

# Performa Algoritma *Data Mining* Untuk Klasifikasi Data Perceraian

Elvin Leander Hadisaputro<sup>1\*</sup>, Joy Nashar Utamajaya<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, STMIK Borneo Internasional, Balikpapan, Indonesia

Email: <sup>1</sup>elvin.leander@stmik-borneo.ac.id, <sup>2</sup>joy.nashar@stmik-borneo.ac.id

---

## INFORMASI ARTIKEL

### *Histori artikel:*

Naskah masuk,  
Direvisi,  
Diiterima,

### *Kata Kunci:*

Performa algoritma  
Feature selection  
Klasifikasi  
Data perceraian  
Decision tree  
Naive bayes  
k-nearest neighbour  
neural network  
support vector machine

## ABSTRAK

**Abstract-** Data mining is the process of analyzing data to find patterns from existing data to produce useful information. One type of data mining is classification, which distinguishes a data object from another. There are several algorithms for classification. This study looks for the best performance of data mining algorithms for decision tree classification, naive Bayes, k-nearest neighbors, neural networks and Support Vector Machines. Performance is seen without using feature selection and using feature selection found in the RapidMiner application, namely Backward Elimination and Forward Selection. The data used is data on divorce cases at the Penajam Religious Court from 2018-2020. 2018-2019 data is used as training data and 2020 data is used as test data. The best performance found from the use of test data is the naive Bayes algorithm that uses Backward Elimination with an accuracy value of 81.72% and AUC 0.691.

**Abstrak-** Data mining merupakan proses analisis data untuk menemukan pola dari data yang ada untuk menghasilkan informasi yang bermanfaat. Salah satu jenis data mining ialah klasifikasi, dimana klasifikasi membedakan suatu objek data dengan yang lain. Terdapat beberapa algoritma untuk klasifikasi. Penelitian ini mencari performa terbaik dari algoritma data mining klasifikasi *decision tree*, *naïve bayes*, *k-nearest neighbour*, *neural network*, dan *Support Vector Machine*. Performa dilihat tanpa menggunakan *feature selection* dan dengan menggunakan *feature selection* yang ditemukan pada aplikasi RapidMiner yaitu *Backward Elimination* dan *Forward Selection*. Data yang digunakan adalah data kasus perceraian di Pengadilan Agama Penajam dari tahun 2018-2020. Data tahun 2018-2019 digunakan sebagai data latih dan data tahun 2020 digunakan sebagai data uji. Performa terbaik yang ditemukan dari penggunaan data uji adalah algoritma *naïve bayes* yang menggunakan *Backward Elimination* dengan nilai *accuracy* 81,72% dan AUC 0,691.

Copyright © 2019 LPPM - STMIK IKMI Cirebon  
This is an open access article under the CC-BY license

---

### *Penulis Korespondensi:*

**Elvin Leander Hadisaputro**

Program Studi Sistem Informasi,  
STMIK Borneo Internasional

Jl. AW. Syahrani No.04, RT.32, Batu Ampar, Kota Balikpapan, Indonesia

Email: elvin.leander@stmik-borneo.ac.id

## 1. Pendahuluan

Perceraian merupakan sebuah permasalahan dalam keluarga/ kependudukan yang sebenarnya sangat ingin dihindari oleh pemerintah. Secara umum tidak ada seorang pun yang menginginkan perceraian, karena pada hakikatnya pernikahan adalah usaha sepasang manusia untuk membentuk keluarga yang harmonis. Meskipun diperbolehkan secara hukum, namun perceraian tetap dianggap sebagai masalah sosial. Dilihat dari prosentase kasus perceraian di Kabupaten Penajam Paser Utara, yang ternyata cenderung meningkat setiap tahunnya

Adanya data digital kasus perceraian pada Pengadilan Agama Penajam di Kabupaten Penajam Paser Utara, sebenarnya dapat dipergunakan untuk menggali lebih dalam informasi mengenai perceraian yang terjadi, agar pemerintah kabupaten dapat merencanakan program pembimbingan dan peningkatan kualitas keluarga, dengan memanfaatkan informasi klasifikasi perceraian. Informasi yang akurat dapat diperoleh dengan metode data mining.

