

**PENERAPAN TEKNIK RANDOM OVERSAMPLING UNTUK
MENGATASI IMBALANCED CLASS DATA PADA
KLASIFIKASI TINGKAT KEBUGARAN TUBUH MANUSIA
MENGGUNAKAN METODE K-NEAREST NEIGHBOR**

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh:

Rachman Dimas Saputra
NIM: 09021281823055

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2023**

LEMBAR PENGESAHAN SKRIPSI

**PENERAPAN TEKNIK RANDOM OVERSAMPLING UNTUK
MENGATASI IMBALANCED CLASS DATA PADA
KLASIFIKASI TINGKAT KEBUGARAN TUBUH MANUSIA
MENGGUNAKAN METODE K-NEAREST NEIGHBOR**

Oleh :

Rachman Dimas Saputra

NIM : 09021281823055

Palembang, 16 Januari 2023

Pembimbing I,



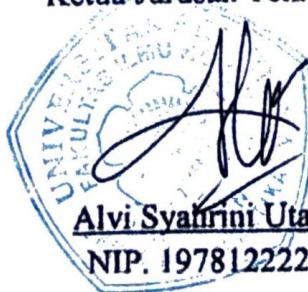
Dian Palupi Rinri, M.Kom, Ph.D.
NIP. 197802232006042002

Pembimbing II,



Anggina Primanita, M.IT, Ph.D.
NIP. 198908062015042002

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Selasa, 3 Januari 2023 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Rachman Dimas Saputra

NIM : 09021281823055

Judul : Penerapan Teknik *Random Oversampling* untuk Mengatasi *Imbalanced Class Data* pada Klasifikasi Tingkat Kebugaran Tubuh Manusia Menggunakan Metode *K-Nearest Neighbor*

Dan dinyatakan **LULUS**.

1. Ketua Penguji

Dr. Muhammad Fachrurrozi, M.T.
NIP. 198005222008121002

2. Pembimbing I

Dian Palupi Rini, M.Kom, Ph.D.
NIP. 197802232006042002

3. Pembimbing II

Anggina Primanita, M.IT, Ph.D.
NIP. 198908062015042002

4. Penguji

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Rachman Dimas Saputra
NIM : 09021281823055
Program Studi : Teknik Informatika
Judul : Penerapan Teknik *Random Oversampling* untuk Mengatasi *Imbalanced Class Data* pada Klasifikasi Tingkat Kebugaran Tubuh Manusia Menggunakan Metode *K-Nearest Neighbor*

Hasil Pengecekan Software iThenticate/Turnitin : 7%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan saya buat dengan sebenarnya dan tidak ada paksaan dari siapapun.



Palembang, 19 Januari 2023

Rachman Dimas Saputra
NIM.09021281823055

MOTTO DAN PERSEMBAHAN

“Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum, sebelum mereka mengubah keadaan diri mereka sendiri.”

(Q.S. Ar Rad : 11)

“Akan selalu ada jalan menuju sebuah kesuksesan bagi siapapun, selama orang tersebut mau berusaha dan bekerja keras untuk memaksimalkan kemampuan yang ia miliki.”

(Bambang Pamungkas)

“Tidak mustahil bagi orang biasa untuk memutuskan menjadi luar biasa.”

(Elon Musk)

Kupersembahkan Karya Tulis Ini Kepada:

- **Orang Tua dan Kakak saya**
- **Keluarga Besar**
- **Sahabat dan Teman-teman saya**
- **Dosen Pembimbing dan Penguji**
- **Teman-teman Seperjuangan**

Teknik Informatika

- **Fakultas Ilmu Komputer**
- **Universitas Sriwijaya**

**APPLICATION OF RANDOM OVERSAMPLING TECHNIQUE TO
OVERCOME IMBALANCED CLASS DATA IN HUMAN FITNESS LEVEL
CLASSIFICATION USING K-NEAREST NEIGHBOR METHOD**

by:

Rachman Dimas Saputra
09021281823055

ABSTRACT

Physical fitness is the body's ability to carry out physical activities without causing excessive fatigue, knowing a level of body fitness is useful for determining the right solution to overcome problems related to body fitness. A classification process with machine learning is needed to determine the level of body fitness, one method that is often used is the K-NN method. Classification using machine learning often has a problem with class imbalance that causes errors in classification, to overcome this problem a data balancing method is needed. The random oversampling technique is a data balancing technique by randomly adding minority class samples until the number of samples is equal to the majority class. The results obtained after applying random oversampling decreased the average accuracy by 4% and the average precision by 6%, but there was an increase in the average recall value by 4%. This research proves that the random oversampling technique can make the recall value higher.

