**ALGORITHM TO LEARN CLASIFICATION RULES BASED ON THE EXTENDED ROUGH SET THEORY**

Yaima Filiberto; R Bello; M Frias – 2011 – Dyna Revista de la Facultad de Minas

**RINGKASAN:** The set kasar telah terbukti untuk efektif untuk mengembangkan teknik pembelajaran mesin, termasuk metode untuk penemuan aturan klasifikasi. Dalam makalah ini kami menyajikan sebuah algoritma untuk menghasilkan aturan klasifikasi berdasarkan hubungan kemiripan, sehingga berlaku dalam kasus di mana fitur yang domain diskrit atau kontinu. Hasil percobaan menunjukkan kinerja yang memuaskan dibandingkan dengan algoritma lainnya yang dikenal sebagai C4.5 dan MODLEM.

**KATA KUNCI:** Klasifikasi Aturan, hubungan kesamaan, teori set kasar, data mining

**Abstraksi:** Apakah diperbolehkan Rough set Beberapa teknik pembelajaran mesin Mengembangkan, Diantaranya metode untuk menemukan aturan klasifikasi. Dalam tulisan ini, kami menyajikan sebuah algoritma untuk menghasilkan aturan klasifikasi berdasarkan hubungan kemiripan, esta Memungkinkan untuk Terapkan metode ini dalam kasus aktual atau fitur Dengan domain diskrit. Hasil percobaan menunjukkan kinerja yang memuaskan dari algoritma esta dibandingkan Dengan sejenis lainnya yang C4.5 dan MODLEM.

**KATA KUNCI:** aturan Klasifikasi, hubungan kesamaan, teori himpunan Rough.

**1. PENDAHULUAN**

Aturan keputusan adalah salah satu bentuk khas representasi pengetahuan untuk meresmikan pengetahuan ditemukan, karena ekspresi simbolis dianggap lebih dimengerti dan alami bagi manusia dari formalisme lainnya; aturan tambahan blok pengetahuan dan domain ahli dapat dengan mudah menganalisa mereka secara individu [1], [2]. Aturan mewakili fungsi yang membangun hubungan antara contoh (dijelaskan oleh seperangkat sifat) dan jenis keputusan. Mereka dinyatakan dalam bentuk Jika P maka sebuah Q, di mana P adalah bagian bersyarat biasanya terdiri dari kombinasi dari kondisi dasar (p1 dan p2 dan ... pk), dan Q adalah bagian dari keputusan yang menetapkan keputusan nilai (kelas ) ke obyek yang memenuhi kondisi tersebut. Aturan pola yang membentuk ketergantungan antara nilai atribut kondisi P dan Q. nilai keputusan Dalam kebanyakan metode pembelajaran mesin untuk bagian aturan induksi P dibangun secara bertahap, kondisi berturut-turut menambahkan pi SD sampai kondisi tertentu terpenuhi; istilah pi adalah bukti dari nilai atribut (menggunakan operator kesetaraan, kurang dari, lebih besar dari, dll). Pencarian untuk kondisi SD terbaik yang akan ditambahkan ke kombinasi tergantung pada kriteria evaluasi, yang mencoba untuk mendorong aturan dengan kondisi yang lebih sedikit dan dengan cakupan terbesar mungkin; sehingga jumlah alternatif yang mungkin besar.

Belajar aturan klasifikasi adalah masalah klasik mesin belajar. Sebagian besar metode mencoba untuk menghasilkan aturan berikut strategi cakupan berurutan. Metode ini menggunakan satu set training (atau belajar) yang terdiri dari benda-benda yang digambarkan oleh kondisi atribut (dengan yang kondisi terbentuk pi) dan keputusan sifat (kelas). Pembangunan pengklasifikasi adalah salah satu teknik yang biasa digunakan dalam data mining [3]. Di antara algoritma klasik untuk memecahkan masalah ini adalah ID3, dan ekstensi untuk kasus kondisi sifat domain dengan C4.5 terus [4], yang menyebabkan pohon keputusan. Kehadiran sifat dengan domain terus menerus memperkenalkan kompleksitas yang lebih besar dalam proses menemukan aturan. Alternatif telah diikuti ciri discretized domain terus menerus sebelum pengolahan atau generasi aturan dilakukan secara bersamaan dan aturan diskritisasi induksi [5], seperti algoritma C4.5 [6], yang itu telah menjadi standar antara teknik pembelajaran mesin untuk menemukan aturan klasifikasi; dianggap di antara 10 algoritma yang paling berpengaruh dan memiliki penggunaan yang lebih besar dalam data mining [3].

