

# KLASIFIKASI ALAT MUSIK CYMBAL MENGGUNAKAN METODE NAIVE BAYES BERBASIS FORWARD SELECTION

Yudhi Handono Warih<sup>1</sup>, Pulung Nurtantio Andono<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro  
Jl. Nakula 1 No. 5-11 Semarang 50131, Telp. (024) 3520165 Fax : 3569684

E-mail : 1, [maspapu@gmail.com](mailto:maspapu@gmail.com)<sup>2</sup>

---

## ABSTRAK

Alat musik merupakan suatu instrumen yang dibuat atau dimodifikasi untuk tujuan menghasilkan musik. cymbal adalah alat musik tak bernada yang dimainkan dengan cara dipukul. Cymbal merupakan salah satu instrument yang jadi bagian dalam alat musik drum, cymbal sendiri terbuat dari bahan logam dan memiliki beberapa jenis ketebalan, seperti pada penilaian kali ini dataset diambil dari salah satu produsen terbesar di dunia yaitu Meinl Cymbals. Pengklasifikasian cymbal sendiri yang sesuai genre musik cukup rumit dan cenderung subyektif. Saat ini masih sangat sulit untuk mendefinisikan perbedaan tiap genre untuk alat musik tersebut, tetapi ada beberapa ciri khusus yang dapat digunakan untuk pengkategorian genre musik. Karena itu mulai banyak berbagai penelitian dan pendekatan klasifikasi secara otomatis yang telah diusulkan dalam beberapa tahun terakhir. Pada penelitian kali ini selain bertujuan mendapatkan nilai akurasi yang baik juga bertujuan mendapatkan model atribut dengan cara menerapkan Feature Selection. Feature Selection adalah salah satu cara untuk menentukan atribut yang paling berpengaruh di dalam dataset. Feature Selection berperan memilih subset yang tepat dari set fitur asli, karena tidak semua fitur/atribut relevan dengan masalah. Bahkan beberapa dari fitur atau atribut tersebut mengganggu dan dapat mengurangi akurasi. Dalam hal ini Naive Bayes memanfaatkan fungsi seleksi fitur dari Forward Selection untuk pemilihan atribut data dengan karakteristik data itu sendiri, dan meningkatkan ketepatan klasifikasi Naive Bayes. Forward Selection berbasis Naive Bayes lebih akurat dan efektif dalam mengklasifikasikan genre musik dari dataset yang bersifat Class Imbalance dengan data yang besar dengan hasil akurasi 93.43% dan termasuk dalam kategori "Kappa excellent". Dibanding dengan menggunakan algoritma Naive Bayes saja dengan hasil akurasi hanya 89.30%.

**Kata kunci:** Alat musik Cymbal, Data Mining, Klasifikasi, Naive Bayes, Forward Selection.

## **ABSTRACT**

*Musical instrument is an instrument that is created or modified for the purpose of generating music. cymbals are not pitched musical instrument played by striking. Cymbal is one instrument that is a part in the musical instrument drum, cymbal itself is made of metal and have some kind of thickness, such as on the assessment of this dataset is taken from one of the largest manufacturers in the world, Meinl Cymbals. Own cymbal appropriate classification genre of music is quite complicated and tends subjective. While this is still very difficult to define the difference of each genre for the instrument, but there are some special features that can be used for categorization of music genres. Since it started many studies and automatically classification approach that has been proposed in recent years. In the present study aims to get a value other than good accuracy also aims to get the model attributes by implementing the Feature Selection. Feature Selection is one way to determine the most influential attributes in the dataset. Feature Selection instrumental selecting the appropriate subset of the original feature set, because not all features / attributes relevant to the problem. Even some of the features or attributes are annoying and can reduce accuracy. In this case Naive Bayes utilize the functionality of Forward Selection feature selection for selecting data attributes to the characteristics of the data itself, and improve the accuracy of classification based Selection Bayes. Forward Naïve Bayes Naive more accurate and effective in classifying musical genres from datasets that are Class Imbalance data large with the result of 93.43% accuracy and are included in the category of "Kappa excellent". Compared with Naive Bayes algorithm using only the results of only 89.30% accuracy.*

