

**Pengaruh *Random Under-sampling* Terhadap Klasifikasi
Kecurangan Transaksi *Mobile* Menggunakan *Gradient Boost Tree***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Darmawan Abinugroho

NIM : 09021181722010

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2022

LEMBAR PENGESAHAN SKRIPSI

PENGARUH *RANDOM UNDER-SAMPLING* TERHADAP KLASIFIKASI KECURANGAN TRANSAKSI *MOBILE* MENGUNAKAN *GRADIENT OOST TREE*

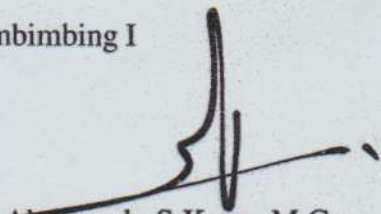
Oleh:

Darmawan Abinugroho

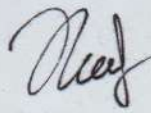
NIM : 09021181722010

Palembang, 27 Januari 2022

Pembimbing I


Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005


Pembimbing II,


Kanda Januar Miraswan, M.T.
NIP. 199001092019031012

Mengetahui,

Ketua Jurusan Teknik Informatika




Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

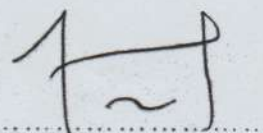
TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Senin tanggal 28 Maret 2022 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Darmawan Abinugroho
NIM : 09021181722010
Judul : Pengaruh *Random Under-sampling* Terhadap Klasifikasi Kecurangan Transaksi Secara *Mobile* Menggunakan *Gradient Boost Tree*

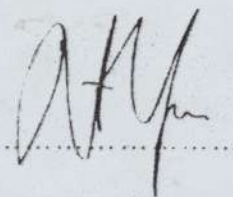
1. Ketua Penguji

Muhammad Fachurrozi, S.Si., M.T.
NIP. 198005222008121002



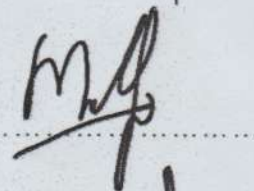
2. Penguji I

Novi Yusliani, M.T.
NIP. 198211082012122001



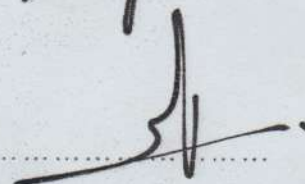
3. Penguji II

M. Qurhanul Rizqie, M.T., Ph.D.
NIP. 1671060312870008



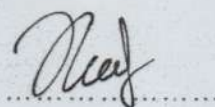
4. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005



5. Pembimbing II

Kanda Januar Miraswan, M.T.
NIP. 199001092019031012



Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan dibawah ini:

Nama : Darmawan Abinugroho
NIM : 09021181722010
Program Studi : Teknik Informatika Reguler
Judul Skripsi : Pengaruh *Random Under-sampling* Terhadap Klasifikasi Kecurangan Transaksi Secara *Mobile* Menggunakan *Gradient Boost Tree*

Hasil Pengecekan Software *iThenticate/Turnitin* : 17%

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

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun



Darmawan Abinugroho
NIM. 09021181722010

MOTTO DAN PERSEMBAHAN

MOTTO:

“It doesn't matter how slowly you go as long as you don't stop”

- Confucius

Kupersembahkan Karya Tulis ini kepada:

- Kedua Orang Tuaku
- Kedua Kakakku
- Keluarga Besar
- Sahabat dan Teman Seperjuangan
- Para Guru dan Dosen
- Fakultas Ilmu Komputer Universitas
Sriwijaya
- Diri sendiri

THE EFFECT OF RANDOM UNDERSAMPLING ON THE CLASSIFICATION OF MOBILE TRANSACTION FRAUD USING A GRADIENT BOOST TREE

By:

