

Peningkatan Akurasi Pembobotan Attribute Importance Weights pada Deteksi Fraud

Weighting Accuracy Improvement of Attribute Importance Weights in Fraud Detection

Erba Lutfina¹, Solichul Huda²

¹Program Studi Sistem Informasi, Universitas Nasional Karangturi, Semarang

²Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang

E-mail: ¹erba.lutfina@unkartur.ac.id, ²solichul.huda@dinus.ac.id

Abstrak

Kerugian miliaran dollar setiap tahunnya dialami oleh bank yang disebabkan oleh Fraud. Salah satu solusi untuk mengatasi kasus fraud yang dialami dunia perbankan dapat dilakukan dengan proses deteksi fraud. Pada proses deteksi Fraud, terdapat berbagai atribut PBF (Process Based Fraud) yang setiap atributnya memiliki dampak yang berbeda dalam mendeteksi fraud. Untuk menentukan bobot setiap atribut PBF digunakan metode MDL (Modified Digital Logic). Metode MDL menghasilkan attribute importance weights yang sesuai dengan dampak atribut PBF. Namun peran pakar masih sangat signifikan dalam menilai setiap attribute importance weights. Penelitian ini bertujuan untuk mengubah prosedur penentuan bobot attribute importance weights dalam metode MDL dengan menambahkan metode Multiple Linear Regression (MLR). Dengan mengganti inputan yang sebelumnya diberikan oleh pakar menjadi perbandingan bobot atribut secara otomatis. Kemudian hasil dari kedua metode dievaluasi menggunakan confusion matrix. Berdasarkan hasil eksperimen, metode MLR menunjukkan persentase klasifikasi menggunakan semua attribute importance weights menunjukkan hasil yang lebih baik dengan akurasi sebesar 99,5%.

Kata kunci: Fraud Detection, Process Mining, Process Based Fraud, Modified Digital Logic, Multiple Linear Regression

Abstract

Bank loses billions of dollars every year due to Fraud. One solution to overcome fraud cases experienced by the banking world can be done with the fraud detection process. In the Fraud detection process, there are various PBF (Process-Based Fraud) attributes, each of which has a different impact in detecting fraud. To determine the weight of each PBF attribute the MDL (Modified Digital Logic) method is used. The MDL method generates attribute importance weights that correspond to the impact of PBF attributes. But the role of experts is still very significant in assessing each attribute's importance weights. This study aims to change the procedure of determining the attribute importance weights in the MDL method by adding the Multiple Linear Regression (MLR) method. By replacing the input previously provided by the expert, it automatically compares the attribute weight. Then the results of both methods are evaluated using a confusion matrix. Based on experimental results, the RLB method shows the percentage of classifications using all the attributes of importance weights have better results with an accuracy of 99.5%.

Keywords: Fraud Detection, Process Mining, Process Based Fraud, Modified Digital Logic, Multiple Linear Regression

1. PENDAHULUAN

Fraud menjadi fokus penelitian pada beberapa tahun terakhir ini. *Fraud* juga menyebabkan bank mengalami kerugian miliaran dollar setiap tahunnya [1]. Metode deteksi *fraud* menjadi salah satu solusi untuk mengatasi kasus fraud yang terjadi di dunia perbankan.

Fraud bisa terjadi dalam data transaksi, namun bisa juga terjadi dalam proses bisnis yang sedang berjalan. Untuk menganalisis data transaksi digunakan pendekatan *data mining*,

sedangkan analisis pada proses bisnis menggunakan pendekatan process mining. Karena metode yang diusulkan dalam penelitian ini menganalisis proses bisnis, maka penelitian ini menggunakan pendekatan *process mining*.

