

## KLASIFIKASI OPPORTUNITY MENGGUNAKAN ALGORITMA C4.5, C4.5 DAN NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION

Endang Sri Palupi <sup>1)</sup> Said Mirza Pahlevi<sup>2)</sup>

<sup>1)</sup> Sistem Informasi  
Universitas Bina Sarana Informatika  
<http://www.bsi.ac.id>  
[endang.epl@bsi.ac.id](mailto:endang.epl@bsi.ac.id)

<sup>2)</sup> Program Pascasarjana Magister Ilmu Komputer  
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri  
<http://www.nusamandiri.ac.id>  
[smirzap@gmail.com](mailto:smirzap@gmail.com)



**Abstract** - Predicting an opportunity whether to be successful (buy) or not (no), with the aim to increase selling for target achievement of marketing. Marketing is required to find a good opportunity so that it can be a prospect of sales in great value and the long term. In this research, the dataset is taken from a CRM application (Customer Relationship Management) at PT. XYZ, sales of telesales team in January - March 2016. From these results, the PSO-based C4.5 algorithm has the highest accuracy and AUC value. In this research comparative algorithms C4.5, C4.5 based PSO, and Naïve Bayes using Confusion Matrix testing methods and ROC curves. The highest accuracy value using PSO-based C4.5 algorithm is 80,90% with AUC value 0.833 is good classification, next is Naïve Bayes based PSO algorithm with accuracy value equal to 83,15% and value of AUC 0,894 is good classification, the last C4.5 algorithm the lowest accuracy value of 66.67% with AUC 0.592 is failure classification. From these results, the PSO-based C4.5 algorithm has the highest accuracy.

**Keywords** : Algoritma C4.5, Naïve Bayes, Particle Swarm Optimization.

**Abstrak** - Membuat klasifikasi untuk memprediksi opportunity apakah akan membeli atau tidak, dengan tujuan untuk meningkatkan penjualan guna pencapaian target sebagai marketing. Marketing diharapkan menemukan opportunity yang baik sehingga bisa menjadi prospek penjualan yang bernilai besar dalam jangka panjang. Dalam penelitian ini dataset diambil dari aplikasi CRM (Customer Relationship Management) di PT. XYZ, data penjualan tim telesales pada Januari - Maret 2016. Dalam penelitian ini komparasi algoritma C4.5, algoritma C4.5 berbasis PSO, dan algoritma Naïve bayes menggunakan metode pengujian Confusion Matrix dan kurva ROC. Dari hasil penelitian ini algoritma C4.5 berbasis PSO memiliki nilai AUC dan akurasi tertinggi, yaitu nilai akurasi 80,90% dan nilai AUC 0,833 termasuk dalam klasifikasi yang baik. Nilai tertinggi kedua adalah algoritma Naïve bayes berbasis PSO dengan nilai akurasi 83,15% dan nilai AUC 0,894 termasuk dalam klasifikasi yang baik, selanjutnya nilai terendah adalah algoritma C4.5 dengan nilai akurasi 66,67% dan nilai AUC 0,592 termasuk dalam klasifikasi kegagalan. Maka hasil penelitian ini algoritma C4.5 berbasis PSO memiliki akurasi paling tinggi.

**Kata kunci** : Algoritma C4.5, Naïve Bayes, Particle Swarm Optimization.

### PENDAHULUAN

Dalam dunia marketing, *opportunity* adalah hal yang sangat penting dalam database nama-nama calon pelanggan yang harus di *follow up*. *Opportunity* merupakan pelanggan yang tertarik atau ada kebutuhan barang/jasa yang kita tawarkan sehingga menunjukkan potensial pelanggan untuk melakukan pembelian barang/jasa yang kita tawarkan. Divisi marketing

pada umumnya mempunyai jadwal meeting minimal 1 bulan sekali untuk membahas pencapaian penjualan selama ini, bagaimana rencana penjualan ke depan untuk mencapai target dan salah satunya adalah memasukkan nama pelanggan yang berniat membeli atau ada rencana melakukan pembelian dalam waktu dekat. Banyak faktor yang menentukan suatu *opportunity* akan *closing* (membeli) atau tidak, yaitu : kompetitor, *lock harga/project*, diskon harga dari *principle*,

kedekatan dan kepercayaan dari pelanggan, dan lain sebagainya (Setiawan et al., 2016).

