



**POLITECNICO
MILANO 1863**

**SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE**

Data quality assessment di log estratti dal sistema SICID
Preparazione dei dati a task di process mining
Maggio, 2024

Autore: Puccia Niccolò
Matricola: 987595
Codice Persona: 10829496

Docente di riferimento: Pernici Barbara
Docente responsabile: Fugini Mariagrazia

Contents

1	Introduzione	2
1.1	Panoramica	2
1.2	Contesto e stato dell'arte	2
1.3	Obiettivi	2
1.4	Descrizione degli strumenti utilizzati e discussione delle scelte effettuate	3
2	Sviluppi	4
2.1	Descrizione del DB e delle principali tabelle	4
2.1.1	Storico	4
2.1.2	Fascicoli	4
2.1.3	Defi	4
2.1.4	StatiEvento	4
2.2	Data exploration: Analisi della macchina stati/eventi	5
2.2.1	Filtro su ritualità 4O	5
2.2.2	Rapporto archi/autoanelli	5
2.2.3	Generalità sul dominio degli attributi	6
2.2.4	Analisi degli stati pozzo	6
2.3	Data understanding: Ctipos e Riformata	8
2.3.1	Studio del significato del dominio di CTIPOS e RIFORMATA	8
2.3.2	Quando Riformata è riempita Ctipos presenta lo stesso valore	10
2.3.3	Comparazione CDESCR, RIFORMATA e CTIPOS	11
2.4	Data understanding: Idgrpev	12
2.4.1	Descrizione valori	13
2.5	Analisi eventi di correzione	14
2.6	Studio dei cambi di ritualità	15
2.7	Analisi dei piani temporali	15
3	Conclusione	16
3.1	Risultati ottenuti	16
3.2	Future analisi e approfondimenti	16

1 Introduzione

1.1 Panoramica

Lo studio di cui si riportano i risultati commentati nel presente report è stato condotto in occasione della partecipazione dello studente al Progetto di Ingegneria Informatica nell'anno universitario 2023/24.

1.2 Contesto e stato dell'arte

Questa analisi si sviluppa sulla scia di studi precedentemente condotti all'interno del dipartimento di Elettronica, Informazione e Bioingegneria del Politecnico di Milano circa l'impatto di strumenti di data mining [1] e deep learning [2] sugli indicatori di performance dei processi nel settore giudiziario.

Indagini condotte dalla commissione europea per l'efficienza della giustizia (Cepej) riguardo i processi civili e commerciali svolti in Italia nell'anno 2020 sembrano confermare il nostro paese tra i primi posti nella classifica degli stati europei per la durata dei processi giudiziari.

Il Disposition Time italiano, ovvero l'indice che stima il tempo che mediamente intercorre tra la presa in carico di un processo e la sua deposizione, supera ampiamente quello degli altri stati membri arrivando, nel solo caso di Corte di Cassazione, a 1156 giorni contro i 172 che rappresentano la mediana europea.

Con l'approvazione del PNRR, l'Italia si è impegnata nei confronti della Commissione europea a raggiungere obiettivi specifici, con particolare riferimento alla riduzione dei tempi e aprendo a tal proposito una linea di ricerca che coinvolge le maggiori università italiane in ambito giudiziario e tecnologico.

Nell'abbracciare il progetto *"NEXT GENERATION UPP: nuovi schemi collaborativi tra Università e uffici giudiziari per il miglioramento dell'efficienza e delle prestazioni della giustizia nell'Italia Nord Ovest"* il Politecnico di Milano ha specializzato il campo di ricerca sui seguenti temi:

1. Tecniche di data mining all'avanguardia possono consentire ai giudici di prendere decisioni più consapevoli in merito alla distribuzione dei casi e del personale?
2. E' possibile sviluppare un cruscotto di monitoraggio dei processi di tipo predittivo, in cui si visualizzi non solo l'andamento corrente dei processi, ma si possa avere una stima dei tempi attesi, che consenta di monitorare quei processi che potrebbero raggiungere tempi critici?
3. Le normative messe in atto per consentire la velocizzazione delle pratiche consentono ai giudici di esplorare soluzioni alternative a quelle storicamente esistenti (nominate *"varianti"*). Tali soluzioni sono state percorse?

In seguito all'informatizzazione progressiva dei sistemi di acquisizione e elaborazione dei dati e all'affinamento delle tecniche di process mining per la gestione dei processi giudiziari, i risultati ottenuti negli ultimi anni sembrano confermare un leggero miglioramento del trend registrato nel 2020 ma sono ancora ampi i margini d'azione.

Il rilascio della prima versione del cruscotto predittivo nel giugno 2024 ha rappresentato un significativo passo avanti nella ricerca condotta dal dipartimento. Tuttavia, sono emersi importanti dubbi riguardo alla consistenza e al significato intrinseco dei dati che costituiranno il training set del modello. Questi dati provengono dai database relativi ai processi, sia archiviati che non, della Corte di Appello di Milano nel periodo compreso tra il 2004 e il 2024.

1.3 Obiettivi

Il progetto i cui sviluppi rappresentano il contenuto del presente report nasce con i seguenti obiettivi:

1. Studiare la consistenza del log rispetto alla descrizione del processo in termini di stati e transizioni.
2. Studiare la qualità dei dati estratti in termini di consistenza, completezza e coerenza per dimostrare che le analisi che su questi si fondano siano valide.
3. Ampliare il dizionario dati.
4. Individuare eventuali problematiche derivanti da errori nel processo di acquisizione e registrazione degli stessi.

Una volta raggiunta una piena comprensione del contesto:

1. Studiare l'impatto di diverse tecniche di cleaning sul risultato finale utilizzando il predittore.

1.4 Descrizione degli strumenti utilizzati e discussione delle scelte effettuate

Per raggiungere l'obiettivo preposto, è stata inizialmente condotta una fase esplorativa dei dati a disposizione utilizzando strumenti che ne consentissero l'analisi e la visualizzazione. Jupyter Notebook è stato il tool software scelto per svolgere questo compito in quanto permette di disporre di:

- **Interattività e Flessibilità:** Jupyter Notebook permette di scrivere ed eseguire codice in modo incrementale, facilitando la sperimentazione e la verifica immediata dei risultati. Questo approccio interattivo è fondamentale per l'analisi esplorativa dei dati, dove è necessario testare rapidamente diverse ipotesi e visualizzare i risultati.
- **Visualizzazione Integrata:** Grazie alla possibilità di integrare grafici e tabelle direttamente nel notebook, Jupyter facilita un'interpretazione visiva immediata dei dati. Strumenti come Matplotlib e Seaborn, spesso utilizzati con Jupyter, permettono di creare visualizzazioni complesse che aiutano a comprendere meglio i pattern e le tendenze nei dati.