Darmawan Abinugroho
09021181722010

ABSTRACT

Mobile Money Transfer can be defined as a payment transaction process in the form of digital money. This process can make payments, money transfers, and transactions via mobile devices. In the traditional economy toward a digital economy, violations often occur. Control of fraud in the financial sector is not enough to commit fraud crimes. This study builds a mobile transaction classification system that uses the Gradient Boost Tree algorithm. There are problems with the data used, the data is unbalanced. Random undersampling was chosen to solve the data problem. In the application of this algorithm, 10 tests were carried out with the evaluation of k-fold cross-validation on the dataset. From and testing the data, it was found that the average training accuracy is 99.84% and the computation time is 1433.6 seconds for the Gradient Boost Tree algorithm with a parameter value of 100 while the 1000 parameter gets an accuracy of 99.89% with a time of 14044 seconds. In the next scenario, the average accuracy is 97.64% and the computation time is 2.98 seconds for Random Under-sampling and Gradient Boost Trees with a parameter value of 100. The last scenario obtained an average accuracy of 98.57% and computation time of 30.64 seconds for Random Under-sampling and Gradient Boost Tree with a parameter value of 1000. There is a decrease inaccuracy due to the Random Under-sampling method that removes data randomly to make the data balanced. Random data deletion causes some important data to be deleted and results in the classification results.

Key Word: Transaction Fraud Classification System, Fraud Transaction, Random Under-sampling, Gradient Boost Tree

PENGARUH *RANDOM UNDER-SAMPLING* TERHADAP KLASIFIKASI KECURANGAN TRANSAKSI *MOBILE* MENGGUNAKAN *GRADIENT BOOST TREE*

Oleh :

Darmawan Abinugroho

09021181722010

ABSTRAK

Mobile Money Transfer dapat diartikan sebagai proses transaksi pembayaran dalam bentuk uang digital. Proses tersebut dapat melakukan pembayaran, transfer uang, serta transaksi melalui perangkat selular. Pada peralihan ekonomi tradisional menuju ekonomi digital seringkali terjadi sebuah tindak pelanggaran. Kontrol terhadap penipuan pada bidang finansial belum cukup untuk menghentikan tindak pelaku kecurangan. Penelitian ini membangun sebuah sistem klasifikasi kecurangan transaksi secara *mobile* yang menggunakan algoritme *Gradient Boost Tree*. Terdapat permasalahan dalam data yang digunakan, data bersifat tidak seimbang. *Random Under-sampling* dipilih untuk mengatasi permasalahan data tersebut. Dalam penerapan algoritme ini, dilakukan pengujian sebanyak 10 kali dengan evaluasi *k-fold cross validation* pada *dataset*. Dari pelatihan dan pengujian data yang dilakukan, didapatkan rata - rata akurasi sebesar 99,84% dan waktu komputasi 1433,6 detik untuk algoritme *Gradient Boost Tree* dengan nilai parameter 100 sedangkan parameter 1000 mendapat akurasi sebesar 99,89% dengan waktu 14044 detik. Pada skenario berikutnya, didapatkan rata – rata akurasi sebesar 97,64% dan waktu komputasi 2,98 detik untuk *Random Under-sampling* dan *Gradient Boost Trees* dengan nilai parameter 100. Skenario terakhir didapatkan rata – rata akurasi sebesar 98,57% dan waktu komputasi 30,64 detik untuk *Random Under-sampling* dan *Gradient Boost Tree* dengan nilai parameter 1000. Adanya penurunan akurasi disebabkan karena cara kerja *Random Under-sampling* yang menghapus data secara acak untuk membuat data menjadi seimbang. Penghapusan data secara acak menyebabkan beberapa data penting juga terhapus dan berakibat pada hasil klasifikasi yang dilakukan.

Kata Kunci : Sistem Klasifikasi Kecurangan Transaksi, Kecurangan Transaksi, *Random Under-sampling*, *Gradient Boost Tree*

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas segala rahmat dan karuniaNya penulis mampu menyelesaikan skripsi dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan Pendidikan program Strata-1 Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan banyak bantuan dan dukungan baik materil dan moril selama proses penyusunan skripsi ini. Oleh karena itu, penulis ingin menyampaikan secara khusus kepada pihak yang telah membantu sebagai berikut:

1. Kedua orang tua saya Abdullah dan Inurmini, saudara – saudaraku Didi Pratama Adhiguna dan Ani Mutia Kurniasih yang telah memberikan dukungan materil dan moril serta doa dan restu yang terus menerus demi kelancara penyusunan skripsi ini.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya beserta jajarannya.
3. Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
4. Bapak Abdiansah, S.Kom., M.Cs selaku dosen pembimbing I dan Bapak Kanda Januar Miraswan, M.T selaku dosen pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi kepada penulis dalam proses perkuliahan dan penyusunan skripsi ini.
5. Ibu Novi Yusliani, M.T. selaku dosen penguji I dan Bapak M. Qurhanul Rizqie, M.T., Ph.D selaku dosen penguji II yang telah memberikan saran dan masukan untuk laporan skripsi ini.
6. Ibu Dian Palupi Rini, M.Kom., Ph.D selaku dosen pembimbing akademik yang telah membimbing, mengarahkan, dan memberikan motivasi penulis selama masa perkuliahan.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

