

# DETEKSI ANOMALI MENGGUNAKAN *ISOLATION FOREST* PADA BELANJA BARANG PERSEDIAAN SATKER POLRI

Ahmad Zulfikar, Farhan Ariq Rahmani, Nurul Azizah  
Kantor Wilayah Direktorat Jenderal Perbendaharaan Provinsi Bangka Belitung

## **Abstract**

*With relative small transaction value but frequent, manual supervision of goods purchasing for operational needs such as inventory purchasing (account 521811) becomes very expensive and time-consuming. As a result, spending on this 521811 account becomes vulnerable to fraudulent practices. With the highest realization purchasing value for account 521811, the Budget Section (BA) 060-Polri faces the greatest risk of fraud on that account compared to others. Machine Learning Modeling can be used to automate internal control over this risk with much greater efficiency and effectiveness. One of the early indications of fraudulent transactions is usually from the irregularity of the transaction pattern from normal behavior (anomalies). This study tries to detect anomalies using the Isolation Forest method at BA 060. The experiment results show that the model is able to detect anomalies of inventory purchasing (account 521811) in working units of BA 060-Polri optimally 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 to automate anomaly detection when the operator performs the transaction inputting process.*

## **Abstrak**

Dengan nilai yang transaksi yang relatif kecil namun sering, pengawasan belanja barang untuk kebutuhan operasional seperti belanja barang persediaan konsumsi (akun 521811) secara manual oleh manusia menjadi sangat mahal dan menyita waktu. Akibatnya, belanja akun 521811 ini menjadi rentan praktik *fraud*. Dengan nilai realisasi belanja akun 521811 paling tinggi, Bagian Anggaran (BA) 060-Polri menghadapi risiko *fraud* realisasi akun 521811 paling besar daripada Bagian Anggaran (BA) lainnya. Pemodelan *Machine Learning* dapat digunakan untuk mengotomatisasi pengawasan internal atas risiko ini dengan efisiensi dan efektivitas yang jauh lebih tinggi. Salah satu indikasi awal transaksi *fraud* biasanya ada pada ketidakwajaran pola transaksi dari perilaku normal (anomali). 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* sebesar 0,3%. Pengembangan lebih lanjut pemodelan ini dapat dijadikan fitur tambahan dalam modul pembayaran aplikasi SAKTI untuk mengotomatisasi deteksi anomali pada saat operator melakukan proses perekaman transaksi.

**Keywords:** Akun 521811, *Fraud*, *Internal Control System*, *Unsupervised Machine Learning*

**JEL Classification:** H50

## PENDAHULUAN

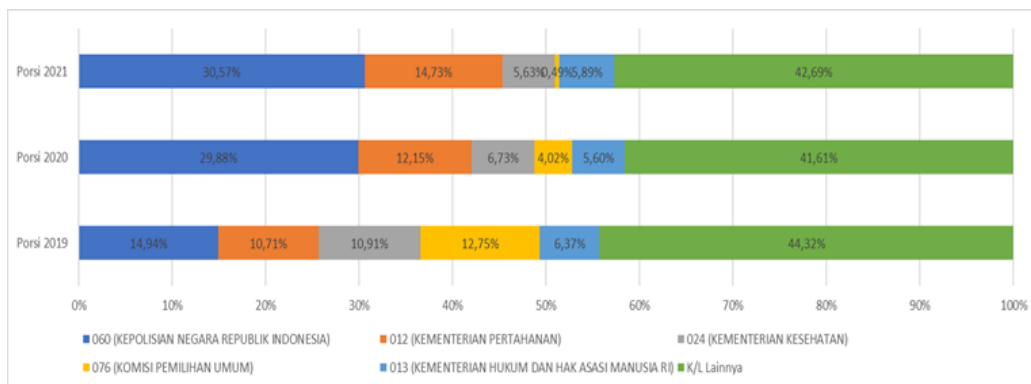
Objek belanja persediaan seperti Alat Tulis Kantor (ATK) paling umum disalahgunakan. Menurut hasil penelitian Sholihah, Alim, dan Musyarofah (2016) atas temuan audit BPK terhadap Laporan Keuangan Pemerintah Daerah (LKPD) Tahun 2014 Kabupaten tertentu (nama disamarkan), menemukan bahwa indikasi *fraud* objek belanja ATK terjadi pada 31 dari 33 Satuan Kerja Perangkat Daerah (SKPD). Dalam konteks Pemerintah Pusat, Anggraeni (2018) menemukan adanya salah saji saldo persediaan pada salah satu Badan Layanan Umum (BLU) Universitas karena inkonsistensi operator staf dalam memasukkan harga/nilai per unit untuk persediaan yang sama antar transaksi yang berbeda.

Adapun Grafik 1 di bawah ini membandingkan realisasi akun 521811 lima Kementerian/Lembaga (K/L) dengan porsi terbesar dengan 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.

