POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Matematica

Politecnico di Torino

Tesi di Laurea Magistrale

AI Maturity Assessment nel settore pubblico: un approccio data-driven

Dal Framework Nolan Norton all'analisi multivariata verso la definizione della strategia IA di un apparato di governo territoriale

Relatore:
Prof. Danilo Bazzanella

Candidata:
M. Virginia Zura-Puntaroni

In collaborazione con:
Nolan Norton Italia

Andrea Carbone
Maria Giovanna Ienco

Anno Accademico 2025-2026

Abstract

L'adozione dell'intelligenza artificiale (IA) nel settore pubblico richiede un complesso bilanciamento tra innovazione tecnologica e conformità normativa, in particolare alla luce dell'AI Act europeo e del quadro legislativo italiano. Questo elaborato propone un approccio analitico e *data-driven* per la misurazione della maturità IA di un apparato di governo territoriale. L'obiettivo principale è definire l'attuale profilo di maturità degli Enti coinvolti, applicando il *framework* strutturato di Nolan Norton e traducendo le valutazioni qualitative in un modello matematico rigoroso. A tale scopo, il nucleo metodologico sfrutta l'Analisi delle Componenti Principali (PCA) per ridurre la dimensionalità dei dati raccolti ed estrarne i fattori latenti dominanti. Infine, tramite tecniche di clustering applicate a questo spazio ridotto, gli Enti sono stati raggruppati in classi dal comportamento omogeneo, fornendo una mappatura utile a confrontare la maturità attuale con il target strategico prefissato.

Parole chiave: AI Maturity Model, Nolan Norton, Pubblica Amministrazione, Analisi Data-Driven, PCA, clustering.

Indice

Capitolo 1

Introduzione

1.1 Contesto e motivazioni

L'Italia, l'Europa e l'intera comunità globale si trovano ad affrontare una sfida complessa e di vasta portata: gestire e controllare l'intelligenza artificiale (di seguito, IA).

L'IA è uno degli strumenti tecnologici più potenti e trasformativi del nostro tempo e la sua adozione sta rivoluzionando il rapporto con i processi, le attività di verifica e la ricerca in ogni settore. La sua capacità di generare valore aggiunto, sia nell'ambito privato che in quello pubblico, alimenta grandi aspettative per il futuro prossimo.

Parallelamente a questa evoluzione tecnologica, anche il mondo del lavoro e la natura stessa dei suoi compiti stanno mutando: se i processi decisionali vengono trasferiti anche solo in parte da una mente umana a un sistema artificiale, emerge con forza la questione etica. L'adozione acritica o non regolamentata dell'IA, infatti, porta con sé rischi significativi, tra cui l'amplificazione di bias cognitivi e discriminazioni algoritmiche, l'opacità dei sistemi decisionali e nuove minacce alla sicurezza delle informazioni. È pertanto fondamentale allineare norme, obiettivi strategici e sviluppi pratici in ambito IA.

Ciò si traduce in un'attenta analisi e in un adeguamento delle normative vigenti (specialmente in materia di privacy e sicurezza dei dati), in un investimento continuo in ricerca e sperimentazione, e nell'esplorazione di un terreno comune che faciliti la collaborazione internazionale e gli sviluppi futuri.

In questo scenario dinamico e fortemente innovativo, una delle priorità per le organizzazioni è mantenersi aggiornate non solo sulle regolamentazioni nazionali e sovranazionali, ma anche sui contributi pratici e le *best*

practice sviluppate a livello globale. Inoltre, diventa essenziale individuare strumenti metodologici che consentano alle organizzazioni di misurare in modo strutturato il proprio grado di preparazione e maturità nell'adozione dell'IA, mappando i punti di forza ed evidenziando le lacune da colmare.

1.2 Obiettivi della tesi

Con il presente elaborato si vogliono perseguire molteplici obiettivi.

In primo luogo, si intende fornire una disamina del quadro normativo italiano ed europeo in materia IA, con particolare riferimento alle iniziative rivolte alla Pubblica Amministrazione. Tale studio preliminare costituisce il punto di partenza per declinare i concetti teorici nello specifico contesto territoriale analizzato. Quest'ultimo si configura come un vero e proprio ecosistema istituzionale, caratterizzato da una struttura articolata e dotato della lungimiranza necessaria per affrontare la sfida dell'innovazione tecnologica. Ci si riferirà a tale apparato utilizzando anche i termini "Soggetto", "Organizzazione" o "Territorio".

In secondo luogo, il presente lavoro di tesi si pone l'obiettivo di approfondire e applicare sul campo il *framework* metodologico proprietario sviluppato da Nolan Norton Italia (2025). L'utilizzo e lo studio di questo avanzato strumento per l'analisi e la misurazione della maturità dell'intelligenza artificiale sono stati resi possibili grazie alla stretta collaborazione instaurata con l'azienda.

Nolan Norton Italia, società di consulenza strategica in ambito Information and Communication Technology (ICT) appartenente al network KPMG, rappresenta un punto di riferimento per le organizzazioni che intendono valorizzare strategicamente l'innovazione tecnologica e guidare processi di trasformazione digitale. La società trae origine dalla Nolan Norton & Company, fondata negli Stati Uniti negli anni Settanta da Richard L. Nolan e David C. Norton, studiosi e consulenti di rilievo nel campo della gestione dei sistemi informativi e della strategia aziendale. Il contributo dei due autori è stato particolarmente significativo nello sviluppo di modelli teorici per l'evoluzione dell'ICT nelle organizzazioni; tra questi, uno dei più noti è il concetto delle "curve di discontinuità tecnologica", utilizzato per descrivere come le innovazioni digitali si sviluppino secondo cicli evolutivi caratterizzati da fasi di crescita progressiva seguite da momenti di rottura tecnologica che richiedono cambiamenti radicali nei modelli organizzativi e nelle strategie aziendali.

Nel corso degli anni, tali contributi teorici hanno influenzato profondamente il modo in cui le organizzazioni affrontano i processi di

innovazione e trasformazione digitale. In questo contesto, Nolan Norton Italia opera come centro di competenza specializzato nella definizione di strategie ICT e nell'accompagnamento di imprese e istituzioni nei percorsi di evoluzione tecnologica, supportandole nell'integrazione strutturata delle tecnologie digitali nei processi di business e nel rafforzamento del loro posizionamento competitivo (Nolan Norton Italia, 2024).

Nel seguente elaborato, il modello viene impiegato per valutare la maturità IA attuale (definita anche maturità As-Is) dell'apparato di governo territoriale. Tale misurazione serve a fotografare lo stato odierno relativo all'adozione, all'integrazione e all'utilizzo dell'IA a supporto dei processi e dei servizi erogati, ed è condotta tramite un questionario di *self-assessment* somministrato in modo capillare ai rappresentanti del Territorio. Il livello di maturità così ottenuto viene successivamente confrontato con un *benchmark* di riferimento, ovvero la Maturità Target, che rappresenta le ambizioni e gli obiettivi strategici da raggiungere.

Sebbene, per esigenze di sintesi, il profilo target non sia oggetto di trattazione all'interno dell'elaborato, esso funge da fondamentale termine di paragone. Il confronto tra questi due piani è infatti essenziale per comprendere il posizionamento strategico dell'Organizzazione, identificare il divario esistente tra obiettivi e realtà operativa, e formulare una *roadmap* per il raggiungimento dei traguardi prefissati.

Considerando la complessa architettura del territorio in esame, l'analisi non si limita a restituire una visione aggregata dei risultati, ma declina la valutazione della maturità IA scendendo nel dettaglio di ogni singolo Ente, agenzia e organismo che lo compone.

1.3 Struttura della tesi

Il presente elaborato è strutturato in sei capitoli, organizzati come segue:

- **Capitolo 2 – Quadro normativo in Italia e in Europa.** Analizza il contesto legale di riferimento, esaminando il *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce norme armonizzate sull'intelligenza artificiale (Artificial Intelligence Act)* e le normative nazionali per comprendere il perimetro d'azione e gli impatti dell'adozione dell'IA nella Pubblica Amministrazione.

- **Capitolo 3 – Strumento e metodologia di indagine.** Presenta il *framework* metodologico utilizzato per la valutazione della maturità IA, descrivendone le dimensioni di analisi e i livelli. Una sezione specifica

illustra come lo strumento sia stato declinato per la progettazione del questionario somministrato ai rappresentanti dell'apparato territoriale.

- **Capitolo 4 – Campione analizzato e creazione del dataset pulito.** Descrive l'intero processo di acquisizione e trattamento dei dati derivanti dal *self-assessment*. Include le fasi tecniche di *data engineering*, codifica delle variabili e gestione della privacy, evidenziando i limiti metodologici.

- **Capitolo 5 – Modello matematico e architettura dell'analisi.** Approfondisce gli aspetti quantitativi della tesi, formalizzando lo spazio vettoriale delle risposte e applicando metodi di riduzione della dimensionalità come l'Analisi delle Componenti Principali e tecniche di clustering per interpretare la struttura latente dei dati.

- **Capitolo 6 – Analisi dei risultati.** Discute gli esiti dell'elaborazione matematica, con il confronto tra la Maturità As-Is e la Maturità Target territoriale, fornendo una clusterizzazione dei rappresentanti del Soggetto.

Capitolo 2

Quadro normativo in Italia e in Europa

L'evoluzione tecnologica dell'intelligenza artificiale ha imposto alle istituzioni europee e nazionali la necessità di definire un perimetro legale chiaro, volto a bilanciare le opportunità di innovazione con la tutela dei diritti fondamentali. Questo capitolo analizza il complesso ecosistema normativo che disciplina l'adozione dell'IA, partendo dal *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce norme armonizzate sull'intelligenza artificiale (Artificial Intelligence Act)* fino alle iniziative legislative italiane, con particolare riguardo all'impatto sulla Pubblica Amministrazione (PA).

2.1 L'AI Act: obiettivi, struttura e implicazioni

L'AI Act rappresenta la prima normativa organica al mondo volta a disciplinare lo sviluppo e l'utilizzo dei sistemi di intelligenza artificiale. Pubblicato nella Gazzetta Ufficiale dell'Unione Europea il 12 luglio 2024, ha l'obiettivo di garantire che i sistemi di IA immessi sul mercato europeo siano sicuri, trasparenti, tracciabili e non discriminatori.

L'impianto normativo dell'AI Act si fonda su un approccio basato sul rischio (*risk-based approach*), classificando i sistemi di IA in quattro livelli, a ciascuno dei quali sono associati obblighi specifici:

- **Rischio inaccettabile.** Include tutte le pratiche considerate una minaccia evidente per la sicurezza e i diritti fondamentali delle persone. L'impiego di tali sistemi è severamente proibito. Rientrano in questa

categoria le tecniche di manipolazione cognitivo-comportamentale, l'identificazione biometrica remota in tempo reale in spazi pubblici (salvo eccezioni strettamente regolamentate per finalità di pubblica sicurezza) e i sistemi di *social scoring*. Quest'ultimo consiste nella valutazione e classificazione delle persone fisiche, da parte di soggetti pubblici o privati, in base al loro comportamento sociale o a tratti della personalità, tale da generare un trattamento ingiustificato, sproporzionato o discriminatorio.

- **Alto rischio.** Riguarda le applicazioni IA impiegate in settori critici quali le infrastrutture, l'istruzione, l'occupazione, i servizi pubblici essenziali e l'applicazione della legge. Per questi sistemi sono previsti obblighi severi di *compliance*, tra i quali la valutazione d'impatto sui diritti fondamentali, la garanzia di elevati standard di qualità per i dati di addestramento, la redazione di una documentazione tecnica dettagliata e l'implementazione di meccanismi di supervisione umana (*human-in-the-loop*).

- **Rischio limitato.** Include i sistemi soggetti a specifici vincoli di trasparenza, come i *chatbot* o i modelli generativi di contenuti, come i *deepfake*, ovvero immagini, audio o video manipolati artificialmente per risultare altamente realistici. In questi casi, il requisito normativo centrale impone che gli utenti siano resi pienamente consapevoli di interagire con un'entità artificiale o di fruire di contenuti sinteticamente manipolati.

- **Rischio minimo.** Comprende la maggior parte dei sistemi di IA attualmente in uso, come ad esempio i filtri antispam, per i quali non sono previsti obblighi aggiuntivi oltre a quelli contenuti nel regolamento generale sulla protezione dei dati (Regolamento UE 2016/679, anche noto come GDPR, pubblicato sulla Gazzetta Ufficiale dell'Unione Europea il 4 maggio 2016 dal Parlamento Europeo e Consiglio dell'Unione Europea, 2016).

L'AI Act introduce inoltre norme specifiche per i modelli di IA con finalità generali (*General Purpose AI*), imponendo ai fornitori di tali modelli obblighi di trasparenza sulla documentazione tecnica e sul rispetto delle norme sul diritto d'autore.

2.2 La normativa italiana sull'intelligenza artificiale

In parallelo al percorso normativo europeo, l'Italia ha introdotto una disciplina nazionale in materia di intelligenza artificiale con la *Legge 23 settembre 2025, n. 132. Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, pubblicata nella Gazzetta Ufficiale della Repubblica Italiana. Il provvedimento definisce i principi generali relativi alla ricerca, allo sviluppo, alla sperimentazione e all'applicazione dei sistemi di IA, promuovendo un utilizzo corretto, trasparente e responsabile di tali tecnologie. La legge mira inoltre a garantire un equilibrio tra l'innovazione tecnologica e la tutela dei diritti fondamentali, prevedendo strumenti di *governance* e deleghe al Governo per l'adozione di misure attuative nei diversi settori di applicazione dell'intelligenza artificiale.

