



**PENYEIMBANGAN KELAS SMOTE DAN SELEKSI FITUR *ENSEMBLE*
FILTER PADA *SUPPORT VECTOR MACHINE* UNTUK KLASIFIKASI
PENYAKIT *LIVER***

SKRIPSI

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

**Oleh
MUHAMMAD AMIR NUGRAHA
NIM 1911016110002**

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

JUNI 2023

SKRIPSI

PENYEIMBANGAN KELAS SMOTE DAN SELEKSI FITUR *ENSEMBLE* FILTER PADA *SUPPORT VECTOR MACHINE* UNTUK KLASIFIKASI PENYAKIT *LIVER*

Oleh:

MUHAMMAD AMIR NUGRAHA
NIM. 1911016110002

telah dipertahankan di depan Dosen Penguji pada tanggal 22 Juni 2023
Susunan Penguji:

Pembimbing Utama



M. Itqan Mazdadi, S.Kom., M.Kom.
NIP. 199006122019031013

Dosen Penguji I



Muliadi, S.Kom., M.Sc.
NIP. 197804222010121002

Pembimbing Pendamping



Andi Farmadi, S.Si., M.T.
NIP. 197307252008011006

Dosen Penguji II



Triando H. Saragih, S.Kom., M.Kom.
NIP. 199308242019031012



Banjarbaru, 22 Juni 2023
Koordinator Program Studi Ilmu Komputer

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

PERNYATAAN

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 diterbitkan oleh orang lain, kecuali yang secara tertulis diacu di dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 22 Juni 2023
Yang menyatakan,



Muhammad Amir Nugraha
NIM. 1911016110002

ABSTRAK

PENYEIMBANGAN KELAS SMOTE DAN SELEKSI FITUR *ENSEMBLE* FILTER PADA *SUPPORT VECTOR MACHINE* UNTUK KLASIFIKASI PENYAKIT *LIVER*

(Oleh: Muhammad Amir Nugraha; Pembimbing: Muhammad Itqan Mazdadi, S.Kom., M.Kom. dan Andi Farmadi, S.Si., M.T; 2023; 69 halaman)

Liver merupakan salah satu organ penting dalam tubuh manusia yang berperan dalam proses metabolisme tubuh. Mengutip artikel dari *American Liver Foundation*, pada tahun 2020 sebanyak 51.642 orang dewasa di Amerika Serikat meninggal akibat penyakit *liver*. Data hasil tes fungsi *liver* dari laboratorium dapat digunakan untuk mendiagnosis penyakit *liver*. Klasifikasi penyakit *liver* pada pasien perlu dilakukan dengan baik karena hasilnya dapat membantu dalam diagnosis awal apakah seorang pasien mengidap penyakit *liver*. Berdasarkan penelitian sebelumnya, metode *Support Vector Machine* (SVM) paling baik dalam mengklasifikasikan pasien penyakit *liver*. Namun, SVM memiliki kelemahan ketika diterapkan pada *dataset* dengan kelas yang tidak seimbang dan tidak bekerja secara akurat ketika terlalu banyak fitur yang tidak relevan digunakan. Untuk menyeimbangkan kelas pada *dataset*, digunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Sedangkan untuk seleksi fitur dilakukan menggunakan metode *Ensemble Filter*, terdiri dari metode *Information Gain*, *Gain Ratio*, dan Relief-F untuk menangani fitur-fitur tidak relevan. Berdasarkan hasil pengujian, penerapan SMOTE dan *Ensemble Filter* pada metode klasifikasi SVM memberikan hasil terbaik dengan nilai *accuracy* sebesar 85% dan AUC sebesar 0.850. Dari pengujian tersebut dapat diketahui jika SMOTE pada penyeimbangan kelas dan *Ensemble Filter* pada seleksi fitur dapat meningkatkan performa klasifikasi dari metode SVM.

Kata kunci: *Liver*, Klasifikasi, SVM, SMOTE, *Ensemble Filter*

ABSTRACT

SMOTE CLASS BALANCING AND ENSEMBLE FILTER FEATURE SELECTION IN SUPPORT VECTOR MACHINE FOR LIVER DISEASE CLASSIFICATION

(By: Muhammad Amir Nugraha; Supervisors: Muhammad Itqan Mazdadi, S.Kom, M., Kom and Andi Farmadi, S.Si., M.T; 2023; 69 pages)

The liver is one of the important organs in the human body that plays a role in the body's metabolic processes. Quoting an article from the American Liver Foundation, in 2020, as many as 51,642 adults in the United States died from liver disease. Liver function test data from the laboratory can be used to diagnose liver disease. Classification of liver disease in patients needs to be done well because the results can help in the initial diagnosis of whether a patient has liver disease. Based on previous research, the Support Vector Machine (SVM) method best classifies liver disease patients. However, SVM has weaknesses when applied to datasets with unbalanced classes and does not work accurately when too many irrelevant features are used. To class-balance the dataset, the Synthetic Minority Oversampling Technique (SMOTE) method is used. Meanwhile, feature selection is performed using the Ensemble Filter method, which consists of Information Gain, Gain Ratio, and Relief-F methods to handle irrelevant features. Based on the test results, the application of SMOTE and Ensemble Filter in SVM classification gives the best results with an accuracy value of 85% and an AUC of 0.850. The test can prove if SMOTE on class balancing and Ensemble Filter on feature selection can improve the classification performance of the SVM method.

