



DETEKSI ANOMALI MENGGUNAKAN *ISOLATION FOREST* BELANJA BARANG PERSEDIAAN KONSUMSI PADA SATUAN KERJA KEPOLISIAN REPUBLIK INDONESIA

Ahmad Zulfikar¹, Farhan Ariq Rahmani², Nurul Azizah³

^{1,2,3}Direktorat Jenderal Perbendaharaan, Kementerian Keuangan, Pangkal Pinang

Alamat Korespondensi: azulfikar@kemenkeu.go.id

Abstract

With frequent transaction but relatively variative in nominal value, manual supervision of goods purchasing for operational needs such as operational inventory expense (account code 521811) becomes very expensive and time-consuming. As a result, expenditure business process on this account is vulnerable to fraudulent practices. The Budget Section (BA) 060-The Indonesian National Police faces the greatest risk of fraud on that account because the realization value of 521811 account is the highest compared to other institutions. One of the early indications of fraudulent transactions is usually from the irregularity of the transaction pattern from normal behavior (anomalies). Machine Learning model can be used to automate internal control over this risk with much greater efficiency and effectiveness. This study tries to detect anomalies at BA 060 using the Isolation Forest method. The experiment results show that the most optimal model to detect anomalies of operational inventory expense (account 521811) in working units of BA 060-Polri at the contamination parameter value of 0.3%. Further development of this model can be used as an additional feature in the payment module of the SAKTI application system to automate anomaly detection instantly when the operator input the inventory transaction.

Abstrak

Dengan transaksi yang sering namun dalam nilai yang cukup bervariasi, pengawasan belanja barang untuk kebutuhan operasional seperti belanja barang persediaan konsumsi (kode akun 521811) secara manual menjadi sangat mahal dan menyita waktu. Akibatnya, proses bisnis belanja akun 521811 ini menjadi rentan praktik *fraud*. Bagian anggaran (BA) 060-Kepolisian Republik Indonesia menghadapi risiko *fraud* realisasi akun 521811 paling besar karena memiliki total realisasi belanja akun 521811 paling tinggi daripada BA lainnya. Salah satu indikasi awal transaksi *fraud* biasanya ada pada ketidakwajaran pola transaksi dari perilaku normal (anomali). Pemodelan *machine learning* dapat digunakan untuk mengotomatisasi pengawasan internal atas risiko anomali ini dengan lebih efektif dan efisien. Penelitian ini mencoba melakukan deteksi anomali menggunakan metode *isolation forest* pada BA 060. Hasil yang didapat, model mampu mendeteksi anomali



transaksi belanja barang persediaan pada satker BA 060-Polri secara optimal pada nilai *contamination* parameter 0,3%. Pengembangan lebih lanjut dari model ini dapat dijadikan fitur tambahan dalam modul pembayaran aplikasi SAKTI untuk mengotomatisasi deteksi anomali pada saat operator melakukan proses perekaman transaksi persediaan.

Keywords: *fraud, internal control system, unsupervised machine learning*

JEL Classification: H50

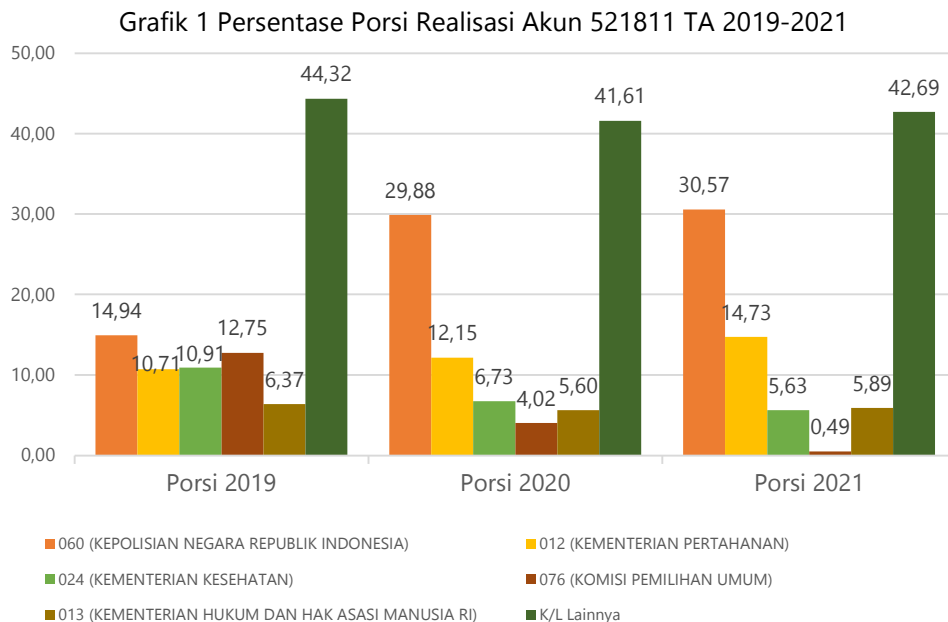
PENDAHULUAN

Transaksi yang sering dan rutin namun dalam nilai yang bervariasi, objek belanja barang persediaan konsumsi (akun 521811) seperti alat tulis kantor (ATK) rentan disalahgunakan. Menurut hasil penelitian Sholihah et al. (2016) atas audit badan pemeriksa keuangan (BPK) terhadap laporan keuangan pemerintah daerah (LKPD) tahun 2014 pada salah satu pemerintah kabupaten, terdapat temuan indikasi *fraud* objek belanja ATK pada 31 dari 33 satuan kerja perangkat daerah (SKPD) di kabupaten tersebut. Dalam konteks pemerintah pusat, Anggraeni (2018) menemukan adanya salah saji saldo persediaan pada salah satu badan layanan umum (BLU) Universitas karena inkonsistensi staf operator dalam memasukkan harga/nilai per unit untuk persediaan yang sama antar transaksi yang berbeda.