Keywords: Classification, Physical fitness, Imbalanced Data, Random Oversampling, K-Nearest Neighbor

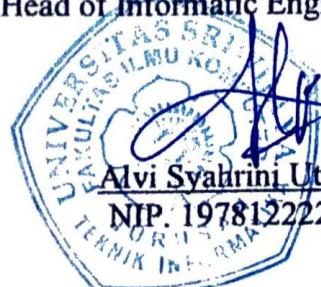
Supervisor I

Dian Palupi Rini, M.Kom, Ph.D.
NIP. 197802232006042002

Palembang, January 16th 2023
Supervisor II

Anggina Primanita, M.IT, Ph.D.
NIP. 198908062015042002

Approve,
Head of Informatic Engineering Department



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

**PENERAPAN TEKNIK RANDOM OVERSAMPLING UNTUK
MENGATASI IMBALANCED CLASS DATA PADA KLASIFIKASI
TINGKAT KEBUGARAN TUBUH MANUSIA MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

Oleh:
Rachman Dimas Saputra
09021281823055

ABSTRAK

Kebugaran jasmani adalah kemampuan tubuh dalam menjalankan aktivitas fisik tanpa menimbulkan kelelahan secara berlebihan, mengetahui suatu tingkat kebugaran tubuh berguna untuk menentukan solusi yang tepat untuk mengatasi permasalahan yang berkaitan dengan kebugaran tubuh. Diperlukan suatu proses klasifikasi dengan *machine learning* untuk menentukan tingkat kebugaran tubuh, salah satu metode yang sering digunakan adalah metode K-NN. Klasifikasi menggunakan *machine learning* seringkali memiliki masalah pada ketidakseimbangan kelas yang menyebabkan kekeliruan dalam melakukan klasifikasi, untuk mengatasi permasalahan ini diperlukan suatu metode penyeimbangan data. Teknik *random oversampling* merupakan teknik penyeimbangan data dengan menambahkan sampel kelas minoritas secara acak sampai jumlah sampel tersebut sama dengan kelas mayoritas. Hasil yang didapat setelah diterapkan *random oversampling* terdapat penurunan pada rata-rata akurasi sebesar 4% dan rata-rata presisi 6%, namun terdapat peningkatan pada rata-rata nilai recall sebesar 4%. Penelitian ini membuktikan bahwa teknik *random oversampling* dapat membuat nilai *recall* menjadi lebih tinggi.

Kata Kunci : Klasifikasi, Kebugaran Tubuh, Data Tidak Seimbang, *Random Oversampling*, *K-Nearest Neighbor*

Pembimbing I

Dian Palupi Rini, M.Kom, Ph.D.
NIP. 197802232006042002

Palembang, 16 Januari 2023
Pembimbing II

Anggina Primanita, M.I.T, Ph.D
NIP. 198908062015042002



KATA PENGANTAR

Puji syukur kepada Allah Subhanahu Wa Ta’ala, atas segala karunia dan rahmat-Nya kepada penulis sehingga dapat menyelesaikan Tugas Akhir yang berjudul “Penerapan Teknik *Random Oversampling* untuk Mengatasi *Imbalanced Class Data* pada Klasifikasi Tingkat Kebugaran Tubuh Manusia Menggunakan Metode *K-Nearest Neighbor*” guna memenuhi salah satu syarat untuk menyelesaikan pendidikan program strata-1 di Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Pada kesempatan ini Penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi selama proses penelitian ini dilaksanakan. Secara khusus Penulis ingin mengucapkan terima kasih kepada :

1. Kedua Orang Tua, Edy Susiyanto dan Masdah serta kakak Rachmania Putri Utami, yang telah memberikan doa serta dukungan yang sangat besar selama mengikuti dan melaksanakan perkuliahan di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya hingga penulis dapat menyelesaikan Skripsi ini.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
3. Ibu Alvi Syahrini Utami, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Nabila Rizky Oktadini, M.T., selaku Dosen Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Dian Palupi Rini, M.Kom.,Ph.D. selaku dosen pembimbing I dan ibu Anggina Primanita, M.IT, Ph.D. selaku Dosen Pembimbing II yang telah membimbing serta mengarahkan saya untuk menyelesaikan Tugas Akhir ini