Keywords: *Liver, Classification, SVM, SMOTE, Ensemble Filter*

PRAKATA

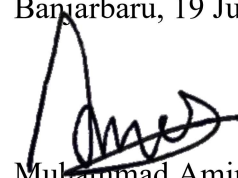
Bismillaahirrahmaanirrahiim, Puji dan Syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas limpahan Rahmat dan Petunjuk dari-Nya penulis dapat menyelesaikan Skripsi ini. Adapun Skripsi dengan judul “Penyeimbangan Kelas SMOTE dan Seleksi Fitur *Ensemble Filter* pada *Support Vector Machine* untuk Klasifikasi Penyakit *Liver*” ditulis untuk memenuhi syarat dalam penyelesaian program Sarjana S1 Ilmu Komputer di FMIPA ULM.

Pada kesempatan ini, penulis menyampaikan ucapan terima kasih kepada seluruh pihak yang terlibat dalam penyelesaian Skripsi ini. Pihak-pihak yang dimaksud yaitu:

1. Keluarga dirumah atas doa, kejujuran, dan dukungan selama penulis mengerjakan Skripsi.
2. Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom selaku Dosen Pembimbing Utama dan Bapak Andi Farmadi, S.Si., M.T selaku Dosen Pembimbing Pendamping atas bantuan, dukungan, dan waktu yang diberikan kepada penulis selama pengerjaan Skripsi berlangsung.
3. Bapak dan Ibu Dosen serta Staff Ilmu Komputer FMIPA ULM atas semua ilmu yang diberikan sehingga penulis dapat mengerjakan Skripsi ini.
4. Kawan-kawan Ilmu Komputer 2019 atas dukungan, semangat, dan doa yang diberikan kepada penulis selama pengerjaan Skripsi.
5. Orang-orang yang belum dapat dituliskan satu persatu dikesempatan ini.

Penulis menyadari, jika Skripsi ini masih memiliki kekurangan, maka dari itu penulis mengharapkan kritik dan saran yang positif dari para pembaca, sehingga Skripsi ini menjadi lebih baik. Semoga Skripsi ini mendapat Ridho dari Tuhan Yang Maha Esa, dan bermanfaat bagi perkembangan ilmu pengetahuan.

Banjarbaru, 19 Juni 2023



Muhammad Amir Nugraha

DAFTAR ISI

SKRIPSI.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu	5
2.2 Landasan Teori	14
2.2.1 Kelas Tidak Seimbang	14
2.2.2 SMOTE	14
2.2.3 Seleksi Fitur.....	16
2.2.4 <i>Ensemble</i> Filter.....	17
2.2.5 <i>Information Gain</i>	19

2.2.6	<i>Gain Ratio</i>	20
2.2.7	Relief-F.....	21
2.2.8	<i>Entropy Based Discretization</i>	22
2.2.9	<i>Support Vector Machine</i>	24
2.2.10	Penyakit <i>Liver</i>	28
2.2.11	Evaluasi	28
2.2.12	<i>Missing Value</i>	30
2.2.13	<i>Z-Score</i>	30
2.2.14	<i>Label Encoding</i>	31
2.2.15	Pembagian Data.....	32
BAB III METODE PENELITIAN.....		33
3.1	Bahan Penelitian	33
3.2	Alat Penelitian	33
3.3	Variabel Penelitian.....	33
3.4	Prosedur Penelitian	33
BAB IV HASIL DAN PEMBAHASAN		39
4.1	Hasil.....	39
4.1.1	Pengumpulan Data	39
4.1.2	<i>Preprocessing</i> Data	40
4.1.3	SMOTE	42
4.1.4	<i>Ensemble</i> Filter.....	46
4.1.5	Pembagian Data.....	50
4.1.6	Klasifikasi SVM.....	50
4.1.7	Evaluasi	53
4.2	Pembahasan	55

BAB V PENUTUP	62
5.1 Kesimpulan	62
5.2 Saran	62
DAFTAR PUSTAKA	63
LAMPIRAN	
RIWAYAT HIDUP	