8. Mba Winda Kurnia Sari dan Kak Ricy Firnando selaku admin Program Studi Teknik Informatika Indralaya, kak Cokro selaku kepala admin lab Indralaya, serta seluruh staf dan pegawai di Fakultas Ilmu Komputer yang telah membantu dalam kelancara adminstrasi dan akademik selama masa perkuliahan.
9. Ahmad, Agung, Faiz, Hafiz, Dzakiyyah, Rini, Suci, Nurhidayah serta teman – teman jurusan Teknik Informatika yang tidak dapat disebutkan satu persatu yang telah berbagi keluh kesah, motivasi, semangat, dan canda tawa selama masa perkuliahan.
10. Tommy, Defrian, Putri, Fahmi, Taufik serta teman – teman di bedeng kades yang tidak dapat disebutkan satu persatu yang telah memberi semangat, motivasi, dan canda tawa.
11. Kak Arfah Anggraina yang telah siap sedia untuk membantu penulis.
12. Semua teman – teman organisasi baik di Fakultas maupun Univeristas yang telah memberikan banyak pelajaran dan pengalaman berharga.

Penulis menyadari dalam penyusunan skripsi ini masih banyak terdapat kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun akan sangat diharapkan untuk kemajuan penelitian berikutnya.

Akhir kata semoga skripsi ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Januari 2022

Darmawan Abinugroho

DAFTAR ISI

Halaman

LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 RUS (<i>Random Under-sampling</i>)	II-1
2.2.2 <i>Gradient Boost Tree</i>	II-3
2.2.3 <i>Decision Tree</i>	II-6

2.2.4 <i>Confusion Matrix</i>	II-8
2.2.5 Rational Unified Process (RUP)	II-10
2.3 Penelitian Lain Yang Relevan	II-12
2.4 Kesimpulan	II-14
BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan.....	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 Jenis dan Sumber Data.....	III-1
3.2.2 Metode Pengumpulan Data.....	III-4
3.3 Tahapan Penelitian.....	III-4
3.3.1 Kerangka Kerja.....	III-4
3.3.2 Kriteria Pengujian	III-5
3.3.3 Format Data Pengujian	III-6
3.3.4 Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-7
3.3.5 Pengujian Penelitian	III-7
3.3.6 Analisis Hasil Pengujian dan Kesimpulan.....	III-9
3.4 Metode Pengembangan Perangkat Lunak	III-9
3.4.1 Fase Insepsi.....	III-9
3.4.2 Fase Elaborasi	III-10
3.4.3 Fase Konstruksi.....	III-11
3.4.4 Fase Transisi	III-11
3.5 Kesimpulan	III-12
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-2
4.1 Pendahuluan.....	IV-2
4.2 Fase Insepsi.....	IV-2
4.2.1 Pemodelan Bisnis.....	IV-2
4.2.2 Kebutuhan Sistem	IV-3
4.2.3 Analisis dan Desain	IV-6
4.3 Fase Elaborasi.....	IV-21
4.3.1 Pemodelan Bisnis.....	IV-21

4.3.2	Kebutuhan Sistem	IV-22
4.3.3	Diagram	IV-22
4.4	Fase Konstruksi	IV-34
4.4.1	Kebutuhan	IV-35
4.4.2	Implementasi	IV-36
4.5	Fase Transisi	IV-42
4.5.1	Pemodelan Bisnis	IV-42
4.5.2	Kebutuhan Sistem	IV-43
4.5.3	Rencana Pengujian	IV-43
4.5.4	Implementasi	IV-48
4.6	Kesimpulan	IV-42
 BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan	V-1
5.2	Hasil Program	V-1
5.3	Hasil penelitian	V-7
5.3.1	Konfigurasi Percobaan	V-7
5.4	Analisis Hasil Pengujian dan Perbandingan	V-18
5.5	Kesimpulan	V-20
 BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran	VI-2
 DAFTAR PUSTAKA		xiv
LAMPIRAN		xvii

