

Check 45. Pengembangan Learning Characteristic Rule Pada Algoritma Data Mining Attribute Oriented Induction.pdf

by Arta Sundjaja

Submission date: 30-Apr-2019 03:55PM (UTC+0700)

Submission ID: 1102516845

File name: Rule_Pada_Algoritma_Data_Mining_Attribute_Oriented_Induction.pdf (423.04K)

Word count: 6200

Character count: 36392

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/307637951>

PENGEMBANGAN LEARNING CHARACTERISTIC RULE PADA ALGORITMA DATA MINING ATTRIBUTE...

Article · September 2016

CITATIONS

0

READS

44

2 authors, including:



[Harco Leslie Hendric Spits Warnars](#)

Binus University

67 PUBLICATIONS 49 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Computer Science for healthcare benefit in Indonesia [View project](#)

All content following this page was uploaded by [Harco Leslie Hendric Spits Warnars](#) on 06 September 2016.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

PENGEMBANGAN LEARNING CHARACTERISTIC RULE PADA ALGORITMA DATA MINING ATTRIBUTE ORIENTED INDUCTION

Adi Wibowo, Harco Leslie Hendric Spits Warnars

Abstract— This paper shows the improvement of current characteristic rule learning in Attribute Oriented Induction (AOI) data mining technique. The proposed algorithm was applied with improvement upon current algorithm with 3 steps where the first step is elimination for checking condition if there is no higher level concept in concept hierarchy for attribute. The second step is elimination of attribute removal if fulfill for checking condition if there is no higher level concept. The third step is elimination of attributes in input dataset which no higher level concept in concept hierarchy. The development of these data mining algorithm applied Knowledge Data Discovery (KDD) methodology which consist 7 steps. Current and proposed AOI characteristic rule learning were implemented with server programming such as PHP Hypertext Preprocessor (PHP) and using 4 input datasets such as adult, breast cancer, census and IPUMS from University of California, Irvine (UCI) machine learning repository. The experiments showed that proposed AOI characteristic rule are better than current AOI characteristic rule, where experiments upon adult, breast cancer, census, IPUMS datasets have average 11, 3.8, 7.2, 7.2 respectively times better performance. The experiments were carried on AMD A10-7300(1.90 GHz) processor with 8.00 GB RAM.

Index Terms— Data Mining, Attribute Oriented Induction, characteristic rule, Knowledge Data Discovery.

I. PENDAHULUAN

Pendekatan Attribute Oriented Induction (AOI) dikembangkan untuk mempelajari berbagai knowledge rules seperti

First Author is with IT department, PT. Jaya Prima Abadi, Semarang, Indonesia (e-mail: adi.wibowo@kopiluwak.org).
Second Author is with Doctor of Computer Science, Bina Nusantara university, Jakarta, Indonesia (e-mail: shendric@binus.edu).

characteristic rules, discrimination atau classification rules, quantitative rules, data evolution regularities [1], qualitative rules [2], dan cluster description rules [3]. Attribute Oriented Induction memiliki konsep hirarki sebagai keuntungan dimana konsep hirarki yang dipakai sebagai latar belakang knowledge didapatkan oleh knowledge insinyur atau ahli dibidangnya [3,4,5]. Konsep disusun dan diurutkan dalam konsep hirarki menurut levelnya dari konsep tingkat tertentu (spesifik) atau rendah ke tingkat yang lebih umum atau lebih tinggi dan generalisasi dicapai dengan naik ke konsep tingkat yang lebih tinggi mengikuti alur dari konsep hirarki [6].

DBLearn adalah sistem data mining prototipe yang dikembangkan di Simon Fraser University mengintegrasikan metodologi machine learning dengan teknologi database dan dengan secara efektif dan efisien melakukan ekstraksi aturan karakteristik dan diskriminan dari database relasional [7]. Sejak tahun 1993 DBLearn telah melahirkan sebuah sistem baru yang disebut DBMiner dengan fitur berikut:

- a) Menggabungkan beberapa teknik data mining seperti Attribute Oriented Induction, analisis statistik, pendalaman progresif untuk mining multiple-level rules dan meta-rule guided knowledge mining [8] data cube dan teknologi OLAP[9].
- b) Mining aturan baru dari database besar termasuk multiple level association rules, classification rules, cluster description rules and prediction.
- c) Generasi otomatis hirarki numerik dan perbaikan konsep hirarki.
- d) Seperti High level SQL dan interface data mining grafis.
- e) Arsitektur client server dan peningkatan kinerja untuk aplikasi yang lebih besar.
- f) Bahasa query data mining DMQL seperti SQL dan Grafis antarmuka pengguna

telah ditingkatkan untuk mining knowledge secara interaktif.

- g) Melakukan roll-up dan drill-down di beberapa tingkat konsep
- h) dengan multiple dimensional data cubes.

DBMiner telah dikembangkan dengan mengintegrasikan database OLAP dan teknologi data mining [9] yang sebelumnya disebut DBLearn yang memiliki arsitektur database sendiri. Konsep hirarki disimpan sebagai relasi dalam database yang menyediakan latar belakang penting untuk generalisasi data dan multiple level data mining.

Konsep hirarki dapat ditentukan berdasarkan hubungan antara atribut database atau dengan set pengelompokan dan disimpan dalam bentuk hubungan dalam database yang sama[8]. Konsep hirarki dapat disesuaikan secara dinamis berdasarkan distribusi dari himpunan data yang relevan dengan data mining task dan hirarki untuk atribut numerik dapat dibangun secara otomatis berdasarkan data analisis distribusi [8].

Untuk mempermudah implementasinya konsep hirarki hanya akan didasarkan pada non rule based concept hierarchy dan characteristic rule saja. Characteristic rule adalah sebuah pernyataan yang mencirikan konsep yang ada pada data yang tersimpan dalam database. Memberikan konsep umum tentang properti yang dapat membantu orang mengenali fitur umum dari data dalam kelas, misalnya gejala dari penyakit tertentu[6]. Untuk melakukan generalisasi ada 8 langkah strategi harus dilakukan [4], di mana langkah 1 sampai 7 sebagai untuk characteristic rule dan langkah 1 sampai 8 untuk discriminant rule.

