



Università  
Ca' Foscari  
Venezia

Corso di Laurea  
Magistrale

in Economia e  
Gestione delle Arti e  
delle Attività Culturali

Tesi di Laurea

**Machine Learning, Deep Neural Networks e IA generativa  
nel settore delle Digital Humanities**

**Relatore**

Ch. Prof. Sebastiano Vascon

**Correlatrice**

Dott.ssa Marina Khoroshiltseva

**Laureanda**

Gaia Pavan

Matricola 886862

**Anno Accademico**

2023 / 2024

# Machine Learning, Deep Neural Networks e IA generativa nel settore delle Digital Humanities

## Sommario

<b>Introduzione</b> .....	2
<b>Capitolo 1 - Definizioni e cenni storici</b> .....	5
1.1 Cenni storici .....	11
1.2 Gli anni '70 e '80 e il consolidamento delle Digital Humanities.....	13
1.3 Machine Learning: tipologie.....	16
1.4 Best practices per la costruzione di un modello di machine learning da impiegare nel settore culturale .....	19
<b>Capitolo 2 - Il Machine Learning Supervisionato</b> .....	21
2.1 Applicazioni del machine learning supervisionato al settore culturale.....	23
2.1.1 La regressione lineare e la regressione logistica.....	23
2.1.2 Support Vector Machines .....	28
2.1.3 Decision Trees e random forests .....	40
<b>Capitolo 3 - Machine Learning non supervisionato</b> .....	45
3.1 Il clustering.....	46
3.2 Impieghi dell'apprendimento non supervisionato nel settore culturale e umanistico .....	48
<b>Capitolo 4 - Deep neural networks e IA generativa in ambito culturale</b> .....	62
4.1 Applicazioni in ambito culturale: convolutional neural network e generative adversarial network.....	68
<b>Capitolo 5 - Applicazioni del Machine Learning alla Gestione Culturale</b> .....	86
<b>Conclusioni</b> .....	105
<b>Bibliografia</b> .....	108

## Introduzione

Il presente elaborato si propone di illustrare le possibilità di applicazione delle tecnologie di machine learning nel settore delle digital humanities, allo scopo di comprenderne le potenzialità e dimostrarne l'efficacia nel supportare il lavoro degli operatori del settore culturale. Il machine learning, una branca dell'intelligenza artificiale, si concentra sull'addestramento di modelli statistici per migliorare le loro prestazioni su compiti specifici attraverso l'apprendimento dai dati. Queste tecniche vengono utilizzate per creare grandi dataset, estrarne informazioni e utilizzarle per evidenziare pattern e tendenze. Negli ultimi anni il machine learning ha trovato applicazioni crescenti nel settore culturale, quali l'analisi di testi storici, la classificazione di opere artistiche, la previsione del posizionamento di siti archeologici e la gestione di collezioni museali digitalizzate. Nel 2020 un briefing redatto per il think tank del parlamento europeo evidenziava come l'impiego crescente di sistemi di intelligenza artificiale nei settori culturali e creativi permetta di aumentare le possibilità di conservazione e fruizione del patrimonio, fornendo il potenziale per creare modi innovativi per navigare nei contenuti culturali (Caramiaux, 2020). Questa tesi, fornendo una panoramica dei sistemi di machine learning applicati maggiormente nel settore culturale, ha come obiettivo quello di contribuire alla comprensione e all'avanzamento dell'uso delle tecnologie computazionali nel campo delle scienze umane e culturali, osservandone soprattutto le applicazioni a supporto della ricerca e della gestione di organizzazioni culturali. In particolare, il desiderio è di evidenziare come, in molti dei casi studio riportati, l'impiego di intelligenza artificiale si sia dimostrato un mezzo per permettere di snellire il lavoro del ricercatore, che non dovendo impiegare grandi quantità di ore a raccogliere dati avrebbe la possibilità di concentrarsi sul lavoro di analisi velocizzando la fase di preparazione dei materiali di studio (delegando, ad esempio, attività di catalogazione e identificazione a sistemi automatizzati). La motivazione alla base di un elaborato di questo tipo è il desiderio di approfondire il tema dell'integrazione del machine learning in un'area di studio altamente interdisciplinare come le digital humanities, evidenziando come il machine learning possa essere utilizzato per lo sviluppo di nuove tecniche che possono essere applicate in modo più efficiente e accurato rispetto ai metodi tradizionali. Una tesi che

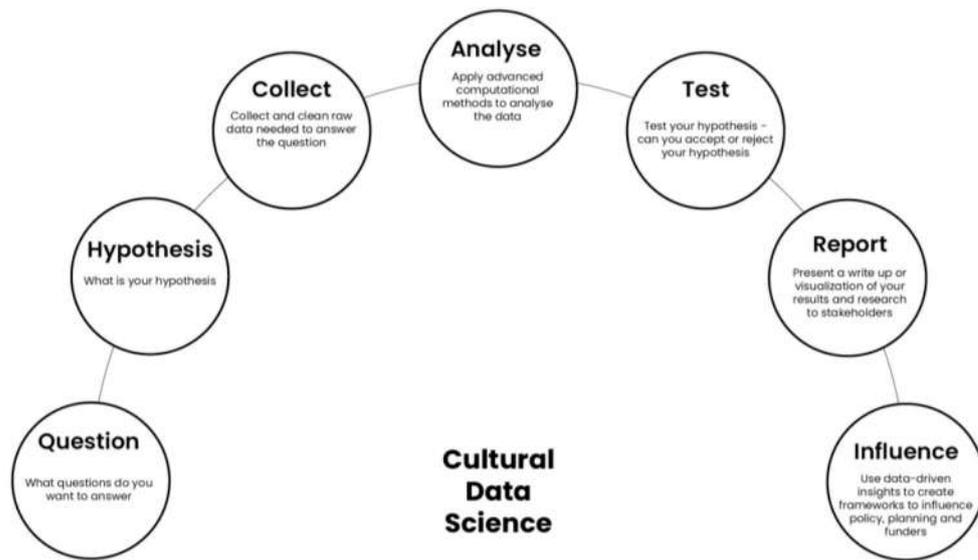
fornisce una panoramica su questo argomento può aprire nuove linee di ricerca, contribuendo a una migliore comprensione delle applicazioni del machine learning in contesti non convenzionali e ampliando così l'orizzonte scientifico del campo. Le tecnologie basate sul machine learning possono essere utilizzate per personalizzare l'esperienza dei visitatori, offrendo contenuti e percorsi su misura, migliorando così la soddisfazione e l'engagement del pubblico e facilitando l'accesso al patrimonio culturale a una più ampia gamma di persone, compresi individui con disabilità, grazie a strumenti la realtà virtuale e adattamenti personalizzati: inoltre lo studio di applicazioni del machine learning specifiche per il settore culturale, come il riconoscimento automatico delle opere d'arte, la previsione del deterioramento dei beni culturali o la creazione di tour virtuali personalizzati apre le porte ad una sempre maggiore collaborazione tra settori. Studiare come integrare tecnologie emergenti come le reti neurali convoluzionali, le reti generative avversarie e altri algoritmi avanzati di machine learning nel contesto culturale può portare a innovazioni significative che trasformano il modo in cui il patrimonio culturale viene conservato, gestito e presentato. L'interesse è evidenziare le sinergie tra il progresso tecnologico e la valorizzazione culturale, mostrando come l'adozione di tecnologie avanzate possa arricchire e proteggere il patrimonio culturale, garantendone la fruizione alle generazioni future. In sostanza, la scrittura di una tesi sull'utilizzo del machine learning a supporto del settore culturale è motivata dalla possibilità di proporre un contributo scientifico con un impatto pratico sull'innovazione tecnologica, creando nuove opportunità per la conservazione e la valorizzazione del patrimonio culturale. Per poter analizzare in dettaglio i diversi sistemi di machine learning il presente elaborato si divide in sei capitoli. Il primo capitolo fornisce definizioni e dei cenni storici, introducendo i concetti chiave dell'intelligenza artificiale, della data science e del cultural analytics, oltre che del concetto di cultural data science elaborato da Vishal Kumar nel 2020. Si discute inoltre dell'evoluzione storica dell'integrazione tra settore umanistico e digitale, delineando le principali tappe e sviluppi che hanno portato alla nascita delle Digital Humanities. Il secondo capitolo si concentra sul machine learning supervisionato, esplorando i principali algoritmi e le loro applicazioni nel settore culturale. Vengono presentati in particolare esempi di utilizzo di regressione logistica, support vector machines e decision trees per l'analisi e l'interpretazione di dati culturali.

Nel terzo capitolo viene indagato l'uso del machine learning non supervisionato nel settore culturale e umanistico, in particolare l'uso di sistemi di clustering per la classificazione di immagini di opere d'arte e manufatti archeologici, e il loro impiego nel campo della fotogrammetria. Nel quarto capitolo vengono invece definite e illustrate le reti neurali profonde, e in particolare vengono proposti casi studio che permettono di osservare l'uso di reti neurali convoluzionali e di reti generative avversarie per supportare attività di conservazione e restauro. Il quinto capitolo non si concentra su uno specifico modello di machine learning, ma attraverso alcuni casi studio indaga l'impiego di questi sistemi allo scopo di supportare la gestione di un'organizzazione culturale, in particolare monitorando il comportamento del pubblico al suo interno e sviluppando sistemi di realtà virtuale per implementare e rendere più accattivante la fruizione culturale. Il sesto capitolo conclude questo elaborato tracciando il percorso fatto ed evidenziando vantaggi e problematiche nell'impiego di sistemi di intelligenza artificiale emerse nella redazione di questa survey.

## Capitolo 1 – Definizioni e cenni storici

Per iniziare a parlare di machine learning è importante partire dal concetto più ampio di intelligenza artificiale. L'intelligenza artificiale è una disciplina che studia le possibilità di riproduzione dei meccanismi che governano i processi mentali umani più complessi. La ricerca ha lo scopo di avvicinare il funzionamento dei computer alle capacità dell'intelligenza umana tramite l'acquisizione e l'elaborazione di dati che vengono trasformati in informazioni, ambienti hardware e software idonei e la costruzione di un modello composto da obiettivi e vincoli del sistema. La scienza dei dati, comunemente conosciuta come data science, è strettamente correlata all'intelligenza artificiale, anche se non ne è una diretta branca. Si tratta dello studio dei dati con l'obiettivo di estrarne informazioni dettagliate e si integra spesso con l'IA per affrontare una vasta gamma di problemi complessi. È un approccio multidisciplinare che combina principi e pratiche nei campi della matematica, statistica, ingegneria informatica e, appunto, intelligenza artificiale. Si tratta di un settore relativamente giovane, che deriva dai campi dell'analisi statistica e del data mining, pratica utilizzata per esplorare ed elaborare grandi insiemi di dati allo scopo di individuarvi schemi o relazioni (Aiello Carlucci, 2015). Secondo Vishal Kumar, tra i primi data scientist a lavorare per la casa d'aste londinese Sotheby's, ora insegnante al Sotheby's Institute of Art, prima di parlare di cultural data science è importante soffermarsi sul concetto di cultural analytics. Le due discipline infatti, per quanto simili, hanno obiettivi diversi (Kumar, 2020). Cultural analytics è una disciplina la cui definizione viene formulata nel 2005 da Lev Manovich in riferimento all'analisi di grandi dataset contenenti dati culturali tramite l'utilizzo di tecniche computazionali e di visualizzazione dei dati. Nel suo saggio del 2017, Manovich descrive la cultural analytics come l'unione dell'analisi di due tipologie di dati: i dati che descrivono le attività degli utenti sui principali social network, e i dati riguardanti specifiche aree culturali e storiche. Questo tipo di analisi viene portato avanti in due settori principali, social computing e digital humanities (Manovich, 2016). In sintesi, la data science ha impatti e conseguenze sul mondo reale, perché si occupa dell'indagine di dati allo scopo di architettare soluzioni di governance e di supportare il decision making; la cultural analytics invece si limita ad analizzare dati per potervi individuare schemi, trends e patterns (Manovich, 2016).

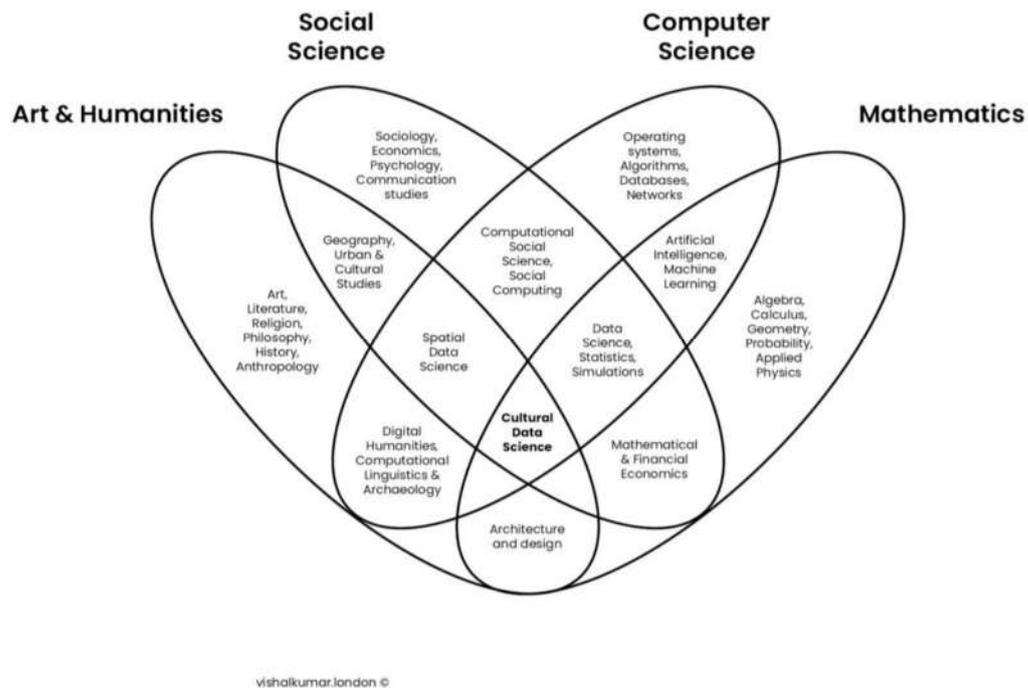
Sebbene gli impieghi della data science siano molteplici, al momento non vi è un connubio forte tra disciplina e settore culturale come per altre tipologie di settori; Vishal Kumar è in questo senso pioniere. Nel 2020 ci fornisce la prima definizione di cultural data science in occasione della conferenza “Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world”, a Dubai. Secondo Kumar, la cultural data science è un “termine ombrello” usato per descrivere l’intero complesso, stratificato processo usato per estrarre valore da dati culturali che hanno implicazioni nel mondo reale per influenzare le politiche, dinamiche di mercato e processi decisionali nelle industrie culturali e creative. Secondo Kumar, il processo seguito prevede sette fasi, illustrate nella figura 1. Per iniziare un processo di ricerca utilizzando la cultural data science, è necessario prima di tutto formulare una domanda chiara alla quale si intende rispondere. Questa domanda rappresenta il punto di partenza e la guida per tutto il processo. Una volta che la domanda è stata definita, è importante stabilire<sup>1</sup> un’ipotesi, ossia una supposizione iniziale su ciò che si crede possa essere la risposta alla domanda posta. Successivamente, è necessario raccogliere e pulire i dati grezzi necessari per rispondere alla domanda. La qualità dei dati è cruciale, quindi questo passaggio richiede cura e attenzione. Una volta che i dati sono stati raccolti e preparati, è possibile applicare metodi computazionali avanzati per analizzarli in dettaglio. Questa fase di analisi è fondamentale per trarre informazioni significative dai dati. Dopo l’analisi, è importante mettere alla prova l’ipotesi iniziale. Ciò significa determinare, tramite test, se i dati raccolti supportano o confutano l’ipotesi stessa. Una volta ottenuti i risultati, è essenziale presentarli in modo chiaro e comprensibile agli stakeholders attraverso una relazione scritta o la creazione di dashboards per la visualizzazione di dati. Infine, l’ultima fase comporta l’uso delle intuizioni basate sui dati per sviluppare un sistema che possa influenzare decisioni, politiche, piani e finanziamenti. In questo modo, i risultati della ricerca contribuiranno a guidare l’azione e il cambiamento; i dati presentati vengono poi sfruttati allo scopo di fare osservazioni e riflessioni che permetteranno di prendere decisioni di policy, governance, funding e pianificazione (Kumar, 2020).



vishalkumar.london ©

*1 -Illustrazione dei sette step necessari per intraprendere un processo di ricerca che si serve della cultural data science (Kumar, 2020)*

Sempre nel contesto della conferenza di Dubai, Kumar evidenzia come la cultural data science, al pari della data science, sia figlia dell'intersezione tra diversi ambiti accademici: il settore umanistico con l'arte e la cultura, le scienze sociali, la matematica e l'informatica (figura 2).



2 - Diagramma di Venn che spiega l'appartenenza della cultural data science a più settori (Kumar, 2020).

Si potrebbe anche dire, secondo lo studioso, che sia un'estensione della cultural analytics, degli studi urbani e dei social studies computazionali. Questa multisetorialità comporta una certa difficoltà da parte degli esperti del settore nell'applicazione ad un settore già di per se variegato come quello culturale (Manovich, 2016). La data science e la cultural data science utilizzano spesso algoritmi di Machine Learning, una tecnica all'interno dell'IA che si concentra sull'addestramento di algoritmi e modelli statistici per migliorare le loro prestazioni su compiti specifici attraverso l'apprendimento dai dati. In sostanza, il machine learning utilizza dati di addestramento per insegnare a un sistema informatico a fare previsioni o prendere decisioni, in base a modelli matematici e statistiche. Questi modelli possono essere applicati in una varietà di campi, tra cui il riconoscimento di pattern, l'analisi dei dati, la classificazione, la regressione e molto altro. Nel momento in cui viene impiegato nel settore culturale si serve di dati che hanno quindi delle peculiarità e caratterizzazioni specifiche. Nel loro saggio del 1997, gli studiosi Handwerker e Wozniak descrivono i dati culturali come processi sociali interattivi che portano alla conoscenza l'uno dell'altro nei gruppi umani:

“Cultural data thus should reflect the social (interactive) processes by which we construct our knowledge of each other and the way these processes work (Handwerker & Wozniak, 1997).”

Kumar amplia la definizione del 1997 individuando diverse categorie di dato culturale:

1. dati sulla storia culturale;
2. dati riguardanti il patrimonio culturale tradizionale, ad esempio dati di collezioni museali e librerie, che possono essere visivi o testuali;
3. dati di cultura contemporanea, ad esempio attività, contenuti e interazioni generati dagli utenti di social media;
4. dati sulle transazioni del mercato culturale, ovvero dati che riguardano oggetti artistici e culturali che vengono scambiati a scopo di profitto;
5. dataset di stampo culturale e socio-economico, contenenti dati che riguardano l'economia e le preferenze della società che vengono generalmente raccolti tramite survey e questionari;
6. dataset contenenti dati spazio-culturali, quindi le collocazioni di edifici culturali, luoghi pubblici, land art e in generale arte pubblica;
7. dati sulle nuove forme di cultura, come gli strumenti digitali utilizzati per la realtà aumentata, virtuale e internet of things.

Secondo James Dobson, ricercatore del Dartmouth College, è importante tenere conto dell'ambiguità nelle materie di studio del settore umanistico e culturale quando ci si affida a metodi computazionali nella ricerca umanistica. Egli ritiene esistano due categorie principali in cui far rientrare la maggior parte del lavoro computazionale in ambito umanistico. La prima categoria è interessata alla critica letteraria e la considera non come secondaria rispetto ad un lavoro creativo ma come un atto creativo in se e per se. La critica creativa computazionale non è strettamente legata in maniera esplicita ad un rigore scientifico che si basa sulle regole proprie della statistica e della computazione; la trasformazione algoritmica del testo è definita creativa. In questo senso la critica può essere argomentativa ma non soggetta ad un sistema di critica che potrebbe definire il testo come errato, obsoleto o non valido. Queste forme di computazione trasformano e deformano un testo di input. La seconda categoria invece si serve di metodi computazionali per fare dichiarazioni empiriche sui dati. Queste pratiche hanno lo scopo

di dare dimostrazione dell'importanza delle scoperte nel settore umanistico; le interpretazioni dei risultati sono derivate da assunti che possono essere testati e verificati, con dichiarazioni ermeneutiche che vengono dall'utilizzo di modelli computazionali che per l'analisi dei dati. In questa seconda categoria, gli studiosi sono generalmente interessati ad affinare e a migliorare i modelli da loro utilizzati per ottenere risultati più significativi. Mentre la ricerca basata sui dati è spesso considerata in opposizione all'ambiguità, perché finalizzata all'individuazione di una risposta precisa ad una domanda preesistente, Dobson sostiene che proprio i dati possono consentire la scoperta e la messa in evidenza dell'incertezza, aprendo la strada ad interpretazioni nuove. È quindi importante integrare le nuove tecnologie a supporto delle attività del settore culturale senza perdere l'accesso alla dimensione sfumata e ambigua dell'informazione del settore culturale, evitando quindi la ricerca di risposte granitiche e mantenendo una flessibilità mentale che apra a diverse interpretazioni. Il rischio che la dimensione ambigua del dato culturale venga persa è presente tanto per il ricercatore che utilizza questi strumenti ai fini delle proprie ricerche quanto per il lettore o il revisore di tali interpretazioni. Le rappresentazioni grafiche dei dati, l'alta precisione nella classificazione e l'accuratezza delle statistiche di sintesi rischiano di essere strutturati facendo riferimento a retoriche preesistenti e a preconetti e comprensioni condivisi comunemente accettati. È quindi importante sottolineare come l'argomentazione e l'interpretazione efficaci e accademicamente responsabili nelle discipline umanistiche richiedano non solo accesso senza limitazioni ai metodi e ai dati utilizzati nelle ricerche per generare risultati e dimostrare l'accuratezza, ma anche che i set di dati e i modelli selezionati siano interpretabili. Ciò include la selezione di modelli di dati interpretabili che preservino il maggior numero possibile di punti di accesso alla complessità e alla diversità dei dati stessi, affinché i modelli interpretabili possano essere utilizzati per garantire che le interpretazioni non siano limitate a una singola prospettiva. Diversi algoritmi applicati agli stessi dati possono produrre simili risultati nella classificazione e nell'accuratezza, ma le caratteristiche utilizzate da questi algoritmi per determinare l'appartenenza ad una classe o ad un'altra possono essere profondamente diversi. Questa forte complessità rappresenta un problema importante nel settore umanistico, generando domande come "Cosa significa se un algoritmo assegna maggiore importanza ad un subset

di caratteristiche invece che ad un altro?”, “Se entrambi gli algoritmi identificano una distinzione significativa tra le classi, significa che la distinzione tra dati contenuti nel dataset è reale?”, oppure “se esiste una differenza che viene individuata con un diverso set di caratteristiche, perché è importante la caratteristica del singolo dato?”. Rispondere a queste domande può non essere immediato: un algoritmo che ricerca una risposta molto accurata priorizzerà e troverà le caratteristiche distintive che sono compatibili con il modello. Un esempio è il dataset Text-Encoding Initiative (TEI), che analizza testi presi dal Text Creation Partnership (TCP), derivato dalla collezione Early English Books Online (EEBO), consentendo di sperimentare con diversi classificatori per esaminare le caratteristiche utilizzate per classificare testi lirici brevi. Non per forza, quindi, l'utilizzo di modelli di machine learning porterebbe ad interpretazioni granitiche e a conclusioni certe; l'utilizzo di modelli interpretabili può aiutare a garantire che l'interpretazione dei dati non sia limitata a una singola prospettiva, ma anzi permetta di giungere ad un'argomentazione e di un'interpretazione responsabili. I modelli interpretabili sono quelli che consentono ai ricercatori di capire come funziona il modello e come è arrivato alle sue conclusioni (Dobson J., 2021).

### **1.1 Cenni storici**

I primi anni di approccio al connubio fra settore umanistico e digitale vengono definiti, fino a circa il 2002, come “prima ondata”. Gli studi di settore sono concordi nel collocare l'inizio della disciplina delle Digital Humanities nel 1949 con l'*Index Thomisticus* di Roberto Busa (Dalbello, 2011). Il progetto prevedeva la messa in relazione dell'opera omnia di Tommaso d'Aquino per un totale di quasi 11 milioni di parole indicizzate su schede perforate IBM e circa 18 milioni di ore di lavoro. Un programma puramente meccanico, che cercasse similitudini tra parole sulla base della loro forma grafica (quindi analizzando la sequenza di lettere che le compone) avrebbe permesso di produrre un risultato in un tempo molto più breve. L'intenzione di Busa era però quella di produrre una concordanza basata sui lemmi, in cui le parole venivano associate sulla base del loro significato (Hockey, 2004). A distanza di mezzo secolo è sorprendente notare la grande differenza nei “big data” rispetto ai giorni nostri: a partire dalle macchine utilizzate, come il VariaSpecima, un progetto che nel 1951 si occupava di utilizzare componenti elettriche e meccaniche per ordinare, replicare e stampare su cartoncini. Quello tra gli anni '40 e '50

fu un momento di passaggio tra computazione elettro-meccanica e digitale, e il metodo delle “punched cards” era una metodologia di trasporto delle informazioni affidabile a sufficienza da essere standardizzata. Il progetto di Busa, come accennavamo, usava carte standard IBM in cartoncino rigido con una tacca nel lato in alto a sinistra che permetteva di orientarla nel verso giusto: nonostante questi cartoncini seguissero delle misure standard, l'enormità del progetto di Busa rese necessaria la creazione di una nuova misura più grande. Grazie al progetto venne analizzata la totalità dei testi prodotti dal santo, consentendo così agli studiosi di conoscere il suo pensiero e la sua filosofia derivata dalla messa in relazione di tutti i singoli scritti di san Tommaso (Rockwell, 2019). Il progetto cominciò come una collaborazione tra istruzione, industria meccanica e ricerca accademica diventando trampolino di lancio per la prima generazione di macchine di calcolo digitale su larga scala quasi vent'anni prima che le stesse si diffondessero negli uffici e venissero impiegate nella vita di tutti i giorni. *L'Index* venne pubblicato circa trent'anni più tardi, tra il 1974 e il 1979, affermandosi come punto di riferimento per teologi, filosofi, medievalisti, storici, latinisti e studiosi del settore (Dalbello, 2011).

A seguito dei numerosi avanzamenti nel campo dello studio di *corpus* linguistici, nel 1966 Joseph Raben fonda la rivista *Computers and the Humanities*, che vede l'analisi di progetti che inizialmente sono fortemente limitati soprattutto da costi elevati delle strutture hardware, processi di elaborazione ancora poco avanzati, e scarsa possibilità di archiviazione (Sula, 2019). Le analisi, inoltre, richiedevano molto tempo per essere elaborate. I dati da analizzare erano principalmente testuali o numerici, dati in input alle macchine a mano o tramite le schede perforate. Queste schede contenevano fino a ottanta caratteri o una linea di testo maiuscolo. I risultati delle analisi potevano essere accessibili per l'utente solo nel momento in cui venivano stampati, alla fine di processi che potevano richiedere anche diverse ore. La maggior parte dei dataset, inoltre, era archiviata su nastri magnetici che potevano essere processati serialmente grazie a software implementati per ridurre al minimo il movimento di questi nastri. Questo sistema rendeva sostanzialmente impossibile la consultazione randomica di dati e necessaria un'archiviazione seriale, che risultava problematica soprattutto per i dati di natura storica, i quali prevedevano la rappresentazione di diversi aspetti di uno stesso oggetto in una sola linea. Gli storici erano quindi restii ad affidarsi alla tecnologia e a farsi coinvolgere in progetti che si servivano

di processi computerizzati (Hockey, 2004). Un problema ulteriore era rappresentato dalla rappresentazione dei caratteri. Fino all'avvento del sistema Unicode, vennero ideati metodi diversi per la rappresentazione di lettere maiuscole e minuscole e caratteri speciali (come le lettere accentate), per esempio attraverso l'utilizzo di asterischi o tramite traslitterazione. Non erano solo i caratteri speciali a risultare problematici: per i sistemi di retrieval e di concordanza era necessario individuare le citazioni tramite il loro posizionamento nel testo, perché testi poetici e letterari non presentavano strutture simili agli articoli di giornale, documenti per i quali questi sistemi erano stati costruiti. Il sistema di riconoscimento strutturale dei documenti umanistici più avanzato fu il COCOA, ideato da D.B. Russell. Il sistema consentiva di generare concordanze a partire da testi di ogni genere, anche letterari. Si trattava di una tabella che conteneva tutte le parole contenute all'interno di un testo specifico, accompagnate da un dato numerico che indicava quante volte una stessa parola venisse ripetuta. La tabella veniva poi emessa in output in tre diverse versioni: una in ordine alfabetico, una in ordine di frequenza delle parole (la parola più utilizzata si trovava al primo posto) e una in ordine alfabetico in base alle rime, cioè in base alla parte finale di ogni parola. Questo è il periodo in cui è soprattutto la linguistica ad affermarsi come oggetto dei principali studi condotti. Gli anni '70 e '80 vedranno, infatti, una concentrazione maggiore del settore sull'utilizzo di questa disciplina. La "seconda ondata" è caratterizzata da un utilizzo del digitale non più solo come mero strumento, ma come mezzo di creazione di dati umanistici digitali. È durante questa seconda ondata che nasce Wikipedia, esempio lampante di oggetto che racchiude in sé i settori digitale e umanistico. Con la terza ondata, a partire circa dal 2009, si riconosce e definisce il nuovo settore delle Digital Humanities, dotandolo di metodologie e definizioni teoriche. Gli studiosi si interrogano sulle potenzialità del computer come mezzo di comunicazione che va al di là della semplice parola stampata. In questo periodo si assiste ad una consapevolezza più profonda degli impatti socioculturali e politici che i progetti del settore potrebbero avere (Helmi, 2021).

