INFOTECH journal
ISSN : 2460-1861 (Print), 2615-4250 (Online)
Vol. 9 No. 1, pp. 54-61

PENGGUNAAN OTIMASI ATRIBUT DALAM PENINGKATAN AKURASI PREDIKSI DEEP LEARNING PADA BIKE SHARING DEMAND

Syarif Hidayatulloh[1], Muhammad Amar Mustaja[2],Yudi Ramdhani[3]
[1,2,3]Program Studi Teknik Informatika, Fakultas Teknologi Informasi,
Universitas Adhirajasa Reswara Sanjaya
Email: m.amar.mustajab@gmail.com

ABSTRACT
Cycling is becoming popular again in the aftermath of the Covid 19 pandemic that occurred in Indonesia yesterday. In this study, the most commonly used algorithms were tested, including Neural Nets, Generelized Linear Models, Support Vector Machines, Random Forests, and Deep Learning. The study was conducted in five predictive algorithm models with ten trials using cross validation, and the best accuracy value was chosen. According to the comparison of these algorithms, the Deep Learning algorithm has an Accuracy value of 90% and an AUC of 0.770. Croos Validation with X fold is the foundation of this comparison algorithm. The deep learning algorithm was discovered to have an accuracy value of 90%, which is 4-5% lower than the other four algorithms. With a significant increase in the research, the accuracy value for weight optimization using the Forward optimize algorithm was 95.63%. Based on the results of the experiments, the researchers concluded that the experiments were successful in increasing the accuracy value of the deep learning algorithm.

Keywords: Attribute Optimization, Deep Learning, Feature Selection, Feature Weights, Prediction.

ABSTRAK

Bersepeda kembali populer pasca pandemi Covid 19 yang terjadi di Indonesia kemarin. Dalam studi ini, algoritma yang paling umum digunakan diuji, termasuk Neural Nets, Generelized Linear Models, Support Vector Machines, Random Forests, dan Deep Learning. Penelitian dilakukan dalam lima model algoritma prediktif dengan sepuluh percobaan menggunakan validasi silang, dan dipilih nilai akurasi terbaik. Berdasarkan perbandingan algoritma tersebut, algoritma Deep Learning memiliki nilai Accuracy sebesar 90% dan AUC sebesar 0,770. Validasi Croos dengan X fold adalah dasar dari algoritma perbandingan ini. Algoritma pembelajaran mendalam ditemukan memiliki nilai akurasi 90%, yang 4-5% lebih rendah dari empat algoritma lainnya. Dengan peningkatan yang signifikan pada penelitian tersebut, maka nilai akurasi untuk optimasi bobot menggunakan algoritma Forward optimize adalah sebesar 95,63%. Berdasarkan hasil percobaan, peneliti menyimpulkan bahwa percobaan tersebut berhasil meningkatkan nilai akurasi dari algoritma deep learning.

Kata Kunci: Deep Learning, Fitur Bobot, Fitur Seleksi, Optimasi Atribut, Prediksi.

Riwayat Artikel :

Tanggal diterima : 01-02-2023

Tanggal revisi : 12-02-2023

Tanggal terbit : 15-02-2023

DOI :

https://doi.org/10.31949/infotech.v9i1.4530

1. PENDAHULUAN

1.1. Latar Belakang

Di era pandemi Covid 19 kemarin di Indonesia, tren olahraga sepeda mulai diminati kembali oleh banyak orang. Ini terjadi sebab bersepeda dapat dilakukan sambil bermain bersama keluarga atau teman. Bukan hanya itu saja, bersepeda juga dapat meningkatkan daya tahan tubuh seseorang.

Kegiatan bersepeda dapat dilakukan di waktu luang misalnya saja pada akhir pekan. Untuk sekedar menjaga kebugaran dan sebagai sarana rekreasi, bersepeda tidak memerlukan jarak tempuh yang jauh. Masyarakat dapat melakukannya disekitar kota. Pada masa pandemi covid 19, bersepeda menjadi alternatif olahraga yang diminati oleh masyarakat, baik dilakukan secara individu maupun kelompok, karena dianggap dapat meningkatkan imunitas tubuh dimasa pandemi [1].

Bersepeda juga menjadi salah satu olahraga yang aman untuk lansia, salah satu upaya untuk menjaga, mencegah, meningkatkan kesehatan dan kesegaran jasmani bagi lansia (lanjut usia) adalah dengan melakukan olahraga. Olahraga bagi lansia yang dilakukan secara terprogram akan mempunyai beberapa manfaat, diantaranya adalah untuk mempertahankan kesehatan, meningkatkan kekuatan otot jantung, meningkatkan sirkulasi darah dalam tubuh, menurunkan kadar lemak, menguatkan otot-otot tubuh, mengurangi stress dan ketegangan batin, meningkatkan sistem kekebalan tubuh [2].

