

COMPARISON OF CLASSIFICATION ALGORITHM AND FEATURE SELECTION IN BITCOIN SENTIMENT ANALYSIS

Indri Tri Julianto^{*1}, Dede Kurniadi², Muhammad Rikza Nashrulloh³, Asri Mulyani⁴

^{1,2,3,4}Jurusan Ilmu Komputer, Institut Teknologi Garut, Indonesia

Email: ¹indritrijulianto@itg.ac.id, ²dede.kurniadi@itg.ac.id, ³rikza@itg.ac.id, ⁴asrimulyani@itg.ac.id

(Naskah masuk: 26 Mei 2022, Revisi : 02 Juni 2022, diterbitkan: 28 Juni 2022)

Abstract

Sentiment analysis is a process for extracting data in the form of textual, with the aim of obtaining information about the tendency to evaluate an object under study. Sentiments given by the general public can be used as a reference in making product decisions. Sentiment given can be in the form of positive, negative and neutral sentiments. One of the information technology products that has stolen enough attention in the last decade is Bitcoin. The purpose of this study is to compare several classification algorithms using Feature Selection. There are several classification algorithms that can be used for sentiment analysis, such as Deep Learning, Decision Tree, KNN, Naïve Bayes. Textual sentiment classification has constraints on datasets that have high dimensions. Feature Selection is a solution to reduce the dimensions of a dataset by reducing attributes that are less relevant. Feature Selection used is Information Gain and Chi Square. The method used to perform the comparison is by comparing the four classification algorithms to find the best algorithm, then comparing the Feature Selection to get the best between the two, then integrating the best classification algorithm and the best Feature Selection. The results showed that the best classification algorithm was Deep Learning with an accuracy value of 78.43% and a kappa of 0.625. The results of the comparison of Feature Selection, Information Gain get the best results with an average accuracy value of 63.79% and an average kappa of 0.382. The results of the integration of the best classification algorithm with the best Feature Selection obtained an accuracy value of 78.63% and a kappa of 0.626 where the value was included in the Fair Classification category.

Keywords: *algorithms, bitcoin, classification, data mining, feature selection, sentiment analysis.*

PERBANDINGAN ALGORITMA KLASIFIKASI DAN FEATURE SELECTION PADA ANALISIS SENTIMEN BITCOIN

Abstrak

Analisis sentimen merupakan proses untuk melakukan ekstraksi terhadap data yang berbentuk tekstual, dengan tujuan untuk mendapatkan informasi mengenai kecenderungan penilaian terhadap suatu objek yang diteliti. Sentimen yang diberikan oleh khalayak umum dapat dijadikan acuan dalam pengambilan keputusan suatu produk. Sentimen yang diberikan dapat berupa sentimen positif, negatif dan netral. Salah satu produk teknologi informasi yang cukup mencuri perhatian dalam satu dekade terakhir adalah *Bitcoin*. Tujuan penelitian ini adalah untuk membandingkan beberapa algoritma klasifikasi dengan menggunakan *Feature Selection*. Terdapat beberapa Algoritma klasifikasi yang dapat digunakan untuk analisis sentimen, seperti *Deep Learning*, *Decision Tree*, *K-Nearest Neighbor* dan *Naïve Bayes*. Klasifikasi sentimen tekstual memiliki kendala pada *dataset* yang memiliki dimensi yang tinggi. *Feature Selection* merupakan solusi untuk mengurangi dimensi dari suatu dataset dengan cara mengurangi atribut yang kurang relevan. *Feature Selection* yang digunakan adalah *Information Gain* dan *Chi Square*. Metode yang digunakan untuk melakukan perbandingan adalah dengan cara membandingkan keempat algoritma klasifikasi untuk mencari algoritma terbaik, selanjutnya perbandingan *Feature Selection* untuk mendapatkan yang terbaik diantara keduanya, kemudian dilakukan integrasi antara algoritma klasifikasi terbaik dan *Feature Selection* terbaik. Hasil penelitian menunjukkan untuk Algoritma klasifikasi terbaik adalah *Deep Learning* dengan nilai akurasi sebesar 78,43% dan kappa 0,625. Hasil perbandingan *Feature Selection*, *Information Gain* mendapat hasil paling baik dengan nilai rata-rata akurasi 63,79% dan rata-rata kappa 0,382. Hasil integrasi Algoritma klasifikasi terbaik dengan *Feature Selection* terbaik memperoleh nilai akurasi 78,63% dan kappa 0,626 dimana nilai tersebut termasuk kedalam kategori *Fair Classification*.