## **1.2 Gli anni '70 e '80 e il consolidamento delle Digital Humanities**

La rivista fondata da Raben nel 1966 aiutò moltissimo la causa del digitale utilizzato nel settore umanistico, ma fu soprattutto grazie ad un ciclo di conferenze, svoltosi nel 1970 a

Cambridge, che si creò un centro di riferimento collocato in UK. Fu proprio qui che si svolsero le conferenze di Edimburgo (1972), Cardiff (1974), Oxford (1976), Birmingham (1978) e Cambridge (1980), che produssero paper di altissima qualità, e al King's College di London nel 1973 venne fondata l'Associazione per l'informatica letteraria e linguistica, che pubblicava una propria rivista. L'associazione si occupava di organizzare incontri e conferenze annuali invitando oratori diversi e a partire dal 1986 si dotò di una rivista, *Literary and Linguistic Computing*. A partire dalla metà degli anni '70 cominciò ad essere organizzato un altro ciclo di conferenze in Nord America, intitolato "International Conference on Computing in the Humanities (ICCH)". Questa conferenza annuale veniva alternata alle conferenze britanniche, che nell'arco del tempo cominciarono ad essere focalizzate soprattutto sul tema della linguistica (Hockey, 2004). L'analisi degli articoli delle riviste *Computer and the Humanities* e *Literary and Linguistic Computing* permette di osservare come, dagli anni '60 agli anni '90 circa, gli articoli dedicati all'analisi di oggetti di natura testuale fossero predominanti rispetto a quelli che analizzavano altre branche settore umanistico (Sula, 2019). Pur continuando ad essere il testo il focus di analisi, tra il 2005 e il 2011 si assiste ad un crescente interesse per l'analisi di video, audio, immagini e videogiochi (Nyhan & Flynn, 2016). Le Digital Humanities oggi comprendono una vasta area di pratiche e studi quali corpus linguistici, archivi digitali e non solo; alcuni tra i progetti di questo ambito (come l'*Index Thomisticus*) sono diventati fonti cardine nelle rispettive discipline. Grazie ai progressi nel settore, non solo gli specialisti di settore hanno potuto servirsi di strumenti che hanno permesso avanzamenti notevoli, ma anche acquisire una serie di competenze che li ha resi grandi innovatori tecnologici. Ecco che arriviamo agli anni '20 del 2000, anni in cui come accennavamo all'inizio di questo capitolo si assiste ad un impiego della data science sempre crescente in ambito umanistico. Tra le diverse applicazioni di intelligenza artificiale di cui si serve la data science, sicuramente una delle più rilevanti è il machine learning, branca dell'intelligenza artificiale grazie alla quale si creano sistemi che apprendono o migliorano performance sulla base dei dati che utilizzano. In sostanza, il sistema apprende dai dati e migliora la sua performance con l'esperienza; l'algoritmo viene "addestrato" a far emergere schemi e correlazioni da grandi set di dati e a formulare previsioni sulla base delle analisi svolte. Lo studio delle reti neurali e dell'intelligenza artificiale dell'era moderna iniziò nel 1943

con gli studi di Warren Sturgis McCulloch e Walter Pitts. Le reti di McCulloch e Pitts sono degli automi a stati finiti, ovvero dei modelli matematici di calcolo che permettono di descrivere precisamente e formalmente il funzionamento dei sistemi consentendo di formulare ipotesi precise sulla natura dei meccanismi cerebrali, in una forma simile a quella dei programmi per computer. La ricerca permise di stabilire una volta per tutte la validità dei modelli formali dei meccanismi cerebrali. L'origine del machine learning vero e proprio viene attribuita invece allo psicologo Frank Rosenblatt della Cornell University, che tra il 1957 e il 1960 condusse una serie di studi finalizzati alla costruzione del Perceptron, una macchina in grado di riconoscere le lettere dell'alfabeto. La macchina diventò il prototipo delle reti neurali artificiali moderne, con un modello di apprendimento molto simile a quello umano e animale sviluppato in psicologia. Nel 1965-66 fu Vladimir Andreevich Yakubovich a sviluppare un approccio sistematico alla teoria matematica della pattern recognition, tecnologia finalizzata a sviluppare sistemi automatici in grado di riconoscere forme o profili attraverso i dati grezzi acquisiti, in questo caso chiamato "metodo delle ineguaglianze ricorrenti". Si basava sulla risoluzione di problemi legati ad un sistema di ineguaglianze costruite allo scopo di essere funzionali al sistema e consentiva di trovare la soluzione ad un numero infinito di ineguaglianze precedentemente nascoste. Fu con l'avvio del XXI secolo che ebbe sviluppo ciò che Fradkov definisce "la corsa all'oro" del Machine Learning, verificatasi grazie alla presenza sinergica di tre nuovi trend. Il primo fra questi fu l'avvento dei Big Data: la quantità di dati iniziava a diventare tale che le nuove scoperte e gli studi scientifici furono guidati dalla necessità pratica più che dalla curiosità degli scienziati. Il secondo trend era la tendenza a ridurre i costi di calcolo simultaneo e di memoria. Insieme a nuove scoperte tecnologiche che permisero la produzione di GPU più efficienti, si ebbe un significativo calo dei costi delle RAM, che permise di tenere in memoria grandi quantità di dati aprendo la strada allo sviluppo di nuovi database non relazionali (NoSQL)<sup>2</sup> (Fradkov, 2020). Il terzo trend fu rappresentato dall'avanzamento degli studi di nuovi algoritmi di deep machine learning, che riprendeva e ampliava gli studi del perceptron. Dopo diversi anni di studio, vide la

---

<sup>2</sup> I database non relazionali contengono dati non strutturati.

luce il concetto tecnologico del deep neural networks. Nel 1986 fu poi la studiosa Rina Dechter a proporre il termine deep learning.

### 1.3 Machine Learning: tipologie

L'obiettivo del machine learning è quello di ideare e sviluppare algoritmi che, servendosi di dati empirici, esperienza e training, permettano ai sistemi di evolversi adattandosi ai cambiamenti dei contesti in cui sono inseriti. Il focus principale del machine learning è l'individuazione di patterns e regole che portino alla creazione di modelli derivati dai dati analizzati. A questo scopo vengono combinati approcci e tecniche di diverse aree, come mostrato in figura:

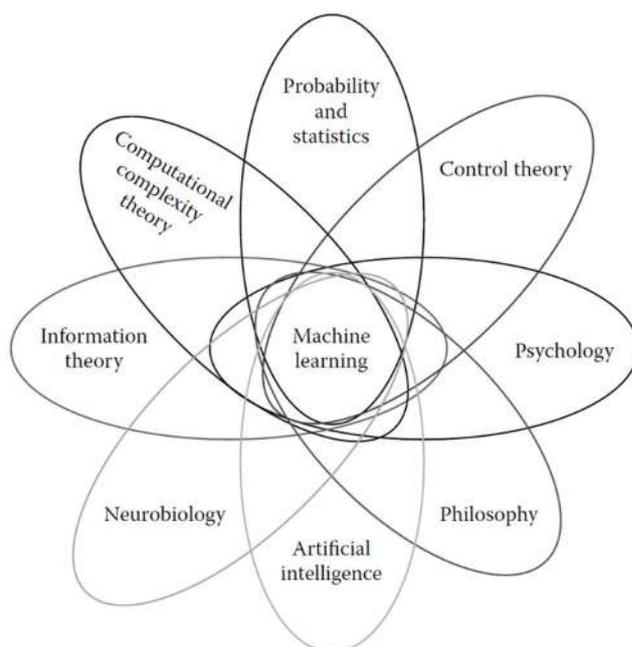


Figura 3 - Il machine learning si colloca al centro dell'intersezione fra diversi ambiti di studio (Kapitanova & Son, 2013)

Esiste una grande quantità di algoritmi di machine learning, raggruppabili in tre classi principali in base al tipo di addestramento utilizzato: apprendimento supervisionato, apprendimento non supervisionato e apprendimento semi-supervisionato. Nell'apprendimento supervisionato, il modello è addestrato su un insieme di dati di addestramento etichettati, cioè dati in cui sia le caratteristiche (input) che l'output desiderato (etichetta) sono noti. L'obiettivo principale dell'apprendimento supervisionato è quello di apprendere una funzione o una mappatura che permetta di fare

previsioni quando il modello è presentato con nuovi dati di input. Esempi comuni di applicazioni di apprendimento supervisionato includono classificazione e regressione. Nell'apprendimento non supervisionato, il modello è addestrato su un insieme di dati di addestramento in cui le etichette non sono disponibili. L'obiettivo dell'apprendimento non supervisionato è scoprire schemi, strutture o raggruppamenti intrinseci nei dati senza alcuna guida sull'output atteso. Esempi comuni di algoritmi di apprendimento non supervisionato includono il clustering (raggruppamento) e la riduzione della dimensionalità. L'apprendimento semi-supervisionato è una combinazione di apprendimento supervisionato e non supervisionato. In questo caso, solo una parte dei dati di addestramento è etichettata, mentre la maggior parte rimane senza etichetta. L'obiettivo è sfruttare le informazioni contenute nelle etichette disponibili insieme ai dati non etichettati per migliorare le prestazioni del modello. L'apprendimento semi-supervisionato è utile quando ottenere un gran numero di etichette può essere costoso o richiede molto tempo, ma è comunque possibile disporre di alcune etichette. Ognuno di questi approcci ha le proprie applicazioni e vantaggi. L'apprendimento supervisionato è ampiamente utilizzato quando ci sono dati etichettati disponibili e l'obiettivo è la predizione precisa, l'apprendimento non supervisionato è utile per scoprire informazioni nascoste nei dati, l'apprendimento semi-supervisionato è una scelta quando si hanno limitate etichette ma si desidera sfruttare al meglio i dati a disposizione, sfruttando quindi dati etichettati e non etichettati. In questo ultimo caso, tipicamente, i dati etichettati rappresentano una piccola percentuale dell'intero dataset. Nell'apprendimento supervisionato, secondo Kapitanova e Son, sono individuabili cinque step che vanno perseguiti quando si decide di addestrare un modello. Il primo step, dopo aver individuato la tipologia di dati di training necessari, consiste nella raccolta dei dati stessi. È importante capire quale tipo di dati serve e in quale quantità: la costruzione di un dataset può richiedere molto tempo ed energie (Kapitanova & Son, 2013). L'utilizzo del machine learning nel settore culturale è ancora piuttosto limitato: la letteratura scientifica in ambito culturale si basa ancora molto su strumenti statistici, comunemente utilizzati su datasets relativamente piccoli che spesso non sono disponibili al pubblico. Per ciò che concerne il settore culturale, abbiamo esempi di utilizzo di tecniche di machine learning supervisionato che comprendono tecniche quali l'utilizzo di regressione lineare e

logistica, modelli di deep learning e alberi decisionali. Il machine learning non supervisionato nel settore culturale si serve principalmente di tecniche di clustering, dette anche “cluster analysis”, ovvero raggruppamenti di dati con caratteristiche simili tra loro sulla base di alcune metriche prestabilite. Le tecniche di clustering comprendono il k-Means e il DBSCAN Clustering. Per quanto riguarda l’apprendimento semi-supervisionato, l’obiettivo è trasferire la conoscenza acquisita da un dominio sorgente ad un dominio target per il quale le etichette non sono normalmente disponibili individuando un sistema di mapping nella distribuzione di dati sorgente e di target. I sistemi più utilizzati quando si parla di machine learning semi-supervisionato sono le reti neurali semi-supervisionate (semi-supervised deep neural networks) e algoritmi di clustering. Come detto in precedenza, raccogliere dati etichettati è un processo lungo e dispendioso, per cui in anni recenti le potenzialità delle deep neural networks semi-supervisionate sono state indagate sempre più spesso. Per il settore dei beni culturali questo si è tradotto in applicazioni nell’ambito dell’analisi di contenuti visuali e con un focus specifico sugli algoritmi di clustering, che sono tra i più utilizzati tra le tecniche di machine learning (Fiorucci, et al., 2020). Come anticipato, il settore culturale ha iniziato solo negli ultimi anni a rendere disponibili e accessibili al pubblico grandi quantità di dati. Diversi musei soprattutto hanno messo a disposizione del pubblico le proprie collezioni in formato digitale. Molte delle applicazioni del Machine learning nel settore culturale sono impostate utilizzando impostazioni di computer vision, che permette una più agevole acquisizione dei dati, ma sono presenti barriere sociali e tecniche che rendono difficoltosa un’incrocio di influenze tra i due settori. Tra questi, primo fra i problemi è la difficoltà nell’avere accesso a dati in quantità sufficienti (i datasets sono molto spesso piccoli e non disponibili al pubblico). Gli sviluppi recenti permettono però di superare in alcuni casi questi problemi e sfruttare appieno tecniche di deep learning. Utilizzando tecniche di machine learning, l’attore del settore culturale ha la possibilità di condurre analisi più snelle ed efficienti, risparmiando grandi quantità di tempo dedicate ad esempio alla lettura di testi o all’osservazione di immagini, attività che grazie a sistemi di text analysis o di clustering possono essere svolte dalla macchina permettendo di concentrarsi sugli output di queste analisi. Secondo James Dobson, l’analisi di pratiche e metodi di text mining supportate dalla macchina rappresentano la forma più esemplare di digital

humanities. L'algoritmo di machine learning che permette la maggior parte degli sforzi di test mining nel settore sono fortemente usati in altre discipline e non rappresentano una nicchia di un modus operandi arbitrario ma una pratica centrale nella ricerca umanistica legata a protocolli di ricerca universitari (Dobson J. E., 2015). Matthew Jockers sostiene che la text analysis computazionale sia la vera e propria fondazione del settore delle digital humanities, riferendosi all'indice tomistico di Busa degli anni '40 del '900 (Jockers, 2013). Nei prossimi capitoli analizzeremo nel dettaglio alcuni dei principali sistemi di machine learning utilizzati nel settore indagando le possibilità di connubio tra settore culturale e informatico.

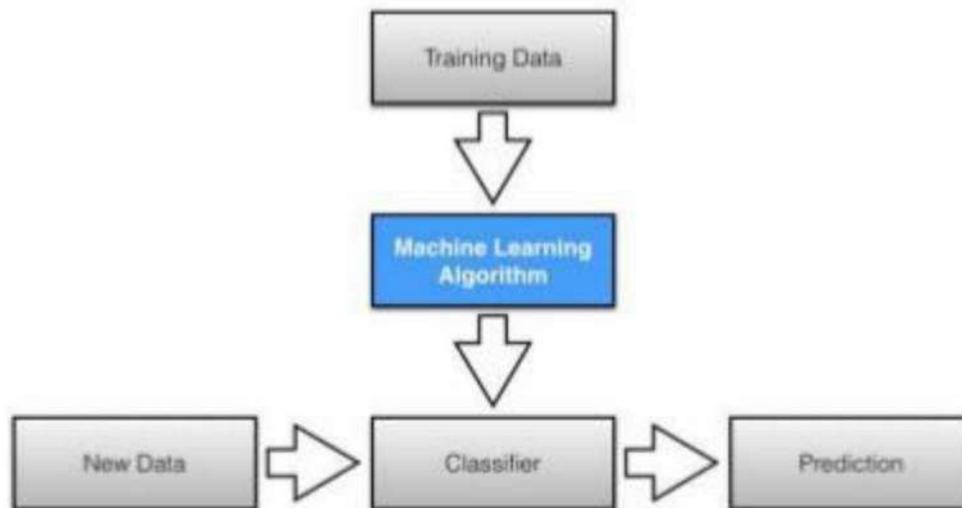
#### **1.4 Best practices per la costruzione di un modello di machine learning da impiegare nel settore culturale**

Ci Viene fornita nel 2018 da Tobias Blake una specifica sugli step esplorativi da intraprendere nell'approcciarsi al dato per la costruzione di un modello di analisi predittiva. Anzitutto è importante avere una panoramica chiara dei casi e delle relative osservazioni, così come le caratteristiche proprie di ciascuna osservazione. Solamente con una adeguata quantità di osservazioni è possibile ottenere sufficienti combinazioni di caratteristiche per la costruzione di un modello. In secondo luogo, bisogna determinare quante siano le caratteristiche categoriche e numeriche. Infine è necessario definire una strategia per lavorare con valori mancanti, problema molto comune in dataset di stampo storico. Per facilitare la riesamina di datasets e le eventuali integrazioni future, si sono diffuse riviste scientifiche che si occupano del tema dell'utilizzo dei dati nel settore umanistico. Uno dei primi cosiddetti *humanities data journals* è il Journal of Open Humanities Data, che pubblicò il primo dataset nel 2015 (Blanke, 2018). Nel 2023 un team di studiosi sottolinea come l'avvento sempre crescente del Machine Learning nei settori commerciali e accademici abbia portato ad un interesse sempre crescente all'utilizzo di modelli di ML che utilizzavano dati del settore culturale. Alimentate dalla crescita di diversi approcci computazionali finalizzati all'elaborazione di grandi quantità di dati, il movimento cosiddetto "Collections as Data" ha guadagnato sempre più popolarità. Questo movimento incoraggia l'utilizzo computazionale di beni culturali digitalizzati o nati digitali rendendo accessibili questi beni sotto forma di dati e disponibili per l'utilizzo computazionale. Ad oggi anche il mondo big tech si impegna nell'elaborazione di grandi

quantità di dati del patrimonio culturale, fattore sicuramente positivo perché determina un utilizzo di queste informazioni che non è più solo strettamente legato al settore umanistico e culturale; ciononostante, questo comporta anche una gestione del dato talvolta in contesti che non sono familiari con le specifiche e la complessità del settore culturale, che richiede una particolare attenzione al contesto per evitare utilizzi impropri (Alkemade, et al., 2023).

## Capitolo 2 - Il Machine Learning Supervisionato

Come accennato nel capitolo precedente, il Machine Learning cosiddetto “supervisionato” prevede algoritmi che si occupano di mappare una certa quantità di input in modo che generino determinati output. Uno dei classici compiti del modello supervisionato è la classificazione: al sistema viene chiesto di assegnare ad un dato una in particolare tra le diverse classi a disposizione dopo aver osservato una certa quantità di esempi di input e output. Il processo di apprendimento a cui è sottoposta la macchina altri non è che il processo di scoperta di regolarità statistiche o altri pattern di dati. Fra tante, la programmazione adattiva è molto popolare: viene utilizzata laddove i modelli sono in grado di imparare dall’esperienza o estraendo nuove informazioni dai dati ottimizzando sempre di più questo processo di input e output. Le classi degli algoritmi supervisionati sono predeterminate, create in un set definito dall’utente umano. Questo significa che i segmenti di dati saranno etichettati in base alle classificazioni predefinite. Il grande paniere dei sistemi di machine learning supervisionato si divide a sua volta in due modelli principali: modelli di classificazione e modelli di regressione. Il primo modello mappa gli input per inserirli in classi predeterminate dall’uomo; il secondo mappa gli input in domini realmente esistenti. In un semplice modello di Machine Learning abbiamo due momenti principali: una fase di apprendimento e una fase di addestramento. Durante il processo di testing, o addestramento, il modello utilizza i dati per fare delle previsioni; i dati etichettati sono il risultato di modelli di apprendimento che possono restituire una previsione o una classificazione (Nasteski, 2017).



*Figura 4 - Una rappresentazione del sistema di ML supervisionato a cura di Vladimir Nasteski (Nasteski, 2017)*

Sovente il sistema appena descritto nel settore culturale è utilizzato per la creazione di dataset con caratteristiche ed etichette ben precisi. Il compito principale è la costruzione di un modello che sia in grado di prevedere l'etichetta di un oggetto sulla base delle sue caratteristiche. L'algoritmo riceve in input un set di caratteristiche a cui corrisponde un determinato output, e impara comparando l'output a cui è giunto con l'output corretto per identificare eventuali errori e modificare il modello di conseguenza. In questo modo, se alcuni dei valori in input sono mancanti, non è però possibile dedurre un output. Questo modello è particolarmente utilizzato per fare delle previsioni di eventi che potrebbero verificarsi sulla base di dati storici. L'obiettivo principale del machine learning supervisionato è, quindi, la formulazione di un prototipo compatto della distribuzione delle etichette in termini di tipologie di previsioni. Il classificatore che ne risulta destina le varie etichette ai test set in cui sono note le stime previsionali, ma i valori delle etichette che identificano le classi non sono identificabili. Come detto in precedenza, la necessità di avere dati etichettati fa sì che sia necessario l'intervento di numerosi specialisti e studiosi dell'ambito in cui si desidera formulare una previsione. I problemi di machine learning supervisionato vengono raggruppati in due grandi famiglie: problemi di regressione e problemi di classificazione. Nei problemi di classificazione, il risultato assume un valore

specifico; l'obiettivo è prevedere questo valore e identificarne la classe di appartenenza. Nella regressione, il valore viene previsto sulla base del dataset etichettato, e i valori di output sono previsti sulla base degli ultimi dati che vengono presi in esame dall'algoritmo. La scelta dell'algoritmo da impiegare nella costruzione di un modello di machine learning supervisionato va quindi condotta tenendo a mente le caratteristiche di eterogeneità, precisione, eccesso, linearità del dato che deve essere utilizzato. Recenti avanzamenti nel machine learning supervisionato hanno permesso di calcolare le proprietà di determinati materiali e prevederne la struttura (Shetty, Singh, & Rao, 2022).

## **2.1 Applicazioni del machine learning supervisionato al settore culturale**

### **2.1.1 La regressione lineare e la regressione logistica**

La regressione lineare rappresenta una classe di modelli matematici per l'apprendimento supervisionato. È utilizzata in molteplici ambiti gestionali per studiare la correlazione matematica esistente fra una variabile detta "target" ed un insieme di variabili esplicative. È, in altre parole, una tecnica di analisi dei dati che prevede il valore di dati sconosciuti utilizzando un altro valore di dati correlato e noto (Milanato, 2008). Molto semplicemente, è un approccio lineare che collega un output a più variabili descrittive: pensiamo un modello con una relazione lineare tra una serie di input  $x$  e un output  $y$ , il quale può essere dedotto proprio dalla combinazione di variabili  $x$ . Quando un input lineare è singolo si parla di regressione lineare semplice, mentre se un input lineare ha valori multipli si parla di regressione lineare multipla (Shetty, Singh, & Rao, 2022). Tra i modelli di machine learning supervisionato, viene utilizzato anche il modello di regressione lineare. L'obiettivo principale è quello di comprendere e quantificare la relazione lineare tra queste variabili. Il modello cerca di adattare una linea retta che rappresenti al meglio la relazione tra le variabili. La linea retta è definita dall'equazione della retta, che può essere espressa come:

$$y=mx+c$$

Dove  $y$  è la variabile dipendente o target,  $x$  è la variabile indipendente o feature,  $m$  è il coefficiente di pendenza della retta e  $c$  è l'intercetta con l'asse  $y$ . Il processo di costruzione di un modello di regressione lineare coinvolge il trovare i valori ottimali per i coefficienti

m e c in modo che la linea retta si adatti al meglio ai dati osservati. Una volta costruito, il modello può essere utilizzato per fare previsioni su nuovi dati, stimando i valori della variabile dipendente in base alle variabili indipendenti fornite. È importante notare che il modello di regressione lineare ha delle assunzioni, come la linearità della relazione tra variabili e l'indipendenza degli errori, che devono essere valutate prima di applicare il modello e interpretarne i risultati. L'obiettivo della regressione lineare è l'individuazione di relazioni e dipendenze tra variabili, in modo da prevedere un target di variabili in continuo aggiornamento. Esistono diversi modelli di apprendimento dei dati per la regressione lineare, come ad esempio la regressione lineare semplice, per cui un singolo input viene usato per stimare i coefficienti tramite calcoli matematici come media, deviazione standard, correlazione e covarianza; ordinary least square (OLS), ovvero il metodo dei minimi quadrati ordinari, utilizzato quando c'è più di un input. Il metodo OLS cerca di minimizzare la somma dei residui al quadrato. Cioè, per una data linea di regressione attraverso gli input, viene calcolata la distanza da ogni punto dati rispetto alla linea di regressione, quindi si eleva al quadrato e si sommano tutti gli errori al quadrato. Supponendo i dati come una matrice, questo approccio utilizza l'algebra lineare per calcolare i valori dei coefficienti. È necessaria una quantità sufficiente di memoria e dati per adattare i dati e completare le operazioni della matrice. Altro metodo di apprendimento di regressione lineare prevede che per più di un valore di input, il processo di ottimizzazione dei valori dei coefficienti può essere ottenuto minimizzando iterativamente gli errori sui dati di addestramento. Questa procedura è chiamata discesa del gradiente e funziona con valori casuali per ogni coefficiente. Per ogni coppia di dati di input e output, viene stimata la somma degli errori al quadrato. I valori dei coefficienti vengono aggiornati nella direzione della riduzione dell'errore. Questa procedura è ripetitiva fino a quando non si raggiunge una somma degli errori al quadrato minima o non è più possibile ottenere ulteriori progressi. Vi è poi il metodo di regolarizzazione, che cerca di minimizzare l'errore al quadrato sui dati di addestramento (utilizzando OLS) e allo stesso tempo ridurre la complessità nel modello. Questi approcci sono considerati efficaci quando i valori di input sono collineari e OLS sovradata il dataset di addestramento.

La regressione logistica è un noto algoritmo di machine learning. Viene utilizzato per prevedere il fattore dipendente facendo uso di un insieme dato di fattori indipendenti. È impiegato per problemi di classificazione e si basa sull'idea di probabilità. La regressione logistica calcola l'output di una variabile dipendente. Di conseguenza, l'esito è un valore discreto. Potrebbe essere sia sì che no, zero o uno, valido o non valido. Tuttavia, anziché fornire un valore definito come 0 e 1, fornisce valori probabilistici che si collocano nell'intervallo tra 0 e 1 (Shetty, Singh, & Rao, 2022). Un articolo del 2018 analizza come, grazie all'utilizzo di un modello di regressione logistica, sia stato possibile prevedere il posizionamento di siti archeologici. Assumendo che la collocazione di questi siti non fosse casuale, ma dipendesse da scelte umane, già dal 1980 il settore archeologico ha iniziato a fare uso di strumenti quali i modelli predittivi. I modelli predittivi archeologici sono strumenti che aiutano a calcolare le probabilità che un sito sia posizionato in un luogo specifico. Il modello predittivo conduce delle analisi quantitative dei settori ambientali e le proietta sulle aree i cui siti sono sconosciuti come aree di probabilità. Queste superfici vengono rappresentate graficamente come mappe colorate o immagini che mostrano valori alti o bassi per la probabilità che il sito ha di essere scoperto. La tecnica statistica predominante per modelli predittivi di stampo archeologico è la regressione logistica. Il modello predittivo che prenderemo ad esempio si serve di variabili binarie, in cui il risultato ci viene fornito come un valore probabilistico tra 0 e 1, o come sì e no. Le variabili ambientali sono definite come indipendenti e il modello esamina la relazione tra variabili indipendenti e dipendenti calcolando la probabilità legata alle variabili dipendenti in ogni punto della mappa. Secondo il principio della regressione logistica, andrebbero utilizzati dati che si basano sul binomio presenza/assenza: se i dati sono assenti, c'è un significato. Nel caso del settore archeologico, però, questo assunto non è sempre valido. L'assenza di dati che indichino la presenza di un sito in un determinato territorio, infatti, non significa che in quel territorio non ci siano siti di interesse archeologico: semplicemente, non sono stati raccolti dati che indichino questa possibilità. Già dagli anni '60 viene utilizzato nella ricerca archeologica un modello di massima entropia, ma nella maggior parte dei casi viene utilizzato per processi dinamici e interazioni spaziali. L'aspettativa di una variabile

è parallela al concetto matematico di media: un buon primo passo potrebbe essere la comparazione delle aspettative di più variabili. Il principio di massima entropia ci chiede di scegliere tra le distribuzioni più appropriate quella con il massimo valore entropico. L'assunzione di base è che il fenomeno naturale tenda al disordine e al caos; nella realtà, il sito archeologico è il risultato di una scelta fatta dalle persone che cercano di instaurare una relazione con il luogo in cui desiderano stabilirsi. Nello studio in analisi, vengono utilizzati regressione logistica e MaxEnt, abbreviazione di "Maximum Entropy," un metodo statistico utilizzato per modellare la distribuzione di probabilità di un insieme di dati. In archeologia, MaxEnt viene utilizzato per prevedere la localizzazione di siti archeologici basandosi su variabili ambientali e altre caratteristiche del paesaggio. Il principio di base del MaxEnt è che, tra tutte le distribuzioni di probabilità che potrebbero spiegare i dati osservati, viene scelta quella con la massima entropia, ossia quella più "uniforme" o meno informativa, a meno che non ci siano informazioni specifiche che suggeriscano altrimenti. MaxEnt non è esattamente un sistema di regressione, anche se condivide alcune somiglianze con metodi di regressione per quanto riguarda l'analisi dei dati e la costruzione di modelli predittivi. L'area di studio si trova nella regione montuosa dell'Alta Galilea, nel nord di Israele. Questa è la parte più alta della catena montuosa centrale di Israele, con un'elevazione massima di 1208 metri. La regione ha un paesaggio diversificato, composto da tre principali unità geografiche: un altopiano basaltico a est, un massiccio centrale elevato e altipiani calcarei moderati a ovest. La regione si trova nella zona climatica mediterranea, con inverni umidi ed estati secche, con una precipitazione media annuale di 960 mm sulla cresta del Meron e di 550-600 mm in altre aree. Geologicamente, l'area è costituita da diverse formazioni calcaree dal Turoniano-Cenomaniano, Senoniano ed Eocene, con calcare, dolomite, gesso e selce. I tipi di suolo predominanti in questa regione sono le terre rosse (sulle montagne), il suolo basaltico e le terre brune e pallide sui gessi. Le pendici moderate dell'area montuosa sono particolarmente adatte per l'agricoltura mediterranea e sono generalmente piantate con colture invernali come grano e orzo insieme a uliveti e vigneti. I campi agricoli si trovano principalmente su terreni pianeggianti e in piccole valli interne intorno alla catena centrale; di solito in aree di gesso tenero o sull'altopiano basaltico. Dove le pendici sono ripide e il suolo è poco profondo, è necessaria la terrazzatura per la coltivazione. Le

risorse idriche di questa area sono principalmente piccole sorgenti e acqua piovana raccolta in cisterne. L'area di studio copre 378 km<sup>2</sup>, suddivisi in quadrati di 100x100 metri, che servono come unità base di analisi. Nelle aree di studio sono state utilizzate immagini satellitari a pixel di 12.5 metri della Japanese Aerospace Exploration Agency. Queste immagini contengono dati di elevazione, che hanno permesso di creare strati raster di pendenze, orientamento e terreno locale (curvatura del terreno). Si tratta di variabili comuni nei modelli predittivi archeologici, basati sull'assunzione logica che le persone tendevano a vivere in aree moderate. Il parametro della curvatura del terreno è stato prodotto utilizzando la funzione di curvatura del GIS (ArcGIS 10.4 di ESRI), che calcola e mette in evidenza i diversi aspetti della forma della pendenza, e presenta le differenze tra ogni zona e quelle vicine. Per valutare quanto sia efficace un modello predittivo, ci sono vari approcci. Il metodo più affidabile è usare un gruppo di controllo, cioè un campione di siti archeologici che non è stato utilizzato per creare il modello. Se l'80-85% dei siti del gruppo di controllo si trova nelle aree indicate come ad alta probabilità dal modello, allora il modello è considerato efficace. Inoltre, si confronta questo gruppo di controllo con i dati di addestramento usati per costruire il modello per assicurarsi che preveda correttamente l'80-85% dei siti all'interno di meno del 33% dell'area di studio totale. Un concetto importante è la probabilità di trovare un sito archeologico in una posizione casuale all'interno dell'area di studio, che chiamiamo probabilità casuale. Questa probabilità si calcola dividendo l'area totale dei siti conosciuti per l'area totale di studio. In entrambi i modelli presentati, questa probabilità è simile: 0,4% nella Galilea Superiore e 0,37% a Fuxin. Studi regionali in Nord America hanno mostrato valori simili, tra 0,1% e 0,4%. L'efficacia di un modello si valuta anche considerando quanto terreno viene classificato come ad alta probabilità di contenere siti. Un modello è più preciso se riesce a identificare correttamente le aree ad alta probabilità senza includere troppa area inutile. Un modo per misurare la precisione del modello è vedere quanto sia migliore rispetto alla casualità. Ad esempio, se il modello è molto più bravo a trovare siti rispetto a una ricerca casuale, allora è considerato molto efficace. Infine, un altro metodo per valutare l'efficacia è l'area sotto la curva ROC (AUC). Questa è una misura che indica quanto bene il modello predice i siti rispetto a un modello casuale. Valori alti indicano un'alta precisione del modello. In questo caso, l'efficacia relativa di

ciascun metodo per ogni caso di studio è stata verificata utilizzando diverse soglie di previsione, che coprono dall'80% al 95% dei dati di addestramento (75% del dataset) rispetto a un gruppo di controllo (25%). Per ciascuna soglia, viene indicata la percentuale del gruppo di controllo che è correttamente prevista dal modello e la percentuale di area dell'intero studio coperta da ciascun modello. Nello studio vengono presentati due modelli specifici per due aree distinte. L'area di studio dell'Alta Galilea copre 37.829 ettari e contiene 54 siti dell'età del Bronzo e del Ferro, con una superficie totale di 153 ettari. La probabilità di trovare un sito in un campione casuale è dello 0,4%. In entrambi i modelli, sono state utilizzate soglie che prevedono dall'80% al 95% dei siti noti dell'area di studio all'interno delle aree di alta probabilità. Quando viene utilizzata una soglia dell'85%, il modello di regressione logistica copre il 37% dell'area totale e prevede correttamente l'80% dei siti del gruppo di controllo. Il modello MaxEnt, con la stessa soglia, copre solo il 13% dell'area totale ma prevede correttamente l'86% dei siti del gruppo di controllo. L'area di studio di Fuxin copre 21.842 ettari e contiene 111 unità di raccolta neolitiche con una superficie totale di 81,9 ettari. La probabilità casuale di trovare un sito in un campione casuale è dello 0,37%. Anche in questo caso, vengono utilizzate soglie che prevedono dall'80% al 95% dei siti noti. Quando utilizziamo una soglia dell'85%, il modello di regressione logistica copre il 37% dell'area totale e prevede correttamente l'84% dei siti del gruppo di controllo. Il modello MaxEnt, con la stessa soglia, copre solo il 10% dell'area totale e prevede correttamente l'84% dei siti del gruppo di controllo (Wachtel, Zidon, Garti, & Shelach-Lavi, 2018).

