KOMPARASI TINGKAT AKURASI *INFORMATION GAIN* DAN *GAIN RATIO* PADA METODE *K-NEAREST NEIGHBOR*

Fajriyan Nur, Moh. Ahsan, Wahyudi Harianto
Program Studi Teknik Informatika S1, Fakultas Sains dan Teknologi
Universitas PGRI Kanjuruhan Malang, Jl.S. Supriadi No.48 Malang, Indonesia
fajriyan20@gmail.com, ahsan@unikama.ac.id, wahyou@unikama.ac.id

ABSTRAK

Seleksi fitur merupakan salah satu metode yang digunakan untuk mengurangi dimensi antar atribut sehingga menghasilkan atribut yang relevan, metode ini dianggap dapat mengurangi anomali pada atribut dalam dataset. Metode *K-Nearest Neighbor* diangap metode klasifikasi dengan akurasi yang rendah, terlebih lagi apabila dibandingkan dengan metode klasifikasi lain. Solusi dari permasalahan kecilnya nilai klasifikasi tersebut menggunakan seleksi fitur, teknik ini digunakan sebagai solusi untuk mengurangi kompleksitas dari atribut. Pada penelitian ini digunakan teknik seleksi fitur dengan *information gain* dan pengembangannya yaitu *gain ratio*. Teknik *gain ratio* dianggap lebih baik, hal itu terlihat pada nilai pembobotan *gain ratio* yang lebih baik dengan contoh perolehan hasil pada atribut "Ba" mendapatkan nilai 0,7204 pada *gain ratio,* sedangkan pada *information gain* hanya mendapatkan nilai 0,41235 sehingga didapatkan selisih 0,3081. Oleh karena itu dapat disimpulkan bahwa *gain ratio* mendapat hasil lebih tinggi dibanding pendahulunya yakni *information gain*. Berdasarkan pengujian menggunakan dataset *glass* terlihat *K-Nearest Neighbor* menggunakan *information gain* dan *gain ratio* memberikan hasil nilai akurasi yang lebih baik dengan perolehan nilai rata-rata 70,06% dibandingkan nilai akurasi ketika menggunakan metode *K-Nearest Neighbor* konvensional yang hanya mendapatkan nilai 66,24% selisih 8,75%, dengan perubahan metode pengujian yang dilakukan berupa variasi nilai *k*= 3, 5,7 dan *splitdata*.

Kata kunci : *K Nearest Neighbor, Seleksi Fitur, Information Gain, Gain Ratio.*

1. PENDAHULUAN

Beberapa penelitian tentang data mining menyatakan bahwa *K-Nearest Neighbor* dianggap memiliki akurasi yang tergolong rendah, terlebih lagi apabila dibandingkan secara langsung dengan metode klasifikasi lainnya. Farid Naufal pada penelitiannya membandingkan kinerja antara metode *K-Nearest Neighbor* dengan *Support Vector Machine*. Hasil dari penelitian ini didapati nilai rata-rata dari akurasi yang dihasilkan *K-Nearest Neighbor* 0.754% sedangkan nilai rata rata yang dihasilkan oleh *Support Vector Machine* sebesar 0.857% [1].

Menurut Syaliman rendahnya akurasi *K-Nearest Neighbor* karena setiap karakteristik atribut dalam metode ini dianggap memiliki pengaruh yang sama terhadap penentuan jarak antar tetangga. Sementara beberapa atribut yang kurang relevan menyebabkan kesalahan dalam pengenalan pola yang digunakan sebagai penentuan kelas untuk data baru [2].

Solusi dari permasalahan kedua metode dengan menggunakan seleksi fitur. Andyana menyatakan bahwa teknik ini digunakan sebagai solusi untuk memperkecil/ mengurangi kompleksitas atribut yang akan dikelola pada saat klasifikasi dengan *K-Nearest Neighbor* [3], sedangkan Mutmainnah mengatakan seleksi fitur dapat membantu mengenali atribut yang kurang relevan dan jumlahnya banyak yang dapat mempengaruhi hasil klasifikasi [4].

