

**PENERAPAN DATA MINING CLASSIFICATION DALAM
PENGECEKAN KECURANGAN TRANSAKSI YANG AKURAT
(STUDI KASUS : BANK BENGKULU)**

SKRIPSI

Program Studi Sistem Informasi

Jenjang Sarjana



Oleh:

Zacky Eka Putra

09031181621031

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2021

LEMBAR PENGESAHAN

PENERAPAN DATA MINING CLASSIFICATION DALAM
PENGECEKAN KECURANGAN TRANSAKSI YANG AKURAT (STUDI
KASUS : BANK BENGKULU)

Sebagai salah satu syarat untuk penyelesaian
Studi di Program Studi Sistem Informasi SI

Oleh:

Zacky Eka Putra

09031181621031

Indralaya, 19 Januari 2021

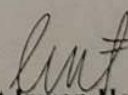
Mengetahui,

Ketua Jurusan Sistem Informasi



Endang Lestari Ruskan, M.T
NIP. 197811172006042001

Pembimbing I,



Rahmat Izwan Heroza, M.T
NIP. 198706302015041001

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini:

Nama : Zacky Eka Putra

NIM : 09031181621031

Program Studi : Sistem Informasi Reguler

Judul Skripsi : Penerapan Data Mining Classification dalam Pengecekan
Kecurangan Transaksi yang Akurat (Studi Kasus : Bank Bengkulu)

Hasil pengecekan *software authenticate/Turnitin* : 4%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya dengan ketentuan yang berlaku.

Demikianlah pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Indralaya, 19 Januari 2021



Zacky Eka Putra
NIM. 09031191621031

HALAMAN PERSETUJUAN

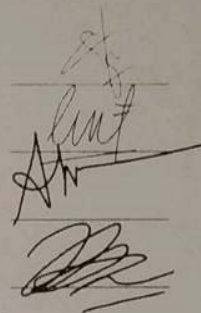
Telah diuji dan lulus pada

Hari : Selasa

Tanggal : 12 Januari 2021

Tim Penguji

- | | |
|--------------|--------------------------|
| 1 Ketua | Ari Wedasmara, M.T |
| 2 Pembimbing | Rahmat Izwan Heroza, M.T |
| 3 Anggota I | Allsela Meiriza, M.T |
| 4 Anggota II | Pacu Putra, M.Cs |



Mengetahui,

Ketua Jurusan Sistem Informasi,



Pandang Lestari Ruskan, M.T
NIP. 197811172006042001

HALAMAN PERSEMBAHAN

MOTTO :

*“You Can Be a Master, Don't Wait for Luck,
Dedicate Yourself and You Gonna Find Yourself”*

Skripsi ini ku persembahkan untuk :

- ❖ Kedua orang tua serta keluarga yang sangat kucintai
- ❖ Teman-teman seperjuangan
- ❖ Dosen Pembimbing yang telah sabar membimbingku
- ❖ Dosen-Dosen Fakultas Ilmu Komputer
- ❖ Almamaterku tercinta Universitas Sriwijaya

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan penyusunan Tugas Akhir. Tugas Akhir ini diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Program Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya. Judul yang penulis ajukan adalah “Penerapan Data Mining *Classification* dalam Pengecekan Kecurangan Transaksi yang Akurat (Studi Kasus : Bank Bengkulu)”.

Dalam penyusunan dan penulisan Tugas Akhir ini, penulis menyadari bahwa tanpa adanya bimbingan, bantuan, dan dorongan serta petunjuk dari semua pihak tidak mungkin Penelitian Tugas Akhir ini dapat diselesaikan. Pada kesempatan ini, dengan segala kerendahan hati penulis ingin menyampaikan rasa terima kasih kepada :

1. Kedua Orang Tua serta keluarga yang selalu memberi dukungan, semangat, uang dan doa kepada penulis.
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
3. Ibu Endang Lestari Ruskan, M.T selaku Ketua Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Rahmat Izwan Heroza, M.T. selaku pembimbing yang selalu sabar dalam membimbing penulis serta memberikan masukan yang membangun sehingga dapat selesainya Penelitian Tugas Akhir ini.
5. Pihak PT Bank Bengkulu yang telah mengizinkan pengambilan data, wawancara dan mempermudah proses penyusunan skripsi ini
6. Para Dosen yang telah membimbing, mengajari, dan membagi ilmunya kepada penulis selama menuntut ilmu di Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya

7. Saudara penulis, Julian Dwi Putra dan Afifah Astri Nina serta Tete Atmi yang sering sekali bertanya kapan wisuda.
8. Beberapa orang yang spesial yang terus mengingatkan dan memberikan semangat sekaligus tekanan dalam menyelesaikan Tugas Akhir ini.
9. Teman-teman bermain game online (LocalHost) tapi sangat berjasa dalam proses skripsi ini.
10. Teman – teman SIAKECE dan teman seperjuangan jurusan Sistem Informasi angkatan 2016.

Penulis menyadari bahwa penulisan tugas akhir ini masih jauh dari sempurna. Maka dari itu, penulis sangat mengharapkan kritik dan saran dari semua pihak dalam perbaikan dan penyempurnaan tugas akhir ini. Dengan keterbatasan yang ada semoga tugas akhir ini dapat bermanfaat bagi kita semua.

Palembang, 19 Januari 2021

Penulis,

A handwritten signature in black ink, appearing to read 'Zacky Eka Putra', with a long horizontal stroke at the bottom.

Zacky Eka Putra

**PENERAPAN DATA MINING CLASSIFICATION DALAM PENGECEKAN
KECURANGAN TRANSAKSI YANG AKURAT
(STUDI KASUS : BANK BENGKULU)**

Oleh :

Zacky Eka Putra

09031181621031


ABSTRAK

Fraud merupakan masalah umum dalam dunia finansial dan perbankan tidak terkecuali Bank Bengkulu. *Fraud* merupakan tindakan penyimpangan yang disengaja untuk memperoleh keuntungan baik secara langsung maupun tidak langsung dan mengakibatkan kerugian bagi pihak Bank seperti *lost profit* ataupun menurunnya kredibilitas perusahaan. Dalam mengatasi masalah bisnis tersebut, berbagai penelitian menyebutkan *data mining* merupakan salah satu rekomendasi terbaik dalam menangani mendeteksi *fraud* yang ada. Penelitian ini bertujuan untuk menerapkan dan mengetahui seberapa baik *data mining* dalam mendeteksi transaksi *fraud* dan menganalisa berbagai algoritma dan pendekatan yang dapat menghasilkan model dengan skor AUC dengan kategori *good* atau *excellent* yang dapat mendeteksi *fraud* pada transaksi perbankan dengan akurat. Karena *fraud* merupakan kasus langka dimana jumlah *fraud* pastinya akan sangat sedikit dibandingkan kasus transaksi normal dan dapat menyebabkan performa model kurang baik dalam mendeteksi *fraud* sehingga perlu berbagai pendekatan untuk masalah *imbalance* tersebut. Berbagai model dengan berbagai pendekatan yang dilatih dengan data yang ada menghasilkan hasil evaluasi yang bervariasi dan dapat dirangking berdasarkan skor AUC. Hasil dalam penelitian yaitu model *data mining* ternyata mampu mendeteksi transaksi *fraud* hingga 80 persen dari total kasus transaksi *fraud* yang ada. Pendekatan untuk kasus *imbalance classification problem* ternyata sangat membantu dalam meningkatkan performa model *data mining* dalam mendeteksi *fraud*. Model yang dihasilkan dari percobaan yang dilakukan memiliki performa yang cukup baik dengan *AUC score* 0.853 atau dapat dikategorikan *good* menuju *excellent* yang mana dengan hasil tersebut model dapat dikatakan cukup akurat dalam mendeteksi transaksi *fraud*.