**Keywords:** *Musical instruments Cymbals , Data Mining , Classification , Naive Bayes , Forward Selection.*

## **1. PENDAHULUAN**

Alat musik merupakan suatu instrumen yang dibuat atau dimodifikasi untuk tujuan menghasilkan musik. Pada prinsipnya, segala sesuatu yang memproduksi suara, dan dengan cara tertentu bisa diatur oleh musisi, dapat disebut sebagai alat musik. Alat musik pukul menghasilkan suara sewaktu dipukul atau ditabuh. Alat musik pukul dibagi menjadi dua yakni bernada dan tidak bernada. Bentuk dan bahan bagian-bagian instrumen serta bentuk rongga getar, jika ada, akan menentukan suara yang dihasilkan instrumen [1]. Contohnya adalah kolintang (bernada), drum (tak bernada), dan cymbal (tak bernada).

Pengklasifikasian simbal sendiri yang

sesuai genre musik cukup rumit dan cenderung subyektif [3]. Saat ini masih sangat sulit untuk mendefinisikan perbedaan tiap genre untuk alat musik tersebut, tetapi ada beberapa ciri khusus yang dapat digunakan untuk pengkategorian genre musik [4] [5]. Karena itu mulai banyak berbagai penelitian dan pendekatan klasifikasi secara otomatis yang telah diusulkan dalam beberapa tahun terakhir.

Pada penelitian kali ini selain bertujuan mendapatkan nilai akurasi yang baik juga bertujuan mendapatkan model atribut dengan cara menerapkan Feature Selection. Feature Selection adalah salah satu cara untuk menentukan atribut yang paling berpengaruh di dalam dataset. Feature Selection berperan memilih subset yang

tepat dari set fitur asli, karena tidak semua fitur/atribut relevan dengan masalah [12]. Bahkan beberapa dari fitur atau atribut tersebut mengganggu dan dapat mengurangi akurasi. Noisy Features atau fitur yang tidak terpakai tersebut harus dihapus untuk meningkatkan akurasi. Selain itu dengan fitur atau atribut yang banyak akan memperlambat proses komputasi. Pada penelitian ini juga akan menggunakan Forward Selection. Forward Selection atau seleksi kedepan dalam analisisnya pemilihan ke depan di mulai dengan tidak ada prediktor dalam model untuk membantu meningkatkan hasil akurasi dan menentukan atribut yang berpengaruh.

## 2. METODE

Jenis penelitian yang dilaksanakan ini merupakan penelitian eksperimen.

### 2.1 Pengumpulan Data

Tahap ini dilakukan sebagai langkah awal dari suatu penelitian. Untuk memperoleh data yang benar-benar akurat, maka penentuan jenis dan sumber data sangatlah penting. Sumber data pada penelitian ini adalah dataset konten fitur audio yang didapat dari *Meinl Cymbal 2015* [ [HYPERLINK](#) | "meinlcymbal" 2 ].

### 2.2 Teknik Analisis Data

Tahap pengolahan awal data dilakukan untuk mempersiapkan data yang benar-benar valid sebelum diproses pada tahap berikutnya namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (preparation data). Jumlah data awal yang diperoleh dari pengumpulan data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui

beberapa tahap pengolahan awal data (preparation data). Untuk mendapatkan data yang berkualitas, menurut Vercellis [26] dilakukan beberapa teknik:

1.Data integration and transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam software RapidMiner. Tabel kategorikal atribut.

2.Data size reduction, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi tetap bersifat informatif.