Chen & Hao menyarankan untuk menggunakan seleksi fitur *information gain* karena dinilai lebih efektif daripada menggunakan *K-Nearest Neighbor* konvensional [5], sedangkan pada penelitian Duneja & Puyalnithi menggunakan *gain ratio* karena dinilai lebih intuitif dan cenderung lebih mudah untuk dipahami [6]. Praveena Priyadarsini et al., mengatakan pada penelitiannya bahwa *gain ratio* merupakan modifikasi perkembangan dari *information gain* yang mengurangi biasnya [7]. Oleh karena itu pada penelitian ini akan dilakukan perbandingan antara kedua metode seleksi fitur antara *information gain* dan *gain ratio* pada metode *K-Nearest Neighbor* dengan harapan mengetahui kecocokan seleksi fitur pada klasifikasi *K-Nearest Neighbor* dan membandingkan seberapa jauh perbedaan rata-rata nilai akurasi antara seleksi fitur *information gain* dan Perkembangannya *gain ratio*.

2. TINJAUAN PUSTAKA

2.1 *K-Nearest Neighbor*

Metode *K-Nearest Neighbor* adalah salah satu metode yang paling banyak digunakan untuk pengenalan pola, pengklasifikasian, dan lain-lain. Hal ini dikarenakan beberapa faktor diantaranya Metode *K-Nearest Neighbor* dianggap cukup atraktif, mudah diterapkan dan diaplikasikan diberbagai *platform* aplikasi.

Metode *K-Nearest Neighbor* juga merupakan sebuah metode *supervised* yang berarti pada metode ini membutuhkan data latih untuk melakukan klasifikasikan objek yang jaraknya paling dekat. Prinsip kerja dari metode *K-Nearest Neighbor* adalah mencari jarak terdekat antara data yang akan di evaluasi dengan *k* tetangga (*neighbor*) dalam data pelatihan [8].

2.2 Seleksi Fitur

Seleksi fitur adalah salah satu metode yang biasa digunakan untuk mempersempit atau mengurangi dimensi antar atribut sehingga menghasilkan beberapa

atribut yang dianggap relevan untuk dilakukan proses klasifikasi. Jumlah atribut yang terlalu banyak dan jika jumlah atribut itu tidak relefan maka akan mempengaruhi dalam pengenalan pola ketika tahap *processing* dengan metode yang digunakan.

2.3 *Information Gain*

Information gain merupakan salah satu seleksi fitur yang populer, seleksi fitur ini biasanya digunakan untuk meranking atribut mana yang dianggap paling berpengaruh terhadap kelasnya. *information gain* digunakan untuk membantu menghasilkan atribut yang relevan untuk digunakan terhadap kelas target, karena setiap atribut memiliki nilai dan dapat dipilih sesuai dengan nilai yang terbaik.

2.4 *Gain Ratio*

Gain ratio merupakan modifikasi perkembangan dari seleksi fitur *information gain* dengan mengurangi daya biasnya. Pada seleksi fitur *gain ratio* memperbaiki dengan mengambil informasi intrinsik dari atribut [7]. *Gain ratio* juga dapat memperbaiki data yang tidak stabil, oleh karena itu cenderung cocok pada data numerik dua kelas, sederhana sehingga komputasi lebih cepat [9].

2.5 *Confusion Matrix*

Confusion matrix adalah sebuah kombinasi berbeda dari nilai sebuah prediksi dan nilai yang sebenarnya. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu *true positif, true negatif, false positif,* dan *false negatif*.

Pengujian suatu hasil dari klasifikasi pada metode yang digunakan diperlukan suatu metode perhitungan untuk evaluasi kinerja. Tahap yang di perlukan untuk evaluasi kinerja yaitu dengan melakukan perhitungan nilai *accuracy*, *precision*, *recall* dan *f1 score*.

