

Deteksi Anomali pada Proses Bisnis dengan *Fuzzy Multi Criteria Decision Making* TOPSIS dan *Decision Tree*

Uswatun Hasana Kunio, Riyanarto Sarno dan Abdul Munif

Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

Jl. AriefRahman Hakim, Surabaya 60111 Indonesia

e-mail: riyaranto@if.its.ac.id

Abstrak—*Internal fraud* atau kecurangan dalam perusahaan menjadi salah satu penyebab kerugian yang paling signifikan. Fakta ini mendorong perusahaan untuk memiliki kebijakan dalam hal keamanan dan sistem informasi yang kuat dalam deteksi *fraud*. Penelitian ini fokus kepada *fraud* berbasis proses. *Fraud* berbasis proses adalah suatu kecurangan yang terjadi dalam proses bisnis. Terdapat sepuluh tipe kriteria anomali terjadinya *fraud* yaitu *Skip Sequence*, *Skip Decision*, *Throughput Time Maximum*, *Throughput Time Minimum*, *Wrong Pattern*, *Wrong Decision*, *Wrong Duty Decision*, *Wrong Duty Sequence*, *Wrong Duty Combine* serta *Wrong Resource*. Untuk menangkap anomali ini dilakukan pemodelan ontologi terhadap *Standard Operating Procedure* dan *event logs*. Pembuatan *rules* pada model ontologi menggunakan *SWRL rules* digunakan untuk mendapatkan kasus yang terindikasi terjadi anomali. *Fuzzy Multi Criteria Decision Making* TOPSIS digunakan untuk perhitungan *rating* anomali dan *Decision Tree* digunakan untuk klasifikasi status anomali dari tiap kasus yang ada pada *event logs*. Dalam artikel ini dilakukan uji coba terhadap studi kasus pengajuan kredit bank dan menghasilkan pendeteksian anomali yang akurat dengan akurasi sebesar 98%, *specificity* sebesar 1, dan *sensitivity* sebesar 0,71.

Kata Kunci—*Decision Tree*, deteksi anomali, *event logs*, model ontologi, TOPSIS.

I. PENDAHULUAN

INTERNAL *fraud* atau kecurangan dalam perusahaan telah menjadi salah satu penyebab kerugian yang paling signifikan. Terdapat 1.483 kasus *fraud* yang telah menyebabkan kerugian sebesar US\$1.700.000.000 di 100 negara [1]. Rata-rata suatu perusahaan telah mengalami kehilangan laba kotor sebesar 7% dalam satu tahun, 20% pegawai pada perusahaan tidak pernah melakukan kecurangan, 60% bergantung pada kesempatan untuk dapat melakukan kecurangan, dan 20% benar-benar melakukan kecurangan [2]. Fakta ini mendorong perusahaan untuk memiliki kebijakan dalam hal keamanan dan sistem informasi yang kuat dalam penanggulangan *fraud*.

Definisi dari *fraud* sendiri adalah penyalahgunaan dari sistem organisasi tanpa melanggar hukum [3]. Menurut referensi lain, *fraud* adalah tindak pidana menggunakan informasi palsu untuk mendapatkan keuntungan dengan cara tidak benar [4]. Penelitian ini fokus kepada *fraud* berbasis proses. *Fraud* berbasis proses adalah suatu kecurangan yang

terjadi dalam proses bisnis [5]. Berdasarkan penelitian sebelumnya, ada beberapa tipe kriteria terjadinya *fraud* [6]. Kemunculan dari kriteria *fraud* ini didefinisikan sebagai anomali atau penyimpangan. Namun, anomali yang terjadi tidak selalu mengindikasikan perusahaan mengalami kerugian olehnya.

Process mining dari *event logs* dapat mengurangi kejadian *fraud* di dalam internal proses bisnis [7]. Namun, metode *process mining* saat ini belum dapat memenuhi kebutuhan pengecekan data yang tersimpan dalam *event logs* (atribut). Untuk itu dalam penelitian ini *event logs* dan *standard operating procedure* (SOP) diubah ke dalam model ontologi sehingga dapat dilakukan pengecekan atribut dan kesesuaian dengan SOP menggunakan Pellet Reasoner. Selain itu dengan melakukan perubahan ke model ontologi berguna untuk mendapatkan anomali yang samar.

Pellet Reasoner yang menggunakan *SWRL rules* digunakan untuk mengambil *event* dalam *event logs* yang terdapat anomali serta melakukan pengecekan data yang tersimpan dalam *event logs*. Pengambilan *event* menggunakan *rules* dilakukan dengan cara membandingkan *event logs* dengan SOP yang telah dimodelkan ke model ontologi. *Fuzzy Multi Criteria Decision Making* TOPSIS digunakan untuk perhitungan *rating* anomali dan *Decision Tree* digunakan untuk mendapatkan pola-pola yang mengindikasikan terjadinya anomali.