Kata kunci: *algoritma, analisis sentimen, bitcoin, data mining, feature selection, klasifikasi.*

1. PENDAHULUAN

Data Mining merupakan proses untuk menemukan keterkaitan baru, pola dan juga tren dengan istilah menambang sejumlah *repository* data dalam jumlah besar [1] dengan menggunakan bidang ilmu teknik matematika serta statistika, kemudian data mining juga sering disebut *Knowledge Discovery in Database* (KDD) ataupun *Pattern Recognition* [2], dimana *Data Mining* ini dapat digolongkan menjadi lima kelompok, pertama adalah *Estimation*, kedua *Forecasting*, ketiga *Classification*, keempat *Clustering* dan kelima adalah *Association* [3].

Bitcoin semakin populer dalam kurun waktu satu dekade terakhir, dimana *bitcoin* memiliki jaminan keamanan dalam transaksinya dengan teknik kriptografi yang diklaim sangat aman [4]. *Bitcoin* memberikan transparansi, kecepatan, keamanan dan kemudahan transaksi yang sekaligus menjadi jawaban atas kebutuhan transaksi keuangan pada zaman sekarang [5].

Analisis sentimen merupakan proses untuk melakukan ekstraksi terhadap data yang berbentuk tekstual, dengan tujuan untuk mendapatkan informasi mengenai kecenderungan penilaian terhadap suatu objek yang diteliti. Sentimen yang diberikan oleh khalayak umum dapat dijadikan acuan dalam pengambilan keputusan suatu produk [6]. Sentimen yang diberikan dapat berupa sentimen positif, negatif dan netral [7].

Permasalahan yang akan ditemukan ketika melakukan analisis sentimen adalah banyaknya atribut yang digunakan dalam *dataset* [8]. Pada dasarnya jumlah atribut yang banyak akan mempengaruhi akurasi menjadi rendah.

Feature Selection sering digunakan untuk pengurangan dimensi model [9], dan menjadikan pengklasifikasian lebih efektif dengan mengurangi jumlah data yang dianalisa, maupun mengidentifikasi fitur yang sesuai untuk dipertimbangkan dalam proses pembelajaran [8]. Cara kerjanya berdasar pada pengurangan ruang fitur yang besar, yaitu dengan cara mengeliminasi atribut yang kurang relevan serta dengan menggunakan penggunaan algoritma *feature selection* yang tepat sehingga dapat meningkatkan akurasi [10].

Terdapat beberapa penelitian yang menjadi rujukan untuk penelitian ini. Pertama adalah penelitian mengenai penggunaan *Feature Selection Information Gain* dan Algoritma *Naive Bayes* Untuk *Review Opini Konsumen* [11], dimana hasil penelitian menunjukkan nilai akurasi sebesar 78,4% dengan menggunakan metode *Naive Bayes*, untuk meningkatkan nilai akurasi dengan menggunakan seleksi fitur *Information Gain* diperoleh tingkat akurasi sebesar 81,2%. Kedua adalah penelitian mengenai Analisis media sosial *twitter* mengenai