6. Mba Wiwin selaku admin Jurusan Teknik Informatika yang telah membantu mengurus seluruh berkas
7. Seluruh dosen dan staff Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Pratama Yanuarta yang senantiasa mengingatkan penulis untuk selalu berjuang menghadapi beratnya skripsi.
9. M Dimas Firmansyah dan Fadhl Syakur karena senantiasa meneman penulis dalam menghadapi kesulitan mengerjakan skripsi
10. Yuni Saputri yang senantiasa membantu dan memotivasi penulis dalam mengerjakan skripsi.
11. Teman satu kelas Kgs. M. Rusdiansyah Muharrom, Denta Mustofa, Hafizh Shafwan Rafa, Ucok Rahmat Hidayat, Muhammad Wahyu Pratama dan teman-teman IFBILA serta seluruh teman teman seperjuangan angkatan 2018 Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan Skripsi ini. Oleh karena itu, segala saran dan kritik sangatlah penting bagi penulis. Akhir kata, semoga Skripsi ini dapat bermanfaat dan berguna bagi khalayak, khususnya mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya.

Palembang, 28 Desember 2022



Rachman Dimas Saputra

DAFTAR ISI

Halaman

HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI.....	ii
HALAMAN TANDA LULUS UJIAN SIDANG SKRIPSI.....	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN MOTTO DAN PERSEMBERAHAN.....	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xvi
DAFTAR LAMPIRAN.....	,,, xviii
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah.....	I-4
1.4 Tujuan	I-5
1.5 Manfaat.....	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan.....	II-1
2.2 Landasan Teori	II-1
2.2.1 Kebugaran Tubuh	II-1
2.2.2 <i>Machine learning</i>	II-4
2.2.3 Klasifikasi.....	II-5

2.2.4	<i>Imbalanced Class</i>	II-6
2.2.5	Normalisasi Data	II-6
2.2.6	<i>K-Nearest Neighbor (K-NN)</i>	II-7
2.2.7	<i>Random Oversampling</i>	II-9
2.2.8	<i>Split Validation</i>	II-9
2.2.9	<i>Confusion Matrix</i>	II-9
2.2.10	<i>Rational Unified Process (RUP)</i>	II-12
2.3	Penelitian Terkait.....	II-13
2.3.1	Penerapan Algoritma <i>Naive Bayes</i> untuk Klasifikasi Tingkat Kebugaran Jasmani Berdasarkan Hasil Pengukuran pada Pegawai	II-13
2.3.2	<i>Improving k-Nearest Neighbors Algorithm for Imbalanced Data Classification</i>	II-14
2.3.3	Penerapan Algoritma <i>K-Nearest Neighbor</i> Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada “Lombok Vape On”	II-15
2.3.4	Penanganan <i>Imbalance Data</i> Pada Klasifikasi Kemungkinan Penyakit Stroke.....	II-16
2.4	Kesimpulan.....	II-17
BAB III	METODOLOGI PENELITIAN	III-1
3.1	Pendahuluan.....	III-1
3.2	Pengumpulan Data.....	III-1
3.2.1	Jenis dan Sumber Data	III-1
3.2.2	Metode Pengumpulan Data	III-2
3.3	Tahapan Penelitian	III-2
3.3.1	Menetapkan Kerangka Kerja Penelitian.....	III-3
3.3.2	Kriteria Pengujian.....	III-6
3.3.3	Format Data Pengujian	III-7
3.3.4	Alat Yang Digunakan dalam Pelaksanaan Penelitian	III-10
3.3.5	Pengujian Penelitian	III-10
3.3.6	Analisis Hasil Pengujian dan Kesimpulan	III-10
3.4	Metode Pengembangan Perangkat Lunak	III-11