Grafik 1 membandingkan realisasi belanja akun 521811 pada lima kementerian/lembaga (K/L) dengan porsi realisasi terbesar dibandingkan dengan total realisasi K/L lainnya selama tahun 2019–2021. Lima K/L yang memiliki rata-rata porsi belanja akun 521811 tertinggi pada tahun anggaran 2019-2021 adalah Kepolisian Negara Republik Indonesia (BA 060), Kementerian Pertahanan (BA 012), Kementerian Kesehatan (BA 024), Komisi

Pemilihan Umum (BA 076), dan Kementerian Hukum dan Hak Asasi Manusia (BA 013). Selama tiga tahun tersebut, gabungan kelima K/L ini memiliki porsi antara 55,68%-58,39% dari total realisasi belanja akun 521811 seluruh K/L di Indonesia.

Dari tahun 2019-2021, Kepolisian Negara Republik Indonesia (Polri) secara konsisten menempati posisi tertinggi sebagai K/L dengan porsi realisasi akun 521811 terbesar secara nasional dan terus meningkat dari tahun ke tahun. Porsi realisasi akun 521811 BA 060 (Polri) pada tahun 2019 sebesar 14,94% dari total realisasi nasional, kemudian meningkat menjadi 29,88% pada tahun 2020, dan makin meningkat menjadi 30,57% porsi realisasi nasional pada tahun 2021. Dengan porsi realisasi terbesar dan makin meningkat dari tahun ke tahun, risiko-risiko penyimpangan belanja akun 521811 pada satker BA Polri pasti juga lebih tinggi dibandingkan satker BA lainnya. Adapun potensi penyimpangan yang dapat terjadi seperti adanya ketidaksesuaian belanja terhadap kebutuhan atau bahkan risiko tindakan kecurangan (*fraud*). Akan tetapi, sistem belanja pemerintah saat ini belum tindak penyimpangan pada belanja barang operasional khususnya Belanja mampu



Sumber: Direktorat Jenderal Perbendaharaan (2022), diolah peneliti

secara maksimal mencegah barang persediaan konsumsi (akun 521811). Akun belanja 521811 digunakan untuk mencatat belanja persediaan keperluan operasional sehari-hari sehingga cenderung memiliki pola perilaku belanja bulanan yang cukup konsisten dari tahun ke tahun. Dengan kondisi ini, potensi *fraud* dapat dilihat dari anomali atau penyimpangan dari pola yang wajar. Oleh karena itu, penggunaan metode *anomaly detection* dapat menjadi salah satu pilihan untuk mengenali potensi *fraud* pada belanja barang persediaan konsumsi yang sifatnya rutin tersebut.

Pada saat ini belum terdapat standar yang jelas batas nominal transaksi belanja akun 521811 yang tergolong anomali. Oleh karena itu, pembahasan dalam tulisan ini menggunakan metode *unsupervised machine learning anomaly detection* sebagai salah satu alternatif. Dengan

metode *unsupervised* ini, alat analisis mengolah seluruh data transaksi belanja dari akun 521811 yang termasuk dalam transaksi BA 060 Polri di tingkat nasional. Selanjutnya, pemodelan *machine learning* secara mandiri mengolah informasi dari data tersebut untuk menentukan kondisi informasi belanja yang dianggap normal maupun yang termasuk dalam kriteria anomali.

Salah satu teknik dalam metode *unsupervised machine learning anomaly detection* yang sering digunakan adalah *isolation forest*. Menurut Young (2020), setidaknya ada dua alasan *isolation forest* disebut sebagai algoritma pendeteksi anomali yang paling baik. Pertama, kinerja pendeteksian metode ini relatif lebih unggul dibandingkan metode lain seperti *local outliers factors* (LOF) dan *k-nearest-neighbors* (KNN). Pendapat ini mengacu pada tes kinerja dan presisi *receiver operating*

characteristic (ROC) yang dilakukan menggunakan *benchmark* data dari *python outlier detection package* (PyOD). Kedua, kecepatan algoritma dalam memproses data dibandingkan model lain. Selain cepat, *isolation forest* tidak memerlukan memori perangkat yang besar ketika memproses data. Karena efisiensinya, model ini sering digunakan untuk pemrosesan data berskala besar (*big data*) sehingga sejalan dengan potensi dari volume data yang terus meningkat.

Penelitian ini ditujukan untuk mendeteksi anomali transaksi belanja barang persediaan konsumsi (akun 521811) pada satuan kerja bagian anggaran Kepolisian Republik Indonesia (BA-060) menggunakan metode *isolation forest*. Model deteksi anomali yang terbentuk diharapkan dapat mempermudah para pemangku kepentingan dalam melakukan pengendalian, termasuk membantu meminimalisasi potensi *fraud* atas belanja barang persediaan tersebut.

TINJAUAN LITERATUR

Persediaan

Persediaan dalam penelitian ini merupakan persediaan dalam konteks pemerintah pusat. Menurut pernyataan standar akuntansi pemerintah nomor 5 (PSAP No. 5) tentang akuntansi persediaan, persediaan adalah aset lancar dalam bentuk barang atau perlengkapan yang dimaksudkan untuk mendukung kegiatan operasional pemerintah dan barang-barang yang dimaksudkan untuk dijual dan/atau diserahkan dalam rangka pelayanan kepada masyarakat.