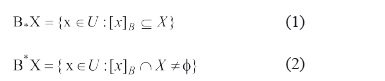
Teori set kasar (teori himpunan Rough, RST) menyediakan alat yang berguna untuk analisis data; termasuk algoritma untuk penemuan aturan klasifikasi (keputusan). Di antara yang paling dikenal adalah LEM2 (Belajar dari Contoh Modul v2), sebuah algoritma yang merupakan bagian dari sistem data mining lers (Belajar dari Contoh berdasarkan Set Rough) [7-9], dan dua algoritma berdasarkan LEM2, MODLEM [ 10-12], dan MLEM2 [13], yang mencoba untuk menemukan satu set minimal aturan yang memungkinkan untuk klasifikasi (yang berarti bahwa contoh training set ditutupi dengan jumlah minimum non - aturan berlebihan). LEM2 algoritma, mirip dengan ID3, hanya mempertimbangkan kondisi dasar dari bentuk atribut = nilai, sehingga dalam kasus domain sifat berkelanjutan memerlukan fase preprocessing di mana domain ini diskretisasi. The MODLEM dan MLEM2 algoritma dapat mempertimbangkan atribut sebagai diskritisasi domain terus menerus dilakukan secara bersamaan dan induksi dan menghasilkan P bagian dari aturan sebagai konjungsi dengan sintaks yang lebih umum. Algoritma MODLEM berusaha penutup pada perkiraan keputusan kelas menetapkan aturan minimum. Dalam [10] studi banding antara MODLEM dan LEM2 data numerik yang disajikan, hasil penelitian menunjukkan bahwa MODLEM mencapai hasil dengan akurasi yang sebanding dengan yang dicapai oleh varian terbaik LEM2 dicapai dengan mempertimbangkan beberapa alternatif diskritisasi. Hasil yang disajikan dalam [14] menunjukkan bahwa MODLEM mencapai algoritma yang sama kinerja C4.5. Penelitian lain pada penemuan aturan keputusan menggunakan set kasar disajikan di [11 : 15-20].

Artikel ini menyajikan sebuah metode untuk menginduksi aturan klasifikasi untuk sistem keputusan dengan fitur kondisi heterogen, yaitu, dapat eksis sebagai fitur diskrit kontinyu. Hal ini dibedakan dari atas yang tidak memerlukan discretize domain terus menerus, dan bagian bersyarat dari aturan tersebut tidak dinyatakan sebagai kombinasi dari kondisi dasar. Algoritma ini didasarkan pada penggunaan rasio kesamaan untuk membangun kelas kesamaan benda. Membangun hubungan kesamaan didasarkan pada teori diperpanjang [21] set kasar. Studi eksperimental yang dikembangkan menunjukkan kinerja yang lebih baik dari algoritma ini bila dibandingkan dengan C4.5 dan MODLEM.

**2. TEORI PAKET perkiraan**

Teori perkiraan Sets (Rough Set Theory, RST) diusulkan oleh Pawlak pada tahun 1982 [22]. Dengan mengembangkan berbagai teknik untuk menganalisa set kasar berdasarkan data, informasi diwakili oleh sebuah meja di mana setiap baris mewakili sebuah objek dan setiap kolom mewakili suatu sifat. Tabel ini disebut Sistem Informasi; lebih formal, itu adalah sepasang file gambar yang hilangDimana U adalah himpunan berhingga tak kosong dari objek yang disebut alam semesta dan A adalah himpunan berhingga tak kosong dari atribut. Sebuah sistem keputusan apapun sistem informasi sebagai file gambar yang hilangDimana hari adalah atribut keputusan. Konsep dasar RST adalah konsep dari bawah dan atas perkiraan. Definisi klasik pendekatan bawah dan atas pada awalnya diperkenalkan dengan mengacu pada hubungan ketidakterpisahan (indiscernibility hubungan) yang merupakan relasi ekivalen.