DAFTAR TABEL

Halaman

Tabel II-1. Confusion Matrix	II-9
Tabel III-1. Sampel Data Asli Synthetic Financial Dataset For Fraud Detection.....	III-2
Tabel III-2. Sampel Data Setelah Prapengolahan yang Dipakai Dalam Penelitian	III-2
Tabel III-3. Rancangan Tabel Confusion Matrix Hasil Klasifikasi	III-6
Tabel III-4. Rancangan Tabel Hasil Klasifikasi.....	III-6
Tabel III-5. Rancangan Tabel Perbandingan Rata - rata Hasil Pengujian Klasifikasi ...	III-9
Tabel IV-1. Kebutuhan Fungsional.....	IV-5
Tabel IV-2. Kebutuhan Non Fungsional.....	IV-6
Tabel IV-3. Definisi Aktor Use Case.....	IV-10
Tabel IV-4. Definisi Use Case	IV-10
Tabel IV-5. Skenario Use Case Memuat Data	IV-12
Tabel IV-6. Skenario Use Case Menampilkan Data	IV-13
Tabel IV-7. Skenario Use Case Proses Train dengan Gradient Boost Tree.....	IV-15
Tabel IV-8. Skenario Use Case Proses Train Random Under-sampling dengan Gradient Boost Tree.....	IV-16
Tabel IV-9. Skenario Use Case Proses Test dengan Gradient Boost Tree	IV-18
Tabel IV-10. Skenario Use Case Proses Test Random Under-sampling dengan Gradient Boost Tree.....	IV-20
Tabel IV-11. Keterangan Implementasi Kelas.....	IV-37
Tabel IV-12. Rencana Pengujian Use Case Memuat Data	IV-43
Tabel IV-13. Rencana Pengujian Use Case Menampilkan Data.....	IV-44
Tabel IV-14. Rencana Pengujian Use Case Train dengan Gradient Boost Tree.....	IV-44
Tabel IV-15. Rencana Pengujian Use Case Test dengan Gradient Boost Tree	IV-45
Tabel IV-16. Rencana Pengujian Use Case Train dengan Random Under-sampling dan Gradient Boost Tree.....	IV-46
Tabel IV-17. Rencana Pengujian Use Case Test dengan Random Under-sampling dan Gradient Boost Tree.....	IV-47
Tabel IV-18. Pengujian Use Case Memuat Data	IV-34
Tabel IV-19. Pengujian Use Case Menampilkan Data	IV-34
Tabel IV-20. Pengujian Use Case Train dengan Gradient Boost Tree	IV-36

Tabel IV-21. Pengujian <i>Use Case Test</i> dengan Gradient Boost Tree	IV-37
Tabel IV-22. Pengujian Use Case Train dengan Random Under-sampling dan Gradient Boost Tree.....	IV-38
Tabel IV-23. Pengujian Use Case Test dengan Random Under-sampling dan Gradient Boost Tree.....	IV-40
Tabel V-1. Confusion Matrix Hasil Klasifikasi Gradient Boost Tree dengan N_Estimators 100.....	V-8
Tabel V-2. Hasil Evaluasi Klasifikasi Gradient Boost Tree dengan N_Estimators 100.	V-9
Tabel V-3. Confusion Matrix Hasil Klasifikasi Gradient Boost Tree dengan N_estimators 1000	V-11
Tabel V-4. Hasil Evaluasi Klasifikasi Gradient Boost Tree dengan N_Estimators 1000.....	V-11
Tabel V-5. Confusion Matrix Hasil Klasifikasi Dengan Random Under-sampling dan Gradient Boost Tree (n_estimator = 100).....	V-13
Tabel V-6. Hasil Evaluasi Klasifikasi dengan Random Under-sampling dan Gradient Boost Tree (n_estimator = 100).....	V-14
Tabel V-7. Confusion Matrix Hasil Klasifikasi Dengan Random Under-sampling dan Gradient Boost Tree (n_estimator = 1000).....	V-16
Tabel V-8. Hasil Evaluasi Klasifikasi dengan Random Under-sampling dan Gradient Boost Tree (n_estimator = 1000).....	V-16
Tabel V-9. Hasil Evaluasi Gradient Boost Tree dengan parameter 50 dan 500	V-18
Tabel V-10. Analisa Perbandingan Hasil Akurasi dan Waktu Komputasi	V-19

