



**MENDETEKSI ANOMALI PADA DATA TRANSAKSI JUAL BELI AKTA TANAH
DENGAN MENGGUNAKAN ALGORITMA DECISION TREE, ISOLATION
FOREST, DAN RANDOM FOREST**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer

Oleh:

Yonathan Hartoko Dwi Ariputro

41517120076

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022



**MENDETEKSI ANOMALI PADA DATA TRANSAKSI JUAL BELI AKTA
TANAH DENGAN MENGGUNAKAN ALGORITMA DECISION TREE,
ISOLATION FOREST, DAN RANDOM FOREST**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Yonathan Hartoko Dwi Aripuro

41517120076

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

LEMBAR PERNYATAAN ORISINALITAS

Yang bertandatangan dibawah ini:

NIM : 41517120076

Nama : Yonathan Hartoko Dwi Aripuro

Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 23 Agustus 2022



UNIVERSITAS Yonathan Hartoko Dwi Aripuro
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

NIM : 41517120076
Nama : Yonathan Hartoko Dwi Aripuro
Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

MERCU BUANA

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 23 Agustus 2022



Yonathan Hartoko Dwi Aripuro

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

NIM : 41517120076

Nama : Yonathan Hartoko Dwi Aripuro

Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	v
		Jurnal Nasional Terakreditasi	v		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Teknologi Informasi dan Komunikasi			
	ISSN	: 2549-8037			
	Link Jurnal	: https://ejournal.ikado.ac.id/index.php/teknika/index			
	Link File Jurnal Jika Sudah di Publish				

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 16 Juli 2022



Yonathan Hartoko Dwi Aripuro

LEMBAR PERSETUJUAN

NIM : 41517120076
Nama : Yonathan Hartoko Dwi Aripuro
Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Tugas Akhir ini telah diperiksa dan disetujui.

Jakarta, 16 Juli 2022



UNIVERSITAS
(Misbahul Fajri Dr. M.TI)
Dosen Pembimbing
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517120076
Nama : Yonathan Hartoko Dwi Aripuro
Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah
Dengan Menggunakan Algoritma Decision Tree, Isolation Forest,
dan Random Forest

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 23 Agustus 2022

Digitally signed by: SUKMA WARDHANA
Signed at: Aug 25, 2022 10:09:34

Sukma Wardhana
(Sukma Wardhana, S.Kom., M.Kom.)

Dosen Penguji

UNIVERSITAS
MERCU BUANA

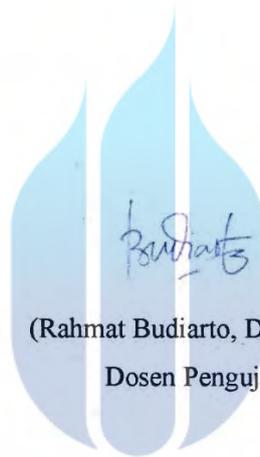


LEMBAR PERSETUJUAN PENGUJI

NIM : 41517120076
Nama : Yonathan Hartoko Dwi Aripuro
Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 23 Agustus 2022



(Rahmat Budiarto, Dr. Prof)
Dosen Penguji

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517120076
Nama : Yonathan Hartoko Dwi Aripuro
Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Agustus 2022



(Wawan Gupawan, S.Kom, MT)
Dosen Penguji

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41517120076
Nama : Yonathan Hartoko Dwi Aripuro
Judul Tugas Akhir : Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Agustus 2022



(Misbahul Fajri, Dr. M.TI)
Dosen Pembimbing



(Wawan Gunawan, S.Kom, MT)

Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)

Ka. Prodi Teknik Informatika

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Tuhan Yang Maha Esa, atas limpahan Rahmat dan Karunia-Nya, Sehingga penulis dapat menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa tanpa bantuan dan bimbingan Dosen Pembimbing dan Berbagai Pihak, Tugas Akhir ini tidak dapat terselesaikan hingga saat ini dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Orang Tua yang selalu memberikan doa dan dukungan penuh sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan semangat, baik dan lancar.
2. Bapak Dr. Misbahul Fajri, M.T.I selaku Dosen Pembimbing Tugas Akhir.
3. Bapak Ir. Emil R. Kaburuan, Ph.D., I.P.M. selaku Dosen Pembimbing Akademik dan Kepala Program Studi Teknik Informatika.
4. Bapak Wawan Gunawan, S.Kom, MT selaku Koordinator Tugas Akhir Teknik Informatika.
5. Oki Uwe Oviska, S.Kom selaku Kepala Divisi Software PT IP Network Solusindo.
6. Teman – Teman Mahasiswa Teknik Informatika Angkatan 2017 Universitas Mercu Buana.

Akhir kata, penulis berharap penulis menyadari bahwa Tugas Akhir ini masih jauh dari kata sempurna dikarenakan terbatasnya pengetahuan dan pengalaman yang dimiliki. Semoga melalui Tugas Akhir ini dapat memberikan manfaat dan menambah pengetahuan daripada penulis dan pembaca yang budiman.

Jakarta, 16 Juli 2022

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR ...	iii
LEMBAR PERSETUJUAN	v
LEMBAR PERSETUJUAN PENGUJI.....	vi
LEMBAR PENGESAHAN.....	ix
ABSTRAK.....	x
ABSTRACT	xi
KATA PENGANTAR	xii
DAFTAR ISI	xiii
NASKAH JURNAL.....	1
KERTAS KERJA	10
BAB 1. LITERATUR REVIEW	14
BAB 2. ANALISIS DAN PERANCANGAN.....	35
BAB 3. SOURCE CODE.....	40
BAB 4. DATASET	45
BAB 5. TAHAPAN EKSPERIMEN.....	48
BAB 6. HASIL DAN KESIMPULAN.....	57
DAFTAR PUSTAKA	60
LAMPIRAN DOKUMEN HAKI	63
LAMPIRAN SCAN KTP.....	64
LAMPIRAN KORESPONDENSI	65

Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest

Yonathan Hartoko^{1*}, Misbahul Fajri², Emil Robert Kaburuan³

^{1*}Program Studi Informatika, Universitas Mercu Buana, Meruya, Jakarta Barat

² Program Studi Informatika, Universitas Mercu Buana, Meruya, Jakarta Barat

³ Program Studi Informatika, Universitas Mercu Buana, Meruya, Jakarta Barat

Email: ^{1*}41517120076@student.mercubuana.ac.id, ² fajri@mercubuana.ac.id, ³ emil.kaburuan@mercubuana.ac.id

(Naskah masuk: 26 Agustus 2022)

Abstrak

Instansi pertanahan saat ini telah memanfaatkan Teknologi Informasi Komunikasi dalam menyimpan, mengolah dan memberikan pelayanan pertanahan kepada masyarakat sehingga menghasilkan database pertanahan sebagai alat pendukung pengambilan keputusan. Namun demikian terdapat beberapa permasalahan antara lain masih terdapat nilai akta abnormal pada berbagai daerah. Nilai akta ini dianggap tidak sesuai dengan nilai akta rata-rata pada daerah tersebut. Dari permasalahan tersebut muncul inisiatif untuk menemukan anomali pada data transaksi jual beli dengan menggunakan sebuah model machine learning yang dapat digunakan untuk mengklasifikasikan nilai yang tidak sesuai. Kertas kerja berjudul Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest yang berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat atau disertakan di artikel jurnal. Dalam kertas kerja ini akan dijelaskan mengenai literature review, dataset yang digunakan, serta langkah-langkah perancangan, tahapan implementasi dan hasil pengujian penelitian.

