Mathematical Modeling
Department of Mathematics
Faculty of Mathematics and Natural Sciences
University of Indonesia

Analisis Tingkat Risiko Keamanan Data Cloud Berdasarkan Perilaku Pengguna

Dosen Fasilitator: Prof. Kiki Ariyanti, M.Si., Ph.D.


Raditya Fauzan


Subhan Irsyaduddien Alhaq


Khadijah Nurul Izzah


Irfan Hanif Yamashita

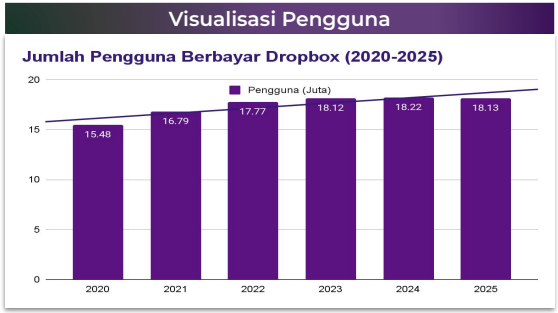

Muhammad Daffa


Muhammad Abyan Laksamana

Present by: Kelompok 2

Background

Keamanan Data Berdasarkan Perilaku Pengguna



KEAMANAN DATA

- Jaringan
- Enkripsi
- Perilaku

Perilaku yang tidak normal

- Aktivitas Login
- Akses file
- Perubahan konfigurasi

Indikasi awal

Keamanan Data

Background

Rumusan Masalah, Tujuan Penelitian, & Asumsi

Rumusan Masalah

1. Bagaimana menganalisis **risiko keamanan** data cloud berdasarkan **perilaku pengguna**?
2. Apa saja **jenis perilaku pengguna** yang berpengaruh terhadap tingkat keamanan data pada layanan cloud?

Tujuan Penelitian

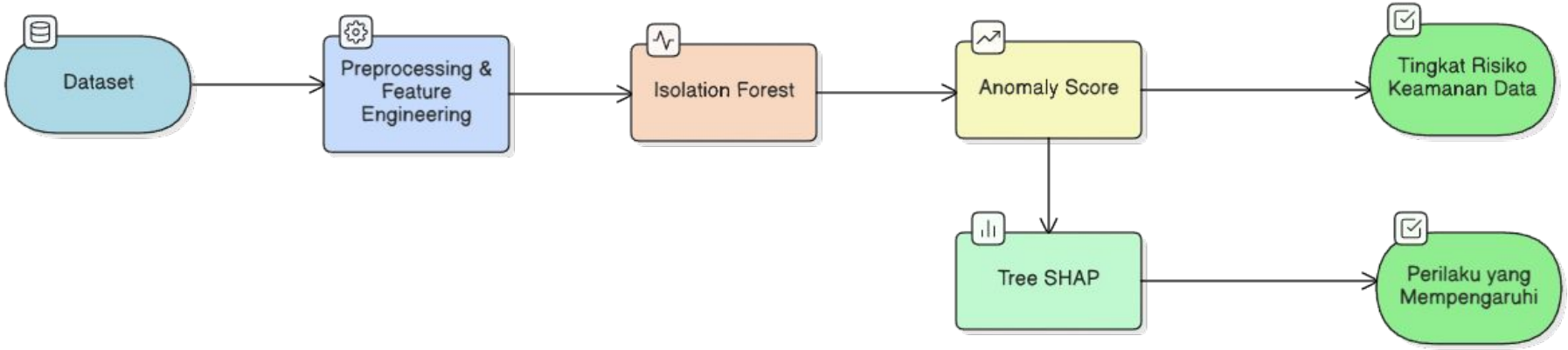
1. Menganalisis **risiko keamanan data** dengan anomali score **berdasarkan perilaku pengguna**.
2. Menentukan **jenis perilaku pengguna** yang berpengaruh terhadap tingkat keamanan data pada layanan cloud.

Batasan Masalah

1. Penelitian berfokus pada **Data Layanan penyimpanan Data Cloud**.
2. Penelitian hanya **berfokus** pada **perilaku pengguna** dalam menggunakan **layanan penyimpanan data cloud**, **bukan** pada **aspek teknis infrastruktur cloud** itu sendiri.