DAFTAR GAMBAR

	Halaman
Gambar II-1. Flowchart Random Under-Sampling	II-3
Gambar II-2. Flowchart Boosting	II-4
Gambar II-3. Arsitektur RUP (Kruchten, 2013)	II-11
Gambar III-1. Tahapan Penelitian Pengujian.....	III-8
Gambar IV-1. Diagram Use Case	IV-9
Gambar IV-2. Diagram Aktivitas Memuat Data.....	IV-23
Gambar IV-3. Diagram Aktivitas Menampilkan Data.....	IV-24
Gambar IV-4. Diagram Aktivitas Proses Training Gradient Boost Tree	IV-24
Gambar IV-5. Diagram Aktivitas Proses Testing Gradient Boost Tree.....	IV-25
Gambar IV-6. Diagram Aktivitas Proses Training Random Under-sampling dan Gradient Boost Tree	IV-26
Gambar IV-7. Diagram Aktivitas Proses Testing Random Under-sampling dan Gradient Boost Tree	IV-27
Gambar IV-8. Diagram Alur Memuat Data	IV-28
Gambar IV-9. Diagram Alur Menampilkan Data	IV-29
Gambar IV-10. Proses Train dengan Gradient Boost Tree	IV-30
Gambar IV-11. Proses <i>Test</i> dengan Gradient Boost Tree	IV-31
Gambar IV-12. Proses Train dengan Random Under-sampling dan Gradient Boost Tree	IV-32
Gambar IV-13. Proses Test dengan Random Under-sampling dan Gradient Boost Tree	IV-33
Gambar IV-14. Rancangan Antarmuka Perangkat Lunak	IV-34
Gambar IV-15. Diagram Kelas	IV-36
Gambar IV-16. Implementasi Antarmuka Perangkat Lunak	IV-38
Gambar IV-17. Hasil Klasifikasi GBT dengan Nilai Parameter 100.....	IV-39
Gambar IV-18. Hasil Klasifikasi RUS dan GBT dengan Nilai Parameter 100	IV-40
Gambar IV-19. Hasil Klasifikasi GBT dengan Nilai Parameter 1000.....	IV-41
Gambar IV-20. Hasil Klasifikasi RUS dan GBT dengan Nilai Parameter 1000	IV-42
Gambar V-1. Open Dataset Menggunakan Random Under-sampling.....	V-2
Gambar V-2. Proses Pelatihan Model Gradient Boost Tree dengan Random Under-sampling	V-3

Gambar V-3. Hasil Pengujian Gradient Boost Tree dengan Random Under-sampling dengan Masukkan Nilai N_Estimators 100	V-4
Gambar V-4. Hasil Pengujian Gradient Boost Tree dengan Random Under-sampling dengan Masukkan Nilai N_Estimators 1000	V-5
Gambar V-5. Hasil Pengujian Gradient Boost Tree dengan Masukkan Nilai N_Estimators 100	V-6
Gambar V-6 Hasil Pengujian Gradient Boost Tree dengan Masukkan Nilai N_Estimators 1000	V-7
Gambar V-7. Grafik GBT dengan N_Estimators = 100	V-10
Gambar V-8. Grafik GBT dengan N_Estimators = 1000	V-12
Gambar V-9. Grafik RUS dan GBT dengan N_Estimators = 100.....	V-15
Gambar V-10. Grafik RUS dan GBT dengan N_Estimators = 1000.....	V-17
Gambar V-11. Grafik Perbandingan Hasil Akurasi	V-19

DAFTAR LAMPIRAN

	Halaman
Tabel L-1. Contoh Data True Positive GBT (100)	xvii
Tabel L-2. Contoh Data False Positive GBT (100).....	xviii
Tabel L-3. Contoh Data False Negative GBT (100)	xix
Tabel L-4. Contoh Data True Negative GBT (100).....	xx
Tabel L-5. Contoh Data True Positive RUS + GBT (100)	xxi
Tabel L-6. Contoh Data False Positive RUS + GBT (100)	xxii
Tabel L-7. Contoh Data False Negative RUS + GBT (100).....	xxiii
Tabel L-8. Contoh Data True Negative RUS + GBT (100).....	xxiv
Tabel L-9. Contoh Data True Positive RUS + GBT (1000)	xxv
Tabel L-10. Contoh Data False Positive RUS + GBT (1000)	xxvi
Tabel L-11. Contoh Data False Positive RUS + GBT (1000)	xxvii
Tabel L-12. Contoh Data False Positive RUS + GBT (1000)	xxviii

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini memperkenalkan latar belakang pertanyaan penelitian, rumusan pertanyaan Penelitian, tujuan penelitian, manfaat penelitian, dan batasan masalah dalam penelitian. Bab ini juga berisi penjelasan umum dari keseluruhan penelitian yang dilakukan dan menjelaskan teori algoritme yang akan digunakan pada penelitian ini. Penjelasan algoritme *Gradient Boost Tree Classification*, Metode untuk mengatasi kesalahan klasifikasi teknis acak *Under-sampling*, dan studi terkait yang telah dipelajari sebelumnya.