Deteksi *fraud* dengan *data mining* telah banyak dilakukan pada berbagai penelitian. Contohnya dengan menggunakan metode *deep learning* [2] dalam penelitian tersebut menunjukkan bahwa *deep learning* adalah model yang menjanjikan untuk mendeteksi penipuan kartu kredit dibandingkan dengan model pembelajaran mesin tradisional seperti ensemble. Namun, metode yang diusulkan dalam penelitian tersebut kurang sesuai untuk *Fraud Detection System* di dunia nyata yang beroperasi dengan beberapa kendala. Kemudian contoh lainya dengan menggunakan algoritma *Neural network* [3] yang dibandingkan dengan metode deteksi penipuan kartu kredit lainnya. Algoritma *Self-Organizing Maps* (SOM) [4] digunakan untuk mencegah transaksi penipuan online, dalam penelitian masih ditemukan kendala pada kompleksitas komputasi SOM. Sedangkan pada penelitian [5] analisis empiris digunakan untuk mendeteksi kecurangan transaksi e-commerce masih memerlukan validitas eksternal untuk mengevaluasi efektivitas framework yang diusulkan. Pada penelitian deteksi *fraud* kartu kredit menggunakan *Hidden Markov Model* [6] dapat digunakan untuk membantu mendapatkan cakupan penipuan yang tinggi dikombinasikan dengan *false alarm rate* yang rendah.

Selain itu, pada penelitian [7] deteksi *fraud* dilakukan dengan menggunakan metode *hybrid* algoritma ARL dan *process mining* untuk mendapatkan jumlah variabel *fraud* dalam menghasilkan *association rule*. Pada penelitian tersebut *process mining* digunakan untuk mencocokkan *rule* yang ada di SOP, akan tetapi dalam metode *hybrid* algoritma ARL peran pakar masih sangat signifikan untuk mengekstrak item set dari *association rule*. Selanjutnya pada penelitian [8] memperkenalkan pendekatan baru berbasis process mining untuk mengatasi masalah pembuatan predictive *Complex Event Processing* (CEP) rules secara otomatis pada proses deteksi *fraud*, dalam hasil penelitian masih diperlukan pengoptimalan prosedur ketika terdapat aliran data yang besar.

Pada penelitian [9] deteksi *fraud* dilakukan dengan menggunakan metode *process mining*. Penelitian mencocokkan *business rules* dengan SOP untuk mendapatkan atribut PBF (*Process Based Fraud*). Pada proses deteksi *Fraud*, setiap atribut PBF memberikan dampak yang berbeda untuk mengindikasi *fraud*. Untuk melakukan pembobotan setiap atribut PBF digunakan metode MDL (*Modified Digital Logic*). Metode MDL menghasilkan *attribute importance weights* yang sesuai dengan dampak atribut PBF. Akan tetapi, dalam metode MDL peran pakar masih sangat signifikan untuk menilai setiap *attribute importance weights*. Hal ini akan bermasalah jika terjadi perubahan objek penelitian atau perubahan pakar. Untuk menangani masalah tersebut dibutuhkan suatu metode untuk menentukan bobot pentingnya atribut dengan mengurangi peran pakar. Metode atau skema yang dipilih yaitu dengan memodifikasi penilaian *attribute importance weights* pada metode MDL secara otomatis.

Penelitian mengenai pemberian *attribute importance weights* menggunakan metode MDL telah dilakukan pada beberapa penelitian. Pada penelitian [10] metode *Modified Digital Logic* (MDL) digunakan untuk menghitung bobot semua parameter *renewable-power plant*. Penilaian prioritas setiap parameter dalam metode MDL masih ditentukan oleh pakar secara subjektif. Pada penelitian [9] proses pembobotan atribut juga dilakukan dengan menggunakan metode MDL (*Modified Digital Logic*). Proses pembobotan menggunakan MDL dilakukan secara subjektif oleh pakar dengan membandingkan setiap atribut dengan atribut lainnya.

Berdasarkan review dari jurnal terkait mengenai metode MDL, didapatkan bahwa proses pembobotan atribut menggunakan metode MDL masih dilakukan secara subjektif dan peran pakar masih sangat signifikan. Untuk mengatasi masalah pembobotan *attribute importance weights* dalam metode MDL, akan diusulkan metode *Multiple Linear Regression* (MLR).

