

## Deteksi Inefisiensi pada Klaim BPJS Kesehatan dengan menggunakan Machine Learning

Hanif Noer Rofiq<sup>1</sup>

<sup>1</sup>Kementerian Keuangan, [hanif.noer94@gmail.com](mailto:hanif.noer94@gmail.com)

---

**Abstract:** Inefficiency in claims submitted by health facilities to BPJS Kesehatan is a problem that needs attention. With the growing number of claims, traditional method such as manual verification is unable to handle enormous amounts of data swiftly. Simultaneously, there is a demand to accelerate the settlement of claims with a limited number of verifiers. One way that can be adopted is to use Machine Learning to detect potential inefficient transactions rapidly. This study compares several Machine Learning algorithms: Random Forest, Gradient Boosting Classifier, Decision Tree, Support Vector Machine, Naive Bayes, CatBoost, and XGBoost. In addition, oversampling and under sampling methods are also used because the dataset is imbalanced. The best results were obtained using the Random Forest + Tomek Links model, which produced an F1 score of 19,53, with the five most influential variables: location of the health facility, participant's age, first-level health facilities diagnosis, participant's primary diagnosis, and health facility types.

**Keywords:** Machine Learning; Random Forest; Tomek Links; Inefisiensi; BPJS Kesehatan

**Abstrak:** Inefisiensi merupakan salah satu masalah yang perlu diperhatikan pada klaim yang diajukan oleh fasilitas kesehatan ke BPJS Kesehatan. Dengan jumlah klaim yang semakin meningkat, cara tradisional seperti verifikasi manual tidak dapat menangani jumlah data yang sangat besar dalam waktu singkat, sementara terdapat tuntutan untuk mempercepat penyelesaian klaim dengan jumlah verifikator yang terbatas. Salah satu cara yang dapat diadopsi adalah dengan menggunakan *Machine Learning* untuk mendeteksi transaksi yang berpotensi inefisien dengan cepat. Penelitian ini membandingkan beberapa algoritma *Machine Learning* berupa *Random Forest*, *Gradient Boosting Classifier*, *Decision Tree*, *Support Vector Machine*, *Naive Bayes*, *CatBoost*, dan *XGBoost*. Selain itu juga digunakan metode *oversampling* dan *undersampling* karena dataset yang digunakan tidak seimbang. Hasil terbaik didapat dengan menggunakan model *Random Forest + Tomek Links* yang menghasilkan *F1 Score* sebesar 19,53, dengan lima variabel yang paling berpengaruh yaitu lokasi fasilitas kesehatan, usia peserta, diagnosa dari fasilitas kesehatan tingkat pertama, diagnosa primer peserta, dan tipe fasilitas kesehatan.

**Kata kunci:** Machine Learning; Random Forest; Tomek Links; Inefisiensi; BPJS Kesehatan

## PENDAHULUAN

Kesehatan merupakan salah satu sektor penting yang harus dapat diakses oleh seluruh masyarakat. Untuk mewujudkan hal tersebut, terdapat sistem *Universal Health Coverage* yang mengamankan bahwa seluruh masyarakat harus bisa mendapatkan akses pelayanan kesehatan tanpa perlu mengalami kesulitan dalam pembayaran. Di Indonesia, hal ini terwujud melalui penyelenggaraan Program Jaminan Kesehatan Nasional (JKN) oleh Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan sesuai dengan Undang-Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN). Namun demikian, terdapat masalah yang dihadapi oleh BPJS Kesehatan. Dalam periode 2014-2018, terjadi defisit Dana Jaminan Sosial (Annisa dkk, 2020). Dalam penelitiannya tersebut, Annisa dkk, 2020 menyoroti aspek kecurangan dalam pelaksanaan Program Jaminan Kesehatan Nasional. Di tingkat global, inefisiensi juga ditemukan dalam studi yang dilakukan oleh Shrank dan Rogstad (2019) yang menyatakan bahwa sekitar 20-25% biaya pelayanan kesehatan di Amerika merupakan pemborosan. Chandra dan Staiger (2020) juga menemukan banyak rumah sakit yang mengalami inefisiensi dengan mayoritas rumah sakit memberikan pelayanan kesehatan secara berlebihan dikarenakan pemahaman yang keliru mengenai kelebihan komparatif yang dimiliki oleh masing-masing rumah sakit. Inefisiensi juga pernah ditemukan dalam penelitian Eni dkk (2020) terhadap industri rumah sakit di Indonesia.

Di Indonesia, pencegahan kecurangan telah diatur dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 16 Tahun 2019 tentang Pencegahan dan Penanganan Kecurangan (*Fraud*) serta Pengenaan Sanksi Administrasi Terhadap Kecurangan (*Fraud*) dalam Pelaksanaan Program Jaminan Kesehatan Nasional. Peraturan tersebut mengidentifikasi area-area rawan sesuai dengan alur kerja BPJS Kesehatan, salah satunya adalah inefisiensi yang disebabkan oleh klaim palsu yang diajukan oleh peserta, ataupun petugas BPJS Kesehatan yang bekerja sama dengan fasilitas kesehatan (Annisa dkk, 2020). Beberapa penelitian telah meneliti mengenai potensi *fraud* dan inefisiensi klaim BPJS Kesehatan, di antaranya oleh Pramono (2018) yang mengkaji mengenai potensi *fraud* atau inefisiensi klaim JKN KIS dan pelaksanaan kredensialing pada proses mitra BPJS Kesehatan Kantor Cabang Purwokerto. Tantangan yang dihadapi saat ini adalah mengidentifikasi upaya yang bisa dilakukan untuk dapat memprediksi terjadinya pemborosan atau inefisiensi pada klaim yang diajukan oleh fasilitas kesehatan (faskes) ke BPJS Kesehatan tersebut.