Kata Kunci: Classification, Outlier Detection, Decision-Tree, Isolation Forest, Random Forest, Tree-Based Classifier

Identifying Anomalies in Land Certificate Transaction Data using Decision Tree, Isolation Forest, and Random Forest Algorithm

Abstract

Currently, Land Agency had been utilize information technology and communication to keep, manage, and provide land services for the public so that resulting in land databases as decisions making tool. However there has been several issue found along the way such as land certificate value abnormality in several region. Those land certificate value had been recognize as contrasted as the average land certificate value in each region. Along with the issue, come into view the initiative for finding outliers in transactin data using a machine learning model that kan be utilized for classifying the contrasted value. This paper that has the title Identifying Anomalies in Land Certificate Transaction Data using Decision Tree, Isolation Forest, and Random Forest Algorithm that contains all materials regarding Thesis's research that didn't published in the article. In this paperwork, will be explained about literature review, used dataset, along with planing steps, implementation phase, and research report results.

Keywords: Classification, Outlier Detection, Decision-Tree, Isolation Forest, Random Forest, Tree-Based Classifier

I. PENDAHULUAN

Pendeteksian *outlier* direferensikan dari menemukan data yang tidak sesuai dengan pola normal pada sekumpulan data [8]. Tugas utama dari pengamatan *outlier* adalah untuk mengidentifikasi penyimpangan pada dataset [3]. Pengamatan *outlier* tidak dapat dikesampingkan karena dapat mempengaruhi pemrosesan lebih lanjut pada sebuah dataset [6]. Data yang dianggap sebagai noise merupakan subjek utama dalam sebuah *outlier*, dimana *outlier* dapat mengubah hasil perhitungan dataset secara ekstrim [6] dan dapat menuntun analisa data dan relasinya kearah kesimpulan yang salah [2]. *Outlier* dapat dianggap sebagai bentuk kejahatan juga, karena melalui *outlier* ini dapat diambil sebagai keuntungan dari berbagai modus kecurangan [4]. *Outlier* biasanya muncul karena diakibatkan oleh kesalahan pada tahap pengumpulan data [1]. Seperti pada penelitian [12], *outlier* muncul dikarenakan kesalahan pencatatan data.

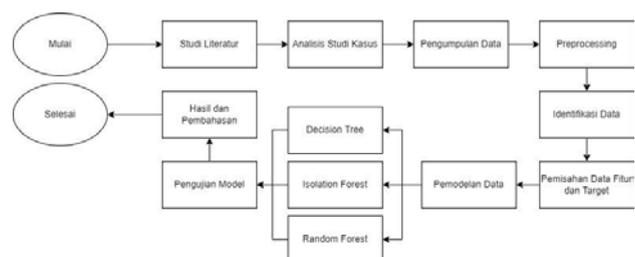
Outlier banyak ditemukan pada berbagai bidang seperti pada penelitian [10], [14], [27], *outlier* ditemukan pada lalu lintas jaringan, dimana *outlier* disini dikategorikan sebagai aktifitas menyimpang pada sebuah sistem yang dapat diidentifikasi dengan adanya tindakan peretasan. Pendeteksian *outlier* banyak digunakan untuk mengetahui adanya kecurangan, anomali pada data medis, perangkat *surveillance*, dan industri dengan IoT *big data* [8]. Kemudian *outlier* juga dapat ditemukan pada data yang menyangkut proses bisnis [12], [15], yang berdampak secara langsung pada kerugian yang dapat dialami oleh suatu perusahaan. Kasus ini juga dialami oleh Instansi Pertanahan. *Outlier* yang terdapat pada nilai akta dapat menyebabkan nilai akta rata-rata pada suatu daerah menjadi tidak pasti hal ini dapat berdampak pada kerugian negara di bidang ekonomi.

Pendeteksian *outlier* dapat dilakukan berbagai cara, salah satunya menggunakan *tools*. Pada penelitian [10], *outlier* dideteksi dengan menggunakan tool *Security Information and Event Management*(SIEM) yang disediakan oleh platform *Splunk*. Kemudian pada penelitian [14], digunakan tool *Intrusion Detection System* untuk mendeteksi adanya aktivitas mencurigakan/*outlier*. Sedangkan pada penelitian ini, pendeteksian *outlier* dilakukan dengan melakukan pengujian model yang dibangun dari algoritma. Berbagai Teknik yang dapat digunakan antara lain klasifikasi, prediksi, estimasi, asosiasi, dan pengelompokan [16]. Pemodelan yang dilakukan dapat membantu untuk menandai penyimpangan yang terjadi pada sebuah data [9].

Penggunaan algoritma pendeteksian *outlier* dapat membandingkan suatu pengamatan dengan pengamatan dengan metode lain pada sebuah dataset[6]. Menurut Pasek Agus Setiawan pada penelitian [16], studi perbandingan metode deteksi *outlier* ini dilakukan untuk mengetahui mana dari algoritma deteksi *outlier* yang lebih dapat diterapkan