Di Korea Selatan (Korsel), khususnya di Seoul, bersepeda menjadi pilihan transportasi jarak dekat dan olahraga. Kota Seoul dan beberapa kota baru lain seperti Incheon Metropolitan City dan Sejong telah memiliki jaringan jalur khusus sepeda yang dapat digunakan untuk berkeliling kota tanpa khawatir bersinggungan dengan kendaraan bermesin. Bagi pesepeda dengan waktu terbatas, kegiatan bersepeda di Seoul dan travelling ke kota terdekat lain seperti Incheon dan Suwon juga tidak kalah menyenangkan [3].

1.2. Tinjuan Pustaka

Kesehatan tidak terpisahkan dalam kehidupan manusia, karena kesehatan merupakan keadaan yang paling penting dalam menjalankan berbagai aktivitas. Tanpa kesehatan manusia akan mengalami hambatan dan mengalami penurunan kondisi fisik. Kesehatan adalah keadaan seimbang yang dinamis, dipengaruhi faktor genetik, lingkungan dan pola hidup sehari-hari seperti makan, minum, seks, kerja, istirahat, hingga pengelolaan kehidupan emosional. Ada empat faktor terpenting yang harus diperhatikan dalam kaitannya dengan kesehatan yaitu 1) makanan sehat, 2) perbanyak minum air putih, 3) tidur cukup, 4) serta selalu aktif [4].

Seiring dengan perkembangan ilmu, pengetahuan dan teknologi yang semakin pesat, maka meningkat pola dan gaya hidup masyarakat. Hal ini berdampak juga terhadap kesehatan, aktivitas gerak menjadi berkurang sebaliknya berbagai penyakit menyerang individu yang masih produktif, seperti penyakit jantung, diabetes, kolesterol, hipertensi dan lainnya [4].

Olahraga bersepeda ini muncul sebagai akibat dari penerapan PSBB di Indonesia yang mana masyarakat merasa jenuh di dalam rumah kemudian melakukan aktivitas bersepeda di luar rumah sebagai salah satu cara untuk mengurangi kejenuhan, mencari udara segar dan bersih, ramah lingkungan, aman, serta dapat menjaga jarak antara pesepeda satu dengan pesepeda lainnya [5].

Teknik Deep Learning akan diterapkan dalam penelitian ini untuk prediksi permintaan bike sharing Graph Convolutional Neural Network dengan model Filter Grafik Berbasis Data yang berhasil mempelajari korelasi heterogen berpasangan tersembunyi antara stasiun untuk memprediksi permintaan per jam tingkat stasiun dalam jaringan berbagi sepeda skala luas, diusulkan oleh Lin et al [6]. Deep Learning termasuk kedalam kecerdasan buatan. Algoritma pembelajaran mesin membangun matematika model berdasarkan data sampel, yang dikenal sebagai “data pelatihan”, untuk membuat prediksi atau keputusan tanpa harus secara eksplisit diprogram untuk melakukan tugas [7].

Teknologi data mining merupakan salah satu alat bantu yang dapat mengoptimalkan data basis yang berukuran besar, dengan menggunakan bermacam spesifikasi tingkat kerumitan, penggunaan dari data mining sejauh ini sudah dapat digunakan untuk memperoleh informasi optimal dari sekelompok besar database besar yang memiliki kompleksitas [8].

1.3. Metodologi Penelitian

Saat ini persewaan sepeda mulai diperkenalkan di banyak kota perkotaan untuk peningkatan kenyamanan mobilitas. Penting untuk menyediakan sepeda sewaan dan dapat diakses oleh publik pada waktu yang tepat karena dapat mengurangi waktu tunggu. Pada akhirnya, menyediakan pasokan sepeda sewaan yang stabil di kota menjadi perhatian utama. Bagian penting adalah prediksi jumlah sepeda yang dibutuhkan setiap jam untuk pasokan sepeda sewaan yang stabil [9].