3.4.1	Fase Insepsi	III-11
3.4.2	Fase Elaborasi.....	III-12
3.4.3	Fase Konstruksi	III-12
3.4.4	Fase Transisi.....	III-13
3.5	Manajemen Proyek Penelitian	III-13
3.6	Kesimpulan.....	III-18
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan.....	IV-1
4.2	Fase Insepsi.....	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem.....	IV-2
4.2.3	Analisis dan Desain	IV-3
4.3	Fase Elaborasi.....	IV-25
4.3.1	Pemodelan Bisnis	IV-25
4.3.2	Perancangan Data.....	IV-25
4.3.3	Perancangan Antarmuka.....	IV-25
4.3.4	Kebutuhan Sistem.....	IV-28
4.3.5	Diagram Aktivitas	IV-28
4.3.6	Diagram <i>Sequence</i>	IV-33
4.4	Fase Konstruksi	IV-36
4.4.1	Kebutuhan Sistem.....	IV-37
4.4.2	Diagram Kelas	IV-37
4.4.3	Implementasi	IV-38
4.5	Fase Transisi	IV-42
4.5.1	Pemodelan Bisnis	IV-43
4.5.2	Rencana Pengujian	IV-43
4.5.3	Implementasi	IV-44
4.4	Kesimpulan.....	IV-47
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan.....	V-1
5.2	Data Hasil Penelitian	V-1

5.2.1	Konfigurasi penelitian	V-1
5.2.2	Hasil pengujian Metode KNN	V-2
5.2.3	Hasil Pengujian Metode ROS-KNN.....	V-5
5.3	Analisis Hasil Penelitian.....	V-9
5.4	Kesimpulan.....	V-19
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Kesimpulan.....	VI-1
6.2	Saran	VI-3

DAFTAR PUSTAKA

DAFTAR TABEL

	Halaman
Tabel II- 1. Confusion Matrix Klasifikasi Biner.....	II-10
Tabel II- 2. Confusion Matrix Multiclass	II-11
Tabel III-1. Rancangan Tabel Hasil Pengujian.....	III-7
Tabel III-2. Tabel Confusion Matrix	III-8
Tabel III-3. Tabel Hasil Analisis Klasifikasi KNN	III-11
Tabel III-4. Tabel Hasil Analisis Klasifikasi ROS-KNN	III-11
Tabel III-5. Perencanaan Aktivitas Penelitian dalam bentuk WBS	III-14
Tabel IV-1. Tabel Kebutuhan Fungsional.....	IV-3
Tabel IV-2. Tabel Kebutuhan Non-Fungsional.....	IV-3
Tabel IV- 3. Data Latih K-Nearest Neighbor (KNN)	IV-6
Tabel IV- 4. Data Uji K-Nearest Neighbor	IV-6
Tabel IV- 5. Tabel Hasil Perhitungan Jarak	IV-9
Tabel IV- 6. Hasil Pengurutan Jarak	IV-10
Tabel IV- 7. Hasil Pengurutan Jarak Berdasarkan Nilai K	IV-10
Tabel IV- 8. Data Latih ROS-KNN	IV-12
Tabel IV- 9. Data Uji ROS-KNN.....	IV-13
Tabel IV- 10. Data Training Hasil proses Random Oversampling	IV-13
Tabel IV- 11. Hasil Perhitungan Jarak pada Metode ROS-KNN	IV-17
Tabel IV- 12. Hasil Pengurutan Jarak pada Metode ROS-KNN	IV-18
Tabel IV- 13. Pengurutan Jarak pada Metode ROS-KNN Berdasarkan Nilai K	IV-19
Tabel IV- 14. Definisi Aktor	IV-20
Tabel IV- 15. Definisi Use Case	IV-20
Tabel IV- 16. Skenario Menentukan Rasio Data Training.....	IV-21
Tabel IV- 17. Skenario Menentukan Nilai K	IV-21
Tabel IV- 18. Skenario Klasifikasi dengan Metode KNN	IV-23
Tabel IV- 19. Skenario Klasifikasi dengan Metode ROS-KNN	IV-24
Tabel IV- 20. Implementasi Kelas	IV-38