menjadi file gambar yang hilangsistem keputusan dan file gambar yang hilangdan file gambar yang hilang. B mendefinisikan relasi ekivalen dan X adalah bagian di alam semesta U. rasio konsep B didefinisikan sebagai berikut: obyek (x, y) tidak dapat dipisahkan sebagai B jika file gambar yang hilangUntuk semua fitur i di B, di mana ai (x) menunjukkan nilai dari fitur i dalam objek x (yaitu, dua benda yang tak terpisahkan sesuai dengan subset dari fitur B jika mereka memiliki nilai yang sama untuk semua sifat di B). Satu set X dapat didekati hanya menggunakan informasi yang terkandung dalam B dengan membangun bagian bawah dan perkiraan B-B-top, dilambangkan dengan B \* X \* X dan B masing-masing, dan didefinisikan oleh ekspresi:



dimana file gambar yang hilangmenunjukkan kelas x sesuai dengan hubungan ketidakterpisahan B. benda B \* X adalah anggota tertentu X, sedangkan benda yang mungkin B \* X X A anggota benteng adalah satu set minimal atribut seperti Bia bahwa file gambar yang hilang, Ie baik menghasilkan partisi yang sama semesta U.

Ketika domain fitur di B tidak diskrit, hubungan ketidakterpisahan didefinisikan dengan cara ini tidak berlaku. Rasio kesetaraan sangat ketat dalam hal domain berkelanjutan sebagai sedikit perbedaan antara nilai-nilai dari objek untuk suatu sifat mungkin tidak signifikan dalam menganalisis ketidakterpisahan mereka; misalnya, suhu 37,8 derajat dapat dianggap sama dengan 37,9 derajat lainnya, untuk mengukur suhu tubuh dua. Hal ini sangat penting dalam kasus fitur numerik di mana kesalahan kecil dalam pengukuran daripadanya dapat menghasilkan perbedaan-perbedaan ini. Dalam hal ini ada dua alternatif, fitur domain kontinyu discretized atau jenis lain dari hubungan antara objek tak terpisahkan U. Gunakan rasio kesamaan daripada relasi ekivalen alam semesta lebih tepat dalam kasus ini digunakan. Mengganti rasio kesetaraan dengan hubungan biner lemah, perpanjangan dari pendekatan klasik RST diperoleh. Beberapa ekstensi yang dibahas dalam [23-27].

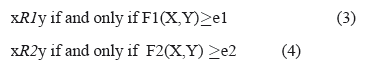
Hal ini dilakukan oleh memperluas konsep ketidakterpisahan objek sehingga mereka dikelompokkan dalam objek yang sama kelas yang sama, nonidentical, sebagai rasio rasio R. kesamaan kesamaan tidak menginduksi partisi dari alam semesta U, tapi menghasilkan kelas kesamaan untuk objek apapun file gambar yang hilang. Kelas kesamaan x sesuai dengan hubungan kesamaan R dinotasikan dengan R (x) dan didefinisikan sebagai file gambar yang hilang. Ini dibaca sebagai "set objek di alam semesta U yang mirip dengan yang objek x sesuai dengan rasio R". Contohnya adalah kasus hubungan toleransi (toleransi hubungan) [28], dimana rasio file gambar yang hilangHal ini reflektif ( file gambar yang hilang) Untuk setiap file gambar yang hilangdan simetris ( file gambar yang hilang) Untuk setiap pasang file gambar yang hilang.

Sementara ekuivalensi hubungan menginduksi partisi alam semesta, hubungan kesamaan menginduksi tudung dari alam semesta. Sebuah penutup dari alam semesta U adalah keluarga dari non - himpunan bagian kosong yang serikat sama dengan yang semesta. Sebuah partisi U adalah meliputi, sehingga konsep cakupan adalah perluasan dari konsep partisi.

