

UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Corso di Laurea magistrale in Ingegneria Informatica e dell'Automazione

Privacy Fault Detection: rilevamento dello snooping negli accessi alle cartelle cliniche elettroniche

Privacy Fault Detection: Detecting Snooping in Electronic Health Records.

Relatore: Chiar.mo

Prof. Dragoni Aldo Franco

Tesi di Laurea di:

Antonini Antonio

A.A. 2025/2026

INDICE

ABSTRACT 6

CAPITOLO I 8

Il problema dello snooping sanitario 8

1.1 Il dato sanitario e la particolarità della sua violazione **8**

1.2 Il quadro normativo di riferimento: dal GDPR al dossier sanitario elettronico 10

1.3 Lo snooping: la minaccia che viene da dentro 15

1.4 Tipologie di illecito e condotte abusive 18

1.5 Anatomia dello snooping 21

1.6 Perché i controlli tradizionali non bastano e l'ipotesi di lavoro 23

1.7 Obiettivo, contributo e perimetro del lavoro 28

CAPITOLO II 33

Strumenti di lavoro 33

2.1 I criteri che hanno guidato la scelta degli strumenti 33

2.2 L'ambiente di sviluppo: Python e le librerie impiegate 34

2.3 La raccolta alla fonte: il filtro applicativo su Tomcat 36

2.4 I formati dei dati e la loro archiviazione 38

2.5 Lo standard di riferimento: HL7 FHIR e la risorsa AuditEvent 39

2.6 Gli strumenti statistici e gli algoritmi di rilevamento 40

2.7 Le metriche di valutazione e la tutela dei dati trattati 44

CAPITOLO III 47

Pipeline e Validazione *47*

3.1 Dalle firme comportamentali alle grandezze misurabili 47

3.2 Esigenze e architettura della pipeline 49

3.3 Acquisizione e normalizzazione dei log 53

3.4 Dal significato delle operazioni alle caratteristiche comportamentali 57

3.5 Il rilevamento delle anomalie 62

3.5.1 Dall'evento al profilo 63

3.5.2 Il modello e la spiegabilità delle segnalazioni 65

3.5.3 Dalla stranezza alla gravità: la priorità operativa 67

3.6 L'esecuzione della pipeline e i suoi risultati 70

3.7 Prototipo di rilevamento in tempo reale 76

3.8 La validazione 77

3.8.1 Il problema della valutazione e il metodo adottato 77

3.8.2 Il confronto tra i cinque algoritmi 79

3.8.3 Stabilità delle segnalazioni ed esclusione delle anomalie reali 82

3.9 I limiti del sistema 84

CAPITOLO IV 91

Conclusione e Sviluppi futuri 91

4.1 Sintesi del percorso e contributo del lavoro 91

4.2 Esito della validazione e natura di proof of concept 93

4.3 Arricchire il segnale: relazione di cura e ruolo dell'operatore 94

4.4 Verso il rilevamento in tempo reale e il confronto con i sistemi a regole 96

4.5 Dalla rilevazione all'esercizio: condizioni per un impiego reale 98

BIBLIOGRAFIA 100

SITOGRAFIA 106

INDICE DELLE FIGURE

Figura 3.1: Pipeline completa 52

Figura 3.2: Fase 1- da log grezzi ad AuditEvent 55

Figura 3.3: Fase 2 - Studio e arricchimento catalogo 60

Figura 3.4: Fase 3 - Isolation Forest 63

Figura 3.5: Fase 4 - Divisione priorità 67

Figura 3.6: Mappa dei profili con le diverse priorità 74

Figura 3.7: Prestazione dei cinque algoritmi al variare delle anomalie 81

Figura 8: Effetto dell'esclusione delle anomalie reali stabili 84

ABSTRACT

La protezione dei dati sanitari deve fronteggiare una minaccia che i controlli di sicurezza tradizionali non riescono a intercettare: lo ***healthcare snooping***, cioè la consultazione di dati clinici da parte del personale ospedaliero che, seppur dotato di credenziali legittime, nei fatti risulta privo di un titolo effettivo che ne legittimi l'accesso. Essendo l'operatore regolarmente autenticato e autorizzato, dal punto di vista del sistema, un simile accesso non è facilmente distinguibile da uno lecito; infatti, *firewall*, autenticazione e regole deterministiche non riescono ad intercettarlo, e i provvedimenti del Garante mostrano come la violazione emerga quasi sempre dal reclamo della vittima anziché da un controllo automatico.

Il lavoro parte da una premessa: se non è possibile stabilire a priori quando un accesso sia illecito, visto che la valutazione richiederebbe informazioni sulle finalità e sulle relazioni dell'operatore non disponibili al sistema, è però possibile misurare quanto un accesso sia anomalo rispetto al comportamento abituale dell'utente e a quello dei suoi pari. In virtù di ciò è stata progettata e realizzata una catena di elaborazione completa che, a partire dai *log* grezzi prodotti da un applicativo sanitario e raccolti tramite un filtro *Tomcat*, li normalizza traducendoli in *AuditEvent* e ne arricchisce le operazioni con un significato clinico e un livello di sensibilità derivato dalle categorie particolari di cui all'articolo 9 del Regolamento (UE) 2016/679 (GDPR).

Inoltre, estrae un insieme di caratteristiche comportamentali e individua le anomalie mediante un algoritmo non supervisionato, l'*Isolation Forest*, corredando ogni segnalazione della spiegazione dei fattori che l'hanno determinata e di una priorità operativa governata da regole dichiarate.

L'elemento distintivo non risiede nei singoli stadi, tutti fondati su tecnologie consolidate, bensì nel modo in cui essi vengono messi in relazione tra loro: da un lato la continuità dell'intera catena, che inizia dal *log* applicativo fino ad arrivare alla segnalazione leggibile da un ufficio *privacy*; dall'altro lato, il legame esplicito tra le firme comportamentali osservate dal modello e i comportamenti già qualificati come illeciti dall'autorità di controllo, dai quali ciascuna caratteristica è stata derivata.

In assenza di violazioni realmente etichettate, il metodo è stato validato attraverso l'iniezione di anomalie sintetiche sia a intensità fissa che a intensità crescente, e la scelta dell'algoritmo è stata giustificata sperimentalmente, confrontando cinque metodi non supervisionati sugli stessi dati.

Sviluppato su un ambiente dimostrativo popolato da anagrafiche fittizie – con la conseguenza, tutt'altro che secondaria in un lavoro dedicato alla protezione dei dati sanitari, che nessun dato clinico reale è stato trattato – l'elaborato costituisce una *proof of concept* completa e funzionante, corredata dai requisiti necessari a un'eventuale messa in esercizio.

CAPITOLO I

IL PROBLEMA DELLO SNOOPING SANITARIO

SOMMARIO: 1.1 IL DATO SANITARIO E LA PARTICOLARITÀ DELLA SUA VIOLAZIONE; 1.2 IL QUADRO NORMATIVO DI RIFERIMENTO: DAL GDPR AL DOSSIER SANITARIO ELETTRONICO; 1.3 LO SNOOPING: LA MINACCIA CHE VIENE DA DENTRO; 1.4 TIPOLOGIE DI ILLECITO E CONDOTTE ABUSIVE; 1.5 ANATOMIA DELLO SNOOPING; 1.6 PERCHÉ I CONTROLLI TRADIZIONALI NON BASTANO E L'IPOTESI DI LAVORO; 1.7 OBIETTIVO, CONTRIBUTO E PERIMETRO DEL LAVORO.

1.1 Il dato sanitario e la particolarità della sua violazione

Quando si parla di **violazione di dati personali**, è molto probabile che nel pensiero si figuri immediatamente un danno economico: una carta di credito clonata, un conto svuotato, o un'identità usata per aprire un finanziamento. Non vi è alcun dubbio che si tratti di danni gravi, ma sono accomunati da una caratteristica idonea a limitarne gli effetti negativi. Difatti, in un certo senso, sono tutte situazioni gestibili e, in quanto tali, reversibili: la carta si blocca, il conto si ripristina, il finanziamento si contesta.

Tuttavia, tali considerazioni non possono essere estese al **dato sanitario**: una diagnosi di sieropositività, un ricovero in un reparto di salute mentale, un'interruzione di gravidanza, o un percorso di dipendenza, costituiscono informazioni che, una volta uscite, non possono più rientrare.

Non esistono né un'operazione di ripristino, né un rimborso, e la conseguenza non è patrimoniale ma esistenziale, perché tocca la reputazione di una persona, i suoi rapporti di lavoro, le sue relazioni familiari e, in molti casi, la sua disponibilità futura a rivolgersi al servizio sanitario.

È esattamente per questo motivo che il legislatore europeo ha collocato i dati relativi alla salute all'interno delle cosiddette **categorie particolari** di dati personali[1], disciplinate

[1] Si tratta dei dati c.d. "sensibili", cioè quelli che rivelano l'origine razziale od etnica, le convinzioni religiose, filosofiche, le opinioni politiche, l'appartenenza sindacale, relativi alla salute o alla vita sessuale.

dall'articolo 9 del Regolamento (UE) 2016/679 (di seguito GDPR)[2], per le quali vige un divieto generale di trattamento salvo condizioni tassative.[3]

Il Considerando[4] n. 35 del GDPR chiarisce che rientrano in questa nozione tutti i dati che rivelano informazioni sullo stato di salute passato, presente o futuro dell'interessato, compresi quelli generati nel corso della prestazione di cura. Non si tratta quindi di una categoria residuale, ma del nucleo più protetto dell'intero impianto normativo[5].

Non è tutto. A questa tutela rafforzata corrisponde, sul piano tecnico, l'obbligo previsto dall'articolo 32 del GDPR[6], che impone al titolare del trattamento di adottare

---

[2] Regolamento generale sulla protezione dei dati, Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE.

[3] Art. 9, par. 1, del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Trattamento di categorie particolari di dati personali": «*È vietato trattare dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona*».

[4] I "considerando" sono introdotti dalla formula "considerando che" e sono numerati, a meno che non ve ne sia soltanto uno. Si collocano all'inizio di un atto dell'UE ed espongono le motivazioni delle sue disposizioni operative, evitando un linguaggio normativo e argomentazioni politiche. Definizione consultabile online nel sito https://eupati.eu/.

[5] Considerando n. 35 del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016: «*Nei dati personali relativi alla salute dovrebbero rientrare tutti i dati riguardanti lo stato di salute dell'interessato che rivelino informazioni connesse allo stato di salute fisica o mentale passata, presente o futura dello stesso. Questi comprendono informazioni sulla persona fisica raccolte nel corso della sua registrazione al fine di ricevere servizi di assistenza sanitaria o della relativa prestazione di cui alla direttiva 2011/24/UE del Parlamento europeo e del Consiglio; un numero, un simbolo o un elemento specifico attribuito a una persona fisica per identificarla in modo univoco a fini sanitari; le informazioni risultanti da esami e controlli effettuati su una parte del corpo o una sostanza organica, compresi i dati genetici e i campioni biologici; e qualsiasi informazione riguardante, ad esempio, una malattia, una disabilità, il rischio di malattie, l'anamnesi medica, i trattamenti clinici o lo stato fisiologico o biomedico dell'interessato, indipendentemente dalla fonte, quale, ad esempio, un medico o altro operatore sanitario, un ospedale, un dispositivo medico o un test diagnostico in vitro.*».

[6] Art. 32, par. 1, del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Sicurezza del trattamento": «*Tenendo conto dello stato dell'arte e dei costi di attuazione, nonché della natura, dell'oggetto, del contesto e delle finalità del trattamento, come anche del rischio di varia probabilità e gravità per i diritti e le libertà delle persone fisiche, il titolare del trattamento e il responsabile del trattamento mettono in atto misure tecniche e organizzative adeguate per garantire un livello di sicurezza adeguato al rischio, che comprendono, tra le altre, se del caso:*

a) la pseudonimizzazione e la cifratura dei dati personali;

b) la capacità di assicurare su base permanente la riservatezza, l'integrità, la disponibilità e la resilienza dei sistemi e dei servizi di trattamento;

c) la capacità di ripristinare tempestivamente la disponibilità e l'accesso dei dati personali in caso di incidente fisico o tecnico;

misure tecniche e organizzative adeguate al rischio, tenendo conto della natura e della gravità delle conseguenze per i diritti e le libertà delle persone fisiche.

Sebbene la norma non prescriva quali misure debbano essere adottate, stabilisce comunque un criterio generale: più il dato è sensibile, più il presidio deve essere robusto. Di conseguenza, nel caso di un *dossier* sanitario, che raccoglie in un unico contenitore l'intera storia clinica di un cittadino all'interno della medesima struttura sanitaria, il sistema di protezione dovrebbe essere il più solido possibile.

1.2 Il quadro normativo di riferimento: dal GDPR al *dossier* sanitario elettronico

Nel corso del tempo, la disciplina della protezione dei dati personali nell'ambito delle professioni e delle strutture sanitarie ha subito una profonda trasformazione.

Prima dell'entrata in vigore del GDPR, la vigilanza del Garante e l'attenzione giurisprudenziale si concentravano su **aspetti prevalentemente formali e burocratici**, come la corretta adozione dei moduli di informativa e la raccolta del consenso cartaceo.

A partire dal 25 maggio 2018, invece, con la piena applicabilità del GDPR, il baricentro normativo si è spostato sulla sostanza della **sicurezza** dei trattamenti e sulla **progettazione** dei sistemi informativi. Non a caso, infatti, l'art. 25 del GDPR ha introdotto i due principi cardine di *Privacy by Design* e *Privacy by Default*[7].

Con il primo principio, si sancisce l'obbligo per il titolare di incorporare i principi e le garanzie di protezione dei dati fin dalla fase di progettazione del trattamento e lungo il suo ciclo di vita, adottando misure tecniche e organizzative adeguate[8].

---

d) una procedura per testare, verificare e valutare regolarmente l'efficacia delle misure tecniche e organizzative al fine di garantire la sicurezza del trattamento.».

[7] Cfr. M. HILDEBRANDT, *Law for Computer Scientists and Other Folk*, Oxford University Press, Oxford 2020, p. 164 ss.

[8] Art. 25, par. 1, del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Protezione dei dati fin dalla progettazione e protezione dei dati per impostazione predefinita": «*Tenendo conto dello stato dell'arte e dei costi di attuazione, nonché della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, come anche dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche costituiti dal trattamento, sia al momento di determinare i mezzi del trattamento sia all'atto del trattamento stesso il titolare del trattamento mette in atto misure tecniche e organizzative adeguate, quali la pseudonimizzazione, volte ad attuare in modo efficace i principi*

Il secondo principio, invece, prevede che le impostazioni predefinite garantiscano il trattamento dei soli dati personali necessari per ciascuna specifica finalità, tenendo conto della quantità dei dati, della portata del trattamento, del periodo di conservazione e della loro accessibilità[9].

In questo contesto, la ripartizione dei ruoli di responsabilità all'interno delle organizzazioni sanitarie appare fondamentale per determinare le rispettive sfere di imputabilità.

L'Azienda Sanitaria (ASL o Azienda Ospedaliera), la clinica privata o il professionista autonomo, assumono la veste di **Titolare del Trattamento**[10], al quale spetta il compito istituzionale di determinare le finalità e i mezzi dell'elaborazione dei dati personali.

Peraltro, il Titolare individua e delega specifiche funzioni a **Dirigenti e Preposti**, responsabili dell'analisi dei rischi, dell'adozione delle opportune misure di sicurezza e dell'istruzione del personale.

Infine, opera in qualità di **Autorizzato al Trattamento**[11] tutto il personale medico, infermieristico e amministrativo, con l'obbligo tassativo di accedere e trattare i dati attenendosi esclusivamente alle istruzioni operative e alle mansioni ricevute[12].

---

di protezione dei dati, quali la minimizzazione, e a integrare nel trattamento le necessarie garanzie al fine di soddisfare i requisiti del presente regolamento e tutelare i diritti degli interessati.».

[9] Art. 25, par. 2, del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Protezione dei dati fin dalla progettazione e protezione dei dati per impostazione predefinita": «*Il titolare del trattamento mette in atto misure tecniche e organizzative adeguate per garantire che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento. Tale obbligo vale per la quantità dei dati personali raccolti, la portata del trattamento, il periodo di conservazione e l'accessibilità. In particolare, dette misure garantiscono che, per impostazione predefinita, non siano resi accessibili dati personali a un numero indefinito di persone fisiche senza l'intervento della persona fisica.*».

[10] Art. 4, n. 7), del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Definizioni": «*«titolare del trattamento»: la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo che, singolarmente o insieme ad altri, determina le finalità e i mezzi del trattamento di dati personali; quando le finalità e i mezzi di tale trattamento sono determinati dal diritto dell'Unione o degli Stati membri, il titolare del trattamento o i criteri specifici applicabili alla sua designazione possono essere stabiliti dal diritto dell'Unione o degli Stati membri.*».

[11] Art. 29 del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Trattamento sotto l'autorità del titolare del trattamento o del responsabile del trattamento": «*Il responsabile del trattamento, o chiunque agisca sotto la sua autorità o sotto quella del titolare del trattamento, che abbia accesso a dati personali non può trattare tali dati se non è istruito in tal senso dal titolare del trattamento, salvo che lo richieda il diritto dell'Unione o degli Stati membri.*».

[12] Art. 2-quaterdecies del Codice Privacy per "l'Autorizzato al Trattamento": «*1. Il titolare o il responsabile del trattamento possono prevedere, sotto la propria responsabilità e nell'ambito del proprio assetto*

Questa architettura di responsabilità poggia su una premessa sostanziale, già richiamata nel paragrafo precedente: i dati sanitari appartengono alle categorie particolari di cui all'articolo 9 del GDPR, per le quali il paragrafo 1 pone un divieto generale di trattamento. Tuttavia, in forza del successivo paragrafo del medesimo articolo, tale divieto viene meno per le finalità di cura. Difatti, la lettera h) autorizza l'accesso perché necessario alla cura, e non copre la consultazione dettata dalla semplice curiosità, che di tale necessità è priva[13]. È proprio qui che nasce il rischio al centro di questo lavoro: essere autorizzati ad accedere per curare non significa essere autorizzati ad accedere sempre. La distanza tra le due cose è lo spazio in cui si colloca l'abuso.

Il quadro appena delineato non si applica però in modo uniforme a qualunque forma di documentazione sanitaria: il Garante e la giurisprudenza distinguono in maniera netta gli strumenti tradizionali di documentazione clinica dalle piattaforme che raccolgono la storia sanitaria nel tempo. A tale distinzione corrisponde, di conseguenza, una diversa regolamentazione giuridica.

La **cartella clinica** è tradizionalmente qualificata come un atto pubblico[14], obbligatorio per legge, che è destinato a documentare in maniera completa e cronologicamente coerente le attività sanitarie svolte durante un singolo ricovero o una specifica prestazione ambulatoriale. Poiché la sua redazione costituisce un adempimento previsto dalla disciplina sanitaria e legato alla cura immediata[15], il trattamento dei dati personali necessario alla prestazione di cura non richiede, quale base giuridica, il consenso dell'interessato.

---

organizzativo, che specifici compiti e funzioni connessi al trattamento di dati personali siano attribuiti a persone fisiche, espressamente designate, che operano sotto la loro autorità.
2. Il titolare o il responsabile del trattamento individuano le modalità più opportune per autorizzare al trattamento dei dati personali le persone che operano sotto la propria autorità diretta.».

[13] Art. 9, par. 2, lettera h), del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016: "Il paragrafo 1 non si applica se si verifica uno dei seguenti casi […] *Il trattamento è necessario per finalità di medicina preventiva o di medicina del lavoro, valutazione della capacità lavorativa del dipendente, diagnosi, assistenza o terapia sanitaria o sociale […]*".

[14] CORTE DI CASSAZIONE, Sentenza n. 17647 del 9/05/2025. La Cassazione ha ribadito che la cartella clinica redatta dal medico di un ospedale pubblico ha natura di atto pubblico e documenta le attività compiute dal pubblico ufficiale.

[15] Art. 7 del D.P.R. n. 128 del 27 marzo 1969, che attribuisce al responsabile del reparto la responsabilità della regolare compilazione e conservazione delle cartelle cliniche.

La consultazione delle informazioni contenute nella cartella deve, tuttavia, avvenire nel rispetto dei principi di necessità e pertinenza, restando riservata all'interessato[16] e ai soggetti legittimati, nonché ai professionisti sanitari per i quali l'accesso sia necessario allo svolgimento delle attività di cura[17].

Il ***Dossier*** **Sanitario Elettronico (DSE)**, invece, è una banca dati che viene predisposta da una struttura per raccogliere in un unico strumento l'intera storia clinica, passata e presente, di un paziente presso quell'istituto. Proprio perché aggrega molto più di un singolo episodio, il *dossier* è considerato un trattamento ulteriore, autonomo e facoltativo rispetto alla cura del singolo evento: la sua costituzione richiede quindi un consenso specifico, che deve essere libero e ben spiegato all'assistito, distinto da quello prestato per i singoli atti medici. Il paziente resta comunque libero di rifiutarlo, e il rifiuto non può in alcun modo limitare l'accesso alle cure[18].

A questo consenso di base se ne aggiunge uno ancora più specifico che risulta necessario per i dati particolarmente delicati: si tratta di informazioni concernenti, ad

---

[16] Art 15, par. 1, del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Diritto di accesso dell'interessato": «*1. L'interessato ha il diritto di ottenere dal titolare del trattamento la conferma che sia o meno in corso un trattamento di dati personali che lo riguardano e in tal caso, di ottenere l'accesso ai dati personali e alle seguenti informazioni:*

a. le finalità del trattamento;

b. le categorie di dati personali in questione;

c. i destinatari o le categorie di destinatari a cui i dati personali sono stati o saranno comunicati, in particolare se destinatari di paesi terzi o organizzazioni internazionali;

d. quando possibile, il periodo di conservazione dei dati personali previsto oppure, se non è possibile, i criteri utilizzati per determinare tale periodo;

e. l'esistenza del diritto dell'interessato di chiedere al titolare del trattamento la rettifica o la cancellazione dei dati personali o la limitazione del trattamento dei dati personali che lo riguardano o di opporsi al loro trattamento;

f. il diritto di proporre reclamo a un'autorità di controllo;

g. qualora i dati non siano raccolti presso l'interessato, tutte le informazioni disponibili sulla loro origine;

h. l'esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato.».

[17] Cfr. art. 92 del d.lgs. 30 giugno 2003, n. 196, come modificato dal d.lgs. 10 agosto 2018, n. 101 (Codice privacy), in combinato disposto con l'art. 15 del Regolamento (UE) 27 aprile 2016, n. 679; artt. 22, 24, comma 7, e 25 della legge 7 agosto 1990, n. 241; art. 4 della legge 8 marzo 2017, n. 24.

[18] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Linee guida in materia di Dossier sanitario*, Provvedimento del 4 giugno 2015, doc. web n. 4084632, pubblicato in G.U. n. 164 del 17 luglio 2015.

esempio, l'infezione da HIV, un'interruzione volontaria di gravidanza o dati relativi a violenza sessuale.

L'accesso al *dossier*, inoltre, non può essere esteso indistintamente a tutto il personale, ma va riservato ai soli operatori che hanno in cura l'assistito nel momento in cui lo consultano. Conformemente, le linee guida chiedono alle strutture di differenziare i profili di accesso in base al ruolo di ciascun operatore; un principio che, come sarà illustrato nel capitolo III, è alla base dell'idea stessa di confrontare il comportamento di un utente con quello dei suoi pari[19].

A tutela del paziente restano, infine, due garanzie importanti.

Innanzitutto, il **diritto di oscuramento**, cioè la possibilità di rendere invisibili determinati referti o episodi a chiunque, tranne che a sé stesso e al medico che li ha prodotti. Peraltro, l'oscuramento deve essere a sua volta nascosto, così da evitare che colui che consulta il *dossier* si renda conto che il dato è stato oscurato.

La seconda garanzia riguarda il **diritto di conoscere i propri accessi**. In sostanza, il paziente può chiedere di sapere chi ha consultato il suo *dossier*, con, altresì, l'indicazione del reparto, della data e dell'ora. La tracciabilità degli accessi, quindi, non è solo una misura di sicurezza interna, ma costituisce un vero e proprio diritto riconosciuto all'interessato[20].

[19] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Linee guida in materia di Dossier sanitario*, op. cit.
[20] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Linee guida in materia di Dossier sanitario*, op. cit.

1.3 Lo snooping: la minaccia che viene da dentro

Quando si progetta la sicurezza di un sistema informativo si tende a immaginare un attacco proveniente da un agente esterno, cioè da un soggetto che non ha alcun titolo per accedere a dati sensibili e che, quindi, deve necessariamente forzare un perimetro. Contro questo tipo di minaccia esiste un armamentario consolidato, fatto di firewall, cifratura[21], autenticazione, segmentazione di rete[22] e sistemi di rilevamento delle intrusioni[23].

Quanto premesso è utile a comprendere come il fenomeno di cui si tratterà in questa tesi sia di natura del tutto diversa.

Con il termine ***healthcare snooping***, o più semplicemente ***snooping***, si indica la consultazione di dati clinici da parte del personale che possiede credenziali legittime e che, sebbene di norma sia formalmente autorizzato ad accedere al sistema, in quello specifico caso non possiederebbe alcun titolo per farlo[24]. Ad esempio, perché non esiste una relazione di cura con il paziente, o perché non sussiste una necessità amministrativa, o, ancora, perché non vi è alcuna richiesta da parte del diretto interessato. Alla base di un siffatto utilizzo, potrebbe esserci una mera curiosità, o, al contrario, l'obiettivo di ottenere un'informazione da utilizzare contro qualcuno.

I casi che saranno affrontati e analizzati nei paragrafi che seguono[25] mostrano come una simile "curiosità" abbia destinatari ricorrenti. Nello specifico, di regola si tratta di

---

[21] La cifratura è il processo di trasformazione di un testo in chiaro (plaintext) in un testo cifrato (ciphertext) tramite l'uso di un algoritmo crittografico e di una chiave. In sostanza, rende i dati comprensibili solo a chi possiede la chiave di decifrazione corretta. Consultabile online sul sito del *National Institute of Standards and Technology* (NIST) https://www.nist.gov/.

[22] La segmentazione di rete è una pratica di sicurezza che consiste nel suddividere una rete informatica di grandi dimensioni in sottoreti più piccole o zone distinte per controllare il traffico, isolare le risorse critiche e impedire il movimento laterale delle minacce. Consultabile online sul sito del NIST https://www.nist.gov/.

[23] Un sistema di rilevamento delle intrusioni, o Intrusion Detection System (IDS), è un software o un hardware che automatizza il processo di monitoraggio e analisi degli eventi in un sistema informatico o in una rete alla ricerca di segnali di problemi di sicurezza, violazioni delle policy o attività anomale. Per questo motivo, genera allarmi e notifiche per gli amministratori di sicurezza quando rileva un potenziale attacco in corso. Consultabile online sul sito del NIST https://www.nist.gov/.

[24] Il termine inglese *healthcare snooping* è di uso corrente e riconosciuto a livello internazionale. Cfr. La pagina *dell'Association of Health Care Journalists* (AHCJ), un'associazione professionale di giornalisti sanitari statunitensi, consultabile online al sito https://healthjournalism.org/.

[25] La casistica è ricostruita ed esaminata nei parr. 1.4 e 1.5 di questo capitolo, dove i provvedimenti del Garante vengono ricondotti a cinque ricorrenze: la sfera affettiva (Azienda Provinciale per i Servizi Sanitari di Trento, Provvedimento del 21 aprile 2021, doc. web 9678507; ASL, Provvedimento del 22 febbraio 2024, doc. web 10001279), il collega come paziente (Azienda Ospedaliera di Verona, Provvedimento del 23

colleghi di lavoro (in particolare le donne in maternità, coloro che hanno patologie stigmatizzanti, e i dipendenti coinvolti in procedimenti disciplinari), di ex *partner* e di loro nuovi compagni, di familiari, di vicini di casa, o di personaggi noti. A conferma di ciò basta vedere i provvedimenti del Garante: in quasi tutti, la vittima non è un paziente a caso, ma qualcuno che l'autore della violazione conosce personalmente.

Ora, ai fini della progettazione di un sistema di rilevamento, appare chiaro come questo tipo di accesso sia formalmente ineccepibile: l'utente è chi dice di essere, si è autenticato correttamente, ha i permessi necessari, opera dalla propria postazione e, molto spesso, lo fa in pieno orario di servizio. Non c'è nessuna regola violata, nessun errore restituito, nessun tentativo fallito. Dunque, dal punto di vista del sistema, quell'accesso è indistinguibile da uno lecito. E infatti, con riferimento ai tre pilastri del controllo degli accessi, cioè il c.d. **modello «AAA»**[26], l'autenticazione ha funzionato, l'autorizzazione ha funzionato e persino la tracciabilità ha funzionato, poiché l'accesso è stato regolarmente registrato. Ciò che, invece, non risulta è la legittimità della finalità, ossia l'unica cosa che il sistema non è in grado di valutare, e che non rientra in nessuno dei tre pilastri.

