

Classification Based Association (CBA) Menggunakan R

Alesia Arum Frederika^{a1}, I Putu Agung Bayupati^{a2}, Putu Wira Buana^{b3}

^aProgram Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana, Bali
Email: ¹alesiaarum@student.unud.ac.id, ²bayupati@unud.ac.id, ³wb@unud.ac.id

Abstrak

Metode baru yang digunakan untuk membuat pengklasifikasian berdasarkan aturan yang ditemukan melalui aturan asosiasi. Metode tersebut disebut dengan associative classification dan salah satu algoritma yang digunakan adalah algoritma CBA atau classification based association. Classification based Association atau yang dikenal dengan algoritma CBA merupakan salah satu algoritma yang terdapat pada metode associative classification yang merupakan strategi baru dalam pengolahan data dalam membangun model klasifikasi. Classification based association merupakan metode yang menggabungkan teknik data mining asosiasi dan klasifikasi pada penerapannya, dengan melakukan pengklasifikasian berdasarkan aturan asosiasi. Penelitian ini menjelaskan bagaimana classification based association atau algoritma CBA dapat diterapkan pada R.

Kata kunci: R, Associative Classification, Algoritma Classification based Association (CBA)

Abstract

A new method for creating classifications based on rules discovered through association rules. This method is called associative classification and one of the algorithms used is the CBA algorithm or association-based classification. Classification based on Association or known as the CBA algorithm is one of the algorithms contained in the associative classification method which is a new strategy in data processing in building a classification model. Classification based association is a method that combines data mining techniques of association and classification in its application, by classifying based on association rules. This study explains how classification-based associations or CBA algorithms can be applied to R.

Keywords: R, Associative Classification, Algoritma Classification based Association (CBA)

1. Pendahuluan

Perkembangan teknologi informasi yang begitu pesat membentuk perusahaan semakin mudah dalam melakukan proses bisnisnya dan tentunya ketersediaan data di segala bidang semakin melimpah. Banyaknya data yang didapat pada suatu perusahaan atau organisasi memerlukan waktu yang lama untuk melakukan pengolahan data. Kemajuan teknologi informasi tersebut merupakan suatu hal yang saling terkait, dalam ketatnya persaingan pasar untuk memenuhi tuntutan pelanggan yang semakin tinggi. Perusahaan memerlukan strategi dan kecerdasan bisnis untuk dapat terus memenuhi keinginan pelanggan dan tuntutan pasar. Sehingga kemajuan teknologi sangat dibutuhkan untuk mengembangkan bisnis perdagangan.

Data Mining merupakan serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang belum diketahui secara manual dari suatu kumpulan data [8]. Data mining, juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang berisi pengumpulan maupun pemakaian data untuk menemukan pola atau hubungan dalam set data berukuran besar [9]. KDD terdiri dari banyak metode dan teknik yang dapat diterapkan pada data yang berbeda dalam pengolahan data. Beberapa metode tersebut meliputi asosiasi, klasifikasi, dan clustering. Salah satu metode yang dijelaskan adalah associative classification yang merupakan gabungan dari teknik asosiasi dan classification.

Associative classification adalah salah satu metode terbaru dalam data mining yang menggabungkan teknik asosiasi dan klasifikasi untuk membangun model klasifikasi.

Pendekatan Associative Classification merupakan pendekatan gabungan terbaru yang mencapai akurasi yang lebih tinggi daripada pendekatan klasifikasi tradisional. Associative classification adalah pendekatan machine learning yang bertujuan untuk membangun model klasifikasi (pengklasifikasi) yang akurat, efektif dan kompak dengan menggabungkan paradigma dari klasifikasi dan pendekatan aturan asosiasi. Associative classification memiliki dua fase dalam proses pengolahan data, fase ini berguna untuk memperoleh hasil model klasifikasi dari suatu data. Fase *rule generation* merupakan fase untuk membangkitkan *frequent itemset* pada metode AC (*associative classification*). Fase ini digunakan untuk mengekstrak rule yang bertujuan untuk membangkitkan himpunan class association rule (CAR). Kedua ada fase building classifier. Fase building classifier adalah fase untuk memprediksi atau mengklasifikasi data yang belum dikategorikan dalam kelas atau label tertentu. Algoritma dari metode associative classification adalah CBA, CMAR, MCAR, L3G, dan sebagainya.

