Journal of Applied Informatics and Computing (JAIC)
Vol.5, No.1, Juli 2021, pp. 37~42
e-ISSN: 2548-6861

Perbandingan Performa Teknik *Sampling* Data untuk Klasifikasi Pasien Terinfeksi Covid-19 Menggunakan *Rontgen* Dada

Akhmad Rezki Purnajaya [1]*, Fuad Dwi Hanggara [2]**
* Teknik Perangkat Lunak, Universitas Universal
** Teknik Industri, Universitas Universal
rezki.purnajaya@uvers.ac.id [1], fuaddh@uvers.ac.id [2]

Article Info

Article history:

Received 31-05-2021
Revised 17-06-2021
Accepted 29-06-2021

Keyword:

Covid-19,
Klasifikasi,
Rontgen Dada,
Sampling Data,
SMOTE.

ABSTRACT

The COVID-19 virus became a virus that was deadly and shocked the world. One of the consequences caused by the COVID-19 virus is a respiratory infection. The solution put forward for this problem is with a prediction of the COVID-19 virus infection. This prediction was made based on the classification of chest X-ray data. One challenging issue in this field is the imbalance on the amount of data between infected chest X-rays and uninfected chest X-rays. The result of imbalanced data is data classification that ignores classes with fewer data. To overcome this problem, the data sampling technique becomes a mechanism to make the data balanced. For this reason, several data sampling techniques will be evaluated in this study. Data sampling techniques include Random Undersampling (RUS), Random Oversampling (ROS), Combination of Over-Undersampling (COUS), Synthetic Minority Over-sampling Technique (SMOTE), and Tomek Link (T-Link). This study also uses the Support Vector Machines (SVM) data classification, because it has high accuracy. Furthermore, the evaluation is carried out by selecting the highest accuracy and Area Under Curve (AUC). The best sampling technique found was SMOTE with an accuracy value of 99% and an AUC value of 99.32%. The SMOTE technique is the best data sampling technique for the classification of COVID-19 chest x-ray data.

I. PENDAHULUAN

Di penghujung tahun 2019, pemberitaan tentang virus mematikan yang dikenal dengan nama virus Corona (COVID-19) menghebohkan dunia. COVID-19 adalah sekelompok virus dalam subfamili *Orthocronavirinae* dalam famili *Coronaviridae* dan nama virusnya adalah *Nidovirales*. COVID-19 dapat menyebabkan infeksi saluran pernapasan ringan, bahkan lebih mematikan. [1]. COVID-19 merupakan virus jenis baru yang menular dan diketahui muncul pertama kali di Wuhan, Cina [2]. Pandemi tersebut diawali dengan sejumlah kasus pneumonia di Wuhan mengancam dunia dengan perkiraan 2% - 5% rata-rata kematian [3-4]. Penyebab dari penyebaran COVID-19 terjadi akibat adanya kontak langsung dengan penderita serta melalui wujud cairan [5]. Penyakit COVID-19 menyebabkan ribuan korban meninggal dunia diberbagai negara [6]. Organisasi kesehatan dunia atau *World Health Organization* (WHO) mengumumkan bahwa COVID-19 sebagai pandemi yang membawa risiko yang besar bagi negara, khususnya dengan sistem kekebalan tubuh yang rentan [7].

Berdasarkan permasalahan tersebut, sebuah solusi dibutuhkan untuk klasifikasi pasien yang terkena COVID-19. Hal tersebut dilakukan untuk memprediksi status apakah pasien tersebut aman atau terinfeksi COVID-19 berdasarkan hasil data rontgen dada. Salah satu implementasi dari solusi tersebut adalah menggunakan data rontgen dada sebagai data input untuk klasifikasi data. Implementasi yang serupa dilakukan oleh Bergtholdt, Wiemker dan Klinder (2016) dalam pembuatan sistem deteksi nodul paru-paru dengan klasifikasi. Nilai *Area Under Curve* (AUC) yang didapatkan adalah sebesar 90% [8]. Klasifikasi data dalam penelitian ini akan dilakukan dengan menggunakan model *Support Vector Machines* (SVM). Klasifikasi data SVM dipilih, karena sangat unggul dalam hasil yang didapatkan berdasarkan penelitian mengenai perbandingan model klasifikasi data. Hasil penelitian Harefa and Pratiwi (2016) terkait citra mammogram menunjukkan SVM mendapatkan keseluruhan

tingkat akurasi 93.98% dibandingkan dengan tingkat akurasi *k-Nearest Neighbour* (k-NN) sebanyak 63.86% [9].

