

**KLASIFIKASI ARTIKEL BERITA INDONESIA MENGGUNAKAN
METODE *NAÏVE BAYES CLASSIFIER***

SKRIPSI



disusun oleh

Ahmad Muzakki

15.11.9031

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**

**KLASIFIKASI ARTIKEL BERITA INDONESIA MENGGUNAKAN
METODE NAÏVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Ahmad Muzakki

15.11.9031

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**

PERSETUJUAN

SKRIPSI

**KLASIFIKASI ARTIKEL BERITA INDONESIA MENGGUNAKAN
METODE NAÏVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Ahmad Muzakki

15.11.9031

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 30 April 2018

Dosen Pembimbing,


M. Rudyanto Ariel, S.T, M.T

NIK. 190302098

PENGESAHAN
SKRIPSI
KLASIFIKASI ARTIKEL BERITA INDONESIA MENGGUNAKAN
METODE NAÏVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Ahmad Muzakki

15.11.9031

telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Oktober 2018

Susunan Dewan Penguji

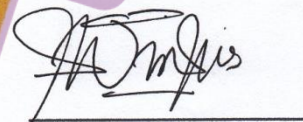
Nama Penguji

M. Rudyanto Arief, S.T, M.T
NIK. 190302098

Erni Seniwati, S.Kom, M.Cs
NIK. 190302231

Wiwi Widayani, M.Kom
NIK. 190302272

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Oktober 2018

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 10 November 2018



Ahmad Muzakki

NIM. 15.11.9031

MOTTO

”Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya” (QS. 2:286)

”... Boleh jadi kamu tidak menyenangi sesuatu, padahal itu baik bagimu, dan boleh jadi kamu menyenangi sesuatu padahal itu tidak baik bagimu. Allah mengetahui, sedangkan kamu tidak mengetahui” (QS. 2:216)

”Orang-orang yang beriman dan hati mereka menjadi tentram dengan mengingat Allah. Ingatlah, hanya dengan mengingat Allah hati menjadi tentram”
(QS. 13:28)

PERSEMBAHAN

Alhamdulillah penulis panjatkan puji syukur kepada Allah SWT atas segala rahmat, taufiq, serta hidayah-Nya, sehingga diberi kesempatan untuk dapat menyelesaikan skripsi ini dengan sebaik-baiknya dengan segala kekurangan penulis. Segala syukur penulis ucapkan kepada-Mu karena telah menghadirkan mereka yang memberikan semangat dan doa disaat menjalani proses pembuatan skripsi ini. Dengan segala kerendahan hati saya persembahkan skripsi ini kepada :

1. Kedua Orang Tua, Bapak Munair, S.Pd dan Ibu Rokhimah yang selalu mendoakan, memberi semangat serta motivasi supaya dapat menyelesaikan skripsi ini dengan lancar serta bermanfaat bagi semua.
2. Adikku Nadhifatul Arofah yang tidak berhenti untuk tetap menghibur hingga memberi semangat pada dalam mengerjakan skripsi.
3. Pak M. Rudyanto Arief, S.T, M.T selaku dosen pembimbing dalam skripsi ini yang tidak lelah untuk tetap membimbing dan mengingatkan penulis dari awal hingga akhir proses pembuatan skripsi.
4. Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu baik ilmu akademik maupun ilmu non-akademik selama kuliah.
5. Keluarga besar kelas 15-S1IF-08 yang telah bersama-sama menemani selama kuliah. Semoga silaturahmi kita tetap terjaga.
6. Keluarga besar UKM UKI Jashtis yang selalu bersama-sama berjalan di dalam dakwah selama kuliah di Amikom. Semoga kita tetap bersama sampai di surga nanti.
7. Serta orang-orang yang selalu membantu peneliti dalam mengerjakan skripsi yang tidak bisa sebut namanya satu-satu.

Saya ucapkan terima kasih yang sebesar-besarnya untuk kalian semua. Mohon maaf jika ada salah kata atau perbuatan baik yang disengaja maupun tidak disengaja selama ini. Sukses untuk kalian semua, semoga Allah SWT memberikan rahmat dan hidayah-Nya kepada kita semua. Dan semoga skripsi ini dapat bermanfaat dan berguna untuk kemajuan ilmu pengetahuan kedepannya. Aamiin...