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	8
Tabel 2. Perancangan Penelitian	13
Tabel 3. <i>Confusion Matrix</i>	29
Tabel 4. Fitur pada <i>Dataset</i>	39
Tabel 5. Jumlah Data di Setiap Label Kelas	39
Tabel 6. Contoh Data pada ILPD	40
Tabel 7. Jumlah <i>Missing Value</i>	40
Tabel 8. Contoh Data Hasil Normalisasi <i>Z-Score</i>	41
Tabel 9. Contoh Data Hasil Transformasi dengan Label <i>Encoding</i>	42
Tabel 10. Contoh Data Hasil Diskretisasi dengan <i>Entropy Based Discretization</i>	46
Tabel 11. Hasil Penilaian Metode Filter Tunggal	47
Tabel 12. Proses dan Hasil Seleksi Fitur <i>Ensemble</i>	48
Tabel 13. Kumpulan Fitur Optimal Berdasarkan <i>Rate of Occurrence</i>	49
Tabel 14. Distribusi Data dari Setiap Kelas Pada Setiap Model Klasifikasi	50
Tabel 15. <i>Confusion Matrix</i> Model Klasifikasi SVM.....	51
Tabel 16. <i>Confusion Matrix</i> Model Klasifikasi SMOTE-SVM.....	52
Tabel 17. <i>Confusion Matrix</i> Model Klasifikasi SMOTE- <i>Ensemble</i> Filter-SVM dengan Kumpulan Fitur Optimal1	53
Tabel 18. <i>Confusion Matrix</i> Model Klasifikasi SMOTE- <i>Ensemble</i> Filter-SVM dengan Kumpulan Fitur Optimal2.....	53
Tabel 19. <i>Confusion Matrix</i> Model Klasifikasi SMOTE- <i>Ensemble</i> Filter-SVM dengan Kumpulan Fitur Optimal3	53
Tabel 20. Hasil Evaluasi Model Klasifikasi SVM.....	54
Tabel 21. Hasil Evaluasi Model Klasifikasi SMOTE-SVM	54
Tabel 22. Hasil Evaluasi Model Klasifikasi SMOTE- <i>Ensemble</i> Filter-SVM dengan Kumpulan Fitur optimal1	55
Tabel 23. Hasil Evaluasi Model Klasifikasi SMOTE- <i>Ensemble</i> Filter-SVM dengan Kumpulan Fitur optimal2.....	55

Tabel 24. Hasil Evaluasi Model Klasifikasi SMOTE- <i>Ensemble</i> Filter-SVM dengan Kumpulan Fitur optimal3	55
Tabel 25. Nilai Parameter <i>Default Library</i> SVM	58

DAFTAR GAMBAR

Gambar 1. Ilustrasi pembuatan data sintetis dengan SMOTE	16
Gambar 2. Contoh visualisasi <i>margin</i> kecil	24
Gambar 3. Contoh visualisasi <i>margin</i> maksimal	25
Gambar 4. Visualisasi pencarian <i>hyperplane</i> pada data tidak linier	25
Gambar 5. Prosedur penelitian	34
Gambar 6. Tahapan dalam <i>Ensemble Filter</i>	36
Gambar 7. Alur Klasifikasi SVM.....	37
Gambar 8. Alur Klasifikasi SMOTE-SVM.....	37
Gambar 9. Alur Klasifikasi SMOTE- <i>Ensemble Filter</i> -SVM	38
Gambar 10. Persentase Jumlah Data pada masing-masing Kelas.....	43
Gambar 11. Sebaran Data dari Beberapa Fitur pada <i>Dataset</i>	44
Gambar 12. Sebaran Data dari Beberapa Fitur <i>Dataset</i> Pasca SMOTE	45
Gambar 13. Persentase Jumlah Data Setiap Kelas Pasca SMOTE	45
Gambar 14. Perbandingan Nilai <i>Accuracy</i> dan AUC antar Kumpulan Fitur Optimal Hasil Seleksi <i>Ensemble Filter</i>	60
Gambar 15. Perbandingan Nilai <i>Accuracy</i> dan AUC antar Model Klasifikasi.....	61

DAFTAR LAMPIRAN

- Lampiran 1. *Source Code* Pengumpulan Data
- Lampiran 2. *Source Code Preprocessing* Data
- Lampiran 3. *Source Code* Model Klasifikasi SVM
- Lampiran 4. *Source Code* SMOTE
- Lampiran 5. *Source Code* Model Klasifikasi SMOTE+SVM
- Lampiran 6. *Source Code* Unduh *Dataset* pasca SMOTE untuk Seleksi Fitur
- Lampiran 7. Gambar Seleksi Fitur *Ensemble* Filter (Tahap 1) Diskritisasi Data Kontinyu dan Penilaian Fitur Oleh Metode Filter Tunggal
- Lampiran 8. *Source Code* Transformasi Fitur dari Dataset Pasca Seleksi Fitur
- Lampiran 9. *Source Code* Model Klasifikasi SMOTE+*Ensemble* Filter+SVM
- Lampiran 10. *Source Code* Evaluasi Kinerja Model Klasifikasi