

Penerapan Data Mining Classification untuk Data Blogger Menggunakan Metode Naïve Bayes

Recha Abriana Anggraini^{*1}, Galih Widagdo^{*2}, Arief Setya Budi^{*3}, M. Qomaruddin^{*4}

**Program Studi Sistem Informasi STMIK Nusa Mandiri
Jakarta Pusat*

¹anggrainirecha@gmail.com

²galihwidagdo19@gmail.com

³ariefsetya334@gmail.com

⁴qomaruddin.mqn@bsi.ac.id

Abstrak— Jumlah pengguna situs *blogger* yang semakin meningkat menyebabkan perlu dilakukan pengklasifikasian data untuk mengetahui pengguna tersebut masuk dalam kategori pengguna *blogger professional* atau bukan. Sebagai referensi terkait penelitian ini adalah penelitian yang sudah dilakukan oleh peneliti sebelumnya. Teknik pengklasifikasian pemodelan deskriptif dan prediktif dengan algoritma *data mining* yaitu menggunakan metode naïve bayes. Untuk mengelola data digunakan *software rapid miner studio 6.0*, *dataset blogger* diperoleh dari *website UCI Machine learning Repository*, Perhitungan performance vector menunjukkan akurasi klasifikasi metode Naive bayes diperoleh sebesar 86.67%. Sedangkan *class precision* dan *class recall* untuk prediksi *yes* menunjukkan tingkat *precision* sebesar 91.30% dan untuk prediksi *no* sebesar 71.43%. Hasil klasifikasi dari data *blogger* dengan metode naïve bayes membagi 2 kelas klasifikasi PB yaitu *class yes* dan *class no*. Untuk nilai *class yes* yaitu 0.680 dan nilai *class no* yaitu 0.320. Dari hasil pengolahan data dapat diketahui bahwa tingkat akurasi pengklasifikasian data *blogger* mencapai 86.67%.

Kata kunci— *Naïve Bayes, Data Mining, Klasifikasi, Blogger, Rapid Miner*

I. PENDAHULUAN

Perkembangan teknologi dalam bidang informasi yang sangat pesat pada saat ini diharapkan dapat memberikan kemudahan bagi manusia dalam mencari dan menyebarkan informasi yang bermanfaat dan dibutuhkan bagi kehidupan. Pencarian dan penyebaran informasi dapat dilakukan melalui berbagai media, namun yang paling populer adalah melalui media *website* dengan memanfaatkan teknologi internet.

Salah satu layanan *website* yang berguna untuk mencari dan menyebarkan informasi yang paling populer adalah situs *blogger*. Berdasarkan data statistik dari ASEAN *blogger* yang dikutip oleh Tempo 2013 menyebutkan bahwa pada tahun 2008 terdapat 500 ribu *blogger* aktif di Indonesia, sedangkan dilansir dari antaranews.com pengguna *blogger* di Indonesia pada

tahun 2015 sekitar 3 juta pengguna. Berdasarkan data tersebut dapat disimpulkan bahwa pengguna situs *blogger* di Indonesia semakin meningkat, hal itu juga menyebabkan penyebaran informasi semakin beraneka ragam.

Jumlah pengguna situs *blogger* yang semakin meningkat menyebabkan perlu dilakukan pengklasifikasian data berdasarkan beberapa faktor untuk mengetahui pengguna tersebut masuk dalam kategori pengguna *blogger professional* atau bukan.

Berdasarkan fenomena tersebut maka dilakukan penelitian dengan menggunakan teknik *data mining* untuk melakukan klasifikasi data dengan algoritma naïve bayes. Penelitian yang digunakan sebagai referensi panduan terkait penelitian ini adalah penelitian yang dilakukan oleh Ardiyansyah, Panny Agustia Rahayuningsih, dan Reza Maulana yang berjudul analisis perbandingan algoritma klasifikasi *data mining* untuk *dataset blogger* dengan rapid miner (Ardiyansyah, Rahayuningsih, & Maulana, 2018).

