



**KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN *K-MEANS*
SMOTE DAN *MULTINOMIAL NAÏVE BAYES* PADA PESAN BENCANA**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer**

**Oleh
NOR INDRIANI
NIM. 1611016120010**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023



**KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN *K-MEANS*
SMOTE DAN *MULTINOMIAL NAÏVE BAYES* PADA PESAN BENCANA**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer**

**Oleh
NOR INDRIANI
NIM. 1611016120010**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2023**

SKRIPSI

KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN *K-MEANS* *SMOTE* DAN *MULTINOMIAL NAÏVE BAYES* PADA PESAN BENCANA

Oleh:

NOR INDRIANI

NIM.1611016120010

Telah dipertahankan di depan Dosen Penguji pada tanggal 09 Juni 2023

Susunan Dosen Penguji:

Pembimbing I



Mohammad Reza Faisal, S.T., M.T., Ph.D
NIP.197612202008121001

Dosen Penguji I



Dwi Kartini, S.Kom., M.Kom
NIP.198704212012122003

Pembimbing II



Irwan Budiman, S.T., M.Kom
NIP.197703252008121001

Dosen Penguji II



Friska Abadi, S.Kom., M.Kom
NIP.19880913201612110001

Banjarbaru, Juni 2023

Ketua Program Studi Ilmu Komputer



Irwan Budiman, S.T., M.Kom
NIP.197703252008121001

PERYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diberikan orang lain, kecuali yang secara tertulis diacu didalam naskah ini dan disebutkan dalam Daftar Pustaka

Banjarbaru, Juni 2023



Nor Indriani
NIM.1611016120010

ABSTRAK

KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN *K-MEANS SMOTE* DAN *MULTINOMIAL NAÏVE BAYES* PADA PESAN BENCANA

(Oleh : Nor Indriani; Pembimbing : Mohammad Reza Faisal, S.T, M.T, Ph.D dan Irwan Budiman, S.T., M.Kom; 2023; 90 halaman)

Bencana dapat terjadi kapan saja dan di mana saja, bencana banjir dan kebakaran hutan adalah dua diantara jenis bencana yang terjadi Indonesia. Provinsi Kalimantan Selatan merupakan wilayah yang tidak luput dari bencana banjir dan kebakaran hutan. Dataset yang digunakan untuk penelitian ini adalah data bencana banjir dan kebakaran hutan dari penelitian sebelumnya oleh Zahra (2020) dan data yang di gunakan tidak seimbang. Kondisi data tidak seimbang bisa menyulitkan metode klasifikasi dalam melakukan proses pada data mining. Metode sampling pada pendekatan level data yang bisa digunakan untuk menyelesaikan permasalahan ketidakseimbangan adalah oversampling, salah satu turunan dari oversampling yaitu *SMOTE*. Metode *K-Means SMOTE* merupakan hasil modifikasi dari *SMOTE*. Salah satu model dari *Naïve Bayes* yang sering digunakan dalam klasifikasi teks adalah *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* memiliki performa yang baik dalam melakukan klasifikasi teks. Hasil penelitian ini pada data bencana banjir menggunakan *K-Means SMOTE* dengan *Multinomial Naïve Bayes* menghasilkan kinerja f1 score 66,04% dan data bencana kebakaran hutan menggunakan *K-Means SMOTE* dengan *Multinomial Naïve Bayes* menghasilkan kinerja f1 score 66,31%.

Kata Kunci : Data Tidak Seimbang, Pesan Bencana, *K-means SMOTE*, Klasifikasi, *Multinomial Naïve Bayes*

ABSTRACT

UNBALANCED DATA CLASSIFICATION USING *K-MEANS SMOTE* AND *MULTINOMIAL NAÏVE BAYES* IN DISASTER MESSAGES

(By : Nor Indriani; Supervisor : Mohammad Reza Faisal, S.T, M.T, Ph.D and Irwan Budiman, S.T., M.Kom; 2023; 90 pages)

Disasters can occur anytime and anywhere, floods and forest fires are two types of disasters that occur in Indonesia. South Kalimantan Province is an area that is not immune from floods and forest fires. The dataset used for this study is flood and forest fire data from previous research by Zahra (2020) and the data used is not balanced. Unbalanced data conditions can make it difficult for classification methods to process data mining. The sampling method for the data level approach that can be used to solve imbalance problems is oversampling, one of the derivatives of oversampling, namely *SMOTE*. The *K-Means SMOTE* method is a modification of *SMOTE*. One model of *Naïve Bayes* that is often used in text classification is *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* has a good performance in classifying text. The results of this study on flood disaster data using *K-Means SMOTE* with *Multinomial Naïve Bayes* yielded an f1 score of 66.04% and forest fire disaster data using *K-Means SMOTE* with *Multinomial Naïve Bayes* produced an f1 score of 66.31%.

