

Penggunaan Metode Naïve Bayes Dalam Mengklasifikasi Pengangguran Pada Desa Bojong Kulur

Yosua Sitoto Tandi Allo¹, Verra Sofica, Noor Hasan³, Minda Septiani⁴
 Universitas Nusa Mandiri^{1,2}, Universitas Bina Sarana Informatika^{3,4}
 yosuasitoto123@gmail.com¹, verra.vsc@nusamandiri.ac.id², noor.nhs@bsi.ac.id³,
 minda.mdt@bsi.ac.id⁴

Abstrak - Pengangguran adalah keadaan dimana seorang yang tergolong pada angkatan kerja yang ingin mendapatkan kerja. Hal ini dapat dicermati untuk membantu mengetahui taraf kesejahteraan masyarakat, mengingat tingginya tingkat pengangguran pada Desa Bojong Kulur, banyaknya warga pada Desa Bojong Kulur, dan sedang terjadinya pandemik Covid-19. Oleh karena itu penulis melakukan klasifikasi pengangguran dengan metode *Naïve Bayes* pada Desa Bojong Kulur. Penulis melakukan klasifikasi dengan metode *Naïve Bayes* menggunakan aplikasi pendukung Rapidminer untuk pengujian akurasi, presisi, dan recall terhadap data yang tersedia. Pengujian dilakukan dengan menyiapkan data *training* sebanyak 40 data dan data *testing* sebanyak 10 data yang dipilih secara acak. Data *testing* tersebut akan dianalisa menggunakan aplikasi pendukung Rapidminer. Hasil pengujian akurasi, presisi, dan recall klasifikasi pengangguran pada Desa Bojong Kulur dengan metode *Naïve Bayes* cukup tinggi yaitu sebesar 80%, presisi sebesar 100%, dan recall sebesar 50%. Jadi, dapat disimpulkan bahwa metode klasifikasi *Naïve Bayes* yang digunakan memberikan proses seleksi yang cepat dan metode yang mudah dipahami dengan tingkat akurasi yang tidak dapat disangkal.

Kata Kunci : *Pengangguran, Naïve Bayes, Klasifikasi, Data Mining*

Abstract - *Unemployment is a condition in which a person belonging to the labor force wants to get a job. This can be observed to help determine the level of community welfare, given the high level of unemployment in Bojong Kulur Village, the large number of residents in Bojong Kulur Village, and the ongoing Covid-19 pandemic. Therefore, the authors classify unemployment using the Naïve Bayes method in Bojong Kulur Village. The author performs the classification using the Naïve Bayes method using the Rapidminer support application for testing accuracy, precision, and recall of the available data. The test is carried out by preparing 40 training data and 10 testing data randomly selected. The testing data will be analyzed using the Rapidminer supporting application. The results of testing the accuracy, precision, and recall of unemployment classification in Bojong Kulur Village using the Naïve Bayes method are quite high at 80%, precision at 100%, and recall at 50%. So, it can be concluded that the Naïve Bayes classification method used provides a fast selection process and an easy-to-understand method with an undeniable level of accuracy.*

Keywords: *Unemployment, Nave Bayes, Classification, Data Mining*

I. PENDAHULUAN

Peningkatan pembangunan lapangan pekerjaan sangat diperlukan dalam pembangunan nasional untuk peningkatan taraf kesejahteraan sosial. Pada dasarnya untuk melaksanakan Undang-undang Dasar 1945 tentang hak setiap warga untuk mendapatkan pekerjaan dan kehidupan yang layak. Di Negara berkembang, pengangguran merupakan masalah utama karena berdampak bagi keadaan ekonomi dan sosial Negara tersebut (Franita, 2016). Melalui taraf pengangguran bisa dicermati taraf kesejahteraan masyarakat. Pengangguran dapat terjadi ditimbulkan oleh ketidakseimbangan dalam pasar tenaga kerja. ini menunjukkan bahwa jumlah pekerja yang ditawarkan melebihi jumlah pekerja yang diminta (Dongoran et al., 2016).