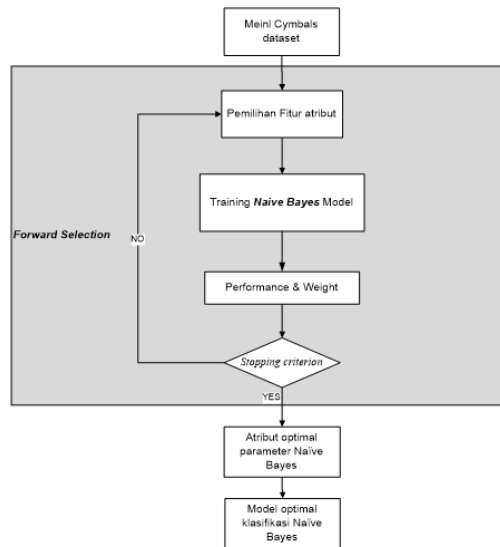
Type	Polynominal	Attribute
Style	Polynominal	Label
Timbre	Polynominal	Attribute
Character	Polynominal	Attribute
Pitch	Polynominal	Attribute
Volume	Polynominal	Attribute
Sustain	Polynominal	Attribute
Weight	Polynominal	Attribute
Finish	Polynominal	Attribute
Material	Binominal	Attribute
Lathe	Polynominal	Attribute
SKU	Polynominal	Attribute

Tabel 2.1 Tipe atribut data

### 2.3 Metode Penelitian

Tahap ini akan membahas metode yang akan digunakan untuk penelitian nanti. Berikut ini adalah tahap yang akan dilakukan dalam penelitian. Dataset diklasifikasikan menggunakan algoritma Naïve Bayes, hasil proses klasifikasi di evaluasi dengan menggunakan Confussion Matrix dan Kappa untuk mengukur performan atau

tingkat akurasi.



Gambar 2.1 Tahapan proposed method

Tahap ini akan membahas metode yang akan digunakan untuk penelitian. Berikut ini adalah tahap yang akan dilakukan dalam penelitian. Tahapan dilakukan mengikuti langkah-langkah metode Forward Selection dengan algoritma Naive Bayes yaitu:

1. Dataset dari Meinl cymbals [2] diseleksi fitur menggunakan Forward Selection, Metode Forward Selection adalah pemodelan dimulai dari nol peubah (empty model).
2. Pemilihan fitur seleksi forward selection diuji menggunakan training atau metode Naive Bayes.
3. Dari training Naive Bayes yang diujikan mendapatkan hasil dan pembobotan.
4. Apabila proses tersebut lolos maka akan mendapatkan suatu atribut/model yang optimal dari klasifikasi Naive Bayes.
5. Sedangkan bila proses tersebut berhenti pada stopping criterion maka proses tersebut diulang dari awal (pemilihan fitur seleksi forward selection) sampai mendapatkan atribut/model optimal.
6. Setelah mendapatkan atribut/model

yang optimal pada klasifikasi Naive Bayes maka akan muncul hasil akurasi dari klasifikasi Naive Bayes yang sudah di fitur seleksi.

## 2.4 Pengujian Model/Metode

Pada tahap ini, dijelaskan tentang teknik pengujian yang digunakan. Metode eksperimen dan pengujian ini mengikuti cara pengklasifikasian menggunakan RapidMiner, dataset diklasifikasi menggunakan algoritma Naive Bayes dan dihitung akurasi dan nilai Kappa.

## 2.5 Evaluasi Dan Validasi Hasil

Pada tahap ini akan dibahas tentang hasil evaluasi dari eksperimen yang telah dilakukan. Model yang terbentuk akan diuji dengan menggunakan Confusion Matrix untuk mengetahui tingkat akurasi. Confusion Matrix akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negative yang benar, dan prediksi negative yang salah. Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif). Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan.

