

**METODE *MULTINOMIAL NAÏVE BAYES CLASSIFIER* DAN
REDUKSI FITUR MENGGUNAKAN *DF THRESHOLDING*
UNTUK KLASIFIKASI GENRE BERDASARKAN *BLURB*
PADA MEDIA BERBAGI CERITA *ONLINE***

SKRIPSI

Dijukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana
Komputer Program Studi Teknik Informatika



Oleh :

LUTHFIYATUL 'AZIZAH

NPM. 1534010002

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

2019

LEMBAR PENGESAHAN SKRIPSI

Judul : METODE *MULTINOMIAL NAÏVE BAYES CLASSIFIER* DAN REDUKSI FITUR MENGGUNAKAN *DF THRESHOLDING* UNTUK KLASIFIKASI GENRE BERDASARKAN *BLURB* PADA MEDIA BERBAGI CERITA *ONLINE*

Oleh : LUTHFIYATUL 'AZIZAH

NPM : 1534010002

Telah Diseminarkan Dalam Ujian Skripsi Pada:
Hari Jumat, Tanggal 23 Agustus 2019

Mengetahui

Dosen Pembimbing

Dosen Penguji

1.

1.

Intan Yuniar P., S.Kom, M.Sc
NPT. 3 8006 04 0198 1

Budi Nugroho, S.Kom, M.Kom
NPT. 3 8009 05 0205 1

2.

2.

Eva Yulia P., S.Kom, M.Kom
NPT. 3 8907 13 0 346 1

Henni Endah Wahanani, ST, M.Kom
NPT. 3 7809 13 0348 1

3.

Wahyu S. J. Saputra, S.Kom, M.Kom
NPT. 3 8608 10 0295 1

Menyetujui

Dekan
Fakultas Ilmu Komputer

Koordinator Program Studi
Teknik Informatika

Dr. Ir. Ni Ketut Sari, MT
NIP. 19650731 199203 2 001

Budi Nugroho, S.Kom, M.Kom
NPT. 3 8009 05 0205 1

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Teknik Informatika UPN "Veteran" Jawa Timur, yang bertandatangan dibawah ini:

Nama : LUTHFIYATUL 'AZIZAH

NPM : 1534010002

Menyatakan bahwa Judul Skripsi / Tugas Akhir yang Saya ajukan dan akan dikerjakan, yang berjudul:

"METODE *MULTINOMIAL NAÏVE BAYES CLASSIFIER* DAN REDUKSI FITUR MENGGUNAKAN *DF THRESHOLDING* UNTUK KLASIFIKASI GENRE BERDASARKAN *BLURB* PADA MEDIA BERBAGI CERITA *ONLINE*"

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan merupakan produk dan atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 5 September 2019

Hormat Saya,




LUTHFIYATUL 'AZIZAH
NPM. 1534010002

KATA PENGANTAR

Puji syukur penulis panjatkan pada Tuhan Yang Maha Esa yang telah menganugerahkan rahmat serta karunia-Nya kepada kita, sehingga penulis diberikan kekuatan dan kesabaran untuk menyelesaikan laporan skripsi dengan judul **“METODE *MULTINOMIAL NAÏVE BAYES CLASSIFIER* DAN REDUKSI FITUR MENGGUNAKAN *DF THRESHOLDING* UNTUK KLASIFIKASI GENRE BERDASARKAN *BLURB* PADA MEDIA BERBAGI CERITA *ONLINE*”** tepat pada waktunya. Laporan ini dibuat untuk memenuhi persyaratan untuk menyelesaikan tugas akhir pada program studi Teknik Informatika di Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.

Penulis menyadari bahwa laporan skripsi ini masih jauh dari kesempurnaan. Atas segala kekurangan penulis sangat mengharapkan masukan, kritik, dan saran yang bersifat membangun kearah perbaikan dan penyempurnaan laporan skripsi ini.