Successivamente, è stato creato un database locale utilizzando i dati forniti opportunamente riorganizzati. Questa scelta è stata motivata da diversi vantaggi:

- **Ottimizzazione dell'Accesso ai Dati:** Avere un database locale permette un accesso rapido e ottimizzato ai dati. Nel caso in analisi ciò ha consentito di eseguire queries complesse o ripetitive, riducendo i tempi di latenza associati a database remoti. Nonostante questo, nell'ottica di rendere il software utilizzabile anche per coloro i quali non desiderano avere una copia locale dell'intera base di dati, particolare attenzione è stata prestata a parametrizzare il codice in modo tale da consentire anche l'accesso remoto al DB (qualora l'organizzazione che lo gestisce ne consenta i permessi).
- **Flessibilità nella Manipolazione dei Dati:** Riorganizzare i dati in un database locale consente di strutturarli in modo più appropriato per le analisi da condurre. Questo processo di riorganizzazione aiuta a migliorare la coerenza e l'integrità dei dati, facilitando la loro successiva manipolazione tramite script Python. Nel nostro caso le operazioni di modifica della base di dati sono state minime nell'interesse di preservarne il contenuto, sul quale non c'è possibilità d'intervento. Le modifiche effettuate sono state principalmente la ridenominazione di attributi equivoci e l'eliminazione di colonne non utili ai fini dell'analisi.

Per l'estrazione e la manipolazione delle informazioni contenute nel database, sono state utilizzate le librerie Pandas e Numpy con l'obiettivo di sfruttarne:

- **Funzionalità avanzate:** Pandas e Numpy sono ottimizzati per operazioni su grandi quantità di dati. Numpy fornisce supporto per array e matrici multidimensionali, mentre Pandas offre strumenti potenti per la manipolazione e l'analisi di dati tabulari. Ciò permette di eseguire analisi approfondite e ottenere insights significativi dai dati.

Il risultato finale del processo di analisi è stata la realizzazione di una libreria di funzioni Python che ricevono come argomento i risultati di query SQL e li elaborano per trarne valore e significato. Questo approccio è stato scelto per diversi motivi:

- **Modularità e Riutilizzabilità:** Creare una libreria di funzioni modulari consente di riutilizzare il codice in diverse parti del progetto e in progetti futuri. Questa modularità migliora la manutenibilità del codice e facilita l'estensione delle funzionalità.
- **Separazione delle Responsabilità:** Separando l'estrazione dei dati (tramite SQL) dalla loro elaborazione (tramite funzioni Python), si ottiene un'architettura più chiara e gestibile semplificando anche il debug e l'evoluzione del sistema.

2 Sviluppi

2.1 Descrizione del DB e delle principali tabelle

Di seguito viene riportata una breve descrizione seguita dallo schema delle principali tabelle analizzate. Qualora l'intero schema risulti contenere troppi attributi viene presentata una selezione di quelli maggiormente utilizzati nelle analisi condotte. Gli attributi che costituiscono la chiave primaria di ogni schema sono stati sottolineati.

2.1.1 Storico

La tabella storico contiene i log dei processi nell'ordine in cui sono stati registrati dal sistema. Ogni entry della tabella corrisponde a un evento relativo a un particolare processo in corso del quale si specificano numero di fascicolo, data di accadimento, data di registrazione in piattaforma, codice utente e descrizione dell'evento.

<u>NUMPRO</u>	<u>NUMPRV</u>	CCDOEV	CTIPSE	NUMGIU	DATAEV	CCODST	CNPARA	CDESCR	DATARE
CODUTE	CEVPAD	CSTAPR	ISVIS	CRONO	TIPOATTO	ULTIMAMOD	PARAMS	NOTA	IDDOCS

Table 1: Stor Table schema

2.1.2 Fascicoli

La tabella Fasc contiene informazioni riguardanti il fascicolo cui il singolo processo fa riferimento. In particolare ogni entry riporta: l'anno di ruolo, il numero di ruolo, l'identificativo del giudice, l'esito (se il processo concluso), la ritualità e il grado di giudizio. Viene riportato di seguito il sottoinsieme degli attributi maggiormente utilizzati in fase di analisi:

<u>NUMPRO</u>	CANRUO	CNURUO	DISCRI	CTIPSE	NUMGIU	CCODST	CSTAPR	ANNOGIUDICE
<u>RUOLOGGIUDIC</u>	<u>COLLEG</u>	<u>CODICEOGG</u>	<u>RIASSCASSAZ</u>	<u>IDREPFASC</u>	<u>ULTIMAMOD</u>	<u>CLASS_ACT</u>		

Table 2: Fasc Table schema

2.1.3 Defi

La tabella Defi contiene informazioni riguardanti il deposito del fascicolo e la chiusura del processo. Fornisce informazioni riguardo: la data di deposito del provvedimento, l'esito del provvedimento, l'indicazione sull'eventuale conferma o riforma della sentenza.

<u>NUMPRO</u>	<u>NUMPRV</u>	CNUMRV	DATUDE	DACONC	DATPUB	CTIPOS	NUMGIU	RIFORMATA

Table 3: Defi Table schema

2.1.4 StatiEvento

L'insieme dei processi che coinvolgono un tribunale può essere modellato come una macchina a stati per la quale il verificarsi di eventi provoca la transizione da uno stato a un altro. Le informazioni relative alla composizione di tale macchina sono contenute in una tabella che associa a ogni coppia (stato di partenza, evento verificatosi) uno stato destinazione.

RIT	<u>CURSTATE</u>	IDGRPEV	<u>IDEVENTO</u>	<u>NEWSTATE</u>	FKEVENT	IDSTATIEV	ISVISIBLE

Table 4: StatiEvento Table schema

2.2 Data exploration: Analisi della macchina stati/eventi

Come introdotto nel paragrafo precedente l'iter che un processo segue una volta presentato alla corte è modellabile tramite una macchina a stati.