Ne consegue che nessun controllo perimetrale può intercettare lo *snooping*, visto che lo stesso non deve attraversare il perimetro, dato che è già al suo interno; le ragioni per cui nessuno dei presidi tradizionali riesca a coglierlo saranno esaminate nei paragrafi successivi.

La tutela contro lo *snooping* non si esaurisce esclusivamente sul piano amministrativo del GDPR. Lo stesso comportamento assume rilievo anche da un punto di vista **penale**. L'articolo 615-ter del Codice penale, infatti, punisce l'accesso abusivo a un sistema informatico o telematico. Più nel dettaglio, la norma stabilisce che *«chiunque abusivamente si introduce in un sistema informatico o telematico protetto da misure di*

gennaio 2020, doc. web 9269629; Ospedale di Bolzano, Provvedimento del 3 luglio 2014, doc. web 3325808), la ripetizione ossessiva (Azienda Ospedaliero-Universitaria di Trieste, Provvedimento del 19 gennaio 2017, doc. web 6521868), la curiosità sporadica e il movente strumentale (Cassazione penale, sentenza n. 33972 del 2023). Qui se ne anticipano soltanto i tratti comuni.

[26] La dicitura "AAA" deriva dai tre termini inglesi "*Authentication, Authorization, Accounting*", che significano, rispettivamente, "autenticazione, autorizzazione e contabilità (o anche tracciabilità)".

sicurezza ovvero vi si mantiene contro la volontà espressa o tacita di chi ha il diritto di escluderlo, è punito con la reclusione fino a tre anni».

Ora, ai fini di questo lavoro, il punto decisivo risiede nella seconda condotta descritta dalla norma, e cioè il fatto di **mantenersi** nel sistema contro la volontà di chi ha diritto di escludere.

La giurisprudenza ha chiarito come questa fattispecie non richieda affatto di forzare una barriera tecnica: commette il reato anche l'utente **regolarmente autorizzato** che accede al sistema per finalità estranee a quelle per cui l'autorizzazione gli è stata concessa[27]. Si tratta proprio del caso dello *snooping*: sebbene l'operatore possegga credenziali valide, le impiega oltre il titolo che le giustifica, e, quindi, abusivamente.

Il diritto penale coglie così, attraverso altre parole, la stessa distinzione che percorre l'intero lavoro: quella tra l'accesso formalmente permesso e l'accesso legittimo nella finalità.

Continuando con l'analisi dell'articolo, si nota come i successivi due commi rendano lo stesso particolarmente pertinente al contesto sanitario.

Nello specifico, il secondo comma prevede un aggravamento della pena quando il fatto è commesso da un pubblico ufficiale o da un incaricato di pubblico servizio con abuso dei poteri o violazione dei doveri d'ufficio, condizione che ricorre per il personale di una struttura sanitaria pubblica[28].

---

[27] Corte di cassazione, Sezioni Unite penali, sentenza 7 febbraio 2012, n. 4694 (ud. 27 ottobre 2011). Le Sezioni Unite hanno chiarito che integra il reato anche la condotta di chi, pur essendo formalmente abilitato ad accedere al sistema, vi si introduce o vi si mantiene violando le condizioni e i limiti posti dal titolare dello ius excludendi, senza che occorra il superamento di alcuna barriera tecnica. È esattamente la situazione descritta in questo lavoro: l'abuso non risiede nell'accesso in sé, che resta tecnicamente consentito, bensì nel suo esercizio oltre il titolo in forza del quale è stato concesso.

[28] Art. 615-ter, comma 2, del Codice penale: «*La pena è della reclusione da due a dieci anni:*
1) se il fatto è commesso da un pubblico ufficiale o da un incaricato di un pubblico servizio, con abuso dei poteri, o con violazione dei doveri inerenti alla funzione o al servizio, o da chi esercita anche abusivamente la professione di investigatore privato, o con abuso della qualità di operatore del sistema;
2) se il colpevole per commettere il fatto usa minaccia o violenza sulle cose o alle persone, ovvero se è palesemente armato;
3) se dal fatto deriva la distruzione o il danneggiamento ovvero la sottrazione, anche mediante riproduzione o trasmissione, o l'inaccessibilità al titolare del sistema o l'interruzione totale o parziale del suo funzionamento, ovvero la distruzione o il danneggiamento dei dati, delle informazioni o dei programmi in esso contenuti.».

Per quanto riguarda il terzo comma, invece, in esso viene stabilito un aggravamento di pena quando l'accesso riguarda sistemi informatici di interesse pubblico, tra i quali sono espressamente compresi quelli relativi alla sanità[29].

Non è tutto. La Legge n. 90 del 28 giugno 2024[30], in materia di cybersicurezza, è intervenuta inasprendo le pene per questa fattispecie, a conferma della crescente attenzione del legislatore verso la protezione dei sistemi informatici, anche in ambito sanitario.

Ne consegue che un medesimo accesso improprio a un *dossier* sanitario può integrare, contemporaneamente, sia un illecito nel trattamento dei dati ai sensi del GDPR, sia un reato alla stregua di quanto disposto dall'articolo 615-ter. Sebbene si tratti di due piani distinti – da un lato quello amministrativo, relativo alla protezione dei dati, e, dall'altro lato, quello penale, concernente la tutela del sistema informatico – alla fine ambedue convergono nel qualificare come illecita la stessa condotta.

1.4 Tipologie di illecito e condotte abusive

Non tutte le violazioni della *privacy* in ambito sanitario hanno la stessa natura. Per delimitare il problema di cui si occupa questo lavoro è utile operare una distinzione, perché è proprio da qui che si delinea il perimetro del sistema di rilevamento. Per questo motivo, in sede di inizializzazione del lavoro, sono state identificate tre **tipologie di violazioni**.

La prima è quella delle **falle strutturali di progettazione**. Qui il problema non risiede nel singolo operatore, bensì nel sistema che mette a disposizione più di quanto dovrebbe.

Rientrano in questa categoria tutte quelle situazioni in cui il *software* consente a chiunque, nell'intera struttura o addirittura nell'intera regione, di consultare il *dossier* di qualsiasi cittadino.

---

[29] Art. 615-ter, comma 3, del Codice penale: «*Qualora i fatti di cui ai commi primo e secondo riguardino sistemi informatici o telematici di interesse militare o relativi all'ordine pubblico o alla sicurezza pubblica o alla sanità o alla protezione civile o comunque di interesse pubblico, la pena è, rispettivamente, della reclusione da tre a dieci anni e da quattro a dodici anni.*».

[30] Legge n. 90 del 28 giugno 2024, "Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici", in G.U. n. 153 del 2 luglio 2024.

Al riguardo, va menzionata la pronuncia del Garante in merito a un caso concernente il Servizio Sanitario del Friuli-Venezia Giulia. In sostanza, il sistema di refertazione della regione era organizzato in modo da consentire l'accesso, senza alcuna restrizione, a tutto il personale, anche con riferimento ai *dossier* dei colleghi.

Il Garante ravvisò criticità proprio con riferimento a informativa e consenso, modalità di accesso al *dossier* e diritto di oscuramento, arrivando a qualificare come illecito il trattamento effettuato e imponendo misure correttive[31].

Un altro caso, sempre in materia, ha riguardato la disattivazione dei filtri *privacy* disposta durante l'emergenza Covid-19. Tale scelta aveva reso accessibili al personale sanitario anche *dossier* relativi a pazienti che non avevano prestato il consenso e consentito la consultazione da parte di operatori non coinvolti nel relativo percorso di cura. Il Garante ha ritenuto illegittime le modalità con cui tale misura era stata attuata, rilevando che il contesto emergenziale poteva giustificare semplificazioni nell'accesso ai dati solo se indispensabili e proporzionate alle esigenze di cura, nonché accompagnate da adeguate garanzie tecniche e organizzative. Nel caso concreto, invece, l'ampliamento degli accessi era stato generalizzato e non era stato affiancato da un sistema idoneo a rilevare comportamenti anomali o potenzialmente illeciti.[32]

Come anticipato, in questi casi la responsabilità è a monte, nella configurazione del sistema: un rilevatore comportamentale come quello proposto in questo studio non li previene, anzi li presuppone risolti, perché ha bisogno che gli accessi siano quantomeno tracciati.

La seconda categoria è quella degli **attacchi informatici e dei difetti *software***. Qui la violazione nasce da un errore di programmazione o da un'azione ostile esterna: ad esempio, un attacco *ransomware*[33] che esfiltra dati[34], un codice QR generato in modo

---

[31] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Dossier sanitario e trattamento dei dati personali dei pazienti*, Provvedimento n. 3 del 10 gennaio 2013, doc. web n. 2284708.

[32] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza-ingiunzione nei confronti dell'Azienda USL Valle d'Aosta*, Provvedimento n. 371 del 10 novembre 2022, doc. web n. 9819792.

[33] Un *ransomware* è un attacco informatico dannoso in cui i criminali bloccano o cifrano i dati di un sistema e chiedono un pagamento in denaro per restituire l'accesso. Consultabile online sul sito del NIST https://www.nist.gov/.

[34] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Provvedimento n. 83 del 13 febbraio 2025, doc. web n. 10116834.

sequenziale che permette di risalire ai referti altrui cambiando una cifra[35], un bug di stampa che allega alla cartella di un paziente i documenti di un altro[36], etc. Sono tutte violazioni reali e spesso gravi, che, però, non appartengono al dominio dell'abuso di privilegi legittimi, ma a quello della sicurezza informatica tradizionale, cioè quella dei firewall, della cifratura, del controllo del codice.

Si arriva, infine, all'ultima tipologia di violazioni, quella specifica di questo lavoro: l'**abuso di privilegi legittimi**, cioè lo *snooping*. Si tratta della fattispecie, già citata, in cui non c'è né una falla nel sistema, né un attacco. Al contrario, vi è un operatore regolarmente autenticato e autorizzato ad accedere; il quale, sebbene consulti dati che formalmente e tecnicamente potrebbe consultare, in realtà, nel caso concreto, non possiede alcun titolo per farlo.

In proposito, due casi possono essere menzionati.

Il primo è quello del personale di radiologia dell'APSS di Trento che, tra il 2017 e il 2019, ha consultato ripetutamente i *dossier* di due persone legate da vincoli affettivi, pur non essendo coinvolto nella loro cura[37].

Il secondo caso riguarda l'Azienda Ospedaliera di Verona, in cui si sono verificati tre distinti episodi: un tecnico di radiologia è entrato nei *dossier* di sette colleghi "per vedere come funzionava l'applicazione", uno specializzando ha consultato per "mera curiosità" le cartelle di tre colleghi neurologi non suoi pazienti, e, infine, sono stati aperti i *dossier* di sei ostetriche, alcune delle quali in congedo di maternità, da una postazione lasciata incustodita durante un turno notturno[38].

In tutti questi casi il Garante ha qualificato l'accesso come illecito non perché violasse una barriera tecnica, dato che non ne era stata forzata nessuna, ma perché mancava la sola cosa che il sistema non sa vedere: la **finalità di cura**.

---

[35] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza ingiunzione nei confronti di Azienda sanitaria unica regionale Marche*, Provvedimento n. 9 del 13 gennaio 2022, doc. web n. 9744496.

[36] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza ingiunzione nei confronti di Ospedale Pediatrico Bambino Gesù*, Provvedimento n. 250 del 24 giugno 2021, doc. web n. 9689566.

[37] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza ingiunzione nei confronti di Azienda provinciale per i servizi sanitari di Trento*, Provvedimento n. 155 del 21 aprile 2021, doc. web n. 9678507.

[38] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Provvedimento correttivo e sanzionatorio nei confronti di Azienda Ospedaliero Universitaria Integrata di Verona*, Provvedimento n. 18 del 23 gennaio 2020, doc. web n. 9269629.

1.5 Anatomia dello snooping

Questo paragrafo è dedicato all'analisi dettagliata delle violazioni ricomprese nella terza categoria. I casi di abuso di privilegi legittimi non sono tutti uguali: cambiano il movente, il bersaglio e, soprattutto, per ciò che qui rileva maggiormente, la traccia che l'accesso lascia nei log.

Ricostruire queste ricorrenze non è un esercizio di catalogazione: serve a capire la forma di un accesso sospetto, perché è quella forma che il modello dovrà imparare a riconoscere.

Dai provvedimenti del Garante emergono cinque ricorrenze, che tornano con particolare frequenza in situazioni di questo tipo.

Innanzitutto, assume una posizione rilevante la **sfera affettiva.** Il bersaglio è spesso un ex partner, il suo nuovo compagno o la sua nuova compagna, oppure una persona con cui l'autore ha o ha avuto un legame personale.

Se si riprende il caso dell'APSS di Trento, si trova conferma di ciò: un radiologo ha consultato ripetutamente, per due anni, i *dossier* dell'ex fidanzata e del nuovo compagno di lei. O ancora, il caso della dottoressa che ha ripetutamente visionato esami e storia clinica del marito, mai stato suo paziente[39].

La traccia comportamentale è netta: l'operatore accede a un oggetto che sta fuori dal proprio bacino abituale di pazienti, e spesso vi ritorna nel tempo.

È il tipo di scostamento che, nel modello, verrà colto dal confronto con il comportamento dei colleghi di reparto e dalla ripetizione su più giornate.

La seconda ricorrenza si collega a tutte quelle situazioni in cui il bersaglio è un **collega**, cioè un dipendente della stessa struttura, che però, in quel dato momento, è **anche paziente**. Oltre ad essere la fattispecie più frequente, è anche quella più dolorosa, perché mette in gioco i rapporti di lavoro.

Rientrano qui numerosi casi, primo fra tutti quello già menzionato con riferimento all'Azienda Ospedaliera di Verona, in cui alcuni dipendenti consultarono i *dossier* delle ostetriche e dei colleghi neurologi. Un altro caso è quello di Bolzano, dove una dipendente

[39] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Provvedimento n. 97 del 22 febbraio 2024, doc. web n. 10001279.

spiò la cartella di una collega scoprendone la sieropositività all'HIV e diffondendo poi l'informazione in reparto[40].

Non è tutto. A dimostrazione dell'assunto iniziale si può citare anche il reclamo dell'Arcispedale di Ferrara, in cui i dipendenti vennero a conoscenza di dati sensibili sulla salute di una collega[41].

Qui, il focus della violazione riguarda l'accesso al *dossier* di un pari da parte di chi non lo ha in cura, aggravato dal fatto che spesso i dati toccati appartengono ad aree particolarmente delicate. La sensibilità dei dati toccati è una delle grandezze che il sistema utilizza per pesare le segnalazioni prodotte dal modello.

Proseguendo con l'elencazione delle ricorrenze più frequenti, è bene parlare della **ripetizione ossessiva**: a volte, anche se il bersaglio è uno solo, il *dossier* viene consultato molte volte.

Il caso emblematico è quello della radiologa dell'Azienda Ospedaliera di Trieste[42], che accedette quaranta volte alla cartella di una donna con cui non aveva alcun rapporto, né di cura, né personale.

In questa situazione particolare, la caratteristica rilevante non è l'ampiezza dell'accesso, visto che l'operatore accede a pochi oggetti distinti, ma la sua intensità: un volume anomalo concentrato su un singolo oggetto. Si tratta di un'altra faccia dello *snooping* "esplorativo", che mostra come contare solo quanti oggetti distinti tocchi un utente non basti, serve anche misurare quanto si accanisce sullo stesso.

La quarta ricorrenza riguarda il caso più sfuggente: la **curiosità sporadica**. Pochi accessi, mirati, distribuiti nel tempo, motivati da una curiosità momentanea, più che da un intento.

È il caso, già menzionato, dello specializzando di Verona che aprì per mera curiosità i *dossier* di tre colleghi. Proprio perché il volume è basso, questi accessi stanno sotto qualunque soglia fissa e sfuggono ai controlli. È qui che l'approccio comportamentale

---

[40] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Trattamento di dati tramite il dossier sanitario aziendale*, Provvedimento n. 340 del 3 luglio 2014, doc. web n. 3325808.

[41] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Provvedimento n. 533 del 25 settembre 2025, doc. web n. 10185490.

[42] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Ordinanza ingiunzione nei confronti di Perrone Rosaria e Azienda Universitaria Ospedaliera Ospedali Riuniti di Trieste, Registro dei provvedimenti n. 15 del 19 gennaio 2017, doc. web n. 6521868.

mostra la sua ragion d'essere: non contano i numeri assoluti, ma lo scostamento dalla norma dell'utente e dei suoi pari.

Infine, assume rilevanza anche il **movente strumentale**: in alcuni casi l'accesso non nasce da curiosità, ma serve a procurarsi un'informazione da usare contro qualcuno.

È il caso di una dirigente la quale, dopo aver prelevato il referto di una dipendente, ne usò il contenuto per una segnalazione disciplinare che portò al trasferimento della vittima[43].

Questa ricorrenza merita attenzione proprio perché segna un confine del sistema: il modello può rilevare l'accesso anomalo, ma non ciò che dell'informazione verrà fatto successivamente. L'uso strumentale del dato resta invisibile a un'analisi dei soli *log* di accesso.

Alla luce di quanto sinora esposto, è bene specificare che queste cinque ricorrenze non sono compartimenti stagni. Al contrario, è possibile che un singolo caso appartenga contemporaneamente a più di una.

Ad esempio, l'accesso al *dossier* dell'ex partner è spesso anche ripetuto. Oppure, l'accesso al *dossier* del collega è spesso anche spinto da una certa curiosità.

Pertanto, poiché lo *snooping* può manifestarsi in forme differenti, un sistema progettato per riconoscerne una sola configurazione rischierebbe di non individuare le altre.

È da qui che nasce l'esigenza di descrivere ogni giornata di lavoro non con un unico indicatore, ma con un insieme di caratteristiche comportamentali: la traduzione di ciascuna ricorrenza in una firma comportamentale misurabile, e di ogni firma in una grandezza calcolabile, sarà oggetto del capitolo III.

1.6 Perché i controlli tradizionali non bastano e l'ipotesi di lavoro

Le ricorrenze appena descritte spiegano anche perché i presidi su cui si fonda tradizionalmente la sicurezza di un sistema informativo non riescano a intercettarle.

[43] CORTE DI CASSAZIONE, Sez. III penale, sentenza 2 agosto 2023, n. 33972.

Nessuno di essi, infatti, è costruito per valutare la finalità di un accesso, che è invece l'unico elemento capace di distinguere lo *snooping* da una consultazione legittima.

Tra questi rientra, innanzitutto, il ***firewall***, cioè un sistema di sicurezza (*hardware* o *software*) che regola il traffico di rete.[44]

Nello specifico, agisce da filtro tra una rete interna fidata e reti esterne non sicure, monitorando e bloccando il traffico di dati in entrata e in uscita grazie all'esistenza di regole predefinite stabilite dall'amministratore. In questo modo, permette di impedire l'accesso non autorizzato e di contenere la diffusione di minacce[45]. Tuttavia, nel caso di specie, il sistema viene eluso perché il dipendente ospedaliero si trova già all'interno del perimetro difensivo. Il suo traffico è considerato legittimo a livello di rete, quindi i controlli IP, di porta o di firma del *firewall* non bloccano le sue azioni, anche se l'uso che fa di quelle risorse è anomalo.

Un altro elemento utile a valutare la liceità di un accesso è l'**autenticazione**, cioè il processo con cui un sistema verifica l'identità dichiarata da un utente o da un dispositivo o da un'applicazione che richiede l'accesso a una risorsa. In particolare, richiede credenziali valide (come *password*, *token* o dati biometrici) o l'autenticazione a più fattori (MFA), al fine di garantire che solo le persone o i sistemi autorizzati possano entrare nell'*account* o nella rete[46]. Essendo però tipicamente eseguita una sola volta a inizio sessione e indipendente dal dato consultato, non è di per sé sufficiente a impedire gli abusi da parte di utenti legittimamente autenticati.

Infine, meritano menzione anche le **regole deterministiche**, cioè istruzioni fisse e prevedibili applicate senza margini di dubbio. In altre parole, si tratta di regole decisionali predefinite nelle quali, al verificarsi di una determinata condizione o combinazione di condizioni, il sistema produce una conseguenza prestabilita. Nei sistemi *rule-based* esse

---

[44] W. STALLINGS, L. BROWN, *Computer Security: Principles and Practice*, V ed., Pearson, New York, 2024, p. 52 ss.

[45] L. AYALA, *Cybersecurity for Hospitals and Healthcare Facilities: A Guide to Detection and Prevention,* I ed., Apress, Berkeley (California), 2016, p. 58.

[46] E. BERTINO, *Data Protection from Insider Threats*, I ed., Springer, Cham (Svizzera), 2012, p. 7 ss.

sono generalmente espresse secondo una struttura binaria del tipo *IF–THEN*[47] (se X allora Y)[48] e, dunque, un utente ha o non ha il permesso di compiere un'azione.

Proprio per questo motivo, se l'operatore interno ha il permesso di visualizzare le cartelle cliniche per svolgere il proprio lavoro, una regola deterministica non bloccherà mai la visualizzazione dei file dopo un certo numero, perché il permesso formale esiste. La regola non comprende il contesto, cioè che consultare, ad esempio, 1.000 file in un'ora è un'anomalia rispetto alla media giornaliera dell'utente.

Su regole di questo tipo si fonda, in effetti, la reazione più naturale di fronte a un problema di tale portata. Si stabilisce, ad esempio, l'obbligo di segnalare l'accesso oltre un certo orario, oppure il limite per un utente di consultare più di un certo numero di cartelle al giorno, salvo la sussistenza di ragionevoli motivi, o, ancora, che un profilo amministrativo non ha il permesso di accedere a determinate aree cliniche.

Tuttavia, sebbene le regole siano utili e vadano mantenute, corrono il rischio di rivelarsi insufficienti in considerazione di tre limiti strutturali[49].

Innanzitutto, **colgono soltanto ciò che è già stato immaginato**: una regola nasce da una violazione nota, quindi è certamente efficace se applicata al caso da cui origina, ma ciò non è altrettanto sicuro rispetto ad altre situazioni, ognuna connotata da proprie peculiarità. Di conseguenza, ogni comportamento abusivo non previsto in fase di scrittura potrebbe rimanere invisibile.

In secondo luogo, le regole sono potenzialmente **aggirabili proprio da chi le conosce**: il personale interno sa qual è la soglia oltre la quale scatta il controllo, e le violazioni descritte nella casistica non sono quasi mai massive. Si tratta, infatti, di manciate di accessi, mirati e distribuiti nel tempo, esattamente il tipo di traffico che risiede al di sotto di qualunque soglia ragionevole.

---

[47] A. A. HOPGOOD, *Intelligent Systems for Engineers and Scientists*, IV ed., CRC Press, Boca Raton (Florida), 2021, p. 33 ss.

[48] Cfr. L. SÁ CORREIA, R. J. CRUZ-CORREIA, P. P. RODRIGUES, *Illegitimate HIS access by healthcare professionals: scenarios, use cases and audit trail-based detection model*, in *Procedia Computer Science*, vol. 164, 2019, pp. 629–636.

[49] Cfr. A. A. BOXWALA, J. KIM, J. M. GRILLO, L. OHNO-MACHADO, *Using statistical and machine learning to help institutions detect suspicious access to electronic health records*, in *Journal of the American Medical Informatics Association*, vol. 18, n. 4, 2011, pp. 498–505.

Da ultimo, la **rilevanza del contesto**: venti accessi in un'ora possono essere del tutto normali per un operatore del pronto soccorso e completamente anomali per un tecnico amministrativo. Una regola con soglia fissa, applicata indistintamente a entrambi, o genera una valanga di falsi allarmi sul primo, o lascia passare inosservato il secondo. Nella pratica, come sarà illustrato nel capitolo IV, rischia di generare entrambi gli scenari.

C'è poi un ulteriore aspetto, più di natura organizzativa che tecnica, che emerge con grande evidenza dai provvedimenti del Garante[50]: nella quasi totalità dei casi la violazione non è stata scoperta da un sistema di controllo, ma dal **reclamo della vittima**, che ha sospettato qualcosa perché, magari, una voce aveva iniziato a circolare in reparto. Solo dopo il reclamo l'azienda sanitaria ha analizzato i propri *file* di *log* e ha trovato le tracce dell'accesso abusivo.

Questo rivela due aspetti: da un lato, che i *log* ci sono, e, dall'altro lato, che nessuno li guarda. In altre parole, il tracciamento esiste come strumento di prova a posteriori[51], non come strumento preventivo di rilevamento.

In definitiva, un *audit trail*[52] consultato solo quando qualcuno si lamenta serve unicamente a confermare, una volta già verificatosi, se un accesso sia effettivamente avvenuto o meno: non è un presidio di sicurezza, è un archivio.

Da queste considerazioni deriva l'assunto alla base dell'intero elaborato che può essere così esplicato: **se non è possibile definire a priori che cosa sia un accesso illecito, è però possibile misurare che cosa sia un accesso anomalo**.

La differenza tra i due concetti è sostanziale e va chiarita subito, perché condiziona tutto ciò che segue.

---

[50] Provvedimento del 21 aprile 2021, doc. web 9678507; Provvedimento del 22 febbraio 2024, doc. web 10001279, Provvedimento del 23 gennaio 2020, doc. web 9269629; Provvedimento del 3 luglio 2014, doc. web 3325808, Provvedimento del 19 gennaio 2017, doc. web 6521868.

[51] Cfr. A. K. MENON, X. JIANG, J. KIM, J. VAIDYA, L. OHNO-MACHADO, *Detecting Inappropriate Access to Electronic Health Records Using Collaborative Filtering*, in *Machine Learning*, vol. 95, n. 1, 2014, pp. 87-101.

[52] Un *audit trail* è un registro cronologico che esamina la sequenza di attività che portano a una specifica operazione, procedura o evento in una transazione rilevante per la sicurezza. Consultabile online sul sito del NIST https://www.nist.gov/.

Un **accesso illecito**[53] è una nozione giuridica che richiede di conoscere le finalità dell'operatore e la sua relazione con il paziente, cioè informazioni che il sistema informativo, di norma, non possiede.

Un **accesso anomalo**[54] è, invece, una nozione statistica, che richiede soltanto di sapere come si comporta abitualmente quell'operatore e come si comportano i suoi pari[55].

Le due nozioni non coincidono, e sarebbe scorretto sostenere il contrario: esistono accessi anomali perfettamente leciti, come una sostituzione in un reparto diverso dal proprio o una guardia notturna straordinaria, ed esistono, almeno in linea di principio, accessi illeciti del tutto ordinari nella forma.

Quello che si può ragionevolmente affermare è che le violazioni descritte nella casistica lasciano una traccia comportamentale, e che questa traccia è misurabile. Ad esempio, il medico che consulta quaranta volte la cartella di una persona che non è sua paziente produce una ripetizione anomala; l'operatore che consulta il *dossier* di una collega di un altro reparto produce uno scostamento dal proprio gruppo di pari; il tecnico che accede a dati clinici per "provare il *software*" produce una combinazione anomala tra il proprio profilo e la sensibilità dei dati consultati. Nessuno di questi segnali, preso da solo, dimostra un illecito. Tutti insieme, però, individuano un insieme ristretto di comportamenti che meritano di essere guardati e controllati da un essere umano.

È opportuno chiarire, a questo punto, un secondo aspetto: il sistema si limita a rilevare gli accessi anomali, non a impedirli. Potrebbe sembrare preferibile un meccanismo che blocchi l'accesso sospetto nel momento stesso in cui viene tentato, ma in ambito sanitario questa soluzione non è praticabile.

---

[53] Nell'ordinamento italiano viene definito come "Accesso abusivo a un sistema informatico o telematico" ed è un vero e proprio reato. Difatti, l'art. 615-ter, comma 1, del Codice penale, stabilisce: «*Chiunque abusivamente si introduce in un sistema informatico o telematico protetto da misure di sicurezza ovvero vi si mantiene contro la volontà espressa o tacita di chi ha il diritto di escluderlo, è punito con la reclusione fino a tre anni.*».

[54] Per accesso anomalo può intendersi un accesso a un sistema informatico che, per modalità, frequenza, orario, posizione geografica, indirizzo IP, dispositivo utilizzato o altre caratteristiche, si discosta dal normale profilo di utilizzo dell'utente e risulta pertanto meritevole di verifica sotto il profilo della sicurezza. L'Agenzia per l'Italia Digitale (AgID) parla espressamente di "comportamenti anomali" degli utenti e fa anche esempi concreti: utilizzo del sistema fuori dai normali orari di lavoro oppure numero di accessi ripetuti in un breve lasso di tempo. Cfr.
Linee guida sul punto di accesso telematico ai servizi della Pubblica Amministrazione, consultabile online al sito https://www.agid.gov.it/it.

[55] E. BERTINO, *op. cit.*, p. 37 ss.

Da un lato, non si può bloccare ciò che non si è ancora in grado di riconoscere, perché l'anomalia non risiede nel singolo evento ma nel suo accumulo, e al momento del primo accesso non esiste alcun elemento che lo distingua da uno legittimo.