Bagan Akun Standar

Berdasarkan Pasal 1 Nomor 1 pada Peraturan Menteri Keuangan Nomor 214/PMK.05/2013, bagan akun standar (BAS) adalah daftar kodifikasi dan klasifikasi terkait transaksi keuangan yang disusun secara sistematis sebagai pedoman dalam perencanaan, penganggaran, pelaksanaan anggaran, dan pelaporan keuangan pemerintah. Ada 12 segmen BAS, yaitu: segmen satker, KPPN, akun, program, *output*, dana, bank, kewenangan, lokasi, anggaran, antar entitas, dan segmen cadangan. Khusus untuk kodifikasi segmen akun dalam BAS ditetapkan dengan Keputusan Direktur Jenderal Perbendaharaan Nomor KEP-211/PB/2018 sebagaimana telah beberapa kali dimutakhirkan, terakhir dengan Keputusan Direktur Jenderal Perbendaharaan Nomor KEP-331/PB/2021.

Belanja Barang Persediaan Konsumsi

Berdasarkan Keputusan Dirjen Perbendaharaan Nomor KEP-211/PB/2018 j.o. KEP-331/PB/2021 tentang kodifikasi segmen akun BAS, transaksi belanja barang persediaan konsumsi dicatat dengan kode akun 521811. Hal tersebut karena akun ini menghasilkan barang milik negara (BMN) berupa aset lancar persediaan, akun 521811 ini hanya dicatat dalam jurnal kas, sementara jurnal akrualnya dicatat menggunakan kode akun 117111 (persediaan barang konsumsi).

Persediaan Barang Konsumsi

Persediaan barang konsumsi (akun 117111) adalah aset lancar untuk keperluan operasional sehari-hari pemerintah yang ditujukan untuk disimpan terlebih dahulu atau belum untuk digunakan/dikonsumsi segera setelah dibeli. Kategori barang yang

termasuk ke dalam persediaan konsumsi antara lain ATK, bahan cetakan, alat-alat rumah tangga, dan suplai komputer seperti *cartridge/tinta/toner printer*.

Kecurangan (*Fraud*)

Bologna et al. (1993) mendefinisikan *fraud* sebagai "*criminal deception intended to financially benefit the deceiver*" atau kecurangan adalah penipuan kriminal yang bermaksud untuk memberi manfaat keuangan kepada pelaku. Amrizal (2004) mencantumkan tiga langkah kecurangan (*fraud*), yaitu (1) tindakan/*the act*, (2) penyembunyian/*the concealment*, dan (3) konversi/*the conversion*. Contoh *fraud* pada aset persediaan seperti pelaku melakukan tindakan (*act*) membuat transaksi fiktif pembelian persediaan, tetapi karena fisik persediaannya tidak ada, pelaku akan menyembunyikan (*concealment*) kecurangan tersebut dengan membuat bukti transaksi pengeluaran/konsumsi persediaan fiktif. Selanjutnya, pelaku dapat dengan leluasa mengonversi (*conversion*) manfaat finansial dari transaksi pengadaan dan konsumsi fiktif tersebut.

Red Flag

Menurut Amrizal (2004), petunjuk adanya kecurangan biasanya ditunjukkan oleh munculnya gejala-gejala (*symptoms*) yang di luar kebiasaan seperti adanya perubahan gaya hidup atau perilaku seseorang, dokumentasi yang mencurigakan, keluhan dari pelanggan ataupun kecurigaan dari rekan sekerja. Gejala yang menyimpang dari kenormalan tersebut dinamakan *red flag*. Meskipun tidak semua indikasi *red flag* merupakan tindak kecurangan

(*fraud*), namun *red flag* biasanya selalu muncul di setiap kasus kecurangan yang terjadi.

Penyalahgunaan Aset

Menurut *Association of Certified Fraud Examiners* (2000), ada tiga kategori *fraud*, yaitu kecurangan laporan keuangan (*financial statement fraud*), penyalahgunaan aset (*asset misappropriation*), dan Korupsi (*Corruption*). Kecurangan belanja persediaan termasuk ke dalam kategori penyalahgunaan aset. Amrizal (2004) menekankan bahwa pemahaman yang tepat atas pengendalian intern yang baik dalam pos-pos aset akan sangat membantu dalam mendeteksi kecurangan penyalahgunaan aset.

Sistem Pengendalian Internal

Menurut Mulyadi (2014), ada empat unsur pokok dalam sistem pengendalian internal (SPI), yaitu (1) pemisahan tanggung jawab secara tegas pada struktur organisasi, (2) sistem wewenang dan prosedur pencatatan yang dapat melindungi aset, utang, pendapatan, dan biaya secara cukup, (3) praktik yang sehat dalam pelaksanaan tugas dan fungsi organisasi, serta (4) kualitas SDM yang sesuai dengan tanggung jawabnya. Dari empat unsur pokok SPI tersebut, unsur pertama (pemisahan kewenangan) menjadi kelemahan utama dalam proses transaksi belanja barang persediaan, terutama pada satker kecil dengan jumlah personel yang terbatas. Dalam konteks transaksi belanja persediaan konsumsi, risiko *fraud* akan makin besar apabila tidak ada pemisahan kewenangan antara pihak yang melakukan kegiatan pengadaan atau

melakukan pembelian persediaan dengan pihak yang menyimpan dan menjaga persediaan pada suatu satker.

Anomali

Alla & Adari (2019) mendefinisikan anomali sebagai hasil atau nilai yang menyimpang dari yang diharapkan, meskipun kriteria yang jelas untuk menentukan sebuah anomali bisa berbeda-beda antarsituasi. Terminologi *red flag* sama dengan anomali. Data yang bersifat anomali dapat dideteksi keberadaannya dengan melihat perbedaan karakteristiknya dengan kondisi data yang dianggap normal.

