



Perbandingan Algoritma *Stochastic Gradient Descent* dan *Naïve Bayes* Pada Klasifikasi *Diabetic Retinopathy*

Ryan Rinaldi Hadistio^{1*}, Herman Mawengkang², Muhammad Zarlis¹

¹Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara Medan, Indonesia

²Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sumatera Utara Medan, Indonesia

Email: ¹ry4n.rinald1@gmail.com, ²hmawengkang@yahoo.com, ³m.zarlis@usu.ac.id

Email Penulis Korespondensi: ry4n.rinald1@gmail.com

Abstrak— Tujuan pada riset ini yaitu untuk melakukan perbandingan terhadap kinerja dari algoritma *Stochastic Gradient Descent* dan *Naïve Bayes* dalam melakukan klasifikasi pada penyakit *Diabetic Retinopathy*. Adapun *Diabetic Retinopathy* merupakan komplikasi dari diabetes yang mengakibatkan gangguan pada retina mata. Gangguan tersebut dapat diketahui dengan deteksi awal melalui data yang diekstraksi dari citra mata. Pada riset ini menggunakan data sumber yang berasal dari *UCI Machine Learning Repository* yaitu *Diabetic Retinopathy Debrecen* yang berjumlah 1.151 record data dengan 19 atribut yang terdiri dari 18 atribut dan 1 target atribut. Pengujian validasi menggunakan metode *Cross Validation* dengan jumlah k sebanyak 10. Dari perbandingan pengujian dua metode yang diusulkan, algoritma *Stochastic Gradient Descent* menghasilkan akurasi rata-rata pengujian sebesar 70.16 %, sementara pada *Naïve Bayes* menghasilkan akurasi rata-rata sebesar 56.74 %. Dari perbandingan kedua algoritma tersebut, algoritma *Stochastic Gradient Descent* diketahui lebih unggul dalam melakukan klasifikasi pada *Diabetic Retinopathy Debrecen Dataset*.

Kata Kunci: Klasifikasi; Stochastic Gradient Descent; Naïve Bayes; Diabetic Retinopathy; Cross Validation

Abstract— The purpose of this research is to compare the performance of the Stochastic Gradient Descent and Naïve Bayes algorithms in classifying Diabetic Retinopathy. Diabetic retinopathy is a complication of diabetes that causes damage to the retina of the eye. These disturbances can be detected by early detection through data extracted from eye images. This research uses source data from the UCI Machine Learning Repository, namely Diabetic Retinopathy Debrecen, totaling 1,151 data records with 19 attributes consisting of 18 attributes and 1 target attribute. The validation test uses the Cross Validation method with a total of 10 k. From the comparison of the two proposed methods, the Stochastic Gradient Descent algorithm produces an average test accuracy of 70.16%, while Naïve Bayes produces an average accuracy of 56.74%. From the comparison of the two algorithms, the Stochastic Gradient Descent algorithm is known to be superior in classifying the Diabetic Retinopathy Debrecen Dataset.

Keywords: Classification; Stochastic Gradient Descent; Naïve Bayes; Diabetic Retinopathy; Cross Validation

1. PENDAHULUAN

Diabetic Retinopathy adalah salah satu bentuk komplikasi dari penyakit *Diabetes Melitus* yang dimana mengakibatkan kadar gula yang tinggi pada tubuh sehingga akhirnya dapat mengakibatkan kerusakan pada pembuluh darah retina mata, terutama di jaringan-jaringan yang sensitif terhadap cahaya. Kondisi ini dapat diderita oleh siapapun yang menderita diabetes tipe 1 maupun 2, terutama mereka yang gula darahnya tidak terkontrol dan telah menderita diabetes dalam jangka waktu yang lama. Hal ini dikarenakan penderita diabetes memiliki resiko yang tinggi untuk memunculkan berbagai masalah Kesehatan [1]. Menurut Atlas of the International Diabetes Federation (IDF) 2017, prevalensi diabetes di Indonesia masih terus meningkat. Indonesia merupakan negara terbesar keenam di dunia setelah China, India, Amerika Serikat, Brazil dan Meksiko. Jumlah penderita diabetes usia 20-79 tahun sekitar 10,3 juta. Riset Kesehatan Dasar (Riskesdas) juga menunjukkan peningkatan prevalensi diabetes yang signifikan, dari 6,9% pada tahun 2013 menjadi 8,5% pada tahun 2018, sehingga perkiraan jumlah penderita diabetes di Indonesia mencapai 22 juta. [2].

Dalam melakukan diagnosis secara dini pada penyakit *Diabetes Retinopathy* dapat dilakukan dengan menerapkan salah satu fitur pada ilmu *Artificial Intelligent* yaitu *Data Mining*. Proses pada metode data mining bertujuan untuk menguraikan temuan baru di dalam dataset dan menjelaskan suatu proses yang menggunakan Teknik Statistik, Matematis, *Artificial Intelligence*, dan *Machine Learning* untuk melakukan ekstrak dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database tersebut.