### **2.1.2 Support Vector Machines**

Le SVM (Support Vector Machines) sono un tipo di machine learning utilizzato sia per la classificazione che per la regressione. Sono principalmente impiegate per problemi di classificazione. Questi algoritmi creano un iperpiano utilizzato per categorizzare le diverse classi<sup>3</sup>. L'iperpiano è generato iterativamente dal SVM con l'obiettivo di

---

<sup>3</sup> Un iperpiano è un concetto geometrico che generalizza il concetto di un piano nello spazio multidimensionale. In termini più semplici, un iperpiano è la generalizzazione di un piano in uno spazio di dimensioni superiori a tre. Nello spazio tridimensionale, un piano è una superficie piatta di dimensione due. Tuttavia, quando si passa a spazi di dimensioni superiori, un piano diventa un iperpiano. Ad esempio, in uno spazio a quattro dimensioni, un iperpiano sarebbe una superficie tridimensionale, mentre in uno spazio a cinque dimensioni, sarebbe una struttura a quattro dimensioni e così via.

minimizzare l'errore. Lo scopo è dividere i dataset in diverse classi per individuare un iperpiano marginale massimo (MMH). Il MMH può essere individuato seguendo i seguenti passaggi: L'SVM crea iperpiani in modo iterativo che separano le classi nel modo più ideale. Successivamente, seleziona l'iperpiano che separa accuratamente le classi. Ad esempio, consideriamo due etichette, blu e nero, con caratteristiche di dati  $p$  e  $q$ . Il classificatore è specificato da una coppia di coordinate  $(p, q)$  che restituisce sia blu che nero. L'SVM considera i punti dati che generano l'iperpiano che separa le etichette. Questa linea è definita come un confine decisionale. Tutto ciò che cade da un lato della linea verrà classificato come blu, mentre tutto ciò che cade dall'altro lato sarà classificato come nero (Shetty, Singh, & Rao, 2022). Questo metodo è stato utilizzato per l'analisi di manoscritti medievali, per i quali nel corso degli anni si è sviluppata una banca dati di manoscritti digitali consistente e pubblica, che permette l'estrazione rapida di informazioni che facilita enormemente il lavoro di studiosi e ricercatori. Negli ultimi anni, un gran numero di manoscritti storici è stato digitalizzato e reso pubblico, creando numerose biblioteche digitali in tutto il mondo. Per queste vaste collezioni, c'è una crescente necessità di tecniche automatiche assistite dal computer che siano in grado di eseguire un'analisi rapida e intelligente dei documenti per estrarre vari tipi di informazioni, come le linee di testo e le lettere maiuscole. Una volta estratte queste informazioni di interesse, gli studiosi possono condurre studi sui manoscritti in modo più efficiente e approfondito. Sebbene le tecniche manuali o semi-automatiche possano essere utilizzate per analizzare dataset più piccoli, una volta che questi raggiungono una certa dimensione, tali tecniche diventano proibitivamente costose in termini di tempo e lavoro. Ad esempio, quando si trattano dataset su larga scala che contengono una notevole variabilità nella struttura fisica dei manoscritti, le tecniche non automatiche non sono desiderabili poiché richiedono generalmente impostazioni di parametri distintivi per produrre buoni risultati. Sebbene alcuni metodi automatici per l'analisi dei manoscritti medievali siano stati recentemente proposti, essi funzionano solo su base per libro/manoscritto. In altre parole, richiedono la disponibilità di più pagine dello stesso manoscritto per addestrare un classificatore dipendente dal manoscritto. Questo problema di dipendenza potrebbe limitare il loro campo di applicazione. Pertanto, sono molto richiesti algoritmi che possano funzionare su base per pagina. Vengono evidenziati in particolare due fattori complicanti: i manoscritti

medievali generalmente presentano strutture fisiche complesse, come uno stile di scrittura flessibile e fori, e spesso hanno subito un significativo degrado a causa dell'invecchiamento, della manipolazione frequente e delle condizioni di conservazione. Di conseguenza, è generalmente più difficile progettare un algoritmo per l'analisi dei manoscritti medievali rispetto ai documenti moderni stampati a macchina. Tuttavia, sono stati fatti tentativi per sviluppare tecniche assistite dal computer per l'analisi del layout dei manoscritti medievali, concentrandosi principalmente sul riconoscimento delle parole, la segmentazione delle parole e l'estrazione delle linee di testo. Sono stati sviluppati tre algoritmi completamente automatici, basati su singole pagine, per l'analisi dei manoscritti medievali, composti da tre fasi: calcolo dei blocchi di testo, segmentazione delle linee di testo ed estrazione di componenti speciali come lettere maiuscole, decorazioni e miniature. Con "manoscritti medievali" ci si riferisce a libri preparati professionalmente prima dell'avvento della stampa meccanica. Tali libri sono stati preparati da professionisti che, come fase della preparazione del manoscritto, rigavano la pergamena prima di scrivere, garantendo layout generalmente regolari e caratteristiche stabili. Migliaia di questi libri, prodotti professionalmente a mano, sono sopravvissuti e sono oggetto di studio da parte degli studiosi. Ad esempio, per il Libro delle Ore, utilizzato per i test, esistono almeno 800 copie sopravvissute. Gli studiosi sono interessati a trovare variazioni in queste copie che riflettano l'individualità nella loro produzione. L'obiettivo del lavoro in analisi è creare un framework affidabile per eseguire l'analisi del layout dei documenti sui manoscritti medievali. Sebbene vengano fatte alcune assunzioni riguardo all'altezza/larghezza del testo medievale, il framework è altamente modulare e alcuni passaggi sono indipendenti dalle assunzioni, rendendolo adattabile ad altri stili di scrittura trovando, ad esempio, una nuova assunzione per un particolare tipo di manoscritto. Il processo per estrarre i blocchi di testo da un'immagine di un manoscritto medievale avviene in due fasi. Nella prima fase, si ottengono blocchi di testo grezzi analizzando i profili di proiezione. Partendo da un'immagine a colori, viene calcolata la sua corrispondente immagine binaria. In questa immagine, il profilo di proiezione, ottenuto sommando i valori dei pixel lungo le righe, differisce significativamente tra le aree di testo e le aree non di testo.

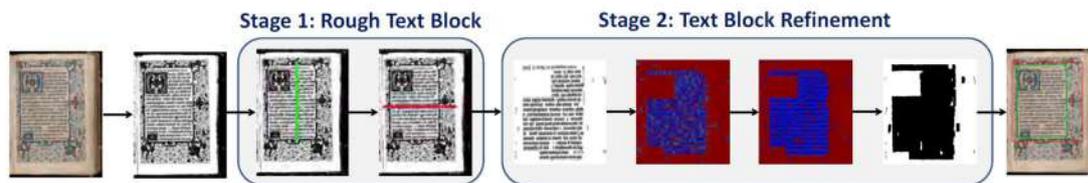


Figura 5 Il processo di calcolo dei blocchi di testo avviene in due fasi. Inizialmente, si analizzano i profili di proiezione per ottenere il blocco di testo grezzo, i cui intervalli verticali e orizzontali sono evidenziati rispettivamente da linee verdi e rosse. Successivamente, viene proposta una strategia di raffinamento per rimuovere le regioni non desiderate che non contengono testo dal blocco grezzo ottenuto, al fine di produrre un blocco di testo finale. L'immagine più a destra mostra i blocchi di testo prima (in rosso) e dopo (in verde) il raffinamento (Yang, Pintus, Gobbetti, & Rushmeier).

Analizzando questo profilo, si possono identificare le aree di testo grezze. Una volta ottenuti i blocchi di testo grezzi, è necessario rimuovere le regioni non desiderate, come decorazioni o margini, per affinare i blocchi di testo. Questo si fa utilizzando una tecnica di template matching basata sull'idea che il testo di una pagina di manoscritto abbia una forma rettangolare. Si crea un'immagine binaria del testo e si utilizzano mappe di distanza in direzione orizzontale e verticale per identificare e unire i pixel del testo, migliorando così l'accuratezza dei blocchi di testo finali. Il testo della pagina del manoscritto viene identificato usando una tecnica di template matching che sfrutta la forma rettangolare del testo. Una volta definita un'immagine binaria del testo, si calcolano le mappe di distanza in direzione orizzontale e verticale. Queste mappe aiutano a identificare i pixel del testo e a chiudere eventuali spazi vuoti tra i pixel non connessi, formando così componenti connessi più grandi e accurati. Infine, si estraggono tutti i componenti connessi dall'immagine binaria e si considerano come blocchi di testo validi se soddisfano determinati criteri dimensionali. Questo processo garantisce che i blocchi di testo siano ben definiti e privi di regioni non desiderate.

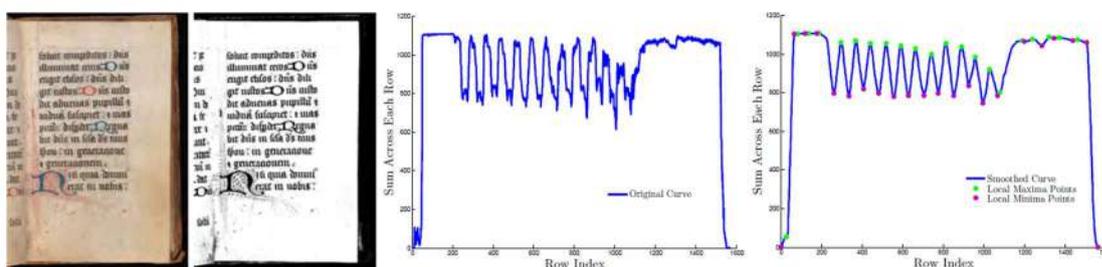


Figura 6 Da sinistra a destra: immagine originale, immagine binaria, profilo di proiezione originale ottenuto sommando i valori dei pixel lungo le righe e profilo di proiezione levigato (Yang, Pintus, Gobbetti, & Rushmeier).

L'estrazione delle linee di testo si basa sull'analisi dei blocchi di testo calcolati nella sezione precedente. Specificamente, le linee di testo vengono estratte dalle immagini

binarie dei blocchi di testo, anziché dall'intera immagine. L'osservazione principale è che lo spazio tra due linee di testo adiacenti genera una fluttuazione a onde nel profilo di proiezione basato sulle righe. Seguendo la stessa idea e utilizzando gli stessi parametri menzionati nella fase precedente, si sommano i valori dei pixel lungo le righe dell'immagine binaria di ciascun blocco di testo, si levigano questi valori di somma e si calcolano i punti massimi locali. Successivamente, si estraggono le regioni dell'immagine delimitate da due punti massimi adiacenti e si considerano queste regioni come linee di testo. Sebbene l'algoritmo di segmentazione delle linee di testo sia basato su immagini binarie, si ottengono prestazioni soddisfacenti (valori di precisione fino al 93,20% e 99,62%). Questo è possibile perché l'algoritmo lavora su blocchi di testo con la conoscenza a priori della struttura fisica della pagina, ovvero l'altezza di ciascun segmento di linea di testo è approssimativamente costante. Successivamente, si passa alla parte più complessa, che si concentra sull'identificazione e l'estrazione delle componenti speciali, se presenti. Poiché sono generalmente colorate e con forme diverse dal testo, si formula questo problema come un problema di clustering. Il processo algoritmico inizia con la selezione di caratteristiche specifiche per ogni pixel, come il colore e le proprietà statistiche, e poi raggruppa i pixel in base a queste caratteristiche. Viene utilizzata una combinazione di algoritmi di clustering, tra cui k-means e l'algoritmo EM per il modello di miscelazione gaussiana, per classificare i pixel. Successivamente, le SVM (Support Vector Machines) vengono addestrate sui dati di training derivati dai risultati del clustering per migliorare la precisione della classificazione. Le immagini di maschera risultanti dai vari metodi di clustering vengono combinate per formare un'immagine binaria finale, dalla quale si estraggono i componenti speciali come componenti connesse che soddisfano determinati criteri dimensionali. Questo processo garantisce che le componenti speciali, come figure e lettere maiuscole, siano correttamente identificate e separate dal testo normale. L'approccio descritto permette di gestire con efficacia la variabilità presente nei manoscritti medievali, migliorando così l'accuratezza dell'analisi del layout dei documenti (Yang, Pintus, Gobbetti, & Rushmeier). Un ulteriore esempio di support vector machine lo si trova in uno studio condotto nel 2016. Nell'ultima decade ci sono stati notevoli progressi nello sviluppo di algoritmi di computer vision per diversi problemi legati al riconoscimento degli oggetti, tra cui il riconoscimento di istanze, la

categorizzazione, il riconoscimento di scene, la stima della posa, ecc. Quando osserviamo un'immagine, non solo riconosciamo le categorie degli oggetti e la categoria della scena, ma possiamo anche dedurre vari aspetti estetici, culturali e storici. Ad esempio, quando guardiamo un dipinto d'arte, un esperto o anche una persona comune può dedurre informazioni sullo stile di quel dipinto (ad esempio, Barocco vs Impressionismo), il genere del dipinto (ad esempio, un ritratto o un paesaggio) o addirittura indovinare l'artista che lo ha dipinto. Le persone possono guardare due dipinti e trovare somiglianze tra di essi in diversi aspetti (composizione, colore, texture, argomento, ecc.). Questa è un'abilità impressionante della percezione umana per apprendere e valutare concetti visivi estetici complessi, che a lungo si pensava non fossero un processo logico. In contrasto, il problema può essere affrontato utilizzando un approccio metodologico computazionale, per dimostrare che le macchine possono effettivamente apprendere tali concetti estetici. Misurare l'influenza artistica è un compito molto difficile a causa dei criteri ampi su cosa possa significare l'influenza tra gli artisti. Come già menzionato, ci sono molti modi diversi in cui i dipinti possono essere descritti. Alcune di queste descrizioni possono essere tradotte in input per un computer. Ad esempio non ci concentriamo su un elemento specifico dell'arte o su un principio dell'arte, ma invece ci concentriamo su trovare e suggerire nuovi confronti sperimentando con diverse misure di somiglianza e caratteristiche. Con l'aumento dei volumi di database d'arte digitalizzati su Internet, si presenta la difficile sfida dell'organizzazione e del recupero dei dipinti. Per gestire correttamente i database di questi dipinti, diventa essenziale classificare i dipinti in diverse categorie e sottocategorie. Sebbene il significato di un dipinto sia unico per ogni artista e completamente soggettivo, può essere in qualche modo misurato dai simboli e dagli oggetti presenti nel dipinto. I simboli sono parole visive che spesso esprimono qualcosa sul significato di un'opera. Ad esempio, le opere degli artisti del Rinascimento come Giovanni Bellini e Jan Van-Eyck utilizzano simboli religiosi come croci, ali e animali per raccontare storie della Bibbia. Ciò dimostra la necessità di una rappresentazione basata sugli oggetti delle immagini. Dovremmo essere in grado di descrivere il dipinto da un elenco di molte classi di oggetti diverse. Avendo una rappresentazione basata sugli oggetti, l'immagine viene descritta a un livello semantico elevato, diversamente da caratteristiche a basso livello come colore e texture, il che facilita suggerimenti di

influenze basati sull'argomento. I dipinti non devono necessariamente assomigliarsi, ma se lo fanno, o hanno oggetti ricorrenti (semantica di alto livello), potrebbero essere considerati simili. Se l'influenza viene individuata osservando caratteristiche simili dei dipinti, diventa importante trovare una buona misura di similarità. Il tempo è anche un fattore essenziale nella determinazione dell'influenza. Un artista non può influenzare un altro artista nel passato. Pertanto, la linearità dei dipinti riduce le possibilità di influenza. Dal punto di vista dell'apprendimento automatico, il problema come un problema di scoperta di conoscenza non supervisionata supportata dall'utilizzo di support vector machines. La metodologia di questo studio si basa su tre componenti: lo studio di diverse rappresentazioni della pittura per determinare quale sia più utile per il rilevamento dell'influenza; la misurazione della similarità tra dipinti; lo studio di diverse misure di similarità tra artisti. I ricercatori hanno raccolto un set di dati contenente 1710 immagini ad alta risoluzione di dipinti di 66 artisti nel periodo 1412-1996 e comprende 13 stili pittorici. Viene raccolto un set di dati di verità fondamentale per il compito delle influenze artistiche, contenente principalmente influenze positive identificate dagli storici dell'arte. Questa verità fondamentale viene utilizzata solo per la valutazione complessiva delle influenze scoperte o suggerite e non viene impiegata nell'apprendimento o nella scoperta di conoscenze. Si sostiene che una rappresentazione semantica di alto livello della pittura sarebbe più utile per il rilevamento delle influenze. Tuttavia, per valutare questa ipotesi, è necessario confrontare le prestazioni di diverse caratteristiche e rappresentazioni nel rilevare influenze rispetto a una verità fondamentale di influenze artistiche, che contenga esempi sia positivi che negativi. A causa delle dimensioni limitate dei dati di verità fondamentale disponibili e della mancanza di esempi negativi, non è utile per confrontare diverse caratteristiche e rappresentazioni. Invece, si fa riferimento a un compito altamente correlato, la classificazione dello stile pittorico. L'ipotesi è che le caratteristiche e le rappresentazioni utili per la classificazione dello stile (un problema di apprendimento supervisionato) siano anche valide per determinare le influenze (un problema non supervisionato). Viene condotto uno studio comparativo completo di diverse caratteristiche e modelli di classificazione per il compito di classificare lo stile pittorico tra sette stili diversi. La conclusione conferma l'ipotesi che le caratteristiche semantiche di alto livello siano più utili per la classificazione dello stile e quindi per determinare le

influenze. L'uso delle caratteristiche corrette per rappresentare il dipinto permette di giudicare la similarità tra i dipinti in modo quantificabile. Un esempio è illustrato dalla nostra metodologia automatizzata, che ha rilevato la somiglianza tra "Studio 9 Rue de la Condamine" di Frédéric Bazille (1870) e "Shueton's Barber Shop" di Norman Rockwell (1950).



Figura 7 F.Bazille, Studio 9, Rue de La Condamine, 1870 (Saleh, Abe, Arora, & Elgammal, 2016)



Figura 8 Norman Rockwell, Shueton's Barber Shop, 1950 (Saleh, Abe, Arora, & Elgammal, 2016)

Questo confronto non è mai stato fatto prima da uno storico dell'arte. Anche se i dipinti potrebbero non sembrare simili a prima vista, un'analisi più attenta rivela una notevole somiglianza nella composizione e nel tema, rilevata dalla metodologia automatizzata. Altri esempi di somiglianza possono essere osservati in ulteriori figure. Molto lavoro è stato fatto sul riconoscimento delle categorie delle immagini, tuttavia, la classificazione dei dipinti si dimostra diversa poiché i dipinti sono differenziati non solo per contenuto, ma anche per lo stile applicato da un particolare pittore o scuola di pittura, o per l'epoca in cui sono stati dipinti. Questo rende la classificazione dei dipinti più sfidante rispetto al normale riconoscimento delle categorie delle immagini. Lo studio affronta il problema della classificazione dello stile pittorico da una prospettiva di apprendimento supervisionato, eseguendo un confronto a due livelli. Il primo livello esamina le prestazioni dei modelli discriminativi rispetto a quelli generativi, mentre il secondo livello confronta le caratteristiche a livello semantico con quelle a basso e medio livello. Sono stati utilizzati sette stili di belle arti: Rinascimento, Barocco, Impressionismo, Cubismo, Astrattismo, Espressionismo e Pop Art. Sono stati eseguiti vari esperimenti per valutare l'accuratezza della classificazione per ciascuna metodologia, che includevano un modello discriminativo con approccio Bag-of-Words (BoW), un modello generativo con approccio BoW e un modello discriminativo utilizzando caratteristiche a livello semantico. Questi modelli differiscono per la metodologia di classificazione e per il tipo di caratteristiche utilizzate per rappresentare il dipinto. Un modello generativo specifica una distribuzione di probabilità congiunta sui campioni osservati e le loro etichette, mentre un modello discriminativo stima direttamente la distribuzione delle etichette dai dati di addestramento, evitando il calcolo della distribuzione di probabilità congiunta. Le caratteristiche a basso livello catturano direttamente elementi formali come colore, texture, bordi e luce. Le caratteristiche a livello intermedio applicano descrittori locali su varie regioni dell'immagine, creando una rappresentazione intermedia con un modello Bag of Words. Le caratteristiche a livello semantico catturano classi di contenuto come acqua, sabbia o automobili presenti nell'immagine, utilizzando un vettore di caratteristiche in cui ogni elemento indica la probabilità di esistenza di una classe semantica.

Lo studio si concentra su caratteristiche a livello intermedio e semantico, con l'ipotesi che queste ultime siano particolarmente utili per la classificazione dei dipinti. Si utilizza Latent Dirichlet Allocation (LDA) per modellare i temi nei dipinti, dove ogni immagine è rappresentata da una miscela di argomenti, e ogni argomento è caratterizzato da una distribuzione sulle parole visive. I vari esperimenti condotti hanno dimostrato che il modello discriminativo con caratteristiche di livello semantico ha ottenuto l'accuratezza più elevata, supportando l'ipotesi che le informazioni di livello semantico siano più adatte per la classificazione fine degli stili artistici. Per la classificazione, il classificatore addestrato viene applicato al vettore delle caratteristiche BoW di un'immagine di test. Il risultato finale di un modello Bag of Words è quindi un istogramma di parole, utilizzato come caratteristica a livello intermedio per rappresentare un dipinto. In questo studio, è stato applicato un classificatore Support Vector Machine (SVM) su un codebook addestrato su immagini del dataset. Sono state utilizzate due varianti delle caratteristiche Scale Invariant Feature Transform "SIFT", chiamate Color SIFT (CSIFT) e opponent SIFT (OSIFT), come caratteristiche locali. Il SIFT è invariante alla scala dell'immagine, alla rotazione, alla distorsione affine e all'illuminazione, utilizzando le orientazioni dei bordi per definire una regione locale e il gradiente dell'immagine. Inoltre, il descrittore SIFT è normalizzato, rendendolo immune ai cambiamenti nella magnitudine del gradiente. CSIFT e opponent SIFT estendono le caratteristiche SIFT alle immagini a colori, essenziali per la classificazione dello stile pittorico. Uno studio precedente ha dimostrato che opponent SIFT supera le altre varianti di colore SIFT nei compiti di categorizzazione delle immagini. Nella ricerca delle caratteristiche da parte del modello discriminativo, è stato estratto il vettore di caratteristiche Classeme come caratteristica visiva per ogni dipinto. Le caratteristiche Classeme sono l'output di un insieme di classificatori corrispondenti a un insieme di etichette di categoria, tratte da un elenco di termini appropriato e non correlate al contesto delle belle arti. Per ogni categoria, è stato raccolto un insieme di immagini di addestramento emettendo una query sull'etichetta di categoria a un motore di ricerca di immagini. Dopo l'estrazione di un insieme di descrittori delle caratteristiche grezze, è stato selezionato un sottoinsieme di dimensioni delle caratteristiche. Utilizzando queste

caratteristiche a dimensioni ridotte, è stato addestrato un classificatore uno-contro-tutti per ciascuna categoria. L'output del classificatore è a valori reali e indica la somiglianza di un'immagine con una classe specifica. Date un'immagine, il vettore delle caratteristiche utilizzato per rappresentarla è il vettore Classeme. La caratteristica Classeme ha una dimensionalità significativa. Tali vettori di caratteristiche sono stati utilizzati per addestrare un classificatore Support Vector Machine (SVM) per ogni genere pittorico. Si ipotizza che le caratteristiche Classeme siano adatte a rappresentare e riassumere i contenuti complessivi di un dipinto, poiché catturano informazioni a livello semantico sulla presenza di oggetti in un dipinto, codificati implicitamente nell'output dei classificatori pre-addestrati. Per la task di classificazione di stile delle pitture, ci concentriamo su un sottoinsieme del nostro dataset che contiene sette categorie di pitture, ovvero Astratto, Barocco, Rinascimentale, Pop-art, Espressionismo, Impressionismo e Cubismo. In ogni categoria sono inserite 70 opere pittoriche. Per ciascuno degli esperimenti seguenti è stata eseguita una cross-validazione a cinque fold, con il 20% delle immagini scelte a scopo di test in ogni fold. Per la formazione del codebook, un rilevatore è stato utilizzato per individuare i punti di interesse. Per un calcolo efficiente, il numero di punti di interesse per ogni pittura è limitato a 3000. L'algoritmo di clustering K-means standard viene utilizzato per costruire un codebook di dimensione 600 parole. Il classificatore SVM è addestrato su descrittori di livello intermedio e semantico. Per determinare i parametri per SVM, è impiegato l'algoritmo di ricerca del grigliato, il quale utilizza la cross-validation per individuare i valori ottimali dei parametri. Inoltre, questo processo è preceduto dalla ridimensionamento dei descrittori del dataset. L'accuratezza complessiva raggiunta è del 65,4%. I risultati mostrano che lo stile Barocco è sempre classificato con la massima accuratezza in tutte le tecniche. È interessante notare anche che lo stile Pop-art è classificato con un'accuratezza superiore al 70% in tutti gli approcci discriminanti mentre l'approccio generativo si è comportato male in quel caso (Saleh, Abe, Arora, & Elgammal, 2016).

### **2.1.3 Decision Trees e random forests**

I decision trees hanno avuto origine già dai primi sviluppi della scrittura: risultati eccezionalmente interpretabili con una struttura simile ad un albero, permettono una

migliore comprensione del significato e della distribuzione dei risultati (Ville, 2013). Un albero decisionale classifica i dati ponendo una serie di domande sulle caratteristiche associate ai dati stessi. I gruppi di decision trees dipendono dai valori dei singoli elementi. Utilizzando strategie specifiche di raccolta informazioni, sono in grado di scoprire quali elementi del dataset riescano a fornire il dato migliore, fino a realizzare un modello che splitta un problema complesso in partizioni che, basandosi su regole di decisioni binarie, considerano un certo sotto-gruppo di valori della caratteristica considerata. La limitazione principale di questo sistema è che si tende alla creazione di modelli complessi che non sono facilmente generalizzabili; questo problema può essere aggirato utilizzando le cosiddette random forests. Una random forests è, semplicemente, un insieme di alberi decisionali che ha lo scopo di aumentare la robustezza e affidabilità delle previsioni.

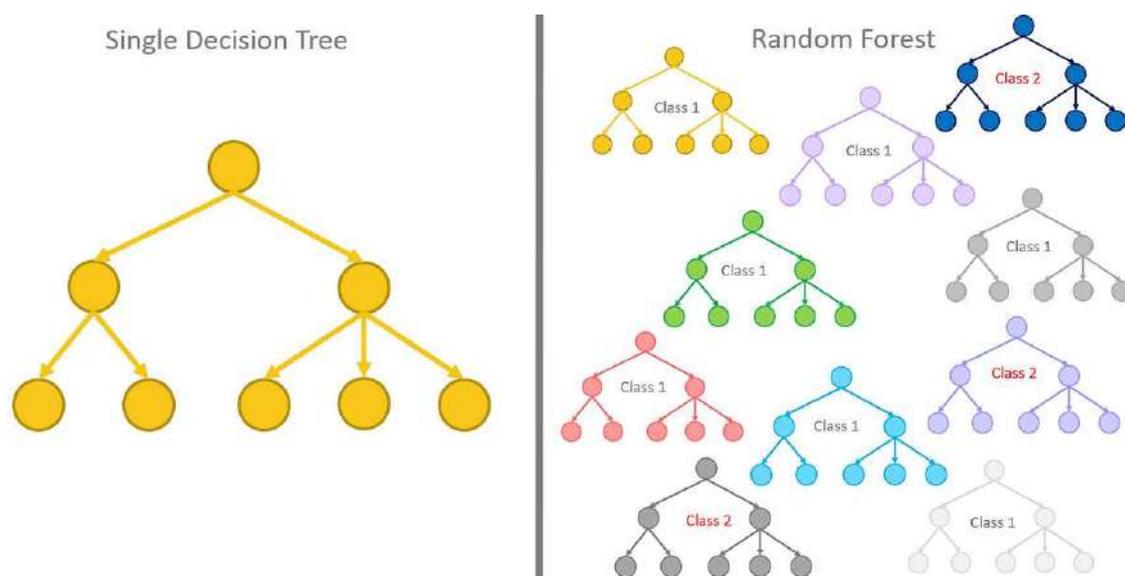
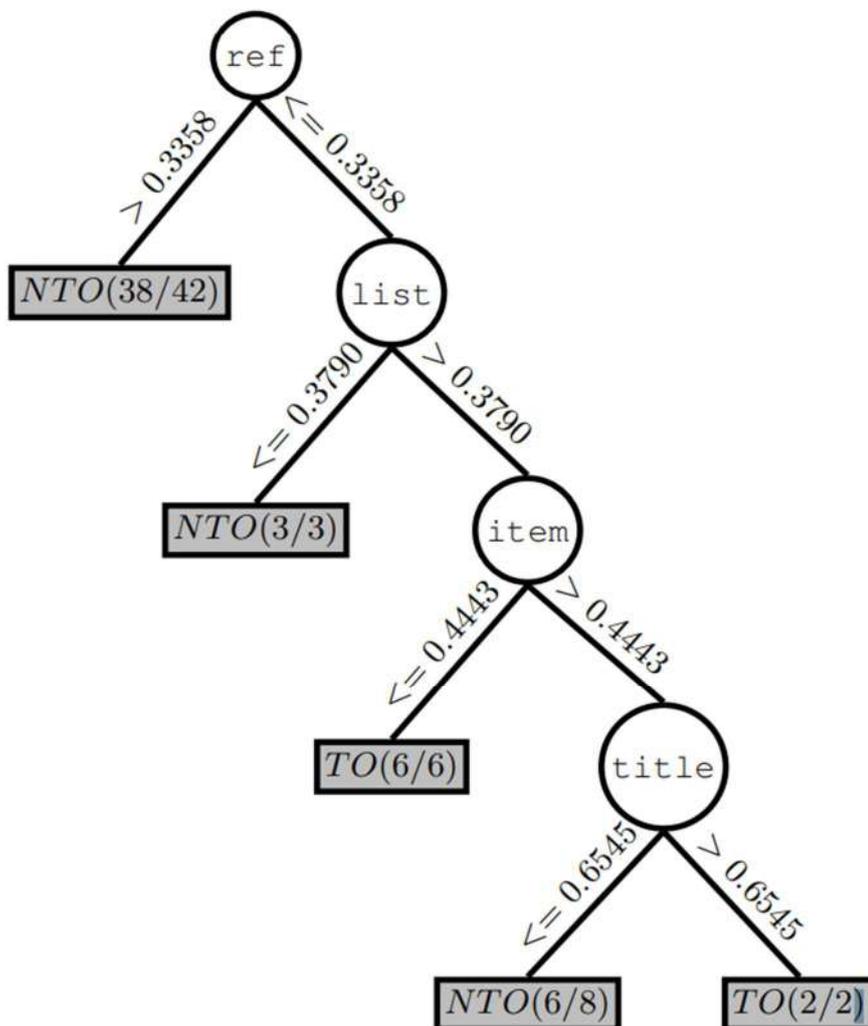


Figura 9 - Decision trees e Random forests (Ville, 2013)

Un albero decisionale è una struttura a cascata composta da nodi e rami. In corrispondenza dei nodi, viene effettuata una spaccatura nei dati in corrispondenza delle caratteristiche di input, generando così due ulteriori rami in output. Questo procedimento continua fino a che tutti o quasi tutti i dati sono appartenenti ad una classe e non sono possibili ulteriori spaccature. Il primo nodo è definito radice; gli ultimi sono chiamati foglie e vengono etichettati. Se è possibile solamente una spaccatura binaria, si parla di alberi binari (Kingsford, 2008). Nell'ambito delle digital humanities, gli alberi decisionali

hanno diversi impieghi. Uno studio presentato nel 2016 alla Language Resources and Evaluation Conference vede l'impiego di decision trees nell'analisi dell'ambiguità di alcuni termini allo scopo di facilitare l'estrazione di informazioni e il riassunto di testi in maniera automatica. Il processo di disambiguazione dei termini viene avviato dopo il task di acquisizione dei termini. Questo task estrae i termini prendendo in considerazione criteri e indici sulla base di un intero corpus di testi. A questo scopo sono stati impiegati diversi modelli di machine learning che hanno estratto informazioni da un dataset comprensivo di 55 documenti, 42 articoli di conferenze e 197000 tokens estratti dal corpus SCIENTEXT. Per la lemmatizzazione del corpus è stato usato il tool TermSuite. Per raggiungere gli scopi prefissati vengono impiegati tre metodi di disambiguazione: approccio su base ipotetica, approccio di specificità di Lafon e approccio in base alla rilevanza. Allo scopo di osservare l'impiego di decision tree nello studio, ci concentreremo sull'approccio di rilevanza. Sono presi ad esame diversi termini ricorrenti, detti termini candidati, per i quali viene controllato se si tratta di termini tecnici (TO), e quindi rilevanti ai fini della comprensione del significato intrinseco del documento, o termini non tecnici (NTO). Le caratteristiche, i lemmi e i temi del discorso si appoggiano a proprietà del genere che sono dette salienti. L'assunto è che i termini siano usati più spesso in posizioni salienti. I testi scientifici contengono solo alcuni termini importanti, i quali compaiono in posizioni salienti allo scopo di facilitare la comprensione del testo stesso per il lettore. Al contrario, gli NTO sono distribuiti più equamente nel documento. Inoltre, il numero di posizioni salienti è limitato, per cui è a maggior ragione difficile che gli NTO occupino posizioni salienti.