Peranan sumber daya manusia (SDM) yang berkualitas dan produktif menjadi penting untuk mengurangi tingkat pengangguran. Pendidikan

dipercaya sebagai sarana meningkatkan sumber daya manusia yang berkualitas lantaran pendidikan dianggap mampu untuk menghasilkan tenaga kerja yang berkualitas, mempunyai pola pikir, dan attitude. Pendidikan merupakan salah satu cara meningkatkan kualitas sumber daya manusia (Utama et al., 2015). Pertumbuhan penduduk yang signifikan, yang tidak dibarengi dengan penyebaran secara merata dan kurangnya pasar kerja menyebabkan penurunan kesempatan dalam memperoleh pekerjaan (Mada & Ashar, 2015). Berdasarkan data dari kantor Desa Bojong Kulur pengangguran warga Desa Bojong Kulur pada tahun 2021 total pengangguran sebanyak 8.005 dengan jumlah total warga sebanyak 48.810. Banyaknya pengangguran pada Desa Bojong Kulur dan terjadinya pandemik Covid-19 yang terjadi membuat penulis melakukan klasifikasi. Pada penelitian ini akan menggunakan metode *Naïve Bayes* untuk melakukan klasifikasi pengangguran.

Pengangguran yang dimaksud adalah warga yang tidak bekerja. Studi kasus dilakukan pada Desa Bojong Kulur.

1. Data Mining

Data mining merupakan interaksi yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengeluarkan dan membedakan data bermanfaat dan informasi terkait dari kumpulan data yang sangat besar (Mardi, n.d.).

Secara analogi, data mining seharusnya lebih sempurna disebut "knowledge mining from data," yang sayangnya relatif panjang. Namun, pada jangka pendek, pengetahuan mining mungkin tidak mencerminkan adanya fokus dalam mining berdasarkan sejumlah besar data (Han & Kamber, 2006).

Data Mining adalah metode pengolahan buat menemukan suatu pola yang tersembunyi buat bisa di olah sebagai pengetahuan dan ilmu pengetahuan baru dan informasi dari data dan hasil buat keputusan pada masa depan (Putri, 2017).

2. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah proses non-trivial untuk mengidentifikasi pola bersifat baru, sah, berpotensi berguna, dan polanya akhirnya jelas dimengerti dalam data (Gullo, 2015).

Knowledge Discovery In Database(KDD) adalah metode untuk mendapatkan pengetahuan dari database yang ada (Mardi, n.d.).

3. Klasifikasi

Klasifikasi mirip seperti estimasi, kecuali bahwa variabel sasaran bersifat kategoris daripada numerik. Dalam klasifikasi, ada variabel kategoris sasaran, seperti: sebagai braket pendapatan, yang, misalnya, bisa dipartisi sebagai tiga kelas atau kategori: berpenghasilan tinggi, berpenghasilan menengah, dan berpenghasilan rendah (Larose & Larose, 2014).

Klasifikasi *Naïve Bayes* merupakan klasifikasi dalam data mining berdasarkan teorema Bayes yang digunakan dengan metode probabilitas dan statistik yang independen (Wibawa, 2018).

Teorema Bayes memiliki persamaan umum sebagai berikut:

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

Ciri pokok pada Klasifikasi *Naïve Bayes* merupakan taksiran yang sangat kuat (naïf) akan independensi sesuai dengan masing-masing kondisi / kejadian (Wibawa, 2018)

Keterangan:

X = Data pada class yang belum ditemukan

H = Hipotesis data X menunjukkan suatu class spesifik

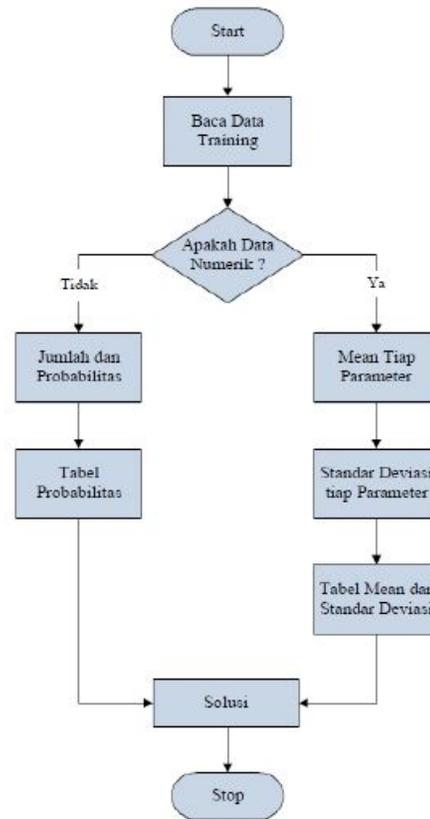
P(H|X) = Probabilitas hipotesis H beralaskan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Alur Metode *Naïve Bayes* sebagai berikut:

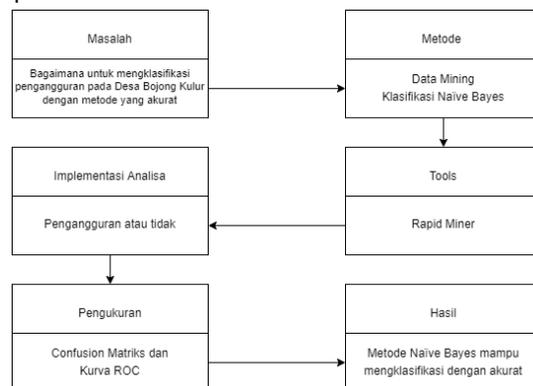


Sumber: (Saleh, 2014)

Gambar 1 Alur Metode Naïve Bayes

II. METODOLOGI PENELITIAN

Berikut adalah gambar tahapan metode penelitian.



Sumber : (Sofica et al., 2022)

Gambar 2. Tahapan Metode Penelitian

Tahapan metodologi penelitian digambarkan secara umum sebagai berikut:

1. Masalah

Pada penelitian ini terdapat masalah untuk mengklasifikasi pengangguran pada Desa Bojong Kulur sehingga perlu dicari Bagaimana untuk mengklasifikasi pengangguran pada Desa Bojong Kulur dengan metode yang akurat.

2. Metode

Metode yang digunakan untuk mengklasifikasi pengangguran pada desa Bojong Kulur adalah metode Data Mining Klasifikasi Naïve Bayes

3. Tool

Aplikasi yang digunakan untuk mengklasifikasi pengangguran pada desa Bojong Kulur adalah RapidMiner

4. Implementasi

Melakukan klasifikasi dengan menggunakan data training, dianalisa menggunakan metode Data Mining Klasifikasi Naïve Bayes untuk menganalisis pengangguran.

5. Pengukuran

Melakukan pengukuran dengan menggunakan *Confusion Matrix* dan Kurva ROC

6. Hasil

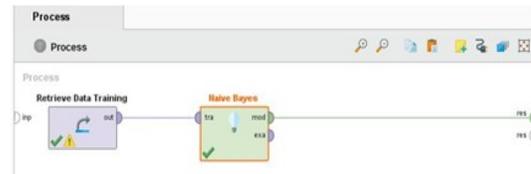
Menganalisis hasil pengolahan data dan mengukur tingkat akurasi dari metode Data Mining Klasifikasi Naïve Bayes.

Menjelaskan kronologis penelitian, termasuk desain penelitian, prosedur penelitian (dalam bentuk algoritma, Pseudocode atau lainnya), bagaimana untuk menguji dan akuisisi data. Deskripsi dari program penelitian harus didukung referensi, sehingga penjelasan tersebut dapat diterima secara ilmiah.

III. HASIL DAN PEMBAHASAN

Pengujian dilakukan untuk memutuskan apabila perhitungan yang telah dilakukan diatas tepat untuk memutuskan status pengangguran dengan menggunakan metode klasifikasi *Naïve Bayes*. Uji coba dilakukan dengan menentukan 10 data testing yang telah dipilih. Data testing akan mencari nilai prediksinya menggunakan RapidMiner.

Melakukan pemilihan attributes yaitu untuk mengetahui hasil analisa, *confusion matrix*, dan kurva ROC dari RapidMiner, apakah hasil perhitungan manual sama atau tidak dengan hasil RapidMiner. Pada gambar 3 adalah proses data training menggunakan RapidMiner.