Selama tiga tahun tersebut, Bagian Anggaran (BA) 060 Kepolisian Negara Republik Indonesia (Polri) secara konsisten menempati posisi pertama sebagai K/L dengan porsi tersebar terhadap realisasi akun 521811 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 semakin meningkat menjadi 30,57% porsi realisasi nasional pada tahun 2021. Dengan porsi realisasi terbesar dan semakin meningkat dari tahun ke tahun, risiko-risiko penyimpangan terkait akun 521811 pada Satker BA Polri tentu juga lebih tinggi dibandingkan Satker BA lainnya. Adapun potensi penyimpangan

**Grafik 1. Persentase Porsi Realisasi Akun 521811 TA 2019–2021**



yang dapat terjadi seperti adanya ketidaksesuaian belanja terhadap kebutuhan, bahkan rentan terhadap tindakan kecurangan (*fraud*). Akan tetapi, sistem belanja pemerintah saat ini belum mampu secara maksimal mencegah tindak penyimpangan pada belanja barang operasional khususnya belanja barang persediaan konsumsi (521811).

Karena akun belanja 521811 ini termasuk belanja keperluan operasional sehari-hari, biasanya terdapat pola perilaku belanja bulanan yang cukup konsisten dari tahun ke tahun. Potensi *fraud* dapat dilihat dari anomali atau penyimpangan dari perilaku normal. Oleh karena itu, untuk mengenali potensi *fraud* pada belanja barang persediaan yang sifatnya rutin dilakukan, kami menerapkan metode *anomaly detection*.

Akan tetapi, belum ada standar berupa data yang jelas berapa nominal transaksi belanja akun 521811 yang tergolong anomali dan normal. Oleh karena itu, metode deteksi anomali yang paling tepat adalah metode *unsupervised machine learning anomaly detection*. Dengan metode *unsupervised* ini, mesin diberikan data seluruh transaksi belanja akun 521811 yang berasal dari seluruh satker BA 060 Polri se-Indonesia, kemudian *machine learning model* secara mandiri mengolah informasi dari data yang diberikan untuk menentukan mana belanja yang anomali dan normal.

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* merupakan algoritma pendeteksi anomali yang paling baik. Pertama

kinerja pendeteksiannya yang unggul dibandingkan metode lain seperti *Local Outliers Factors* (LOF) dan K-Nearest-Neighbors (KNN)) berdasarkan tes kinerja dan presisi ROC yang dilakukan menggunakan *benchmark data* dari *Python Outlier Detection package* (PyOD). Kedua, kecepatan algoritma memproses data dibandingkan model lain. Selain cepat, *Isolation Forest* tidak memerlukan memori yang banyak ketika memproses data. Karena efisiensinya, model ini sering digunakan untuk pemrosesan data berskala besar (*big data*) sehingga prospek kedepannya terjamin ketika data semakin tumbuh.

Hasil deteksi anomali pada akun 521811 ini diharapkan dapat mempermudah para *stakeholder* melakukan pengendalian, termasuk membantu meminimalisasi potensi *fraud* terhadap belanja barang tersebut.

## TINJAUAN LITERATUR

### Persediaan

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. Penelitian ini membahas tentang persediaan barang konsumsi yang masuk ke dalam golongan persediaan untuk mendukung kegiatan operasional pemerintah dengan kode akun 117111 berdasarkan Kodifikasi Segmen Akun pada Bagan Akun Standar yang diatur dalam 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**

Akun belanja barang persediaan konsumsi (kode akun 521811) digunakan untuk mencatat transaksi belanja barang persediaan konsumsi. Karena akun ini menghasilkan aset/BMN dalam bentuk persediaan, akun 521811 ini hanya dicatat dalam jurnal kas, sementara jurnal akrualnya langsung dicatat menggunakan kode akun Persediaan Barang Konsumsi (kode akun 111711).

### **Persediaan Barang Konsumsi**

Persediaan Barang Konsumsi (111711) adalah aset lancar untuk keperluan operasional sehari-hari Pemerintah yang ditujukan untuk disimpan terlebih dahulu atau belum untuk digunakan/dikonsumsi segera setelah dibeli. Yang termasuk ke dalam persediaan konsumsi adalah ATK, bahan cetakan, alat-alat rumah tangga, maupun suplai komputer seperti *cartridge/tinta/toner printer*.

### **Kecurangan (Fraud)**

Bologna, et al. (1993) mendefinisikan "*Fraud is criminal deception intended to financially benefit the deceiver*" atau kecurangan adalah penipuan kriminal yang bermaksud untuk memberi manfaat keuangan kepada si penipu. Amrizal (2004) mencantumkan tiga langkah kecurangan (*fraud*) yaitu (1) tindakan/*the act*, (2) menyembunyian/*the concealment*, dan (3) konversi/*the conversion*. Dia mencontohkan *fraud* pada aset persediaan seperti pencurian atas harta persediaan adalah tindakan (*act*) pelaku,

kemudian pelaku akan menyembunyikan (*concealment*) kecurangan tersebut misalnya dengan membuat bukti transaksi pengeluaran fiktif. Selanjutnya setelah perbuatan pencurian dan menyembunyian dilakukan, pelaku akan melakukan konversi (*conversion*) dengan cara memakai sendiri atau menjual persediaan tersebut.

### **Red Flag**

Menurut Amrizal (2004), petunjuk adanya kecurangan biasanya ditunjukkan oleh munculnya gejala-gejala (*symptoms*) seperti adanya perubahan gaya hidup atau perilaku seseorang, dokumentasi yang mencurigakan, keluhan dari pelanggan ataupun kecurigaan dari rekan sekerja. Pada awalnya, kecurangan ini akan tercermin melalui timbulnya karakteristik tertentu, baik yang merupakan kondisi / keadaan lingkungan, maupun perilaku seseorang. Karakteristik yang bersifat kondisi/ situasi tertentu, perilaku/kondisi seseorang personal tersebut dinamakan *red flag*. Meskipun timbulnya *red flag* tersebut tidak selalu merupakan indikasi adanya kecurangan (*fraud*), namun *red flag* ini biasanya selalu muncul di setiap kasus kecurangan yang terjadi.