Akhir kata penulis mengucapkan terimakasih kepada semua pihak yang berperan dalam penyusunan laporan skripsi ini. Penulis berharap semoga laporan skripsi ini dapat bermanfaat bagi semua pihak.

Surabaya, 5 September 2019

Penulis

UCAPAN TERIMA KASIH

Segala puji hanyalah milik Allah SWT, sumber segala hikmah dan ilmu pengetahuan karena berkat rahmat dan berkah-Nya penulis dapat menyusun dan menyelesaikan laporan Skripsi ini hingga selesai tepat pada waktunya. Dengan segala hormat, Ucapan terima kasih saya sampaikan ke berbagai pihak yang turut membantu memperlancar penyelesaian laporan Skripsi ini, yaitu kepada:

1. Kedua orang tua yang selalu memberi semangat dan mendoakan keberhasilan penulis serta selalu memberi motivasi untuk menyelesaikan semua tugas, dan selalu memberikan fasilitas yang mendukung untuk mempercepat penyelesaian laporan skripsi ini.
2. Ibu Dr. Ir. Ni Ketut Sari, MT selaku dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
3. Bapak Budi Nugroho, S.Kom. M.Kom selaku kepala program studi Teknik Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
4. Ibu Intan Yuniar Purbasari, S.Kom., M.Sc selaku dosen pembimbing 1 yang dengan sabar membimbing, mengarahkan serta memberikan masukan sejak awal penelitian ini berlangsung hingga akhir.
5. Ibu Eva Yulia Puspaningrum, S.Kom, M.Kom selaku dosen pembimbing 2 yang dengan sabar membimbing, mengarahkan serta memberikan masukan sejak awal penelitian ini berlangsung hingga akhir.

6. Bapak dan Ibu dosen program studi Teknik Informatika UPN “Veteran” Jawa Timur yang telah mendidik dan memberikan ilmunya sehingga penulis memiliki bekal untuk dapat melakukan penelitian ini.
7. Kakak dan adik yang selalu memberikan semangat dan motivasi untuk segera menyelesaikan pengerjaan penelitian ini.
8. Teman seperjuangan khususnya Devi Anugrah P P, Nuraini Hariyanti dan Ririn Puji A yang telah membantu banyak dalam pengerjaan penelitian ini hingga selesai.
9. Teman-teman jurusan Teknik Informatika, serta semua pihak yang telah membantu proses perkuliahan di Teknik Informatika yang tidak bisa penulis sebutkan satu persatu.

Semoga Tuhan Yang Maha Esa memberikan balasan yang setimpal kepada semua pihak yang telah memberikan bantuan, bimbingan maupun motivasi yang bermanfaat bagi penulis. Sehingga penulis mampu dan berhasil melaksanakan penelitian dan proses penyusunan laporan penelitian skripsi dengan baik dan tepat.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	Error! Bookmark not defined.
SURAT PERNYATAAN ANTI PLAGIAT	Error! Bookmark not defined.
ABSTRAK	Error! Bookmark not defined.
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH	v
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	Error! Bookmark not defined.
1.1 Latar Belakang	Error! Bookmark not defined.
1.2 Perumusan Masalah	Error! Bookmark not defined.
1.3 Batasan Masalah	Error! Bookmark not defined.
1.4 Tujuan Penelitian	Error! Bookmark not defined.
1.5 Manfaat Penelitian	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA	Error! Bookmark not defined.
2.1 Penelitian Terdahulu	Error! Bookmark not defined.
2.2 Media Berbagi Cerita <i>Online</i> (Wattpad).	Error! Bookmark not defined.
2.3 <i>Text Mining</i>	Error! Bookmark not defined.
2.4 Klasifikasi Teks	Error! Bookmark not defined.
2.5 <i>Text Preprocessing</i>	Error! Bookmark not defined.