Keywords : Unbalanced Data, Disaster Message, *K-means SMOTE*, Classification, *Multinomial Naïve Bayes*

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan kepada Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **“KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN *K-MEANS SMOTE* DAN *MULTINOMIAL NAÏVE BAYES* PADA PESAN BENCANA”** untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiama.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, diantaranya:

1. Diri saya sendiri yang tidak pernah patah semangat walaupun banyak menemui kesulitan baik disebabkan oleh diri sendiri maupun dari hal lain.
2. Keluarga terutama kedua orang tua saya bapak Sairi dan Ibu Kaswati yang selalu memberikan dukungan, doa, dan bantuan dalam proses penyelesaian skripsi ini.
3. Bapak Mohammad Reza Faisal, S.T., M.T., Ph.D selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku dosen pembimbing pendamping dan Koordinator Program Studi Ilmu Komputer FMIPA ULM yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
5. Ibu Dwi Kartini, S.Kom, M.Kom dan bapak Friska Abadi, S.Kom, M.Kom selaku dosen penguji yang banyak memberi saran dan meluangkan waktu untuk penyelesaian skripsi ini.
6. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.

7. Teman-teman saya yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.
8. Teman-teman keluarga Ilmu Komputer angkatan 2016 yang memberikan dukungan dan bantuan semasa kuliah dan selama proses penyelesaian skripsi.
9. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru Juni 2023

Nor Indriani

DAFTAR ISI

LEMBAR JUDUL	i
LEMBAR PENGESAHAN	ii
PERYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
BAB II TINJAUAN PUSTAKA	4
2.1 Kajian Terdahulu	4
2.2 Keaslian Penelitian.....	5
2.3 Bencana.....	8
2.4 Klasifikasi	10
2.5 Data Mining	11
2.6 Text Mining	12
2.7 TF-IDF	14
2.8 Data Tidak Seimbang.....	15
2.9 Oversampling.....	16
2.10 Cross Validation.....	23
2.11 Multinomial Naïve Bayes	24
2.12 Confusion Matrix	25
BAB III METODE PENELITIAN	29
3.1 Alat Penelitian.....	29
3.2 Bahan Penelitian	29
3.3 Prosedur Penelitian	29

BAB V HASIL DAN PEMBAHASAN	33
4.1 Hasil.....	33
4.1.1 Pengumpulan Data.....	33
4.1.2 Preprocessing Data.....	36
4.1.3 Pembobotan TF-IDF.....	43
4.1.4 Cross Validation.....	45
4.1.5 Klasifikasi.....	46
4.2 Pembahasan.....	61
BAB IV PENUTUP	64
5.1 Kesimpulan.....	64
5.2 Saran.....	64
DAFTAR PUSTAKA	65
LAMPIRAN	69
RIWAYAT HIDUP	76

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian	6
Tabel 2. Perancangan Penelitian	7
Tabel 3. Confusion Matrix	25
Tabel 4. Confusion matrix multiclass	27
Tabel 5. Data Banjir dan Kebakaran Hutan	30
Tabel 6. Contoh data bencana banjir.....	33
Tabel 7. Jumlah data bencana banjir	34
Tabel 8. Contoh data bencana kebakaran hutan.....	34
Tabel 9. Jumlah data kebakaran hutan	35
Tabel 10. Penerapan cleansing.....	36
Tabel 11. Penerapan case folding	37
Tabel 12. Penerapan tokenizing	38
Tabel 13. Kamus stopwords	40
Tabel 14. Penerapan filtering	41
Tabel 15. Penerapan stemming	42
Tabel 16. Contoh perhitungan TF-IDF	43
Tabel 17. Hasil pembobotan kata TF-IDF	45
Tabel 18. 10 K-fold cross validation.....	46
Tabel 19. <i>Multinomial Naïve Bayes</i> data bencana banjir	48
Tabel 20. Perhitungan presisi <i>Multinomial Naïve Bayes</i> data bencana banjir	48
Tabel 21. Hasil presisi <i>Multinomial Naïve Bayes</i> data bencana banjir	48
Tabel 22. Perhitungan recall <i>Multinomial Naïve Bayes</i> data bencana banjir.....	49
Tabel 23. Hasil recall <i>Multinomial Naïve Bayes</i> data bencana banjir.....	49
Tabel 24. Perhitungan f1 score <i>Multinomial Naïve Bayes</i> data bencana banjir.....	49
Tabel 25. Hasil f1 score <i>Multinomial Naïve Bayes</i> data bencana banjir.....	49
Tabel 26. <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan	50
Tabel 27. Perhitungan presisi <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan	50
Tabel 28. Hasil presisi <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan	51
Tabel 29. Perhitungan recall <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan.....	51
Tabel 30. Hasil recall <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan.....	51
Tabel 31. Perhitungan f1 score <i>Multinomial Naïve Bayes</i> data kebakaran hutan	51
Tabel 32. Hasil f1 score <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan.....	51