II. KAJIAN PUSTAKA

A. *Fraud* dan Anomali

Fraud merupakan suatu bentuk kecurangan yang terjadi dalam sebuah perusahaan dan merugikan perusahaan. *Fraud* dapat terjadi dikarenakan adanya kesempatan yang dimiliki oleh pelaku yang memiliki otorisasi tertentu dalam sebuah perusahaan. *Fraud* memiliki definisi sebagai penyalahgunaan jabatan untuk melakukan kecurangan tanpa melanggar aturan [8].

Anomali merupakan penyimpangan pada proses-bisnis-yang-sedang-terjadi dari model-standar-proses-bisnis-yang-ada atau SOP. Anomali inilah yang dapat mengindikasikan terjadinya *fraud*. Berdasarkan penelitian yang telah dilakukan [6], terdapat enam tipe atribut atau kriteria anomali

indikasi akan terjadinya *fraud*, yaitu *Skipped Activity*, *Wrong Throughput Time*, *Wrong Resource*, *Wrong Duty*, *Wrong Pattern*, dan *Wrong Decision*. Beberapa kriteria kemudian dipecah ke dalam beberapa subkriteria. *Skipped Activity* memiliki subkriteria *Skip Decision* dan *Skip Sequence*. *Wrong Throughput Time* memiliki subkriteria *Throughput Time Maximum* dan *Throughput Time Minimum*. Dan *Wrong Duty* memiliki subkriteria *Wrong Duty Sequence*, *Wrong Duty Decision*, serta *Wrong Duty Combine*.

B. Model Ontologi

Model Ontologi adalah sebuah struktur hirarki dari istilah-istilah untuk menggambarkan taksonomi dan jaringan klasifikasi. *The Web Ontology Language* (OWL) merupakan salah satu bahasa untuk pemodelan ontologi.

Ontologi OWL merupakan kumpulan dari aksiom yang dapat memberikan pernyataan tersurat mengenai tiga hal yaitu kelas, individu dan *property*. Individu merupakan anggota dari kelas, sedangkan *property* adalah sifat atau atribut yang dimiliki oleh individu tersebut. Terdapat dua jenis *property* di ontologi OWL. Yang pertama adalah *data property*, yaitu relasi biner yang menghubungkan individu ke suatu data tertulis. Yang kedua adalah *object property*, yaitu relasi biner yang menghubungkan individu satu dengan yang lain.

Untuk dapat menarik suatu fakta relasi yang secara implisit terkandung di dalam ontologi dibutuhkan suatu perangkat lunak yang disebut *reasoner*. Fakta tersebut ditarik berdasarkan *rules* yang telah dibuat. Salah satu *reasoner* yang digunakan dalam penelitian ini adalah Pellet Reasoner. Pellet Reasoner menggunakan *Semantic Web Rule Language* (SWRL) sebagai bahasa *rules* [9].

C. Event Logs

Event logs adalah sekumpulan rekaman kejadian yang dieksekusi pada saat proses bisnis dijalankan. *Event logs* berisi informasi mengenai aktivitas apa yang dijalankan oleh siapa, kapan, serta data apa saja yang diterima. *Event logs* tersusun atas *event id*, *case id*, *activity*, *start time*, *complete time*, serta *resource* seperti yang ditunjukkan pada Tabel 1. *Event id* merupakan nomor *event*. *Case id* menunjukkan satu proses instan yang berjalan dari awal sampai akhir. *Activity* merupakan aktivitas apa yang dieksekusi pada suatu *event*. *Start time* menunjukkan waktu sebuah aktivitas mulai dieksekusi. *Complete time* menunjukkan waktu berakhir dari aktivitas yang telah dieksekusi. *Resource* adalah aktor yang melakukan aktivitas yang dieksekusi [7].

Tabel 1. Contoh *event logs*

Event ID	Case ID	Activity	Start Time	Complete Time	Resource
1	1	Receive	1/23/2015 8:00	1/23/2015 8:07	Ani
2	1	Check_SID	1/23/2015 8:17	1/23/2015 8:30	Budi
3	2	Check_SID	1/23/2015 8:18	1/23/2015 8:25	Chika
...