pada suatu dataset. Pada penelitian [14], metode perbandingan yang dibandingkan adalah algoritma Naïve Bayes dan Support Vector Machine(SVM) dan penelitian [16] membandingkan algoritma K-means biasa dengan K-means yang dikombinasikan dengan Local *Outlier Factor*(LOF). Penelitian ini melakukan perbandingan antara algoritma Decision Tree, Isolation Forest, dan Random Forest untuk mengenali *outlier* pada dataset akta tanah. Dari sekian banyak algoritma klasifikasi, algoritma klasifikasi berbasis tree memiliki kehandalan dalam mendeteksi *outlier*. Klasifikasi sendiri sebenarnya sering digunakan untuk mengidentifikasi pola tidak biasa dan dapat mengurangi tingkat salah prediksi [22]. Secara teoritis, penelitian [7] menyimpulkan bahwa algoritma berbasis tree mendekati distribusi probabilitas yang tinggi dengan rate error yang rendah dengan kemampuan mengisolasi *outlier* yang ditemukan pada garis edar yang cukup singkat. Hal ini membuktikan bahwa metode berbasis *tree* sangatlah cocok untuk melakukan estimasi percampuran pada distribusi data yang didalamnya mengandung *outlier*. Mengacu juga pada penelitian [2], [7] yang menggunakan 3 algoritma, disimpulkan bahwa Isolation Forest menunjukkan hasil yang paling baik dalam mendeteksi *outlier*. Isolation Forest merupakan algoritma yang dibangun dari beberapa algoritma trees [19]. Sedangkan pada penelitian [27], Decision Tree menunjukkan akurasi 100% untuk mendeteksi *outlier* pada dataset serangan siber. Decision Tree dapat memperkirakan properti yang tidak diketahui dengan mempertanyakan berturut-turut mengenai properti yang sudah diketahui [18] karena Decision Tree memiliki struktur yang memiliki alur yang mengikuti aturan IF-THEN [25]. Kemudian penelitian [5] menunjukkan algoritma pemodelan Random Tree dan Random Forest menghasilkan akurasi yang paling baik dan keakurasian yang stabil pada berbagai macam dataset. Random Forest ini dapat menciptakan sebuah *forest* dari subset mandiri dari sebuah dataset [21]. Pada penelitian lain [24], [26], dilakukan penggabungan antara beberapa algoritma untuk menghasilkan klasifikasi yang lebih baik. Hasil pendeteksian *outlier* ini dapat digunakan sebagai peramalan akurat terhadap suatu dataset lain untuk memprediksi adanya kemunculan *outlier* lain [15]. Dimana dijelaskan melalui penelitian tersebut, dapat digunakan sebagai pengambilan keputusan.

II. METODOLOGI PENELITIAN



Gambar 1 Diagram Tahapan Penelitian

A. Studi Literatur

Pada proses ini, peneliti melakukan pengumpulan literatur yang berkaitan dengan topik bahasan yang akan diteliti dari berbagai macam sumber literatur seperti situs jurnal internasional hingga situs jurnal pada beberapa instansi pendidikan. Literatur yang digunakan dapat membantu penelitian agar lebih berdasar dan memudahkan proses penelitian baik pada saat pengumpulan data, pemrosesan, hingga penarikan kesimpulan hasil penelitian.

B. Analisis Studi Kasus

Analisis studi kasus dibutuhkan untuk menemukan masalah yang terjadi di lapangan terkait dengan penyimpangan [12], [15]. Dalam fase analisis studi kasus, peneliti melakukan survei dan wawancara di salah satu instansi pertanahan dengan tim teknis maupun pemegang keputusan mengenai topik yang sedang diteliti. Peneliti melakukan diskusi dengan tim teknis dan pemegang keputusan terkait pendeteksian *outlier* pada database pertanahan.

C. Pengumpulan Data

Dalam penelitian *outlier*, berbagai sumber dapat digunakan untuk mendapatkan dataset yang memiliki *outlier*, seperti sistem *surveillance* [6], [9], data historis [8], [13], hingga website [15], [17]. Pada penelitian ini pengeumpulan data dilakukan melalui tahap survey dan wawancara. Peneliti mulai menggali data yang dianggap memiliki atribut yang menyimpang bersama dengan tim teknis, melakukan perhitungan mengenai metode pencarian data yang memiliki atribut menyimpang. Untuk penelitian yang dilakukan ini, peneliti diberikan sejumlah dataset oleh pihak salah satu instansi pertanahan yang sudah dibersihkan sehingga data yang akan diproses tidak memiliki unsur konfidensial.

D. Preprocessing Data

Preprocessing Data merupakan bentuk pemurnian dan generalisasi data yang merupakan tahap penting dalam analisa dataset [2]. Pada tahap preprocessing, dilakukan data cleaning. Data Cleaning ini mencakup proses penghapusan atribut yang tidak perlu digunakan sebagai data fitur, melakukan konversi data yang bertipe objek atau string menjadi number, dan memilah data berdasarkan Tahun. Karena keterbatasan jumlah data, peneliti mengambil data paling banyak, yaitu data tahun 2018.

E. Identifikasi Data

Tahap pertama dalam melakukan pendeteksian anomali adalah dengan analisa eksplorasi data pada masing-masing variabel [1]. Untuk membentuk pemodelan dibutuhkan data yang memiliki atribut yang dapat digunakan sebagai data target. Pada penelitian [12], atribut anomali ditemukan

terlebih dahulu dengan melakukan perhitungan selisih antar bulan. Dalam penelitian [11] juga dilakukan analisa terhadap hubungan antar atribut yang ada untuk dapat menarik kesimpulan yang akan digunakan untuk proses klasifikasi. Sedangkan pada penelitian [27], atribut anomali diidentifikasi terlebih dahulu dengan menghitung *gini index*. Identifikasi data pada penelitian ini dilakukan dengan melakukan perhitungan terhadap nilai akta dan luas terhitung. Dengan menghitung kedua data dengan atribut tersebut, nilai akta permeter dapat ditemukan. Kemudian dari situ peneliti dapat mengidentifikasi distribusi data dengan menggunakan standar deviasi. Standar deviasi adalah nilai statistik yang dimanfaatkan untuk menentukan bagaimana sebaran data dalam sampel, serta seberapa dekat titik data individu ke mean atau rata-rata nilai sampel. Varian dan simpangan baku merupakan ukuran-ukuran variasi yang sering dipakai oleh semua orang. Dasar penghitungan varian dan simpangan baku merupakan tujuan untuk mengetahui variasi dari kelompok data[28]. Rumus dari standar deviasi adalah[28]:

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

S = Standar Deviasi

x_i = Nilai x ke i

\bar{x} = Nilai rata – rata data

n = Jumlah data

Untuk menghitung standar deviasi ini digunakan perhitungan nilai mean, median, min, dan max. Setelah standar deviasi ditemukan, digunakan asumsi simpangan sebesar 25% dari kuartil atas dan kuartil bawah. Untuk menentukannya, data dibagi dalam 3 kuartil dimana kuartil 1 adalah nilai $\frac{1}{4}$ dari jumlah data, kuartil 2 adalah nilai tengah, dan kuartil 3 adalah $\frac{3}{4}$ dari jumlah data. Apabila nilai akta permeter lebih kecil dari batas bawah, data tersebut diidentifikasi sebagai anomali, dan apabila nilai akta permeter lebih besar dari batas atas, maka data tersebut juga diidentifikasi sebagai anomali.