I punti cardine del provvedimento italiano possono essere sintetizzati come segue:

- **Governance e Autorità competenti:** la Legge italiana n.132/2025 individua nell'Agenzia per l'Italia Digitale (AgID) e nell'Agenzia per la Cybersicurezza Nazionale (ACN) le autorità nazionali responsabili, rispettivamente, della promozione dell'innovazione e della vigilanza sui sistemi di IA, in raccordo con il Garante per la protezione dei dati personali.

- **Settori strategici:** il testo pone l'accento sull'utilizzo dell'IA in ambito sanitario, lavorativo e giudiziario, stabilendo che l'uso di algoritmi non deve mai pregiudicare il ruolo decisionale umano, specialmente in contesti sensibili come la diagnosi medica o le decisioni giudiziarie.

- **Investimenti e Sanzioni:** vengono previsti fondi per supportare le *start-up* e le piccole e medie imprese che promuovono l'innovazione nel settore dell'IA. Inoltre, viene recepito il sistema sanzionatorio europeo, adattandolo al contesto giuridico nazionale e introducendo sanzioni penali per usi illeciti volti a manipolare o danneggiare terzi.

2.3 Impatti normativi sull'adozione dell'IA nella Pubblica Amministrazione

Nel contesto italiano, un primo riferimento operativo è rappresentato dalle *Linee guida per l'adozione dell'intelligenza artificiale nella Pubblica Amministrazione* elaborate dall'Agenzia per l'Italia Digitale, 2025. Tali linee guida forniscono indicazioni metodologiche e organizzative per supportare le amministrazioni nell'introduzione responsabile di sistemi di IA, con particolare attenzione agli aspetti di *governance*, gestione dei dati, trasparenza e valutazione dei rischi.

L'intersezione tra l'AI Act europeo e la normativa nazionale ha effetti profondi sulla Pubblica Amministrazione, chiamata non solo a rispettare le nuove regole, ma a fare da traino per l'innovazione responsabile.

Per la PA l'adozione di soluzioni di IA comporta una serie di adempimenti preliminari imprescindibili:

1. **Valutazione d'impatto sui diritti fondamentali.** Prima di implementare un sistema di IA classificato come "ad alto rischio", l'organizzazione è tenuta a condurre una valutazione specifica per mitigare i rischi di discriminazione e i bias. È il caso di algoritmi per la selezione del personale o per l'assegnazione di benefici sociali.

2. **Trasparenza e Registro pubblico.** Le PA devono rendere noto ai cittadini l'utilizzo di sistemi di IA nei procedimenti amministrativi. Questo obbligo si ricollega al principio di spiegabilità, secondo cui le decisioni automatizzate o semi-automatizzate devono essere comprensibili e motivabili agli occhi dell'interessato.

3. **Qualità e *Governance* dei Dati.** L'efficacia e la liceità dell'IA dipendono dai dati di addestramento, quindi la PA deve garantire che i dataset utilizzati siano rappresentativi e privi di errori, in stretta conformità con il GDPR. La gestione del dato diventa quindi un prerequisito abilitante per qualsiasi progetto di IA.

4. **Formazione e Competenze.** L'AI Act introduce l'obbligo per gli operatori di possedere un adeguato livello di alfabetizzazione in materia di IA. Ciò implica per le organizzazioni non solo investimenti in tecnologia, ma anche nella formazione continua del personale, affinché i dipendenti siano in grado di supervisionare i sistemi e interpretarne correttamente gli output.

In conclusione, il quadro normativo non va inteso come un freno burocratico, bensì come una guida per un'adozione consapevole dell'IA. In questo scenario emerge la necessità di uno strumento di *assessment* basato su molteplici standard normativi, come quello esaminato. Tale strumento risulta fondamentale per valutare la capacità dell'Organizzazione di rispettare i requisiti richiesti e per pianificare le azioni necessarie ad adeguarsi alla legislazione vigente.

Capitolo 3

Strumento e metodologia di indagine

3.1 Il *Framework* proprietario di Nolan Norton

Per valutare il proprio livello di maturità IA e tracciarne i progressi nel tempo, il Soggetto ha adottato un *framework* metodologico proprietario di Nolan Norton. Questo strumento permette di misurare in modo oggettivo la capacità dell'apparato di governo territoriale di introdurre e mettere a sistema soluzioni di Intelligenza Artificiale, partendo dal presupposto che la consapevolezza del proprio livello iniziale sia fondamentale per raggiungere gli obiettivi prefissati.

Il modello poggia su un solido impianto che integra certificazioni internazionali (come la ISO/IEC 42001) con i principali standard europei e nazionali: il Regolamento Europeo sull'Intelligenza Artificiale (AI Act 2024/1689), la Legge italiana 23 settembre 2025, n. 132, e la Bozza di linee guida AgID per l'adozione dell'IA nella Pubblica Amministrazione. A questi si affiancano le *best practice* di settore dettate dalla *Data Management Association* (DAMA) e dall'*Information Technology Infrastructure Library* (ITIL).

Concepito come uno strumento scalabile, dinamico e riutilizzabile, il *framework* è messo a disposizione di tutti gli istituti, società, organizzazioni, agenzie e amministrazioni che compongono il Territorio. Il suo obiettivo è misurare la Maturità As-Is dell'Organizzazione, orientando l'impiego dell'IA verso soluzioni utili e sostenibili e rispondendo a esigenze concrete e ambizioni strategiche. La sua natura intrinsecamente dinamica e adattabile è, inoltre,

indispensabile per non risultare obsoleti di fronte a tecnologie, normative e modelli organizzativi in rapida e continua trasformazione.

Nella pratica, il *framework* si basa su un *assessment* mirato e dettagliato, con quesiti sviluppati a partire dagli standard e dalle certificazioni di settore nominati precedentemente, e consente di mappare l'intero ecosistema IA e comprenderne tutta la complessità. I quesiti sono organizzati su due livelli strettamente interconnessi, tramite cui vengono analizzati tutti i possibili aspetti inerenti all'IA.

Sul livello superiore si collocano sette Dimensioni IA: Strategia, Infrastruttura, Dati, Processi, Organizzazione, Persone e competenze, Privacy e compliance. Queste costituiscono le macro-aree di indagine e sono i veri abilitatori dell'intero sistema, e saranno oggetto di un approfondimento dedicato nelle sezioni successive dell'elaborato.

Ad un livello più granulare, invece, l'ecosistema è articolato in circa 40 topic specifici, strutturati per affrontare in maniera olistica tutte le tematiche afferenti all'IA. Per garantire una valutazione aderente alle realtà territoriali del Soggetto, a ciascun topic vengono assegnati pesi differenziati, calibrati in base alla loro rilevanza strategica.

Questa natura trasversale, basata sulla doppia categorizzazione in topic e dimensioni, rappresenta una caratteristica distintiva del modello: i singoli topic, infatti, possono estendersi e influenzare più dimensioni contemporaneamente. Ogni quesito è quindi costruito per indagare in modo mirato l'intersezione tra questi due elementi, descrivendo al meglio la specifica coppia topic-dimensione.

Grazie alla struttura del questionario — articolato per dimensioni IA e caratterizzato da cinque livelli di risposta per ciascun quesito, con *score* da 1 (basso) a 5 (alto) — i risultati si prestano a essere mappati in modo naturale sulla Scala di Maturità, presente nel *framework* proprietario di Nolan Norton Italia e rappresentata in figura Fig. 3.1.

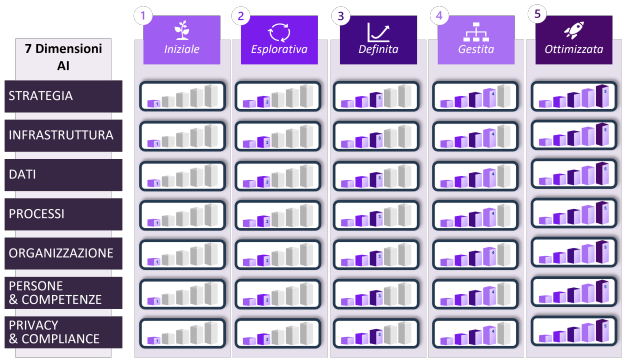


Figura 3.1: Scala di maturità del *framework* Nolan Norton.

Tale scala costituisce il fulcro dell'intero modello, poiché mette in relazione le sette macro-aree di analisi con cinque stadi progressivi di evoluzione dell'IA: da un approccio iniziale e non strutturato (livello 1) fino a una piena ottimizzazione, accompagnata da una solida capacità di innovazione (livello 5).

3.1.1 Sette dimensioni IA

Di seguito viene presentata nel dettaglio una panoramica delle sette Dimensioni IA e dei cinque livelli di maturità.

- **Strategia**: valuta il grado di maturità con cui l'Organizzazione integra l'IA all'interno della propria visione e dei propri processi organizzativi. L'analisi si concentra sulla presenza di una direzione strategica in grado di definire politiche chiare, responsabilità precise e obiettivi misurabili e aggiornati. Un elemento centrale di questa valutazione è la definizione e formalizzazione dell'ambito di applicazione del sistema di gestione dell'IA e l'impegno nel garantire e allocare le risorse umane, infrastrutturali, finanziarie e informative necessarie per l'intera implementazione. Sotto il profilo finanziario e degli investimenti strategici, la dimensione indaga la capacità dell'Organizzazione di pianificare e gestire le iniziative IA in stretta coerenza con le strategie digitali. Viene esaminato il processo metodologico utilizzato per l'identificazione, la classificazione e la prioritizzazione dei progetti all'interno del portafoglio investimenti. Si verifica, inoltre, l'esistenza di criteri strutturati per la stima preventiva dei costi e per la valutazione del ritorno atteso, sia in termini economici che di efficacia dei servizi. Un'attenzione particolare è dedicata alla valutazione dell'allineamento agli standard internazionali di settore e la capacità di integrare organicamente la gestione dell'IA con i sistemi di *governance* preesistenti, come quelli dedicati alla qualità, alla sicurezza e alla privacy. Infine, viene esplorato l'approccio dell'Organizzazione nel coltivare l'innovazione, valutando il bilanciamento tra le iniziative sviluppate internamente e le sinergie costruite all'esterno. Si analizzano il livello di conoscenza e la disponibilità a partecipare agli *AI regulatory sandbox*, sia come opportunità di collaborazione con partner tecnologici e fornitori, sia come leva strategica per sperimentare soluzioni innovative e accelerare l'adeguamento ai requisiti normativi dei sistemi di IA.

- **Infrastruttura**: valuta l'adeguatezza, la scalabilità e la solidità dell'ambiente IT preposto allo sviluppo e alla gestione dei sistemi di

IA. L'analisi si concentra sulla capacità dell'organizzazione di allocare le risorse computazionali in modo flessibile e di garantire prestazioni di rete ottimali per supportare carichi di lavoro complessi. Un aspetto centrale riguarda l'architettura dei dati e l'interoperabilità: il *framework* indaga la propensione delle nuove soluzioni IA a integrarsi fluidamente con i sistemi *legacy* e con le infrastrutture pubbliche preesistenti, nel rispetto degli standard nazionali ed europei. Parallelamente, la valutazione esamina l'implementazione di presidi di cybersecurity, la continuità operativa e la resilienza complessiva delle architetture. Vengono verificati i meccanismi di protezione delle informazioni sensibili e le contromisure tecniche adottate per prevenire anomalie o derive dei modelli. Infine, si analizzano i processi di monitoraggio, accertando la capacità dell'infrastruttura di tracciare le performance e gestire i *log* di sistema in piena conformità ai requisiti normativi.

- **Dati:** analizza la capacità di gestire, valorizzare e governare il patrimonio informativo dell'Organizzazione. Un elemento centrale è rappresentato dalla *data governance*, intesa come l'insieme di strategie, policy, ruoli e processi volti a garantire qualità, affidabilità, tracciabilità e conformità dei dati. In questo contesto, si valuta l'adozione di strumenti quali *business glossary*, indicatori di monitoraggio e l'integrazione delle regole di gestione nei processi operativi e decisionali. Strettamente connessi sono i temi di *data management* e *data architecture*. Viene valutata la capacità di descrivere e catalogare i dati tramite modelli coerenti e sistemi di *metadata management*, nonché la gestione dei dati master per assicurare uniformità organizzativa. Particolare attenzione è posta all'adozione di un'architettura chiara e formalizzata, in grado di supportare l'integrazione tra sistemi eterogenei e la comprensione delle trasformazioni del dato mediante attività di *discovery* e tracciamento del *data lineage*. La dimensione include inoltre la valutazione della qualità dei dati e la capacità di migliorarla in modo continuo in base al contesto di utilizzo. A ciò si affianca la preparazione dei dati in funzione dell'impiego nei sistemi di IA, assicurandone rappresentatività, pertinenza e adeguatezza agli obiettivi d'uso. Si esamina poi il livello di strutturazione delle infrastrutture tecnologiche (sistemi di *storage*, piattaforme di elaborazione e *business intelligence*) e la sicurezza delle informazioni, includendo la gestione attiva degli accessi e dei rischi. Infine, valuta la capacità dell'Organizzazione di garantire un utilizzo affidabile ed etico dei dati, con particolare riferimento alla trasparenza dei contenuti

generati e alla prevenzione di bias e discriminazioni nei dataset impiegati con l'IA.