Background
Flow Chart

Flowchart Alur Pengerjaan

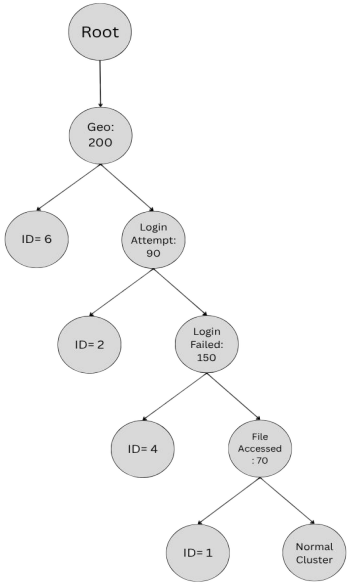


Modelling
Isolation Forest

Isolation Forest	**Isolation Forest** adalah teknik pendeteksian anomali dengan **mengisolasi anomali** dari data lainnya dengan **menggunakan Isolation Tree**. Isolation Forest melakukan pendekatan ***ensamble* Isolation Tree** dan **sub-sampling**.

Langkah-Langkah Isolation Tree

1. **Pilih sub-sample** data **secara acak**.
2. **Pilih fitur secara acak** dari data.
3. **Pilih titik belah** (split value) **secara acak pada range** dari fitur yang dipilih untuk data yang ada pada simpul tersebut.
4. **Bagi data menjadi dua *child nodes*** berdasarkan titik belah (split value).
5. **Ulangi proses** ini hingga:
   - **Tiap simpul daun** hanya **memiliki satu data**;
   - **Kedalaman maksimum** pohon **tercapai**.



Contoh Isolation Tree

Modelling
Anomaly Score

Anomaly Score

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

Rata-Rata Panjang Jalur

$$E(h(x)) = \frac{1}{t}\sum_{i=1}^{t} h_i(x)$$

Faktor Normalisasi

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}$$

$h(x)$	Panjang jalur dari titik x, dihitung dengan banyaknya busur yang menghubungkan titik x telah terisolasi dengan simpul akarnya.
x	Data dari subsample
t	Banyaknya tree yang terdapat titik x
$E(h(x))$	Rata-rata panjang jalur h(x) dari koleksi iTree

Deret harmonik, dimana $H(i) = \sum_{k=1}^{i} \frac{1}{k}$	$H(i)$
Banyaknya *sub-sample* data yang diambil secara acak dari data set yang kita punya	n
Ekspektasi panjang jalur rata-rata pada dataset	$c(n)$

Interpretasi

Jika nilai s(x,n) mendekati 1 maka Anomaly, s(x,n): 0,5<=x<1 mendekati anomali, s(x,n): 0<x<0,5 sebagai normal

Modelling

SHAP (SHapley Additive exPlanations)

Setelah kita mengetahui Anomaly Score dari x maka kita akan mencari tahu apa perilaku yang mempengaruhi Score Anomaly dengan menggunakan SHAP.

SHAP	**SHAP** atau **SHAPLEY ADDITIVE EXPLANATIONS** merupakan metode untuk menjelaskan *output* model *machine learning* dengan **mengukur pengaruh fitur terhadap output model**. Metode **SHAP** ini berasal dari *game theory* yang mengukur kontribusi tiap pemain dalam suatu permainan terhadap *outcome* permainan tersebut.

