

**IMPLEMENTASI ALGORITMA *RANDOM FOREST*
DAN *MULTINOMIAL NAIVE BAYES* DENGAN
MENGUNAKAN SELEKSI FITUR *INFORMATION
GAIN* UNTUK KLASIFIKASI BERITA MEDIA
MONITORING KAWASAN GEOPARK CILETUH**

TUGAS AKHIR



**UNIVERSITAS
BAKRIE**

**RAFI RAMADHAN BAHARI
1162001024**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2021**

**IMPLEMENTASI ALGORITMA *RANDOM FOREST*
DAN *MULTINOMIAL NAIVE BAYES* DENGAN
MENGUNAKAN SELEKSI FITUR *INFORMATION
GAIN* UNTUK KLASIFIKASI BERITA MEDIA
MONITORING KAWASAN GEOPARK CILETUH**

TUGAS AKHIR

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
Komputer



**UNIVERSITAS
BAKRIE**

**RAFI RAMADHAN BAHARI
1162001024**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2021**

Halaman Pernyataan Orisinalitas

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

Nama : Rafi Ramadhan Bahari

NIM : 1162001024

Tanda Tangan :



Tanggal : 31 Agustus 2021

Halaman Pengesahan

Laporan Tugas Akhir ini diajukan oleh :

Nama : Rafi Ramadhan Bahari
NIM : 1162001024
Program Studi : Teknik Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul : Implementasi Algoritma *Random Forest* dan *Multinomial Na-
ive Bayes* dengan Menggunakan Seleksi Fitur *Information Gain*
untuk Klasifikasi Berita Media *Monitoring* Kawasan Geopark Ci-
letuh

Telah berhasil dipertahankan di hadapan Dewan Penguji sebagai persya-
ratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer,
Universitas Bakrie.


DEWAN PENGUJI

Pembimbing : Guson P. Kuntarto, S.T., M.Sc.



(...02.Sep.2021...)

Penguji 1 : Iwan Adhicandra, M.I.E.E.E., M.I.E.T., M.B.C.S.



(.....)

Ditetapkan di : Jakarta

Tanggal : 31 Agustus 2021

Ungkapan Terima Kasih

Assalamualaikum Warahmatullahi Wabarakatuh. Dengan memanjatkan puji syukur kepada Allah SWT yang telah melimpahkan rahmat, karunia, dan hidayah-Nya sehingga Tugas Akhir ini dapat terselesaikan dengan judul “Implementasi Algoritma *Random Forest* dan *Multinomial Naive Bayes* dengan Menggunakan Seleksi Fitur *Information Gain* untuk Klasifikasi Berita Media *Monitoring* Kawasan *Geopark* Ciletuh”, sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer Universitas Bakrie.

Proses Penelitian dan Penyusunan tugas akhir ini tentunya tidak dapat terselesaikan tanpa bantuan, dukungan, bimbingan ataupun nasihat dari banyak pihak. Oleh karena itu, pada kesempatan ini dengan segala ketulusan dan kerendahan hati penulis menyampaikan ucapan terima kasih kepada:

1. Kedua Orang Tua penulis yaitu Bapak Saiful Bahri, Ibu Ratna Jayawati, dan juga Kakak Aditya Mirza Bahari, yang selalu memberikan doa, kasih sayang, semangat, nasihat, dan bimbingannya dalam setiap langkah hidup yang penulis ambil dan jalani sampai saat ini.
2. Bapak Guson Prasamuarso Kuntarto, S.T., M.Sc. selaku dosen pembimbing Tugas Akhir yang selalu sedia untuk memberikan bimbingan, motivasi, nasihat, doa, serta semangat selama proses penelitian dan penyusunan tugas akhir ini.
3. Bapak Irwan Prasetya Gunawan, S.T., M.Eng., Ph.D. selaku dosen pembimbing akademik dan pembahas seminar proposal yang sudah memberikan bimbingan, nasihat dan saran yang sangat membangun selama masa studi dan penyusunan tugas akhir ini.
4. Bapak Iwan Adhicandra, M.I.E.E.E., M.I.E.T., M.B.C.S. selaku dosen penguji sidang tugas akhir yang sudah memberikan saran dan perbaikan yang sangat membangun dalam penyusunan tugas akhir ini.
5. Rizky Novriyedi dan Alhamsyah Bintang Dhyasta yang selalu memberikan nasihat, semangat, serta saran yang membangun selama penulis menyusun tugas

akhir dan menempuh studi pada Program Studi Teknik Informatika Universitas Bakrie.