Tabel IV- 21. Rencana Pengujian Use Case Menentukan Rasio Data Training	IV-43
Tabel IV- 22. Rencana Pengujian Use Case Menentukan Nilai K.....	IV-43
Tabel IV- 23. Rencana Pengujian Use Case Klasifikasi KNN	IV-44
Tabel IV- 24. Rencana Pengujian Use Case Klasifikasi ROS-KNN	IV-44
Tabel IV- 25. Pengujian Memilih Rasio Data Training	IV-44
Tabel IV- 26. Pengujian Menentukan Nilai K	IV-45
Tabel IV- 27. Pengujian Klasifikasi Menggunakan Metode KNN	IV-46
Tabel IV- 28. Pengujian Klasifikasi Menggunakan Metode ROS-KNN	IV-46
Tabel V- 1. Tabel Hasil Pengujian KNN Rasio Data Training 70%	V-3
Tabel V- 2. Tabel Hasil Pengujian KNN Rasio Data Training 80%	V-4
Tabel V- 3. Tabel Hasil Pengujian KNN Rasio Data Training 90%	V-5
Tabel V- 4. Tabel Hasil Pengujian ROS-KNN Rasio Data Training 70%.....	V-6
Tabel V- 5. Tabel Hasil Pengujian ROS-KNN Rasio Data Training 80%.....	V-7
Tabel V- 6. Tabel Hasil Pengujian ROS-KNN Rasio Data Training 90%.....	V-8
Tabel V- 7. Tabel Waktu Eksekusi Pada Sistem Klasifikasi.....	V-18

DAFTAR GAMBAR

	Halaman
Gambar II- 1. Rational Unified Process (RUP)	II-12
Gambar III-1. Tahapan Penelitian.....	III-3
Gambar III-2. Kerangka Kerja.....	III-5
Gambar IV- 1. Diagram Use-case	IV-19
Gambar IV- 2. Rancangan Antarmuka Halaman pilih Rasio Data Training..	IV-26
Gambar IV- 3. Rancangan Antarmuka Dashboard	IV-26
Gambar IV- 4. Rancangan Antarmuka Hasil Klasifikasi KNN	IV-27
Gambar IV- 5. Rancangan Antarmuka Hasil Klasifikasi ROS-KNN	IV-27
Gambar IV- 6. Diagram Aktivitas Menentukan Rasio Data Training	IV-29
Gambar IV- 7. Diagram Aktivitas Menentukan Nilai K pada Klasifikasi KNN	IV-29
Gambar IV- 8. Diagram Aktivitas Menentukan Nilai K pada Klasifikasi ROS-KNN.....	IV-30
Gambar IV- 9. Diagram Aktivitas Klasifikasi Menggunakan Metode KNN	IV-31
Gambar IV- 10. Diagram Aktivitas Klasifikasi Menggunakan ROS-KNN... IV-32	
Gambar IV- 11. Sequence Diagram Menentukan Rasio Data Training.....	IV-33
Gambar IV- 12. Sequence Diagram Menentukan Nilai K pada Klasifikasi KNN	IV-34
Gambar IV- 13. Sequence Diagram Menentukan Nilai K pada Klasifikasi ROS- KNN.....	IV-34
Gambar IV- 14. Sequence Diagram Klasifikasi Menggunakan Metode KNN.....	IV-35
Gambar IV- 15. <i>Sequence Diagram</i> Klasifikasi Menggunakan Metode ROS- KNN.....	IV-36
Gambar IV- 16. Diagram Kelas	IV-37

Gambar IV- 17. Antarmuka Halaman Pilih Rasio Data Training	IV-39
Gambar IV- 18. Antarmuka Halaman Dashboard.....	IV-40
Gambar IV- 19. Antarmuka Halaman Hasil Pengujian KNN	IV-41
Gambar IV- 20. Antarmuka Halaman Hasil Pengujian ROS-KNN	IV-42
Gambar V- 1. Gambar Perbandingan Akurasi Rasio Data Training 70%	V-9
Gambar V- 2. Gambar Perbandingan Akurasi Rasio Data Training 80%	V-10
Gambar V- 3. Gambar Perbandingan Akurasi Rasio Data Training 90%	V-11
Gambar V- 4. Rata-Rata Presisi Rasio Data <i>Training</i> 70%	V-12
Gambar V- 5. Rata-Rata Presisi Rasio Data Training 80%	V-13
Gambar V- 6. Rata-Rata Presisi Rasio Data Training 90%	V-14
Gambar V- 7. Rata-Rata Recall Rasio Data Training 70%	V-15
Gambar V- 8. Rata-Rata Recall Rasio Data Training 80%	V-16
Gambar V- 9. Rata-Rata Recall Rasio Data Training 90%	V-17

DAFTAR LAMPIRAN

1. Dataset Tingkat Kebugaran Tubuh Manusia

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan membahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan yang berkaitan dengan penelitian ini.