- a) Generalization on the smallest decomposable components
- b) Attribute removal
- c) Concept tree Ascension
- d) Vote propagation
- e) Threshold control on each attribute
- f) Threshold control on generalized relations
- g) Rule transformation
- h) Handling overlapping tuples

II. IDENTIFIKASI PERMASALAHAN

Penelitian sebelumnya menunjukkan kelemahan teknik AOI characteristic rule[14,15] yaitu :

- Algoritma AOI ini hanya dapat menyediakan sebuah snapshot dari generalized knowledge dan bukan merupakan gambaran umum dari data. Gambaran umum dari data bisa didapatkan dengan mencoba threshold yang berbeda berkali – kali
- Mengubah threshold yang berbeda akan menghasilkan generalized tuples yang berbeda pula. Tetapi, menggunakan threshold yang berbeda tentu saja memakan waktu yang cukup lama apabila data yang digeneralisir banyak.
- Terdapat masalah dalam memilih generalized rules yang terbaik antara threshold kecil dan threshold besar. Dimana threshold besar akan menunjukkan hasil yang terlalu spesifik dan sebaliknya threshold kecil akan menunjukkan hasil yang terlalu generalisir sehingga ada kemungkinan kehilangan beberapa informasi yang berharga

Berdasarkan kelemahan yang didapatkan dari studi literatur maka didapatkan informasi bahwa algoritma ini sangat tergantung kepada proses pencarian generalisasi data berdasarkan threshold dan dilakukan berulang – ulang. Oleh karena itulah diperlukan proses generalisasi yang lebih cepat sehingga proses generalisasi dengan menggunakan threshold yang berbeda dapat dilakukan tanpa waktu yang terlalu lama.

III. ANALISA KELEMAHAN AOI CHARACTERISTIC RULE YANG ADA SEKARANG

Dari pemaparan dan hasil implementasi algoritma AOI characteristic rule yang ada sekarang maka peneliti melihat ada sebuah langkah yang kurang efisien dalam penggunaannya yaitu langkah ke 3 dan ke 4 pada algoritma characteristic rule dibawah ini

1. For each of attribute A_i ($1 \leq i \leq n$, where $n = \#$ of attributes) in the generalized relation GR


```

2. { While #_of
   distinct_values_in_attribute_Ai > threshold
3. { If no higher level concept in concept
   hierarchy for attribute_Ai
4. Then remove attribute_Ai
5. Else substitute the value of Ai by its
   corresponding minimal generalized
   concept
6. Merge identical tuples
7. }
8. }
9. While #_of_tuples in GR > threshold
10. { Selective generalize attributes
11. Merge identical tuples
12. }

```

Algoritma ini akan melakukan pengecekan berulang pada setiap atributnya terhadap file concept hierarchy-nya. Seperti yang terlihat pada langkah ke 3 yaitu bila atribut tersebut tidak ada pada file concept hierarchy maka akan dibuang dari generalisasi yang terlihat pada langkah ke 4. Karena concept hierarchy tidak memiliki keseluruhan generalisasi dari tiap atributnya dimana hanya beberapa atribut saja yang dipakai sebagai learning. Maka perulangan ini akan memakan banyak waktu apabila semakin banyak atribut yang dicek maka semakin lama pula waktu yang dibutuhkan dalam melakukan generalisasi.

Selain itu, dalam penelitian yang didapat melalui studi literatur juga didapatkan kesulitan dalam menentukan hasil generalisasi yang sesuai dengan threshold yang diinginkan. Karena dalam algoritma tersebut tidak disebutkan secara jelas apa yang harus dilakukan untuk memenuhi kriteria sesuai dengan threshold tersebut.

Setelah menganalisa kelemahan dari algoritma AOI characteristic rule yang ada sekarang maka algoritma tersebut diubah langkahnya menjadi lebih pendek dan dengan memasukan data yang lebih spesifik terlebih dahulu sebelum dilakukan generalisasi. Dimana data yang dibacapada perulangan langkah ke 2 dibatasi pada hanya pada atribut yang hanya akan di-learning saja. Yang dilakukan adalah melakukan transformasi terhadap data terlebih dahulu sehingga, hanya atribut yang terdapat pada file concept hierarchy saja yang akan diproses lebih lanjut

pada proses generalisasi menggunakan algoritma AOI. Data tersebut ditransformasikan saat user akan memilih atribut mana yang akan di learn dari data tersebut. Pada intinya langkah ke 4 yaitu menghilangkan atribut dilakukan sebelum data diproses, sehingga mengurangi waktu terbuang untuk mengecek atribut yang tidak mempunyai nilai concept di generalisasi nilai diatasnya. Pada akhirnya, langkah ke-3 yaitu mengecek untuk atribut yang tidak ada nilai concept diatasnya pada file concept hierarchy tidak dilakukan, karna semua atribut yang digunakan adalah atribut yang mempunyai nilai concept diatasnya pada file concept hierarchy.

Dikarenakan data sudah dilakukan proses transformasi, maka ada beberapa langkah yang dihilangkan karena data sudah pasti ada pada file concept hierarchy. Berikut algoritma yang sudah disesuaikan dengan data yang sudah di transformasikan dimana langkah ke 3 dan ke 4 dieliminasi, sehingga langkah algoritma yang mempunyai 12 baris menjadi 10 baris. Selain itu, dilakukan juga generalisasi lanjutan atas data yang masih belum sesuai dengan threshold yang diberikan. Jadi bila pada langkah 8 jumlah data masih lebih besar dari pada rule thresholdnya maka akan dilakukan selective generalize attributes dengan cara melakukan generalisasi lanjutan terhadap high level atributnya bila ada seperti yang dilakukan pada langkah 2 sampai dengan 6. Sehingga bila dijabarkan lebih detail maka akan didapatkan algoritma sebagai berikut :

```

1. For each of attribute Ai (1 ≤ i ≤ n, where n = #
   of attributes) in the generalized relation
   GR
2. { While #_of
   distinct_values_in_attribute_Ai > threshold
3. { substitute the value of Ai by its
   corresponding minimal generalized
   concept
4. Merge identical tuples
5. }
6. }
7. While #_of_tuples in GR > threshold
8. { Repeat step 2-6
9. Merge identical tuples