6. Mutiara Julia Ifra dan Clara Velita yang saling memberikan nasihat, doa, dan semangat untuk berjuang bersama menyelesaikan penelitian dan penyusunan tugas akhir ini.
7. Seluruh staff pengajar Program Studi Teknik Informatika Universitas Bakrie yang telah memberikan ilmu pengetahuan yang sangat berguna selama penulis menempuh studi pada Program Studi Teknik Informatika Universitas Bakrie.
8. Seluruh rekan Teknik Informatika angkatan 2016 telah bersama selama ini.

Semoga Allah SWT membalas segala kebaikan kepada semua pihak yang telah membantu sampai saat ini. Penulis menyadari bahwa penulisan Tugas Akhir ini masih jauh dari kata sempurna, karena itu segala kritik dan saran yang membangun akan menyempurnakan tugas akhir ini yang kelak dapat bermanfaat bagi banyak pihak.

Jakarta, 31 Agustus 2021

Penulis



Rafi Ramadhan Bahari

Halaman Pernyataan Persetujuan Publikasi

Sebagai civitas akademik Universitas Bakrie, saya yang bertanda tangan dibawah ini:

Nama : Rafi Ramadhan Bahari
NIM : 1162001024
Program Studi : Informatika
Fakultas : Teknk dan Ilmu Komputer

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty Free Right)** atas karya ilmiah saya yang berjudul:

Implementasi Algoritma *Random Forest* dan *Multinomial Naive Bayes* dengan Menggunakan Seleksi Fitur *Information Gain* untuk Klasifikasi Berita Media *Monitoring* Kawasan Geopark Ciletuh

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya sebagai penulis/pencipta dan pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 31 Agustus 2021

Jakarta, 31 Agustus 2021



Rafi Ramadhan Bahari

IMPLEMENTASI ALGORITMA *RANDOM FOREST* DAN *MULTINOMIAL NAIVE BAYES* DENGAN MENGGUNAKAN SELEKSI FITUR *INFORMATION GAIN* UNTUK KLASIFIKASI BERITA MEDIA *MONITORING* KAWASAN GEOPARK CILETUH

Rafi Ramadhan Bahari

ABSTRAK

Pada era teknologi informasi dan komunikasi seperti saat ini, proses penyebaran informasi menjadi lebih masif dan cepat. Hal ini menyebabkan proses media *monitoring* yang dilakukan praktisi *Public Relations* (PR) untuk mengembangkan dan mempertahankan reputasi kawasan *Geopark* Ciletuh menjadi kurang efektif dan efisien apabila dalam melakukan identifikasi dan analisis berita masih dilakukan secara manual. Penelitian ini mengusulkan sistem klasifikasi berita dengan menggunakan algoritma *machine learning* untuk membuat proses analisis berita yang dilakukan lebih efektif dan efisien dengan menggunakan dua algoritma *Machine Learning* yaitu *Random Forest* dan *Multinomial Naive Bayes* serta *Information Gain* sebagai metode pemilihan fitur dengan nilai *threshold* 0.05 dan 0.01 untuk melakukan klasifikasi berita *hard news* atau *soft news*. Dalam sistem *media monitoring*, proses ini masuk kedalam proses *Analysis Backend*. Dataset dibentuk berdasarkan dua model yang berbeda yaitu *single dimensional* dan *multidimensional*. Hasil penelitian untuk model *dataset single dimensional* dengan algoritma *Random Forest* memperoleh rata-rata tertinggi untuk nilai akurasi sebesar 81.42% menggunakan pemilihan fitur *Information Gain* dengan *threshold* 0.01, sedangkan algoritma *Multinomial Naive Bayes* memperoleh rata-rata tertinggi untuk nilai akurasi sebesar 74.18% menggunakan *Information Gain* dengan *threshold* 0.01 dan tanpa pemilihan fitur. Untuk model *dataset multidimensional* algoritma *Random Forest* memperoleh rata-rata tertinggi untuk nilai akurasi sebesar 93.8%, sedangkan algoritma *Multinomial Naive Bayes* sebesar 72.72%. Hasil penelitian menunjukkan bahwa penggunaan *Information Gain* untuk pemilihan fitur menunjukkan performa yang tidak optimal khususnya pada saat diterapkan di dalam algoritma *Multinomial Naive Bayes*. Hal ini dikarenakan algoritma tersebut memperlakukan seluruh fitur sebagai fitur independen. Selain itu, penggunaan metode *laplacian smoothing* tidak optimal di dalam melakukan seleksi fitur.