II. TINJAUAN PUSTAKA

A. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah penemuan atau pencarian pengetahuan (nilai tambah) di dalam sebuah *database*, karena *data mining* adalah suatu rangkaian proses (Sari & Saro, 2018).

B. Data Mining

Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Darmawan, Kustian, Rahayu, & Tabebuya, 2018).

Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang (Hermawati, 2013).

C. Dataset

Himpunan data (dataset) merupakan kumpulan dari objek dan atributnya (Hermawati, 2013). Dalam dataset, jenis data dapat dibagi menjadi 2 bagian yaitu:

1. Data Training

Data training adalah data yang digunakan untuk perhitungan probabilitas dari data berdasarkan data pembelajaran yang dilakukan (Pratiwi & Nugroho, 2016). Data training merupakan data yang sebelumnya sudah ada sesuai dengan fakta.

2. Data Testing

Data testing merupakan data yang akan atau sedang terjadi dan dipergunakan sebagai bahan uji yang sebelumnya sudah didapatkan pada data training (Pratiwi & Nugroho, 2016). Data ini digunakan untuk mengukur sejauh mana klasifikasi berhasil melakukan klasifikasi dengan benar.

D. Naïve Bayes

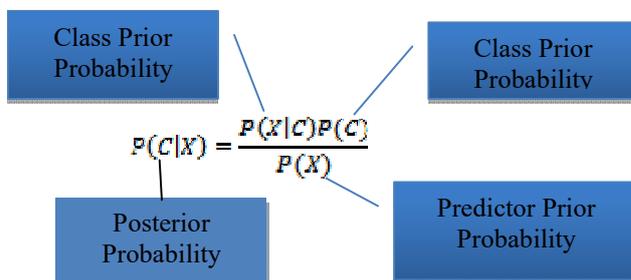
Teori bayes adalah kondisi probabilitas suatu kejadian hipotesis bergantung pada kejadian lain sebagai bukti (Moriesta & Ibrahim, 2017). Metode yang diadopsi dari nama penemunya yaitu Thomas Bayes pada tahun 1950. Pada dasarnya, teori tersebut menyatakan bahwa kejadian dimasa depan dapat diprediksi dengan syarat kejadian sebelumnya telah terjadi. Teori naïve bayes memiliki kemampuan klasifikasi yang serupa dengan decision tree dan neural network bahkan algoritma naïve bayes memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Kusrini dan Luthfi, 2013). Secara umum teori naïve bayes ditulis dalam bentuk (Saleh & Nasari, 2018) :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots [1]$$

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- P(H) : Probabilitas hipotesis H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

Untuk menemukan kelas yang cocok bagi sampel yang dianalisis dalam proses klasifikasi, metode naïve bayes di atas disesuaikan sebagai berikut :



Gambar 1. Rumus Naive Bayes

Keterangan :

- x : Data dengan class yang belum diketahui
- c : Hipotesis data merupakan suatu class spesifik
- P(c|x) : Probabilitas hipotesis berdasar kondisi (posteriori probability)
- P(c) : Probabilitas hipotesis (prior probability)
- P(x|c) : Probabilitas berdasarkan kondisi pada hipotesis
- P(x) : Probabilitas c

Rumus diatas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus diatas dapat pula ditulis sebagai berikut :

$$Posterior = \frac{prior \times likelihood}{evidence} \dots [2]$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan (c|x1,...,xn) menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned}
 P(C|X_1, \dots, X_n) &= P(C)P(X_1, \dots, X_n|C) \\
 &= P(C)P(X_1|c)(X_2, \dots, X_n|C, X_1) \dots [3] \\
 &= P(C)P(X_1|c)P(X_2|C, X_1)(X_3, \dots, X_n|C, X_1, X_2) \\
 &= P(C)P(X_1|c)P(X_2|C, X_1)P(X_3|C, X_1, X_2) \dots P(X_n|C, X_1, X_2, \dots, X_{n-1})
 \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang

sangat tinggi (naif), bahwa masing masing petunjuk saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(c|X_1, \dots, X_n) = P(c) \prod_{i=1}^n P(X_i|c)$$

$$P(c|X) = P(X_1|c) \dots P(X_n|c)P(c) \dots [4]$$

Persamaan diatas merupakan model dari Teorema Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss :

$$P = (X_i - \mu_i | Y_i - y_i - \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots [5]$$