SWOT (*Strengths Weaknesses Opportunity threats*) adalah metode perencanaan strategis yang digunakan untuk mengevaluasi kekuatan, kelemahan, peluang, dan ancaman, dalam suatu proyek atau suatu spekulasi bisnis. Aplikasi yang paling banyak digunakan untuk menunjang SWOT dalam mengelola pelanggan pada dunia marketing adalah CRM (*Customer Relation Management*) (So & Sheila, 2011). CRM sebagai sebuah sistem informasi yang terintegrasi yang digunakan untuk merencanakan, menjadwalkan, dan mengendalikan aktivitas – aktivitas pra-penjualan dalam sebuah organisasi. (Pratikto & Wibisono, 2012). CRM juga membantu para marketing mengatur *account* dan *track opportunity* mereka. Semua *opportunity* diinput ke dalam CRM untuk di *follow up* lebih dalam sampai closing (*Purchase Order*).

*Opportunity* ada yang kuat ada juga yang lemah, tidak ada faktor yang benar – benar memastikan sebuah *opportunity* bisa berakhir dengan pembelian atau tidak. Banyak juga *opportunity* yang dianggap kuat ternyata tidak berakhir dengan pembelian, dan sebaliknya *opportunity* yang dianggap lemah ternyata bisa berakhir dengan pembelian. Untuk itu penulis melakukan penelitian untuk memprediksi *opportunity* yang akan membeli (*Purchase Order*) atau tidak.

*Opportunity* atau peluang, menurut kamus besar Bahasa Indonesia adalah kemungkinan. Sedangkan peluang dalam ilmu matematika adalah kejadian yang mungkin terjadi. Analisis lingkungan eksternal dapat membuahkan peluang baru bagi sebuah perusahaan untuk meraih keuntungan dan pertumbuhan (Fauziah, 2008). Beberapa contoh kesempatan tersebut adalah :

- a. Kebutuhan pelanggan yang tidak dipenuhi di pasar
- b. Kedatangan teknologi baru
- c. Pelonggaran peraturan
- d. Penghapusan hambatan perdagangan internasional (Fauziah, 2008)

*Data mining* adalah suatu proses pengambilan pengetahuan menggunakan data dengan jumlah besar yang disimpan dalam basis data, data warehouse atau informasi dalam repository (Han, Kamber, & Pei, 2012). *Data mining* adalah menerapkan metode statistik dan logika untuk mengolah dataset yang sangat besar (Siegel, Naishadham, & Jemal, 2012). *Data mining* adalah proses menggali informasi atau pola dalam data berukuran besar yang sudah ada dalam database untuk keperluan tertentu, salah satu teknik dalam *data mining* adalah klasifikasi yang tujuannya membagi objek untuk ditugaskan hanya ke salah satu nomor kategori yang disebut kelas (Wang &

Bramer, 2007). variable target dari klasifikasi adalah variable kategori (Ratelle, Larose, Guay, & Senécal, 2005).

Beberapa penelitian dalam bidang penjualan marketing untuk mengklasifikasi menggunakan algoritma banyak dilakukan sebelumnya.

Teguh Budi Santoso pada tahun 2015 yang menggunakan algoritma C4.5 untuk memprediksi loyalitas pelanggan. Hasil dari penelitian tersebut adalah nilai akurasi untuk data training sebesar 97,5% dan data testing nilai akurasinya sebesar 93,3% (Santoso, 2015)

Penelitian lain dilakukan oleh Syaeful Mujab pada tahun 2009 yaitu pencarian model terbaik antara algoritma C4.5 dan algoritma C4.5 berbasis PSO untuk prediksi promosi deposito. Hasil dari penelitian tersebut akurasi algoritma C4.5 sebesar 88,83% dan AUC 0.868 sedangkan akurasi C4.5 berbasis PSO sebesar 89,26% dan AUC 0,874 (Mujab, 2013).

Penelitian selanjutnya adalah penerapan algoritma *Naïve bayes* berbasis PSO untuk memprediksi nasabah yang berpotensi membuka simpanan deposito oleh Prabowo dan Muljono yang dikomparasi dengan Algoritma C4.5 dan C4.5 berbasis PSO. Hasilnya akurasi tertinggi menggunakan *Naïve bayes* berbasis PSO dengan nilai 89,70% (Prabowo & Muljono, 2018).