```

10. }

IV. IMPLEMENTASI ALGORITMA AOI USULAN

Setiap implementasi dari sebuah algoritma data mining haruslah digunakan sebuah metodologi dalam implementasinya, oleh karena itulah implementasi AOI Characteristic Rule usulan juga akan diimplementasikan dengan metodologi Knowledge Discovery in Database (KDD) dengan 7 langkah – langkah sebagai berikut :

- A. Data Cleaning.
- B. Data Integration.
- C. Data Selection.
- D. Data Transformation.
- E. Proses Mining.
- F. Pattern Evaluation.
- G. Knowledge presentation.:

E. Data Cleaning

Terdapat empat buah dataset yang akan dijadikan sumber data pada penelitian ini masing – masing adalah dataset adult, dataset breast cancer, dataset census 1990, dan dataset IPUMS yang didapatkan dari data public pada UCI Machine learning university of California (<https://archive.ics.uci.edu/ml/>) [16].

1) Dataset Adult

Dataset ini didapatkan pada data publik pada UCI Machine Learning University of California (<https://archive.ics.uci.edu/ml/datasets/Adult>), dataset ini didonorkan oleh Ronny Kohavi dan Barry Becker pada tahun 1996 pada situs ini dengan data ini merupakan hasil ekstraksi yang dilakukan oleh Barry Becker dari data sensus tahun 1994. Data ini diekstrak dengan menggunakan kondisi berikut: ((AAGE>16) && (AGI>100) && (AFNLWGT>1)&& (HRSWK>0)). Dataset ini memiliki jumlah record 48842 dan memiliki 14 atribut.

2) Dataset Breast Cancer

Dataset ini didapatkan pada sumber yang sama yaitu UCI Machine Learning University of California (<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29>), dataset ini didonorkan oleh

Dr. William H. Wolberg, W. Nick Street dan Olvi L. Mangasarian pada tahun 1995. Data ini terdiri dari 699 record dan 32 atribut.

3) Dataset Census 1990

Dataset ini didapatkan pada sumber yang sama yaitu UCI Machine Learning University of California ([https://archive.ics.uci.edu/ml/datasets/US+Census+Data+\(1990\)](https://archive.ics.uci.edu/ml/datasets/US+Census+Data+(1990))), Dataset ini didonorkan oleh Chris Merk, Bo Thiesson, dan David Heckerman. Dataset ini didapatkan dari (U.S. Department of Commerce) Census Bureau website menggunakan Data Extraction System. Data ini merupakan bagian dari sensus penduduk amerika tahun 1990. Dataset ini memiliki jumlah record 2458285 yang terdiri dari 68 atribut.

4) Dataset IPUMS

Dataset ini didapatkan pada sumber yang sama yaitu UCI Machine Learning University of California (<https://archive.ics.uci.edu/ml/datasets/IPUMS+Census+Database>). Dataset ini didonorkan oleh Stephen Bay tahun 1999, sumber asli dari dataset ini adalah IPUMS project yang dilakukan oleh Ruggles Sobek tahun 1997. IPUMS project adalah pengumpulan dari beberapa sensus data yang distandarisasikan sehingga mempermudah perbandingan. Dataset ini memiliki jumlah record 256932 data dan terdiri dari 61 atribut.

Pada tahapan ini sebelum data diproses dengan menggunakan algoritma AOI maka data akan dipilah – pilah, dibersihkan terlebih dahulu dari data – data yang tidak relevan dan tidak dapat digunakan karena mengganggu performa dari operational database serta mengganggu hasil generalisasi dari program karena data tersebut tidak ada di file concept hierarchy-nya. Contoh data yang melalui proses pembersihan adalah data dengan value ? (tanda tanya) yang berarti data tersebut tidak jelas nilainya, sehingga data yang nantinya digunakan adalah data yang integrasinya baik.

F. Data Integration

Dikarenakan data diambil dari satu sumber yaitu data publik pada UCI Machine Learning University of California dan data diambil dalam bentuk teks dalam satu file maka data sudah pasti homogen jadi langkah ini tidak dilakukan dan dapat dilanjutkan ke langkah berikutnya dalam Knowledge Discovery in Database.

G. Data Selection

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Pada penelitian ini digunakan dua buah algoritma yaitu current AOI dan modified AOI. Pada penelitian modified AOI maka data yang digunakan adalah data yang ada dalam concept hierarchy dikarenakan semakin sedikit data yang diproses maka semakin cepat pula proses algoritma dalam menyelesaikan generalisasi data. Sesuai dengan algoritma yang digunakan maka pemilihan data tidak akan memberikan hasil akhir yang berbeda dengan penggunaan keseluruhan data sebagai input proses.

1. Dataset Adult

Pada dataset ini terdapat 48842 record dan memiliki 14 atribut. Karena implementasi menggunakan algoritma modified AOI maka atribut yang terpakai adalah atribut yang berada dalam concept hierarchy yaitu Workclass, education, marital_status, occupation dan native country. Setelah dataset ini dicleaning maka tersisa 32562 record yang bersih dari noise.

2. Dataset Breast Cancer

Pada dataset ini terdapat 699 record dan memiliki 10 atribut. Tidak semua atribut dipakai, tetapi hanya atribut yang terdapat pada concept hierarchy saja yaitu Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Bare Nuclei, dan Normal Nucleoli. Setelah dataset ini dicleaning ternyata tidak terdapat noise pada dataset ini sehingga data yang didapat tetap 699 record.

3. Dataset Census 1990

Pada dataset ini memiliki jumlah record 2458285 yang terdiri dari 68 atribut. Pada penelitian ini digunakan data yang sama sebagai pembandingan dengan algoritma current AOI sejumlah 9860 record tetapi dengan menggunakan atribut yang terdapat pada concept hierarchy saja yaitu iClass, iMarital, iMeans, iRelat1, dan iYearSch.