Rumus SHAP yang akan diImplementasikan pada Isolation Forest

treeSHAP

$$\phi_j = \frac{1}{T}\sum_{t=1}^{T}\phi_{j,t}$$

T	Menyatakan total pohon yang ada di dalam model ensemble
ϕ_j	Rata-rata nilai kontribusi dari fitur ke-j terhadap prediksi model secara keseluruhan
$\phi_{j,t}$	Mewakili nilai kontribusi spesifik terhadap prediksi pada satu pohon tunggal ke-t

$$\phi_{j,t}(x) = \sum_{v \in \text{Path}(x)} \phi_j(v, x)$$

j	Fitur yang sedang kita ukur
t	Menyatakan banyak total pohon yang ada di dalam model ensemble
x	Data spesifik yang sedang kita jelaskan
v	Node "titik keputusan" (percabangan) di dalam pohon
$\phi_j(v,x)$	Menyatakan Kontribusi Lokal dari fitur j yang dihitung hanya node v

$$\phi_j(v,x) = \begin{cases} E[s|v_{\text{anak}}] - E[s|v_{\text{induk}}] & \text{jika node } v \text{ membelah pada fitur } j \\ 0 & \text{selainnya, fitur } j \text{ tidak dipakai} \end{cases}$$

$E[s \mid v_{\text{induk}}]$	Menyatakan skor anomali rata-rata dari semua data training yang berada sebelum fitur v
$E[s \mid v_{\text{anak}}]$	Menyatakan skor anomali rata-rata dari semua data training yang berada setelah fitur v

Metric Evaluasi

F1-Score

Setelah kita membuat model matematika, kita akan lihat apakah model tersebut bagus untuk mencari anomaly pada datasetnya?

F1-Score	**F1-Score** adalah sebuah **metrik** yang digunakan **untuk mengukur kinerja model klasifikasi** dalam *machine learning*.

Rumus F1-Score mempunyai rumus sebagai berikut.

F1 Score

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recal

$$\text{Recall} = \frac{TP}{TP + FN}$$

Dengan...

TP	Model **benar** memprediksi sesuatu sebagai **anomali**.
FP	Model **salah** memprediksi data **normal** sebagai **anomali**.
FN	Model **salah** memprediksi **anomali** sebagai data **normal**.

Interpretasi

F1-Score yang **mendekati 1** menunjukkan **model sangat efektif dan seimbang**, sedangkan yang **mendekati 0** berarti **model gagal menemukan sebagian besar target** atau **terlalu sering membuat kesalahan identifikasi.**

Referensi

Sunarto, M. W., Kurniawan, D., Siswanto, E., & Huda, H. I. (2021). *Deteksi Anomali Menggunakan Extended Isolation Forest (EIF)*. **Jurnal Ilmu Teknik dan Informatika (TEKNIK)**, 1(2), 96–111. Universitas Sains dan Teknologi Komputer Semarang.
p-ISSN: 2808-8751 | e-ISSN: 2798-2513.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). *Isolation Forest*. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining (ICDM) (pp. 413-422). IEEE.

Salih, A. M., Raisi-Estabragh, Z., Boscolo Galazzo, I., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2023). *A perspective on explainable artificial intelligence methods: SHAP and LIME.* arXiv. https://arxiv.org/abs/2305.02012v3

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc.

Landauer, M., Skopik, F., Höld, G., & Wurzenberger, M. (2022). *A user and entity behavior analytics log data set for anomaly detection in cloud computing.* In *Proceedings of the 2022 IEEE International Conference on Big Data (Big Data)* (pp. 4285–4294). IEEE. https://doi.org/10.1109/BigData55660.2022.10020672

Nugroho, K. A., Hariguna, T., & Barkah, A. S. (2025). *Deteksi anomali trafik jaringan dan aktivitas pengguna menggunakan Isolation Forest untuk meningkatkan keamanan jaringan. Jurnal Pendidikan dan Teknologi Indonesia (JPTI), 5*(5), 1365–1376. https://doi.org/10.52436/1.jpti.790

Amar, C. S., & Bensaber, B. A. (2025). *Detecting attacks in V2G environments using an isolation forest based anomaly detection model.* In *Proceedings of the 2025 IEEE International Conference on Communications (ICC): Mobile and Wireless Networks Symposium* (pp. 4263–4269). IEEE. https://doi.org/10.1109/ICC52391.2025.11161178

Bulut, O., Gorgun, G., & He, S. (2024). *Unsupervised Anomaly Detection in Sequential Process Data.* Journal of Educational Measurement (ISSN 2190-8370). https://doi.org/10.1027/2151-2604/a000558


Terimakasih