### **Penyalahgunaan Aset (Asset Misappropriation)**

Menurut *Association of Certified Fraud Examination* (ACFE-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.

### **Deteksi Anomali**

Lee (2021) mendefinisikan deteksi anomali (*anomaly detection*) sebagai suatu proses untuk mengidentifikasi titik data (*data point*) yang berada di luar pola perilaku normal. Alla dan Adari (2019) mendefinisikan anomali sebagai hasil atau nilai yang menyimpang dari yang diharapkan, tetapi kriteria yang jelas untuk menentukan sebuah anomali bisa berbeda-beda antar situasi. Terminologi *red flag* sama dengan anomali. Anomali kadang-kadang disebut *outlier*, *novelties* atau kebaruan, *noise* atau kebisingan, deviasi atau penyimpangan, dan eksepsi atau pengecualian (Jayawickrama, 2020).

Menurut Yulei Wang, et al. (2018), untuk mendeteksi anomali diperlukan ambang batas yang sesuai untuk menentukan dataset merupakan anomali atau bukan anomali. Dalam penelitian ini, ambang batas yang ditentukan untuk mendeteksi anomali data disebut *contamination parameter*.

### **Unsupervised Learning Technique**

Menurut Nurhayati, et al. (2019) Kumpulan data yang besar atau dikenal dengan istilah *big data* dapat dianalisis dengan berbagai macam teknik. Teknologi yang dapat digunakan dalam mengolah big data salah satunya adalah *Machine Learning* (ML). *Machine Learning* merupakan salah satu varian dari sistem kecerdasan buatan yang memungkinkan komputer dapat belajar tanpa diprogram secara eksplisit.

Menurut Ben Auffrath dalam bukunya yang berjudul *Machine Learning for Time-Series With Python* (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.*"

Menurut Jason Brownlee (2016) Machine Learning mempunyai 2 tipe teknik yaitu *supervised learning* dan *unsupervised learning*. *Unsupervised learning* adalah salah satu tipe algoritma *machine learning* yang digunakan untuk menarik kesimpulan dari *dataset* yang terdiri dari input data *labeled response*. Metode *unsupervised learning* yang paling umum adalah analisis *cluster* yang digunakan pada analisa data untuk mencari pola-pola tersembunyi atau pengelompokan dalam data.

### **Isolation Forest**

Chun hui Xiao, et al. (2018) dalam penelitiannya yang berjudul *Anomaly Detection in Network Management System Based on Isolation Forest*, menyebutkan bahwa *Isolation Forest* adalah *Unsupervised Machine Learning* yang digunakan untuk mendeteksi anomali. *Isolation Forest* membangun sebuah *ensemble* dari iTrees yang memiliki struktur setacara dengan pohon pencarian biner.

Selanjutnya, Yufei Song, et al. (2019) dalam penelitiannya yang berjudul *Isolation Forest based Detection for False Data Attacks in Power Systems*, mengungkapkan prinsip dasar dari pendeteksian *Isolation Forest* ini adalah dengan memisahkan ruang data menjadi satu set subruang dengan pola acak. Data anomali biasanya terletak pada data yang jarang ada pada suatu dataset, algoritma ini dapat mengidentifikasi titik anomali dengan sangat cepat.

### **Sistem Pengendalian Internal Belanja Persediaan Konsumsi**

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 (tusi) organisasi, serta (4) kualitas SDM yang sesuai dengan tanggung jawabnya. Dari empat unsur pokok SPI tersebut, unsur pertama (pemisahan tugas) menjadi kelemahan utama dalam proses transaksi belanja barang persediaan. Dalam suatu satuan kerja, biasanya kewenangan penerimaan dan pengeluaran persediaan dipegang oleh satu orang atau satu bagian saja. Misalnya, dalam konteks pembelian persediaan *toner printer* yang harganya cukup mahal, pelaku bisa saja membuat kuitansi fiktif (*concealment*) dimana sebenarnya tidak ada *toner* yang dibeli. Kemudian pelaku meminta penggantian (*conversion*) kwitansi tersebut kepada Bendahara. Karena pelaku juga yang mencatat pengeluaran persediaan, dia kemudian membuat bon pengeluaran fiktif (*concealment*) untuk mengeluarkan persediaan fiktif yang telah dicatat pada pembukuan persediaan agar seolah-olah *toner* tadi sudah digunakan oleh unit tertentu. Persediaan masuk sama dengan persediaan keluar sehingga saldo persediaan *toner* kembali ke posisi semula.

## METODOLOGI PENELITIAN

### 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, kami dapatkan dari data realisasi belanja APBN pada Direktorat Jenderal Perbendaharaan.