1.2 Latar Belakang Masalah

Kebugaran jasmani adalah kemampuan tubuh dalam menjalankan aktivitas fisik tanpa menimbulkan kelelahan secara berlebihan (Sinuraya & Barus, 2020). Dengan menjaga kebugaran tubuh dengan baik akan membuat aktivitas kita menjadi lebih baik karena tubuh menjadi lebih segar dan tidak cepat lelah, menjaga kebugaran tubuh juga memberikan banyak manfaat diantaranya memperkecil resiko terkena berbagai penyakit, mengurangi stres dan depresi, dan membuat kinerja otak lebih baik.

Untuk mengetahui suatu tingkat kebugaran tubuh diperlukan suatu proses klasifikasi yang berguna untuk menentukan suatu kelompok atau kelas tingkat kebugaran tubuh yang ditentukan berdasarkan beberapa atribut, sehingga dapat menentukan solusi yang tepat untuk mengatasi permasalahan kebugaran tubuh (Sinuraya & Barus, 2020).

Metode klasifikasi yang paling umum digunakan pada bidang ilmu statistika adalah analisis diskriminan dan regresi logistik, namun semakin populer dan

berkembangnya era data maka terjadi pertumbuhan yang signifikan pada volume data yang menghasilkan set data besar (*big data*). Untuk mengatasi permasalahan tersebut sangat dibutuhkan metode analisis yang berguna untuk mendeskripsikan informasi yang terdapat pada set data besar dan untuk mengubah set data besar tersebut menjadi informasi yang terorganisir (Yuliati & Sihombing, 2021).

Seiring dengan perkembangan teknologi yang pesat proses klasifikasi dapat dilakukan dengan semakin efektif dan efisien. Klasifikasi yang efektif dan efisien sangat diperlukan untuk menghasilkan tingkat ketepatan yang baik, hal ini diperlukan untuk menghindari adanya kesalahan diagnosis (Mutmainah, 2021). Dengan perkembangan teknologi maka berkembanglah metode machine learning, yaitu mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunanya, yang dikembangkan berdasarkan disiplin ilmu seperti statistika, matematika, dan data mining (Yuliati & Sihombing, 2021). Terdapat banyak metode klasifikasi pada *machine learning* yang dapat digunakan seperti *support vector machine*, *naive bayes*, dan *k-Nearest Neighbor* (k-NN). Salah satu metode yang paling populer adalah metode k-NN (Yuliska & Syaliman, 2020).

Pada penelitian yang berjudul “Penerapan Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada Lombok Vape On” oleh (Yahya & Hidayanti, 2020) yang meneliti tentang penerapan metode k-NN pada klasifikasi efektivitas penjualan vape. Berdasarkan permasalahan pada penelitian tersebut, akan dikembangkan suatu penelitian tentang penerapan klasifikasi dengan metode k-NN untuk menentukan tingkat kebugaran tubuh manusia.

Klasifikasi menggunakan *machine learning* seringkali memiliki masalah pada persebaran data yang tidak seimbangan atau *imbalanced class* yang menyebabkan kekeliruan dalam melakukan klasifikasi (Syukron & Subekti, 2018). *Imbalanced class* terjadi karena pada proses pengumpulan data kelas mayoritas lebih mendominasi dibandingkan kelas minoritas, yang membuat *classifier learning* cenderung memprediksi data ke kelas mayoritas (Shi, 2020). Kesalahan pada pengklasifikasian akan berdampak pada kesalahan dalam diagnosis dan salah penanganan (Mutmainah, 2021).

Untuk mengatasi permasalahan pada *imbalanced class* diperlukan suatu metode untuk menyeimbangkan data, salah satu teknik yang dapat digunakan adalah teknik *resampling*. Teknik *resampling* dilakukan dengan menyeimbangkan data berdasarkan serangkaian algoritma *sampling* (Syukron & Subekti, 2018). Pada teknik *resampling* terdapat tiga kategori yaitu *over-sampling*, *under-sampling*, dan hibrida. Salah satu teknik *resampling* yang digunakan adalah *random oversampling*. Teknik *random oversampling* merupakan teknik penambahan *sample* dari kelas minoritas sampai jumlah *sample* tersebut sama dengan kelas mayoritas dengan cara menduplikasi kelas secara acak *sample* yang terdapat pada kelas minoritas.

