bayes document classification in doing text Indonesia-themed education based on words which do & document Indonesia education themed. This is apparent from the results of the calculation, i.e. the most words appear is the word "program" with the results of the probability of 13.4% and said that at least appears is the word "science" with the results of the probability of 0.3% with the number of the document that was tested as many as 35 of the document.

**Keyword** : Classification, Naïve Bayes, Probability

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Topik yang dipilih dalam penelitian yang dilakukan sejak bulan Februari 2015 ini adalah **KLASIFIKASI KATA MENGGUNAKAN METODE NAÏVE BAYES UNTUK DOKUMEN BERBAHASA INDONESIA BERTEMA PENDIDIKAN**. Dalam kesempatan ini, dengan penuh rasa hormat penulis ingin mengucapkan ucapan terima kasih yang sebesar-besarnya atas bimbingan dan kerja sama yang diberikan kepada penulis, dan untuk itu penulis ucapkan terima kasih kepada :

1. Allah SWT, atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi ini.
2. Keluarga penulis khususnya Ibu (Euis Suaedah) yang senantiasa mendoakan kelancaran dan kesuksesan dalam pelaksanaan skripsi ini.
3. Bapak Dr.Nidjo Sandjojo,M.Sc, selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jakarta.
4. Ibu Yuni Widiastiwi, S.Kom.,M.Si. sebagai Kepala Program Studi S-1 Teknik Informatika Fakultas Ilmu Komputer UPN “Veteran” Jakarta.
5. Ibu Iin Ernawati, S.Kom., M.Si. dosen pembimbing yang telah memberikan masukan-masukan yang berguna bagi penulis.
6. Bapak Bayu Hananto, S.Kom. selaku Dosen Pembimbing Akademik penulis.
7. Teman terbaik, Virna, Pratiwi, Astri, Puput, Andika, Bagus, dan Imam Dermawan. Terima kasih atas dukungan dan sarannya demi terselesaikannya skripsi ini.
8. Dan segenap teman-teman Program Studi Teknik Informatika angkatan 2011 UPN “Veteran” Jakarta, khususnya TI Lokal A.

Saran dan kritik yang membangun dari semua pihak sangat diharapkan demi penyempurnaan selanjutnya.

Jakarta, Juli 2015

Tita Suharyati



## DAFTAR ISI

HALAMAN JUDUL .....	i
PERNYATAAN ORISINALITAS .....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI .....	iii
PENGESAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	1
I.1 Latar Belakang .....	1
I.2 Rumusan Masalah .....	1
I.3 Tujuan Penelitian.....	1
I.4 Batasan Masalah.....	1
I.5 Luaran dan Manfaat .....	2
I.6 Sistematika Penulisan .....	2
BAB II LANDASAN TEORI.....	4
II.1 Berita Pendidikan .....	4
II.2 Klasifikasi Teks .....	4
II.3 Metode Pengklasifikasian Teks .....	5
II.4 Text Mining .....	7
II.5 Naïve Bayes untuk Klasifikasi Teks.....	9
II.6 Database .....	11
II.7 Review Riset yang Relevan .....	14
BAB III METODOLOGI PENELITIAN .....	15
III.1 Desain Penelitian.....	15
III.2 Tahapan Penelitian.....	16
III.3 Tata Laksana .....	17
III.4 Waktu dan Tempat Penelitian.....	18
III.5 Alat dan Bahan Penelitian .....	18
BAB IV PEMBAHASAN.....	20
IV.1 Analisis .....	20
IV.2 Praproses .....	24
IV.3 Implementasi Penelitian .....	29
BAB V PENUTUP.....	37
V.1 Kesimpulan .....	37
V.2 Saran .....	37

DAFTAR PUSTAKA.....	38
RIWAYAT HIDUP	
LAMPIRAN	

## DAFTAR TABEL

Tabel 1 Review Riset Yang Relevan.....	14
Tabel 2 Jadwal Kegiatan.....	18
Tabel 3 Pencarian Nilai Tf.....	22
Tabel 4 Pencarian Nilai Idf.....	23
Tabel 5 Nilai Tf.....	27
Tabel 6 Nilai Idf.....	28
Tabel 7 Kata Kunci per Dokumen.....	30
Tabel 8 Hasil Probabilitas Naïve Bayes.....	35

## DAFTAR GAMBAR

Gambar 1	Pembelajaran mesin dengan klasifikasi terpandu.....	4
Gambar 2	Diagram proses klasifikasi teks.....	7
Gambar 3	Kerangka Pemikiran.....	15
Gambar 4	Proses Tokenisasi.....	24
Gambar 5	Proses Filtering.....	25
Gambar 6	Penentuan Pola/ Kata Kunci.....	26
Gambar 7	Hasil Probabilitas Sekolah.....	31
Gambar 8	Hasil Probabilitas Siswa.....	31
Gambar 9	Hasil Probabilitas Kuliah.....	31
Gambar 10	Hasil Probabilitas Beasiswa.....	32
Gambar 11	Hasil Probabilitas Program.....	32
Gambar 12	Hasil Probabilitas Studi.....	32
Gambar 13	Hasil Probabilitas Prodi.....	32
Gambar 14	Hasil Probabilitas Perguruan.....	32
Gambar 15	Hasil Probabilitas Pendidik.....	33
Gambar 16	Hasil Probabilitas Biaya.....	33
Gambar 17	Hasil Probabilitas Akademik.....	33
Gambar 18	Hasil Probabilitas SD.....	33
Gambar 19	Hasil Probabilitas SMP.....	34
Gambar 20	Hasil Probabilitas SMA.....	34
Gambar 21	Hasil Probabilitas Lulus.....	34
Gambar 22	Hasil Probabilitas Keilmuan.....	34
Gambar 23	Hasil Probabilitas Jurusan.....	34
Gambar 24	Hasil Probabilitas Swasta.....	35
Gambar 25	Hasil Probabilitas Ujian.....	35
Gambar 26	Hasil Probabilitas Mahasiswa.....	35

## **DAFTAR LAMPIRAN**

Lampiran 1 Perhitungan pembobotan dokumen

Lampiran 2 Artikel yang digunakan

Lampiran 3 Kata Kunci Per Dokumen