1.2 Latar Belakang

Transaksi pembayaran yang mengizinkan penggunaan uang dalam bentuk virtual dalam melakukan pembayaran, transfer uang, serta transaksi melalui perangkat seluler dapat disebut sebagai *Mobile Money Transfer* (MMT) (Coppolino et al., 2015). Pada saat peralihan ekonomi tradisional menuju ekonomi digital, akses terhadap ekonomi digital menjadi lebih mudah dan seringkali rentan pada sebuah pelanggaran. Kontrol terhadap penipuan dalam finansial belum cukup untuk menghentikan pelaku kecurangan (Lopez-Rojas & Barneaud, 2019). Selain kemudahannya, layanan pembayaran elektronik telah menjadi target dalam penipuan selama beberapa tahun. Antara 2015 dan 2016, telah terjadi tindak penipuan yang dilakukan oleh satu kelompok pada akun *Apple Pay* yang

menghabiskan lebih dari \$1,5 juta. Tahun 2015, penipu menyerang akun *PayPal* dan menyebabkan kerugian sebesar \$1 juta. Diperkirakan bahwa kerugian akibat dari tindak kejahatan ini akan meningkat dua kali lipat pada tahun 2020. Oleh karena itu, sistem deteksi untuk kejahatan seperti ini sangat penting (Kang, 2019).

Berbagai macam metode telah digunakan untuk deteksi kecurangan transaksi *mobile*. Menurut Botchey et al.(2020) mengusulkan metode *Support Vector Machines*, *Gradient Boosted Decision Trees*, dan *Naïve Bayes*. Hasil akurasi *Support Vector Machine* yang dipadukan dengan *Random Under-sampling* adalah 86,34%. Sedangkan untuk *Naïve Bayes* dengan *Random Under-sampling* menghasilkan akurasi sebesar 88,97%. *Gradient Boosted Tree* mendapatkan akurasi sebesar 99,90%. Penelitian lain terkait deteksi kecurangan *Mobile Money Transaction* pernah dilakukan oleh Kang (2019) menggunakan metode *machine learning* seperti *Random Forest* dan *Gradient Boosting*. *Random Forest* memiliki hasil akurasi sebesar 98,98% sedangkan *Gradient Boosting* dengan hasil akurasi 99,24%.

Support Vector Machine (SVM) mempunyai kelebihan dalam menyediakan generalisasi *out-of-sample* yang baik dengan menentukan nilai generalisasi yang sesuai. SVM juga dapat menjadi kuat dalam modelnya bahkan ketika sampel pelatihan memiliki beberapa bias. Kelemahan yang ada pada SVM adalah kurangnya transparansi hasil. Metode ini tidak dapat mewakili semua kumpulan nilai sebagai fungsi parametrik sederhana (Auria dan Rouslan., 2008). *Naïve Bayes* (NB) punya keunggulan, yaitu hanya memerlukan sedikit waktu untuk melatih data dan meningkatkan kinerja klasifikasi dengan cara menghapus fitur yang tidak

relevan (Jadhav & Channe, 2016). Klasifikasi dengan metode ini juga memiliki kekurangan seperti membutuhkan file catatan dengan jumlah besar untuk mendapatkan hasil yang bagus (Amra, 2017).

Algoritme *Gradient Boost Tree* (GBT) bisa digunakan untuk pengklasifikasian serta dalam hal regresi (Zhou et al., 2016). GBT dapat membangun sebuah model *decision tree* didasarkan dengan perbaikan struktur pohon dengan pelatihan yang lemah untuk memperbaiki kesalahan di pohon dan mencegah terjadinya *overfitting*. Algoritme ini mampu menembus masalah dengan menyesuaikan pemebelajaran lemah dengan gradien negatif dari fungsi kerugian (*loss function*) dan memperbaiki pohon (*trees*) dengan parameter yang mewakili variabel split yang dipasang pada masing-masing *node* terminal dalam *tree*. Dalam mengatasi *misclassification* digunakan pendekatan secara data dalam menyelesaikan masalah ketidakseimbangan data. *Misclassification* dapat terjadi akibat dari data yang tidak seimbang akibat distribusi data yang tidak merata (Bisri & Rachmatika, 2019).