Omnibus Law dengan menggunakan *Feature Selection Chi Square* dan algoritma *K-Nearest Neighbor* (K-NN) [12], dimana hasil penelitian menunjukkan hasil pemodelan menggunakan *K-Fold Cross Validation*, akurasi tertinggi diperoleh dengan skema penggunaan fitur 25% (seleksi fitur *Chi Square*), dan nilai $k = 5$ pada K-NN adalah 81,4%, kemudian setelah dilakukan evaluasi dengan melakukan sebuah pengujian model klarifikasi dengan K-NN dan *K Fold Cross Validation* serta fitur seleksi *Chi Square* hasil akurasi meningkat menjadi 83%. Ketiga adalah penelitian mengenai analisis sentimen berbahasa Indonesia dengan menggunakan *Feature Selection* dan Algoritma *Deep Belief Network* yang termasuk dalam metode *Deep Learning* [13], dimana hasil penelitian menunjukkan akurasi 93,31% yang lebih baik jika dibandingkan dengan *Naive Bayes* 79,10% dan *Support Vector Machines* 92,18%. Keempat adalah penelitian mengenai perbandingan Algoritma *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbor* dalam analisis sentimen transportasi KRL *Commueter Line* [14], dimana hasil penelitian menunjukkan nilai akurasi untuk *Decision Tree* 100%, *Naive Bayes* 80% dan K-NN 80%. Secara singkat penelitian-penelitian tersebut disajikan dalam bentuk tabel, sebagaimana tampak pada Tabel 1.

Tabel 1. Roadmap Penelitian

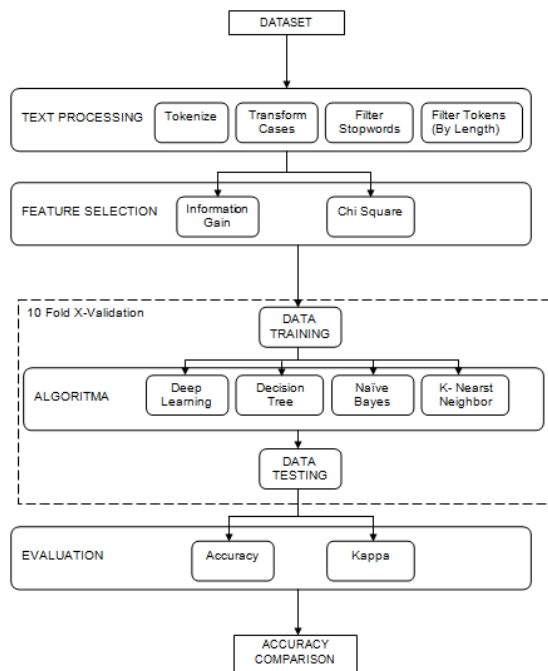
Penelitian	Dataset	Feature Selection	Algoritma	Tujuan Penelitian
1	<i>Review Opini Konsumen</i>	<i>Information Gain</i>	NB	Analisis Sentimen
2	<i>Omnibus Law</i>	<i>Chi Square</i>	K-NN	Analisis Sentimen
3	<i>Tweet Bahasa Indonesia KRL</i>	-	DBN, NB, SVM	Analisis Sentimen
4	<i>Commueter Line</i>	-	DT, NB, KNN	Analisis Sentimen

Keterangan: NB (*Naive Bayes*), K-NN (*K-Nearest Neighbor*), DBN (*Deep Belief Network*), SVM (*Support Vector Machines*).

Berbeda dengan penelitian sebelumnya, penelitian ini akan membandingkan empat Algoritma sekaligus yaitu *Deep Learning*, *Decision Tree*, *Naive Bayes*, *K-Nearest Neighbor*. Penambahan *Deep Learning* dalam penelitian dikarenakan memiliki keunggulan pada bagian efisiensi ekstraksi fitur secara otomatis serta kemampuannya dalam menangkap fitur sintaksis dan semantik teks tanpa persyaratan rekayasa fitur tingkat tinggi (*high-level feature engineering*) [15]. *Feature Selection* yang digunakan adalah *Information Gain* dan *Chi Square*

2. METODE PENELITIAN

Kerangka penelitian yang diusulkan untuk penelitian ini disajikan sebagaimana tampak pada Gambar 1.