2.6	<i>Document Frequency Thresholding</i>	Error! Bookmark not defined.
2.7	Pembobotan (<i>Term Weighting</i>).....	Error! Bookmark not defined.
2.8	<i>Multinomial Naïve Bayes Classifier</i>	Error! Bookmark not defined.
2.9	Uji Coba	Error! Bookmark not defined.
2.10	Evaluasi Performa	Error! Bookmark not defined.
BAB III METODOLOGI.....		Error! Bookmark not defined.
3.1	Metode Penelitian	Error! Bookmark not defined.
3.2	Pengumpulan Data.....	Error! Bookmark not defined.
3.3	Preprocessing Data.....	Error! Bookmark not defined.
3.3.1	<i>Case Folding</i>	Error! Bookmark not defined.
3.3.2	<i>Tokenizing</i>	Error! Bookmark not defined.
3.3.1	<i>Filtering</i>	Error! Bookmark not defined.
3.3.1	<i>Stemming</i>	Error! Bookmark not defined.
3.4	Reduksi Fitur	Error! Bookmark not defined.
3.5	Pembobotan Kata.....	Error! Bookmark not defined.
3.6	Klasifikasi	Error! Bookmark not defined.
3.7	Skenario Uji Coba.....	Error! Bookmark not defined.
3.8	Evaluasi Performa.....	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN		Error! Bookmark not defined.
4.1	Pengumpulan Dataset.....	Error! Bookmark not defined.
4.2	Implementasi Sistem.....	Error! Bookmark not defined.

4.2.1	Text Prepossesing	Error! Bookmark not defined.
4.2.2	Pembobotan <i>TF IDF</i>	Error! Bookmark not defined.
4.2.3	<i>DF Thresholding</i>	Error! Bookmark not defined.
4.2.4	Klasifikasi	Error! Bookmark not defined.
4.2.5	Implementasi Antarmuka	Error! Bookmark not defined.
4.3	Uji Coba	Error! Bookmark not defined.
4.3.1	Uji Coba <i>TF IDF</i> + <i>MNBC</i>	Error! Bookmark not defined.
4.3.2	Uji coba <i>TF IDF</i> + <i>DF Thresholding</i> + <i>MNBC</i>	Error! Bookmark not defined.
4.3.3	Analisa Hasil Uji Coba	Error! Bookmark not defined.
4.4	Validasi	Error! Bookmark not defined.
BAB V KESIMPULAN DAN SARAN.....		Error! Bookmark not defined.
5. 1	Kesimpulan.....	Error! Bookmark not defined.
5. 2	Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA		Error! Bookmark not defined.
BIODATA PENULIS		Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian.....	Error! Bookmark not defined.
Gambar 3. 2 Proses Preprocessing	Error! Bookmark not defined.
Gambar 3. 3 Proses Reduksi Fitur	Error! Bookmark not defined.
Gambar 3. 4 Proses Pembobotan Kata.....	Error! Bookmark not defined.
Gambar 4. 1 Struktur data_cerita.....	Error! Bookmark not defined.
Gambar 4. 2 Contoh Case Folding	Error! Bookmark not defined.
Gambar 4. 3 Contoh Hasil Tokenizing	Error! Bookmark not defined.
Gambar 4. 4 Hasil Filtering	Error! Bookmark not defined.
Gambar 4. 5 contoh hasil stemming	Error! Bookmark not defined.
Gambar 4. 6 Halaman Utama	Error! Bookmark not defined.
Gambar 4. 7 Halaman Dataset.....	Error! Bookmark not defined.
Gambar 4. 8 Halaman Preprocessing	Error! Bookmark not defined.
Gambar 4. 9 Halaman Proses Preprocessing	Error! Bookmark not defined.
Gambar 4. 10 Halaman Validasi.....	Error! Bookmark not defined.
Gambar 4. 11 Halaman Perhitungan Klasifikasi ...	Error! Bookmark not defined.
Gambar 4. 12 Jumlah Term Sebelum Reduksi	Error! Bookmark not defined.
Gambar 4. 13 Jumlah Term Setelah Reduksi Threshold 2-15	Error! Bookmark not defined.