Metode MLR dapat digunakan untuk memprediksi suatu nilai dan mencari koefisien bobot setiap atribut [11]. Prediksi didapatkan dari perkalian data latih dengan koefisien regresi. Koefisien regresi menunjukkan nilai bobot setiap atribut pada proses prediksi. Koefisien regresi setiap variabel inilah yang dapat digunakan sebagai inputan pada metode MDL untuk mengatasi masalah *attribute importance weights* secara otomatis.

Pendekatan yang mungkin dilakukan berdasarkan analisis masalah diatas dapat diselesaikan dengan menggunakan koefisien regresi. Koefisien regresi merupakan hasil dari proses MLR untuk menghitung *attribute importance weights* pada metode MDL. Usulan tersebut diharapkan dapat digunakan untuk menghitung *attribute importance weights* PBF dalam menentukan nilai deteksi *fraud*.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data *fraud* dari sistem deteksi *Process-Based Fraud*. Data tersebut dikelompokkan dalam data testing sebanyak 806 *case*. Data yang digunakan merupakan data *case* yang menyimpang dari SOP. Penyimpangan teridentifikasi berdasarkan kriteria atribut PBF.

Dataset yang digunakan tidak mengandung nilai kosong, sehingga dapat digunakan secara langsung untuk proses prediksi. Contoh dari dataset ditunjukkan pada tabel dibawah ini.

Tabel 1 Dataset PBF

| Case | Skip sequence | Skip decision | Throughput time min | Throughput time max | Wrong resource | Wrong duty sequence | Wrong duty decision | Wrong duty combine | Wrong pattern | Wrong decision | Event parallel | Fraud |
|------|---------------|---------------|---------------------|---------------------|----------------|---------------------|---------------------|--------------------|---------------|----------------|----------------|-------|
| 102 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 200 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,878 |
| 201 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,878 |
| 203 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 204 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,878 |
| 205 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,228 |
| 211 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,228 |
| 214 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,228 |
| 580 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,878 |
| 583 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,878 |

Dari tabel diatas terdapat 11 variabel yang mengindikasikan atribut PBF mulai dari atribut *Skip sequence* hingga *Event parallel*. Pelanggaran yang terjadi dalam setiap *case* ditunjukkan dengan frekuensi munculnya nilai tiap attribute PBF.

Contohnya pada case 102, pelanggaran muncul satu kali pada atribut skip sequence dan satu kali pada atribut wrong pattern. *Case* 102 memiliki nilai deteksi *fraud* sebesar 1 yang mengindikasikan bahwa *case* 102 terindikasi fraud. Jika dibandingkan dengan *case* 200 yang muncul tiga kali pada atribut *Throughput time min*, nilai deteksi *fraud* hanya bernilai 0,878. Dari contoh diatas menunjukkan bahwa setiap atribut PBF memiliki bobot yang berbeda-beda untuk menghasilkan nilai deteksi *fraud*.

2.2 Atribut PBF

Pada proses deteksi *fraud* dilakukan identifikasi SOP dengan menganalisis proses bisnis menggunakan metode dalam *Process Mining* [12]. Metode tersebut adalah sebagai berikut:

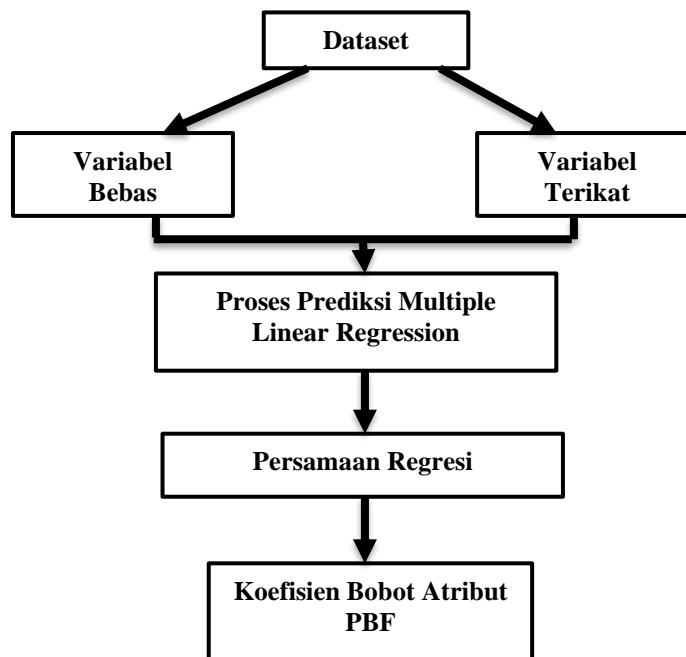
Tabel 2 Atribut PBF

| Atribut | Deskripsi |
|----------------------------|--|
| <i>Throughput time min</i> | Waktu eksekusi <i>event</i> yang lebih cepat dari waktu <i>standard event</i> |
| <i>Throughput time max</i> | Waktu eksekusi <i>event</i> yang lebih lambat dari waktu <i>standard event</i> |
| <i>Skip sequence</i> | Sebuah <i>event</i> dilewati satu atau lebih pada urutan proses |
| <i>Skip decision</i> | Sebuah <i>event</i> dilewati satu atau lebih pada urutan proses yang memiliki percabangan |
| <i>Wrong duty decision</i> | <i>Originator</i> mengeksekusi lebih dari satu <i>event</i> pada <i>event decision</i> |
| <i>Wrong resource</i> | <i>Event</i> dieksekusi oleh <i>originator</i> yang tidak memiliki wewenang |
| <i>Wrong pattern</i> | <i>event</i> yang dieksekusi tidak sesuai dengan pola urutan SOP |
| <i>Wrong duty sequence</i> | <i>Originator</i> mengeksekusi lebih dari satu <i>event</i> pada <i>event sequence</i> |
| <i>Wrong duty combine</i> | <i>Originator</i> mengeksekusi lebih dari satu <i>event</i> pada <i>event sequence</i> dan <i>decision</i> |
| <i>Wrong decision</i> | Event yang memiliki keputusan dieksekusi tidak sesuai dengan SOP |
| <i>Event paralel</i> | Lebih dari satu <i>event</i> dieksekusi secara bersamaan |

2.3 Multiple Linear Regression

Analisis regresi memiliki tujuan utama untuk memprediksi atau memperkirakan suatu nilai dari hubungan tiap-tiap variable [13]. Multiple Linear Regression (MLR) merupakan analisis regresi yang variabel terikatnya (Y) memiliki hubungan dengan lebih dari satu variabel bebas (x) namun diagramnya masih menunjukkan hubungan yang linier [14].

Tahap awal penelitian yaitu dengan membentuk Persamaan Regresi dengan melakukan prediksi menggunakan data training sebanyak 806 data. Persamaan Regresi akan menghasilkan bobot koefisien atribut PBF yang dapat digunakan sebagai inputan metode MDL. Tahapan membentuk Persamaan regresi digambarkan dengan alur sebagai berikut:



Gambar 1 Alur implementasi MLR

Tahap awal dilakukan dengan input dataset yang terdiri dari variabel bebas (atribut PBF) dan variabel terikat (nilai *fraud*). Langkah selanjutnya melakukan proses prediksi MLR dengan inputan dataset yang telah dijelaskan sebelumnya. Hasil dari proses MLR didapatkan persamaan regresi dari perkalian koefisien regresi dengan setiap variabel bebas. Persamaan matematis MLR

dinyatakan sebagai berikut :

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 \dots \beta_nx_n + \varepsilon \quad (1)$$

Dimana nilai β dan ε dapat didapatkan dengan rumus:

$$\beta = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \quad (2)$$

$$\varepsilon = \frac{\sum y - b \sum x}{n} \quad (3)$$

Keterangan :

Y : Variabel terikat (nilai prediksi)

$\beta_0, \beta_1, \beta_2, \beta_3 \dots \beta_n$: koefisien regresi

$x_1, x_2, x_3 \dots x_n$: variabel bebas

ε : deviasi kesalahan

Misalkan pada proses MLR dihasilkan persamaan regresi [13]:

$$y = \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \varepsilon \quad (4)$$