- **Processi**: valuta la maturità dell'Organizzazione nel governare l'intero ciclo di vita delle soluzioni IA, dall'identificazione dei fabbisogni fino all'impiego della tecnologia per l'ottimizzazione dei servizi. Un elemento centrale è la gestione strutturata della conformità normativa e documentale, essenziale per garantire il rispetto degli standard di sicurezza. Parallelamente, viene esaminata la capacità dell'organizzazione di presidiare l'ecosistema dei rischi, definendo chiare responsabilità lungo l'intera catena del valore e valutando preventivamente gli impatti etici, sociali e sui diritti fondamentali. La dimensione indaga inoltre i livelli di trasparenza operativa, assicurando un'adeguata supervisione umana e una comunicazione chiara verso gli utenti finali. Infine, si analizza la robustezza post-rilascio: l'Organizzazione deve disporre di procedure standardizzate per il monitoraggio continuo delle prestazioni, la gestione tempestiva delle anomalie e l'analisi sistematica dei *Key Performance Indicator* (KPI), al fine di alimentare un ciclo virtuoso di miglioramento e gestione del cambiamento.

- **Organizzazione**: analizza l'assetto strutturale, operativo e procedurale dell'Organizzazione in relazione all'adozione e alla gestione dei sistemi di IA. L'attenzione è posta sulla definizione di un'architettura organizzativa solida, con una chiara assegnazione formale di ruoli e responsabilità lungo l'intero ciclo di vita dei progetti di IA. In particolare, si valuta la presenza di una *leadership* dedicata all'IA, di team responsabili della *data governance* e di referenti incaricati di interfacciarsi con le autorità di sorveglianza in caso di audit o ispezioni. Viene inoltre considerato il corretto inquadramento del ruolo legale dell'Organizzazione (ad esempio come fornitore, importatore, distributore o utilizzatore di sistemi ad alto rischio), elemento fondamentale per definire il perimetro operativo e le relative responsabilità amministrative e legali. Un ulteriore aspetto riguarda la presenza di strutture organizzative interne volte a garantire la conformità normativa e l'allineamento ai principi etici. In questo ambito si analizzano i meccanismi di monitoraggio delle evoluzioni normative e degli standard di armonizzazione, l'adozione di codici di condotta interni e l'istituzione di presidi dedicati alla valutazione d'impatto sui diritti fondamentali e all'analisi etica dei modelli, con particolare attenzione ai sistemi di IA generativa. Si verifica inoltre la

disponibilità di personale e procedure adeguate a garantire un'effettiva supervisione umana sui processi automatizzati. Infine, la dimensione valuta la maturità dei processi interni e dei *framework* operativi, considerando la definizione di procedure per la gestione degli incidenti, l'integrazione nel sistema di gestione della qualità e l'adozione di metodologie di sviluppo flessibili, come l'approccio Agile. Viene inoltre analizzata la capacità dell'organizzazione di documentare e monitorare i test in ambienti controllati (*sandbox*) o in contesti reali, nonché la presenza di ruoli dedicati al *risk management*, con particolare attenzione alla mitigazione dei bias e alla corretta classificazione dei sistemi di IA.

- **Persone e Competenze**: valuta la capacità dell'Organizzazione di formare e valorizzare il proprio capitale umano in funzione di un'adozione sicura e responsabile dell'IA. L'analisi si concentra sulla disponibilità di talenti interni e sulla strutturazione di percorsi formativi differenziati per ruolo. Si verifica la presenza di programmi volti a garantire un'adeguata alfabetizzazione sull'IA per tutto il personale, affiancati da moduli specialistici per le figure direttamente coinvolte nello sviluppo, utilizzo o supervisione dei sistemi. L'obiettivo è dotare l'Organizzazione delle competenze tecniche e manageriali necessarie per governare l'innovazione, gestire eventuali criticità e interpretare correttamente il mutevole quadro normativo. Parallelamente, si misura la consapevolezza diffusa rispetto al mantenimento della conformità etica e legale. Assume particolare rilevanza la definizione chiara e documentata delle responsabilità, assicurando che la supervisione e le decisioni finali nei processi supportati dall'IA rimangano sempre in capo al personale umano e siano pienamente tracciabili. Infine, viene indagata la preparazione dei team nel partecipare proattivamente ad ambienti di sperimentazione tecnologica controllata.

- **Privacy e Compliance**: misura il grado di maturità dell'Organizzazione nell'adozione e gestione di soluzioni IA secondo principi di sicurezza, trasparenza e responsabilità etica. L'analisi si focalizza sull'allineamento continuo al quadro normativo, declinato in base ai diversi profili di rischio dei sistemi di IA. Un primo asse di indagine riguarda la strutturazione dei processi di *compliance*, che includono la gestione documentale, la tracciabilità del ciclo di vita dei modelli e la capacità di cooperare proattivamente con le autorità di vigilanza. Parallelamente, valuta l'impegno verso la trasparenza e la

tutela dei diritti fondamentali. Si verifica che l'Organizzazione adotti *policy* chiare per informare gli *stakeholder* in merito alle decisioni automatizzate, applicando metodologie strutturate per la mitigazione dei rischi legali e sociali. Un ulteriore pilastro è costituito dalla rigorosa *governance* dei dati (*privacy by design*), fondamentale per proteggere le informazioni sensibili e garantire la sicurezza negli ambienti di test. Infine, la dimensione esamina la robustezza del monitoraggio post-rilascio, accertando la presenza di indicatori di performance e processi di *reporting* volti ad assicurare un miglioramento continuo, la standardizzazione delle conoscenze e la piena reversibilità delle decisioni algoritmiche.

3.1.2 Cinque livelli di maturità

- **Livello 1 - Iniziale**: è caratterizzato da capacità IA limitate o del tutto inesistenti. Si evidenzia la totale assenza di una strategia dedicata, di *governance* e di criteri di prioritizzazione. Sul fronte tecnico, si osserva un'infrastruttura inadeguata, priva di scalabilità e *business continuity*, unita alla mancanza di presidi strutturati di *data management* e *data governance*. Operativamente, si registra un approccio frammentario e puramente reattivo, assenza di strutture organizzative, di competenze specifiche e di piani formativi. Infine, si riscontra la carenza di misure atte a garantire la conformità normativa.

- **Livello 2 - Esplorativo**: è caratterizzato dall'avvio delle prime sperimentazioni e da un'iniziale integrazione dell'IA con le infrastrutture esistenti. Pur permanendo l'assenza di una strategia organica e di linee guida documentate, si evidenzia l'applicazione di parziali criteri di prioritizzazione alle singole iniziative. Sul fronte tecnico, si osserva un'allocazione di risorse su richiesta e l'introduzione di soluzioni basilari di *business continuity*, sebbene manchino ancora scalabilità e automazione. La gestione dei dati risulta frammentaria, non sistematica e supportata da processi prevalentemente manuali. Operativamente, si registra l'identificazione delle prime figure responsabili e un interesse crescente verso la formazione, affidata tuttavia a iniziative individuali e non strutturate. Infine, si riscontra l'avvio delle prime attività di *compliance*, con un iniziale riconoscimento dell'importanza della trasparenza.

- **Livello 3 - Definito**: è caratterizzato da una prima fase di industrializzazione, con progetti IA avanzati che prevedono una

gestione attiva. Si evidenzia un approccio olistico in cui la strategia e i criteri di prioritizzazione sono in fase di formalizzazione. Sul fronte tecnico, si osserva la presenza di risorse dedicate con *provisioning* automatizzato di base, supportate da una rete performante e da adeguate misure di *business continuity*, unitamente all'avvio delle prime iniziative strutturate di *data management* e *data governance*. I processi IA risultano documentati e gestiti in modo strutturato tramite ruoli specifici, sebbene l'integrazione con le funzioni impattate non sia ancora completa. Sul piano del personale, si rileva l'erogazione di percorsi formativi e la presenza di competenze, benché risultino ancora parziali rispetto al reale fabbisogno. Infine, si riscontra l'implementazione di attività strutturate per la *compliance* e il monitoraggio dell'IA, pur con un'adozione ancora parziale.

- **Livello 4 - Gestito**: si rileva una fase di piena industrializzazione e consolidamento, caratterizzata da processi definiti e dal supporto attivo del *top management*. Si evidenzia una strategia IA pienamente formalizzata, con criteri di prioritizzazione chiari a tutte le funzioni e sottoposti a revisione periodica. Sul fronte tecnico, si osserva un'infrastruttura ottimizzata, dotata di *provisioning* e scalabilità dinamici e di un'elevata affidabilità, unita a una gestione del ciclo di vita del dato strutturata. Si registra l'integrazione strutturale dell'IA nei processi, con una *governance* chiara, ruoli formalizzati e un controllo sistematico delle attività. Sul piano organizzativo, si riscontra una diffusione capillare delle competenze, sostenuta da un piano formativo regolare, adattivo e coerente con la *capacity*. Infine, la *compliance* risulta pienamente integrata nella *governance*, supportata da politiche complete e periodicamente aggiornate.

- **Livello 5 - Ottimizzato**: si rileva una gestione in grado di governare la massima complessità, caratterizzata da un'adozione e da una standardizzazione completa dell'IA. Si evidenzia una strategia pienamente formalizzata, in cui l'IA costituisce parte integrante e fondante delle scelte direzionali. Sul fronte tecnico, si osserva un'infrastruttura intelligente con risorse automatizzate, scalabili e accessibili in modalità *self-service*, unita a una gestione dei dati proattiva, strategica e interamente integrata. Si misura una *governance* matura e automatizzata, supportata da processi adattivi per la gestione di tutti gli aspetti dell'adozione. Sul piano organizzativo, si riscontra l'implementazione di una formazione continua perfettamente allineata alla strategia e una gestione proattiva della *capacity*.

Infine, la conformità normativa risulta garantita in modo predittivo, automatizzato e strutturalmente fuso nella *governance* generale.

3.2 Dall'*assessment* generale al *self-assessment*

A partire dal questionario originario approfondito, sono state selezionate e riadattate 46 domande ritenute particolarmente rilevanti per il contesto territoriale in esame. Da tale processo è stato costruito un questionario di *self-assessment*, i cui risultati costituiscono il principale oggetto di analisi della presente tesi. Attraverso questo mezzo, si è voluto indagare il livello di competenze — informatiche, organizzative e strutturali — degli organismi interni al Soggetto, nonché ottenere una visione aggregata della maturità IA a livello territoriale.

Il questionario è stato infatti somministrato ai principali rappresentanti del Territorio, ai quali è stata richiesta la compilazione in autonomia, pur garantendo supporto per eventuali chiarimenti e/o dubbi.

I risultati dell'autovalutazione hanno consentito di:

1. Valutare il livello di Maturità As-Is e posizionarlo sulla scala di maturità, sia con una visione complessiva territoriale, sia per le singole parti che lo compongono;

2. Analizzare la presenza di eventuali cluster tra i rispondenti, al fine di comprendere meglio la composizione del Soggetto e distinguere tra anticipatori, ritardatari e profili atipici.

Nel capitolo 4 si approfondisce la struttura del campione di indagine, con la definizione del dataset derivante dal *self-assessment* e le relative modalità di gestione.

Capitolo 4

Campione analizzato e creazione del dataset pulito

4.1 Struttura e mappatura del campione

Il dataset oggetto di analisi è costituito dalle risposte fornite da un campione di 29 Enti, raccolte attraverso la compilazione di un questionario di *self-assessment* composto da circa 46 domande. Il numero effettivo di quesiti può variare in funzione delle risposte fornite, poiché alcune domande sono condizionate e vengono presentate solo qualora pertinenti (ad esempio, solo se l'Ente utilizza sistemi di IA o tratta dati personali con l'IA).

Il disegno di ricerca originario prevede il coinvolgimento di un numero più ampio di Enti, selezionati con l'obiettivo di garantire un elevato livello di rappresentatività del contesto oggetto di analisi. Al momento della stesura del presente lavoro, tuttavia, la fase di raccolta dei dati risulta ancora in corso e non tutti gli Enti invitati hanno completato la compilazione del questionario di autovalutazione.

Per tale ragione, le analisi presentate in questa sezione si basano esclusivamente sui questionari effettivamente pervenuti entro il momento dell'elaborazione. Pur trattandosi di un dataset parziale rispetto al disegno campionario originario, il numero di risposte raccolte è stato ritenuto adeguato per consentire una prima esplorazione empirica del fenomeno e per procedere con le analisi statistiche sul campione disponibile.

Per l'indagine sono stati contattati i rappresentanti degli Enti, suggerendo loro di affidare la compilazione a personale tecnico o amministrativo qualificato. Tale procedura ha fatto sì che per ogni Ente vi fosse un solo rispondente, garantendo di fatto una corrispondenza biunivoca tra il numero dei questionari compilati e quello degli Enti. Di conseguenza,

l'unità statistica di osservazione è rappresentata dall'Ente stesso, e le analisi successive sono state condotte considerando tale specificità metodologica.

La struttura informativa del campione si articola in due caratterizzazioni principali, che rappresentano delle possibili chiavi interpretative per l'analisi della Maturità As-Is:

1. **Dimensione funzionale (ecosistema):** identifica il settore o ambito operativo in cui l'Ente svolge la propria attività.
2. **Dimensione giuridica (PA vs non-PA):** distingue gli Enti in base alla loro natura, pubblica oppure privata/partecipata.