### Metode Deteksi Anomali

Untuk mendeteksi anomali pada realisasi belanja akun tersebut, kami menggunakan modul *machine learning* dengan teknik *unsupervised anomaly detection* yang disebut *Isolation Forest*. *Isolation Forest* mengisolasi observasi dengan memilih *feature* secara acak kemudian secara acak memilih nilai pemisah antara nilai maksimum dan minimum dari *feature* yang terpilih tersebut. *Isolation Forest* merupakan salah satu modul dalam *library Machine Learning scikit-learn* yang berjalan dalam bahasa pemrograman *Python*. Untuk menjalankan *Python*, kami menggunakan *coding environment Google Colab*.

### Data Cleansing

Sebelum melakukan deteksi anomali menggunakan *Isolation Forest*, kami melakukan tahapan proses *data cleansing* terlebih dahulu, dengan rincian sebagai berikut:

1. mengedit nama satker pada data. data yang kami dapatkan dalam format csv (*comma separated values*). Ketika diupload ke dalam *Google Colab*, terjadi *error* pada beberapa baris data. Setelah ditelusuri ternyata ada nama satker yang mengandung "koma". Untuk mengatasi ini, kami melakukan replace nama satker dengan mengganti tanda koma menjadi karakter "spasi" agar tidak terbaca sebagai pemisah data pada csv.
2. *data filtering*. Data yang diterima mengandung seluruh realisasi per akun per bulan pada suatu satker.

Oleh karena itu, kami melakukan *filter* data hanya untuk akun 521811 saja. Selain itu data yang kami terima berisi seluruh Kementerian/Lembaga, Maka dari itu kami melakukan *filter* data dengan mengambil data Kepolisian Republik Indonesia menggunakan kode BA 060.

3. mengeluarkan data satker yang terdapat nilai realisasi minus. Terdapat beberapa realisasi pada bulan tertentu yang nilainya minus. Prediksi kami, Ini disebabkan oleh adanya koreksi terhadap realisasi belanja akun 521811 bulan sebelumnya yang melebihi nilai realisasinya. Karena koreksi yang dilakukan bisa saja untuk mengkoreksi realisasi bulan-bulan sebelumnya pada tahun bersangkutan sehingga realisasi bulan sebelumnya bisa dikatakan salah akun, Atas dasar ketidakpastian bulan mana yang dikoreksi dalam setahun, maka kami memutuskan untuk menghapus seluruh realisasi 521811 selama satu tahun pada satker yang terdapat realisasi bulanannya minus.
4. menjumlahkan data pagu belanja barang akun 52 per satker. Pada beberapa satker terdapat *row* atau baris data pagu belanja 52 lebih dari satu dengan kode lokasi yang berbeda. Berdasarkan validasi nilai pagu tersebut ke OM SPAN, penjumlahan pagu merupakan nilai untuk pagu satker tersebut. Maka, kami menjumlahkan pagu belanja barang akun 52 tersebut agar hanya terdapat data pagu belanja barang dalam satu *row* untuk setiap satker.
5. penggabungan *data set*. Data tahun 2019-2021 yang telah dibersihkan

diberi *tag* tambahan berupa kolom "Tahun" dan "Tanggal" untuk kepentingan pembuatan grafik,, kemudian dilakukan proses penggabungan data menjadi satu set data.

### **Data Transformation**

Nilai rupiah realisasi antar satker sangat variatif tergantung besar pagu satker bersangkutan. Semakin besar pagunya, semakin besar juga nilai realisasinya. Oleh karena itu, perlu dilakukan transformasi data realisasi menjadi persentase belanja akun 521811 terhadap total pagu akhir tahun belanja barang, selanjutnya disebut "Persentase Belanja Persediaan". Kami tidak memilih rasio dengan pagu akun 521811 secara langsung karena selama tahun berjalan, pagu per akun dimungkinkan untuk digeser (ditambah/dikurang) melalui mekanisme revisi POK sesuai dengan kebutuhan satker. Sementara itu, total pagu belanja barang biasanya jarang berubah karena membutuhkan proses revisi DIPA.

Penggunaan transformasi data menjadi nilai persentase ini merupakan upaya normalisasi data agar interval antar realisasi tidak jauh yang disebabkan perbedaan skala nominal belanja persediaan Satker Polri yang berbeda antar kantor dengan pagu besar dan pagu kecil. Transformasi data ini dilakukan untuk memudahkan dalam melihat pola persebaran data *outliers* atau data anomali per bulan nantinya.

### **Menentukan Nilai Contamination Parameter**

*Contamination parameter* yang kami tentukan dalam penelitian ini digunakan untuk mengajarkan model berapa proporsi anomali terhadap dataset yang diinginkan. Dapat dipahami juga bahwa parameter ini merupakan ambang batas

suatu data dapat dikatakan anomali atau bukan. Penentuan *contamination parameter* merupakan kunci utama dalam menentukan akurasi model. Karena dalam menentukannya diperlukan keahlian (*expertise*) dalam mengenali data yang sedang diolah. Metode pengujian beberapa nilai *contamination parameter* untuk menentukan nilai yang paling akurat untuk sebuah model *unsupervised machine learning* disebut "*Hyperparameter tuning*".

### Menghitung Anomaly Score

*Anomaly score* atau nilai yang menunjukkan sebuah *row data* merupakan suatu anomali atau bukan. Secara matematis *anomaly score* dapat dituliskan sebagai berikut.