Pada penelitian yang berjudul “*Improving k-Nearest Neighbors Algorithm for Imbalanced Data Classification*” oleh (Shi, 2020) yang meneliti tentang penerapan teknik penyeimbangan data yang digunakan untuk meningkatkan kinerja klasifikasi k-NN. Pada penelitian yang berjudul “*Penanganan Imbalance Data Pada Klasifikasi Kemungkinan Penyakit Stroke*” yang diteliti oleh

(Mutmainah, 2021) yang meneliti tentang penanganan *imbalance data* pada klasifikasi penyakit stroke dengan membandingkan teknik *random oversampling* dan *random undersampling*, dan mendapatkan hasil teknik *random oversampling* memiliki performa lebih tinggi dibandingkan *dengan random undersampling*.

Kebugaran tubuh saat ini diklasifikasi dengan menggunakan metode yang masih bersifat konvensional dan membutuhkan waktu yang lama jika data dan atribut yang digunakan banyak dan bervariasi. K-NN adalah suatu metode klasifikasi populer yang sederhana dan efektif untuk klasifikasi data yang berjumlah banyak dan penerapan teknik *random oversampling* diharapkan mampu untuk meningkatkan akurasi dan meningkatkan kinerja klasifikasi k-NN untuk membantu menentukan tingkat kebugaran tubuh manusia. Oleh karena itu, berdasarkan permasalahan diatas akan dikembangkan suatu penelitian tentang penerapan teknik *random oversampling* untuk mengatasi *imbalanced class* pada klasifikasi tingkat kebugaran tubuh menggunakan metode k-NN.

1.3 Rumusan Masalah

Berdasarkan latar belakang sebelumnya, maka dalam penelitian ini terdapat beberapa rumusan masalah sebagai berikut:

1. Bagaimana cara melakukan klasifikasi menggunakan metode k-NN?
2. Apakah teknik *random oversampling* dapat diterapkan untuk mengatasi permasalahan yang disebabkan oleh *imbalanced class* pada klasifikasi k-NN dalam menentukan klasifikasi tingkat kebugaran tubuh?
3. Bagaimana kinerja dan tingkat akurasi klasifikasi k-NN setelah diterapkan teknik *random oversampling*?

1.4 Tujuan

Adapun tujuan dilakukan penelitian adalah sebagai berikut:

1. Menerapkan teknik *random oversampling* untuk mengatasi *imbalanced class* pada klasifikasi k-NN dalam menentukan tingkat kebugaran tubuh.
2. Mengetahui tingkat keakuratan sistem klasifikasi k-NN setelah diterapkan metode *random oversampling*.

1.5 Manfaat

Manfaat yang diperoleh pada penelitian ini sebagai berikut:

1. Dapat mengetahui kinerja dan tingkat akurasi klasifikasi k-NN setelah diterapkan teknik *random oversampling* untuk menentukan tingkat kebugaran tubuh.
2. Dapat memberikan informasi tentang tingkatan kebugaran tubuh manusia.

1.6 Batasan Masalah

Pada penelitian ini terdapat batasan-batasan masalah, diantaranya:

1. Data bersumber dari *Korea Sport Promotion Foundation* yang diambil dari situs kaggle.com.
2. Perancangan aplikasi sistem klasifikasi dirancang terbatas berbasis *website*.

1.7 Sistematika Penulisan

Sistematika penulisan pada penelitian ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab pendahuluan akan menguraikan latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian ,batasan masalah, dan sistematika penulisan.

BAB II. LANDASAN TEORI

Pada bab landasan teori akan membahas tentang definisi kebugaran tubuh, klasifikasi, *k-Nearest Neighbor*, *imbalanced class*, *random oversampling*, dan semua aspek yang digunakan dalam proses penelitian.

BAB III METODOLOGI PENELITIAN

Pada bab metodologi penelitian akan membahas tahapan yang akan diterapkan pada penelitian. Seluruh tahapan yang dilakukan pada penelitian akan dideskripsikan secara rinci berdasarkan pada suatu kerangka kerja.