L'elenco degli ecosistemi e la loro caratterizzazione sono riassunti nella tabella 4.1:

Ecosistema	**Ambito di riferimento**
Agricoltura e foreste	Gestione del territorio agro-forestale e sostenibilità ambientale.
Amministrazione	Supporto, servizi istituzionali e gestione territoriale.
Arte e cultura	Valorizzazione del patrimonio storico e artistico.
Costruire e abitare	Edilizia, urbanistica e infrastrutture fisiche.
Formazione e lingue	Istruzione ed educazione.
Informatica e digitalizzazione	Gestione di tecnologie digitali, infrastrutture informatiche e servizi ICT.
Innovazione e ricerca	Sviluppo scientifico e innovazione tecnologica.
Lavoro ed economia	Sviluppo imprenditoriale, occupazione e politiche di mercato.
Salute e benessere	Servizi sanitari, prevenzione e assistenza pubblica.
Sicurezza e protezione civile	Tutela dell'ordine pubblico e gestione delle emergenze.
Turismo e mobilità	Promozione del territorio, trasporti e reti di collegamento.

Tabella 4.1: Classificazione degli Enti per ecosistema di riferimento.

La doppia categorizzazione permette di condurre un approfondimento dell'analisi dei risultati: la suddivisione per ecosistema consente di verificare se la maturità IA dipende dalla vocazione funzionale. Un'ipotesi sottostante è che la varianza tra gli ecosistemi sia significativa; ad esempio, si attende un punteggio superiore nel settore "Informatica e digitalizzazione", rispetto

ad altri settori. La distinzione tra PA e non PA consente, invece, di investigare l'impatto dei vincoli strutturali interni sull'adozione tecnologica, valutando, per esempio, l'impatto burocratico, ma anche la diversa prontezza al cambiamento e le risorse economiche a disposizione.

4.2 *Pre-processing* e costruzione del dataset

La trasformazione dei dati grezzi raccolti tramite questionario in un formato idoneo all'elaborazione matematica ha richiesto una rigorosa fase di *data engineering*. La fase di *pre-processing* risulta fondamentale per garantire l'affidabilità dei modelli analitici successivi: la qualità dei dati in ingresso, infatti, determina direttamente la validità dell'intero processo di *knowledge discovery* come descritto in *Introduction to Data Mining* di Tan et al. (2016). Di seguito vengono riportate le diverse procedure eseguite.

Protocollo di anonimizzazione e gestione della privacy. Data la natura puntuale dell'analisi e la ridotta cardinalità dell'insieme degli Enti, il protocollo di privacy adottato si basa sulla pseudonimizzazione, tecnica che consiste nel nascondere l'identità delle persone all'interno di un database, sostituendo i loro dati identificativi con etichette artificiali.

- **Livello Ente:** ogni Ente reale è sostituito da un codice univoco: E01, E02, . . . , E29.
- **Livello Ecosistema:** le etichette macroscopiche sono mantenute in chiaro per permettere l'interpretazione dei risultati.

Tale procedura è applicata per tutelare l'identità istituzionale dei partecipanti, pur mantenendo inalterata la struttura delle correlazioni nei dati.

Codifica delle risposte e composizione della stratificazione. Le risposte fornite sono di natura categorica ordinale. Al fine di renderle computabili all'interno di un modello matematico, è stato definito un processo di codifica basato su una mappatura esplicita.

Nello specifico, attraverso un file Excel denominato `Mapping_Score`, ogni domanda del questionario è stata associata alle sue cinque possibili opzioni di risposta. Le variabili sono state quindi mappate all'interno della colonna *score*, su una scala numerica lineare con valori interi in $[1, 5]$, dove:

$$1 = \text{maturità minima}, \qquad 5 = \text{maturità massima}.$$

Metodologicamente, questa operazione si traduce in un processo di *label encoding* supervisionato, costruito in modo da preservare l'intrinseca gerarchia semantica e ordinale delle risposte.

Parallelamente alla codifica delle risposte, si è resa necessaria la definizione formale delle due stratificazioni teoriche precedentemente formalizzate: per ecosistema e per natura giuridica. A tale scopo, è stato predisposto un secondo file di *mapping* denominato `ente_ecosistema_PA`. Nello specifico, la classificazione avviene mediante una colonna descrittiva per l'ecosistema e un indicatore dicotomico per la natura giuridica

Come verrà approfondito nel capitolo successivo dedicato al modello matematico, questi due file hanno ricoperto ruoli complementari ma distinti nella fase di modellazione: il `Mapping_Score` è stato impiegato per operazioni di *merge* con il contenuto delle risposte, mentre il *mapping* `ente_ecosistema_PA` per l'anagrafica è stato progettato per fungere da tabella dimensionale indipendente. L'azione combinata di queste due componenti ha permesso di filtrare e aggregare i risultati, ponendo le basi per l'esplorazione visiva, l'analisi comparativa e la reportistica finale.

Gestione dei *missing values*. Durante la fase di *data cleaning* è stata affrontata la gestione dei *missing values*. In un contesto di *self-assessment*, basato sulle capacità della propria organizzazione, la mancata risposta è stata interpretata come un segnale informativo di assenza di competenza o di non applicabilità. È stata quindi adottata una strategia di imputazione cautelativa: i valori mancanti sono stati sostituiti con il valore minimo della scala, cioè 1.

Definendo X la matrice dei dati, l'elemento generico x_{ij}, che rappresenta la risposta dell'Ente i alla domanda j, è definito come:

$$x_{ij} = \begin{cases} x_{ij} & \text{se presente,} \\ 1 & \text{se mancante.} \end{cases}$$

Questa scelta evita di sovrastimare artificialmente il livello di digitalizzazione in materia IA dell'Ente, preferendo una valutazione prudenziale della maturità attuale.

4.3 Limiti metodologici e considerazioni sulla robustezza

L'impianto metodologico della ricerca presenta alcune limitazioni intrinseche, derivanti principalmente dalla natura auto-dichiarativa dello strumento di indagine e dalle caratteristiche del campione:

Vincoli campionari e bias di auto-valutazione. Ci sono alcune criticità legate alla natura auto-dichiarativa del *self-assessment*:

- ***Social desirability bias*:** nonostante ai partecipanti sia stata garantita l'assenza di risposte strettamente giuste o sbagliate, permane il rischio che i rispondenti tendano a sovrastimare la maturità del proprio Ente. Tale dinamica può scaturire dalla volontà di allinearsi alle presunte aspettative del progetto o dal desiderio di proiettare un'immagine organizzativa più innovativa.
- **Bias del singolo rispondente:** la compilazione del questionario affidata a un unico referente espone i risultati a un rischio di parzialità. La valutazione finale riflette inevitabilmente la percezione, il livello di informazione e il ruolo specifico del singolo individuo e potrebbe non catturare appieno l'effettiva complessità dell'intera struttura organizzativa.
- **Numerosità campionaria ridotta:** la dimensione contenuta del campione vincola l'applicabilità di modelli di inferenza statistica avanzata. Ciò conferisce ai risultati emersi una valenza prevalentemente esplorativa e descrittiva del contesto specifico.

Al fine di mitigare l'incidenza dei primi due bias, sono stati programmati incontri di *follow-up* con alcuni Enti chiave del Territorio. Tali sessioni sono volte a valutare, validare e analizzare in dettaglio le risposte fornite durante la fase di *self-assessment*.

Nel capitolo successivo verranno presentati gli strumenti metodologici e analitici adottati per l'elaborazione dei dati. Nello specifico, si descriveranno i processi di aggregazione e sintesi delle risposte implementati tramite *dashboard* in Power BI, unitamente alle tecniche di riduzione della dimensionalità. Tra queste ultime verranno approfondite l'Analisi delle Componenti Principali (PCA) e un approccio alternativo basato sui sette abilitatori teorici del *framework*, le cui logiche computazionali sono state sviluppate in ambiente Python.

Capitolo 5

Modello matematico e architettura dell'analisi

5.1 Obiettivi dell'analisi

Il presente studio si pone un duplice obiettivo, analizzando il dataset da una prospettiva sia descrittiva che inferenziale. Tali scopi sono stati raggiunti grazie all'impianto di categorizzazione e ai file di *mapping* introdotti nel capitolo precedente, uniti al modello matematico che verrà presentato di seguito. Nello specifico, l'analisi si propone di:

1. **Valutare la Maturità As-Is e confrontarla con la Maturità Target.** L'attuale livello di maturità degli Enti viene quantificato rispetto a sette macro-dimensioni teoriche dell'IA e confrontato con il livello atteso di Maturità IA Target rilevato dal Soggetto[1]. In particolare, il livello di Maturità As-Is è stato definito considerando differenti livelli di analisi: sia in forma aggregata, cioè considerando l'Organizzazione nel suo insieme, sia attraverso suddivisioni predefinite del campione, quali l'appartenenza a specifici ecosistemi e la distinzione tra Enti della Pubblica Amministrazione (PA) e non-PA.

2. **Clusterizzare il campione.** L'analisi puramente aggregata, definita in base a gerarchie predefinite o su una base strettamente individuale, viene superata identificando matematicamente ulteriori sotto-gruppi di Enti accomunati da profili di maturità simili, definendo così dei "prototipi" di adozione dell'IA.

[1]Si ricorda che le dinamiche di misurazione della Maturità IA Target esulano dagli scopi di questa trattazione.

5.2 Valutazione della Maturità As-Is

La prima fase del modello è finalizzata alla misurazione quantitativa della maturità IA dell'organizzazione. Come definito in dettaglio nel capitolo 4, l'architettura di base si appoggia su specifici file di *mapping* costruiti per tradurre le informazioni qualitative in metriche quantitative elaborabili. Le operazioni di modellazione e integrazione dei dati sono state condotte direttamente all'interno di Power BI, software utilizzato per la *Business Intelligence*.

Dal punto di vista architetturale, l'integrazione dei dati in Power BI è stata strutturata seguendo il paradigma del modello relazionale a stella. Nello specifico, il dataset contenente le risposte al questionario (definito dalle tuple: Ente, domanda, risposta testuale) è stato inizialmente sottoposto a un'operazione di *unpivoting*. Questa trasformazione ha convertito i dati in formato lungo, generando la tabella dei fatti centrale.

In seguito, tale tabella è stata unita, tramite un'operazione di *merge* interna a Power BI, al file "`Mapping_Score`", composto dalle colonne: domanda, risposta, *score* e dimensione teorica. Tale operazione ha permesso di assegnare a ciascuna risposta la corretta dimensione teorica di appartenenza e il relativo valore numerico su una scala ordinale discreta nell'intervallo $[1, 5]$, dove 1 rappresenta un livello "Iniziale" e 5 un livello "Ottimizzato".

Parallelamente, il file anagrafico "`ente_ecosistema_PA`", contenente le informazioni relative all'ecosistema di appartenenza e alla classificazione dicotomica in PA o non-PA, è stato importato nel sistema come tabella dimensionale indipendente. Anziché procedere con un ulteriore *merge*, la tabella dimensionale anagrafica è stata collegata alla tabella dei fatti principale sfruttando il motore relazionale di Power BI, completando così lo schema a stella.

Questa scelta strutturale consente di utilizzare le partizioni del Soggetto come filtri logici per raggruppare i dati. A partire da questa struttura, per ciascun Ente $i = 1, \ldots, N$ e per ciascuna dimensione teorica $k = 1, \ldots, 7$, il modello definisce uno *score* aggregato $D_{i,k}$, calcolato come media aritmetica dei punteggi delle variabili afferenti a tale dimensione.

Sfruttando il modello relazionale appena descritto, le operazioni di aggregazione avvengono dinamicamente all'interno delle visualizzazioni al variare dei filtri applicati dall'utente. Nello specifico, sono stati sviluppati dei cruscotti interattivi in Power BI che supportano una duplice prospettiva di analisi:

- **Analisi aggregata, di sottoinsieme e singola:** è possibile visualizzare il vettore delle medie riferito all'intero campione globale, per ogni dimensione IA:

$$\frac{1}{N}\sum_{i=1}^{N} D_{i,k}, \quad \forall k = 1, \ldots, 7.$$

  Inoltre, tramite l'applicazione dei filtri derivanti dalla tabella anagrafica, è possibile aggregare i dati su specifici sottoinsiemi con un approccio *drill-down*, incrociando dinamicamente l'ecosistema di riferimento o la tipologia di Ente tra PA e non-PA. Le operazioni di aggregazione si adattano dinamicamente all'interno di Power BI per calcolare la media parziale: il parametro N e l'indice i della sommatoria vanno a rappresentare esclusivamente la cardinalità e gli Enti del sottogruppo filtrato.

- **Maturità As-Is vs Target:** l'interfaccia permette di confrontare direttamente il punteggio As-Is, ricavato empiricamente dai dati, con il livello target desiderato. Questa analisi evidenzia le aree di maggiore allineamento e le eventuali carenze rispetto agli obiettivi prefissati sulle sette dimensioni d'indagine.

5.3 Clusterizzazione

Sebbene la visualizzazione su Power BI fornisca uno strumento descrittivo potente, essa presenta due limiti analitici speculari:

- L'analisi del campione aggregato appiattisce la varianza: la media aritmetica tra un Ente altamente innovativo, con *score* medio tra il 4 e il 5 e uno poco maturo con *score* tra l'1 e il 2, restituisce un valore intermedio che non rappresenta fedelmente nessuna delle due realtà, occultando la reale eterogeneità del territorio.

- Il focus sulle sole peculiarità del singolo Ente preclude la generalizzazione dei risultati, limitando la capacità di far emergere *pattern* sistemici e tendenze trasversali all'intero apparato territoriale.