Pembelajaran Mesin (*Machine Learning*) dapat didefinisikan sebagai proses dimana sistem dapat meningkatkan performa dari pengalaman. Terdapat tiga jenis *Machine Learning* yaitu *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning* [3]. Salah satu *Task* dalam *Supervised Learning* adalah klasifikasi, dimana ini menjadi fokus permasalahan dari penelitian ini[4].

Diabetic Retinopathy dapat didiagnosis dengan melihat ciri-ciri pada mata. Dataset *Diabetic Retinopathy Debrecen* yang terdiri dari 1151 model mata dalam bentuk data numerik yang merupakan hasil ekstraksi dari citra mata. Dataset ini juga digunakan pada penelitian yang dilakukan oleh Bahrami and Sajedi pada tahun 2018 untuk menguji performa algoritma *Random Forest* [5]. Deteksi dini *Diabetic Retinopathy* dilakukan pada penelitian yang dilakukan oleh Heryawan pada tahun 2017 dengan morfologi matematika dengan teknik pengolahan citra untuk mengolah dataset gambar yaitu DIARETDB1 sebanyak 89 citra fundus mata dan e-optha sebanyak 200 fundus mata. Pengujian menggunakan dataset DIARETDB1 tersebut diperoleh



hasil akurasi 86,51%, sensitivitas 90% dan spesifitas 55%, sementara pengujian menggunakan dataset e-opta diperoleh hasil akurasi 70.5%, sensitivitas 80% dan spesifitas sebesar 60% [6].

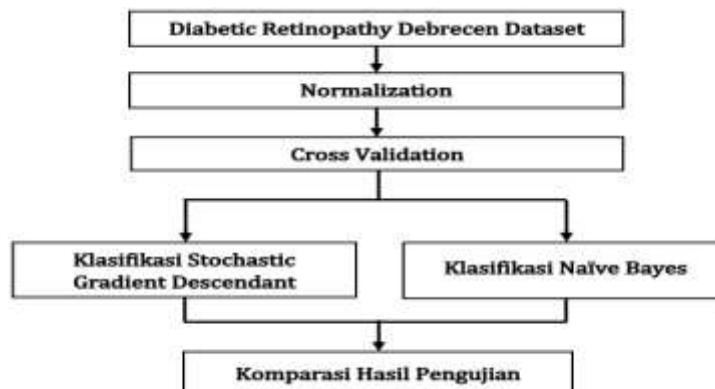
Penelitian yang dilakukan oleh Al Rivan *et.al* (2020) yang membahas tentang diagnosis penyakit Diabetic Retinopathy dengan menerapkan metode *clustering* optimasi metode Fuzzy C-Means dan K-Means menggunakan algoritma Genetika dengan hasil yang didapatkan yaitu [1]. Pada penelitian yang dilakukan oleh Subarkah, *et.al* (2021) dalam melakukan komparasi akurasi algoritma CART dan *Neural Network* dalam mendiagnosis penyakit Diabetic Retinopathy dengan menggunakan data pengujian yang diperoleh dari *UCI Machine Learning Repository* dengan hasil pengujian akurasi yang diperoleh yaitu pada metode CART sebesar 63.4231 % dan dengan Neural Network sebesar 72.285 % [7]. Pada penelitian Tyasnurita & Pamungkas (2020) membahas mengenai penerapan metode regresi logistic dalam mendiagnosis penyakit *Diabetic Retinopathy* dengan akurasi yang diperoleh yaitu sebesar 80.17 % [7].

Penelitian yang dilakukan oleh Zulfikar & Lukman pada tahun 2016, perbandingan metode *Naïve Bayes* dan K-NN dalam melakukan identifikasi penyakit mata untuk melihat kinerja dari algoritma *Naïve Bayes Classifier* dan *Nearest Neighbor* memiliki keakuratan keakuratan yang sebanding. Dalam hal kecepatan, *Nearest Neighbor* memiliki rata-rata catatan lebih cepat 0.027 detik dari pada *Naïve Bayes Classifier* [8]. Penelitian Umar *et.al* (2018) meneliti tentang penerapan *Stochastic Gradient Descent* (SGD) pada proses klasifikasi kinerja programmer berdasarkan aktivitas media sosial dengan hasil yang diperoleh yaitu menunjukkan nilai akurasi sebesar 80%, nilai precision 81% dan nilai recall 80% [9].

Pada penelitian ini penulis akan menerapkan metode *Stochastic Gradient Descent* (SGD) dan *Naïve Bayes* untuk memprediksi hasil akurasi dari klasifikasi penyakit *Diabetic Retinopathy* dengan menggunakan dataset yang berasal dari *UCI Machine Learning Repository* yaitu *Diabetic Retinopathy Debrecen Dataset*. Dan juga untuk melihat hasil yang lebih baik dari perbandingan kinerja kedua metode tersebut dalam klasifikasi penyakit *Diabetic Retinopathy*.