In particolare in ogni momento un processo si trova in uno stato e da quello stato passa a un altro al sopraggiungere di uno specifico evento (ad esempio la delibera di un giudice o il deposito di documenti la cui attesa obbligava alla permanenza in un certo stato).

Si noti come la macchina riflette la normativa vigente in quanto essa regola quali transizioni sono possibili da un certo stato verso un altro. Come conseguenza la macchina è un oggetto dinamico poiché risente di continui cambiamenti dall'esterno.

Per inquadrare il contesto di analisi e per ricercare eventuali anomalie all'interno dei log che rappresentano gli eventi che scatenano le transizioni tra stati sono state condotte analisi esplorative della macchina stati-eventi.

2.2.1 Filtro su ritualità 4O

E' stato scelto di limitare l'analisi ai processi di ritualità 4O. Sono state pertanto eliminate le righe riguardanti altre ritualità.

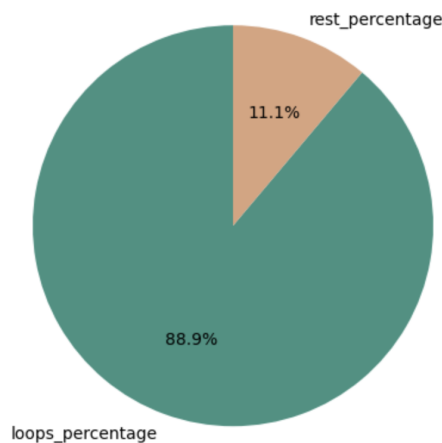
RIT	CURSTATE	IDGRPEV	IDEVENTO	NEWSTATE	FKEVENT	IDSTATIEV	ISVISIBLE
11578	40	DF	PP	33	++	MOD_DOM	4ODFPP33
23664	40	RZ	-	RD	++	69	4ORZ-RD
23665	40	RZ	-	RQ	UT	210	4ORZ-RQ
23666	40	RZ	-	RR	CP	16	4ORZ-RR
24223	40	D1	PP	DQ	++	106	4OD1PPDQ

Table 5: Esempio tabella dei processi

2.2.2 Rapporto archi/autoanelli

Una volta acquisite le necessarie informazioni di contesto si è passato ad analizzare la macchina a stati da una prospettiva di alto livello.

Il primo passo fatto in tale direzione è stato quello di determinare quanti eventi generano autoanelli sul totale delle transizioni. Di seguito vengono riportati i risultati ottenuti:



Risulta interessante osservare come la maggior parte degli eventi non generano transizioni verso altri stati ma, al contrario, mantengono il processo nello stato corrente.

2.2.3 Generalità sul dominio degli attributi

Successivamente è stata condotta un'analisi di tipo descrittivo sui singoli attributi della tabella studiando per ognuno di essi la distribuzione dei valori contenuti, contando valori assenti o nulli e calcolando medie e mediane dei valori numerici.

RIT	CURSTATE	IDGRPEV	IDEVENTO	NEWSTATE	FKEVENT	IDSTATIEV
count	3643	3643	3643	3643	3643	3643
unique	1	44	17	306	39	174
top	40	UT	-	DECO	++	1
freq	3643	245	1805	44	3237	620

Table 6: Informazioni dalla tabella stati-eventi

2.2.4 Analisi degli stati pozzo

Una classificazione rilevante ai fini dell'analisi predittiva riguarda la distinzione tra stati terminali e non. A tal proposito è stata condotta un'analisi che ha permesso di identificare gli stati detti "pozzo" ovvero quelli per i quali, una volta entrati, non esiste un evento che vi conduca fuori. Per la ritualità considerata sono sotto elencati gli identificativi degli stati che presentano questa caratteristica:

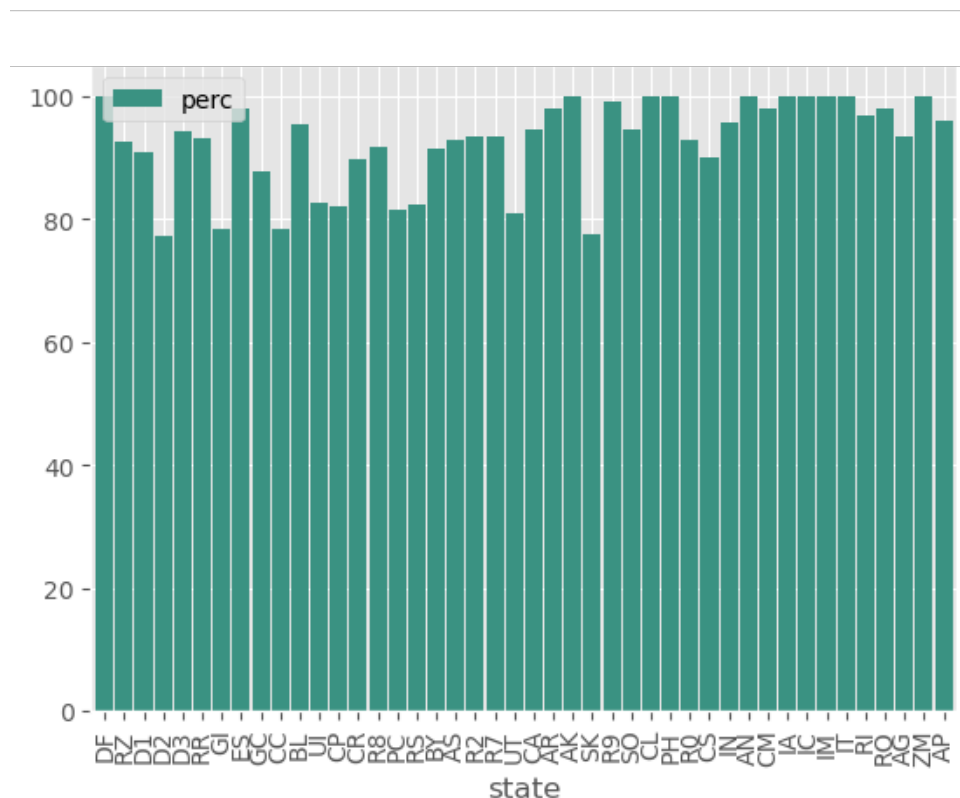


Figure 1: Rapporto autoanelli e normali transizioni in ritualità 40

DF	AK	CL	PH	AN	IA	IC	IM	IT	ZM
----	----	----	----	----	----	----	----	----	----

Table 7: Identificativi degli stati pozzo

Risulta a questo punto cruciale definire se gli eventi che conducono ai suddetti siano in qualche modo riconducibili ad eventi finali. Se ciò non dovesse rivelarsi necessariamente vero potrebbero sussistere casi in cui la caduta in uno stato pozzo nel mezzo dello svolgimento di un processo determini l'impossibilità di arrivare a una chiusura dello stesso data l'inesistenza di un evento che possa fare da esso uscire. Sono di seguito riportati in tabella gli eventi che conducono agli stati pozzo sopra identificati:

CCDOEV	CDESCR
DRIG	DECRETO DI RIGETTO
TE	INCOMPETENZA PER TERRITORIO
IS	INAMMISSIBILITA'
CO	CONCILIAZIONE
QA	ORDINANZA DEFINITIVA GENERICA
O2	SENTENZA A VERBALE
IM	IMPROCEDIBILITA'
IN	ORDINANZA DI INCOMPETENZA
9C	DEPOSITO SENTENZA - PUBBLICAZIONE (DISPOSITIVO LETTO IN UDIENZA)
AP	ANNULLAMENTO PROCEDIMENTO
7E	TRASMISSIONE ATTI/FASCICOLO AD ALTRO UFFICIO GIUDIZIARIO
2E	DEPOSITO SENTENZA - PUBBLICAZIONE

Table 8: Tabella dei codici eventi

Come si evince chiaramente dalla tabella riportata sopra tutti gli eventi sembrano essere quelli che determinano la chiusura di un processo e, a ragione, conducono pertanto in uno stato dal quale non è possibile uscire.

2.3 Data understanding: Ctipos e Riformata

2.3.1 Studio del significato del dominio di CTIPOS e RIFORMATA

Data la natura delle informazioni riportate al loro interno le tabelle STOR e DEFI si prestano ad essere utilizzate per una prima determinazione dell'esito dei processi da verificare o completare con l'analisi dei testi delle sentenze.

Un primo studio condotto in questa direzione all'interno del laboratorio ha evidenziato come dalle colonne CTIPOS e RIFORMATA delle suddette tabelle fosse possibile derivare l'esito dell'appello. Tuttavia al dominio dei valori assumibili dai suddetti attributi non è stato associato un significato chiaro ne tantomeno univoco cosa, quest'ultima, che ha reso impossibile condurre statistiche in materia.

Value	Meaning
A	Confermata
B	Riforma totale
C	da stabilire
D	da stabilire
E	Riforma parziale
Z	da stabilire (altro)

Table 9: Riformata legacy

Value	Meaning
A	da stabilire
B	da stabilire
C	da stabilire
D	da stabilire
E	da stabilire

Table 10: Ctipos legacy

Per far fronte alla necessità di ampliare il dizionario dati si è scelto nel corso del progetto di incrociare i dati di DEFI con quelli di STOR eseguendo un'operazione di join naturale sull'attributo NUMPRO in modo da associare l'attributo CDESCR in STOR ai valori di RIFORMATA e CTIPOS in DEFI.

Tra gli eventi analizzati si è scelto di prestare particolare attenzione a quelli che conducono al deposito della sentenza (2E) o alla sua pubblicazione a verbale (O2) in modo da derivare dalla descrizione di sentenze concluse un'indicazione sull'esito del processo.

Questa ricerca precedentemente condotta in modo manuale è stata automatizzata su tutta la tabella STOR con i risultati esposti nelle tabelle sottostanti:

Outcome	Count
rigetto	0
accoglimento	44
conferma	10219
riforma parziale	17
riforma totale	27
rinvio al primo grado	1
accoglimento parziale	0
accoglimento totale	0
n/a	0
altro	39

Table 11: RIFORMATA = A

Outcome	Count
rigetto	2
accoglimento	3174
conferma	29837
riforma parziale	13
riforma totale	9
rinvio al primo grado	0
accoglimento parziale	5
accoglimento totale	233
n/a	3
altro	2

Table 12: CTIPOS = A

Outcome	Count
rigetto	0
accoglimento	0
conferma	23
riforma parziale	51
riforma totale	2662
rinvio al primo grado	1
accoglimento parziale	3
accoglimento totale	0
n/a	10
altro	18

Table 13: RIFORMATA = B

Outcome	Count
rigetto	0
accoglimento	1
conferma	0
riforma parziale	8
riforma totale	7409
rinvio al primo grado	0
accoglimento parziale	313
accoglimento totale	0
n/a	422
altro	6

Table 14: CTIPOS = B

Outcome	Count
rigetto	72
accoglimento	4
conferma	13
rimborso parziale	4512
rimborso totale	45
rinvio al primo grado	2
accoglimento parziale	0
accoglimento totale	0
n/a	30
altro	13

Table 15: RIFORMATA = C

Outcome	Count
rigetto	1142
accoglimento	0
conferma	3
rimborso parziale	12557
rimborso totale	5
rinvio al primo grado	0
accoglimento parziale	0
accoglimento totale	0
n/a	407
altro	6

Table 16: CTIPOS = C

Outcome	Count
rigetto	0
accoglimento	0
conferma	0
rimborso parziale	0
rimborso totale	1
rinvio al primo grado	34
accoglimento parziale	0
accoglimento totale	0
n/a	0
altro	9

Table 17: RIFORMATA = D

Outcome	Count
rigetto	0
accoglimento	0
conferma	0
rimborso parziale	0
rimborso totale	0
rinvio al primo grado	181
accoglimento parziale	0
accoglimento totale	0
n/a	8
altro	1

Table 18: CTIPOS = D

Outcome	Count
rigetto	0
accoglimento	1
conferma	20
rimborso parziale	13
rimborso totale	19
rinvio al primo grado	13
accoglimento parziale	0
accoglimento totale	0
n/a	12
altro	2373

Table 19: RIFORMATA = E

Outcome	Count
rigetto	2
accoglimento	4
conferma	7
rimborso parziale	19
rimborso totale	16
rinvio al primo grado	0
accoglimento parziale	5
accoglimento totale	0
n/a	1841
altro	6632