Dalam studi yang dilakukan pada kelompok data siap pakai, proses prediksi dan klasifikasi biasanya dilakukan oleh algoritma pembelajaran mesin [10]. Hasil yang diperoleh dapat bervariasi tergantung pada algoritma yang dipilih dan karakteristik kumpulan data yang dipelajari [11]. Dalam penelitian ini, studi eksperimental dilakukan pada klasifikasi kebutuhan stok sepeda sewaan menggunakan algoritma pembelajaran mesin sesuai dengan nilai pengukuran dan informasi kehidupan yang diperoleh dari individu. Klasifikasi adalah salah satu kasus yang menggunakan data berlabel dalam Supervised Learning. Klasifikasi membagi dataset ke dalam pelatihan data dan uji data

kemudian membentuk kelas baru prediksi oleh kelas kategori dan kelas label [11]. Penelitian ini akan menggunakan Deep Learning sebagai model utama algoritma untuk pengujian.

Deep Learning merupakan salah satu bidang dari Machine Learning yang memanfaatkan jaringan syaraf tiruan untuk implementasi permasalahan dengan dataset yang besar. Teknik Deep Learning memberikan arsitektur yang sangat kuat untuk Supervised Learning. Dengan menambahkan lebih banyak lapisan maka model pembelajaran tersebut bisa mewakili data citra berlabel dengan lebih baik. Pada Machine Learning terdapat teknik untuk menggunakan ekstraksi fitur dari data pelatihan dan algoritma pembelajaran khusus untuk mengklasifikasi citra maupun untuk mengenali suara. Namun, metode ini masih memiliki beberapa kekurangan baik dalam hal kecepatan dan akurasi [12].

Aplikasi konsep jaringan syaraf tiruan yang dalam (banyak lapisan) dapat ditangguhkan pada algoritma Machine Learning yang sudah ada sehingga komputer sekarang bisa belajar dengan kecepatan, akurasi, dan skala yang besar. Prinsip ini terus berkembang hingga Deep Learning semakin sering digunakan pada komunitas riset dan industri untuk membantu memecahkan banyak masalah data besar seperti Computer vision, Speech recognition, dan Natural Language Processing [12]. Feature Engineering adalah salah satu fitur utama dari Deep Learning untuk mengekstrak pola yang berguna dari data yang akan memudahkan model untuk membedakan kelas. Feature Engineering juga merupakan teknik yang paling penting untuk mencapai hasil yang baik pada tugas prediksi. Namun, sulit untuk dipelajari dan dikuasai karena kumpulan data dan jenis data yang berbeda memerlukan pendekatan teknik yang berbeda juga [13].

Pada penelitian ini akan menggunakan algoritma Deep Learning yang nantinya akan dioptimasikan menggunakan optimasi atribut dan optimasi bobot untuk meningkatkan tingkat akurasinya dalam prediksi kebutuhan stok sewa sepeda. Tujuan utama dari penelitian ini adalah untuk meningkatkan akurasi dari dataset Seoul Bike Sharing Demand Data Set dengan menerapkan algoritma deep learning untuk memprediksi kebutuhan stok sepeda disetiap harinya dan pemilihan fitur seleksi untuk mencapai akurasi yang lebih baik dibandingkan hasil prediksi sebelumnya.

2. PEMBAHASAN

Penelitian ini didasari oleh paradigma data mining yang menjelaskan proses pencarian atau penambangan knowledge. Tujuan dari penelitian ini adalah untuk mengusulkan fungsi optimasi berdasarkan algoritma Deep Learning. Fungsi objektif ini yang nantinya digunakan dalam optimasi algoritma pada dataset “Seoul Bike Sharing Demand Data Set”.

A. Identifikasi Masalah.

Sama seperti yang telah dijelaskan pada pendahuluan dalam penelitian ini, akan menggunakan algoritma Deep Learning yang nantinya akan dioptimasikan menggunakan optimasi atribut dan optimasi bobot untuk meningkatkan tingkat akurasinya dalam prediksi kebutuhan stok sewa sepeda. meningkatkan akurasi dari dataset Seoul Bike Sharing Demand Data Set dengan menerapkan algoritma deep learning untuk memprediksi kebutuhan stok sepeda disetiap harinya dan pemilihan fitur seleksi untuk mencapai akurasi yang lebih baik dibandingkan hasil prediksi sebelumnya.

