

ANALISA METODE NAÏVE BAYES DENGAN GAIN RATIO PADA DATA IDENTIFIKASI KACA

Oleh:

Alvian ¹⁾

Zulham Sitorus ²⁾

Arpan ³⁾

Universitas Pembangunan Pancabudi, Medan ^{1,2,3)}

E-mail :

alvianno@gmail.com ¹⁾

zulhamsitorus@gmail.com ²⁾

arpan@gmail.com ³⁾

ABSTRACT

Tests that have been carried out on all data show that the Gain Ratio classification model can provide better accuracy values because there is a change in the weighting of the attribute values in the dataset used. The value of the weighted Gain ratio is used to calculate the probability in Nave Bayes, which is a parameter to see the relationship between each attribute in the data, and is used as the basis for the weighting of each attribute of the dataset. The higher the Gain ratio of an attribute, the greater the relationship to the data class. So that the accuracy value increases than the accuracy value generated by the Naïve Bayes classification model. The increase in accuracy in the Naïve Bayes classification model is due to the number of weight accuracy from the attribute selection in the Gain ratio.

Keywords: Accuracy, Gain Ratio, Classification, Nave Bayes

ABSTRAK

Pengujian yang telah dilakukan pada semua data menunjukkan bahwa model klasifikasi Gain Ratio dapat memberikan nilai akurasi yang lebih baik karena terdapat perubahan bobot nilai atribut pada dataset yang digunakan. Nilai weighted Gain ratio digunakan untuk menghitung probabilitas pada Nave Bayes, yang merupakan parameter untuk melihat hubungan antar setiap atribut dalam data, dan digunakan sebagai dasar pembobotan setiap atribut dari dataset. Semakin tinggi rasio Gain suatu atribut, semakin besar hubungannya dengan kelas data. Sehingga nilai akurasinya meningkat dibandingkan dengan nilai akurasi yang dihasilkan oleh model klasifikasi Naïve Bayes. Peningkatan akurasi pada model klasifikasi Naïve Bayes disebabkan oleh banyaknya akurasi bobot dari pemilihan atribut pada Gain ratio.

Kata kunci: Akurasi, Gain Ratio, Klasifikasi, Nave Bayes

1. PENDAHULUAN

Pada penelitian klasifikasi merupakan metode yang dipakai dalam membuat ketetapan pada keberadaan kelas dengan penelitian tertentu. Sedangkan pada penelitian memanfaatkan klasifikasi dalam

membuat sebuah deskripsi data dengan membuat prediksi pada sebuah data dimasa datang merupakan salah satu masalah yang perlu dihadapi pada algoritma klasifikasi yang memiliki fitur yang terkadang tidak

relevan, maka perlu dilakukan penambahan fitur pada metode tersebut.

Klasifikasi merupakan langkah pembelajaran secara terarah. Klasifikasi berfungsi untuk memprediksi kelas dari objek yang kelas-kelasnya belum diketahui. Naïve bayes adalah model klasifikasi yang bersifat probabilitas yang lebih mudah dalam *machine learning* dengan cara melakukan perhitungan dari dataset yang bertujuan memprediksi probabilitas dalam sebuah kelas dengan asumsi ketidakketergantungan yang kuat.

Menurut beberapa metode klasifikasi pada datamining yaotu *Naïve Bayes* yang memiliki tujuan dalam membangun model klasifikasi dengan memakai dataset yang diuji yang dapat diproses dalam penentuan kelas. Sedangkan menurut *Naïve Bayes* adalah teknik yang dipakai dalam melabel kelas dalam sebuah permasalahan klasifikasi sehingga dapat direpresentasikan dengan vector nilai dengan menggunakan label kelas dari himpunan-himpunan datanya.

Gain ratio merupakan metode yang di modifikasi dari *information gain* pada penelitian. *Information Gain* telah diperbaiki oleh *Gain Ratio* dengan pengambilan informasi intrinsic pada tiap atribut dengan hasil yang telah diperbaiki oleh algoritma *Gain*. *Gini Index* yaitu criteria yang tifak murni atau data yang pincang dalam engukur divergensi antara

probabilitas dan nilai atribut pada target. Sedangkan menurut yang menyatakan *Gini Index* dapat mengklasifikasi data yang diinginkan melalui data yang sesuai secara dinamis.