2. METODOLOGI PENELITIAN

Pengujian dari dua metode yang diusulkan dilakukan dengan tahapan yang alurnya dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Metode Penelitian

2.1 Dataset

Pada tahap ini, sumber data yang akan digunakan adalah dataset Diabetic Retinopathy Debrecen yang berasal dari database *UCI Machine Learning Repository*. Kumpulan data berisi 1151 record dan 19 atribut (18 atribut dan 1 atribut target). Atribut target terdiri dari dua output, yaitu retinopati diabetik normal dan retinopati diabetik indikatif. Dua dari atribut ini akan digunakan sebagai data keluaran untuk retinopati diabetik. Atribut retinopati diabetik ditunjukkan pada Tabel 1 dan Tabel 2.

Tabel 1. Rincian Atribut Diabetic Retinopathy Debrecen Dataset

Atribut	Keterangan	Penanda
Attribute 0	Hasil Biner Pada Kualitas Penilaian	X1
Attribute 1	Hasil Biner Pada Pre-Screening	X2
Attribute 2	Hasil Deteksi MA	X3
Attribute 3	Hasil Deteksi MA	X4
Attribute 4	Hasil Deteksi MA	X5
Attribute 5	Hasil Deteksi MA	X6
Attribute 6	Hasil Deteksi MA	X7
Attribute 7	Hasil Deteksi MA	X8
Attribute 8	Hasil Normalisasi Pada Eksudat	X9



Atribut	Keterangan	Penanda
Attribute 9	Hasil Normalisasi Pada Eksudat	X10
Attribute 10	Hasil Normalisasi Pada Eksudat	X11
Attribute 11	Hasil Normalisasi Pada Eksudat	X12
Attribute 12	Hasil Normalisasi Pada Eksudat	X13
Attribute 13	Hasil Normalisasi Pada Eksudat	X14
Attribute 14	Hasil Normalisasi Pada Eksudat	X15
Attribute 15	Hasil Normalisasi Pada Eksudat	X16
Attribute 16	Jarak Euclidean pada pusat manula dan pusat cakram optic	X17
Attribute 17	Diameter cakram optic	X18
Attribute 18	Hasil biner pada klasifikasi berbasis AM/FM	X19
Target Attribute 19	1=mengandung tanda-tanda DR. 0=tidak ada tanda-tanda DR	-

Tabel 2. Attribute Target Diabetic Retinopathy Debrecen Dataset

No.	Target Attribute	Jumlah Record Data
1	<i>Contains Signs</i>	611
2	<i>No Signs</i>	540
	Jumlah	1151

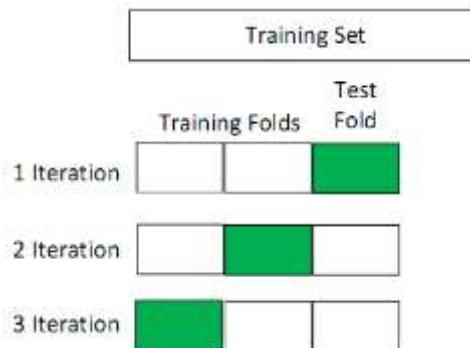
2.2 Normalisasi

Normalisasi bertujuan untuk mendapatkan data yang bersih dan mempersiapkan pengujian, termasuk identifikasi dan pemilihan atribut, pemrosesan atribut yang hilang atau tidak lengkap, dan proses diskritisasi nilai. Dalam penelitian ini, normalisasi *min-max* diterapkan. Fitur age dinormalisasi karena nilai pada fitur lainnya berupa 0 dan 1. Jadi normalisasi fitur age adalah untuk mendapatkan nilai antara 0 dan 1. Persamaan yang digunakan untuk menghitung normalisasi min max yaitu dengan persamaan (1) berikut: [7] [8]

$$\text{Normalisasi Min Max} = \frac{N - \text{Min}(n)}{\text{Max}(n) - \text{Min}(n)} \quad (1)$$

2.3 Cross Validation

Cross Validation adalah proses acak yang membagi kumpulan data menjadi K titik diskontinyu dengan ukuran yang kira-kira sama, dan setiap lipatan digunakan secara bergantian untuk menguji model yang diinduksi oleh algoritma klasifikasi x oleh lipatan K-1 lainnya[10][12]. Pada Gambar 2 berikut adalah contoh skema proses *Cross Validation*.



