

## ANALISIS DAN IMPLEMENTASI KLASIFIKASI DATA DENGAN MENGGUNAKAN ALGORITMA AVERAGED ONE-DEPENDENCE ESTIMATORS (AODE)

Liska Yunika<sup>1</sup>, Arie Ardiyanti Suryani<sup>2</sup>, Intan Nurma Yulita<sup>3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

---

### Abstrak

Perkembangan teknologi informasi yang sangat pesat menyebabkan pertumbuhan data menjadi sangat cepat. Dari sekumpulan data tersebut dapat digali nilai tambah atau pengetahuan melalui proses data mining, salah satunya melalui proses klasifikasi. Salah satu algoritma klasifikasi yaitu algoritma Averaged One-Dependence Estimator (AODE). AODE merupakan metode classifier yang berdasarkan probabilitas dan Teorema Bayesian dimana setiap atribut bergantung kepada kelas dan atribut tunggal lainnya. Algoritma AODE mengurangi kerugian dalam hasil akurasi akibat penggunaan asumsi atribut saling bebas pada Naive Bayes. Namun, algoritma AODE hanya dapat menangani data bertipe kategorikal. Pada tugas akhir ini, dilakukan implementasi klasifikasi data bertipe kontinu pada AODE dengan melakukan preprocessing diskritisasi dan perhitungan probabilitas dengan Gaussian Density Function. Pengujian akan dilakukan dengan membandingkan performansi pemrosesan atribut kontinu dengan diskritisasi dalam AODE atau menggabungkan diskritisasi dan Gaussian Density Function dalam AODE (Hybrid AODE). Parameter yang dianalisis adalah akurasi dengan f-measure dan waktu klasifikasi. AODE menunjukkan akurasi yang lebih baik pada data yang tidak distribusi normal, sedangkan Hybrid AODE menunjukkan akurasi yang lebih baik pada data yang terdistribusi normal. Waktu klasifikasi yang dibutuhkan oleh Hybrid AODE cenderung lebih rendah daripada AODE.

**Kata Kunci :** AODE, averaged one-dependence estimator, bayes, classifier, diskritisasi, gaussian density function, hybrid AODE, klasifikasi, probabilitas.

---

### Abstract

The development of information technology very rapidly causing a very rapid data growth. From the collection of such data can be extracted or added value of knowledge through data mining process, one through the classification process. One of the classification algorithm is an algorithm Averaged One-Dependence Estimator (AODE). AODE classifier is a probability-based methods and Bayesian theorem where each attribute depends on the class and other single attribute. AODE algorithm reduces the loss in accuracy results from use of the assumption of independent attributes on Naive Bayes. However, AODE algorithm can only handle discrete data type. In this thesis, carried out the implementation of continuous classification data type on AODE to perform preprocessing discretization and calculation with a Gaussian probability density function. Testing will be done by comparing the processing performance with the discretization of continuous attributes in the AODE or combining discretization and Gaussian Density Function in the AODE (Hybrid AODE). The parameters analyzed were the f-measure accuracy and classification time. AODE showed better accuracy on data that are not normally distributed, while the Hybrid AODE showed better accuracy in the data normally distributed. Classification time required by Hybrid AODE tends to be lower than the AODE.

**Keywords :** AODE, averaged one-dependence estimator, bayes, classifier, discretization, gaussian density function, hybrid AODE, classification, probability.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Perkembangan yang pesat dalam bidang teknologi informasi menyebabkan terjadinya akumulasi data dalam jumlah yang sangat besar atau problem *data explosion*. Di dalam data yang berukuran besar tersebut sebenarnya terkandung berbagai informasi yang berguna untuk pengambilan keputusan. Namun, untuk menggali informasi tersebut tentunya akan sulit dilakukan secara manual. Kebutuhan inilah yang melatarbelakangi hadirnya data mining. Data mining merupakan proses ekstraksi informasi atau pola penting atau menarik dari data dengan jumlah besar [9]. Salah satu teknik yang terdapat pada data mining adalah klasifikasi.

Klasifikasi adalah suatu fungsi pembelajaran yang memetakan sebuah unsur (item) data ke dalam salah satu dari beberapa kelas yang telah didefinisikan. Klasifikasi akan menghasilkan suatu model untuk memprediksi kelas atau kategori dari objek-objek di dalam basisdata. Model tersebut yang akan digunakan untuk memprediksi kelas dari objek-objek yang kelasnya belum diketahui [9]. Terdapat banyak algoritma yang tersedia untuk teknik klasifikasi, salah satu yang banyak digunakan adalah algoritma Naïve Bayes Classifier.