Nella figura sopra vediamo un esempio di decision tree ottenuto dall'analisi del termine *aspect*. Questo set di regole è specifico per le occorrenze del termine che non sono classificate grazie all'utilizzo di regole generiche. In questo specifico caso, grazie ad un decision tree, possiamo vedere come la parola *aspect* difficilmente sia un TO se non sono vicine a riferimenti bibliografici rappresentati da un tag del nodo 1 (Daille, Jacquy, Lejeune, Melo, & Touissaint, 2016). Nel corso dei prossimi anni, il machine learning supervisionato rappresenterà il miglior metodo per poter divulgare informazioni tratte dall'enorme quantità di dati che avremo a disposizione anche nel settore culturale. Numerosi ingegneri e figure specializzate entreranno a far parte in maniera sempre più

rilevante dei team che operano nelle digital humanities, con sviluppi che porteranno ad un'efficienza sempre maggiore del settore e delle tecnologie in esso impiegate.

## Capitolo 3 - Machine Learning non supervisionato

Al contrario di ciò che succede per il Machine Learning supervisionato, nel sistema non supervisionato la macchina non riceve degli esempi di output atteso. L'obiettivo di un modello di ML non supervisionato è quello di trovare pattern e strutture in ciò che considereremmo caos non strutturato, allo scopo di costruire una rappresentazione che permetta di prendere decisioni, immaginare input futuri e comunicarli ad altri software. La gran parte del lavoro di un modello di machine learning non supervisionato può essere considerato per la sua possibilità di fornirci un modello probabilistico dei dati (Ghahramani, 2003). Il machine learning non supervisionato è una classe di algoritmi di apprendimento che aiuta a scoprire la struttura intrinseca dei dati senza alcuna supervisione esterna. L'obiettivo di un modello di questo tipo impiegato in ambito culturale è principalmente quello di scoprire specifiche similarità in un gruppo di dati, tramite sistemi di clustering, oppure determinarne la distribuzione, stimandone la densità in aree specifiche (Tyagi, Rane, Sriram, & Manry, 2022). L'apprendimento automatico non supervisionato, a differenza dell'apprendimento supervisionato, non ha bisogno dell'intervento umano a supervisione del modello. Il data scientist abilita il sistema a imparare da solo guardando i dati e identificando i pattern. In altre parole, questa categoria dell'apprendimento automatico consente a un sistema di operare su dati specifici senza istruzioni esterne. Gli approcci di apprendimento non supervisionato sono essenziali per la costruzione di sistemi di intelligenza artificiale che simulano il funzionamento dell'apprendimento umano, poiché le macchine intelligenti devono essere in grado di trarre conclusioni indipendenti basate su enormi quantità di dati non etichettati. Gli algoritmi di machine learning non supervisionato sono migliori nel risolvere compiti considerati complicati rispetto agli algoritmi di apprendimento supervisionato; tuttavia, i modelli di apprendimento supervisionato forniscono risultati più accurati perché un programmatore insegna esplicitamente al sistema cosa cercare nei dati presentati. I modelli non supervisionati, come detto, possono essere utilizzati per scoprire pattern nascosti in enormi quantità di dati, ma anche per categorizzare ed etichettare i punti dati. Le somiglianze e le differenze tra i punti dati non ordinati vengono utilizzate per raggrupparli. Esiste una grande quantità di dati non etichettati, la cui

etichettatura è un'operazione che richiede tempo e che necessita dell'intervento umano. L'intelligenza artificiale può essere utilizzata per gestire questo stesso processo, rendendo la codifica di un grande dataset più semplice. In altre parole, nell'apprendimento supervisionato, uno scienziato dei dati fornisce dati etichettati al sistema, ad esempio fotografie di soggetti artistici etichettate come tali, in modo che la macchina ragioni e impari tramite esempi. Nell'apprendimento non supervisionato vengono presentate le stesse immagini al software ed è compito del sistema esaminare i dati e determinare di quale soggetto artistico di tratti. Grandi quantità di dati sono necessarie per l'apprendimento automatico non supervisionato. Quando gli scienziati dei dati utilizzano set di dati per addestrare gli algoritmi, inizia il processo di apprendimento non supervisionato. Questi set di dati non includono punti dati etichettati o classificati. Lo scopo dell'algoritmo di apprendimento è trovare pattern nel set di dati e valutare i punti dati in base a tali pattern. I problemi di clustering, associazione, rilevamento di anomalie e autoencoder sono quattro tipi di sfide di apprendimento non supervisionato. L'apprendimento non supervisionato è il modo in cui gli esseri umani imparano a identificare e classificare gli oggetti di cui pensano. Le macchine utilizzano l'apprendimento non supervisionato per trovare pattern e classificare in modo simile i punti dati.

### 3.1 Il clustering

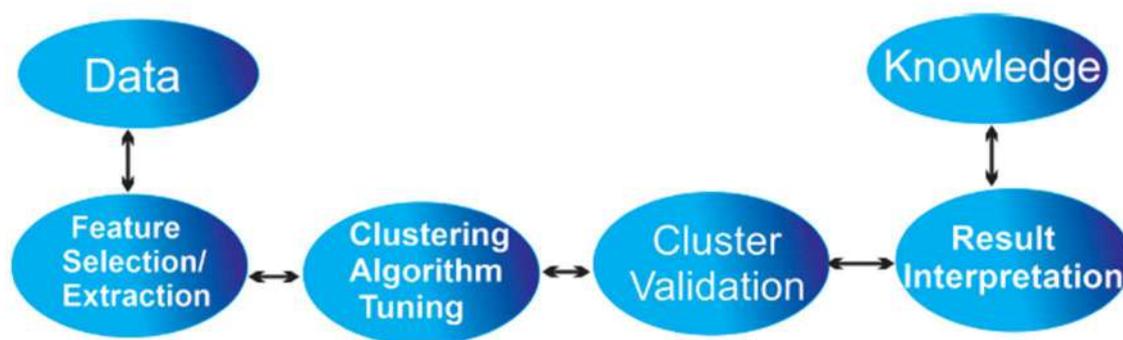


Figura 10 Flusso di lavoro del clustering nell'apprendimento automatico non supervisionato (Naeem, Ali, Anam, & Ahmed, 2023)

Le tecniche classiche per analizzare e/o riassumere set di dati includono il clustering, cioè la suddivisione dei dati in sottoinsiemi con caratteristiche simili (Feldman, Schmidt, & Sohler, 2020). Il clustering può essere suddiviso in diverse tipologie: partizionato, gerarchico, sovrapposto e probabilistico. Nel clustering partizionato, o exclusive pooling, i dati vengono divisi in modo che ogni informazione possa appartenere solo a un cluster; il clustering k-means è un esempio di partizionamento. Il clustering gerarchico, come suggerisce il termine, ha l'obiettivo di costruire una gerarchia di cluster. Al suo interno sono distinti due sottogruppi: clustering agglomerativo, per cui tramite un approccio "bottom up" (dal basso verso l'alto) si parte dall'inserimento di ciascun dato in un cluster differente e si procede all'accorpamento graduale di cluster a due a due; clustering divisivo, nel quale tramite un approccio "top down" (dall'alto verso il basso) tutti gli elementi che si trovano inizialmente in un singolo cluster vengono suddivisi ricorsivamente in sotto-cluster. Il numero di cluster viene ridotto tramite connessioni iterative tra i due cluster più vicini. Viene utilizzato per organizzare i dati in insiemi fuzzy sovrapposti. L'approccio di apprendimento non supervisionato dell'Association Rule Learning (ARL) viene utilizzato per scoprire le associazioni tra le variabili in enormi set di dati. Questo sistema può accettare dati non numerici, a differenza di alcuni metodi di apprendimento automatico specifici. In breve, si occupa di analizzare se e come alcune particolari variabili siano collegate. Le regole di associazione possono evidenziare connessioni tra insiemi di dati disparati. Un sistema di association rule consente l'analisi del carrello virtuale degli utenti; ad esempio è possibile verificare se e quanto spesso all'acquisto del biglietto di ingresso di un museo è legato l'acquisto di un souvenir, o di un biglietto per un altro luogo di cultura della stessa città. Un ulteriore modello di machine learning non supervisionato permette di rilevare anomalie. Queste anomalie potrebbero suggerire un'attività di rete insolita, un sensore difettoso o dati che devono essere riordinati e ripuliti prima di essere utilizzati a scopo di analisi. Quando i modelli di dati vanno oltre o si discostano dai modelli normali, si tratta di un'anomalia. L'individuazione delle anomalie viene effettuata trovando o prevedendo punti dati che differiscono dal modello standard (Naeem, Ali, Anam, & Ahmed, 2023). Come abbiamo accennato in apertura, gli algoritmi di clustering sono tra i metodi di apprendimento supervisionato più utilizzati nel settore culturale. I metodi di clustering assegnano dati a gruppi, chiamati

appunto cluster, in modo che le similarità a coppie tra i punti assegnati allo stesso cluster tendano ad essere più alte rispetto a quelle in cluster diversi. È importante notare che la riduzione delle dimensioni dei cluster è ampiamente utilizzata all'interno del settore culturale, sebbene venga principalmente impiegata come passaggio di pre-elaborazione o come strumento di visualizzazione per rappresentare informazioni (Fiorucci, et al., 2020). Il clustering k-means è uno degli algoritmi più basilari e utilizzati. Il k-means implica il calcolo di una misura di distanza e il calcolo dei vettori centrali per K diverse misure di distanza (Tyagi, Rane, Sriram, & Manry, 2022).

### **3.2 Impieghi dell'apprendimento non supervisionato nel settore culturale e umanistico**

Un esempio di impiego di apprendimento supervisionato si trova nello studio del 2022 condotto da Giovanna Castellano e Gennario Vessio, che si sono occupati di indagare un sistema di data retrieval e successivo clustering per l'analisi di opere pittoriche allo scopo di inserire le informazioni in un modello di computer vision codificato tramite knowledge graph. La capacità di riconoscere caratteristiche, similarità, ecc., in e fra opere d'arte digitalizzate ricade inerentemente nel dominio della percezione estetica umana. Giacché tale percezione è fortemente soggettiva, e influenzata da diversi fattori, fra cui la conoscenza storica relativa all'opera, o l'emozione che l'opera suscita in chi l'osserva. Nelle analisi delle arti visive, un aspetto fondamentale consiste nella ricerca di relazioni di "similarità" tra opere di artisti e scuole pittoriche diverse. Queste relazioni aiutano gli storici dell'arte a individuare e comprendere meglio le influenze artistiche e i cambiamenti culturali tra i vari movimenti artistici. Gli esperti d'arte tendono a studiare le opere visive non come singole creazioni isolate, ma all'interno di contesti più ampi, considerando le influenze e le connessioni tra diverse scuole. Tradizionalmente, questa analisi viene condotta manualmente attraverso l'esame di grandi collezioni di fotografie annotate. Tuttavia, questo processo può essere difficile e dispendioso, data la vastità delle immagini disponibili e la loro distribuzione tra periodi e scuole pittoriche diverse. Tramite reti neurali convoluzionali profonde sono state estratte caratteristiche significative dalle immagini delle opere d'arte digitalizzate, e grazie ad un meccanismo non supervisionato di nearest neighbor sono stati trovati collegamenti tra tali opere. La ricerca dei collegamenti "visuali" è completamente non supervisionata, il che rende il

metodo particolarmente utile nei casi in cui i metadati siano scarsi, non disponibili o difficili da ottenere. Questo tipo di approccio non solo restituisce le immagini più simili alla query in input, ma consente anche di studiare modelli storici analizzando un "grafo delle influenze" costruito sui collegamenti ritrovati. Il metodo non è vantaggioso solo per gli storici dell'arte, ma può essere utile anche per gli appassionati che desiderano esplorare collezioni digitalizzate di musei e gallerie d'arte online. Il recupero automatico di collegamenti facilita l'accesso e la fruizione delle collezioni digitali, migliorando così l'esperienza degli utenti.



Figura 11 A sinistra, immagini in input al sistema; a destra, corrispondenti immagini restituite in output (Castellano & Vessio, 2022)

Sebbene il metodo descritto nel paragrafo precedente sia efficace nel trovare opere d'arte collegate visivamente, mostra limiti nel raggruppare le opere in cluster distinti. Questo

perché i dati tendono a distribuirsi uniformemente all'interno di un unico cluster omogeneo nello spazio delle caratteristiche. Un modello in grado di raggruppare le opere d'arte senza dipendere da etichette difficili da ottenere o da un giudizio soggettivo umano potrebbe essere vantaggioso per molte applicazioni. Ad esempio, potrebbe aiutare a individuare periodi diversi nella produzione di uno stesso artista o a classificare l'arte contemporanea, che spesso non è annotata in modo esaustivo. Per ovviare al problema, gli studiosi hanno sviluppato un metodo che combina l'uso di una rete neurale convoluzionale profonda pre-addestrata per l'estrazione delle caratteristiche con una componente di clustering profondo basata su un auto-encoder. Questa soluzione è stata adottata a causa delle difficoltà nell'applicare metodi tradizionali di clustering o riduzione delle dimensioni sia allo spazio dei pixel in input, a causa della sua alta dimensionalità, sia allo spazio delle caratteristiche risultante dall'embedding di una rete neurale, specialmente quando si lavora con immagini d'arte complesse. I risultati sperimentali, sia quantitativi che qualitativi, hanno dimostrato che il metodo proposto è in grado di individuare cluster ben distinti sia in insiemi di dati ampi e eterogenei che comprendono diversi periodi, sia in opere di uno stesso artista. In particolare, il modello sembra considerare non solo lo stile artistico per raggruppare le opere, ma anche gli attributi semantici relativi al contenuto della scena rappresentata, come il genere dell'opera. Questa capacità sembra promettente nel risolvere il problema della cross depiction, che rimane una sfida aperta per la comunità scientifica. Infatti, potrebbe essere utilizzata per individuare somiglianze tra opere d'arte indipendentemente dal loro stile di rappresentazione.



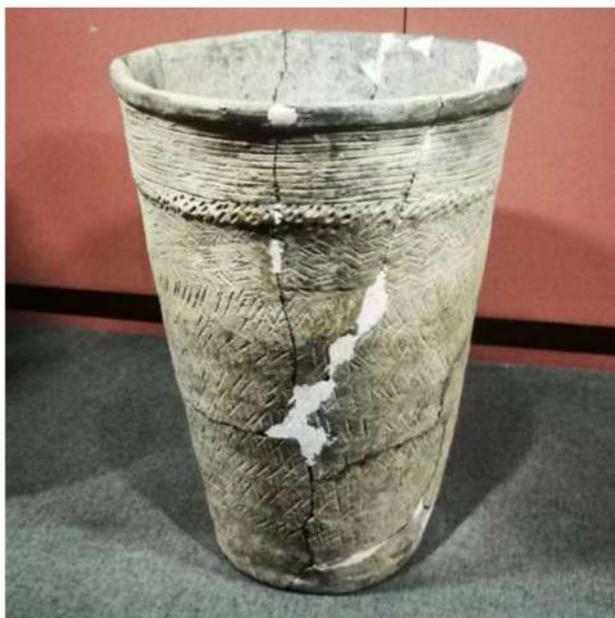
*Figura 12 Cluster individuati automaticamente dal sistema tra opere prodotte da Pablo Picasso.*

La ricerca ha evidenziato un'altra considerazione importante: gran parte degli studi nella letteratura si basa esclusivamente sull'informazione codificata dai pixel caratteristici dei dipinti digitalizzati. Questo approccio, purtroppo, trascura una vasta quantità di conoscenza del settore, nonché le relazioni e le connessioni note tra opere d'arte e/o artisti, che potrebbero migliorare la qualità delle soluzioni esistenti. Le opere d'arte non possono essere comprese solo attraverso il loro aspetto visivo, ma devono essere considerate anche in relazione a vari altri fattori storici, sociali e contestuali che le inquadrano in un contesto più ampio. Pertanto, avere una base di conoscenza che unifichi non solo le opere d'arte, ma anche una vasta gamma di metadati, informazioni contestuali, descrizioni testuali, ecc., all'interno di un unico framework strutturato potrebbe essere una risorsa preziosa per lo sviluppo di modelli più efficaci. Questo framework sarebbe utile non solo per gli utenti generici, che potrebbero utilizzare le informazioni codificate

per navigare nella base di conoscenza, ma anche per gli esperti d'arte, che sono interessati a scoprire nuove relazioni tra opere d'arte e/o artisti per una migliore comprensione dell'arte passata e contemporanea. Vessio e Castellano si sono occupati anche dello sviluppo di ArtGraph, un knowledge graph di stampo artistico. Un knowledge graph fornisce una rappresentazione più espressiva e flessibile per incorporare relazioni di complessità arbitraria tra entità che riguardano l'arte, che non possono essere ottenute considerando solo il loro aspetto visivo. In particolare, il knowledge graph proposto integra le informazioni raccolte da WikiArt e DBpedia e sfrutta le potenzialità di un database NoSQL, Neo4. In questo modo, il database NoSQL aiuta già a fornire uno strumento per la scoperta di conoscenza senza addestrare esplicitamente un algoritmo di apprendimento. La conoscenza codificata in ArtGraph può essere poi integrata con feature apprese automaticamente per affrontare vari task, tra cui la previsione di attributi come lo stile e il genere. Man mano che le applicazioni degli algoritmi di Computer Vision nel dominio artistico diventano sempre più mature, un'implementazione interessante di queste tecniche in contesti reali è l'incorporamento all'interno dei cosiddetti "robot sociali". Questi robot rappresentano un campo di ricerca emergente focalizzato sullo sviluppo di una "intelligenza sociale" che mira a mantenere l'illusione di interagire con un essere umano. In questo contesto, i recenti progressi nella Computer Vision consentono ai ricercatori di dotare i robot di nuove e potenti capacità. Nella ricerca in analisi è stato utilizzato un robot sociale, Pepper, come guida turistica di un museo, con un approccio basato sulla visione per supportare i turisti durante la visita. Pepper è un robot semi-umanoide su ruote, dotato di diverse telecamere e sensori. Il modulo di visione consente a Pepper di percepire la presenza dei visitatori e di localizzarli nello spazio, stimandone età e genere. Inoltre, lo stesso modulo di visual link retrieval descritto in precedenza offre a Pepper la capacità di utilizzare l'immagine del dipinto osservato dal visitatore come query visiva per cercare dipinti visivamente simili all'interno dello stesso museo. Il robot utilizza questi dati e altre informazioni acquisite durante il dialogo per fornire ai visitatori consigli su opere d'arte simili a cui potrebbero essere interessati. Progettare i comportamenti di un robot sociale che funga da guida museale richiede di dotarlo di diverse competenze per fornire ai visitatori un'esperienza coinvolgente ed efficace durante la visita. Queste funzionalità hanno lo scopo di consentire al robot di rilevare e

localizzare le persone nel museo, riconoscere l'opera d'arte che il visitatore sta guardando, profilare l'utente durante la visita al fine di generare raccomandazioni adeguate e coinvolgere le persone nell'interazione utilizzando adeguate capacità di conversazione (Castellano & Vessio, 2022). Non solo all'interno di musei è possibile applicare sistemi non supervisionati, ma anche in ambito di ricerca. Un ulteriore esempio di impiego di apprendimento non supervisionato, e in particolare di clustering, è stato tentato anche in ambito archeologico: lo si trova infatti nello studio per la classificazione di manufatti archeologici cinesi presentato quest'anno alla International Conference on Software Engineering and Information Management. Wang Hao sviluppa un modello di apprendimento non supervisionato che si serve di sistemi di clustering. Il gruppo di ricerca non riesce ad ottenere una classificazione degli artefatti museali con precisione e recall superiori al 70%, ma è interessante osservare come viene utilizzato questo strumento a scopo di ricerca. Come sappiamo, la classificazione dei reperti museali è un lavoro tedioso e professionale che richiede un'ampia quantità di lavoro per mano di studiosi esperti. Senza una profonda esperienza, infatti, i neofiti non sarebbero in grado di classificare facilmente diversi reperti nelle categorie corrette. Per sviluppare lo studio di seguito riportato, sono state raccolte 215 foto di reperti museali esposti presso il Museo Archeologico Cinese in Cina e applicati modelli non supervisionati per classificare i dati delle immagini. L'obiettivo è stato quello di classificare automaticamente i reperti museali con il minor lavoro manuale possibile. Per classificare le foto si è ricorsi inizialmente alla classificazione diretta delle foto senza alcun passaggio di preelaborazione, come input per l'algoritmo di clustering k-means. Tuttavia, dallo studio emerge come lo sfondo e l'illuminazione delle foto influenzino profondamente la qualità dei risultati: è stato quindi aggiunto un passaggio di preelaborazione consistente nella sottrazione dello sfondo prima dell'esecuzione dell'algoritmo.

Le figure sotto mostrano il confronto della foto del reperto prima e dopo la sottrazione dello sfondo:



*Figura 13 (Wang, 2024)*



*Figura 14 (Wang, 2024)*

Dall'elaborazione fotografica emerge come algoritmi moderni di sottrazione dello sfondo possono portare a prestazioni molto elevate sulle foto scattate con smartphone, come illustrato nella Figura 2. Grazie alla sottrazione dello sfondo, si nota un miglioramento delle prestazioni del nostro approccio. Dopo aver sottratto lo sfondo dalle foto dei reperti museali, è stato impostato il numero di cluster a 5 (Esperimento 1), 8 (Esperimento 2) e 9 (Esperimento 3). Nell'esperimento 1 (Numero di cluster: 5) le giare a forma di botte nella Figura 1 e nella Figura 2 sono classificate in 2 categorie diverse: una con 5 giare del tipo, l'altra con 10 giare del tipo. Tuttavia, la categoria di 10 reperti comprende un totale di 37 reperti. Se si considera questa categoria come la categoria vera per questo tipo, l'accuratezza è del 27% e il richiamo è del 67%. L'accuratezza è troppo bassa, mentre il richiamo è solo un numero sufficiente nel campo dell'apprendimento automatico. Se prendiamo i 5 reperti come la categoria vera, le prestazioni sono ancora peggiori. Controlliamo un'altra categoria classificata dal nostro algoritmo. Questa categoria è categorizzata da reperti simili alla Figura 3, ovvero scodelle di ceramica. L'accuratezza di questa categoria è del 30%, e il richiamo è del 67%. La nostra conclusione per questa categoria è la stessa di prima: l'accuratezza è troppo bassa mentre il richiamo è solo sufficiente. Questo non sembra molto utile per gli archeologi nel loro lavoro.

	Barrel-shaped Jar	Pottery Bowl	Bronze Object
Accuracy	27%	30%	32%
Recall	67%	67%	28%

Tabella 1 (Wang, 2024)

Viene poi controllata la distribuzione degli oggetti in bronzo in diverse categorie. Gli oggetti in bronzo hanno un valore di accuratezza del 32% e un valore di richiamo del 28%. Entrambi i valori sono troppo bassi perché gli oggetti in bronzo sono distribuiti su 5 categorie con distribuzioni quasi uniformi. È molto difficile identificare una singola categoria che sia composta in maniera dominante da manufatti in bronzo. Sulla base delle osservazioni precedenti, risulta quasi impossibile categorizzare correttamente diversi

reperiti utilizzando solo la metrica di accuratezza quando consideriamo solo i materiali e (principalmente) la forma dei reperti. Ripristinando il numero di cluster a 8, otteniamo i seguenti risultati: Per le giare a forma di botte illustrate in Fig. 1 e Fig. 2, l'accuratezza è del 44% e il richiamo è del 67%. Questo rappresenta un notevole miglioramento rispetto all'esperimento precedente, tuttavia, l'accuratezza rimane un numero significativamente basso. Al contrario, per le scodelle di ceramica (Fig. 3), l'accuratezza aumenta al 53%, e il richiamo sale all'89%.

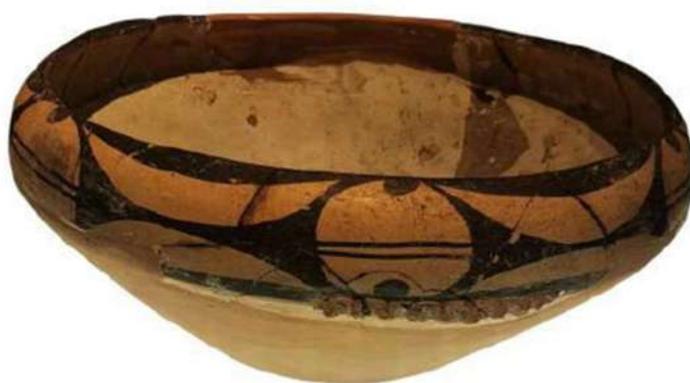


Figura 15 (Wang, 2024)

Evidentemente, le prestazioni dell'algorithm sono state notevolmente migliorate.

	Barrel-shaped Jar	Pottery Bowl	Bronze Object
Accuracy	44%	53%	22%
Recall	67%	89%	33%

Tabella 2 (Wang, 2024)

Per gli oggetti in bronzo (Fig. 4) l'accuratezza è del 22% e il richiamo è del 33%. L'accuratezza peggiora mentre il richiamo diventa più alto. Dalle osservazioni emerse da questa analisi, la classificazione degli oggetti in bronzo è il compito più difficile, perché sono oggetti distribuiti tra diverse categorie. È persino molto difficile identificare la categoria che raccolga in maggior quantità gli oggetti in bronzo.

	Barrel-shaped Jar	Pottery Bowl	Bronze Object
Accuracy	60%	35%	22%
Recall	38%	78%	43%

Tabella 3 (Wang, 2024)

Impostando il numero di cluster a 9, i risultati vengono aggiornati come segue: Per gli artefatti del tipo illustrato in Fig. 1 e Fig. 2, l'accuratezza è del 60% e il richiamo è del 38%. Per le ciotole di terracotta (Fig. 3), l'accuratezza diventa del 35% e il richiamo del 78%, quindi un peggioro rispetto all'esperimento precedente. Se esaminiamo la categoria degli oggetti in bronzo l'accuratezza è del 22% e il richiamo del 43%. Abbiamo affrontato le stesse difficoltà nella classificazione degli oggetti in bronzo come prima, eccetto che il richiamo è migliorato notevolmente.