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

Dengan keterangan  $h(x)$  adalah *path length* dan  $E(h(x))$  adalah rata-rata *path length* dari setiap *isolation tree*. *Path length* sendiri dapat dipahami sebagai jarak yang mengukur titik dimana data tidak dapat dibagi lagi dengan *root node* dari struktur algoritma pohon keputusan. Sedangkan  $c(n)$  adalah rata-rata dari  $h(x)$  yang diketahui nilai variabel  $n$ . (Eugenia Anolla, 2021)

*Anomaly score* sendiri memiliki rentang nilai dari -1 sampai dengan 1. *Anomaly score* yang semakin mendekati -1 dianggap anomali dan semakin mendekati nilai 1 dianggap normal.

## HASIL PENELITIAN DAN PEMBAHASAN

### Melatih Model *Isolation Forest* dengan dataset

Setelah dilakukan *data cleansing* dan *data transformation*, didapatkan set data

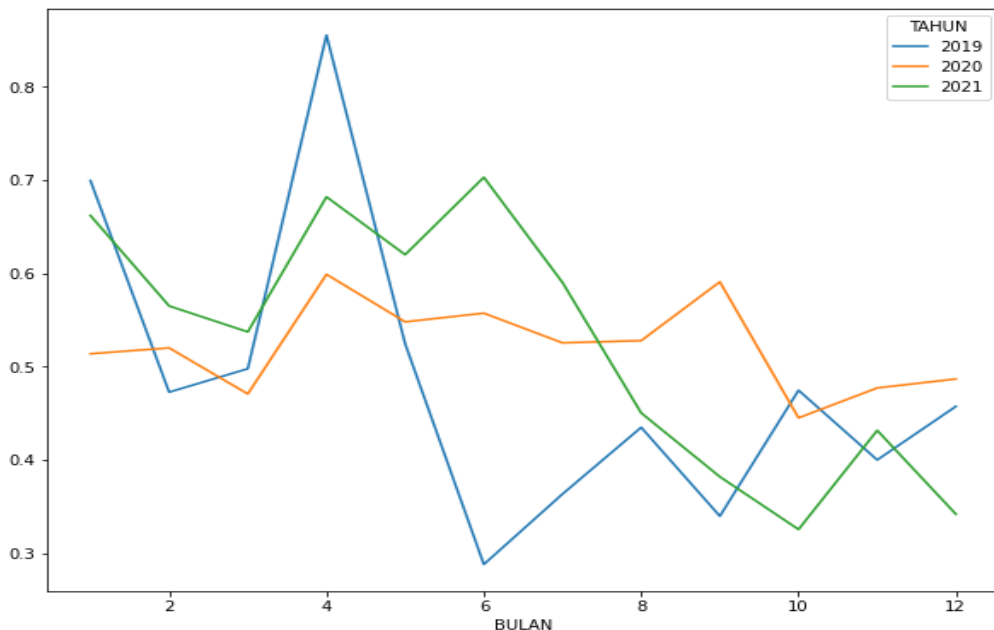
yang dapat ditampilkan sebagiannya seperti Tabel 1 di bawah. Dataset tersebut tidak bisa langsung digunakan untuk melatih model *machine learning*. Hal ini dikarenakan terdapat beberapa kolom data yang yang tidak diperlukan dalam melatih pendeteksian anomali.

**Tabel 1. Set Data setelah Proses *Data Cleansing* dan *Data Transformation***

	KANWIL_DJPB	TAHUN	BULAN	SATKER	REAL_PERSEDIAAN	PAGU_BARANG	PERSENTASE	TANGGAL	THSATKER
0	K01	2019	1	640075	95984400	35071963000	0.273678	2019-01-01	2019640075
1	K01	2019	2	640075	1400000	35071963000	0.003992	2019-02-01	2019640075
2	K01	2019	3	640075	2400000	35071963000	0.006843	2019-03-01	2019640075
3	K01	2019	4	640075	3552000	35071963000	0.010128	2019-04-01	2019640075
4	K01	2019	6	640075	3206500	35071963000	0.009143	2019-06-01	2019640075
5	K01	2019	7	640075	2707000	35071963000	0.007718	2019-07-01	2019640075
6	K01	2019	8	640075	1434000	35071963000	0.004089	2019-08-01	2019640075
7	K01	2019	9	640075	2007000	35071963000	0.005723	2019-09-01	2019640075
8	K01	2019	11	640075	6851000	35071963000	0.019534	2019-11-01	2019640075
9	K01	2019	1	640101	5800000	1140081000	0.508736	2019-01-01	2019640101



**Grafik 2. Rata-Rata Bulanan Realisasi Belanja Akun 521811 pada BA 060-Polri**



Dari hasil analisis kami terhadap data persentase belanja persediaan di Kepolisian RI yang tergambar pada Grafik 2 di atas, terlihat adanya *seasonality* (pola) sepanjang tahun 2017-2021. Pola yang muncul menunjukkan adanya pola antar bulan yang sama antar tahun. Pola ini dapat dilihat dari kesamaan persentase belanja persediaan di bulan yang sama antar tahun. Contohnya bulan April menunjukkan pola persentase belanja persediaan yang tertinggi kemudian turun di bulan Mei. Pola tersebut cenderung sama antar tahun 2019 sampai 2021.