Dall'altro lato, negare un accesso legittimo in un contesto clinico può avere conseguenze sulla salute di un paziente: un falso positivo che impedisce a un medico di consultare una cartella in situazione di urgenza produce un danno incomparabilmente più grave di una consultazione impropria. Non è un caso che in sanità si adotti il paradigma del *break the glass*[56], in base al quale l'accesso eccezionale non viene negato, bensì concesso e tracciato: la prevenzione, in questo modello, non è tecnica ma deterrente, perché si fonda sulla consapevolezza di essere tracciati e sulla certezza della conseguenza disciplinare.

Il rilevamento resta comunque la premessa di qualunque forma di blocco, dal momento che nessun sistema può impedire un comportamento che non sia prima in grado di riconoscere. Proprio per tale ragione, il capitolo III riserva un'attenzione specifica alla spiegabilità dei risultati[57], visto che un allarme di sicurezza privo di una chiara giustificazione contestuale è inutilizzabile ai fini di un *audit*[58].

1.7 Obiettivo, contributo e perimetro del lavoro

L'obiettivo di questa tesi è la progettazione e la realizzazione di una catena di elaborazione completa che arrivi a segnalare in modo spiegabile i comportamenti di accesso anomali, partendo dai *log* applicativi grezzi prodotti da un applicativo sanitario[59].

---

[56] M. BISHOP, COMPUTER SECURITY: ART AND SCIENCE, Boston, 2019, p. 249 ss.

[57] Vedi cap. III.

[58] L'*audit* è un processo di revisione e un esame indipendente di registri, attività e controlli di un sistema per verificare
il rispetto di determinati standard. Consultabile online sul sito del NIST https://www.nist.gov/.

[59] Con "applicativo sanitario" si intende la versione digitale della cartella clinica del paziente, contenente informazioni sanitarie longitudinali e resa disponibile agli operatori sanitari autorizzati per finalità assistenziali.
Cfr. OFFICE OF THE NATIONAL COORDINATOR FOR HEALTH INFORMATION TECHNOLOGY (ONC), *What is an Electronic Health Record (EHR)?*, consultabile online sul sito https://healthit.gov/.

La *pipeline*[60] si articola nelle seguenti fasi, ciascuna delle quali corrispondente a un problema che è stato necessario affrontare e risolvere[61]:

1. Innanzitutto, la **raccolta dei *log* grezzi** attraverso un filtro applicativo (*interceptor*) posizionato in ingresso, che intercetta le richieste in transito e le registra in un database documentale;
2. In secondo luogo, la **normalizzazione degli accessi nello *standard* FHIR *AuditEvent***[62], cioè la traduzione di un formato proprietario e disomogeneo in una risorsa standardizzata, interoperabile e semanticamente esplicita;
3. Successivamente, l'**attribuzione di significato alle operazioni**, mediante la costruzione di un catalogo semantico che associa a ogni operazione tracciata l'area clinica interessata e un livello di sensibilità derivato dalle categorie particolari di cui all'articolo 9 del GDPR;
4. Poi, l'**estrazione di caratteristiche comportamentali** (*feature engineering*), progettate a partire dalle firme ricorrenti individuate nell'analisi della casistica;
5. Infine, **il rilevamento delle anomalie** grazie a un algoritmo non supervisionato (*Isolation Forest*), con esplicitazione dei fattori che hanno determinato ciascuna segnalazione.

A partire dalla scelta di produrre per ogni segnalazione un punteggio motivato e leggibile, l'architettura è stata pensata per potersi estendere verso una sesta fase non ancora realizzata: il recapito dell'allarme a una piattaforma SIEM (*Security Information and Event Management)*[63], che ne curerebbe la correlazione con le regole deterministiche

---

[60] Per pipeline si intende una sequenza organizzata di fasi attraverso cui i dati vengono acquisiti, trasformati e trasferiti verso sistemi o processi successivi di elaborazione e analisi. Cfr.
IBM, *What Is a Data Pipeline?*, consultabile online sul sito https://www.ibm.com/it-it.

[61] Vedi cap. III, parr. 3.3 - 3.5.

[62] È una risorsa dello *standard* internazionale *Fast Healthcare Interoperability Resources* (FHIR) utilizzata per registrare e tracciare qualsiasi evento rilevante per la sicurezza, la *privacy* o l'integrità dei dati all'interno di un sistema informativo sanitario.

[63] Il *National Institute of Standards and Technology* (NIST) definisce uno strumento SIEM come: "*Application that provides the ability to gather security data from information system components and present that data as actionable information via a single interface*.". In altre parole, le piattaforme SIEM sono strumenti di sicurezza informatica finalizzati alla raccolta centralizzata dei dati provenienti dai diversi componenti di un sistema informativo, alla loro analisi e alla trasformazione degli stessi in informazioni utili ai fini del monitoraggio e dell'individuazione di eventi rilevanti sotto il profilo della sicurezza. Consultabile
online sul sito del *National Institute of Standards and Technology* https://www.nist.gov/.

già in uso e l'eventuale attivazione di una risposta. Questa integrazione non rientra nel perimetro di questo lavoro.

Da un lato, una piattaforma SIEM basata su regole deterministiche non è di per sé in grado di rilevare lo *snooping* comportamentale, per le stesse ragioni discusse in questo capitolo: la sua efficacia dipende dal segnale comportamentale che questo lavoro produce, non lo sostituisce.

Dall'altro, costruire e validare un'integrazione reale tra i due sistemi richiederebbe una raccolta di dati e una collaborazione con chi gestisce la piattaforma in uso in azienda, non compatibili con i tempi di questo lavoro. Un primo confronto preliminare tra i due approcci è comunque in corso di valutazione, ed è discusso nel capitolo dedicato agli sviluppi futuri[64].

Nessuno dei singoli stadi appena elencati costituisce, da solo, un elemento di novità, dal momento che FHIR è uno standard pubblico sviluppato da *Health Level Seven* (HL7) *International*[65] per lo scambio elettronico di informazioni sanitarie, basato su risorse modulari e tecnologie *web*; *Isolation Forest* è un algoritmo consolidato di *anomaly detection* che individua le anomalie isolandole tramite partizioni casuali dei dati, visto che i punti anomali tendono a richiedere meno suddivisioni per essere isolati rispetto ai dati normali[66]; e la raccolta dei *log* applicativi è una pratica ordinaria di acquisizione e conservazione degli eventi generati dalle applicazioni, finalizzata al monitoraggio, all'analisi, all'*audit* e alla sicurezza dei sistemi informativi[67].

Ciò che questo lavoro aggiunge risiede nel modo in cui tali elementi vengono messi in relazione tra loro, e si può riassumere in due punti, costituenti altresì i due fili conduttori dell'elaborato.

---

[64] Vedi cap. IV.

[65] *HL7 International* è un'organizzazione globale che crea standard internazionali per lo scambio e l'integrazione di dati elettronici tra i diversi sistemi informativi sanitari.

[66] F. T. LIU, K. M. TING, Z.-H. ZHOU, *Isolation Forest*, in *Proceedings of the 8th IEEE International Conference on Data Mining* (ICDM), IEEE, 2008.
Successivamente approfondito in: F. T. Liu
, K. M. TING, Z.-H. ZHOU, *Isolation-Based Anomaly Detection*, in *ACM Transactions on Knowledge Discovery from Data*, vol. 6, n. 1, 2012, pp. 1–39.

[67] K. KENT, M. SOUPPAYA, *Guide to Computer Security Log Management*, NIST Special Publication 800-92, 2006, consultabile online sul sito del NIST https://www.nist.gov/.

Il primo è la **continuità della catena**. Il percorso che porta da un *log* applicativo, scritto per finalità di diagnostica, fino a una segnalazione leggibile da un ufficio *privacy*, viene qui compiuto per intero: ogni stadio produce un formato che lo stadio successivo è in grado di interpretare, e le informazioni necessarie all'analisi vengono conservate lungo tutto il tragitto. Il capitolo III è dedicato in larga parte a questo aspetto, perché i *log* di partenza non erano stati concepiti per essere letti in chiave di *audit*, e ricostruire da essi il contenuto informativo richiesto dallo standard ha rappresentato la parte più consistente dello sviluppo.

Il secondo punto è il **legame esplicito tra la parte giuridica e la parte tecnica**. Le caratteristiche comportamentali che il modello osserva non sono state scelte in base a criteri puramente statistici, ma sono state derivate, una per una, dai comportamenti che il Garante ha già ritenuto illeciti nei propri provvedimenti.

La tassonomia delle violazioni da cui sono state estratte le firme comportamentali ricorrenti è stata elaborata nei parr. 1.4 e 1.5 di questo capitolo, mentre nel capitolo III verrà mostrato come ciascuna di queste firme sia stata tradotta in una grandezza calcolabile a partire da un *AuditEvent*.

Un sistema di rilevamento è credibile nella misura in cui risulta chiaro che cosa non sia in grado di fare. In questa sede, quindi, si anticiperanno i limiti che saranno poi discussi in modo più esteso nel capitolo III.

Il sistema **rileva abusi di privilegi legittimi**, e non è progettato per rilevare né gli attacchi informatici esterni, né i difetti di progettazione dei software gestionali, né gli errori umani. Sono tutte fattispecie ampiamente presenti nella casistica, ma appartengono a famiglie diverse di violazione e richiedono presidi diversi.

Inoltre, il sistema **attribuisce un comportamento a una credenziale, non a una persona**. Diversi provvedimenti del Garante riguardano accessi effettuati con credenziali altrui, sottratte, ad esempio, da un post-it o da una postazione lasciata incustodita, oppure con *account* generici di reparto condivisi da più operatori. In tutti questi casi il *log* registra un utente legittimo, e nessuna analisi comportamentale può sapere chi fosse davvero seduto davanti alla tastiera. Si tratta di un limite strutturale, comune a qualunque sistema di questo tipo, e costituisce un'ulteriore ragione per cui la segnalazione non può mai essere considerata una prova.

Infine, va dichiarato il **contesto** in cui il lavoro è stato svolto. La *pipeline* è stata sviluppata e collaudata su un **ambiente dimostrativo aziendale**, popolato da anagrafiche fittizie e utilizzato internamente per attività di collaudo e di formazione del personale di nuovo inserimento. Questa scelta ha una conseguenza positiva e una limitante.

Quella positiva è che nessun dato sanitario reale è stato trattato nel corso di questo lavoro, il che è tutt'altro che secondario, visto che la tesi si occupa proprio della protezione di tali dati.

La conseguenza limitante, invece, riguarda il fatto che il comportamento "normale" appreso dal modello costituisce il comportamento di un ambiente di prova, e non quello di un reparto ospedaliero: la validazione, che sarà trattata nel capitolo III, e che è stata condotta mediante iniezione di anomalie sintetiche, dimostra quindi la capacità della catena di elaborazione di isolare i profili anomali, ma non sostituisce una sperimentazione sul campo.

Pertanto, come anticipato, e in definitiva, il lavoro in questo senso va inteso come una *proof of concept* completa e funzionante, corredata dai requisiti necessari a un'eventuale messa in esercizio.[68]

[68] Si tratta, in altre parole, di una realizzazione dimostrativa e funzionante volta a provare la fattibilità dell'approccio, non ancora destinata all'impiego operativo.

CAPITOLO II

STRUMENTI DI LAVORO

SOMMARIO: 2.1 I CRITERI CHE HANNO GUIDATO LA SCELTA DEGLI STRUMENTI; 2.2 L'AMBIENTE DI SVILUPPO: PYTHON E LE LIBRERIE IMPIEGATE; 2.3 LA RACCOLTA ALLA FONTE: IL FILTRO APPLICATIVO SU TOMCAT; 2.4 I FORMATI DEI DATI E LA LORO ARCHIVIAZIONE; 2.5 LO STANDARD DI RIFERIMENTO: HL7 FHIR E LA RISORSA AUDITEVENT; 2.6 GLI STRUMENTI STATISTICI E GLI ALGORITMI DI RILEVAMENTO; 2.7 LE METRICHE DI VALUTAZIONE E LA TUTELA DEI DATI TRATTATI.

2.1 I criteri che hanno guidato la scelta degli strumenti

Il capitolo precedente ha definito il problema e l'ipotesi di lavoro. Questo capitolo descrive gli strumenti con cui quell'ipotesi è stata messa alla prova: i linguaggi, i formati, lo standard di interoperabilità, gli algoritmi e le metriche impiegati. Si è scelto di tenerli distinti dal racconto del loro impiego: qui si dice che cosa ciascuno strumento sia e per quale ragione sia stato preferito ad altri, mentre il modo in cui è stato usato, le difficoltà incontrate e i risultati ottenuti sono oggetto del capitolo III.

La selezione non è avvenuta per familiarità con l'uno o con l'altro strumento, ma seguendo cinque criteri, tutti discendenti dalla natura del problema descritta nel capitolo precedente.

Il primo criterio è la preferenza per standard aperti e documentati. Un sistema destinato al controllo degli accessi in una struttura sanitaria non è un esercizio che si esaurisce con chi lo scrive: deve poter essere ispezionato, mantenuto ed esteso da altri, e deve poter dialogare con i sistemi già in uso. Ogni volta che esisteva uno standard pubblico in luogo di un formato proprietario, è stato scelto lo standard.

Il secondo criterio discende dall'assenza di etichette. Poiché nessun insieme di accessi è mai stato dichiarato lecito o illecito da un'autorità competente, non esistono esempi di abuso su cui addestrare un modello. Questo esclude in radice l'intera famiglia degli

algoritmi supervisionati e impone strumenti non supervisionati, capaci di isolare i profili fuori norma senza che nessuno abbia mostrato loro, prima, che cosa sia un abuso.

Il terzo criterio è la spiegabilità. Un sistema che esprime un sospetto su una persona deve poter essere interrogato su ogni sua decisione, e questo restringe il campo degli strumenti ammissibili: sono stati preferiti metodi il cui risultato resta riconducibile a grandezze leggibili, e ogni soglia è stata tenuta fuori dal modello, dichiarata nel codice anziché appresa dai dati.

Il quarto criterio è la robustezza ai valori estremi. È un punto che attraversa l'intero impianto statistico del lavoro e che vale la pena enunciare subito: i dati su cui il sistema lavora contengono, per definizione, proprio le anomalie che si vogliono trovare. Uno strumento che calcolasse il riferimento di normalità con la media e la deviazione standard si lascerebbe spostare dagli stessi valori che deve individuare. Per questa ragione, ovunque occorra un riferimento — nel riscalamento delle grandezze, nel calcolo degli scostamenti, nella definizione delle soglie di gravità — sono state impiegate la mediana e lo scarto interquartile, che quei valori estremi li ignorano.

Il quinto criterio, infine, è la riproducibilità. Gli strumenti sono stati scelti anche in funzione della possibilità di ripetere ogni passaggio ottenendo lo stesso risultato: formati intermedi leggibili da un essere umano, componenti indipendenti gli uni dagli altri, e componenti aleatorie governate da un valore di inizializzazione dichiarato.

2.2 L'ambiente di sviluppo: Python e le librerie impiegate

L'intera catena di elaborazione è scritta in Python[69]. La ragione della scelta non è la velocità di esecuzione, rispetto alla quale altri linguaggi offrirebbero prestazioni migliori, bensì la disponibilità di un ecosistema maturo e consolidato per l'analisi dei dati e per l'apprendimento automatico, che consente di impiegare implementazioni verificate degli

[69] Python è un linguaggio di programmazione interpretato, a tipizzazione dinamica e di uso generale. La documentazione ufficiale e le specifiche del linguaggio sono consultabili online al sito https://www.python.org/.

algoritmi anziché riscriverle, e la leggibilità del codice, che in un lavoro destinato a essere discusso e verificato conta quanto il risultato.

Le librerie impiegate sono sei, ciascuna con un compito preciso.

La prima è *pandas*[70], che fornisce la struttura dati tabellare sulla quale gli accessi vengono raggruppati per utente e per giornata e dalla quale si ricavano le caratteristiche comportamentali. La seconda è *NumPy*[71], su cui *pandas* stesso si fonda, impiegata per il calcolo numerico e in particolare per le statistiche di posizione — mediane e quartili — che ricorrono in tutto il lavoro.

La terza, e la più rilevante, è *scikit-learn*[72], la libreria di apprendimento automatico da cui provengono tanto l'algoritmo di rilevamento adottato quanto i quattro metodi con cui è stato confrontato, oltre agli strumenti di preelaborazione e alle metriche di valutazione. Il fatto che algoritmi e metriche provengano da un'unica libreria non è un dettaglio organizzativo: garantisce che il confronto tra i cinque metodi avvenga a parità di condizioni, senza che differenze di implementazione o di convenzione sui punteggi ne alterino l'esito.

Le ultime tre hanno un ruolo di servizio. *PyMongo*[73] è il driver con cui gli script dialogano con il *database*, *Matplotlib*[74] produce i grafici con cui i risultati della validazione sono rappresentati, e *Watchdog*[75], impiegata unicamente nello stadio di acquisizione, resta in ascolto della comparsa di nuovi *file* nella cartella monitorata.

---

[70] W. McKinney, *Data Structures for Statistical Computing in Python*, in *Proceedings of the 9th Python in Science Conference*, 2010, pp. 56–61. La documentazione della libreria è consultabile online al sito https://pandas.pydata.org/.

[71] C. R. Harris et al., *Array programming with NumPy*, in *Nature*, vol. 585, 2020, pp. 357–362. La documentazione della libreria è consultabile online al sito https://numpy.org/.

[72] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, in *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825–2830. La documentazione delle singole implementazioni è consultabile online al sito https://scikit-learn.org/.

[73] PyMongo è il driver ufficiale con cui un programma Python dialoga con MongoDB. La documentazione è consultabile online al sito https://pymongo.readthedocs.io/.

[74] J. D. Hunter, *Matplotlib: A 2D Graphics Environment*, in *Computing in Science & Engineering*, vol. 9, n. 3, 2007, pp. 90–95. La documentazione della libreria è consultabile online al sito https://matplotlib.org/.

[75] Watchdog è una libreria Python per il monitoraggio degli eventi del file system, che consente a un programma di reagire alla comparsa o alla modifica di un file senza interrogare ciclicamente la cartella. La documentazione è consultabile online al sito https://python-watchdog.readthedocs.io/.

2.3 La raccolta alla fonte: il filtro applicativo su Tomcat

I *log* su cui l'intera *pipeline* lavora, non sono un sottoprodotto casuale dell'applicativo, ma sono generati da un componente dedicato di tracciamento, sviluppato da un collaboratore esterno e in modo indipendente da questo progetto.

Sebbene il presente lavoro riparta dal suo output, è comunque necessario conoscerne il funzionamento, perché alcune garanzie introdotte a questo livello si ritrovano, coerenti, nei vari stadi della *pipeline*.

Sul piano tecnico si tratta di un servlet filter[76], cioè di un componente previsto dalla specifica Java per i server che ospitano applicazioni web: il filtro si interpone tra la richiesta in arrivo e l'applicazione che deve servirla, e può osservarla senza alterarne il comportamento. Il contenitore in cui esso opera è Apache Tomcat[77], il server applicativo su cui l'applicativo sanitario è in esecuzione; la registrazione del filtro avviene nella configurazione del contenitore, e non richiede di intervenire sul codice dell'applicativo osservato.

Il suo compito è osservare le richieste che arrivano al sistema. Tali richieste non vengono registrate nella loro totalità, bensì solo quelle scelte in anticipo: un *file* di configurazione dedicato, o anche *file* di *properties*[78] che viene tenuto separato dal resto del codice, elenca gli indirizzi (URL) da intercettare e, per ciascuno, indica quali parametri della richiesta catturare. Una richiesta viene quindi registrata solo se il suo indirizzo corrisponde a una delle regole configurate; e poiché queste vivono in un *file* a parte, cambiare ciò che si osserva non richiede di toccare il programma, ma solo di aggiornare quel *file*.

Quando una richiesta corrisponde a una di queste regole, il programma non si limita a salvare le sole informazioni che gli sono servite per riconoscerla, ma raccoglie anche

---

[76] Il filtro è un componente previsto dalla specifica Servlet, che ne disciplina il ciclo di vita e l'ordine di applicazione alle richieste in transito. Cfr. la *Jakarta Servlet Specification*, consultabile online al sito https://jakarta.ee/.

[77] Apache Tomcat è un contenitore open source che implementa le specifiche Jakarta Servlet e Jakarta Server Pages, e ospita l'esecuzione di applicazioni web scritte in Java. Consultabile online al sito https://tomcat.apache.org/.

[78] Un *file* di *properties* è un *file* di configurazione testuale, tipico dell'ambiente Java, che memorizza parametri come coppie chiave/valore, tenute separate dal codice del programma. Cfr. ORACLE, *The Java Tutorials*, consultabile online al sito https://docs.oracle.com/.

un insieme definito, sebbene ampio, di dati su quella richiesta, come l'indirizzo, il metodo, l'esito e la durata, o ancora, i dati dell'operatore e i parametri configurati per quell'indirizzo, e li scrive come un evento in un *file* di testo.

Non ne viene letto il corpo, ma solo la vista aggregata dei parametri. Il formato usato per questi *file*, chiamato **JSONL**[79], è pensato appositamente per contenere tanti eventi diversi, uno per ogni riga.

La scrittura avviene in locale e in modo asincrono, per introdurre il minimo *overhead* possibile sulla richiesta applicativa che la sta generando, dato che tracciare un accesso non deve mai rallentare, in modo percepibile, l'operazione che l'utente sta effettivamente compiendo.

Un componente separato, chiamato *worker*[80], legge periodicamente i *file* già scritti e li invia allo *storage* esterno configurato, di *default* MongoDB. Il canale che alimenta effettivamente la *pipeline* è un altro: i *file* JSONL prodotti dal filtro vengono prelevati manualmente tramite un *client* SFTP[81] e depositati nella cartella monitorata da cui parte la catena di elaborazione descritta nel capitolo III.

Il filtro è progettato attorno a **tre garanzie**, coerenti con la delicatezza del ruolo che occupa.

Innanzitutto, un errore nel tracciamento non deve mai bloccare la richiesta applicativa che sta tracciando: la funzione di *audit* è accessoria rispetto al servizio, non può comprometterlo.

Inoltre, un'indisponibilità temporanea dello *storage* esterno non deve comportare la perdita di eventi, perché i *file* restano in locale fino a quando il loro invio non sia andato a buon fine.

Infine, l'*overhead*[82] introdotto sulla richiesta va tenuto al minimo, per non degradare l'esperienza dell'applicativo che si sta osservando.

---

[79] JSON *Lines* (JSONL) è un formato testuale in cui ogni riga del file è un oggetto JSON autonomo e indipendente, adatto a contenere lunghe sequenze di eventi. Consultabile online al sito https://jsonlines.org/.
[80] Un *worker* è un processo che opera in *background*, separato dal flusso principale, incaricato di svolgere periodicamente un compito ripetitivo — qui, il trasferimento dei file già scritti verso lo storage esterno.
[81] SFTP (SSH *File Transfer Protocol*) è un protocollo per il trasferimento sicuro di file tra client e server su una connessione cifrata SSH. Consultabile online sul sito del NIST https://www.nist.gov/.
[82] Con *overhead* si intende l'insieme delle risorse accessorie consumate in sovrappiù da un sistema per gestire un processo, senza però produrre direttamente il risultato finale.

2.4 I formati dei dati e la loro archiviazione

Gli strumenti descritti nei paragrafi precedenti si scambiano dati, e il formato con cui lo fanno è esso stesso una scelta.

Il formato di base è JSON (JavaScript Object Notation)[83], un formato testuale leggero per la rappresentazione e lo scambio di dati strutturati, organizzati come coppie chiave-valore ed eventualmente annidati, leggibile sia dall'uomo sia dalla macchina. È questa doppia leggibilità a renderlo adatto a una catena in cui ogni stadio deve poter essere ispezionato: il contenuto di un passaggio intermedio si verifica aprendo un file, senza strumenti dedicati. La sua variante JSONL, già incontrata a proposito del filtro, ne conserva le proprietà e vi aggiunge l'indipendenza di ciascuna riga, utile quando gli eventi arrivano in sequenza continua.

Un formato testuale, però, porta con sé una questione che in un contesto reale non è affatto secondaria: la codifica dei caratteri, cioè la convenzione con cui le lettere vengono rappresentate in byte. Lo standard attuale è UTF-8 (Unicode Transformation Format, 8 bit)[84], che rappresenta ogni carattere del repertorio universale Unicode mediante una sequenza di uno o più byte e conserva la piena compatibilità con il set US-ASCII. Accanto ad esso sopravvive però Windows-1252[85], la codifica a byte singolo adottata storicamente come code page predefinita nei sistemi Microsoft Windows per l'Europa occidentale. Le due convenzioni coincidono sui caratteri non accentati e divergono su tutti gli altri, ed è da questa divergenza che nasce una delle difficoltà concrete affrontate nello stadio di acquisizione, descritta nel capitolo III.

Quanto alla conservazione, gli eventi non restano in file ma vengono depositati in un database. Lo strumento impiegato è MongoDB[86], un database documentale, appartenente cioè alla famiglia NoSQL: anziché organizzare i dati in tabelle dalle colonne fisse,

---

[83] La specifica del formato è consultabile online al sito https://www.json.org/.

[84] F. YERGEAU, *UTF-8, a transformation format of ISO 10646*, RFC 3629, 2003, consultabile online al sito https://www.rfc-editor.org/.

[85] La tabella dei caratteri della code page 1252, che estende lo standard ISO/IEC 8859-1 (Latin-1), è consultabile online al sito https://learn.microsoft.com/.

[86] La documentazione del database e del suo modello dei dati è consultabile online al sito https://www.mongodb.com/.

memorizza ciascun record come un documento in formato JSON, senza imporre uno schema comune a tutti.

La scelta di un database documentale non è casuale: i *log* dell'applicativo non hanno una struttura fissa, ma variano i parametri presenti nella richiesta a seconda dell'operazione. Per questo motivo, imporre qui uno schema rigido avrebbe voluto dire scartare informazioni prima ancora di sapere se sarebbero servite.

Un documento JSON per evento, senza schema imposto, conserva ogni campo esattamente com'è arrivato, e rimanda ogni decisione di struttura agli stadi successivi della catena.

2.5 Lo standard di riferimento: HL7 FHIR e la risorsa AuditEvent

Lo standard su cui si fonda la normalizzazione degli accessi è *Fast Healthcare Interoperability Resources* (FHIR).

Più nel dettaglio, costituisce lo *standard* internazionale per lo scambio di dati sanitari, sviluppato da *Health Level Seven International* (HL7)[87]. A differenza dei formati proprietari, scompone il patrimonio informativo sanitario in **risorse** discrete e standardizzate – il paziente, la prestazione, il farmaco – ciascuna con una struttura definita e interrogabile allo stesso modo ovunque venga usata.

Nell'ambito di tali risorse, è possibile rinvenirne una pensata proprio per la sicurezza e la tracciabilità. Si tratta dell'*AuditEvent*, che registra un evento di accesso rilevante per la *privacy* o, appunto, per la sicurezza: ogni volta che qualcuno consulta una cartella, scarica un referto, o modifica un dato, il sistema genera un record che risponde alle cosiddette **5 W** dell'informatica forense, e cioè a **chi** (*who*, l'agente che ha compiuto l'azione), **cosa** (*what*, l'oggetto coinvolto), **quando** (*when*, il timestamp), **dove** (*where*, l'indirizzo di rete), e **perché** (*why*, la finalità dichiarata dell'accesso)[88].

---

[87] Per la specifica ufficiale della risorsa, HL7 *International*, *FHIR – AuditEvent Resource*, consultabile online al sito https://www.hl7.org/.

[88] Il riferimento è al metodo delle «5 W» (*Who*, *What*, *When*, *Where*, *Why*), mutuato dalla prassi investigativa e giornalistica e applicato all'analisi forense per strutturare la ricostruzione di un evento rilevante. Cfr. J. HAN, J. KIM, S. LEE, *5W1H-based Expression for the Effective Sharing of Information in Digital Forensic Investigations*, 2020, consultaible *online* al sito https://doi.org/10.32877/bt.v9i1.4053.

Peraltro, la risorsa distingue, da un lato, i vari **tipi di azione** che possono essere compiute, e cioè tra creazione, lettura, modifica, cancellazione, esecuzione (C/R/U/D/E)[89], e, dall'altro lato, i possibili **esiti** dell'operazione, e, dunque, se è andata a buon fine o è fallita.

2.6 Gli strumenti statistici e gli algoritmi di rilevamento

Prima degli algoritmi veri e propri conviene fissare i due strumenti statistici elementari su cui l'intero impianto poggia, in applicazione del quarto criterio enunciato in apertura.

Il primo è la coppia formata dalla mediana e dallo scarto interquartile[90]. La mediana è il valore che divide a metà un insieme ordinato di osservazioni; lo scarto interquartile è la differenza tra il terzo e il primo quartile, cioè l'ampiezza della fascia centrale entro cui ricade la metà dei dati. A differenza della media e della deviazione standard, nessuna delle due si sposta se qualche osservazione assume un valore estremo, ed è esattamente ciò che serve quando i valori estremi sono l'oggetto della ricerca.