Kata Kunci : *Machine Learning, Random Forest, Soft News, Hard News, Multinomial Naive Bayes, Information Gain, Single Dimensional, Multidimensional*

**IMPLEMENTATION OF THE RANDOM FOREST AND
MULTINOMIAL NAIVE BAYES ALGORITHM BY USING
INFORMATION GAIN FEATURE SELECTION FOR NEWS
CLASSIFICATION OF MEDIA MONITORING FOR THE CILETUH
GEOPARK AREA**

Rafi Ramadhan Bahari

ABSTRACT

In the era of information and communication technology as it is today, the process of disseminating information is becoming more massive and faster. This causes the media monitoring process carried out by Public Relations (PR) practitioners to develop and maintain the reputation of the Ciletuh Geopark area to be less effective and efficient if the identification and analysis of news are still done manually. In the media monitoring system, this process is included in the Analysis Backend process. This study proposes a news classification system using machine learning algorithms to make the news analysis process more effective and efficient by using two Machine Learning algorithms, namely Random Forest and Multinomial Naive Bayes and Information Gain as a feature selection method with threshold values of 0.05 and 0.01 to perform classification of news hard news or soft news. The dataset is formed based on two different models, namely single-dimensional and multidimensional. The results for the single-dimensional dataset model with the Random Forest algorithm obtained the highest average for an accuracy value of 81.42% using the Information Gain feature selection with a threshold of 0.01, while the Multinomial Naive Bayes algorithm obtained the highest average for an accuracy value of 74.18% using Information Gain with a threshold of 0.01 and without feature selection. For the multidimensional dataset model, the Random Forest algorithm obtained the highest average for an accuracy value of 93.8%, while the Multinomial Naive Bayes algorithm was 72.72%. The results showed that the use of Information Gain for feature selection showed suboptimal performance, especially when applied to the Multinomial Naive Bayes algorithm. This is because the algorithm treats all features as independent features. In addition, the use of the laplacian smoothing method is not optimal in performing feature selection.

Keywords: Machine Learning, Random Forest, Soft News, Hard News, Multinomial Naive Bayes, Information Gain, Feature Selection, Public Relations, Single Dimensional, Multidimensional

Daftar Isi

Halaman Pernyataan Orisinalitas	i
Halaman Pengesahan	ii
Ungkapan Terima Kasih	iii
Halaman Pernyataan Persetujuan Publikasi	v
Abstrak	vi
Abstract	vii
Daftar Gambar	xi
Daftar Tabel	xiii
Daftar Singkatan	xv
Daftar Rumus	xvi
I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan dan Manfaat Penelitian	6
1.3.1 Tujuan Penelitian	6
1.3.2 Manfaat Penelitian	7
1.4 Ruang Lingkup Penelitian	7
1.5 Kontribusi Penelitian	7
1.6 Sistematika Penelitian	8
II Tinjauan Pustaka	10
2.1 Penelitian Terkait	10
2.2 <i>Media Monitoring</i>	14
2.3 <i>Term Frequency Inverse Document Frequency</i>	19
2.4 <i>Feature Selection</i>	20
2.4.1 <i>Information Gain</i>	22
2.5 <i>Machine Learning</i>	23
2.5.1 <i>Random Forest</i>	25
2.5.2 <i>Multinomial Naïve Bayes</i>	27
2.6 Pengukuran Performa dan Evaluasi	29
III Metodologi Penelitian	31
3.1 Tahapan Penelitian	31