KATA PENGANTAR

Alhamdulillah penulis panjatkan puji syukur kepada Allah SWT atas segala rahmat, taufiq, serta hidayah-Nya kepada penulis sehingga dapat menyelesaikan skripsi yang berjudul “Klasifikasi Artikel Berita Indonesia Menggunakan Metode Naïve Bayes Classifier”.

Selama proses pengerjaan skripsi ini penulis menyadari bahwa dalam proses penulisan skripsi ini banyak mengalami kendala, namun berkat bantuan, bimbingan, kerjasama dari berbagai pihak dan berkah dari Allah SWT sehingga kendala-kendala tersebut bisa diatasi. Selanjutnya ucapan terimakasih penulis sampaikan kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak M. Rudyanto Arief, S.T, M.T selaku dosen pembimbing yang telah memberikan banyak masukan yang membantu membimbing dalam menyelesaikan skripsi ini.
3. Ibu Krisnawati Krisnawati, S.Si, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Dosen Penguji (Erni Seniwati, S.Kom, M.Cs dan Wiwi Widayani, M.Kom) yang telah memberikan masukan terhadap penelitian ini.
5. Kedua Orang Tua dan keluarga yang selalu memberikan doa, dukungan dan semangat.
6. Keluarga Besar UKM UKI Jashtis dan Keluarga Besar Kelas 15-S1IF-08.
7. Serta Semua pihak yang tidak bisa penulis sebutkan satu-persatu yang telah membantu dalam penyusunan skripsi ini.

Penulis menyadari bahwa masih banyak terdapat kekurangan-kekurangan dalam mengerjakan skripsi ini, sehingga penulis mengharapkan adanya saran dan kritik yang membangun demi kesempurnaan skripsi ini.