3. METODE PENELITIAN

Metodologi penelitian merupakan penjabaran akan hal yang akan digunakan dalam penelitian. Metodologi penelitian digunakan sebagai pedoman dalam melakukan tahapan-tahapan dalam melakukan penelitian, berikut beberapa tahapan penelitian.

3.1. Teknik Pengumpulan Data

Data yang digunakan adalah data sekunder. Data sekunder merupakan data yang dalam penelitiannya diperoleh secara tidak langsung melainkan diperoleh dari pihak ketiga/ diunggah oleh pihak lain. Data yang digunakan adalah data *glass* yang berasal dari UCI *machine learning repository*.

Data *glass* merupakan data yang berasal dari "*Central Research Establishment German*". Pada data ini mengidentifikasikan kaca dari kandungan bahan pembentuknya, data glass berjumlah sebanyak 214 dengan jumlah atribut sebanyak 9 atribut (*RI, Na, Mg, Al, Si, K, Ca, Ba, Fe*), 1 atribut (*Type*) sebagai *label* dan 1 atribut sebagai *id*.

3.2. Pembagian Data

Data yang diperoleh dari teknik pengumpulan data akan dibagi menjadi data *training* dan data *testing* menggunakan metode *split validation* dengan rasio perbandingan 90:10 dan 80:20. Berikut adalah penjelasan kedua pembagian:

a) Data *training* merupakan data yang diperoleh 90% dan 80% dari dataset yang diacak. Data *training* nantinya akan digunakan sebagai pengenalan pola dengan menggunakan metode *K-Nearest Neighbor*.
b) Data *testing* merupakan data yang diperoleh 10% dan 20% dari dataset yang diacak. Nantinya data *testing* akan digunakan sebagai data uji dari metode *K-Nearest Neighbor* untuk mendapatkan nilai *confusion matrix*.

3.3. Implementasi Seleksi Fitur *Information Gain* dan *Gain Ratio*

Implementasi seleksi fitur merupakan tahapan awal sebagai *pre-processing* yang digunakan sebagai pembobotan pada tiap atribut. Pada dataset *glass* setiap atribut akan dipilih sesuai dengan nilai *gain* yang didapat dari penghitungan *entropy*. Setelah nilai *entropy* setiap kelas didapatkan kemudian dilakukan penghitungan *information gain* dan *gain ratio* untuk mengetahui atribut mana yang paling berpengaruh. Implementasi seleksi fitur dilakukan didalam aplikasi *rapidminer*.

3.4. Implementasi Metode *K-Nearest Neighbor*

Implementasi metode penelitian menggunakan metode *K-Nearest Neighbor* yang digunakan untuk klasifikasi. Tahapannya dimulai dari *split* data *training* dan *testing*. Pada penelitian ini nilai k yang digunakan adalah 3, 5 dan 7. Data *training* yang klasifikasi dengan menggunakan metode *K-Nearest Neighbor* mengahasilkan pola yang akan diujikan dengan data testing kemudian didapatkan nilai *confusion matrix*. Implementasi metode dilakukan didalam aplikasi *rapidminer*.

3.5. Pengujian Komparasi Seleksi Fitur pada Metode *K-Nearest Neighbor*

Pengujian komparasi yang dilakukan dalam rangka mendapat hasil dari klasifikasi nantinya akan dianalisis dengan mengukur nilai akurasi yang didapatkan dari klasifikasi dengan metode *K-Nearest Neighbor*, kemudian pada percobaan kedua dataset akan melewati seleksi fitur *information gain* dan *gain ratio* yang nantinya akan dipakai sebagai komparasi.

Pengujian juga dilakukan dengan skenario merubah nilai *k* pada metode *K-Nearest Neighbor* untuk melihat seberapa jauh perbedaan nilai akurasi dengan pilihan nilai *k*=3, 5 dan 7.