Per superare questa dicotomia tra l'eccessiva sintesi del dato aggregato e la frammentazione dell'analisi puntuale, occorre individuare un livello di indagine intermedio basato sulla reale similarità dei profili. Si rende pertanto necessaria una tecnica di apprendimento non supervisionato per partizionare

gli N Enti in C gruppi, massimizzando l'omogeneità interna (*intra-cluster*) e l'eterogeneità esterna (*inter-cluster*), seguendo l'approccio consolidato in letteratura descritto in *Introduction to Data Mining* da Tan et al. (2016).

5.3.1 Riduzione della dimensionalità: dalla PCA all'approccio con le sette dimensioni teoriche

Inizialmente, per affrontare l'elevata dimensionalità del dataset originale $X \in \mathbb{R}^{29\times46}$, si è esplorato l'utilizzo dell'Analisi delle Componenti Principali abbinata all'algoritmo K-Means. L'intento è stato quello di estrarre le prime due componenti principali (PC1 e PC2) per mappare i rispondenti su un piano cartesiano $\mathbb{R}^2$ e individuare visivamente la presenza di possibili cluster.

Tuttavia, questo approccio è stato scartato per tre criticità fondamentali:

1. **Instabilità algebrica.** La presenza di un numero di variabili maggiore rispetto alle osservazioni ($P > N$) determina la singolarità della matrice Σ di covarianza campionaria, rendendo l'estrazione degli autovettori instabile e il modello esposto al rischio di *overfitting*.

2. **Perdita di interpretabilità.** Le componenti principali perdono il legame semantico con le sette dimensioni d'indagine, rendendo impossibile spiegare i cluster finali in termini, ad esempio, di strategia, dati o infrastruttura.

3. **Necessità di definire a priori il numero di cluster.** L'algoritmo K-means è un metodo che raggruppa i punti dati in un numero predefinito di cluster basato sulla loro distanza dal centroide di ciascun cluster.

Si è quindi optato per una riduzione della dimensionalità guidata dalle sette dimensioni IA già impostate nel modello, seguita da un algoritmo di clustering gerarchico agglomerativo, una tecnica di clustering utilizzata per organizzare i dati in una struttura ad albero chiamata dendrogramma, che rappresenta le relazioni di similarità tra i dati. Le caratteristiche principali del metodo possono essere elencate nei seguenti punti:

- Approccio *Bottom-Up*: l'algoritmo inizia con ogni punto dati come un singolo cluster e li unisce gradualmente in cluster più grandi.

- Distanza e similarità: la fusione tra cluster è basata su una misura di distanza. Nel caso in questione, si è usata la distanza euclidea.

- Dendrogramma: la struttura gerarchica creata aiuta a visualizzare i cluster e a scegliere il numero ottimale di gruppi.

5.3.2 Modellizzazione del clustering gerarchico e metriche di distanza

Aggregazione dei dati e standardizzazione. Sfruttando il *mapping* presente in "`Mapping_Score`", la matrice originale X è stata ridotta a una nuova matrice più compatta e interpretabile $X' \in \mathbb{R}^{29\times 7}$, la quale assume un significato chiaro per le sette dimensioni. Affinché l'algoritmo di clustering non fosse influenzato da eventuali differenze di varianza tra le diverse dimensioni, si è proceduto alla standardizzazione: per ogni elemento della matrice, si è calcolato il valore standardizzato $z_{i,k}$ come:

$$z_{i,k} = \frac{D_{i,k} - \mu_k}{\sigma_k}, \forall i = 1, \ldots, N, \forall k = 1, \ldots, 7,$$

con, rispettivamente, μ_k e σ_k media e deviazione standard di ogni dimensione. Si ottiene così la matrice standardizzata Z, dove ogni colonna ha media nulla e varianza unitaria.

Calcolo della distanza. Per quantificare la similarità tra gli Enti nel nuovo spazio vettoriale $\mathbb{R}^7$, si è adottata la metrica della distanza euclidea. Dati due Enti rappresentati dai vettori riga u e v, la loro distanza nello spazio è:

$$d(u, v) = \sqrt{\sum_{k=1}^{7} (z_{u,k} - z_{v,k})^2}$$

Questo calcolo, reiterato per tutte le combinazioni possibili, genera una matrice delle distanze simmetrica $M \in \mathbb{R}^{29\times 29}$.

Clustering gerarchico con metodo di Ward. Sulla matrice delle distanze M è stato applicato l'algoritmo di clustering gerarchico agglomerativo. Il processo iterativo ha inizio considerando ogni singola osservazione data dagli $N = 29$ Enti come un cluster indipendente, caratterizzato da una varianza interna iniziale pari a zero. Ad ogni step, l'algoritmo valuta tutte le possibili combinazioni di fusione e unisce la coppia di cluster considerata più affine.

Per definire tale affinità, lo studio adotta il metodo di Ward o metodo della minima varianza. A differenza delle metriche classiche basate sulla pura distanza geometrica tra i punti, come il *single* o *complete linkage*, il criterio di Ward adotta un approccio orientato all'ottimizzazione: il suo obiettivo è minimizzare l'incremento della devianza totale dentro i gruppi (Error Sum of Squares, ESS) ad ogni passaggio di aggregazione.

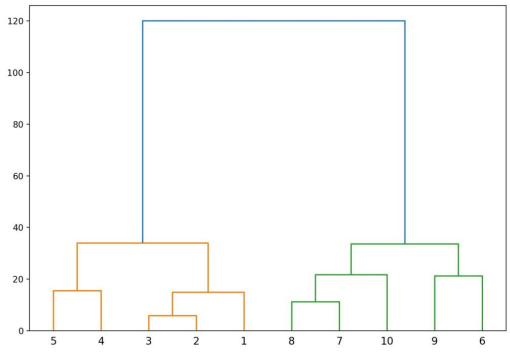


Figura 5.1: Esempio di dendrogramma.

In termini pratici, l'algoritmo sceglie di fondere esclusivamente la coppia di cluster che produce il minor aumento possibile della varianza totale. Questa logica garantisce la creazione di gruppi compatti, dimensionalmente bilanciati e minimizza la perdita di informazione sui profili degli Enti. Dati due cluster A e B, l'incremento di devianza Δ generato dalla loro potenziale unione è calcolato secondo la seguente formula:

$$\Delta(A, B) = \frac{n_A n_B}{n_A + n_B} ||C_A - C_B||^2,$$

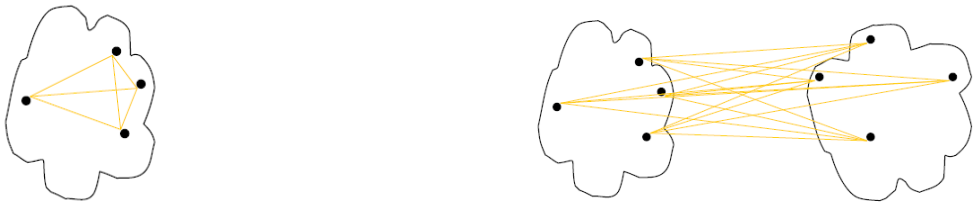
dove n_A e n_B indicano la numerosità dei rispettivi cluster, mentre C_A e C_B ne rappresentano i centroidi.

Taglio del dendrogramma e definizione dei cluster. Il processo iterativo di aggregazione genera una struttura ad albero denominata dendrogramma. In questa rappresentazione grafica, mostrata in Fig. 5.1, l'asse delle ordinate illustra la distanza, ovvero l'incremento di devianza Δ, a cui si verifica ogni singola unione. A differenza di algoritmi partizionali come il K-Means, nel clustering gerarchico il numero ottimo di cluster non viene imposto a priori. La sua determinazione avviene ricercando i punti in cui si registra il gradiente $\Delta h_m = h_m - h_{m-1}$ massimo, dove h_m rappresenta l'altezza dell'm-esima fusione.

Osservando i rami del grafico, un valore marcatamente elevato di Δh_m (cioè un notevole dislivello verticale) segnala visivamente un vero e proprio salto di eterogeneità: ciò indica che l'algoritmo sta forzando l'unione di gruppi di Enti con caratteristiche radicalmente dissimili. L'analisi visiva di queste distanze ha permesso di stabilire l'altezza ottimale a cui tracciare la linea di "taglio" orizzontale sul dendrogramma, arrestando il processo di fusione un attimo prima di unire cluster troppo eterogenei. In questo modo si è

definita la partizione definitiva del campione, unendo la valutazione empirica del grafico al rigore della metrica di Ward.

Validazione dei cluster ottenuti: Silhouette Score. Per validare la bontà della partizione ottenuta tramite il metodo di aggregazione di Ward, l'analisi si avvale di una misura di validazione interna, che viene utilizzata per misurare la qualità della struttura di clustering ottenuta, senza doversi basare su informazioni o etichette esterne preesistenti. Tra le diverse metriche possibili (*rand index*, *adjusted rand-index*, indici di coesione e indici di separazione), lo studio adotta il Silhouette Score. Tale indice permette di valutare il grado di coerenza con cui ogni oggetto si colloca all'interno del proprio cluster, considerando simultaneamente sia la coesione interna del gruppo che la sua separazione rispetto agli altri cluster. Una rappresentazione grafica del significato di coesione e separazione è data dalle due immagini in Fig. 5.2.



(a) Coesione. (b) Separazione.

Figura 5.2: Metriche per valutare la qualità dei cluster.

Nello specifico, per ogni singola osservazione i, l'algoritmo calcola due parametri: $a(i)$, ovvero la dissimilarità media di i rispetto a tutti gli altri oggetti all'interno dello stesso cluster (dove un valore minore indica un'assegnazione migliore), e $b(i)$, che rappresenta la minima dissimilarità media di i rispetto a un qualsiasi altro cluster di cui l'oggetto non fa parte. Il coefficiente per il singolo punto è definito matematicamente dalla formula:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}.$$

Questo punteggio è compreso in un intervallo tra -1 e $+1$, dove valori più prossimi a 1 indicano una migliore qualità dell'assegnazione.

Per valutare infine la robustezza complessiva del modello generato, si calcola la media di $s(i)$ su tutte le osservazioni del dataset: tale indicatore

aggregato restituisce una misura globale di quanto appropriatamente l'intero campione di dati sia stato segmentato.

5.4 Implementazione computazionale in Python

L'architettura matematica del modello relativa alla fase di segmentazione degli Enti è stata implementata computazionalmente utilizzando il linguaggio di programmazione *Python*. Il codice, riportato in Appendice A, è stato sviluppato avvalendosi di librerie *open-source* standard per la manipolazione dei dati (`pandas`, `numpy`) e per il *machine learning* (`scikit-learn`, `scipy`). Nello specifico, la stesura degli script si è articolata in due macro-fasi distinte: una prima fase esplorativa che ha visto l'uso della PCA e una seconda fase definitiva basata sulla riduzione dimensionale nello spazio teorico di valutazione.

Di seguito si riporta una sintesi delle funzioni più rilevanti impiegate nelle diverse fasi della pipeline:

1. **Gestione e manipolazione dei dati:**
    - `pd.read_excel()` e `to_excel()`: funzioni fondamentali per il caricamento del dataset originale e per l'esportazione strutturata degli output finali, garantendo l'interoperabilità con strumenti di Business Intelligence come Power BI.
    - `groupby()`: utilizzata nella fase di profilazione per aggregare le osservazioni in base all'etichetta del cluster assegnato, permettendo il calcolo rapido di metriche sintetiche come le medie e le deviazioni standard (dispersione *intra-cluster*).
2. **Pre-processing e machine learning:**
    - `StandardScaler.fit_transform()`: modulo cruciale per la preparazione dei dati. Standardizza le feature rimuovendo la media e scalando la varianza a 1. Questo passaggio è indispensabile prima di applicare algoritmi basati sulla distanza (PCA, K-Means, Ward) per evitare che variabili con scale diverse dominino il calcolo della similarità.
    - `PCA.fit_transform()`: istanzia ed esegue l'Analisi delle Componenti Principali, proiettando lo spazio multidimensionale dei dati standardizzati in un nuovo spazio a dimensionalità ridotta (2 componenti), massimizzando la varianza spiegata.
    - `silhouette_score()`: funzione di valutazione statistica che calcola il coefficiente di Silhouette medio di tutte le osservazioni, quantificando la bontà del partizionamento mettendo a rapporto

la distanza media *intra-cluster* (coesione) con la distanza media rispetto al cluster più vicino (separazione).

3. **Clustering gerarchico e calcolo spaziale:**

   - `sch.linkage(..., method='ward')`: motore computazionale del clustering gerarchico. Calcola la matrice delle distanze e restituisce una matrice di *linkage* che codifica l'albero gerarchico delle fusioni. L'opzione `method='ward'` unisce ad ogni step i cluster che minimizzano l'incremento della varianza interna totale.
   - `sch.fcluster()`: funzione di estrazione dei cluster dal dendrogramma. Consente di tagliare l'albero gerarchico a un livello specifico (criterio `maxclust` per ottenere 3 cluster) restituendo un vettore con l'assegnazione definitiva di ciascun Ente.
   - `pdist()` e `squareform()`: funzioni per il calcolo delle distanze euclidee tra i centroidi. `pdist()` calcola le distanze *pairwise*, mentre `squareform()` trasforma l'*array* in una matrice quadrata simmetrica per una visualizzazione leggibile.

Capitolo 6

Analisi dei risultati

In apertura di questo capitolo, è opportuno precisare che i risultati di seguito riportati si basano su un campione specifico e circoscritto e non rappresentano la totalità del Territorio. Pertanto, le evidenze emerse hanno una valenza esplorativa e non hanno la pretesa di descrivere in modo esaustivo la totalità del dominio d'indagine.