Laporan Pengelolaan Program dan Keuangan BPJS Kesehatan Tahun 2021 menunjukkan kenaikan cakupan kepesertaan JKN setiap tahun. Di tahun 2020, jumlah cakupan kepesertaan JKN adalah 222.461.906 jiwa, kemudian di tahun 2021 naik menjadi 235.719.262 jiwa atau sebesar 86,07% dari jumlah penduduk Indonesia per 31 Desember 2021. Dilansir dari situs web BPJS Kesehatan, per 30 Juni 2022 peserta JKN telah mencapai 241,7 juta jiwa, sedangkan sampai dengan

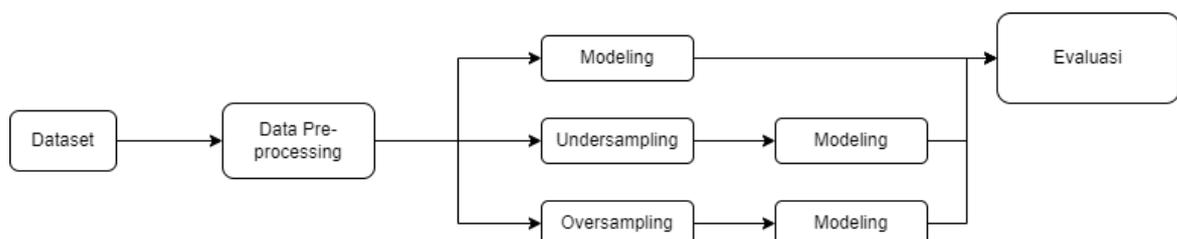
1 Maret 2023 peserta JKN telah mencapai 252.235.864 jiwa. Sejalan dengan penambahan jumlah peserta, klaim jaminan kesehatan kepada BPJS Kesehatan akan meningkat juga. Untuk menangani potensi naiknya klaim jaminan, BPJS Kesehatan melakukan penambahan pegawai verifikator. Tenaga manusia yang bertambah bisa mempercepat proses klaim, namun belum sebanding dengan potensi jumlah klaim yang masuk setiap harinya (Putri dkk, 2019). Selain itu, penggunaan sumber daya manusia juga cenderung memakan waktu yang relatif lama karena proses klaim dilakukan melalui serangkaian proses panjang (Putra dan Kusumo, 2016). Di sisi lain, klaim yang diajukan oleh fasilitas kesehatan kepada BPJS Kesehatan termasuk dalam area rawan terjadinya kecurangan akibat adanya inefisiensi yang mungkin terjadi. Dengan jumlah klaim yang semakin meningkat, cara tradisional seperti verifikasi manual oleh verifikator tidak dapat menangani jumlah data yang sangat besar dalam waktu singkat, sementara terdapat tuntutan untuk mempercepat penyelesaian klaim dengan jumlah verifikator yang terbatas.

Dengan melakukan klasifikasi menggunakan *machine learning*, prediksi potensi inefisiensi pembayaran klaim dapat dilakukan dengan cepat sebanyak apapun jumlah transaksi yang ada. Contoh penggunaan *machine learning* untuk deteksi kecurangan pada penelitian yang ditulis oleh Lu dkk (2020) yang menggunakan klasifikasi untuk mengidentifikasi *fraud* pada Telecom (Lu dkk, 2020). Klasifikasi termasuk dalam *Supervised Machine Learning* dan digunakan untuk melakukan prediksi karena bertujuan untuk membuat hasil klasifikasi berdasarkan hasil tertentu (Jiang dkk, 2020). Pengklasifikasian menggunakan *Machine Learning* tersebut kemudian memunculkan sebuah pertanyaan mengenai model algoritma yang memberikan performa paling baik untuk memprediksi transaksi yang efisien dan yang inefisien. Penelitian ini mencoba meneliti hal tersebut.

## METODE

Penelitian ini menggunakan pengklasifikasian otomatis dengan *Machine Learning*, subkategori *supervised learning* yang memerlukan pelabelan *dataset* untuk melatih algoritma dengan tujuan memprediksi *outcome* secara akurat. *Dataset* yang digunakan berasal dari data BPJS *Healthkathon 2022* Kategori *Machine Learning*, di mana pada *dataset* tersebut telah terdapat label ‘Potensi inefisiensi’.

Langkah penelitian dilakukan sesuai dengan Gambar 1, yang terdiri dari *data preprocessing*, *modelling*, dan *data evaluation*.



Gambar 1. Langkah Penelitian

### **Data Preprocessing**

*Data preprocessing* dilakukan sebelum *modelling* dijalankan. Secara umum, tujuan *data preprocessing* adalah untuk menghilangkan variabel atau efek yang tidak diinginkan sehingga pemodelan bisa dijalankan secara efisien (Mishra, 2020). *Data preprocessing* yang dilaksanakan antara lain menggabungkan tiga *dataset* menjadi satu, melakukan *data cleansing*, serta *feature engineering*.

### **Modelling**

Penelitian ini dilakukan dengan membandingkan algoritma *Random Forest*, *Gradient Boosting Classifier*, *Decision Tree*, *Support Vector Machine*, *Naive Bayes*, *CatBoost*, dan *XGBoost* untuk melihat algoritma yang menghasilkan performa terbaik berdasarkan matriks evaluasi. Selain itu karena *dataset* yang digunakan bersifat *imbalance*, algoritma tersebut juga dikombinasikan dengan metode *oversampling* dan *under sampling*. Metode *oversampling* yang digunakan adalah *Adaptive Synthetic Sampling* (ADASYN) dan *Synthetic Minority Over-sampling Technique* (SMOTE). Penggunaan metode ini didasari oleh penelitian yang dilakukan oleh Lu dkk (2020) untuk memprediksi *fraud* terhadap *dataset imbalance* menggunakan algoritma ADASYN. Sedangkan penggunaan SMOTE didasari oleh penelitian Wang dkk (2022) yang memanfaatkan SMOTE untuk menyeimbangkan *dataset imbalance*. Metode *under sampling* yang digunakan adalah *NearMiss* dan *Tomek Links*. Penggunaan *NearMiss* didasari oleh penelitian Mqadi dkk (2021) yang menerapkan metode *NearMiss* dalam menyeimbangkan *dataset imbalance* kartu kredit untuk mendeteksi *fraud* atau pola anomali dalam transaksi finansial. Sedangkan *Tomek Links* digunakan atas dasar penelitian Sawangarrearak dan Thanathamthee (2020) untuk *dataset imbalance* prediksi depresi mahasiswa/i universitas. Pemodelan dalam Penelitian ini menggunakan algoritma dasar tanpa dilakukan *hyperparameter tuning*.