Il secondo strumento è lo z-score, o punteggio standardizzato[91], che esprime la distanza di un valore dal centro della distribuzione misurata in unità di dispersione. Nella variante robusta qui adottata, centro e dispersione sono appunto la mediana e lo scarto interquartile. La sua utilità non è soltanto statistica: poiché lo scarto viene calcolato allo stesso modo per ogni grandezza, caratteristiche di natura diversa — un conteggio, un rapporto, una percentuale — diventano confrontabili sulla stessa scala, e possono quindi

---

[89] Nello standard FHIR il tipo di azione registrato in un *AuditEvent* è codificato secondo il code system «AuditEventAction»: C = *Create* (creazione di un nuovo oggetto), R = *Read*/*View*/*Print* (lettura, visualizzazione o stampa), U = *Update* (modifica), D = *Delete* (cancellazione), E = Execute (esecuzione di una funzione applicativa, comprese le operazioni di interrogazione o ricerca). Cfr. HL7 INTERNATIONAL, *FHIR – AuditEventAction Code System*, consultabile online al sito https://www.hl7.org/fhir/.

[90] Per le definizioni delle statistiche di posizione e di dispersione qui richiamate, cfr. NIST/SEMATECH, *e-Handbook of Statistical Methods*, consultabile online al sito https://www.itl.nist.gov/.

[91] Nella forma classica lo z-score esprime la distanza di un valore dalla media misurata in numero di deviazioni standard ($z = (x - \mu)/\sigma$); nella variante robusta qui adottata media e deviazione standard sono sostituite da mediana e scarto interquartile. Cfr. NIST/SEMATECH, *e-Handbook of Statistical Methods*, consultabile online al sito https://www.itl.nist.gov/.

essere ordinate tra loro. È questa proprietà a rendere possibile la spiegazione delle segnalazioni, illustrata nel capitolo III.

Prima di essere sottoposte a un algoritmo, le caratteristiche che descrivono un profilo devono essere portate sulla stessa scala, perché altrimenti una grandezza che assume valori grandi peserebbe più di una che varia tra zero e uno non in quanto più importante, ma soltanto perché i suoi numeri sono più grandi.

Lo strumento impiegato per questa operazione è un RobustScaler[92], che usa la mediana e lo scarto interquartile al posto della media e della deviazione *standard* classiche.

La ragione è la stessa che guida l'intero impianto statistico del sistema: i dati contengono anomalie, e uno *scaling* calcolato con media e deviazione *standard* si lascerebbe distorcere proprio dai valori estremi che si vogliono individuare, comprimendo il resto dei profili e rendendo più difficile distinguerli.

L'algoritmo adottato per il rilevamento delle anomalie è l'Isolation Forest.[93]

Il principio su cui si fonda il modello selezionato è diverso da quello della maggior parte degli altri metodi di rilevamento delle anomalie, che tipicamente modellano cosa sia "normale" e misurano la distanza di un punto da quel modello.

L'*Isolation Forest* compie l'operazione opposta: costruisce alberi di isolamento[94] che suddividono ripetutamente lo spazio dei dati tramite scelte casuali – una caratteristica e una soglia scelte a caso a ogni passo – fino a isolare ogni singolo punto in una partizione

---

[92] Il *RobustScaler* è uno strumento di preelaborazione dei dati nel *machine learning* che riscala ogni caratteristica usando la mediana e lo scarto interquartile (la differenza tra il terzo e il primo quartile) al posto della media e della deviazione *standard*, risultando meno sensibile ai valori estremi. Consultabile online al sito https://scikit-learn.org/.

[93] F. T. LIU, K. M. TING, Z.-H. ZHOU, *Isolation Forest*, in Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), 2008; approfondito in *Isolation-Based Anomaly Detection*, in ACM Transactions on Knowledge Discovery from Data, vol. 6, n. 1, 2012.

[94] Un albero di decisione «rappresenta una funzione che prende in ingresso un vettore di valori di attributi e restituisce una "decisione", cioè un singolo valore di uscita; i valori di ingresso e di uscita possono essere discreti o continui» (traduzione propria). È un metodo di apprendimento supervisionato non parametrico, che apprende dai dati semplici regole decisionali, dedotte dalle caratteristiche e organizzate in una struttura ad albero. Nell'Isolation Forest se ne impiega una variante non supervisionata, gli isolation tree, che isolano i punti con partizioni casuali anziché predire una variabile obiettivo. Cfr. S. J. RUSSELL, P. NORVIG, *Artificial Intelligence: A Modern Approach,* IV ed., *Global Edition, Pearson, Harlow* (Regno Unito), 2022; per l'implementazione impiegata, SCIKIT-LEARN, *Decision Trees*, consultabile online al sito https://scikit-learn.org/.

propria. L'idea è che un punto anomalo, per costruzione distante dagli altri, richieda in media un numero di suddivisioni molto più basso per essere isolato rispetto a un punto immerso nella massa dei dati normali. Bastano poche divisioni casuali per separarlo dal resto, mentre un punto tipico va isolato scavando molto più in profondità nell'albero.

Il punteggio di anomalia di un profilo è quindi funzione della sua distanza media di isolamento, calcolata su una foresta di alberi costruiti in questo modo (da qui il nome del metodo) e non richiede alcuna assunzione sulla forma o sulla distribuzione dei dati normali.

L'Isolation Forest non è stato adottato sulla base della sola letteratura, ma confrontato sperimentalmente con altri quattro metodi non supervisionati di uso corrente nel rilevamento delle anomalie. I quattro termini di paragone sono descritti qui di seguito, mentre l'esito del confronto è riportato nel capitolo III.[95]

Innanzitutto, il primo metodo è il ***Local Outlier Factor*** **(LOF)**[96]: per ogni profilo, l'algoritmo guarda quanto è "affollata" la zona in cui si trova, confrontandola con quanto sono affollate le zone dei suoi vicini più prossimi.

Se un profilo si trova in una regione molto più rada di quella dei suoi vicini, cioè è isolato rispetto al proprio vicinato, anche se in valore assoluto non è lontanissimo da nessuno, riceve un punteggio di anomalia alto. Un punteggio vicino al valore "normale" indica, invece, un profilo che vive in una zona densa quanto quella dei suoi vicini: dunque, si comporta come loro.

Il secondo metodo messo a confronto è il ***kNN-distance***[97].

Si tratta dell'idea più diretta dei cinque metodi. In questo caso, infatti, si misura semplicemente quanto un profilo sia lontano, in media, dai suoi vicini più prossimi tra i profili di riferimento.

---

[95] L'insieme dei cinque algoritmi qui posti a confronto ricalca quello adottato in un recente studio comparativo su dati simulati: E. F. AGYEMANG, *Anomaly detection using unsupervised machine learning algorithms: a simulation study*, in Scientific African, vol. 26, 2024, e02386.

[96] M. M. BREUNIG, H.-P. KRIEGEL, R. T. NG, J. SANDER, *LOF: Identifying Density-Based Local Outliers*, in *Proceedings of the* 2000 ACM SIGMOD *International Conference on Management of Data*, ACM, New York, 2000, pp. 93–104.

[97] S. RAMASWAMY, R. RASTOGI, K. SHIM*, Efficient Algorithms for Mining Outliers from Large Data Sets*, in *Proceedings of the* 2000 ACM SIGMOD *International Conference on Management of Data*, ACM, New York, 2000, pp. 427–438.

Un profilo distante dai propri pari riceve un punteggio alto; un profilo immerso in un gruppo di profili simili riceve, invece, un punteggio basso.

L'unica accortezza tecnica consiste nell'escludere il profilo da sé stesso quando fa parte dei dati di riferimento, altrimenti il suo "vicino più vicino" sarebbe sé stesso, a distanza zero, e falserebbe la media verso il basso.

Proseguendo nell'illustrazione dei metodi messi a confronto, si arriva al ***One-Class SVM***[98], cioè quell'algoritmo che cerca un confine attorno alla regione occupata dalla maggior parte dei profili normali, disegnato in modo da lasciare all'esterno solo una piccola frazione di essi.

Per fare questo, proietta i dati in uno spazio con più dimensioni, dove diventa più facile tracciare un confine netto attorno al gruppo compatto dei profili normali.

Un nuovo profilo che cade dentro quel confine è giudicato normale, mentre uno che cade al di fuori è giudicato anomalo. Nel modello utilizzato, la proiezione avviene tramite un *kernel* a base radiale (RBF)[99], che permette al confine di essere curvo anziché una semplice retta.

L'ultimo termine di paragone è il Robust Covariance.[100]

Il metodo assume che i profili normali si distribuiscano attorno a un centro secondo una forma "a nuvola" tipica di una distribuzione gaussiana, e ne stima centro e forma in modo robusto, cioè cercando di non farsi influenzare proprio dagli *outlier* che vuole individuare, usando solo il sottoinsieme di profili più coerente tra loro per calcolare centro e dispersione.

Qui, un profilo è giudicato anomalo quando si trova troppo lontano da quel centro, tenendo conto anche della forma della nuvola: non pesa allo stesso modo uno scostamento lungo una direzione in cui i dati normali variano molto e uno lungo una direzione in cui variano pochissimo.

---

[98] B. SCHÖLKOPF, J. C. PLATT, J. SHAWE-TAYLOR, A. J. SMOLA, R. C. WILLIAMSON, *Estimating the Support of a High-Dimensional Distribution*, in Neural Computation, vol. 13, n. 7, 2001, pp. 1443–1471.

[99] Un *kernel* è una funzione che misura la somiglianza tra due punti proiettandoli implicitamente in uno spazio a più dimensioni, senza calcolarne le coordinate; quello a base radiale (*Radial Basis Function*, RBF) fa dipendere tale somiglianza soltanto dalla distanza tra i due punti e consente di tracciare confini di separazione curvi anziché lineari.

[100] Il metodo stima centro e dispersione tramite lo stimatore *Minimum Covariance Determinant* (MCD). Cfr. P. J. ROUSSEEUW, K. VAN DRIESSEN, *A Fast Algorithm for the Minimum Covariance Determinant Estimator*, in Technometrics, vol. 41, n. 3, 1999, pp. 212–223.

Tra i cinque metodi, quest'ultimo costituisce quello con l'assunzione più forte sulla forma dei dati, un'assunzione che, come mostrerà il confronto sperimentale riportato nel capitolo III, i profili di questo lavoro violano.

I cinque metodi condividono l'assenza di esempi etichettati, ma differiscono nell'idea di normalità su cui si fondano: la densità del vicinato per il Local Outlier Factor, la distanza assoluta per il kNN-distance, un confine tracciato attorno alla massa dei dati per il One-Class SVM, una distribuzione di forma nota per il Robust Covariance, e la facilità di separazione per l'Isolation Forest. È proprio questa diversità di premesse a rendere il confronto informativo.

2.7 Le metriche di valutazione e la tutela dei dati trattati

Scegliere un algoritmo e dichiararlo efficace richiede un criterio di misura. Le due grandezze elementari, in un problema di rilevamento, sono la precision, cioè la quota di segnalazioni prodotte dal sistema che risultano effettivamente corrette, e il recall, cioè la quota di casi realmente anomali che il sistema riesce a segnalare. Le due sono in tensione tra loro: ampliare le segnalazioni tende ad aumentare il recall e a ridurre la precision, e viceversa. Nessuna delle due, però, può essere calcolata senza sapere quali casi siano davvero anomali, ed è un limite che il capitolo III affronta nel merito.[101]

Misurare quanto un sistema di rilevamento funzioni bene richiede una metrica indipendente da una soglia arbitraria, visto che la soglia è essa stessa una decisione operativa successiva alla valutazione del modello.

La metrica principale adottata è il **ROC-AUC**[102], che corrisponde alla probabilità che il modello assegni, a una coppia scelta a caso composta da un profilo anomalo e uno

[101] La *precision* (precisione) è la quota di segnalazioni prodotte dal sistema che risultano effettivamente corrette; il *recall* (richiamo, o sensibilità) è la quota di casi realmente anomali che il sistema riesce a segnalare. Le due grandezze sono in tensione fra loro: ampliare le segnalazioni tende ad aumentare il recall e a ridurre la *precision*, e viceversa.

[102] L'area sotto la curva ROC (*Receiver Operating Characteristic*) misura, in un unico valore compreso tra 0 e 1, la capacità di un modello di ordinare correttamente i casi anomali rispetto a quelli normali, indipendentemente dalla soglia di decisione adottata.

normale, un punteggio più alto al primo. Vale 0,5 quando il modello non fa altro che indovinare a caso, 1,0 quando ordina perfettamente ogni anomalo sopra ogni normale, indipendentemente da dove si scelga di tracciare la soglia operativa. Costituisce una misura della capacità di separazione, non di una singola decisione.

Una seconda misura più severa è l'***Average Precision***[103], che valuta quanto le anomalie siano concentrate nelle primissime posizioni della classifica, non semplicemente separate dal resto. Si tratta di una metrica più **sensibile** tutte le volte in cui, come in questo caso, la classe di interesse è rara, cioè pochi profili anomali su un numero molto più grande di profili normali, perché penalizza duramente ogni normale che si infili prima di un'anomalia, anche se la separazione complessiva resta buona.

L'**accuratezza** è stata deliberatamente esclusa, poiché, con una prevalenza di anomalie di pochi punti percentuali, un modello che non segnala mai nulla otterrebbe un'accuratezza altissima pur essendo del tutto inutile: la metrica premierebbe l'inazione, non la capacità di rilevamento.

Una terza scelta metodologica riguarda la soglia usata quando si vuole calcolare *precision* e *recall* a un singolo punto operativo, anziché lungo tutte le soglie possibili come fa il ROC-AUC.

La **soglia** non viene fissata a un percentile[104] arbitrario, ma allineata alla prevalenza delle anomalie nel campione valutato: si segnala una quota di profili pari, in proporzione, al numero di anomalie effettivamente presenti.

La **ragione** è **aritmetica**, non statistica: se si segnalassero meno profili rispetto a quante sono le anomalie iniettate, nessun modello, pur essendo perfetto, potrebbe raggiungere un *recall* pieno, perché mancherebbe fisicamente la possibilità di includerle

---

Cfr. T. FAWCETT, *An introduction to ROC analysis*, in Pattern Recognition Letters, vol. 27, n. 8, 2006, pp. 861–874.

[103] L'*Average Precision* riassume in un unico valore la curva *precision-recall*, cioè l'andamento della precisione al variare del recall; su insiemi fortemente sbilanciati, in cui la classe di interesse è rara, tale curva è più informativa di quella ROC.
Cfr. J. DAVIS, M. GOADRICH, *The Relationship Between Precision-Recall and ROC Curves*, in Proceedings of the 23rd International Conference on Machine Learning (ICML), 2006, pp. 233–240.

[104] Il percentile è il valore al di sotto del quale ricade una data percentuale delle osservazioni di una distribuzione ordinata: il novantesimo percentile, ad esempio, è il valore che lascia sotto di sé il 90% dei dati. Fissare una soglia «a un percentile» significa quindi segnalare una quota prestabilita di profili, in questo caso, i più anomali, indipendentemente dal loro punteggio effettivo.

tutte nella lista dei segnalati. Una soglia troppo stretta misurerebbe, in quel caso, la soglia stessa, non la qualità del modello.

Un ultimo strumento non riguarda il rilevamento, bensì la protezione dei dati che il rilevamento stesso maneggia. Un sistema che analizza i *log* di accesso è a sua volta un trattamento di dati personali: i *log* registrano chi ha consultato quale paziente, quando e da quale postazione, e da essi si risale a comportamenti e relazioni delle persone osservate.

Per questa ragione la catena comprende un componente dedicato alla pseudonimizzazione degli identificativi degli operatori[105]. La pseudonimizzazione è definita dal GDPR come il trattamento dei dati personali in modo tale che non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, conservate separatamente e soggette a misure che ne impediscano l'attribuzione; è inoltre espressamente richiamata dall'articolo 32 tra le misure tecniche idonee a garantire la sicurezza del trattamento. Non è anonimizzazione: il legame con la persona non viene distrutto, ma reso indisponibile a chi non debba conoscerlo.

Nel lavoro, i nomi utente reali sono sostituiti da etichette convenzionali stabili, in modo che lo stesso operatore resti riconoscibile da una giornata all'altra — condizione necessaria, perché il confronto con la propria storia non avrebbe altrimenti senso — mentre la corrispondenza tra etichetta e identità reale è mantenuta separata. L'analisi e la discussione dei profili anomali avvengono così senza esporre l'identità delle persone coinvolte, e il nome reale viene ricongiunto all'etichetta soltanto nel momento in cui un caso debba essere effettivamente verificato da chi ne ha titolo.

[105] Art. 4, n. 5), del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, "Definizioni": «"pseudonimizzazione": il trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile».

CAPITOLO III

PIPELINE E VALIDAZIONE

SOMMARIO: 3.1 DALLE FIRME COMPORTAMENTALI ALLE GRANDEZZE MISURABILI; 3.2 ESIGENZE E ARCHITETTURA DELLA PIPELINE; 3.3 ACQUISIZIONE E NORMALIZZAZIONE DEI LOG; 3.4 DAL SIGNIFICATO DELLE OPERAZIONI ALLE CARATTERISTICHE COMPORTAMENTALI; 3.5 IL RILEVAMENTO DELLE ANOMALIE; 3.5.1 (SEGUE) DALL'EVENTO AL PROFILO; 3.5.2 (SEGUE) IL MODELLO E LA SPIEGABILITÀ DELLE SEGNALAZIONI; 3.5.3 (SEGUE) DALLA STRANEZZA ALLA GRAVITÀ: LA PRIORITÀ OPERATIVA; 3.6 L'ESECUZIONE DELLA PIPELINE E I SUOI RISULTATI; 3.7 PROTOTIPO DI RILEVAMENTO IN TEMPO REALE; 3.8 LA VALIDAZIONE; 3.8.1 (SEGUE) IL PROBLEMA DELLA VALUTAZIONE E IL METODO ADOTTATO; 3.8.2 (SEGUE) IL CONFRONTO TRA I CINQUE ALGORITMI; 3.8.3 (SEGUE) STABILITÀ DELLE SEGNALAZIONI ED ESCLUSIONE DELLE ANOMALIE REALI; 3.9 I LIMITI DEL SISTEMA.

3.1 Dalle firme comportamentali alle grandezze misurabili

Le cinque ricorrenze descritte nel capitolo precedente sono state finora considerate dalla prospettiva del soggetto che commette l'abuso.

Ai fini della costruzione di un sistema di rilevamento, è però necessario adottare una prospettiva diversa e individuare, per ciascun movente, quali tracce osservabili esso lasci nei *log* e in che modo tali tracce possano essere tradotte in grandezze misurabili. La tabella seguente sintetizza questo passaggio, associando a ciascuna firma comportamentale i moventi da cui può originare e la relativa variabile utilizzata dal modello per rilevarla.

Le definizioni tecniche di tali grandezze saranno trattate nei paragrafi che seguono. In questa sede, è sufficiente mostrare il legame tra ciascuna firma comportamentale e la grandezza misurabile che il modello impiega per coglierla. Difatti, ogni grandezza nasce da un comportamento realmente osservato nei provvedimenti del Garante, e non da una scelta statistica astratta.

Firma comportamentale	Moventi generanti	Grandezza misurabile
Accesso a oggetti fuori dal proprio bacino abituale	Sfera affettiva (Trento); collega-paziente (Verona, Bolzano)	Scostamento dai pari del reparto (*pg_ratio*) e novità dell'oggetto
Volume di accessi superiore alla propria routine	Curiosità esplorativa; collega-paziente	Scostamento dalla propria storia (*self_ratio*)
Accanimento su un singolo oggetto	Ripetizione ossessiva (Trieste, 40 accessi)	Rapporto tra eventi e oggetti distinti (*fanout*)
Consultazione di aree particolarmente sensibili	Collega-paziente con patologie stigmatizzanti (Bolzano, HIV)	Livello di sensibilità dei dati toccati (quota_sensibili)
Attività fuori dall'orario ordinario	Accessi non giustificati da turni di servizio	Quota di accessi in fascia anomala (quota_fuori_orario)
Ricorrenza su più giornate*	Sfera affettiva; movente strumentale	— (aggravante di priorità, non feature)

Tabella 3.1: Dalle firme comportamentali alle grandezze misurabili

* la ricorrenza su più giornate, come le scritture anomale (n_scritture) e l'uso di più postazioni/IP nella stessa giornata (ip_distinti), non è una feature del modello di rilevamento, ma un **aggravante** applicato nello strato di priorità operativa, descritto al 3.5.3.

Fa eccezione la curiosità sporadica: proprio perché si manifesta attraverso un numero limitato di accessi, non produce una propria firma di volume e può essere intercettata solo indirettamente, attraverso eventuali scostamenti rispetto al comportamento dei propri pari o alla storia individuale dell'utente. È per questo motivo che, come già evidenziato, rappresenta la ricorrenza più difficile da rilevare.

Si nota che nessuna singola grandezza esaurisce il fenomeno: una stessa firma comportamentale può essere riconducibile a moventi diversi e, allo stesso tempo, un singolo episodio di abuso può lasciare più tracce contemporaneamente.

È proprio questa molteplicità a rendere necessario descrivere ciascuna giornata di attività, non attraverso un unico indicatore, ma mediante un vettore di caratteristiche comportamentali, ciascuna sensibile a una diversa dimensione dell'abuso. La costruzione di tali caratteristiche a partire dai dati grezzi e la loro rappresentazione in una forma che ne consenta il calcolo costituiscono l'oggetto dei paragrafi che seguono.

3.2 Esigenze e architettura della pipeline

Il problema è stato finora osservato da due prospettive distinte.

Il capitolo I ha mostrato che, di fronte a un operatore curioso ma regolarmente autenticato e autorizzato, non è possibile stabilire a priori se un accesso sia illecito, ma è possibile misurare quanto esso si discosti dal comportamento abituale dell'utente e da quello dei suoi pari.

Il paragrafo precedente ha tradotto questa intuizione in termini concreti, ricavando dai provvedimenti del Garante un insieme di firme comportamentali ricorrenti e associando a ciascuna la grandezza misurabile che la rende osservabile.

Una volta chiarito questo passaggio, rimane da affrontare la parte più impegnativa, quella di trasformare tali firme in un sistema funzionante, a partire non da dati ideali, ma dai *log* così come vengono prodotti da un applicativo sanitario reale. Le scelte progettuali della *pipeline* che sarà descritta in questo capitolo nascono proprio dalla distanza tra ciò che le firme richiederebbero e ciò che i dati grezzi realmente offrono.

Procedendo per ordine, occorre specificare quale sia la natura dei dati di partenza.

I ***log* applicativi** sono scritti per finalità di diagnostica, non di *audit*, e non dicono in modo esplicito e uniforme chi ha fatto cosa, su quale paziente e con quale grado di delicatezza. Prima di poter cercare qualunque anomalia occorre dunque ricondurli a una forma comune e semanticamente esplicita: è la ragione per cui il primo tratto della catena acquisisce i *log* senza alterarli e li normalizza nello standard **FHIR *AuditEvent***.

Le firme, però, non parlano soltanto di accessi, ma di accessi a dati di diversa sensibilità, come la cartella di un collega, o il referto di un'area stigmatizzante. Perché il sistema possa pesare un accesso occorre allora attribuire a ogni operazione un significato clinico e un livello di sensibilità, derivato dalle categorie particolari di cui all'articolo 9 del GDPR illustrate nel capitolo I. Questo sarà il compito del **catalogo semantico**.

Come anticipato, c'è poi un tratto comune a tutte le firme, il fatto che nessuna di esse sia un **valore assoluto**. Un accesso non è anomalo perché numeroso, ma perché si discosta dalla routine di quell'operatore e da quella dei suoi colleghi. Questo impone di non giudicare il singolo evento, bensì di aggregare gli accessi per utente e per giornata e di descriverli con un insieme di caratteristiche comportamentali, costruite proprio per rendere quel confronto possibile.

Infine, due vincoli attraversano l'intero impianto.

Il primo è l'**assenza di etichette**: siccome nessun insieme di accessi è mai stato dichiarato lecito o illecito, il rilevamento non può che essere non supervisionato, affidato a un algoritmo capace di isolare i profili fuori norma senza bisogno di esempi di abuso.

Il secondo vincolo è la **spiegabilità**, cioè la capacità di un sistema che esprime un sospetto su una persona di poter essere interrogato su ogni sua decisione. Infatti, ogni segnalazione è accompagnata dai fattori che l'hanno determinata e una priorità governata da regole dichiarate, non apprese dai dati.

Le esigenze di normalizzare, dare significato, confrontare, isolare, spiegare, si traducono in diversi stadi che saranno descritti nei paragrafi successivi.

Sul piano architetturale, queste cinque esigenze prendono la forma di una catena di elaborazione a stadi: ogni fase riceve in ingresso l'output della precedente, applica una trasformazione precisa e produce un dato più strutturato e più ricco di significato di quello che ha ricevuto.

Il percorso completo parte dall'acquisizione dei *log* applicativi grezzi, scritti dall'*interceptor* per finalità di diagnostica e registrati in formato JSON in una **collezione di MongoDB**, fino ad arrivare alla classifica di profili comportamentali, ciascuno corredato dalla spiegazione dei motivi che lo rendono anomalo, passando per la loro traduzione nello standard *FHIR AuditEvent*, per l'arricchimento tramite il catalogo

semantico, per l'aggregazione in caratteristiche comportamentali e, infine, per il rilevamento delle anomalie con l'assegnazione di una priorità operativa.

Questa suddivisione in stadi distinti, ciascuno realizzato da uno script autonomo, non è una scelta di comodo organizzativo bensì una condizione di correttezza.

Nessuna fase modifica i dati che riceve, infatti ciascuna scrive il proprio risultato in una sede separata, lasciando intatto quello prodotto a monte. Se la traduzione in FHIR contenesse un errore, i *log* grezzi originali resterebbero intatti in MongoDB, e la traduzione potrebbe essere corretta e rieseguita senza perdita di informazione.

Lo stesso vale per il catalogo delle operazioni, infatti, la classificazione automatica può essere rigenerata in qualunque momento senza cancellare le decisioni verificate a mano, che vivono in un file separato e vengono riapplicate a ogni esecuzione.

Dato che ogni stadio è ripetibile, ispezionabile e correggibile in isolamento, è una proprietà che si potrebbe chiamare di **non distruttività**: non esiste un solo punto della catena in cui un errore a valle possa danneggiare il dato a monte.

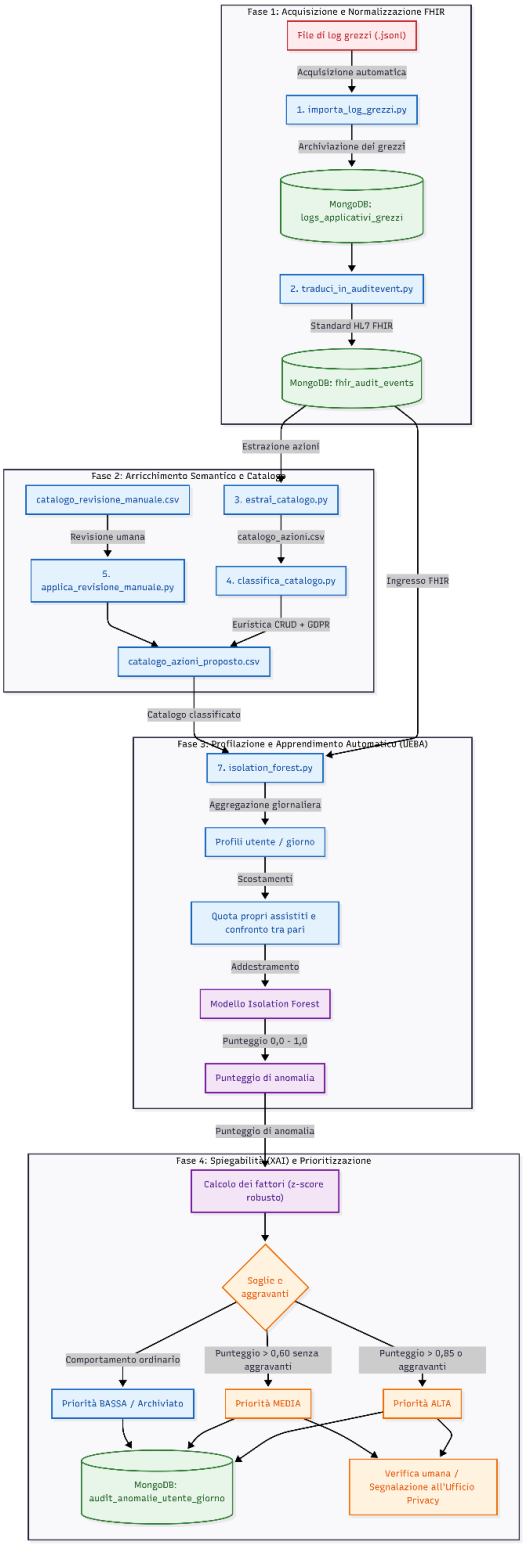


Figura 3.1: Pipeline completa

La Figura 3.1 riassume, tramite un diagramma a blocchi, l'intero percorso, dai *log* grezzi fino alla segnalazione finale; ciascuna delle fasi che vi compaiono sarà ripresa e approfondita nelle sezioni che seguono.