Analisis perbandingan dilakukan berdasarkan dari nilai *accuracy*, *recall*, *precision* dan *f1 score* yang didapat dari nilai *confusion matrix*. Berikut persamaan yang digunakan untuk menentukan hasil.

3.5.1 *Accuracy*

Nilai *accuracy* merupakan sebuah rasio keakuratan prediksi dengan nilai benar (positif dan negatif) dalam sebuah keseluruhan dataset. Penentuan nilai *accuracy* dapat dilakukan seperti pada persamaan (1) berikut.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}\ x\ 100\% \quad (1)$$

3.5.2 *Recall*

Nilai *recall* merupakan rasio prediksi dengan nilai benar positif dibandingkan dengan seluruh data yang benar positif. Penentuan nilai *recall* dapat dilakukan seperti pada persamaan (2) berikut.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

3.5.3 *Precision*

Nilai *precision* merupakan rasio prediksi dengan nilai benar positif dibandingkan dengan seluruh hasil yang diprediksi positif. Penentuan nilai *precision* dapat dilakukan seperti pada persamaan (3) berikut.

$$Precison = \frac{TP}{TP+FP} \quad (3)$$

3.5.4 *F1 score*

Nilai *F1 score* merupakan perbandingan rata-rata antara nilai *precision* dan *recall* yang dibobotkan. Penentuan nilai *F1 score* dapat dilakukan seperti pada persamaan (4) berikut.

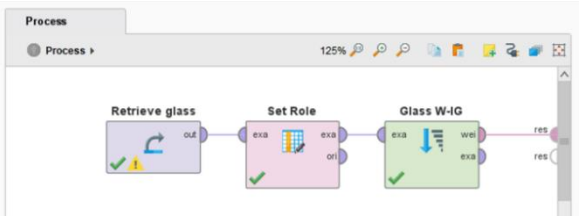
$$F1 = \frac{Precision\ x\ Recall}{Precision+\ Recall} \quad (4)$$

4. HASIL DAN PEMBAHASAN

Terdapat 3 hal yang ditekankan dalam pembahasan penelitian ini yakni hasil dari metode *K-Nearest Neighbor* konvensional dan metode *K-Nearest Neighbor* dengan menggunakan seleksi fitur *information gain* dan *gain ratio*.

3.1. Seleksi Fitur *Information Gain*

Pada proses ini dilakukan sebuah pembobotan pada dataset *glass* dengan menggunakan bantuan aplikasi *rapidminer*, pembobotan disini menggunakan *information gain* dengan skema penyusunan pada Gambar 1 berikut.



Gambar 1. Skema seleksi fitur *information gain*

Hasil dari perhitungan menggunakan *RapidMiner* dengan operator "*Weight By Information Gain*" menggunakan *parameter sort direction* "*Descending*" sesuai dengan skema Gambar 1 didapatkan hasil bahwa atribut Mg memiliki bobot tertinggi dengan nilai 0.5627, hal itu menunjukan bahwa atribut Mg merupakan atribut yang palling berpengaruh pada dataset *glass*. Sedangkan atribut Fe dengan bobot 0.0990 menjadi nilai terkecil pada dataset glass sehingga dapat dianggap sebagai atribut dengan pengaruh terkecil. Berikut hasil dari pembobotan dari dataset glass dengan menggunakan *Information Gain*.