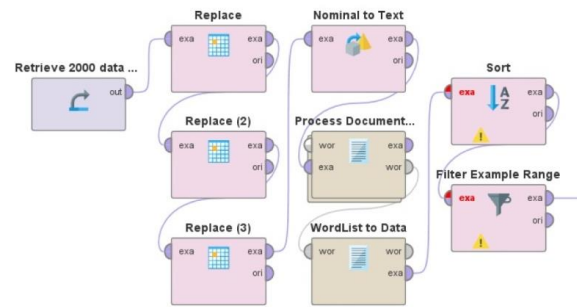


Gambar 1. Kerangka Penelitian

Tahapan pertama sebelum dilakukan komparasi adalah tahapan *text processing*. *Dataset* yang digunakan adalah data cuitan di *Twitter* dengan kata kunci *Bitcoin*, dimana data tersebut diperoleh dari situs web www.data.world.com. Data tersebut terdiri dari 2000 data cuitan yang telah memiliki label berupa positif, negatif dan netral untuk setiap cuitan di *Twitter*. *Text Processing* bertujuan untuk membuat data menjadi terstruktur [16]. Adapun tahapannya adalah sebagai berikut:

1. *Tokenize*, adalah merupakan proses untuk memisahkan kata, dimana potongan kata itu disebut dengan token atau *term* [17][18], [19];
2. *Transform Cases*, adalah proses mengubah secara otomatis semua huruf pada teks menjadi huruf kecil atau sebaliknya [17]. Pada penelitian ini semua huruf dirubah kedalam huruf kecil semua;
3. *Filter Stopword*, adalah proses menghilangkan kata yang sering muncul namun tidak berpengaruh terhadap ekstraksi sentimen. Kata yang termasuk seperti kata penunjuk waktu, kata tanya [20];
4. *Filter Token (By Length)*, adalah proses menghapus kata dengan jumlah huruf tertentu melalui dengan parameter *min-chars* 4 dan *max-chars* 25 untuk membatasi jumlah huruf pada kata minimal 4 dan maksimal 25 pada teks [21].

Rapidminer Studio digunakan untuk *Text Processing*. Adapun untuk proses tersebut disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 2.



Gambar 2. *Text Processing* Rapidminer

Gambar 2, menunjukkan bahwa pada *dataset* yang digunakan dilakukan tahapan mulai dari *Tokenize*, sampai dengan *Filter Token (By Length)* yang mana proses tersebut berada di dalam operator yang namanya *Process Document from Data*, karena operator tersebut merupakan operator sub-proses yang artinya didalam operator tersebut terdapat operator-operator yang lain.

Tahapan kedua adalah perbandingan *Feature Selection* yang digunakan yaitu *Information Gain* dan *Chi Square*. Disini akan dilihat nilai akurasi dan kappa yang lebih baik diantara keduanya dengan menggunakan dataset yang ada. Kemudian di dalam tahapan kedua ada proses lagi yang terhubung yaitu melakukan perbandingan mengenai Algoritma terbaik diantara keempat algoritma yang dipilih dengan menggunakan *10 fold X-Validation*.

Tahapan ketiga adalah Evaluasi dimana akan dilakukan integrasi dengan menggunakan Algoritma klasifikasi terbaik dan *Feature Selection* terbaik. Disini akan dilihat apakah terdapat peningkatan akurasi dan kappa atau tidak. Akurasi merupakan persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [22]. Akurasi memiliki tingkat diagnosa yaitu