### **Evaluasi**

Model yang dihasilkan oleh masing-masing algoritma *machine learning* kemudian dievaluasi dengan menggunakan matriks evaluasi: *Precision*, *Recall*, *F1 Score*, dan Akurasi.

#### **Precision**

*Precision* merupakan proporsi antara benar positif (*True Positive*) dengan banyaknya data yang diprediksi positif, secara matematis dapat dituliskan sebagai berikut:

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)}$$

#### **Recall**

*Recall* merupakan proporsi benar positif (*True Positive*) yang dapat diprediksi dengan benar, secara matematis dapat dituliskan sebagai berikut:

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)}$$

### F1 Score

*F1 Score* adalah *harmonic mean* dari *precision* dan *recall* yang secara matematis dapat dituliskan sebagai berikut:

$$F1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision}$$

### Accuracy

Akurasi adalah proporsi jumlah data yang dapat diprediksi dengan benar.

$$Akurasi = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{True\ Positive\ (TP) + True\ Negative\ (TN) + False\ Positive\ (FP) + False\ Negative\ (FN)}$$

## HASIL

Setelah model dijalankan tanpa; dan dikombinasikan dengan metode *oversampling* atau *under sampling*, dihasilkan bahwa model *Random Forest + Tomek Links* menghasilkan *F1 Score* yang paling baik, yakni sebesar 19,53.

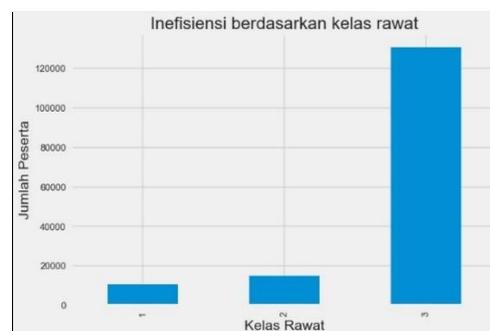
### Exploratory Data Analysis (EDA)

Sebelum melakukan *data preprocessing*, dilakukan EDA untuk melihat gambaran dataset yang dimiliki. Penelitian ini menggunakan tiga *dataset*, yakni data kunjungan peserta JKN ke fasilitas kesehatan rujukan tingkat lanjut dengan jumlah 11.401.882; data diagnosa penyakit peserta dengan jumlah 17.308.536; dan data prosedur yang didapatkan peserta JKN sejumlah 12.202.871.

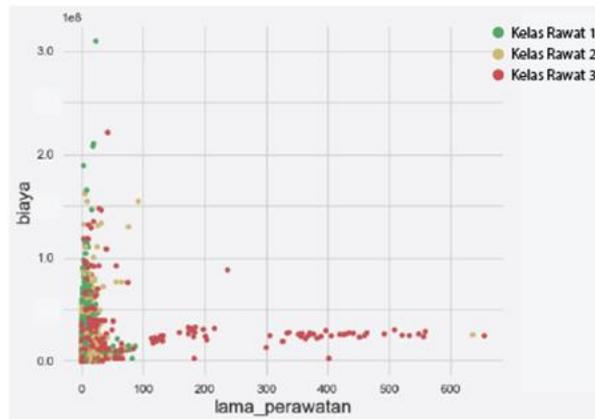
EDA menampilkan grafik yang berisi informasi umum *dataset*. Gambar 3 menunjukkan bahwa dari dataset yang dimiliki, jumlah inefisiensi lebih tinggi pada jenis pelayanan rawat jalan. Sedangkan berdasarkan kelas rawat, jumlah inefisiensi tertinggi terdapat pada Kelas Rawat 3 seperti pada Gambar 4. Sebaran inefisiensi juga bisa dilihat dari segi biaya, lama perawatan, dan kelas rawat. Hasil EDA terhadap ketiga segi tersebut dapat dilihat pada Gambar 5. Inefisiensi tersebar di lama perawatan pasien, namun lebih dominan di kasus dengan durasi perawatan yang rendah, khususnya pada kelas rawat 3.



Gambar 3. Jumlah Inefisiensi pada Jenis Pelayanan



Gambar 4. Jumlah Inefisiensi pada Kelas Rawat Inap



Gambar 5. Inefisiensi dilihat dari biaya, lama perawatan, dan kelas rawat

### Data Preprocessing

Penelitian ini menggunakan 11.402.036 data, dengan data berlabel “1” yang berarti memiliki potensi inefisiensi pada klaim sebanyak 156.899 atau 1,4% dari total data yang ada. *Dataset* yang digunakan sesuai dengan Tabel 1, yang kemudian digabungkan sehingga menghasilkan satu *dataset* sesuai Tabel 2.

Tabel 1 *Dataset* Awal

Nomor	<i>Dataset</i>
1	Data kunjungan peserta JKN ke fasilitas kesehatan rujukan tingkat lanjut
2	Data diagnosa penyakit peserta, di mana dalam satu kunjungan peserta bisa memiliki lebih dari satu diagnosa. Terdiri dari satu diagnosa primer dan atau satu atau lebih diagnosa sekunder.
3	Data prosedur yang didapatkan peserta JKN. Dalam setiap kunjungan, peserta bisa mendapatkan satu atau lebih prosedur/tindakan medis.