Data mining merupakan proses penggalian dan penambangan pengetahuan dari data yang besar dengan menggunakan teknik-teknik matematis, kecerdasan buatan dan pembelajaran mesin[1][2]. Data mining dalam penerapannya dapat berupa diantaranya: klusterisasi, klasifikasi, asosiasi, neural network, dan algoritma genetik. Klasifikasi merupakan suatu proses pemodelan untuk menjelaskan dan membedakan kelas atau label suatu objek data[3].

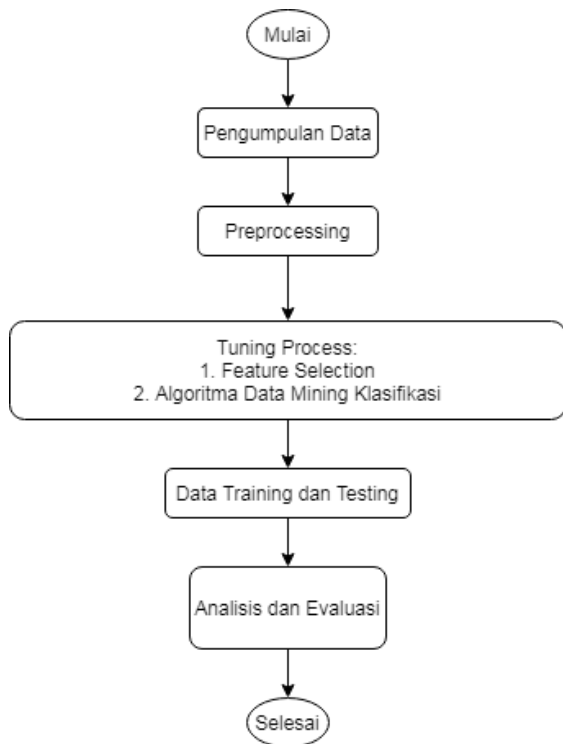
Penelitian yang dilakukan oleh Chandani et.al mengkomparasi algoritma klasifikasi *Artificial Neural Network* (ANN), *Naïve Bayes* (NB) dan *Support Vector Machine* (SVM) untuk analisis sentiment review film, penelitian dilakukan dengan membandingkan nilai Accuracy dan AUC dari ketiga algoritma tersebut. Hasil penelitian ini menemukan bahwa SVM merupakan algoritma yang paling baik untuk pengujian data IMDB review film, dengan nilai accuracy dan AUC tertinggi, serta feature selection terbaik untuk analisis data tersebut adalah feature information gain [4]. Sartika dan Sensuse membandingkan algoritma *Naïve Bayes*, *Nearest Neighbour* dan *Decision Tree* untuk pemilihan pola pakaian dengan metode pengujian *use training test*, *percentage split*, dan *supplied test*, menemukan bahwa algoritma klasifikasi decision tree merupakan algoritma klasifikasi dengan akurasi tertinggi diantara ketiga algoritma[5]. Penelitian dari Darmansyah dan Kusuma menemukan bahwa algoritma *neural network* dengan *feature selection backward elimination* merupakan algoritma dengan akurasi terbaik dibandingkan algoritma *naïve bayes*, *decision tree* dan k-NN dengan feature

selection *Optimize Selection*, *forward selection* dan *backward elimination* pada pada klasifikasi nutrisi balita berdasarkan kesadaran gizi keluarga dan indeks pembangunan manusia[6]. Penelitian dari Windarti dan Suradi menemukan bahwa algoritma *Bayessian Network* memiliki kinerja akurasi, presisi, *recall* dan nilai AUC terbaik dibandingkan algoritma klasifikasi *naïve bayes*, *neural network*, k-NN, *decision tree*, dan SVM untuk prediksi masa studi mahasiswa[7]. Yuniyal dalam penelitiannya menggunakan software WEKA dan melihat hasil dari confusion matrix setiap algoritma menemukan bahwa algoritma decision tree memiliki nilai akurasi tertinggi dibandingkan algoritma *support vector machine* dan *Naïve Bayes* pada dataset penyakit ginjal kronis dan kanker payudara[8]. Penelitian yang dilakukan oleh Sunaryo [9], yang membandingkan algoritma C4.5, Naive Bayes, k-NN, SVM dan Neural Network dalam menentukan website palsu, menemukan bahwa algoritma Neural Network dan SVM memiliki akurasi dan nilai AUC tertinggi. Dari penelitian yang dilakukan oleh Sari Dewi mengenai komparasi algoritma *neural network*, *Naïve Bayes*, pohon keputusan, k-NN, dan *Logistic Regression* menggunakan data pemasaran pada bank, menyimpulkan algoritma neural network yang memiliki akurasi tertinggi dalam memberikan pemecahan untuk permasalahan dalam mengidentifikasi tingkat keberhasilan telemarketing pada pemasaran bank[10].