3.3 Acquisizione e normalizzazione dei log

Il primo stadio della *pipeline* ha un compito semplice: portare i *file* di log, così come li produce l'*interceptor* applicativo, dentro un *database*, con l'obiettivo specifico di **non modificarne il contenuto**.

Si tratta di una scelta di prudenza. Difatti, i *file* arrivano via **SFTP** in cartelle monitorate, in formato JSON o JSONL, e sono l'unica traccia di ciò che è realmente accaduto sul sistema. Un errore introdotto qui si trascinerebbe, senza possibilità di correzione, in tutti gli stadi successivi. Per questo l'acquisizione non interpreta né filtra nulla: legge, controlla che ogni riga sia un JSON valido, e scrive.

Il primo ostacolo trovato qui sembrava piccolo, e invece bloccava vari JSONL: alcuni *file* non venivano importati, senza che l'errore dicesse in quale riga risiedesse il problema. La causa è stata individuata nella **codifica dei caratteri**: alcuni *log* contengono note testuali con lettere accentate salvate non in UTF-8 ma nella vecchia codifica *Windows*, la cp1252. Bastava **un solo carattere non valido**, in un punto qualsiasi del *file*, per far fallire la lettura di tutto il contenuto. In questo modo, il rischio sarebbe stato quello di perdere migliaia di eventi a causa di una singola lettera accentata.

Di conseguenza, la soluzione è stata rinvenuta in una lettura tollerante: anziché interrompere l'elaborazione, i *byte* non leggibili vengono sostituiti con un segnaposto e viene tenuto il conto di quante righe abbiano avuto bisogno di questa correzione.

Si tratta di un rimedio a posteriori per un problema che nasce alla base, nella qualità del dato così com'è prodotto. Appare chiaro come la soluzione più corretta sarebbe quella di uniformare la codifica in fase di esportazione, ma fino a quando ciò non succede la *pipeline* deve comunque riuscire ad andare avanti senza buttare via un intero *file* per un difetto puntuale.

Anche il **formato dei *file* non è sempre lo stesso**. Alcuni sono un unico *array* JSON tra parentesi quadre, altri sono una sequenza di oggetti separati da una virgola, secondo una convenzione JSONL non del tutto rigorosa. Lo *script* riconosce la forma del *file* dal suo contenuto e si adatta di conseguenza, gestendo anche righe troncate o vuote, che vengono contate e segnalate, non ignorate in silenzio.

Un secondo problema riguarda la **ripetibilità** dell'importazione.

I *file* possono essere scaricati più di una volta, oppure il processo può fermarsi e ripartire. Pertanto, è necessario garantire che la scrittura in MongoDB non crei mai duplicati, neppure rieseguendo tutto da capo.

È per effetto delle garanzie del filtro applicativo, descritte nel capitolo II, che lo stadio di importazione, primo stadio effettivamente sviluppato nell'ambito di questo lavoro, non può assumere che ogni *file* arrivi una volta sola: deve poter riconoscere un evento già acquisito ed evitarne la duplicazione. In questo progetto, infatti, l'obiettivo è quello di arrivare a garantire l'idempotenza, cioè la proprietà di un'operazione che, ripetuta più volte sugli stessi dati, produce nel caso di specie sempre il medesimo risultato di una singola esecuzione senza mai generare duplicati.

La soluzione usa l'identificativo univoco già presente in ogni *log* come chiave del documento: prima di scrivere, lo *script* controlla quali identificativi sono già in MongoDB, e inserisce solo quelli nuovi. I *log* non cambiano mai una volta registrati, quindi non c'è alcun bisogno di aggiornarli, ma solo di inserirli, a patto che non siano già presenti.

Una volta importato, ogni *file* viene spostato in una cartella separata, così da evitare che sia riletto la volta successiva. Se un *file* con lo stesso nome è già stato archiviato in precedenza, gli viene aggiunto un suffisso con la data e l'ora, invece di sovrascriverlo senza dire nulla.

Il processo, infine, può girare in due modi: **a blocchi**, scandendo una volta la cartella all'avvio, oppure **in continuo**, restando in ascolto di nuovi arrivi tramite un osservatore di *file system*. È questa seconda modalità quella pensata per l'uso reale, dove i *log* arrivano di continuo e non a pacchetti.

I *log* grezzi, una volta in MongoDB, mantengono comunque il loro **formato originale**: nascono per la diagnostica dell'applicativo, non per il controllo degli accessi. Ogni evento porta con sé decine di campi tecnici, come *cookie* di sessione, parametri interni, codici applicativi, etc. ma non dice, in modo esplicito e uniforme, ciò che serve davvero: **chi ha fatto cosa, quando, e su quale paziente**.

Il secondo stadio della pipeline traduce ogni *log* in questa informazione servendosi della risorsa AuditEvent dello standard FHIR, descritta nel capitolo II. Delle cinque dimensioni che la risorsa prevede, però, i *log* disponibili ne valorizzano soltanto quattro.

La finalità dell'accesso, pur prevista dallo standard, non è valorizzata nei *log* analizzati. Come già osservato, essi registrano automaticamente le sole tracce tecniche della richiesta, cioè chi, cosa, quando, dove, ma non includono la motivazione dell'accesso, visto che l'applicativo non chiede mai all'operatore di esplicitarla.

In altre parole, quindi, gli altri quattro elementi sono sempre presenti, mentre il **perché** resterebbe rilevabile solo se il sistema imponesse una dichiarazione esplicita di finalità, ad esempio tramite un meccanismo di *break-the-glass*.

È proprio una tale assenza a rendere necessario l'approccio comportamentale adottato in questo lavoro: non potendo conoscere la finalità dichiarata di un accesso, il sistema la inferisce indirettamente, misurando quanto quell'accesso si discosti dalla norma.

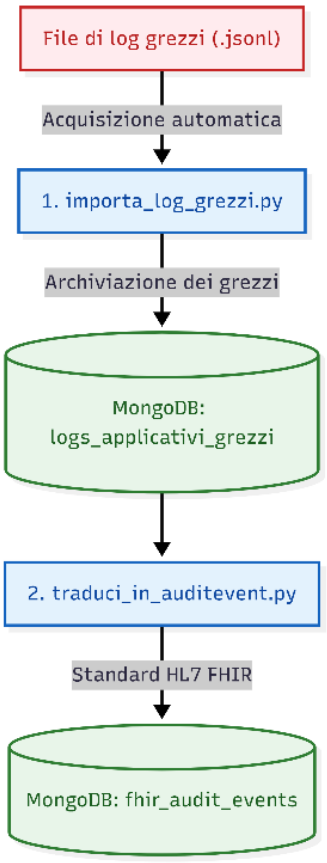


Figura 3.2: Fase 1- da log grezzi ad AuditEvent

Tradurre un *log* grezzo in un *AuditEvent* significa quindi mappare ogni campo proprietario sul campo FHIR corrispondente: l'*username* diventa l'**agente**, il *timestamp*

diventa il campo **recorded**, il codice di risposta HTTP[106] diventa l'**esito**. Buona parte di questa traduzione è meccanica. Tuttavia, in due casi questo non accade, perché i dati reali non si comportano come ci si aspetterebbe da uno schema pulito.

Innanzitutto, il **primo problema** riguarda il fatto che l'oggetto acceduto non è sempre il paziente. La logica più naturale sarebbe registrare, per ogni accesso, l'identificativo del paziente coinvolto. Ma nei dati reali, rilevati nel corso della stesura della tesi, questo non è quasi mai il caso: solo il 20% degli accessi porta con sé l'identificativo diretto del paziente (ID_ASSISTITO); il 77% porta invece l'identificativo della pratica, come il fascicolo amministrativo aperto per un ricovero o un ciclo di esami; mentre il restante 3% non comprende nessuno dei due. Se il sistema avesse cercato solo il paziente, avrebbe perso quasi otto accessi su dieci.

La soluzione è quella di rappresentare sempre l'oggetto acceduto, qualunque esso sia: se è presente il paziente si usa il suo identificativo, perché è l'informazione più diretta per capire chi è stato guardato; se è assente, si confida nell'identificativo della pratica.

Nel codice, tale priorità è resa esplicita con una struttura a cascata: prima il paziente, successivamente la pratica, e infine nessun oggetto identificabile. Inoltre, il risultato viene trascritto in due campi uniformi (*accessed-object-type* e *accessed-object-id*) che vengono letti dagli stadi successivi senza la necessità di doversi preoccupare di quale delle due fonti abbia fornito il dato.

Il **secondo problema** risiede nella consapevolezza per cui l'operazione, in realtà, non sempre è dove dovrebbe essere. Per questo motivo, l'applicativo espone un parametro dedicato, chiamato ***sqlcmd***, pensato per contenere il nome dell'operazione applicativa invocata.

Tuttavia, in molte richieste, questo parametro manca: nei dati analizzati, il **39%** degli accessi non lo riporta, perché gli stessi non passano attraverso il punto del sistema in cui viene registrata l'operazione. In tutti questi casi, quindi, il nome dell'operazione si trova altrove, e cioè, nell'**indirizzo *web*** della richiesta, come ultimo segmento del percorso, prima di un'eventuale estensione come .jsp.

---

[106] I codici di stato HTTP sono codici numerici a tre cifre con cui un *server* indica l'esito di una richiesta, raggruppati in cinque classi (1xx informativi, 2xx successo, 3xx reindirizzamento, 4xx errore del *client*, 5xx errore del *server*) e definiti dallo standard IETF RFC 9110. Cfr. MOZILLA DEVELOPER NETWORK (MDN), *HTTP response status codes*, consultabile online al sito https://developer.mozilla.org/.

La soluzione cerca l'operazione in entrambi i posti: prima nel parametro dedicato, e ove esso sia assente, nell'URL, tenendo traccia di quale delle due fonti sia stata utilizzata. Grazie a questo doppio tentativo, l'operazione risulta identificata nel **100%** degli accessi analizzati: nessuno resta privo di nome e, senza un'operazione tracciata, l'intero rilevamento successivo, che si basa proprio sul confronto tra le operazioni compiute, non avrebbe potuto funzionare.

Oltre a questi due problemi, la traduzione conserva anche il **contesto organizzativo** dell'operatore — reparto, unità, profilo utente — che nei *log* grezzi era presente ma non veniva utilizzato. È un'informazione che a prima vista sembra secondaria, ma è quella che nei paragrafi che seguono permette di confrontare il comportamento di un utente con quello dei suoi pari: senza il reparto, non ci sarebbe nulla con cui confrontarlo.

3.4 Dal significato delle operazioni alle caratteristiche comportamentali

Un *AuditEvent*, a questo punto della *pipeline*, sa chi ha fatto cosa, quando e su quale oggetto. Ma non sa ancora cosa significhi quell'operazione dal punto di vista clinico.

Un nome come *a32an_qry_infettivologia* non dice nulla, di per sé, a chi legge un report: è necessario sapere che riguarda l'area delle malattie infettive, che è quindi un dato particolarmente delicato, e che, a giudicare dal prefisso qry, è probabilmente una lettura. Senza questo significato, il sistema potrebbe sapere che un'operazione è avvenuta, ma non quanto conta.

Il terzo, il quarto e il quinto stadio della *pipeline* costruiscono questo significato.

Innanzitutto, viene estratto l'**inventario** di tutte le operazioni distinte presenti nei dati, insieme alla loro frequenza.

Successivamente, vengono associate a ognuna di esse tre informazioni: l'**azione** compiuta – se si tratta di una lettura, una scrittura, una modifica, o una cancellazione – l'**area clinica** a cui appartiene, e un **livello di sensibilità**.

Il criterio scelto per la sensibilità non guarda al tipo di dato in sé, ma a **quanto sarebbe grave se qualcuno lo consultasse senza titolo**. Non tutte le informazioni cliniche pesano allo stesso modo: un dato anagrafico e una diagnosi di infezione da HIV sono ambedue

“dati sanitari”, ma le conseguenze di un accesso improprio sono di gravità del tutto diversa.

Questo criterio riprende direttamente le **categorie particolari** di cui all’articolo 9 del GDPR: le aree cliniche più delicate, come quelle concernenti le malattie infettive, il fine vita, o la salute mentale, ricevono sensibilità alta. Tutte le altre, invece, a seconda dei casi sono classificate come standard o media.

Con oltre trecento operazioni distinte nei dati, classificarle una per una a mano non è pensabile. Il sistema procede quindi in modo **semi-automatico**.

In un primo momento, ogni operazione viene classificata basandosi sulla **convenzione di *naming***: se il nome contiene “qry”, o sue varianti come “qy”/ “qty”/ “q_”, è probabilmente una lettura. Se contiene, invece, “infett” o “gravidanz”, appartiene con ogni probabilità a un’area clinica specifica.

È un’euristica, non una certezza, e proprio per questo lascia scoperto un caso sistematico: quando il nome di un’operazione non contiene nessuna delle parole chiave riconosciute, l’euristica la classifica genericamente come “esecuzione indeterminata”, anche quando si tratta chiaramente di una scrittura.

È il caso di *a32_compilazione_sdac_amm*: il nome indica con chiarezza un’azione di compilazione, quindi di scrittura, ma non contiene nessuno dei verbi che l’euristica riconosce esplicitamente. Così, l’automatismo la marca come incerta e la segnala per la verifica. In questo modo, la revisione manuale le assegna poi l’azione corretta, che nel caso di specie è l’aggiornamento.

Il meccanismo funziona anche nella direzione opposta, quando l’automatismo è già affidabile: un’operazione come *a32an_qry_infettivologia* viene riconosciuta correttamente, senza bisogno di intervento umano, sia come lettura (dal prefisso “qry”), sia come appartenente all’area delle malattie infettive, ricevendo quindi sensibilità alta per effetto della regola di area.

Per questo esiste un **secondo livello**, che ha la precedenza sul primo: un catalogo di **decisioni verificate a mano**, dove le operazioni classificate in maniera non esatta dall’euristica vengono corrette.

Il *file* che raccoglie queste correzioni non viene mai toccato dalla *pipeline* automatica: resta separato e viene **riapplicato** ogni volta che il catalogo automatico viene rigenerato, così il lavoro di verifica non si perde mai.

Sono ammesse sia correzioni puntuali, su una singola operazione, sia regole più generali[107], che si applicano automaticamente anche a operazioni non ancora comparse nei dati, se dovessero farlo in futuro.

Questa classificazione, automatica o corretta a mano che sia, si basa interamente sul **nome** dell'operazione e sulle convenzioni con cui è stata scritta.

Va chiarito che non si tratta di una verifica di ciò che l'operazione fa realmente, visto che ciò richiederebbe l'accesso al codice sorgente del framework applicativo, di cui non si è disposto in questo lavoro. È quindi una stima, non un accertamento: dove il nome è ambiguo, la classificazione resta incerta, e viene segnalata come tale piuttosto che presentata con una sicurezza che in realtà non possiede.

Il meccanismo appena descritto solleva una domanda naturale: nella pratica, quanto lavoro di verifica umana serve affinché il sistema funzioni correttamente?

Per rispondere, la *pipeline* è stata eseguita due volte sugli stessi dati, cambiando soltanto il *file* delle decisioni manuali: una prima volta con il *file* effettivamente usato – ventitré correzioni puntuali e tre regole di area, concentrate sui casi in cui l'euristica automatica risultava errata –, una seconda volta con un *file* arricchito artificialmente, che aggiungeva più di cento righe, riportanti, però, le stesse decisioni già prese dall'automatismo, senza introdurre alcuna correzione ulteriore.

Con riferimento ai profili analizzati, ai punteggi di anomalia, alle priorità assegnate, e alle metriche di validazione, i risultati delle due esecuzioni si sono rivelati **identici** in ogni cifra.

La conclusione è che le righe aggiunte non modificavano nulla, perché si limitavano a ripetere ciò che il catalogo automatico calcolava già correttamente da solo: un *file* di revisione più corto ma mirato produce esattamente lo stesso risultato di uno più lungo ma ridondante.

[107] Un esempio di regola generale può essere quella che stabilisce quanto segue: "tutte le operazioni dell'area Malattie infettive sono ad alta sensibilità".

Questo conferma, con un riscontro empirico, il principio con cui il meccanismo di revisione manuale è stato concepito: il suo valore sta nel correggere le eccezioni che l'automatismo sbaglia, non nel duplicare le decisioni che l'automatismo già prende correttamente. Un *file* di revisione più esteso non è, di per sé, un *file* più efficace.

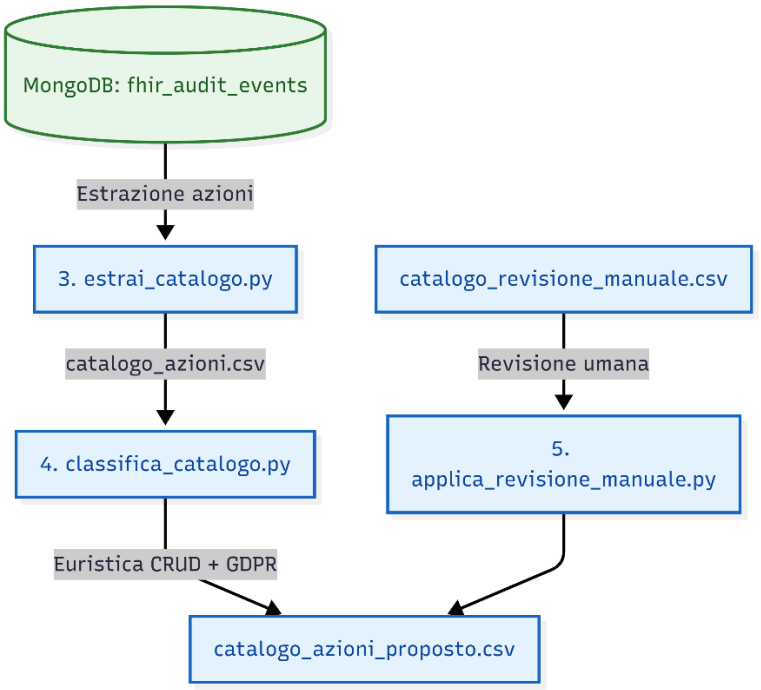


Figura 3.3: Fase 2 - Studio e arricchimento catalogo

Un *AuditEvent*, tradotto e arricchito di significato, descrive un singolo accesso. Tuttavia, un singolo accesso, da solo, **non dice quasi nulla**: uno *snooper* e un operatore che lavora normalmente possono generare, in un istante isolato, *log* del tutto identici.

Il comportamento anomalo emerge solo guardando l'**insieme degli accessi** di una persona in un certo periodo, e confrontandolo con un riferimento.

Questo stadio della *pipeline* trasforma gli *AuditEvent* in un insieme di **caratteristiche numeriche**, le c.d. *feature*[108], pensate apposta per rendere quel confronto possibile.

[108] Una *feature* è una proprietà misurabile del dato utilizzata come *input* di un modello; il *feature engineering* è il processo di selezione, trasformazione e costruzione di tali caratteristiche a partire dai dati grezzi, allo scopo di migliorare le prestazioni del modello. Cfr. IBM, *What is Feature Engineering?*, consultabile online al sito https://www.ibm.com/.

Le *feature* sono state progettate per restituire sempre un valore continuo – quanti accessi, quanti oggetti distinti, quale percentuale – e mai un indicatore già ridotto a "sì/no" da una soglia fissa incorporata nel calcolo.

La ragione risiede nel fatto che una soglia è già una decisione, e decidere non è compito della *feature*: se il numero che fa scattare un allarme fosse scelto una volta per tutte dentro il calcolo della caratteristica, cambiarlo richiederebbe modificare il codice che produce il dato. Tenendo la *feature* come valore continuo, la soglia si sposta più avanti, nel livello che si occupa davvero di decidere. Si tratta della stessa logica già vista per la priorità operativa, dove le regole di gravità restano separate dal punteggio del modello.

Per quanto riguarda l'unità di analisi, la *pipeline* aggrega gli accessi per giornata, non limitandosi al singolo evento.

I portali applicativi generano un fenomeno di ***fanout***: aprire una singola cartella clinica comporta decine di chiamate distinte al sistema, una per ogni sezione caricata. Contare gli eventi grezzi, quindi, non misura quanto un operatore ha lavorato, ma quanto il portale ha risposto. Ad esempio, due operatori con lo stesso numero di eventi possono aver avuto comportamenti opposti: uno che ha aperto una sola cartella e l'ha esaminata a fondo, l'altro che ne ha scorse venti superficialmente.

Il segnale onesto non è il numero di chiamate, ma **quanti oggetti distinti** un operatore ha toccato, come pazienti o pratiche. Per questo, come anticipato, la *pipeline* aggrega gli accessi **per utente e per giornata**: unità sufficientemente ampia da neutralizzare il *fanout* e abbastanza piccola da restare leggibile in un *report* di *audit*.

Il rapporto tra le due grandezze, cioè tra eventi e oggetti distinti, è esso stesso una caratteristica utile. Difatti, un valore alto descrive un accesso concentrato, che si sofferma a lungo su pochi oggetti; un valore basso, invece, descrive un accesso esplorativo, distribuito su molti. Non è di per sé un indicatore di gravità, ma di **stile**: rivela come un profilo sia anomalo, anche se non di quanto lo sia.

Da ogni giornata di lavoro di ogni utente vengono calcolate tredici caratteristiche, raggruppabili in sei macrocategorie:

Innanzitutto, la prima categoria fa riferimento al **volume e** all'**ampiezza**, in particolare al numero di eventi, alla quantità di oggetti distinti, alla presenza di operazioni distinte, e al *fanout*.

In secondo luogo, si presta attenzione alla **sensibilità dei dati consultati**. Nello specifico, la quota di accessi che tocca almeno dati di sensibilità media e il numero di oggetti di sensibilità alta. Si tratta di due soglie diverse, perché la prima misura il peso complessivo della giornata, la seconda l'esposizione ai dati più delicati in senso stretto.

Inoltre, è importante considerare la **novità**, e cioè quanti oggetti, tra quelli consultati, non erano mai stati toccati prima da quello stesso utente.

In quarto luogo, si rinviene nel **contesto temporale** un'altra categoria. È bene osservare la quota di attività svolta fuori dall'orario ordinario o nel fine settimana.

Non è tutto. Anche i **segnali operativi e di sicurezza** possono rivelarsi utili. Infatti, vanno considerati anche il numero di operazioni di scrittura, il numero di postazioni distinte da cui l'utente ha operato (indirizzi IP) e la quota di operazioni fallite.

Infine, l'ultima macrocategoria è quella dello **scostamento dai pari e dalla propria routine**, che valuta quanto gli oggetti consultati da parte dello stesso utente nel giorno attuale si discostano dalla mediana dei giorni passati e dalla mediana dei colleghi del suo reparto.

Le ultime due, rispettivamente lo scostamento dalla propria routine e quello dai colleghi di reparto, sono i segnali più forti dell'intero insieme.

Entrambe richiedono alcune cautele di calcolo: da un lato, un confronto costruito in modo causale (guardando solo al passato), e, dall'altro lato, l'esclusione degli *account* artificiali dal riferimento di normalità, da cui dipende la loro correttezza.

Trattandosi di scelte che riguardano il momento del rilevamento, più che la costruzione della singola caratteristica, il loro dettaglio implementativo, con il relativo codice, sarà discusso nei paragrafi che seguono.

3.5 Il rilevamento delle anomalie

A questo punto della *pipeline*, ogni accesso è un *AuditEvent* completo di significato: si sa chi lo ha compiuto, su cosa, quando, e con quale grado di sensibilità clinica.

Il settimo e ultimo stadio prende questi accessi e li trasforma in una classifica di comportamenti anomali, ciascuno con un punteggio, una spiegazione e una priorità operativa.

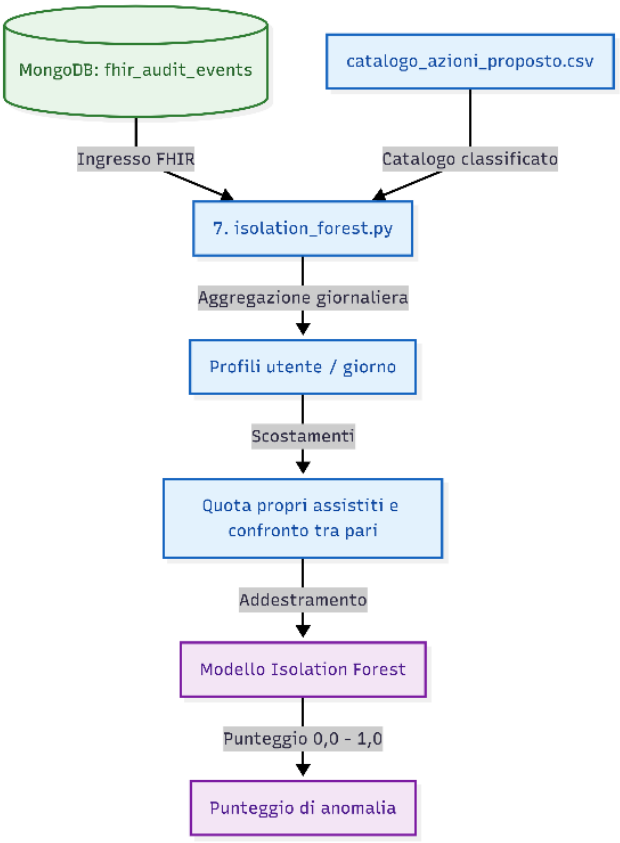


Figura 3.4: Fase 3 - Isolation Forest

3.5.1 Dall'evento al profilo

Il primo passo aggrega gli accessi per **utente e per giornata**: da migliaia di eventi singoli si passa a un numero molto più piccolo di profili, uno per ogni combinazione di persona e giorno lavorato. Si tratta della scelta già discussa nel paragrafo precedente con il *fanout* dei portali applicativi, per cui il segnale su cui aggregare sono gli oggetti distinti consultati, come pazienti o pratiche, e non il numero di chiamate.

Alla luce di quanto detto precedentemente, le migliori caratteristiche da utilizzare tra le tredici calcolate per ogni profilo, sono lo scostamento dalla propria *routine* e lo scostamento dai colleghi di reparto. Entrambi meritano un dettaglio implementativo, perché è da esso che dipende la loro correttezza.

Il confronto con la propria *routine* – quanti oggetti oggi, rispetto alla mediana dei giorni passati dello stesso utente – è calcolato in modo **causale**: per ogni giornata vengono considerati solo i giorni **precedenti** ad essa, mai i successivi.

Si caratterizza per il fatto di essere una condizione necessaria, non un dettaglio tecnico: se il calcolo usasse anche i giorni futuri di un utente, il sistema misurerebbe uno scostamento che in produzione non potrebbe mai conoscere al momento in cui quella giornata viene valutata. Pertanto, il risultato della validazione non sarebbe replicabile sul campo.

Il confronto con i pari del reparto, dal canto suo, esclude dal calcolo della mediana di reparto gli *account* di collaudo: il loro traffico artificiale, se lasciato dentro, sposterebbe verso l'alto il riferimento di normalità, e renderebbe meno visibile lo scostamento di un abuso reale. Gli *account* che non concorrono a definire il comportamento normale continuano comunque a essere valutati e a comparire nei risultati: pur non entrando nel calcolo del riferimento, restano soggetti allo stesso giudizio di tutti gli altri.

Nel codice, questa condizione si traduce in un accorgimento preciso: la mediana di riferimento viene aggiornata dopo aver calcolato lo scostamento, così ogni giornata è confrontata solo con quelle che la precedono.

```
    # self-baseline causale: ogni giornata è confrontata solo con i giorni precedenti
P = P.sort_values([“username”, “giorno”]).reset_index(drop=True)
self_ratio = np.ones(len(P), float)
for user, gu in P.groupby(“username”, sort=False):
   prec = []                      # oggetti distinti dei giorni già visti
   for idx in gu.index:
      if prec:
         med = np.median(prec)
         self_ratio[idx] = P.at[idx, “oggetti_distinti”] / med if med > 0 else 1.0
      prec.append(P.at[idx, “oggetti_distinti”])   # aggiornata DOPO l’uso
P[“self_ratio_oggetti”] = self_ratio
```

Listato 1: calcolo dello scostamento dalla propria routine.

Nel Listato 1, si nota che la riga finale, che registra il valore del giorno corrente solo dopo averlo usato, garantisce che nessuna giornata sia mai confrontata con il proprio futuro.

L'identificativo dell'*account* escluso non è dedotto automaticamente da qualche euristica, ad esempio, riconoscendo un pattern nel nome utente, ma è indicato esplicitamente nel codice, come nome fisso conosciuto in anticipo.

È una scelta deliberatamente semplice: in un ambiente di collaudo con un numero limitato di *account* di test, individuarli a mano è più affidabile che tentare di riconoscerli automaticamente, ed evita il rischio opposto, cioè un'euristica troppo aggressiva che escluda per errore un *account* reale dalla baseline[109], indebolendo così il confronto tra pari.