Penerapan *Practicle Swarm Optimization* untuk menyeleksi fitur pada analisis sentimen review perusahaan penjualan online menggunakan *Naïve bayes*. Penelitian yang dilakukan oleh Siti Ernawati ini hasil akurasinya 86,88% dengan nilai AUC 0,705 (Siti, 2016).

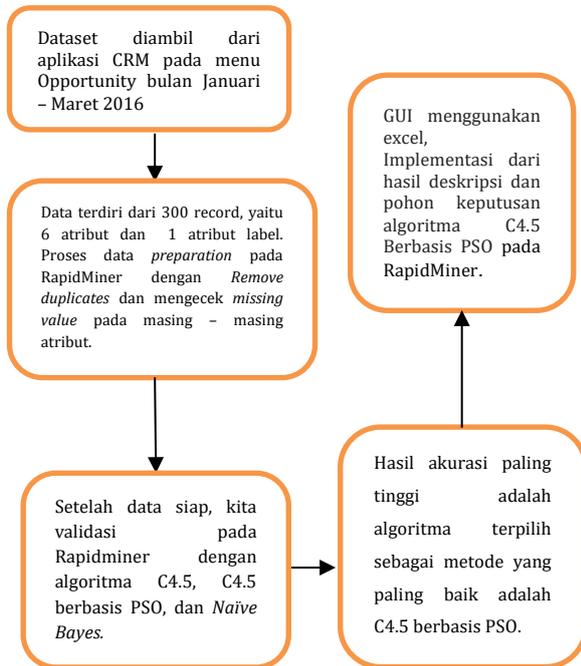
Implementasi data mining dengan *Naïve bayes Classifier* untuk mendukung strategi pemasaran di bagian Humas STMIK Amikon Yogyakarta. Hasil dari uji coba 1000 data testing yang diambil secara acak dengan data training menggunakan aplikasi Smart Marketing untuk prediksi minat studi mendapatkan akurasi sebesar 92,7% dan eror sebesar 7,3% (Saputra & Mukhtar, 2014).

Penulisan dalam penelitian ini akan menggunakan pendekatan metode algoritma C4.5 berbasis PSO (Mujab, 2013) yang berfungsi memilih dan mengoptimasi atribut di saat bersamaan sehingga mendapatkan nilai tingkat akurasi yang lebih baik untuk memprediksi *opportunity* yang akan membeli (*buy*), kemudian penulis juga akan menganalisa komparasi algoritma C4.5, C4.5 berbasis *Particle Swarm Optimization* dan algoritma *Naïve bayes*.

Tujuan dari penelitian ini adalah untuk memprediksi secara akurat *opportunity* yang akan membeli atau tidak membeli dengan menggunakan algoritma C4.5, C4.5 berbasis PSO, dan *Naïve bayes* berbasis PSO.

**BAHAN DAN METODE**

**1. Rancangan Penelitian**



Sumber : (Palupi & Pahlevi, 2019)  
Gambar 1. Kerangka Kerja Penelitian

**2. Pengambilan Data dan Teknik Pengambilan Data**

Data diambil dari aplikasi CRM (*Customer Relation Management*) team telesales PT. XYZ menu opportunity pada bulan Januari - Maret 2016 sebanyak 300 dataset yang selama ini di input ketika ada *Opportunity*. Data akan di export ke dalam Excel.

\* My Open Opportunities \*

Opportunity ID	Opportunity Sub-Topic	Est. Close Date #	Est. Close Quarter	Est. Revenue	Amount Available	Amount Remaining	Open Amount	Contact	Account	Industry
7012004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	WALDI SULA	MANULAND	Vegetation
7022004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	AMAL SUDARNO	TERANG	Vegetation
7007004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	ALIF	WARTANA SENT.	Vegetation
7017004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	ABRIANTHUS	INDONESIA S.C.	Vegetation
7027004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	TENDI	DAK PERKOTA	Vegetation
7036004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	AGUS SUDARNO	SUPERVAL KAL.	Vegetation
7014004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	HAZRI	PT. MITSUB.	Vegetation
7039004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	WALYUDI	DEWILATI 79	Vegetation
7026004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	RENY	INDONESIA KOD.	Vegetation
7027004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	ALYANESS	ARISTOKRASI	Vegetation
7012004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	ANGGA	INDONESIA BAK.	Vegetation
7014004002	Trade	12/31/2016	2016-Q4	\$2,500,000.00	\$0.00	\$0.00	\$0.00	WALDI SULA	WALDI SULA	Vegetation