Table 20: CTIPOS = E

Outcome	Count
rigetto	1
accoglimento	53
conferma	0
rimborso parziale	0
rimborso totale	0
rinvio al primo grado	0
accoglimento parziale	0
accoglimento totale	0
n/a	32
altro	0

Table 21: RIFORMATA = R

Outcome	Count
rigetto	4215
accoglimento	3
conferma	1
rimborso parziale	8
rimborso totale	3
rinvio al primo grado	0
accoglimento parziale	1
accoglimento totale	0
n/a	4
altro	1

Table 22: CTIPOS = R

Outcome	Count
rigetto	0
accoglimento	0
conferma	1
rimforma parziale	0
rimforma totale	1
rinvio al primo grado	0
accoglimento parziale	0
accoglimento totale	0
n/a	0
altro	0

Table 23: RIFORMATA = Z

Outcome	Count
rigetto	0
accoglimento	0
conferma	0
rimforma parziale	0
rimforma totale	0
rinvio al primo grado	0
accoglimento parziale	0
accoglimento totale	0
n/a	0
altro	0

Table 24: CTIPOS = Z

Pertanto il significato che con maggiore probabilità è attribuibile ai valori di Riformata e Ctipos è il seguente:

Value	Meaning
A	Conferma/Accoglimento
C	Riforma parziale
E	Altro
B	Riforma totale
R	Rigetto
D	Rinvio al primo grado
Z	(Non compare)
None	(Non compare)

Table 25: Associazione valore-significato

2.3.2 Quando Riformata è riempita Ctipos presenta lo stesso valore

Da un'osservazione delle Tabelle 20 e 21 risulta chiaro come le informazioni riportate in Riformata e in Ctipos risultino spesso fornire informazioni contraddittorie circa il significato dell'esito del processo. Inoltre osservando la distribuzione delle occorrenze riportate nella tabella sottostante notiamo come, a parità di valore, il numero di occorrenze di Riformata nella tabella Defi risulti essere significativamente inferiore a quelle di Ctipos. Inoltre la maggior parte delle volte in cui Riformata presenta un valore esso è null.

Value	Meaning
A	10352
C	5098
B	2778
E	2459
R	258
D	44
Z	2
None	56219

Table 26: Distribuzione Riformata in DEFI

Value	Meaning
A	36107
C	16353
E	10793
B	9111
R	4595
D	243
Z	4
None	4

Table 27: Distribuzione Ctipos in DEFI

Nella matrice riportata in basso il contenuto di ogni cella rappresenta quante volte un valore di RIFORMATA corrisponde a un valore di CTIPOS a parità di processo considerato (nelle colonne vediamo riportati i valori di RIFORMATA mentre sulle righe quelli di CTIPOS).

La diagonale principale (dove le categorie coincidono) è colorata di azzurro, indicando le corrispondenze dirette tra i valori.

Notiamo come i match prevalgano per i valori A, B, C, D, E mentre per i restanti le discordanze sono più marcate.

	A	B	C	D	E	R	Z	NONE
A	276481	744	522	33	1086	23	0	0
B	616	72939	1557	31	333	0	0	0
C	765	1078	138483	52	1610	4360	0	0
D	0	32	0	932	192	0	0	0
E	358	313	244	349	56901	0	0	0
R	2548	535	91	30	1030	147	0	0
Z	30	27	0	0	0	0	0	0
NONE	661117	168041	316038	5224	183362	92505	76	120

Table 28: Matrice dei match e mismatch tra Riformata e Ctipos

2.3.3 Comparazione CDESCR, RIFORMATA e CTIPOS

A conclusione dell'analisi nelle tabelle seguenti viene riportata, per ogni evento culminante con esito nello storico, la corrispondente configurazione di RIFORMATA e CTIPOS.

Infatti se nella precedente ricerca abbiamo fissato il valore di RIFORMATA e CTIPOS e dedotto il significato degli attributi da CDESCR ora al contrario viene fissato CDESCR e osservata la configurazione degli altri due attributi.

Troviamo riportate sulle colonne i valori assumibili da RIFORMATA mentre sulle righe i valori assumibili da CTIPOS. Lo scopo è quello di verificare che le ipotesi sul significato di questi attributi corrispondano alla descrizione che dell'evento è data.

	A	B	C	D	E	R	Z	NULL
A	47	0	0	0	0	0	0	3405
B	0	6	0	0	0	0	0	323
C	4	0	4	0	0	0	0	6
D	0	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	11
R	53	0	0	0	0	0	0	4
Z	0	0	0	0	0	0	0	0
NULL	0	0	0	0	0	0	0	0

Table 29: Accoglimento

	A	B	C	D	E	R	Z	NULL
A	1	0	17	0	0	0	0	19
B	0	4	50	0	0	0	0	9
C	0	0	4521	0	0	0	0	8627
D	0	0	0	0	0	0	0	0
E	0	0	9	0	6	0	0	21
R	0	0	0	0	0	0	0	8
Z	0	0	0	0	0	0	0	0
NULL	0	0	0	0	0	0	0	0

Table 32: Riforma parziale

	A	B	C	D	E	R	Z	NULL
A	1	0	0	0	0	0	0	4
B	0	1	0	0	0	0	0	4
C	0	0	5	0	0	72	0	1160
D	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	5
R	0	0	0	0	0	1	0	4203
Z	0	0	0	0	0	0	0	0
NULL	0	0	0	0	0	0	0	0

Table 30: Rigetto

	A	B	C	D	E	R	Z	NULL
A	1	27	0	0	0	0	0	9
B	0	2688	0	0	0	0	0	5145
C	0	42	6	0	0	0	0	5
D	0	1	0	0	0	0	0	0
E	0	14	1	0	7	0	0	14
R	0	0	0	0	0	0	0	3
Z	0	1	0	0	0	0	0	0
NULL	0	0	0	0	0	0	0	0

Table 33: Riforma totale

	A	B	C	D	E	R	Z	NULL
A	10228	0	0	0	0	0	0	21883
B	25	0	0	0	0	0	0	2
C	14	0	0	0	0	0	0	7
D	0	0	0	0	0	0	0	0
E	18	0	0	0	3	0	0	6
R	0	0	0	0	0	0	0	1
Z	1	0	0	0	0	0	0	0
NULL	0	0	0	0	0	0	0	0