Berdasarkan hasil analisis *seasonality* persentase belanja persediaan tersebut, kami menggunakan logika deteksi anomali dengan membandingkan data transaksi suatu bulan dengan bulan yang sama antar tahun 2019-2021. Sebagai contoh, suatu transaksi Satker A

di bulan Maret Tahun 2021 akan ditentukan sebagai anomali atau tidak dengan membandingkannya terhadap seluruh transaksi Satker Polri yang terjadi di bulan Maret sepanjang 2019-2021.

Maka dari itu, *features* atau variabel yang digunakan berjumlah dua yaitu "BULAN" dan "PERSENTASE". Kolom "BULAN" menampilkan data waktu bulan transaksi terjadi. Sedangkan "PERSENTASE" menampilkan data persentase belanja persediaan tiap transaksi.

Untuk melatih model *Isolation Forest*, maka hanya diambil dua kolom tersebut dari dataset. Dataset yang hanya berisi *features* yang diperlukan model ini diberi nama "data nasional". Proses selanjutnya adalah memasukkan dataset ini ke dalam fungsi *Isolation Forest* dengan menuliskan baris pemrograman sebagai berikut.

**Gambar 1. Kode Pelatihan *Isolation Forest***

```
✓ [5] datanasional=dfpolri_nominus[['BULAN','PERSENTASE']]
0s

✓ [6] model = IsolationForest(contamination=0.003, max_samples='auto', n_estimators=100)
1s      model.fit(datanasional)

IsolationForest(contamination=0.003)
```

Parameter dalam *Isolation Forest* yang digunakan antara lain *contamination*, *max\_samples*, dan *n\_estimators* seperti yang ditunjukkan pada Gambar 1 di atas. Parameter *contamination* akan dibahas di bagian selanjutnya karena terkait dengan penentuan nilainya mempengaruhi akurasi model mendeteksi anomali. Parameter kedua adalah *max\_samples* yang menentukan berapa banyak sampel yang boleh 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* disetel ke angka 100 yang merupakan angka yang standar digunakan (*default*).

#### **Menentukan *contamination parameter* yang tepat**

Setelah membangun model, kami menguji akurasi dengan menyetel

berbagai nilai *contamination parameter*. Proses ini dinamakan *Hyperparameter tuning* dengan tujuan mendapatkan kinerja akurasi model yang terbaik.

Kami menguji dataset pada nilai *contamination* parameter sebesar 0,005; 0,003; dan 0,001. Penentuan ketiga nilai tersebut berdasarkan wawasan yang kami miliki atas toleransi ketidakwajaran belanja persediaan. Hasil pengujian ketiga parameter kemudian dikonversi dari *anomaly score* menjadi *yes or no* dengan menggunakan kode yang ditampilkan pada Gambar 2 di bawah ini. Langkah selanjutnya, melakukan visualisasi data 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.

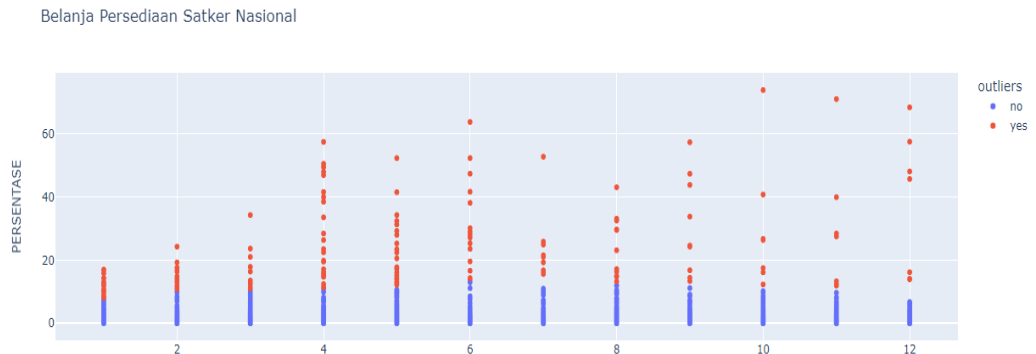
**Gambar 2. Kode Konversi *Anomaly Score* dan Visualisasi Data**

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

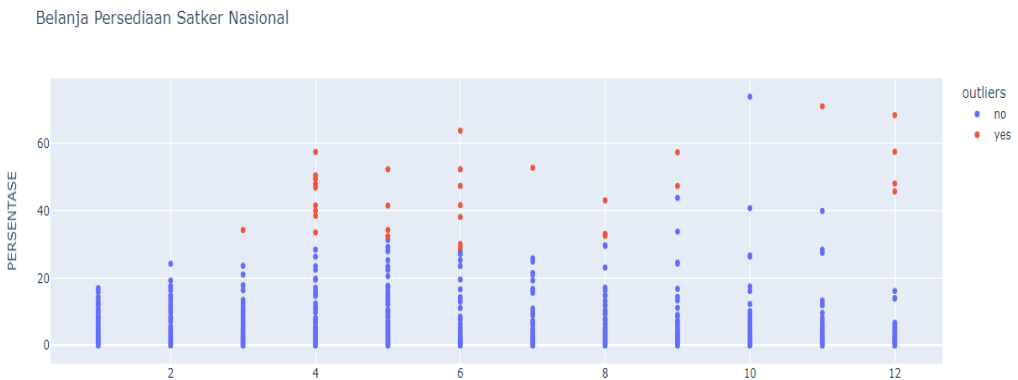
fig = px.scatter(dfpolri_nominus, x='BULAN', y='PERSENTASE', color='outliers', hover_data=['TAHUN','SATKER','PERSENTASE'], title='Belanja Persediaan Satker Nasional')