Tabel 1. Hasil pembobotan *information gain*

No	*Attribute*	*Information Gain*
1	Mg	0.5627820064312257
2	Ba	0.4123500379467499
3	Al	0.3856994912509743
4	Na	0.3346396618715928
5	K	0.3224777255429660
6	RI	0.1820007402678387
7	Ca	0.1641575566001214
8	Si	0.1089260536339016
9	Fe	0.0990833389051539

Setelah didapatkan nilai *information gain* kemudian tahapan selanjutnya adalah pemilihan atribut dengan menggunakan sebagai berikut

$$\log_2 9\ (Jumlah\ atribut) = 3.169$$

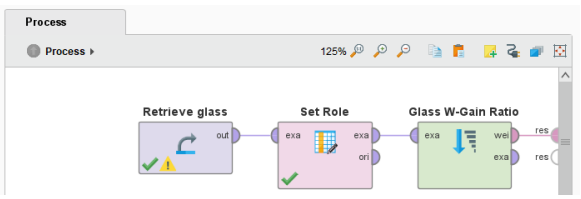
Dataset *glass* memiliki 9 atribut kemudian akan dipilih 3 atribut teratas yang dianggap sebagai atribut paling berpengaruh berdasarkan nilai pembobotan tertinggi untuk dilakukan proses klasifikasi.

Tabel 2. Atribut terpilih pada *information gain*

No	*Attribute*	*Information Gain*
1	Mg	0.5627820064312257
2	Ba	0.4123500379467499
3	Al	0.3856994912509743

3.2. Seleksi Fitur *Gain Ratio*

Pada proses ini dilakukan pembobotan pada dataset *glass* dengan menggunakan aplikasi *rapidminer*, pembobotan disini menggunakan *gain ratio*. Proses pembobotan dilakukan dengan menggunakan operator "*Weight By Information Gain Ratio*" dengan *parameter sort direction* "*Descending*" seperti pada Gambar 2.



Gambar 2. Skema seleksi fitur *gain ratio*

Hasil dari pembobotan didapatkan bahwa atribut "Ba" merupakan atribut yang paling berpengaruh pada dataset *glass* dikarenakan mendapat bobot paling tinggi dengan nilai 0.7204 sedangkan atribut "Ca"

mendapatkan bobot paling rendah dengan nilai "0.3580" oleh karena itu bisa dikatakan merupakan atribut dengan pengaruh terkecil dari dataset *glass*. Berikut hasil dari pembobotan dari dataset *glass* dengan menggunakan *gain ratio*.

Tabel 3. Hasil pembobotan *gain ratio*

No	*Attribute*	*Gain Ratio*
1	Ba	0.720427217670637
2	Mg	0.652700258459587
3	Al	0.554992092458355
4	K	0.520069023959777
5	RI	0.506651715660458
6	Na	0.506651715660458
7	Si	0.506651715660458
8	Fe	0.445998813468101
9	Ca	0.358018005695456

Setelah didapatkan hasil dari pembobotan pada tabel 2 menggunakan *gain ratio* kemudian dilakukan pemilihan atribut sesuai dengan persamaan sebagai berikut.