Penelitian penggunaan *ingormation gain ratio* yang dapat melakukan peningkatan terhadap metode naïve bayes yang berdimensi tinggi, dengan meningkatkan pembobotan setiap atribut dengan data yang dipakai. Kemudian bobot itu menghasilkan hitungan dari *information gain ratio*. penelitian ini menggunakan 10 dataset dengan menggunakan nilai akurasi sebesar 4,25%.

Salah satu dari banyak metode yang dipakai dalam penklasifikasian data yang tidak terlihat dan untuk memperoleh sebuah informasi dari beberapa data. Jurnal ini membahas tentang metode klasifikasi Naïve Bayes yang dikombinasikan dengan Sistem Pendukung Keputusan yang tidak perlu penggunaan bobot pada proses perhitungan, akan tetapi dapat menggunakan nilai probabilitas pada data yang ada. Penulis memilih Naïve Bayes dalam melakukan klasifikasi dalam penelitiannya karena peluang klasifikasi yang mudah untuk dipelajari, efisien, akurasi akurat, dan memiliki kemampuan dalam klasifikasi yang berdasarkan kategori matematika sederhana.

2. TINJAUAN PUSTAKA

2.1. Klasifikasi

Beberapa peneliti menyebutkan klasifikasi dengan istilah ramalan dalam melakukan pengkategorian kelas dan memproses data yang digolongkan dalam beberapa bagian atau membuat sebuah pola atau model dengan melakukan proses pada data latih yang ada agar membuat sebuah nilai pada kelas-kelas pada tiap atribut pada dataset dan menggunakan kelompok data baru. Biasanya klasifikasi digunakan dalam memproses pengguna bank dalam melakukan pengkreditan, target penjualan, diagnosa penyakit dirumah sakit, dan analisa beberapa masalah-masalah yang membutuhkan hasil keputusan.

Pada proses klasifikasi ini membutuhkan kelas, data yang untuk diprediksi, serta data yang untuk dilatih, dan data uji. Proses ini dilakukan dengan cara melakukan penguraian kepada himpunan kelas yang ada dan menggunakan pola yang bertujuan agar klasifikasi tuple data yang pada kelasnya belum dideteksi. Kemudian predictor bertujuan dalam penentuan pola-pola pada klasifikasi data, *Decision Tree*, atau rumus matematika. Jadi tujuan dari data latih adalah agar mendapatkan hasil pengelompokan data yang telah ada kelas dan prediktornya. Sedangkan data uji dihasilkan dari pengelompokan dari model klasifikasi yang telah diproses untuk mendapatkan data yang baru.

2.2. Naïve Bayes

Naïve bayes merupakan salah satu model klasifikasi. Naïve bayes memiliki probabilitas yang lebih mudah dalam *machine learning* dengan cara melakukan perhitungan dari dataset yang bertujuan memprediksi probabilitas dalam sebuah kelas dengan asumsi ketidakketergantungan yang kuat.

Naïve bayes adalah model klasifikasi yang bersifat probabilitas yang lebih mudah dalam *machine learning* dengan cara melakukan perhitungan dari dataset yang bertujuan memprediksi probabilitas dalam sebuah kelas dengan asumsi ketidakketergantungan yang kuat.

Naïve Bayes adalah alat mesin belajar untuk memecahkan masalah masalah probabilitas. Naïve bayes lebih kompleks dan memperlambat penyebaran siklus ketika tidak ada sampel pelatihan yang meningkat karena ketika jumlah pelatihan sampel yang meningkatkan Naïve Bayes membutuhkan lebih banyak penyimpanan. Naïve Bayes awalnya dikembangkan untuk masalah klasifikasi biner dan menemukan langsung perumusan kasus multi-kelas tidak mudah dan masih menjadi masalah penelitian yang sedang berlangsung. Pada Naïve Bayes terdapat 2 cara untuk mengklasifikasi multiclass; yang pertama adalah mempertimbangkan semua data dalam satu formulasi optimasi, dan yang kedua adalah menyederhanakan masalah multi-kelas ke

beberapa masalah biner. Solusi kedua dianggap lebih dari pendekatan pertama karena pengklasifikasi biner lebih mudah untuk diterapkan dan terlebih lagi beberapa Algoritma yang kuat seperti Naïve Bayes secara inheren biner.