*Figura 16 Ding rettangolare in bronzo (Wang, 2024)*

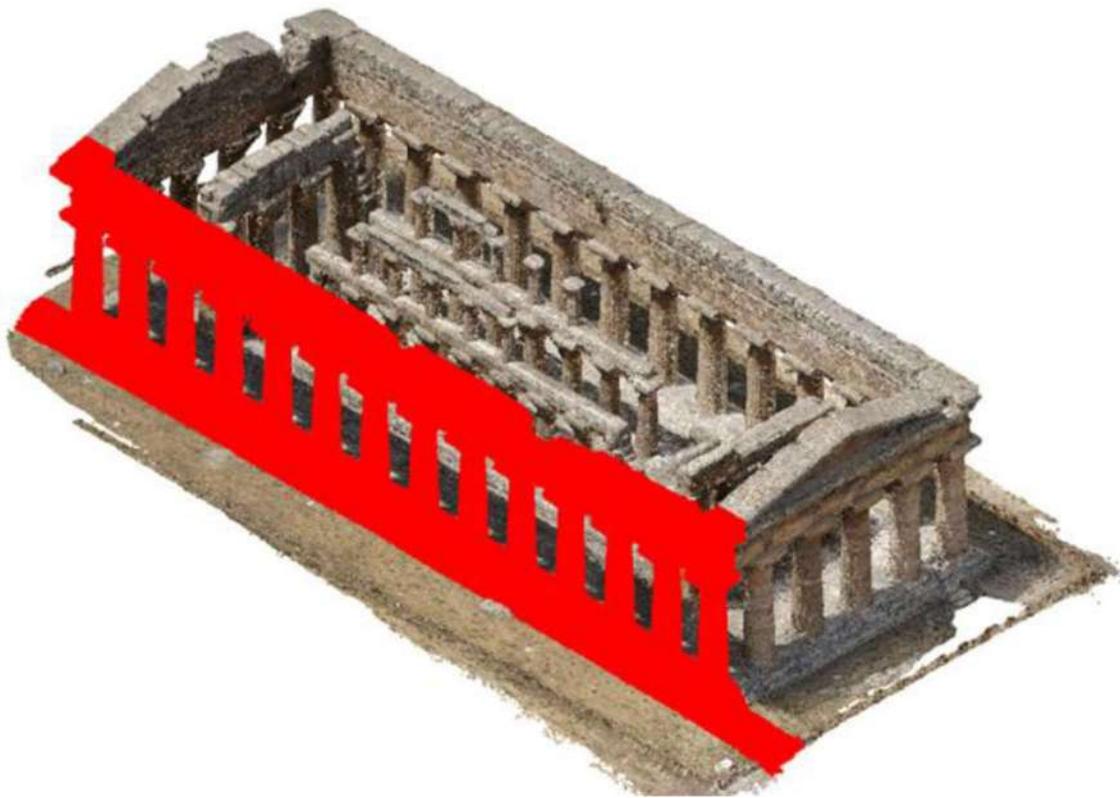
Possiamo osservare che i risultati della clusterizzazione k-means non tengono conto dell'età. Analizzando tre dei cinque cluster classificati prodotti dall'algoritmo k-means nel primo caso notiamo come la classe più probabile in termini di età contiene 18 manufatti neolitici, 10 manufatti della dinastia Zhou, 7 manufatti della dinastia Shang, 3 manufatti della dinastia Han e 3 manufatti della dinastia Tang. Il secondo cluster include invece 17 manufatti neolitici, 10 manufatti della dinastia Zhou, 6 manufatti della dinastia Shang, 2 manufatti della dinastia Han e 2 manufatti della dinastia Tang. Il terzo cluster comprende 15 manufatti neolitici, 11 manufatti della dinastia Zhou, 8 manufatti della dinastia Shang, 4 manufatti della dinastia Han e 2 manufatti della dinastia Tang. Questi insiemi non sembrano mostrare un chiaro pattern cronologico. Lo studio sembra indicare che l'algoritmo k-means da solo potrebbe non essere sufficiente per la classificazione dei manufatti in base all'età. Inoltre, gli artefatti neolitici possono appartenere a più di un cluster. È evidente che i risultati della classificazione non tengono conto dell'età. La

classificazione automatica degli artefatti museali è ancora una novità negli studi di settore. Pur avendo ottenuto un'accuratezza massima del 70%, Hao osserva come questo e altri studi dimostrino che esiste una concreta possibilità di sviluppo dell'impiego di modelli supervisionati in ambito culturale, per cui sarà possibile avanzare con modelli altamente complessi in futuro (Wang, 2024). Il clustering è un sistema utilizzato anche per l'applicazione della fotogrammetria nel settore culturale. Negli ultimi dieci anni, l'applicazione della fotogrammetria nella registrazione digitale 3D è cresciuta enormemente. Grazie allo sviluppo della tecnologia della visione artificiale e delle nuove tecnologie informatiche, la fotogrammetria ha accelerato i tempi di elaborazione e ha raggiunto un certo livello di automazione. Con l'uso della tecnologia Structure from Motion (SfM), l'uso scientifico delle misurazioni 3D con tecnologia laser scanner è stato gradualmente sostituito dalla fotogrammetria. La tecnologia UAV ha reso la fotogrammetria più accessibile e affidabile, promuovendo il suo uso in applicazioni di medie e grandi dimensioni, specialmente nel settore culturale. La valutazione della qualità del progetto fotogrammetrico può essere effettuata utilizzando diverse caratteristiche, derivate sia nella fase di acquisizione (ad esempio, il numero di immagini che contribuiscono alla ricostruzione 3D di un punto di collegamento o l'angolo di intersezione dei raggi omologhi) sia nella fase di elaborazione delle immagini (ad esempio, l'errore di riproiezione). L'errore di riproiezione è un errore geometrico corrispondente alla distanza dell'immagine tra il punto proiettato e il punto misurato. Viene utilizzato per misurare quanto accuratamente una stima del punto 3D ricrea la vera proiezione del punto. Per calcolare le coordinate 3D del punto di collegamento, vengono utilizzati i parametri di orientamento interno ed esterno della fotocamera e le coordinate dell'immagine del punto. L'errore di riproiezione è la differenza tra il punto dell'immagine e il punto riproiettato su una singola immagine. Questo errore è anche chiamato residuo RMS dell'immagine. Un altro parametro analizzato è l'angolo di intersezione tra le linee di proiezione omologhe. Questo angolo è direttamente proporzionale alla qualità metrica della nuvola di punti. Un angolo maggiore, entro certi limiti, tende a migliorare l'accuratezza del modello 3D. Per estrarre e calcolare questo angolo, sono state utilizzate librerie come NumPy e Pandas in un ambiente Python. Vi è poi la ridondanza delle immagini, ovvero il numero di scatti fotogrammetrici utilizzati nel processo SfM per la

ricostruzione di un punto 3D. Si presume che un numero maggiore di immagini migliori la qualità metrica della nuvola di punti, sempre supponendo che l'angolo di intersezione sia soddisfacente. L'accuratezza della proiezione dipende dalla scala alla quale i punti di collegamento sono stati individuati. Metashape utilizza informazioni sulla scala per ponderare gli errori di riproiezione dei punti di collegamento. Il caso che viene considerato per osservare l'utilizzo di clustering in attività di fotogrammetria è il Tempio di Nettuno di Paestum, rilevato tramite fotogrammetria aerea da UAV nel 2017. La complessa articolazione spaziale delle geometrie rende questo sito adatto per l'analisi qualitativa condotta. Durante il volo, l'accesso al Tempio e all'area circostante è stato chiuso al pubblico per 2 ore. Sono state acquisite un totale di 908 immagini, elaborate poi in Agisoft Metashape per la generazione del modello 3D.



*Figura 17 (Gujski, Di Filippo, & Limongiello, 2022)*

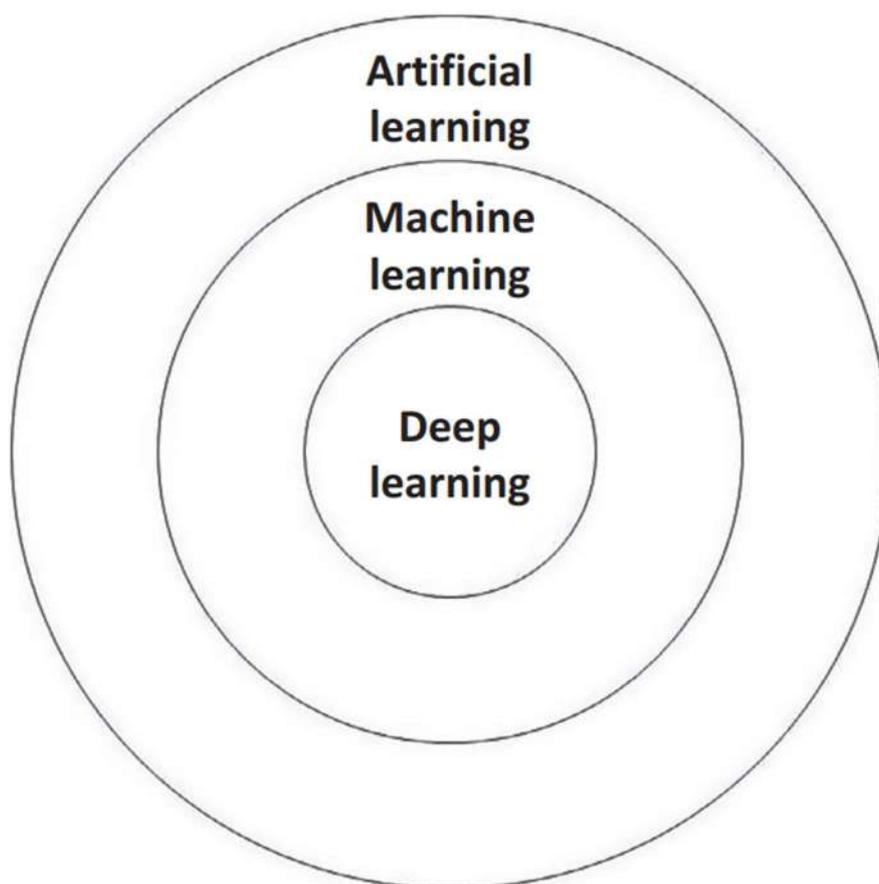


*Figura 18 In rosso, la sezione considerata a scopo di analisi (Gujski, Di Filippo, & Limongiello, 2022)*

In questo lavoro sono stati utilizzati due algoritmi di clustering: Self-Organizing Map (SOM) e K-means. SOM è una rete neurale artificiale utilizzata per la visualizzazione e l'analisi di dati ad alta dimensionalità. Viene impiegata non solo per la visualizzazione, ma anche per il clustering dei dati. Come la scalatura multidimensionale, la SOM può essere integrata con approcci di riduzione della dimensionalità. K-means è una tecnica che divide un dataset in un set di gruppi basati sul numero di cluster forniti dall'utente. L'algoritmo converge quando non ci sono ulteriori cambiamenti nell'assegnazione dei punti ai cluster. L'intero processo è stato implementato utilizzando la libreria Scikit-learn per Python. L'analisi delle caratteristiche dei cluster ha mostrato che i valori medi tra le classi di alta accuratezza e rumore variano notevolmente per l'angolo di intersezione, confermando che questo parametro ha un'influenza significativa sul rumore nella nuvola di punti fotogrammetrica (Gujski, Di Filippo, & Limongiello, 2022).

## Capitolo 4 - Deep neural networks e IA generativa in ambito culturale

Il deep learning è una sottocategoria dell'apprendimento automatico che utilizza reti neurali artificiali con molteplici strati per analizzare e interpretare dati complessi. Questa tecnica è alla base di molte innovazioni moderne nel campo dell'intelligenza artificiale (AI), come il riconoscimento delle immagini, la traduzione automatica e la guida autonoma. L'apprendimento non supervisionato può essere alla base delle reti neurali artificiali, rendendo possibile il deep learning, ma anche le tecniche di apprendimento supervisionato per le reti neurali possono essere utilizzate se l'output desiderato è già noto (Jakhar & Kaur, 2020).



*Figura 19 Il deep learning è una sotto categoria del Machine Learning (Jakhar & Kaur, 2020).*

Il deep learning ha rivoluzionato il modo in cui affrontiamo problemi complessi grazie alla sua capacità di apprendere rappresentazioni gerarchiche dai dati. Si basa su reti neurali artificiali, strutture ispirate al cervello umano che elaborano le informazioni attraverso una serie di strati interconnessi. Questi strati permettono al modello di apprendere rappresentazioni di alto livello dai dati grezzi, migliorando l'accuratezza delle previsioni e delle decisioni. Le reti neurali profonde sono costituite da tre tipi principali di strati. Uno strato di input riceve i dati grezzi e li trasmette agli strati successivi, gli strati nascosti, che processano le informazioni ricevute dallo strato di input, estraendone delle caratteristiche; infine, uno strato di output fornisce il risultato finale dell'elaborazione, come una classificazione o una previsione. Il deep learning comprende diverse tecniche, ciascuna con specifiche applicazioni e vantaggi (LeCun, Bengio, & Hinton, 2015). Ne vediamo alcune:

- Le Reti Neurali Convoluzionali, utilizzate principalmente per il riconoscimento delle immagini e la visione artificiale. Sono costituite da strati convoluzionali che applicano filtri alle immagini per estrarre caratteristiche significative. Un esempio classico è il riconoscimento di oggetti nelle immagini, dove la rete impara a identificare bordi, forme e texture per riconoscere oggetti complessi. È una delle tecniche di deep learning maggiormente impiegata nel settore culturale. Sono molto sfruttate in ambito culturale proprio per la loro capacità di snellire significativamente il lavoro di classificazione dei ricercatori.
- Le Reti Neurali Ricorrenti sono progettate per elaborare dati sequenziali, come il testo o le serie temporali. In parole povere, mantengono una "memoria" delle informazioni precedenti, permettendo di catturare dipendenze temporali. Un'applicazione comune delle Reti neurali ricorrenti è la traduzione automatica, dove la rete deve comprendere il contesto delle frasi per tradurre accuratamente da una lingua all'altra.
- Vi è poi la tecnica delle Short-Term Memory, una variante delle reti neurali ricorrenti che migliora la capacità della rete di apprendere dipendenze a lungo termine. Le LSTM sono utilizzate in applicazioni come la generazione di testi e il riconoscimento vocale.

- Gli Autoencoder vengono utilizzati per l'apprendimento non supervisionato e la riduzione dimensionale. Gli autoencoder codificano i dati di input in una rappresentazione a bassa dimensione e poi li decodificano.
- Le Reti Neurali Spaziali sono utilizzate per elaborare dati strutturati, importanti per applicazioni come la previsione di interazioni chimiche e social network (Noor & Ige, 2024).

Nel settore culturale, le reti neurali convoluzionali sono il sistema di deep learning maggiormente usato, poichè particolarmente efficace per l'elaborazione e l'analisi delle immagini. La convoluzione è il processo fondamentale delle convolutional neural network, o CNN. Un filtro, o kernel<sup>4</sup>, scorre sull'immagine di input e, a ogni posizione, esegue un'operazione matematica (dot product) tra il filtro e la porzione dell'immagine coperta dal filtro. Questo genera una mappa di attivazione o feature map, che mette in evidenza caratteristiche specifiche dell'immagine come bordi, angoli, e texture. Gli strati convolutivi: Applicano diversi filtri all'immagine di input per generare diverse feature map, ognuna delle quali estrae diverse caratteristiche. Avviene poi un'attività di pooling, un'operazione di riduzione dimensionale che riduce la risoluzione delle feature map, mantenendo comunque le informazioni essenziali. Le due operazioni di pooling più comuni sono il max pooling, che prende il valore massimo da un sottoinsieme della feature map, e l'average pooling, che prende il valore medio da un sottoinsieme della feature map. Il pooling aiuta a ridurre il numero di parametri, diminuire il rischio di overfitting e rendere la rete più robusta a variazioni minori nell'immagine. Dopo una serie di strati convolutivi e di pooling, le feature map risultanti sono "appiattite" in un vettore. Questo vettore viene poi passato attraverso uno o più strati completamente connessi (fully connected layers), simili a quelli delle reti neurali tradizionali. Ogni neurone di uno strato completamente connesso è collegato a ogni neurone dello strato precedente. Questo consente la combinazione delle caratteristiche estratte dai livelli precedenti per effettuare la classificazione o la regressione finale (O'shea & Nash, 2015). Nel complesso, il processo si svolge in questa maniera:

---

<sup>4</sup>Il kernel è una matrice più piccola dell'immagine originale, utilizzata per trasformare l'immagine attraverso l'operazione di convoluzione.

1. Input: Un'immagine di input viene fornita alla rete.
2. Convoluzione: Filtri convolutivi estraggono caratteristiche locali dall'immagine.
3. Pooling: Le feature map vengono ridotte dimensionalmente.
4. Strati Completamente Connessi: Le caratteristiche estratte vengono utilizzate per produrre un output finale, come una classificazione.

Le reti neurali convoluzionali offrono diversi vantaggi, come la riduzione dei parametri, che grazie ai filtri convolutivi che condividono i pesi, rendono le CNN più efficienti delle reti neurali tradizionali; l'invarianza alle traslazioni, ovvero la capacità di riconoscere caratteristiche indipendentemente dalla loro posizione nell'immagine; la capacità di costruire rappresentazioni gerarchiche delle caratteristiche, dalle linee semplici alle forme complesse (Li, Liu, Yang, Peng, & Zhou, 2021).

Il concetto di reti neurali artificiali risale agli anni '40 e '50, quando i ricercatori iniziarono a esplorare modelli matematici ispirati al cervello umano. Tuttavia, il campo ha visto un notevole progresso solo negli ultimi decenni grazie agli avanzamenti tecnologici e alla disponibilità di grandi quantità di dati. Nel 1943 Warren McCulloch e Walter Pitts pubblicano un modello di neurone artificiale che costituisce la base teorica delle reti neurali. Nel 1958 Frank Rosenblatt sviluppa il perceptron, un modello di rete neurale semplice che può apprendere attraverso l'addestramento supervisionato. Successivamente, nel 1986, la retropropagazione del gradiente (backpropagation), una tecnica per addestrare reti neurali multi-strato, viene popolarizzata da David Rumelhart, Geoffrey Hinton e Ronald Williams. Questo metodo permette di addestrare reti neurali profonde con maggiore efficienza. Nel Alex Krizhevsky, Ilya Sutskever e Geoffrey Hinton vincono la competizione ImageNet con una rete neurale convoluzionale, dimostrando il potenziale del deep learning nel riconoscimento delle immagini. Nel 2014, infine, Ian Goodfellow e i suoi colleghi introducono le reti neurali avversarie, aprendo nuove frontiere nella generazione di dati sintetici (Dong, Wang, & Abbas, 2021). Il deep learning rappresenta una rivoluzione nel campo dell'intelligenza artificiale, con applicazioni che spaziano dal riconoscimento delle immagini all'elaborazione del linguaggio naturale, particolarmente utili in ambito culturale. Legata al deep learning è l'intelligenza artificiale generativa (in inglese GAI, Generative Artificial Intelligence), rappresenta una branca

dell'intelligenza artificiale focalizzata sulla capacità delle macchine di generare nuovi contenuti. Questi contenuti possono includere testi, immagini, suoni e altre forme di dati, creati in modo autonomo a partire da modelli addestrati su set di dati preesistenti. La GAI ha rivoluzionato numerosi settori, offrendo strumenti potenti per la creazione automatizzata e la trasformazione dei dati. Si basa su algoritmi avanzati che permettono alle macchine di creare nuovi dati che imitano le caratteristiche dei dati di addestramento. A differenza dei modelli discriminativi, che imparano a classificare i dati, i modelli generativi imparano a capire la distribuzione dei dati e a generarne di nuovi che rispettino la stessa distribuzione. Comprende diverse tecniche, tra cui le Reti Neurali Generative Avversarie (in inglese GAN, Generative Adversarial Network): Composte da due reti neurali, una generativa e una discriminativa, che competono tra loro. La rete generativa crea dati falsi che cercano di ingannare la rete discriminativa, che invece cerca di distinguere tra dati reali e falsi. Questo processo di competizione migliora gradualmente la qualità dei dati generati. I Varianti Auto-Encoder, infine, sono reti neurali che apprendono una rappresentazione compressa dei dati di input e possono generare nuovi dati campionando da questa rappresentazione latente. Questa tecnica è utile per generare varianti di dati esistenti mantenendo una certa diversità e realismo. Abbiamo infine i Trasformatori Generativi Pre-addestrati (GPT), utilizzati principalmente per la generazione di testo. Questi modelli imparano a prevedere il token successivo in una sequenza di testo, permettendo loro di generare testo coerente e convincente su una vasta gamma di argomenti. I GPT possono essere addestrati su grandi corpora di testo e poi raffinati per specifiche applicazioni (Feuerriegel, Hartmann, Janiesch, & Zsche, 2024). Sebbene abbiano avuto un fortissimo sviluppo soprattutto negli ultimi anni, i modelli generativi hanno una lunga storia nell'intelligenza artificiale, risalente agli anni '50 con lo sviluppo dei Modelli di Markov Nascosti e dei Modelli di Miscela Gaussiane. Questi modelli erano utilizzati principalmente per generare dati sequenziali come il parlato e le serie temporali. Tuttavia, è stato solo con l'avvento del deep learning che i modelli generativi hanno visto miglioramenti significativi nelle prestazioni. Nell'ambito del deep learning, le diverse aree inizialmente non avevano molta sovrapposizione. Nel processing del linguaggio naturale, un metodo tradizionale per generare frasi era l'uso del modello linguistico N-gram, che imparava la distribuzione delle parole e cercava la sequenza

migliore. Questo metodo, però, non era efficace per le frasi lunghe. Per risolvere questo problema, sono state introdotte le reti neurali ricorrenti, seguite dallo sviluppo delle Long Short-Term Memory e delle Gated Recurrent Unit, che utilizzano meccanismi di gating per controllare la memoria durante l'addestramento. Questi metodi potevano gestire dipendenze lunghe fino a circa 200 token, migliorando notevolmente rispetto ai modelli N-gram. Nel campo della visione artificiale, prima dell'avvento dei metodi basati sul deep learning, gli algoritmi tradizionali di generazione delle immagini utilizzavano tecniche come la sintesi delle texture e il mapping delle texture. Questi algoritmi si basavano su caratteristiche progettate a mano e avevano capacità limitate nella generazione di immagini complesse e diversificate. Nel 2014, un team di studiosi capitanati da Ian Goodfellow propone le Reti Generative Avversarie, segnando una pietra miliare in questo settore grazie ai risultati impressionanti in varie applicazioni. Inoltre, gli Autoencoder Variazionali (VAE) e altri metodi come i modelli generativi di diffusione sono stati sviluppati per offrire un controllo più fine sul processo di generazione delle immagini e la capacità di generare immagini di alta qualità. L'articolo "Generative Adversarial Networks" di Ian Goodfellow introduce un framework innovativo per stimare modelli generativi attraverso un processo avversariale. Questo framework coinvolge l'addestramento simultaneo di due modelli: un modello generativo (G) che cattura la distribuzione dei dati e un modello discriminativo (D) che stima la probabilità che un campione provenga dai dati di addestramento piuttosto che dal modello generativo. La procedura di addestramento per G consiste nel massimizzare la probabilità che D commetta un errore, corrispondente a un gioco a somma zero a due giocatori. Il generatore cerca di migliorare continuamente per ingannare il discriminatore, mentre il discriminatore si allena per diventare sempre più abile nel distinguere i dati reali da quelli generati. Questo processo di competizione migliora la qualità dei campioni generati dal generatore. L'intelligenza artificiale generativa ha radici profonde nella storia dell'intelligenza artificiale e dell'apprendimento automatico. Tra il 1970 e il 1980: Le prime tecniche generative utilizzavano modelli statistici semplici per generare testo e altre forme di dati. Nel 2013 Kingma e Welling presentano i VAE, che combinano i concetti di codifica variabile e apprendimento non supervisionato per generare nuovi dati da distribuzioni continue. Questi metodi erano limitati dalla capacità computazionale e dalla

manca di grandi set di dati. Come già detto, è nel 2014 che Ian Goodfellow e alcuni suoi colleghi introducono le GAN, rivoluzionando il campo della generazione automatica di dati grazie alla loro capacità di creare immagini realistiche. Nel 2019, infine, OpenAI introduce GPT-2, un modello di trasformatori generativi pre-addestrati capace di generare testo di alta qualità e coerente su una vasta gamma di argomenti. In alcuni casi queste tecniche di intelligenza artificiale generativa sono state usate per la generazione di testi, sceneggiature, e immagini artistiche, creando opere d'arte che fondono input umani e machine learning suscitando anche diverse polemiche tra gli attori del settore culturale (Cao, et al., 2023).

#### **4.1 Applicazioni in ambito culturale: convolutional neural network e generative adversarial network**

Le tecniche di deep learning e di intelligenza artificiale generativa hanno visto una crescita generale negli ultimi anni. Il patrimonio culturale digitale consiste di risorse digitali e informazioni create per rappresentare e riprodurre il patrimonio culturale tangibile o intangibile in forma digitale. La digitalizzazione del patrimonio culturale ha aperto la strada a una vasta gamma di applicazioni di strumenti tecnologici che non erano possibili in precedenza. Le tecnologie dell'informazione e della comunicazione (ICT) hanno trasformato il modo in cui il patrimonio culturale viene documentato, interpretato e presentato (Portalés, Rodrigues, & Rodrigues Gonçalves, 2018). Pertanto, per esplorare appieno il loro potenziale nascosto, diverse nuove applicazioni del digitale sono dedicate alla promozione degli asset culturali o per affrontare sfide legate alla loro cura e ripristino. Diverse applicazioni avanzate, come abbiamo già visto, includono lo scambio di dati, la cura, la gestione e la conservazione a lungo termine. L'obiettivo finale è raggiungere la loro conservazione a lungo termine. Per quanto riguarda le applicazioni relative all'esplorazione del pieno potenziale di modelli nascosti nei dati culturali, le collezioni di dati culturali possono essere utilizzate per affrontare sfide legate alla cura, ricostruzione e erronella annotazione di asset culturali trascurati. Ciò potrebbe aumentare il loro valore e migliorare la loro attrattività. Nel 2023, un team di ricercatori ha studiato una soluzione volta ad arricchire gli asset del patrimonio culturale con l'uso della classificazione dei dati e della ricostruzione delle immagini. A questo scopo i ricercatori si sono serviti di tre set di dati contenenti principalmente dipinti. Il passaggio di pre-elaborazione è consistito

nell'omissione dei campioni di dati non ottimali e duplicati. La galleria web WikiArt ospita una vasta collezione di opere d'arte di migliaia di artisti. I dati comprendono più di 140.000 dipinti, che non sono direttamente disponibili per il download pubblico. I dati sono stati raccolti utilizzando uno script Python personalizzato con la libreria Beautiful Soup. È stato poi preso in considerazione il dataset del Met di New York, che ha pubblicato metà della sua collezione ha fornito un file comma-separated value (CSV) della collezione pubblicata, comprendente alcuni metadati e link web. Infine, l'ultima fonte è il dataset del Rijksmuseum di Amsterdam, che dispone di un'interfaccia di programmazione delle applicazioni (API) pubblica per raccogliere la sua collezione. Lo studio propone due sistemi di annotazione dati: la classificazione multimodale e la classificazione gerarchica. L'approccio di classificazione multimodale combina informazioni visive e testuali come input del modello e utilizza più output per consentire l'apprendimento multitask tramite condivisione rigida dei parametri. Per la parte di utilizzo di reti neurali convoluzionali, viene utilizzata la rete ResNet50 a causa delle sue prestazioni superiori rispetto ad altre reti in una sfida di classificazione a singolo task. Per la classificazione dei dipinti, sono stati considerate come etichette più importanti l'artista, il mezzo pittorico o scultoreo utilizzato, il genere, lo stile e l'anno di creazione come le etichette più importanti. Genere, stile e medium sono spesso disponibili, mentre rappresentano una sfida maggiore l'artista e l'anno. Dal punto di vista tecnico, le etichette testuali utilizzate sono state codificate utilizzando array one-hot. Per la rete ResNet50, sono stati rimossi gli array superiori e mantenuto tutti i livelli fino all'output dell'ultimo livello convolutivo. È stato poi concatenato questo livello con un livello denso di 256 unità collegato agli input testuali. I livelli concatenati sono poi collegati a un livello di rappresentazione condiviso incaricato di codificare una rappresentazione congiunta delle informazioni testuali e visive. Anche il formato a due livelli di output è stato collegato. Per valutare e testare questa rete, il confronto è avvenuto direttamente con una rete con la stessa architettura ma con l'informazione visiva come unico input e mantenendo i due livelli di output. Nel caso dei dipinti, il compito di attribuzione dell'artista è il compito di classificazione più importante. Di conseguenza, sono stati utilizzati sottogruppi segregati in base al numero di dipinti o opere d'arte per artista. A seguito dei confronti con uno scenario in cui l'unico input è l'informazione visiva, è stato notato che la classificazione multimodale è più performante

rispetto al modello a singolo input in termini di accuratezza. Le metriche di classificazione per questo modello rispetto a un modello a singolo input sugli stessi dati, infatti, mostrano che utilizzare più dati sotto forma di etichette testuali ha un impatto positivo sull'accuratezza della classificazione (tra il 6% e il 15% di miglioramento).

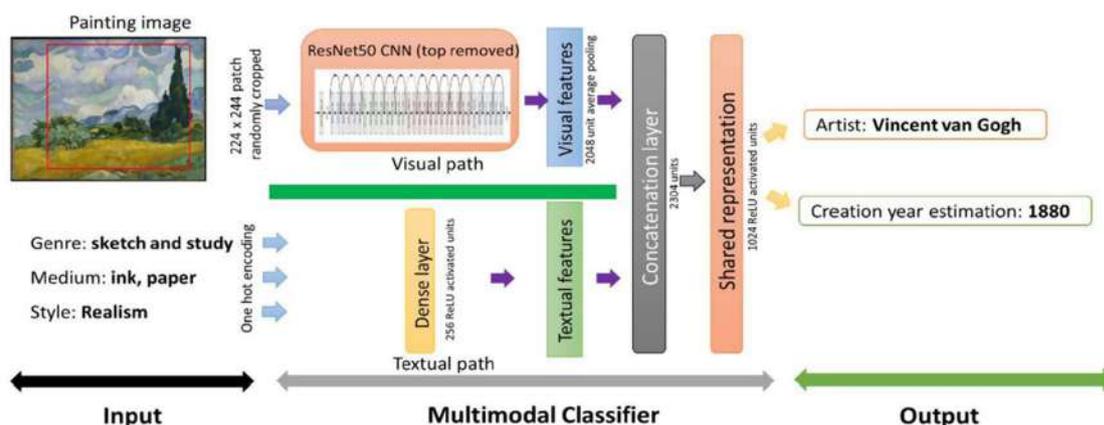


Figura 20 Architettura del classificatore multimodale (Belhi, Al-Ali, & Fofou, 2023)

A seguito dell'analisi dei set di dati culturali esistenti, è emerso scoperto che i metadati degli oggetti variano da un tipo di oggetto all'altro. Utilizzare un modello generico per la classificazione di tali set di dati non strutturati non è efficiente. Di conseguenza, è stato proposto e sviluppato un approccio di classificazione gerarchica multitask che esegue la classificazione in due fasi. Nella prima fase, un classificatore di tipo generale prevede la categoria o il tipo dell'oggetto. Una volta che questa categoria è prevista con buona sicurezza, l'annotazione finale viene assegnata a un modello di seconda fase che si occupa della classificazione del tipo assegnato. Sono state introdotte varie ottimizzazioni ai modelli, come l'uso transfer learning. Complessivamente, questo approccio è migliore per uno scenario di classificazione nel mondo reale. Queste classificazioni sono state effettuate allo scopo di condurre un'attività di completamento delle immagini culturali, detta Inpainting. Questa attività si rende necessaria perché molti beni culturali possiedono un grande valore storico e morale, ma a causa del loro degrado, questo valore è fortemente compromesso poiché la loro attrattiva è persa. Una delle soluzioni che la maggior parte delle organizzazioni e dei musei del patrimonio attualmente scelgono è sfruttare la conoscenza degli esperti d'arte e storia per recuperare e restaurare i beni danneggiati. Questo processo è, in realtà, laborioso, costoso e spesso risulta in una

semplice ipotesi sulla regione danneggiata o mancante. È quindi particolarmente efficace utilizzare il machine learning e l'intelligenza artificiale per ricostruire digitalmente le specifiche regioni mancanti. La ricostruzione digitale può essere utilizzata come ricostruzione finale, e l'oggetto può essere esposto in un museo virtuale o in una galleria web. Nel condurre questa attività sono emersi due principali tipi di problemi. La maggior parte delle tecniche di completamento e ricostruzione delle immagini basate sull'apprendimento sono addestrate su grandi set di dati, con l'obiettivo di recuperare la distribuzione originale dei campioni di addestramento. Tuttavia, questa distribuzione diventa più complessa quando i dati di addestramento sono eterogenei (caso che, come abbiamo visto, si verifica spesso in ambito culturale), rendendo il processo di addestramento difficile e la ricostruzione inefficiente. Inoltre, concentrandosi sulle tecniche basate sul deep learning, si scopre che i metodi esistenti funzionano molto bene quando si tratta di completare immagini con lo stesso contesto visivo del loro set di dati di addestramento, ma mancano della capacità di diversificazione in termini di completamento. Infatti, quando cerchiamo di addestrare queste tecniche su contesti visivi diversificati, il processo di addestramento diventa più difficile, richiede un considerevole lasso di tempo e la qualità dell'output è fortemente degradata, poiché i molteplici contesti ampliano lo spazio di ricerca per il completamento. La maggior parte delle tecniche di ricostruzione visiva richiede modelli complessi per ottenere risultati di qualità migliore. Per superare le limitazioni evidenziate, viene proposto un framework che utilizza architetture più veloci e metodi di ricostruzione delle immagini migliori, per contesti visivi limitati. Il framework mira a superare la mancanza di generalizzazione di questi metodi, ricostruendo al contempo gli input danneggiati con un ritardo minimo. Il nucleo dell'approccio di inpainting culturale si basa sulla strategia "divide et impera", che utilizza il clustering dei dati visivi per suddividere i dati di addestramento in cluster più piccoli, raggruppando campioni con contesti visivi simili. I cluster risultanti vengono quindi utilizzati per addestrare più istanze di modelli di inpainting invece di addestrare una singola istanza di modello sull'intero set di dati di addestramento.

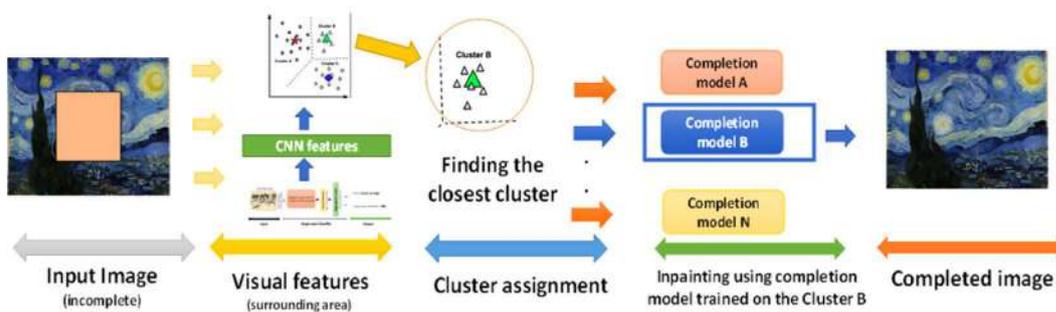


Figura 21 Architettura del framework di completamento delle immagini (Belhi, Al-Ali, & Fougou, 2023).