Machine Learning

Menurut Buslim & Iswara (2019), kumpulan data yang besar atau dikenal dengan istilah *big data* dapat dianalisis dengan berbagai macam teknik. Salah satu teknik yang dapat digunakan dalam mengolah *big data* adalah *machine learning*. Teknik *machine learning* merupakan salah satu varian dari sistem kecerdasan buatan (*artificial intelligence*) yang memungkinkan komputer dapat belajar tanpa diprogram secara eksplisit. Auffarth (2021) mengungkapkan bahwa "*Machine learning is the study of algorithms that improve through experience. These algorithms or models can make systematic, repeatable, validated decisions based on data.*"

Supervised vs Unsupervised Machine Learning

Menurut Brownlee (2016), *Machine Learning* mempunyai dua tipe teknik yaitu *supervised learning* dan *unsupervised learning*. Umumnya, metode *supervised learning* digunakan untuk melakukan dua tipe analisis, yaitu: klasifikasi (*classification*) dan regresi (*regression*). Klasifikasi memisahkan data

menjadi beberapa kategori sesuai dengan label yang diberikan, sedangkan regresi digunakan untuk memahami hubungan/pengaruh antara variabel dependen dan independen.

Sementara itu, Delua (2021) mendefinisikan *unsupervised learning* sebagai pendekatan pembelajaran mesin dari set data tanpa memberikan label tertentu. Mesin akan secara mandiri mengungkap pola-pola tersembunyi di dalam data tanpa intervensi manusia. Ada tiga tipe analisis yang dapat dilakukan menggunakan metode *unsupervised learning*, yaitu: klasterisasi (*clustering*), asosiasi (*association*), dan pengurangan dimensi (*dimensionality reduction*). Klasterisasi mengelompokkan data yang tidak diberikan label berdasarkan kesamaan atau perbedaannya. Asosiasi mencari hubungan tersembunyi dalam set data yang besar. Pengurangan dimensi digunakan untuk mengurangi jumlah fitur/dimensi/variabel dalam suatu set data dengan cara mengelompokkan dimensi-dimensi yang mempunyai kemiripan dan hubungan yang kuat.

Baik *supervised* maupun *unsupervised learning* dapat digunakan untuk melakukan deteksi anomali. Sebagaimana disebutkan pada bagian sebelumnya, saat ini belum ada standar yang disepakati terkait berapa nilai transaksi belanja barang persediaan konsumsi yang dianggap wajar sehingga tidak bisa ditentukan (*labelling*) mana transaksi normal dan mana transaksi anomali. Oleh karena itu, penelitian ini melakukan deteksi anomali transaksi belanja barang persediaan konsumsi menggunakan metode *unsupervised machine learning anomaly detection*.

Isolation Forest

Metode *isolation forest* adalah salah satu metode *unsupervised machine learning anomaly detection* yang pertama kali diperkenalkan oleh Liu et al. (2008). Metode *isolation forest* adalah algoritma *unsupervised learning* dan *nonparametric* yang didasarkan atas algoritma pohon keputusan atau *trees*. Deteksi anomali menggunakan *isolation forest* melewati dua tahap proses, yaitu tahap pelatihan (*training*) dan tahap evaluasi (*evaluation*). Pada tahap pelatihan, mesin membuat pohon isolasi menggunakan sampel dari seluruh set data. Kemudian pada tahap evaluasi, seluruh individu/*instance* dalam set data melalui pohon isolasi untuk menghitung skor anomali setiap *instance*.

Penelitian Terdahulu

Metode *isolation forest* banyak digunakan untuk mendeteksi *fraud* transaksi keuangan, seperti *fraud* kartu kredit dan pencucian uang. Ounacer et al. (2018) membandingkan kinerja empat teknik *unsupervised anomaly detection* pada 284.807 transaksi kartu kredit dari Kaggle menemukan bahwa *isolation forest* memberikan kinerja yang paling baik jika dibandingkan dengan tiga metode lainnya, yaitu *one class SVM* (*support vector machine*), *local outlier factor* (LOF), dan *k-means*. Vijayakumar et al. (2020) yang membandingkan performa *isolation forest* dengan LOF juga menemukan bahwa *isolation forest* merupakan metode terbaik untuk melakukan deteksi anomali pada transaksi kartu kredit. Shokry et al. (2020) juga menemukan bahwa *isolation forest* merupakan model terbaik untuk mendeteksi pencucian uang pada transaksi perbankan jika dibandingkan dengan *one class SVM*. Belum ada

penelitian yang khusus melakukan deteksi anomali pada transaksi belanja pemerintah. Oleh karena itu, penelitian ini berusaha untuk menggunakan *isolation forest* sebagai metode deteksi anomali transaksi belanja pemerintah, khususnya belanja barang persediaan.

METODE PENELITIAN

Penelitian ini menggunakan teknik *unsupervised machine learning anomaly detection* dengan metode *isolation forest*. metode *isolation forest* merupakan salah satu modul dalam *library machine learning scikit-learn* yang berjalan dengan bahasa pemrograman Python. Aplikasi Google Colab digunakan sebagai *coding environment* untuk menjalankan Python.

Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder realisasi belanja per bulan akun 521811 pada seluruh satker BA 060-Polri selama tahun 2019 s.d. 2021. Data tersebut bersumber dari data realisasi belanja APBN pada Direktorat Jenderal Perbendaharaan, Kementerian Keuangan RI.

Data Cleansing

Data yang diterima masih kotor (*dirty/noisy*) sehingga dilakukan data *cleansing* sebelum proses deteksi anomali dilakukan. Ridzuan & Zainon (2019) mendefinisikan data kotor sebagai data yang tidak akurat, tidak konsisten, dan tidak lengkap karena kesalahan yang ditemukan dalam kumpulan data. Data yang diterima merupakan gabungan transaksi realisasi akun 521811, baik realisasi belanja maupun pengembalian belanja atau

koreksi akun. Indikasi dari hal tersebut, ada beberapa data realisasi yang bernilai minus. Data realisasi bulanan akun 521811 yang bernilai minus bisa diakibatkan oleh adanya koreksi kurang atau pengembalian belanja yang melebihi realisasi belanja pada bulan bersangkutan.