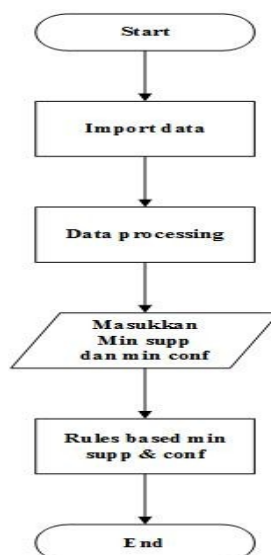
Salah satu algoritma pada metode associative classification yaitu, CBA (Classification based Association). Classification based Association merupakan algoritma yang memberikan keseimbangan yang sangat baik antara akurasi, kecepatan, dan pemahaman model. Tidak seperti banyak pendekatan yang lebih baru, pengklasifikasi classification based association mudah untuk ditafsirkan dan diterapkan. Kumpulan aturan yang dihasilkan relatif kecil, aturannya tegas, dan aturan diurutkan menurut prediksi kekuatan. Batasan lain yang berlaku khusus untuk CBA adalah bahwa bahkan ketika pengguna menentukan ambang batas yang wajar, CBA biasanya menghasilkan lebih banyak aturan daripada pendekatan terkait lainnya [2]. Keterbatasan ini mungkin menjadi alasan mengapa Implementasi CBA belum tersedia di banyak lingkungan komputasi untuk machine learning dan statistik. Terdapat tiga packages yang digunakan pada imlementasi CBA dan ketiga packages tersebut terdapat pada bahasa R, salah satunya ialah arulesCBA. Jurnal ini menjelaskan penerapan algoritma CBA menggunakan bahasa R.

2. Metodologi Penelitian

Metodologi penelitian menjelaskan gambaran umum associative classification dari algoritma CBA yang digunakan dalam pengolahan data dengan bahasa R.

2.1 Gambaran Umum

Gambaran umum classification based association menjelaskan proses penerapan algoritma CBA yang dilakukan pada penelitian yang dapat dilihat pada gambar 1.



Gambar 1. Gambaran Umum CBA

Gambar 1 merupakan gambaran umum classification based association yang dimulai dengan import data pada proses penerapannya. Data yang di import akan diproses pada tahap data processing untuk diubah menjadi data yang dapat dibaca pada proses penerapan

classification based association. Penerapan classification based association perlu memasukkan minimum support dan minimum confidence untuk menghasilkan rules yang didapatkan pada proses classification based association.

3. Studi Literatur

Studi literatur dilakukan berdasarkan pengumpulan data dan informasi untuk digunakan pada penelitian.

3.1. Data Mining

Data mining dapat diartikan sebagai pengolahan data atau penggalian pengetahuan dari data yang berjumlah besar.[1] *Data mining* berdasarkan fungsinya berarti proses pengambilan keputusan dari volume data yang besar yang disimpan dalam *repository* [4].

Data mining juga disebut dengan istilah populer yaitu *knowledge discovery from data* atau KDD. Tahapan proses *data mining* sama dengan proses yang dilakukan pada *knowledge discovery*. Tahapan dimulai dari tahap seleksi data dari data sumber ke data target, tahap *preprocessing* untuk memperbaiki kualitas data dan transformasi data *mining* serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Penjelasan mengenai 5 tahapan *data mining* atau *knowledge discovery* yaitu sebagai berikut.

1. Seleksi Data

Tahapan seleksi data bertujuan untuk mengekstrak data dari gudang data yang besar menjadi data yang relevan dengan analisis data mining. Proses ini dapat membantu untuk merampingkan dan mempercepat proses.

2. Data Preprocessing

Tahapan data *preprocessing* merupakan tahapan pembersihan data dan persiapan tugas untuk memastikan hasil dengan benar, seperti *missing value* data dan memastikan bahwa tidak ada nilai data palsu.

3. Transformasi Data

Tahapan transformasi data yaitu tahapan dalam mengubah data dalam bentuk atau format yang sesuai untuk kebutuhan data *mining*. Proses normalisasi biasanya diperlukan dalam tahap ini.

4. Data mining

Tahapan data mining bertujuan untuk menganalisis data sesuai algoritma yang digunakan sehingga menemukan hasil berupa pola atau aturan yang bermakna serta menghasilkan model prediksi. Tahapan ini adalah elemen inti dari siklus KDD.