Table 31: Conferma

	A	B	C	D	E	R	Z	NULL
A	0	0	0	1	0	0	0	0
B	0	0	0	1	0	0	0	0
C	0	0	0	2	0	0	0	0
D	0	0	0	34	0	0	0	155
E	0	0	0	14	0	0	0	0
R	0	0	0	0	0	0	0	0
Z	0	0	0	0	0	0	0	0
NULL	0	0	0	0	0	0	0	0

Table 34: Rinvio al primo grado

	A	B	C	D	E	R	Z	NULL
A	1	0	0	0	41	0	0	7
B	0	6	1	0	16	0	0	15
C	0	0	7	0	13	0	0	16
D	0	0	0	0	9	0	0	1
E	0	0	0	0	2398	0	0	5781
R	0	0	0	0	0	0	0	1
Z	0	0	0	0	0	0	0	0
NULL	0	0	0	0	0	0	0	0

Table 35: Altro

In generale possiamo notare come la somma dei valori interni alla tabella rappresenti il numero di volte che un esito del tipo specificato si verifica nella tabella stor. In particolare, per la prima tabella possiamo ad esempio notare come nella maggioranza dei casi un evento di accoglimento presenti una configurazione del tipo RIFORMATA = null e CTIPOS = "A". Questo risultato sembra confermare le ipotesi precedentemente formulate sul significato del valore A per l'attributo CTIPOS. Considerazioni analoghe valgono per i risultati emersi dalle altre tabelle.

2.4 Data understanding: Idgrpev

L'attributo IDGRPEV, presente nella tabella StatiEvento, consente di raggruppare gli eventi in categorie. Esso assume 58 valori distinti i quali, tuttavia, non erano ancora stati organizzati in un dizionario dati. La stessa tipologia di indagine condotta in precedenza per i valori di Riformata e Ctipos è stata utilizzata per ricavare il significato dei dieci valori di IDGRPEV che con maggior frequenza appaiono nello storico.

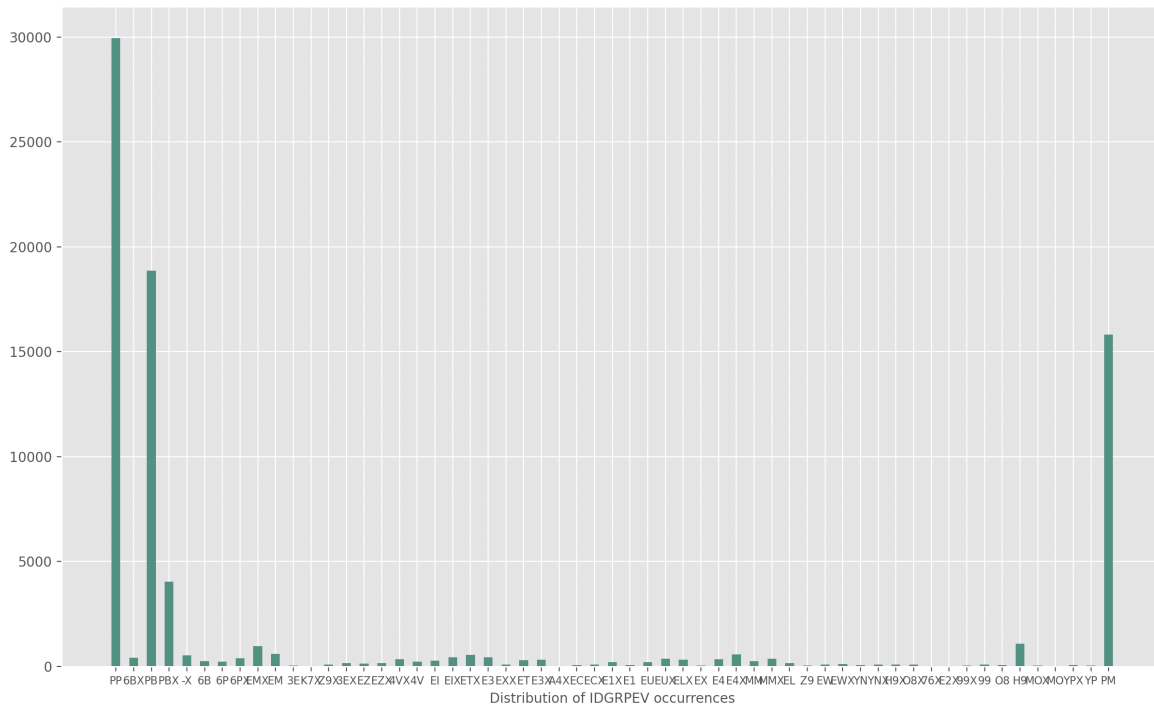


Figure 2: Occorrenze dei valori di IDGRPEV nella tabella Stor

In particolare dal momento che più eventi possono appartenere alla stessa categoria, sono stati selezionati gli eventi associati a uno stesso IDGRPEV ed è stato contato il loro numero di occorrenze tra i log della tabella Stor.

In questo modo è stato possibile collegare all'etichetta IDGRPEV un significato dato dalla descrizione dell'evento ad essa associato che con maggior frequenza si ripresenta tra le entries del sistema.

2.4.1 Descrizione valori

Vengono di seguito riportati per ognuno dei 10 valori significativi di IDGRPEV gli eventi a esso maggiormente associati e il conto delle loro occorrenze tra i log:

CCDOEV	Count	Descrizione
DQ	121480	DEPOSITO COMPARSE CONCLUSIONALI
IA	113373	ISCRIZIONE RUOLO GENERALE
J1	101692	DEPOSITO ATTO NON CODIFICATO
DD	89202	DEPOSITO MEMORIE DI REPLICA
NH	50520	DEPOSITO NOTA SPESE
XV	17633	ACCOGLIMENTO RICHIESTA DI VISIBILITA'
1H	13255	ATTO NON CODIFICATO
NI	11323	DEPOSITO FASCICOLO DI PARTE
DN	7259	DEPOSITO MEMORIE
I7	6843	PRECISAZIONE DELLE CONCLUSIONI

Table 36: Descrizione associata a PP

CCDOEV	Count	Descrizione
DM	58351	IN DECISIONE
YB	48592	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
MI	44363	RINVIO AD ALTRA UDIENZA
RS	31521	RISERVA
SW	20690	SOSTITUZIONE GIUDICE PER SURROGA
RB	20288	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
R3	19425	RINVIO D'UFFICIO
1H	13255	ATTO NON CODIFICATO
KD	11266	DICHIARAZIONE DI CONTUMACIA
DU	9968	DIFFERIMENTO UDIENZA