Terdapat kemungkinan data yang bernilai positif masih bercampur dengan data realisasi non belanja seperti koreksi tambah. Akan tetapi, data kotor yang bernilai positif tidak bisa dieliminasi secara visual atau manual sehingga tidak dieliminasi dalam proses data *cleansing*. Koreksi tambah yang menimbulkan nilai yang tidak wajar akan otomatis tereliminasi dari proses deteksi anomali.

Data Transformation

Nilai rupiah realisasi antar satker sangat bervariasi sesuai dengan nilai pagu satker bersangkutan. Oleh karena itu, perlu dilakukan transformasi data realisasi menjadi persentase belanja akun 521811 terhadap total pagu belanja barang agar dapat dibandingkan realisasi antar satker dengan lebih relevan. Rasio persentase realisasi belanja akun 521811 terhadap total pagu belanja barang lebih *reliable* dibandingkan rasio terhadap akun 521811 secara langsung karena selama tahun berjalan, pagu per akun dimungkinkan untuk berubah dan mengalami penyesuaian melalui mekanisme revisi petunjuk operasional kegiatan (POK). Sementara itu, total pagu belanja barang biasanya jarang berubah karena harus melalui proses revisi daftar isian pelaksanaan anggaran (DIPA) dengan melakukan pergeseran pagu antar jenis belanja atau penambahan/pengurangan pagu DIPA dari satker lain.

Menentukan Nilai Contamination Parameter

Nilai *contamination parameter* digunakan untuk mengajarkan model berapa proporsi anomali dalam dataset. Nilai *contamination parameter* ditentukan secara *trial and error* melalui beberapa kali percobaan nilai parameter. Secara visual akan terlihat model yang paling sesuai dan mewakili dari *scatterplot* yang terbentuk.

Menghitung Anomaly Score

Anomaly score adalah nilai yang menunjukkan kriteria anomali *row* data. Secara matematis, *anomaly score* didapat dari hasil perhitungan formula di bawah ini:

$$s(x, n) = 2 \frac{E(h(x))}{c(n)}$$

Ket.

$H(x)$: *path length*

$E(h(x))$: rata-rata *path length* dari setiap *isolation tree*.

$C(n)$: rata-rata $h(x)$ dari nilai variabel ke- n

Path length adalah jarak yang mengukur titik di mana data tidak dapat dibagi lagi dengan *root node* dari struktur algoritma pohon keputusan. *Anomaly score* sendiri memiliki rentang nilai dari -1 sampai dengan 1. *Anomaly score* yang makin mendekati -1 dianggap anomali dan makin mendekati nilai 1 dianggap normal.

HASIL PENELITIAN DAN PEMBAHASAN

Untuk melatih model *isolation forest*, maka hanya diambil dua kolom dari *dataset*, yaitu "BULAN" dan "PERSENTASE". Kolom "BULAN" menampilkan data periode bulan

terjadinya transaksi. Sedangkan kolom "PERSENTASE" menampilkan data periode bulanan atas persentase belanja persediaan konsumsi (akun 521811) terhadap total pagu belanja barang. Set data yang hanya berisi *features* yang diperlukan model ini kemudian diberi nama "data nasional".

Dari proses *data cleansing* dan *data transformation*, didapatkan set data dengan statistik deskriptif seperti yang ditampilkan oleh Tabel 1. Dari tahun 2019 sampai dengan tahun 2021, rata-rata tertinggi persentase realisasi belanja barang persediaan konsumsi (akun 521811) terhadap total pagu

belanja barang di satker BA 060-Polri terjadi pada bulan April dengan rata-rata persentase 0,6643%. Selama tiga tahun tersebut, total satker yang melakukan realisasi belanja akun 521811 paling sedikit terjadi pada bulan Januari, yaitu hanya 1.300 satker. Nilai minimum dan maksimum per bulan menunjukkan tingginya variasi persentase realisasi belanja akun 521811 Satker BA 060-Polri, dari nilai minimum 0,0000% sampai dengan 73,946%.

Proses selanjutnya adalah memasukkan *dataset* ke dalam fungsi *isolation forest* sebagaimana yang ditampilkan oleh Gambar 1. Parameter

Tabel 1 Statistik Deskriptif Persentase Bulanan Realisasi Belanja Akun 521811
BA 060-Polri Tahun 2019-2021

Bulan	Mean	Standard Deviation	Minimum	Maximum	Count
Jan	0,6116495	1,6984140	0,0003394	17,0165897	1.300
Feb	0,5146982	1,2677632	0,0008167	24,2775211	2.808
Mar	0,4992162	1,3477575	0,0007967	34,3017570	2.941
Apr	0,6642793	3,0948643	0,0009004	57,5222310	2.911
Mei	0,5518261	2,2867625	0,0004091	52,3638541	2.831
Jun	0,5190978	2,6490202	0,0005149	63,8406067	2.810
Jul	0,4764332	1,6553660	0,0008207	52,8147396	2.919
Agu	0,4569261	1,8012051	0,0001533	43,1229356	2.833
Sep	0,4301082	1,9934281	0,0000004	57,3973523	2.857
Okt	0,4300444	1,8816261	0,0003273	73,9463839	2.910
Nov	0,4310028	1,7895555	0,0002664	71,0918830	2.974
Des	0,4348209	2,2469765	0,0000038	68,4648888	2.728