4. Dataset IPUMS

Dataset ini memiliki jumlah record 256932 data dan terdiri dari 61 atribut. Dikarenakan pada penelitian ini akan dibandingkan kecepatan penyelesaian generalisasi antar kedua algoritma maka data yang digunakan sama sejumlah 7520 record dan dengan atribut yang sesuai dengan concept hierarchy yaitu RELATE, MARST, EDUCREC, MIGRATES, dan TRANWORK

H. Data Transformation

Data Transformation dilakukan dengan menyederhanakan data sehingga data yang akan dicari knowledge-nya akan menjadi lebih ringkas dan mudah dipahami. Perubahan data / Transformasi data ini akan mengacu kepada concept hierarchy-nya sehingga tetap tidak mengubah maksud dari data tersebut.

Pada implementasi algoritma AOI yang diusulkan ini, yang pertama dilakukan adalah melakukan transformasi terhadap data terlebih dahulu sehingga, hanya atribut yang terdapat pada file concept hierarchy saja yang akan diproses lebih lanjut pada proses generalisasi menggunakan algoritma AOI. Data tersebut ditransformasikan saat user akan memilih atribut mana yang akan di learn dari data tersebut. Dikarenakan implementasi algoritma ini juga menggunakan data yang sama dengan implementasi algoritma AOI Characteristic Rule yang ada sekarang maka concept hierarchy yang digunakan juga sama dengan implementasi sebelumnya.

I. Proses Mining

Pada tahap inilah algoritma akan diimplementasikan dalam menemukan knowledge baru yang bisa didapat dalam data yang akan dimining.

1. Proses Mining pada dataset adult

Pada pengujian dataset adult ini terdapat 32562 buah data dan learning yang dipilih adalah education dan characteristic rule yang akan dipelajari adalah basic yang terdiri dari 2316 data.

Dari hasil penelitian menggunakan beberapa threshold mulai dari 2 sampai dengan 6 dan apabila data semakin besar dengan menggunakan threshold yang kecil maka didapatkan data yang kurang menarik dikarenakan memiliki banyak nilai ANY, seiring dengan penambahan threshold maka data yang didapatkan akan semakin memiliki nilai yang menarik.

2. Proses Mining pada dataset breast cancer

Pada pengujian dataset adult ini terdapat 699 buah data dan learning yang dipilih adalah clump thickness dan characteristic rule yang akan dipelajari adalah aboutAverClump yang terdiri dari 533 data.

Dari hasil penelitian menggunakan beberapa threshold mulai dari 2, 4, 8, 14 dan apabila data semakin besar dengan menggunakan threshold yang kecil maka didapatkan data yang kurang menarik dikarenakan memiliki banyak nilai ANY, seiring dengan penambahan threshold maka data yang didapatkan akan semakin memiliki nilai yang menarik

3. Proses Mining pada dataset Census 1990

Pada pengujian dataset Census ini terdapat 9860 buah data dan learning yang dipilih adalah means dan characteristic rule yang akan dipelajari adalah Green yang terdiri dari 5760 data.

4. Proses Mining pada dataset IPUMS

Pada pengujian dataset adult ini terdapat 7519 buah data dan learning yang dipilih adalah relate dan characteristic rule yang akan dipelajari adalah family yang terdiri dari 6931 data.

J. Pattern Evaluation

Dari Running Program baik current AOI maupun modified AOI menghasilkan generalisasi yang sama dan bisa didapatkan informasinya.

1. Hasil generalisasi dataset adult

- Untuk threshold 2 maka didapatkan generalisasi $V(x) = \text{Basic}(x) \rightarrow (\text{Occupation}(x) \in \text{Indoor}) [47.63\%] \vee (\text{Occupation}(x) \in \text{Outdoor}) [52.37\%]$

Dari distinct attributnya terlihat bahwa keseluruhan atribut menuju ke high level atribut tertingginya dikarenakan threshold yang kecil (2) sehingga “selective generalize attribute” dilakukan sehingga lebih mengerucut lagi ke high level atributnya.

- Untuk threshold 3 maka didapatkan generalisasi $V(x) = \text{Basic}(x) \rightarrow (\text{Occupation}(x) \in B) [35.19\%] \vee (\text{Occupation}(x) \in C) [52.37\%] \vee (\text{Occupation}(x) \in A) [12.44\%]$

Atribut “Occupation” tidak memiliki populasi distinct di level tertinggi pada concept hierarchynya, hal ini terlihat dengan nilai 0 pada 2nd pada tabel distinct atributnya. Hal ini menandakan bahwa generalisasi berakhir pada level ke 3rd dari concept hierarchy pada occupation karena sudah sesuai dengan threshold yang diberikan yaitu 3

- Untuk threshold 4 maka didapatkan generalisasi $V(x) = \text{Basic}(x) \rightarrow ((\text{Marital Status}(x) \in \text{Married}) \wedge (\text{Occupation}(x) \in \text{Indoor})) [17.56\%] \vee ((\text{Marital Status}(x) \in \text{Married}) \wedge (\text{Occupation}(x) \in \text{Outdoor})) [32.69\%] \vee ((\text{Marital Status}(x) \in \text{Unmarried}) \wedge (\text{Occupation}(x) \in \text{Indoor})) [30.27\%] \vee ((\text{Marital Status}(x) \in \text{Unmarried}) \wedge (\text{Occupation}(x) \in \text{Outdoor})) [19.69\%]$

Dari distinct atributnya terlihat bahwa keseluruhan atribut menuju ke high level atribut tertingginya sehingga populasi distinct atributnya terdapat pada keseluruhan level.