## 3. HASIL DAN PEMBAHASAN

Pada penelitian ini menguji keakuratan klasifikasi alat musik cymbal dengan menggunakan algoritma *Naive Bayes*, setelah itu *Naive Bayes* dengan *Forward Selection* sebagai fitur seleksi. Penelitian ini menggunakan dataset yang diambil dari *Meinl Cymbals 2015* [ [HYPERLINK \l "meinlcymbal" 2](#) ], yaitu dataset dari hasil ekstraksi fitur konten audio yang memiliki 18 *class* atau 18 kategori genre musik, dengan data yang besar (memiliki 486 *record* dan 12 *attribute*).

### 3.1 Algoritma Naïve Bayes

Naïve Bayes adalah metode yang baik karena mudah dibuat, tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar [20].

Berikut teorema bayes :

$$P(X|H) = P(H|X)P(H)$$

Keterangan :

$X$ : Data dengan class yang belum diketahui

$H$ : Hipotesis data  $x$  merupakan suatu class spesifik

$P(H|X)$ : Probabilitas hipotesis  $H$  berdasarkan kondisi  $X$  (posteriori probability)

$P(H)$  : Probabilitas hipotesis  $H$  (prior probability)

$P(X|H)$ : Probabilitas  $X$  berdasar kondisi pada hipotesis  $H$

$P(X)$ : Probabilitas dari  $X$

### 3.2 Evaluasi Naïve Bayes dengan data sampel

Pengujian menggunakan data sampel yang diambil dari IASol dataset dengan: 2 label class (tepat dan terlambat), 10 record (7 class tepat dan 3 class terlambat) dan 21 attribute seperti yang dapat dilihat pada halaman lampiran.

Berikut ini adalah contoh perhitungan mencari nilai akurasi dari atribut kelompok dengan menggunakan metode Cross-Validation (X-Validation).

Training 1:

Style	Weight
Rock, Pop, Fusion, Jazz, Funk, RNB, Reggae, Studio, World	Extra thin
Rock, Pop, Fusion, Jazz, Funk, RNB, Reggae, Studio, World	Thin
Rock, Pop, Fusion, Jazz, Funk, RNB, Reggae, Studio, World	Thin
Rock, Pop, Fusion, Jazz, Funk, RNB, Reggae, Studio, World	Heavy

Rock, Pop, Fusion, Jazz, Funk, RNB, Reggae, Studio, World	Heavy
Rock, Pop, Fusion, Jazz, Funk, RNB, Reggae, Studio, World	Medium
Rock, Pop, Fusion, Jazz, Funk, RNB, Reggae, Studio, World	Medium
Orchestra	Extra Heavy
Orchestra	Extra Heavy

Tabel 3.1 data training cross validation naïve bayes

Dari data diatas didapatkan Probabilitas kelas:

$$P(\text{Tepat}) = 7/9 = 0.777777777$$

$$P(\text{Terlambat}) = 2/9 = 0.222222222$$

Dari data diatas didapatkan Probabilitas kelompok terhadap masing masing kelas:

$$P(\text{Akademik}|\text{Tepat}) = 3/7 = 0.428571429$$

$$P(\text{Reguler}|\text{Tepat}) = 4/7 = 0.571428572$$

$$P(\text{Parsial}|\text{Tepat}) = 0/7 = 0$$

$$P(\text{Akademik}|\text{Terlambat}) = 0/2 = 0$$

$$P(\text{Reguler}|\text{Terlambat}) = 0/2 = 0$$

$$P(\text{Parsial}|\text{Terlambat}) = 2/2 = 1$$

Testing 1:

Data testing dari status kelulusan dengan kelompok parsial:

$$\text{Prediction parsial: } P(X|\text{Tepat}) = 0/7 = 0$$

$$P(X|\text{Terlambat}) = 2/2 = 1$$

Perhitungan dilakukan 10 kali sampai *training 10* dan *testing 10* sesuai metode *Cross-Validation (X-Validation)*.