Kata Kunci: *Fraud, Fraud Detection, Data Mining, CRISP-DM, Classification, Imbalanced Problem, Resampling, AUC Score.*

Indralaya, 22 Januari 2021

Pembimbing I,


Rahmat Izwan Heroza, M.T
NIP. 198706302015041001

**IMPLEMENTATION OF DATA MINING CLASSIFICATION IN ACCURATE
FRAUD TRANSACTION CLASSIFICATION
(CASE STUDY: BANK BENGKULU)**

By

Zacky Eka Putra

09031181621031

ABSTRACT

Fraud is a common problem in the financial and banking, including Bank Bengkulu. Fraud is a deliberate act of deviation in order to gain profit, either directly or indirectly, and results in losses for the bank, such as lost profits and decreased corporate credibility. In overcoming these business problems, various studies have stated that data mining is one of the best recommendations in dealing with detecting existing fraud. This study aims to apply and find out how well data mining is in detecting fraudulent transactions and analyze various algorithms and approaches that can produce a model with an AUC score with the good or excellent category that can accurately detect fraud in banking transactions. Because fraud is a rare case where the number of frauds is certainly very small compared to cases of normal transactions and can cause the model's performance to be less good in detecting fraud, so it needs various approaches to the problem of returns. Models with different approaches trained with available data yield varied evaluation results and can be ranked based on AUC scores. The result of the research is that the data mining model is able to detect fraudulent transactions up to 80 percent of the total fraud transaction cases that exist. Then the approach to the case of imbalance classification problems turned out to be very helpful in improving the performance of the data mining model in detecting fraud. And the model produced from the experiments conducted has a fairly good performance with an AUC score of 0.853 or can be categorized as good to excellent, which means that the model is quite accurate in detecting fraudulent transactions.

Keywords : Fraud, Fraud Detection, Data Mining, CRISP-DM, Classification, Imbalanced Problem, Resampling, AUC Score.

Indralaya, 22 Januari 2021

Supervisor I,


Rahmat Izwan Heroza, M.T
NIP. 198706302015041001

DAFTAR ISI

Halaman

LEMBAR PENGESAHAN	ii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan.....	6
1.3 Manfaat.....	7
1.5 Batasan Masalah.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Profil Bank Bengkulu.....	9
2.1.1 Sejarah.....	9
2.1.2 Struktur Organisasi	10
2.1.3 Visi dan Misi.....	11
2.2 Kajian Pustaka	12
2.3 <i>Credit Card Fraud Detection</i>	15
2.4 <i>Data Mining</i>	16
2.4.1 <i>Anomaly / Outlier Detection</i>	18
2.4.2 <i>Classification</i>	18

2.5	<i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i>	18
2.5.1	<i>Business Understanding</i>	19
2.5.2	<i>Data Understanding</i>	19
2.5.3	<i>Data Preparation</i>	19
2.5.4	<i>Modeling</i>	20
2.5.5	<i>Evaluation</i>	20
2.5.6	<i>Deployment</i>	20
2.6	<i>Resampling</i>	20
2.6.1	<i>Random Under Sampling</i>	21
2.6.2	<i>Random Over Sampling</i>	21
2.6.3	<i>SMOTE</i>	21
2.7	<i>Data Mining Classification Algorithm</i>	22
2.8	<i>Data Mining Evaluation</i>	23
2.9	Tools yang digunakan	25
2.9.1	<i>Python Programming Language</i>	25
2.9.2	<i>Jupyter Notebook</i>	25
BAB III	METODE PENELITIAN	27
3.1	Objek Penelitian	27
3.2	Metode Pengumpulan Data	27
3.2.1	Jenis Data	27
3.2.2	Sumber Data.....	27
3.2.3	Pengumpulan Data	27
3.3	Tahap Penelitian	28
3.4	<i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i>	30
3.4.1	<i>Business Understanding</i>	30
3.4.2	<i>Data Understanding</i>	31
3.4.3	<i>Data Preparation</i>	31
3.4.4	<i>Modeling</i>	31
3.4.5	<i>Evaluation</i>	31
3.4.6	<i>Deployment</i>	32
BAB IV	HASIL DAN PEMBAHASAN	33