Sebelum klasifikasi data dilakukan, data sampel harus dijadikan pertimbangan terlebih dahulu. Dataset dipastikan terlebih dahulu apakah jumlah setiap kategori data seimbang atau *imbalanced*. Apabila jumlah dataset *imbalanced*, maka kelas minor (dengan jumlah lebih sedikit) dalam klasifikasi data akan diabaikan [10]. Hal tersebut menyebabkan rata-rata misklasifikasi menjadi lebih tinggi dalam kelas minor. Mekanisme untuk mengatasi permasalahan tersebut adalah penggunaan metode *sampling*. Metode tersebut mengubah dataset yang tidak seimbang dengan prosedur yang berbeda untuk menghasilkan persebaran data yang seimbang. Data yang seimbang dapat meningkatkan performa keseluruhan proses klasifikasi dibandingkan dengan data yang tidak diproses [10]. Untuk itu, penelitian ini dirumuskan untuk mendapatkan pengetahuan mengenai performa yang dimiliki oleh beberapa teknik *sampling* data. Hasil evaluasi dari perbandingan teknik *sampling* data akan diukur berdasarkan AUC serta tingkat akurasi sebagai parameter tambahan.

II. METODE

Untuk mendapatkan hasil dari perbandingan performa setiap teknik *sampling* data, maka data rontgen dada akan digunakan pada setiap teknik tersebut. Perbandingan tersebut juga dilakukan terhadap data yang *imbalanced* dengan ke 5 teknik *sampling* data. Dalam penelitian ini, 5 teknik *sampling* data yang akan dibandingkan mencakup Random *Undersampling* (RUS), Random *Oversampling* (ROS), *Combination of OverUndersampling* (COUS), *Synthetic Minority Over-sampling Technique* (SMOTE), dan *Tomek Link* (T-Link). Bagian ini akan membahas tahap untuk mendapatkan hasil pengukuran performa dari teknik *sampling* data tersebut.

A. Persiapan Data Rontgen Dada

Tahap ini dilakukan untuk mempersiapkan dataset yang digunakan untuk *sampling* dan klasifikasi berdasarkan hasil rontgen dada. Pengujian dalam penelitian ini menggunakan dua kelompok data, yaitu data rontgen dada untuk pasien yang terinfeksi dan aman dari COVID-19. Data rontgen dada untuk *infected* menggunakan data yang bersumber dari Cohen, Morrison and Dao (2020) [11]. Kemudian, data rontgen dada untuk *uninfected* menggunakan sumber dari Wang et al. (2019) [12]. Kedua data tersebut memiliki format gambar jpg yang dikonversi dalam ukuran 50 x 50 pixel. Total data citra adalah sebanyak 402 dimana data *infected* sebanyak 282 dan data *uninfected* sebanyak 120.

B. Penggunaan Sampling

Sebelum memulai *sampling* data, persiapan juga dilakukan pada *sampling* yang akan digunakan dalam RStudio. Penjelasan dari *sampling* tersebut adalah sebagai berikut:

- **EBImage** merupakan sebuah *package* atau *toolbox* yang digunakan untuk pemrosesan dan analisis terhadap citra dalam R [13]. Fungsionalitas yang disediakan EBImage mempermudah pemrosesan sinyal, pemodelan statistik, *machine learning*, dan visualisasi dengan menggunakan data citra atau image [14]. *Sampling* ini akan digunakan dalam penelitian ini untuk mendapatkan fitur dari citra.
- ***unbalanced*** merupakan *package* R yang memiliki beberapa teknik *sampling* untuk klasifikasi data *imbalanced*. Teknik *sampling* yang dimiliki *package* ini mencakup ubOver, ubUnder, SMOTE, dan Tomek [15]. *Sampling* ini akan digunakan untuk teknik *sampling* data RUS, ROS, SMOTE, dan T-Link.
- **ROSE** merupakan *package* yang menyediakan fungsi untuk menyelesaikan permasalahan klasifikasi biner dalam kelas *imbalanced* [16]. *Sampling* ini akan digunakan untuk teknik *sampling* data COUS.
- **e1071** merupakan *package* yang menyediakan fungsionalitas untuk analisis kelas laten atau *Latent Class Analysis* (LCA), SVM (*Support Vector Machines*), dan lainnya [17]. *Sampling* ini akan digunakan untuk klasifikasi data menggunakan model SVM.