Prima di procedere con l'interpretazione dei risultati di maturità IA e la successiva segmentazione del Soggetto, è stato eseguito un protocollo di *data quality check* per valutare il tasso di copertura delle risposte fornite dai 29 Enti del campione.

6.1 Analisi preliminare e qualità del dato

La tabella 6.1, prodotta tramite lo script di analisi dei dati riportato in Appendice A, conferma un'eccellente qualità complessiva: ben 22 Enti su 29 hanno infatti raggiunto un tasso di compilazione superiore al 95% sulle 46 domande previste.

Sebbene siano stati identificati due Enti (E01 ed E03) con una copertura lievemente inferiore all'80%, è fondamentale sottolineare che tali valori mancanti non rappresentano un'omissione o una mancanza da parte dei rispondenti. Essi sono, al contrario, il risultato diretto della struttura adattiva del questionario. Tramite apposite logiche di salto, alcune domande non sono state somministrate a quegli Enti che, nelle domande filtro, dichiaravano di non possedere determinati prerequisiti.

Poiché la mancata visualizzazione di tali quesiti riflette oggettivamente uno stadio di maturità iniziale in quegli specifici ambiti, si è optato per un'imputazione logica e cautelativa. A tutte le domande bypassate a causa delle logiche di salto è stato assegnato d'ufficio il punteggio

minimo (*score* = 1). Tale approccio ha permesso di non escludere alcun Ente dall'analisi, preservando la numerosità campionaria e garantendo una valutazione coerente ed equa della loro effettiva maturità IA.

Ente	Missing	Copertura (%)	Ente	Missing	Copertura (%)
E01	10	78.3	E16	2	95.7
E02	8	82.6	E17	1	97.8
E03	10	78.3	E18	5	89.1
E04	8	82.6	E19	0	100.0
E05	3	93.5	E20	1	97.8
E06	1	97.8	E21	2	95.7
E07	1	97.8	E22	1	97.8
E08	0	100.0	E23	6	87.0
E09	2	95.7	E24	2	95.7
E10	1	97.8	E25	1	97.8
E11	0	100.0	E26	2	95.7
E12	1	97.8	E27	2	95.7
E13	0	100.0	E28	2	95.7
E14	1	97.8	E29	1	97.8
E15	0	100.0			

Tabella 6.1: Report sulla qualità dei dati: copertura per Ente.

6.2 Valutazione della Maturità As-Is e *gap analysis*

6.2.1 Il profilo di maturità: As-Is vs Target

La prima fase dell'analisi si concentra sulla valutazione aggregata del campione. Al fine di garantire la riservatezza del Soggetto e, al contempo, fornire una rappresentazione realistica e sfaccettata della sua maturità, le valutazioni puntuali emerse dall'analisi sono state tradotte in range corrispondenti. I range sono stati creati in modo da comunicare al meglio la variabilità del dato e presentano un'ampiezza fissa pari a 0.4, ad eccezione del primo e dell'ultimo intervallo, definiti rispettivamente tra $0 - 0.2$ e $4.8 - 5.0$.

Il livello medio di maturità As-Is si attesta su un punteggio nel range $1.8 - 2.2$, posizionando il Soggetto in una fase definita come "Esplorativa" nella scala di maturità. Questo livello è caratterizzato da una sperimentazione guidata dell'IA, con prime fasi di industrializzazione.

Tale evidenza empirica è stata confrontata con il livello target, fissato a un punteggio medio nel range $4.3 - 4.7$, che corrisponde a una fase "Gestita-Ottimizzata" dell'adozione dell'IA. La Maturità Target delinea l'ambizione di un'IA avanzata, strutturale e diffusa, capace di supportare tutti gli Enti sfruttandone la piena capacità trasformativa, nel rispetto degli standard di etica, trasparenza e affidabilità.

La costruzione dei range è stata calibrata per rispettare e mantenere intatto il divario matematico reale tra lo stato attuale e l'obiettivo target. In questo modo, le proporzioni rimangono inalterate e le priorità di intervento emergono con chiarezza. Nella Tabella 6.2 viene illustrato come le dimensioni Dati, Infrastruttura, Privacy e Compliance e Strategia richiedano lo sforzo evolutivo più ampio, presentando i divari medi più marcati.

7 Dimensioni IA	Maturità As-Is	Maturità Target	Divario Medio
Strategia	2.3–2.7	4.8–5.0	2.4
Infrastruttura	1.8–2.2	4.3–4.7	2.5
Dati	1.8–2.2	4.8–5.0	2.9
Processi	1.8–2.2	3.8–4.2	2.0
Organizzazione	2.3–2.7	3.8–4.2	1.5
Persone e Competenze	2.3–2.7	3.8–4.2	1.5
Privacy e Compliance	1.8–2.2	4.3–4.7	2.5

Tabella 6.2: Livelli di maturità IA As-Is e Target per le sette dimensioni.

6.2.2 Stratificazione del campione: ecosistemi e natura giuridica

La seconda analisi svolta ha previsto l'utilizzo di *dashboard* create in Power BI, tramite cui è stato possibile esplorare i dati grazie all'uso di filtri applicati alle due stratificazioni presentate precedentemente. Di seguito si presenta una panoramica del livello di maturità IA, declinata per ecosistema di riferimento e in base all'appartenenza o meno alla Pubblica Amministrazione.

Analizzando le differenze tra ecosistemi, emergono polarizzazioni significative mostrate in tabella 6.3.

Gli Enti legati a Innovazione e Ricerca e Informatica e Digitalizzazione registrano le performance più elevate, agendo da pionieri nell'adozione dell'IA. Di contro, ecosistemi tradizionalmente più operativi sul territorio, come Agricoltura e Foreste e Amministrazione, presentano i *gap* più severi.

7 Dimensioni IA	Maturità As-Is per Ecosistema				Maturità Target
	Informatica e Digitalizzazione	Innovazione e Ricerca	Amministrazione	Agricoltura e Foreste	
Strategia	3.8–4.2	2.8–3.2	1.8–2.2	1.8–2.2	4.8–5.0
Infrastruttura	3.8–4.2	2.3–2.7	1.8–2.2	0.8–1.2	4.3–4.7
Dati	3.8–4.2	1.8–2.2	1.3–1.7	0.8–1.2	4.8–5.0
Processi	3.8–4.2	1.8–2.2	1.3–1.7	1.3–1.7	3.8–4.2
Organizzazione	3.8–4.2	2.8–3.2	1.8–2.2	1.3–1.7	3.8–4.2
Persone e Competenze	3.8–4.2	2.8–3.2	2.3–2.7	1.8–2.2	3.8–4.2
Privacy e Compliance	3.8–4.2	2.3–2.7	1.8–2.2	1.3–1.7	4.3–4.7

Tabella 6.3: Livelli di Maturità As-Is per dimensione ed ecosistema confrontati con il livello target territoriale.

Un'ulteriore chiave di lettura è fornita dalla natura giuridica degli Enti. Filtrando i risultati per appertenenza o meno alla Pubblica Amministrazione, i dati evidenziano una maggiore prontezza tecnologica degli Enti non appartenenti alla PA. Nella tabella 6.4 è possibile vedere come gli Enti non-PA ottengano punteggi sistematicamente superiori in tutte le dimensioni, ad eccezione di Privacy e Compliance, dove raggiungono lo stesso range di risultati $(1.8 - 2.2)$. L'analisi potrebbe suggerire la presenza di alcuni vincoli burocratici all'interno della PA che rallentano l'adozione di soluzioni IA su scala strutturale.

7 Dimensioni IA	PA	Non-PA
Strategia	1.8–2.2	2.3–2.7
Infrastruttura	1.3–1.7	2.3–2.7
Dati	1.3–1.7	1.8–2.2
Processi	1.3–1.7	1.8–2.2
Organizzazione	1.8–2.2	2.3–2.7
Persone e Competenze	2.3–2.7	2.8–3.2
Privacy e Compliance	1.8–2.2	1.8–2.2

Tabella 6.4: Confronto dei livelli di maturità IA tra Pubblica Amministrazione e Enti non-PA.

6.3 Esplorazione e segmentazione del campione

6.3.1 PCA e K-means per un'analisi esplorativa

Mentre nella sezione precedente l'analisi si è focalizzata su alcuni partizionamenti derivanti dalla natura intrinseca del Soggetto, in questa fase la ricerca si pone l'obiettivo di far emergere strutture latenti basate direttamente sulle risposte fornite dagli Enti.

Per raggiungere tale scopo, è stata applicata l'Analisi delle Componenti Principali (PCA) in combinazione con l'algoritmo di clustering K-means. La PCA ha permesso di sintetizzare l'informazione contenuta nelle domande del questionario in un numero ridotto di variabili ortogonali. La scelta di limitare l'analisi alle prime due componenti principali è stata guidata principalmente dalla rappresentabilità grafica.

I risultati ottenuti mostrano una netta dominanza della prima componente principale:

- La componente PC1 spiega il 61.88% della varianza totale.
- La componente PC2 spiega il 6.01% della varianza totale.

L'inclusione della seconda componente principale (PC2), pur offrendo un contributo quantitativamente minore rispetto alla prima, risulta preziosa poiché cattura un'ulteriore quota di varianza. Questo arricchisce il contenuto informativo del modello, portando la varianza complessiva spiegata al 67.89%, e offre contestualmente il vantaggio di poter proiettare e interpretare i dati in un intuitivo spazio bidimensionale, come mostrato in Fig. 6.1.

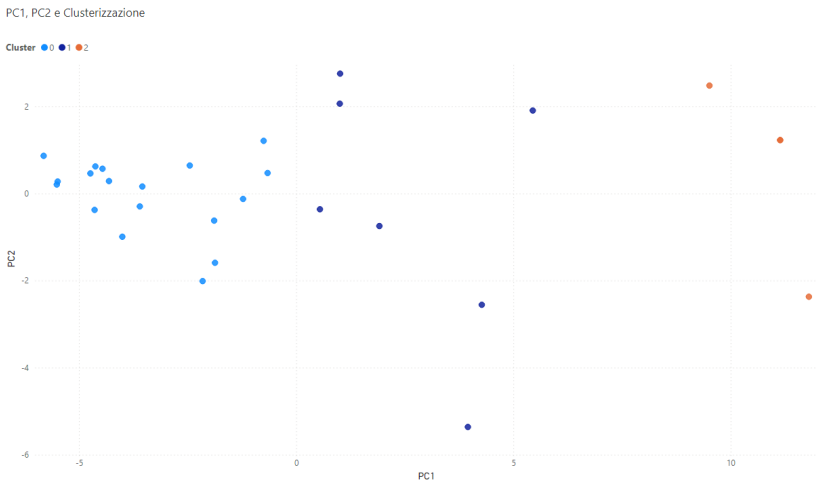


Figura 6.1: Proiezione bidimensionale della PCA con applicazione del K-Means ($k = 3$).

Per determinare il numero k di cluster da utilizzare con l'algoritmo K-means, non essendoci una soluzione intrinsecamente esatta a priori, sono stati confrontati due scenari. Sono state messe a confronto le configurazioni esplorative con $k = 3$ e $k = 4$, valutando il compromesso tra la coesione interna dei gruppi (Inerzia - SSE) e la loro separazione (Silhouette Score), come riportato in tabella 6.5.

Numero di Cluster (k)	Silhouette Score	Inerzia (SSE)
3	0.534	149.56
4	0.487	94.91

Tabella 6.5: Valutazione della qualità dei cluster K-means combinati con PCA per $k = 3$ e $k = 4$.

La scelta metodologica di procedere con la configurazione a 3 cluster non deriva da un vincolo matematico stringente, ma rappresenta il compromesso ottimale tra il rigore statistico e l'interpretabilità:

1. **Qualità della segmentazione:** Il valore di Silhouette più elevato (0.534 contro 0.487) indica che con tre cluster i punti sono meglio assegnati al proprio gruppo e più distanti dai cluster limitrofi. Un passaggio a $k = 4$, pur riducendo l'inerzia, provocherebbe una frammentazione eccessiva del campione, creando gruppi meno distinti tra loro.

2. **Interpretabilità dei risultati:** In un'ottica di analisi della maturità IA degli Enti, una ripartizione in tre livelli (ad esempio: *base, intermedio, avanzato*) risulta più coerente con le finalità della ricerca e permette una descrizione dei profili più nitida, evitando le sovrapposizioni concettuali che una maggiore granularità avrebbe inevitabilmente introdotto.

Nonostante la discreta quota di varianza spiegata dal modello, l'approccio puramente esplorativo della PCA presenta un limite intrinseco sotto il profilo dell'interpretabilità: le componenti principali, in quanto combinazioni lineari di diverse variabili, non sempre permettono una caratterizzazione semantica immediata e univoca dei fattori latenti.

In questa prospettiva, i risultati derivanti dall'applicazione della PCA e dell'algoritmo K-means sono stati utilizzati come test esplorativo preliminare per osservare la propensione del campione alla segmentazione, valutando diverse configurazioni di raggruppamento attraverso il confronto di metriche oggettive.

Per la successiva fase di profilazione, si è deciso di utilizzare il clustering gerarchico applicato direttamente alle sette dimensioni teoriche definite nel *framework*. Tale passaggio garantisce una maggiore aderenza agli obiettivi dello studio, permettendo di caratterizzare i gruppi non più attraverso coordinate matematiche sintetiche (come riscontrato nell'analisi PCA), bensì in relazione ai sette abilitatori fondamentali dotati di una precisa e immediata valenza semantica.