Tabel 2 Data kunjungan peserta JKN ke fasilitas kesehatan rujukan tingkat lanjut

No	Nama	<i>dtypes</i>	<i>Missing %</i>	<i>Uniques</i>	<i>Entropy</i>	Keterangan
1	Id	int64	0.0	11401868	23.44	Id kunjungan
2	id_peserta	int64	0.0	8527919	22.74	Id peserta
3	dati2	int64	0.0	489	7.78	Lokasi Fasilitas Kesehatan
4	typefaskes	object	0.0	26	3.13	Tipe Fasilitas Kesehatan
5	usia	int64	0.0	111	6.21	Usia Peserta
6	jenkel	object	0.0	2	1.0	Jenis Kelamin Peserta
7	pisat	float64	0.0	5	1.83	Hubungan Kepesertaan
8	tgldatang	datetime	0.0	1096	9.98	Tanggal Kedatangan
9	tglpulang	datetime	0.0	1247	9.99	Tanggal Kepulangan

10	jenispel	int64	0.0	2	0.94	Jenis Pelayanan
11	politujuan	object	35.4	261	4.03	Poli Tujuan
12	diagfktp	object	0.0	17012	9.65	Diagnosa dari FKTP
13	biaya	float64	0.5	57854	10.56	Biaya
14	jenispulang	float64	0.0	5	0.33	Kondisi Peserta saat Pulang
15	Cbg	object	0.0	1034	6.15	Kode <i>Case Based Group</i>
16	kelasrawat	int64	0.0	3	0.75	Kelas Perawatan
17	kdsa	object	1.7	36	0.02	Kode <i>Special Sub-Acute Group</i>
18	kdsp	object	1.7	13	0.07	Kode <i>Special Procedures</i>
19	kdsr	object	1.7	6	0.01	Kode <i>Special Prothesis</i>
20	kdsi	object	1.7	4	0.01	Kode <i>Special Investigations</i>
21	kdsd	object	1.7	12	0.14	Kode <i>Special Drugs</i>
22	label	int64	0.0	2	0.1	Label Potensi Inefisiensi

Tabel 3 Data diagnosa penyakit peserta

No	Nama	dtypes	Missing %	Uniques	Entropy	Keterangan
1	Id	int64	0	11403151	23.24	Id kunjungan
2	diag	object	0	13273	9.43	Kode Diagnosa FKRTL
3	levelid	int64	0	2	0.93	Jenis Diagnosa (Primer, Sekunder)

Tabel 4 Data prosedur yang didapatkan peserta JKN

No	Nama	dtypes	Missing %	Uniques	Entropy	Keterangan
1	id	int64	0	6454758	22.28	Id kunjungan
2	proc	object	0	3695	6.88	Kode Prosedur

### Data Cleansing

*Data cleansing* dilakukan dengan cara menghapus data duplikat dan menghapus kolom biaya untuk menghindari bias prediksi karena biaya yang ada pada dataset adalah biaya yang sudah dibayarkan, sehingga biaya 0 (nol) dan *null* berarti bahwa klaim tersebut tidak layak sehingga tidak dibayarkan.. Kolom *diagfktp* dengan nilai kosong, kolom *jmlh\_diag\_sec* dengan nilai “0”, dan

sisanya diisi dengan nilai yang paling sering muncul. Kemudian membuang kolom *kdsa*, *kdsp*, *kdsr*, *kdsi*, dan *kdsd* karena memiliki nilai *entropy* yang rendah serta politujuan dan jumlah prosedur karena memiliki *missing value* yang sangat tinggi.

### Feature Engineering

*Feature engineering* dilakukan terhadap tiga dataset awal dengan cara membuat kolom baru berupa lama perawatan yang didapat dari tanggal kedatangan dikurangi tanggal kepulangan peserta; serta memisahkan kolom “*cbg*” menjadi 4 di mana *cbg-1* menggambarkan kode CMG (*Casemix Main Groups*), *cbg-2* menggambarkan tipe kelompok kasus (*Case Groups*), *cbg-3* menggambarkan spesifikasi kelompok kasus dan *cbg-4* menggambarkan tingkat keparahan kelompok kasus (Permenkes No. 76 tahun 2016). Kemudian dilakukan proses transformasi label kata menjadi bentuk numerik dengan menggunakan *label encoder* dan memisahkan data menjadi 80% untuk data *train* dan 20% untuk data *test* (*The Pareto principle*). Hasil dari *feature engineering* tersebut dapat dilihat pada Tabel 5 berikut.

Tabel 5 *Dataset* Akhir Beserta Karakteristiknya

No	Name	dtypes	Missing %	Uniques	Keterangan
1	<i>dati2</i>	<i>int64</i>	0	489	Lokasi Fasilitas Kesehatan
2	<i>typefaskes</i>	<i>object</i>	0	26	Tipe Fasilitas Kesehatan
3	<i>Usia</i>	<i>int64</i>	0	111	Usia Peserta
4	<i>Jenkel</i>	<i>object</i>	0	2	Jenis Kelamin Peserta
5	<i>Pisat</i>	<i>float64</i>	0	5	Hubungan Kepersetaan
6	<i>jenispel</i>	<i>int64</i>	0	2	Jenis Pelayanan
7	<i>diagfktp</i>	<i>object</i>	0	17013	Diagnosa dari FKTP
8	<i>jenispulang</i>	<i>float64</i>	0	5	Kondisi Peserta Saat Pulang
9	<i>kelasrawat</i>	<i>int64</i>	0	3	Kelas Perawatan
10	<i>Diag</i>	<i>object</i>	0	10946	Diagnosa Primer Peserta
11	<i>jumlah_diagnosa_sekunder</i>	<i>float64</i>	0	23	Jumlah Diagnosa Sekunder Peserta
12	<i>cbg1</i>	<i>object</i>	0	23	<i>CMG (Casemix Main Groups)</i> ,
13	<i>cbg2</i>	<i>object</i>	0	9	Tipe Kelompok Kasus ( <i>Case Groups</i> )
14	<i>cbg3</i>	<i>object</i>	0	61	Spesifikasi Kelompok Kasus
15	<i>cbg4</i>	<i>object</i>	0	4	Tingkat Keparahan KelompokKasus
16	<i>lama_perawatan</i>	<i>int64</i>	0	508	Lama Perawatan
17	<i>label</i>	<i>int64</i>	0	2	Label Potensi Inefisiensi

**Data Evaluation**

Setelah model dijalankan, dan dilakukan pengujian dengan empat matriks evaluasi tanpa menggunakan *oversampling* dan *undersampling*, didapatkan hasil sesuai Tabel 3 dengan hasil tertinggi diberikan *highlight* warna hijau.