Dengan nilai $\varepsilon = -0,05$; $\beta_1 = 0,299$; $\beta_2 = 0,427$; $\beta_3 = 0,322$; $\beta_4 = 0,273$; $\beta_5 = 0,405$

Variabel bebas masing-masing memiliki nilai $x_1 = 2$; $x_2 = 1$, $x_3 = 1$, $x_4 = 3$, $x_5 = 1$. Dengan persamaan MLR tersebut, nilai y dapat diprediksi sebagai berikut:

$$\begin{aligned} y &= 0,299 (2) + 0,427 (1) + 0,322 (1) + 0,273 (3) + 0,405 (1) -0,05 \\ &= 2,521 \end{aligned}$$

Dari proses MLR tersebut didapatkan koefisien regresi masing-masing atribut PBF sebesar 0,299; 0,427; 0,322; 0,273; 0,405 yang selanjutnya dapat digunakan sebagai inputan di metode MDL.

2.3 Modified Digital Logic

Atribut PBF memiliki dampak yang berbeda untuk mendeteksi fraud. Oleh karena itu setiap atribut diberikan bobot yang berbeda-beda sesuai dampak yang dihasilkan. *Attribute importance weights* digunakan untuk memberikan bobot sesuai dengan dampak masing-masing atribut untuk menghasilkan nilai deteksi fraud. *Attribute importance weights* dapat dihitung dengan menggunakan metode Modified Digital Logic.

Modified Digital Logic (MDL) merupakan salah satu metode yang sering digunakan untuk memperkirakan bobot parameter. Metode ini memasukan pendapat pakar untuk menetapkan prioritas masing-masing parameter [15]. Pakar akan menentukan prioritas dengan memberikan nilai 1 untuk parameter yang kurang penting, nilai 2 untuk prioritas yang seimbang dan nilai 3 untuk parameter yang lebih penting. Berdasarkan pendapat pakar, *decision table* dibentuk untuk menunjukkan perbandingan setiap parameter. Setelah itu memperkirakan jumlah keputusan positif yang mungkin dengan $N = n(n-1) / n$. Dimana n adalah jumlah atribut / parameter. P adalah jumlah dari semua *positive decision* untuk parameter tertentu dan W_j adalah bobot akhir dengan persamaan [16]:

$$W_j = \frac{P_j}{\sum_{j=1}^n P_j} \quad (5)$$

Contoh penerapan metode MDL adalah sebagai berikut:

Misalkan diketahui koefisien bobot atribut A1 sampai A5 adalah masing-masing berikut 0,299; 0,427; 0,322; 0,273; 0,405. Koefisien atribut tersebut didapatkan dari proses prediksi pada metode MLR.

Perbandingan atribut dilakukan dengan memberikan nilai 1 untuk parameter yang kurang penting, nilai 2 untuk prioritas yang seimbang dan nilai 3 untuk parameter yang lebih penting.

Maka matriks perbandingan atributnya adalah sebagai berikut :

Tabel 3 Matriks perbandingan *attribute importance weights*

| | | | | | |
|----|----|----|----|----|----|
| | A1 | A2 | A3 | A4 | A5 |
| A1 | 2 | 1 | 1 | 3 | 1 |
| A2 | 3 | 2 | 3 | 3 | 3 |
| A3 | 3 | 1 | 2 | 3 | 1 |
| A4 | 1 | 1 | 1 | 2 | 1 |
| A5 | 3 | 1 | 3 | 3 | 2 |

Selanjutnya menghitung nilai *Positive Decision* yang didapatkan dari penjumlahan hasil perbandingan bobot dalam satu atribut.

Dari perhitungan nilai *Positive Decision*, selanjutnya menghitung *attribute importance weights* dengan cara membagi tiap atribut dengan jumlah *Positive Decision* seluruh atribut.