Gambar 2. Skema Proses Cross Validation

2.4 Stochastic Gradient Descent

Stochastic Gradient Descent (SGD) merupakan metode optimasi sederhana berbasis statistik yang efisien digunakan mencari nilai koefisien untuk meminimalkan *loss (error)* function pada skala besar [9][13]. Proses algoritma SGD adalah dengan menemukan nilai θ yang dapat meminimalkan fungsi $J(\theta)$. Untuk menentukan nilai awal θ digunakan algoritma pencarian, kemudian pada setiap iterasi nilai θ agar terus diperbarui sampai menemukan titik minimum atau nilai J yang paling minimum. Proses pembaharuan nilai θ pada setiap iterasi menggunakan Persamaan (2). Pembaharuan dilakukan secara bersamaan untuk semua nilai $j = 0, \dots, n$. Variable α merupakan *learning rate* yang mengatur seberapa besar pembaharuan nilai. Persamaan nilai $J(\theta)$ dapat dilihat pada Persamaan (3) [13] [14], di mana L merupakan loss function yang digunakan pada data pelatihan $(x_1, y_1), \dots, (x_n, y_n)$ dan R merupakan regularisasi atau penalty terhadap kompleksitas model.

$$\theta_j = \theta_j - \alpha \frac{\partial y}{\partial x} J(\theta) \quad (2)$$



$$J(\theta) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \alpha R(w) \quad (3)$$

2.5 Naïve Bayes

Naïve Bayes adalah algoritma data mining yang menerapkan teorema Bayes dalam proses klasifikasi. *Naïve Bayes* sendiri memiliki definisi klasifikasi, dengan teknik probabilitas dan statistik, yang dapat memprediksi kejadian di masa depan berdasarkan kejadian yang sudah ada sebelumnya [10][11][18]. Persamaan teorema Bayes dan interpretasinya ditunjukkan pada Persamaan (4) berikut [19][20].

$$P(H|X) = \frac{P(H|X)P(X)}{P(X)} \quad (4)$$

Keterangan:

- H : Hipotesis data X
- $P(H|X)$: Probabilitas H berdasarkan X
- $P(X|H)$: Probabilitas X berdasarkan H
- $P(H)$: Probabilitas H
- $P(X)$: Probabilitas X

3. HASIL DAN PEMBAHASAN

3.1 Normalisasi Dataset

Tahapan pertama yang dilakukan yaitu melakukan normalisasi pada dataset yang akan diujikan. Untuk proses normalisasi dataset dilakukan dengan perhitungan normalisasi *min-max*. Normalisasi *min-max* pada penelitian dibantu dengan menggunakan *RapidMiner Studio*. Adapun hasil yang diperoleh sebelum dataset dinormalisasi dengan dataset setelah dinormalisasi yaitu pada Tabel 3 dan Tabel 4 berikut.

Tabel 3. Diabetic Retinopathy Debrecen Dataset Sebelum Normalisasi

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	...	X17	X18	X19	Class
1	1	1	22	22	22	19	18	14	49.89	17.77	...	0.486	0.100	1	Class 0
2	1	1	24	24	22	18	16	13	57.70	23.79	...	0.520	0.144	0	Class 0
3	1	1	62	60	59	54	47	33	55.83	27.99	...	0.530	0.128	0	Class 1
4	1	1	55	53	53	50	43	31	40.46	18.44	...	0.483	0.114	0	Class 0
5	1	1	44	44	44	41	39	27	18.02	8.570	...	0.475	0.123	0	Class 1
6	1	1	44	43	41	41	37	29	28.35	6.935	...	0.502	0.126	0	Class 1
7	1	0	29	29	29	27	25	16	15.44	9.113	...	0.541	0.139	0	Class 1
8	1	1	6	6	6	6	2	1	20.67	9.497	...	0.576	0.071	1	Class 0
9	1	1	22	21	18	15	13	10	66.69	23.54	...	0.500	0.116	0	Class 1
10	1	1	79	75	73	71	64	47	22.14	10.05	...	0.560	0.109	0	Class 1
11	1	1	45	45	45	43	40	32	84.35	50.97	...	0.546	0.112	0	Class 0
12	1	0	25	25	25	23	22	18	22.48	13.94	...	0.551	0.139	1	Class 0
13	1	1	70	69	65	63	63	50	10.56	3.108	...	0.534	0.089	0	Class 1
14	1	1	48	43	39	32	27	18	23.01	6.737	...	0.501	0.138	1	Class 1
15	1	1	94	93	92	89	86	77	8.610	1.981	...	0.541	0.124	0	Class 0
16	1	1	20	18	16	15	13	9	65.11	33.12	...	0.569	0.089	1	Class 0
17	1	1	105	95	81	66	46	32	123.0	70.57	...	0.524	0.134	1	Class 1
18	1	1	25	25	24	23	22	19	17.03	9.976	...	0.552	0.108	0	Class 0
19	1	1	64	64	63	58	55	40	19.67	6.064	...	0.551	0.098	0	Class 0
20	1	0	46	41	39	32	23	15	115.5	21.29	...	0.540	0.104	1	Class 1
:	:	:	:	:	:	:	:	:	:	:	...	:	:	:	:
1151	1	1	7	7	7	7	7	69.42	7.031	1.750	...	0.556	0.088	0	Class 0