Nilai Akurasi 0.90 – 1.00 = *Excellent classification*

Nilai Akurasi 0.80 – 0.90 = *Good classification*

Nilai Akurasi 0.70 – 0.80 = *Fair classification*

Nilai Akurasi 0.60 – 0.70 = *Poor classification*

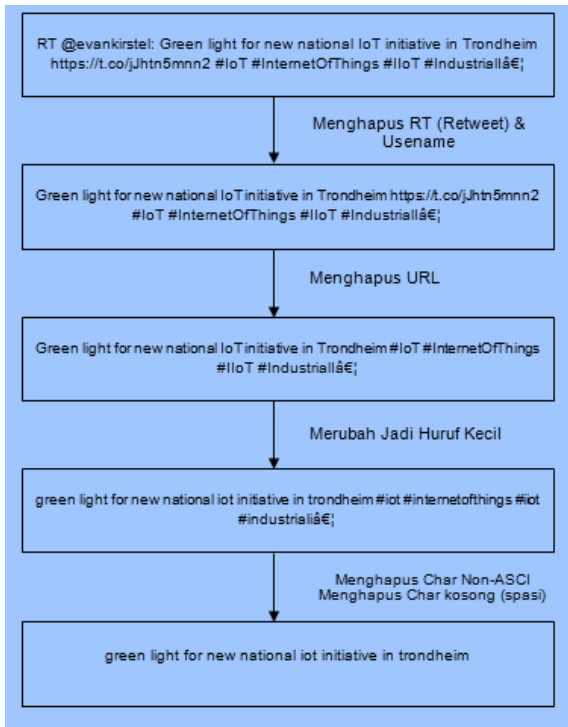
Nilai Akurasi 0.50 – 0.60 = *Failure*

Pada dasarnya kappa menunjukkan analisis diantara kelas-kelas yang berbeda, dimana semakin tinggi kappa *statistic*, maka akan dipertimbangkan sebagai performa yang mempunyai kriteria atau kinerja yang bagus [22]. Dikarenakan label dalam dataset ini berupa *Polynomial* bukan *Binomial* maka pengukuran yang digunakan adalah Kappa bukan AUC.

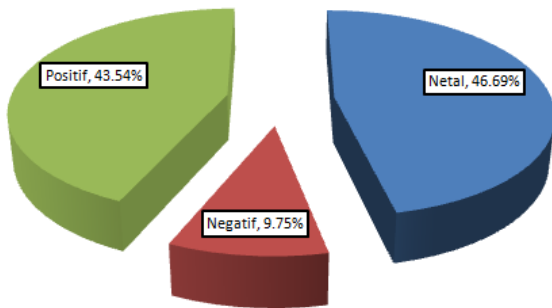
3. HASIL DAN PEMBAHASAN

Penelitian yang dilakukan menggunakan komputer dengan spesifikasi CPU Intel Pentium 1.86GHz, RAM 3GB, dan sistem operasi *Microsoft Windows 7 Professional* 32-bit. Aplikasi yang digunakan adalah RapidMiner 9.10.1.

Hasil dari tahapan pertama yang dilakukan, disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 3.



Gambar 3. Hasil Text Processing

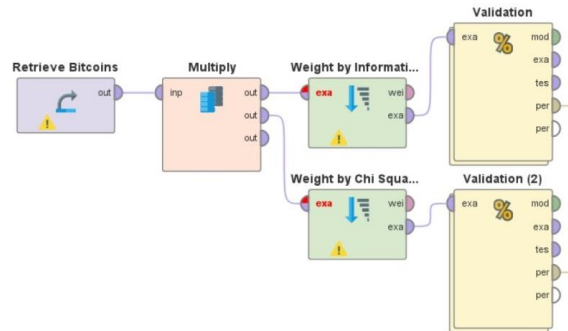


Gambar 4. Persentase Sentimen Terhadap Bitcoin

Tahapan ini berhasil menjadikan data yang digunakan menjadi data yang terstandarisasi. Terlihat pada bagan paling atas merupakan kalimat yang belum dibersihkan atau masih kalimat aslinya. Kemudian setelah dilakukan penghapusan tulisan *retweet*(RT), *username*, lalu menghapus URL, merubah semua huruf menjadi huruf kecil dan menghapus karakter Non-ASCII serta karakter kosong seperti spasi, maka pada bagan paling bawah dapat terlihat kalimat yang sudah bersih dan terstandarisasi. Pemberian sentimen yang dilakukan menunjukkan hasil nilai untuk cuitan Positif 43,54%, Negatif 9,75%, dan Netral sebesar 46,69%. Adapun hasil tersebut divisualisasikan dalam bentuk diagram, sebagaimana tampak pada Gambar 4.

Hasil tahapan kedua adalah membangun model proses untuk melakukan perbandingan keempat algoritma klasifikasi. Adapun model proses tersebut

disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 5.