C. Pembuatan Fitur dan Kelas Citra

Tahap ini dilakukan untuk mendapatkan ekstrak fitur berdasarkan gambar dari rontgen dada. Fitur dari citra didapatkan dengan menggunakan *sampling* EBImage. Perulangan dilakukan untuk mengambil setiap data dari citra dalam path tempat citra tersimpan. Kemudian, citra akan dilakukan proses pembacaan oleh *sampling* dengan *channel* warna abu-abu (*gray*). Proses dari perulangan dengan *sampling* dilakukan untuk mengekstrak fitur dari citra. Karena terdapat dua jenis citra, yakni *uninfected* dan *infected*, maka perulangan tersebut akan dilakukan sebanyak 2 kali. Setelah ekstrak fitur dilakukan, tahap selanjutnya adalah memberi kelas pada masing-masing kategori data citra. Pemberian kelas pada data citra disesuaikan dengan *sampling unbalanced* dimana 0 untuk kelas mayor dan 1 untuk kelas minor. Kelas mayor berada pada data *infected*, karena memiliki jumlah yang lebih banyak dibandingkan dengan data *uninfected*. Dengan kata lain, kelas data *infected* adalah 0 dan kelas data *uninfected* adalah 1 [18].

D. Sampling Data

Tahap ini dilakukan untuk melihat efektivitas dari masing-masing teknik *sampling* data. Perbandingan tersebut dilakukan pada data yang *imbalanced* serta 5 teknik *sampling* data. Dalam melakukan perbandingan, masing-masing teknik *sampling* data dilakukan dengan menggunakan bantuan *sampling*. Penjelasan dari perbandingan yang dilakukan adalah sebagai berikut:

- Data *Imbalanced*

  Sebuah dataset dikatakan *imbalanced* apabila distribusi kelas tidak seimbang dimana terjadi saat jumlah salah satu kelas lebih rendah dibandingkan kelas lainnya. Permasalahan ini dapat menghambat performa dari klasifikasi data yang membuat kelas

minor diabaikan [10]. Data citra rontgen dada menunjukkan adanya data yang tidak seimbang dimana jumlah data sampel tidak seimbang. Jumlah data citra *infected* (282 data) lebih banyak dibandingkan dengan jumlah data citra *uninfected* (120 data). Dari jumlah tersebut, data citra rontgen dada dapat dikatakan sebagai data *imbalanced*. Data *frame* akan dibuat pada tahap ini dengan mengkombinasikan fitur dan kelas citra. Fungsi *table*() dapat dimanfaatkan untuk menampilkan kelas 0 (*infected*) dan 1 (*uninfected*) agar memastikan hanya ada 2 kelas dalam dataset.

- *Random Undersampling* (RUS)

  Random Undersampling (RUS) merupakan metode non-heuristik yang menyeimbangkan distribusi kelas melalui penghapusan kelas mayor secara acak untuk mendapatkan instance set yang seimbang [10]. RUS bekerja dengan menghapus sampel secara acak untuk menyeimbangkan distribusi skew dalam masing-masing dataset atau dengan sampel kelas minor [19]. Teknik *sampling* data ini akan dilakukan dengan menggunakan fungsi *ubUnder()* dari *sampling unbalanced* [18]. *Sampling* data RUS dimulai dengan memasukkan fitur dan kelas citra sebagai parameter dari ubUnder(). Selanjutnya, data yang diolah melalui algoritma RUS dihasilkan dan dapat dicek dengan table() untuk mendapatkan jumlah dari setiap kelas citra. Fitur dan kelas data citra rontgen dada dimasukkan sebagai argument dari fungsi *ubUnder().* Selanjutnya, dataset citra akan diproses dengan algoritma RUS dan tabel baru dibuat untuk menyimpan fitur dan kelas yang telah diproses RUS. Jumlah dataset berdasarkan kelas dapat dicek dengan fungsi *table().*

- *Random Oversampling* (ROS)