6.3.2 Cluster gerarchico sulle dimensioni teoriche

L'applicazione dell'algoritmo di clustering gerarchico con metodo di Ward ha consentito di segmentare il campione agendo sui punteggi medi delle sette dimensioni già note. Il metodo, basandosi sulla minimizzazione della varianza all'interno dei gruppi, favorisce la creazione di cluster particolarmente omogenei, garantendo la piena interpretabilità dei profili emergenti.

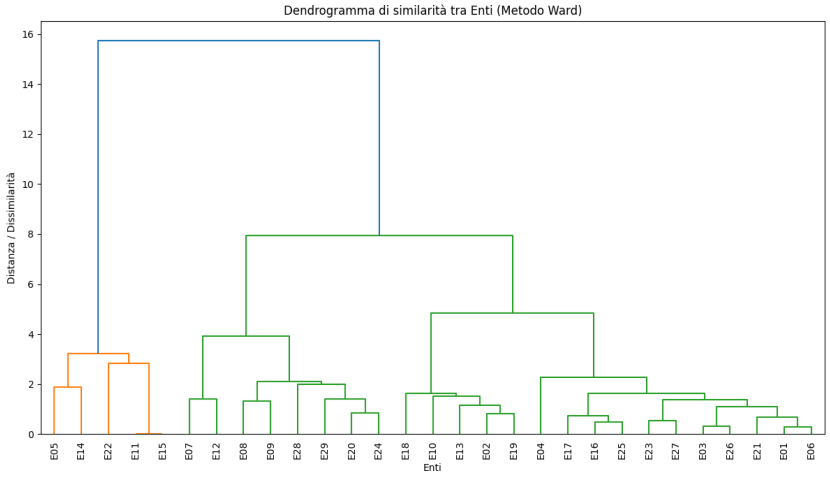


Figura 6.2: Dendrogramma generato tramite cluster gerarchico (metodo di Ward).

L'analisi visiva del dendrogramma in Fig. 6.2 permette di identificare la struttura dei legami tra gli Enti. Il "salto" di dissimilarità più significativo si registra nella parte superiore del grafico, suggerendo che una divisione del campione in tre raggruppamenti rappresenti una soluzione robusta. Tale partizione permette di mappare efficacemente l'eterogeneità del campione, riconducendola a tre macro-profili distinti per maturità e attitudine rispetto agli abilitatori indagati.

6.3.3 Profilazione dei cluster

Nella prima fase esplorativa, la valutazione del livello di maturità IA degli Enti è stata condotta attraverso l'analisi della media campionaria delle sette dimensioni del modello. Tale indicatore fornisce una sintesi immediata del fenomeno osservato, ma dal punto di vista statistico può risultare poco rappresentativo qualora la distribuzione dei dati presenti elevata varianza o una struttura multimodale. In tali contesti, il valore medio tende infatti a mascherare la presenza di sottogruppi caratterizzati da comportamenti significativamente differenti.

L'applicazione del clustering gerarchico ha consentito di superare questo limite analitico, permettendo di decomporre la variabilità del campione e individuare gruppi omogenei di Enti con caratteristiche strutturalmente simili. I risultati mostrano come il valore medio globale osservato in precedenza sia in realtà il risultato della sovrapposizione di tre popolazioni organizzative nettamente distinte in termini di maturità IA.

Per interpretare e caratterizzare tali gruppi, è stata effettuata una profilazione dei cluster attraverso l'analisi dei loro centroidi, ovvero i vettori costituiti dai valori medi delle sette dimensioni considerate nel *framework* di maturità.

Dimensioni	Cluster 1	Cluster 2	Cluster 3
Strategia	3.8–4.2	2.8–3.2	1.3–1.7
Infrastruttura	3.8–4.2	2.3–2.7	0.8–1.2
Dati	3.8–4.2	1.8–2.2	1.3–1.7
Processi	3.3–3.7	1.8–2.2	1.3–1.7
Organizzazione	3.8–4.2	2.3–2.7	1.3–1.7
Persone e Competenze	3.8–4.2	2.8–3.2	1.8–2.2
Privacy e Compliance	3.3–3.7	1.8–2.2	1.3–1.7

Tabella 6.6: Maturità As-Is delle sette dimensioni per ogni cluster.

L'analisi comparativa dei raggruppamenti, illustrata in tabella 6.6, permette di delineare le eccellenze e le criticità che caratterizzano ciascun gruppo rispetto ai sette abilitatori:

Cluster 1: Enti leader nell'intelligenza artificiale. Il primo cluster rappresenta il gruppo più avanzato del campione. Gli Enti appartenenti a questa categoria presentano livelli di maturità IA elevati e relativamente omogenei su tutte le dimensioni considerate, con valori nel range $3.8-4.2$ per quasi tutta la totalità delle dimensioni. Risultati leggermente inferiori, nel range $3.3-3.7$, si misurano nelle dimensioni Processi e Privacy e Compliance, restando comunque ben oltre la media territoriale.

Dal punto di vista interpretativo, questo profilo suggerisce la presenza di organizzazioni in cui la trasformazione digitale è stata internalizzata come processo sistemico. L'innovazione non si manifesta come intervento isolato, ma come risultato di un allineamento strutturale tra strategia, infrastruttura tecnologica, gestione dei dati e capitale umano.

Cluster 2: Enti in fase di transizione. Il secondo cluster identifica una classe intermedia di Enti che possono essere interpretati come in fase di transizione verso livelli più elevati di maturità IA. I valori medi spaziano tra 1.8 e 3.2, evidenziando un livello di sviluppo parziale e non ancora pienamente strutturato.

Un elemento particolarmente significativo riguarda il divario tra le dimensioni Strategia e Persone e Competenze, che mostrano valori

relativamente più elevati, nel range $2.8 - 3.2$, e le dimensioni tecnologiche e operative, come Dati, Processi e Privacy e Compliance, che presentano livelli sensibilmente inferiori, nel range $1.8 - 2.2$.

Questo squilibrio suggerisce che, in tali organizzazioni, esista una consapevolezza strategica della trasformazione in relazione all'IA e una certa disponibilità di competenze, ma che tali risorse non siano ancora pienamente supportate da infrastrutture tecnologiche e modelli organizzativi adeguati. In altre parole, la visione dell'innovazione riguardo all'IA appare superiore rispetto alla capacità effettiva di implementazione.

Cluster 3: Enti nelle fasi iniziali. Il terzo cluster raggruppa gli Enti caratterizzati dai livelli di maturità più bassi dell'intero campione. I valori medi risultano frequentemente prossimi al livello minimo della scala di valutazione, con punte particolarmente ridotte nella dimensione Infrastruttura, la quale si colloca nel range $0.8 - 1.2$.

In questo contesto emerge tuttavia un altro elemento interessante: anche negli Enti meno maturi, la dimensione Persone e Competenze, che presenta valori nel range $1.8 - 2.2$, risulta relativamente superiore rispetto agli altri abilitatori. Ciò suggerisce che il capitale umano possa rappresentare un fattore di resilienza organizzativa, in grado di sostenere percorsi di sviluppo anche in presenza di limitate risorse tecnologiche e di una scarsa strutturazione dei processi digitali.

Validazione della struttura di clustering Per valutare la robustezza della partizione ottenuta è stato calcolato il Silhouette Score nello spazio delle sette dimensioni del modello, ottenendo un valore pari a 0.395.

In questo contesto di analisi, un valore positivo prossimo a 0.4 risulta indicativo di una struttura di raggruppamento coerente, evidenziando livelli di separazione e coesione interna sufficienti a validare la distinzione dei tre profili. La lieve riduzione rispetto al valore osservato nello spazio delle componenti principali può essere attribuita alla maggiore complessità del dataset originale: mentre l'analisi delle componenti principali tende a ridurre il rumore statistico concentrando l'informazione nelle dimensioni più rilevanti, l'analisi nello spazio completo degli abilitatori preserva l'intera variabilità del fenomeno osservato.

In un contesto di ricerca basato su un questionario e su valutazioni percettive delle organizzazioni, un valore di Silhouette prossimo a 0.4 può essere considerato indicativo di una segmentazione coerente e metodologicamente solida, senza ricorrere a forzature algoritmiche nella definizione dei gruppi.

Dispersione *intra-cluster* L'analisi della dispersione *intra-cluster*, misurata attraverso la deviazione standard media delle dimensioni, evidenzia valori pari a 0.61 per il Cluster 1, 0.52 per il Cluster 2 e 0.41 per il Cluster 3, come risportato in tabella 6.7.

Cluster	Deviazione standard media
Cluster 1	0.61
Cluster 2	0.52
Cluster 3	0.41

Tabella 6.7: Dispersione *intra-cluster* misurata tramite deviazione standard media delle dimensioni.

Il risultato indica che il Cluster 3, caratterizzato dai livelli di maturità più bassi, presenta anche la maggiore omogeneità interna, mentre gli Enti più avanzati mostrano una maggiore variabilità nei profili dimensionali. Questo fenomeno suggerisce che, mentre i livelli iniziali di maturità tendono a configurarsi secondo modelli organizzativi relativamente simili, gli Enti più evoluti possono raggiungere livelli elevati di sviluppo in materia di IA, attraverso percorsi differenziati di integrazione tra strategia, infrastrutture, dati, organizzazione e competenze.

Distanza tra i centroidi Il calcolo della distanza euclidea tra i centroidi dei cluster conferma la separazione progressiva tra i gruppi identificati. Come evidenziato in tabella 6.8, la distanza maggiore si osserva tra il Cluster 1 e il Cluster 3 con una distanza pari a 6.23, mentre il Cluster 2 si colloca in posizione intermedia, con una distanza pari a 3.83 dal Cluster 1 e 2.46 dal Cluster 3.

Questo risultato risalta la struttura graduale dei livelli di maturità IA: il Cluster 2 rappresenta una configurazione di transizione tra gli Enti a maturità iniziale e quelli caratterizzati da livelli più avanzati di sviluppo.

	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0.00	3.83	6.23
Cluster 2	3.83	0.00	2.46
Cluster 3	6.23	2.46	0.00

Tabella 6.8: Matrice delle distanze euclidee tra i centroidi dei cluster.

Il confronto grafico tra i vettori dei centroidi dei cluster per i sette abilitatori è illustrato nel *radar chart* in Fig. 6.3. I tre profili assumono una configurazione annidata che delinea una gerarchia netta: le performance di

un gruppo superano sistematicamente quelle del gruppo di livello inferiore su ciascun dominio, senza presentare intersezioni grafiche. Un ulteriore elemento rilevante è la quasi linearità della progressione tra i cluster. I profili risultano infatti sostanzialmente paralleli, confermando la totale assenza di inversioni di tendenza. Questa dinamica suggerisce che gli abilitatori della maturità IA tendano a evolvere in modo sincrono e interdipendente, avvalorando l'ipotesi secondo cui il processo di trasformazione organizzativa possieda una natura intrinsecamente sistemica.

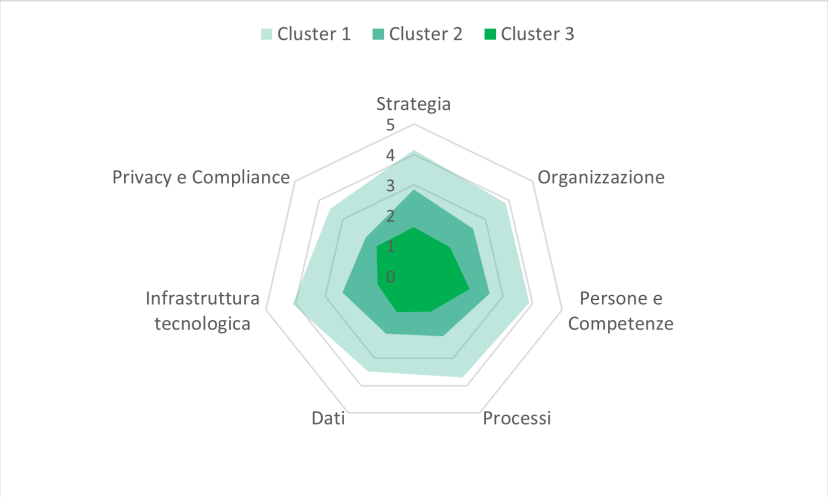


Figura 6.3: Profilo dei centroidi dei cluster nelle sette dimensioni di maturità.

6.4 Sintesi delle evidenze e considerazioni conclusive

Al termine dell'analisi quantitativa descritta nel presente capitolo, emergono alcune evidenze fondamentali che definiscono il reale stato di Maturità As-Is del Soggetto indagato.

In primo luogo, l'integrazione tra le metriche aggregate e i modelli di clustering ha permesso di decostruire la visione parziale data dalla media campionaria. Il livello medio di maturità, stimato nel range $1.8 - 2.2$, maschera in realtà un ecosistema profondamente frammentato e caratterizzato da un marcato *digital divide*. Il Territorio non si muove a una velocità uniforme verso l'obiettivo target, ma risulta polarizzato in tre macro-archetipi organizzativi: le eccellenze sistemiche (Cluster 1), le realtà in fase di transizione frenate da limiti infrastrutturali (Cluster 2) e gli Enti ancora in fase embrionale (Cluster 3).

In secondo luogo, l'osservazione dei centroidi e la conformazione parallela dei profili nel *radar chart* (Fig. 6.3) confermano inequivocabilmente la natura sistemica dell'innovazione. Nessun Ente registra picchi isolati in singole dimensioni tecnologiche senza un corrispettivo sviluppo strategico e organizzativo; l'adozione dell'IA, per essere scalabile, richiede un'evoluzione sincrona di tutti gli abilitatori.