- Untuk threshold 5 maka didapatkan generalisasi $V(x) = \text{Basic}(x) \rightarrow (\text{Occupation}(x) \in B) [35.19\%] \vee (\text{Occupation}(x) \in C) [52.37\%] \vee (\text{Occupation}(x) \in A) [12.44\%]$

Hanya terdapat 3 buah distinct atribut pada atribut “Occupation” yaitu A, B, dan C hal ini sesuai dengan generalisasi yang berakhir pada level 3rd dari concept hierarchy jadi karena

threshold dari atributnya sudah sesuai maka generalisasi tidak dilanjutkan sampai ke level 2nd

- Untuk threshold 6 maka didapatkan generalisasi Basic(x) $\rightarrow ((\text{Marital Status}(x) \in \text{Married}) \wedge (\text{Occupation}(x) \in \text{B}))$ [11.70%] $\vee ((\text{Marital Status}(x) \in \text{Married}) \wedge (\text{Occupation}(x) \in \text{C}))$ [32.69%] $\vee ((\text{Marital Status}(x) \in \text{Unmarried}) \wedge (\text{Occupation}(x) \in \text{A}))$ [6.78%] $\vee ((\text{Marital Status}(x) \in \text{Unmarried}) \wedge (\text{Occupation}(x) \in \text{C}))$ [19.69%] $\vee ((\text{Marital Status}(x) \in \text{Unmarried}) \wedge (\text{Occupation}(x) \in \text{B}))$ [23.49%] $\vee ((\text{Marital Status}(x) \in \text{Married}) \wedge (\text{Occupation}(x) \in \text{A}))$ [5.66%]

Hasil dari distinct atributnya sama dengan threshold sebelumnya, terlihat dari tidak adanya level 2nd dari atribut "Occupation."

2. Hasil generalisasi dataset breast cancer

- Untuk threshold 2 maka didapatkan generalisasi $V(x) = \text{aboutAverClump}(x) \rightarrow (\text{Normal Nucleoli}(x) \in \text{AboutAverNucleoli})$ [91.56%] $\vee (\text{NormalNucleoli}(x) \in \text{aboveAverNucleoli})$ [8.44%]

Dari populasi distinct atributnya terlihat bahwa keseluruhan atribut sampai kepada high level atributnya (2nd) dikarenakan threshold yang kecil.

- Untuk threshold 4 maka didapatkan generalisasi $V(x) = \text{aboutAverClump}(x) \rightarrow ((\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [83.30%] $\vee ((\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [8.26%] $\vee ((\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboveAverNucleoli}))$ [3.19%] $\vee ((\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboveAverNucleoli}))$ [5.25%]

Dari populasi distinct atributnya terlihat bahwa keseluruhan atribut sampai kepada high level atributnya (2nd).

- Untuk threshold 8 maka didapatkan generalisasi $V(x) = \text{aboutAverClump}(x) \rightarrow ((\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [82.55%] $\vee ((\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [5.07%] $\vee ((\text{Uniformity of Cell Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [1.13%] $\vee ((\text{Uniformity of Cell Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [3.19%] $\vee ((\text{Uniformity of Cell Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboveAverNucleoli}))$ [3.56%] $\vee ((\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboveAverNucleoli}))$ [2.06%] $\vee ((\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboveAverNucleoli}))$ [1.69%] $\vee ((\text{Uniformity of Cell Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [0.75%]

Sama seperti threshold sebelumnya, keseluruhan high level atribut (2nd) memiliki anggota didalamnya.

- Untuk threshold 16 maka didapatkan generalisasi $V(x) = \text{aboutAverClump}(x) \rightarrow ((\text{Uniformity of Cell Size}(x) \in \text{aboutAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [82.36%] $\vee ((\text{Uniformity of Cell Size}(x) \in \text{aboutAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [82.36%] $\vee ((\text{Uniformity of Cell Size}(x) \in \text{aboutAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboveAverNucleoli}))$ [82.36%] $\vee ((\text{Uniformity of Cell Size}(x) \in \text{aboveAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [82.36%] $\vee ((\text{Uniformity of Cell Size}(x) \in \text{aboveAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [82.36%] $\vee ((\text{Uniformity of Cell Size}(x) \in \text{aboveAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboutAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [82.36%] $\vee ((\text{Uniformity of Cell Size}(x) \in \text{aboveAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboutAverNucleoli}))$ [82.36%] $\vee ((\text{Uniformity of Cell Size}(x) \in \text{aboveAverShape}) \wedge (\text{Uniformity of Cell Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare Nuclei}(x) \in \text{aboveAverNuclei}) \wedge (\text{Normal Nucleoli}(x) \in \text{aboveAverNucleoli}))$ [82.36%]

$\text{aboutAverNucleoli}) [4.69\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboveAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboutAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [1.13\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboveAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboveAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboutAverNucleoli})) [2.81\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboveAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboveAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [2.63\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboutAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboutAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [1.50\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboutAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboveAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [1.13\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboutAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboutAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [0.56\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboutAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboveAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [0.94\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboveAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboveAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [0.56\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$

$\text{aboutAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboveAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboutAverNucleoli})) [0.38\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboveAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboveAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboutAverNucleoli})) [0.38\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboveAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboveAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboutAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboveAverNucleoli})) [0.75\%] \vee$
 $((\text{Uniformity of Cell Size}(x) \in$
 $\text{aboveAverShape}) \wedge (\text{Uniformity of Cell}$
 $\text{Shape}(x) \in \text{aboutAverShape}) \wedge (\text{Bare}$
 $\text{Nuclei}(x) \in \text{aboutAverNuclei}) \wedge$
 $(\text{Normal Nucleoli}(x) \in$
 $\text{aboutAverNucleoli})) [0.19\%]$

Sama seperti threshold sebelumnya, ternyata threshold ini juga mengakibatkan keseluruhan atribut bertransformasi ke high level atributnya (2nd)

3. Hasil generalisasi dataset census

- Untuk threshold 2 maka didapatkan generalisasi $V(x) = \text{Green}(x) \rightarrow (\text{iMarital}(x) \in \text{No-Family}) [54.93\%] \vee (\text{iMarital}(x) \in \text{Have-family}) [45.07\%]$

Dari populasi distinct atributnya terlihat bahwa keseluruhan atribut sampai kepada high level atributnya (2nd). Threshold yang terlalu kecil akan mengakibatkan data sampai mengerucut melebihi high level atributnya dan menjadi ke level tertinggi yaitu ANY sehingga malah tidak menarik.