Sumber : (Palupi & Pahlevi, 2019)  
Gambar 2. Contoh Opportunity pada Aplikasi CRM

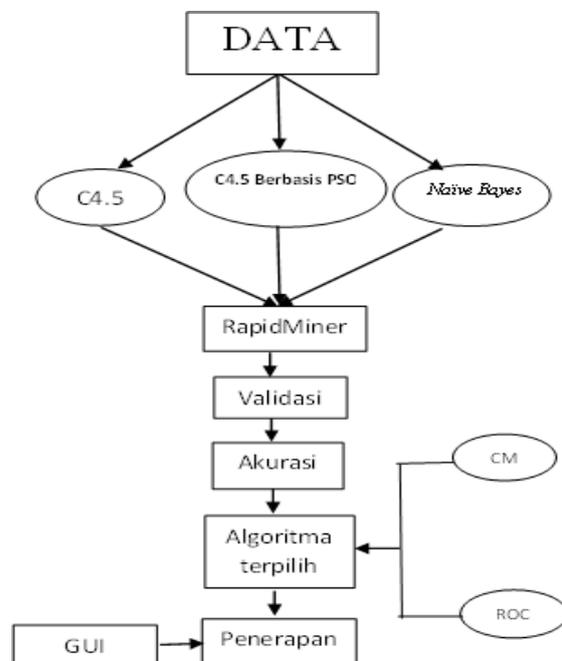
**3. Metode Penelitian**

CRISP-DM adalah standarisasi *data mining* yang disusun oleh tiga pengagagas *data mining* market. Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR. Kemudian dikembangkan pada berbagai workshops (antara 1997-1999). Lebih

dari 300 organisasi yang berkontribusi dalam proses modelling ini dan akhirnya CRISP-DM 1.0 dipublikasikan pada 1999.

CRISP-DM merupakan singkatan dari *Cross-Industry Standard Proses for Data Mining*, secara umum menjelaskan tentang proses *data mining* dalam enam tahap. Proses ini salah satu tujuannya untuk menemukan pola yang menarik dan bermakna dalam data. Serta melibatkan beberapa disiplin ilmu seperti *Statistika*, *Machine learning*, *Artificial Intelligence*, *Pattern Recognition*, dan *Data Mining* (Budiman, Prahasto, & Christyono, 2014).

*Cross-Industry Standard Proses for Data Mining (CRISP-DM)* dikembangkan pada tahun 1996 oleh analis dari beberapa industri. CRISP-DM menyediakan standard proses data mining sebagai pemecahan masalah secara umum dari bisnis atau unit penelitian (Hand, 2008).



Sumber : (Palupi & Pahlevi, 2019)  
Gambar 3. Langkah - langkah Penelitian

**HASIL DAN PEMBAHASAN**

**Pengujian Model dengan Algoritma C4.5**

Dalam klasifikasi ini digunakan pendekatan metode algoritma C4.5. Perhitungan *Entropy* dan *Gain* berdasarkan atribut yang telah ditentukan dengan menggunakan data *training* yang berjumlah 300 *record*, sebanyak 205 *customer* yang membeli (*buy*) atau PO (*Purchase Order*) dan 95 *customer* tidak PO. Berikut nilai entropi nya :

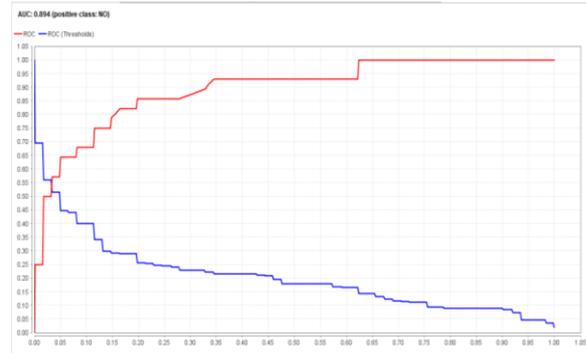
$$Entropy (S) = (-95/300) * \text{IMLOG2}(95/300) + (-205/300) * \text{IMLOG2}(205/300) = 0,090071968$$