Teknik *Undersampling* merupakan teknik yang secara sembarang memilih contoh dari kelas mayoritas dan menghapusnya sebanyak jumlah kelas mayoritas sehingga membentuk data latih yang baru (He et al., 2018). *Random Under-sampling* (RUS) menghasilkan subsampel sembarang dari *instance* kelas mayoritas. Metode ini secara acak memilih sampel di kelas mayoritas dan menghapusnya. Teknik *Random Under-sampling* ini dipilih sebagai penyelesaian pada masalah ketidakseimbangan kelas.

Pada penelitian ini, pemilihan Teknik *Undersampling* akan membantu dalam menangani *misclassification*. Metode *Random Under-sampling* dipilih untuk menangani masalah pada data dengan tujuan menjadikan hasil prediksi menjadi lebih baik.

1.3 Rumusan Masalah

Pada penelitian ini terdapat rumusan masalah yaitu bagaimana peranan *Random Under-sampling* dalam menangani *misclassification* terhadap cara kerja algoritme *Gradient Boost Trees*. Untuk menjawab rumusan masalah tersebut, maka penulis akan mengemukakan permasalahan yang berkaitan dengan latar belakang, yaitu sebagai berikut:

1. Bagaimana penerapan *Random Under-sampling* pada klasifikasi kecurangan transaksi *mobile* dengan *Gradient Boost Tree*?
2. Bagaimana pengaruh *Random Under-sampling* dengan *Gradient Boost Tree* dalam menangani *misclassification* untuk meningkatkan akurasi algoritme?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Mengetahui cara penerapan *Random Under-sampling* pada klasifikasi kecurangan transaksi *mobile* dengan algoritme *Gradient Boost Trees*;
2. Mengetahui pengaruh *Random Under-sampling* terhadap *Gradient Boost Tree* dalam menangani *misclassification* untuk peningkatan akurasi.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Mengerti cara kerja algoritme *Gradient Boost Tree* untuk klasifikasi kecurangan transaksi *mobile*;
2. Memahami peranan *Random Under-sampling* yang berfungsi sebagai penanganan atas data yang tidak seimbang;
3. Mampu mengaplikasikan metode *Random Under-sampling* dengan *Gradient Boost Tree* untuk klasifikasi kecurangan transaksi *mobile*.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Data diambil dari situs *Kaggle*¹ yang merupakan data sintetik dibidang finansial yang membahas tentang kecurangan transaksi secara *mobile*.
2. Format berkas yang dipakai hanya *.csv*

1.7 Sistematika Penulisan

Sistematika penulisan skripsi dijelaskan dibawah ini.

BAB I. PENDAHULUAN

Bab I menjelaskan tentang latar belakang penelitian, perumusan masalah penelitian, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, metodologi penelitian, dan sistematika penulisan.

¹ <https://www.kaggle.com>

BAB II. KAJIAN LITERATUR

Bab II menjelaskan seluruh dasar-dasar teori yang dipakai dalam penelitian, algoritme yang digunakan, serta penelitian sebelumnya yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab III menjelaskan semua tahapan yang digunakan pada penelitian. Mulai dari pengumpulan data yang dipakai pada penelitian, tahapan rencana penelitian dijelaskan secara rinci yang merujuk pada suatu kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab IV menjelaskan keseluruhan proses pengembangan perangkat lunak dengan metode RUP yang dipakai sebagai alat dalam penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab V menjelaskan hasil dari pengujian yang telah dilakukan berdasarkan urutan Langkah – Langkah yang telah direncanakan pada Bab sebelumnya serta menganalisa dan membandingkan hasil dari setiap skenario yang menjadi landasan kesimpulan untuk penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab VI menjelaskan semua kesimpulan dan saran dalam penelitian ini yang berguna untuk penelitian selanjutnya.

1.8 Kesimpulan

Bab I telah menjelaskan secara global tentang masalah – masalah pada latar belakang dalam penelitian ini. Bab I ini juga telah dijelaskan terkait rumusan masalah penelitian, tujuan dari penelitian, manfaat dari penelitian, batasan – batasan masalah pada penelitian, dan sistematika penulisan pada penelitian.