Tabel 4. Diabetic Retinopathy Debrecen Dataset Setelah Normalisasi

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	...	X17	X18	X19	Class
1	0.059	0.298	-0.64	-0.61	-0.57	-0.62	-0.55	-0.47	-0.24	-0.24	...	-1.29	-0.46	1.404	Class 0
2	0.059	0.298	-0.56	-0.53	-0.57	-0.67	-0.65	-0.53	-0.10	0.032	...	-0.08	2.005	-0.71	Class 0
3	0.059	0.298	0.920	0.957	1.046	1.027	0.935	0.784	-0.14	0.227	...	0.274	1.121	-0.71	Class 1
4	0.059	0.298	0.646	0.667	0.783	0.838	0.730	0.652	-0.40	-0.21	...	-1.42	0.354	-0.71	Class 0
5	0.059	0.298	0.217	0.294	0.388	0.412	0.525	0.387	-0.78	-0.67	...	-1.68	0.843	-0.71	Class 1
6	0.059	0.298	0.217	0.252	0.256	0.412	0.423	0.519	-0.61	-0.74	...	-0.72	1.020	-0.71	Class 1
7	0.059	-3.35	-0.36	-0.32	-0.26	-0.25	-0.19	-0.34	-0.83	-0.64	...	0.660	1.735	-0.71	Class 1
8	0.059	0.298	-1.26	-1.28	-1.27	-1.24	-1.37	-1.33	-0.74	-0.62	...	1.892	-2.08	1.404	Class 0
9	0.059	0.298	-0.64	-0.65	-0.75	-0.81	-0.80	-0.73	0.044	0.021	...	-0.82	0.465	-0.71	Class 1
10	0.059	0.298	1.583	1.580	1.660	1.832	1.806	1.711	-0.71	-0.60	...	1.345	0.039	-0.71	Class 1
11	0.059	0.298	0.256	0.335	0.432	0.506	0.576	0.718	0.346	1.291	...	0.812	0.219	-0.71	Class 0
12	0.059	-3.35	-0.52	-0.49	-0.44	-0.44	-0.34	-0.20	-0.71	-0.42	...	1.014	1.740	1.404	Class 0
13	0.059	0.298	1.232	1.331	1.309	1.454	1.755	1.910	-0.91	-0.92	...	0.398	-1.05	-0.71	Class 1