B. Data Set

Untuk melakukan suatu penelitian tentunya dataset merupakan bahan utama yang akan diolah menggunakan suatu algoritma. Dataset yang digunakan dalam penelitian ini adalah data Seoul Bike Sharing Demand yang diambil dari website repositori UCI. Dataset ini berisikan rekam jejak peminjaman sepeda sejumlah 8760 data yang disebarkan secara publik berdasarkan jumlah sepeda yang dipinjam, tanggal bulan dan tahun peminjaman, suhu udara, rata-rata waktu pinjam sepeda per jam dalam satu hari, tanggal bulan dan tahun peminjaman, suhu udara, tingkat kelembaban udara, kecepatan angin, jarak penglihatan mata dalam satuan 10m, suhu embun, radiasi sinar uv, tingkat rintik hujan, tingkat ketebalan salju, data set ini dibagi menjadi 4 musim yaitu, musim dingin, musim semi, musim panas dan musim gugur, lalu dibagi berdasarkan hari kerja dan hari libur dan terakhir waktu-waktu tertentu orang-orang meminjam sepeda. Penjelasan lebih mudahnya ada pada table 1.

Table 1. Informasi Atribut Dalam Data Set Seoul Bike Sharing Demand

No	Nama Data	Deskripsi Data	Tipe Data
1	Date	Tanggal, bulan dan tahun	Date
2	Dew point temperature(°C)	Suhu embun pagi	Real
3	Functioning Day	Waktu-waktu tertentu orang-orang meminjam sepeda	Binominal
4	Holiday	Hari kerja dan hari libur	Binominal
5	Hour	Rata-rata waktu pinjam sepeda per jam dalam satu hari	Integer

6	Humidity(%)	Tingkat kelembaban uda	Integer
7	Rainfall(mm)	Tingkat rintik hujan	Integer
8	Rented Bike count	Jumlah sepeda yang dipinjam	Integer
9	Seasons	Terdiri dari 4 musim, musim dingin, musim semi, musim panas, musim gugur	Polynominal
10	Snowfall (cm)	Tingkat ketebalan salju	Integer
11	Solar Radiation (MJ/m2)	Radiasi sinar uv	Real
12	Temperature(°C)	Suhu udara	Integer
13	Visibility (10m)	Jarak penglihatan mata	Integer
14	Wind speed (m/s)	Kecepatan angin	Real

C. Pre-procesing

Kumpulan teknik yang digunakan sebelum penerapan metode data mining disebut sebagai data preprocessing untuk data mining dan dikenal sebagai salah satu masalah yang paling berarti dalam Penemuan Pengetahuan yang terkenal dari proses Data [14]. Karena data kemungkinan besar tidak sempurna, mengandung inkonsistensi dan redundansi tidak dapat diterapkan secara langsung untuk memulai proses penambangan data. Jumlah data yang dikumpulkan lebih besar membutuhkan mekanisme yang lebih canggih untuk menganalisisnya. Pre-procesing data mampu menyesuaikan data dengan persyaratan yang diajukan oleh masing-masing algoritma penambangan data, memungkinkan untuk memproses data yang tidak mungkin dilakukan[15].

Data dalam penelitian ini tidak dapat terbaca seluruhnya sebab aplikasi data mining yang digunakan dalam eksperimen ini yaitu rapidminer tidak bisa membaca data set lebih dari 5000 data, sehingga peneliti terpaksa mengurangi sejumlah data agar hasil yang diberikan seimbang.

D. Komparasi Algoritma

Klasifikasi, model pembelajaran mesin yang diawasi digunakan untuk memprediksi hasil dari data. Penelitian ini mengusulkan suatu teknik untuk memprediksi kebutuhan persediaan sepeda sewaan dengan menggunakan prosedur klasifikasi dan untuk meningkatkan akurasi klasifikasi melalui penggunaan pengklasifikasi. Dalam penelitian ini, pengujian dilakukan pada beberapa algoritme yang paling umum digunakan, yaitu Neural Net, Generelized Linear Model, Support Vector Machine, Random Forest dan Deep Learning.

E. Klasifikasi

Tahap klasifikasi merupakan tahap untuk mengklasifikasikan kualitas dari dataset "Seoul Bike Sharing Demand Data Set". Pada penelitian ini dilakukan validasi menggunakan Cross Validation yang sebelumnya telah dilakukan pembagian data training dan data testing untuk mentukan model mana yang memiliki tingkat akurasi yang terbaik. Kemudian validasi menggunakan Split Validation untuk menguji model yang telah diambil menggunakan cross validation.

F. Evaluasi Matriks

Perhitungan performansi suatu model diperlukan untuk mengetahui apakah model tersebut benar atau tidak dalam proses klasifikasi. Ada beberapa metode evaluasi kinerja untuk model pembelajaran mesin. Perpaduan alat evaluasi yang berbeda diharapkan mendukung pengembangan penelitian analitis[16].