5. Interpretasi dan Evaluasi

Algoritma data mining biasanya menghasilkan jumlah yang tidak terbatas dari hasil aturan pada proses tersebut yang mungkin tidak bermakna atau berguna. Maka dari itu tahap ini bertujuan untuk memilih model-model yang berguna dalam membuat keputusan bisnis masa depan.

3.2. Asosiasi

Asosiasi dalam data mining merupakan proses untuk menemukan atribut yang muncul dalam suatu waktu. Tugas asosiasi bertujuan untuk mencari hubungan antar suatu kombinasi item dalam suatu set data yang telah ditentukan. Asosiasi atau yang bisa juga disebut dengan *association rules* digunakan untuk menemukan pola-pola yang terjadi dalam suatu kumpulan data. Asosiasi memiliki istilah *antecedent* dan *consequent*. *Antecedent* mewakili bagian "jika" dan *consequent* untuk mewakili bagian "maka". asosiasi berbentuk $C \rightarrow D$, dimana C dan D adalah dua *itemset* terpisah (*disjoint*) yang masing-masing disebut sebagai LHS (*Left-Hand Side*) dan RHS (*Right-Hand Side*). *Antecedent* juga bisa disebut dengan LHS (*Left Hand Side*) sedangkan *consequent* bisa disebut dengan RHS (*Right Hand Side*), dengan interpretasi dimana setiap pembelian *item* pada LHS memungkinkan adanya pembelian pada RHS.

3.3. Klasifikasi

Klasifikasi adalah pengelompokan yang sistematis mengenai objek atau gagasan ke dalam kelas atau golongan tertentu berdasarkan kesamaan karakteristik [7]. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. Klasifikasi secara umum terdiri dari dua tahap menurut Han dalam bukunya. Tahapan *learning* (proses belajar) merupakan model yang dibuat dalam menggambarkan himpunan kelas atau konsep yang sudah ditentukan sebelumnya. Model klasifikasi dibuat berdasarkan analisa *record* pada database yang digambarkan dalam bentuk atribut. *Record* yang diasumsikan akan masuk ke dalam suatu kelas yang telah ditentukan sebelumnya, yang disebut atribut kelas.

3.4. Associative Classification

Associative classification merupakan gabungan antara aturan asosiasi dengan klasifikasi. Associative classification adalah kasus khusus dari association rule mining dimana atribut kelas yang digunakan terdapat di bagian sisi kanan rule, sebagai contoh, $X \Rightarrow Y$, Y harus merupakan atribut kelas [3]. Contoh dataset training E mempunyai atribut berbeda, $I = \{i_1, i_2, \dots, i_m\}$ dan C adalah daftar dari label kelas. Banyaknya baris dalam E dinotasikan sebagai $|E|$. Atribut dapat berupa kategori atau kontinu. Pada kasus atribut bertipe kategori, semua kemungkinan nilai dipetakan ke himpunan bilangan integer positif. Untuk atribut yang kontinu, atribut ini harus dideskritisasi untuk diubah ke bentuk kategori. Dalam tahap associative classification, setiap transaksi mengandung subset dari item - item yang dipilih dari I (di) dan label kelas C . Kumpulan lebih item disebut itemset. Itemset disebut frequent itemset jika nilai support dari itemset tersebut melampaui nilai minimum support (minsupp) threshold. Rule r disebut frequent rule jika nilai rule support dari r melewati nilai minsupp threshold. Sebuah rule di ekstrak sebagai Class Association Rules (CAR) jika nilai support dan confidence dari r melewati nilai minsupp and minimum confidence (minconf) threshold. Algoritma Associative Classification telah diusulkan seperti Classification based Association (CBA), Classification based on Multiple Association Rules (CMAR), Class based Associative Classification (CAC), dan Classification based on Predicted Association Rule (CPAR).[5]

3.5. Classification Based Association (CBA)

Algoritma CBA (Classification based Association) yang merupakan pemanfaatan dari algoritma Apriori [6]. CBA menemukan dan menghasilkan aturan dengan menerapkan Apriori yang menghasilkan karakteristik kandidat. Perbedaan utama antara sebuah itemset dan ruleitem adalah bahwa itemset hanya muncul sebagai nilai atribut melalui dirinya sendiri, di mana sebagai item aturan terdiri dari nilai klasifikasi yang dilampirkan dengan nilai atribut. Algoritma CBA merupakan algoritma yang memiliki proses klasifikasi yang sedikit lebih akurat dan efektif dengan dasar pendekatan asosiatif. CBA telah terbukti sangat berguna ketika tujuan utamanya adalah untuk menghasilkan hasil yang sangat dapat diinterpretasikan. Pemodelan algoritma CBA atau classification based association memiliki dua tahap dalam prosesnya.