Keterangan :

- P* : Peluang
- X_i* : Atribut ke i
- x_i* : Nilai atribut ke i
- Y* : Kelas yang dicari
- y_j* : Sub kelas Y yang dicari
- μ* : Mean, menyatakan rata rata dari seluruh atribut
- σ* : Deviasi standar, menyatakan varian dari seluruh atribut

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \dots [6]$$

Deviasi Standar

$$\sigma = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2 \right]^{0.5} \dots [7]$$

Algoritma *data mining* naive bayes juga banyak digunakan pada penelitian-penelitian sebelumnya terutama untuk mengklasifikasikan data. Beberapa contoh penelitian terdahulu yang menggunakan algoritma naive bayes antaralain yaitu penelitian yang berjudul *opinion mining* untuk ulasan produk dengan klasifikasi naive bayes. Pada penelitian tersebut diketahui tingkat akurasi dari pengolahan data menggunakan metode naive bayes sebesar 77,7% (Jeremy, A., Christanti, V., & Mulyawan, 2018). Selain penelitian tersebut, penelitian lain yang menggunakan algoritma naive bayes berjudul algoritma naive bayes *classifier* pada aplikasi *data mining* berbasis *web*. Pada penelitian tersebut, diketahui hasil pengolahan datanya mencapai tingkat akurasi sebesar 83,00% (Murdiyansyah & Siswanto, 2018). Berdasarkan penelitian-penelitian yang terdahulu, yang menggunakan algoritma naive bayes peneliti mengetahui bahwa belum ada penelitian yang juga membuktikan hasilnya dengan pengujian manual. Sehingga pada penelitian ini, penulis mencoba untuk membuktikan hasil pengujian naive bayes dengan rapid miner dan juga hasil perhitungan naive bayes secara manual.

III. METODOLOGI

Penelitian ini menggunakan konsep *data mining* dengan bantuan *software* rapid miner studio 6.0. *Dataset* yang digunakan adalah *dataset* blogger yang diperoleh dari sebuah *website* yang menyediakan data publik yaitu UCI Machine learning Repository. *Dataset* blogger bersumber dari Kohkiloye dan Boyer Ahmad Province di Iran. Data tersebut akan digunakan sebagai *data training* dan *data testing*. Atribut yang terdapat pada *dataset* blogger yaitu *Local Political Social Space (LPSS)*, *Local Media Turnover (LMT)*, *topics*, *Caprice*, dan *Degree*. Sedangkan untuk *class* yang terdapat pada *dataset* tersebut yaitu *Professional Blogger (PB)* dengan nilai *yes* dan *no*. Teknik pengolahan *dataset* dalam penelitian ini menggunakan teknik klasifikasi dengan permodelan deskriptif dan prediktif menggunakan salah satu algoritma *data mining* yaitu metode naive bayes.

IV. HASIL DAN PEMBAHASAN

Dalam permodelan algoritma naive bayes pada penelitian ini, *data training* diolah menggunakan model *10 fold* serta *cross validation*. Jumlah responden dalam penelitian ini sebesar 100 responden, dengan *dataset* blogger ini peneliti akan mengklasifikasikan jenis *blogger* kedalam 2 kelompok yaitu *Blogger Professional (BP)* dan *Blogger Musiman (BM)*. Berikut merupakan data training dari *dataset* blogger:

TABEL 1.
DATA TRAINING DATASET BLOGGER

Degree	Caprice	Topics	Lmt	Lpss	Pb
high	left	impressi on	yes	yes	yes
high	left	political	yes	yes	yes
medium	middle	tourism	yes	yes	yes
high	left	political	yes	yes	yes
medium	middle	news	yes	yes	yes
medium	middle	news	yes	yes	yes
high	left	political	yes	yes	yes
high	right	political	yes	no	yes
high	right	political	yes	no	no
.
.
.
low	right	impressi on	yes	no	no
medium	right	news	yes	yes	no
medium	left	impressi on	yes	yes	yes

Tabel 1 merupakan *dataset* blogger yang dipakai sebagai *data training* dalam penelitian ini. Data tersebut berjumlah 100 data yang direpresentasikan dalam bentuk tabel.