3.1.1	Studi Literatur	32
3.1.2	Perumusan Masalah dan Tujuan Penelitian	32
3.1.3	Pengumpulan Data	32
3.1.4	Pelaksanaan Penelitian	33
3.1.5	Hasil Implementasi dan Pembahasan	33
3.1.6	Penyusunan Laporan Hasil Penelitian	33
3.2	Kerangka Kerja Penelitian	34
3.2.1	Pengumpulan Data	34
3.2.2	Pembentukan <i>Dataset</i>	35
3.2.3	Data <i>Preprocessing</i>	36
3.2.4	Penentuan Fitur dengan Implementasi <i>Bag Of Words</i>	38
3.2.5	Pembobotan TF-IDF	38
3.2.6	Pemilihan Fitur <i>Information Gain</i>	38
3.2.7	Implementasi Algoritma <i>Random Forest</i> dan <i>Multinomial Naive Bayes</i>	39
3.2.8	Uji Skenario Hasil dan Pengukuran Performa	40
3.3	Instrumen Penelitian	40
IV	Implementasi dan Hasil Penelitian	41
4.1	Pengumpulan Data	41
4.2	Pembentukan <i>Dataset</i>	42
4.3	Pembentukan <i>Dataset Multidimensional Model</i>	42
4.4	Pembentukan <i>Dataset Single Dimensional Model</i>	44
4.4.1	Data <i>Preprocessing</i>	45
4.4.2	Implementasi <i>Bag-of-Words</i> dalam Pembentukan Fitur	46
4.4.3	Pembobotan Kata dengan TF-IDF	48
4.5	Pemilihan Fitur dengan <i>Information Gain</i>	50
4.6	Pengembangan Model Klasifikasi Berbasis Algoritma <i>Random Forest</i>	54
4.6.1	<i>Boostrapping Data</i>	55
4.6.2	Pembentukan Pohon Keputusan (<i>Decision Tree</i>)	56
4.7	Pengembangan Model Klasifikasi Berbasis Algoritma <i>Multinomial Naive Bayes</i>	57
4.7.1	<i>Train Multinomial Naive Bayes</i>	58
4.7.2	<i>Testing Multinomial Naive Bayes Model</i>	61
4.8	Eksperimen	62
4.8.1	Hasil Pemilihan Fitur dengan <i>Information Gain</i>	64
4.8.2	Hasil Uji Coba Klasifikasi dengan <i>Random Forest</i>	65
4.8.3	Hasil Uji Coba Klasifikasi dengan <i>Multinomial Naive Bayes</i>	68
4.9	Pembahasan	71
V	Simpulan dan Saran	76
5.1	Simpulan	76
5.2	Saran	77
	Daftar Pustaka	78
A	<i>Raw Dataset</i>	82

B	Data Hasil <i>PreProcessing</i>	97
C	Hasil Perhitungan Nilai <i>Information Gain</i>	103
D	Hasil Perhitungan Nilai <i>Chi-Square</i>	106
E	<i>Dataset Single Dimensional</i>	109
F	<i>Raw Dataset Multidimensional</i>	113
G	<i>Dataset Multidimensional</i> Setelah Pemberian Bobot Manual	131
H	Hasil Penelitian Dengan Algoritma <i>Random Forest</i>	139
I	Hasil Penelitian Dengan Algoritma <i>Multinomial Naive Bayes</i>	149
J	<i>Source Code</i>	158

Daftar Gambar

1.1	Rancangan Arsitektur <i>Media Monitoring System Geopark Ciletuh</i> [5]	3
2.1	Rancangan Arsitektur <i>Media Monitoring System Geopark Ciletuh</i> [5]	16
2.2	Ilustrasi metode <i>Supervised Learning</i> [27]	24
2.3	Ilustrasi <i>random feature selection</i> [43]	25
2.4	Struktur <i>Random Forest Classification</i> [10]	26
3.1	Tahapan Penelitian Prediksi Berita <i>Online data Media Monitoring Geopark Ciletuh</i>	32
3.2	Kerangka Kerja Penelitian Implementasi Algoritma <i>Random Forest</i> dan <i>Multinomial Naive Bayes</i> dengan Menggunakan Seleksi Fitur <i>Information Gain</i> untuk Klasifikasi Berita <i>Media Monitoring</i> Kawasan Geopark Ciletuh	34
3.3	Tahapan <i>Preprocessing</i> , pembobotan kata dengan TF-IDF dan Seleksi Fitur dengan <i>Information Gain</i>	37
3.4	Proses klasifikasi yang diusulkan dalam penelitian berbasis Algoritma <i>Random Forest</i> dan <i>Multinomial Naive Bayes</i>	39
4.1	Contoh Potongan <i>Raw Data</i> Berita Terkait <i>Geopark Ciletuh</i>	41
4.2	Contoh Potongan Hasil Pembentukan <i>Dataset</i> dengan teknik <i>Multidimensional</i>	43
4.3	Contoh Potongan Hasil Pembentukan <i>Dataset</i> dengan Teknik <i>Multidimensional</i> Setelah dilakukan Skema Pembobotan	44
4.4	Contoh Salah Satu Output Hasil Data <i>Preprocessing</i>	46
4.5	Kumpulan Kata Hasil Implementasi BoW	47
4.6	Potongan hasil perhitungan nilai TF-IDF	49
4.7	Contoh Salah Satu <i>Decision Tree</i> yang Terbentuk	57
4.8	Hasil Uji Coba Model RF Terhadap <i>Testing Data</i> untuk <i>Dataset Single Dimensional</i> dengan <i>Grid Search</i> : Akurasi	67
4.9	Hasil Uji Coba Model MNB Terhadap <i>Testing Data</i> dengan <i>Dataset Single Dimensional</i> : Akurasi	69
4.10	Rata-rata Nilai Akurasi Terhadap <i>Testing Data</i> yang Diperoleh Kedua Algoritma RF dan MNB	72
4.11	Rata-rata Nilai Akurasi Terhadap <i>Training Data</i> yang Diperoleh Kedua Algoritma RF dan MNB	73
H.1	<i>Confusion Matrix</i> Algoritma <i>Random Forest</i> tanpa pemilihan fitur (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	142
H.2	<i>Confusion Matrix</i> Algoritma <i>Random Forest</i> dengan <i>Information Gain Threshold</i> 0.01 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	143
H.3	<i>Confusion Matrix</i> Algoritma <i>Random Forest</i> dengan <i>Information Gain Threshold</i> 0.05 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	144
H.4	<i>Confusion Matrix</i> Algoritma <i>Random Forest</i> dengan <i>Chi-Square Threshold</i> 0.01 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	145