F. Pemisahan Data Fitur dan Target

Pada tahap ini peneliti melakukan pembagian dataset yang telah diolah dengan menghitung nilai akta kedalam dua kelompok yaitu data fitur dan target. Data fitur memiliki atribut nilai akta, luas terhitung, dan nilai permeter yang bertipe number. Sedangkan data target memiliki atribut *isAnomaly* yang bertipe boolean. Tidak semua fitur digunakan untuk mengidentifikasi dataset karena tidak semua fitur dianggap relevan untuk digunakan pada pemoelan data [27]. Seleksi fitur ini dibutuhkan untuk mengurangi dimensi data dan menjamin performa yang lebih baik [20]. Pada penelitian [16], fitur yang digunakan adalah nomor induk

pegawai, nilai perilaku, nilai kehadiran, dan nilai kerja. Fitur ini akan digunakan sebagai mengenali pola normal untuk dapat membedakan antara data normal dan data tidak normal [13]. Pada penelitian lain [23], seleksi fitur ini dapat menggunakan metode khusus seperti Correlation-Based Feature Selection (CFS) atau Particle Swarm Organization (PSO) yang juga dapat membantu memudahkan seleksi fitur apabila terdapat dataset dengan fitur yang cukup banyak.

G. Pemodelan Data

Pemodelan data pada penelitian ini menggunakan 3 algoritma berbeda yaitu:

1. Algoritma Decision Tree
2. Algoritma Isolation Forest
3. Algoritma Random Forest

Sebelum dilakukan pemodelan data dibagi kedalam 2 kelompok yaitu data training dan testing [14]. Untuk train, data mengandung 70% dari jumlah data secara keseluruhan, sedangkan untuk test, data mengandung 30% dari jumlah data secara keseluruhan [17].

H. Pengujian Model

Pada tahap pengujian model, peneliti menggunakan train data sebagai inputan pemodelan, kemudian dimasukan kedalam algoritma yang digunakan untuk menghasilkan *outlier* yang diidentifikasi oleh masing-masing algoritma. Evaluasi dilakukan untuk membandingkan jumlah *outlier* pada dataset [16] yang dideteksi oleh masing-masing algoritma. Setelah itu hasil pengujian algoritma untuk mengetahui *outlier* pada dataset, divisualisasikan menggunakan grafik plot untuk mendapatkan persebaran data *outlier* [14] dan data normal.

I. Hasil dan Pembahasan

Hasil dari penelitian merupakan pembahasan mengenai perbandingan ketiga algoritma yang digunakan dalam pengujian untuk mendeteksi *outlier* pada dataset nilai akta, manakah algoritma yang cocok digunakan untuk menemukan *outlier* pada studi kasus nilai akta.

III. HASIL DAN PEMBAHASAN

Data yang didapatkan merupakan data mentah yang dituangkan dalam bentuk CSV dengan 7 atribut seperti gambar 5.1 dibawah ini.

TANGGAL	KELURAHAN	NILAIAKTA	TAHUN	BULAN	LUASTERTULIS	LUASTERHITUNG
13/10/2017 00.00	Cisaranten Kulon	460.000.000	2017	10	78	77.99
09/06/2017 00.00	Cisaranten Kulon	550.000.000	2017	6	197	197.611
13/04/2018 00.00	Cisaranten Kulon	500.000.000	2018	4	186	186
12/04/2018 00.00	Cisaranten Kulon	0	2018	4	120	120.001
15/06/2017 00.00	Cisaranten Kulon	140.000.000	2017	6	100	100

Gambar 2 Dataset dalam format CSV

Kemudian data diimport kedalam program dengan menggunakan library Panda.

```
from io import StringIO
import pandas as pd
import numpy as np

file_tanah = "C:/Users/John/Kuliah/TA/dokumentanah.csv"
dt = pd.read_csv(file_tanah, sep=';')

dt.head()
```

	TANGGAL	KELURAHAN	NILAIAKTA	TAHUN	BULAN	LUASTERTULIS	LUASTERHITUNG
0	13/10/2017 00.00	Cisaranten Kulon	460.000.000	2017	10	78	77.99
1	09/06/2017 00.00	Cisaranten Kulon	550.000.000	2017	6	197	197.611
2	13/04/2018 00.00	Cisaranten Kulon	500.000.000	2018	4	186	186
3	12/04/2018 00.00	Cisaranten Kulon	0	2018	4	120	120.001
4	15/06/2017 00.00	Cisaranten Kulon	140.000.000	2017	6	100	100

Gambar 3 Import data

Data masih dalam bentuk mentah, sehingga saat data diimport, sebagian besar data angka masih dalam bentuk teks inputan seperti pada nilai akta 550.000.000 dan juga dapat dilihat nilai desimal pada luas terhitung menggunakan anotasi titik(.) seperti 197.611. Peneliti melakukan perubahan format nilai akta dengan menghapus titik dan pembulatan luas terhitung menjadi bilangan non-desimal.

```
In [5]: dt['LUASTERHITUNG'] = dt['LUASTERHITUNG'].str.extract(r'(\d{1,3})').squeeze().str.zfill(3)
dt['NILAIAKTA'] = dt['NILAIAKTA'].str.replace('.', '')
```

Gambar 4 Normalisasi data

Selanjutnya, atribut dataset dengan value yang tidak digunakan dalam penelitian ini dihapus agar tidak menambahkan unsur noise pada dataset seperti luas tertulis karena sudah menggunakan luas terhitung), tanggal, dan bulan karena penelitian dilakukan berdasarkan data tahunan.

```
In [6]: dt = dt.drop(['LUASTERTULIS', 'TANGGAL', 'BULAN'], axis=1)
dt = dt.dropna()
```

Gambar 5 Penghapusan kolom yang tidak digunakan

Kemudian sebagian besar data yang akan digunakan dalam penelitian masih dalam bentuk objek atau string, pada tahap ini peneliti mengkonversi tipe data nilai akta dan luas terhitung menjadi numerik dan tahun menjadi string.

```
dt['NILAIAKTA'] = dt['NILAIAKTA'].astype('int64')
dt['LUASTERHITUNG'] = dt['LUASTERHITUNG'].astype('int64')
dt['TAHUN'] = dt['TAHUN'].astype('str')
```

```
dt.head()
```