Infine, emerge un dato di grande rilevanza sociologica e gestionale: anche negli ecosistemi più arretrati e carenti dal punto di vista dell'infrastruttura e della gestione dei dati, la dimensione Persone e Competenze si mantiene costantemente come il valore più alto.

Alla luce di queste evidenze empiriche, risulta chiaro che eventuali politiche di supporto all'adozione dell'intelligenza artificiale non potranno assumere una logica standardizzata. Per colmare il divario evidenziato dalla *gap analysis*, sarà necessario modulare gli interventi in base al reale profilo di appartenenza degli Enti: fornendo alfabetizzazione digitale e infrastrutture di base ai soggetti ritardatari, colmando il *gap* tecnologico-architetturale della fascia intermedia e permettendo ai leader di agire come catalizzatori per la condivisione di *best practice* sull'intero Territorio.

Appendice A

Appendice A: Codice Python

Script Python per la costruzione della pipeline di analisi della maturità IA: data preparation, analisi esplorativa (PCA e K-Means) e modello finale di clustering gerarchico basato su dimensioni teoriche.

```
"""
APPENDICE A: Script Python per l'Analisi di Maturita' As-Is e Clustering
Il seguente script esegue la pipeline completa di Data Preparation,
l'analisi esplorativa (PCA + K-Means) e il modello definitivo
(Riduzione dimensionale teorica + Clustering Gerarchico di Ward).
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.cluster.hierarchy as sch

from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, pairwise_distances
from scipy.spatial.distance import pdist, squareform

# ============================================================
# 1. CARICAMENTO DATI
# ============================================================
print("="*60)
print("1. CARICAMENTO DATI")
print("="*60)

try:
    df_survey = pd.read_excel("Dataset_Anonimo_Self_Assessment.xlsx")
    df_map = pd.read_excel("Mapping_Score.xlsx")
    print("Dati caricati con successo.")
except FileNotFoundError as e:
    print(f"ERRORE: File non trovato. Assicurati che i file Excel
    siano nella stessa cartella dello script.\nDettagli: {e}")
    exit()

# ============================================================
# 2. CREAZIONE DEL DIZIONARIO DI CODIFICA
# ============================================================
mapping_dict = {}
```

```
39
40 for index, row in df_map.iterrows():
41     domanda = str(row['Domanda']).strip()
42     risposta = str(row['Risposta']).strip()
43     score = row['Score']
44     mapping_dict[(domanda, risposta)] = score
45
46 # ============================================================
47 # 3. APPLICAZIONE DELLA CODIFICA E DATA QUALITY CHECK
48 # ============================================================
49 df_numeric = df_survey.copy()
50 colonne_domande = df_survey.columns[1:]
51 TOTALE_DOMANDE = len(colonne_domande)
52
53 errori_per_ente = {str(nome): 0 for nome in df_survey.iloc[:, 0]}
54 VALORE_DEFAULT = 1
55
56 print(f"\nInizio conversione su {TOTALE_DOMANDE}
57         domande per {len(df_survey)} Enti...")
58
59 for col in colonne_domande:
60     for i in df_survey.index:
61         nome_ente = str(df_survey.iloc[i, 0])
62         risposta_txt = str(df_survey.at[i, col]).strip()
63         key = (col.strip(), risposta_txt)
64
65         if key in mapping_dict:
66             df_numeric.at[i, col] = mapping_dict[key]
67         else:
68             df_numeric.at[i, col] = VALORE_DEFAULT
69             errori_per_ente[nome_ente] += 1
70
71 # --- REPORT COPERTURA ---
72 print("\n" + "="*60)
73 print(f"REPORT QUALITA DATI (Totale domande: {TOTALE_DOMANDE})")
74 print("="*60)
75 print(f"{'ENTE':<40} | {'MISSING':<8} | {'COPERTURA %'}")
76 print("-" * 65)
77
78 enti_problematici = []
79
80 for ente, num_errori in errori_per_ente.items():
81     risposte_ok = TOTALE_DOMANDE - num_errori
82     copertura_perc = (risposte_ok / TOTALE_DOMANDE) * 100
83     print(f"{ente[:40]:<40} | {num_errori:<8} | {copertura_perc:.1f}%")
84
85     if copertura_perc < 80.0:
86         enti_problematici.append(ente)
87
88 print("-" * 65)
89 if enti_problematici:
90     print(f"\nATTENZIONE:
91             I seguenti Enti hanno scarsa qualita' dei dati (<80%):")
92     for e in enti_problematici:
93         print(f"- {e}")
94 else:
95     print("\nOttimo!Tutti gli Enti hanno una copertura sufficiente (>80%).")
96
97 # ============================================================
98 # 4. ANALISI ESPLORATIVA: PCA E K-MEANS
99 # ============================================================
100 print("\n" + "="*60)
```

```
print("4. ANALISI ESPLORATIVA (PCA + K-MEANS)")
print("="*60)

enti = df_numeric.iloc[:, 0].values
X = df_numeric.iloc[:, 1:].values

# Standardizzazione
scaler_pca = StandardScaler()
X_std = scaler_pca.fit_transform(X)

# PCA
pca = PCA(n_components=2)
principal_components = pca.fit_transform(X_std)
var_spiegata = pca.explained_variance_ratio_

print(f"Risultati PCA:")
print(f"- PC1 spiega il {var_spiegata[0]*100:.2f}% della varianza")
print(f"- PC2 spiega il {var_spiegata[1]*100:.2f}% della varianza")

# Export PCA
df_pca = pd.DataFrame(data=principal_components, columns=['PC1', 'PC2'])
df_pca.insert(0, 'Codice_Ente', enti)
df_pca.to_excel("Output_Pca_Powerbi.xlsx", index=False)
print("\n> Salvato: Output_Pca_Powerbi.xlsx")

# K-Means
K_MEANS_CLUSTERS = 3
kmeans = KMeans(n_clusters=K_MEANS_CLUSTERS, random_state=42, n_init=20)
cluster_labels_kmeans = kmeans.fit_predict(principal_components)

df_cluster_kmeans = df_pca.copy()
df_cluster_kmeans['Cluster'] = cluster_labels_kmeans
df_cluster_kmeans.to_excel("Output_Cluster_Kmeans.xlsx", index=False)
print("> Salvato: Output_Cluster_Kmeans.xlsx")

# Valutazione K-Means
sil_score_kmeans = silhouette_score(
    principal_components,
    cluster_labels_kmeans
)
inerzia = kmeans.inertia_
print(f"\nValutazione K-Means (K={K_MEANS_CLUSTERS}):")
print(f"- Silhouette Score: {sil_score_kmeans:.3f}")
print(f"- Inerzia (SSE): {inerzia:.2f}")

# ============================================================
# 5. SCORING: AGGREGAZIONE SULLE DIMENSIONI (La matrice X')
# ============================================================
print("\n" + "="*60)
print("5. SCORING SULLE 7 DIMENSIONI TEORICHE")
print("="*60)

mappatura_dimensioni = {}
df_unique_mapping = df_map[['Dimensione', 'Domanda']].drop_duplicates()

for index, row in df_unique_mapping.iterrows():
    dimensione = str(row['Dimensione']).strip()
    domanda = str(row['Domanda']).strip()
    if dimensione not in mappatura_dimensioni:
        mappatura_dimensioni[dimensione] = []
    mappatura_dimensioni[dimensione].append(domanda)

```

```
print(f"Trovate {len(mappatura_dimensioni)} dimensioni.")

df_dimensioni = pd.DataFrame()
df_dimensioni['Nome_Ente'] = df_numeric.iloc[:, 0]

for dim, col_domande in mappatura_dimensioni.items():
    col_valide = [col for col in col_domande if col in df_numeric.columns]
    if col_valide:
        df_dimensioni[dim]=df_numeric[col_valide].astype(float).mean(axis=1)
    else:
        print(f"Attenzione:
                Nessuna domanda valida trovata per la dimensione '{dim}'")

print("\nPrime 5 righe della nuova matrice X' (Dimensioni):")
print(df_dimensioni.head())

# ============================================================
# 6. CLUSTERING GERARCHICO
# ============================================================
print("\n" + "="*60)
print("6. CLUSTERING GERARCHICO (METODO WARD)")
print("="*60)

colonne_dim = df_dimensioni.columns[1:]

# Standardizzazione delle dimensioni (Matrice Z)
scaler_hier = StandardScaler()
dati_scalati_hier = scaler_hier.fit_transform(df_dimensioni[colonne_dim])

df_scalato_hier = pd.DataFrame(
    dati_scalati_hier,
    columns=colonne_dim,
    index=df_dimensioni['Nome_Ente']
)

# Calcolo linkage (Matrice delle distanze M e fusioni)
Z_hier = sch.linkage(df_scalato_hier, method='ward')

# Dendrogramma
print("Generazione Dendrogramma...")
plt.figure(figsize=(12, 7))
plt.title("Dendrogramma di similarita tra Enti (Metodo Ward)")
plt.xlabel("Enti")
plt.ylabel("Distanza Euclidea / Dissimilarita")
sch.dendrogram(
    Z_hier,
    labels=df_scalato_hier.index.values,
    leaf_rotation=90,
    leaf_font_size=10
)
plt.tight_layout()
plt.show()

# Assegnazione Cluster
NUM_CLUSTERS_HIER = 3
etichette_cluster_hier = sch.fcluster(
    Z_hier,
    t=NUM_CLUSTERS_HIER,
    criterion='maxclust'
)

df_finale = df_dimensioni.copy()
```

```
df_finale['Cluster_Assegnato'] = etichette_cluster_hier

print(f"\nAssegnazione in {NUM_CLUSTERS_HIER} cluster completata.")
df_finale.to_excel("Risultati_Clustering_Enti.xlsx", index=False)
print("> Salvato: Risultati_Clustering_Enti.xlsx")

# ============================================================
# 7. PROFILAZIONE E VALUTAZIONE CLUSTER GERARCHICO
# ============================================================
print("\n" + "="*60)
print("7. PROFILAZIONE E VALUTAZIONE (GERARCHICO)")
print("="*60)

# Dispersione Intra-Cluster (Deviazione Standard media per cluster)
print("Dispersione Intra-Cluster (Std Dev media):")
dispersione_media = (
    df_finale.groupby('Cluster_Assegnato')[colonne_dim]
    .std()
    .mean(axis=1)
)
print(dispersione_media)

# Calcolo Centroidi
centroidi = df_finale.groupby('Cluster_Assegnato')[colonne_dim].mean()
print("\nCentroidi dei Cluster:")
print(centroidi)
centroidi.to_excel("Centroidi_Cluster.xlsx")
print("> Salvato: Centroidi_Cluster.xlsx")

# Matrice delle distanze tra i centroidi
print("\nMatrice delle Distanze Euclidee tra i Centroidi (Separazione):")
distanze_centroidi = pd.DataFrame(
    squareform(pdist(centroidi, metric='euclidean')),
    index=centroidi.index,
    columns=centroidi.index
)
print(distanze_centroidi)

# Valutazione Statistica
sil_score_hier = silhouette_score(df_scalato_hier, etichette_cluster_hier)
print("\n" + "-"*40)
print(
    f"Silhouette Score (Gerarchico K={NUM_CLUSTERS_HIER}): "
    f"{sil_score_hier:.3f}"
)
print("(Valori > 0 indicano buona separazione, verso l'1 ottimali)")
print("-"*40)

print("\nProcesso completato con successo")
```

Listing A.1: Script Python per data preparation, analisi esplorativa (PCA e K-Means) e modello di clustering gerarchico basato su dimensioni teoriche

Bibliografia

Agenzia per l'Italia Digitale (2025). *Linee guida per l'adozione dell'Intelligenza Artificiale nella Pubblica Amministrazione*. Consultazione pubblica. Bozza in consultazione pubblica. URL: https://www.agid.gov.it/sites/agid/files/2025-02/Linee_Guida_adozione_IA_nella_PA.pdf.

Nolan Norton Italia (2024). *Presentazione Nolan Norton Italia*. Company presentation. URL: https://www.nolannorton.it/wp-content/uploads/2023/09/Presentazione-NOLAN-NORTON-WEB-SETT-2024.pdf.

Nolan Norton Italia (2025). *AI Maturity Model*. White paper. URL: https://kdocs.kpmg.it/Marketing_Coms/NolanNorton/AI-Strategy&Governance.pdf.

Parlamento Europeo e Consiglio dell'Unione Europea (2016). *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 e alla libera circolazione di tali dati (General Data Protection Regulation)*. Gazzetta ufficiale dell'Unione europea. L 119, 4 maggio 2016. URL: https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32016R0679.

Parlamento Europeo e Consiglio dell'Unione Europea (2024). *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce norme armonizzate sull'intelligenza artificiale (Artificial Intelligence Act)*. Gazzetta ufficiale dell'Unione europea. L 1689, 12 luglio 2024. URL: https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32024R1689.

Repubblica Italiana (2025). *Legge 23 settembre 2025, n. 132. Disposizioni e deleghe al Governo in materia di intelligenza artificiale*. Gazzetta

Ufficiale della Repubblica Italiana, Serie Generale n. 223 del 25 settembre 2025. Entrata in vigore: 10 ottobre 2025. URL: https://www.gazzettaufficiale.it/eli/id/2025/09/25/25G00143/sg.

Tan, Pang-Ning, Michael Steinbach e Vipin Kumar (2016). *Introduction to data mining*. Pearson Education India.