Dari hasil klasifikasi menggunakan data sample (2 label class, 10 record dan 21 attribute) dengan metode *Naïve Bayes* diperoleh hasil nilai akurasi sebesar 60.00%, berikut ini hasil perhitungannya seperti dapat dilihat pada gambar berikut:

	true Rock, Pop, Fusion, Jazz, Funk, RnB, true Orchestra	class precision
pred. Rock, Pop, Fusion, Jazz, Funk, RnB: 4	1	80.00%
pred. Orchestra	3 2	40.00%
class recall	57.14%	66.67%

Gambar 3.1 Validasi Naïve bayes data Sampel

$$\begin{aligned}
 &= \frac{4+2}{4+1+3+2} \\
 &= 0.6 \\
 &= 60\%
 \end{aligned}$$

Dari eksperimen diatas berdasarkan metode Cross-Validation, didapat nilai akurasi dari satu atribut, yaitu atribut *Weight*.

Cross-validation	Predicted Class	
	<i>Rock, Pop, Fusion, Jazz, Funk, RnB, Reggae, Studio, World</i>	Orchestra
True Class <i>Rock, Pop, Fusion, Jazz, Funk, RnB, Reggae, Studio, World</i>	7	1
Orchestra	0	2

Tabel 3.2 : Validasi Atribut Weight

$$\begin{aligned}
 &= \frac{7+2}{7+0+1+2} \\
 &= 0.9 \\
 &= 90\% \text{ (nilai akurasi atribut } \textit{Weight})
 \end{aligned}$$

Nilai akurasi tersebut dijadikan nilai *subset* pada atribut *Weight* atau dijadikan nilai prediktor untuk perhitungan yang digunakan dalam metode *Forward Selection-Naïve Bayes*.

#### 4. KESIMPULAN DAN SARAN

Algoritma Naive Bayes terbukti efektif dalam mengklasifikasikan genre musik dari dataset dengan dimensi data yang

besar dan memiliki keadaan kelas yang tidak seimbang antara kelas yang satu dengan kelas yang lain atau bersifat Class Imbalance.

Metode Forward Selection dapat mereduksi dimensi dataset yang besar dan dapat membantu meningkatkan hasil akurasi klasifikasi Naïve Bayes.

Dalam hal ini Naive Bayes memanfaatkan fungsi seleksi fitur dari Forward Selection untuk pemilihan atribut data dengan karakteristik data itu sendiri, dan meningkatkan ketepatan klasifikasi Naïve Bayes.

Forward Selection berbasis Naive Bayes lebih akurat dan efektif dalam mengklasifikasikan genre musik dari dataset yang bersifat Class Imbalance dengan data yang besar dengan hasil akurasi 93.43% dan termasuk dalam kategori “Kappa excellent”. Dibanding dengan menggunakan algoritma Naive Bayes saja dengan hasil akurasi hanya 89.30%.

#### 4.1 Saran

Metode Forward Selection berbasis Naive Bayes terbukti akurat dalam klasifikasi genre musik dari dataset yang bersifat Class Imbalance dengan dimensi data yang besar, tetapi dalam penelitian ini terdapat beberapa saran dalam pengembangannya antara lain prosedur ini tidak selalu mengarahkan ke model pemilihan atribut yang terbaik. Forward Selection berbasis Naive Bayes hanya mempertimbangkan sebuah subset kecil dari semua model-model yang mungkin, sehingga resiko melewati atau kehilangan model terbaik akan bertambah, seiring dengan penambahan jumlah variabel bebas.