III. METODOLOGI

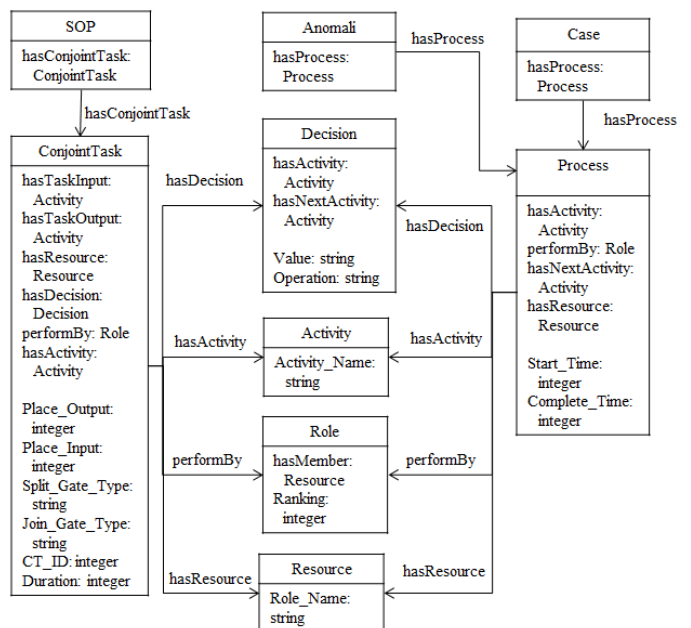
A. Dataset

Dataset merupakan data simulasi yang digunakan sebagai data latih dan data uji. Data ini merupakan data simulasi dengan studi kasus pengajuan kredit bank. Data ini merupakan hasil dari penelitian sebelumnya [10].

Label untuk tiap baris data yang ada pada *dataset* merupakan label dari pakar. Label pakar ini didapatkan dari hasil empat metode, yaitu *Fuzzy Multi Attribute Decision Making*, *Multi Criteria Decision Making Weighted Sum Method*, *Fuzzy Analytical Hierarchy Process Multi Criteria Decision Making*, dan *Fuzzy Multi Criteria Decision Making TOPSIS*. Penyesuaian label dilakukan dengan melihat kecenderungan label dari setiap pakar. Kecenderungan yang dimaksud adalah apabila jumlah label Anomali lebih banyak untuk suatu *case*, maka label dari pakar akan menjadi Anomali. Berlaku juga hal yang sama untuk Semi Anomalian dan Tidak Anomali. Apabila tidak terdapat kecenderungan pada label, maka dibuatkan aturan untuk memutuskan label pakar tersebut. Aturan tersebut adalah jika Anomali dan Semi Anomali seimbang, maka label dari pakar menjadi Semi Anomali. Jika Semi Anomali dan Tidak Anomali seimbang maka label dari pakar menjadi Tidak Anomali. Dan jika Anomali dan Tidak Anomali seimbang maka label dari pakar menjadi Tidak Anomali.

B. Konversi SOP dan Event Logs ke Model Ontologi

Proses konversi SOP dan *event logs* ke model ontologi membutuhkan dua masukan. Masukan yang pertama yaitu model proses bisnis SOP berupa *file* Petri Net berekstensi PNML. Masukan yang kedua berupa *file event logs* yang dihasilkan dari proses bisnis yang dijalankan dengan ekstensi XES. Kedua *file* ini disatukan dalam satu *file* model ontologi yang sama. Gambar 1 menggambarkan model ontologi hasil penggabungan model tersebut.



Gambar 1. Pemodelan SOP dan *event logs* ke dalam model ontologi

C. Cek Pelanggaran Menggunakan SWRL Rules

Model ontologi dapat menggambarkan setiap objek sebagai kelas dan memberi relasi antar objek-objek tersebut. Pada penelitian ini salah satu kelebihan model ontologi yaitu dapat mendeteksi Anomali, Semi Anomali, dan Tidak Anomali.

Untuk dapat menangkap anomali yang ada di model ontologi, dibutuhkan pembuatan SWRL rules. Setiap kriteria anomali memiliki rules yang berbeda untuk dapat menangkap pelanggaran yang ada.

1. Skip Sequence dan Skip Decision

Anomali ini terjadi ketika terdapat aktivitas yang seharusnya dieksekusi berdasarkan model proses bisnis SOP namun berdasarkan event logs tidak dieksekusi. Terdapat dua jenis skipped activity, yang pertama Skip Sequence, apabila aktivitas yang dilewati merupakan aktivitas yang tidak memiliki percabangan, yang kedua Skip Decision, apabila aktivitas yang dilewati memiliki percabangan.

2. Throughput Time Maximum dan Throughput Time Minimum

Anomali yang terjadi ketika durasi eksekusi suatu aktivitas tidak sesuai dengan standar durasi eksekusi pada model proses bisnis SOP. Throughput Time Maximum yaitu apabila durasi eksekusi aktivitas di atas durasi normal dan Throughput Time Minimum yaitu apabila aktivitas yang dikerjakan di bawah durasi normal.

3. Wrong Resource

Pada model proses bisnis SOP telah ditentukan resource yang berhak untuk mengeksekusi suatu aktivitas. Pada satu aktivitas terdapat beberapa orang yang berhak mengeksekusi aktivitas tersebut. Pembagian ini sudah ditentukan di dalam SOP. Pelanggaran WrongResource terjadi ketika ada aktivitas yang dieksekusi oleh resource yang tidak memiliki wewenang untuk mengerjakan aktivitas tersebut. Namun terdapat suatu pengecualian, resource tersebut boleh mengerjakan aktivitas tersebut apabila resource tersebut memiliki role yang lebih tinggi dari role yang diperbolehkan untuk mengeksekusi aktivitas tersebut. Jika role resource tersebut sama, maka akan dianggap sebagai pelanggaran dan termasuk ke dalam anomali Wrong Resource.