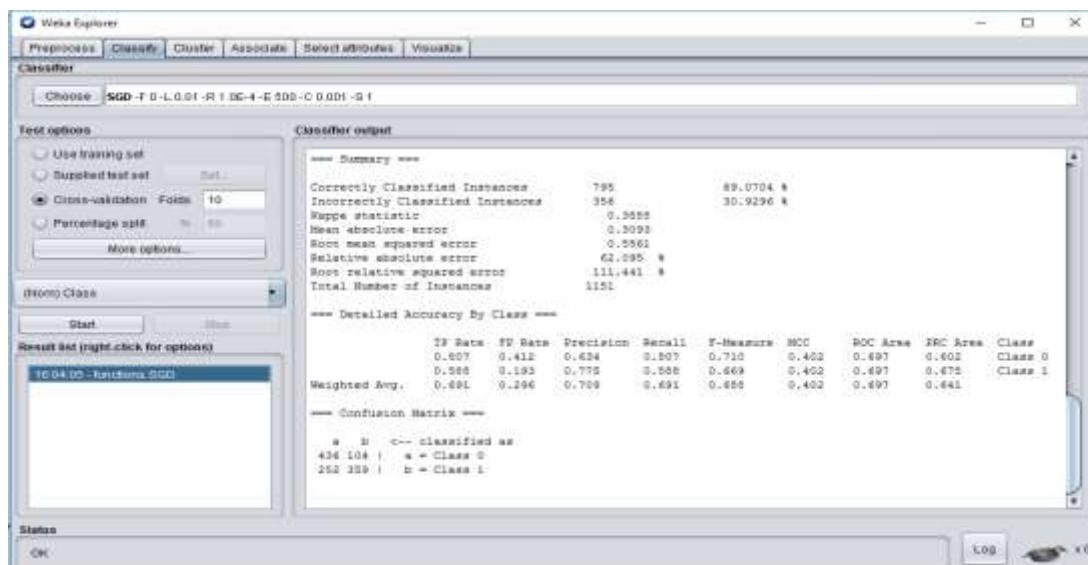


No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	...	X17	X18	X19	Class
14	0.059	0.298	0.373	0.252	0.169	-0.01	-0.08	-0.20	-0.70	-0.75	...	-0.77	1.663	1.404	Class 1
15	0.059	0.298	2.168	2.326	2.493	2.685	2.934	3.698	-0.94	-0.97	...	0.643	0.895	-0.71	Class 0
16	0.059	0.298	-0.71	-0.78	-0.83	-0.81	-0.80	-0.80	0.017	0.464	...	1.648	-1.03	1.404	Class 0
17	0.059	0.298	2.598	2.409	2.010	1.596	0.884	0.718	1.008	2.198	...	0.044	1.438	1.404	Class 1
18	0.059	0.298	-0.52	-0.49	-0.48	-0.44	-0.34	-0.14	-0.80	-0.60	...	1.026	-0.00	-0.71	Class 0
19	0.059	0.298	0.998	1.123	1.221	1.217	1.345	1.248	-0.75	-0.78	...	0.996	-0.54	-0.71	Class 0
20	0.059	-3.35	0.295	0.169	0.169	-0.01	-0.29	-0.40	0.879	-0.08	...	0.615	-0.19	1.404	Class 1
...
1151	0.059	0.298	-1.26	-1.24	-1.23	-1.19	-1.11	-1.06	0.091	-0.74	...	1.175	-1.08	-0.71	Class 0

3.2 Hasil Klasifikasi dengan Stochastic Gradient Descent

Setelah proses normalisasi data selesai dilakukan, maka selanjutnya dilakukan proses pengujian klasifikasi. Tahapan klasifikasi yang pertama dilakukan yaitu dengan proses perhitungan klasifikasi pada *Diabetic Retinopathy Debrecen Dataset* dengan algoritma *Stochastic Gradient Descent* (SGD). Kemudian untuk pengujian dengan *Stochastic Gradient Descent* menggunakan teknik evaluasi *Cross Validation* dengan pengujian *fold* dari nilai 1 sampai dengan 10 dan masing-masing diujikan untuk memperoleh jumlah data yang benar (*Correct*) dan data yang salah (*Incorrect*) dari hasil klasifikasi untuk menghitung nilai akurasi.

Dalam mempersingkat waktu proses pengujian metode klasifikasi yang diusulkan, pada penelitian ini penulis menggunakan bantuan *Waikato Environment for Knowledge Analysis* (WEKA). Adapun contoh dari tampilan hasil pengujian dengan *Stochastic Gradient Descent* (SGD) menggunakan WEKA pada Gambar 3 dan hasil klasifikasi yang diperoleh dengan *Stochastic Gradient Descent* (SGD) yaitu pada Tabel 5 berikut.



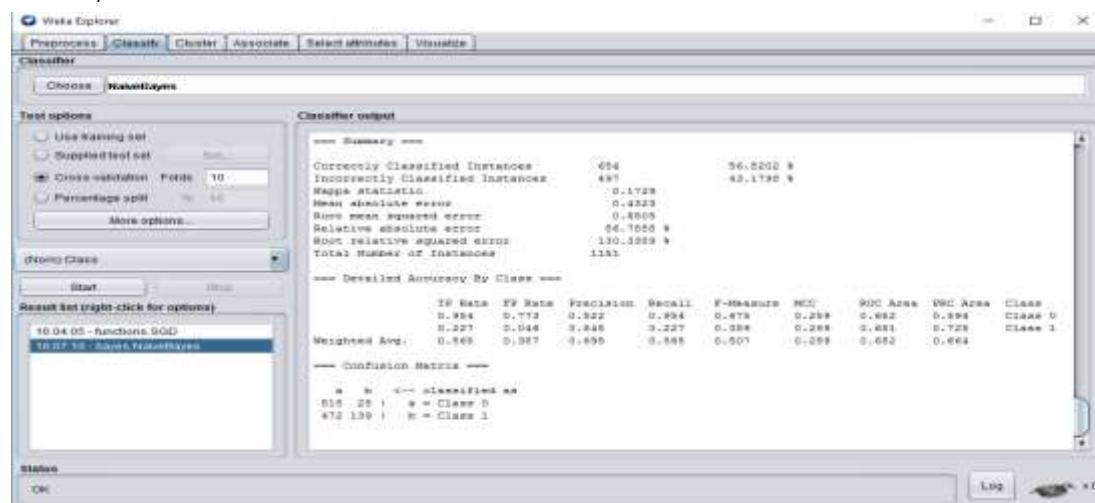
Gambar 3. Pengujian *Stochastic Gradient Descent* (SGD) Menggunakan Nilai *Fold-10*

Tabel 5. Hasil Pengujian Klasifikasi dengan *Stochastic Gradient Descent*

Fold Ke-	Correct	Incorrect	Akurasi Prediksi (%)
1	831	320	72.20
2	800	351	69.50
3	800	351	69.50
4	808	343	70.20
5	801	350	69.59
6	811	340	70.46
7	810	341	70.37
8	804	347	69.85
9	816	335	70.89
10	795	356	69.07
Rata-Rata			70.16

3.3 Hasil Klasifikasi dengan Naïve Bayes

Kemudian dilakukan proses perhitungan klasifikasi pada *Diabetic Retinopathy Debrecen Dataset* dengan metode *Naïve Bayes*. Pengujian dengan *Naïve Bayes* berdasarkan nilai validasi dari *Cross Validation* dari pengujian nilai *Fold 1* sampai dengan 10. Adapun contoh dari tampilan hasil pengujian dengan *Naïve Bayes* menggunakan WEKA pada Gambar 4 dan hasil klasifikasi yang diperoleh dengan *Naïve Bayes* yaitu pada Tabel 6 berikut.