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Dalam penelitian ini, empat metrik dasar (accuracy, precision, recall, F-Score) akan diperiksa untuk mengetahui perbedaan dalam algoritma berbasis pembelajaran mesin. Pada saat menggunakan perhitungan performansi akurasi, presisi, dan recall, digunakan variabel-variabel yang tercantum dalam confusion matrix. Unsur-unsur matriks konfusi adalah true positive (TP), true negative (TN), false positive (FP) and false negative (FN) [17]. Label positif dikategorikan sebagai No Holiday dan label negatif dikategorikan Holiday. Masing-masing variabel pada Tabel 2 memiliki penjelasannya, yaitu:

Tabel 2. Confusion Matrix

	Predicted Positive	**Predicted Negative**
Actual Postive	True Postive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

A. True Positive (TP) menunjukkan data prediksi menyatakan positif atau benar terjadi pada No

Holiday atau hari kerja. Data yang ditunjukkan itu sesuai dengan No Holiday atau hari kerja.

B. True Negative (TN) menunjukkan data prediksi menyatakan negatif atau tidak benar terjadi pada Holiday atau hari libur. Data yang ditunjukkan menyatakan itu sesuai dengan Holiday atau hari libur.

C. False Positive (FP) menunjukkan data prediksi menyatakan positif atau benar terjadi pada No Holiday atau hari kerja. Data yang ditunjukkan tidak sesuai dengan Holiday atau hari libur.

D. False Negative (FN) menunjukkan data prediksi menyatakan negatif atau tidak benar terjadi pada Holiday atau hari libur. Data yang ditunjukkan tidak sesuai dengan No Holiday atau hari kerja.

Kemudian rumus untuk menghitung nilai Accuracy, Precision, dan Recall sebagai berikut:

1. Accuracy (ACC) adalah efektivitas keseluruhan dari hasil klasifikasi.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

2. Precision (PREC) merupakan persentase dari label data dengan label positif yang diberikan klasifkasi [12].

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

3. Recall (REC) atau sensitivity adalah efektivitas dari pengeklasifikasi dalam mengidentifikasi label postif [12].

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

G. Area under The Curve (AUC)

Area Under Curve merupakan salah satu metode umum yang digunakan untuk menghitung nilai under the ROC curve [12]. Area Under the Curve dapat diartikan sebagai nilai probabilitas, jika memilih satu contoh positif dan negatif secara acak, metode klasifikasi akan memberikan skor yang lebih tinggi pada contoh positif dari pada contoh negatif. Oleh karena itu, bilai AUC yang lebih tinggi dapat menunjukan metode klasifikasi yang lebih baik[12], [18]. Berikut ini adalah formula dalam mencari nilai AUC dilihat dari hasil confusion matrix [19]:

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right)$$

Nilai Area under Curve (AUC) akan selalu berada pada range 0-1, karena bagian dari luas persegi satuan dengan sumbu x dan sumbu y memiliki nilai dari 0 sampai 1. Nilai diatas 0,5 dikatakan nilai yang menarik karena prediksi acak menghasilkan garis diagonal antara (0,0) dan (1,1) yang memiliki luas 0,5. Kualitas klasifikasi keakuratan dari tes diagnostik menggunakan nilai AUC ditunjukkan pada Tabel 3 [12].

Tabel 3. Keakuratan Hasil Klasifikasi Berdasarkan AUC

Nilai AUC	***Kategori***
0.90 – 1.00	Sangat Baik
0.80 – 0.90	Baik
0.70 – 0.80	Cukup Baik
0.60 – 0.70	Kurang Baik
0.50 – 0.60	Buruk

3. PROSES DAN HASIL PENELITIAN

Penelitian dilakukan dalam lima model algoritma klasifikasi dengan 10 kali percobaan menggunakan cross validation dan diambil berdasarkan nilai akurasi terbaik. Hasil eksperimen dapat dilihat pada tabel 4. Hasil eksperimen yang tertuang pada tabel 4 menyatakan bahwa algoritma deep learning menghasilkan tingkat akurasi yang lebih rendah dibandingkan dengan algoritma lain. Kemudian nantinya akan dilakukan optimasi menggunakan fitur seleksi genetika algoritma dengan berdasarkan variasi rasio split validation untuk klasifikasi stok kebutuhan sepeda pinjaman. Hasil komparasi algoritma dapat dilihat pada tabel 4. Jika dilihat dari tabel 6 nilai akurasi yang dihasilkan oleh seleksi fitur genetik algoritma memiliki nilai yang lebih baik dari sebelumnya.