1.8 Kesimpulan

Berdasarkan uraian diatas, dapat disimpulkan bahwa penelitian ini akan dilakukan penyeimbangan *class* menggunakan teknik *random oversampling* yang akan diterapkan pada klasifikasi tingkat kebugaran tubuh dengan metode K-NN.

DAFTAR PUSTAKA

- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Faradisa, S. M., Nugrahadi, T. D., Muliadi, Budiman, I., & Kartini, D. (2021). Implementasi *IQR-SMOTE* Untuk Mengatasi Ketidakseimbangan Kelas Pada Klasifikasi Diabetes menggunakan *K-Nearest Neighbors*. 15, 48–60.
- Febriarini, A. S., & Astuti, E. Z. (2019). Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang. *Eksplora Informatika*, 8(2), 95–103. <https://doi.org/10.30864/eksplora.v8i2.156>
- Gaussian, J. (2021). 3 1,2,3. 10, 11–20.
- Giriwijoyo, H. Y. S. S., & Sidik, D. Z. (2010). Konsep Dan Cara Penilaian Kebugaran Jasmani Menurut Sudut Pandang Ilmu Faal Olahraga. *Jurnal Kepelatihan Olahraga*, 2(1), 9. <https://ejournal.upi.edu/index.php/JKO/article/view/16223>.
- Hutahaean, D. J., Wardani, N. H., & Purnomo, W. (2019). Pengembangan Sistem Informasi Penyewaan Gedung Berbasis Web dengan Metode Rational Unified Process (RUP) (Studi Kasus: Wisma Rata Medan). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(Vol. 3, No. 6, Juni), 5789–5798.
- Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, 1(1), 21–2014. www.bluefame.com,
- Kurniawan, Y. I., & Barokah, T. I. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor. *Jurnal Ilmiah Matrik*, 22(1), 73–82. <https://doi.org/10.33557/jurnalmatrik.v22i1.843>
- Marital, R., & Cheat, I. (2004). Data Mining : Data Lecture Notes for Chapter 2 Introduction to Data Mining by What is Data ? *Ratio*.
- Mutmainah, S. (2021). Penanganan Imbalance Data Pada Klasifikasi. *SNATi*, 1, 10–16.
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>

- Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. <http://arxiv.org/abs/1811.12808>
- Shafarindu, A. I., Ernawati, L., & Zaidiah, A. (2021). Penerapan Algoritma Naive Bayes untuk Klasifikasi Tingkat Kebugaran Jasmani Berdasarkan Hasil Pengukuran pada Pegawai. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, September, 278–287.
- Shi, Z. (2020). Improving k-Nearest Neighbors Algorithm for Imbalanced Data Classification. *IOP Conference Series: Materials Science and Engineering*, 719(1). <https://doi.org/10.1088/1757-899X/719/1/012072>
- Sihombing, P. R., & Yuliati, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIX : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(2), 417–426. <https://doi.org/10.30812/matrik.v20i2.1174>
- Sinuraya, J. F., & Barus, J. B. N. B. (2020). Tingkat Kebugaran Jasmani Mahasiswa Pendidikan Olahraga Tahun Akademik 2019/2020 Universitas Quality Berastagi. *Kinestetik*, 4(1), 23–32. <https://doi.org/10.33369/jk.v4i1.10359>
- Syukron, A., & Subekti, A. (2018). Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit. *Jurnal Informatika*, 5(2), 175–185. <https://doi.org/10.31311/ji.v5i2.4158>
- Xiao, J., Xie, L., He, C., & Jiang, X. (2012). Dynamic classifier ensemble model for customer classification with imbalanced class distribution. *Expert Systems with Applications*, 39(3), 3668–3675. <https://doi.org/10.1016/j.eswa.2011.09.059>
- Yahya, Y., & Puspita Hidayanti, W. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada “Lombok Vape On.” *Infotek : Jurnal Informatika Dan Teknologi*, 3(2), 104–114. <https://doi.org/10.29408/jit.v3i2.2279>
- Yuliska, Y., & Syaliman, K. U. (2020). Peningkatan Akurasi K-Nearest Neighbor Pada Data Index Standar Pencemaran Udara Kota Pekanbaru. *IT Journal Research and Development*, 5(1), 11–18. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).4680](https://doi.org/10.25299/itjrd.2020.vol5(1).4680)