	KELURAHAN	NILAIAKTA	TAHUN	LUASTERHITUNG
0	Cisaranten Kulon	460000000	2017	77
1	Cisaranten Kulon	550000000	2017	197
2	Cisaranten Kulon	500000000	2018	186
4	Cisaranten Kulon	140000000	2017	100
5	Cisaranten Kulon	2319000	2017	99

Gambar 6 Pengubahan tipe data

Setelah itu, peneliti melakukan pengelompokan berdasarkan tahun. Didapatkan 3 kelompok dalam proses tersebut, yaitu tahun 2017, 2018, dan 2019.

```
dt['TAHUN'].value_counts()
```

```
2018    357
2017    203
2019    171
Name: TAHUN, dtype: int64
```

Gambar 7 Pengelompokan data berdasarkan tahun

Karena tahun 2018 memiliki data terbanyak, maka peneliti akan menggunakan data tahun 2018 sebagai data tahunan yang akan diproses.

```
thn2018 = dt.groupby(dt['TAHUN'])
dt2018 = thn2018.get_group('2018')
```

```
dt2018.head()
```

	KELURAHAN	NILAIAKTA	TAHUN	LUASTERHITUNG
2	Cisaranten Kulon	500000000	2018	186
12	Cisaranten Kulon	480000000	2018	132
19	Cisaranten Kulon	300000000	2018	100
22	Cisaranten Kulon	230000000	2018	100
23	Cisaranten Kulon	900000000	2018	340

Gambar 8 Pengambilan data tahun 2018

Setelah data dikelompokkan dan diambil data tahun 2018, peneliti melakukan perhitungan terhadap nilai akta dan luas terhitung untuk mengetahui nilai permeter dari akta tersebut.

```
dt2018['NILAIpermeter'] = dt2018['NILAIAKTA']/dt2018['LUASTERHITUNG']
dt2018['NILAIpermeter'] = dt2018['NILAIpermeter'].apply(lambda x: '{:.2f}'.format(x))
dt2018['NILAIpermeter'] = dt2018['NILAIpermeter'].str.extract(r'(\d{3,7})').squeeze().str.zfill(7)
dt2018['NILAIpermeter'] = dt2018['NILAIpermeter'].astype('int64')
```

```
dt2018.head(5)
```

	KELURAHAN	NILAIAKTA	TAHUN	LUASTERHITUNG	NILAIpermeter
0	Cisaranten Kulon	500000000	2018	186	2688172
1	Cisaranten Kulon	480000000	2018	132	3636363
2	Cisaranten Kulon	300000000	2018	100	3000000
3	Cisaranten Kulon	230000000	2018	100	2300000
4	Cisaranten Kulon	900000000	2018	340	2647058

Gambar 9 Perhitungan nilai permeter

Pada tahap berikutnya, data yang memiliki nilai atribut null atau nol(0) dihapus dari dataset.

```
columns = ['NILAIpermeter']
dt2018 = dt2018.replace(0, np.nan).dropna(axis=0, how='any', subset=columns).fillna(0)
```

Gambar 10 Penghapusan nilai null

Dataset perlu dilakukan pengurutan berdasarkan nilai permeter dari akta untuk memudahkan penentuan nilai kuartil.

```
dt2018 = dt2018.sort_values(by=['NILAIpermeter'])
dt2018['NILAIpermeter'] = dt2018['NILAIpermeter'].astype('int64')
```

Gambar 11 Pengurutan data

Lalu dicari mean dan median dari nilai permeter yang telah dihitung.

```
Pencarian nilai Mean
```

```
rht2018 = int(dt2018['NILAIpermeter'].mean(axis=0, skipna=True))
rht2018
```

```
3067538
```

```
Pencarian nilai Median
```

```
mht2018 = int(dt2018['NILAIpermeter'].median(axis=0, skipna=True))
mht2018
```

```
2689119
```

Gambar 12 Pencarian nilai Mean dan Median

Untuk menemukan nilai kuartil pada dataset, peneliti perlu mencari nilai minimum dan nilai maksimum dari nilai permeter.

```
Pencarian nilai tertinggi dari dataset

maxht2018 = int(dt2018['NILAIPERMETER'].max(axis=0, skipna=True))
maxht2018
9500000

Pencarian nilai terendah dari dataset

minht2018 = int(dt2018['NILAIPERMETER'].min(axis=0, skipna=True))
minht2018
5714
```

Gambar 13 Pencarian nilai maksimum dan minimum

Untuk kuartil 2, digunakan nilai tengah. Kuartil 1 ditentukan dengan menghitung kuantil 0.25 dari dataset yang telah diurutkan. Untuk kuartil 3, dataset dihitung dengan kuantil 0.75.

```
Pencarian nilai q1, q2, dan q3

q2_2018 = mht2018
q1_2018 = int(dt2018['NILAIPERMETER'].quantile(0.25))
q3_2018 = int(dt2018['NILAIPERMETER'].quantile(0.75))

q1_2018
1699757

q3_2018
4184432
```

Gambar 14 Pencarian Q1, Q2, dan Q3

Setelah nilai kuartil ditemukan, maka Langkah selanjutnya adalah perhitungan batas atas dan batas bawah. Perhitungan batas atas ditentukan dengan menghitung simpangan kuartil 3 sebesar 25% dan untuk perhitungan batas bawah dilakukan dengan menghitung simpangan kuartil 1 sebesar 25% pula.

```
Menghitung batas bawah

qb_2018 = int((q1_2018 * 75) / 100)
qb_2018
1274817

Menghitung batas atas

qa_2018 = int((q3_2018 * 125) / 100)
qa_2018
5230540
```

Gambar 15 Pencarian batas atas dan batas bawah

Kemudian penilaian atribut anomali dilakukan berdasarkan batas atas dan batas bawah yang telah dihitung.