Vengono così addestrati più modelli di inpainting, un modello per cluster. Durante il completamento, viene identificato il cluster più vicino e la ricostruzione viene assegnata al modello di completamento del cluster correlato. Le caratteristiche visive vengono estratte dalle reti neurali convoluzionali per costruire il modello di clustering utilizzando l'algoritmo k-means. Successivamente, viene addestrato un modello di inpainting delle immagini basato su un approccio generativo che si serve di encoder contestuali e addestramento avversario. La Figura 3 mostra il risultato del completamento di un'immagine in cui una regione centrale è stata rimossa artificialmente e il framework aveva il compito di completarla. Ovviamente, eseguire il clustering sui dati di addestramento è meglio che utilizzare dati di addestramento grezzi. Nello studio viene sottolineata l'importanza di adottare una soluzione basata su più clustering, perché avere una varietà di contesti visivi limita la capacità dei modelli di inpainting di fornire ipotesi accurate. L'obiettivo principale di questo framework è aumentare l'attrattività dei beni culturali utilizzando il machine learning per affrontare alcune sfide legate alla mancanza di informazioni o ai danni fisici che potrebbero essere stati subiti da manufatti o immobili di valore culturale. Nelle intenzioni dei ricercatori potrebbe trattarsi di uno strumento di supporto per utenti professionali in istituzioni culturali e storiche che si occupano della restaurazione e della cura delle collezioni di dati culturali; il framework è stato al momento validato solo per le analisi di dipinti (Belhi, Al-Ali, & Fougou, 2023). Un'ulteriore indagine dell'impiego del deep learning nel settore culturale e umanistico è stata condotta nel 2022, con una ricerca volta ad esaminare l'uso delle reti neurali profonde per l'analisi di testi nel settore delle Digital Humanities. L'obiettivo principale è esplorare come possano migliorare la ricerca e l'analisi in questi campi interdisciplinari, affrontando le

sfide della scarsa disponibilità dei dati di addestramento e dell'adattamento al settore. Spesso, i testi del settore digital humanities sono specifici dell'ambito e contengono pochi esempi di addestramento, rendendo difficile per le reti neurali profonde un corretto processo di apprendimento. Queste risorse inoltre richiedono spesso un lavoro di pulizia e processing significativo prima di poter essere utilizzate dai modelli di deep neural networks (Suissa, Elmalech, & Zhitomirsky-Geffet, 2022). In particolare, nel 2020 è stata analizzata la Bibbia di Avila, un manoscritto latino del XII secolo. Questo manoscritto è composto da 870 pagine scritte a mano da 12 copisti diversi. Ogni pagina è stata digitalizzata ad alta risoluzione e annotata da un esperto paleografo. Il dataset si compone di immagini digitali ad alta qualità di questa Bibbia, custodita nella Biblioteca Nacional di Madrid. Ogni pagina contiene due colonne, scritte a mano in Italia nella terza decade del XII secolo. L'analisi paleografica ha individuato 12 mani scriventi, ma le pagine scritte da ogni copista non sono equamente distribuite (variano da 1 a 143). Ci sono anche casi (circa il 2% del dataset) in cui parti della stessa pagina sono scritte da copisti diversi. Per avere un numero sufficiente di dati utilizzabili in un approccio di deep learning, gli studiosi hanno considerato solo 749 pagine in cui sono identificabili 8 scrittori. Ogni pagina è stata digitalizzata con una risoluzione di 6000 × 4000 pixel e etichettata da un esperto paleografo con le lettere A, D, E, F, G, H, I, X, identificando un singolo scrittore. Vale la pena notare che, nel caso di una pagina scritta da copisti diversi, i paleografi hanno etichettato ciascuna pagina indicando lo scrittore "prevalente" (cioè lo scrittore che ha scritto la maggior parte della pagina). Per questo motivo, il sistema end-to-end fornisce solo una classe per ogni pagina. Tutte le righe di testo in 96 pagine, 12 per scrittore, sono state annotate. Le restanti 653 pagine sono state utilizzate per valutare le prestazioni del sistema di identificazione dello scrittore della pagina.

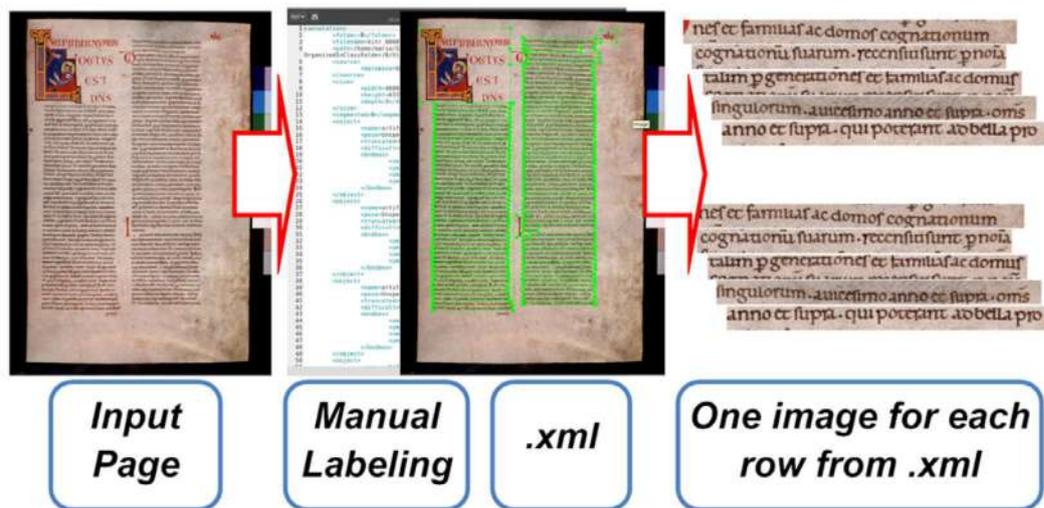


Figura 22 Procedura per l'etichettamento manuale delle righe (Cilia, De Stefano, Fontanella, & Marr, 2020)

Il sistema proposto è costituito da un modello di classificazione a tre fasi: un rilevatore di oggetti rileva automaticamente le righe di testo contenute in ciascuna pagina, una rete neurale profonda ne estrae le caratteristiche e classifica ciascuna singola riga, e infine un combinatore di decisioni a maggioranza ponderata per la opera una classificazione delle pagine. Durante la fase di test, il sistema viene utilizzato in maniera end-to-end, il che significa che il sistema riceve in input l'immagine RGB di una singola pagina e produce una decisione corrispondente allo scrittore di quella pagina. Per l'addestramento, al fine di aumentare il numero di campioni per ciascun scrittore, vengono estratte le righe di testo da ciascuna pagina e addestrate sia il rilevatore di oggetti sia la rete neurale con immagini di singole righe. L'addestramento delle due reti è stato eseguito separatamente e, in particolare, entrambe sono state addestrate utilizzando tecniche di "transfer learning" e "fine tuning". L'obiettivo del rilevatore di righe è identificare automaticamente le righe di testo in ciascuna pagina del manoscritto. Questo è strettamente correlato a un problema di visione artificiale in cui lo scopo è rilevare e riconoscere un oggetto in una scena. Uno dei modelli più potenti e versatili è il MobileNetV2, un'architettura progettata per ambienti mobili e con risorse limitate. La caratteristica principale di MobileNetV2 è l'uso delle convoluzioni separabili in profondità, rendendo necessari meno parametri e meno memoria per l'addestramento, con il risultato di avere un modello molto piccolo ed efficiente. Per adattare la rete MobileNetV2 ai problemi di rilevamento degli oggetti, è stato utilizzato un Single Shot Detector per generare una mappa di rilevamento. Per

l'addestramento, a causa del basso numero di campioni, è stato applicato un approccio di transfer learning, in cui un modello addestrato su un dataset di base viene quindi riaddestrato su un dataset e un compito target. Il secondo passo dell'approccio è la classificazione delle righe di testo rilevate. A tal fine, viene definita un'architettura deep neural network a due fasi. Inizialmente viene applicato un estrattore di caratteristiche convolutive alle immagini di input per ottenere caratteristiche di alto livello. La scelta dell'estrattore di caratteristiche è cruciale poiché il numero di parametri e strati influisce direttamente sul tempo di elaborazione e sulle prestazioni del rilevatore. In secondo luogo, viene proposta una meta-architettura per la fase di classificazione. La meta-architettura del classificatore è costituita da due strati completamente connessi con 2048 nodi. Un livello di dropout è utilizzato dopo il primo completamente connesso per migliorare la capacità di generalizzazione della rete. Infine, viene utilizzato uno strato Softmax per generare i gradi di confidenza della classe di output nell'intervallo  $[0, 1]$ . Grazie alla meta-architettura del classificatore, l'addestramento del modello del classificatore di righe può essere effettuato da zero o con fine tuning. Nel primo caso, l'intera rete viene addestrata direttamente sul compito a portata di mano, inizializzando tutti i pesi in tutti gli strati attraverso la procedura di inizializzazione di Xavier. Nel secondo caso, l'addestramento del modello del classificatore di righe viene realizzato in due fasi. La prima è un riaddestramento del modello tramite transfer learning che consiste nel regolare i parametri solo negli strati di classificazione finali piuttosto che in tutti i pesi all'interno degli strati dell'estrattore di caratteristiche. Questo significa che tutti i parametri dell'estrattore di caratteristiche sono congelati con valori presi da un modello pre-addestrato, mentre i parametri del classificatore della meta-architettura sono inizializzati casualmente e addestrati. Nella seconda fase, viene eseguito un fine tuning in cui tutti i parametri sia dell'estrattore di caratteristiche sia del classificatore vengono sbloccati e riaddestrati. L'obiettivo finale dell'approccio proposto è assegnare ogni pagina del manoscritto a un singolo scrittore. A tal fine, le decisioni rigide del classificatore di righe ottenute per tutte le righe in una singola pagina vengono combinate per ottenere una singola decisione. Per evitare qualsiasi parità tra diversi scrittori, la regola di combinazione proposta è un voto di maggioranza in cui ciascuna riga è ponderata con il grado di confidenza dell'output del classificatore di righe. Di conseguenza, una pagina

viene assegnata al scrittore a cui il classificatore di righe ha assegnato il maggior numero di righe ponderate. Sono stati condotti tre esperimenti in totale. Il primo, il rilevatore di righe su un dataset etichettato manualmente, ha ottenuto un'accuratezza del 25% nel rilevamento corretto delle righe. Applicato al dataset delle pagine, il rilevatore è riuscito a riconoscere 21.071 righe nelle 653 pagine di test, stimando una media di 126 righe per pagina. Il tasso di rilevamento apparentemente basso è dovuto alla scelta di rilevare solo righe quasi complete (almeno il 90%), garantendo che le fasi successive del sistema non classificassero campioni non affidabili. La classificazione delle righe ha visto diversi test su vari modelli di reti neurali convoluzionali. A livello di accuratezza, i modelli addestrati con fine tuning hanno generalmente superato quelli addestrati da zero. Il modello InceptionResNetV2 ha ottenuto le migliori prestazioni con un'accuratezza globale del 95.58% e un F1 score del 94.85% sul dataset etichettato, e un'accuratezza dell'89.34% e un F1 score dell'89.74% sul dataset delle pagine. L'ultimo esperimento prevedeva la combinazione delle decisioni del classificatore di righe per identificare lo scrittore delle pagine. Il modello InceptionResNetV2 ha ottenuto le migliori prestazioni con un'accuratezza del 96.48% e un F1 score del 96.56%. Le prestazioni variano tra i modelli, con il modello di fine tuning che ha dimostrato una maggiore stabilità e indipendenza dall'architettura dell'estrattore di caratteristiche (Cilia, De Stefano, Fontanella, & Marr, 2020). Un esempio simile di applicazione di tecniche di deep learning è individuabile anche nel settore del patrimonio culturale architettonico. La documentazione accurata del patrimonio culturale architettonico è fondamentale per la sua conservazione, manutenzione e restauro. La digitalizzazione di queste informazioni facilita la gestione e la condivisione, permettendo un monitoraggio continuo dello stato dei beni e contribuendo alla loro preservazione. Tuttavia, la classificazione manuale delle immagini raccolte durante questo processo può essere laboriosa e soggetta a errori. Le tecniche di deep learning, in particolare le reti neurali convoluzionali, offrono una soluzione promettente per automatizzare questa digitalizzazione, che si svolge in due fasi principali: una di raccolta dei dati, che include la misurazione, la fotografia sistematica e l'archiviazione delle immagini e dei dati raccolti e una di classificazione e gestione, che riguarda l'interpretazione e l'organizzazione delle informazioni per facilitare ricerche efficienti e specifiche, migliorando l'accessibilità e l'usabilità dei dati. Uno studio del 2017

ha avuto proprio come obiettivo principale dello studio l'applicazione delle reti neurali convoluzionali per la classificazione automatica delle immagini del patrimonio architettonico. La corretta classificazione delle immagini è essenziale per una gestione efficiente dei dati digitali, permettendo di migliorare le ricerche e facilitando le attività di studio e interpretazione dei beni culturali. Gli autori hanno creato un nuovo dataset di immagini del patrimonio architettonico, dato che non all'epoca dello studio non esistevano dataset pubblici adeguati per l'addestramento di questo tipo di reti in questo contesto (. Hanno esplorato due approcci principali: l'addestramento da zero delle reti neurali, in cui le reti sono addestrate utilizzando esclusivamente il nuovo dataset, senza alcun pre-addestramento, e il fine tuning di reti pre-addestrate, nel quale le reti neurali pre-addestrate su grandi dataset generali, come ImageNet, vengono ulteriormente addestrate (fine-tuned) utilizzando il nuovo dataset specifico per il patrimonio architettonico. I risultati sono promettenti in termini di accuratezza della classificazione delle immagini. Le reti neurali si sono dimostrate efficaci nel distinguere tra diversi elementi di interesse nelle immagini architettoniche. I risultati indicano che l'uso di reti pre-addestrate con un ulteriore fine tuning sul dataset specifico del patrimonio architettonico produce prestazioni migliori rispetto all'addestramento da zero. Questo suggerisce che le caratteristiche apprese dalle reti su grandi dataset generali possono essere utilmente trasferite e adattate a contesti specifici. La classificazione automatica delle immagini, secondo gli autori, è cruciale per migliorare l'efficienza della documentazione digitale. Riduce il carico di lavoro umano e minimizza gli errori, permettendo una gestione più rapida e accurata dei dati. Inoltre, l'uso di reti pre-addestrate consente di risparmiare tempo e risorse computazionali, offrendo al contempo una maggiore accuratezza. Il fine tuning permette alle reti di adattarsi meglio alle specifiche peculiarità del patrimonio architettonico. L'applicazione delle CNN per la classificazione delle immagini del patrimonio architettonico rappresenta un'importante innovazione nel campo della documentazione digitale.

Category	Examples
Altar (829 images)	
Apse (514 images)	
Bell tower (1059 images)	
Column (1919 images)	
Dome (inner) (616 images)	
Dome (outer) (1177 images)	
Flying buttress (407 images)	
Gargoyle (and Chimera) (1571 images)	
Stained glass (1033 images)	
Vault (1110 images)	

Figura 23 Esempio di immagini utilizzate per la creazione del dataset (Llamas, Lerone, Medina, & Zalama, 2017).

I risultati ottenuti dimostrano che queste tecniche, unitamente alla creazione e al mantenimento di dataset specifici, possono migliorare significativamente l'efficienza e l'accuratezza del processo di documentazione, contribuendo alla conservazione e alla valorizzazione del patrimonio culturale (Llamas, Lerone, Medina, & Zalama, 2017).

L'articolo "Generative Digital Humanities" di Fabian Offert e Peter Bell esplora l'applicazione delle tecniche di machine learning generativo, in particolare le reti generative avversarie (GAN), nel campo delle digital humanities. Queste tecniche possono facilitare l'analisi e l'esplorazione di grandi corpus di immagini, offrendo nuove prospettive critiche e metodologiche. Come anticipato, per parlare di reti generative avversarie è fondamentale la distinzione fondamentale tra approcci discriminativi e generativi nel machine learning. Gli approcci discriminativi cercano di classificare i dati distinguendo tra classi, mentre gli approcci generativi apprendono la distribuzione completa delle caratteristiche dei dati, permettendo di generare nuovi campioni simili ai dati di addestramento. Questa distinzione offre una prospettiva critica alternativa sul machine learning nel contesto delle digital humanities. Le GAN infatti, come detto in precedenza, consistono in due reti neurali che lavorano in tandem: una rete generativa che crea nuovi dati e una rete discriminativa che valuta la qualità dei dati generati. Questo processo competitivo migliora progressivamente la qualità dei dati prodotti. Le GAN sono state applicate con successo in vari compiti complessi nelle scienze, e gli autori propongono che possano essere altrettanto utili nel contesto delle humanities digitali, specialmente nella storia dell'arte digitale. Gli autori hanno utilizzato una GAN per generare nuovi campioni da due corpora storici dell'arte. Il primo corpus comprendeva 20.000 scene di adorazione di Cristo, mentre il secondo includeva 50.000 immagini dalla collezione online del Museum of Modern Art (MoMA) di New York. Utilizzando l'architettura StyleGAN, hanno addestrato la rete neurale generativa sui corpora e poi hanno analizzato le direzioni più significative nello spazio latente appreso, identificando concetti semantici e sintattici nei dati generati. L'esperimento ha dimostrato che le GAN possono generare immagini che combinano caratteristiche di diverse opere d'arte, introducendo una continuità nei corpora visivi che non esiste nei dati originali. Questo approccio potrebbe offrire nuove prospettive per l'analisi dei dati nelle humanities digitali, permettendo di studiare i corpus in modo più approfondito e di esplorare le soglie

semantiche che definiscono una collezione di immagini. L'articolo conclude affermando che sono necessarie ulteriori ricerche sia sul potenziale pratico che sulle implicazioni teoriche delle tecniche di machine learning generativo nelle digital humanities. Gli autori sostengono che, mentre gli studi letterari computazionali hanno già implicitamente adottato approcci generativi, le humanities digitali visive mancano ancora di strumenti equivalenti. Propongono di esplorare ulteriormente l'uso delle GAN nella storia dell'arte digitale, suggerendo che queste tecniche potrebbero diventare uno strumento di esplorazione non supervisionata per grandi corpora di immagini (Offert & Bell, 2020). Un ultimo esempio arriva da uno studio cinese sulla ricostruzione digitale di manufatti artistici antichi. Il framework proposto utilizza una GAN (Generative Adversarial Network) per riempire i vuoti nei modelli 3D degli artefatti culturali. La cosiddetta MUGAN è composta da un generatore e un discriminatore. Il generatore aumenta la densità della nuvola di punti nell'area del foro, mentre il discriminatore determina se i dati della nuvola di punti provengono dai dati reali o sono stati generati. La rilevazione accurata dei fori è un prerequisito fondamentale per completare i modelli 3D degli artefatti culturali (CHarts). A causa dell'erosione fisica e chimica delle superfici di contatto, le superfici fratturate dei frammenti spesso non sono completamente coerenti come quando si sono appena rotte, causando la formazione di fori nei modelli 3D riassembleati. I fori possono essere classificati come fori semplici e fori complessi, come mostrato nella Figura 6. Basandosi sul principio del triangolo, è stato proposto un metodo efficace per la

rilevazione dei fori nelle mesh 3D. Il metodo proposto si basa su un algoritmo di ricerca in profondità per identificare i bordi dei fori nelle mesh 3D.

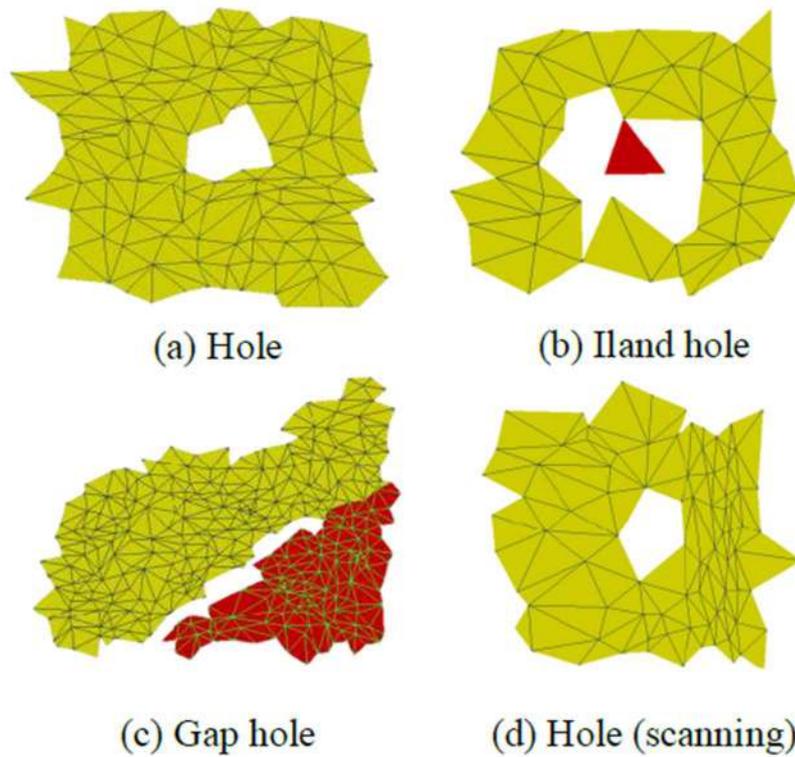
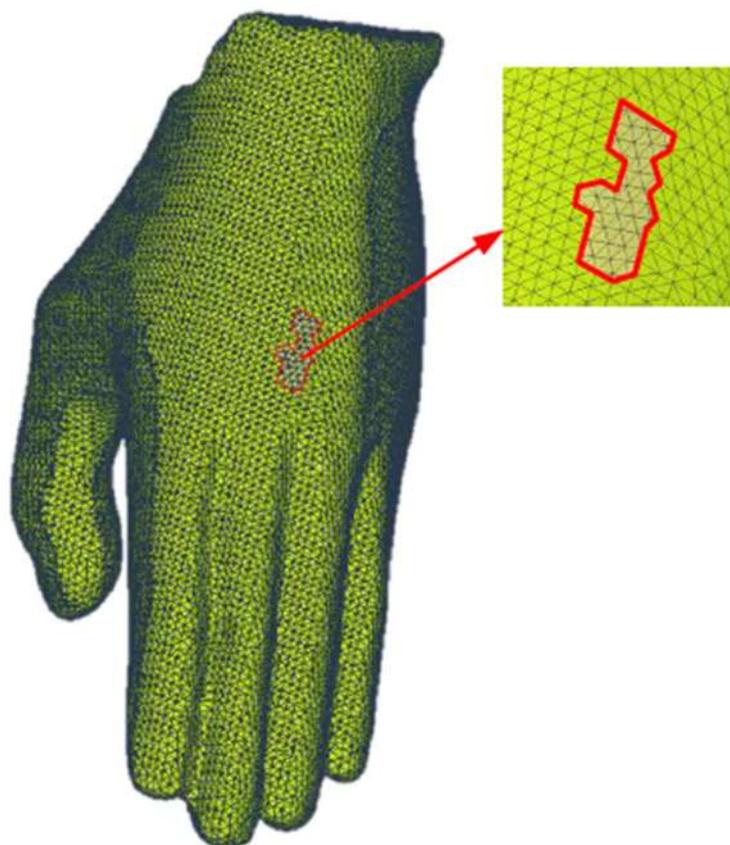


Figura 24 Esistono diversi tipi di fori sulla superficie dei modelli 3D degli artefatti culturali. (a) è un semplice foro chiuso formato collegando i bordi della mesh triangolare; (b) è un foro semplice con un piccolo numero di mesh triangolari indipendenti all'interno del foro, chiuso dai bordi della mesh triangolare; (c) è un foro complesso non chiuso; (d) è un foro semplice nei dati di scansione ottenuti dallo scanner (Ren, et al., 2020)



*Figura 25 Risultato dell'attività di individuazione dei fori: lo stesso è delineato in rosso (Ren, et al., 2020).*

L'architettura della MU-GAN include un modulo di estrazione delle caratteristiche, un modulo di espansione delle caratteristiche e un modulo di ricostruzione del set di punti. Il modulo di estrazione delle caratteristiche utilizza connessioni dense intra-livello per estrarre le caratteristiche dalla nuvola di punti di input. Il modulo di espansione delle caratteristiche aumenta il numero di punti, mentre il modulo di ricostruzione del set di punti genera una nuvola di punti densa. Il modulo di estrazione utilizza una rete neurale convoluzionale per estrarre caratteristiche significative dalla nuvola di punti. Le connessioni dense intra-livello permettono di mantenere un flusso di informazioni ricco e dettagliato attraverso la rete. Vi sono poi un modulo di espansione delle caratteristiche, che aumenta la densità dei punti nella nuvola di punti e utilizza le caratteristiche estratte per generare nuovi punti che riempiono le aree mancanti, mantenendo la coerenza con la struttura originale, e un modulo di ricostruzione del set di punti, che combina le

informazioni dal modulo di espansione per ricostruire una nuvola di punti densa e dettagliata. Questo modulo si assicura che i nuovi punti siano posizionati in modo accurato per riempire i fori identificati. Durante la fase di addestramento, il generatore e il discriminatore vengono addestrati alternativamente. La perdita complessiva della MU-GAN è composta dalla perdita del generatore, dalla perdita di ricostruzione e dalla perdita del discriminatore. Il generatore cerca di produrre nuvole di punti realistiche che possano ingannare il discriminatore. Viene addestrato minimizzando la perdita generata dal discriminatore, cercando di migliorare continuamente la qualità delle nuvole di punti. Il discriminatore viene addestrato per distinguere tra nuvole di punti reali e generate. Viene addestrato massimizzando la capacità di riconoscere correttamente le nuvole di punti reali e falsi, migliorando la sua precisione. La perdita totale del sistema include la perdita del generatore (che cerca di ingannare il discriminatore), la perdita di ricostruzione (che valuta quanto accuratamente i punti generati riempiono i fori) e la perdita del discriminatore (che cerca di distinguere tra reale e generato). Sono stati condotti esperimenti per valutare le prestazioni della MU-GAN utilizzando dati reali degli artefatti di alcuni Guerrieri di Terracotta, facenti parte di un esercito di statue realizzato tra il 246 e il 206 a.C con lo scopo di essere collocati nel mausoleo dell'imperatore Qin Shi Huang per servirlo nell'aldilà. I risultati hanno dimostrato che il framework proposto può riempire efficacemente i fori nei modelli 3D con dettagli fini e alta fedeltà. Per verificare la robustezza della rete proposta, è stato generato un piccolo foro in ciascun modello per i test. La Figura 7 mostra i risultati completi dei dati di input. La rete ha ottenuto risultati di riparazione più ragionevoli sia sulla testa, con più punti caratteristici, sia sul corpo, con meno punti caratteristici. Nello specifico, sono state estratte le patch della regione del foro dalle tre diverse scale generate da MU-GAN; la nuvola di punti completa finale è stata generata unendo le patch della regione del foro con la nuvola di punti incompleta. Con questo metodo, non solo i fori vengono riempiti, ma anche le caratteristiche ad alta fedeltà del modello vengono preservate. Per superare il problema della mancanza di informazioni disponibili per i fori di grandi dimensioni, è necessario unire le caratteristiche ottenute dalle tre diverse scale per riparare il modello incompleto. I risultati mostrano che il metodo è in grado di mantenere bene le informazioni strutturali della nuvola di punti di

input anche con diversi rapporti di mancanza, ma se i fori sono di grandi dimensioni il framework ha difficoltà a completarli del tutto.

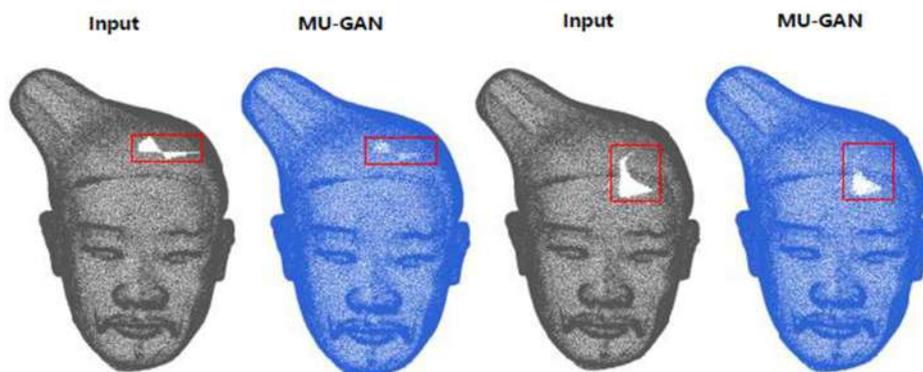


Figura 26 Esempi di fori nelle immagini completati dal framework (Ren, et al., 2020)

Il framework MU-GAN è stata in grado di generare punti che riempiono i fori in modo realistico, mantenendo l'integrità strutturale degli artefatti. Il framework ha dimostrato di essere capace di preservare i dettagli fini del modello originale, producendo ricostruzioni di alta qualità che sono visivamente indistinguibili dai dati reali. Rispetto ai metodi tradizionali di riempimento dei fori, la MU-GAN ha mostrato una superiorità significativa in termini di accuratezza e realismo delle ricostruzioni (Ren, et al., 2020).

In conclusione, L'impiego di reti neurali convoluzionali (CNN) e reti generative avversarie (GAN) nel settore culturale offre numerosi vantaggi significativi che contribuiscono a rivoluzionare la conservazione, la valorizzazione e la fruizione del patrimonio culturale. Queste tecnologie avanzate, basate sul deep learning, consentono di affrontare sfide complesse in modi innovativi e altamente efficaci. Le CNN e le GAN migliorano significativamente la conservazione digitale dei manufatti culturali. Le CNN, con la loro capacità di riconoscere e analizzare dettagli complessi nelle immagini, permettono di creare rappresentazioni digitali accurate di opere d'arte e manufatti storici. Questo è essenziale per documentare e preservare il patrimonio culturale per le future generazioni, soprattutto in situazioni in cui gli originali sono a rischio di deterioramento o distruzione. Le GAN, in particolare, sono straordinariamente utili per il restauro virtuale. Queste reti possono generare nuove parti di immagini o modelli 3D mancanti, riparando digitalmente opere d'arte danneggiate o incomplete. Questo processo non solo riduce la necessità di

interventi fisici, che possono essere invasivi o irreversibili, ma permette anche di esplorare diverse ipotesi di restauro in modo sicuro e non distruttivo. Nel campo della ricerca, le CNN e le GAN offrono strumenti potenti per l'analisi dei dati culturali. Le CNN possono essere utilizzate per l'analisi stilistica, l'attribuzione di opere e la datazione di manufatti, mentre le GAN possono aiutare nella simulazione e nello studio di scenari storici. Queste tecnologie permettono ai ricercatori di estrarre nuove informazioni e comprendere meglio le dinamiche storiche e culturali. Guardando al futuro, l'integrazione delle CNN e delle GAN nel settore culturale promette di aprire nuove frontiere nella conservazione e valorizzazione del patrimonio culturale. Con l'avanzamento continuo delle tecnologie di deep learning, possiamo aspettarci una sempre maggiore accuratezza e capacità di queste reti, rendendo possibile la creazione di soluzioni ancora più sofisticate e innovative per preservare e promuovere la nostra eredità culturale. L'adozione di queste tecnologie non solo migliora la gestione e la protezione del patrimonio culturale, ma arricchisce anche l'esperienza del pubblico, promuovendo una maggiore comprensione e apprezzamento del patrimonio storico e culturale.