Tabel 3. Hasil Evaluasi terhadap Model tanpa Menggunakan *Oversampling* dan *Undersampling*

Algoritma	Precision	Recall	F1 Score	Accuracy
Decision Tree	16.2169	18.8068	17.4161	97.5459
Random Forest	54.2362	10.3643	17.4029	98.6463
CatBoost	70.0943	5.214	9.706	98.6652
XGboost	70.3304	3.5281	6.7191	98.6521
Naive Bayes	1.4185	21.8631	2.6641	78.018
Gradient Boosting Classifier	25	0.0032	0.0064	98.624
Support Vector Classifier	0	0	0	98.624

Setelah itu, dengan menggunakan *oversampling* berupa SMOTE dan ADASYN didapatkan hasil sebagaimana Tabel 4 berikut:

Tabel 4. Hasil Evaluasi terhadap Model Menggunakan *Oversampling*

Algoritma	SMOTE				ADASYN			
	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy
Random Forest	20.157	17.2706	18.6025	97.9204	19.8437	17.2356	18.4479	97.9032
CatBoost	11.2585	29.7734	16.3387	95.8046	11.1496	28.897	16.0908	95.8531
Decision Tree	10.3301	21.5126	13.9578	96.3506	10.2442	21.299	13.8344	96.3494
XGboost	8.2168	25.4964	12.4282	95.0561	7.8869	22.9499	11.7395	95.2517
Gradient Boosting Classifier	3.86	16.9232	6.2862	93.0572	4.215	10.5045	6.016	95.484
Naive Bayes	1.5811	40.8229	3.0443	64.2214	1.5248	47.0121	2.9538	57.4943
Support Vector Classifier	1.3692	56.551	2.6736	43.3485	1.3788	46.2186	2.6777	53.7721

Sedangkan dengan menggunakan *undersampling* berupa *Tomek Links* dan *NearMiss*, didapatkan hasil sebagaimana Tabel 5. berikut:

Tabel 5. Hasil Evaluasi terhadap Model Menggunakan *Undersampling*

Algoritma	<i>Tomek Links</i>				<i>NearMiss</i>			
	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy
Random Forest	54.1437	11.91	19.5251	98.6491	1.9852	68.0849	3.8579	53.3067
Decision Tree	16.9673	20.413	18.5314	97.5304	1.693	71.699	3.3079	42.3257
CatBoost	68.8356	6.406	11.7211	98.6723	1.9746	67.677	3.8373	53.3277

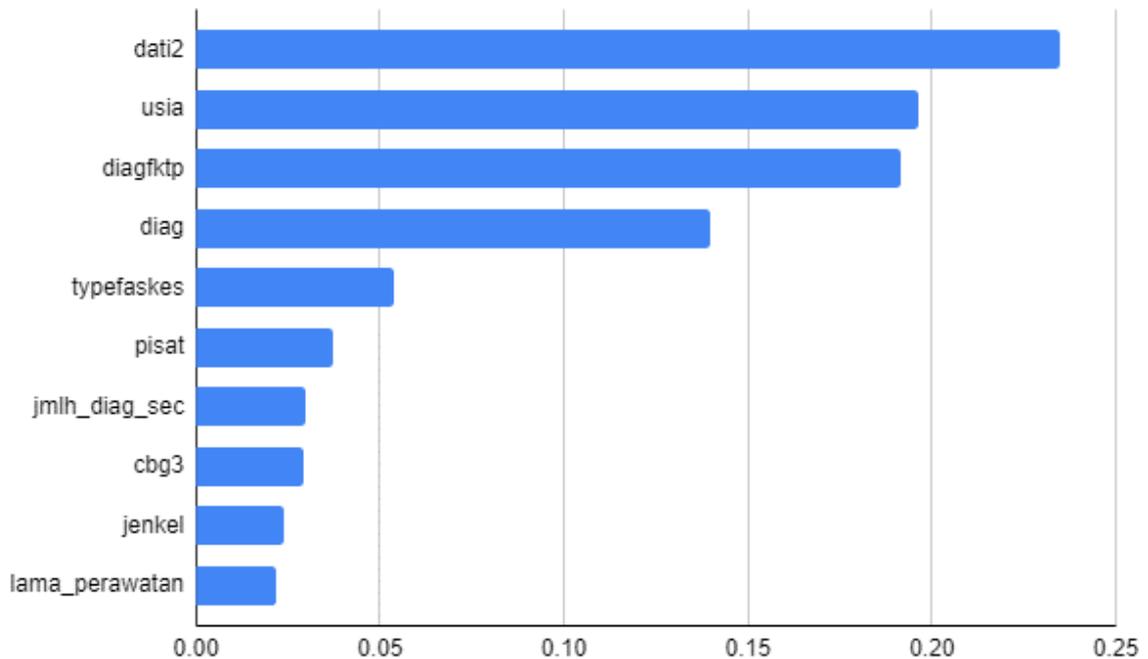
<i>XGboost</i>	70.0319	3.493	6.6541	98.6515	1.9783	66.9471	3.843	53.9024	
<i>Naive Bayes</i>	1.4195	21.8759	2.666	78.0214	1.4703	43.3343	2.8442	59.2635	
<i>Gradient Classifier</i>	<i>Boosting</i>	20	0.0032	0.0064	98.6239	1.7878	61.0798	3.4738	53.2945
<i>Support Vector Classifier</i>	0	0	0	98.624	1.3892	98.4192	2.7397	3.8514	

Tabel 4 menunjukkan hasil evaluasi Model Oversampling memberikan nilai Precision, F1 Score, dan Accuracy tertinggi pada model menggunakan SMOTE daripada ADASYN. Sedangkan hasil evaluasi Model Undersampling sesuai Tabel 5 menunjukkan Tomek Links lebih unggul daripada NearMiss di Precision, F1 Score, dan Accuracy. Merujuk pada metrik evaluasi Precision, F1 Score, dan Accuracy pada penelitian ini, model menggunakan Tomek Links menunjukkan nilai yang lebih bagus daripada SMOTE.