Apabila nilai permeter lebih kecil dari batas bawah atau lebih besar dari batas atas, maka data akan diberikan nilai anomali. Setelah nilai atribut anomali ditemukan pada keseluruhan dataset, didapatkan data dengan nilai anomali berjumlah 91 data dan 265 data tidak anomali atau normal.

```
Penentuan nilai Anomali(True or False) berdasarkan batas atas dan batas bawah

dt2018['isAnomaly'] = dt2018['NILAIPERMETER'].apply(lambda x: 'True' if x < qb_2018 or x > qa_2018 else 'False')

dt2018.head(5)

```

	KELURAHAN	NILAI AKTA	TAHUN	LUASTERHITUNG	NILAIPERMETER	isAnomaly
40	Cisaranten Kulon	10000000	2018	175	5714	True
42	Cisaranten Kulon	4500000	2018	300	15000	True
328	Cisaranten Kulon	15000000	2018	300	50000	True
179	Cisaranten Kulon	190000000	2018	399	476190	True
207	Cisaranten Kulon	200000000	2018	399	501253	True

```
dt2018['isAnomaly'].value_counts()
False    265
True      91
Name: isAnomaly, dtype: int64
```

Gambar 16 Penentuan nilai anomali

Sebelum dilakukan pemodelan, data terlebih dahulu dibagi kedalam data fitur dan target. Data fitur memiliki atribut nilai akta, luas terhitung, dan nilai permeter. Sedangkan data target memiliki atribut isAnomaly saja.

```
dt2018.head()

```

	KELURAHAN	NILAI AKTA	TAHUN	LUASTERHITUNG	NILAIPERMETER	isAnomaly
40	Cisaranten Kulon	10000000	2018	175	5714	True
42	Cisaranten Kulon	4500000	2018	300	15000	True
328	Cisaranten Kulon	15000000	2018	300	50000	True
179	Cisaranten Kulon	190000000	2018	399	476190	True
207	Cisaranten Kulon	200000000	2018	399	501253	True

```
target2018 = pd.DataFrame(data = dt2018['isAnomaly'])
features2018 = dt2018.drop(['KELURAHAN', 'TAHUN', 'isAnomaly'], axis=1)
```

Gambar 17 Pemisahan data target dan fitur

Pada tahap pemodelan, dataset kemudian dibagi lagi menjadi data train dan data test. Data train dipakai untuk model mengenai pola dataset dan data test digunakan oleh model untuk menguji pola yang telah dipelajari. Data train berisi 70% dari jumlah data dan data test 30% dari jumlah data.

```
test_size = 0.3
random_state = 42
X_train, X_test, y_train, y_test = train_test_split(features2018, target2018, test_size=test_size, random_state=random_state)
```

Gambar 18 Pembagian data train dan data test

Pemodelan dilakukan dengan menggunakan algoritma Decision Tree, Isolation Forest, dan Random Forest. Data train dimasukkan kedalam masing-masing fungsi model untuk dilakukan pemrosesan.

Pemodelan Decision Tree

```
clf = DecisionTreeClassifier()
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
```

Gambar 19 Pemodelan menggunakan Decision Tree

Pemodelan Isolation Forest

```
clf = IsolationForest()
clf = clf.fit(X_train,y_train)
```

Gambar 20 Pemodelan menggunakan Isolation Forest

Pemodelan Random Forest

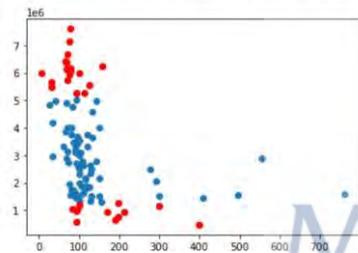
```
clf = RandomForestClassifier()
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
```

Gambar 21 Pemodelan menggunakan Random Forest

Setelah pemodelan dilakukan, peneliti melakukan visualisasi terhadap hasil pemodelan untuk mengidentifikasi persebaran data normal dan data anomali dengan menggunakan grafik plot. Maka diketahui persebaran data sebagai berikut untuk masing-masing model.

Visualisasi persebaran data berdasarkan pemodelan Decision Tree

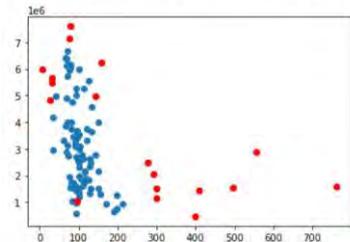
```
plt.scatter(X_test["LUASTERHITUNG"], X_test["NILAIPEMETER"])
plt.scatter(outlier_values["LUASTERHITUNG"], outlier_values["NILAIPEMETER"], c = "r")
```



Gambar 22 Visualisasi persebaran data Decision Tree

Visualisasi persebaran data berdasarkan pemodelan Isolation Forest

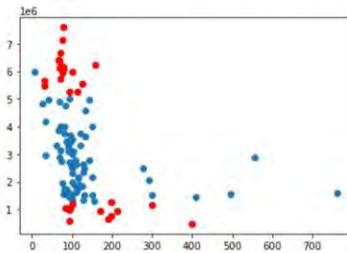
```
plt.scatter(X_test["LUASTERHITUNG"], X_test["NILAIPEMETER"])
plt.scatter(outlier_values["LUASTERHITUNG"], outlier_values["NILAIPEMETER"], c = "r")
```



Gambar 23 Visualisasi persebaran data Isolation Forest

Visualisasi persebaran data berdasarkan pemodelan Decision Tree

```
plt.scatter(X_test["LUASTERHITUNG"], X_test["NILAIPEMETER"])
plt.scatter(outlier_values["LUASTERHITUNG"], outlier_values["NILAIPEMETER"], c = "r")
```



Gambar 24 Visualisasi persebaran data Random Forest

Pada grafik diatas, data dengan nilai normal ditandai dengan warna biru, sedangkan data dengan nilai anomali ditandai dengan warna merah.

Hasil akurasi yang didapatkan dengan menggunakan algoritma Decision Tree adalah sebesar 100%, Isolation Forest sebesar 57%, sedangkan Random Forest sebesar 99%.

Tingkat akurasi prediksi model Decision Tree

```
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
Accuracy: 1.0
```

Gambar 25 Akurasi prediksi Decision Tree

Tingkat akurasi prediksi model Isolation Forest

```
print("Accuracy percentage:", 100*list(data['y_pred']).count(-1)/(outliers_counter))
Accuracy percentage: 57.57575757575758
```

Gambar 26 Akurasi prediksi Isolation Forest

Tingkat akurasi prediksi model Random Forest

```
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
Accuracy: 0.9906542056074766
```

Gambar 27 Akurasi prediksi Random Forest

IV. KESIMPULAN DAN SARAN

Sehingga dapat disimpulkan melalui pemodelan menggunakan 3 algoritma yang diterapkan pada studi kasus pendeteksian outlier pada data nilai akta, Decision Tree memiliki akurasi prediksi tertinggi dengan presentasi 100%, kemudian disusul dengan Random Forest dengan akurasi 99% dan terakhir, algoritma Isolation Forest memiliki presentasi akurasi prediksi cukup rendah di 57%. Dari sini dapat diketahui bahwa ketiga algoritma ini selanjutnya dapat digunakan untuk studi kasus pendeteksian anomali/outlier pada data data nilai akta tanah dan dapat dikembangkan untuk menggunakan atribut lebih banyak.