Gambar 4. Pengujian *Naïve Bayes* dengan Menggunakan Nilai *Fold-10*

Tabel 6. Hasil Pengujian Klasifikasi dengan *Naïve Bayes*

Fold Ke-	Correct	Incorrect	Akurasi Prediksi (%)
1	652	499	56.64
2	652	499	56.64
3	654	497	56.82
4	656	495	56.99
5	651	500	56.55
6	652	499	56.64
7	655	496	56.91
8	652	499	56.64
9	653	498	56.73
10	654	497	56.82
Rata-Rata			56.74

3.4 Hasil Perbandingan Stochastic Gradient Descent dan *Naïve Bayes*

Berdasarkan hasil pengujian dan analisis yang bertujuan untuk mengetahui diantara model algoritma *Stochastic Gradient Descent* (SGD) dan *Naïve Bayes* yang memiliki akurasi paling tinggi untuk memprediksi penyakit *Diabetic Retinopathy*. Berdasarkan hasil pengukuran tingkat akurasi kedua algoritma tersebut, diketahui bahwa nilai akurasi *Stochastic Gradient Descent* (SGD) adalah 70.16%, sedangkan nilai akurasi *Naïve Bayes* 56.74% dan. Adapun komparasi hasil klasifikasi pada Tabel 7 berikut.

Tabel 7. Perbandingan Hasil Klasifikasi

Algoritma	Rata-Rata Akurasi Prediksi (%)
<i>Stochastic Gradient Descent</i>	70.16
<i>Naïve Bayes</i>	56.74

Hasil perbandingan tingkat akurasi tersebut menunjukkan bahwa dalam melakukasi klasifikasi penyakit *Diabetic Retinopathy*, nilai tingkat akurasi algoritma *Stochastic Gradient Descent* (SGD) lebih tinggi dibanding nilai tingkat akurasi *Naïve Bayes*. Adapun hal mendasari mengapa akurasi dari algoritma *Stochastic Gradient Descent* (SGD) memperoleh hasil akurasi yang lebih baik dibandingkan dengan hasil akurasi algoritma *Naïve Bayes* yaitu dikarenakan pada algoritma SGD dengan proses perhitungan yang efisien pada proses mencari nilai koefisien dalam meminimalkan *loss (error) function* pada data yang diujikan sehingga kesalahan (*error*) pada saat pengujian dapat diminimalisir dan akurasi yang diperoleh menjadi lebih maksimal. Berbeda dengan algoritma *Naïve Bayes* yang hanya dapat menghitung probabilitas dari data yang diujikan dengan data yang tersimpan sebelumnya tanpa adanya proses untuk meminimalkan kesalahan (*error*) dari proses perhitungan klasifikasinya.

4. KESIMPULAN

Berdasarkan hasil penelitian didapatkan kesimpulan dari pengujian yang telah dilakukan pada penelitian ini. Diketahui bahwa dari hasil pengujian untuk memprediksi akurasi pada klasifikasi penyakit *Diabetic Retinopathy* menggunakan algoritma *Stochastic Gradient Descent* (SGD) dan *Naïve Bayes*. Kedua algoritma tersebut menghasilkan nilai akurasi prediksi yang berbeda. Pada algoritma *Stochastic Gradient Descent* (SGD)



memperoleh rata-rata akurasi sebesar 70.16 % berdasarkan pengujian *Cross Validation* sebanyak jumlah 10 jumlah *Fold* pengujian. Sedangkan pada algoritma *Naïve Bayes* memperoleh rata-rata akurasi sebesar 56.74 % berdasarkan pengujian *Cross Validation* sebanyak jumlah 10 jumlah *Fold* pengujian. Maka selisih akurasi dari kedua algoritma tersebut yaitu sebesar 13.42 %. Dari hasil perbandingan tingkat akurasi tersebut menunjukkan bahwa dalam melakukasi klasifikasi penyakit *Diabetic Retinopathy*, algoritma *Stochastic Gradient Descent* (SGD) lebih unggul jika dibandingkan dengan *Naïve Bayes*.