V. PROSES DATA MINING

A. Pemrosesan Awal

Dalam pemrosesan ini yang menjadi sarannya adalah data *blogger* pada tahun 2013 yang akan menghasilkan grafik atau pola pengelompokan *blogger* serta memperlihatkan prediksi pengguna *blogger* pada tahun yang akan datang.

Pembuatan model diawali dengan pembacaan file data (*read excel*). *Data training* disimpan dalam satu file excel kemudian data tersebut divalidasi menggunakan *cross validation*.

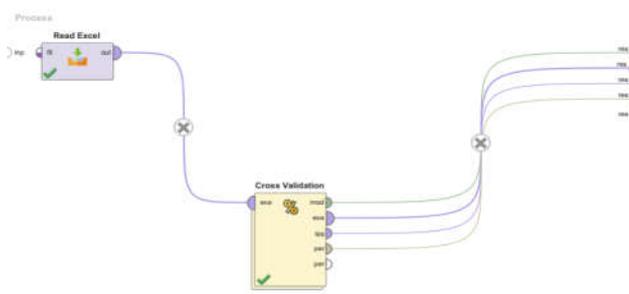
TABEL 2. DATA TRAINING

Degree	Caprice	Topics	Lmt	Lpss	Pb
high	left	impression	yes	yes	yes
high	left	political	yes	yes	yes
medium	middle	tourism	yes	yes	yes
high	left	political	yes	yes	yes
medium	middle	news	yes	yes	yes
medium	middle	news	yes	yes	yes
high	left	political	yes	yes	yes
high	right	political	yes	no	yes
high	right	political	yes	no	no
.
.
low	right	impression	yes	no	no
medium	right	news	yes	yes	no
medium	left	impression	yes	yes	yes

Tabel 2 merupakan *data training* yang akan diolah dengan rapid miner untuk mendapatkan hasil klasifikasi data.

B. Validation

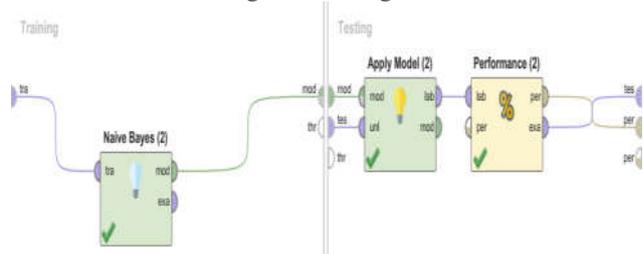
Validation dilakukan dengan analisis berbagai model dan memilih model dengan kinerja prediksi yang baik. Gambar berikut merupakan proses validasi menggunakan *cross validation*, setelah pembacaan file data, blok *read excel* dihubungkan dengan blok *cross validation*. Proses *data mining* berada didalam proses (*cross validation*) dapat ditampilkan dengan cara melakukan *double klik* pada *mouse* sehingga muncul proses *training* dan *testing*.



Gambar 2. Validation

Gambar 2 merupakan proses validasi menggunakan model olah data *10 fold* dan *cross validation*.

C. Proses Training dan Testing



Gambar 3. Proses Training dan Testing

Gambar 3 merupakan proses *training* dan *testing* data menggunakan model *10 fold* dalam rapid miner.

D. Hasil Performance Vector

Proses klasifikasi dengan rapid miner studio dengan metode *naïve bayes* yang digunakan untuk mengklasifikasikan data *blogger* pada penelitian ini menghasilkan *accuracy*, *precision*, dan *recall*.

accuracy: 76.00% +/- 10.25% (micro average: 76.00%)

	true yes	true no	class precision
pred. yes	82	19	76.54%
pred. no	6	13	68.42%
class recall	91.18%	40.62%	

Gambar 4. Hasil Performance Vector

Perhitungan *performance vector* pada gambar 4 menunjukkan akurasi klasifikasi menggunakan metode *Naive bayes* diperoleh sebesar 75,00% +/- 10,25%. Sedangkan *class precision* dan *class recall* untuk prediksi *yes* menunjukkan tingkat *precision* sebesar 76,54% dan untuk prediksi *no* sebesar 68,42%.