3. METODE PENELITIAN

Pada rancangan penelitian yang dilakukan pada diatas tersebut, penulis menguji algoritma Naïve Bayes dengan penginputan data kemudian data tersebut dilakukan proses klasifikasi dengan algoritma Naïve Bayes dengan menggunakan *Gain Ratio*.

Pada klasifikasi menggunakan *Naïve Bayes* dengan data uji dan data latih dilakukan perhitungan dalam melakukan uji data yang kemudian akan ditampilkan dalam tabel *confusion matrix*. Pada klasifikasi yang dihasilkan oleh data uji memiliki kelas yang berbeda yaitu positif dan negative. Berikut

merupakan table *confusion matrix* sebagai berikut

Pada data yang didalamnya memiliki kelas yang bernilai positif yang hasil prediksi klasifikasinya sesuai dengan actual maka memiliki nilai positif disebut dengan *True Positive*. Pada data yang didalamnya memiliki kelas yang bernilai negative yang hasil prediksi klasifikasinya sesuai dengan actual maka memiliki nilai negative disebut dengan *True Negative*.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini emnggunakan dataset Identifikasi Kaca yang berasal dari Kaggle.com, setelah melakukan proses preprocessing makan mendapatkan hasil data uji yang selanjutnya akan diuji dengan metode Naïve Bayes yang berdasarkan *Confusion Matrix*. Dibawah ini merupakan data yang digunakan pada penelitian:

Tabel 2. Dataset Kaca dari Kaggle.com

No.	RI	Na	Mg	Al	...	Type
1	0.344	0.288	1.254	-0.704	...	1
2	0.336	0.592	0.637	-0.180	...	1
3	0.333	0.154	0.603	0.182	...	1
4	0.336	-0.237	0.700	-0.321	...	1
5	0.336	-0.164	0.651	-0.422	...	1
6	0.332	-0.751	0.644	0.343	...	1

Atribut Informasi:

RI : indeks bias

Na : Natrium

mg : magnesium

Al : Aluminium

Si : silikon

K : Kalium

Ca : Kalsium

Ba : Barium

Fe : Besi

Ciri kaca: (atribut kelas)

1: jendela mengapung

2: jendela non float

3: jendela kendaraan mengapung

4: jendela kendaraan non float (tidak ada dalam database ini)

5: wadah

6: peralatan makan

7: lampu depan

Setelah diperoleh *Preprocessing*, maka selanjutnya model klasifikasi tersebut diuji menggunakan data uji dataset pada Identifikasi Kaca. Selanjutnya melakukan pemberian bobot terhadap atribut dari dataset berdasarkan *Gain ratio*. Tabel dibawah ini merupakan hasil dari algoritma gain ratio untuk pembobotan atribut Identifikasi Kaca. Berikut adalah hasil output dari pengujian:

Tabel 3. Confussion Matrix Naïve Bayes Gain Ratio

Actual	Predicted	Predicted	Predicted	Predicted	Predicted	Predicted
	Class 1	Class 2	Class 3	Class 5	Class 6	Class 7
Actual Class 1	59	0	0	0	0	0
Actual Class 2	0	76	0	0	0	0
Actual Class 3	0	0	17	0	0	0

Actual Class 5	0	0	0	13	0	0
Actual Class 6	0	0	0	0	9	0
Actual Class 7	0	0	0	0	0	29

Tabel 5. Confussion Matrix Naïve Bayes

Actual	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 5	Predicted Class 6	Predicted Class 7
Actual Class 1	55	31	13	0	1	1
Actual Class 2	9	40	4	3	0	1
Actual Class 3	5	0	0	0	0	0
Actual Class 5	0	2	0	9	0	3
Actual Class 6	0	2	0	0	7	0
Actual Class 7	0	1	0	1	1	24