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

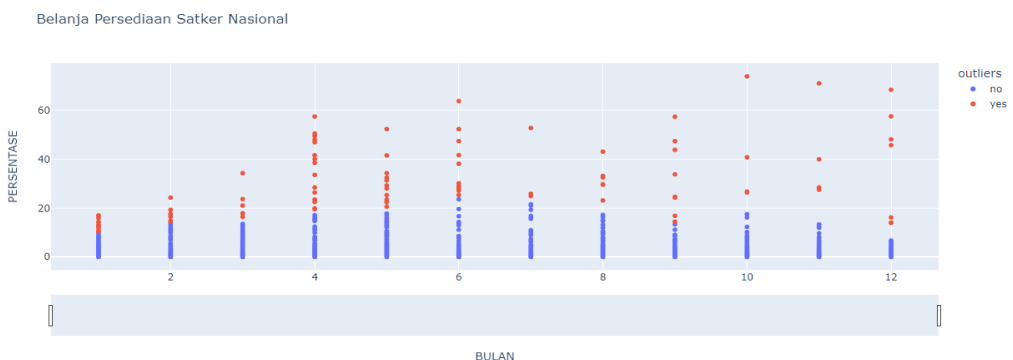
**Grafik 3. Hasil Prediksi Model Isolation Forest dengan parameter kontaminasi 0,005**



**Grafik 4. Hasil Prediksi Model Isolation Forest dengan parameter kontaminasi 0,001**



**Grafik 5. Hasil Prediksi Model Isolation Forest dengan parameter kontaminasi 0,003**



Berdasarkan hasil penelitian sebagaimana terlihat dari Grafik 3 di atas,, apabila menggunakan nilai *contamination parameter* 0,005 jumlah anomali data transaksi yang terdeteksi sebanyak 165 transaksi dari 32.822

transaksi yang terdeteksi sebagai anomali.

Sedangkan, setelah *dataset* diuji pada nilai *contamination parameter* 0,001 terlihat dari Grafik 4, jumlah anomali data transaksi yang terdeteksi sebanyak

33 transaksi dari 32.822 transaksi. Kemudian jika menguji model dengan nilai *contamination parameter* 0,003 atau 0,3% seperti yang diperlihatkan pada Grafik 5, sebanyak 99 transaksi terdeteksi sebagai anomali. Angka tersebut sama dengan proporsi hasil anomali terhadap *dataset* yaitu 99 transaksi anomali dari total 32822.

Berdasarkan hasil pengujian pada model *isolation forest* dengan *contamination parameter* yang telah dilakukan. Apabila menggunakan nilai 0,5%, terlalu banyak data yang dideteksi sebagai anomali. Hal ini akan berdampak pada *kesalahan deteksi* yaitu data-data yang wajar, dideteksi sebagai anomali. Kemudian apabila menggunakan nilai 0,1%, data yang terdeteksi anomali terlalu sedikit. Hal ini dapat menyebabkan beberapa transaksi yang merupakan anomali tidak terdeteksi pada hasil pengujian. Maka dari itu, kami menentukan nilai *contamination parameter* 0,003 atau 0,3% sebagai parameter yang tepat untuk menentukan batas toleransi data dideteksi sebagai anomali.

Tabel 2 di bawah ini 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%.

Dari sebagian data anomali yang disebutkan dapat disimpulkan model berhasil mendeteksi anomali baik anomali belanja terlalu tinggi maupun terlalu rendah. Langkah selanjutnya adalah menyimpan model *Isolation Forest* yang telah dilatih dan disetel (*tuning*) kedalam *file* menggunakan *library joblib*. Penyimpanan model *machine learning* ini dilakukan untuk mempermudah ketika suatu dataset baru hendak dideteksi anomalnya, cukup memanggil *file* tersebut.

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

KANWIL_DJPB	TAHUN	BULAN	SATKER	REAL_PERSEDIAAN	PAGU_BARANG	PERSENTASE
K02	2019	10	640338	43368003498	58647903000	7.394638e+01
K17	2019	11	645852	14182474490	19949499000	7.109188e+01
K02	2020	12	640338	53412887116	78015006000	6.846489e+01
K02	2021	6	640338	43276268580	67787997000	6.384061e+01
K31	2019	12	669282	9897169604	17194095000	5.756145e+01
...	...	...	...	...	...	...
K15	2020	9	644044	124000	113906388000	1.088613e-04
K03	2019	12	679389	18278	18572210000	9.841586e-05
K15	2019	12	644400	18000	42513071000	4.233992e-05
K15	2021	12	644470	800	20978058000	3.813508e-06
K23	2020	9	646307	500	111615217000	4.479676e-07

## KESIMPULAN DAN SARAN

Dari hasil percobaan, terbukti bahwa penggunaan model *Isolation Forest* dapat 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 melatih model mendeteksi anomali.

Saran yang diberikan penulis adalah perlu adanya tindak lanjut dari hasil data transaksi yang telah ditandai anomali oleh model. Data tersebut harus dikonfirmasi ke Satker bersangkutan terkait apakah transaksi anomali merupakan tindak kecurangan (*fraud*) atau transaksi yang normal seperti koreksi transaksi. Indikasi *fraud* ini penting untuk meminimalisir potensi penyalahgunaan aset.