E Hasil Klasifikasi Class dengan Metode Naïve Bayes

1. Simple Distribution Models

Dengan menggunakan *software* rapid miner studio untuk menganalisis tabel data *blogger* dengan menggunakan metode *naïve bayes* dapat menghasilkan beberapa kelas utama pembagian sebagai berikut :

```

SimpleDistribution
Distribution model for label attribute pb

Class yes (0.680)
5 distributions

Class no (0.320)
5 distributions
    
```

Gambar 5. Simple Distribution

Gambar 5 merupakan hasil klasifikasi dari data *blogger* dengan metode *naïve bayes* membagi 2 kelas klasifikasi PB yaitu *class yes* dan *class no*. Untuk nilai *class yes* yaitu 0.680 dan nilai *class no* yaitu 0.320.

2. Distribution Table

Attribute	Parameter	yes	no
Degree	value=high	0.441	0.281
Degree	value=medium	0.500	0.406
Degree	value=low	0.059	0.312
Degree	value=unknown	0.000	0.000
caprice	value=left	0.617	0.312
caprice	value=middle	0.147	0.126
caprice	value=right	0.235	0.562
caprice	value=unknown	0.000	0.000
topic	value=impression	0.255	0.250
topic	value=political	0.412	0.219
topic	value=tourism	0.147	0.166
topic	value=news	0.191	0.187
topic	value=scientific	0.015	0.187
topic	value=unknown	0.000	0.000
int	value=yes	0.882	0.812
int	value=no	0.118	0.188
int	value=unknown	0.000	0.000
loss	value=yes	0.706	0.750
loss	value=no	0.294	0.260
loss	value=unknown	0.000	0.000

Gambar 6. Distribution table

Gambar 6 merupakan *distribution table* hasil dari olah data rapid miner. Dari *distribution table* yang ada dapat diketahui hasil klasifikasi data *blogger professional* dan *blogger* musiman.

Selain menggunakan rapid miner, hasil klasifikasi data *blogger* yang ada pada *distribution table* juga dibuktikan dengan dihitung secara manual menggunakan microsoft excel. Berikut merupakan gambar perhitungan manual menggunakan microsoft excel:

job	predict ??	No	Yes	degree	No	Yes	caprice	No	Yes	topic	No	Yes	int	No	Yes	loss	No	Yes
yes	no	0.920405	0.280594	low	0.312	0.059	right	0.312	0.059	impression	0.312	0.059	no	0.312	0.059	no	0.312	0.059
		725	2747							on								
no	yes	0.9597925	0.0402074	medium	0.406	0.5	right	0.312	0.059	tourism	0.312	0.059	yes	0.312	0.059	yes	0.312	0.059
		966	305															
no	yes	0.264594	0.735405	medium	0.406	0.5	middle	0.312	0.059	impression	0.312	0.059	no	0.312	0.059	yes	0.312	0.059
		603	998							on								
yes	yes	1.309090	0.690909	medium	0.406	0.5	left	0.312	0.059	tourism	0.312	0.059	yes	0.312	0.059	no	0.312	0.059
		2728	728															
yes	yes	0.227821	0.772178	medium	0.406	0.5	middle	0.312	0.059	news	0.312	0.059	yes	0.312	0.059	yes	0.312	0.059
		9368	363															
yes	yes	0.314096	0.685903	medium	0.406	0.5	left	0.312	0.059	news	0.312	0.059	no	0.312	0.059	yes	0.312	0.059
		563	436															
no	no	0.675342	0.324657	low	0.312	0.059	middle	0.312	0.059	impression	0.312	0.059	yes	0.312	0.059	no	0.312	0.059
		551	449							on								
no	no	0.863490	0.136510	low	0.312	0.059	right	0.312	0.059	impression	0.312	0.059	yes	0.312	0.059	no	0.312	0.059
		638	373							on								