  Random Oversampling (ROS) merupakan metode non-heuristik yang menyeimbangkan distribusi kelas melalui penambahan data pada kelas minor secara acak. Teknik *sampling* data ini akan dilakukan dengan menggunakan fungsi *ubOver()* dari *sampling unbalanced* [10]. Penelitian Johnson dan Khoshgoftaar (2019) mendapatkan hasil bahwa teknik ini memiliki performa yang lebih baik dibandingkan dengan RUS [20]. Pang et al. (2019) menjelaskan bahwa teknik ROS dapat menyebabkan terjadinya permasalahan *overfitting* selama proses learning [21]. Selanjutnya, data yang diolah melalui algoritma ROS dihasilkan dan dapat dicek dengan table() untuk mendapatkan jumlah dari setiap kelas citra. Fitur dan kelas data citra rontgen dada dimasukkan sebagai argument dari fungsi *ubOver().* Selanjutnya, dataset citra akan diproses dengan algoritma ROS dan tabel baru dibuat untuk menyimpan fitur dan kelas yang telah diproses ROS. Jumlah dataset berdasarkan kelas dapat dicek dengan fungsi *table().*

- *Combination of Over-Undersampling* (COUS)

  Combination of Over-Undersampling (COUS) merupakan kombinasi dari algoritma teknik *sampling* ROS dan RUS. Dengan kata lain, teknik ini terbentuk secara hybrid yang menggabungkan metode *over-sampling* dan *under-sampling*. Algoritma teknik gabungan ini adalah dengan melakukan *over-sampling* terhadap kelas minor dan *under-sampling* terhadap kelas mayor hingga kelas memiliki jumlah sampel yang sama [20]. Teknik *sampling* data ini akan dilakukan dengan menggunakan fungsi *ovun.sample()* dari *sampling* ROSE. Argumen method dalam fungsi tersebut harus menggunakan parameter *both* agar menggabungkan ROS dan RUS dalam *sampling* [16]. Berbeda dengan *sampling unbalanced*, tabel akan langsung menyimpan data yang telah diproses dan diseimbangkan dengan COUS. Dengan kata lain, tabel baru untuk menyimpan data yang telah diproses COUS tidak perlu dibuat. Jumlah dataset berdasarkan kelas dapat dicek dengan fungsi *table().*

- *Synthetic Minority Over-sampling Technique* (SMOTE)

  Synthetic Minority Over-sampling Technique (SMOTE) merupakan metode over*sampling* yang paling banyak digunakan (Raghuwanshi and Shukla, 2020) serta diperkenalkan oleh Chawla et al. (2002) [22-23]. Teknik SMOTE melakukan over*sampling* tanpa duplikasi atau penambahan berdasarkan k-NN (*k-Nearest Neighbors*) dari kelas minor [24]. SMOTE bertujuan untuk memperkaya batas kelas minor dengan membuat contoh buatan dalam kelas minor dadripada menambah contoh yang sudah ada untuk menhindari permasalahan overfitting [25]. Cara pembuatan data sampel baru dalam teknik ini adalah dengan memilih baris yang matching secara acak, kemudian mempersiapkan kombinasi konveks untuk sampel baru [19]. Teknik *sampling* data ini akan dilakukan dengan menggunakan fungsi *ubSMOTE()* dari *sampling unbalanced* [18]. Fitur dan kelas data citra rontgen dada dimasukkan sebagai argument dari fungsi *ubSMOTE().* Selanjutnya, dataset citra akan diproses dengan algoritma SMOTE dan tabel baru dibuat untuk menyimpan fitur dan kelas yang telah diproses SMOTE. Jumlah dataset berdasarkan kelas dapat dicek dengan fungsi *table().*

- *Tomek Link* (T-Link)

  Tomek Link (T-Link) dianggap sebagai peningkatan dari *Nearest-Neighbor Rule* (AT et al. 2016). Dalam dataset, terdapat *instance* yang merupakan data tetangga terdekat dan berada pada kelas yang berbeda. Teknik T-Link ini mencari *instance* tersebut menggunakan 1-NN (*One-Nearest-Neighbor*) dalam dataset. Untuk mengatasi *imbalanced*, pada kelas mayor *instance* tersebut dihapus [18]. Teknik *sampling* data ini akan dilakukan dengan menggunakan fungsi *ubTomek()* dari *sampling unbalanced* [18]. Fitur dan kelas data citra rontgen

dada dimasukkan sebagai argument dari fungsi *ubTomek()*. Selanjutnya, dataset citra akan diproses dengan algoritma T-Link dan tabel baru dibuat untuk menyimpan fitur dan kelas yang telah diproses T-Link. Jumlah dataset berdasarkan kelas dapat dicek dengan fungsi *table()*.