- Untuk threshold 5 maka didapatkan generalisasi $V(x) = \text{Green}(x) \rightarrow (\text{iMarital}(x) \in \text{Non Married}) [54.93\%] \vee (\text{iMarital}(x) \in \text{Married}) [31.67\%] \vee (\text{iMarital}(x) \in \text{Failed Marriage}) [13.40\%]$

Populasi atribut iMarital hanya sampai ke 3rd level saja, terlihat dari

iMarital yang terdiri dari Non Married, Married, dan Failed Marriage. Hal ini diakibatkan generalisasi telah sesuai threshold pada atribut level 3rd tersebut.

- Untuk threshold 6 maka didapatkan generalisasi $V(x) = \text{Green}(x) \rightarrow ((i\text{Marital}(x) \in \text{No-Family}) \wedge (i\text{Relat1}(x) \in \text{Family})) [47.64\%] \vee ((i\text{Marital}(x) \in \text{Have-Family}) \wedge (i\text{Relat1}(x) \in \text{Family})) [41.01\%] \vee ((i\text{Marital}(x) \in \text{No-Family}) \wedge (i\text{Relat1}(x) \in \text{Not-Family})) [3.44\%] \vee ((i\text{Marital}(x) \in \text{Have-Family}) \wedge (i\text{Relat1}(x) \in \text{Extend-Family})) [1.93\%] \vee ((i\text{Marital}(x) \in \text{No-Family}) \wedge (i\text{Relat1}(x) \in \text{Extend-Family})) [3.85\%] \vee ((i\text{Marital}(x) \in \text{Have-Family}) \wedge (i\text{Relat1}(x) \in \text{Not-Family})) [2.14\%]$

Pada threshold ini keseluruhan atribut mencapai high level atribut (2nd)

- Untuk threshold 8 maka didapatkan generalisasi $V(x) = \text{Green}(x) \rightarrow ((i\text{Marital}(x) \in \text{No-Family}) \wedge (i\text{YearSch}(x) \in \text{Intermediate})) [7.34\%] \vee ((i\text{Marital}(x) \in \text{Have-Family}) \wedge (i\text{YearSch}(x) \in \text{Basic})) [16.65\%] \vee ((i\text{Marital}(x) \in \text{No-Family}) \wedge (i\text{YearSch}(x) \in \text{Basic})) [33.85\%] \vee ((i\text{Marital}(x) \in \text{Have-Family}) \wedge (i\text{YearSch}(x) \in \text{Intermediate})) [26.39\%] \vee ((i\text{Marital}(x) \in \text{No-Family}) \wedge (i\text{YearSch}(x) \in \text{No-Education})) [13.51\%] \vee ((i\text{Marital}(x) \in \text{Have-Family}) \wedge (i\text{YearSch}(x) \in \text{Advance})) [1.35\%] \vee ((i\text{Marital}(x) \in \text{No-Family}) \wedge (i\text{YearSch}(x) \in \text{Advance})) [0.23\%] \vee ((i\text{Marital}(x) \in \text{Have-Family}) \wedge (i\text{YearSch}(x) \in \text{No Education})) [0.68\%]$

Dari populasi distinct atributnya terlihat bahwa keseluruhan atribut sampai kepada high level atributnya (2nd).

4. Hasil generalisasi dataset IPUMS

- Untuk threshold 2 maka didapatkan generalisasi $V(x) = \text{Family}(x) \rightarrow (\text{Tranwork}(x) \in \text{vehicle}) [36.20\%] \vee (\text{Tranwork}(x) \in \text{No-Vehicle}) [63.80\%]$

Dari populasi distinct atributnya terlihat bahwa keseluruhan atribut sampai kepada high level atributnya (2nd). Threshold yang terlalu kecil akan mengakibatkan data sampai mengerucut melebihi high level atributnya dan menjadi ke level tertinggi yaitu ANY sehingga malah tidak menarik.

- Untuk threshold 4 maka didapatkan generalisasi $V(x) = \text{Family}(x) \rightarrow (\text{Migrate5}(x) \in \text{Not-moved}) [43.27\%] \vee (\text{Migrate5}(x) \in \text{No-applicable}) [8.04\%] \vee (\text{Migrate5}(x) \in \text{No-applicable}) [48.69\%]$

Pada threshold ini keseluruhan atribut mencapai high level atribut (2nd)

- Untuk threshold 6 maka didapatkan generalisasi $V(x) = \text{Family}(x) \rightarrow ((\text{Marst}(x) \in \text{Married}) \wedge (\text{Tranwork}(x) \in \text{vehicle})) [24.74\%] \vee ((\text{Marst}(x) \in \text{Married}) \wedge (\text{Tranwork}(x) \in \text{No-vehicle})) [21.02\%] \vee ((\text{Marst}(x) \in \text{Unmarried}) \wedge (\text{Tranwork}(x) \in \text{No-vehicle})) [42.78\%] \vee ((\text{Marst}(x) \in \text{Unmarried}) \wedge (\text{Tranwork}(x) \in \text{vehicle})) [11.46\%]$

Sama seperti threshold sebelumnya, keseluruhan atribut juga mencapai high level atribut (2nd)

- Untuk threshold 8 maka didapatkan generalisasi $V(x) = \text{Family}(x) \rightarrow ((\text{Migrate5}(x) \in \text{Not-moved}) \wedge (\text{Tranwork}(x) \in \text{vehicle})) [12.94\%] \vee ((\text{Migrate5}(x) \in \text{Not-moved}) \wedge (\text{Tranwork}(x) \in \text{No-vehicle})) [10.17\%] \vee ((\text{Migrate5}(x) \in \text{No-applicable}) \wedge (\text{Tranwork}(x) \in \text{No-vehicle})) [0.37\%] \vee ((\text{Migrate5}(x) \in \text{Moved}) \wedge (\text{Tranwork}(x) \in \text{vehicle})) [41.96\%] \vee ((\text{Migrate5}(x) \in \text{Moved}) \wedge (\text{Tranwork}(x) \in \text{No-vehicle})) [34.57\%]$