4. Wrong Duty Sequence, Wrong Duty Decision, dan Wrong Duty Combine

Wrong duty adalah anomali ketika terdapat resource yang mengeksekusi aktivitas yang bukan menjadi hak role-nya. Namun apabila role dari resource tersebut lebih tinggi maka tidak menjadi masalah, dengan catatan resource tersebut tidak melakukan dua atau lebih aktivitas yang berbeda sekaligus dalam satu case yang sama. Terdapat tiga jenis WrongDuty, yaitu Wrong Duty Sequence apabila terjadi pada aktivitas sequence, Wrong Duty Decision apabila terjadi pada aktivitas decision, dan Wrong Duty Combine apabila terjadi pada aktivitas sequence dan decision.

5. Wrong Pattern

Anomali ini terjadi ketika urutan aktivitas yang dieksekusi tidak sesuai dengan model proses bisnis SOP.

6. Wrong Decision

Anomali ini terjadi ketika terdapat pengambilan keputusan yang tidak sesuai yang terjadi di aktivitas decision. Keputusan tersebut tidak sesuai dengan yang telah ditetapkan pada SOP. Pada aktivitas decision terdapat keterangan atribut yang akan menentukan ke arah jalur mana proses akan berjalan berikutnya.

D. Perhitungan Bobot Kriteria dengan Pairwise Comparison

Perhitungan bobot kepentingan kriteria dilakukan dengan pendekatan integrasi, yaitu menggabungkan pendekatan subjektif dan objektif. Pendekatan subjektif dalam tugas akhir ini didapatkan dari pendapat para ahli mengenai tingkat kepentingan antar kriteria. Karena hasil dari pendekatan subjektif adalah pernyataan linguistik, maka data diolah menjadi bilangan numerik dengan menggunakan pairwise comparison.

Pairwise comparison diukur menggunakan skala, yaitu Skala Saaty. Skala yang memetakan antara set diskrit dari pilihan linguistik dengan set diskrit dari angka yang merepresentasikan kepentingan atau berat dari pilihan linguistik tersebut. Kesamaran yang ada pada Skala Saaty untuk bilangan fuzzy diperbaiki oleh van Laarhoven dan Pedryz [12] dengan menetapkan fungsi keanggotaan triangular untuk pairwise comparison seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Skala kepentingan dalam fuzzy triangular

Skala Saaty	Definisi	Skala Fuzzy Triangular
1	Equally important	(1, 1, 1)
3	Weakly important	(2, 3, 4)
5	Fairly important	(4, 5, 6)
7	Strongly important	(6, 7, 8)
9	Absolutely important	(9, 9, 9)
2	Nilai tengah antara 2 pendapat	(1, 2, 3)
4	Nilai tengah antara 2 pendapat	(3, 4, 5)
6	Nilai tengah antara 2 pendapat	(5, 6, 7)
8	Nilai tengah antara 2 pendapat	(7, 8, 9)

Tabel 3. Data tingkat kepentingan antar kriteria

Kriteria	Penilaian			
	Pakar 1	Pakar 2	Pakar 3	Pakar 4
Skip Sequence	W	W	W	W
Skip Decision	VI	VI	VI	VI
Throughput Time Minimum	F	I	I	I
Throughput Time Maximum	F	F	I	F
Wrong Resource	W	W	W	W
Wrong Duty Sequence	VW	VW	VW	VW
Wrong Duty Decision	W	W	W	W
Wrong Duty Combine	W	W	W	W
Wrong Pattern	W	W	W	W
Wrong Decision	VI	VI	VI	VI

Keterangan:
 VW = Very Weak
 W = Weak
 F = Fair
 I = Important
 VI = Very Important

Dalam penelitian ini, data penilaian tingkat kepentingan antar kriteria diambil dari penelitian yang dilakukan oleh Solichul Huda [6] kepada empat pakar audit suatu bank. Data tingkat kepentingan tersebut dapat dilihat pada Tabel 3. Tingkat kepentingan kriteria didefinisikan dengan penilaian secara lingual ke dalam lima kelas. Kelas tersebut dari tingkatan paling rendah ke tinggi yaitu *Very Weak* (VW), *Weak* (W), *Fair* (F), *Important* (I), dan *Very Important* (VI).