E. Membuat Data Uji

Tahap ini dilakukan untuk membuat data uji yang digunakan dalam klasifikasi data. Data uji atau *testing data set* merupakan data yang akan digunakan untuk pengujian. Penelitian ini menggunakan 100 buah sampel data uji yang dipilih secara acak. Selanjutnya, data aktual atau hasil asli dari dataset disimpan didalam tabel terpisah untuk pengujian.

F. Klasifikasi Data

Setelah melakukan *sampling* data, tahap ini dilakukan untuk klasifikasi terhadap data tersebut. Metode yang dipilih untuk klasifikasi data dalam penelitian ini adalah *Support Vector Machines* (SVM). Model klasifikasi SVM pertama kali diperkenalkan oleh Cortes dan Vapnik (1995) [26]. SVM merupakan metode klasifikasi yang banyak digunakan, karena akurasi klasifikasi yang sangat dipengaruhi oleh pengaturan parameter kernel dan seleksi fitur [27]. SVM digolongkan sebagai teknik *Supervised Machine Learning* yang digunakan untuk klasifikasi dan regresi. SVM mengelompokkan data dengan mencari *hyperplane* cocok yang dapat memisahkan data berdasarkan margin tertinggi [28-29]. Pembuatan model SVM dilakukan dengan memanfaatkan fungsi *svm()* dalam *package* e1071. Fungsi ini digunakan untuk melakukan pelatihan data terhadap model SVM [17]. Argumen yang digunakan dalam fungsi *svm()* adalah kelas dan dataset yang telah diproses berdasarkan setiap teknik *sampling* data.

G. Evaluasi Teknik Sampling

Tahap ini dilakukan untuk membandingkan hasil dari performa setiap teknik *sampling* data serta data *imbalanced* setelah klasifikasi data dilakukan. Evaluasi dilakukan dengan menghitung akurasi, spesifisitas, sensitivitas, dan AUC untuk setiap teknik *sampling* data. Untuk mengukur kriteria tersebut, maka prediksi akan dilakukan dengan menggunakan data *imbalanced* dan yang telah melalui proses *sampling*. Prediksi dilakukan dengan menggunakan fungsi *predict*() dalam R. Argumen dalam fungsi tersebut adalah menggunakan model SVM dari setiap teknik *sampling* dan data pengujian. Selanjutnya, dataframe dibuat untuk membandingkan hasil prediksi dengan aktual. Hasil yang didapatkan dari proses prediksi adalah *confussion matrix*. Hasil *confussion matrix* digunakan untuk menghitung Akurasi, Spesifisitas, Sensitivitas, dan AUC. Menurut Apostolopoulos and Mpesiana (2020) serta (Purnajaya and Kusuma, 2019), persamaan yang akan digunakan kriteria tersebut adalah sebagai berikut:

$$Akurasi = \frac{True\ Positive + True\ Negative}{(Length\ Positive + Length\ Negative)} \quad (1)$$

$$Spesifisitas = \frac{True\ Negative}{(True\ Negative + False\ Positive)} \quad (2)$$

$$Sensitivitas = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \quad (3)$$

$$AUC = \frac{Sensitivity + Specificity}{2} \quad (4)$$

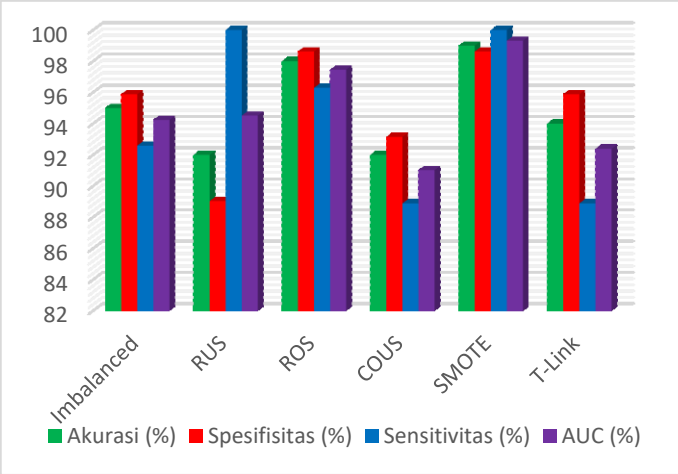
Persamaan tersebut diterapkan pada *confussion matrix* yang didapatkan untuk setiap prediksi yang dilakukan. Kemudian, teknik *sampling* dan kriteria tersebut disajikan dalam sebuah tabel. Tabel tersebut menjadi tabel untuk evaluasi terhadap kriteria performa yang didapatkan oleh setiap teknik *sampling* agar dapat dibandingkan [30-31].