Sama seperti threshold sebelumnya, keseluruhan atribut juga mencapai high level atribut (2nd)

V. PERBANDINGAN ALGORITMA AOI SEKARANG DENGAN ALGORITMA AOI USULAN

Untuk menganalisis perbandingan AOI yang ada sekarang dengan AOI yang diusulkan maka akan dibahas sesuai dengan kelemahan AOI yang didapatkan pada proses studi literatur, yaitu algoritma ini memerlukan proses repetitif berulang yang tentu saja membutuhkan waktu yang cepat dalam eksekusi generalisasinya sehingga knowledge dari data tersebut dapat lebih cepat ditemukan saat mencoba beberapa threshold yang berbeda. Untuk membandingkan kedua algoritma ini maka akan dibandingkan waktu eksekusi dalam satuan detik di dalam menyelesaikan proses generalisasi dari keempat data yang diujikan.

A. Perbandingan Algoritma AOI Sekarang Dengan AOI Usulan Pada Dataset Adult

Dari hasil penelitian ini juga didapatkan data waktu program untuk menyelesaikan generalisasi sebagai berikut :

Attribute Threshold	Rule Threshold	Algorithm Current AOI	Algorithm Modified AOI	Percentage
2	2	4.4862	0.4502	996.49%
3	3	4.3523	0.4104	1060.5%
4	4	3.8506	0.3509	1097.34%
5	5	3.0738	0.2801	1097.39%
6	6	2.7658	0.2215	1248.66%

Table 1. Perbandingan Antara Current AOI dan Modified AOI 1

Dari tabel diatas maka didapatkan data bahwa perbandingan antara current AOI dan modified AOI berselisih mulai dari 996.49% dan mempunyai rata-rata prosentase 1125.98% atau 11.25 kali lipat. Hasil tabel 1 menunjukkan bahwa dengan semakin besar threshold yang diberikan selisih tersebut akan semakin jauh. Algoritma current AOI akan semakin lama

bila dibandingkan dengan modified AOI dalam menyelesaikan generalisasi data.

B. Perbandingan Algoritma AOI Sekarang Dengan AOI Usulan Pada Dataset Breast Cancer

Dari hasil penelitian ini juga didapatkan data waktu program untuk menyelesaikan generalisasi sebagai berikut :

Attribute Threshold	Rule Threshold	Algorithm Current AOI	Algorithm Modified AOI	Percentage
2	2	0.2855	0.0850	335.88%
4	4	0.2230	0.0708	314.97%
8	8	0.2500	0.0724	345.30%
14	14	0.2750	0.0500	550.00%

Table 2. Perbandingan Antara Current AOI dan Modified AOI 2

Dari tabel diatas maka didapatkan data bahwa perbandingan antara current AOI dan modified AOI berselisih mulai dari 335.88% dan mempunyai rata-rata 386.54% atau 3.86 kali lipat. Dikarenakan jumlah record yang diproses lebih sedikit dibanding dengan data sebelumnya yaitu adult dataset yaitu 533 record, maka akan membuat algoritma current AOI dan modified AOI memproses data mendekati sama dalam prosesnya. Sama dengan penelitian atas data sebelumnya yaitu adult dataset, semakin besar threshold yang diberikan selisih tersebut akan semakin jauh dari yaitu dari 335.88% hingga 550%.

C. Perbandingan Algoritma AOI Sekarang Dengan AOI Usulan Pada Dataset Cencus 1990

Dari hasil penelitian ini juga didapatkan data waktu program untuk menyelesaikan generalisasi sebagai berikut :

Attribute Threshold	Rule Threshold	Algorithm Current AOI	Algorithm Modified AOI	Percentage
---------------------	----------------	-----------------------	------------------------	------------

old		t AOI	ed AOI	
2	2	41.41 17	0.555 6	7453.5 0%
5	5	48.24 84	0.586 8	8222.2 9%
6	6	34.77 70	0.543 9	6394.0 0%
8	8	36.10 67	0.525 1	6876.1 5%

Table 3. Perbandingan Antara Current AOI dan Modified AOI 3

Dari tabel diatas didapatkan data ternyata threshold yang besar tidak membuat perbandingan antara algoritma current AOI dan modified AOI semakin membesar perbedaannya. Persentase yang besar antara algoritma current AOI dan modified AOI diakibatkan oleh banyaknya attribute dari data census ini yaitu 68 attribute. Dari sini terlihat jelas langkah ke 3 dan 4 dari algoritma current AOI yang ditampilkan pada algoritma pertama diatas, akan menghabiskan waktu banyak untuk melakukan eliminasi data apabila data tidak ditemukan higher levelnya di concept hierarchy. Melakukan eliminasi terhadap 59 attribute tentu saja akan mempengaruhi kinerja dari algoritma ini bila dibandingkan dengan melakukan generalisasi terhadap data yang ada di concept hierarchy saja (4 attribute saja) seperti yang dilakukan oleh algoritma modified AOI. Perbandingan antara current AOI dan modified AOI berselisih mulai dari 6394.00% dan mempunyai rata-rata 7236.49% atau 72.36 kali lipat.

D. Perbandingan Algoritma AOI Sekarang Dengan AOI Usulan Pada Dataset IPUMS

Dari hasil penelitian ini juga didapatkan data waktu program untuk menyelesaikan generalisasi sebagai berikut :

Attribute Threshold	Rule Threshold	Algoritma Current AOI	Algoritma Modified AOI	Persentase
2	2	51.40 56	0.655 8	7838.6 0%
4	4	40.75	0.551	7387.9

		18	6	2%
6	6	39.96 01	0.605 5	6599.5 2%
8	8	37.76 01	0.529 3	7133.9 6%

Table 4. Perbandingan Antara Current AOI dan Modified AOI 2

Sama seperti dataset sebelumnya yang memiliki atribut banyak, dataset IPUMS ini juga memiliki atribut sebanyak 61 sehingga perbandingan kecepatan melakukan generalisasi menjadi jauh antara current AOI dan modified AOI. Perbandingan antara current AOI dan modified AOI berselisih mulai dari 6599.52% dan mempunyai rata-rata 7240.01% atau 72.40 kali lipat.