Sumber: Direktorat Jenderal Perbendaharaan (2022), diolah peneliti

Gambar 1 Kode Eksekusi *Isolation Forest*

```
datamodel=dftrain[['BULAN','PERSENTASE']]
```

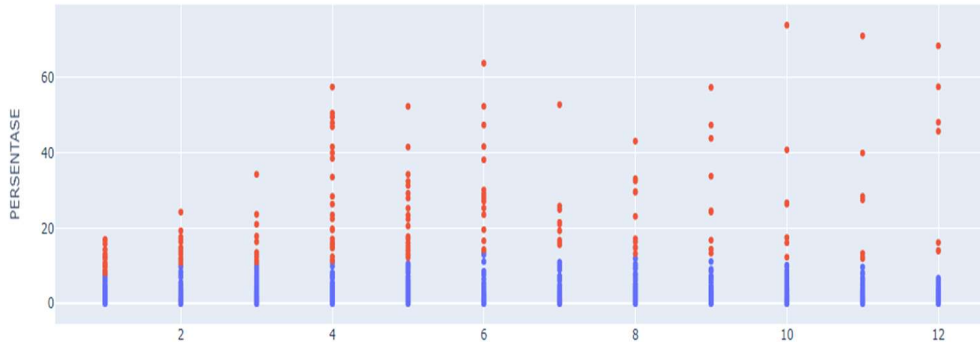
```
model = IsolationForest(contamination=float(0.003), max_samples='auto',  
n_estimators=100)  
model.fit(datamodel)
```

Sumber: diolah peneliti

dalam *isolation forest* yang digunakan antara lain *contamination*, *max_samples*,

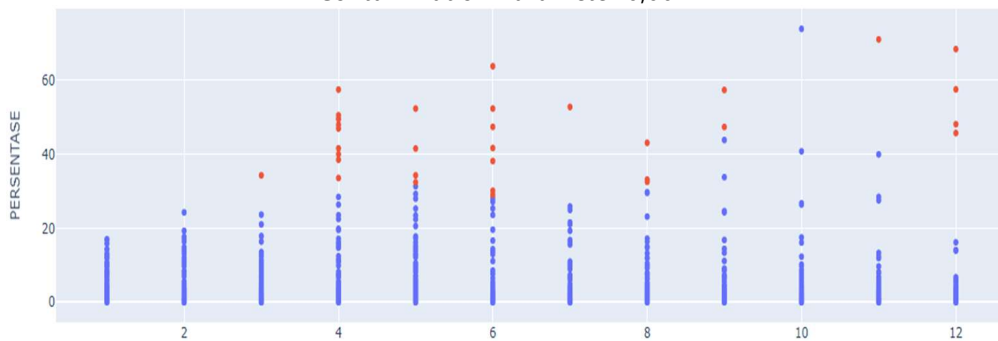
dan *n_estimators*. Nilai *contamination* parameter akan dibahas di bagian

Grafik 2 Hasil Prediksi Model *Isolation Forest* dengan *Contamination Parameter* 0,005



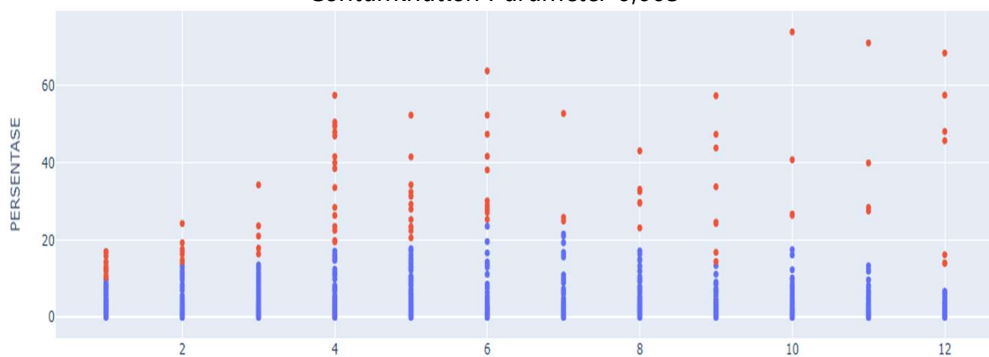
Sumber: diolah peneliti

Grafik 3 Hasil Prediksi Model *Isolation Forest* dengan *Contamination Parameter* 0,001



Sumber: diolah peneliti

Grafik 4 Hasil Prediksi Model *Isolation Forest* dengan *Contamination Parameter* 0,003



Sumber: diolah peneliti

Gambar 2 Kode Konversi *Anomaly Score* dan Visualisasi Data

```
dftrain['outliers']=pd.Series(model.predict(datamodel)).apply(lambda
x:'yes' if (x == -1) else 'no')

fig = px.scatter(dftrain, x='TANGGAL', y='PERSENTASE',
color='outliers', title='Belanja Persediaan Satker Nasional')

fig.update_xaxes(
    rangeslider_visible=True,
)
fig.show()
```

Sumber: diolah peneliti

selanjutnya, karena penentuan nilainya memengaruhi akurasi model mendeteksi anomali. Parameter kedua adalah *max_samples* yang menentukan berapa banyak sampel yang dapat digunakan model untuk berlatih. Di model ini, *max_samples* dipasang 'auto' yang memiliki arti semua data (populasi) digunakan untuk latihan model. Terakhir, parameter *n_estimators* dipilih ke angka 100 yang merupakan angka standar (default) dari model *isolation forest*.