H.5	<i>Confusion Matrix</i> Algoritma <i>Random Forest</i> dengan <i>Chi-Square Threshold</i> 0.05 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	146
H.6	<i>Confusion Matrix</i> Algoritma <i>Random Forest</i> dengan <i>Model Dataset Multidimensional</i> (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	147
I.1	<i>Confusion Matrix</i> Algoritma <i>Multinomial Naive Bayes</i> tanpa pemilihan fitur (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	151
I.2	<i>Confusion Matrix</i> Algoritma <i>Multinomial Naive Bayes</i> dengan <i>Information Gain Threshold</i> 0.01 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	152
I.3	<i>Confusion Matrix</i> Algoritma <i>Multinomial Naive Bayes</i> dengan <i>Information Gain Threshold</i> 0.05 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	153
I.4	<i>Confusion Matrix</i> Algoritma <i>Multinomial Naive Bayes</i> dengan <i>Chi-Square Threshold</i> 0.01 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	154
I.5	<i>Confusion Matrix</i> Algoritma <i>Multinomial Naive Bayes</i> dengan <i>Chi-Square Threshold</i> 0.05 (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	155
I.6	<i>Confusion Matrix</i> Algoritma <i>Multinomial Naive Bayes</i> dengan <i>Model Dataset Multidimensional</i> (a) 90:10% (b) 80:20% (c) 70:30% (d) 60:40% (e) 50:50%	156