Karena keterbatasan akses terhadap dataset pertanahan, untuk penelitian selanjutnya diharapkan dapat disediakan akses terhadap dataset pendukung seperti dataset Nilai Objek Pajak, sehingga klasifikasi anomaly dalam nilai akta tanah dapat lebih akurat dan berdasar.

REFERENSI

- [1] P. Dey, Z. Zhang and D. B. Dunson, "Outlier Detection for Multi-Network Data," p. 2, 2022.
- [2] T. Mecheva, "Outlier Detection in Traffic Data Set," pp. 1-5, 2021.
- [3] D. P and S. Abraham, "Fair Outlier Detection," pp. 1-3, 2020.
- [4] D. Rahmawati, "Deteksi Anomali Menggunakan Flow Control Patterns dan Fuzzy Regression di Terminal Peti Kemas," pp. 1-8, 2017.
- [5] K. A. Harris and V. , "Komparasi Performa Tree-Based Classifier Untuk Deteksi Anomali Pada Data Berdimensi Tinggi dan Tidak Seimbang," *Jurnal Media Informatika Budidarma*, vol. 6, pp. 370-377, 2022.
- [6] R. Bettinger, "Outlier Detection and Treatment," pp. 1-26, 2020.
- [7] S. Buschjager, P.-J. Honysz and K. Morik, "Randomized Outlier Detection with Trees," *International Journal of Data Science and Analytics*, pp. 91-103, 2020.
- [8] Z. Chen, Z. Peng, X. Zou and H. Sun, "Deep Learning Based Anomaly Detection for Multi-Dimensional Time Series: A Survey," pp. 73-86, 2022.
- [9] D. Hendrycks, M. Mazeika and T. Dietterich, "Deep Anomaly Detection with Outlier Exposure," pp. 1-8, 2019.
- [10] M. R. Kamal and M. A. Setiawan, "Deteksi Anomali dengan Security Information and Event Management (SIEM) Splunk pada Jaringan UII," pp. 1-5, 2020.
- [11] H. Lv, Chunyan Xu and Zhen Chui, "Global Information Guided Video Anomaly Detection," pp. 4679-4682, 2020.
- [12] A. Muhaimin, "Deteksi Anomali Pada Pemakaian Air Pelanggan PDAM Surya Sembada Kota Surabaya Menggunakan Algoritma Kohonen Self Organizing Maps (SOM) dan Local Outlier Factor (LOF)," pp. 1-9, 2018.
- [13] H. Park, J. Noh and B. Ham, "Learning Memory-Guided Normality for Anomaly Detection," pp. 14372-14379, 2020.
- [14] I. Riadi, "Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes dan Support Vector Machine(SVM)," pp. 17-23, 2019.
- [15] I. Thaib, G. Thabrani and S. Netsyah, "Peramalan dan Deteksi Outlier Saham Perusahaan Angkatan Laut Umum di Masa COVID-19 dengan pemodelan ARIMA," *Jurnal Kajian Manajemen dan Wirausaha*, vol. 3, pp. 8-17, 2021.
- [16] "Optimasi Data Pada Metode K-Means dengan Analisis Outlier," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 05, pp. 88-95, 2019.
- [17] F. Rahman, T. E. Sutanto and N. Fitriyati, "Web Traffic Anomaly Detection Using Stacked Long Short-Term Memory," *Indonesian Journal of Pure and Applied Mathematics*, vol. 3, pp. 112-121, 2021.
- [18] A. Criminisi, J. Shotton and E. Konukoglu, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Foundation and Trends in Computer Graphics and Vision*, vol. 7, pp. 81-227, 2011.
- [19] Alaa Abd Ali Hadi, "Performance Analysis of Big Data Intrusion Detection System over Random Forest Algorithm," *International Journal of Applied Engineering Research*, vol. 13, pp. 1520-1527, 2018.
- [20] N. Appaswamy, N. A. N. D H and P. D. Shenoy, "ERCR TV: Ensemble of Random Committee and Random Tree for Efficient Anomaly Classification Using Voting," 2018.
- [21] I. S. Atawodi, "A Machine Learning Approach to Network Intrusion Detection System Using K Nearest Neighbor and Random Forest," 2019.
- [22] S. Aljawarneh, M. B. Yassein and M. Aljundi, "An enhanced J48 classification algorithm for the anomaly intrusion detection systems," p. S10549-S10565, 2019.
- [23] M. N. Aziz, "Data preprocessing and feature selection for machine learning intrusion detection systems," 2019.
- [24] Y. Liu, Y. Wang, X. Ren, H. Zhou and S. Diao, "A Classification Method Based on Feature Selection for Imbalanced Data," *IEEE Access*, vol. 7, 2019.
- [25] E. Popoola and A. Adewumi, "Efficient feature selection technique for network intrusion detection system using discrete differential evolution and decision tree," *International Journal of Network Security*, vol. 19, pp. 660-669, 2017.
- [26] M. Aldwairi and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," 2017.
- [27] M. Fadhlurrohman, A. Muliawati and B. Hananto, "Analisis Kinerja Intrusion Detection System pada Deteksi Anomali dengan Metode Decision Tree Terhadap Serangan Siber," *Ilmu-Komputer Agri-Informatika*, vol. 8, pp. 91-93, 2021.
- [28] S. R. Setiawan, "Analisis Penerapan Standar Deviasi Dalam Penentuan Persediaan Pengaman Pada UD Mirama Kota Gorontalo," *Gorontalo Management Research*, vol. 1, p. 1, 2018.



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan kelengkapan artikel jurnal dengan judul Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, dan Random Forest yang berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat atau disertakan di artikel jurnal. Dalam kertas kerja ini akan dijelaskan mengenai literatur review, dataset yang digunakan, serta langkah-langkah perancangan, tahapan implementasi dan hasil pengujian penelitian.

Pendahuluan

Pendeteksian *outlier* direferensikan dari menemukan data yang tidak sesuai dengan pola normal pada sekumpulan data [8]. Tugas utama dari pengamatan *outlier* adalah untuk mengidentifikasi penyimpangan pada dataset [3]. Pengamatan *outlier* tidak dapat dikesampingkan karena dapat mempengaruhi pemrosesan lebih lanjut pada sebuah dataset [6]. Data yang dianggap sebagai noise merupakan subjek utama dalam sebuah *outlier*, dimana *outlier* dapat mengubah hasil perhitungan dataset secara ekstrim [6] dan dapat menuntun analisa data dan relasinya kearah kesimpulan yang salah [2]. *Outlier* dapat dianggap sebagai bentuk kejahatan juga, karena melalui *outlier* ini dapat diambil sebagai keuntungan dari berbagai modus kecurangan [4]. *Outlier* biasanya muncul karena diakibatkan oleh kesalahan pada tahap pengumpulan data [1]. Seperti pada penelitian [12], *outlier* muncul dikarenakan kesalahan pencatatan data.