## Capitolo 5 - Applicazioni del Machine Learning alla Gestione Culturale

Il machine learning sta rapidamente trasformando il modo in cui le aziende operano, migliorando l'efficienza, aumentando le vendite e ottimizzando la gestione economica. Questa tecnologia consente alle macchine di apprendere dai dati e migliorare le loro performance nel tempo, offrendo vantaggi significativi in vari contesti di business. Nel settore culturale, come abbiamo visto, il machine learning comporta molti vantaggi, migliorando l'efficienza dei processi attraverso diverse modalità. L'automazione basata su ML riduce la necessità di interventi manuali in processi ripetitivi, permettendo alle aziende di risparmiare tempo e risorse. A livello gestionale generale, l'automazione delle operazioni di back-office può ridurre significativamente i costi operativi. Può consentire di analizzare dati di mercato, comportamenti dei clienti e trend stagionali per ottimizzare i prezzi dei prodotti e servizi in tempo reale; permette di condurre analisi predittive anticipando problemi e opportunità e migliorando la pianificazione strategica. Il settore culturale, comprendente istituzioni come musei, biblioteche, teatri, gallerie d'arte, e siti archeologici, svolge un ruolo cruciale nella conservazione del patrimonio culturale, nella promozione della conoscenza e nell'arricchimento culturale delle comunità. La gestione di queste istituzioni presenta sfide uniche e richiede una comprensione approfondita della loro economia e delle loro caratteristiche operative. In questa introduzione, esploreremo le peculiarità del settore culturale, le sfide affrontate dalle istituzioni e come il machine learning possa supportare la loro gestione. Le istituzioni culturali operano spesso come enti senza scopo di lucro, sostenuti da fondi pubblici, donazioni private, e entrate generate da biglietti d'ingresso e vendite di merchandising. Queste istituzioni devono bilanciare la loro missione educativa e conservativa con la necessità di sostenibilità finanziaria. Le istituzioni culturali affrontano una costante pressione per dimostrare il loro valore economico e sociale ai finanziatori e al pubblico (Berg & Rachleff Burt, 2023). Oltre alla conservazione, le istituzioni culturali hanno un ruolo educativo significativo. Inoltre le istituzioni culturali hanno il compito di supportare la produzione artistica contemporanea, offrendo spazi espositivi e risorse per artisti emergenti. Questo contribuisce alla dinamica creativa e all'innovazione nel campo culturale, creando un ambiente che favorisce l'espressione artistica e la scoperta di nuovi talenti. Con l'avvento

delle tecnologie digitali, le istituzioni culturali stanno esplorando nuove modalità di interazione con il pubblico. L'uso di realtà virtuale e aumentata, archivi digitali e piattaforme online sta trasformando il modo in cui il patrimonio culturale è fruibile. Queste tecnologie permettono di raggiungere un pubblico più ampio e di migliorare l'esperienza educativa e interattiva offerta dalle istituzioni (Hesmondhalgh & Meier, 2018). Il machine learning sta trasformando il modo in cui le istituzioni culturali gestiscono e preservano il patrimonio culturale. Dall'analisi dei dati alla conservazione preventiva, l'apprendimento automatico offre strumenti potenti per migliorare l'efficienza e l'efficacia delle operazioni culturali. Nella gestione culturale, possiamo trovare esempi di applicazioni dell'apprendimento per scopi di marketing. Nello specifico, troviamo un esempio di questo genere nello studio condotto nel La segmentazione del mercato è un processo di divisione del mercato in gruppi interni omogenei che appaiono distinti rispetto agli altri gruppi. La segmentazione riconosce che la domanda totale per le offerte artistiche è essenzialmente eterogenea e può essere disaggregata in segmenti con esigenze e preferenze diverse. Poiché le organizzazioni delle arti performative tendono ad applicare un approccio di marketing incentrato sul prodotto, l'obiettivo della segmentazione è identificare quei segmenti di mercato più suscettibili alle offerte attuali e distinguerli da quelli che potrebbero essere raggiunti solo utilizzando tecniche di marketing più impegnative. In questo contesto il machine learning può essere impiegato proprio nella segmentazione del mercato delle arti performative. Diversi modelli di segmentazione sono stati sviluppati in passato per raggruppare il pubblico delle arti performative, ma nel tempo hanno evidenziato problemi nell'applicazione pratica dell'approccio di segmentazione all'interno del mercato delle arti. I principali errori possono risiedere nelle analisi insufficienti della struttura del mercato, con l'assunzione che il mercato sia segmentato quando in realtà non lo è, e viceversa (Colbert, 2020). Tuttavia, in maniera simile ad altri settori, vengono utilizzate quattro variabili di segmentazione di base per dividere il mercato delle arti: variabili geografiche, variabili sociodemografiche, variabili psicografiche e variabili comportamentali. Dato che il consumo di arti performative è principalmente legato a un particolare luogo, la segmentazione geografica può essere vista come un approccio abbastanza naturale per le organizzazioni delle stesse. Tuttavia, con lo sviluppo delle nuove tecnologie

dell'informazione che consentono l'acquisto di biglietti online e nuove possibilità di trasporto economico, l'area di influenza geografica per le arti performative è stata notevolmente ampliata. Per quanto riguarda la segmentazione sociodemografica, il pubblico è diviso in termini di età, sesso, reddito, istruzione, razza, professione, ecc. Alcuni fattori sociodemografici devono essere evidenziati come predittori di partecipazione. Abbiamo poi una segmentazione psicografica: questo metodo implica raggruppare il pubblico basandosi su linee di personalità e stili di vita, coinvolgendo un'analisi delle caratteristiche psicologiche del pubblico, dei loro atteggiamenti, valori e opinioni. Infine, la segmentazione comportamentale si concentra sulle abitudini di partecipazione, rispondendo alle domande su quando, perché, in quali circostanze e con quale frequenza le persone partecipano agli eventi artistici. Diversi studi empirici sono stati condotti per rivelare il comportamento dei consumatori e le preferenze relative alle arti performative. Gli algoritmi e le architetture che apprendono dai dati osservati sono continuamente sviluppati nella ricerca sul machine learning. Gli algoritmi di apprendimento possono apprendere modelli di somiglianza dai dati in modo "non supervisionato". Tuttavia, per la ricerca di mercato, quando si cerca di scomporre i set di dati in gruppi o cluster senza criteri predefiniti, gli algoritmi di apprendimento potrebbero non produrre i risultati desiderati di performance per la segmentazione del mercato. Il data mining applica tecniche di machine learning per estrarre informazioni utili da grandi set di dati o database. Nella ricerca di marketing, è particolarmente utile per analizzare le relazioni nelle intenzioni di acquisto per diversi prodotti e servizi. Questa informazione può aiutare i professionisti del marketing a individuare i prodotti correlati e offrire servizi correlati nello stesso pacchetto. Per valutare l'affidabilità di un albero decisionale come modello per la segmentazione del mercato, è stata proposta una strategia in quattro fasi. Come prima cosa, è stata effettuata l'inclusione dei set di esempi con quante più informazioni disponibili, considerando le variabili ottenute attraverso i diversi modelli esperti attuali sulla segmentazione del mercato di una popolazione. Sono stati sfruttati algoritmi per apprendere i classificatori da diversi approcci e calcolare la loro accuratezza predittiva per un determinato set di dati. È stato generato poi un albero decisionale per tutti i casi del campione rispondenti ad una predizione accurata, adottandolo come modello di segmentazione gerarchica. Viene poi sfruttato un algoritmo di selezione o estrazione delle

caratteristiche vengono per tutti gli algoritmi di apprendimento utilizzati nel secondo passo. Tramite una valutazione empirica viene descritto il campione e i set di dati utilizzati, i dettagli tecnici relativi all'applicazione della strategia in quattro fasi e i risultati dell'accuratezza e del modello di segmentazione creato utilizzando questa strategia (Abad-Grau, Tajtakova, & Arias-Aranda, 2009). Nel settore del patrimonio culturale, la natura delle visite ai musei si presta alla categorizzazione di modelli e comportamenti che i visitatori esibiscono mentre attraversano tali spazi. Lo studio dei visitatori dei musei risale ai primi anni del XX secolo. Questi studi possono portare a una migliore comprensione delle esigenze dei visitatori e a generare servizi personalizzati. I visitatori dei musei rappresentano anche una classe particolare di utenti, presentando molte sfide per il monitoraggio e la modellazione efficace del comportamento: da questo punto di vista, sta diventando sempre più importante sia per i musei tradizionali che per le versioni tecnologicamente assistite o aumentate di essi. Nonostante la diffusa disponibilità di ogni tipo di dispositivi fissi e mobili, la raccolta sistematica dei dati dei musei si basa in molti casi sulla vendita dei biglietti e sui questionari. Questi approcci comportano molte limitazioni, tra cui le dimensioni ridotte dei campioni, la riluttanza di alcune persone a rispondere, le difficoltà nel garantire una corretta randomizzazione, il tempo e il costo delle interviste e la bassa frequenza delle indagini e delle analisi globali (Bollo, 2017). Infatti, tali questionari includono informazioni che possono essere raccolte o dedotte, in gran parte, utilizzando dispositivi tecnologici: orari preferiti e durata della visita, attività svolte durante la visita, canali di informazione preferiti utilizzati, tipi di materiali e contenuti informativi letti o ascoltati. A questo scopo, i sistemi di raccolta dati non invasivi possono tracciare la presenza e i movimenti dei visitatori senza alcuna forma di preregistrazione o fornitura esplicita di dispositivi per l'identificazione e l'interazione. In alternativa, i sistemi interattivi possono offrire molteplici tipi di esperienze sensoriali e servizi agli utenti finali, affidandosi anche a dispositivi mobili che possono monitorare e salvare informazioni relative alle posizioni degli utenti, alle azioni eseguite e ai tempi impiegati. Ogni tipo di sistema precedentemente menzionato ha i suoi specifici vantaggi e limitazioni, mentre entrambi consentono la raccolta e il monitoraggio continui, massicci e automatizzati di prove, nonché analisi molto frequenti (anche in tempo reale) e studi longitudinali. A partire dalle considerazioni sopra riportate, è possibile valutare

l'applicazione delle tecniche di apprendimento non supervisionato e metodologie di Analisi dei Dati e Visualizzazione dei Dati su dati raccolti da dispositivi interattivi, mobili e multimediali al fine di costruire conoscenze utili, come alcuni indicatori comportamentali a supporto del personale e degli stakeholder dei musei. Grazie a queste metodologie, è possibile comprendere meglio le dinamiche delle visite e di conseguenza apportare, ad esempio, modifiche appropriate all'interno di un museo o offrire servizi nuovi e personalizzati. Considerando l'uso di una guida museale elettronica e IoT, è anche possibile effettuare azioni di personalizzazione in base ai tipi di museo, come un percorso specifico customizzato in base al tipo di visitatore. Un esempio di questo tipo di ricerca si trova nello studio effettuato nel 2019 nel Museo Civico di Castel Nuovo, situato all'interno del Maschio Angioino di Napoli. Durante il periodo di raccolta dati, che va da maggio 2017 a ottobre 2017, ogni visitatore ha avuto l'opportunità di noleggiare un dispositivo interattivo mobile con un sovrapprezzo sul costo del biglietto; in totale 4719 visitatori hanno accettato di noleggiarlo e utilizzarlo. Grazie a questi dispositivi, i visitatori potevano scegliere tra oltre 160 elementi culturali con molteplici funzioni multimediali. Il dispositivo mobile incorpora un'applicazione personalizzata, sviluppata allo scopo di fornire informazioni aggiuntive al visitatore attraverso la realtà aumentata (file audio, foto, testo per ciascun'opera d'arte) e tracciare le azioni del visitatore.

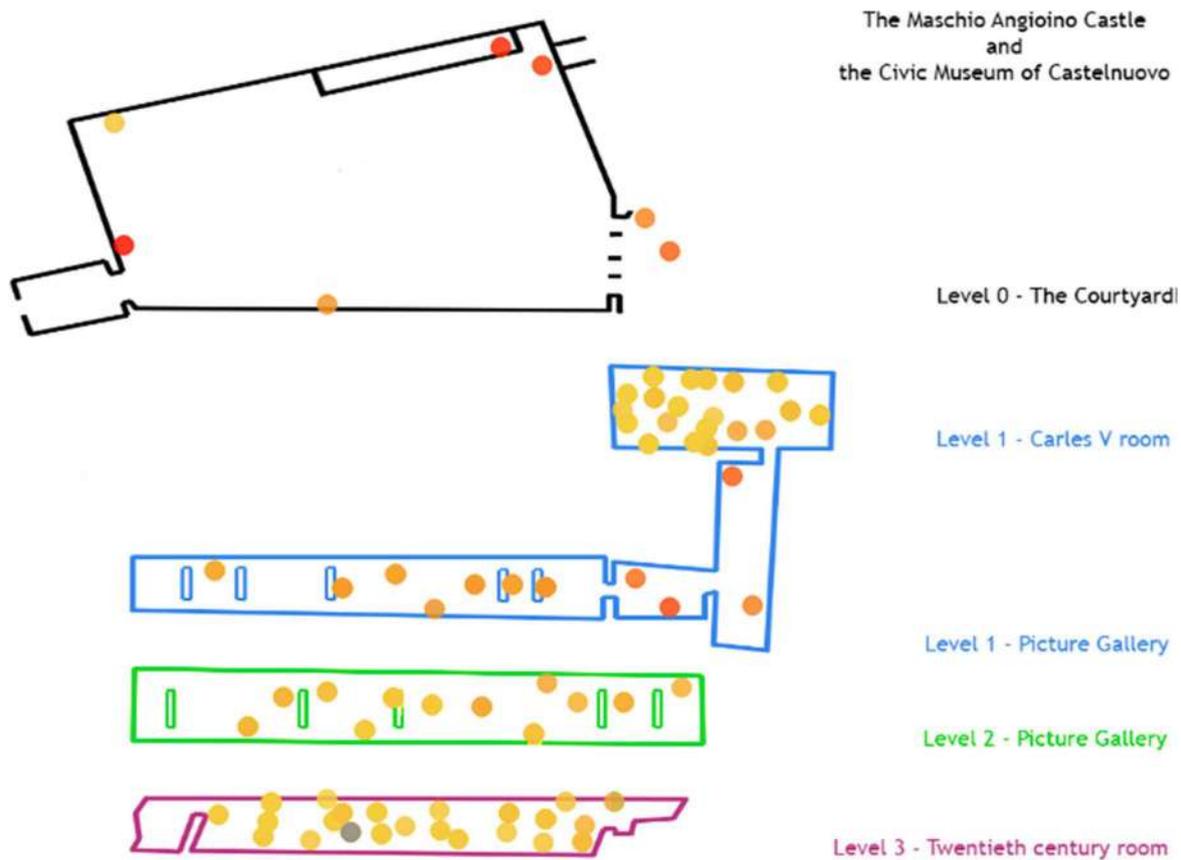


Figura 27 La figura mostra la pianta del Museo Civico di Castel Nuovo di Napoli. I punti rappresentano il posizionamento di beacon, mentre il colore dei punti (giallo, arancio e rosso) rappresenta il numero di interazioni dei visitatori con ciascun dispositivo.

Diversi aspetti e altre azioni vengono registrati come dati in un dataset. Nel campo del patrimonio culturale si può parlare di laboratori artistici o storici, ma più specificamente di laboratori del patrimonio come processo di innovazione nelle relazioni tra società/patrimonio, alla continua ricerca di un forte coinvolgimento sociale. Questa ricerca per la creazione di sistemi culturali, sociali e politici si basa sul favorire pratiche eque e sostenibili attraverso esperienze partecipative. I dati sono stati raccolti nell'arco di un periodo durato cinque mesi. Dopo un'attività preliminare di pre-elaborazione e pulizia dei dati, sono stati estratti più di 30 caratteristiche. Successivamente, sono state selezionate solo le caratteristiche multimediali relative alle opere d'arte, risultando in un totale di 16 attributi. L'elenco e la descrizione di questi sono presentati nella tabella seguente:

<b>Campo</b>	<b>Contenuto</b>
lang	Lingua preferita scelta sul tablet
month	Mese della visita
n_audio_full	Conteggio distinto delle opere d'arte di cui il visitatore ha ascoltato per intero i file audio
ratio_n_artwork	Conteggio distinto delle opere d'arte visitate, percentuale rispetto al massimo (169)
sec_visit	Tempo trascorso dalla selezione della prima opera d'arte alla fine della visita basata sul tablet, misurato in secondi
sec_hole	Tempo di pausa in secondi, somma di tutti i tempi tra l'ultima azione di log su un'opera d'arte e la selezione della prossima opera
n_text_invisible	Numero di azioni con text_button = invisible
n_text_visible	Numero di azioni con text_button = visible
n_audio_stop	Numero di azioni che interrompono l'ascolto della traccia audio (es. play_button_playing = false)
n_audio_active	Numero di azioni che mantengono attiva la traccia audio (es. play_button_playing = true)
n_audio_seektrack	Numero di azioni di ricerca sulla barra della traccia audio
ratio_audio_full	Rapporto tra il numero di audio ascoltati interamente e il numero di opere d'arte visitate
sec_action	Tempo di utilizzo del dispositivo (calcolato come sec_visits - sec_hole)
n_total_operations	Numero totale di azioni basate su tablet eseguite dal visitatore sulle opere d'arte
n_photo_action	Numero di azioni su foto/immagini (nextpic_photo, prevpic_photo)

ratio_artwork_F*	Percentuale di opere d'arte visitate per ogni piano (dove * può essere 0, 1, 2, 3)
------------------	--

I primi due attributi (lang e month) sono di tipo categorico. Le altre caratteristiche considerate sono legate a variabili dipendenti dal tempo e al numero di azioni sul dispositivo (che possono essere correlate a audio, foto o testo). Il passo di profilazione delle informazioni dalle caratteristiche selezionate è stato eseguito con una metodologia di apprendimento non supervisionato. Per questo studio è stato sfruttato un algoritmo di K-medoids<sup>5</sup>. L'algoritmo di k-medoids è una tecnica di clustering che è simile all'algoritmo k-means, ma differisce nella modalità di determinazione dei rappresentanti dei cluster. Mentre k-means usa le medie dei punti nei cluster (i centroidi), k-medoids utilizza punti reali come rappresentanti (i medoid). Un medoid è un punto all'interno di un cluster che ha la minore distanza media da tutti gli altri punti nel cluster. Questo rende i medoid più resistenti agli outlier rispetto ai centroidi usati in k-means. L'obiettivo è minimizzare la somma delle dissimilarità tra i punti e i loro medoid. Questo implica che la funzione obiettivo da minimizzare è spesso la somma delle distanze assolute (manhattan distance) o altre misure di dissimilarità. L'algoritmo ha quattro fasi principali:

- Inizializzazione: Selezionare casualmente dei punti dal dataset come medoid iniziali.
- Assegnazione: Assegnare ogni punto del dataset al medoid più vicino, formando k cluster.
- Aggiornamento: Per ogni cluster, calcolare il costo di scambiare il medoid corrente con uno qualsiasi dei punti assegnati al cluster. Se il costo del nuovo medoid è inferiore, effettuare lo scambio.
- Iterazione: Ripetere i passi di assegnazione e aggiornamento fino a quando non ci sono più cambiamenti significativi nei medoid o una condizione di arresto è soddisfatta (ad esempio, un numero massimo di iterazioni).

---

<sup>5</sup> F. Piccialli, G. Casolla, S. Cuomo, F. Giampaolo, E. Prezioso, V. Schiano di Cola, *Unsupervised Learning on multimedia data: a Cultural Heritage Case Study*

Il sistema k-medoids ha alcuni vantaggi, come la robustezza agli outlier; ha però anche lo svantaggio di essere più computazionalmente costoso rispetto al clustering k-means, specialmente per grandi dataset, poiché il calcolo del nuovo medoid richiede considerazioni su molti punti del cluster e quindi maggiore lentezza (Sureja, Chawda, & Vasant, 2022). Nella scienza dei dati, sono comunemente utilizzati diversi round del metodo k-means. L'algorithmo di clustering k-means divide i componenti in gruppi in base alla loro similarità..

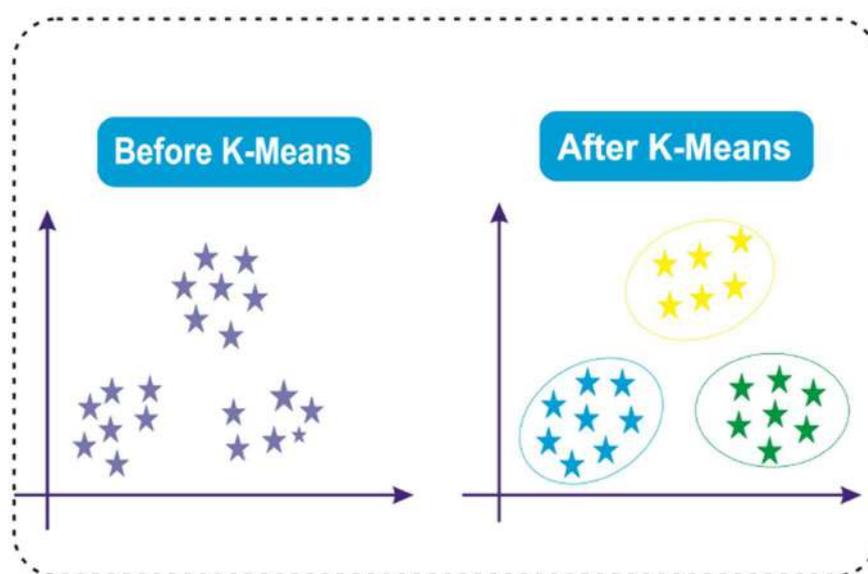


Figura 28 rappresentazione grafica del clustering k-means (Naeem, Ali, Anam, & Ahmed, 2023)

La lettera k indica il numero di gruppi. Di conseguenza, se k è 3, ci saranno tre raggruppamenti. Questo algorithmo di clustering divide il set di dati non etichettato in cluster unici con qualità comparabili per ciascun dato. Prima di procedere con l'attività di clustering, poiché alcune caratteristiche possono avere diverse unità di misura, è stata eseguita una fase di normalizzazione. L'algorithmo è stato implementato utilizzando il pacchetto "cluster" del software di analisi R. Per analizzare la semantica dei cluster ottenuti e proporre un significato utile per comprendere appieno il comportamento dei visitatori, sono state effettuate alcune considerazioni osservando alcuni risultati nei diagrammi riportati in seguito, che rappresentano come le caratteristiche multimediali possano variare la popolazione dei cluster. Le figure mostrano come i valori delle caratteristiche variano all'interno dei cluster attraverso i diagrammi a violino. La scelta di

questi grafici è motivata dalla geometria delle distribuzioni degli elementi all'interno di ciascun cluster. Per eseguire l'attività di clustering, è scelto un numero di cluster uguale a 5 in modo empirico, equivalenti ognuno ad un tipo di visitatore particolare. I gruppi ottenuti rappresentano diversi tipi di visitatori, che possono essere riassunti come segue:

- Visitatori estranei (cluster color salmone)
- Visitatori fortemente coinvolti (cluster color verde chiaro)
- Visitatori moderati (cluster color verde brillante)
- Visitatori di con un livello di interessamento base (cluster color azzurro)
- Visitatori appassionati di fotografie (cluster color magenta)

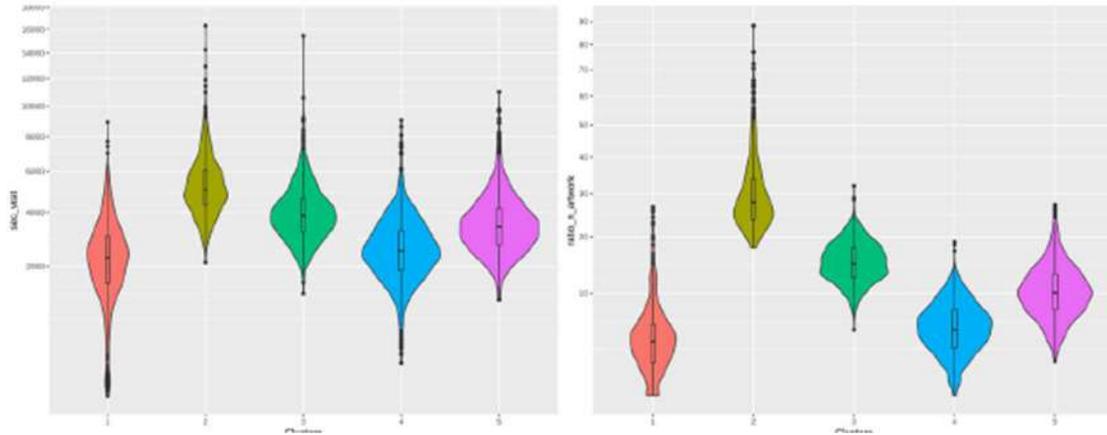


Figura 29 a sx, violin plot della distribuzione della visita nel museo, in secondi; a dx, percentuale delle opere osservate sul totale (Piccialli, et al., 2020)

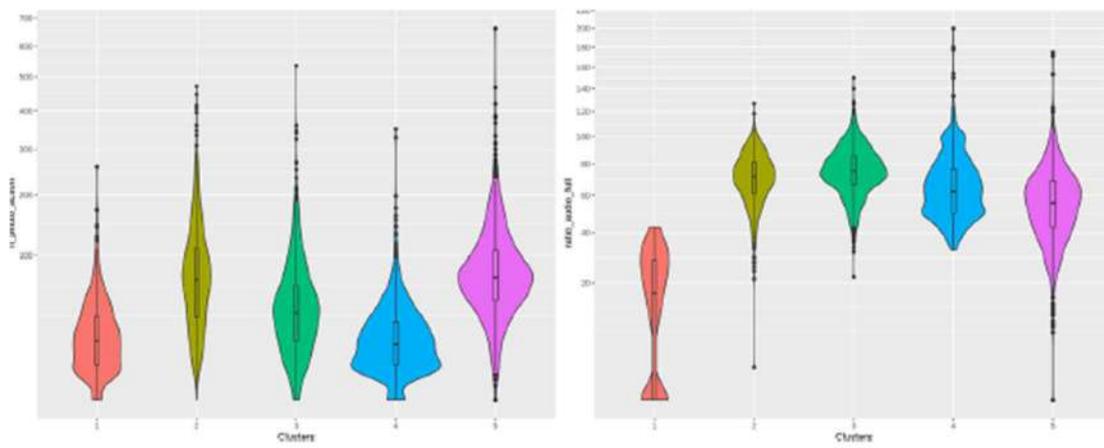


Figura 30 a sx, distribuzione delle attività di fotografia; a dx, audio ascoltati in full rispetto al totale (Piccialli, et al., 2020).

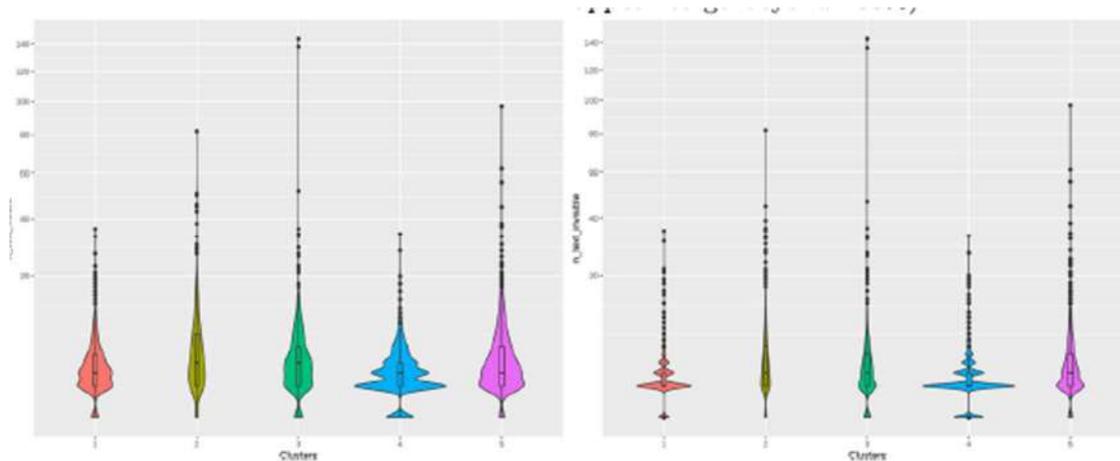


Figura 31 a sx, quantità di testi resi visibili sul tablet, a dx quantità di testi nascosti dal visitatore (Piccialli, et al., 2020)

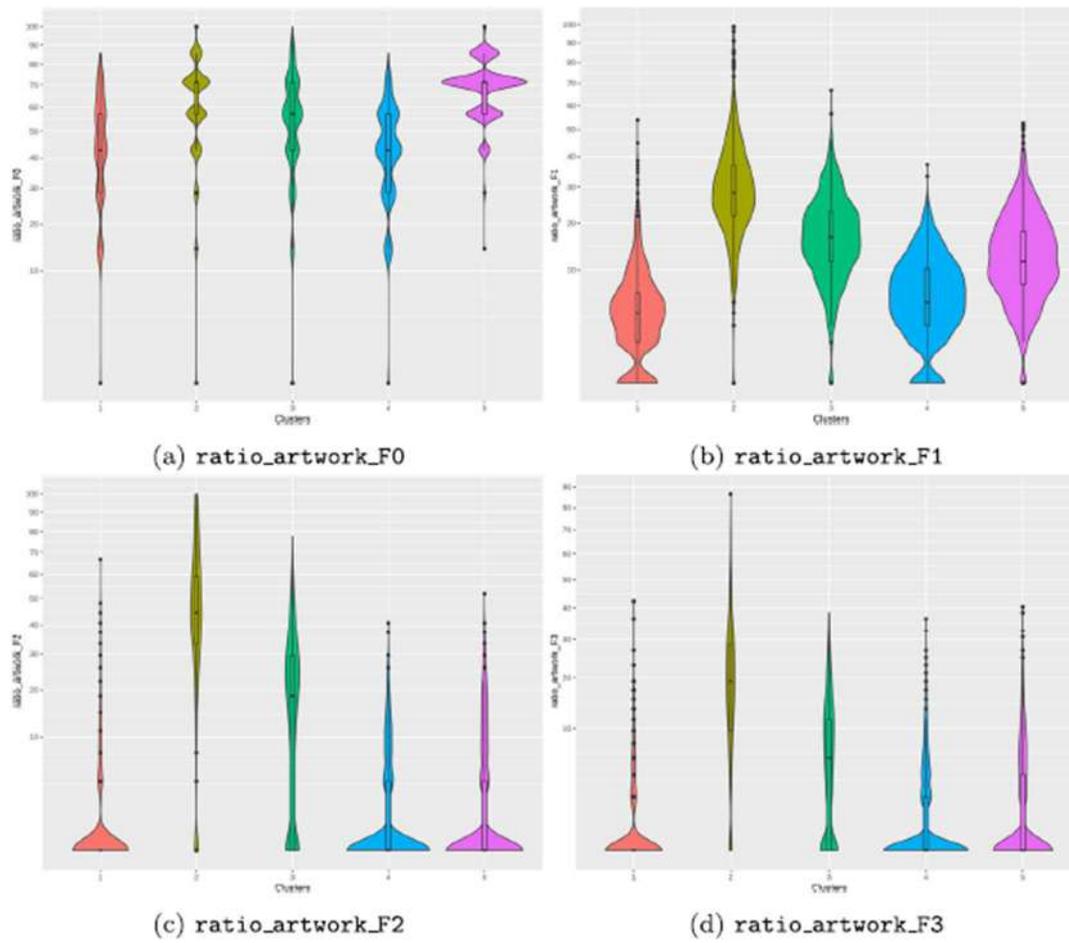


Figura 32 Rappresentazione delle opere osservate sul totale delle opere, suddiviso per piano (Piccialli, et al., 2020).