Naïve Bayes merupakan metode *classifier* yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap atribut bersifat bebas (*independence*) [3]. Namun, kondisi atribut *independent* pada Naïve Bayes sangat jarang terjadi pada dunia nyata. Sehingga, berkembanglah algoritma AODE untuk mengatasi kelemahan dalam penanganan ketergantungan atribut. Pada algoritma AODE, setiap atribut bergantung kepada kelas dan atribut tunggal lainnya. Algoritma AODE dapat mengurangi kerugian dalam hasil akurasi akibat penggunaan asumsi atribut saling bebas. Itu berarti, AODE memiliki tingkat akurasi tinggi dan *error rate* yang rendah dalam klasifikasi. Selain itu, AODE memiliki *training* dan *classification efficiency* yang tinggi [10]. Namun, AODE hanya dapat diimplementasikan pada data beratribut kategorikal. Sehingga, muncullah algoritma Hybrid AODE yang dapat menangani atribut kontinu dengan penggabungan diskritisasi dan Gaussian Density Function. Oleh karena itu, pada Tugas Akhir ini, metode yang digunakan dan dianalisis dalam pembangunan klasifikasi adalah algoritma Averaged One-Dependence Estimator (AODE).

### 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka permasalahan yang akan dibahas adalah :

1. Bagaimana mengimplementasikan algoritma Averaged One-Dependence Estimators (AODE) dan Hybrid AODE dalam pengklasifikasian data?

2. Bagaimana perilaku algoritma AODE terhadap performansi pengklasifikasian data berdasarkan perbedaan jenis atribut?
3. Bagaimana performansi pemrosesan atribut kontinu dengan diskritisasi menggunakan dalam AODE?
4. Bagaimana performansi pemrosesan atribut kontinu dengan penggabungan diskritisasi dan Gaussian Density Function dalam AODE (Hybrid AODE)?

### 1.3 Batasan Masalah

Adapun batasan masalah dalam Tugas Akhir ini adalah :

1. Dataset yang digunakan merupakan dataset pada database UCI Machine Learning Repository dan Keel Dataset Repository.
2. Tidak menangani missing value
3. Tidak menangani data beratribut campuran pada sistem. Data beratribut campuran merupakan data yang mengandung atribut kontinu dan kategorikal dalam satu file. Data yang bisa diproses hanya data beratribut kontinu atau kategorikal secara keseluruhan.
4. Penanganan diskritisasi menggunakan tools WEKA.
5. Untuk analisis hasil, parameter yang digunakan untuk mengevaluasi hasil adalah f-measure serta waktu klasifikasi.

### 1.4 Tujuan

Tujuan yang ingin dicapai dari tugas akhir ini adalah:

1. Melakukan implementasi algoritma Averaged One-Dependence Estimators (AODE) dan Hybrid AODE untuk klasifikasi data.
2. Menganalisis perilaku algoritma AODE terhadap performansi pengklasifikasian data berdasarkan perbedaan jenis atribut.
3. Melakukan implementasi pemrosesan atribut kontinu dengan diskritisasi menggunakan Entropy-Minimum Description Length dalam AODE
4. Melakukan implementasi pemrosesan atribut kontinu dengan penggabungan diskritisasi dan Gaussian Density Function dalam AODE (Hybrid AODE)
5. Melakukan analisis performansi sistem berdasarkan pemrosesan atribut kontinu dalam klasifikasi.

### 1.5 Metodologi Penyelesaian Masalah

Metodologi yang akan digunakan untuk menyelesaikan permasalahan pada tugas akhir ini adalah sebagai berikut:

1. Tahap Studi Literatur

Pada tahap ini dilakukan pencarian dan pembelajaran literatur yang berhubungan dengan permasalahan yang ada. Juga dilakukan pencarian data yang akan digunakan dalam tugas akhir ini.

## 2. Tahap Implementasi

Perancangan dan pembangunan *software* untuk mengimplementasikan klasifikasi dengan algoritma yang digunakan yaitu Averaged One-Dependence Estimators (AODE) dan Hybrid Averaged One Dependence Estimator (Hybrid AODE).

## 3. Tahap Analisis

Pada tahap ini, penulis mempelajari tentang klasifikasi dan melakukan analisis pada algoritma yang akan digunakan yaitu Averaged One-Dependence Estimators (AODE) dan Hybrid Averaged One Dependence Estimator (Hybrid AODE).

## 4. Tahap Penyusunan Laporan Penelitian

Pada tahap ini akan dilakukan pembuatan laporan tugas akhir dan pengumpulan dokumentasi serta lampiran-lampiran yang diperlukan.