Dari Tabel 3 dilakukan perhitungan rerata geometrik (\tilde{r}_i) dengan menggunakan (1) [13]. Dengan n merupakan jumlah kriteria, dan d_{ij} merupakan nilai perbandingan fuzzy untuk kriteria ke- ij .

$$\tilde{r}_i = \left(\prod_{j=1}^n \tilde{d}_{ij} \right)^{\frac{1}{n}}, i = 1, 2, \dots, n \tag{1}$$

Rerata geometrik dari nilai perbandingan fuzzy untuk semua kriteria ditampilkan seperti pada Tabel 4. Kemudian, nilai total dan nilai *reverse* juga ditampilkan. Karena angka fuzzytriangular harus dalam bentuk urutan maju, maka urutan angka *reverse* dibalik. Untuk mencari bobot fuzzy \tilde{w}_i dari tiap kriteria ke- i , kalikan tiap \tilde{r}_i dengan nilai *reverse* seperti dalam (2). Dengan lw_i adalah bobot fuzzy lower ke- i , mw_i adalah bobot fuzzy middle ke- i , dan uw_i adalah bobot fuzzy upper ke- i .

Tabel 4.
Rerata geometrik dari nilai perbandingan fuzzy

Kriteria	\tilde{r}_i		
	Lower	Middle	Upper
<i>Skipped Activity</i>	1,26	1,82	2,29
<i>Wrong Throughput Time</i>	0,83	1,12	1,44
<i>Wrong Resource</i>	0,46	0,66	1,07
<i>Wrong Duty</i>	0,66	0,93	1,28
<i>Wrong Pattern</i>	0,35	0,47	0,74
<i>Wrong Decision</i>	1,12	1,70	2,38
Total	4,68	6,70	9,20
Reverse (power of -1)	0,21	0,15	0,11
Urutan maju	0,11	0,15	0,21

$$\begin{aligned} \tilde{w}_i &= \tilde{r}_i \times (\tilde{r}_1 \oplus \tilde{r}_2 \oplus \dots \oplus \tilde{r}_n)^{-1} \\ &= (lw_i, mw_i, uw_i) \end{aligned} \tag{2}$$

Dari perhitungan dengan (2) didapatkan bobot relatif fuzzy untuk tiap kriteria pada level 1 yang ditunjukkan pada Tabel 5. Perhitungan bobot kriteria pada level 2 (subkriteria) dapat dilakukan seperti langkah pengerjaan dalam perhitungan bobot kriteria pada level 1. Bobot akhir relatif fuzzy untuk tiap kriteria dan subkriteria ditunjukkan pada Tabel 6.

E. Perhitungan Rating Menggunakan FMCDM TOPSIS

Fuzzy Multi Criteria Decision Making TOPSIS (FMCDM TOPSIS) digunakan untuk menghitung *rating* dari masing masing *case* yang ada pada *event logs*. Tiap *case* ini kemudian diberikan label Anomali, Semi Anomali, atau Tidak Anomali.

Data masukan untuk perhitungan ini diubah terlebih dahulu ke dalam bentuk fuzzy trapezoidal sehingga akan terdapat empat nilai yaitu “lower”, “middle 1”, “middle 2”, dan

Tabel 5.
Bobot relatif fuzzy untuk tiap kriteria

Kriteria	\tilde{w}_i		
	Lower	Middle	Upper
<i>Skipped Activity</i>	0,137	0,271	0,490
<i>Wrong Throughput Time</i>	0,090	0,168	0,308
<i>Wrong Resource</i>	0,050	0,099	0,229
<i>Wrong Duty</i>	0,072	0,139	0,275
<i>Wrong Pattern</i>	0,038	0,070	0,158
<i>Wrong Decision</i>	0,122	0,254	0,508

Tabel 6.
Bobot akhir relatif fuzzy untuk tiap kriteria dan subkriteria

Kriteria	\tilde{w}_i		
	Lower	Middle	Upper
<i>Skip Sequence</i>	0,0288	0,0904	0,311
<i>Skip Decision</i>	0,0501	0,1809	0,5386
<i>Throughput Time Min</i>	0,0452	0,0838	0,1542
<i>Throughput Time Max</i>	0,0452	0,0838	0,1542
<i>Wrong Resource</i>	0,050	0,099	0,229
<i>Wrong Duty Sequence</i>	0,0069	0,0227	0,0914
<i>Wrong Duty Decision</i>	0,02	0,0752	0,2636
<i>Wrong Duty Combine</i>	0,0109	0,0414	0,1661
<i>Wrong Pattern</i>	0,038	0,070	0,158
<i>Wrong Decision</i>	0,122	0,254	0,508

Tabel 7.
Derajat pelanggaran

Range Value(x)	Atribut	Lower	Middle1	Middle2	Upper
0	NORMAL (N)	0,0	0,0	0,0	0,0
0<x≤0,1	VERY WEAK (VW)	0,0	0,0	0,1	0,2
0,1<x≤0,2	BETWEEN VERY WEAK AND WEAK (BVW&W)	0,0	0,1	0,2	0,3
0,2<x≤0,3	WEAK (W)	0,1	0,2	0,3	0,4
0,3<x≤0,4	BETWEEN WEAK AND FAIR (BW&F)	0,2	0,3	0,4	0,5
0,4<x≤0,5	FAIR (F)	0,3	0,4	0,5	0,6
0,5<x≤0,6	BETWEEN FAIR AND STRONG (BF&S)	0,4	0,5	0,6	0,7
0,6<x≤0,7	STRONG (S)	0,5	0,6	0,7	0,8
0,7<x≤0,8	BETWEEN STRONG AND VERY STRONG (BS&VS)	0,7	0,8	0,9	1,0
0,8<x≤1	VERY STRONG (VS)	0,8	0,9	1,0	1,0