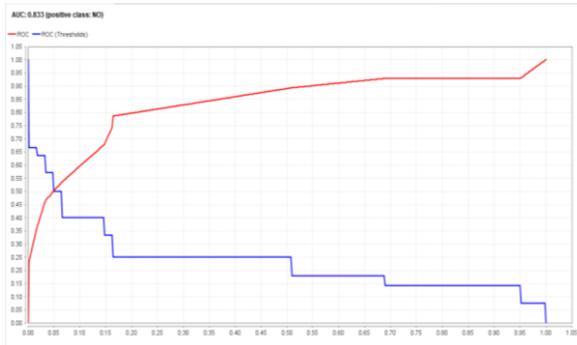
accuracy: 80.9%

	true YES	true NO	class precision
pred YES	59	15	79.73%
pred NO	2	13	86.67%
class recall	96.72%	46.43%	

Sumber : (Palupi & Pahlevi, 2019)  
Gambar 10. Akurasi Metode Algoritma C4.5 Berbasis PSO



Sumber : (Palupi & Pahlevi, 2019)  
Gambar 14. Nilai AUC Algoritma Naive Bayes Berbasis PSO



Sumber : (Palupi & Pahlevi, 2019)  
Gambar 11. Nilai AUC Algoritma C4.5 Berbasis PSO

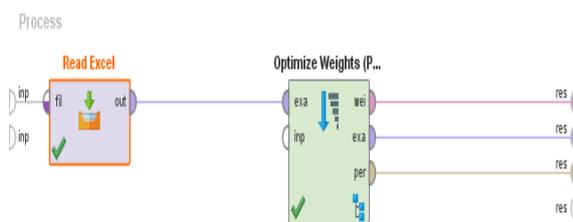
Dari hasil pengujian model *Naive bayes* mendapatkan nilai Area Under Curve sebesar 0.894 termasuk dalam *good classification*.

Dari hasil pengujian model C4.5 berbasis PSO mendapatkan nilai *Area Under Curve* sebesar 0.833 termasuk dalam *good classification*.

Sumber : (Palupi & Pahlevi, 2019)  
Gambar 15. Tampilan Hasil ANNOVA

3. Pengujian Model dengan Algoritma *Naive Bayes* Berbasis PSO.

Dari hasil Annova prob sebesar 0.000 artinya significant  $0.000 < \alpha = 0.050$



Sumber : (Palupi & Pahlevi, 2019)  
Gambar 12. Desain Model Algoritma Naive Bayes Berbasis PSO

PREDIKSI OPPORTUNITY		PREDIKSI?
INDUSTRY	GOVERNMENT	YES
DISKON	RENDAH	
JUMLAH PO SEBELUMNYA	BANYAK	
ESTIMASI HARGA	1.000.000.000	
KEBUTUHAN	LICENSE	
RESOURCE	NO	

Sumber : (Palupi & Pahlevi, 2019)  
Gambar 16. Tampilan Prediksi Opportunity

accuracy: 83.15%

	true YES	true NO	class precision
pred YES	58	12	82.86%
pred NO	3	16	84.21%
class recall	95.00%	57.14%	

Sumber : (Palupi & Pahlevi, 2019)  
Gambar 13. Akurasi Metode Algoritma Naive Bayes Berbasis PSO

**KESIMPULAN**

Dalam penelitian ini dilakukan pengujian model dengan membandingkan tiga metode *data mining* yaitu algoritma C4.5, algoritma C4.5 berbasis *Particle Swarm Optimization*, algoritma *naive bayes*. Hasil evaluasi dan validasi, diketahui bahwa algoritma C4.5 berbasis *Particle Swarm Optimization* memiliki nilai akurasi paling tinggi sebesar 80.90% dibandingkan dengan algoritma

*Naive bayes* yang ada pada urutan kedua akurasiya yaitu sebesar 75.67%, dan paling rendah akurasiya algoritma C4.5 dengan akurasi sebesar 66.67%. Dengan demikian, metode C4.5 berbasis *Particle Swarm Optimization* merupakan metode yang cukup baik dalam memprediksi *opportunity* yang akan *Purchase Order (Buy)*.