4.1	Pemahaman Data (<i>Data Understanding</i>)	33
4.2	Persipan Data (<i>Data Preparation</i>)	36
4.2.1	<i>Data Cleaning</i>	36
4.2.2	<i>Data Encode</i>	37
4.2.3	<i>Scale Data</i>	38
4.2.4	<i>Split Data</i>	38
4.2.5	<i>Resampling</i>	39
4.3	Pemodelan (<i>Modeling</i>)	39
4.4	Evaluasi (<i>Evaluation</i>)	41
4.5	<i>Deployment</i>	48
4.6	Rekomendasi ke Perusahaan	50
4.7	Pembahasan	51
BAB V	HASIL DAN PEMBAHASAN	53
5.1	Kesimpulan	53
5.2	Saran	53
	DAFTAR PUSTAKA	55
	LAMPIRAN	57

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Struktur Organisasi Bank Bengkulu.....	10
Gambar 2.2 <i>Data Mining Techniques</i>	17
Gambar 2.3 <i>CRISP-DM Life Cycle</i>	19
Gambar 2.4 <i>Confusion Matrix</i>	24
Gambar 3.1 Tahap Penelitian.....	28
Gambar 4.1 Eksplorasi Data	33
Gambar 4.2 Statistik deskriptif dari data.....	34
Gambar 4.3 Pemilihan <i>Feature Data</i>	36
Gambar 4.4 Pembersihan Data dari <i>Missing Value</i>	36
Gambar 4.5 <i>Encoding</i> data ketegori kedalam bentuk numerik	37
Gambar 4.6 <i>Scale Independent Feature</i>	38
Gambar 4.7 StratifiedKFold Validation Skema	39
Gambar 4.8 Penyiapan <i>Classifier</i> atau Algoritma.....	40
Gambar 4.9 Pembuatan <i>function TrainModel</i>	40
Gambar 4.10 Menjalankan <i>function TrainModel</i>	41
Gambar 4.11 Pohon keputusan dari model terbaik	49
Gambar 4.12 Pohon keputusan dari model terbaik dengan kedalam tiga	49
Gambar 4.13 Website dengan inplementasi model terpilih	50

DAFTAR TABEL

	Halaman
Tabel 4.1 Hasil evaluasi model tanpa <i>resampling</i>	42
Tabel 4.2 Hasil evaluasi model dengan Oversampling.....	44
Tabel 4.3 Hasil evaluasi model dengan SMOTE.....	45
Tabel 4.4 Hasil evaluasi model dengan Undersampling.....	46
Tabel 4.5 Hasil evaluasi model dengan AUC Score tertinggi	47

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam dunia finansial, miliaran dollar hilang setiap tahunnya karena penipuan kartu kredit. Laporan *online fraud* oleh CyberSource menunjukkan hal itu meskipun persentase kehilangan pendapatan (*lost profit*) telah stabil sebesar 1,4% dari pembayaran online, dan pertumbuhannya sejalan meningkatnya pertumbuhan penjualan online. Mengacu pada Nilson Report (2015), kerugian akibat penipuan kartu kredit dan debit mencapai 16,31 juta dollar pada tahun 2014, dan naik 14%. Cybersource (2013) melaporkan bahwa *online fraud* menyebabkan kerugian 3,5 juta dollar pada tahun 2012 (Zhang et al. 2019). Dengan pertumbuhan transaksi kartu kredit sebagai sistem pembayaran, hal itu berkorelasi peningkatan penipuan kartu kredit. Selama bertahun-tahun, seiring dengan evolusi metode deteksi penipuan, pelaku penipuan juga telah berkembang praktik penipuan mereka untuk menghindari deteksi. Dan survey terbaru mengenai *fraud* menunjukkan bahwa perilaku *fraud* merupakan masalah global dan menghabiskan sekitar 5% pendapatan tahunan organisasi pada umumnya (Baader and Kremer 2018).

Fraud atau tindakan penyimpangan yang sengaja dilakukan untuk mengelabui, menipu, atau memanipulasi bank, nasabah, atau pihak lain, yang terjadi di lingkungan Bank sehingga pelaku *fraud* memperoleh keuntungan baik secara langsung atau tidak langsung dan mengakibatkan bank, nasabah, atau pihak lain menderita kerugian (Bank Indonesia 2011). Dampak yang ditimbulkan oleh *fraud* cukup besar baik bagi nasabah & pihak lain maupun pihak bank itu sendiri.

Selain *lost profit* yang cukup merugikan bagi pihak bank, masalah lain yang juga dapat timbul yaitu turunnya reputasi atau kredibilitas bank.