Tabel 4. Komparasi Algortima

Algoritma	Akurasi	AUC
Neural Net	95.71%	0.841
GLM	94%	0.707
SVM	94%	0.5
RF	95%	0.99
DL	90%	0.770

Berdasarkan komparasi algoritma tersebut diketahui bahwa algoritma Deep Learning menghasilkan nilai Accuracy yaitu sebesar 90% dan AUC sebesar 0,770. Berikut adalah tabel confusion matrix yang dihasilkan model klasifikasi deep learing dapat dilihat pada tabel 5.

Tabel 5. Confusion Matrix

	True 1	**True 0**	**Class precision**
Pred. 1	4248	192	95.68%
Pred. 0	265	96	26.59%
Class Recall	94.13%	33.33%	

1. Accuracy

Dari hasil pengujian nilai akurasi yaitu 90%. Berikut hasil perhitungan manual:

$$Accuracy = \frac{4248 + 96}{4248 + 96 + 265 + 192} 100\%$$
$$= \frac{4344}{4801} \times 100\%$$
$$= 0.904 \times 100\%$$
$$= 90\%\%$$

2. Precision

Dari hasil pengujian nilai Precision yaitu 95.68% untuk precision *No Holiday* = 192 dan 26.59% untuk precision *Holiday* = 96. Berikut hasil perhitungan manual:

$$Precision\ 1 = \frac{4248}{4248 + 265} \times 100\%$$
$$= \frac{4248}{4513} \times 100\%$$
$$= 0.94 \times 100\%$$
$$= 94\%$$

$$Precision\ 2 = \frac{96}{96 + 265} \times 100\%$$
$$= \frac{96}{361} \times 100\%$$
$$= 0.2659 \times 100\%$$
$$= 26.59\%$$

3. Recall

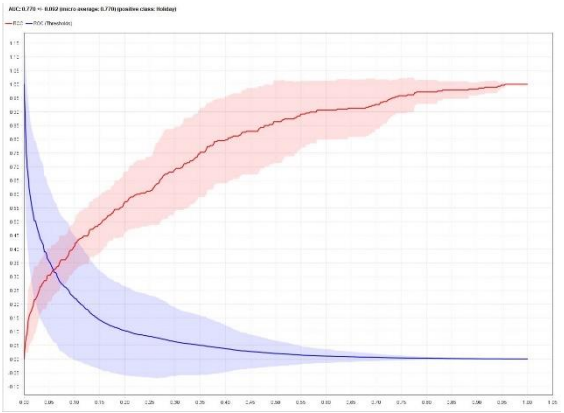
Dari hasil pengujian nilai Recall yaitu 95.68% untuk precision *No Holiday* = 192 dan 26.69% untuk precision *Holiday* = 92. Berikut hasil perhitungan manual:

$$Recall\ 1 = \frac{4248}{4248 + 265} \times 100\%$$
$$= \frac{4248}{4513} \times 100\%$$
$$= 0.94 \times 100\%$$
$$= 94\%$$

$$Recall\ 2 = \frac{96}{96 + 192} \times 100\%$$
$$= \frac{96}{288} \times 100\%$$
$$= 0{,}33 \times 100\%$$
$$= 33\%$$

4. Area Under Curve

Dari hasil pengujian nilai AUC model algoritma Deep Learning adalah 0.770. Berdasarkan nilai pengujian tersebut menunjukan bahwa algoritma Deep Learning mencapai klasifikasi Cukup Baik. Kinerja algoritma dapat dilihat dari kurva ROC-AUC seperti yang ditampilkan pada gambar 1 [12]. Gambar 1 menunjukan kurva ROC yang menggambarkan hubungan antara data testing dan datra prediksi. Dari kurva tersebut didapatkan nilai Area Under the Curve (AUC) yaitu luas daerah dibawah kurve ROC [12]. Nilai AUC yang diperoleh sebesar 0,770. Dapat dikatakan bahwa hasil klasifikasi yang didapat cukup baik. Selanjutnya adalah tahap pendeketan menggunakan fitur optimasi.



Gambar 1. ROC Average

Setelah didapatkan model yang terbaik tahap selanjutnya menerapkan optimasi fitur. Pada penelitian ini dilakukan dalam lima eksperimen berdasarkan variasi rasio Split Validation. Pada setiap eksperimen akan diterapkan akan diterapkan algoritma terbaik dari hasil komparasi yaitu Random Forest dan optimasi fitur algoritma genetika. Hasil eksperimen dapat dilihat pada tabel 6.