Outlier banyak ditemukan pada berbagai bidang seperti pada penelitian [10][14][27], *outlier* ditemukan pada lalu lintas jaringan, dimana *outlier* disini dikategorikan sebagai aktifitas menyimpang pada sebuah sistem yang dapat diidentifikasi dengan adanya tindakan peretasan. Pendeteksian *outlier* banyak digunakan untuk mengetahui adanya kecurangan, anomali pada data medis, perangkat *surveillance*, dan industri dengan IoT *big data* [8]. Kemudian *outlier* juga

dapat ditemukan pada data yang menyangkut proses bisnis [12][15], yang berdampak secara langsung pada kerugian yang dapat dialami oleh suatu perusahaan. Kasus ini juga dialami oleh salah satu instansi pertanahan. *Outlier* yang terdapat pada nilai akta dapat menyebabkan nilai akta rata-rata pada suatu daerah menjadi tidak pasti hal ini dapat berdampak pada kerugian negara di bidang ekonomi.

Pendeteksian *outlier* dapat dilakukan berbagai cara, salah satunya menggunakan *tools*. Pada penelitian [10], *outlier* dideteksi dengan menggunakan tool *Security Information and Event Management*(SIEM) yang disediakan oleh platform *Splunk*. Kemudian pada penelitian [14], digunakan tool *Intrusion Detection System* untuk mendeteksi adanya aktivitas mencurigakan/*outlier*. Sedangkan pada penelitian ini, pendeteksian *outlier* dilakukan dengan melakukan pengujian model yang dibangun dari algoritma. Berbagai Teknik yang dapat digunakan antara lain klasifikasi, prediksi, estimasi, asosiasi, dan pengelompokan [16]. Pemodelan yang dilakukan dapat membantu untuk menandai penyimpangan yang terjadi pada sebuah data [9].

Penggunaan algoritma pendeteksian *outlier* dapat membandingkan suatu pengamatan dengan pengamatan dengan metode lain pada sebuah dataset [6]. Menurut Pasek Agus Setiawan pada penelitian [16], studi perbandingan metode deteksi *outlier* ini dilakukan untuk mengetahui mana dari algoritma deteksi *outlier* yang lebih dapat diterapkan pada suatu dataset. Pada penelitian [14], metode perbandingan yang dibandingkan adalah algoritma Naïve Bayes dan Support Vector Machine(SVM) dan penelitian [16] membandingkan algoritma K-means biasa dengan K-means yang dikombinasikan dengan Local Outlier Factor(LOF). Penelitian ini melakukan perbandingan antara algoritma Decision Tree, Isolation Forest, dan Random Forest untuk mengenali *outlier* pada dataset akta tanah. Dari sekian banyak algoritma klasifikasi, algoritma klasifikasi berbasis tree memiliki kehandalan dalam mendeteksi *outlier*. Klasifikasi sendiri sebenarnya sering digunakan untuk mengidentifikasi pola tidak biasa dan dapat mengurangi tingkat salah prediksi [22]. Secara teoritis, penelitian [7] menyimpulkan bahwa algoritma berbasis tree mendekati distribusi probabilitas yang tinggi dengan rate error yang rendah dengan kemampuan mengisolasi *outlier* yang ditemukan pada garis edar

yang cukup singkat. Hal ini membuktikan bahwa metode berbasis *tree* sangatlah cocok untuk melakukan estimasi percampuran pada distribusi data yang didalamnya mengandung *outlier*. Mengacu juga pada penelitian [2][7] yang menggunakan 3 algoritma, disimpulkan bahwa Isolation Forest menunjukkan hasil yang paling baik dalam mendeteksi *outlier*. Isolation Forest merupakan algoritma yang dibangun dari beberapa algoritma trees [19]. Sedangkan pada penelitian [27], Decision Tree menunjukkan akurasi 100% untuk mendeteksi *outlier* pada dataset serangan siber. Decision Tree dapat memperkirakan properti yang tidak diketahui dengan mempertanyakan berturut-turut mengenai properti yang sudah diketahui [18] karena Decision Tree memiliki struktur yang memiliki alur yang mengikuti aturan IF-THEN [25]. Kemudian penelitian [5] menunjukkan algoritma pemodelan Random Tree dan Random Forest menghasilkan akurasi yang paling baik dan keakurasian yang stabil pada berbagai macam dataset. Random Forest ini dapat menciptakan sebuah *forest* dari subset mandiri dari sebuah dataset [21]. Pada penelitian lain [24][26], dilakukan penggabungan antara beberapa algoritma untuk menghasilkan klasifikasi yang lebih baik. Hasil pendeteksian *outlier* ini dapat digunakan sebagai peramalan akurat terhadap suatu dataset lain untuk memprediksi adanya kemunculan *outlier* lain [15]. Dimana dijelaskan melalui penelitian tersebut, dapat digunakan sebagai pengambilan keputusan.

Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan diatas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana hasil pembentukan model data dengan menggunakan metode Decision Tree, Isolation Forest, dan Random Forest?
2. Bagaimanakah perbandingan efektifitas model data yang dibangun dengan metode Decision Tree dan Random Forest dalam menemukan anomali dalam sebuah data transaksi

Tujuan dan Manfaat

Tujuan

Tujuan dalam melakukan penelitian ini adalah:

1. Melakukan pengolahan pada data transaksi jual beli akta untuk menemukan anomali pada nilai akta dengan membuat sebuah model data
2. Melakukan perbandingan model data yang dibangun dari tiga algoritma machine learning yang berbeda yaitu antara Decision Tree, Isolation Forest, Random Forest.

Manfaat

Manfaat yang didapatkan melalui penelitian ini adalah:

1. Data transaksi jual beli akta dapat menjadi lebih valid
2. Mendapatkan simpulan dari hasil perbandingan tiga algoritma yang berbeda mengenai mana yang lebih efektif dalam menemukan anomali pada sebuah data transaksi

Batasan Masalah

Penelitian ini dilakukan untuk membandingkan pendeteksian anomali pada nilai akta yang dilakukan dengan tiga algoritma yang berbeda sehingga hasil yang didapatkan adalah sejauh mana ketiga algoritma tersebut mampu untuk mendeteksi anomali pada dataset yang diberikan.

UNIVERSITAS
MERCU BUANA