Tahapan selanjutnya adalah menentukan klasifikasi variabel untuk setiap nilai deteksi *fraud* dengan menggunakan metode klasifikasi k-nn. Hasil klasifikasi deteksi nilai *fraud* menggunakan *attribute importance weights* dibandingkan dengan hasil klasifikasi deteksi nilai *fraud* tanpa menggunakan perhitungan *attribute importance weights*. Perbandingan hasil klasifikasi dilakukan untuk membandingkan pengaruh *attribute importance weights*.

2.4 Penentuan kelas *Attribut Importance Weight*

Tahap penentuan kelas ditentukan dengan menggunakan sebuah rule. Jika nilai bobot atribut *importance weights* bernilai lebih atau sama dengan 0,2 maka bobot termasuk pada kelas *very importance*. Jika nilai bobot bernilai kurang dari 0,2 maka bobot termasuk pada kelas *importance*. Berikut merupakan tabel hasil penentuan kelas *attribute importance weights* berdasarkan rule :

Tabel 4 Hasil penentuan kelas *attribute importance weights*

| Data | Atribut 1 | Kelas |
|--------------|-----------|----------------|
| Weight Uji 1 | 0,24 | Very Important |
| Weight Uji 2 | 0,2 | Very Important |
| Weight Uji 3 | 0,12 | Important |
| Weight Uji 4 | 0,28 | Very Important |
| Weight Uji 5 | 0,16 | Important |

2.4 Evaluasi Klasifikasi

Evaluasi klasifikasi dilakukan dengan confusion matrix. Evaluasi menggunakan confusion matrix digunakan untuk menghitung akurasi, presisi dan recall. Evaluasi dilakukan dengan membandingkan klasifikasi nilai *fraud* tanpa menggunakan *attribute importance weights* yang termasuk dalam kelas “Important” dan klasifikasi yang menggunakan seluruh *attribute importance weights*.

Evaluasi hasil menggunakan confusion matrix, contoh perhitungan sebagai berikut:

Tabel 5 Confusion Matrix

| Kelas Klasifikasi | Prediksi Kelas | |
|-------------------|------------------|------------------|
| | Very Important | Important |
| Very Important | (True Positive) | (False Negative) |
| Important | (False Positive) | (True Negative) |

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

3. HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan dengan menggunakan 806 data case penyimpangan SOP, didapatkan hasil berupa *attribute importance weights*. *Attribute importance weights* dihitung dengan menggunakan metode MDL. Inputan pada metode MDL berupa koefisien bobot yang dihasilkan oleh metode Regresi Linier Berganda sebagai berikut:

Tabel 6 Hasil koefisien bobot

| Atribut | Koefisien bobot |
|---------------------|-----------------|
| Skip sequence | 1 |
| Skip decision | 0,9951 |
| Throughput time min | 0,2861 |
| Throughput time max | 0,286 |
| Wrong resource | 0,9756 |
| Wrong duty sequence | 0,9624 |
| Wrong duty decision | 0,884 |
| Wrong duty combine | 0,9187 |
| Wrong pattern | 0,467 |
| Wrong decision | 0,878 |
| Event parallel | 0,427 |

Dari tabel diatas, setiap atribut memiliki bobot yang selanjutnya menjadi inputan pada metode MDL untuk menghasilkan bobot penting setiap atribut. Seperti perhitungan pada sub-bab implementasi metode MDL dihasilkan matriks perbandingan bobot atribut sebagai berikut:

Tabel 7 Matriks perbandingan bobot

| | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 | A11 |
|-----|----|----|----|----|----|----|----|----|----|-----|-----|
| A1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| A2 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| A3 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| A4 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| A5 | 1 | 1 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 |
| A6 | 1 | 1 | 3 | 3 | 1 | 2 | 3 | 3 | 3 | 3 | 3 |
| A7 | 1 | 1 | 3 | 3 | 1 | 1 | 2 | 1 | 3 | 3 | 3 |
| A8 | 1 | 1 | 3 | 3 | 1 | 1 | 3 | 2 | 3 | 3 | 3 |
| A9 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 2 | 1 | 3 |
| A10 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 3 | 2 | 3 |
| A11 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |

Dari matriks perbandingan bobot atribut diatas, didapatkan *attribute importance weights* ditunjukkan pada tabel Weight. *Attribute importance weights* selanjutnya dikelompokkan dengan hasil sebagai berikut :