### 1.6 Sistematika Penulisan

Penulisan tugas akhir ini dibagi ke dalam 5 bab, antara lain :

1. **Bab 1 Pendahuluan**  
Berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan masalah, metodologi penyelesaian masalah dan sistematika penulisan
2. **Bab 2 Landasan Teori**  
Berisi landasan teori tentang data mining, klasifikasi, algoritma AODE, diskritisasi Entropy-Minimum Length Description Principle, Gaussian Density Function, Laplace-Estimate, algoritma Hybrid AODE dan evaluasi performansi.
3. **Bab 3 Analisa dan Perancangan Sistem**  
Berisi analisa dan perancangan aplikasi yang akan dibuat dengan bahasa permodelan DFD dan PSPEC
4. **Bab 4 Implementasi dan Pengujian**  
Berisi penjelasan mengenai implementasi hasil perancangan, uji coba terhadap sistem, dan analisa perangkat lunak yang dibangun.
5. **Bab 5 Kesimpulan dan Saran**  
Berisi kesimpulan dan saran dari hasil penelitian untuk pengembangan lebih lanjut.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dari hasil penelitian di atas dapat diambil beberapa kesimpulan sebagai berikut:

1. Hasil dari klasifikasi dengan AODE dan Hybrid AODE sangat tergantung pada karakteristik persebaran data itu sendiri.
2. Jika klasifikasi dilakukan terhadap data kontinu yang persebarannya normal, maka akan lebih baik jika data tetap diperlakukan sebagai atribut kontinu. Perhitungan probabilitas antar atribut menggunakan Gaussian Density Function pada Hybrid AODE.
3. Jika klasifikasi dilakukan terhadap data kontinu yang persebarannya tidak merata, maka akan lebih baik jika data tersebut diperlakukan sebagai diskrit dengan preprocessing diskritisasi dan diklasifikasikan dengan AODE.
4. Waktu training pada AODE cenderung lebih singkat dibandingkan Hybrid AODE. Namun, pada data yang memiliki atribut banyak (twonorm dan ringnorm, 22 atribut), Hybrid AODE memiliki waktu lebih singkat dibanding AODE. Hal itu terjadi karena dengan jumlah atribut yang banyak, lamanya proses training yang dibutuhkan AODE meningkat pesat. Sehingga dapat dikatakan bahwa jumlah atribut akan mempengaruhi waktu training.
5. Waktu testing pada Hybrid AODE lebih singkat daripada AODE. Hal itu disebabkan proses testing Hybrid AODE yang hanya membutuhkan mean dan stdev untuk perhitungan probabilitas dengan Gaussian Density Function, sedangkan AODE membutuhkan probabilitas yang banyak berdasarkan semua kombinasi class, superparent dan child.
6. Penurunan jumlah kategori pada atribut akan mengakibatkan kenaikan akurasi. Hal ini disebabkan pada saat data diperlakukan sebagai data kontinu dan didiskritisasi, nilai kategori pada atribut akan menurun dan akurasi pengklasifikasian akan meningkat.

#### 5.2 Saran

1. Dapat dilakukan pengembangan dengan penanganan data campuran.
2. Dapat dilakukan pengembangan AODE untuk menangani data kontinu secara murni tanpa diskritisasi.

## DAFTAR PUSTAKA

- [1] Fayyad, U. M. And Irani, K.B. 1993. *Multi-interval discretization of countinuous- valued attributes for classification learning*. Artificial Intelligence, 13, 1022-1027.
- [2] Flores, M. J., Gamez, J. A., Martinez, A. M., and Puerta, J. M. *GAODE and HAODE: Two Proposal based on AODE to deal with continuous variable*. Spanyol
- [3] G. I., J. Boughton, and Z. Wang (2002). *Averaged One-Dependence Estimator : Preliminary Results*. Australia : Monash University.
- [4] Han, Jiawei and Micheline Kamber. 2006. *Data Mining: Concepts and Techniques* Second Edition, Morgan Kaufmann Publishers, San Fransisco.
- [5] Mancill, P.A. 2010. *An Exploration of Naive Bayesian Classification Augmented with Confidence Intervals*. Washington State University
- [6] Moertini, Veronika S. 2002. *Data Mining Sebagai Solusi Bisnis*. Integral, vol.7 no.1 : 44-56.
- [7] Moran, Stuart., He, Yulan., and Liu, Kecheng. 2009. *Choosing the Best Bayesian Classifier : An Empirical Study*. IAENG International Journal of Computer Science, 36:4.
- [8] Santosa, Budi. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu.
- [9] Tan, Pang-Ning., Steinbach, M., and Kumar, V. 2006. *Introduction of data mining*. Boston : Pearson Education.
- [10] Wang, Qing., Chuan-hua Zhou, and Jian-kui Guo. *Learning Selective Averaged One-Dependence Estimators for Probability Estimation*. China : Anhui University of Technology and Fudan University.
- [11] Webb, G. I., J. Boughton, and Z. Wang. 2005. *Not So Naive Bayes: Aggregating One-Dependence Estimators*. Machine Learning 58(1). Netherlands: Springer, pages 5-24.
- [12] Zheng, Fei., and Webb, G. I. *Efficient Lazy Elimination for Averaged One-Dependence Estimator*. Australia : Monash University.