Sumber : (Sofica et al., 2022)

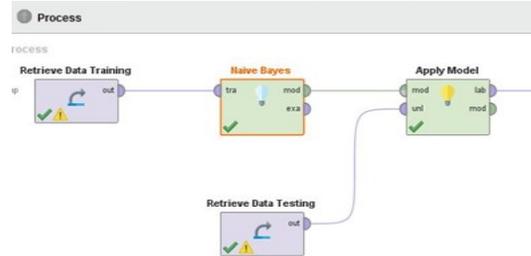
Gambar 3 Proses Data Training Menggunakan Klasifikasi Naïve Bayes



Sumber : (Sofica et al., 2022)

Gambar 4 Hasil Distribusi

Dari gambar 4 mendapatkan hasil jumlah dari simple distribusi kelas bekerja sebanyak 0,625 dan untuk kelas tidak bekerja sebanyak 0,375.



Sumber : (Sofica et al., 2022)

Gambar 5 Proses Memilih Atribut

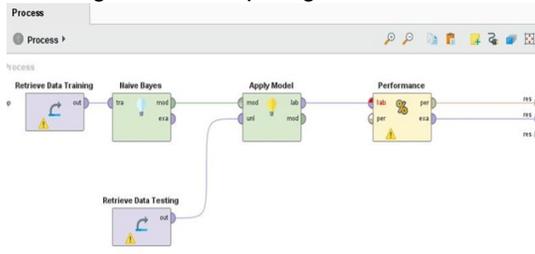
Row No.	Pekerjaan	prediction(P...	confidence...	confidence...	Usia	Pendidikan...	Jenis Kelamin
1	Bekerja	Bekerja	0.735	0.265	40 <	S1 =>	Perempuan
2	Tidak Bekerja	Bekerja	0.541	0.459	40 =>	S1 <	Laki-Laki
3	Bekerja	Bekerja	0.603	0.397	40 <	S1 <	Laki-Laki
4	Bekerja	Bekerja	0.735	0.265	40 <	S1 =>	Perempuan
5	Tidak Bekerja	Bekerja	0.787	0.213	40 =>	S1 =>	Laki-Laki
6	Tidak Bekerja	Tidak Bekerja	0.470	0.530	40 <	S1 <	Perempuan
7	Bekerja	Bekerja	0.541	0.459	40 =>	S1 <	Laki-Laki
8	Bekerja	Bekerja	0.787	0.213	40 =>	S1 =>	Laki-Laki
9	Tidak Bekerja	Tidak Bekerja	0.470	0.530	40 <	S1 <	Perempuan
10	Bekerja	Bekerja	0.603	0.397	40 <	S1 <	Laki-Laki

Sumber : (Sofica et al., 2022)

Gambar 6 Hasil Prediksi Menggunakan RapidMiner

Proses klasifikasi oleh RapidMiner memakai metode Klasifikasi *Naïve Bayes* dalam data pengangguran ini untuk mencocokkan *data testing* dengan *data training* yang telah ditemukan sebelumnya.

Pada tampilan *process* masukan masing-masing *data training* dan *data testing*, selanjutnya masukan operator *Naïve Bayes*, *Apply Model*, dan *Performance* lalu sambungkan kabel seperti gambar 7 :



Sumber : (Sofica et al., 2022)
 Gambar 7 Proses Prediksi *Confusion Matrix*

accuracy: 80.00%

	true Bekerja	true Tidak Bekerja	class precision
pred. Bekerja	6	2	75.00%
pred. Tidak Bekerja	0	2	100.00%
class recall	100.00%	50.00%	

Sumber : (Sofica et al., 2022)
 Gambar 8 Hasil Akurasi Data Testing

precision: 100.00% (positive class: Tidak Bekerja)

	true Bekerja	true Tidak Bekerja	class precision
pred. Bekerja	6	2	75.00%
pred. Tidak Bekerja	0	2	100.00%
class recall	100.00%	50.00%	

Sumber : (Sofica et al., 2022)
 Gambar 9 Hasil Presisi Data Testing

recall: 50.00% (positive class: Tidak Bekerja)

	true Bekerja	true Tidak Bekerja	class precision
pred. Bekerja	6	2	75.00%
pred. Tidak Bekerja	0	2	100.00%
class recall	100.00%	50.00%	

Sumber : (Sofica et al., 2022)
 Gambar 10 Hasil Recall Data Testing

Hasil pengukuran data akurasi yang diperoleh dari *data training* mencapai 80.00%. Jumlah

prediksi bekerja yang diklasifikasikan sebagai bekerja oleh *classifier* yaitu 6 data, dan jumlah prediksi tidak bekerja yang diklasifikasikan sebagai bekerja oleh *classifier* yaitu 0.