Tabel 8 Kelas *attribute importance weights*

| Atribut | Attribute importance weights | Kelas |
|---------------------|------------------------------|----------------|
| Skip sequence | 0,132 | Very Important |
| Skip decision | 0,124 | Very Important |
| Throughput time min | 0,058 | Important |
| Throughput time max | 0,050 | Important |
| Wrong resource | 0,116 | Very Important |
| Wrong duty sequence | 0,107 | Very Important |
| Wrong duty decision | 0,091 | Very Important |
| Wrong duty combine | 0,099 | Very Important |
| Wrong pattern | 0,074 | Important |
| Wrong decision | 0,083 | Very Important |
| Event parallel | 0,066 | Important |

Evaluasi hasil menggunakan confusion matrix adalah sebagai berikut:

Tabel 9 Hasil Confusion Matrix

| Kelas Klasifikasi | Prediksi Kelas | | | |
|----------------------|----------------------|-----------------|-------|-----------|
| | Very Confident Fraud | Confident Fraud | Fraud | Not Fraud |
| Very Confident Fraud | 37 | 0 | 0 | 0 |
| Confident Fraud | 0 | 0 | 0 | 0 |
| Fraud | 0 | 0 | 0 | 0 |
| Not Fraud | 1 | 0 | 0 | 204 |

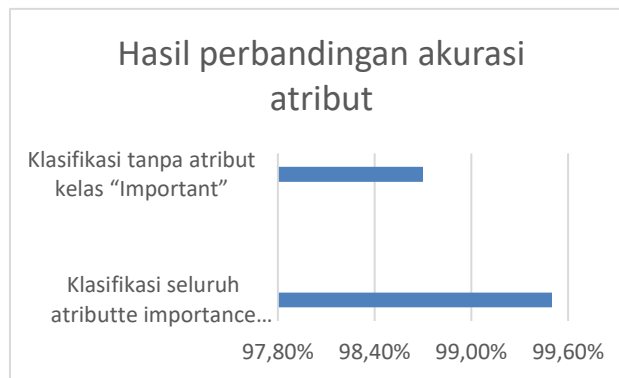
$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{241}{242} = 0,995 = 99,5 \%$$

Hasil perbandingan menunjukkan bahwa klasifikasi k-nn dengan nilai k = 3 tanpa menggunakan *attribute importance weights* yang termasuk dalam kelas “Important” memiliki hasil yang lebih rendah dengan nilai akurasi sebesar 98,7% dibandingkan dengan klasifikasi yang menggunakan seluruh *attribute importance weights* dengan nilai akurasi sebesar 99,5%.

Tabel 10 Perbandingan akurasi klasifikasi

| Klasifikasi | Akurasi |
|--|---------|
| Klasifikasi seluruh atributte importance weights | 99,5% |
| Klasifikasi tanpa atribut kelas "Important" | 98,7% |

Berikut merupakan gambar perbandingan hasil klasifikasi *attribute importance weights*:



Gambar 2 Hasil Perbandingan Akurasi Atribut

4. KESIMPULAN DAN SARAN

Berdasarkan hasil uji coba yang telah dilakukan menunjukkan bahwa pembobotan *attribute importance weights* dengan menggunakan metode MLR pada metode MDL terbukti dapat dilakukan secara otomatis. Hasil pembobotan memiliki hasil sebanyak tujuh variabel *attribute importance weights* termasuk dalam kelas "Very Important" dan empat variabel lainnya termasuk dalam kelas "Important".

Berdasarkan hasil uji coba yang telah dilakukan menunjukkan bahwa metode MLR cocok digunakan untuk menghasilkan bobot setiap atribut. Hasil evaluasi menunjukkan bahwa klasifikasi dengan menggunakan seluruh *attribute importance weights* memiliki hasil yang lebih baik dengan nilai akurasi sebesar 99,5%. Hal ini menunjukkan bahwa seluruh *attribute importance weights* mempengaruhi akurasi pada klasifikasi nilai fraud.