REFERENCES

- [1] M. E. Al Rivan, S. Steven, and W. Tanzil, "Optimasi Fuzzy C-Means dan K-Means Menggunakan Algoritma Genetika untuk Pengklasteran Dataset Diabetic Retinopathy," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 5, p. 993, 2020, doi: 10.25126/jtiik.2020711872.
- [2] P. Subarkah, "Penerapan Algoritme Klasifikasi Classification And Regression Trees (Cart) Untuk Diagnosis Penyakit Diabetes Retinopathy," *MATRIX : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 294–301, 2020, doi: 10.30812/matrik.v19i2.676.
- [3] R. Tyasnurita and A. Y. M. Pamungkas, "Deteksi Diabetik Retinopati menggunakan Regresi Logistik," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 130–135, 2020, doi: 10.33096/ilkom.v12i2.578.130-135.
- [4] R. T. Prasetyo, "Genetic Algorithm to Optimize k-Nearest Neighbor Parameter for Benchmarked Medical Datasets Classification," *Jurnal Online Informatika*, vol. 5, no. 2, p. 153, 2020, doi: 10.15575/join.v5i2.656.
- [5] M. Bahrami and H. Sajedi, "Prediction of diabetic retinopathy based on a committee of random forests," *International Journal of Intelligent Machines and Robotics*, vol. 1, no. 2, p. 133, 2018, doi: 10.1504/ijimr.2018.10016284.
- [6] L. Heryawan, "Deteksi Dini Retinopati Diabetik dengan Pengolahan Citra Berbasis Morfologi Matematika," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 11, no. 2, p. 209, 2017, doi: 10.22146/ijccs.24761.
- [7] P. Subarkah, M. M. Abdallah, and S. O. N. Hidayah, "Komparasi Akurasi Algoritme CART Dan Neural Network Untuk Diagnosis Penyakit Diabetes Retinopathy," *CogITO Smart Journal*, vol. 7, no. 1, p. 121, 2021, doi: 10.31154/cogito.v7i1.304.121-134.
- [8] W. B. Zulfikar and N. Lukman, "Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata," *Jurnal Online Informatika*, vol. 1, no. 2, pp. 82–86, 2016, doi: 10.15575/join.v1i2.33.
- [9] R. Umar, I. Riadi, and P. Purwono, "Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial dengan Metode Support Vector Machines," *Cybernetics*, vol. 4, no. 01, p. 32, 2020, doi: 10.29406/cbn.v4i01.2042.
- [10] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," vol. 5, pp. 1580–1584, 2021, doi: 10.30865/mib.v5i4.3347.
- [11] R. A. Wijayanti, M. T. Furqon, and S. Adinugroho, "Penerapan Algoritme Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIK) Universitas Brawijaya*, vol. 2, no. 10, pp. 3500–3507, 2018, [Online]. Available: <http://j-ptik.ub.ac.id/index.php/j-ptik/article/download/2647/991>.
- [12] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015, doi: 10.1016/j.patcog.2015.03.009.
- [13] S. Widodo, "DETEKSI COVID-19 PADA CITRA CT-SCAN MENGGUNAKAN ALEXNET DAN STOCHASTIC GRADIENT DESCENT DENGAN," *Prosiding Seminar Informatika Kesehatan Nasional (SIKesNas)*, vol. 2, pp. 241–251, 2019.
- [14] S. G. Descent, "DETEKSI SMS SPAM BERBAHASA INDONESIA MENGGUNAKAN TF-IDF DAN STOCHASTIC GRADIENT DESCENT CLASSIFIER (Indonesian SMS Spam Detection using TF-IDF and Stochastic Gradient Descent)," vol. 3, no. 2, pp. 200–207, 2021.
- [15] V. Sari, F. Firdausi, and Y. Azhar, "Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 4, no. 2, pp. 1–9, 2020, doi: 10.29408/edumatic.v4i2.2202.
- [16] A. Budianita and F. I. Pratama, "Penerapan Algoritma Klasifikasi Dengan Fitur Seleksi Weight By Information Gain Pada Pemodelan Prediksi Kelulusan Mahasiswa," *Infotekmesin*, vol. 11, no. 2, pp. 80–86, 2020, doi: 10.35970/infotekmesin.v11i2.255.
- [17] H. Göker, H. I. Bülbül, and E. Irmak, "The estimation of students' academic success by data mining methods," *Proceedings - 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013*, vol. 2, pp. 535–539, 2013, doi: 10.1109/ICMLA.2013.173.
- [18] R. Satria, "Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naive Bayes," *IlmuKomputer.com Journal of Software Engineering*, vol. 1, no. 2, pp. 101–108, 2015.
- [19] A. Mukminin and D. Riana, "Komparasi Algoritma C4 . 5 , Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah," *Jurnal Informatika*, vol. 4, no. 1, pp. 21–31, 2017, [Online]. Available: <https://ejurnal.bsi.ac.id/ejurnal/index.php/ji/article/view/1002>.
- [20] M. A. Maricar and Dian Pramana, "Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali," *Jurnal Sistem dan Informatika (JSI)*, vol. 14, no. 1, pp. 16–22, 2019, doi: 10.30864/jsi.v14i1.233.