Bank Indonesia selaku bank sentral juga telah membuat kebijakan untuk menangani dan mengurangi dampak *fraud* dan harus dipatuhi oleh bank-bank umum yang ada di Indonesia melalui surat edaran Bank Indonesia no. 13 tahun 2011 perihal Penerapan Strategi Anti Fraud bagi Bank Umum. Dalam surat edaran tersebut, bank umum wajib menerapkan strategi anti *fraud* yang dalam penerapannya berupa sistem pengendalian *fraud* yang terdiri dari empat pilar yaitu, pencegahan, deteksi, investigasi, dan evaluasi.

Pada saat ini, berbagai bank telah menerapkan sistem deteksi *fraud* dengan berbagai teknik dan pendekatan. Bank Bengkulu merupakan salah satu bank yang memiliki layanan kartu kredit dan telah memiliki sistem deteksi *fraud* pada transaksi kartu kredit dan belum berbasis *machine learning*, namun dalam penanganan *fraud* tersebut dibutuhkan *resource* biaya yang cukup besar dan waktu yang cukup lama dan belum berbasis *behaviour*. Berdasarkan *annual report* yang diterbitkan oleh Bank Bengkulu 4 tahun terakhir, ada beberapa kasus *fraud* yang terdeteksi oleh sistem. Dan melihat dari jumlah kasus yang terdeteksi cukup sedikit dan teknik tradisional yang belum terlalu efektif untuk menemukan pola dari data yang berjumlah besar, dikhawatirkan masih ada beberapa kasus *fraud* yang mungkin lolos atau terdeteksi sebagai transaksi normal. Hal ini perlu diperhatikan mengingat jumlah transaksi yang terus meningkat dan juga potensi transaksi *fraud* terus meningkat dan terus berkembang polanya. Untuk itu Bank

Bengkulu perlu menilai kembali sistem yang ada dan berbagai rekomendasi yang ada guna meningkatkan kemampuan dari sistem saat ini.

Dalam beberapa tahun terakhir perkembangan teknologi informasi telah sangat berkembang. Hampir disetiap bidang kehidupan telah memanfaatkan teknologi informasi dalam mengefisiensikan proses bisnisnya mulai dari pertanian, perekonomian, kedokteran, pendidikan, dan berbagai bidang lainnya. Dampak penggunaan teknologi informasi yaitu terdokumentasinya berbagai data yang ada dalam gudang data (*Data Warehouse*) maupun dalam wadah yang jauh lebih besar lagi yaitu *Big Data*. Data yang banyak tersebut yang awalnya hanya sekedar data historis biasa namun sekarang dapat diolah / diekstrak untuk mendapatkan informasi dan pengetahuan yang bermanfaat dan juga pola tersembunyi (*Hidden Pattern*) yang ada. *Knowledge* ataupun *hidden pattern* yang ada dapat dimanfaatkan oleh eksekutif atau manajer tingkat atas sebagai masukan dan rekomendasi dalam pengambilan keputusan strategis bisnis dan juga dapat digunakan dalam melihat pola-pola transaksi *fraud* yang terdapat dalam data transaksi yang besar.

Sistem pendeteksi kecurangan transaksi perbankan yang memiliki performa lebih baik akan berbanding lurus dengan penurunan *lost profit*. Hal ini disebabkan oleh kemampuan yang baik dalam mencegah transaksi *fraud*, sehingga dampak atau kerugian akibat *fraud* dapat dikurangi. Selain itu dengan sistem yang baik dapat berdampak terhadap performa bisnis yang lebih baik, meningkatkan efisiensi operasional dalam upaya pemantauan dan investigasi *fraud*, dan dapat mengidentifikasi dan bertindak terhadap perilaku (*behaviour*) baru yang tidak

biasa. Dengan melihat potensi teknologi yang ada dan salah satu solusi untuk pihak Bank Bengkulu dalam pengembangan sistem pendeteksi *fraud* tersebut yaitu *data mining* dan *machine learning* karena memiliki potensi yang mungkin dapat menurunkan *cost* atau biaya penanganan, waktu deteksi kecurangan yang lebih cepat serta menghasilkan performa lebih baik prediksi yang lebih akurat dibandingkan dengan teknik tradisional yang ada (Nian et al. 2016).