Tabel 33. <i>SMOTE</i> dengan <i>Multinomial Naïve Bayes</i> data bencana banjir.....	52
Tabel 34. Perhitungan presisi data bencana banjir menggunakan <i>SMOTE</i>	53
Tabel 35. Hasil presisi data bencana banjir menggunakan <i>SMOTE</i>	53
Tabel 36. Perhitungan recall data bencana banjir menggunakan <i>SMOTE</i>	53
Tabel 37. Hasil recall data bencana banjir menggunakan <i>SMOTE</i>	53
Tabel 38. Perhitungan f1 score data bencana banjir menggunakan <i>SMOTE</i>	53
Tabel 39. Hasil f1 score data bencana banjir menggunakan <i>SMOTE</i>	54
Tabel 40. <i>SMOTE</i> dengan <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan.....	54
Tabel 41. Perhitungan presisi data bencana kebakaran hutan menggunakan <i>SMOTE</i>	55
Tabel 42. Hasil presisi data bencana kebakaran hutan menggunakan <i>SMOTE</i>	55
Tabel 43. Perhitungan recall data bencana kebakaran hutan menggunakan <i>SMOTE</i>	55
Tabel 44. Hasil recall data bencana kebakaran hutan menggunakan <i>SMOTE</i>	55
Tabel 45. Perhitungan f1 score data bencana kebakaran hutan menggunakan <i>SMOTE</i>	56
Tabel 46. Hasil f1 score data bencana kebakaran hutan menggunakan <i>SMOTE</i>	56
Tabel 47. <i>K-Means SMOTE</i> dengan <i>Multinomial Naïve Bayes</i> data bencana banjir	57
Tabel 48. Perhitungan presisi menggunakan <i>K-Means SMOTE</i> data bencana banjir	57
Tabel 49. Hasil presisi data bencana banjir menggunakan <i>K-Means SMOTE</i>	57
Tabel 50. Perhitungan recall menggunakan <i>K-Means SMOTE</i> data bencana banjir.....	58
Tabel 51. Hasil recall data bencana banjir menggunakan <i>K-Means SMOTE</i>	58
Tabel 52. Perhitungan f1 score menggunakan <i>K-Means SMOTE</i> data bencana banjir.....	58
Tabel 53. Hasil f1 score data bencana banjir menggunakan <i>K-Means SMOTE</i>	58
Tabel 54. <i>K-Means SMOTE</i> dengan <i>Multinomial Naïve Bayes</i> data bencana kebakaran hutan	59
Tabel 55. Perhitungan presisi menggunakan <i>K-Means SMOTE</i> data bencana kebakaran hutan.....	60
Tabel 56. Hasil presisi data bencana kebakaran hutan menggunakan <i>K-Means SMOTE</i>	60
Tabel 57. Perhitungan recall menggunakan <i>K-Means SMOTE</i> data bencana kebakaran hutan	60
Tabel 58. Hasil recall data bencana kebakaran hutan menggunakan <i>K-Means SMOTE</i>	60
Tabel 59. Perhitungan f1 score menggunakan <i>K-Means SMOTE</i> data bencana kebakaran hutan.....	60
Tabel 60. Hasil f1 score data bencana kebakaran hutan menggunakan <i>K-Means SMOTE</i>	61

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Tahapan Proses Text Mining	12
Gambar 2. Ilustrasi algoritma <i>SMOTE</i>	19
Gambar 3. Alur dari algoritma <i>K-means SMOTE</i>	21
Gambar 4. Prinsip kerja <i>K-means SMOTE</i>	22
Gambar 5. Penjelasan 10 Folds Cross Validation.....	24
Gambar 6. Alur penelitian.....	30
Gambar 7. Perbandingan kinerja f1 score data bencana banjir.....	61
Gambar 8. Perbandingan kinerja f1 score data bencana kebakaran hutan.....	62
Gambar 9. Hasil rata-rata kinerja f1 score data bencana banjir dan kebakaran hutan	63

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1. Source code cleansing dan case folding	70
Lampiran 2. Source code tokenizing, filtering dan stemming	71
Lampiran 3. Source code pembobotan kata TF-IDF.....	71
Lampiran 4. Source code klasifikasi <i>Multinomial Naïve Bayes</i>	72
Lampiran 5. Source code <i>SMOTE</i> dengan klasifikasi <i>Multinomial Naïve Bayes</i>	73
Lampiran 6. Source code <i>K-Means SMOTE</i> dengan klasifikasi <i>Multinomial Naïve Bayes</i>	74