Terlepas dari kebenaran indikasi *fraud*, keberhasilan model ini mendeteksi transaksi anomali sudah dapat dirasakan manfaatnya oleh Bendahara Umum Negara maupun institusi pengendalian internal. Model ini dapat dijadikan fitur tambahan yang berfungsi sebagai *early warning* pada aplikasi otorisasi belanja Satker. Selain itu hasil deteksi anomali dapat menjadi bahan untuk evaluasi atau pemeriksaan kepada Satker agar tidak terjadi lagi belanja barang persediaan yang tidak wajar.

## IMPLIKASI DAN KETERBATASAN

### Implikasi

Penelitian deteksi anomali dengan menggunakan *isolation forest* dapat diimplementasikan sebagai pencegahan deteksi *fraud*. Satuan kerja sebagai manajemen dapat menggunakan model ini untuk melaksanakan pengendalian

intern atas belanja barang persediaan di lingkungannya. Pemodelan ini bisa 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 transaksi per bulan, bukan data per transaksi SP2D. Dalam nilai transaksi per bulan ini selain transaksi pencairan SP2D, juga terdapat transaksi koreksi SP2D. Hal ini terlihat dari adanya nilai transaksi minus pada bulan dan satker tertentu, yang berarti total koreksi "kurang" pada bulan tersebut lebih besar daripada nilai realisasi belanjanya. Akan tetapi, kejadian koreksi "tambah" tidak bisa dideteksi karena nilainya sama-sama positif. Ada beberapa indikasi kecurigaan terkait koreksi tambah ini yaitu adanya nilai realisasi per bulan sampai dengan lebih dari 70% pada satker tertentu. Penelitian selanjutnya disarankan untuk menggunakan nilai transaksi per SP2D saja sehingga koreksi akun tidak termasuk ke dalam model.

## REFERENSI

- ACFE. (2000). *Fraud Examiners Manual*. Third Edition. Texas: Association of Certified Fraud Examination.
- Aisyiah, H.N., Ahzar, F.A. (2017). Ex Ante Audit Sebagai Upaya Pencegahan Fraud. *Akrual : Jurnal Akuntansi*, 9(1), 55.
- Amrizal. (2004). Pencegahan dan Pendeteksian Kecurangan oleh Internal Auditor, 2.
- Anello, Eugenia. (2021). Better Programming, *Anomaly Detection With Isolation Forest*. <https://betterprogramming.pub/anomaly-detection-with-isolation-forest-e41f1f55cc6>.
- Auffarth, Ben. (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. Jack., Robert J. Lindquist and Joseph T. Wells. (1993). *Investigation Audit*. 4th Edition, New Jersey: Prentice-Hall.
- Chaidir, Hatta., Putrada, A.G., Abdurrohman, Maman. (2021). Perbandingan Metode One Class SVM dan Isolation Forest Dalam Mendeteksi Anomali Dalam Activity Recognition Pada Rumah Dengan PIR Sensor. *e-Proceeding of Engineering*, 8(5), 11080.
- Khrisnan, Adithya. (2019, Agustus 24). *Anomaly Detection : Isolation Forest with Statistical Rules*. Diakses dari <https://towardsdatascience.com/isolation-forest-with-statistical-rules-4dd27dad2da9>.
- Nurhayati., Busman., 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.
- Omar, Julio., Nino, Palacio., Berzal, Fernando. (2019). Evaluation Metrics for Unsupervised Learning Algorithms. *arXiv:1905.05667(2)*, 9.
- Ounacer, S.E.B, Hicham Ait., Oubrahim, Younes., Ghoumari, M.Y., & Azzouazi, Mohamed. (2018). Using Isolation Forest in anomaly detection : the case of credit card transactions. *Periodicals of Engineering and Natural Sciences*, 6(2), 394 - 397.
- Republik Indonesia. (2018). Keputusan Direktur Jenderal Perbendaharaan Nomor 211 Tahun 2018 tentang Kodifikasi Bagan Akun Pada Bagan Akun Standar. Jakarta: Republik Indonesia
- Republik Indonesia. (2021). Keputusan Direktur Jenderal Perbendaharaan Nomor 331 Tahun 2018 tentang Kodifikasi

Bagan Akun Pada Bagan Akun Standar. Jakarta: Republik Indonesia

Sholihah, Siti, M. Nizarul Alim, Siti Musyarofah. (2016). Memotret Pola *Fraud* pada Rincian Objek Belanja yang Menjadi Temuan BPK. *Journal of Auditing, Finance, and Forensic Accounting Vol. 4 No. 2*. Bangkalan: Universitas Trunojoyo.  
<https://doi.org/10.21107/jaffa.v4i2.2783>

Wang, Yulei., Lee, L., Xue, B., Wang, L., Song, M., Yu, C., Li, S., & Chang, C. (2018) A Posteriori Hyperspectral Anomaly Detection for Unlabeled Classification. 3-7.

Young, Andre. (2020). Making Sense of Big Data, *Isolation Forest is the best Anomaly for Big Data Right Now*.  
<https://towardsdatascience.com/isolation-forest-is-the-best-anomaly-detection-algorithm-for-big-data-right-now-e1a18ec0f94f>.