Penelitian dari Hairani dkk, 2023 menunjukkan bahwa penggunaan SMOTE-Tomek Links menghasilkan accuracy, sensitivity, precision, dan F1-score lebih tinggi daripada hanya menggunakan SMOTE saja. Hal ini salah satunya karena salah satu kelemahan SMOTE adalah adanya generalisasi berlebihan dan munculnya *noise data* (Hairani, 2020 dalam Hairani, 2023). Kamaladevi (2021) menyebutkan bahwa biasanya noise data menurunkan akurasi dari metode klasifikasi. Sedangkan penggunaan Tomek Links dapat menghilangkan *noise sample* dari kelas mayoritas tanpa menghilangkan informasi penting data (Kamaladevi, 2021)

Matriks evaluasi Akurasi digunakan sebagai acuan untuk mengukur performa model ketika distribusi label *dataset* mirip. Sedangkan terhadap *data imbalance*, menggunakan matriks evaluasi bisa membuat akurasi menjadi sangat bias (Ibrahim, 2018). Penelitian ini yang menggunakan data yang tidak seimbang (*imbalanced dataset*), di mana data potensi inefisiensi hanya sebesar 1,4% dari total data yang ada. Ketika menghadapi data imbalance, muncul *trade off* antara precision dan recall yang mengharuskan pemilihan antara mana yang lebih penting dari precision atau recall (Janet dkk, 2022; Huilgol Purva, 2023). Penggunaan metrik evaluasi Precision dan Recall pada data imbalance dipengaruhi oleh toleransi atas *False Positive* atau *False Negative*. Situasi yang dihadapi pada kasus ini menunjukkan pentingnya titik temu antara precision dan recall, di mana klaim yang salah dikategorikan menjadi inefisien (*False Negative*) sama pentingnya dengan klaim yang salah dikategorikan sebagai efisien (*False Positive*). Ketika berada pada situasi ini, digunakan F1 yang merupakan rata-rata harmonis dari precision dan recall. Dengan karakteristiknya tersebut, F1 Score menjadi metrik evaluasi yang sering digunakan pada data imbalance (He dan Garcia, 2009 pada Zhu dkk, 2018). Sehingga Penulis menggunakan metrik evaluasi berupa *F1 Score* yang bisa memberikan penilaian yang lebih baik atas performa model.

Dari hasil evaluasi model yang digunakan, didapat bahwa model *Random Forest + Tomek Links* menghasilkan *F1 Score* yang paling baik. Dari model tersebut kemudian didapatkan variabel-variabel yang paling berkontribusi terhadap prediksi model, sebagaimana Gambar 6 di bawah ini.



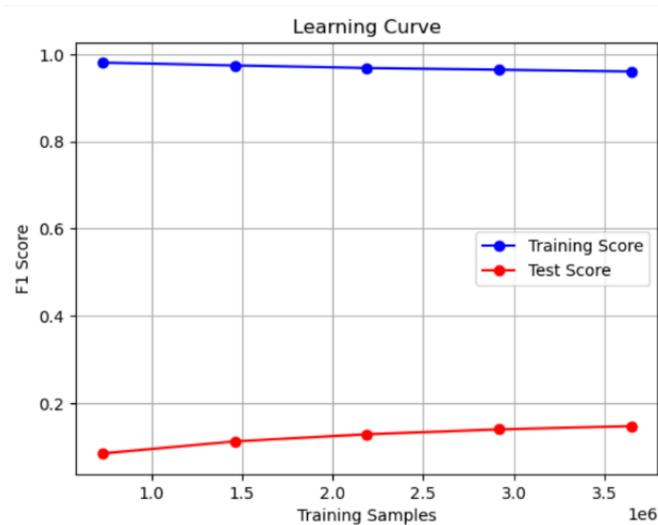
Gambar 2. Variabel paling berkontribusi pada model terpilih

## PEMBAHASAN

Hasil evaluasi dapat dilihat pada Tabel 3, Tabel 4, dan Tabel 5. Berdasarkan Tabel 3, hasil evaluasi terhadap model tanpa menggunakan *oversampling* dan *undersampling* menghasilkan nilai *F1 Score* tertinggi pada algoritma *Decision Tree*, dengan nilai 17,42. Sedangkan berdasarkan Tabel 4, hasil evaluasi terhadap metode *oversampling* menunjukkan bahwa algoritma *Random Forest* menghasilkan nilai evaluasi terbaik pada seluruh matriks evaluasi, kecuali pada *Recall*. Nilai *F1 Score* tertinggi pada metode *oversampling* adalah sebesar 18,60 yang dihasilkan oleh algoritma *Random Forest + SMOTE*. Sedangkan hasil evaluasi terhadap model *undersampling* dapat dilihat pada Tabel 5 yang menghasilkan nilai *F1 score* tertinggi pada algoritma *Random Forest + Tomek Links* dengan nilai 19,53 yang mana penggunaan *Random Forest + Tomek Links* memberikan peningkatan yang cukup signifikan terhadap nilai *F1 score*.

Meskipun *Random Forest + Tomek Links* memberikan hasil *F1 Score* tertinggi dari berbagai model yang telah dilakukan hasil tersebut masih kurang bagus. Jika dibandingkan dengan *F1 score* *Random Forest + Tomek Links* dari *training* yang mencapai 93,48 dapat dikatakan bahwa terjadi *overfitting* pada model tersebut. Untuk mengatasi *overfitting*, Penulis mengulang model *Random*

Forest + Tomek Links dengan menambahkan penggunaan *cross validation* berupa Stratified K-Fold sebanyak lima lipatan. *Cross Validation* merupakan metode *resampling* data yang bertujuan untuk mencegah *overfitting* (Berrar, 2019; Gemci dkk, 2022). *Stratified K-Fold Cross Validation* merupakan salah satu variasi dari *Cross Validation* yang memberikan hasil akurasi estimasi yang lebih andal dibandingkan metode *K-Fold Cross Validation* yang biasa (Zeng dan Martinez, 2000). Dengan menggunakan *5-fold Stratified Cross Validation*, perbandingan rata-rata *F1 Score* antara *training* dan *test* dapat dilihat pada gambar 3 berikut.