Proses Penentuan Nilai Contamination Parameter yang Tepat

Dari hasil *trial and error*, terpilih tiga nilai *contamination parameter*, yaitu 0,005; 0,003; dan 0,001. Hasil pengujian ketiga parameter kemudian dikonversi dari *anomaly score* menjadi label *yes* dan *no* dengan menggunakan kode pemrograman sebagaimana ditampilkan pada Gambar 2. Kode *yes* berarti baris data tersebut merupakan data anomali (*outlier*), sedangkan kode *no* merupakan tanda untuk baris data yang normal (*inlier*). Langkah selanjutnya, melakukan visualisasi data realisasi yang merupakan *outliers*. Sedangkan tanda titik berwarna biru

Tabel 2 Data *Outlier* Hasil Model *Isolation Forest*

KANWIL_DJPB	TAHUN	BULAN	SATKER	REAL_PERSEDIAAN	PAGU_BARANG	PERSENTASE
K02	2019	10	640338	43,368,003,498	58,647,903,000	73.94638389
K17	2019	11	645852	14,182,474,490	19,949,499,000	71.09188301
K02	2020	12	640338	53,412,887,116	78,015,006,000	68.46488881
K02	2021	6	640338	43,276,268,580	67,787,997,000	63.84060674
K31	2019	12	669282	9,897,169,604	17,194,095,000	57.56144539
...	...	...	...	...	...	...
K07	2019	1	641834	3,648,866,069	33,850,556,000	10.77933866
K21	2019	1	644861	3,933,280,720	37,158,924,000	10.58502318
K16	2021	1	645362	5,048,632,280	48,930,076,000	10.31805526
K14	2019	1	643784	3,276,929,769	32,536,497,000	10.07155063
K11	2019	1	417899	192,500,000	1,914,792,000	10.05331127

Sumber: diolah peneliti

merupakan titik data realisasi yang wajar. dalam *scatter plot diagram*. Tanda titik berwarna merah menunjukkan titik data realisasi yang merupakan *outliers*. Sedangkan tanda titik berwarna biru merupakan titik data realisasi yang wajar.

Berdasarkan hasil pengujian deteksi anomali menggunakan nilai *contamination parameter* sebesar 0,005 seperti *scatter plot* yang ditampilkan Grafik 2, jumlah data transaksi yang terdeteksi anomali sebanyak 165 baris data. Secara visual, masih ada *plot* data yang sebenarnya masih dekat dengan data lainnya, dideteksi sebagai anomali oleh model.

Apabila menggunakan nilai *contamination parameter* 0,001 seperti Grafik 3, jumlah anomali data transaksi yang terdeteksi sebanyak 33 transaksi. Berkebalikan dengan Grafik 2, *plot* data yang tampak sudah jauh dari data lainnya masih dianggap normal (*inlier*) oleh model. Secara visual, nilai *contamination parameter* terbaik adalah 0,003 atau 0,3% seperti yang diperlihatkan pada Grafik 4. Model mendeteksi sebanyak 99 baris data sebagai anomali. *Plot* data yang jauh dari data lainnya dianggap anomali, sementara yang masih berdekatan dianggap normal oleh sistem.

Tabel 2 menunjukkan transaksi yang dideteksi sebagai anomali oleh model. Berdasarkan tabel tersebut, data anomali dengan persentase persediaan tertinggi adalah transaksi kode satker 640338 di bulan Oktober 2019 dengan nilai belanja persediaan (5211811) sebesar Rp43,37 miliar dari pagu belanja barang (52) sebesar Rp58,65 miliar atau persentase persediaan sebesar 73,95%. Sedangkan data anomali dengan persentase persediaan terendah adalah transaksi kode satker 646307 di bulan

September 2022 dengan nilai belanja persediaan sebesar Rp400 dari pagu belanja barang Rp111,61 Miliar atau persentase persediaan sebesar 0,0000007%.

Langkah selanjutnya adalah menyimpan data yang sudah memperoleh label *inlier* dan *outlier* menggunakan *library joblib*. Karena sudah memiliki label, data dari hasil model *Isolation Forest* tersebut juga dapat digunakan menjadi data training untuk membuat model deteksi anomali *Supervised Machine Learning*.

KESIMPULAN DAN SARAN

Kesimpulan

Hasil pengujian menunjukkan bahwa model *isolation forest* memiliki potensi untuk mendeteksi anomali yang terjadi pada data transaksi akun belanja persediaan (521811). Berdasarkan hasil uji coba beberapa nilai *contamination parameter*, nilai 0,3% merupakan nilai yang paling optimal dalam uji coba model yang dipilih dalam mendeteksi anomali. Terlepas dari valid atau tidaknya indikasi *fraud* pada data yang dideteksi sebagai anomali oleh model, hasil *learning* dari penelitian ini dapat dimasukkan sebagai fitur tambahan yang berfungsi sebagai *early warning system* pada aplikasi otorisasi belanja satker, seperti aplikasi sistem aplikasi keuangan tingkat instansi (SAKTI).

Saran

Penelitian ini menggunakan data realisasi belanja akun 521811 secara bulanan. Nilai realisasi bulanan ini merupakan nilai neto yang didapat dari nilai realisasi belanja ditambah/dikurangi dengan transaksi koreksi akun dan/atau transaksi pengembalian belanja. Transaksi koreksi

akun dan pengembalian belanja ini biasanya merupakan koreksi dan/atau pengembalian belanja persediaan bulan sebelumnya. Untuk penelitian selanjutnya, disarankan untuk menggunakan nilai per transaksi pada *surat perintah pencairan dana* (SP2D) sehingga koreksi akun dan pengembalian belanja tidak termasuk ke dalam model.

IMPLIKASI DAN KETERBATASAN

Implikasi

Penelitian deteksi anomali dengan menggunakan *isolation forest* dapat diimplementasikan sebagai alternatif pencegahan melalui deteksi *fraud*. Pemodelan ini memiliki potensi untuk dikembangkan dan ditambahkan ke dalam modul pembayaran aplikasi SAKTI sebagai fitur deteksi anomali secara otomatis saat penginputan transaksi oleh operator.