Mengingat volume data yang ada telah sangat besar dan mesin dapat mengolah data lebih cepat, serta lebih akurat dibandingkan teknik tradisional ataupun aplikasi statistik. *Data mining* dan *machine learning* dapat menjadi solusi dan saat ini telah banyak digunakan oleh berbagai perusahaan finansial besar untuk mengelola dataset yang ada untuk mendapatkan informasi, *knowledge*, ataupun *hidden pattern* dari data untuk mendeteksi terjadinya *fraud*. Dan juga perusahaan-perusahaan besar saling berlomba dalam pengembangan teknologi ini dengan tujuan meningkatkan keunggulan kompetitif perusahaan mereka. Oleh karena itu Bank Bengkulu perlu mempertimbangkan untuk memulai menggunakan teknologi tersebut. *Data mining* mengacu pada proses pencarian informasi tersembunyi dari sejumlah besar data melalui algoritma. Penambangan data menggunakan studi multi-disiplin dari matematika, statistik, ilmu informasi, dan ilmu komputer, dengan memanfaatkan teknik seperti *machine learning*, *classification*, *clustering analysis*, *data mining*, *database*, dan *visualization*. (Liu et al. 2019). *Data mining* merupakan solusi hemat biaya dan efisien dibandingkan dengan aplikasi data statistik lainnya, dan juga memiliki proses yang cepat dan

memudahkan penggunanya dalam menganalisis sejumlah besar data dalam waktu yang lebih singkat.

Pemilihan teknik data mining bergantung pada tujuan yang diinginkan dan *dataset* yang ada, namun ada hal yang menarik yaitu pada *fraud detection*. Teknik *data mining* yang ada, mayoritas tidak terlalu dapat menangani kasus *fraud* dikarenakan *imbalance data class*. Kasus *Imbalance data* merujuk pada dataset di mana satu atau beberapa kelas yang memiliki jumlah data yang jauh lebih besar daripada yang lain. Kelas yang paling lazim (*non-fraud*) disebut kelas mayoritas karena sering muncul, sedangkan kelas paling langka (*fraud*) atau jarang terjadi biasa disebut kelas minoritas (Haixiang et al. 2017). Pada kasus ini, performa model akan baik pada kelas mayoritas (*non-fraud*) dan tidak untuk kelas minoritas (*fraud*). Akurasi model pada kasus ini sangat tinggi namun model performa kurang baik.

Beberapa tahun kebelakang, para peneliti telah mengembangkan berbagai cara untuk menangani masalah *machine learning* tersebut. Ada beberapa pendekatan untuk menangani kasus *imbalanced dataset*, antara lain yaitu dengan *data mining anomaly / outlier detection* dan juga *data mining classification* dengan berbagai pendekatan seperti pendekatan *data-level* yaitu *resampling dataset*, pendekatan *algorithm-level*, pendekatan *cost-sensitive learning*, dan pendekatan *boosting / ensemble method* (Dal Pozzolo et al. 2014).

Kedua teknik tersebut (*anomaly* dan *classification*) dapat digunakan untuk menyelesaikan masalah *fraud detection* dengan pendekatan masing-masing. Mengingat jumlah data yang banyak dan pola tindakan kecurangan (*fraud*) yang

terus berubah-ubah seiring waktu, banyak peneliti lebih memilih metode klasifikasi dengan berbagai pendekatan seperti merekayasa distribusi data, penyetelan parameter algoritma, dan *boosting / ensemble method*. Pendekatan *data-level* atau biasa dikenal dengan *resampling* dan dikombinasikan dengan berbagai algoritma klasifikasi yang ada terbukti menghasilkan model dengan performa yang cukup baik. Dari berbagai algoritma data mining klasifikasi, tidak ada algoritma yang benar-benar mutlak menghasilkan model terbaik untuk mendeteksi kasus *fraud* (Haixiang et al. 2017). Mengacu pada penelitian-penelitian terdahulu, terkadang salah satu algoritma bekerja dengan sangat baik pada salah satu dataset namun tidak terlalu baik pada dataset tertentu, karena itu kita tidak dapat hanya mengandalkan salah satu algoritma. Untuk mendapatkan kombinasi algoritma dan pendekatan terbaik untuk itu maka perlu dilakukan pengujian untuk membandingkan teknik dan algoritma yang ada.