III. HASIL DAN PEMBAHASAN

Bagian ini akan membahas hasil evaluasi yang diperoleh dengan kriteria akurasi, spesifisitas, sensitivitas, dan AUC. Hasil evaluasi dari teknik *sampling* data dengan data *imbalanced* dan teknik *sampling* data untuk data citra *infected* dan *uninfected* dapat dilihat pada Tabel 1.

TABEL I
PERBANDINGAN PERFORMA DATA *IMBALANCED* DAN TEKNIK *SAMPLING* DATA

Teknik *Sampling* Data	Akurasi (%)	Spesifisitas (%)	Sensitivitas (%)	AUC (%)
Imbalanced	**95,00**	95,89	92,59	**94,24**
RUS	**92,00**	89,04	100	**94,52**
ROS	**98,00**	98,63	96,30	**97,46**
COUS	**92,00**	93,15	88,89	**91,02**
SMOTE	**99,00**	98,63	100	**99,32**
T-Link	**94,00**	95,89	88,89	**92,39**



Gambar 1. Perbandingan Performa Data *Imbalanced* dan Teknik *Sampling* Data

Pada Tabel 1 dan Gambar 1 dapat dilihat mengenai informasi mengenai kriteria yang didapatkan oleh setiap teknik *sampling* data. Akurasi yang didapatkan oleh data

imbalanced, RUS, ROS, COUS, SMOTE, dan T-Link adalah sebanyak 95%, 92%, 98%, 92%, 99%, dan 94% secara berurutan. Akurasi tertinggi berada pada teknik *sampling* data SMOTE sebanyak 99% dibandingkan dengan teknik *sampling* lainnya. Spesifisitas yang diperoleh data *imbalanced*, RUS, ROS, COUS, SMOTE, dan T-Link adalah sebanyak 95.89%, 89.04%, 98.63%, 93.15%, 98.63%, dan 95.89% secara berurutan. Spesifisitas tertinggi berada pada teknik *sampling* data ROS dan SMOTE sebanyak 98.63%. Kemudian, sensivisitas data *imbalanced*, RUS, ROS, COUS, SMOTE, dan T-Link adalah sebanyak 92.59%, 100%, 96.3%, 88.89%, 100%, dan 88.89% secara berurutan. RUS dan SMOTE memiliki tingkat sensitivitas tertinggi sebanyak 100%. AUC yang diperoleh data *imbalanced*, RUS, ROS, COUS, SMOTE, dan T-Link adalah sebanyak 94.24%, 94.52%, 97.46%, 91.02%, 99.32%, dan 92.39% secara berurutan AUC tertinggi yang didapatkan adalah sebanyak 99.32% oleh teknik *sampling* data SMOTE.

IV. KESIMPULAN

Dalam penelitian ini, evaluasi data *imbalanced* dan teknik *sampling* data telah dilakukan. Dataset yang digunakan dalam penelitian ini adalah data citra rontgen dada *infected* dan *uninfected*. Penelitian ini juga membandingkan performa data *imbalanced* dengan 5 teknik *sampling* data, yakni RUS, ROS, COUS, SMOTE, dan T-Link. Hasil yang didapatkan oleh setiap teknik *sampling* data berbeda-beda. Berdasarkan hasil evaluasi yang dilakukan, maka dapat disimpulkan bahwa teknik *sampling* data SMOTE paling unggul dibandingkan dengan teknik lainnya. Hal tersebut diukur dari kriteria perhitungan AUC yang didapatkan oleh data *imbalanced*, RUS, ROS, COUS, SMOTE, dan T-Link adalah 94.24%, 94.52%, 97.46%, 91.02%, 99.32%, dan 92.39% secara berurutan. SMOTE mendapatkan nilai AUC tertinggi sebesar 99.32% dan akurasi tertinggi sebesar 99%. Selain itu, SMOTE dapat meningkatkan nilai akurasi pada data *imbalanced* sebanyak 4% dan meningkatkan nilai AUC pada data *imbalanced* sebanyak 5.08%. Dengan kata lain, SMOTE merupakan teknik *sampling* yang paling unggul dalam menangani permasalahan *imbalanced* dengan data citra rontgen dada (*infected* dan *uninfected*) untuk prediksi COVID-19.