Berdasarkan penelitian-penelitian terdahulu, ditemukan bahwa didapat algoritma berbeda yang memiliki performa terbaik pada tiap studi penelitian. Penelitian ini bertujuan untuk menemukan performa terbaik diantara algoritma klasifikasi *decision tree*, *naïve bayes*, k-NN, *neural network*, dan SVM untuk klasifikasi pada data perceraian di Pengadilan Agama Penajam.

## 2. Metode

Penelitian ini dilakukan dalam 5 tahapan seperti yang ditunjukkan dalam gambar 1. Tahap pertama adalah pengumpulan data. Data dalam penelitian ini berasal dari data kasus perceraian di Pengadilan Agama Penajam. Data yang digunakan merupakan data kasus perceraian tahun 2018 sampai dengan tahun 2020.



Gambar 1 Metodologi Penelitian

Tahap ketiga dilakukan proses *Tuning* dimana dicari model *feature selection* dan algoritma data mining terbaik[6].

Tahap keempat dilakukan *training data* dan *testing data* dengan algoritma data mining dan *feature selection* yang telah dipilih. *Training data* dilaksanakan dengan data kasus perceraian tahun 2018-2019, dan *testing data* dilakukan menggunakan data kasus perceraian tahun 2020.

Tahap kelima dilakukan analisis dan evaluasi dari hasil pengujian. Analisis dan evaluasi performa algoritma akan difokuskan pada nilai *accuracy* dan *AUC (area under the curve)*. *Accuracy* diperoleh dari nilai pada *Confusion Matrix* yang dihasilkan. *AUC* didapat dari *ROC Curves*[11].

### 3. Hasil dan Pembahasan

Penelitian yang dilakukan menggunakan komputer dengan spesifikasi CPU Intel Core i7 3.40GHz, RAM 8GB, dan sistem operasi Microsoft Windows 10 Enterprise. Aplikasi yang digunakan adalah RapidMiner Studio 9.10.

Hasil tuning mendapatkan parameter model terbaik untuk setiap algoritma data mining dengan *feature selection* masing-masing ditunjukkan pada tabel 1.

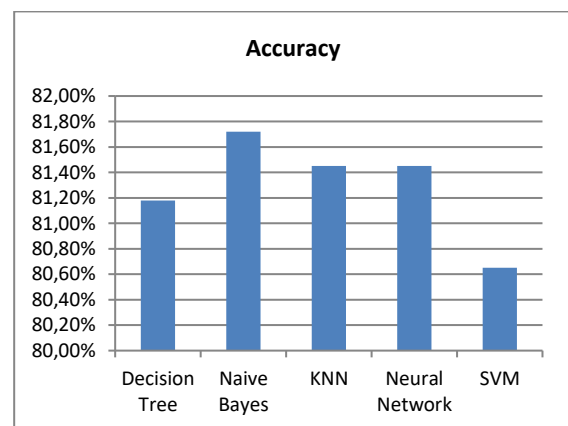
Tabel 1 Hasil Tuning

	Decision Tree	Naive Bayes	k-NN	Neural Network	SVM
Feature Selection	Backward Elimination	Backward Elimination	Forward Selection	Backward Elimination	-
Parameter	criterion gain_ratio max_depth 20 confidence 0.1 min_gain 0.01 min_leaf_size 4	laplace_correction TRUE TRUE	k 1 weighted_vote yes measure_types MixedMeasures	momentum 0.1 learning_rate 0.3	kernel type dot C 0

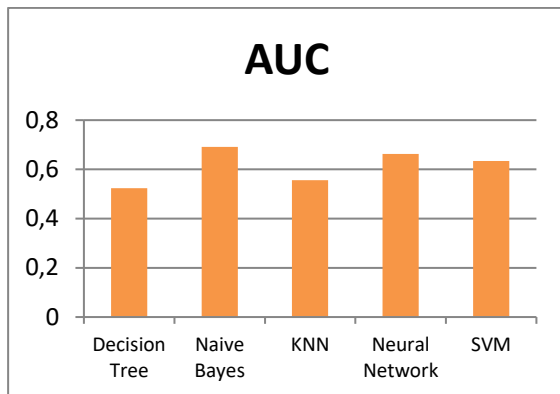
Dengan menggunakan *feature selection* dan parameter model dari tabel 1, kemudian dilakukan *training data* menggunakan data kasus perceraian tahun 2018-2019, dan pengujian data menggunakan data tahun 2020. Tabel 2 merupakan rangkuman hasil performa data testing algoritma klasifikasi.