DAFTAR TABEL

Tabel 3. 1 Contoh Case Folding	Error! Bookmark not defined.
Tabel 3. 2 Contoh Tokenizing	Error! Bookmark not defined.
Tabel 3. 3 Contoh Filtering	Error! Bookmark not defined.
Tabel 3. 4 Pembobotan Kata TF-IDF	Error! Bookmark not defined.
Tabel 3. 5 Contoh Data Testing	Error! Bookmark not defined.
Tabel 4. 1 Contoh Tabel TF IDF	Error! Bookmark not defined.
Tabel 4. 2 Hasil Uji Coba TF IDF + MNBC data training ..	Error! Bookmark not defined.
Tabel 4. 3 Uji Coba TF IDF + DF Thresholding + MNBC (2 - 37).....	Error! Bookmark not defined.
Tabel 4. 4 Uji Coba TF IDF + DF Thresholding + MNBC (2 - 15).....	Error! Bookmark not defined.
Tabel 4. 5 Uji Coba TF IDF + DF Thresholding + MNBC (2 - 10).....	Error! Bookmark not defined.
Tabel 4. 6 Hasil Uji Coba.....	Error! Bookmark not defined.
Tabel 4. 7 Confusion Matrix TF IDF + MNBC	Error! Bookmark not defined.
Tabel 4. 8 Confusion Matrix TF IDF + DF Thresholding + MNBC	Error! Bookmark not defined.
Tabel 4. 9 Hasil Perbandingan.....	Error! Bookmark not defined.

METODE *MULTINOMIAL NAÏVE BAYES CLASSIFIER* DAN REDUKSI FITUR MENGGUNAKAN *DF THRESHOLDING* UNTUK KLASIFIKASI GENRE BERDASARKAN *BLURB* PADA MEDIA BERBAGI CERITA *ONLINE*

Nama Mahasiswa : Luthfiyatul ‘Azizah

NPM : 1534010002

Program Studi : Teknik Informatika

Dosen Pembimbing : 1. Intan Yuniar Purbasari, S.Kom, MSc

2. Eva Yulia Puspaningrum, S.Kom, M.Kom

Abstrak

Saat ini banyak media *online* yang menyediakan bacaan berupa cerita-cerita untuk bisa dinikmati oleh para pembaca. Untuk menentukan sebuah bacaan masuk ke *genre* tertentu bukanlah sesuatu yang mudah apalagi bacaan yang perbedaannya tidak terlalu jelas. Penelitian ini bertujuan untuk melakukan prediksi menggunakan metode *Multinomial Naïve Bayes Classifier (MNBC)* dipadukan dengan pembobotan kata yaitu *Term Frequency Inverse Document Frequency (TF IDF)* dan reduksi fitur menggunakan *Document Frequency Thresholding*.

Data yang digunakan dalam penelitian ini diambil dari media berbagi cerita *online* yaitu <https://www.wattpad.com/> berjumlah 500 data yang terdiri dari lima kategori yaitu humor, horror, roman, misteri dan spiritual. Proses pertama adalah *text processing* yang terdiri dari empat tahap yaitu *case folding*, *tokenizing*, *filtering* dan *stemming*. Proses selanjutnya adalah pembobotan kata menggunakan *TF IDF* dan reduksi fitur menggunakan *DF Thresholding*. Kemudian proses klasifikasi menggunakan *Multinomial Naïve Bayes Classifier*.

Dari hasil penelitian ini diketahui bahwa *DF Thresholding* meningkatkan nilai akurasi, presisi dan *recall*. Rata-rata nilai yang dihasilkan oleh *TF IDF + MNBC* adalah akurasi = 0.80, presisi = 0.80, dan *recall* = 0.81. Dan rata-rata nilai dari *TF IDF + DF Thresholding + MNBC* adalah akurasi = 0.86, presisi = 0.87, dan *recall* = 0.87.

Kata kunci: *DF Thresholding*, *Klasifikasi*, *Multinomial Naive Bayes Classifier*, *TF IDF*