Keterbatasan

Data realisasi belanja akun 521811 yang diuji hanya data total nominal transaksi per bulan, bukan data per transaksi SP2D. Dalam nilai transaksi per bulan yang menjadi objek penelitian, selain transaksi pencairan SP2D, juga terdapat transaksi koreksi SP2D dan pengembalian belanja bulan sebelumnya. Hal ini terlihat dari adanya nilai transaksi minus pada bulan dan satker tertentu yang berasosiasi dengan total koreksi "kurang" pada bulan tersebut lebih besar daripada nilai realisasi belanja bulan berkenaan. Akan tetapi, model dalam penelitian ini tidak dapat mendeteksi dan mengeliminasi koreksi "tambah" karena nilainya sama-

sama positif. Ada beberapa indikasi yang terkait koreksi tambah ini yaitu adanya nilai realisasi per bulan yang lebih dari 70% pada satker tertentu.

REFERENSI

- Alla, S & Adari, S. K. (2019). *Beginning anomaly detection using python-based deep learning with keras and pytorch*. New York: Apress.
- Amrizal. (2004). *Pencegahan dan pendeteksian kecurangan oleh internal auditor*. Jakarta. Badan Pengawasan Keuangan dan Pembangunan.
- Anggraeni, D.P. (2018). *Fraud pada tata kelola penatausahaan persediaan badan layanan umum (studi kasus pada universitas x di Kota Malang)*. Tesis. Universitas Brawijaya.
- Association of Certified Fraud Examiners. (2000). *Fraud examiners manual third edition*. Texas: Association of Certified Fraud Examiners.
- Auffarth, B. (2021). *Machine learning for time-series with python: Forecast, predict, and detect anomalies with state-of-the-art machine learning methods*. Birmingham, UK: Packt Publishing.
- Bologna, G. J., Lindquist, R. J., & Wells, J. T. (1993). *Investigation audit 4th edition*. New Jersey: Prentice-Hall.
- Brownlee, J. (2016). *Machine learning mastery with python: Understand your data, create accurate models and work projects end-to-end*. San Francisco: Machine Learning Mastery.
- Buslim, N. & Iswara, R. P. (2019). Pengembangan algoritma unsupervised learning technique pada big data analysis di media

- sosial sebagai media promosi online bagi masyarakat. *Jurnal Teknik Informatika*, 12(1), 79 – 81. <https://doi.org/10.15408/jti.v12i1.11342>.
- Delua, J. (2021). Supervised vs. unsupervised learning: What's the difference? *IBM*. Retrieved March 22, 2022 from <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>
- Direktorat Jenderal Perbendaharaan. (2018). *Keputusan Direktur Jenderal Perbendaharaan Nomor 211 Tahun 2018 tentang Kodifikasi Bagan Akun pada Bagan Akun Standar*. Jakarta.
- Direktorat Jenderal Perbendaharaan. (2018). *Keputusan Direktur Jenderal Perbendaharaan Nomor 331 Tahun 2018 tentang Kodifikasi Bagan Akun pada Bagan Akun Standar*. Jakarta.
- Direktorat Jenderal Perbendaharaan. (2022). *Dataset realisasi belanja bulanan per satker per akun secara nasional*. Jakarta.
- Kementerian Keuangan. (2013). *Peraturan Menteri Keuangan Nomor 214/PMK.05/2013 tentang Bagan Akun Standar*. Jakarta.
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *Eighth IEEE International Conference on Data Mining*, 413-422. <https://doi.org/10.1109/ICDM.2008.17>.
- Mulyadi. (2014). *Sistem akuntansi*. Cetakan Keempat. Jakarta: Salemba Empat.
- Ounacer, S., El Bour, H. A., Oubrahim, Y., Ghomari, M. Y., & Azzouazi, M. (2018). Using isolation forest in anomaly detection: The case of credit card transactions. *Periodicals of Engineering and Natural Sciences*, 6(2), 394-400. <https://doi.org/10.21533/pen.v6i2.533>
- Republik Indonesia. (2005). *Lampiran VII Peraturan Pemerintah Republik Indonesia Nomor 24 Tahun 2005 tentang Pernyataan Standar Akuntansi Pemerintah Nomor 05 – Akuntansi Persediaan*. Jakarta.
- Ridzuan, F., & Zainon, W. M. N. W. (2019). A review on data cleansing methods for big data. *Procedia Computer Science*, 161, 731-738.
- Shokry, A. M., Rizka, M. A., & Labib, N. M. (2020). Counter terrorism finance by detecting money laundering hidden networks using unsupervised machine learning algorithm. In *International Conferences ICT, Society, and Human Beings*, 89-97. https://doi.org/10.33965/ict_csc_wbc_2020_2020081012
- Sholihah, S., Alim, M. N., & Musyarofah, S. (2016). Memotret pola fraud pada rincian objek belanja yang menjadi temuan BPK. *Journal of Auditing, Finance, and Forensic Accounting*, 4(2), 101-110. <https://doi.org/10.21107/jaffa.v4i2.2783>
- Vijayakumar, V., Divya, N. S., Sarojini, P., & Sonika, K. (2020). Isolation forest and local outlier factor for credit card fraud detection system. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(4), 261-265. <https://doi.org/10.35940/ijeat.d6815.049420>
- Young, A. (2020). *Making sense of big data, isolation forest is the best anomaly for big data right now*.

Retrieved March 10,2022 from
<https://towardsdatascience.com/isolation-forest-is-the-best-anomaly-detection-algorithm-for-big-data-right-now-e1a18ec0f94f>