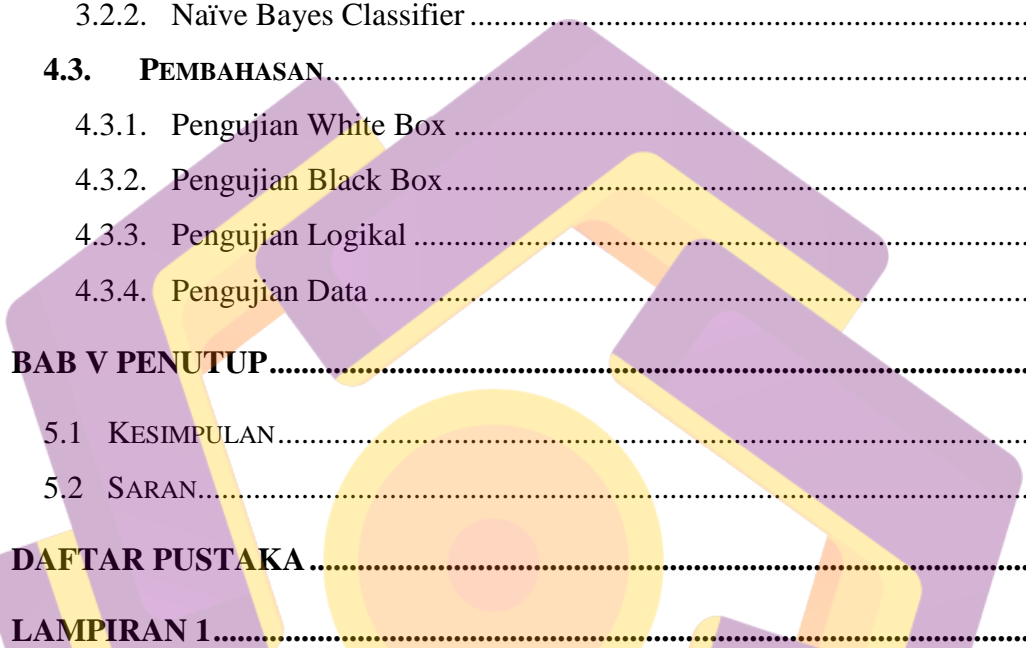
Yogyakarta, 10 November 2018

Ahmad Muzakki
NIM. 15.11.9031

DAFTAR ISI

JUDUL	i
PERSETUJUAN.....	II
PENGESAHAN.....	ERROR! BOOKMARK NOT DEFINED.
PERNYATAAN.....	IV
MOTTO	V
PERSEMBAHAN.....	VI
KATA PENGANTAR.....	VII
DAFTAR ISI.....	VIII
DAFTAR TABEL	XI
DAFTAR GAMBAR.....	XII
DAFTAR ISTILAH	XIII
INTISARI	XIV
ABSTRACT.....	XV
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	2
1.3 BATASAN MASALAH	2
1.4 MAKSUD DAN TUJUAN PENELITIAN	3
1.4.1. Maksud Penelitian.....	3
1.4.2. Tujuan Penelitian	3
1.5 METODE PENELITIAN.....	4
1.5.1. Metode Pengumpulan Data.....	4
1.5.2. Metode Analisis	4
1.5.3. Metode Perancangan	5

1.5.4. Metode Pengujian.....	5
1.5.5. Metode Implementasi.....	6
1.7 SISTEMATIKA PENULISAN	6
BAB II LANDASAN TEORI	8
2.1 TINJAUAN PUSTAKA	8
2.2 DASAR TEORI.....	9
2.2.1. Berita.....	9
2.2.2. Konsep Basis Data	11
2.2.3. System Development Life Cycle (SDCL).....	12
2.2.4. Data Flow Diagram (DFD)	15
2.2.5. Entity Relationship Diagram (ERD)	16
2.2.6. Teks Preprocessing.....	18
2.2.7. Data Mining	21
2.2.8. Naïve Bayes Classifier (NBC)	23
2.2.9. Laplace Smoothing.....	24
2.2.10. Analisis SWOT	25
2.2.11. Confusion Matrix	28
BAB III ANALISIS DAN PERANCANGAN.....	30
3.1 ANALISIS	30
3.1.1. Gambaran Umum.....	30
3.1.2. Analisis Masalah	31
3.1.3. Identifikasi Masalah.....	31
3.1.4. Analisis SWOT	32
3.1.5. Analisis Kebutuhan	34
3.1.6. Analisis Pengetahuan	36
3.2 PERANCANGAN	38
3.2.1. Perancangan Proses.....	38
3.2.2. Perancangan Data.....	41
3.2.3. Perancangan Antar Muka (Interface).....	43
BAB IV HASIL DAN PEMBAHASAN	48



4.1. IMPLEMENTASI	48
4.1.1. Implementasi Data Base.....	48
4.1.2. Implementasi AntarMuka.....	51
4.2. LISTING KODE PROGRAM	55
3.2.1. Preprocessing	55
3.2.2. Naïve Bayes Classifier	56
4.3. PEMBAHASAN	59
4.3.1. Pengujian White Box	59
4.3.2. Pengujian Black Box.....	60
4.3.3. Pengujian Logikal	62
4.3.4. Pengujian Data	65
BAB V PENUTUP	72
5.1 KESIMPULAN.....	72
5.2 SARAN.....	73
DAFTAR PUSTAKA	74
LAMPIRAN 1	76

DAFTAR TABEL

Tabel 2.1	Analisis SWOT	27
Tabel 2.2	Confusion Matrix	28
Tabel 3.1	Hasil Analisis SWOT	33
Tabel 3.2	Tabel DataBase Dictionary	42
Tabel 3.3	Tabel DataBase Kata	42
Tabel 3.4	Tabel DataBase Kategori	42
Tabel 3.5	Tabel DataBase Training	42
Tabel 3.6	Tabel DataBase Prob	43
Tabel 4.1	Uji Black Box	60
Tabel 4.2	Jumlah Kata Data Training	62
Tabel 4.3	Hasil Perhitungan Manual	63
Tabel 4.4	Dataset	65
Tabel 4.5	Pengujian	65
Tabel 4.6	Hasil Pengujian	70