In un contesto di produzione, con un numero di *account* di servizio potenzialmente più ampio e variabile nel tempo, questa lista andrebbe mantenuta come parte della configurazione del sistema, aggiornata dall'organizzazione insieme all'anagrafica del personale, piuttosto che lasciata fissa nel codice.

3.5.2 Il modello e la spiegabilità delle segnalazioni

Il modello adottato per il rilevamento è l'***Isolation Forest***.

La scelta non è stata fatta a priori, ma verificata sperimentalmente confrontandolo con altri quattro metodi non supervisionati sugli stessi dati: il funzionamento dell'algoritmo e quello dei termini di paragone sono stati descritti nel capitolo II, mentre l'esito del confronto è riportato al capitolo III.

Le tredici caratteristiche di ogni profilo vengono portate sulla stessa scala prima di essere date in pasto al modello, perché altrimenti una caratteristica che assume valori grandi, visto che il numero di eventi può arrivare a qualche decina, peserebbe più di una che varia tra zero e uno, come la quota di accessi fuori orario. E questo non in quanto sia più importante, ma solo perché i suoi numeri sono più grandi.

Lo strumento impiegato per questa operazione è il RobustScaler descritto nel capitolo II, che riscala ogni caratteristica servendosi della mediana e dello scarto interquartile in luogo della media e della deviazione standard, così da non lasciarsi distorcere proprio dai valori estremi che il sistema deve individuare.

[109] Con *baseline* si intende il riferimento di normalità, ricavato dal comportamento degli *account* legittimi, rispetto al quale si misurano gli scostamenti dei singoli operatori. Consultabile online sul sito del NIST https://www.nist.gov/

Il modello viene addestrato escludendo gli *account* di collaudo, per lo stesso motivo già visto per la *baseline* di reparto: il loro comportamento artificiale non deve entrare nella definizione di normalità.

Il punteggio grezzo prodotto viene infine riportato in un intervallo tra zero e uno, in modo che il profilo meno anomalo dell'insieme riceva punteggio pari a zero, il più anomalo pari a uno, e tutti gli altri si distribuiscano in proporzione tra i due estremi.

Un punteggio, da solo, non giustifica una segnalazione che riguarda una persona.

Per ogni profilo, il sistema calcola quindi uno scarto robusto, cioè uno *z-score*, costruito con la stessa logica di mediana e scarto interquartile, su ciascuna delle tredici caratteristiche, e di esse riporta le tre che hanno avuto lo scostamento più marcato come ***driver*** della segnalazione.

Queste caratteristiche sono le tre ragioni concrete per cui quel profilo è stato giudicato anomalo: non "questo profilo è sospetto", ma "questo profilo è sospetto perché ha toccato un numero di oggetti sensibili molto superiore alla norma, ha operato da più postazioni, e ha effettuato un numero insolito di operazioni di scrittura".

Poiché lo scarto è calcolato allo stesso modo per tutte le caratteristiche, *driver* di natura molto diversa, come un conteggio, un rapporto, o una percentuale, diventano confrontabili sulla stessa scala, e possono essere ordinati insieme.

3.5.3 Dalla stranezza alla gravità: la priorità operativa

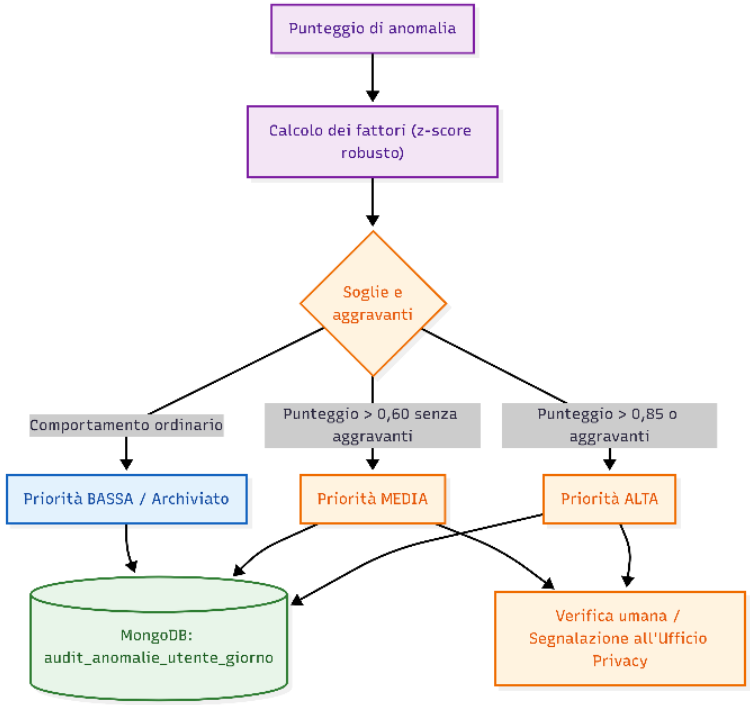


Figura 3.5: Fase 4 - Divisione priorità

Il punteggio del modello misura quanto un profilo è **strano** rispetto alla norma. Ma la stranezza non è ancora la gravità.

Due giornate ugualmente anomale possono avere un peso molto diverso, a seconda che uno scostamento riguardi dati anagrafici o dati clinici sensibili. Per questo, sopra al punteggio del modello, la *pipeline* introduce uno strato di regole che ne combina il valore con quattro **aggravanti**, e cioè: l'accesso a dati sensibili, un numero anomalo di operazioni di scrittura, l'uso di più postazioni nella stessa giornata, e la ricorrenza dello stesso profilo anomalo su più giornate distinte.

Ciascuna aggravante scatta non per la semplice presenza del valore, ma soltanto quando il profilo è realmente fuori norma su quella specifica dimensione, anch'essa misurata con lo scarto robusto. Infatti, un criterio che scattasse per qualunque scrittura o qualunque secondo indirizzo IP non discriminerebbe nulla, perché quasi ogni profilo sopra soglia presenta almeno una scrittura o una seconda postazione nel corso della giornata.

Nell'ambito di queste quattro aggravanti, le scritture meritano una precisazione.

Il fenomeno al centro di questo lavoro, è la **consultazione** impropria, cioè lo *snooping* in senso stretto, che non lascia alcuna modifica nei dati.

Le operazioni di scrittura non appartengono, a rigore, a questa categoria: un profilo che scrive molto più del solito non sta necessariamente consultando cartelle non sue, sta compiendo un numero insolito di modifiche.

Per questo motivo, includerle tra le aggravanti non genera segnalazioni ingiustificate sugli utilizzi legittimi: poiché il punteggio del modello si fonda sullo scostamento dalla propria *routine* e da quella dei pari, un operatore che scrive abitualmente molto, come nel caso di un reparto dedicato alla compilazione di referti, non risulta anomalo, perché quel volume di scritture è la sua norma. La scrittura pesa sul punteggio solo quando è essa stessa fuori dall'ordinario per quella persona e per il suo contesto.

Resta però vero che un profilo segnalato per scritture anomale[110] non rappresenta uno *snooping* nel senso proprio del termine, ma un'altra forma di anomalia di accesso: potenzialmente una manomissione dei dati, un errore operativo ripetuto, o un uso intensivo del sistema al di fuori delle prassi consuete.

Il sistema, quindi, non si limita a rilevare esclusivamente lo *snooping* in senso stretto: individua più in generale le **anomalie di accesso**, di cui lo *snooping* costituisce il caso primario e quello per cui l'intero impianto, dalle firme comportamentali alle *feature*, è stato progettato.

Le altre forme di anomalia che il sistema incidentalmente intercetta, come le scritture fuori norma, sono un beneficio aggiuntivo dell'architettura, non l'obiettivo per cui è stata costruita.

Da questa combinazione nascono tre livelli di priorità.

In primo luogo, un profilo raggiunge la priorità **alta** mediante due distinte modalità: quando il punteggio del modello è già di per sé molto elevato, indipendentemente dalle aggravanti, oppure quando il punteggio supera una soglia più bassa ma è accompagnato da almeno un'aggravante concreta.

[110] Come uno dei casi discussi nel paragrafo di esempio pratico di *output*, dove il punteggio deriva quasi interamente dal numero di modifiche piuttosto che dal volume di consultazioni.

Inoltre, un profilo con punteggio elevato ma privo di qualunque aggravante riceve invece priorità **media**: è statisticamente fuori norma, ma su dimensioni che il sistema non considera di per sé gravi.

Infine, i restanti profili vengono archiviati a priorità **bassa**, pur restando registrati per intero.

La distinzione tra “strano” e “grave” ha un effetto visibile sui dati reali.

Due profili possono avere un punteggio quasi identico e ricevere priorità diversa: da un lato, potrebbe esserci un profilo che resta a priorità media perché il suo scostamento riguarda prevalentemente il volume di oggetti consultati, senza toccare nessuna delle quattro dimensioni aggravanti; dall’altro lato, potrebbe esserci un altro profilo con punteggio simile che, però, sale a priorità alta, perché accompagnato da scritture anomale o da un accesso a dati sensibili. Una soglia calcolata sul solo punteggio avrebbe trattato i due casi allo stesso modo.

Le regole che determinano le aggravanti e le soglie di priorità non sono apprese dai dati: sono scritte esplicitamente nel codice, verificabili e modificabili dall’organizzazione senza dover intervenire sul modello statistico.

È una scelta che si lega direttamente a quanto discusso nel primo capitolo: un sistema che esprime un sospetto su una persona deve poter essere interrogato su ogni sua decisione, e una soglia dichiarata è per definizione più difendibile di una soglia appresa e non ispezionabile.

```
    SOGLIA_ALTA        = 0.85   # oltre questo punteggio si segnala comunque
SOGLIA_ATTENZIONE   = 0.60   # sotto questo punteggio non si segnala
SOGLIA_Z_AGGRAVANTE = 2.0    # scarto oltre cui la caratteristica è fuori norma

AGGRAVANTI = {"n_oggetti_sensibili": "dati sensibili",
            "n_scritture":        "scritture",
            "ip_distinti":        "postazioni multiple"}

def assegna_priorita(P):
   Z = _z_robusti(P, list(AGGRAVANTI.keys()))     # scarto robusto (mediana/IQR)
   sopra = P[P["anomaly_score"] >= SOGLIA_ATTENZIONE]["username"].value_counts()
   ricorrenti = set(sopra[sopra >= 2].index)      # 4° aggravante: più giornate
   for i, r in P.iterrows():
      agg = [et for j, et in enumerate(AGGRAVANTI.values())
            if Z[i, j] >= SOGLIA_Z_AGGRAVANTE]  # scatta solo se fuori norma
      if r["username"] in ricorrenti:
         agg.append("già segnalato in altra giornata")
      if   r["anomaly_score"] >= SOGLIA_ALTA:              p = "ALTA"
      elif r["anomaly_score"] >= SOGLIA_ATTENZIONE and agg: p = "ALTA"
      elif r["anomaly_score"] >= SOGLIA_ATTENZIONE:        p = "MEDIA"
      else:                                                p = "BASSA"
      P.at[i, "priorita"] = p
```

Listato 2: lo strato di priorità operativa.

Le quattro condizioni **aggravanti** (dati sensibili, scritture, postazioni multiple, ricorrenza) scattano solo quando il profilo è statisticamente fuori norma sulla dimensione corrispondente ($z \geq 2$); le tre fasce di priorità derivano da soglie dichiarate, ispezionabili e modificabili senza toccare il modello.

3.6 L'esecuzione della pipeline e i suoi risultati

I sette stadi descritti nelle sezioni precedenti sono *script* indipendenti, pensati per essere eseguiti anche singolarmente durante lo sviluppo e la verifica di ciascuna fase.

Per l'uso corrente, tuttavia, la *pipeline* viene eseguita come un'unica sequenza, orchestrata da uno script dedicato che lancia ciascuno stadio nell'ordine corretto, verifica l'esito di ognuno prima di procedere al successivo, e si ferma con un messaggio esplicito se uno degli stadi termina con un errore, invece di proseguire silenziosamente su dati incompleti.

L'ordine di esecuzione non è arbitrario. La revisione manuale del catalogo, in particolare, è collocata **dopo** la classificazione automatica e **prima** della fase di rilevamento: solo in questa posizione le decisioni verificate a mano possono essere

riapplicate al catalogo appena rigenerato, senza che una nuova esecuzione della classificazione automatica le cancelli. È lo stesso principio di non distruttività discusso in apertura del capitolo, reso concreto nell'ordine con cui gli stadi vengono concatenati.

Un'unica eccezione riguarda il primo stadio, quello di acquisizione: nella sua modalità pensata per la produzione, resta in ascolto indefinitamente di nuovi *file* in arrivo, e non termina mai da solo. Perché l'orchestratore possa proseguire oltre questo stadio, tale modalità va disattivata, lasciando attiva solo l'importazione dei *file* già presenti nella cartella al momento dell'avvio; in alternativa, l'acquisizione può essere eseguita a parte, come processo separato e persistente, mentre il resto della *pipeline* gira periodicamente sui dati già raccolti.

Descritti i singoli stadi e la loro orchestrazione, resta da mostrare che cosa la catena produce quando viene eseguita per intero.

Questo paragrafo riporta l'esito di un'esecuzione completa sull'ambiente dimostrativo, dai *log* grezzi fino alla graduatoria dei profili anomali.

Le metriche che ne quantificano l'affidabilità, ottenute per iniezione di anomalie sintetiche, saranno discusse successivamente. Qui ci si focalizzerà sul prodotto operativo della *pipeline*.

Lanciata dall'orchestratore, la catena attraversa in sequenza i propri stadi[111] e si conclude in poco più di due minuti.

La Tabella 3.2 ne riassume l'andamento: dai *log* grezzi vengono ricavati 17.341 *AuditEvent*, ricondotti a 336 operazioni cliniche distinte e infine aggregati in 206 profili utente-giornata.

Oltre l'ottanta per cento del tempo di calcolo è assorbito dall'ultimo stadio, quello di rilevamento, mentre la costruzione del catalogo e la sua revisione sono pressoché istantanee.

[111] Come anticipato, il prototipo di rilevamento per-evento resta fuori dall'esecuzione ordinaria.

Fase	Esito	Durata
Importazione dei log grezzi	log importati in MongoDB	5,9 s
Normalizzazione in AuditEvent	17.341 AuditEvent	10,9 s
Estrazione del catalogo	336 operazioni distinte	1,9 s
Classificazione automatica	336 operazioni pre-classificate	0,1 s
Revisione manuale	23 correzioni puntuali + 11 regole di area	0,1 s
Rilevamento (Isolation Forest)	206 profili, ordinati per punteggio	107,9 s
Totale		**126,8 s**

Tabella 3.2 – Esecuzione completa della pipeline sull'ambiente dimostrativo.

Dei 336 nomi di operazione presenti nei dati, la classificazione automatica ne ha riconosciuti la maggior parte per convenzione di *naming*.

La revisione manuale è intervenuta con ventitré correzioni puntuali su singole operazioni e undici regole di area, tra cui l'attribuzione di sensibilità alta alle aree «Malattie infettive» e «Fine vita / decesso» e di sensibilità media all'area «Diagnosi».

Le operazioni più frequenti sono, coerentemente con la natura clinica dell'applicativo, *query* di consultazione: i dati generali dell'assistito, il diario clinico, le anamnesi.

Il prodotto finale è la graduatoria dei profili ordinati per punteggio di anomalia.

La seguente Tabella ne riporta i primi dieci, ciascuno con il punteggio, la priorità operativa e i tre driver che ne motivano la posizione.

Utente	Giorno	Ogg. distinti	Score	Priorità	Driver principali (scarto z)
account_collaudo	01/07/2026	71	1,000	ALTA	oggetti mai visti (z=20,6); oggetti distinti (z=18,1); scost. dai pari (z=17,8)

Utente	Giorno	Ogg. distinti	Score	Priorità	Driver principali (scarto z)
operatore2	09/07/2026	26	0,945	ALTA	scost. dalla routine (z=8,3); postazioni distinte (z=6,7); eventi (z=6,6)
operatore1	21/07/2026	5	0,931	ALTA	scritture (z=14,3); fan-out (z=6,9); postazioni distinte (z=6,7)
operatore4	16/07/2026	31	0,903	ALTA	scost. dalla routine (z=15,7); oggetti mai visti (z=7,5); oggetti distinti (z=7,3)
operatore3	02/07/2026	38	0,884	ALTA	scritture (z=13,7); oggetti mai visti (z=10,7); oggetti distinti (z=9,2)
operatore2	02/07/2026	17	0,847	ALTA	oggetti ad alta sensibilità (z=9,3); scost. dalla routine (z=5,1); eventi (z=4,6)
operatore4	21/07/2026	22	0,766	ALTA	scost. dalla routine (z=6,9); postazioni distinte (z=6,7); eventi (z=5,9)
operatore18	11/08/2026	6	0,665	ALTA	scritture (z=11,3); fan-out (z=6,2); oggetti ad alta sensibilità (z=4,7)

Utente	Giorno	Ogg. distinti	Score	Priorità	Driver principali (scarto z)
operatore5	24/07/2026	6	0,627	ALTA	postazioni distinte (z=8,1); fan-out (z=4,2); eventi (z=3,1)
operatore6	01/07/2026	6	0,621	MEDIA	attività fuori orario (z=6,6); operazioni distinte (z=4,2); fan-out (z=3,3)

Tabella 3.3 – I dieci profili utente-giornata più anomali.

Sull'intero insieme, il sistema porta all'attenzione dodici profili su 206, pari a circa il 6%, di cui undici a priorità alta e uno a priorità media. I restanti 194 restano a priorità bassa, registrati ma non segnalati.

Questa netta separazione è resa visibile dalla Figura 3.6, che proietta i 206 profili su due dimensioni: la massa dei profili normali si addensa al centro, mentre i pochi casi segnalati si collocano isolati ai margini.

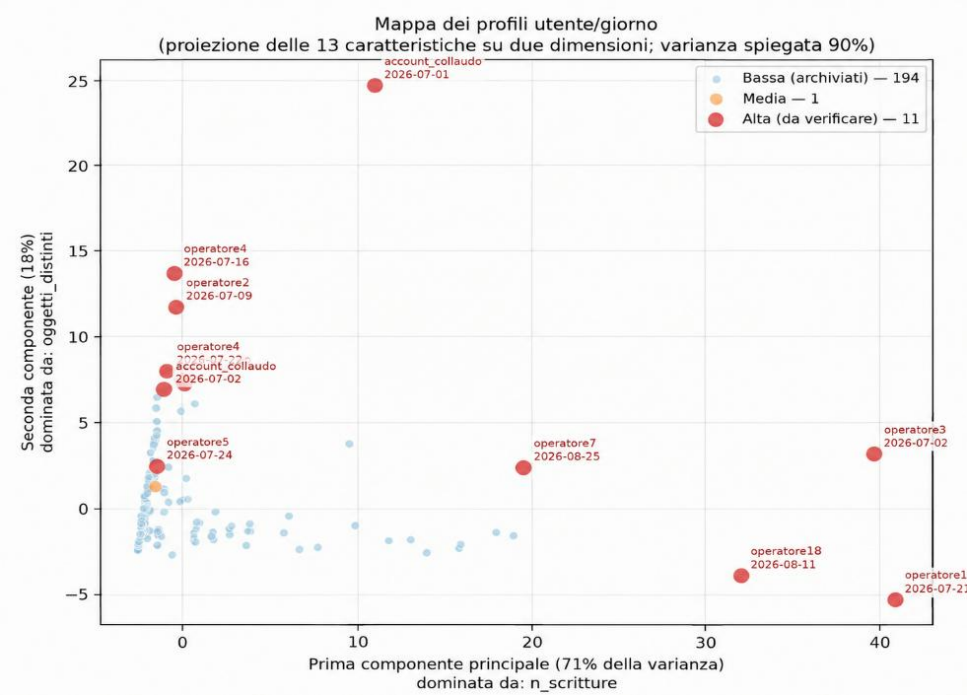


Figura 3.6: Mappa dei profili con le diverse priorità

La concentrazione dell'attenzione su una porzione così ristretta è esattamente l'obiettivo enunciato nel Capitolo I: non un allarme di massa, ma un insieme minimo di casi da sottoporre a verifica umana.

La lettura dei *driver* rivela due profili di anomalia distinti.

Il primo è la consultazione anomala vera e propria, lo *snooping* in senso stretto, riconoscibile nei soggetti in cui tra i *driver* si rinvengono gli oggetti distinti, gli oggetti mai visti prima e lo scostamento dalla propria *routine* o dai pari.

È il caso di operatore4 del 16 luglio, il cui accesso a trentuno oggetti si discosta dalla propria media di un fattore che il modello traduce in uno scarto z pari a 15,7.

Il secondo profilo riguarda l'anomalia di scrittura, in cui il punteggio nasce quasi per intero da un numero insolito di modifiche a fronte di pochissimi oggetti toccati.

Ne sono esempi chiari sia operatore1 (cinque soli oggetti distinti, ma scritture con z pari a 14,3), che operatore3.

Come anticipato nei paragrafi precedenti, questi ultimi non costituiscono *snooping* in senso proprio, ma quella diversa forma di anomalia di accesso che l'architettura intercetta come beneficio aggiuntivo.

La distinzione tra profilo «strano» e profilo «grave» è visibile confrontando operatore5 e operatore6, dal punteggio quasi identico (0,627 e 0,621): il primo sale a priorità alta per la presenza di un'aggravante concreta, le postazioni multiple, mentre il secondo, anomalo soltanto per l'orario e l'ampiezza operativa, resta a priorità media. Una soglia calcolata sul solo punteggio avrebbe trattato i due casi allo stesso modo.

Al vertice della graduatoria compare il profilo account_collaudo, con il punteggio massimo (1,000) e la ricorrenza su due giornate distinte.

Si tratta dell'*account* nominativo assegnato all'autore durante il periodo di tirocinio e utilizzato per prendere confidenza con l'applicativo: in assenza di una *routine* operativa definita, la navigazione ha toccato un gran numero di sezioni e di fascicoli diversi, in modo esplorativo e non sistematico. È proprio questo a produrne il profilo: settantuno oggetti distinti in una sola giornata, quasi tutti mai consultati prima e privi di una regolarità riconducibile a un compito ricorrente.

Coerentemente con quanto illustrato nel paragrafo precedente e proprio perché un simile comportamento è manifestamente non rappresentativo di un utilizzo ordinario, il profilo è escluso dalla costruzione del riferimento di normalità. Difatti, esso non concorre

né al calcolo della mediana del reparto, né all'addestramento del modello, così da non distorcere la nozione di «normale» su cui poggia il rilevamento; continua però a essere valutato e ordinato come tutti gli altri.

Per questo, il sistema lo colloca correttamente in cima alla classifica, e l'analisi di stabilità condotta in sede di validazione[112] lo conferma come anomalia reale nella quasi totalità delle prove: non perché vi sia alcunché di illecito, ma perché quel comportamento è, sul piano statistico, il più distante dalla norma dell'ambiente.

Ne emerge, quasi in forma didascalica, la distinzione tenuta ferma per tutto il lavoro: un accesso può essere massimamente anomalo e al tempo stesso del tutto legittimo, e spetta alla verifica umana e non al punteggio, chiarirne la ragione.

3.7 Prototipo di rilevamento in tempo reale

Il sistema descritto finora è **retrospettivo**: analizza giornate concluse, ed è quindi adatto a un *audit* svolto a posteriori, ma non a intercettare un abuso mentre è ancora in corso. È un limite intrinseco alla scelta di aggregare per giornata, e non elimina il valore del sistema per l'*audit*. Al contrario, lascia aperta una domanda naturale: cosa servirebbe per spostare il rilevamento più vicino al momento dell'accesso?

A questo scopo è stato sviluppato un percorso parallelo, tenuto volutamente separato dalla *pipeline* principale.

In sostanza, anziché aggregare gli accessi per giornata conclusa, calcola per ogni singolo evento un insieme di caratteristiche a **finestra scorrevole**: quanti accessi lo stesso utente ha compiuto nell'ultima ora e nelle ultime ventiquattro ore, calcolate al momento in cui l'evento arriva, guardando solo indietro nel tempo.

Si tratta dello stesso principio di causalità già trattato con riferimento alla *self-baseline* giornaliera, applicato però a una scala temporale molto più fine, compatibile con una valutazione istantanea piuttosto che a fine giornata.

Questo percorso non è incluso nell'esecuzione ordinaria della *pipeline*: resta uno *script* a sé, pensato per essere eseguito separatamente.

[112] Vedi cap. III, par. 3.8.

La scelta è motivata dal fatto che il sistema a giornate resta il nucleo del rilevamento e della sua validazione, mentre l'approccio a finestra scorrevole rappresenta la base concettuale – non ancora integrata nel flusso principale – per un'eventuale evoluzione verso il rilevamento in tempo reale.

Sia il profilo giornaliero, sia tale prototipo, trattano, allo stato attuale, la fascia oraria come un semplice indicatore di attività fuori orario.

Un raffinamento naturale, che è stato sperimentato separatamente sulle caratteristiche per-evento, ma non ancora integrato in questo percorso, consiste nel codificare l'ora in forma ciclica, attraverso una coppia di funzioni seno e coseno, in modo che le 23:59 e le 00:01 risultino vicine sulla scala usata dal modello, e non agli estremi opposti di un intervallo lineare: una precisione che ha senso quando si valuta l'istante preciso di un evento, meno quando si aggregano interi giorni.

3.8 La validazione

I paragrafi precedenti hanno descritto che cosa la pipeline produce. Resta da stabilire quanto ci si possa fidare di ciò che produce, e cioè se la graduatoria dei profili anomali colga davvero i comportamenti che dovrebbe cogliere. È l'oggetto di questo paragrafo, che ne espone il problema di fondo, il metodo adottato per aggirarlo, il confronto sperimentale tra gli algoritmi e la verifica di stabilità delle segnalazioni.

3.8.1 Il problema della valutazione e il metodo adottato

Affinché sia possibile dire che un sistema di rilevamento funzioni bene servirebbe sapere, caso per caso, quale sia la risposta giusta, e, quindi, quali accessi costituissero davvero degli abusi e quali no.

Questa risposta, per il fenomeno qui trattato, semplicemente non esiste. Nessuna delle giornate analizzate è mai stata dichiarata, da un'autorità competente, come *snooping* accertato o come consultazione legittima: non c'è un elenco certificato di violazioni realmente avvenute nell'ambiente su cui il sistema è stato sviluppato.

Una tale assenza non costituisce un difetto della sperimentazione, ma si pone ad un livello strutturale rispetto al problema: se un elenco di violazioni certificate esistesse già, non servirebbe costruire un sistema per trovarle.

Si tratta della stessa osservazione, applicata alla validazione, che nel primo capitolo motiva l'intero lavoro: un accesso anomalo è misurabile, un accesso illecito richiede un accertamento che il sistema informativo, da solo, non può fornire.

Senza etichette, *precision* e *recall* non sono calcolabili direttamente sui dati di produzione: per calcolarli servirebbe sapere, per ogni segnalazione, se sia un vero o un falso positivo che rappresenta esattamente l'informazione mancante.

Il problema non si aggira etichettando a mano un campione, perché stabilire se un accesso fosse legittimo richiederebbe di conoscere la relazione di cura tra operatore e paziente, informazione che, come illustrato quando si è parlato dei limiti del lavoro, non è disponibile in questo contesto.

Di conseguenza, è stata seguita un'altra strada: costruire artificialmente dei casi di cui si conosce la risposta, e misurare se il sistema li riconosce.

Il principio è semplice da enunciare: si costruiscono profili con le caratteristiche note dello ***snooping***, li si mescola ai profili reali, e si misura quanti di essi il modello riesce a ripescare nelle posizioni più alte della classifica. Il dettaglio che rende questo principio applicabile con rigore riguarda come costruire un'anomalia sintetica credibile.

Ogni profilo iniettato parte dalla **mediana** dei profili reali su tutte le caratteristiche, quindi non da valori casuali, così da non ottenere un caso talmente estremo su ogni dimensione da risultare banale da isolare. Da questa base, solo le caratteristiche coerenti con le firme comportamentali[113] vengono spinte verso l'alto: il numero di oggetti distinti toccati, gli oggetti mai visti prima, lo scostamento dalla propria routine e dai pari, la quota di dati sensibili e di accessi fuori orario, oltre al numero assoluto di oggetti sensibili toccati. A quel punto, il numero di eventi e il *fanout* seguiranno di conseguenza, essendo entrambi derivati dal numero di oggetti distinti.

[113] Vd. cap. I.

Il profilo che ne risulta non è un'anomalia generica: è la traduzione numerica dello *snooper* descritto nella casistica, cioè chi accede a molte più persone del solito, in aree cliniche delicate, discostandosi sia da sé stesso, che dal proprio reparto.

Per ogni ciclo di valutazione vengono iniettati cinque profili sintetici, un numero volutamente contenuto, proprio perché lo *snooping* è un evento raro, e una contaminazione troppo alta renderebbe il modello incapace di apprendere correttamente cosa sia la normalità, falsando la stessa validazione che dovrebbe misurare.