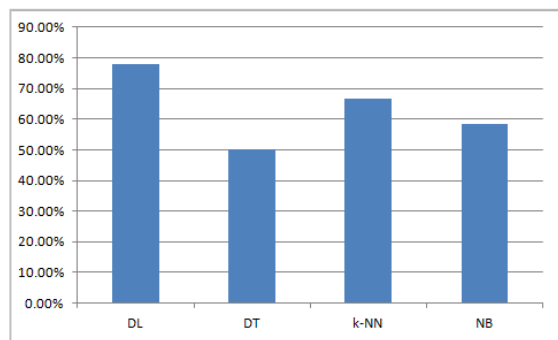
Gambar 5. Model Proses Perbandingan Algoritma Klasifikasi

Dataset diproses dengan menggunakan algoritma yang terpilih, kemudian dihubungkan dengan *Feature Selection Information Gain* dan *Chi Square* lalu dilakukan *10 fold x-validation* untuk mengetahui hasil nilai akurasi dan kappa. Nilai akurasi dan kappa dari setiap algoritma akan dibandingkan dan akan dicari mana yang memiliki persentase tertinggi.

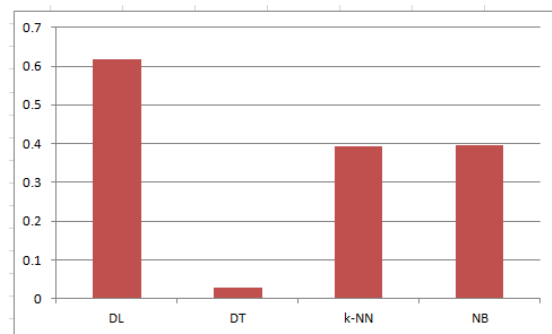
Hasil tahapan ketiga adalah evaluasi, dimana hasil dari model proses yang dibangun akan dibandingkan nilai akurasi dan kappa nya. Adapun nilai perbandingan Algoritma terbaik disajikan dalam bentuk tabel, sebagaimana tampak pada Tabel 1.

Tabel 1. Hasil Perbandingan Algoritma

	Perbandingan Algoritma			
	DL	DT	k-NN	NB
ACC	78.08%	50.10%	66.82%	58.31%
KAPPA	0.618	0.029	0.393	0.396



Gambar 6. Perbandingan Akurasi Algoritma Klasifikasi



Gambar 7. Perbandingan Kappa Algoritma Klasifikasi

Hasil perbandingan menunjukkan bahwa Algoritma *Deep Learning* memiliki nilai akurasi terbaik dengan persentase 78,08% dan kappa 0,618. Adapun visualisasi data perbandingan nilai akurasi dan kappa disajikan dalam bentuk grafik, sebagaimana tampak pada Gambar 6 dan Gambar 7.

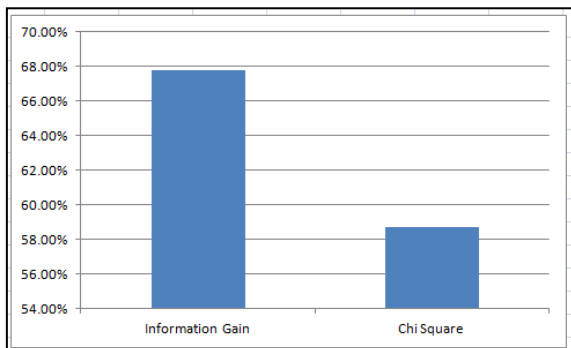
Algoritma *Deep Learning* memiliki nilai perbedaan yang cukup signifikan dibandingkan algoritma-algoritma yang lain. Hal ini mengkonfirmasi penelitian yang telah dilakukan oleh Nuryani [15], dimana pada penelitiannya dikatakan bahwa algoritma *Deep Learning* memiliki keunggulan mudah untuk diuji dan di klasifikasi.

Hasil perbandingan *Feature Selection* terbaik disajikan dalam bentuk tabel, sebagaimana tampak pada Tabel 2.