“upper”. Untuk itu data masukan *value* diklasifikasikan ke dalam sembilan derajat pelanggaran seperti yang ditunjukkan pada Tabel 7. Data masukan ini didefinisikan dalam bentuk matriks D berukuran $i \times j$ seperti (3). Dengan i adalah jumlah kriteria anomali dan j adalah jumlah *case*.

$$D = \begin{matrix} & X_1 & X_2 & \dots & X_j \\ A_1 & \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1j} \\ r_{22} & r_{22} & \dots & r_{2j} \\ \vdots & \vdots & \dots & \vdots \\ r_{i1} & r_{i2} & \dots & r_{ij} \end{bmatrix} \\ A_2 & & & & \\ \vdots & & & & \\ A_i & & & & \end{matrix} \tag{3}$$

Selanjutnya data yang telah difuzzifikasi dikalikan dengan bobot kepentingan kriteria w_j^* menggunakan (4)[11]. Dengan, a merupakan “lower”, b merupakan “middle 1”, c merupakan “middle 2”, dan d merupakan “upper” dari $value_{ke-ij}$. Kemudian pada w_j^* , α merupakan “lower”, β dan γ merupakan “middle”, serta δ merupakan “upper” untuk masing-masing kriteria ke- j sesuai pada Tabel 6.

$$v_{ij} = r_{ij}(\cdot)w_j^* = \left(\frac{a_{ij}}{d_j^*} \alpha_j, \frac{b_{ij}}{c_j^*} \beta_j, \frac{c_{ij}}{b_j^*} \gamma_j, \frac{d_{ij}}{a_j^*} \delta_j \right) \quad (4)$$

Langkah selanjutnya yaitu menentukan Solusi Ideal Positif (A^*) serta Solusi Ideal Negatif (A^-). Solusi ideal didapatkan dengan cara menghitung rerata umum $M(v_{ij})$ dari Chen dan Hwang [14] pada (5) untuk masing-masing v_{ij} . Dengan, a merupakan “lower”, b merupakan “middle 1”, c merupakan “middle 2”, dan d merupakan “upper” dari $value_{ke-ij}$.

$$M(v_{ij}) = \frac{-a_{ij}^2 - b_{ij}^2 + c_{ij}^2 + d_{ij}^2 - a_{ij}b_{ij} + c_{ij}d_{ij}}{[3(-a_{ij} - b_{ij} + c_{ij} + d_{ij})]} \quad (5)$$

Solusi Ideal Positif (A^*) untuk tiap kriteria ke- j adalah yang memiliki rerata umum terbesar di kriteria ke- j . Sedangkan, Solusi Ideal Negatif (A^-) untuk tiap kriteria ke- j adalah yang memiliki rerata umum terkecil di kriteria ke- j . Tahap selanjutnya yaitu mengukur *separation measure* dari Solusi Ideal Positif (S_{i^+}) serta *separation measure* dari Solusi Ideal Negatif (S_{i^-}) dengan menggunakan (6) dan (7) [11].

$$S_{i^+} = \sum_{j=1}^n D_{ij}^* \quad , \quad i = 1, \dots, m \quad (6)$$

$$S_{i^-} = \sum_{j=1}^n D_{ij}^- \quad , \quad i = 1, \dots, m \quad (7)$$

Untuk data *crisp*, selisih dari D_{ij}^* dan D_{ij}^- diukur dengan menggunakan (8) dan (9). Dengan v_{ij} adalah rerata umum untuk tiap v_{ij} , v_j^* adalah A^* untuk tiap kriteria ke- j , dan v_j^- adalah A^- untuk tiap kriteria ke- j . Tahapan terakhir yaitu menghitung kedekatan relatif (c_{i^+}) dengan solusi ideal menggunakan (10) [11].

$$D_{ij}^* = |v_{ij} - v_j^*| \quad (8)$$

$$D_{ij}^- = |v_{ij} - v_j^-| \quad (9)$$

$$c_{i^+} = \frac{S_{i^-}}{(S_{i^+} + S_{i^-})} \quad (10)$$

Selanjutnya *label* didapatkan dengan mengklasifikasikan *rating* ke dalam 3 kelompok, yaitu Anomali, Semi Anomali, dan Tidak Anomali. Klasifikasi tersebut berdasarkan syarat SOP dan aturan verifikasi, yaitu:

1. Kasus yang terdapat *Skipped Activity* dan *Wrong Decision* pasti mengandung anomali.

Tabel 8.
Tipe penyimpangan dan nilainya

Tipe penyimpangan	Rangenilai rating	Nilai penyimpangan
Tidak Anomali	$0 \leq rating \leq 0,1$	1
	$0,1 < rating < 0,2$	$(0,2 - rating) / 0,1$
Semi Anomali	$0,1 < rating < 0,2$	$(rating - 0,1) / 0,1$
	$0,2 \leq rating \leq 0,4$	1
	$0,4 < rating < 0,5$	$(0,5 - rating) / 0,1$
Anomali	$0,4 < rating < 0,5$	$(rating - 0,4) / 0,1$
	$rating \geq 0,5$	1

2. Kasus yang terdapat *Wrong Resource* dan *Wrong Duty* secara bersamaan pasti mengandung anomali.

Klasifikasi data berdasarkan nilai *rating* yang dihasilkan dari perhitungan dapat dilihat pada Tabel 8.

F. Training dan Testing Data dengan Decision Tree

Decision Tree merupakan salah satu metode *data mining* yang digunakan untuk klasifikasi. Konsep dari *Decision Tree* yaitu mengubah data menjadi pohon keputusan dan aturan-aturan keputusan [15]. Gambaran pemakaiannya yaitu untuk membuat aturan (*rule*) yang dapat digunakan untuk menentukan apakah *case* tersebut mempunyai potensi terdapat anomali berdasarkan kriteria-kriteria yang sudah ditentukan sebelumnya. Karena data masukan berupa bilangan kontinu, data terlebih dahulu diklasifikasikan ke dalam tiga kelas diskrit seperti pada Tabel 9.

Tabel 9.
Klasifikasi data kontinu

Range nilai atribut (x)	Kelas
0	0
$0 < x \leq 0,33$	Low
$0,33 < x \leq 0,67$	Medium
$0,67 < x \leq 1$	High

Langkah selanjutnya yaitu pembuatan pohon keputusan. *Root* pertama yang dipilih adalah atribut yang memiliki *Information Gain* paling besar. Untuk itu perhitungan dimulai dengan menghitung entropi dengan menggunakan (11) dari tiap atribut. Dengan S merupakan atribut, n adalah jumlah data, i jumlah atribut, c kelas pelabelan, dan p_i adalah probabilitas untuk tiap kelas c terhadap jumlah n data.

$$Entropi(S) = - \sum_{i=1}^n p_i(c) \times \log_2 p_i(c) \quad (11)$$

Setelah itu dilakukan perhitungan *Information Gain*-nya dengan menggunakan (12). Dengan *Entropi (S)* adalah entropi atribut, *Entropi (S_i)* adalah entropi suatu kelas ke- i , dan n adalah jumlah kelas.

$$Gain(S, A) = Entropi(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropi(S_i) \quad (12)$$

Jumlah data yang digunakan untuk *training* yaitu sebanyak 1000 data. Sedangkan jumlah data yang digunakan untuk *testing* yaitu sebanyak 200 data. *Testing* dilakukan untuk memprediksi kelas pada setiap *case* berdasarkan pola *Decision Tree* yang telah terbentuk.

Rincian terhadap 1000 data untuk kelas diketahui, yaitu Anomali sebanyak 21, Semi Anomali sebanyak 18, dan Tidak Anomali sebanyak 961. Kemudian rincian 200 data kelas diketahui untuk Anomali sebanyak 5, Semi Anomali sebanyak 3, dan Tidak Anomali sebanyak 192.

Berdasarkan prediksi kelas yang dihasilkan dari *testing* untuk 200 data tersebut, dilakukan perhitungan akurasi menggunakan *confusion matrix* dengan (13). Dengan, TP yaitu *True Positive*, TN yaitu *True Negative*, TSP yaitu *True Semi Positive*, FN yaitu *False Negative*, FP yaitu *False Positive*, dan FSP yaitu *False Semi Positive*.

Selain dilakukan perhitungan akurasi, juga dilakukan perhitungan terhadap *sensitivity* dan *specificity*. *Sensitivity* merupakan pengukuran proporsi data positif (*P*), yaitu data berlabel Anomali, yang dapat diprediksi dengan benar (*TP*) seperti pada (14). Sedangkan *specificity* adalah pengukuran proporsi data negatif (*N*), yaitu data berlabel Tidak Anomali, yang dapat diprediksi dengan benar (*TN*) seperti pada (15).

$$Akurasi = \frac{TP + TSP + TN}{TP + TSP + TN + FP + FSP + FN} \quad (13)$$

$$Sensitivity = \frac{TP}{P} \quad (14)$$

$$Specificity = \frac{TN}{N} \quad (15)$$

IV. HASIL PENGUJIAN

Pengujian yang dilakukan yaitu pengujian metode untuk melakukan evaluasi terhadap metode *Fuzzy Multi Criteria Decision Making* TOPSIS dan *Decision Tree*. Dari hasil pengujian terhadap 200 data didapatkan hasil seperti yang ditunjukkan pada Tabel 10. Hal ini menunjukkan tingkat akurasi yang dihasilkan yaitu sebesar 98%. *Sensitivity* sebesar 0,71 dan *specificity* yang didapat yaitu sebesar 1.