CBA-RG atau CBA *rule generator* dilakukan berdasarkan metode Apriori dalam menemukan hasil rules asosiasi pada prosesnya. Tahap ini dilakukan untuk menemukan ruleitems yang memiliki atau memenuhi parameter dari minimum *support*. Tahap menggenerate semua frequent ruleitem yang pertama menghitung *support* dari ruleitem individual dan tentukan statusnya. Subsequence dengan nilai *support* yang memiliki tingkat lebih tinggi dari minimum *support* merupakan *frequent* ruleitem pada tahap sebelumnya. Proses membuat set possibly pada *frequent* ruleitems yang baru disebut candidate ruleitem. Candidate ruleitems tersebut lalu dihitung nilai *support*nya untuk menentukan candidate ruleitem yang mempunyai nilai *support* lebih besar dari minimum *support* untuk menentukan candidate ruleitem yang *frequent*, Setelah itu diproduksi rules nya (CARs).

CBA-CB adalah sebuah *classifier builder* menggunakan CARs atau prCARs yang telah didapatkan sebelumnya pada tahap CBA-RG. Memproduksi *classifier* dilakukan dengan cara melakukan tahap evaluasi pada semua possible subset yang terdapat pada data training dan memilih subset dengan rule sequence yang memiliki error paling sedikit dengan melihat rule yang mempunyai nilai *confidence* yang lebih besar. Tahapan membangun classifier builder adalah dengan mengurutkan seperangkat rules yang dihasilkan menurut hubungan

"precedence". Hal ini bertujuan untuk memastikan agar memilih rules dengan hak tertinggi untuk classifier. Langkah kedua memilih rules untuk classifier diurutkan berdasarkan yang telah diurutkan. Tahapan tersebut dapat di proses jika memasukkan minimum support dan minimum confidence yang digunakan pada classification based association dengan rumus

$$\text{Supp} = \frac{\text{Rulesup}}{|\text{D}|} \times 100\% \quad (1)$$

$$\text{Conf} = \frac{\text{Rulesup}}{\text{condsup}} \times 100\% \quad (2)$$

Dimana rulesup merupakan jumlah item dan kelas y di dataset D. sedangkan Condsup merupakan jumlah item di dataset D dan |D| merupakan jumlah / ukuran dataset

4. Hasil dan Pembahasan

Hasil dan pembahasan pada penelitian ini menjelaskan tentang implementasi algoritma CBA pada bahasa R

4.1. Implementasi

Implementasi dengan menggunakan algoritma CBA dengan menggunakan package arulesCBA pada bahasa R. Aplikasi yang digunakan adalah Rstudio, pengguna dapat melakukan implementasi algoritma CBA dengan menginstall packages arulesCBA pada Rstudio. Pengguna menggunakan contoh data transaksi dalam melakukan implementasi algoritma CBA. Package yang sudah diinstall akan digunakan pada Rstudio, sebelum digunakan pengguna perlu mengimport data yang akan dipakai dengan menggunakan algoritma CBA. Setelah data di import akan melakukan implementasi algoritma CBA dengan memberikan perintah seperti pada gambar

```
train <- sapply(data, as.factor)
train <- data.frame(train, check.names = FALSE)
txns <- as(train, "transactions")

train$class <- as.factor(train$class)
classifier <- arulesCBA::CBA(class ~ ., data = train, supp = 0.01, confidence = 0.1)
classifier
```

Gambar 2. code perintah CBA

Gambar 2 merupakan code perintah penerapan classification based association pada R dimana data yang di import akan di split menjadi data train dan test, sebelum melakukan tahapan penggunaan algoritma CBA. Setelah melakukan splitting data maka akan melakukan tahapan penggunaan algoritma CBA dengan packages arulesCBA lalu pengguna dapat melihat berapa banyak rules yang didapatkan pada proses tersebut. Detail hasil keseluruhan rules dari penerapan algoritma CBA dapat dilihat dengan menggunakan inspect seperti pada gambar 3.

```
ru <- rules(classifier)
inspect(rules(classifier))
```