Berdasarkan pernyataan yang telah diuraikan di atas, maka penyusun tertarik dan ingin menajukan penelitian dengan judul “**Penerapan *Data Mining Classification* dalam Pengecekan Kecurangan Transaksi yang Akurat (Studi Kasus : Bank Bengkulu)**”.

1.2 Tujuan

Adapun tujuan yang ingin dicapai dari pembuatan laporan tugas akhir ini yaitu, untuk menerapkan dan mengetahui seberapa baik *data mining* dalam mendeteksi transaksi kecurangan (*fraud*). Dan menganalisis algoritma dan pendekatan yang menghasilkan model dengan AUC score dengan kategori *good*

atau *excellent* yang dapat mendeteksi kasus kecurangan (*fraud*) pada transaksi perbankan dengan akurat.

1.3 Manfaat

Adapun manfaat yang diharapkan dari pembuatan laporan tugas akhir ini adalah sebagai berikut :

1. Pihak bank mendapatkan rekomendasi sebagai alternatif sistem deteksi *fraud* yang dapat dikembangkan dalam mengklasifikasi / mendeteksi kecurangan (*fraud*) pada transaksi perbankan yang lebih akurat.
2. Dapat mengimplementasikan berbagai teknik dan pendekatan untuk menangani kasus *imbalance data* pada data mining klasifikasi dalam pembuatan sistem deteksi kecurangan (*fraud*) pada transaksi perbankan yang lebih akurat.

1.5 Batasan Masalah

Agar pembahasan yang diteliti tidak menyimpang dan keluar dari konteks serta lebih terarah dan dapat dipahami dengan mudah maka perlu adanya batasan masalah, adapun batasan masalahnya antara lain:

1. Data yang akan digunakan dalam yaitu data transaksi *fraud* perbankan dari IEEE *competition* yang sudah di *transform* dan variabel yang mungkin tidak dijelaskan dikarenakan privasi data pelanggan.
2. Untuk *Classification algorithm* yang akan digunakan yaitu *Decision Tree*, *Logistic Regression*, *Random Forest*, *Neural Network*, *Naive Bayes*, *Random Forest*, dan *XGBoost* dikarenakan kedelapan

algoritma tersebut memiliki performa yang cukup baik dalam data mining klasifikasi berdasarkan penelitian serupa yang telah lalu. Dan pendekatan untuk meningkatkan performa pada kasus *imbalance* yaitu pendekatan *resampling* data (*Random Under Sampling, Random Over Sampling, & SMOTE*). Data tipe *dataframe* yang digunakan yaitu X1 dengan 35 *feature set* dan X2 dengan *feature set* identitas

3. Dalam mengimplementasikan *data mining* akan menggunakan *Python Programming*.
4. Visualisasi hasil uji coba dan model *data mining* akan ditampilkan menggunakan *web-based* menggunakan jupyter notebook.
5. Tahapan *deployment* dari model yang telah dibangun yaitu berupa website sebagai *prototype* implementasi model dan belum implementasi ke sistem yang ada sekarang dikarenakan membutuhkan otoritas, perizinan, dan pengujian lebih lanjut dari pihak bank.