Gambar 3. Learning Curve dari Random Forest + Tomek Links dengan menggunakan *5-fold Stratified Cross Validation*

Gambar 3 menunjukkan *overfitting* masih terjadi pada model yang telah ditambahkan *5-fold Stratified Cross Validation*. Ini merupakan hal yang biasa terjadi, apalagi ketika menggunakan data yang imbalance (Tyagi dan Talbar, 2022; Lee et al, 2020; Shalabi dan Coolen, 2014).

Hasil dari penelitian ini menunjukkan model terbaik dalam deteksi inefisiensi klaim BPJS adalah algoritma *Random Forest + Tomek Links*. Sampai saat ini belum ada penelitian mengenai penggunaan *Machine Learning* untuk mendeteksi inefisiensi pada klaim yang diajukan oleh fasilitas Kesehatan ke BPJS Kesehatan. Salah satu kasus yang mirip dalam pendeteksian inefisiensi adalah dengan penggunaan *Machine Learning* untuk mendeteksi *fraud* yang telah dilaksanakan dalam penelitian terdahulu. Pada deteksi *fraud* di telecom (Lu dkk, 2020) model terbaik menggunakan *Random Forest* yang dikombinasikan dengan ADASYN. Sedangkan model terbaik untuk deteksi *fraud* dalam laporan keuangan oleh Ye dkk, 2019 adalah model *Random Forest* dengan kombinasi SMOTE. Algoritma *Random Forest* juga merupakan model dengan hasil terbaik dibandingkan *Support Vector Machine* dan *Logistic Regression* pada deteksi *fraud* transaksi konsumen pada

dataset *fraud* IEEE-CIS (Du dkk, 2021). Berdasarkan penelitian-penelitian sebelumnya terdapat kesamaan penggunaan algoritma *Random Forest* sebagai model terbaik untuk melakukan klasifikasi atas masalah yang diberikan.

Penggunaan algoritma selain *Random Forest* seperti *CatBoost*, *XGBoost*, *Gradient Boosting Classifier*, *Decision Tree*, *Support Vector Classifier*, dan *Naïve Bayes* membuat penelitian ini berbeda dengan penelitian-penelitian sebelumnya. Selain itu, penelitian ini juga mengkombinasikan metode *oversampling* berupa SMOTE dan ADASYN serta metode *undersampling* berupa *Tomek Links* dan *NearMiss*.

Perbedaan algoritma yang dapat menghasilkan model terbaik dalam setiap penelitian merupakan hasil dari perbedaan *dataset* yang digunakan. *Dataset* merupakan faktor penting yang mempengaruhi performa dari model prediksi (Ding dkk, 2021). Model *Machine Learning* dibangun berdasarkan data yang diberikan. Adanya perubahan data yang digunakan akan mengubah model *Machine Learning* (Zhang dan Ling, 2018). Penelitian ini bertujuan untuk mendeteksi apakah terjadi inefisiensi pada klaim yang diajukan ke BPJS Kesehatan oleh fasilitas kesehatan, dengan demikian *dataset* serta karakteristik data yang digunakan tentunya berbeda dengan *dataset* yang digunakan untuk deteksi *fraud* maupun penelitian lain yang sejenis. Dengan perbedaan *dataset* tersebut, maka diperlukan pembuatan model *machine learning* baru dengan penentuan model terbaiknya didasarkan pada matrik evaluasi yang dipilih sehingga penelitian ini menghasilkan *output* yang berbeda dengan penelitian-penelitian sebelumnya.

## KESIMPULAN

Penelitian ini membangun model deteksi inefisiensi pada klaim yang diajukan oleh fasilitas kesehatan ke BPJS Kesehatan dengan menggunakan *Machine Learning*. Dari model yang telah dievaluasi menggunakan *F1 Score*, didapat *Random Forest + Tomek Links* menghasilkan nilai tertinggi yakni 19,53 serta lima variabel yang paling mempengaruhi dari model tersebut adalah lokasi fasilitas kesehatan (*dati2*), usia peserta, diagnosa dari FKTP (*diagfktp*), diagnosa primer peserta (*diag*) dan tipe faskes. Untuk mengatasi *overfitting*, Penulis mengulang model *Random Forest + Tomek Links* dan menambahkan penggunaan *cross validation* berupa *Stratified K-Fold* dengan lima lipatan. Meskipun demikian, masalah *overfitting* tetap terjadi. Hal ini merupakan tantangan umum yang sering terjadi ketika menggunakan data yang bersifat *imbalance*. Dengan demikian, untuk penelitian selanjutnya, dapat dilakukan penambahan data seperti pendetailan pada komponen pembentuk biaya yang diajukan oleh fasilitas kesehatan dan data terkait obat-obatan yang diberikan karena jika berbicara inefisiensi tentu aspek keuangan akan menjadi faktor penting yang mana hal tersebut tidak ada pada *dataset* yang digunakan dalam penelitian ini.