Daftar Tabel

2.1	Rangkuman Penelitian Terkait dan Kontribusi	14
2.3	<i>Confussion Matrix</i> [17]	30
3.1	Pembagian <i>Training</i> dan <i>Testing</i> Data untuk Klasifikasi Berita <i>Media Monitoring Geopark Ciletuh</i>	35
3.2	Tabel Kategori dan Topik Pembahasan Berita <i>Media Monitoring Geopark Ciletuh</i>	36
4.2	Fitur Pembentuk Dataset <i>Single Dimensional</i>	47
4.3	Tabel Hasil Perhitungan <i>Parent Entropy</i> , <i>Class Entropy</i> , dan <i>Information Gain</i> dari Sebagian Fitur	53
4.4	Tabel Jumlah Fitur yang Relevan atau Kurang Relevan Berdasarkan Pemilihan Fitur <i>Information Gain</i> dan <i>Chi-Square</i>	65
4.5	Hasil Uji Coba Model RF Terhadap <i>Testing</i> Data untuk <i>Dataset Single Dimensional</i> dengan <i>Grid Search: Recall, Precision, F1-Score, OOB</i>	66
4.6	Hasil Uji Coba Model RF Terhadap <i>Testing</i> Data untuk <i>Dataset Multidimensional</i> dengan <i>Grid Search: Recall, Precision, F1-Score, F1-Score, OOB Error Estimated</i> , dan Akurasi	67
4.7	Hasil Uji Coba Model RF terhadap <i>Training</i> Data untuk Seluruh <i>Dataset</i>	68
4.8	Hasil Uji Coba Model MNB dengan <i>Dataset Single Dimensional: Recall, Precision, F1-Score</i>	70
4.9	Hasil Uji Coba Model MNB dengan <i>Dataset Multidimensional</i>	70
4.10	Hasil Uji Coba Model MNB Terhadap <i>Training</i> Data	71
A.1	<i>Raw Dataset Media monitoring Geopark Ciletuh</i>	82
B.1	<i>Dataset Hasil PreProcessing</i>	97
C.1	Hasil Perhitungan Nilai <i>Information Gain</i> Seluruh Fitur	103
C.2	Fitur yang Kurang Relevan atau Informatif Berdasarkan Nilai <i>Threshold 0.01</i>	105
C.3	Fitur yang Kurang Relevan atau Informatif Berdasarkan Nilai <i>Threshold 0.05</i>	105
D.1	Hasil Perhitungan Nilai <i>Chi-Square</i> Seluruh Fitur	106
D.2	Fitur yang Kurang Relevan atau Informatif Berdasarkan <i>ChiSquare</i> dengan Nilai <i>Threshold 0.01</i>	107
D.3	Fitur yang Kurang Relevan atau Informatif Berdasarkan <i>ChiSquare</i> dengan Nilai <i>Threshold 0.05</i>	108
F.1	<i>Raw Dataset Multidimensional</i> Sesuai Model <i>Multidimensional</i> yang diusulkan Carsten Reinemman [32]	113
G.1	<i>Dataset Multidimensional</i> Setelah dilakukan Pembobotan Secara Manual	131
H.1	Hasil Prediksi Algoritma <i>Random Forest</i> Terhadap Data <i>Testing</i>	141

H.2	Hasil Prediksi Algoritma <i>Random Forest</i> Terhadap Data <i>Training</i>	148
I.2	Hasil Prediksi Algoritma <i>Multinomial Naive Bayes</i> Terhadap Data <i>Testing</i>	150
I.3	Hasil Prediksi Algoritma <i>Multinomial Naive Bayes</i> Terhadap Data <i>Training</i>	157

Daftar Singkatan

PR	: <i>Public Relations</i>
APJII	: Asosiasi Penyelenggara Jasa Internet Indonesia
SVM	: <i>Support Vector Machine</i>
MNB	: <i>Multinomial Naive Bayes</i>
RF	: <i>Random Forest</i>
SMO	: <i>Sequential Minimal Optimization</i>
UGG	: UNESCO Global Geopark
LDA	: <i>Latent Dirichlet Allocation</i>
UNESCO	: <i>United Nations of Educational, Scientific, and Cultural Organization</i>
IR	: <i>Information Retrieval</i>
SVD	: <i>Singular Value Decomposition</i>
TF-IDF	: <i>Term Frequency-Inverse Document Frequency</i>
BNB	: <i>Bernoulli Naive Bayes</i>
NLTK	: <i>Natural Language Tools Kit</i>

Daftar Rumus

2.1	TF	20
2.2	IDF	20
2.3	TF-IDF	20
2.4	<i>Entropy</i>	22
2.5	<i>Information Gain</i>	22
2.6	<i>Gini Index</i>	25
2.7	<i>Multinomial Naïve Bayes : Probabilitas term / class</i>	28
2.8	<i>Multinomial Naïve Bayes : Probabilitas prior</i>	28
2.9	<i>Multinomial Naïve Bayes : Probabilitas term atau kata ke-n</i>	28
2.10	<i>Multinomial Naïve Bayes : Maximum a Posteriori</i>	29
2.11	<i>Multinomial Naïve Bayes : Logaritma Maximum a Posteriori</i>	29
2.12	<i>Accuracy</i>	30
2.13	<i>Precision</i>	30
2.14	<i>Recall</i>	30
2.15	<i>F1-Score</i>	30
4.1	Entropy Parent	50
4.2	Perhitungan Entropi Atribut 1 pada Fitur "Banjir"	52
4.3	Perhitungan Entropi Atribut 2 pada Fitur "Banjir"	52
4.4	Perhitungan <i>Class Entropy</i> dari Fitur "Banjir"	53
4.5	Perhitungan Nilai <i>Information Gain</i> dari Fitur "Banjir"	53