## DAFTAR PUSTAKA

- [1] James Blades, *Percussion Instruments and Their History*.: Bold Strummer Ltd, 1992.
- [2] [Online]. HYPERLINK "http://meinlcymbals.com/" <http://meinlcymbals.com/>
- [3] Fabian Holt, "Genre in Popular Music," *Journal of The Society for Musicology in Ireland*, p. 27, 2008.
- [4] George Tzanetakis and Perry Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, July 2002.
- [5] A Meng, P Ahrendt, J Larsen, and L.K Hansen, "Temporal Feature Integration for Music Genre Classification," *IEEE Trans. on Speech and Audio Processing*, vol. 15, 2007.
- [6] Yi Liu , Lei Wei , and Peng Wang, "Regional Style Automatic Identification for Chinese folk Songs," 2009.
- [7] N. Scaringella and G. Zoia, "On The Modeling of Time Information For Automatic Genre Recognition Systems in Audio Signals," in *Symposium on Music Information Retrieval*, London, UK, 2005.
- [8] Yingying Zhu, Zhong Ming, and Qiang Huang, "Automatic Audio Genre Classification Based on Support Vector Machine," *Third International Conference on Natural Computation*, 2007.
- [9] Shih-Hao Chen, Shi-Huang Chen, and Rodrigo Capobianco Guido, "Music Genre Classification Algorithm Based on Dynamic Frame Analysis and Support Vector Machine," *2010 IEEE International Symposium on Multimedia*, 2010.
- [10] A. Berenzweig, D. Ellis, and S. Lawrence, "Using voice segments to improve artist classification of music," in *The AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, 2002.
- [11] Christopher DeCoro, Zafer Barutcuoglu, and Rebecca Fiebrink, "Bayesian Aggregation For Hierarchical Genre Classification," in *Austrian Computer Society (OCG)*, 2007.
- [12] Oded Maimon and Lior Rokach, *Data Mining and Knowledge Discovery Handbook* , 2nd ed.: Springer, 2010.
- [13] Steve Levine, *Automatic Music Genre Categorization using Multiclass Classification*.: 6.867 Machine Learning Final Project, 2010.
- [14] *Naive Bayes for Text Classification with Unbalanced Classes*. Auckland, New Zealand: Computer Science Department, University of Waikato and Xtal Mountain Information Technology, 2011.

- [15] Leung K.Ming, *Naive Bayesian Classifier.*, november 2007. 1960.
- [16] Mark A. Hall and Geoffrey Holmes , "Benchmarking Attribute Selection Techniques for Discrete Class Data Mining," *IEEE Transactions On Knowledge And Data Engineering*, vol. 15, p. 3, May/June 2003.
- [17] Mohamad Fajarianditya Nugroho, Romi Satria Wahono, and Vincent Suhartono , "Penerapan Metode Forward Selection untuk Fitur Seleksi Pada Klasifikasi Genre Musik Menggunakan Algoritma Naive Bayes," Udinus, Mkom Thesis 2013.
- [18] Ian H Witten, Eibe Frank, and Mark A Hall, *Data Mining Practical Machine Learning Tools and Techniques*, 3rd ed.: Morgan Kaufmann , 2011.
- [19] Florin Gorunescu, *Data Mining: Concepts, Model and Techniques*, Prof. Janusz Kacprzyk and Prof. Lakhmi C. Jain, Eds. Berlin, Jerman: Springer, 2011, vol. 12.
- [20] Xindong Wu and Vipin Kumar, *The top ten Algorithms in Data Mining.*, Taylor & Francis Group and LLC, Eds., 2009.
- [21] Budi Santoso, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, 1st ed. Yogyakarta, Indonesia, 2007.
- [22] Jacob Cohen, "A Coefficient Of Agreement For Nominal Scale ," 1960.
- [23] Joseph L. Fleiss , "Measuring Nominal Scale Agreement Among Many Raters," 1971.
- [24] Mikael Berndtsson, Jörgen Hansson, Björn Olsson, and Björn Lundell, *Thesis Projects A Guide for Students in Computer Science and Information Systems*, 2nd ed. London: Springer, 2008.
- [25] Christian W Dawson, *Projects in Computing and Information Systems A Student's Guide*, 2nd ed. England: Pearson Education, 2009.
- [26] Carlo Vercellis, *Business Intelligent: Data Mining and Optimization for Decision Making.* Southern Gate: Chichester: John Willey & Sons, Ltd., 2009.
- [27] Jiawei Han, *Data Mining Concept And Technique*, 2nd ed., Asma Stephan, Ed. Champaign, United States of America: Multiscience Press, 2007.