Untuk penelitian selanjutnya, harapannya adalah penelitian dengan jumlah dataset yang lebih banyak disertai dengan ukuran citra yang lebih besar agar dapat mengetahui perubahan pada kriteria evaluasi. Hal tersebut diharapkan dapat menunjukkan adanya peningkatan pada performa klasifikasi data. Harapan lainnya adalah perbandingan dengan metode teknik *sampling* data lainnya diluar dari 5 teknik *sampling* ini agar dapat mengetahui teknik yang lebih baik dari teknik yang dibandingkan..

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Direktorat Riset dan Pengabdian Masyarakat, Kementerian Riset dan Teknologi/Badan Riset dan Inovasi Nasional atas dukungan yang diberikan kepada peneliti berupa bantuan dana penelitian yang menunjang berlangsungnya penelitian ini dengan baik.

DAFTAR PUSTAKA

[1] Yunus, N. R. and Rezki, A. (2020) 'Kebijakan Pemberlakuan Lock Down Sebagai Antisipasi Penyebaran Corona Virus Covid-19', SALAM: Jurnal Sosial dan Budaya Syar-i, 7(3). doi: 10.15408/sjsbs.v7i3.15083.

[2] Shi, H. et al. (2020) 'Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study', The Lancet Infectious Diseases. Elsevier Ltd, 20(4), pp. 425–434. doi: 10.1016/S1473-3099(20)30086-4.

[3] Guo, H. et al. (2020) 'The impact of the COVID-19 epidemic on the utilization of emergency dental services', Journal of Dental Sciences. Elsevier B.V., (xxxx), pp. 0–3. doi: 10.1016/j.jds.2020.02.002.

[4] Pastor, C. K. L. (2020) 'Sentiment Analysis on Synchronous Online Delivery of Instruction due to Extreme Community Quarantine in the Philippines caused by COVID-19 Pandemic', Asian Journal of Multidisciplinary Studies, 3(1).

[5] Li, Q. et al. (2020) 'Early Transmission Dynamics in Wuhan, China, of Novel8 Coronavirus–*Infected* Pneumonia', New England Journal of Medicine, pp. 1199–1207. doi: 10.1056/nejmoa2001316.

[6] Mahase, E. (2020) 'Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate', BMJ (Clinical research ed.), 368(February), p. m641. doi: 10.1136/bmj.m641.

[7] Sohrabi, C. et al. (2020) 'World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)', International Journal of Surgery. Elsevier, 76(February), pp. 71–76. doi: 10.1016/j.ijsu.2020.02.034.

[8] Bergtholdt, M., Wiemker, R. and Klinder, T. (2016) 'Pulmonary nodule detection using a cascaded SVM classifier', Medical Imaging 2016: Computer-Aided Diagnosis, 9785, p. 978513. doi: 10.1117/12.2216747.

[9] Harefa, J. and Pratiwi, M. (2016) 'Comparison Classifier: Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) In Digital Mammogram Images', Juisi, 02(02), pp. 35–40. Available at: http://peipa.essex.ac.uk/pix/mias/.

[10] Fernández, A. et al. (2018) Learning from *Imbalanced* Data Sets, IEEE Transactions on Knowledge and Data Engineering. Cham: Springer International Publishing. doi: 10.1007/978-3-319-98074-4.

[11] Cohen, J. P., Morrison, P. and Dao, L. (2020) 'COVID-19 Image Data Collection'. Available at: http://arxiv.org/abs/2003.11597.

[12] Wang, X. et al. (2019) 'ChestX-ray: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases', Advances in Computer Vision and Pattern Recognition, pp. 369–392. doi: 10.1007/978-3-030-13969-8_18.