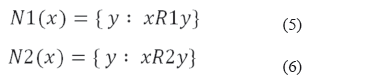
Dalam [29, 30], ukuran diusulkan, yang disebut kualitas kesamaan dengan kasus sistem keputusan terus menerus, ukuran ini memungkinkan untuk membangun hubungan kesamaan dalam RST diperpanjang, dan juga menghitung bobot untuk fitur dalam sistem keputusan terus menerus; menggunakan target heuristik, dalam hal ini berdasarkan partikel (PSO) [31], mengoptimalkan set bobot yang mencari ciri-ciri yang memaksimalkan kualitas ukuran kesamaan, seperti Bagian 3. dijelaskan berdasarkan metode ini membangun hubungan kemiripan telah mengembangkan penemuan algoritma aturan klasifikasi (keputusan), dijelaskan pada bagian 4.

**3. KONSTRUKSI HUBUNGAN KESAMAAN**

Metode membangun hubungan kesamaan diusulkan dalam [29, 30] pada prinsip bahwa "masalah yang sama memiliki solusi yang sama" (vektor input yang sama memiliki nilai output yang sama) Artinya, diberikan dua relasi .:



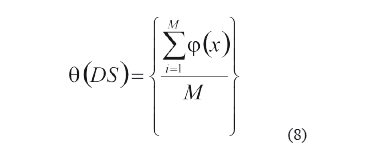
yang membangun hubungan kesamaan antara dua benda (x, y) mengingat kemiripan yang sama dengan sehubungan dengan fitur A (dihitung sesuai dengan fungsi F1 dalam kaitannya R1) dan sifat sasaran (dihitung sesuai dengan fungsi F2 semesta dalam kaitannya R2), tujuannya adalah untuk menemukan hubungan R1 dan R2 sehingga R1 (x) dan R2 (x) adalah sebagai dekat sebagai mungkin untuk setiap elemen alam semesta. Berdasarkan pendekatan ini, set dibangun:



Masalahnya adalah menemukan fungsi F1 dan F2 sehingga N1 (x) = N2 (x), di mana simbol "=" mungkin kesamaan terbesar antara set N1 (x) dan N2 (x) untuk setiap objek di alam semesta. The tingkat kesamaan antara yang dua set untuk sebuah objek x dinyatakan dengan ukuran berikut:

http://www.scielo.org.co/img/revistas/dyna/v78n169/a07eq07.gif

Dari yang didefinisikan oleh ekspresi (8) sebagai kualitas sistem keputusan kesamaan (DS) dengan alam semesta benda M:



Tujuannya adalah untuk memaksimalkan nilai dari ukuran q (DS). Nilai ukuran ini tergantung pada fungsi F1. Menggunakan jumlah tertimbang yang ditetapkan oleh ekspresi (11), dan mengingat fungsi perbandingan untuk setiap sifat, masalah berkurang untuk menemukan set beban W = {w1, w2, ..., wn}, yang menggunakan tujuan heuristik, seperti partikel - optimasi berbasis (Optimization partikel Swarm, PSO) [31].

**4. ATURAN PEMBANGKIT ALGORITMA KLASIFIKASI BERDASARKAN HUBUNGAN KESAMAAN**

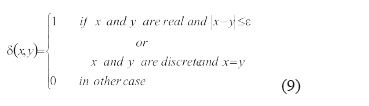
Algoritma untuk induksi aturan klasifikasi disajikan dalam bagian ini untuk menemukan pengetahuan dari sistem keputusan yang mengandung sifat dengan domain diskrit sebagai suatu terus menerus, karena perbedaan antara yang dua jenis domain hanya terletak pada fungsi perbandingan fitur gunakan; yang membuat tidak memerlukan membuat proses diskritisasi, baik sebelum atau belajar sebagai LEM2 ID3 atau C4.5 selama belajar sebagai atau MODLEM. Algoritma menginduksi aturan formulir Jika P maka Q, tetapi dalam kasus ini kondisi P memiliki bentuk Σwi \* di () ≥ε mana wi adalah bobot fitur i, di () adalah fungsi perbandingan untuk sifat i ε adalah ambang batas. Algoritma ini mencari set minimal aturan yang mencakup mengikuti strategi berurutan, yang membangun kelas kesamaan keputusan sistem benda. Algoritma ini terdiri dari modul utama dengan tiga langkah dan dua metode untuk membangun aturan.