Tabel 7. Confussion Matrix Naïve Bayes

Actual	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 5	Predicted Class 6	Predicted Class 7
Actual Class 1	55	31	13	0	1	1
Actual Class 2	9	40	4	3	0	1
Actual Class 3	5	0	0	0	0	0
Actual Class 5	0	2	0	9	0	3
Actual Class 6	0	2	0	0	7	0
Actual Class 7	0	1	0	1	1	24

$$55+40+0+9+7+24+31+13+1+1+9+4+3+1+5+2+3+2+1+1+1 = 213$$

Tingkat prediksi klasifikasi *Naïve Bayes* adalah sebesar 63.3%. Proses klasifikasi menggunakan *Naïve Bayes* dan *Gain Ratio* yang dapat mempengaruhi nilai akurasi terhadap dataset. Hasil pengujian yang dilakukan memiliki perbandingan

akurasi dari kedua metode yang dapat dilihat pada gambar berikut ini:

5. SIMPULAN

Proses klasifikasi dengan algoritma *Naïve Bayes* dengan dataset Identifikasi Kaca dengan akurasi 63.3%. Hasil ini merupakan

hasil yang kurang baik. Maka dari itu penulis melakukan pengujian kembali terhadap gain ratio dengan hasil akurasi algoritma naive bayes meningkat dengan hasil sebesar 100% dengan begitu maka *gain ratio* lebih meningkatkan akurasi yang akurasi 36.7%.

6. DAFTAR PUSTAKA

- Alharan, A. F. H., Fatlawi, H. K., & Ali, N. S. 2019. *A cluster-based feature selection method for image texture classification*. Indonesian Journal of Electrical Engineering and Computer Science, 14(3), 1433–1442. <https://doi.org/10.11591/ijeecs.v14.i3.pp1433-1442>
- Granik, M., & Mesyura, V. 2017. *Fake news detection using naive Bayes classifier*. 2017 IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 - Proceedings, 900–903. <https://doi.org/10.1109/UKRCON.2017.8100379>
- Alhayali, Ibrahim R. A., Ahmed, M. A., Mohialden, Y. M., & Ali, A. H. 2020. *Efficient method for breast cancer classification based on ensemble hoeffding tree and naïve Bayes*. Indonesian Journal of Electrical Engineering and Computer Science, 18(2), 1174–1180. <https://doi.org/10.11591/ijeecs.v18.i2.pp1074-1080>
- Karomi., M. Adib Al, Ivandari. 2019. *Optimasi Algoritma Naïve Bayes dengan Information Gain Ratio untuk Menangani Dataset Berdimensi Tinggi*. STMIK Widya Pratama Pekalongan.
- Raviya. Kaushik H & Gajjar, Biren. 2013. *Performance Evaluation of Different Data Mining Classification Algoritma Using WEKA*. Indian Journal of Research. Volume. 2. Issue.1. ISSN: 2250-1991.
- Seth, H. R., & Banka, H. 2016. *Hardware implementation of Naive Bayes classifier: A cost effective technique*. 3rd International Conference on Recent Advances in Information Technology, RAIT 2016, 264–267. <https://doi.org/10.1109/RAIT.2016.7507913>
- Witten, I.H., Frank, E. & Hall, M. a., 2015. *Data Mining: Practical Machine Learning Tools and Techniques Third Edit.*, USA: Morgan Kaufmann Publishers
- XuLingling ' Pan Zuguang Hu'Bing, Shangfeng Zhang'Juying counterpart Zeng The Dynamic KNN Clustering of Undergraduate Consumption With Itâ s like this greatly enhanced: A Case of Zhejiang 0-7695-2882-1/07 \$25.00 ©2007 IEEE

Yuliana, Erlangga. 2017. *Analysis of Data Mining Methods Naive Bayes Classifier (Nbc)*. Information System, Computer Science Faculty, Bandar Lampung University, Indonesia.