

ABSTRAK

Saat ini pelaksanaan Good Corporate Governance (GCG) dalam komitmen organisasi adalah kunci keberhasilan dalam mencapai kinerja bisnis yang efektif, efisien dan berkelanjutan. Salah satu implementasi dari GCG adalah manajemen penipuan melalui deteksi penipuan.

Dalam penelitian sebelumnya, deteksi penipuan dilakukan pada layanan suara dengan sumber data Call Detail Record. Pembahasan penelitian ini berfokus pada penipuan dalam layanan data dan internet yang belum dilakukan pada penelitian sebelumnya. Penipuan pada layanan data dan internet dilakukan melalui penyalahgunaan konfigurasi jaringan / IT oleh karyawan. Data yang digunakan berdasarkan aktivitas log dari administrator yang bertanggung jawab untuk menjalankan konfigurasi. Penipuan tidak dapat diselesaikan dengan query biasa, sehingga diperlukan proses untuk membentuk pola data yang kemudian diproses melalui data mining. Penelitian memerlukan pengembangan model data yang sesuai dengan kebutuhan deteksi penipuan.

Penelitian ini meliputi beberapa tahap: pengumpulan data, pemilihan data, integrasi data, pembersihan data, transformasi data, identifikasi komponen utama, penanganan data yang tidak seimbang, dan proses klasifikasi. Identifikasi komponen yang signifikan menggunakan Principal Component Analysis. Data yang digunakan memiliki proporsi yang tidak seimbang. Data yang tidak seimbang tersebut ditangani dengan metode Synthetic Minority Over-sampling Technique. Metode klasifikasi yang digunakan adalah klasifikasi terbimbing yaitu Naive Bayes Classifier yang mengasumsikan bahwa tidak ada ketergantungan antar atribut. Proses Principal Component Analysis menunjukkan bahwa atribut yang signifikan yaitu: indication, sid, access_device, username, services, dan timestamp. Naive Bayes Classifier melakukan simulasi dengan 48,361 data dan mendapatkan hasil yang baik dengan akurasi lebih dari 91%.

Kata Kunci : data mining, fraud detection, principal component analysis, classification algorithm, Naive Bayes Classifier.