**IRBASIR algoritma** (induksi aturan berdasarkan hubungan kemiripan)

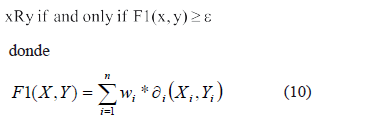
Karena sistem keputusan DS = (U, {d} æ), dengan m benda, dan menetapkan A mengandung n ciri domain kontinyu atau diskrit.

P1. Menentukan langkah-langkah kesamaan lokal.

Membangun fungsi fitur perbandingan di (x, y) untuk masing-masing sifat A, seperti yang membandingkan nilai-nilai sifat tersebut; misalnya ekspresi (9).



P2. Membangun kesamaan rasio R, seperti yang didefinisikan oleh (3):



P3. Membangun aturan klasifikasi sesuai dengan prosedur berdasarkan R. GenRulesRST

**Aktifkan GenRulesRST.**

Untuk menemukan nilai-nilai untuk wi (10) mengusulkan untuk menggunakan pendekatan yang diusulkan dalam [1] dan metode [2] diuraikan dalam bagian 3. Prosedur GenRulesRST menghasilkan aturan klasifikasi dan nilai-nilai yang sesuai mereka kepastian.

**prosedur GenRulesRST**

Berbagai komponen m, disebut Digunakan [digunakan], di mana Digunakan [i] memiliki nilai 1 jika benda itu sudah digunakan oleh prosedur GenRulesRST, atau 0 jika tidak. Dalam prosedur ini keputusan sistem benda yang tidak diperhitungkan sebelumnya dicari, jenisnya kesamaan dibangun menggunakan rasio yang dipilih kesamaan dan aturan yang mencakup benda-benda yang keputusan nilai kelas mayoritas di kelas dibangun kesamaan.

P1: Inisialisasi objek kontra

Digunakan [j] ← 0, untuk j = 1, ..., m   
RulSet ← f   
i ← 0

P2: Mulai pengolahan objek Oi

i ← indeks objek pertama yang tidak digunakan

Jika saya = 0 kemudian mengakhiri proses menghasilkan aturan.

Digunakan tapi [i] ← 1

Q3: Membangun semacam kemiripan Oi objek sebagai R

Hitung [Oi] R [x] R menunjukkan kelas kesamaan objek x

P4: Generasi aturan keputusan

Jika | f ([Oi] R) | = 1 maka {/ \* Membangun aturan yang menjelaskan jenis kesamaan dengan keputusan konsekuen sama dengan nilai objek Oi \* /

← k d (Oi)   
C ← [Oi] R

}

Else {/ \* Membangun aturan yang menjelaskan jenis kemiripan dengan konsekuen sama dengan nilai keputusan mayoritas [Oi] R \* /

← nilai k benda keputusan mayoritas [Oi] R   
C ← benda [Oi] k R berkelas

}

Aktifkan GenRulSim (k, [Oi] R, C, Rul) / \* Prosedur ini membangun aturan keputusan \* /

← RulSet RulSet U {Rul}

P5: Mark semua objek yang digunakan dalam C Rul ditutupi oleh aturan dan k konsekuen.

Q6: Pergi ke P2

Dimana | f ([Oi] R) | menunjukkan jumlah nilai keputusan selain benda-benda di kelas kesamaan [Oi] R.

**Prosedur GenRulSim (k, Cs, C; Rul)**

K, Cs dan C parameter yang input dan output Rul.

P1: Membangun vektor dengan n komponen referensi P ke set dari objek dalam C.

P (i) ← f (Vi) di mana Vi adalah seperangkat nilai-nilai i trait pada objek di C

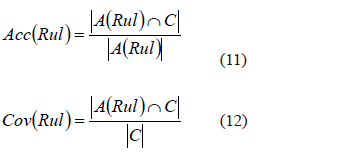
P2: Menghasilkan aturan dari vektor referensi P.

← Rul Jika w1 \* d (X1, P1) + ... + wn \* d (Xn, Pn) ≥ dan kemudian sebuah k d =

Di mana bobot wi diambil dari fungsi F1 (Expression 10); e adalah ambang batas yang digunakan dalam rasio kesamaan R; Pi adalah nilai fitur i dalam vektor referensi P; dan di adalah fungsi perbandingan untuk fitur i.