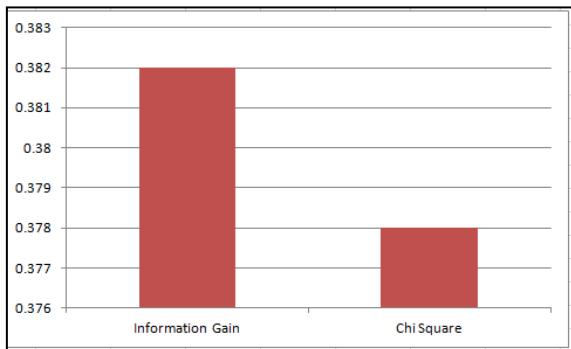
Tabel 2. Hasil Perbandingan *Feature Selection* Terbaik

	<i>Information Gain</i>		<i>Chi Square</i>	
	Akurasi	Kappa	Akurasi	Kappa
<i>Deep Learning</i>	78.63%	0.626	78.43%	0.625
<i>Decision Tree</i>	49.90%	0.087	49.95%	0.090
<i>k-NN</i>	68.32%	0.420	66.82%	0.393
<i>Naïve Bayes</i>	58.31%	0.396	58.71%	0.404
Rata-rata	63.79%	0.382	63.48%	0.378

Untuk visualisasi data mengenai nilai akurasi dan kappa untuk *Feature Selection* terbaik disajikan dalam bentuk grafik, sebagaimana tampak pada Gambar 8 dan Gambar 9.



Gambar 8. Perbandingan Akurasi *Feature Selection*



Gambar 9. Perbandingan Kappa *Feature Selection*

Tabel 2 merupakan hasil perbandingan *Feature Selection* terbaik dimana *Information Gain* mendapatkan nilai terbaik dengan rata-rata akurasi sebesar 63,79% dan rata-rata kappa sebesar 0,382.

Hasil integrasi antara Algoritma terbaik yaitu *Deep Learning* dengan *Feature Selection* terbaik yaitu *Information Gain*, disajikan dalam bentuk tabel, sebagaimana tampak pada Tabel 3.

Tabel 3. Hasil Integrasi Algoritma Klasifikasi dan *Feature Selection* Terbaik

<i>Deep Learning + Information Gain</i>	
ACC	78.63%
KAPPA	0.626

Tabel 3 menunjukkan bahwa hasil integrasi tersebut menghasilkan peningkatan akurasi dan juga nilai kappa. Terlihat dimana sebelumnya nilai akurasi *Deep Learning* sebelum menggunakan *Feature Selection Information Gain* adalah 78,08% dan kappa 0,618, sedangkan setelah diintegrasikan maka nilai akurasi meningkat menjadi 78,63% dan kappa 0,626. Peningkatan akurasi juga terjadi di Algoritma *k-NN* dan *Naïve Bayes*, sedangkan untuk *Decision Tree* justru mengalami penurunan akurasi dan nilai kappa. Hal ini mengkonfirmasi penelitian yang telah dilakukan oleh Aldi Rahmansyah [23], dimana pada penelitian tersebut dikatakan bahwa penggunaan *Feature Selection* di beberapa algoritma, ada yang dapat meningkatkan akurasi dan kappa dan ada juga yang justru menurunkan akurasi dan kappa, karena sekecil apapun pengaruh suatu *feature*, untuk algoritma non-probabilistik cukup berpengaruh untuk meningkatkan akurasi keseluruhan suatu algoritma.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa Algoritma *Deep Learning* merupakan algoritma terbaik untuk klasifikasi *dataset* sentimen *bitcoin*, dengan nilai akurasi 78,08% dan kappa 0,618. Kemudian hasil perbandingan *Feature Selection*, *Information Gain* menjadi yang terbaik dengan rata-rata nilai akurasi sebesar 63,79% dan rata-rata kappa 0,382. Hasil integrasi antara *Deep Learning* dan *Information Gain* terbukti meningkatkan akurasi dan kappa, dimana nilai akurasi naik dari 78,08% menjadi 78,63% dan nilai kappa dari 0,618 menjadi 0,626. Maka dengan nilai tersebut klasifikasi yang dilakukan masuk kedalam kategori *Fair Classification*.

DAFTAR PUSTAKA

- [1] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Comparison Of Data Mining Algorithm For Forecasting Bitcoin Crypto Currency Trends," *JUTIF*, vol. 3, no. 2, pp. 245–248, 2022.
- [2] K. Fatmawati and A. P. Windarto, "Data Mining : Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 3, no. 2, pp.