Per questo motivo, l'intero esperimento viene ripetuto trenta volte, ogni volta con un *random seed*[114] diverso e, così facendo, si ottiene una stima stabile, caratterizzata da media e deviazione *standard*, invece di un singolo valore che dipenderebbe dalla particolare estrazione casuale capitata quella volta.

Il numero di ripetizioni non è arbitrario: trenta estrazioni indipendenti sono sufficienti a stabilizzare la media della metrica principale, mentre aumentarne il numero produrrebbe variazioni trascurabili a fronte di un costo di calcolo sensibilmente maggiore.

3.8.2 Il confronto tra i cinque algoritmi

La decisione di adottare l'***Isolation Forest***[115] non è stata presa facendo riferimento a un singolo lavoro di letteratura. Al contrario, è stata verificata sperimentalmente sui dati di questo progetto, confrontandolo con altri quattro metodi non supervisionati comunemente impiegati per l'anomaly detection e descritti nel capitolo II.

Il primo confronto, condotto iniettando anomalie nette, costruite senza modulazione dell'intensità, restituisce i seguenti risultati:

Algoritmo	ROC-AUC	Average Precision
Robust Covariance	1.000 ± 0.000	1.000 ± 0.000

[114] Con *random seed* si intende il valore che inizializza il generatore di numeri pseudo-casuali su cui si fondano l'estrazione dei profili sintetici e le componenti aleatorie del modello: fissarlo rende l'esperimento esattamente riproducibile, mentre farlo variare permette di osservare la variabilità dei risultati tra un'esecuzione e l'altra. Richiamabile con la funzione di Python *random.seed.*

[115] F. T. LIU, K. M. TING, Z.-H. ZHOU, *Isolation Forest, cit.; sul suo funzionamento,* v. cap. II, par. 2.6.

kNN-distance	0.997 ± 0.004	0.897 ± 0.142
Isolation Forest	0.997 ± 0.002	0.866 ± 0.068
One-Class SVM	0.988 ± 0.004	0.555 ± 0.085
Local Outlier Factor	0.400 ± 0.239	0.053 ± 0.043

Tabella 3.4: Confronto dei cinque algoritmi su anomalie nette

I primi quattro metodi, tutti sopra 0,98 di ROC-AUC, si equivalgono sostanzialmente: su anomalie di questa entità, quasi ogni approccio le riconosce, e il confronto non è discriminante.

Inoltre, è da notare l'ampiezza della deviazione *standard* del *kNN-distance* sull'*Average Precision*: più del doppio di quella dell'*Isolation Forest*, indice di un comportamento meno stabile al variare delle particolari anomalie iniettate a ogni ripetizione.

Per distinguere davvero i cinque metodi serve uno scenario più vicino a quello reale: uno *snooper* accorto non produce un'anomalia eclatante, ma uno scostamento appena percettibile.

In considerazione di ciò, è stato condotto un secondo esperimento, che modula l'intensità dell'anomalia iniettata, anziché costruirla a valori fissi: ogni caratteristica spinta verso l'alto viene collocata a una distanza dalla mediana pari non a un valore assoluto scelto entro un intervallo, ma a un multiplo dello scarto interquartile robusto dei profili reali, cioè la stessa unità di variabilità normale già impiegata altrove nel sistema.

Si tratta di una costruzione dell'anomalia diversa da quella del 5.2.1.8.1.

Infatti, il primo esperimento verifica la capacità del modello scelto di riconoscere uno *snooping* tipico, secondo la stessa costruzione usata nella validazione del sistema nel suo complesso. Questo secondo esperimento, invece, serve a parametrizzare la difficoltà; cosa che una costruzione a valori assoluti non permette di fare in modo continuo.

I risultati sono riportati in termini di ROC-AUC medio su venti ripetizioni per livello di intensità.

Si tratta di un numero ridotto rispetto alle trenta della validazione principale, vista la necessita di contenere il costo del calcolo ripetuto su sei livelli di intensità e cinque algoritmi, dal più sottile (0,5 volte lo scarto normale) al più eclatante (5 volte):

Algoritmo	**x0.5**	**x1.0**	**x1.5**	**x2.0**	**x3.0**	**x5.0**
Isolation Forest	0.466	0.615	0.765	0.843	0.921	0.968
Robust Covariance	0.526	0.526	0.773	0.847	0.905	0.968
Local Outlier Factor	0.185	0.177	0.126	0.177	0.192	0.207
kNN-distance	0.075	0.142	0.230	0.328	0.498	0.696
One-Class SVM	0.030	0.015	0.021	0.075	0.610	0.787

Tabella 3.5: Confronto dei cinque algoritmi su anomalie variabili

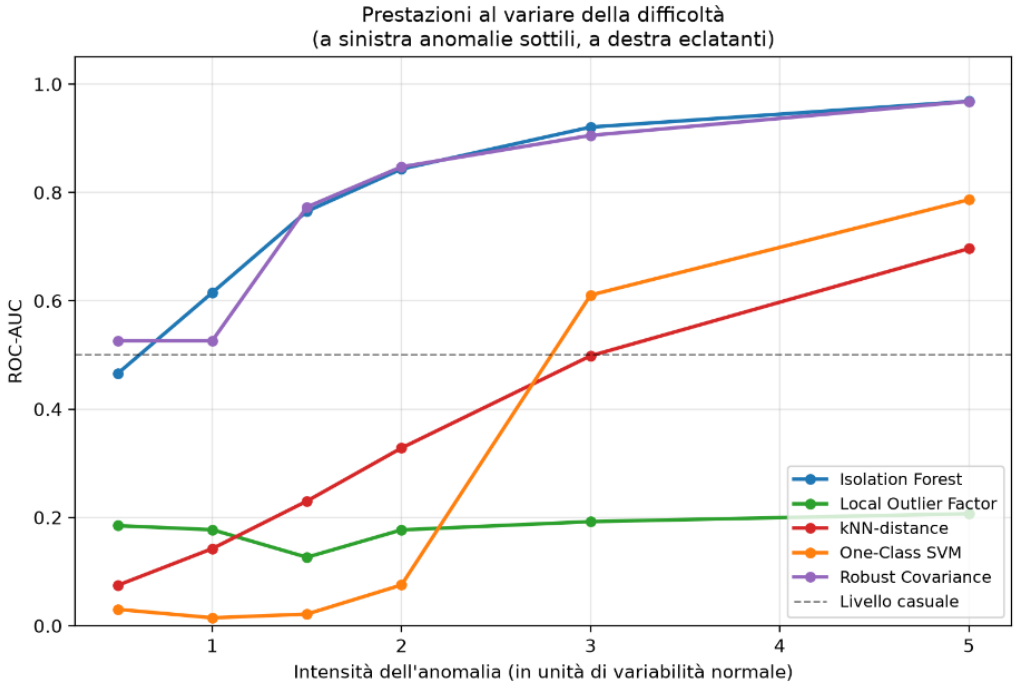


Figura 3.7: Prestazione dei cinque algoritmi al variare delle anomalie

Osservando la Figura 3.7 e la Tabella 3.5 si nota come il quadro si ribalti rispetto al primo confronto.

Alla minima intensità, l'anomalia iniettata è così tenue che nessun metodo riesce a separarla in modo affidabile dai profili normali: i due migliori, *Isolation Forest* e *Robust Covariance*, si collocano appena a ridosso del livello del caso (0,5), mentre gli altri restano perfino al di sotto.

Il vantaggio dell'*Isolation Forest* diventa netto e stabile a partire da uno scostamento pari allo scarto normale (intensità 1,0), dove distacca chiaramente ogni altro metodo e da lì cresce in modo regolare fino alle anomalie più evidenti.

La *Robust Covariance*, che nel test precedente otteneva il punteggio più alto in assoluto, resta invece inchiodata al suo valore iniziale alle due intensità più basse e comincia a salire solo a partire da 1,5, un segnale che l'assunzione di normalità gaussiana su cui si fonda è violata dai dati in esame, coerentemente con le segnalazioni di matrice non a rango pieno che il metodo emette durante l'addestramento.

L'*Isolation Forest,* insieme a *kNN-distance* e *Robust Covariance*, degrada in modo regolare al calare dell'intensità; il *Local Outlier Factor*, al contrario, ha un andamento irregolare che in alcuni tratti risale. Questa regolarità, in un sistema destinato all'uso operativo, conta quanto la prestazione media.

L'*Isolation Forest* è stato scelto per questa combinazione di ragioni: la superiorità dimostrata nel regime in cui uno *snooper* realistico effettivamente si colloca, l'assenza di assunzioni sulla distribuzione dei dati, la robustezza rispetto alla scala delle variabili, l'efficienza computazionale necessaria per un eventuale funzionamento in tempo reale, e l'interpretabilità delle segnalazioni prodotte.

3.8.3 Stabilità delle segnalazioni ed esclusione delle anomalie reali

La validazione descritta finora misura se il sistema riconosce le anomalie sintetiche iniettate. Resta però una domanda distinta, cioè quella in cui ci si chiede se le segnalazioni prodotte sui profili reali sono stabili oppure dipendono dalla particolare estrazione casuale di ciascuna esecuzione.

Un profilo che risulta anomalo una volta su trenta e uno che lo risulta ventotto volte su trenta non hanno lo stesso peso, e distinguerli è essenziale prima di portarli all'attenzione di un revisore.

Per rispondere, l'intera procedura è stata ripetuta trenta volte sui soli profili reali, registrando per ciascuno quante volte è finita tra i casi segnalati.

Un solo profilo, su 206, supera la soglia di stabilità fissata al 90% delle ripetizioni: quello dell'*account* nominativo assegnato all'autore durante il tirocinio[116], segnalato in ventotto prove su trenta. Non si tratta di un artefatto dell'iniezione, bensì di un'anomalia reale di natura del tutto legittima: la navigazione esplorativa di chi prende confidenza con l'applicativo, priva di una *routine* ricorrente.

La presenza di un'anomalia reale così marcata tra i profili di riferimento disturba la stessa validazione: un caso genuinamente fuori norma, lasciato tra i profili normali, si comporta come un'anomalia non etichettata e penalizza le metriche, in particolare l'*Average Precision*, la più sensibile alla presenza di casi anomali nelle prime posizioni della classifica.

Per questo, il profilo è stato messo da parte e le metriche ricalcolate senza di esso.

È bene sottolineare, però, che l'esclusione non costituisce un'etichettatura manuale: il profilo non è stato scelto perché ritenuto a priori anomalo, ma perché è il modello stesso a segnalarlo in modo stabile, secondo il criterio dichiarato nel 90% delle ripetizioni.

L'effetto è coerente con questa lettura: escludendo l'unico profilo reale stabilmente anomalo, il *ROC-AUC* medio passa da 0,997 a 0,999 e l'*Average Precision* da 0,866 a 0,967, con una riduzione della relativa deviazione *standard*. Quanto appena detto è riscontrabile nella Figura 3.8.

Il miglioramento non indica un modello più bravo, ma una misura più pulita, depurata da un caso reale che, confuso tra i profili di riferimento, ne abbassava artificiosamente il valore.

[116] Vedi cap. III, par. 3.6.

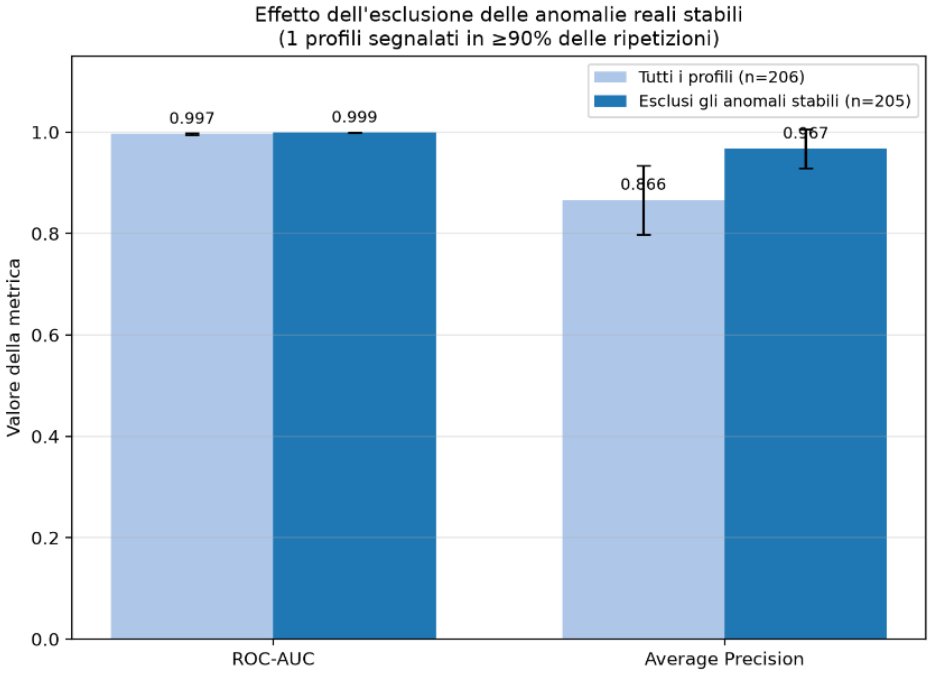


Figura 3.8: Effetto dell'esclusione delle anomalie reali stabili

3.9 I limiti del sistema

Questo paragrafo raccoglie i limiti introdotti precedentemente e li esamina, distinguendo quelli intrinsechi all'approccio, che accompagnerebbero il sistema anche in una versione matura, da quelli legati al contesto specifico in cui il lavoro è stato svolto, e che una messa in esercizio reale supererebbe.

Nel capitolo che segue, gli stessi limiti verranno riletti in positivo, come mappa delle direzioni in cui il sistema può crescere. Alcune di queste direzioni, legate al contesto dimostrativo o all'assenza di un dato, indicano miglioramenti che una messa in esercizio reale renderebbe possibili; altre suggeriscono estensioni del metodo che ne amplierebbero la portata. Ciascuno sviluppo viene ricondotto, dove possibile, al limite specifico che è destinato a rimuovere, distinguendo le direzioni che porterebbero il maggior beneficio con il minor sforzo da quelle di più ampio respiro.

Il primo limite riguarda la base empirica su cui il sistema poggia, ed è forse il più importante da comprendere, perché tocca il cuore stesso del metodo. Il modello non

giudica un accesso in termini assoluti — non esiste, nel sistema, una soglia fissa che dica "più di N accessi sono troppi" — ma sempre in termini relativi, rispetto a una baseline: il comportamento abituale di quello specifico operatore e quello dei suoi pari nello stesso reparto. È esattamente questa relatività a rendere il sistema capace di cogliere lo scostamento là dove una regola fissa fallirebbe, ma è anche ciò che lo rende dipendente dalla qualità e dall'estensione dello storico disponibile.

Una baseline costruita su pochi giorni di osservazione è inaffidabile per una ragione statistica elementare: con pochi dati, la mediana dei giorni passati di un utente e quella del suo reparto sono stime instabili, facilmente distorte da una singola giornata atipica. Un operatore che, nella prima settimana di raccolta, avesse avuto per caso due giornate particolarmente intense vedrebbe la propria "normalità" già sovrastimata, e uno snooping successivo apparirebbe meno anomalo di quanto sia. Solo dopo alcune settimane di attività, la baseline si stabilizza abbastanza da rendere il confronto significativo. Ne discende un requisito operativo che va dichiarato con chiarezza a chi volesse adottare il sistema: esso non è utilizzabile "dal primo giorno", ma richiede un periodo iniziale di apprendimento, dell'ordine di alcune settimane, durante il quale osserva e costruisce le baseline senza ancora emettere segnalazioni su cui si possa fare affidamento. Questo non è un difetto correggibile con una modifica del codice: è una proprietà inevitabile di qualunque sistema che misuri lo scostamento da una norma appresa, e va gestito come un requisito di messa in esercizio, non nascosto.

A questa considerazione sulla quantità dei dati se ne affianca una sulla loro provenienza. Come dichiarato fin dall'introduzione, la *pipeline* è stata sviluppata e collaudata su un ambiente dimostrativo aziendale, popolato da anagrafiche fittizie e utilizzato internamente per attività di collaudo e formazione.

La scelta ha un pregio tutt'altro che secondario in un lavoro che si occupa di protezione dei dati sanitari: nessun dato clinico reale è stato trattato nel corso della tesi. Ma comporta anche una conseguenza da non sottovalutare. Il comportamento che il modello ha appreso come "normale" è quello di un ambiente di prova, generato da *account* di collaudo che riproducono l'attività clinica in modo verosimile ma non identico a quello di un reparto reale, dove convivono turni notturni, sostituzioni, picchi stagionali, urgenze e prassi locali che nessun ambiente demo riproduce fedelmente.

Di conseguenza, i profili che il sistema segnala sui dati attuali hanno valore esemplificativo rispetto al funzionamento del metodo, mostrando sia come il sistema distingua i comportamenti anomali, sia le ragioni che vi pone alla base. Nonostante questo, essi non costituiscono, né potrebbero costituire, evidenza di accessi realmente impropri. Queste conferma la capacità della catena di isolare profili anomali, non l'esistenza di abusi in quell'ambiente.

A questo limite se ne affianca un altro, insito nel metodo stesso con cui il sistema è stato validato.

Per misurare quanto bene un sistema di rilevamento funzioni, servirebbe sapere, per un insieme di casi reali, quali accessi fossero davvero impropri e quali no, che corrisponde all'informazione che permette di calcolare quante anomalie il sistema coglie e quanti falsi allarmi produce. Purtroppo, nell'ambiente disponibile, questa informazione non esiste: non c'è un elenco di violazioni certificate contro cui misurarsi. La capacità del sistema è stata perciò valutata iniettando anomalie sintetiche – profili artificiali costruiti con le caratteristiche note dello *snooping*, mescolati ai dati reali – e verificando quanti di essi il modello riuscisse a collocare in cima alla propria classifica.

Questo approccio è l'unico praticabile in assenza di etichette reali, ma va riconosciuto per ciò che effettivamente dimostra, e per ciò che non dimostra. Esso prova che il modello di rilevamento, cioè l'*Isolation Forest* applicato alle feature comportamentali, è in grado di separare un profilo anomalo da uno normale con elevata affidabilità.

Non prova, invece, che l'intera catena *end-to-end*, dal *log* grezzo alla segnalazione finale, si comporti correttamente di fronte a un caso di abuso reale: le anomalie iniettate entrano nel sistema già in forma di profilo strutturato, saltando di fatto gli stadi a monte – l'acquisizione, la normalizzazione FHIR, la classificazione del catalogo – che su un caso reale potrebbero introdurre imprecisioni proprie.

Vi è inoltre una circolarità che va dichiarata con onestà. Le anomalie sintetiche sono costruite a partire dalle stesse firme comportamentali che il modello è progettato per riconoscere: in un certo senso, si verifica che il sistema trovi ciò che è stato costruito per trovare. Questo non rende la validazione priva di valore — la difficoltà del compito è modulata rendendo le anomalie progressivamente più sottili, fino al limite in cui diventano quasi indistinguibili dal comportamento normale, ed è proprio in quel regime

che il metodo dimostra la propria tenuta — ma impone di non presentare i risultati come una prova di efficacia sul campo. Una validazione piena richiederebbe ciò che, per definizione, il problema non offre: un insieme di casi reali etichettati da un'autorità competente. Il limite, in questo senso, non è una debolezza rimediabile della sperimentazione, ma una conseguenza strutturale della natura non supervisionata del problema affrontato — la stessa che, come discusso nel Capitolo I, ha reso necessario l'intero approccio.

Il limite che incide più direttamente sulla precisione del sistema — e la cui rimozione porterebbe il miglioramento più significativo — riguarda l'assenza della relazione di cura. Il sistema, allo stato attuale, non sa quali pazienti un determinato operatore avrebbe titolo di consultare. Sa riconoscere che un accesso è insolito — che tocca oggetti mai visti prima, che si discosta dalla routine dell'utente e da quella dei suoi pari — ma non può sapere se quel paziente rientri o meno tra quelli effettivamente in carico a chi lo sta consultando. Distingue, in altre parole, l'accesso anomalo dall'accesso ordinario, ma non l'accesso legittimo da quello illecito: la distanza tra queste due coppie di nozioni, come discusso nel Capitolo I, è esattamente ciò che separa una misura statistica da un giudizio giuridico.

La conseguenza pratica è duplice. Da un lato, il sistema può segnalare come anomali accessi che erano in realtà pienamente legittimi — una sostituzione in un reparto diverso dal proprio, il subentro in un turno, la presa in carico di un nuovo paziente — perché privo dell'informazione che li giustificherebbe. Dall'altro, e all'opposto, non può cogliere la gravità piena di un accesso a un paziente che l'operatore non aveva alcun titolo di trattare, se quel comportamento non è anche numericamente anomalo. È un limite che il sistema condivide con qualunque approccio puramente comportamentale, e che non nasce da una scelta implementativa ma dall'assenza, a monte, di un dato: la relazione formale tra operatore e paziente, custodita nei sistemi di gestione dei reparti ma non disponibile in questo lavoro. Proprio perché è un'assenza di dato e non un difetto di metodo, è anche il limite la cui rimozione trasformerebbe il segnale del sistema da indiretto a diretto.

A questo si aggiunge un limite di natura strutturale, che riguarda l'oggetto stesso di ciò che il sistema osserva. Ogni segnalazione è riferita a una credenziale — un nome utente che ha compiuto una certa attività — e non a una persona fisica. Nella maggior

parte dei casi le due cose coincidono, ma non sempre: diversi provvedimenti del Garante riguardano accessi effettuati con credenziali altrui, sottratte da un post-it o da una postazione lasciata incustodita, oppure con *account* generici di reparto condivisi da più operatori. In tutte queste situazioni, il *log* registra un utente formalmente legittimo, e nessuna analisi comportamentale, per quanto raffinata, può stabilire chi fosse realmente seduto davanti alla tastiera nel momento dell'accesso.

Questo limite ha una conseguenza che va oltre la semplice imprecisione, e tocca il valore probatorio delle segnalazioni. Poiché il sistema non può garantire l'identità della persona dietro una credenziale, una sua segnalazione non potrà mai, di per sé, costituire una prova di responsabilità individuale: potrà indicare che da un certo *account* è stato compiuto un accesso anomalo, ma l'attribuzione di quell'accesso a una persona determinata richiede accertamenti che esulano dai dati comportamentali — testimonianze, verifiche sulle postazioni, controlli sugli orari. È una ragione ulteriore, che si aggiunge a quelle già viste, per cui l'esito del sistema va inteso come un elemento che orienta un'indagine umana, non come una conclusione. Il sistema restringe il campo di ciò che merita attenzione; stabilire chi e con quale intento resta compito di chi quella segnalazione la riceve.

Un limite di precisione locale riguarda il catalogo semantico delle operazioni. L'attribuzione a ciascuna operazione di un'azione, di un'area clinica e di un livello di sensibilità si fonda sul nome dell'operazione e sulle convenzioni con cui è stata denominata, non su una verifica di ciò che essa effettivamente compie. È una classificazione euristica: efficace nella maggior parte dei casi, perché le convenzioni di naming dell'applicativo sono regolari, ma non infallibile, soprattutto per le operazioni il cui nome è ambiguo o non segue le convenzioni attese.

La ragione di questo limite è concreta e va dichiarata: verificare cosa un'operazione fa davvero — se legge o modifica un dato, quale dato tocca esattamente — richiederebbe l'accesso al codice sorgente del framework applicativo, che non è stato disponibile nel corso di questo lavoro. In sua assenza, la classificazione resta una stima informata, corretta a mano nei casi in cui il nome dell'operazione lascia ragionevolmente dedurre un errore dell'euristica, ma esplicitamente segnalata come incerta là dove il nome non consente una deduzione affidabile. Le operazioni classificate in modo errato — per

esempio una scrittura scambiata per lettura, perché il suo nome contiene un prefisso tipico delle interrogazioni — non compromettono il funzionamento complessivo del sistema, ma possono indebolire localmente un segnale: una scrittura non riconosciuta come tale, ad esempio, non contribuisce al conteggio delle operazioni di modifica, e un profilo che se ne fosse reso protagonista risulterebbe leggermente meno anomalo di quanto meriterebbe. È un limite che una verifica sistematica sul framework, condotta insieme a chi lo sviluppa e lo mantiene, ridurrebbe sensibilmente.

Un secondo limite, più circoscritto ma metodologicamente interessante, riguarda una specifica firma comportamentale: l'accanimento su un singolo oggetto, cioè la consultazione ripetuta e ossessiva della stessa cartella — la firma che, nella casistica, corrisponde al caso del medico che accede decine di volte al *dossier* di una singola persona. Il sistema non dispone di una caratteristica dedicata a misurare direttamente quante volte uno stesso oggetto viene riaperto: coglie questo comportamento solo in modo indiretto, attraverso il *fanout*, cioè il rapporto tra il numero complessivo di eventi e il numero di oggetti distinti toccati in una giornata.

Il *fanout* è però una media giornaliera, e come ogni media perde informazione sulla distribuzione sottostante. Un profilo che consulta pochi oggetti moltissime volte e uno che li consulta in modo più uniforme possono presentare *fanout* simili, pur corrispondendo a comportamenti diversi; e, all'opposto, un accanimento estremo su un singolo oggetto immerso in un'attività per il resto ordinaria può risultare, come media, meno marcato di quanto il fenomeno reale sia. In altre parole, il *fanout* è un buon indicatore dello stile di un accesso — concentrato o distribuito — ma un indicatore imperfetto dell'intensità con cui un singolo oggetto specifico viene preso di mira. Una caratteristica dedicata, che misurasse direttamente il numero massimo di accessi a uno stesso oggetto nella giornata, catturerebbe questa firma in modo più fedele; la sua assenza è una scelta di semplicità del modello attuale, non un ostacolo strutturale, e rientra tra i raffinamenti possibili.

L'ultimo limite non è, a rigore, un difetto, ma una scelta consapevole di cui è bene rendere esplicite le ragioni. Il sistema rileva gli accessi anomali, non li impedisce: fa emergere un comportamento dopo che si è manifestato, non blocca l'accesso nel momento in cui viene tentato. Si potrebbe ritenere preferibile un meccanismo che neghi l'accesso

sospetto sul nascere, ma in ambito sanitario questa soluzione non è praticabile, per due ragioni distinte.

La prima è tecnica: l'anomalia non risiede nel singolo evento, ma nel suo accumulo. Al momento del primo accesso di una sequenza, non esiste alcun elemento che lo distingua da uno legittimo — è solo osservando la sequenza nel suo insieme che il comportamento diventa anomalo. Non si può quindi bloccare in tempo reale ciò che, istante per istante, appare regolare. La seconda ragione è clinica, ed è più grave: negare a un operatore l'accesso a una cartella sulla base di un sospetto significa accettare il rischio che un falso allarme impedisca una consultazione legittima in una situazione di urgenza, con conseguenze potenzialmente serie sulla salute di un paziente. Il danno di un accesso improprio, per quanto grave, è di natura diversa e generalmente inferiore a quello di una cura ritardata. Non a caso, in ambito sanitario si adotta il paradigma del break the glass: l'accesso eccezionale non viene negato, ma concesso e tracciato, spostando la tutela dal piano dell'impedimento tecnico a quello della deterrenza — la consapevolezza di essere tracciati e la certezza della conseguenza disciplinare. Il sistema qui proposto si colloca precisamente in questo paradigma: non è una barriera, ma lo strumento che rende la tracciatura effettivamente utilizzabile, trasformando un archivio di *log* che nessuno consulta in un insieme ristretto di segnalazioni motivate su cui un ufficio *privacy* può realmente intervenire.

CAPITOLO IV

CONCLUSIONE E SVILUPPI FUTURI

SOMMARIO: 4.1 SINTESI DEL PERCORSO E CONTRIBUTO DEL LAVORO; 4.2 ESITO DELLA VALIDAZIONE E NATURA DI PROOF OF CONCEPT; 4.3 ARRICCHIRE IL SEGNALE: RELAZIONE DI CURA E RUOLO DELL'OPERATORE; 4.4 VERSO IL RILEVAMENTO IN TEMPO REALE E IL CONFRONTO CON I SISTEMI A REGOLE; 4.5 DALLA RILEVAZIONE ALL'ESERCIZIO: CONDIZIONI PER UN IMPIEGO REALE.

4.1 Sintesi del percorso e contributo del lavoro

Il lavoro ha preso le mosse da una difficoltà che i presidi tradizionali della sicurezza informatica non sono in grado di risolvere. Lo snooping sanitario non attraversa alcun perimetro, non forza alcuna barriera e non produce alcun errore: è compiuto da un operatore regolarmente autenticato, regolarmente autorizzato e, quasi sempre, in pieno orario di servizio. Ciò che manca non è un permesso, ma un titolo; e il titolo, a differenza del permesso, non è scritto da nessuna parte nel sistema informativo.