$$\log_2 9\ (Jumlah\ atribut) = 3.169$$

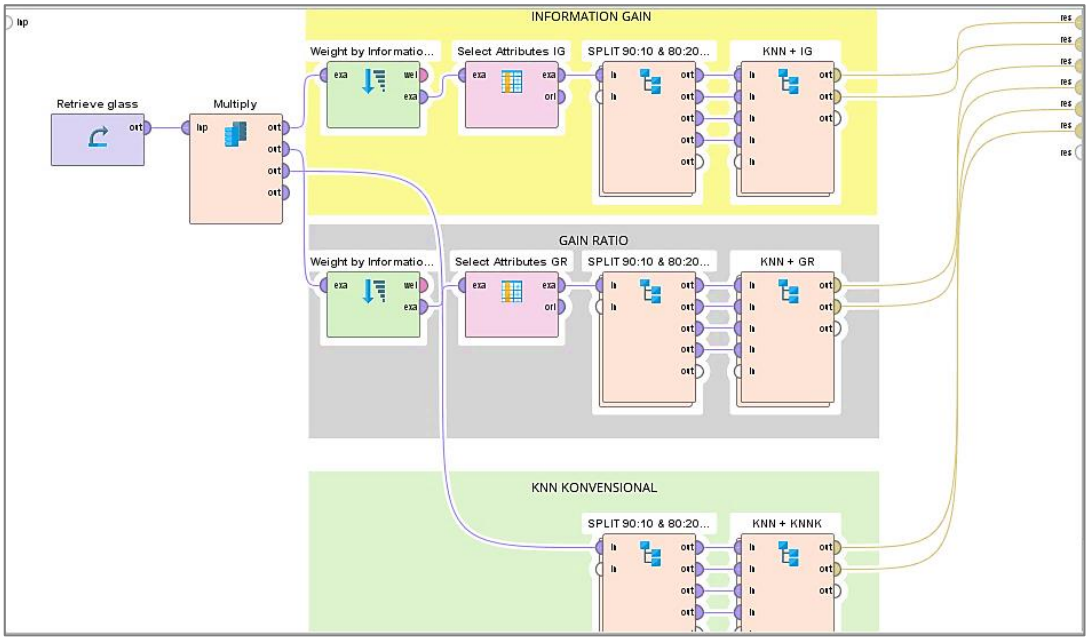
Hasil dari perhitungan diatas berjumlah 3.169 kemudian dibulatkan menjadi 3 yang berarti akan ada 3 atribut teratas yang dipilih sebagai atribut yang dianggap paling berpengaruh berdasarkan hasil nilai *gain ratio* pada tabel 2. Berikut atribut yang terpilih pada tabel 4.

Tabel 4. Atribut terpilih pada *gain ratio*

No	*Attribute*	*Gain Ratio*
1	Ba	0.720427217670637
2	Mg	0.652700258459587
3	Al	0.554992092458355

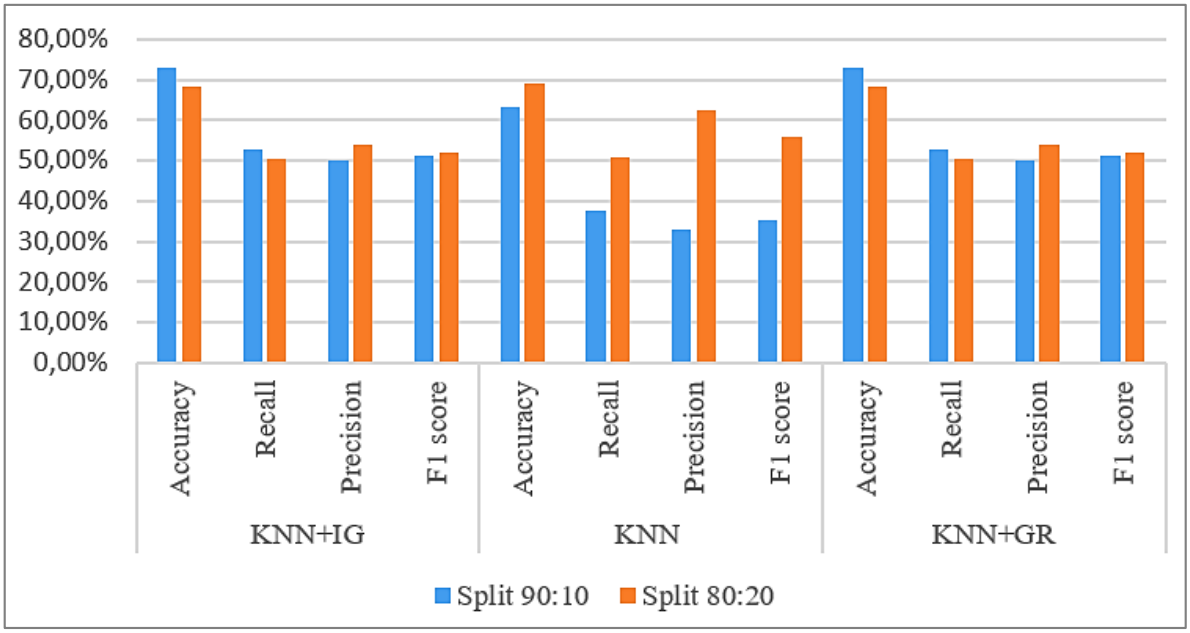
3.1. Hasil Klasifikasi *K-Nearest Neighbor + Information Gain + Gain Ratio*