Gambar 6. Hasil perhitungan manual menggunakan microsoft excel

Gambar 6 merupakan hasil dari perhitungan manual yang dilakukan dengan microsoft excel. Dari hasil perhitungan manual tersebut dapat dilihat bahwa hasil yang ada pada *distribution table* di *rapid miner* sama dengan hasil perhitungan manual yang dilakukan menggunakan microsoft excel, dengan demikian dapat diambil kesimpulan bahwa akurasi klasifikasi data *blogger* ini baik.

VI. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat dikatakan bahwa pengolahan data untuk mengklasifikasikan *dataset blogger* menggunakan algoritma naïve bayes memiliki hasil yang baik dan dapat dijadikan sebagai acuan bagi orang-orang yang ingin mengetahui klasifikasi *blogger* musiman atau *blogger professional*. Selain menggunakan *software rapid miner studio*, peneliti juga mencoba untuk mengolah *dataset blogger* secara manual dengan menggunakan salah satu data testing sebagai prediktor untuk acuan pembuktian tingkat akurasi klasifikasi dari pengelompokan data *blogger*. Dari hasil pengolahan data manual dan pengolahan data menggunakan bantuan *software rapid miner studio*, dapat disimpulkan bahwa hasil akurasi pengolahan data *blogger* menggunakan metode naïve bayes adalah baik.

Pengolahan *dataset blogger* untuk mengklasifikasikan data *blogger profesional* dan *blogger musiman* sebaiknya ditambahkan dengan algoritma klasifikasi *data mining* yang lain agar dapat dilakukan perbandingan hasil penelitian dengan menggunakan beberapa algoritma sehingga dapat diketahui algoritma mana yang menghasilkan tingkat akurasi terbaik dalam proses klasifikasi data *blogger*.

REFERENSI

- [1.] Ardiansyah, Rahayuningsih, P. A., & Maulana. R., (2018). Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner, VI(1), 20–28.
- [2.] Darmawan, A., Kustian, N., Rahayu, W., & Tabeuya. (2018). Implementasi Data Mining Menggunakan Model Svm, 2(3), 299–307.
- [3.] Hermawati, Fajar Astuti. (2013). Data Mining. Yogyakarta: Andi Offset
- [4.] Kusri, Luthfi, Emha Taufiq. (2013). Algoritma Data Mining. Yogyakarta: Andi Offset
- [5.] Moriesta, E., & Ibrahim, A. (2017). Analisis Penyaringan Email Spam Menggunakan Metode Naive Bayes, 3(1), 45–48.
- [6.] Pratiwi Rizki Wijayatun, N. Y. S. (2016). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *Jurnal Teknik Elektro (ISSN 1411-0059)*, 8(2), 60–63. <https://doi.org/10.7454/global.v19i1.136>
- [7.] Saleh, A., & Nasari, F. (2018). Penggunaan Teknik Unsupervised Discretization pada Metode Naive Bayes dalam Menentukan Jurusan Siswa Madrasah Aliyah. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 353. <https://doi.org/10.25126/jtiik.201853705>
- [8.] Sari, F., & Saro, D. (2018). Jurnal Penelitian Pos Dan Informatika Keluarga Berencana Di Kecamatan Dumai Timur Implementation Of Algorithm C4.5 To Determining Location Priority Counseling Family Planning Program In East Dumai Abstrak, 8(1), 63–76. <https://doi.org/10.17933/Jppi.2018.080105>
- [9.] Jeremy, A., Christanti, V., & Mulyawan, B. (2018). Opinion Mining Untuk Ulasan Produk Dengan Klasifikasi Naive Bayes. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 6(1), 9.
- [10.] Murdiansyah, A. O., & Siswanto, S. (2018). Algoritma Naive Baiyes Classifier Pada Aplikasi Data Mining Berbasis WEB. *SKANIKA*, 1(1), 284-290.