DAFTAR PUSTAKA

- Baader, Galina and Helmut Krcmar. 2018. "Reducing False Positives in Fraud Detection: Combining the Red Flag Approach with Process Mining." *International Journal of Accounting Information Systems* 31(July 2016):1–16.
- Bank indonesia. 2011. "Surat Edaran BI - No.13." *Bank Indonesia* (13):7.
- Brownlee, Jason PhD. 2020. "Step-By-Step Framework for Imbalanced Classification Projects." Retrieved August 11, 2020 (<https://machinelearningmastery.com/framework-for-imbalanced-classification-projects/>).
- Carcillo, Fabrizio, Yann-aël Le Borgne, Olivier Caelen, and Yacine Kessaci. 2019. "Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection." *Information Sciences*.
- Cases, Use. 2019. "Chapter 16 - Comparison of Data Science Algorithms." Pp. 523–29 in.
- Dal Pozzolo, Andrea, Olivier Caelen, Yann Aël Le Borgne, Serge Waterschoot, and Gianluca Bontempi. 2014. "Learned Lessons in Credit Card Fraud Detection from a Practitioner Perspective." *Expert Systems with Applications* 41(10):4915–28.
- Dubey, Rashmi, Jiayu Zhou, Yalin Wang, Paul M. Thompson, and Jieping Ye. 2014. "NeuroImage Analysis of Sampling Techniques for Imbalanced Data : An n = 648 ADNI Study." *NeuroImage* 87:220–41.
- Galton, Francis. 2019. *Chapter 5 - Regression Methods*.
- Haixiang, Guo, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. 2017. "Learning from Class-Imbalanced Data: Review of Methods and Applications." *Expert Systems with Applications* 73:220–39.
- Han, Jiawei, Micheline Kamber, and Jian Pei. 2014. *Data Mining: Data Mining Concepts and Techniques*. Morgan Kaufmann.
- Industry, Cross, Standard Process, Data Mining, Cross Industry, Standard Process, Data Mining, and The Crisp-dm. 2019. "Chapter 2 - Data Science Process." Pp. 19–37 in.
- Jagelid, Michelle. 2017. "A Comparison of Resampling Techniques to Handle the Class Imbalance Problem in Machine Learning Conversion Prediction of Spotify Users." KTH ROYAL INSTITUTE OF TECHNOLOGY.
- El Khouli, Riham H., Katarzyna J. Macura, Peter B. Barker, Mohamed R. Habba, Michael A. Jacobs, and David A. Bluemke. 2009. "Relationship of Temporal

- Resolution to Diagnostic Performance for Dynamic Contrast Enhanced MRI of the Breast.” *Journal of Magnetic Resonance Imaging* 30(5):999–1004.
- Kirlidog, Melih and Cuneyt Asuk. 2012. “A Fraud Detection Approach with Data Mining in Health Insurance.” Pp. 989–94 in *Procedia - Social and Behavioral Sciences*. Vol. 62.
- Kotu, Vijay and Bala Deshpande. 2019a. “Chapter 1 - Introduction.” Pp. 1–18 in.
- Kotu, Vijay and Bala Deshpande. 2019b. “Chapter 4 - Classification.”
- Liu, Jiaying, Xiangjie Kong, Xinyu Zhou, Lei Wang, Da Zhang, Ivan Lee, and Bo Xu. 2019. “Data Mining and Information Retrieval in the 21st Century : A Bibliographic Review.” 34.
- Nian, Ke, Haofan Zhang, Aditya Tayal, Thomas Coleman, and Yuying Li. 2016. “Auto Insurance Fraud Detection Using Unsupervised Spectral Ranking for Anomaly.” *The Journal of Finance and Data Science* 2(1):58–75.
- Suh, Yirey, Jaemyung Yu, Jonghoon Mo, Leegu Song, and Cheongtag Kim. 2017. “A Comparison of Oversampling Methods on Imbalanced Topic Classification of Korean News Articles.” *Journal of Cognitive Science* 18(4):391–437.
- Wikipedia, Contributors. 2019. “Python (Bahasa Pemrograman).” *Wikipedia, Ensiklopedia Bebas*. Retrieved December 17, 2019 ([https://id.wikipedia.org/wiki/Python_\(bahasa_pemrograman\)](https://id.wikipedia.org/wiki/Python_(bahasa_pemrograman))).
- Yıldırım, Pınar. 2016. “Pattern Classification with Imbalanced and Multiclass Data for the Prediction of Albendazole Adverse Event Outcomes Pınar Yıldırım A.” Pp. 1013–18 in *Procedia - Procedia Computer Science*. Vol. 83. Elsevier Masson SAS.
- Zhang, Xinwei, Yaoci Han, Wei Xu, and Qili Wang. 2019. “HOBA : A Novel Feature Engineering Methodology for Credit Card Fraud Detection with a Deep Learning Architecture.” *Information Sciences*.