Metode *Decision Tree* dapat mendeteksi seluruh *case* yang diprediksi Tidak Anomali, Semi Anomali dan Anomali.

V. KESIMPULAN

Pemodelan ontologi terhadap *standard operating procedure* (SOP) dan *event logs* serta penggunaan SWRL *rules* untuk penarikan kesimpulan terbukti mampu untuk mendapatkan nilai anomali yang berada di tengah yaitu Semi Anomali. Dengan hasil akurasi sebesar 98% *Fuzzy Multi Criteria Decision Making* TOPSIS dapat menghasilkan label yang baik. *Decision Tree* mampu untuk mendapatkan kasus yang terindikasi Semi Anomali.

DAFTAR PUSTAKA

[1] Association of Certified Fraud Examiners (ACFE), "Report to the Nations on Occupational Fraud and Abuse," 2014.
 [2] P. D. Goldmann, *Anti-Fraud Risk and Control Workbook*. New York: Wiley, 2009.
 [3] E. Lundin, H. Kvarnstrom, and E. Jonsson, "A synthetic fraud data generation methodology," *4th International Conference of Information and Communications Security*, vol. 2513, Singapore, 2002, pp. 265-277.

Tabel10.
Confusion matrix

		Kelas Prediksi		
		Anomali	Semi Anomali	Tidak Anomali
Kelas Diketahui	Anomali	5 (TP)	0 (FSP)	0 (FN)
	Semi Anomali	0 (FP)	3 (TSP)	2 (FN)
	Tidak Anomali	0 (FP)	2 (FSP)	188 (TN)

Keterangan:

- TP = *True Positive*
- FP = *False Positive*
- TN = *True Negative*
- FN = *False Negative*
- TSP = *True Semi Positive*
- FSP = *False Semi Positive*

[4] E. L. Barse, H. Kvarnström, and E. Jonsson, "Synthesizing test data for fraud detection systems," *Proceedings of the 19th Annual Computer Security Applications Conference*, Las Vegas, 2003, pp. 384-394.
 [5] Mieke Jans, Jan Martijn van der Werf, Nadine Lybaert, and Koen Vanhoof, "A business process mining application for internal transaction fraud mitigation," *Expert Systems with Applications* 38, pp. 13351-13359, 2011.
 [6] S. Huda, R. Sarno, T. Ahmad, H. A. Santoso, "Identification of process-based fraud patterns in credit application," *2nd International Conference on Information and Communication Technology (ICoICT)*, Bandung, 2014, pp. 84-89.
 [7] Wil M.P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer-Verlag, 2011.
 [8] R. D. Dewandono, R. Sarno, *Process Sequence Mining For Fraud Detection Using Complex Event Processing*. Surabaya: Institut Teknologi Sepuluh Nopember, 2013.
 [9] Ian Horrocks, Peter F. Patel-Scheider, Harold Boley, Said Tabet, Benjamin Grosf, and Mike Dean. (2004, May) SWRL: A Semantic Web Rule Language Combining OWL and RuleML. [Online]. <http://www.w3.org/Submission/SWRL/>
 [10] Fernandes P. Sinaga, Riyanarto Sarno, Abdul Munif, *Rancang Bangun Fuzzy Association Rule Miner untuk Mendeteksi Fraud pada Proses Bisnis Enterprise Resource Planning (ERP)*. Surabaya: Institut Teknologi Sepuluh Nopember, 2014.
 [11] Cengiz Kahraman, Ihsan Kaya, Sezi Çevik, Nüfer Yasin Ates, and Murat Gülbay, "Fuzzy Multi-Criteria Evaluation of Industrial Robotic Systems Using TOPSIS," *Fuzzy Multi-Criteria Decision Making Theory and Applications with Recent Developments*. New York: Springer Science+Business Media, LLC, 2008, pp. 159-186.
 [12] P.J.M van Laarhoven and W. Pedrycz, "A fuzzy extension of Saaty's priority theory," *Fuzzy Sets and Systems*, vol. 11, no. 1-3, pp. 199-227, Jan. 1983.
 [13] Mustafa Batuhan Ayhan, "A Fuzzy AHP Approach for Supplier Selection Problem: A Case Study in A Gearmotor Company," *International Journal of Managing Value and Supply Chains (IJMVSC)*, vol. 4, no. 3, pp. 11-23, September 2013.
 [14] Shu-Jen Chen and Chin-Lai Hwang, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*. Berlin: Springer-Verlag, 1992.
 [15] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, *Introduction to Data Mining*. New York: Pearson, 2005.