Untuk pengembangan penelitian ke depan, dapat dikembangkan dengan menggunakan variabel bebas lain atau menambah variabel bebas yang memiliki pengaruh yang signifikan terhadap perubahan nilai fraud.

DAFTAR PUSTAKA

- [1] N. Mahyavanshi, "A Novel Idea for Credit Card Fraud Detection using Decision Tree A Novel Idea for Credit Card Fraud Detection using Decision Tree," *Int. J. Comput. Appl.*, no. September, 2017.
- [2] E. Kim, J. Lee, H. Shin, H. Yang, S. Cho, S. Nam, Y. Song, J. Yoon, and J. Kim, "Champion-challenger analysis for credit card fraud detection : Hybrid ensemble and deep learning," *Expert Syst. Appl.*, vol. 128, pp. 214–224, 2019.
- [3] J. O. Awoyemi and S. A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques : A Comparative Analysis," *Int. Conf. Comput. Netw. Informatics*, 2017.
- [4] N. Balasupramanian, B. G. Ephrem, and I. S. Al-barwani, "User Pattern Based Online Fraud Detection and Prevention using Big Data Analytics and Self Organizing Maps," *2017 Int. Conf. Intell. Comput. Instrum. Control Technol.*, pp. 691–694, 2017.
- [5] J. Zhao, R. Y. K. Lau, W. Zhang, K. Zhang, X. Chen, and D. Tang, "Extracting and reasoning about implicit behavioral evidences for detecting fraudulent online transactions in e-Commerce," *Decis. Support Syst.*, 2016.
- [6] V. Bhusari, "Study of Hidden Markov Model in Credit Card Fraudulent Detection," *2016 World Conf. Futur. Trends Res. Innov. Soc. Welf. (Startup Conclave)*, vol. 20, no. 5, pp. 33–36, 2016.
- [7] R. Sarno, R. D. Dewandono, T. Ahmad, and M. F. Naufal, "Hybrid Association Rule Learning and Process Mining for Fraud Detection," no. April, 2015.
- [8] L. Patrima and K. Zeitouni, "Automatic Learning of Predictive CEP Rules : Bridging the Gap between Data Mining and Complex Event Processing," *Proc. 11th ACM Int. Conf. Distrib. Event-based Syst.*, 2017.
- [9] S. Huda, R. Sarno, and T. Ahmad, "Fuzzy MADM Approach for Rating of Process-Based Fraud," vol. 9, no. 2, pp. 111–128, 2015.
- [10] F. S. Mohammad Ali Beheshtinia, "Prioritize the renewable-power plants using a combination of Modified Digital Logic and fuzzy VIKOR methods," *Energy Eng. Manag.*, 2017.
- [11] M. Protić, F. Fathurrahman, and M. Raos, "Modelling Energy Consumption of the Republic of Serbia using Linear Regression and Artificial Neural Network Technique," vol. 3651, pp. 135–141, 2019.
- [12] S. Huda, T. Ahmad, R. Sarno, and H. A. Santoso, "Identification of process-based fraud patterns in credit application," *2014 2nd Int. Conf. Inf. Commun. Technol. ICoICT 2014*, pp. 84–89, 2014.
- [13] M. and G. G. Clausel, "Practical Session: Multiple Linear Regression," vol. 66, no. 2014, pp. 73–75, 2015.
- [14] J. Wang, F. Wang, Y. Liu, J. Xu, H. Lin, B. Jia, W. Zuo, Y. Jiang, L. Hu, and F. Lin, "Multiple Linear Regression and Artificial Neural Network to Predict Blood Glucose in Overweight Patients," *Exp. Clin. Endocrinol. Diabetes*, pp. 34–38, 2016.
- [15] S. Vats, G. Vats, R. Vaish, and V. Kumar, "Selection of optimal electronic toll collection system for India : A subjective-fuzzy decision making approach," vol. 21, pp. 444–452, 2014.
- [16] N. Ogunbiyi and A. Basukoski, "A Context-Aware Process Mining Predictive Model," pp. 1–9.