Sedangkan jumlah prediksi bekerja yang diklasifikasikan sebagai tidak bekerja oleh *classifier* yaitu 2 data, dan jumlah prediksi tidak bekerja yang diklasifikasikan sebagai tidak bekerja oleh *classifier* yaitu 2 data. Dengan pencapaian kelas presisi true tidak bekerja 100.00%. Untuk kelas *recall* dengan true tidak bekerja mencapai 50,00%.

Proses klasifikasi dengan menggunakan metode *Naïve Bayes* pada data pengangguran ini untuk membandingkan *data testing* dengan *data training* yang sudah diketahui sebelumnya. Berikut hasil *confusion matrix* secara manual sebagai berikut :

$$\text{Akurasi} = \frac{6+2}{6+2+0+2} * 100$$

$$\text{Akurasi} = \frac{8}{10} * 100$$

$$\text{Akurasi} = 80\%$$

Hasil dari perhitungan akurasi secara manual mendapatkan 80%.

$$\text{Persisi} = \frac{6}{6+2} * 100$$

$$\text{Presisi} = \frac{6}{8} * 100$$

$$\text{Presisi} = 75\%$$

Hasil dari perhitungan presisi secara manual mendapatkan 75%

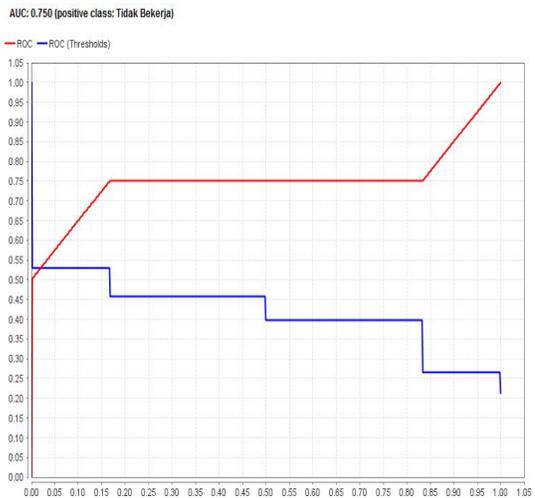
$$\text{Recall} = \frac{6}{6+0} * 100$$

$$\text{Recall} = \frac{6}{6} * 100$$

$$\text{Recall} = 100\%$$

Hasil dari perhitungan recall secara manual mendapatkan 100%

Nilai AUC yang didapatkan dari pengujian berdasarkan kurva ROC menggunakan metode *Naïve bayes* sebesar 0.750, sehingga dari hasil tersebut berdasarkan kriteria diatas menunjukkan klasifikasi yang dihasilkan termasuk kedalam kelompok *fair classification* atau klasifikasi yang sangat baik. Hasil kurva ROC dapat dilihat pada gambar 11.



Sumber : (Sofica et al., 2022)
Gambar 11 Hasil kurva ROC

Untuk mengetahui *performance vector* yang didapat, akan diterangkan pada gambar 12 berikut ini :

PerformanceVector

```
PerformanceVector:
accuracy: 80.00%
ConfusionMatrix:
True: Bekerja Tidak Bekerja
Bekerja: 6 2
Tidak Bekerja: 0 2
precision: 100.00% (positive class: Tidak Bekerja)
ConfusionMatrix:
True: Bekerja Tidak Bekerja
Bekerja: 6 2
Tidak Bekerja: 0 2
recall: 50.00% (positive class: Tidak Bekerja)
ConfusionMatrix:
True: Bekerja Tidak Bekerja
Bekerja: 6 2
Tidak Bekerja: 0 2
AUC (optimistic): 0.792 (positive class: Tidak Bekerja)
AUC: 0.750 (positive class: Tidak Bekerja)
AUC (pessimistic): 0.708 (positive class: Tidak Bekerja)
```