Klasifikasi yang digunakan dalam penelitian ini menggunakan metode *K-Nearest Neighbor* dengan kombinasi "*split data*" (90:10) dan (80:20) menggunakan metode *sampling type* "*shuffled sampling*". Setelah dilakukan *split* kemudian proses dengan menggunakan tambahan seleksi fiur harus digunakan operator "*select atribut*" untuk melakukan *filter* atribut mana saja yang telah terpilih sesuai dengan hasil dari pembobotan seleksi fitur sebelumnya. Berikut skema klasifikasi pada Gambar 3.



Gambar 3. Skema klasifikasi

Berikut hasil setelah dilakukan proses klasifikasi menggunakan metode *K-Nearest Neighbor* dengan nilai *k*=3, 5, 7 menggunakan aplikasi *rapidminer*. Hasil dari klasifikasi dapat dilihat pada grafik yang ada pada Gambar 4 dibawah ini.



Gambar 4. Hasil klasifikasi

Dataset *glass* dengan *split data* (90:10) terlihat nilai *accuracy*, *recall*, *precision* dan *f1 score* mengalami peningkatan ketika menggunakan seleksi fitur. Nilai *accuracy* dari klasifikasi *K-Nearest Neighbor* mendapatkan rata-rata sebesar 73,01% dengan menggunakan *information gain* dan *gain ratio*, sedangkan *K-Nearest Neighbor* tanpa seleksi fitur mendapatkan nilai rata-rata sebesar 63,49% dengan perbedaan selisih rata-rata sebesar 9,52% pada dataset *glass*. Hasil menjukkan perbedaan pada split data 80:20 dengan hasil *accuracy* yang menurun sebesar 0,78% ketika menggunakan seleksi fitur.

Pada dataset *glass* menujukkan bahwa atribut yang paling berpengaruh pada seleksi fitur *information gain* adalah Mg (0.56278), Ba (0.41235) dan Al (0.38569) sedangkan pada seleksi fitur *Gain Ratio* mendapatkan hasil yang sama dengan urutan pemobotan yang berbeda yakni Ba (0.72042), Mg (0.65270) dan Al (0.55499), oleh karenanya pada tabel 5.2 hasil *information gain* dan *gain ratio* terlihat sama antara keduanya. Perbedaan antara kedua seleksi fitur terjadi pada atribut ke 4 dimana pada *information gain* atribut Na (0.33463) terpilih nomor 4 sedangkan pada *gain ratio* atribut yang terpilih nomor 4 adalah K (0.52006).