Da questa constatazione è nata la premessa metodologica dell'intero elaborato: se non è possibile stabilire a priori che cosa sia un accesso illecito, perché la valutazione richiederebbe di conoscere finalità e relazioni che il sistema non possiede, è però possibile misurare quanto un accesso sia anomalo rispetto al comportamento abituale di quell'operatore e a quello dei suoi pari. Le due nozioni non coincidono, e il lavoro non ha mai preteso il contrario: ciò che si è sostenuto, e che la casistica esaminata nel capitolo I conferma, è che le violazioni lasciano una traccia comportamentale, e che quella traccia è misurabile.

Su questa premessa è stata progettata e realizzata una catena di elaborazione completa. A partire dai *log* grezzi prodotti da un applicativo sanitario, il sistema li acquisisce senza alterarli, li traduce nello standard FHIR AuditEvent, attribuisce a ciascuna operazione un'area clinica e un livello di sensibilità derivato dalle categorie particolari dell'articolo 9 del GDPR, aggrega gli accessi in profili utente-giornata descritti da tredici

caratteristiche comportamentali, e infine isola i profili fuori norma con un algoritmo non supervisionato, corredando ogni segnalazione dei fattori che l'hanno determinata e di una priorità operativa governata da regole dichiarate.

Il contributo del lavoro non risiede in nessuno di questi stadi preso singolarmente: FHIR è uno standard pubblico, l'Isolation Forest un algoritmo consolidato, la raccolta dei *log* una pratica ordinaria. Risiede, piuttosto, in due elementi che li tengono insieme.

Il primo è la continuità della catena. Il percorso che va da un *log* scritto per finalità di diagnostica fino a una segnalazione leggibile da un ufficio *privacy* è stato compiuto per intero, senza interruzioni e senza passaggi lasciati all'intuizione: ogni stadio produce un formato che lo stadio successivo sa interpretare, e nessuno distrugge ciò che ha ricevuto. È anche la parte che ha richiesto il lavoro maggiore, e per una ragione che merita di essere dichiarata: i *log* di partenza non erano stati concepiti per essere letti in chiave di audit. Ricostruire da essi il contenuto informativo richiesto dallo standard ha significato risolvere problemi che nessuna descrizione teorica della pipeline lascerebbe prevedere, dall'oggetto acceduto che nel settantasette per cento dei casi non è il paziente ma la pratica, fino al nome dell'operazione assente in quattro richieste su dieci.

Il secondo elemento è il legame esplicito tra la parte giuridica e la parte tecnica. Le caratteristiche che il modello osserva non sono state scelte con criteri statistici astratti, ma derivate una per una dai comportamenti che il Garante ha già ritenuto illeciti nei propri provvedimenti. La tabella del paragrafo 3.1 è il punto in cui questo legame diventa visibile: a sinistra la firma comportamentale ricavata dalla casistica, a destra la grandezza che il sistema calcola per coglierla. È, a giudizio di chi scrive, l'aspetto per cui il lavoro ha senso come tesi di un percorso che tiene insieme diritto e informatica, e non come esercizio esclusivamente tecnico.

Vi è infine un esito che non era stato cercato. Poiché il modello misura lo scostamento e non il contenuto, esso intercetta anche forme di anomalia diverse dallo snooping in senso stretto — in particolare i profili con un numero insolito di operazioni di scrittura, che non sono consultazioni improprie ma possibili manomissioni, errori operativi ripetuti o usi del sistema fuori dalle prassi consuete. Non è l'obiettivo per cui l'impianto è stato costruito, ma è un beneficio dell'architettura che vale la pena registrare.

4.2 Esito della validazione e natura di proof of concept

La verifica di un sistema di rilevamento presuppone di sapere, caso per caso, quale sia la risposta giusta. Per il fenomeno qui trattato quella risposta non esiste: nessuna delle giornate analizzate è mai stata dichiarata, da un'autorità competente, come snooping accertato o come consultazione legittima. È la stessa asimmetria che motiva l'intero lavoro, riproposta sul piano della misura, e non si aggira etichettando a mano un campione, perché stabilire la legittimità di un accesso richiederebbe proprio l'informazione che manca.

La strada seguita è stata quella di costruire artificialmente dei casi di cui la risposta è nota, iniettando profili sintetici che riproducono le firme comportamentali dello snooping e misurando quanti di essi il modello ripesca nelle posizioni più alte della classifica. Escludendo l'unico profilo reale risultato stabilmente anomalo, il ROC-AUC medio si attesta a 0,999 e l'Average Precision a 0,967 su trenta ripetizioni indipendenti. Sull'insieme dei profili reali, il sistema porta all'attenzione dodici casi su duecentosei, pari a circa il sei per cento: non un allarme di massa, ma l'insieme ristretto di situazioni da sottoporre a verifica umana che il capitolo I aveva indicato come obiettivo.

Il confronto tra i cinque algoritmi ha inoltre mostrato che la scelta dell'Isolation Forest non era indifferente. Su anomalie nette quasi ogni metodo funziona, e il confronto non discrimina; ma quando l'intensità dell'anomalia viene modulata fino a renderla appena percettibile — il regime in cui uno snooper accorto effettivamente si colloca — l'Isolation Forest si distacca nettamente a partire da uno scostamento pari allo scarto normale, e degrada in modo regolare al calare dell'intensità. È una proprietà che, in un sistema destinato all'uso operativo, conta quanto la prestazione media: un metodo che peggiora in modo prevedibile è preferibile a uno che, come il Robust Covariance, resta inchiodato proprio nella fascia più difficile.

Va però riconosciuto con precisione che cosa questa validazione dimostri e che cosa non dimostri. Essa prova che il modello, applicato alle caratteristiche comportamentali, separa con elevata affidabilità un profilo anomalo da uno normale. Non prova che l'intera catena, dal *log* grezzo alla segnalazione, si comporti correttamente di fronte a un caso di abuso reale: le anomalie iniettate entrano nel sistema già in forma di profilo, e non

attraversano gli stadi che le avrebbero dovute produrre. Vi è inoltre una circolarità che è giusto dichiarare, poiché i profili sintetici sono costruiti a partire dalle stesse firme che il modello è progettato per riconoscere.

A ciò si aggiunge la natura dell'ambiente. La pipeline è stata sviluppata e collaudata su un ambiente dimostrativo aziendale popolato da anagrafiche fittizie, con la conseguenza — tutt'altro che secondaria in un lavoro dedicato alla protezione dei dati sanitari — che nessun dato clinico reale è stato trattato, ma anche con il limite che la normalità appresa dal modello è quella di un ambiente di prova e non quella di un reparto ospedaliero.

È per l'insieme di queste ragioni che il lavoro va inteso per ciò che è: una proof of concept completa e funzionante, che percorre l'intera catena e ne dimostra la praticabilità, corredata dei requisiti necessari a un'eventuale messa in esercizio, ma non ancora un sistema in produzione. I paragrafi che seguono indicano le direzioni lungo le quali quella distanza potrebbe essere colmata.

4.3 Arricchire il segnale: relazione di cura e ruolo dell'operatore

Esaminati i limiti nel capitolo precedente, i paragrafi che seguono li rileggono in positivo: ciascuna delle direzioni delineate risponde, dove possibile, a uno dei limiti discussi, ordinate dai raffinamenti più immediati e a maggior impatto fino alle estensioni di più ampio respiro.

Il primo e più promettente sviluppo è la rimozione del limite indicato come il più incisivo tra quelli esaminati: l'**assenza della relazione di cura** tra operatore e paziente. Allo stato attuale, il sistema sa riconoscere che un accesso è insolito, ma non se il paziente consultato rientri tra quelli effettivamente in carico a chi lo consulta. Integrare questa informazione trasformerebbe il segnale prodotto dal sistema da indiretto a diretto: non più soltanto "questo operatore ha toccato molti più pazienti del solito", ma "questo operatore ha consultato pazienti che non aveva in carico".

Sul piano pratico, questo sviluppo è più accessibile di quanto la sua importanza possa suggerire, e per una ragione precisa: l'informazione necessaria, in un contesto ospedaliero

reale, esiste già. I sistemi di gestione dei reparti sanno, per definizione, quali pazienti sono assegnati a quale operatore o a quale unità, perché è un dato indispensabile al loro funzionamento ordinario. Non si tratterebbe quindi di costruire da zero una nuova base di conoscenza, ma di collegare al sistema una fonte dati già disponibile. Sul piano del modello, l'integrazione richiederebbe l'aggiunta di una caratteristica costruita sulla falsariga di quelle già presenti — per esempio, la quota di oggetti consultati che ricadono al di fuori dell'insieme dei pazienti in carico — calcolata con la stessa logica con cui il sistema già misura, oggi, la novità di un oggetto rispetto alla storia passata dell'utente.

Vi è però una complessità da non sottovalutare, che riguarda la natura dinamica della relazione di cura: un paziente entra ed esce dalla presa in carico di un operatore nel tempo, e ciò che è legittimo consultare oggi poteva non esserlo un mese fa. Un'integrazione corretta non potrebbe quindi limitarsi a una fotografia dello stato attuale, ma dovrebbe disporre dello storico delle assegnazioni, per poter valutare ogni accesso rispetto alla relazione di cura vigente nel momento in cui esso è avvenuto. È lo stesso principio di causalità già applicato, nel sistema attuale, al calcolo della baseline comportamentale: un accesso va valutato con l'informazione disponibile al suo tempo, non con quella successiva. Con questa cautela, l'integrazione della relazione di cura rappresenta il singolo miglioramento di maggior impatto possibile, capace di avvicinare sensibilmente la nozione statistica di accesso anomalo a quella giuridica di accesso illecito.

Sulla stessa linea si colloca uno sviluppo di portata più contenuta, ma già interamente predisposto dai dati disponibili, che riguarda l'uso del ruolo dell'operatore come ulteriore segnale di anomalia. Nel corso della normalizzazione, il sistema conserva già, per ogni accesso, il profilo e il reparto di chi lo ha compiuto — informazioni che oggi vengono usate per il confronto tra pari, ma non ancora come segnale autonomo. Esse rendono possibile una caratteristica che il modello attuale non sfrutta: la coerenza tra il ruolo di un operatore e la natura dei dati che consulta.

L'intuizione è che certi abbinamenti tra profilo e tipo di dato siano intrinsecamente più sospetti di altri, a prescindere dal volume. Un profilo puramente amministrativo che accede a diagnosi cliniche, o un profilo tecnico che consulta dati appartenenti alle aree più sensibili, costituisce un'anomalia di natura diversa da quella del volume o dello scostamento: è un accesso che può essere numericamente modesto — poche consultazioni

— ma incoerente con la funzione di chi lo compie. Una caratteristica costruita su questa incoerenza coglierebbe una firma di snooping che oggi il sistema intercetta solo indirettamente, e lo farebbe riutilizzando un dato — il profilo utente — che è già presente in ogni AuditEvent, senza bisogno di alcuna fonte aggiuntiva. È un raffinamento del modello, non un'estensione dell'infrastruttura, e come tale rappresenta uno degli sviluppi a più immediata realizzabilità.

4.4 Verso il rilevamento in tempo reale e il confronto con i sistemi a regole

Il sistema descritto in questo lavoro è retrospettivo: ragiona per giornate concluse, aggregando gli accessi di ciascun operatore su base giornaliera. Questa scelta, come discusso, è ciò che neutralizza il fan-out dei portali e rende le segnalazioni leggibili per un audit; ma comporta che il sistema si accorga di un comportamento anomalo solo dopo che la giornata è terminata. È adatto, quindi, a un controllo a posteriori — a rivedere ciò che è accaduto — ma non a intercettare un abuso mentre è ancora in corso.

Una direzione naturale di sviluppo è affiancare al modello giornaliero un secondo livello, capace di valutare gli accessi via via che arrivano. Il lavoro contiene già la base concettuale di questo secondo livello, delineata nel capitolo III: un percorso parallelo, tenuto separato dalla pipeline principale, che invece di aggregare per giornata calcola per ogni singolo evento un insieme di caratteristiche a finestra scorrevole — quante operazioni lo stesso utente ha compiuto nell'ultimo minuto, nell'ultima ora, nelle ultime ventiquattro ore, valutate nell'istante in cui l'evento si presenta e guardando soltanto al passato. È lo stesso principio di causalità del modello giornaliero, applicato però a una scala temporale molto più fine, compatibile con una valutazione immediata anziché a fine giornata.

Un'architettura matura di questo tipo si articolerebbe in due fasi complementari: un addestramento offline periodico, che apprende la normalità e fissa le soglie a partire dallo storico consolidato, e un servizio online continuo, che applica quel modello agli eventi in arrivo e produce un segnale nel momento in cui una sequenza di accessi comincia a discostarsi dalla norma.

Un simile sistema non contraddirebbe la scelta di rilevare anziché prevenire: continuerebbe a non bloccare gli accessi, per le ragioni cliniche già viste, ma ridurrebbe drasticamente il tempo tra l'inizio di un comportamento anomalo e la sua emersione, spostando la segnalazione dal giorno successivo al momento stesso in cui il comportamento si manifesta. Il valore di questo sviluppo non è quindi nel bloccare prima, ma nel sapere prima, restringendo la finestra entro cui un abuso resta invisibile.

L'unica direzione, di natura diversa dalle precedenti, non estende il sistema ma lo mette alla prova rispetto a un termine di paragone. Come discusso fin dal Capitolo I, l'approccio comportamentale adottato in questo lavoro si contrappone a quello, più tradizionale, dei sistemi basati su regole deterministiche – soglie fisse, condizioni note, pattern predefiniti – tipicamente implementati sulle piattaforme SIEM in uso nelle organizzazioni. È in corso di valutazione un confronto sperimentale tra i due approcci, condotto sugli stessi AuditEvent prodotti da questo lavoro, con l'obiettivo di misurare non tanto quale dei due "trovi di più", quanto cosa ciascuno vede che l'altro non vede.

L'ipotesi che il confronto intende verificare discende direttamente dall'argomento del Capitolo I. Un sistema a regole fisse, privo di una baseline storica per ciascun operatore, non può distinguere ciò che è insolito per quella persona da ciò che è semplicemente vero in generale: dovrebbe quindi cogliere bene alcune firme immediate e statiche – un accesso da un indirizzo di rete inatteso, un'operazione in un orario anomalo – ma restare cieco proprio agli scostamenti comportamentali che costituiscono il cuore dello snooping, quelli che solo il confronto con una norma appresa può rivelare. All'opposto, un sistema a regole potrebbe intercettare condizioni che il modello comportamentale non considera affatto, come determinati pattern tecnici o di rete. I due approcci, in questa lettura, non sono in competizione ma complementari: le regole deterministiche colgono l'atteso e l'immediato, il modello comportamentale coglie l'inatteso e il progressivo. Un confronto rigoroso sui dati reali permetterebbe di quantificare questa complementarità, e costituirebbe la premessa naturale a quell'integrazione con una piattaforma SIEM che, come anticipato nell'introduzione, rappresenta l'estensione più ambiziosa dell'intero lavoro: un'architettura in cui il modello comportamentale fornisce il segnale che le regole, da sole, non saprebbero produrre.

4.5 Dalla rilevazione all'esercizio: condizioni per un impiego reale

Un sistema che analizza i *log* di accesso a dati sanitari è esso stesso un trattamento di dati personali, e in parte delicati: i *log* registrano chi ha consultato quale paziente, quando e da dove – informazioni che, pur non contenendo il dato clinico in sé, permettono di risalire a comportamenti e relazioni. Un lavoro che si occupa di protezione dei dati non può ignorare la protezione dei dati che esso stesso maneggia, e questo apre una direzione di sviluppo che è insieme tecnica e organizzativa.

Sul piano tecnico, una prima cautela è già stata adottata in questo lavoro: la pseudonimizzazione degli identificativi degli operatori nelle segnalazioni prodotte. È l'applicazione concreta, al sistema stesso, del principio di minimizzazione: trattare in chiaro solo i dati indispensabili, e solo dove servono.

Sul piano organizzativo, invece, la messa in esercizio richiederebbe passaggi che esulano dalla competenza tecnica e vanno dichiarati come prerequisiti. Il trattamento dei *log* per finalità di controllo degli accessi impropri, pur essendo non solo lecito ma in una certa misura dovuto – le stesse linee guida sul *dossier* sanitario impongono di tracciare gli accessi e riconoscono al paziente il diritto di conoscerli – andrebbe formalmente inquadrato: definito nella sua base giuridica e nella sua finalità, coperto dalle autorizzazioni interne, e valutato dal responsabile della protezione dei dati dell'organizzazione, cui spetterebbe stabilire se sia necessario un aggiornamento del registro dei trattamenti o una valutazione d'impatto. Si tratta di adempimenti che non incidono sul funzionamento del sistema, ma senza i quali la sua adozione in un contesto reale non sarebbe corretta. Dichiararli come parte integrante dei requisiti di messa in esercizio, e non come un dettaglio successivo, è coerente con lo spirito dell'intero lavoro: un sistema che sorveglia il rispetto della *privacy* deve essere, per primo, esemplare nel rispettarla.

Sul piano operativo, un ulteriore miglioramento riguarda il primo stadio della pipeline, quello di acquisizione dei log. Nella configurazione attuale, i file prodotti dal filtro applicativo vengono prelevati manualmente e depositati nella cartella da cui parte l'elaborazione. È già stato sviluppato, e verificato singolarmente, uno strumento che automatizza questo prelievo, collegandosi al server remoto e scaricando i soli file non

ancora acquisiti; esso non è però ancora stato integrato nell'orchestratore che governa l'esecuzione automatica della pipeline. Il suo aggancio alla sequenza costituisce un passo immediato — non una progettazione da affrontare, ma un'integrazione già delineata — che eliminerebbe l'ultimo intervento manuale rimasto nel flusso, rendendo la catena interamente automatica dall'arrivo del *log* fino alla segnalazione.

Resta, al di là di ogni sviluppo tecnico, la ragione per cui il problema merita di essere affrontato. I provvedimenti esaminati nel primo capitolo raccontano quasi sempre la stessa vicenda: una persona scopre, per una voce che ha iniziato a circolare in reparto, che qualcuno ha guardato la sua storia clinica senza averne titolo, e solo allora l'organizzazione va a cercare nei propri *log* una traccia che era lì da sempre. Un sistema come quello descritto in queste pagine non impedisce quell'accesso, e non pretende di stabilire se sia stato illecito: si limita a fare in modo che, quando accade, qualcuno se ne accorga prima della vittima. È un obiettivo più modesto di quanto la tecnologia lasci immaginare, ma è esattamente ciò che oggi manca.

BIBLIOGRAFIA

DOTTRINA

E. F. AGYEMANG, *Anomaly detection using unsupervised machine learning algorithms: a simulation study*, in Scientific African, vol. 26, 2024, e02386, consultabile *online* sul sito https://doi.org/10.1016/j.sciaf.2024.e02386.

L. AYALA, *Cybersecurity for Hospitals and Healthcare Facilities: A Guide to Detection and Prevention*, I ed., Apress, Berkeley (California), 2016.

E. BERTINO, *Data Protection from Insider Threats*, I ed., Springer, Cham (Svizzera), 2012.

M. BISHOP, *Computer Security: Art and Science*, II ed., Addison-Wesley, Boston, 2019.

A. A. BOXWALA, J. KIM, J. M. GRILLO, L. OHNO-MACHADO, *Using statistical and machine learning to help institutions detect suspicious access to electronic health records*, in Journal of the American Medical Informatics Association, vol. 18, n. 4, 2011, pp. 498-505, consultabile *online* al sito https://doi.org/10.1136/amiajnl-2011-000217.

M. M. BREUNIG, H.-P. KRIEGEL, R. T. NG, J. SANDER, *LOF: Identifying Density-Based Local Outliers*, in Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, ACM, New York, 2000, pp. 93-104, consultabile *online* sul sito https://doi.org/10.1145/335191.335388.

J. DAVIS, M. GOADRICH, *The Relationship Between Precision-Recall and ROC Curves*, in Proceedings of the 23rd International Conference on Machine Learning (ICML), 2006, pp. 233-240, consultabile *online* sul sito https://doi.org/10.1145/1143844.1143874.

T. FAWCETT, *An introduction to ROC analysis*, in Pattern Recognition Letters, vol. 27, n. 8, 2006, consultabile *online* sul sito https://doi.org/10.1016/j.patrec.2005.10.010.

J. HAN, J. KIM, S. LEE, *5W1H-based Expression for the Effective Sharing of Information in Digital Forensic Investigations*, 2020, consultaible *online* al sito https://doi.org/10.32877/bt.v9i1.4053.

C. R. HARRIS ET AL., *Array programming with NumPy*, in Nature, vol. 585, 2020, pp. 357-362, consultaibile *online* sul sito *https://doi.org/10.14293/s2199-1006.1.sor-uncat.a7759461.v1.raugfi*.

M. HILDEBRANDT, *Law for Computer Scientists and Other Folk*, Oxford University Press, Oxford, 2020.

A. A. HOPGOOD, *Intelligent Systems for Engineers and Scientists*, IV ed., CRC Press, Boca Raton (Florida), 2021.

J. D. HUNTER, *Matplotlib: A 2D Graphics Environment*, in Computing in Science & Engineering, vol. 9, n. 3, 2007, pp. 90-95, consultabile *online* sul sito https://doi.org/10.1109/mcse.2007.55.

K. KENT, M. SOUPPAYA, *Guide to Computer Security Log Management*, NIST Special Publication 800-92, 2006, consultabile *online* sul sito https://doi.org/10.6028/nist.sp.800-92.

F. T. LIU, K. M. TING, Z.-H. ZHOU, *Isolation Forest*, in Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), IEEE, 2008, consultabile *online* sul sito https://doi.org/10.1109/icdm.2008.17.

F. T. LIU, K. M. TING, Z.-H. ZHOU, *Isolation-Based Anomaly Detection*, in ACM Transactions on Knowledge Discovery from Data, vol. 6, n. 1, 2012, pp. 1-39, consultabile *online* sul sito https://doi.org/10.1145/2133360.2133363.

W. MCKINNEY, *Data Structures for Statistical Computing in Python*, in Proceedings of the 9th Python in Science Conference, 2010, pp. 56-61, consultabile *online* sul sito https://doi.org/10.25080/majora-92bf1922-00a.

A. K. MENON, X. JIANG, J. KIM, J. VAIDYA, L. OHNO-MACHADO, *Detecting Inappropriate Access to Electronic Health Records Using Collaborative Filtering*, in Machine Learning, vol. 95, n. 1, 2014, pp. 87-101, consultabile *online* al sito https://doi.org/10.1007/s10994-013-5376-1.

F. PEDREGOSA ET AL., *Scikit-learn: Machine Learning in Python*, in Journal of Machine Learning Research, vol. 12, 2011, pp. 2825-2830, consultabile *online* al sito https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

S. RAMASWAMY, R. RASTOGI, K. SHIM, *Efficient Algorithms for Mining Outliers from Large Data Sets*, in Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, ACM, New York, 2000, pp. 427-438, consultabile *online* sul sito https://doi.org/10.1145/335191.335437.

P. J. ROUSSEEUW, K. VAN DRIESSEN, *A Fast Algorithm for the Minimum Covariance Determinant Estimator*, in Technometrics, vol. 41, n. 3, 1999, pp. 212-223, consultabile *online* sul sito https://doi.org/10.1080/00401706.1999.10485670.

S. J. RUSSELL, P. NORVIG, *Artificial Intelligence: A Modern Approach,* IV ed., *Global Edition, Pearson, Harlow* (Regno Unito), 2022.

L. SÁ CORREIA, R. J. CRUZ-CORREIA, P. P. RODRIGUES, *Illegitimate HIS access by healthcare professionals: scenarios, use cases and audit trail-based detection model*, in Procedia Computer Science, vol. 164, 2019, pp. 629-636, consultabile *online* sul sito https://doi.org/10.1016/j.procs.2019.12.229.

B. SCHÖLKOPF, J. C. PLATT, J. SHAWE-TAYLOR, A. J. SMOLA, R. C. WILLIAMSON, *Estimating the Support of a High-Dimensional Distribution*, in Neural Computation, vol. 13, n. 7, 2001, pp. 1443-1471, consultabile *online* sul sito https://doi.org/10.1162/089976601750264965.

W. STALLINGS, L. BROWN, *Computer Security: Principles and Practice*, V ed., Pearson, New York, 2024.

F. YERGEAU, *UTF-8, a transformation format of ISO 10646*, RFC 3629, 2003, consultabile *online* sul sito https://doi.org/10.17487/rfc3629.

NORMATIVA

Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (GDPR): Considerando n. 35; artt. 4, 9, 15, 25, 29, 32.

D.lgs. 30 giugno 2003, n. 196, Codice in materia di protezione dei dati personali, come modificato dal d.lgs. 10 agosto 2018, n. 101: artt. 2-quaterdecies, 92.

Codice penale, art. 615-ter, Accesso abusivo a un sistema informatico o telematico.

Legge 28 giugno 2024, n. 90, Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici, in G.U. n. 153 del 2 luglio 2024.

Legge 8 marzo 2017, n. 24, art. 4.

Legge 7 agosto 1990, n. 241, artt. 22, 24, comma 7, e 25.

D.P.R. 27 marzo 1969, n. 128, art. 7.

GIURISPRUDENZA

CORTE DI CASSAZIONE, Sezioni Unite penali, sentenza 7 febbraio 2012, n. 4694 (ud. 27 ottobre 2011).

CORTE DI CASSAZIONE, Sez. III penale, sentenza 2 agosto 2023, n. 33972.

CORTE DI CASSAZIONE, sentenza 9 maggio 2025, n. 17647.

PROVVEDIMENTI DEL GARANTE PER LA PROTEZIONE DEI DATI PERSONALI

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Dossier sanitario e trattamento dei dati personali dei pazienti*, Provvedimento n. 3 del 10 gennaio 2013, doc. web n. 2284708.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Trattamento di dati tramite il dossier sanitario aziendale*, Provvedimento n. 340 del 3 luglio 2014, doc. web n. 3325808.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Linee guida in materia di Dossier sanitario*, Provvedimento del 4 giugno 2015, doc. web n. 4084632.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza ingiunzione nei confronti di Perrone Rosaria e Azienda Universitaria Ospedaliera Ospedali Riuniti di Trieste*, Provvedimento n. 15 del 19 gennaio 2017, doc. web n. 6521868.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Provvedimento correttivo e sanzionatorio nei confronti di Azienda Ospedaliero Universitaria Integrata di Verona*, Provvedimento n. 18 del 23 gennaio 2020, doc. web n. 9269629.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza ingiunzione nei confronti di Azienda provinciale per i servizi sanitari di Trento*, Provvedimento n. 155 del 21 aprile 2021, doc. web n. 9678507.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza ingiunzione nei confronti di Ospedale Pediatrico Bambino Gesù*, Provvedimento n. 250 del 24 giugno 2021, doc. web n. 9689566.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza ingiunzione nei confronti di Azienda sanitaria unica regionale Marche*, Provvedimento n. 9 del 13 gennaio 2022, doc. web n. 9744496.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Ordinanza-ingiunzione nei confronti dell'Azienda USL Valle d'Aosta*, Provvedimento n. 371 del 10 novembre 2022, doc. web n. 9819792.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Provvedimento n. 97 del 22 febbraio 2024, doc. web n. 10001279.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Provvedimento n. 83 del 13 febbraio 2025, doc. web n. 10116834.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Provvedimento n. 533 del 25 settembre 2025, doc. web n. 10185490.

SITOGRAFIA

AGENZIA PER L'ITALIA DIGITALE (AGID), *Linee guida sul punto di accesso telematico ai servizi della Pubblica Amministrazione*, https://www.agid.gov.it/

APACHE SOFTWARE FOUNDATION, *Apache Tomcat*, https://tomcat.apache.org/

ASSOCIATION OF HEALTH CARE JOURNALISTS, https://healthjournalism.org/

ECLIPSE FOUNDATION, *Jakarta Servlet Specification*, https://jakarta.ee/

EUROPEAN PATIENTS' ACADEMY (EUPATI), https://eupati.eu/

HL7 INTERNATIONAL, *FHIR - AuditEvent Resource*, https://www.hl7.org/fhir/

IBM, *What Is a Data Pipeline?*, https://www.ibm.com/

INTERNET ENGINEERING TASK FORCE, *RFC Editor*, https://www.rfc-editor.org/

JSON, *Introducing JSON*, https://www.json.org/

JSON LINES, https://jsonlines.org/

MICROSOFT, *Windows-1252 code page*, https://learn.microsoft.com/

MONGODB, https://www.mongodb.com/

MOZILLA DEVELOPER NETWORK, *HTTP response status codes*, https://developer.mozilla.org/

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST), https://nist.gov/

NIST/SEMATECH, *e-Handbook of Statistical Methods*, https://www.itl.nist.gov/

OFFICE OF THE NATIONAL COORDINATOR FOR HEALTH INFORMATION TECHNOLOGY (ONC), *What is an Electronic Health Record (EHR)?*, https://healthit.gov/

ORACLE, *The Java Tutorials*, https://docs.oracle.com/

Matplotlib, https://matplotlib.org/

NumPy, https://numpy.org/

pandas, https://pandas.pydata.org/

PyMongo, https://pymongo.readthedocs.io/

Python, https://www.python.org/

scikit-learn, https://scikit-learn.org/

Watchdog, https://python-watchdog.readthedocs.io/