Sumber : (Sofica et al., 2022)
Gambar 12 Performa Vector

IV. KESIMPULAN

Dari hasil penelitian yang telah dilakukan terhadap pengangguran pada Desa Bojong Kulu dengan menggunakan metode Klasifikasi *Naïve Bayes*, dapat ditarik kesimpulan sebagai berikut:

1. Metode Klasifikasi *Naïve Bayes* yang digunakan memberikan proses seleksi yang cepat dan metode yang mudah dipahami dengan tingkat akurasi yang tidak dapat disangkal.
2. Metode Klasifikasi *Naïve Bayes* menggunakan data *training* untuk menciptakan probabilitas setiap kriteria untuk berbagai kelas yang berbeda, dengan tujuan agar nilai-nilai

probabilitas dari kriteria ini dapat ditingkatkan untuk memprediksi status pengangguran berdasarkan proses klasifikasi yang dilakukan dengan metode Klasifikasi *Naïve Bayes*.

3. Dari pengujian yang dilakukan dengan aplikasi pendukung Rapidminer didapatkan akurasi sebesar 80%, presisi sebesar 100%, dan recall sebesar 50%.

4. *Naïve Bayes* dapat digunakan untuk mengklasifikasi pengangguran di Desa Bojong Kulu.

V. REFERENSI

- Dongoran, F. R., Nisa, K., Sihombing, M., & Purba, L. D. (2016). Analisis Jumlah Pengangguran dan Ketenagakerjaan Terhadap Keberadaan Usaha Mikro Kecil dan Menengah di Kota Medan. *Jurnal EduTech*, 2, 59–72. http://jurnal.umsu.ac.id/index.php/edutech/article/view/599/pdf_26
- Franita, R. (2016). Analisa Pengangguran Di Indonesia. *Nusantara (Ilmu Pengetahuan Sosiologi)*, 1, 88–93. <https://jurnal.um-tapsel.ac.id/index.php/nusantara/article/viewFile/97/97>
- Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. *Physics Procedia*, 18–22. <https://www.sciencedirect.com/science/article/pii/S1875389215000371>
- Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques. In *Soft Computing* (Vol. 54, Issue Second Edition). <https://doi.org/10.1007/978-3-642-19721-5>
- Larose, D. T., & Larose, C. D. (2014). *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining*.
- Mada, M., & Ashar, K. (2015). Analisis Variabel Yang Mempengaruhi Jumlah Pengangguran Terdidik Di Indonesia. *Jurnal Ilmu Ekonomi Pembangunan (JIEP)*, 15, 50–76. <https://jurnal.uns.ac.id/jiep/article/view/9894/8810>
- Mardi, Y. (n.d.). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *2016*, 2, 213–219. <http://ejournal.stkip-pgri-sumbar.ac.id/index.php/eDikInformatika/article/view/1465>

- Putri, A. N. (2017). Penerapan Naive Bayesian Untuk Perankingan Kegiatan Di Fakultas Tik Universitas Semarang. *Simetris Jurnal Tek. Mesin, Elektro Dan Ilmu Komputer*, 8, 603–610.
<https://jurnal.umk.ac.id/index.php/simet/article/view/1545>
- Saleh, A. (2014). Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa. *Konferensi Nasional Ilmu Pengetahuan Dan Teknologi*, 200–208.
<http://repository.potensi-utama.ac.id/jspui/handle/123456789/402>
- Sofica, V., Hasan, N., Septiani, M., & Allo, Y. S. T. (2022). *Penggunaan Metode Naive Bayes Dalam Mengklasifikasi Pengangguran Pada Desa Bojong Kulur*.
- Utama, S. S., Suparti, S., & Rahmawati, R. (2015). *Pemodelan Tingkat Pengangguran Terbuka di Jawa Tengah Menggunakan Regresi Spline*. *Jurnal Gaussian*, 4, 113–122.
<https://ejournal3.undip.ac.id/index.php/gaussian/article/view/8151/7927>
- Wibawa, A. P. (2018). Metode-metode Klasifikasi. *Prosiding SAKTI (Seminar Ilmu Komputer Dan Teknologi Informasi)*, 134–138.
<http://ejournals.unmul.ac.id/index.php/SAKTI/article/view/2101>