Q3: Hitung kepastian aturan.

Mempertimbangkan langkah-langkah akurasi (Acc) dan cakupan (COV):



Di mana A (Rul) adalah sekumpulan objek Cs yang anteseden dari aturan terpenuhi Rul.

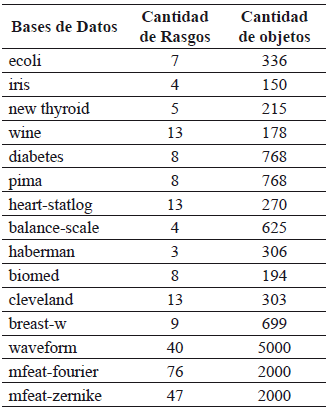
GenRulSim pada langkah P1 fungsi f menunjukkan operator agregasi; misalnya: Jika nilai penggunaan nyata Vi rata, jika mereka terlalu mencolok. Tujuannya adalah untuk membangun sebuah prototipe atau massa untuk satu set objek serupa; di sini telah menggunakan pendekatan yang sama seperti yang dilakukan pada algoritma k-means [3]; di [32] tentang berbagai alternatif disajikan.

**5. HASIL PERCOBAAN**

Dalam rangka untuk menilai akurasi algoritma itu dilakukan studi berikut. 12 database UCI Repository [33] di mana sebagian besar atribut A adalah kontinu dan diskrit fitur domain d digunakan. Untuk memvalidasi hasil silang - proses validasi yang digunakan.

Salib - validasi metode, membagi data asli ditetapkan dalam subset k dengan ukuran yang sama, salah satu yang digunakan sebagai uji ditetapkan sementara yang lain membentuk training set. Kemudian akurasi keseluruhan classifier dihitung sebagai rata-rata dari presisi diperoleh dengan semua tes subset. Dengan menggunakan teknik ini jumlah sub - rakitan dengan yang bekerja diperhitungkan, karena dengan nilai yang lebih tinggi untuk K, tren berbagai kesalahan yang sebenarnya dari estimasi berkurang, perkiraan akan lebih akurat, varians kesalahan kisaran nyata lebih tinggi dan waktu komputasi yang diperlukan juga meningkat karena meningkatkan jumlah percobaan yang dilakukan. Sebagai k menjadi lebih kecil, jumlah percobaan dan akibatnya waktu komputasi berkurang, varians estimasi juga berkurang dan kecenderungan estimasi akan menjadi lebih besar (konservatif atau lebih besar dari kisaran kesalahan yang sebenarnya) . Menempel ini bekerja dengan khas direkomendasikan nilai k = 10. Teknik ini menghilangkan masalah tumpang tindih set pengujian dan membuat penggunaan efektif dari semua data yang tersedia.

Dasar yang digunakan dalam penelitian ini dijelaskan dalam [Tabel 1](https://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=es&sp=nmt4&tl=id&u=http://www.scielo.org.co/scielo.php%3Fscript%3Dsci_arttext%26pid%3DS0012-73532011000500007&usg=ALkJrhgAUPkXAdKZ_DSpTW-Iz9VkXCt5oA#tab01) .

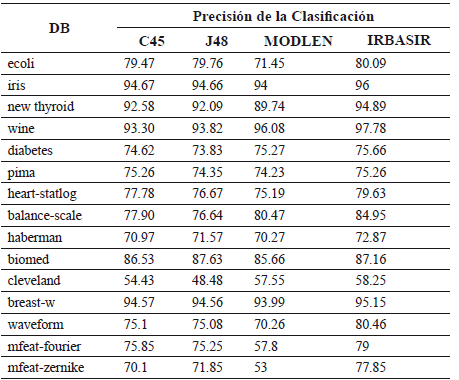
**Tabel 1.** Deskripsi database yang digunakan untuk eksperimen   


Untuk hasil eksperimen yang diperoleh dengan dua versi dari algoritma dibandingkan C4.5 (pengklasifikasi C4.5 alat KEEL [34] dan Weka J48 [35] tool), algoritma alat MODLEM Rose2 [36 ] dan diusulkan algoritma (IRBASIR).