DAFTAR PUSTAKA

- Amra, I. A. A., & Maghari, A. Y. (2017, May). Students performance prediction using KNN and Naïve Bayesian. In *2017 8th International Conference on Information Technology (ICIT)* (pp. 909-913). IEEE.
- Azwanti, N. (2018). Analisa Algoritma C4. 5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning. *Jurnal Ilmiah Ilmu Komputer, Informatika Mulawarman*, 13(1), 33-38.
- Bisri, A., & Rachmatika, R. (2019). Integrasi Gradient Boosted Trees dengan SMOTE dan Bagging untuk Deteksi Kelulusan Mahasiswa. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(4), 309. <https://doi.org/10.22146/jnteti.v8i4.529>
- Botchey, F. E., Qin, Z., & Hughes-Lartey, K. (2020). Mobile money fraud prediction-A cross-case analysis on the efficiency of support vector machines, gradient boosted decision trees, and Naïve Bayes algorithms. *Information (Switzerland)*, 11(8). <https://doi.org/10.3390/INFO11080383>
- Coppolino, L., D'Antonio, S., Formicola, V., Massei, C., & Romano, L. (2015). Use of the Dempster–Shafer theory to detect account takeovers in mobile money transfer services. *Journal of Ambient Intelligence and Humanized Computing*, 6(6), 753–762. <https://doi.org/10.1007/s12652-015-0276-9>
- Cynthia, E. P., & Ismanto, E. (2018). Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 3(July), 1. <https://doi.org/10.30645/jurasik.v3i0.60>
- Düntsche, I., & Gediga, G. (2019). Confusion Matrices and Rough Set Data Analysis. *Journal of Physics: Conference Series*, 1229(1). <https://doi.org/10.1088/1742-6596/1229/1/012055>

- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, 98, 105–117. <https://doi.org/10.1016/j.eswa.2018.01.012>
- Jadhav, S. D., & Channe, H. P. (2016). Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research (IJSR)*, 5(1), 1842–1845. <https://doi.org/10.21275/v5i1.nov153131>
- Jian, C., Gao, J., & Ao, Y. (2016). A new sampling method for classifying imbalanced data based on support vector machine ensemble. *Neurocomputing*, 193, 115–122. <https://doi.org/10.1016/j.neucom.2016.02.006>
- Jiawei Han, M. K. and J. P. (2012). Data Mining: Concepts and Techniques, Third Edition - Books24x7. *Morgan Kaufmann Publishers*, 745. <http://library.books24x7.com/toc.aspx?bkid=44712>
- Kang, H. (2019). *Fraud Detection in Mobile Money Transactions Using Machine Learning*. *Fraud Detection in Mobile Money Transactions Using Machine Learning*.
- Krauss, C., & Huck, N. (2016). *Discussion Papers in Economics random forests : Statistical arbitrage on the S & P 500. 03*.
- Lopez-Rojas, E. A., & Barneaud, C. (2019). Advantages of the PaySim Simulator for Improving Financial Fraud Controls. *Advances in Intelligent Systems and Computing*, 998, 727–736. https://doi.org/10.1007/978-3-030-22868-2_51
- Rahman, M. A. (2015). Algoritma C45 Untuk Menentukan Mahasiswa Penerima Beasiswa (Studi Kasus : Pps Iain Raden Intan Bandar Lampung). *Jurnal Teknologi Informasi Magister Darmajaya*, 1(02), 118–128.

- Rajesh, K. N. V. P. S., & Dhuli, R. (2018). Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. *Biomedical Signal Processing and Control*, *41*, 242–254. <https://doi.org/10.1016/j.bspc.2017.12.004>
- Rani, L. N. (2016). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *INOVTEK Polbeng - Seri Informatika*, *1*(2), 126. <https://doi.org/10.35314/isi.v1i2.131>
- Singh, L., Kaur, N., & Chetty, G. (2018). Customer Life Time Value Model Framework Using Gradient Boost Trees with RANSAC Response Regularization. *Proceedings of the International Joint Conference on Neural Networks, 2018-July*. <https://doi.org/10.1109/IJCNN.2018.8489710>
- Syukron, A., & Subekti, A. (2018). Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit. *Jurnal Informatika*, *5*(2), 175–185. <https://doi.org/10.31311/ji.v5i2.4158>
- Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, *58*, 308–324. <https://doi.org/10.1016/j.trc.2015.02.019>
- Zhou, B., Lu, M., & Wang, Y. (2016). Counting people using gradient boosted trees. *Proceedings of 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2016*, 391–395. <https://doi.org/10.1109/ITNEC.2016.7560388>