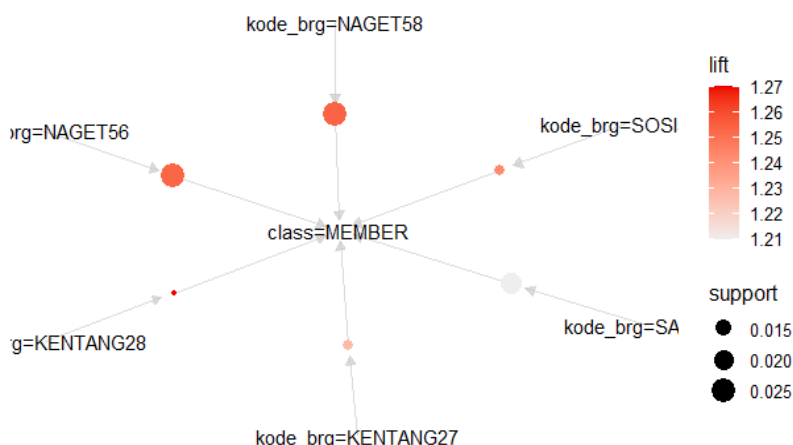
Gambar 3. Code perintah menampilkan rules

Gambar 3 merupakan code perintah untuk menampilkan detail dari keseluruhan rules yang didapat. Hasil dari penerapan yang dilakukan dengan algoritma CBA pada Rstudio dapat dilihat pada gambar 4.

```
CBA Classifier object
Class:
Default Class: NA
Number of rules: 10
Classification method: first
Description: CBA algorithm (Liu et al., 1998)
```

Gambar 4. Tampilan Rules

Gambar 4 menampilkan jumlah rules yang didapatkan setelah proses implementasi algoritma CBA diterapkan dengan menggunakan data yang telah dimasukkan. Rules yang didapat bisa dilihat dengan menggunakan visualisasi pada R seperti pada gambar 5.



Gambar 5. Visualisasi CBA

Gambar 5 merupakan tampilan visualisasi pada penerapan classification based association atau algoritma CBA pada R dengan menggunakan scatter plot pada penerapannya.

5. Kesimpulan

Penelitian ini menjelaskan bahwa Classification based association yang merupakan salah satu algoritma pada metode associative classification yang menghasilkan model klasifikasi dengan aturan asosiasi dapat diterapkan dan digunakan pada machine learning dengan menggunakan bahasa R pada Rstudio dengan menggunakan packages yang terdapat pada R yaitu arulesCBA

References

- [1] Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- [2] Alcalá-Fdez, R. Alcalá, and F. Herrera. A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Transactions on Fuzzy Systems*, 19(5):857–872, 2011. URL <https://doi.org/10.1109/TFUZZ.2011.2147794>. [p254, 256]
- [3] Andikos, A. F., & Andri, H. (2019). Pengujian Association Clasification Dalam Meningkatkan Kualitas Minyak Sawit Sebagai Bahan Dasar Biodisel. *Jurnal Media Informatika Budidarma*, 3(4), 340. <https://doi.org/10.30865/mib.v3i4.1383>
- [4] Devi, A. S., Putra, I. K. G. D., & Sukarsa, I. M. (2015). Implementasi Metode Clustering DB-SCAN pada Proses Pengambilan Keputusan. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, 6(3), 185. <https://doi.org/10.24843/lkjiti.2015.v06.i03.p05>
- [5] Karyawati, E., & Winarko, E. (2011). Class Association Rule Pada Metode Associative Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 5(3),

17. <https://doi.org/10.22146/ijccs.5207>
- [6] Nagar-, N., Pradesh, A., & Pradesh, A. (2020). *Diabetic Disease Prediction System using Associative Classification with Improved Classifier Accuracy*. 13(1), 2821–2827.
- [7] Nyoman, N., & Smrti, E. (2015). Otomatisasi Klasifikasi Buku Perpustakaan Dengan Menggabungkan Metode K-Nn Dengan K-Medoids. *Lontar Komputer*, 4(1), 201–214.
- [8] Pramudiono, I. 2007. Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data. <http://www.ilmukomputer.org/wpcontent/uploads/2006/08/ikodatamining.zip>.
- [9] Santosa, Budi, 2007, ³Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis¥, Graha Ilmu, Yogyakarta.
-