DAFTAR GAMBAR

Gambar 2.1	Flowchart Preprocessing	19
Gambar 3.1	Flowchart Sistem	39
Gambar 3.2	Data Flow Diagram Level 1	39
Gambar 3.3	Data Flow Diagram Level 2	40
Gambar 3.4	ERD	40
Gambar 3.5	Relasi Tabel	40
Gambar 3.6	Rancangan Halama Utama	43
Gambar 3.7	Rancangan Halaman Klasifikasi	44
Gambar 3.8	Rancangan Halaman Daftar Data Training	44
Gambar 3.9	Rancangan Halaman Daftar Data Probabilitas	45
Gambar 3.10	Rancangan Halaman Daftar Kategori	45
Gambar 3.11	Rancangan Halaman Daftar Data Kategori	46
Gambar 3.12	Rancangan Halaman Berita	46
Gambar 4.1	Hasil Query Tabel Dictionary	48
Gambar 4.2	Hasil Query Tabel Kata	49
Gambar 4.3	Hasil Query Tabel Kategori	49
Gambar 4.4	Hasil Query Tabel Training	50
Gambar 4.5	Hasil Query Tabel Prob	51
Gambar 4.6	Hasil Implementasi Database	51
Gambar 4.7	Implementasi Halaman Utama	52
Gambar 4.8	Implementasi Halaman Klasifikasi	52
Gambar 4.9	Implementasi Halaman Daftar Data Training	53
Gambar 4.10	Implementasi Halaman Daftar Probabilitas	53
Gambar 4.11	Implementasi Halaman Daftar Kategori	54
Gambar 4.12	Implementasi Halaman Daftar Data Kategori	54
Gambar 4.13	Implementasi Halaman Berita	55
Gambar 4.14	Contoh Pengujian White Box	59
Gambar 4.15	Contoh Hasil Pengujian White Box	60
Gambar 4.16	Contoh Pengujian Black Box	61
Gambar 4.17	Hasil Uji Perhitungan Sistem	64

DAFTAR ISTILAH



INTISARI

Keuntungan terbesar dari berita internet yang ditulis secara online, oleh beberapa Koran-koran elektronik di portal mereka. Setiap kali ada sesuatu yang dilaporkan di mana saja di seluruh dunia, hal itu akan dipublikasikan di internet dalam hitungan menit. Surat kabar biasanya memiliki batas waktu untuk berita yang akan dilaporkan dalam setiap hal yang mereka publikasikan. Ini adalah salah satu keuntungan terbesar dari Koran online yang cepat menyebarkan beritanya kapan pun.

Salah satu cara yang paling berhasil untuk mengorganisasikan informasi dalam jumlah banyak dan dapat dipahami oleh para pencari informasi adalah dengan melakukan klasifikasi dokumen berdasarkan topiknya. Kebutuhan akan dokumen pembelajaran untuk melakukan klasifikasi dokumen merupakan salah satu permasalahan yang sering muncul dalam topik klasifikasi dokumen. Permasalahan yang timbul menjadi semakin rumit dengan adanya fakta bahwa jumlah simpanan data berita menjadi sangat besar dan tidak terorganisir. Oleh karena itu, diperlukan suatu strategi pengelompokan otomatis dokumen-dokumen berita tersebut.

Oleh karena itu dalam penelitian ini akan dirancang sistem klasifikasi yang dapat mengelompokkan artikel berita menggunakan naive bayes classifier (NBC). Klasifikasi ini ditekankan untuk data artikel berbahasa indonesia. Pada sistem klasifikasi ini menggunakan proses pembelajaran dan untuk proses keterkaitan antar data artikel diukur berdasarkan nilai probabilitas dari data dan kata yang ada.

Kata Kunci : naive bayes classification, berita, teks minning, preprocessing

ABSTRACT

The biggest advantage of internet news is written online, by some electronic newspapers on their portals. Whenever something is reported anywhere in the world, it will be published on the internet in a matter of minutes. Newspapers usually have a time limit for news to be reported in every thing they publish. This is one of the biggest advantages of an online newspaper that is fast-forwarding news at any time.

One of the most successful ways of organizing large amounts of information and can be understood by information seekers is to classify documents based on the topic. The need for document learning to classify documents is one of the problems that often appear in the topic of document classification. The problems that arise become increasingly complicated by the fact that the amount of data storage data becomes very large and not organized. Therefore, an automatic grouping strategy of the news documents is required.

Therefore in this study will be designed a classification system that can classify news articles using naive bayes classifier (NBC). This classification is emphasized for Indonesian language article data. In this classification system using the learning process and for the process of linkage between data articles measured based on the probability value of existing data and words.

Keyword : *naive bayes classification, news, text minning, preprocessing*