Hasil *accuracy*, *recall*, *precision* dan *f1 score* pada dataset *glass* sangat berpengaruh dengan perubahan nilai *k*, dapat dilihat dari 6 kali percobaan perubahan skema pengujian diketahui nilai *k*=3 merupakan nilai *k* dengan hasil paling tinggi dengan perolehan nilai 80,95% pada *splitdata* 90:10 sedangkan pada *splitdata* 80:20 mendapatkan nilai 72,09%, hasil tersebut lebih tinggi dari hasil dengan nilai *k* yang lainnya. Berdasarkan pengujian yang telah dilakukan dengan menggunakan dataset *glass* terlihat bahwa *K-Nearest Neighbor* menggunakan *information gain* dan *gain ratio* dapat memberikan sebuah hasil nilai akurasi yang lebih baik dengan perolehan nilai rata-rata 70,06% dibandingkan nilai akurasi ketika hanya menggunakan metode *K-Nearest Neighbor* konvensional yang hanya mendapatkan nilai 66,24%, dengan perubahan metode pengujian yang dilakukan berupa variasi nilai *k* dan *splitdata*.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil yang diperoleh dari pengujian yang telah dilakukan pada penelitian ini didapatkan kesimpulan bahwa ketika menggunakan seleksi fitur *information gain* dan *gain ratio* terbukti dapat meningkatkan nilai akurasi dibanding dengan menggunakan *K-Nearest Neighbor* konvensional. Berdasarkan dari pengujian menggunakan dataset *glass* dengan berbagai macam kondisi nilai *k* yang berbeda, sehingga dapat disimpulkan penggunaan seleksi fitur *gain ratio* dianggap lebih optimal dibanding dengan *information gain* dikarenakan dalam pembobotan seleksi fitur *gain ratio* dianggap lebih baik dengan contoh perolehan hasil pembobotan pada atribut "Ba" mendapatkan nilai 0,7204 pada *gain ratio*, sedangkan pada *information gain* hanya mendapatkan nilai 0,41235 sehingga didapatkan selisih 0,3081. Oleh karena itu dapat disimpulkan bahwa *gain ratio* mendapat hasil lebih tinggi dibanding pendahulunya yakni *information gain*. Namun hasil akurasi klasifikasi dari kedua seleksi fitur cenderung sama dikarenakan pembobotan dari ketiga dataset mendapatkan urutan yang sama dengan nilai bobot yang berbeda.

Saran untuk pengembangan penelitian selanjutnya yakni dengan melakukan perbandingan seleksi fitur dengan menggunakan metode klasifikasi lain sehingga dapat disimpulkan apakah metode

klasifiksi lain menghasilkan nilai yang lebih baik, dan melakukan penambahan operasi normalisasi dataset sebelum dilakukan pembobotan dengan menggunakan seleksi fitur yang digunakan untuk mendapatkan nilai hasil yang lebih baik.

DAFTAR PUSTAKA

[1] M. Farid Naufal, "Perbandingan, Analisis Svm, Algoritma Untuk, dan CNN," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, pp. 311–318, 2021, doi: 10.25126/jtiik.202184553.

[2] K. U. Syaliman, "Peningkatan Akurasi Pada Metode Klasifikasi K-Nearest Neighbor Menggunakan Local Mean Based Dan Distance Weight K-Nearest Neighbor," *Preeklamsia Berat*, pp. 44–85, 2018, [Online]. Available: repository.usu.ac.id/bitstream/123456789/30230/4/Chapter II.pdf.

[3] I. made B. Adnyana, "Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa," *J. Sist. Dan Inform.*, vol. 13, pp. 72–76, 2019.

[4] U. Mutmainnah, *Pengaruh Seleksi Fitur Information Gain pada K-Nearest Neighbor untuk Klasifikasi Tingkat Kelancaran Pembayaran Kredit Kendaraan*, vol. 3, no. 9. 2019.

[5] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Syst. Appl.*, vol. 80, pp. 340–355, 2017, doi: 10.1016/j.eswa.2017.02.044.

[6] A. Duneja and T. Puyalnithi, "Enhancing Classification Accuracy of K-Nearest Neighbours Algorithm Using Gain Ratio," *Int. Res. J. Eng. Technol.*, vol. 4, no. 9, pp. 1385–1388, 2017, [Online]. Available: https://irjet.net/archives/V4/i9/IRJET-V4I9260.pdf.

[7] Praveena Priyadarsini, V. M.L., and S. S., "Gain Ratio Based Feature Selection Method for Privacy Preservation," *ICTACT J. Soft Comput.*, vol. 01, no. 04, pp. 201–205, 2011, doi: 10.21917/ijsc.2011.0031.

[8] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.

[9] K. Kurniabudi, A. Harris, and A. E. Mintaria, "Komparasi Information Gain, Gain Ratio, CFs-Bestfirst dan CFs-PSO Search Terhadap Performa Deteksi Anomali," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 332, 2021, doi: 10.30865/mib.v5i1.2258.