#### REFERENSI

- Budiman, I., Prahasto, T., & Christyono, Y. (2014). Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma. *JURNAL SISTEM INFORMASI BISNIS*. <https://doi.org/10.21456/vol1iss3pp129-134>
- Fauziah, A. (2008). PELUANG INVESTASI EMAS ..., Anggriani Fauziah, HUKUM EKONOMI SYARIAH FAI, UMP 2016., 14-41.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Hand, D. J. (2008). Data Mining: Methods and Models by D. T. Larose. *Biometrics*. [https://doi.org/10.1111/j.1541-0420.2008.00962\\_9.x](https://doi.org/10.1111/j.1541-0420.2008.00962_9.x)
- Mujab, S. (2013). PENCARIAN MODEL TERBAIK ANTARA ALGORITMA C4.5 DAN C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI PROMOSI DEPOSITO. *Rancang Bangun E-CRM Pada Pasar Murah Solo*, 1, 3-4.
- Palupi, E. S., & Pahlevi, S. M. (2019). *Laporan Akhir Penelitian Mandiri: Klasifikasi Opportunity Menggunakan Algoritma C4.5, C4.5 Dan Naive Bayes Berbasis Particle Swarm Optimization*. Jakarta.
- Prabowo, A. D. R., & Muljono, M. (2018). Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization. *Techno.Com*. <https://doi.org/10.33633/tc.v17i2.1648>
- Pratikto, F. R., & Wibisono, Y. Y. (2012). Analisis Customer Lifetime Value Terhadap Wholesale Customer Perusahaan Telekomunikasi Dengan Memperhitungkan Risiko Potensi Laba. *Research Report - Engineering Science*, 1. Retrieved from <http://journal.unpar.ac.id/index.php/rekayasa/article/view/139>
- Ratelle, C. F., Larose, S., Guay, F., & Senécal, C. (2005). Perceptions of parental involvement and support as predictors of college students' persistence in a science curriculum. *Journal of Family Psychology*. <https://doi.org/10.1037/0893-3200.19.2.286>
- Santoso, T. B. (2015). Analisa Dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan. *CEUR Workshop Proceedings*. <https://doi.org/10.1017/CBO9781107415324.004>
- Saputra, E. H., & Mukhtar, B. A. (2014). Implementasi Data Mining Dengan Naive Bayes Classifier Untuk Mendukung Strategi Pemasaran Di Bagian. *Seminar Nasional Teknologi Informasi Dan Multimedia 2014*.
- Setiawan, H., Minarsih, M. M., Fathoni, A., Jurusan, M., Fakultas, M., Dan, E., ... Ekonomika, F. (2016). PENGARUH KUALITAS PRODUK, KUALITAS PELAYANAN DAN KEPERCAYAAN TERHADAP KEPUASAN NASABAH DAN LOYALITAS NASABAH DENGAN KEPUASAN SEBAGAI VARIABEL INTERVENING ( Studi Kasus Pada Nasabah Koperasi Rejo Agung Sukses Cabang Ngaliyan ). *Journal Of Management*, 2(2). Retrieved from <http://jurnal.unpand.ac.id/index.php/MS/article/view/492>
- Siegel, R., Naishadham, D., & Jemal, A. (2012). Cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*. <https://doi.org/10.3322/caac.20138>
- Siti, E. (2016). Penerapan Particle Swarm Optimization Untuk Seleksi Fitur Pada Analisis Sentimen Review Perusahaan Penjualan Online Menggunakan Naive Bayes. *Evolusi*. <https://doi.org/10.2311/evo.v4i1.276>
- So, I. G., & Sheila, S. (2011). Analisis Perancangan Customer Relationship Management Berbasis Web pada PT ASP Jakarta. *Binus Business Review*, 2(1), 100. <https://doi.org/10.21512/bbr.v2i1.1115>
- Wang, X., & Bramer, M. (2007). Exploring Web search results clustering. In *Research and Development in Intelligent Systems XXIII - Proceedings of AI 2006, the 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*. <https://doi.org/10.1007/978-1-84628-663-6-30>