Tabel 2 Performa Data Testing Algoritma Klasifikasi

	Accuracy	AUC
Decision Tree	81,18%	0,524
Naive Bayes	81,72%	0,691
k-NN	81,45%	0,556
Neural Network	81,45%	0,663
SVM	80,65%	0,634



Gambar 2 Performa Accuracy Algoritma Klasifikasi



Gambar 3 Performa AUC Algoritma Klasifikasi

Berdasarkan Tabel 2, Gambar 2 dan Gambar 3 didapat hasil terbaik adalah *Naïve Bayes* dengan *accuracy* = 81,72% dan AUC = 0,691. Hal ini berbeda dengan penelitian Vinita Chandani [4] dan Sunaryono [9] yang menemukan SVM sebagai algoritma terbaik, atau penelitian dari Darmansyah [6] dan Sari Dewi [10] yang menemukan algoritma *neural network* sebagai algoritma terbaik. Untuk penguian data perceraian di Pengadilan Agama Penajam, *Naïve Bayes* dengan *backward elimination* merupakan algoritma yang paling baik. Performa klasifikasi sangat bergantung pada data latih dan data yang diuji.

#### 4. Kesimpulan

Hasil dari penelitian menunjukkan untuk klasifikasi data perceraian, diantara algoritma *Decision Tree*, *Naïve Bayes*, *k-Nearest Neighbor* (k-NN), *Neural Network*, dan *Support Vector Machine* (SVM), didapatkan bahwa *Naïve Bayes* dengan *Backward Elimination* memberikan performa terbaik dengan nilai *accuracy* = 81,72% dan nilai AUC = 0,691. Hasil penilaian performa sangat dipengaruhi dengan jumlah data pada dataset yang digunakan dalam penelitian.

#### Ucapan Terima kasih

Artikel jurnal ini ditulis berdasarkan penelitian Klasifikasi Data Perceraian di Kabupaten Penajam Paser Utara yang dibiayai oleh Kementerian Riset dan Teknologi / Badan Riset dan Inovasi Nasional melalui Hibah Penelitian Dasar dan Pembinaan

/Kapasitas Tahun Tunggal Tahun Anggaran 2021. Isi sepenuhnya menjadi tanggung jawab penulis.

#### Daftar Pustaka

- [1] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti.," *J. Online Inform.*, vol. 2, no. 1, pp. 36–41, 2017.
- [2] M. J. Zaki and M. J. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, 2014.
- [3] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining concepts and techniques*, 3rd ed. Elsevier, 2012.
- [4] V. Chandani, R. S. Wahono, and Purwanto, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film," *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, 2015.
- [5] D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, pp. 151–161, 2017.
- [6] Darmansyah and G. P. Kusuma, "Data Mining Performance of Toddler Nutrition Classification Based on Family Nutrition Awareness and Human Development Index," *Int. J. Recent Technol. Eng.*, vol. 8, no. 5, pp. 1591–1596, 2020.
- [7] M. Windarti and A. Suradi, "Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa," *Telematika*, vol. 12, no. 1, p. 14, 2019.
- [8] A. H. Yunial, "ANALISA PERBANDINGAN ALGORITMA KLASIFIKASI SUPPORT VECTOR MACHINE , DECESSION TREE DAN NAIVE BAYES," vol. 5, pp. 169–185, 2020.
- [9] Sunaryono, "Penelitian Komparasi Algoritma Klasifikasi dalam Menentukan Website Palsu," *Teknikom*, vol. 1, no. 1, pp. 1–12, 2017.
- [10] S. Dewi, "Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *Techno Nusa Mandiri*, vol. XIII, no. 1, pp. 60–66, 2016.
- [11] J. Han and M. Kamber, *Data Mining Concepts and Theory*, 2nd ed. Moran Kaufmann, 2006.