Nei diagrammi a violino rappresentati nelle pagine precedenti vediamo l'illustrazione delle variabili considerate in diversi contesti, come ad esempio la percentuale del numero di opere d'arte visitate rispetto al numero totale di opere d'arte per ogni piano (0, 1, 2 e 3 rispettivamente). I numeri sull'asse y sono scalati con una scala a radice quadrata. Per comprendere meglio i cluster ottenuti, possiamo analizzare più approfonditamente i grafici. La Figura 4 illustra la distribuzione dei cluster per il totale dei secondi della visita all'interno del museo. Questo mostra che c'è un ordine tra i cluster, anche se la distinzione non è così evidente. Quando viene considerato il numero di opere d'arte visitate, come mostrato nella Figura 5, l'ordine rimane invariato, ma ora è amplificato. Infatti, il cluster 2 ha i valori più alti, anche per il valore minimo considerando l'intervallo di confidenza. Considerando le azioni nelle foto, come mostrato nella Figura 6, la variazione è piuttosto impercettibile, rispettivamente all'ordine dei cluster. Riguardo ai file audio ascoltati integralmente, come mostrato nella Figura 5(dx), il cluster 1 ha i valori più bassi rispetto agli altri. È interessante notare che anche il cluster 5 mostra valori bassi. L'ultima distribuzione dei cluster considerata, mostrata nella Figura 6(sx), riguarda il numero di testi descrittivi per l'opera d'arte che sono stati resi visibili dal visitatore. Non vengono fatte distinzioni tra i cluster, e inoltre, tutti mostrano valori molto bassi. Consideriamo ora la distribuzione dei cluster in base al numero di opere d'arte visitate categorizzate per il numero di piano, come illustrato nella Figura 7: si può osservare che ogni piano fornisce una distribuzione completamente diversa dei cluster. Questo è correlato anche alla Figura 2, dove il numero di visitatori che interagiscono con il beacon è rappresentato con una mappa di calore. Complessivamente, più visitatori sono estremamente coinvolti con le opere d'arte al piano 0 e all'inizio del piano 1, mentre riducono il loro interesse salendo al piano superiore o accedendo alla sala di Carlo V. I cluster al piano 0 non possono essere distinti. Inoltre, tutti sono influenzati da una grande varianza in termini di distribuzione dei visitatori. Per gli altri piani, la narrazione è completamente diversa: ad esempio, i cluster 1 e 4 al piano 1, mostrati nella Figura 4b, presentano valori molto bassi, mentre il cluster 2 continua a presentare i valori più alti rispetto agli altri cluster. È interessante notare che per il secondo e il terzo piano, mostrati nella Figura 7, c'è una netta prevalenza

di persone che non guardano affatto alcuna opera d'arte, tranne che per il cluster 2 e 3. Nella tabella sottostante i cluster sono ordinati per dimensione di ciascun cluster.

Cluster	%	Visitor profile	Visit time	n. artwork	Ratio full audio	Photo click
4	26	Basic interested	Low	Low	Medium	Low
5	25	Photo-enthusiast	Medium	Medium	Medium	High
3	19	Moderate	Medium	High	High	Medium
1	17	Stranger	Low	Low	Very low	Low
2	13	Strongly engaged	High	Very high	High	High

Tabella 4 (Piccialli, et al., 2020)

Per rafforzare la qualità della descrizione del clustering, sono state generate matrici di correlazione. Queste mostrano una forte correlazione tra il numero di volte in cui un file audio viene interrotto prima che finisca e il numero di opere d'arte visitate in totale; il numero di azioni sulla foto non è correlato alla durata della visita al museo, il che suggerisce un interesse selettivo solo per foto specifiche; c'è poi una forte anticorrelazione tra il numero di opere d'arte visitate al primo piano e sia il numero di opere d'arte visitate al secondo piano che il numero di opere d'arte visitate al terzo piano. Inoltre, non c'è correlazione tra la durata della visita all'interno del museo e il numero di opere d'arte visitate, il che suggerisce la possibilità che i visitatori dall'orario. Per il cluster 4, è presente un'anticorrelazione tra il numero di opere d'arte visitate e il numero di file audio ascoltati integralmente. Per il cluster 5 ignorando la relazione di correlazione comune con gli altri cluster, non si riscontra alcuna correlazione. Un processo preliminare di estrazione delle caratteristiche multimediali è stato eseguito dai file di log non strutturati che rappresentano il comportamento dei visitatori durante l'uso dei dispositivi multimediali interattivi all'interno del museo. Considerando il lavoro dei direttori dei musei e gli stakeholder del patrimonio culturale, questo approccio può essere utilizzato per progettare e produrre servizi personalizzati. L'applicazione di metodologie di apprendimento non supervisionato allo scopo di eseguire un'analisi delle correlazioni può essere utile anche per comprendere e individuare pattern nei dati rispetto ai cluster ottenuti. Le tecniche di apprendimento possono essere utilizzate anche per concentrarsi sul colmare la carenza di dati, necessari per prendere decisioni informate (Piccialli, et al., 2020). Un esempio di come i dati raccolti dalle interazioni dei visitatori possano essere analizzati da algoritmi di ML per capire meglio le preferenze del pubblico è individuabile

in uno studio condotto nel 2023. Questo permette alle istituzioni culturali di personalizzare le loro offerte e migliorare la soddisfazione dei visitatori. Inoltre, l'analisi dei dati può aiutare a pianificare mostre e eventi futuri in base agli interessi e ai comportamenti dei visitatori. Un esempio proposto da un team di ricercatori cinesi vede un modello di estrazione tripla entità-relazione per identificare automaticamente entità e relazioni dai dati frammentari per la costruzione di un knowledge graph. Inoltre, un modello di completamento delle conoscenze è presentato per prevedere informazioni mancanti e migliorare la completezza del knowledge graph. Simulazioni comparative dimostrano l'efficacia e l'accuratezza dell'approccio proposto sia per il modello di estrazione delle conoscenze che per il modello di completamento delle conoscenze. L'efficacia dell'applicazione del knowledge graph è corroborata da un caso di studio che utilizza dati sulla collezione di manufatti in ceramica del Palace Museum in Cina. Questo metodo fornisce informazioni automatizzate, interconnesse, visivamente accattivanti e facilmente accessibili sul patrimonio culturale. Il modello identifica automaticamente entità e relazioni definite dagli esperti, consentendo un'integrazione efficiente dei dati. Inoltre, viene presentato un modello di completamento delle conoscenze per affrontare il problema delle informazioni mancanti nei dati del patrimonio culturale. Infine, viene costruito un knowledge graph del patrimonio culturale per soddisfare i requisiti di gestione del patrimonio culturale. Lo studio comprende cinque passaggi critici: definizione dell'ontologia, costruzione del dataset, estrazione del knowledge graph, completamento del knowledge graph e applicazione del knowledge graph. Anzitutto, l'ontologia è stata definita utilizzando un modello di riferimento ampiamente accettato nel dominio del patrimonio culturale. Sono state inizialmente raccolte tutte le informazioni rilevanti sulle ceramiche del patrimonio culturale, comprese descrizioni, caratteristiche fisiche, storie e contesti. Successivamente, queste informazioni sono state organizzate in una struttura coerente, categorizzando le ceramiche per tipo, epoca, provenienza e altre caratteristiche rilevanti. È stato utilizzato il CIDOC, Conceptual Reference Model (CRM), uno standard internazionale, come base per sviluppare l'ontologia, garantendo una struttura completa e standardizzata. Sono state stabilite poi le relazioni tra le diverse entità nel dominio della ceramica, come la relazione tra una ceramica specifica e il periodo storico a cui appartiene, o tra una ceramica e il sito

archeologico in cui è stata trovata. Infine, l'ontologia viene utilizzata per creare un knowledge graph, una rappresentazione visiva e interattiva delle informazioni e delle relazioni tra le entità, facilitando l'accesso, la gestione e l'interpretazione dei dati sul patrimonio culturale. È stato costruito il dataset raccogliendo i dati dal sito web del Palace Museum utilizzando Beautiful Soup. I dati sono stati annotati e trasformati in dataset per l'estrazione e il completamento del knowledge graph. Questo processo ha previsto l'estrazione automatizzata delle conoscenze sui manufatti in ceramica utilizzando modelli di deep learning. Il modello di estrazione è stato addestrato e valutato attraverso metodi di validazione incrociata. Tramite un modello migliorato di attenzione grafica sono state previste le relazioni mancanti nel knowledge graph. Il knowledge graph è stato infine implementato utilizzando un'architettura tecnica a quattro livelli: archiviazione dei dati, architettura del knowledge graph, servizi del knowledge graph e applicazione del knowledge graph. Con l'applicazione di knowledge graphs alla gestione del patrimonio culturale digitale i benefici risultanti sono stati numerosi, tra cui il miglioramento della ricerca, dell'interpretazione e della narrazione, la facilitazione della collaborazione e dell'integrazione dei dati, l'aumento dell'accessibilità e una migliore conservazione e protezione dei manufatti. Sono state individuate anche delle limitazioni: le principali sono la necessità di una grande quantità di dati etichettati, l'incompletezza del framework ontologico attuale rispetto al CIDOC CRM, e la necessità di ulteriori verifiche e test sui dati esistenti (Orea-Giner, Muñoz-Mazón, & Villacé-Molinero, 2022). Le tecnologie di Machine Learning possono anche migliorare l'esperienza dei visitatori nei musei e in altre istituzioni culturali, aumentando così le entrate. Attraverso l'uso di chatbot e assistenti virtuali, le istituzioni possono fornire assistenza in tempo reale ai visitatori, rispondendo a domande frequenti e offrendo informazioni personalizzate sui reperti. Questo tipo di interazione migliora la soddisfazione del visitatore e può incoraggiare visite ripetute (Ivanov, 2023). L'integrazione del machine learning nella gestione economica delle istituzioni culturali offre numerosi vantaggi, dall'ottimizzazione delle operazioni alla personalizzazione dell'esperienza del visitatore, fino al miglioramento delle strategie di marketing. Mentre le sfide tecniche e organizzative devono essere affrontate con attenzione, le potenzialità del ML nel settore culturale sono diverse. Le istituzioni culturali che adottano queste tecnologie possono migliorare significativamente la loro efficienza

operativa e la loro capacità di attrarre e soddisfare i visitatori, garantendo la sostenibilità a lungo termine del patrimonio culturale. Il patrimonio culturale italiano è uno dei più ricchi al mondo, e la sua gestione efficiente è cruciale per la preservazione e la valorizzazione delle risorse storiche e artistiche. Il caso studio di Villa Revedin Bolasco a Castelfranco Veneto (Treviso) rappresenta un esempio significativo di come le tecnologie di machine learning possano essere applicate per migliorare la gestione del patrimonio culturale. Questo progetto ha integrato diverse tecnologie digitali per creare un sistema avanzato di archiviazione e gestione dei dati culturali, affrontando le sfide legate alla conservazione e alla fruizione di un complesso storico di grande valore. Il progetto ha utilizzato un approccio interdisciplinare, combinando tecniche di geomatica, machine learning e gestione dei dati per sviluppare un geodatabase multisource. Questo geodatabase ha permesso di archiviare, gestire e visualizzare dati provenienti da diverse fonti, migliorando la qualità e l'accessibilità delle informazioni relative a Villa Revedin Bolasco. La prima fase del progetto ha previsto la raccolta di dati utilizzando tecniche di rilevamento geomatico come la fotogrammetria e la scansione laser 3D. Questi strumenti hanno permesso di creare modelli dettagliati del complesso architettonico e del parco circostante, catturando informazioni geometriche e radiometriche con alta precisione. I dati raccolti sono stati integrati in un geodatabase multisource, che combina informazioni geospaziali con dati storici e documentali. Questo database è stato progettato per essere accessibile attraverso una piattaforma web, consentendo agli utenti di visualizzare e interagire con i dati in modo intuitivo. Per migliorare l'analisi e la gestione dei dati, sono stati implementati algoritmi di machine learning per la classificazione e il riconoscimento delle caratteristiche architettoniche e strutturali. In particolare, sono stati utilizzati modelli di random forest per la classificazione dei dati di point cloud, identificando elementi come muri, finestre, tetti e colonne.

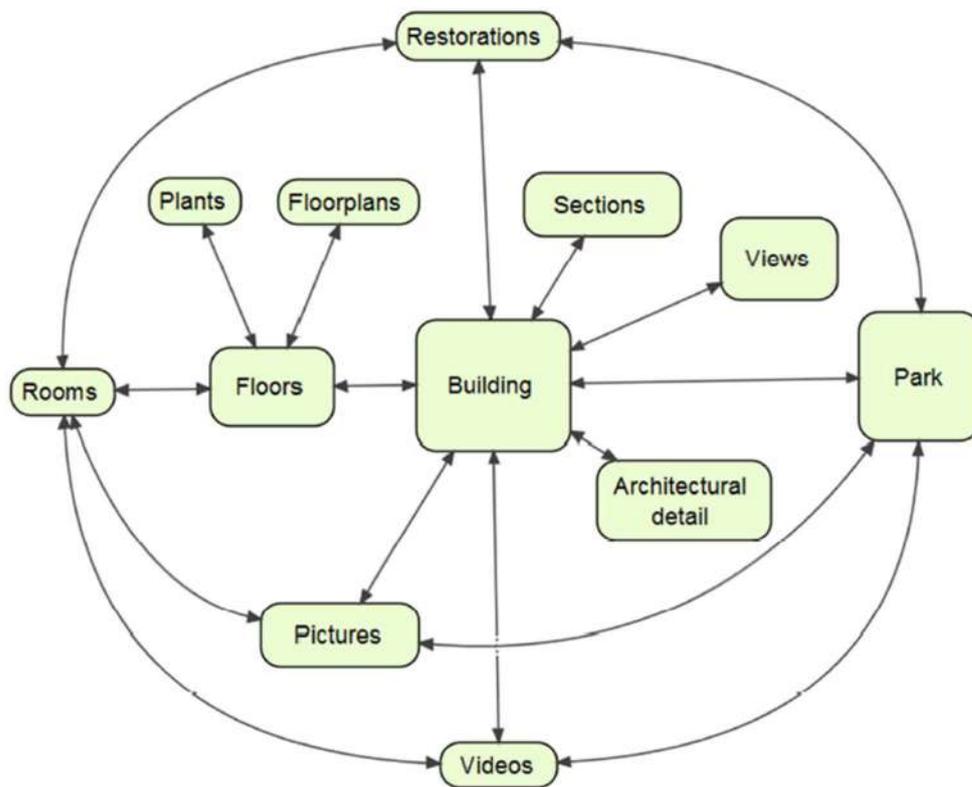


Figura 33 Modello concettuale del database (Guarnieri, Masiero, Piragnolo, & Pirotti, 2016)

I dati classificati sono stati visualizzati attraverso strumenti di realtà aumentata (AR) e realtà virtuale (VR), permettendo agli utenti di esplorare virtualmente Villa Revedin Bolasco e di comprendere meglio la sua struttura e storia. Questa visualizzazione avanzata ha facilitato la comprensione delle relazioni spaziali e delle condizioni di conservazione del sito. Il progetto ha dimostrato l'efficacia dell'utilizzo del machine learning nella gestione del patrimonio culturale, offrendo diversi benefici. L'analisi dei dati ha permesso di identificare aree critiche che necessitano di interventi di conservazione, migliorando la pianificazione e l'efficacia degli interventi di restauro. Ad esempio, l'uso dei modelli predittivi ha aiutato a individuare potenziali problemi strutturali e di degrado, permettendo interventi preventivi. La creazione di un geodatabase accessibile tramite piattaforma web ha migliorato l'accessibilità delle informazioni per ricercatori, storici dell'arte e il pubblico generale. La possibilità di esplorare virtualmente il sito ha aumentato l'interesse e l'engagement del pubblico,

promuovendo una maggiore consapevolezza del patrimonio culturale. L'automazione della classificazione e dell'analisi dei dati ha ridotto significativamente il tempo e le risorse necessarie per la gestione del sito. Questo ha permesso di allocare più risorse per attività di conservazione e promozione, migliorando l'efficienza operativa complessiva. L'adozione di tecnologie avanzate come il deep learning può offrire soluzioni per migliorare l'efficienza e l'efficacia delle operazioni culturali. Esplorare e implementare queste tecnologie, insieme a pratiche di governance responsabili e strategie di coinvolgimento comunitario, è fondamentale per garantire la sostenibilità e il successo a lungo termine delle istituzioni culturali (Guarnieri, Masiero, Piragnolo, & Pirotti, 2016).

## Conclusioni

L'applicazione dell'intelligenza artificiale nel settore culturale rappresenta una delle innovazioni più promettenti degli ultimi anni. Questa tesi ha esplorato come le tecnologie di machine learning possano essere integrate nelle scienze umane e culturali, evidenziandone applicazioni, sfide e potenziali miglioramenti. Uno dei principali vantaggi dell'intersezione tra intelligenza artificiale e settore culturale è la capacità di analizzare vasti insiemi di dati culturali, identificando pattern e tendenze difficili da riconoscere manualmente. L'analisi di testi storici, immagini di opere d'arte e manufatti archeologici può essere automatizzata e resa più precisa, consentendo ai professionisti del settore di risparmiare tempo e fatica sulla fase di preparazione di grandi database fisici. L'intelligenza artificiale offre strumenti avanzati per la ricerca accademica, e in particolare per l'analisi e l'interpretazione dei dati culturali. Gli algoritmi clustering consentono l'associazione di diverse opere d'arte, facilitando la catalogazione e la conservazione del patrimonio culturale e permettendo agli utenti di esplorare collezioni digitali in modo intuitivo. I robot sociali e i sistemi di raccomandazione basati su intelligenza artificiale possono arricchire l'esperienza dei visitatori nei musei, ampliando le possibilità di allargare il pubblico fruitore. I knowledge graph e le reti neurali convoluzionali, come visto, possono essere impiegate nel supporto alla ricerca dello storico dell'arte, facilitando lo studio delle influenze artistiche. Vi sono quindi molte possibilità di applicazioni vantaggiose. Tuttavia, esistono ancora aree di miglioramento e sfide future che devono essere affrontate. La qualità dei risultati ottenuti dai modelli di intelligenza artificiale dipende fortemente dalla qualità dei dati utilizzati per l'addestramento, pertanto è essenziale migliorare la raccolta, l'annotazione e la gestione dei dati culturali per garantire l'accuratezza e l'affidabilità degli algoritmi di machine learning. Negli ultimi anni la crescita di database contenenti dati culturali è stata notevole, ma è importante che le istituzioni continuino a prestare attenzione all'importanza di archiviare e digitalizzare reperti e fonti rendendoli accessibili al pubblico. Inoltre, gli algoritmi di intelligenza artificiale devono essere trasparenti e interpretabili per essere efficacemente utilizzati nel settore culturale, e devono tenere conto delle peculiarità dell'ambito. L'integrazione delle tecnologie di intelligenza artificiale nelle scienze

culturali richiede una stretta collaborazione tra esperti di diverse discipline, tra cui informatica, economia, storia dell'arte, archeologia e museologia: promuovere un dialogo interdisciplinare e sviluppare competenze trasversali è fondamentale per sfruttare appieno il potenziale dell'intelligenza artificiale nel settore culturale. Le istituzioni culturali devono essere supportate nello sviluppo di capacità e competenze per gestire l'implementazione etica e responsabile delle tecnologie di intelligenza artificiale, in accordo con le linee guida dell'AI Act (The AI Act Explorer, 2024). Attraverso un approccio integrato e responsabile, l'intelligenza artificiale può contribuire in modo significativo alla promozione e alla tutela delle eredità culturali. Gli algoritmi supervisionati, come quelli di classificazione e regressione, offrono strumenti potenti per mappare input a output specifici, consentendo previsioni accurate e identificazioni automatizzate basate su dati storici e culturali. Questi modelli sono particolarmente utili per creare dataset etichettati che possono prevedere eventi futuri o classificare oggetti culturali con precisione. Anche in questo caso, la sfida principale si trova nella qualità e la quantità dei dati etichettati, cruciali per il successo degli algoritmi supervisionati. L'impiego del machine learning non supervisionato nel campo dell'arte, come dimostrato dagli studi di Castellano e Vessio, evidenzia il potenziale di queste tecnologie per migliorare la comprensione delle opere d'arte e delle loro interconnessioni. È importante, come detto in precedenza, investire nella raccolta e pulizia dei dati, nonché sviluppare metodologie per gestire i dati mancanti. Nonostante i progressi significativi, l'applicazione del machine learning non supervisionato nel settore culturale presenta diverse sfide da affrontare. La qualità dei dati, la necessità di integrazione di conoscenze contestuali e la continua evoluzione degli algoritmi sono aspetti cruciali che richiedono attenzione. Attraverso l'automazione di processi ripetitivi e l'analisi predittiva dei dati, le tecnologie di apprendimento automatico permettono alle istituzioni di risparmiare tempo e risorse, riducendo i costi operativi e migliorando la pianificazione strategica. Queste capacità sono particolarmente rilevanti in un contesto culturale dove la gestione sostenibile e la valorizzazione del patrimonio sono essenziali per garantire la continuità e la fruizione del patrimonio culturale per le future generazioni. L'integrazione del machine learning nelle istituzioni culturali offre anche nuove opportunità per l'interazione e l'engagement del pubblico. L'uso di tecnologie avanzate come la realtà virtuale e aumentata, insieme ai chatbot e agli

assistenti virtuali, migliora l'esperienza dei visitatori, rendendo il patrimonio culturale più accessibile e coinvolgente. Inoltre, l'analisi dei dati raccolti dalle interazioni dei visitatori consente di personalizzare le offerte e di pianificare mostre ed eventi in modo più efficace, rispondendo meglio agli interessi e ai comportamenti del pubblico. L'analisi di progetti che utilizzano machine learning per la gestione del patrimonio culturale, come quello di Villa Revedin Bolasco, dimostrano l'efficacia di queste tecnologie nel migliorare la fruizione dei beni culturali; nel caso del progetto implementato nel museo civico di Castelnuovo invece emerge la potenzialità offerta da un'analisi fatta sul pubblico e sulle sue preferenze, consentendo di implementare decisioni commerciali e di marketing che permettono a lungo termine alle istituzioni di affrancarsi dalla necessità di appoggiarsi a fondi e supporti statali, permettendoci di immaginare un settore culturale più autonomo.

## Bibliografia

- Abad-Grau, M., Tajtakova, M., & Arias-Aranda, D. (2009). Machine learning methods for the market segmentation of the performing arts audiences. *International journal of business environment*, 2.
- Aiello Carlucci, L. (2015). *Intelligenza Artificiale*. Tratto da Treccani: [https://www.treccani.it/enciclopedia/intelligenza-artificiale\\_res-1eadb35b-dd79-11e6-add6-00271042e8d9\\_\(Enciclopedia-Italiana\)/](https://www.treccani.it/enciclopedia/intelligenza-artificiale_res-1eadb35b-dd79-11e6-add6-00271042e8d9_(Enciclopedia-Italiana)/)
- Alkemade, H., Claeysens, S., Colavizza, G., Freire, N., Lehmann, J., Neudecker, C., van Strien, D. (2023). Datasheets for Digital Cultural Heritage Datasets. *Journal of Open Humanities Data*, 1-11.
- Belhi, A., Al-Ali, A., & Foufou, S. (2023). A machine learning framework for enhancing digital experience in cultural heritage. *Journal of Enterprise Information Management*, 36.
- Berg, K. D., & Rachleff Burt, M. (2023). Socially Engaged Art's Histories and Tactics for the Neoliberal Era. Karen van den Berg and Melissa Rachleff Burt in conversation with Angela Dimitrakaki and Nomusa Makhubu. *Journal of Cultural Management and Cultural Policy/Zeitschrift für Kulturmanagement und Kulturpolitik*, 9.
- Blanke, T. (2018). Predicting the past. *Digital Humanities Quarterly*.
- Bollo, A. (2017). Musei, Società e Partecipazione. Le sfide dell'Audience Development. *Museologia Scientifica Memorie*, p. 61-63.
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., & Su, P. (2023). *A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT*. Tratto da Arxiv Cornell University: <https://arxiv.org/abs/2303.04226>
- Caramiaux, B. (2020, Maggio 15). *Research for CULT Committee - The Use of Artificial Intelligence in the Cultural and Creative Sectors*. Tratto da <https://www.europarl.europa.eu/>: [https://www.europarl.europa.eu/thinktank/en/document/IPOL\\_BRI\(2020\)629220](https://www.europarl.europa.eu/thinktank/en/document/IPOL_BRI(2020)629220)
- Castellano, G., & Vessio, G. (2022). Analisi e valorizzazione del patrimonio artistico mediante Intelligenza artificiale. *Culture digitali. Intersezioni: filosofia, arti, media. Proceedings* (p. 37). Lecce: AIUCD.
- Cilia, N., De Stefano, C., Fontanella, F., & Marr, C. (2020). An end-to-end deep learning system for medieval writer identification. *Pattern Recognition Letters*, 129.
- Colbert, F. (2020). Marketing the arts. In R. Towse, & T. N. Hernández, *Handbook of Cultural Economics, Third Edition*. Edward Elgar Publishing.
- Daille, B., Jacquey, E., Lejeune, G., Melo, L., & Touissaint, Y. (2016). Ambiguity Diagnosis for Terms in Digital Humanities. *Language Resources and Evaluation Conference*. Portoroz.

- Dalbello, M. (2011). A Genealogy of Digital Humanities. *Journal of Documentation*, 480-506.
- Dobson, J. (2021). Interpretable Outputs: Criteria for Machine Learning in the Humanities. *DHQ: Digital Humanities Quarterly*, p. 15(2).
- Dobson, J. E. (2015). Can an Algorithm be disturbed? Machine Learning, Intrinsic Criticism and Digital Humanities. *College Literature: a Journal of Critical Literary Studies*.
- Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review*, p. 40.
- Feldman, D., Schmidt, M., & Sohler, C. (2020). Turning Big data into tiny data: Constant-size coresets for k-means, PCA and projective clustering. *SIAM Journal on Computing*, 49.
- Feuerriegel, S., Hartmann, J., Janiesch, C., & Zsche, P. (2024). Generative ai. *Business & Information Systems Engineering*, p. 66.
- Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., & James, S. (2020). Machine Learning for Cultural Heritage: a Survey. *Pattern Recognition Letters*, p. 133.
- Fradkov, A. (2020). Early history of machine learning. *IFAC-PapersOnLine*, p. 1385-1390.
- Ghahramani, Z. (2003). Unsupervised Learning. *Summer school on machine learning*. Berlin: Springer Berlin Heidelberg.
- Guarnieri, A., Masiero, A., Piragnolo, M., & Pirotti, F. (2016). A geodatabase for multisource data applied to cultural heritage: the case study of Villa Revedin Bolasco. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, p. 41.
- Gujski, L. M., Di Filippo, A., & Limongiello, M. (2022). Machine Learning Clustering for Point Clouds Optimisation via Feature Analysis in Cultural Heritage. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- Helmi, L. (2021). Digital Humanities: a paradigm for the 21st century. *BAU Journal: Society, Culture and Human Behaviour*.
- Hesmondhalgh, D., & Meier, L. M. (2018). What the digitalisation of music tells us about capitalism, culture and the power of the information technology sector. *Information, Communication & Society*, p. 1555-1570.
- Hockey, S. (2004). The History of Humanities Computing. In S. S. Schreibman, *A Companion to Digital Humanities* (p. 1-19). Oxford: Oxford University Press.
- I. Wachtel, R. Z.-L. (2018). Predictive modeling for archaeological site locations: Comparing logistic regression and maximal entropy in north Israel and north-east China. *Journal of Archaeological Science*.

- Ivanov, R. (2023). Exhibitplorer: Enabling personalized content delivery in museums using contextual geofencing and artificial intelligence. *ISPRS International Journal of Geo-Information*, 12.
- J. Nyhan, A. F. (2016). Computation and the Humanities. Towards an Oral History of Digital Humanities. *Springer Series on Cultural Computing*, p. 285.
- Jakhar, D., & Kaur, I. (2020). Artificial intelligence, machine learning & deep learning: Definitions and differences. *Clinical and experimental dermatology*, p. 45.
- Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. Illinois: Urbana: University of Illinois Press.
- K. Kapitanova, S. S. (2013). Machine Learning Basics. In Q. H. F. Hui, *Intelligent Sensor Networks. The integration of Sensor Networks, Signal Processing and Machine Learning*. CRC Press.
- Kingsford, C. S. (2008). What are decision trees? *Nature Biotechnology*, p. 1011–1013.
- Kumar, V. (2020). Defining cultural data science. *Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world*. Dubai.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, p. 521.
- Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, p. 33.
- Llamas, J., Lerone, P., Medina, R. E., & Zalama, G. J. (2017). Classification of Architectural Heritage Images Using Deep Learning Techniques. *Applied Sciences*, 7.
- Manovich, L. (2016). The science of culture? Social Computing, Digital Humanities and Cultural analytics. *Journal of Cultural Analytics*.
- Milanato, D. (2008). *Demand Planning. Processi, metodologie e modelli matematici per la gestione della domanda commerciale*. Milano: Springer.
- Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*.
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*, 51-62.
- Noor, M. H., & Ige, A. O. (2024, Maggio 16). *A Survey on Deep Learning and State-of-the-art Applications*. Tratto da Arxiv Cornell University: <https://arxiv.org/abs/2403.17561>
- Offert, F., & Bell, P. (2020). Generative Digital Humanities. *CHR*, 202-212.

- Orea-Giner, A., Muñoz-Mazón, A., & Villacé-Molinero, T. (2022). Cultural tourist and user experience with artificial intelligence: a holistic perspective from the Industry 5.0 approach. *Journal of Tourism Futures*.
- O'shea, K., & Nash, R. (2015, Dicembre 2). *An introduction to convolutional neural networks*. Tratto da Arxiv Cornell University: <https://arxiv.org/pdf/1511.08458>.
- Piccialli, F., Casolla, G., Cuomo, S., Giampaolo, F., Prezioso, E., & Schiano di Cola, V. (2020). Unsupervised Learning on multimedia data: a Cultural Heritage Case Study. *Multimedia tools and applications*, 79.
- Portalés, C., Rodrigues, J. M., & Rodrigues Gonçalves, A. (2018). Digital cultural heritage. *Multimodal Technologies and Interaction*, p. 2.
- Ren, Y., Chu, T., Jiao, Y., Zhou, M., Geng, G., & Li, K. (2020). Multi-scale upsampling GAN based hole-filling framework for high-quality 3D cultural heritage artifacts. *Applied Sciences*, 12.
- Rockwell, G. P. (2019). The index thomisticus as a big data project. *Umanistica Digitale*.
- Saleh, B., Abe, K., Arora, R., & Elgammal, A. (2016). Toward automated discovery of artistic influence. *Multimedia Tools and Applications*.
- Shetty, S., Singh, C., & Rao, A. (2022). Supervised Machine Learning: Algorithms and Applications. In P. Singh, *Fundamentals and Methods of Machine and Deep Learning: Algorithms Tools and Applications*. Wiley Scrivener Publishing.
- Suissa, O., Elmalech, A., & Zhitomirsky-Geffet, M. (2022). Text analysis using deep neural networks in digital humanities and information science. *Journal of the Association for Information Science and Technology*, 73.
- Sula, C. A. (2019). The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004). *Digital Scholarship in the Humanities*.
- Sureja, N., Chawda, B., & Vasant, A. (2022). An improved K-medoids clustering approach based on the crow search algorithm. *Journal of Computational Mathematics and Data Science*, 3.
- The AI Act Explorer*. (2024, Aprile 19). Tratto da EU Artificial Intelligence Act: <https://artificialintelligenceact.eu/ai-act-explorer/>
- Tyagi, K., Rane, C., Sriram, R., & Manry, M. (2022). Unsupervised learning. In R. Pandey, S. K. Khatri, N. K. Singh, & P. Verma, *Artificial intelligence and machine learning for edge computing* (p. 33-52). Academic Press.
- Ville, B. D. (2013). Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, p. 448-455.

- W. P. Handwerker, D. F. (1997, 12). Sampling strategies for the comparisons with collection of cultural data: an answer to Galton's problem. *Current Anthropology*, p. 869.
- Wang, H. (2024). Automatic Classification of Museum Artifacts based on Unsupervised Models. *Proceedings of the 2024 7th International Conference on Software Engineering and Information Management*, (p. 116 - 119).
- Y. Yang, R. P. (s.d.). Automatic single page-based algorithms for medieval manuscript analysis. *Journal on Computing and Cultural Heritage (JOCCH)*, 1-22.