**DAFTAR RUJUKAN**

- Annisa, R., Winda, S., Dwisaputro, E., & Isnaini, K. N. (2020). Mengatasi Defisit Dana Jaminan Sosial Kesehatan Melalui Perbaikan Tata Kelola. *Integritas: Jurnal Antikorupsi*, 6(2), 209-224.
- Badan Penyelenggara Jaminan Sosial. (2023). *Peserta Program JKN*. Diakses pada 12 April 2023, dari <https://bpjs-kesehatan.go.id/bpjs/>.
- Badan Penyelenggara Jaminan Sosial. (2023). *Sukses Pertahankan WTM, Ini Sejumlah Capaian BPJS Kesehatan di Tahun 2021*. Diakses pada 12 April 2023, dari <https://www.bpjs-kesehatan.go.id/bpjs/post/read/2022/2341/Sukses-Pertahankan-WTM-Ini-Sejumlah-Capaian-BPJS-Kesehatan-di-Tahun-2021>.
- Badan Penyelenggara Jaminan Sosial. (2022). Laporan Pengelolaan Program dan Keuangan BPJS Kesehatan Tahun 2021 (Audit). BPJS Kesehatan.
- Chandra, A., & Staiger, D. O. (2020). Identifying sources of inefficiency in healthcare. *The Quarterly Journal of Economics*, 135(2), 785-843.
- D. Berrar, Cross-Validation. *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, vol. 1, pp. 542-545, 2019
- Ding, Y., Fan, L., & Liu, X. (2021). Analysis of feature matrix in machine learning algorithms to predict energy consumption of public buildings. *Energy and Buildings*, 249, 111208.
- Eni, Y., & Edi Abdurachman, R. (2020). Analysis of Efficiency Level Using Data Envelopment Analysis (Dea) Method At The Indonesian Hospitals. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(7), 2463-2475.
- GEMCİ, F., İBRİKÇİ, T., & ÇEVİK, U. Comparative of Success of KNN With New Proposed K-Split Method And Stratified Cross Validation On Remote Homologue Protein Detection. *Eskişehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering*, 23(1), 87-108.
- Hairani, H., Anggrawan, A., & Priyanto, D. (2023). Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link. *JOIV: International Journal on Informatics Visualization*, 7(1), 258-264.
- Huilgol, P. (last modification 30 Mei 2023). Precision and Recall | Essential Metrics for Data Analysis (Updated 2023). [https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/#What\\_is\\_Recall](https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/#What_is_Recall).
- Ibrahim, M., Torki, M., & El-Makky, N. (2018, December). Imbalanced toxic comments classification using data augmentation and deep learning. *In 2018 17th IEEE international conference on machine learning and applications (ICMLA)* (pp. 875-878). IEEE.

- Janet, B., & Ganesh, D. P. S. (2022, April). Credit Card Fraud Detection with Unbalanced Real and Synthetic dataset using Machine Learning models. In *2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC)* (pp. 73-78). IEEE.
- Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: a brief primer. *Behavior Therapy*, *51*(5), 675-687.
- Lee, T., Kim, M., & Kim, S. P. (2020). Improvement of P300-based brain-computer interfaces for home appliances control by data balancing techniques. *Sensors*, *20*(19), 5576.
- Lu, C., Lin, S., Liu, X., & Shi, H. (2020, May). Telecom fraud identification based on ADASYN and random forest. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)* (pp. 447-452). IEEE.
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, *132*, 116045.
- Mqadi, N. M., Naicker, N., & Adeliyi, T. (2021). Solving misclassification of the credit card imbalance problem using near miss. *Mathematical Problems in Engineering*, *2021*, 1-16.
- Pramono, S. (2018). *Gambaran Proses Klaim Jkn Kis Dan Pelaksanaan Kredensialing Fasilitas Pelayanan Kesehatan Mitra Bpjs Kesehatan Kantor Cabang Purwokerto* (Doctoral dissertation, Universitas Harapan Bangsa).
- Putra, D. A. A., & Kusumo, M. P. (2016). Model Verifikasi Klaim BPJS Pasien Rawat Inap di RS PKU Muhammadiyah Gamping. In *Prosiding Interdisciplinary Postgraduate Student Conference*.tract/135/2/785/5698324
- Putri, N. K. A., Karjono, K., & Uktutias, S. A. (2019). Analisis Faktor-Faktor Penyebab Keterlambatan Pengajuan Klaim BPJS Kesehatan Pasien Rawat Inap Di RSUD Dr. R. Sosodoro Djatikoesoemo Bojonegoro. *Jurnal Manajemen Kesehatan Yayasan RS. Dr. Soetomo*, *5*(2), 134-143.
- Sawangarreerak, S., & Thanathamatee, P. (2020). Random forest with sampling techniques for handling imbalanced prediction of university student depression. *Information*, *11*(11), 519.
- Shalabi, A., Coolen, A. C. C., & de Rinaldis, E. (2014). Overcoming computational inability to predict clinical outcome from high-dimensional patient data using Bayesian methods. *arXiv preprint arXiv:1406.5062*.
- Shaohui, D., Qiu, G., Mai, H., & Yu, H. (2021, January). Customer transaction fraud detection using random forest. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 144-147). IEEE.
- Shrank, W. H., Rogstad, T. L., & Parekh, N. (2019). Waste in the US health care system: estimated costs and potential for savings. *Jama*, *322*(15), 1501-1509.

- Tyagi, S., & Talbar, S. N. (2022). CSE-GAN: A 3D conditional generative adversarial network with concurrent squeeze-and-excitation blocks for lung nodule segmentation. *Computers in Biology and Medicine*, 147, 105781.
- Wang, S., Liu, S., Zhang, J., Che, X., Yuan, Y., Wang, Z., & Kong, D. (2020). A new method of diesel fuel brands identification: SMOTE oversampling combined with XGBoost ensemble learning. *Fuel*, 282, 118848.
- Ye, H., Xiang, L., & Gan, Y. (2019, October). Detecting financial statement fraud using random forest with SMOTE. In IOP Conference Series: *Materials Science and Engineering* (Vol. 612, No. 5, p. 052051). IOP Publishing.
- Zeng X. Martinez TR. Distribution-balanced stratified cross-validation for accuracy estimation. *Journal of Experimental & Theoretical Artificial Intelligence*, 2000.
- Zhang, Y., & Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. *Npj Computational Materials*, 4(1), 25.
- Zhu, R., Wang, Z., Ma, Z., Wang, G., & Xue, J. H. (2018). LRID: A new metric of multi-class imbalance degree based on likelihood-ratio test. *Pattern Recognition Letters*, 116, 36-42.
- Peraturan Menteri Kesehatan Nomor 76 Tahun 2016 tentang Pedoman Indonesian Case Base Group (INA-CBG) dalam Pelaksanaan Jaminan Kesehatan Nasional.
- Peraturan Menteri Kesehatan Republik Indonesia Nomor 16 Tahun 2019 tentang Pencegahan dan Penanganan Kecurangan (*Fraud*) serta Pengenaan Sanksi Administrasi Terhadap Kecurangan (*Fraud*) dalam Pelaksanaan Program Jaminan Kesehatan Nasional.