Tabel 6. Hasil Eksperimen Optimasi Seleksi

Rasio Spilt Validasi	Deep Learning	Optimize Selection (Evolutinary)
0,5	92%	94.12%
0,6	91%	94.38%
0,7	92.56%	95.83%
0,8	88%	94.38%
0,9	94%	93.75%
Rata-rata	92%	94.49%

Berdasarkan hasil eksperimen optimasi seleksi pada tabel 6, Algoritma Deep Learning menghasilkan nilai akurasi rata-rata 92%, sebelum melakukan optimasi atribut memiliki hasil terbaik diperoleh 94% dengan rasio split validation 0,9 dan sebesar 95.83% setelah menerapkan optimize selection (evolutinary). Ini membuat split rasio 0,7 memiliki nilai akurasi tertinggi dibandingkan dengan yang lainnya.

Berdasarkan hasil eksperimen pada tabel 6 dapat disimpulkan bahwa terdapat peningkatan

kinerja pada algoritma deep learning, peningkatan ini cukup tinggi sebab peningkatan eksperimen menggunakan split rasio 0,8 yang memiliki nilai akurasi 88% terendah dibandingkan dengan rasio yang lainnya, setelah menerapkan optimize weight forward hasilnya menjadi 94,38%.

Tabel 7. Hasil Eksperimen Optimasi Bobot

Rasio Spilt Validasi	Deep Learning	Optimize Weight (Forward)
0,5	92%	94.88%
0,6	91%	95.47%
0,7	92,56%	93.47%
0,8	88%	95.63%
0,9	94%	95.83%
Rata-rata	92%	95.06%

Berdasarkan hasil eksperimen optimasi artibut pada tabel 7, Algoritma Deep Learning menghasilkan nilai akurasi rata-rata 92%, sebelum melakukan optimasi bobot memiliki hasil terbaik diperoleh 94% dengan rasio split validation 0,9 dan sebesar 95.83% setelah menerapkan optimize weight (forward). Ini membuat split rasio 0,9 memiliki nilai akurasi tertinggi dibandingkan yang lainnya.

Berdasarkan hasil eksperimen pada tabel 7 dapat disimpulkan bahwa terdapat peningkatan kinerja pada algoritma deep learning, peningkatan ini cukup tinggi sebab peningkatan eksperimen menggunakan split rasio 0,8 yang memiliki nilai akurasi 88% terendah dibandingkan dengan rasio yang lainnya, setelah menerapkan ooptimize weight forward hasilnya menjadi 95.63% yang membuatnya menjadi nilai akurasi terbaik kedua setelah 95.83% dari rasio split validasi 0,9. Tujuan dari penelitian ini adalah untuk mengoptimasi algoritma deep learning agar memiliki nilai akurasi yang lebih baik dari sebelumnya, berdasarkan hasil eksperimen optimasi artibut pada tabel 6, sebelum melakukan optimasi seleksi memiliki hasil terbaik diperoleh 94% dengan rasio split validation 0,9 dan sebesar 95.83% setelah menerapkan optimize selection (evolutinary). Ini membuat split rasio 0,7 memiliki nilai akurasi tertinggi dibandingkan dengan yang lainnya. Berdasarkan hasil eksperimen optimasi artibut pada tabel 7, sebelum melakukan optimasi bobot memiliki hasil terbaik diperoleh 94% dengan rasio split validation 0,9 dan sebesar 95.83% setelah menerapkan optimize weight (forward). Ini membuat split rasio 0,9 memiliki nilai akurasi tertinggi dibandingkan yang lainnya. Penelitian ini juga menerapkan evaluasi matriks untuk menghitung nilai Accuracy, Precision, dan Recall demi memperoleh data yang lebih akurat. Data yang diperlukan untuk perhitungan itu tersedia pada tabel 5, hasil perhitungan yang diperoleh melalui evaluasi matriks cukup sesuai dengan apa yang ditunjukkan dalam tabel 5.

4. KESIMPULAN

Penelitian ini dimulai dengan membandingkan hasil 5 model algoritma prediksi pada "Bike Sharing Demand Data Set" dengan menerapkan cross validation sebesar 10 fold, 5 algortima yang digunakan dalam penelitian ini adalah Neural Net, Generelized Linear Model, Support Vector Machine, Random Forest dan Deep Learning. Setelah menerapkan cross validation 10 fold, ditemukan bahwa algoritma deep learning memiliki nilai akurasi sebesar 90%, 4-5% lebih rendah dibandingkan dengan 4 algortima lainnya yang memiliki nilai akurasi lebih tinggi, akan tetapi nilai AUC yang diperoleh sebesar 0.770 termasuk Cukup Baik. Peningkatan yang signifikan dalam penelitian ini adalah nilai akurasi pada rasio split validasi 0,8 sebesar 88%, ini merupakan nilai akurasi terendah dibandingkan rasio yang lainnya, setelah melakukan eksperimen ini, diperoleh nilai akurasi sebesar 94.38% untuk optimasi seleksi dan nilai akurasi sebesar 95.63% untuk optimasi bobot. Melihat hasil yang diperoleh dari eksperimen yang telah dilaksanakan, peneliti menyatakan eksperiman yang dilakukan berhasil meningkatkan nilai akurasi algoritma deep learning.