- 173–178, 2018.
- [3] D. T. Larose and C. D. Larose, *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining*. Canada: John Wiley & Sons, Inc, 2014.
- [4] A. P. Singh and S. Malani, “Understanding and Predicting Trends In Cryptocurrency Prices Using Data Mining Techniques,” *IIIT Hyderabad*, pp. 1–7, 2018.
- [5] R. C. Noorsanti, H. Yulianton, and K. Hadiono, “Blockchain - Teknologi Mata Uang Cryptocurrency,” *Pros. SENDI_U 2018*, pp. 306–311, 2018.
- [6] D. S. Utami and A. Erfina, “Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, vol. 1, no. 1, pp. 299–305, 2021.
- [7] R. Parlika, S. I. Pradika, A. M. Hakim, and K. R. N. M., “Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob,” *J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, no. 2, pp. 33–37, 2020.
- [8] J. Ipmawati, Kusrini, and E. Taufiq Luthfi, “Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen,” *Indones. J. Netw. Secur.*, vol. 6, no. 1, pp. 28–36, 2017.
- [9] I. M. B. Adnyana, “Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa,” *J. Sist. dan Inform.*, vol. 13, no. 2, pp. 72–76, 2019.
- [10] O. Somantri and M. Khambali, “Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes dan Algoritme Genetika,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 3, pp. 301–306, 2017, doi: 10.22146/jnteti.v6i3.332.
- [11] Nurfaizah, N. Hermanto, and Y. I. Romadon, “Seleksi Fitur Information Gain Dan Algoritma Naïve Bayes Untuk Review Opini Konsumen,” *Comput. Based Inf. Syst. J.*, vol. 8, no. 2, pp. 55–59, 2020.
- [12] A. Y. Pratama, Y. Umaidah, and A. Voutama, “Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja),” *Sains Komput. Inform.*, vol. 5, no. 2, pp. 897–910, 2021, [Online]. Available: https://tunasbangsa.ac.id/ejurnal/index.php/j_sakti/article/view/386/365.
- [13] I. Zulfa and E. Winarko, “Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 187, 2017, doi: 10.22146/ijccs.24716.
- [14] N. Tri Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes, KNN, dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [15] Nuryani and D. Mahayana, “Analisis Sentimen Berbasis Aspek dengan Deep Learning Ditinjau dari Sudut Pandang Filsafat Ilmu,” *Jumanji*, vol. 4, no. 2, pp. 70–85, 2021.
- [16] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi: (Sebuah Ulasan),” *J. JUPITER*, vol. 13, no. 1, pp. 66–78, 2021, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/jupiter/article/view/3249/1396>.
- [17] C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval (2nd edition)*. Cambridge: Cambridge University Press, 2009.
- [18] Han and Kamber, *Data Mining Concepts and Technique*. San Francisco: Diane Cerra, 2006.
- [19] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Technique*. San Francisco: Morgan Kaufmann, 2011.
- [20] F. Rahutomo, A. Retno, T. Hayati, and P. N. Malang, “Evaluasi daftar stopword bahasa indonesia,” vol. 6, no. 1, 2019, doi: 10.25126/jtiik.201861226.
- [21] L. K. Harsono, Y. Alkhalifi, Nurajijah, and W. Gata, “Analisis Sentimen Stakeholder atas Layanan haiDJPb pada Media Sosial Twitter Dengan Menggunakan Metode Support Vector Machine dan Naïve Bayes,” *J. Ilmu-ilmu Inform. dan Manaj.*, vol. 14, no. 1, pp. 36–44, 2020.
- [22] D. Nurlaela, “Penerapan Adaboost untuk Meningkatkan Akurasi Naive Bayes Pada Prediksi Pendapatan Penjualan Film,” *Inti Nusa Mandiri*, vol. 14, no. 2, pp. 181–188, 2020.
- [23] A. Rahmansyah, O. Dewi, P. Andini, T. Hastuti, P. Ningrum, and M. E. Suryana, “Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine,” *Semin. Nas. Apl. Teknol. Inf.*, pp. 1907–5022, 2018