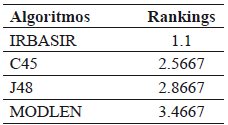
Percobaan 1: Bandingkan hasil akurasi C4.5, J48, MODLEM dan IRBASIR pengklasifikasi untuk setiap BD, selama sepuluh partisi yang dibuat.

Untuk menentukan ketepatan klasifikasi C4.5, J48, MODLEM dan IRBASIR untuk setiap pengklasifikasi BD.

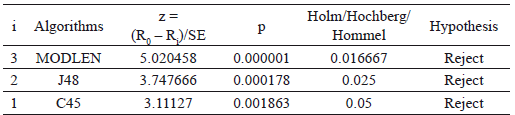
Dalam [Tabel 2](https://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=es&sp=nmt4&tl=id&u=http://www.scielo.org.co/scielo.php%3Fscript%3Dsci_arttext%26pid%3DS0012-73532011000500007&usg=ALkJrhgAUPkXAdKZ_DSpTW-Iz9VkXCt5oA#tab02) hasil perbandingan algoritma ditampilkan.

**Tabel 2.** Hasil ketepatan klasifikasi C45, J48, algoritma MODLEM dan IRBASIR   


Dalam semua kasus akurasi klasifikasi mencapai ketika aturan diinduksi algoritma IRBASIR digunakan lebih tinggi daripada yang dihasilkan oleh C4.5, J48 dan algoritma MODLEM. Membandingkan hasil tes untuk beberapa perbandingan untuk menemukan algoritma yang terbaik digunakan. Dalam [Tabel 3](https://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=es&sp=nmt4&tl=id&u=http://www.scielo.org.co/scielo.php%3Fscript%3Dsci_arttext%26pid%3DS0012-73532011000500007&usg=ALkJrhgAUPkXAdKZ_DSpTW-Iz9VkXCt5oA#tab03) kita dapat melihat bahwa peringkat terbaik memiliki IRBASIR metode yang diusulkan.

**Tabel 3.** Peringkat diperoleh dengan uji Friedman   


Tes Iman-Davenport (F-distribusi dengan 3 dan 42 derajat kebebasan) dipekerjakan untuk menemukan perbedaan yang signifikan antara IRBASIR, C4.5, J48 dan MODLEM algoritma, memperoleh oleh Friedman nilai p- uji nilai: 0,000005. Jadi dalam [Tabel 4](https://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=es&sp=nmt4&tl=id&u=http://www.scielo.org.co/scielo.php%3Fscript%3Dsci_arttext%26pid%3DS0012-73532011000500007&usg=ALkJrhgAUPkXAdKZ_DSpTW-Iz9VkXCt5oA#tab04) hasil prosedur Holm adalah untuk membandingkan algoritma yang diusulkan dengan IRBASIR tersisa. tes menolak semua kasus untuk algoritma peringkat yang lebih baik sehingga kita dapat melihat bahwa algoritma yang diusulkan secara statistik lebih tinggi dari yang lain dalam hal akurasi klasifikasi.

**Tabel 4.** Tabel Holm untuk α = 0,05, dengan IRBASIR sebagai suatu metode pengendalian.   


**6. KESIMPULAN**

Makalah ini menyajikan algoritma IRBASIR (induksi aturan berdasarkan hubungan kemiripan) untuk menghasilkan aturan klasifikasi, hubungan kerja kesamaan dapat mengobati kedua fitur dengan domain diskrit sebagai terus menerus tanpa diskretisasi domain terus menerus. Membangun hubungan kesamaan didasarkan pada teori set kasar diperpanjang. algoritma menginduksi aturan di mana bagian bersyarat daripadanya yang tidak dinyatakan sebagai kombinasi dari kondisi dasar, seperti dalam kebanyakan metode. Bila dibandingkan dengan algoritma lain yang dikenal (C4.5, J48 dan MODLEM) untuk menemukan aturan untuk jenis informasi ini hasilnya menunjukkan kinerja yang unggul dari IRBASIR algoritma.