PUSTAKA

[1] H. Fitroni, “Fenomena peningkatan motivasi bersepeda masyarakat di masa pandemi COVID-19,” *Sporta Saintika*, vol. 6, no. 1, pp. 109–118, 2021.

[2] E. N. Deniati and A. Annisaa, “Hubungan Tren Bersepeda dimasa Pandemi Covid-19 dengan Imunitas Tubuh Lansia,” *Sport Science and Health*, vol. 3, no. 3, pp. 125–132, May 2021, doi: 10.17977/UM062V3I32021P125-132.

[3] “Menjelajahi Seoul dan Korea dengan Bersepeda | kumparan.com.” https://kumparan.com/jifiawan-gana-putra/menjelajahi-seoul-dan-korea-dengan-bersepeda-1vL7Ax2ZLFh (accessed Jan. 17, 2023).

[4] B. S. Pane, “Peranan olahraga dalam meningkatkan kesehatan,” *Jurnal pengabdian kepada masyarakat*, vol. 21, no. 79, pp. 1–4, 2015.

[5] “Pakar UGM Jelaskan Cara Aman Bersepeda di Tengah Pandemi Corona | Universitas Gadjah Mada.” https://ugm.ac.id/id/berita/19632-pakar-ugm-jelaskan-cara-aman-bersepeda-di-tengah-pandemi-corona (accessed Jan. 17, 2023).

[6] L. Lin, Z. He, and S. Peeta, “Predicting station-level hourly demand in a large-scale bike-sharing network: A graph

convolutional neural network approach," *Transp Res Part C Emerg Technol*, vol. 97, pp. 258–276, 2018, doi: https://doi.org/10.1016/j.trc.2018.10.011.

[7] O. O. Oladimeji and O. Oladimeji, "Predicting survival of heart failure patients using classification algorithms," *JITCE (Journal of Information Technology and Computer Engineering)*, vol. 4, no. 02, pp. 90–94, 2020.

[8] D. Riana, Y. Ramdhani, R. T. Prasetio, and A. N. Hidayanto, "Improving Hierarchical Decision Approach for Single Image Classification of Pap Smear.," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 8, no. 6, 2018.

[9] V. E. Sathishkumar, J. Park, and Y. Cho, "Using data mining techniques for bike sharing demand prediction in metropolitan city," *Comput Commun*, vol. 153, pp. 353–366, 2020.

[10] R. T. Prasetio, A. A. Rismayadi, N. Suryana, and R. Setiady, "Features Selection and k-NN Parameters Optimization based on Genetic Algorithm for Medical Datasets Classification," *Heart Disease (SPECTF)*, vol. 267, no. 44, p. 2, 2020.

[11] M. Huljanah, Z. Rustam, S. Utama, and T. Siswantining, "Feature selection using random forest classifier for predicting prostate cancer," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 546, no. 5, p. 052031.

[12] A. N. Rachmi, "Implementasi Metode Random Forest Dan Xgboost Pada Klasifikasi Customer Churn," *Laporan Tugas Akhir. Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Islam Indonesia Yogyakarta*, 2020.

[13] T. Nurhikmat, "Implementasi deep learning untuk image classification menggunakan algoritma Convolutional Neural Network (CNN) pada citra wayang golek," 2018.

[14] S. García, J. Luengo, and F. Herrera, *Data preprocessing in data mining*, vol. 72. Springer, 2015.

[15] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Anal*, vol. 1, no. 1, pp. 1–22, 2016.

[16] A. Ishaq *et al.*, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE access*, vol. 9, pp. 39707–39716, 2021.

[17] A. M. Hay, "The derivation of global estimates from a confusion matrix," *Int J Remote Sens*, vol. 9, no. 8, pp. 1395–1398, 1988.

[18] F. Gorunescu, *Data Mining: Concepts, models and techniques*, vol. 12. Springer Science & Business Media, 2011.

[19] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf Process Manag*, vol. 45, no. 4, pp. 427–437, 2009.