VI. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan maka didapatkan kesimpulan sebagai berikut :

Algoritma AOI akan melakukan generalisasi terhadap kumpulan data berdasarkan dari learning data yang akan dipelajari, concept hierarchy dari dataset tersebut dan threshold yang diberikan. Penentuan set data terbaik dilakukan dengan cara melakukan generalisasi lanjutan sampai kepada threshold yang telah diberikan terpenuhi. Dengan melakukan generalisasi lanjutan didapatkan data yang sudah tergeneralisir sesuai dengan concept hierarchy dari dataset yang diuji cobakan. Generalisasi lanjutan dilakukan dengan cara yang sama seperti yang dilakukan pada generalisasi umum hanya saja generalisasi lanjutan dibatasi sesuai dengan rule thresholdnya.

Untuk mendapatkan hasil yang baik dalam penelitian maka dilakukan percobaan terhadap suatu variabel dilakukan berkali kali dalam menentukan threshold atau batasan yang dapat menghasilkan data yang menarik karena nilai dari suatu threshold akan berbeda tergantung dari data yang akan diolah. Pada suatu data threshold tertentu bisa digunakan dan didapatkan data yang menarik namun pada suatu data yang lain threshold tersebut

tidak dapat digunakan untuk menemukan data yang menarik.

Sesuai dengan alasan tersebut diatas dan sudah dibuktikan dalam penelitian maka diperlukan efisiensi dari algoritma AOI agar lebih cepat dalam eksekusinya sehingga percobaan yang harus dilakukan berkali – kali untuk menentukan threshold yang tepat dapat lebih cepat dilakukan. Algoritma AOI yang diusulkan melalui penelitian didapatkan data dapat mempercepat eksekusi generalisasi terhadap dataset yang diujikan.

REFERENCES

- [1] J Han, O Cai, N Cercone, and Y Huang, "Discovery of Data Evolution Regularities in Large Databases," *Journal of Computer and Software Engineering*, vol. 3, no. 1, pp. 41-69, 1995.
- [2] J Han, Y Cai, and N Cercone, "Data-driven discovery of quantitative rules in relational databases.," *IEEE Trans on Knowl and Data Engin*, pp. 29 - 40, 1993.
- [3] J Han and Y Fu, "Exploration of the power of attribute-oriented induction in data mining in U. Fayyad, G.Piatetsky-Shapiro, P.Symth and R.Uthurasamy, eds," *Advances in Knowledge Discovery and Data Mining*, pp. 399-421, 1995.
- [4] J Han, Y Cari, and N Cercone, "Knowledge discovery in databases: An Attribute-oriented Approach," In *Proceedings of 18th International Conference on Very Large Databases*, pp. 547-559, 1992.
- [5] J Han, "Towards on-line analytical mining in large databases," *SIGMOD Rec*, vol. 27, no. 1, pp. 97-107, 1998.
- [6] J Han and Y Fu, "Dynamic Generation and Refinement of Concept Hierarchies for Knowledge Discovery in Databases.," In *Proceedings of AAAI Workshop on Knowledge Discovery in Databases*, pp. 157-168, 1994.
- [7] J Han, Y Fu, Y Huang, Y Cai, and N Cercone, "DBLearn: a system prototype for knowledge discovery in relational databases," *ACM SIGMOD Record*, vol. 23, no. 2, p. 516, 1994.
- [8] J Han et al., "DBMiner:A system for mining knowledge in large relational databases.," In *Proceedings Int'l Conf. on Data Mining and Knowledge Discovery*, pp. 250-255, 1996.
- [9] J Han et al., "DBMiner: a system for data mining in relational databases and data warehouses," In *Proceedings of the 1997 Conference of the Centre For Advanced Studies on Collaborative Research*, p. 8, 1997.
- [10] Y Cai, *Attribute-oriented induction in relational*, 1989, Master Thesis, Simon Fraser University.
- [11] Cai, Y; Cercone, N; Han, J;, "An Attribute-Oriented Approach for Learning Classification Rules from Relational Database," In *Proceeding of 6th International Conference on Data Engineering*, pp. 281-288, 1990.
- [12] Chen, M.S; Han, J; Yu, P.S., "Data Mining: An Overview from a Database Perspective," *IEEE Trans. on Knowl. and Data Eng*, vol. 8, no. 6, pp. 866-883, 1996.
- [13] D Fudger and H J Hamilton, "A Heuristic for Evaluating Databases for knowledge Discovery with DBLEARN.," In *Proceedings of the International Workshop on Rough Sets and Knowledge Discovery: Rough Sets, Fuzzy Sets and Knowledge Discovery*, pp. 44-51, 1993.

- [14] S Warnars, "Mining frequent pattern with Attribute Oriented Induction High Level Emerging Pattern (AOI-HEP)," In Proceeding IEEE the 2nd International Conference on Information and Communication Technology (IEEE ICoICT 2014), pp. 149-154, 2014.
- [15] S Warnars, "Mining Patterns with Attribute Oriented Induction", In Proceeding of The International Conference on Database, Data Warehouse, Data Mining and Big Data (DDDMBD2015), pp.11-21, 2015.
- [16] A. Frank and A. Asuncion, UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, 2010.

Check 45. Pengembangan Learning Characteristic Rule Pada Algoritma Data Mining Attribute Oriented Induction.pdf

ORIGINALITY REPORT

16%

SIMILARITY INDEX

17%

INTERNET SOURCES

7%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1	library.binus.ac.id Internet Source	5%
2	jsiskom.undip.ac.id Internet Source	4%
3	Submitted to Universitas Diponegoro Student Paper	3%
4	www.ksie.or.kr Internet Source	3%
5	arxiv.org Internet Source	1%
6	fmp.conncoll.edu Internet Source	1%

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On