Table 37: Descrizione associata a PB

CCDOEV	Count	Descrizione
NX	207365	ANNOTAZIONE
CEPM	6030	EMESSO PARERE DEL PM/PG
CVPM	2342	VISTO DEL PM/PG
SPPM	1437	SOSTITUZIONE PM TITOLARE
SPMA	598	SOSTITUZIONE PM ATTIVITA
CPM	593	CONCLUSIONI DEL PM
MEPM	87	DEPOSITO MEMORIA DEL PM/PG
CAPM	17	RICHIESTA APERTURA VISIBILITA' PM
MSPM	4	MODIFICA SEZIONE E PM TITOLARE
IPM	4	INTERVENTO DEL PM

Table 38: Descrizione associata a PM

CCDOEV	Count	Descrizione
DM	58351	IN DECISIONE
YB	48592	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
MI	44363	RINVIO AD ALTRA UDIENZA
RS	31521	RISERVA
RB	20288	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
R3	19425	RINVIO D'UFFICIO
DU	9968	DIFFERIMENTO UDIENZA
MG	9867	RINVIO MANCATA COMPARIZIONE PARTI (art.309 cpc)
6O	8929	FISSAZIONE UDIENZA CAMERALE
PA	8503	ANTICIPAZIONE UDIENZA

Table 39: Descrizione associata a PBX

CCDOEV	Count	Descrizione
DM	58351	IN DECISIONE
YB	48592	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
MI	44363	RINVIO AD ALTRA UDIENZA
RS	31521	RISERVA
MG	9867	RINVIO MANCATA COMPARIZIONE PARTI (art.309 cpc)
X6	8029	PRECISAZIONE CONCLUSIONI IN UDIENZA E CAUSA IN...
7P	5883	IN DECISIONE CON RINUNCIA TERMINI PER CONCLUSI...
CA	5706	CANCELLAZIONE
O2	4990	SENTENZA A VERBALE
LD	3399	LETTURA DISPOSITIVO

Table 40: Descrizione associata a YNX

CCDOEV	Count	Descrizione
DM	58351	IN DECISIONE
YB	48592	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
MI	44363	RINVIO AD ALTRA UDIENZA
RS	31521	RISERVA
RB	20288	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
MG	9867	RINVIO MANCATA COMPARIZIONE PARTI (art.309 cpc)
X6	8029	PRECISAZIONE CONCLUSIONI IN UDIENZA E CAUSA IN...
7P	5883	IN DECISIONE CON RINUNCIA TERMINI PER CONCLUSI...
CA	5706	CANCELLAZIONE
O2	4990	SENTENZA A VERBALE

Table 41: Descrizione associata a MMX

CCDOEV	Count	Descrizione
DM	58351	IN DECISIONE
MI	44363	RINVIO AD ALTRA UDIENZA
RS	31521	RISERVA
RB	20288	RINVIO ALL'UDIENZA DI PRECISAZIONE CONCLUSIONI...
MG	9867	RINVIO MANCATA COMPARIZIONE PARTI (art.309 cpc)
X6	8029	PRECISAZIONE CONCLUSIONI IN UDIENZA E CAUSA IN...
7P	5883	IN DECISIONE CON RINUNCIA TERMINI PER CONCLUSI...
CA	5706	CANCELLAZIONE
O2	4990	SENTENZA A VERBALE
RY	3542	RINVIO ALL'UDIENZA DI TRATTAZIONE (art. 183 cpc)

Table 42: Descrizione associata a E4X

2.5 Analisi eventi di correzione

Un fronte di analisi di notevole rilevanza è quello riguardante gli eventi di correzione. In particolare, data la natura stessa dell'evento, risulta cruciale capire se in seguito al suo verificarsi venga generato un autoanello o una semplice transizione verso uno stato diverso da quello di partenza.

In seguito alle analisi condotte possiamo notare come:

1. Il numero totale di eventi di correzione è 110947 (4.25 % del totale).
2. Il numero di eventi di correzione che determinano una transizione verso uno stato diverso da quello di partenza è: 14651.
3. Tra questi 10145 sono generati dall'evento con etichetta "C9". La descrizione di tale evento riporta: *Stato corretto da %s a %s*.
4. Tra i 4895 eventi di correzione "anomali" che non sono caratterizzati dall'etichetta "C9", 3430 presentano come descrizione: *Correzione evento materiale* e 1461 *Correzione ritualità*.

La condizione anomala rappresentata dalla presenza di eventi di correzione che non generano autoanelli è spiegata dal fatto che tali eventi si riferiscono esplicitamente alla correzione dello stato di un processo (e, di conseguenza, a un suo cambiamento) oppure a errori per i quali una transizione di stato è plausibile (cambio ritualità ed errore materiale).

2.6 Studio dei cambi di ritualità

La figura sottostante ha lo scopo di illustrare la frequenza con cui avvengono i cambi di ritualità evidenziando il rito di partenza e quello di arrivo:

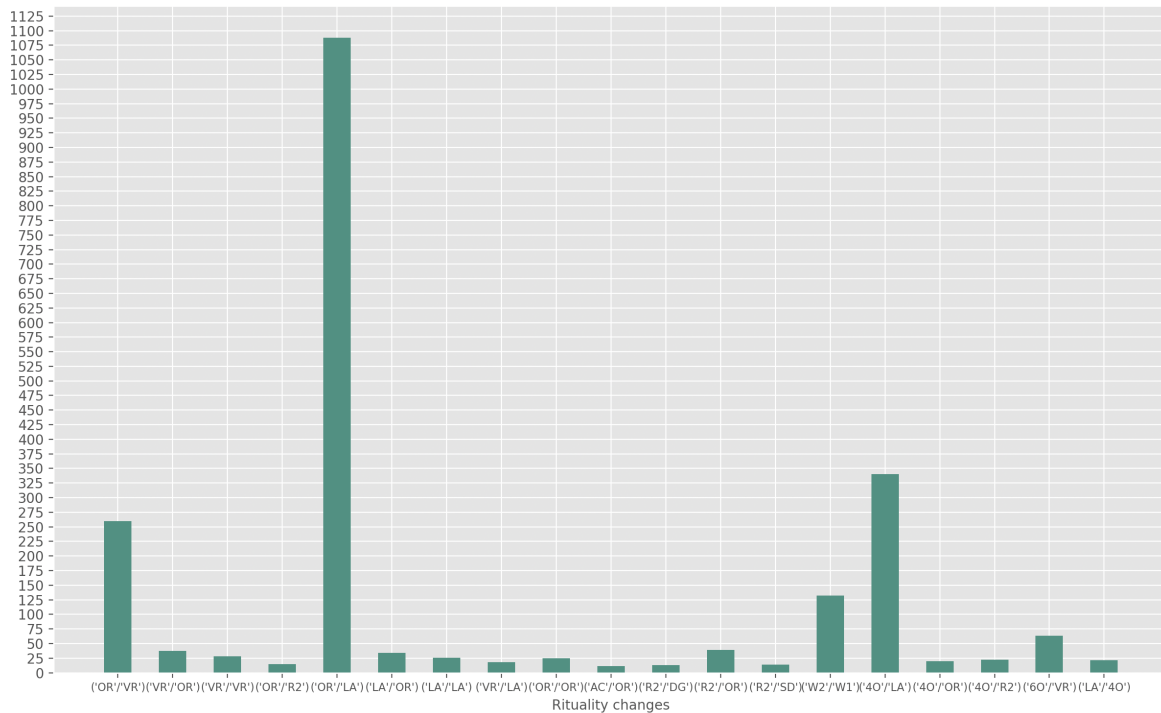


Figure 3: Frequenza dei cambi di ritualità

Come possiamo notare i cambi di ritualità che avvengono più spesso partono dal rito ordinario o da 40 per passare a "lavoro" oppure dal rito ordinario per passare a "VR".

2.7 Analisi dei piani temporali

Una delle analisi di consistenza dei dati che ha condotto a risultati di maggiore rilevanza è stata quella relativa all'allineamento temporale degli eventi all'interno del sistema SICID.

Esso memorizza infatti per ogni evento il timestamp del suo avvenimento, della sua registrazione all'interno della base di dati e della sua ultima modifica.

Nel corso dello studio sono state sviluppate funzioni che permettano di identificare eventi in cui la registrazione precede l'accadimento o l'ultima modifica precede l'avvenimento o la data di registrazione.

Per questi è necessario risalire al processo in cui si sono verificati e, di conseguenza, al fascicolo da cui provengono.

Sono stati riscontrati 57 log che non rispettano i vincoli temporali imposti. Nonostante questo numero, se messo in relazione con il totale delle righe nella tabella stor, non sia grande abbastanza da risultare significativo è comunque sintomo della presenza di ulteriori errori del sistema nella registrazione e modifica degli eventi.

Sono di seguito riportati gli eventi che generano questo tipo di inconsistenza:

numprv	numpro	descrizione
13906	1472	The occurrence of event follows its registration
21294	3282	The occurrence of event follows its registration
21294	3282	The occurrence of event follows its modification
46345	1671	The occurrence of event follows its registration
46345	1671	The occurrence of event follows its modification
46348	1671	The occurrence of event follows its registration
57250	11384	The occurrence of event follows its registration
62872	560	The occurrence of event follows its registration
62872	560	The occurrence of event follows its modification
144203	17367	The occurrence of event follows its registration
144203	17367	The occurrence of event follows its modification
144204	17367	The occurrence of event follows its registration
144204	17367	The occurrence of event follows its modification
144205	17367	The occurrence of event follows its registration
144205	17367	The occurrence of event follows its modification
564333	37232	The occurrence of event follows its registration
564333	37232	The occurrence of event follows its modification
564365	37215	The occurrence of event follows its registration
564365	37215	The occurrence of event follows its modification
564396	37438	The occurrence of event follows its registration
564396	37438	The occurrence of event follows its modification
590601	37355	The occurrence of event follows its registration
651223	21579	The occurrence of event follows its registration
660462	30588	The occurrence of event follows its registration
660462	30588	The occurrence of event follows its modification
668539	38208	The occurrence of event follows its registration
1156080	68383	The occurrence of event follows its registration
1156080	68383	The occurrence of event follows its modification
1173392	50130	The occurrence of event follows its registration
1278207	73547	The occurrence of event follows its registration
1328313	76514	The occurrence of event follows its registration
1557303	87790	The occurrence of event follows its registration
1557303	87790	The occurrence of event follows its modification
1622296	81371	The occurrence of event follows its registration
1622296	81371	The occurrence of event follows its modification
1902287	103480	The occurrence of event follows its registration
2006216	91391	The registration of event follows its modification
2096478	104759	The occurrence of event follows its registration
2096481	105008	The occurrence of event follows its registration
2136547	102457	The occurrence of event follows its registration
2143172	112632	The occurrence of event follows its registration
2143172	112632	The occurrence of event follows its modification
2306871	113711	The occurrence of event follows its registration
2306871	113711	The occurrence of event follows its modification

Table 43: Eventi anomali riscontrati nel sistema

3 Conclusione

3.1 Risultati ottenuti

I dati presenti nella base di dati sono stati analizzati con un grado di profondità sufficiente a considerarli complessivamente consistenti e di buona qualità.

L'analisi riguardante il significato degli attributi ha contribuito ad ampliare i dizionari già esistenti arricchendo i dati di nuove interpretazioni di cui fruire per ricerche successive. Allo stesso tempo la loro categorizzazione ha aperto alla possibilità di condurre statistiche sugli stessi.

3.2 Future analisi e approfondimenti

In sviluppi futuri del processo ci si propone di osservare con la nuova consapevolezza acquisita l'effetto dell'eliminazione di una parte dei log dal training set del cruscotto predittivo e verificarne le prestazioni a valle.

References

- [1] Barbara Pernici, Carlo Alberto Bono, Ludovica Piro and Mattia Del Treste, Giancarlo Vecchi, 2023. Improving the analysis of the judiciary performance - the use of data mining techniques to assess the timeliness of civil trials. *<https://www.emerald.com/insight/0951-3558.htm>*.
- [2] Barbara Pernici, Marco Dilettis, 2023. Valorizzare i dati degli uffici giudiziari: verso un cruscotto previsionale basato su process mining.