[13] Kim, S. et al. (2018) 'Time-resolved fractal dimension analysis in ferroelectric copolymer thin films using R-based image processing', Materials Letters. Elsevier B.V., 230, pp. 195–198. doi: 10.1016/j.matlet.2018.07.125.

[14] Ole´s, A. et al. (2020) 'Image processing and analysis toolbox for R', R *package* version 4.30.0, pp. 1–52. Available at: http://bioconductor.org/*packages*/release/bioc/html/EBImage.html.

[15] Zhu, B. et al. (2019) 'IRIC: An R *sampling* for binary *imbalanced* classification', SoftwareX. Elsevier B.V., 10(October), p. 100341. doi: 10.1016/j.softx.2019.100341.

[16] Lunardon, N., Menardi, G. and Torelli, N. (2015) 'ROSE: Random Over-*Sampling* Examples', R *package* version 0.0-3, pp. 1–19. doi: 10.1007/s10618-012-0295-5.

[17] Meyer, D. et al. (2019) 'Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien', R *package* version 1.7-3, pp. 1–63. Available at: https://cran.r-project.org/*package*=e1071.

[18] Andrea, A. et al. (2015) 'Racing for *Unbalanced* Methods Selection', R *package* version 2.0, pp. 1–18. Available at: https://cran.r-project.org/*package=unbalanced*.

[19] Mehmood, R. and Selwal, A. (2020) Proceedings of ICRIC 2019, Lecture Notes in Electrical Engineering. Edited by P. K. Singh et al. Cham: Springer International Publishing (Lecture Notes in Electrical Engineering). doi: 10.1007/978-3-030-29407-6.

[20] Johnson, J. M. and Khoshgoftaar, T. M. (2019) 'Deep learning and data *sampling* with *imbalanced* big data', Proceedings - 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science, IRI 2019, (July), pp. 175–183. doi: 10.1109/IRI.2019.00038.

[21] Pang, Y. et al. (2019) 'A signature-based assistant random over*sampling* method for malware detection', Proceedings - 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019, pp. 256–263. doi: 10.1109/TrustCom/BigDataSE.2019.00042.

[22] Raghuwanshi, B. S. and Shukla, S. (2020) 'SMOTE based class-specific extreme learning machine for *imbalanced* learning', Knowledge-Based Systems. Elsevier B.V., 187, p. 104814. doi: 10.1016/j.knosys.2019.06.022.

[23] Chawla, N. V et al. (2002) 'SMOTE: Synthetic *minor*ity over-*sampling* technique', Journal of Artificial Intelligence Research, 16, pp. 321–357. doi: 10.1613/jair.953.

[24] Komori, O. and Eguchi, S. (2019) Statistical Methods for *Imbalanced* Data in Ecological and Biological Studies. doi: 10.1007/978-4-431-55570-4.

[25] AT, E. et al. (2016) 'Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-*sampling* (RUS) as a Data Reduction Method', Global Journal of Technology and Optimization, 01(S1). doi: 10.4172/2229-8711.S1111.

[26] Cortes, C. and Vapnik, V. (1995) 'Support-vector networks', Machine Learning, 20(3), pp. 273–297. doi: 10.1007/BF00994018.

[27] Wang, M. and Chen, H. (2020) 'Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis', Applied Soft Computing Journal. Elsevier B.V., 88, p.105946. doi: 10.1016/j.asoc.2019.105946.

[28] Vluymans, S. (2019) Dealing with *imbalanced* and weakly labelled data in machine learning using fuzzy and rough set methods, Studies in Computational Intelligence. doi: 10.1007/978-3-030-04663-7_1.

[29] Jain, M. et al. (2020) 'Speech Emotion Recognition using Support Vector Machine', International Journal of Smart Home, 6(2), pp. 101–108. doi: 10.1109/kst.2013.6512793.

[30] Apostolopoulos, I. D. and Mpesiana, T. A. (2020) 'Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks', Physical and Engineering Sciences in Medicine. Springer International Publishing, (0123456789), pp. 1–6. doi: 10.1007/s13246-020-00865-4.

[31] Purnajaya, A. R. and Kusuma, W. A. (2019) Prediksi Interaksi pada Jejaring Bipartite Senyawa dan Protein pada Data yang Tidak Seimbang. Institut Pertanian Bogor. doi: 10.13140/RG.2.2.28328.52484.

.